Noncommutative algebra, multiple harmonic sums and applications in discrete probability

Christian Costermans

Université Lille 2, 1 Place Déliot, 59024 Lille

Hoang Ngoc Minh

Université Lille 2, 1 Place Déliot, 59024 Lille

Abstract

After having recalled some important results about combinatorics on words, like the existence of a basis for the shuffle algebras, we apply them to some special functions, the *polylogarithms* $\mathrm{Li}_{\mathbf{s}}(z)$ and to special numbers, the *multiple harmonic sums* $\mathrm{H}_{\mathbf{s}}(N)$. In the "good" cases, both objects converge (respectively as $z \to 1$ and as $N \to +\infty$) to the same limit, the *polyzêta* $\zeta(\mathbf{s})$. For the divergent cases, using the technologies of noncommutative generating series, we establish, by techniques "à la Hopf", a theorem "à l'Abel", involving the generating series of polyzêtas. This theorem enables one to give an explicit form to generalized Euler constants associated with the divergent harmonic sums, and therefore, to get a very efficient algorithm to compute the asymptotic expansion of any $\mathrm{H}_{\mathbf{s}}(N)$ as $N \to +\infty$. Finally, we explore some applications of harmonic sums throughout the domain of discrete probabilities, for which our approach gives rise to exact computations, which can be then easily asymptotically evaluated.

Key words: Polylogarithm, polyzêta, harmonic sum, asymptotic analysis, discrete probability.

1. Introduction

For computation of discrete probabilities, we often need the asymptotic evaluation, in the scale of $\{N^\alpha \ln^\beta(N), \alpha, \beta \in \mathbb{Z}\}$, of functions of an integer N, as N becomes very large. For instance, harmonic numbers of order $r \geq 1, \mathrm{H}_r(N)$ (or *generalized harmonic numbers*) $\mathrm{H}_r(N) = \sum_{k=1}^{N} k^{-r}$ appear in the computation of complexities in the analysis of algorithms (Knuth, 1997; Flajolet and Sedgewick, 1996). Euler used his summation formula (also discovered afterwards, and independently, by Mac-Laurin) to obtain

Email addresses: ccostermans@univ-lille2.fr (Christian Costermans), hoang@univ-lille2.fr (Hoang Ngoc Minh).

Preprint submitted to Elsevier 30 May 2008

$$\sum_{n=1}^{N}\frac{1}{n}=\log N+\gamma-\sum_{j=1}^{k-1}\frac{B_j}{j}\frac{1}{N^j}+\mathrm{O}\left(\frac{1}{N^k}\right), \tag{1}$$

$$\sum_{n=1}^{N}\frac{1}{n^r}=\zeta(r)-\sum_{j=r-1}^{k-1}\frac{B_{j-r+1}}{j}\binom{j}{r-1}\frac{1}{N^j}+\mathrm{O}\left(\frac{1}{N^k}\right), \tag{2}$$

where B_i are Bernoulli numbers and γ (resp. $\zeta(r), r\geq 2$) is called Euler-Mac Laurin constant associated to the divergent harmonic number $\mathrm{H}_1(N)$ (resp. to the convergent harmonic number $\mathrm{H}_r(N), r\geq 2$) (Hardy, 2000).

Recently, the application of strategies of type *divide and conquer* to algorithms and hierarchical data structures on trees, led some authors to harmonic sums, associated to a composition $\mathbf{s}=(s_1,\ldots,s_r)$ (Flajolet and Vallée, 2000; Labelle and Laforest, 1995a)

$$\mathrm{H}_{\mathbf{s}}(N)=\sum_{N\geq n_1>\ldots>n_r>0}\frac{1}{n_1^{s_1}\ldots n_r^{s_r}} \tag{3}$$

which ordinary generating function, $\mathrm{P}_{\mathbf{s}}$, is a polylogarithmic function :

$$\mathrm{P}_{\mathbf{s}}(z)=\sum_{N\geq 1}\mathrm{H}_{\mathbf{s}}(N)z^N=\frac{\mathrm{Li}_{\mathbf{s}}(z)}{1-z}\quad\text{with}\quad \mathrm{Li}_{\mathbf{s}}(z)=\sum_{n_1>\ldots>n_r>0}\frac{z^{n_1}}{n_1^{s_1}\ldots n_r^{s_r}}. \tag{4}$$

We recall in Section 3 that the $\mathbb{C}$-algebra generated by polylogarithms $\mathrm{Li}_{s_1,\ldots,s_r}(z)$ and by logarithms $\log^n(z), n\geq 1$, is isomorphic to the $\mathbb{C}$- shuffle algebra, over the two-letters alphabet $X=\{x_0,x_1\}$ which is a free algebra and owns a basis, as recalled (Theorem 1) in Section 2, which sets up the background for the combinatorics on words. This point enables one to get algorithms for computing the monodromy (Hoang Ngoc Minh et al., 1998), the differential Galois group (Hoang Ngoc Minh, 2003a), functional equations of Kummer-type (Hoang Ngoc Minh et al., 1999) for these polylogarithms through their noncommutative generating series.

In Section 4, we now focus on the infinite alphabet $Y=\{y_i\}_{i\geq 1}$. To each composition $(s_1,\ldots,s_r)$, we can associate a word $w=y_{s_1}\ldots y_{s_r}$ over Y. This way, polylogarithms, harmonic sums and their ordinary generating function can be indexed by words : $\mathrm{H}_w=\mathrm{H}_{s_1,\ldots,s_r}$, $\mathrm{Li}_w=\mathrm{Li}_{s_1,\ldots,s_r}$, $\mathrm{P}_w=\mathrm{P}_{s_1,\ldots,s_r}$. It is also proved that the $\mathbb{C}$-Hadamard algebra of ordinary generating functions P_w of harmonic sums is isomorphic to the $\mathbb{C}$-shuffle algebra, over Y, leading so to the isomorphism between the $\mathbb{C}$-algebra of harmonic sums and the same $\mathbb{C}$-shuffle algebra, which is also free and owns a basis (Theorem 2).

Moreover, one important point linking $\mathrm{Li}_w(z)$ to $\mathrm{H}_w(N)$ is the fact that for $w\in Y^*\setminus y_1Y^*$, the limits of $\mathrm{Li}_w(z)$, when $z\to 1$, and of $\mathrm{H}_w(N)$, when $N\to\infty$, exist, and by Abel Theorem, are equal, the common limit being the polyzêta

$$\zeta(w)=\zeta(s_1,\ldots,s_r)=\sum_{n_1>\ldots>n_r>0}\frac{1}{n_1^{s_1}\ldots n_r^{s_r}}. \tag{5}$$

In order to study the divergent cases, *i.e.* for $w\in y_1Y^*$, we consider the noncommutative generating series of polylogarithms, and of multiple harmonic sums $\Lambda(z)=\sum_{w\in Y^*}\mathrm{Li}_w(z)\ w$ and $\mathrm{H}(N)=\sum_{w\in Y^*}\mathrm{H}_w(N)\ w$, and we prove by techniques "à la Hopf" the following theorem "à l'Abel" (Theorem 6)

$$\lim_{z\to 1}e^{-y_1\,\mathrm{Li}_{y_1}(z)}\Lambda(z)=\lim_{N\to\infty}\exp\left(\sum_{k\geq 1}\mathrm{H}_{y_k}(N)\frac{(-y_1)^k}{k}\right)\mathrm{H}(N)=S, \tag{6}$$

S standing for the noncommutative generating series of convergent polyzêtas, factorised as an infinite product indexed by Lyndon words $\mathcal{L}yn(Y) \setminus \{y_1\}$. This enables one in particular to explicit the generalized Euler constants associated to divergent polyzêtas $\{\zeta_{\text{ш+}}(w)\}_{w \in y_1 Y^*}$, and to get the asymptotic expansion of $\mathrm{H}_w(n)$. We show that these constants belong to the $\mathbb{Q}$-algebra generated by Euler constant γ and by convergent polyzêtas $\{\zeta(w)\}_{w \in Y^* \setminus y_1 Y^*}$. In fact, in order to get the asymptotic behaviour of $\mathrm{H}_w(N)$, the structural properties of the generating series of $\mathrm{Li}_w, w \in X^*$, and more particularly its behaviour as $z \to 1$ had already given rise to another algorithm (Costermans et al., 2005a). And the existence of a basis for the shuffle algebra over Y, joint to the isomorphism with the algebra of harmonic sums had also given rise to a third algorithm (Costermans et al., 2005b). We discuss each one of the three algorithms, compare them in terms of computing time, and conclude that the explicitation of generalized Euler constants improve significantly the two previous existing algorithms.

Section 5 is devoted to some applications, concerning various domains met throughout the area of discrete probabilities. In this section, we interpret some results found by different authors, Foata et al. (2001); Labelle and Laforest (1995a); Bai et al. (1998); Ivanin (1976), in terms of harmonic sums, which enables us to use all combinatoric tools previously presented, either to get some asymptotical evaluation, either to get an exact expression (for instance the leading constant κ_d - cf Theorem 10 - involved in the asymptotical expansion of the variance of the number of maxima in an hypercube). The first example deals with the “hyperharmonic numbers”, that we rewrite as a difference of harmonic sums. The second example is interested in the arity of random multidimensional quadtrees and we have a special look at some cases which make appear Euler transforms of harmonic sums. The third one exploits two formulas for the variance of the random number of maxima in a hypercube, and precises the algebraic nature (cf Theorem 11) of the coefficients occuring in its asymptotic expansion.

2. Combinatorics on words

2.1. *Hopf algebra*

Considering a finite alphabet $X = \{x_1, \ldots, x_k\}$ or an infinite alphabet $Y = \{y_i, i \geq 0\}$, we denote the empty word by ϵ. The length of a word $w = x_{i_1} \ldots x_{i_k}$, i.e. the integer k, is denoted by $|w|$. If each letter of the alphabet is associated with an integer constant called *weight*, we call *weight* of a word the sum of the weights of its letters. For instance, the word $w = y_1^2 y_3 y_2 y_1$ built over Y has for weight 8 and for length $|w| = 5$.

The set of words over X is denoted by X^*. A noncommutative formal power series over X, with coefficients in $\mathbb{C}$ is an application $S : w \in X^* \mapsto \langle S|w \rangle \in \mathbb{C}$. By abuse of notation, we will simply write $S = \sum_{w \in X^*} \langle S|w \rangle w$. The set of formal power series over X with coefficients in $\mathbb{C}$ is denoted by $\mathbb{C}\langle\langle X \rangle\rangle$.

Definition 1. Let $y_i, y_j \in Y$ and $u, v \in Y^*$. The *shuffle* (respectively *stuffle* and *minus-stuffle*) product of $u = y_i u'$ and $v = y_j v'$ is the polynomial recursively defined by

$$\begin{cases} \epsilon \,ш\, u = u \,ш\, \epsilon = u & \text{and} \quad u \,ш\, v = y_i(u' \,ш\, v) + y_j(u \,ш\, v') \\ \epsilon \,\text{ш+}\, u = u \,\text{ш+}\, \epsilon = u & \text{and} \quad u \,\text{ш+}\, v = y_i(u' \,\text{ш+}\, v) + y_j(u \,\text{ш+}\, v') + y_{i+j}(u' \,\text{ш+}\, v') \\ \epsilon \,\text{ш-}\, u = u \,\text{ш-}\, \epsilon = u & \text{and} \quad u \,\text{ш-}\, v = y_i(u' \,\text{ш-}\, v) + y_j(u \,\text{ш-}\, v') - y_{i+j}(u' \,\text{ш-}\, v') \end{cases}$$

Remark 1. Hoffman (2000) defines a family of "quasi-shuffle products" over $\mathbb{C}\langle Y\rangle$, Y being a locally finite set of generators, by

$$\epsilon \star u = u \star \epsilon = u \,\text{and}\, u \star v = y_i(u' \star v) + y_j(u \star v') + [y_i, y_j](u' \star v'),$$

where $[\cdot,\cdot]$ is supposed to verify $[y_i, 0] = 0$, to be commutative, associative and also such that $[y_i, y_j]$ either is identically 0 or has for weight $i + j$. The products ⧢ and $\uplus$ can so naturally be seen as "quasi-shuffle products". $\square$

The stuffle product $\uplus$ enables one to define a linear application defined for w_1 and $w_2 \in X^*$ by $st : w_1 \otimes w_2 \in \mathbb{C}\langle X\rangle \otimes \mathbb{C}\langle X\rangle \to w_1 \uplus w_2 \in \mathbb{C}\langle X\rangle$, extended to polynomials by linearity. Then, the linear application $\mathbb{1} : k \in \mathbb{C} \mapsto \mathbb{1}(k) = k \in \mathbb{C}\langle X\rangle$ appears as a *unity*. So $(\mathbb{C}\langle X\rangle, st, \mathbb{1})$ constitutes an associative and graduated $\mathbb{C}$-algebra.This algebra is known as *shuffle algebra*.

We define a coproduct by $\Delta : w \in \mathbb{C}\langle X\rangle \to \sum_{uv=w} u \otimes v \in \mathbb{C}\langle X\rangle \otimes \mathbb{C}\langle X\rangle$. Then $e : S \in \mathbb{C}\langle X\rangle \mapsto e(S) = \langle S|\epsilon\rangle \in \mathbb{C}$ appears as a counit for the coproduct Δ, and so $(\mathbb{C}\langle X\rangle, \Delta, e)$ becomes a (non cocommutative) coalgebra.

For a word $w = y_{s_1} \ldots y_{s_r}$, we can define the action of a composition $I = (i_1, \ldots, i_l)$ of the integer r (i.e. a finite sequence of positive integers summing to r) by

$$I[w] = y_{s_1+\ldots+s_{i_1}} y_{s_{(i_1+1)}+\ldots+s_{(i_1+i_2)}} \cdots y_{s_{(i_1+\ldots+i_{l-1}+1)}+\ldots+s_r}.$$

Example 1. Let $w = y_1^5 y_6 y_2$ and $I = (1, 2, 3, 1)$ a composition of 7 then $I[w] = y_1 y_2 y_8 y_2$.

Then, the bialgebra $(\mathbb{C}\langle X\rangle, st, \mathbb{1}, \Delta, e)$ becomes in fact an Hopf algebra, which antipode S is given by (Hoffman, 2000) $S(w) = (-1)^r \sum_{I\in\text{Comp}(r)} I[y_{s_r} \ldots y_{s_1}]$, $\text{Comp}(r)$ standing for the set of composition of r.

Remark 2. The Hopf algebraic structure remains almost the same when replacing the stuffle product by shuffle or minus-stuffle products, except for the antipode. In the first case, this one is given by $S(w) = (-1)^{|w|}\widetilde{w}$, with $\widetilde{y_{s_1} \ldots y_{s_r}} = y_{s_r} \ldots y_{s_1}$ (*mirror* function). In the case of the minus-stuffle product, the antipode is the same as the one given for the stuffle product, but with the action of the composition $I = (i_1, \ldots, i_l)$ corrected by a factor $(-1)^l$. More generally, for a quasi-shuffle $\star$, we shall define this action by

$$I[w] = [y_{s_1}, \ldots, y_{s_{i_1}}] \ldots [y_{s_{(i_1+\ldots+i_{l-1}+1)}}, \ldots, y_{s_r}]. \qquad \square$$

2.2. Lyndon words and Radford theorem

By definition, a *Lyndon word* is a nonempty word $l \in X^*$ strictly smaller than any of its proper right factors (Reutenauer, 1993) (for lexicographical order) i.e. for all $u, v \in X^* \setminus \{\epsilon\}$, $l = uv \Rightarrow l < v$. The set of Lyndon words over X is denoted by $\mathcal{L}ynX$.

Theorem 1. *(Radford, 1979) Let* $\mathrm{C}_1 = \mathbb{C} \oplus (x_0\mathbb{C}\langle X\rangle x_1)$ *be the set of polynomials, called "convergent", over* X. *Then,* $(\mathbb{C}\langle X\rangle, ⧢) \simeq \mathbb{C}[\mathcal{L}ynX] = \mathrm{C}_1[x_0, x_1]$.

Theorem 2. *(Malvenuto and Reutenauer, 1995) Let* $\mathrm{C}_2 = \mathbb{C} \oplus (Y \setminus y_1\mathbb{C}\langle Y\rangle) \simeq \mathrm{C}_1$ *be the set of polynomials, called "convergent" over* Y. *Then,* $(\mathbb{C}\langle Y\rangle, ⧢) \simeq (\mathbb{C}\langle Y\rangle, \uplus) \simeq \mathbb{C}[\mathcal{L}ynY] = \mathrm{C}_2[y_1]$.

Example 2.

$$x_1x_0x_1x_0x_1 + 2x_0x_1^2x_0x_1 + 2x_0x_1x_0x_1^2 = \frac{1}{2}x_0x_1 ⧢ x_0x_1 ⧢ x_1 - 2x_0^2x_1^2 ⧢ x_1 \in \mathbb{C}[\mathcal{L}ynX]$$
$$= x_0x_1x_0x_1 ⧢ x_1 \in \mathrm{C}_1[x_0, x_1]$$
$$y_2y_4y_1 + y_2y_1y_4 + y_1y_2y_4 + y_2y_5 + y_3y_4 = y_4 ⧢ y_2 ⧢ y_1 - y_4y_2 ⧢ y_1 - y_6 ⧢ y_1 \in \mathbb{C}[\mathcal{L}ynY]$$
$$= y_2y_4 ⧢ y_1 \in \mathrm{C}_2[y_1]$$

2.3. Bracket forms and dual basis

The *bracket form* $\check{S}_l$ of a Lyndon word $l = uv$, with $l, u, v \in \mathcal{L}ynX$ and the word v being as long as possible (factorisation - unique - called *standard* of a Lyndon word) is recursively defined by

$$\begin{cases} \check{S}_l &= [\check{S}_u, \check{S}_v] = \check{S}_u\check{S}_v - \check{S}_v\check{S}_u \\ \check{S}_x &= x \quad \text{for every letter } x \in X, \end{cases}$$

It is known that the set $\mathcal{B}_1 = \{\check{S}_l \ ; \ l \in \mathcal{L}ynX\}$ is a basis for the free Lie algebra. Moreover, each word $w \in X^*$ can be expressed, uniquely, as a decreasing (concatenation) product of Lyndon words:

$$w = l_1^{\alpha_1} l_2^{\alpha_2} \dots l_k^{\alpha_k}, \quad l_1 > l_2 > \dots > l_k, \quad k \geq 0. \tag{7}$$

The Poincaré–Birkhoff–Witt basis $\mathcal{B} = \{\check{S}_w;\ w \in X^*\}$ can be obtained from (7) putting (Reutenauer, 1993) $\check{S}_w = \check{S}_{l_1}^{\alpha_1}\check{S}_{l_2}^{\alpha_2} \dots \check{S}_{l_k}^{\alpha_k}$.

Its dual basis $\mathcal{B}^* = \{\mathcal{S}_w;\ w \in X^*\}$ can be then computed by

$$\begin{cases} \mathcal{S}_w &= \dfrac{\mathcal{S}_{l_1}^{⧢ \alpha_1} ⧢ \dots ⧢ \mathcal{S}_{l_k}^{⧢ \alpha_k}}{\alpha_1!\alpha_2!\dots\alpha_k!}, \\ \mathcal{S}_l &= x\mathcal{S}_w, \quad \forall l \in \mathcal{L}ynX, \text{ where } l = xw,\ x \in X,\ w \in X^*. \end{cases}$$

In (Reutenauer, 1993), it is proved that $\mathcal{B}$ and $\mathcal{B}^*$ are dual basis of $\mathbb{C}\langle X\rangle$ i.e. $(\check{S}_u|\mathcal{S}_v) = \delta_{u,v}$, for all words $u, v \in X^*$ with $\delta_{u,v} = 1$ if $u = v$, else 0.

3. Polylogarithms

3.1. Encoding by words

Let us consider the following two differential forms $\omega_0(z) = dz/z$ and $\omega_1(z) = dz/(1-z)$. The polylogarithm $\mathrm{Li}_{\mathbf{s}}(z)$ is defined for a composition $\mathbf{s} = (s_1, \dots, s_r)$ and for a complex z such that $|z| < 1$ by

$$\mathrm{Li}_{\mathbf{s}}(z) = \sum_{n_1 > \dots > n_r > 0} \frac{z^{n_1}}{n_1^{s_1} \dots n_r^{s_r}}. \tag{8}$$

This expression corresponds to the iterated integral over ω_0, ω_1 and along the path $0 \rightsquigarrow z$,

$$\mathrm{Li}_{\mathbf{s}} = \int_{0\rightsquigarrow z} \omega_0^{s_1-1}\omega_1 \dots \omega_0^{s_r-1}w_1. \tag{9}$$

Let $X = \{x_0, x_1\}$. We shall by now identify any composition $\mathbf{s} = (s_1, \ldots, s_r)$ with its encoding word $w = x_0^{s_1-1}x_1 \cdots x_0^{s_r-1}x_1$ over X^*x_1, identification suggested by the previous formula. We obtain so a concatenation isomorphism from the $\mathbb{C}$-algebra of compositions into the subalgebra $\mathbb{C}\langle X\rangle x_1 \subset \mathbb{C}\langle X\rangle$. In that way, the polylogarithm $\mathrm{Li}_{\mathbf{s}}$ defined by the formula (8) can be also indexed by $w \in X^*x_1$. To extend the definition of polylogarithms over X^*, we put $\mathrm{Li}_{x_0}(z) = \log(z)$. By linearity, the definition of Li_w is extended to polynomials on $\mathbb{C}\langle X\rangle$. One of the most important facts concerning these polylogarithms is the following result, i.e the isomorphism with a shuffle algebra, for the usual product $⧢$.

Theorem 3. *(Hoang Ngoc Minh et al., 1998) The map* $\mathrm{L} : w \mapsto \mathrm{Li}_w$ *is an isomorphism from* $(\mathbb{C}\langle X\rangle, ⧢)$ *to* $(\mathbb{C}\{\mathrm{Li}_w\}_{w\in X^*}, .)$.

3.2. Noncommutative generating series

The noncommutative generating series of polylogarithms, $\mathrm{L} = \sum_{w\in X^*} \mathrm{Li}_w \; w$, satisfies Drinfel'd differential equation (Drinfel'd, 1990)$d\mathrm{L} = (x_0\omega_0 + x_1\omega_1)\mathrm{L}$, with the initial condition $\mathrm{L}(\varepsilon) = e^{x_0 \log \varepsilon} + \mathrm{O}(\sqrt{\varepsilon})$, for $\varepsilon \to 0^+$. This enables one to prove that L is the exponential of a Lie series. From the factorization of a monoid by Lyndon words $l \in \mathcal{L}ynX$, we get the factorization of the series L (Hoang Ngoc Minh et al., 1998) :

$$\mathrm{L}(z) = e^{x_1 \log \frac{1}{1-z}} \left[\prod_{l\in\mathcal{L}ynX\setminus\{x_0,x_1\}}^{\searrow} e^{\mathrm{Li}_{S_l}(z)\check{S}_l} \right] e^{x_0 \log z}. \tag{10}$$

For all $l \in \mathcal{L}ynX \setminus \{x_0, x_1\}$, we have $S_l \in x_0X^*x_1$. So, let us put

$$\mathrm{L}_{\mathrm{reg}} = \prod_{l\in\mathcal{L}ynX\setminus\{x_0,x_1\}}^{\searrow} e^{\mathrm{Li}_{S_l}\check{S}_l} \quad \text{and} \quad Z = \mathrm{L}_{\mathrm{reg}}(1). \tag{11}$$

Let σ be the monoid endomorphism verifying $\sigma(x_0) = -x_1, \sigma(x_1) = -x_0$, we also get (Hoang Ngoc Minh et al., 1999)

$$\mathrm{L}(z) = \sigma[\mathrm{L}(1-z)]Z = e^{x_0 \log z}\sigma[\mathrm{L}_{\mathrm{reg}}(1-z)]e^{-x_1 \log(1-z)}Z. \tag{12}$$

Definition 2. (Hoang Ngoc Minh et al., 2001) Let $\zeta_{⧢} : (\mathbb{C}\langle X\rangle, ⧢) \to (\mathbb{C}, .)$ be the algebra morphism (*i.e.* for $u, v \in X^*, \zeta_{⧢}(u ⧢ v) = \zeta_{⧢}(u)\zeta_{⧢}(v)$) verifying for all convergent word $w \in x_0X^*x_1, \zeta_{⧢}(w) = \zeta(w)$, and such that $\zeta_{⧢}(x_0) = \zeta_{⧢}(x_1) = 0$.

Then, the noncommutative generating series $Z_{⧢} = \sum_{w\in X^*} \zeta_{⧢}(w)\, w$ verifies $Z_{⧢} = Z$ (Hoang Ngoc Minh et al., 2001). In consequence, $Z_{⧢}$ is the unique Lie exponential verifying $\langle Z_{⧢}|x_0\rangle = \langle Z_{⧢}|x_1\rangle = 0$ and $\langle Z_{⧢}|w\rangle = \zeta(w)$, for any $w \in x_0X^*x_1$.

Remark 3. Equation (12) enables one to reach the expansion of any $\mathrm{Li}_w(z)$ around $z = 1$ (a *singular expansion* if $w \in x_1X^*x_1$). For example,

$$\begin{aligned}\mathrm{Li}_{2,1}(1-z) &= -\,\mathrm{Li}_3(z) + \log(z)\,\mathrm{Li}_2(z) - \log^2(z)\,\mathrm{Li}_1(z)/2 - \zeta(2)\,\mathrm{Li}_1(z) + \zeta(3)\\ &= \big(-1-\zeta(2)\big)z + z\log(z) - z\log^2(z)/2 + \zeta(3) + \mathrm{O}\,(|z|)\,. \qquad \square\end{aligned}$$

4. Special values

4.1. Symmetric functions

Let $\{t_i\}_{i\in\mathbb{N}_+}$ be an infinite set of variables, and let us define the (modified) symmetric functions $\lambda_k^{(r)}$ and the sums of powers $\psi_k^{(r)}$ by

$$\lambda_k^{(r)}(\underline{t}) = \sum_{n_1>\ldots>n_k>0} t_{n_1}^r \ldots t_{n_k}^r \quad \text{and} \quad \psi_k^{(r)}(\underline{t}) = \sum_{n>0} t_n^{rk}. \tag{13}$$

Remark 4. In the case $r = 1$, we find the classical elementary symmetric functions $\{\lambda_k^{(1)}\}_{k\geq 1}$ (MacDonald, 1995). □

They are respectively coefficients of the following generating functions

$$\lambda^{(r)}(\underline{t}|z) = \sum_{k>0} \lambda_k^{(r)}(\underline{t})z^k = \prod_{i\geq 1}(1 + t_i^r z) \quad \text{and} \tag{14}$$

$$\psi^{(r)}(\underline{t}|z) = \sum_{k>0} \psi_k^{(r)}(\underline{t})z^{k-1} = \sum_{i\geq 1} \frac{t_i^r}{1 - t_i^r z}. \tag{15}$$

These generating functions satisfy a Newton identity $d/dz \log \lambda^{(r)}(\underline{t}|z) = \psi^{(r)}(\underline{t}| - z)$.

We also recall the Waring formula (putting $\lambda_0^{(r)} \equiv 1$) :

$$\lambda_k^{(r)} = \frac{(-1)^k}{k!} \sum_{\substack{s_1,\ldots,s_k>0 \\ s_1+\ldots+ks_k=k}} \binom{k}{s_1,\ldots,s_k} \left(-\frac{\psi_1^{(r)}}{1}\right)^{s_1} \ldots \left(-\frac{\psi_k^{(r)}}{k}\right)^{s_k} \tag{16}$$

4.2. Quasi-symmetric functions

To the composition $\mathbf{s} = (s_1, \ldots, s_r)$, we now also associate the word $w = y_{s_1} \ldots y_{s_r}$ defined over the alphabet $Y = \{y_i, i \geq 0\}$.

The number of occurrences of letter y_i in the word $w \in Y^*$ is denoted by $|w|_i$.

Let $w = y_{s_1} \ldots y_{s_r} \in Y^*$. The quasi-symmetric functions F_w and G_w, of depth $r = |w|$ and of degree (or weight) $s_1 + \ldots + s_r$, are defined by

$$\mathrm{F}_w(\underline{t}) = \sum_{n_1>\ldots>n_r>0} t_{n_1}^{s_1} \ldots t_{n_r}^{s_r} \quad \text{and} \quad \mathrm{G}_w(\underline{t}) = \sum_{n_1\geq\ldots\geq n_r>0} t_{n_1}^{s_1} \ldots t_{n_r}^{s_r}. \tag{17}$$

Remark 5. There exist explicit relations between the functions G_w and F_w. For simplicity in the expression of the relations, we express them with multi-indexed notations rather than the word-indexed notation we just defined. Precisely, if $I = (i_1, \ldots, i_r)$ (resp. $J = (j_1, \ldots, j_p)$) is a composition of n (resp. of r) then $J \circ I = (i_1 + \ldots + i_{j_1}, i_{j_1+1} + \ldots + i_{j_1+j_2}, \ldots, i_{k-j_p+1} + \ldots + i_k)$ is a composition of n. If $\mathbf{s} = (s_1, \ldots, s_r)$ then we have (Hoffman, 2004) $\mathrm{G}_\mathbf{s} = \sum_{J\in\mathrm{Comp}(r)} \mathrm{F}_{J\circ\mathbf{s}}$ and $\mathrm{F}_\mathbf{s} = \sum_{J\in\mathrm{Comp}(r)} (-1)^{l(J)-r}\mathrm{G}_{J\circ\mathbf{s}}$, the second formula, in which $l(J)$ is the number of parts of J, being derived from the first one by Möbius inversion. For example, $\mathrm{G}_{(1,2,1)} = \mathrm{F}_{(1,2,1)} + \mathrm{F}_{(3,1)} + \mathrm{F}_{(1,3)} + \mathrm{F}_{(4)}$, and $\mathrm{F}_{(1,2,1)} = \mathrm{G}_{(1,2,1)} - \mathrm{G}_{(3,1)} - \mathrm{G}_{(1,3)} + \mathrm{G}_{(4)}$. More generally, for a composition of length r, the conversion from a form to the other one make appear $Card(\mathrm{Comp}(r)) = 2^{r-1}$ terms. □

In particular, $\mathrm{F}_{y_r^k} = \lambda_k^{(r)}$ and $\mathrm{F}_{y_{rk}} = \mathrm{G}_{y_{rk}} = \psi_k^{(r)}$. As a consequence, integrating differential equation given by Newton identity shows that functions $\mathrm{F}_{y_r^k}$ and $\mathrm{F}_{y_{rk}}$ are linked by the formula

$$\sum_{k\geq 0}\mathrm{F}_{y_r^k} z^k = \exp\Big[-\sum_{k\geq 1}\mathrm{F}_{y_{rk}}\frac{(-z)^k}{k}\Big]\ \Big(\text{or}\sum_{k\geq 0}\mathrm{G}_{y_r^k} z^k = \exp\Big[\sum_{k\geq 1}\mathrm{G}_{y_{rk}}\frac{z^k}{k}\Big]\Big). \quad (18)$$

By linearity, the definitions of F_w and G_w are extended to polynomials on $\mathbb{C}\langle Y\rangle$.

If u (resp. v) is a word in Y^*, of length r and of weight p (resp. of length s and of weight q), $\mathrm{F}_{u \uplus v}$ and $\mathrm{G}_{u \sqcup v}$ are quasi-symmetric functions of depth $r+s$ and of weight $p+q$, and one has

$$\mathrm{F}_{u \uplus v} = \mathrm{F}_u\,\mathrm{F}_v \quad \text{and} \quad \mathrm{G}_{u \sqcup v} = \mathrm{G}_u\,\mathrm{G}_v. \quad (19)$$

Since functions $\mathrm{F}_w, w \in Y^*$ are linearly independent (Gessel, 1984), the remarkable identity (16) can be then seen as

$$y_r^k = \frac{(-1)^k}{k!}\sum_{\substack{s_1,\ldots,s_k>0\\ s_1+\ldots+ks_k=k}}\binom{k}{s_1,\ldots,s_k}\frac{(-y_r)^{\uplus s_1}}{1^{s_1}}\uplus\ldots\uplus\frac{(-y_{kr})^{\uplus s_k}}{k^{s_k}}. \quad (20)$$

4.3. Multiple harmonic sums

Definition 3. For any $w = y_{s_1}\ldots y_{s_r} \in Y^*$, let us define the maps H_w and $\underline{\mathrm{H}}_w$ from $\mathbb{N}_+$ to $\mathbb{Q}$ by

$$\mathrm{H}_w(N) = \sum_{N\geq n_1>\ldots>n_r>0}\frac{1}{n_1^{s_1}\ldots n_r^{s_r}},\quad \underline{\mathrm{H}}_w(N) = \sum_{N\geq n_1\geq\ldots\geq n_r>0}\frac{1}{n_1^{s_1}\ldots n_r^{s_r}}.$$

We use the conventions $\mathrm{H}_w(0) = \underline{\mathrm{H}}_w(0) = 0$ for any nonempty word w and $\mathrm{H}_\epsilon = \underline{\mathrm{H}}_\epsilon \equiv 1$.

By linearity, the definitions of H_w and $\underline{\mathrm{H}}_w$ are extended to polynomials on $\mathbb{C}\langle Y\rangle$.

For $N \geq 1$ and $w \in Y^*$, any $\mathrm{H}_w(N)$ (resp. $\underline{\mathrm{H}}_w(N)$) can be obtained by specializing variables $\{t_i\}_{N\geq i\geq 1}$ at $t_i = 1/i$ and, for $i > N, t_i = 0$ in the quasi-symmetric function F_w (resp. G_w) (Hoffman, 1997). In particular, the relations given in Remark 5 turn into

$$\underline{\mathrm{H}}_{\mathbf{s}} = \sum_{J\in\mathrm{Comp}(r)}\mathrm{H}_{J\circ\mathbf{s}} \quad \text{and} \quad \mathrm{H}_{\mathbf{s}} = \sum_{J\in\mathrm{Comp}(r)}(-1)^{l(J)-r}\underline{\mathrm{H}}_{J\circ\mathbf{s}}, \quad (21)$$

Morevover, from relations (19), we get

Proposition 1. (Hoffman, 1997) For $u, v \in Y^*, \mathrm{H}_{u \uplus v} = \mathrm{H}_u\,\mathrm{H}_v$ and $\underline{\mathrm{H}}_{u \sqcup v} = \underline{\mathrm{H}}_u\,\underline{\mathrm{H}}_v$.

Let $w = y_s w' \in Y^*$ such that $|w| = r$. One has

$$\mathrm{H}_w(N) = \sum_{l=r}^{N}\frac{\mathrm{H}_{w'}(l-1)}{l^s} \quad \text{and} \quad \underline{\mathrm{H}}_w(N) = \sum_{l=1}^{N}\frac{\underline{\mathrm{H}}_{w'}(l)}{l^s}. \quad (22)$$

In consequence,

Theorem 4. *For any* $w = y_s w' \in Y^*$, $\mathrm{H}_w(N)$ *and* $\underline{\mathrm{H}}_w(N)$ *converge when* $N \to +\infty$ *if and only if* $s > 1$. *Therefore, if* $s \geq 2$ *then the limits* $\lim_{N\to+\infty} \mathrm{H}_w(N)$ *and* $\lim_{N\to+\infty} \underline{\mathrm{H}}_w(N)$ *are denoted respectively by* $\zeta(w)$ *and by* $\underline{\zeta}(w)$. *This justifies the fact that* w *was said to be* convergent *in this case (otherwise, it is said to be* divergent*).*

4.4. Noncommutative generating series over Y

Proposition 2. (Hoang Ngoc Minh, 2003b) For $w \in X^*$, let $\mathrm{P}_w(z) = (1-z)^{-1}\,\mathrm{Li}_w(z)$. Thus for $u, v \in Y^*, \mathrm{P}_{u \,⧢\, v} = \mathrm{P}_u \odot \mathrm{P}_v$, where $\odot$ denotes the Hadamard product.

Proof. This comes from the fact that, for $u \in Y^*, \mathrm{P}_u(z) = \sum_{N\geq 0} \mathrm{H}_u(N) z^N$, so P_u appears as the generating series of $\{\mathrm{H}_u(N), N \geq 0\}$. The proposition follows then directly from Proposition 1, and from the definition of the Hadamard product. $\square$

As consequences of Theorem 3, we also have

Theorem 5. *(Hoang Ngoc Minh, 2003b) The map* $\mathrm{P} : u \mapsto \mathrm{P}_u$ *is an isomorphism from polynomial algebra* $(\mathbb{C}\langle Y\rangle, ⧢)$ *to the Hadamard algebra* $(\mathbb{C}\{\mathrm{P}_w\}_{w\in Y^*}, \odot)$. *Therefore, the map* $\mathrm{H} : u \mapsto \mathrm{H}_u$ *(resp.* $\underline{\mathrm{H}} : u \mapsto \underline{\mathrm{H}}_u$*) is an isomorphism from* $(\mathbb{C}\langle Y\rangle, ⧢)$ *(resp.* $(\mathbb{C}\langle Y\rangle, \sqcup\!\sqcup)$*) to the algebra* $(\mathbb{C}\{\mathrm{H}_w\}_{w\in Y^*}, .)$ *(resp.* $(\mathbb{C}\{\underline{\mathrm{H}}_w\}_{w\in Y^*}, .)$*).*

The noncommutative generating series $\mathrm{P}(z) = \sum_{w\in X^*} \mathrm{P}_w(z)\, w = (1-z)^{-1}\mathrm{L}(z)$, can be factorised, from the factorisation (10) of the series L, in

$$\mathrm{P}(z) = e^{-(x_1+1)\log(1-z)}\mathrm{L}_{\mathrm{reg}}(z) e^{x_0 \log z}. \tag{23}$$

Let π_Y the projector from $\mathbb{C}\langle\langle X\rangle\rangle$ to $\mathbb{C}\langle\langle Y\rangle\rangle$ erasing the monomials ending with x_0, i.e. for any word $w \in X^*, \pi_Y(wx_0) = 0$ and $\pi_Y(x_0^{s_1-1}x_1 \ldots x_0^{s_r-1}x_1) = y_{s_1} \ldots y_{s_r}$. Then

$$\Lambda(z) = \pi_Y \mathrm{L}(z) \underset{z\to 1}{\widetilde{\quad}} \exp\left(y_1 \log \frac{1}{1-z}\right)\pi_Y Z. \tag{24}$$

Since $\mathrm{P}_{y_1^k}(z) = (1-z)^{-1}\mathrm{Li}_{y_1^k}(z) = (-1)^k(1-z)^{-1}\log^k(1-z)/k!$, we get another expression for the factorisation of P, when projected onto $\mathbb{C}\langle\langle Y\rangle\rangle$:

Lemma 1. Let $\mathrm{Mono}(z) = e^{-(x_1+1)\log(1-z)}$. Then

$$\pi_Y \mathrm{Mono} = \sum_{k\geq 0} \mathrm{P}_{y_1^k}\, y_1^k \quad \text{and} \quad \pi_Y \mathrm{Mono}^{-1} = \sum_{k\geq 0} \mathrm{P}_{y_1^k}\, (-y_1)^k.$$

Looking now at the coefficient of z^N in the Taylor expansion of $\mathrm{P}_{y_1^k}$, which is, using a common notation, $[z^N]\mathrm{P}_{y_1^k}(z) = \mathrm{H}_{y_1^k}(N)$, we derive from Equation (18) the following

Lemma 2. Let $\mathrm{Const} = \sum_{k\geq 0} \mathrm{H}_{y_1^k}\, y_1^k$. Then

$$\mathrm{Const} = \exp\left[-\sum_{k\geq 1} \mathrm{H}_{y_k} \frac{(-y_1)^k}{k}\right] \text{and } \mathrm{Const}^{-1} = \exp\left[\sum_{k\geq 1} \mathrm{H}_{y_k} \frac{(-y_1)^k}{k}\right].$$

With the introduction of the series Mono, we can now sum up equations (10),(12) into

Proposition 3. $\mathrm{P}(z) \underset{z\to 0}{\widetilde{\quad}} e^{x_0 \log z}$ and $\mathrm{P}(z) \underset{z\to 1}{\widetilde{\quad}} \mathrm{Mono}(z) Z$.

Corollary 1. Let $\Pi(z) = \pi_Y \mathrm{P}(z) = \sum_{w\in Y^*} \mathrm{P}_w(z)\, w$. Then $\Pi(z) \underset{z\to 1}{\widetilde{\quad}} \pi_Y \mathrm{Mono}(z) \pi_Y Z$.

Once again, looking at Taylor coefficients of P_w, we derive, as a direct consequence of Lemma 1, the following equivalent for the generating series $\mathrm{H}(N) = \sum_{w\in Y^*} \mathrm{H}_w(N) w$,

Corollary 2. $\mathrm{H}(N) \underset{N\to\infty}{\widetilde{\quad}} \mathrm{Const}(N) \pi_Y Z$.

Theorem 6. *(Hoang Ngoc Minh, 2007)*

$$\lim_{z\to 1} \exp\Big(y_1 \log(1-z) \Big) \Lambda(z) \;=\; \lim_{N\to\infty} \exp\Big(\sum_{k\geq 1} \mathrm{H}_{y_k}(N) \frac{(-y_1)^k}{k} \Big) \mathrm{H}(N) = \pi_Y Z,$$

where the limit shall be understood as a limit word by word.

This rewriting of Formula (24), and of Corollary 2 provides a theorem “à l'Abel”, linking the behaviour of the series Λ as z tends to 1 with the asymptotic behaviour of multiple harmonic sums (seen as Taylor coefficients) as N tends to $+\infty$.

4.5. Generalized Euler constants

Since Z is well known (and of course $\pi_Y Z$ also) and already studied, let us see now how to exploit Theorem 6 to get precise information about divergent harmonic sums.

Definition 4. We define, for any word $w \in Y^*$, the “generalized Euler constant”, denoted by $\zeta_{⧢}(w)$, as the constant occuring in the asymptotic expansion of $\mathrm{H}_w(n)$ in the scale of $\{n^\alpha \log^\beta(n), \alpha \in \mathbb{Z}, \beta \in \mathbb{Z}\}$.

Proposition 4. The map $\zeta_{⧢} : (\mathbb{C}\langle Y \rangle, ⧢) \to (\mathbb{C}, .)$ is an algebra morphism (*i.e.* verifies for all $u, v \in Y^*, \zeta_{⧢}(u ⧢ v) = \zeta_{⧢}(u)\zeta_{⧢}(v)$), such that for a convergent word $w \in Y^* \setminus y_1 Y^*$, $\zeta_{⧢}(w) = \zeta(w)$ and $\zeta_{⧢}(y_1) = \gamma$.

Notation. For convenience, we will by now denote by $\mathcal{Z}$ the $\mathbb{Q}$-algebra generated by $\{\zeta(w), w \in Y^* \setminus y_1 Y^*\}$, and by $\mathcal{Z}'$ the $\mathbb{Q}$-algebra generated by $\mathcal{Z}$ and γ.

This proposition, consequence of the Proposition 1, provides sufficient conditions to compute all constants $\zeta_{⧢}(w), w \in Y^*$. Indeed, Theorem 2 enables one to write any word $w \in Y^*$ as a combination of (stuffle) powers of y_1, with coefficients in the set of convergent words. Precisely, $w = \sum_{j=0}^{|w|} c_j(w) ⧢ y_1^{⧢ j}$, with $c_j(w)$ convergent polynomials (which is equivalent to $\mathrm{H}_w = \sum_{j=0}^{|w|} \mathrm{H}_{c_j(w)} \mathrm{H}_1^j$).

Example 3. Let $w = y_1^2 y_2$, then

$$c_3(w) = 0,\; c_2(w) = 2y_2,\; c_1(w) = -y_2 y_1 - y_3,\; c_0(w) = y_2 y_1^2 + y_3 y_1 + y_4/2,$$

and applying Proposition 4 leads to (using the reduction table of polyzêtas)

$$\zeta_{⧢}(w) = \zeta(c_0(w)) + \zeta(c_1(w))\gamma + \zeta(c_2(w))\gamma^2 = \frac{7}{10}\zeta(2)^2 - 2\zeta(3)\gamma + \frac{\zeta(2)}{2}\gamma^2.$$

Proposition 5. $\zeta_{⧢}(y_1^k) = \sum_{\substack{s_1,\ldots,s_k>0 \\ s_1+\ldots+ks_k=k}} \frac{(-1)^k}{s_1!\ldots s_k!}(-\gamma)^{s_1}\left(-\frac{\zeta(2)}{2}\right)^{s_2}\ldots\left(-\frac{\zeta(k)}{k}\right)^{s_k}.$

Proof. By Formula (20) and applying the (surjective) morphism $\zeta_{⧢}$, we get the expected result. □

In consequence,

Theorem 7. *(Hoang Ngoc Minh, 2003b) For $k > 0$, the constant $\zeta_{⧢}(y_1^k)$ associated to divergent polyzêta $\zeta(y_1^k)$ is a polynomial of degree k in γ with coefficients in $\mathbb{Q}[\zeta(2), \zeta(2i+1)]_{0<i\le(k-1)/2}$. Moreover, for $l = 0, .., k$, the coefficient of γ^l is of weight $k - l$.*

Example 4. $\zeta_{⧢}(y_1^2) = \frac{\gamma^2 - \zeta(2)}{2}, \quad \zeta_{⧢}(y_1^3) = \frac{\gamma^3 - 3\zeta(2)\gamma + 2\zeta(3)}{6}.$

Let us consider the (exponential) partial Bell polynomials in the variables $\{t_l\}_{l\ge1}$, $b_{n,k}(t_1, \ldots, t_{n-k+1})$, defined by the exponential generating series :

$$\sum_{n=0}^{\infty}\sum_{k=0}^{n} b_{n,k}(t_1, \ldots, t_{n-k+1})\frac{v^n u^k}{n!} = \exp\left(u\sum_{l=1}^{\infty} t_l \frac{v^l}{l!}\right). \tag{25}$$

In particular, we have

Lemma 3. Let $t_m = (-1)^{m+1}(m-1)!\zeta_{⧢}(m)$, for $m \ge 1$. Then

$$\exp\left[-\sum_{k\ge1}\zeta_{⧢}(k)\frac{(-y_1)^k}{k}\right] = 1 + \sum_{n\ge1}\left[\sum_{k=1}^{n} b_{n,k}(\gamma, -\zeta(2), 2\zeta(3), \ldots)\right]\frac{y_1^n}{n!}.$$

Let us build the noncommutative generating series of $\zeta_{⧢}(w)$ and let us take the constant part of the two members of $\mathrm{H}(N) \underset{N\to\infty}{\widetilde{\quad}} \mathrm{Const}(N)\pi_Y Z$, we have

Theorem 8. *(Hoang Ngoc Minh, 2007) Let $Z_{⧢} = \sum_{w\in Y^*}\zeta_{⧢}(w)\, w$ be the noncommutative generating series of the constants $\zeta_{⧢}(w)$. Then*

$$Z_{⧢} = \left[1 + \sum_{n\ge1}\left(\sum_{k=1}^{n} b_{n,k}(\gamma, -\zeta(2), 2\zeta(3), \ldots)\right)\frac{y_1^n}{n!}\right]\pi_Y Z.$$

Identifying coefficients of $y_1^k w$ in each member leads to

Corollary 3. For all $w \in Y^* \setminus y_1 Y^*$ and $k \ge 0$, we have

$$\zeta_{⧢}(y_1^k w) = \sum_{i=0}^{k}\frac{\zeta_{ш}(x_1^{k-i}\pi_X w)}{i!}\left[\sum_{j=1}^{i} b_{i,j}(\gamma, -\zeta(2), 2\zeta(3), \ldots)\right],$$

$\pi_X w$ standing for the translation of w in alphabet X.

Finally, using the expression of $\zeta_{ш}(x_1^{k-i}\pi_X w)$ given in (Hoang Ngoc Minh et al., 2001), i.e. for a word $u \in X^* x_1, \zeta_{ш}(x_1^k x_0 u) = \zeta\left(x_0(x_1^k ш\, u)\right)$, we get the following

Corollary 4. For all $w \in Y^* \setminus y_1Y^*$ and $k \geq 0$, we have

$$\zeta_{\uplus}(y_1^k w) = \sum_{i=0}^{k} \frac{\zeta(x_0[(-x_1)^{k-i} ⧢ u])}{i!} \left[\sum_{j=1}^{i} b_{i,j}(\gamma, -\zeta(2), 2\zeta(3), \ldots)\right],$$

with u defined as the suffix such that $\pi_X w = x_0 u$.

Example 5. Consider the word $y_1^2 y_2$, corresponding, with the previous notations, to $k = 2$ and $w = y_2$, then $\pi_X w = x_0 x_1$, so $u = x_1$ and

$$\zeta(x_0[(-x_1)^2 ⧢ u]) = 3\zeta(x_0 x_1^3), \quad \zeta(x_0[(-x_1) ⧢ u]) = -2\zeta(x_0 x_1^2), \quad \zeta(x_0 u) = \zeta(x_0 x_1).$$

$$\begin{aligned} \text{So, } \zeta_{\uplus}(y_1^2 y_2) &= \zeta_{⧢}(x_1^2 \pi_X y_2) + \zeta_{⧢}(x_1 \pi_X y_2) b_{1,1}(\gamma) + \zeta(2)[b_{2,1}(-\zeta(2)) + b_{2,2}(\gamma)]/2 \\ &= 3\zeta(2,1,1) - 2\zeta(2,1)\gamma + \zeta(2)[-\zeta(2) + \gamma^2]/2, \end{aligned}$$

and using the reduction table, we find $\zeta_{\uplus}(y_1^2 y_2) = \zeta(2)\gamma^2/2 - 2\zeta(3)\gamma + 7\zeta(2)^2/10$, a result in agreement with Example 3.

Remark 6. We can in fact derive Proposition 5 from Corollary 3, in the special case $w = \epsilon$, since all values $\zeta_{⧢}(x_1^{k-i})$ are equal to zero, except when $i = k$, giving rise to $\zeta_{⧢}(\epsilon) = 1$, so $\zeta_{\uplus}(y_1^k) = \dfrac{1}{k!}\left[\sum_{j=1}^{k} b_{k,j}\big(\gamma, -\zeta(2), 2\zeta(3), \ldots, (-1)^{k-1}(k-1)!\zeta(k)\big)\right]$ □

In consequence,

Theorem 9. *(Hoang Ngoc Minh, 2003b) For $w \in Y^* \setminus y_1Y^*, k \geq 0$, the constant $\zeta_{\uplus}(y_1^k w)$ associated to $\zeta(y_1^k w)$ is a polynomial of degree k in γ and with coefficients in $\mathcal{Z}$. Moreover, for $l = 0, .., k$, the coefficient of γ^l is of weight $|w| + k - l$.*

4.6. Asymptotic aspects of harmonic sums

In the previous section, we introduced the constant $\zeta_{\uplus}(w)$ as the constant term involved in the asymptotic expansion, of $\mathrm{H}_w(n)$. Let us be more precise about the computation of this expansion.

A first path (see (Costermans et al., 2005a)) to get this expansion is given by the property of L (relation (12)), that we interpret immediately on P by $\mathrm{P}(1-z) = (1-z)z^{-1}[\sigma\mathrm{P}(z)]Z$. Then, for a given word w, we can write $\mathrm{P}_w(z)$ as a $\mathbb{C}\left[z, z^{-1}, (1-z)^{-1}\right]$-linear combination of some $\mathrm{P}_u(1-z)$, that we can expand up to any order, getting so a singular expansion of $\mathrm{P}_w(z)$ around $z = 1$, in the scale of functions $\{(1-z)^\alpha \log(1-z)^\beta, \alpha \in \mathbb{Z}, \beta \in \mathbb{N}\}$. According to singularity analysis principles, this expansion gives rise to an asymptotic expansion of its Taylor coefficient $[z^n]\mathrm{P}_w(z)$ i.e. $\mathrm{H}_w(n)$.

Example 6. Following Remark 3, we have

$$\mathrm{P}_{2,1}(1-z) = \frac{1-z}{z}\left(-\mathrm{P}_3(z) + \log(z)\mathrm{P}_2(z) - \frac{1}{2}\log^2(z)\mathrm{P}_1(z) + \frac{\zeta(3)}{1-z}\right)$$

$$\mathrm{P}_{2,1}(z) = \frac{\zeta(3)}{1-z} + \log(1-z) - 1 - \frac{\log^2(1-z)}{2}$$
$$+ (1-z)\left(-\frac{\log^2(1-z)}{4} + \frac{\log(1-z)}{4}\right) + \mathrm{O}(|1-z|).$$

But $[z^N]\zeta(3)(1-z)^{-1} = \zeta(3),\ [z^N]\log(1-z) = -N^{-1},$
$$[z^N]\frac{\log^2(1-z)}{2} = [z^N]\frac{2!(1-z)\mathrm{P}_{y_1^2}(z)}{2} = \mathrm{H}_{y_1^2}(N) - \mathrm{H}_{y_1^2}(N-1)\ldots$$

We find finally, using Formula (20) which expresses $\mathrm{H}_{y_1^r}$ as a polynomial combination of single-indexed harmonic sums (i.e. *generalized harmonic numbers*),

$$[z^N]\mathrm{P}_{2,1}(z) = \mathrm{H}_{2,1}(N) = \zeta(3) - \frac{\log(N)+1+\gamma}{N} + \frac{1}{2}\frac{\log(N)}{N^2} + \mathrm{O}\left(\frac{1}{N^2}\right).$$

This approach is very efficient, but the main difficulty is to extract from Formula (12) the coefficients of a particular word w, mostly when the *weight* of w grows. We will refer to this algorithm as *Algo* 1.

Another path to get the asymptotic expansion of $\mathrm{H}_w(n)$ is to start from the recursive Formula (22), and to expand the numerator. This gives rise to a recursive algorithm, which initialization is given by the Euler-MacLaurin summation formula giving the expansion of generalized harmonic numbers.

Example 7.

$$\mathrm{H}_{4,2}(N) = \zeta(4,2) - \sum_{i=N+1}^{\infty}\frac{\mathrm{H}_2(i-1)}{i^4}, \quad \text{but} \quad \mathrm{H}_2(i-1) = \zeta(2) - \frac{1}{i} - \frac{1}{2}\frac{1}{i^2} + \mathrm{O}\left(\frac{1}{i^3}\right),$$

$$\text{So,}\quad \mathrm{H}_{4,2}(N) = \zeta(4,2) - \zeta(2)\sum_{i=N+1}^{\infty}\frac{1}{i^4} + \sum_{i=N+1}^{\infty}\frac{1}{i^5} + \frac{1}{2}\sum_{i=N+1}^{\infty}\frac{1}{i^6} + \sum_{i=N+1}^{\infty}\mathrm{O}\left(\frac{1}{i^7}\right)$$

Expanding the sums in N, we finally find

$$\mathrm{H}_{4,2}(N) = \zeta(4,2) - \frac{1}{3}\frac{\zeta(2)}{N^3} + \frac{\frac{1}{2}\zeta(2)+\frac{1}{4}}{N^4} - \frac{\frac{1}{3}\zeta(2)+\frac{2}{5}}{N^5} + \mathrm{O}\left(\frac{1}{N^6}\right).$$

Unfortunately, these basic principles do not work for a divergent word, for example, $w = y_1y_2$, since using Euler-MacLaurin summation formula can give the divergent part and the infinitesimal part, but not the N-free term (we are going to make this point precise). A possible solution to avoid this problem was given in (Costermans et al., 2005b), and consisted in decomposing the divergent word thanks to Theorem 2. For instance, $y_1y_2 = y_1 ⧢ y_2 - y_2y_1 - y_3$, so there are just the computations of the expansions for y_2y_1 and y_3 left, expansions for which the previous principles may be applied. The main problem here is the cost of the decomposition from Theorem 2, mostly when the *length* of the word w grows. We refer to this algorithm as *Algo* 2.

Following an idea suggested by B. Salvy, the authors were interested in seeing what happened when applying naively Euler-MacLaurin formula on the numerator, in Formula (22), without making difference between divergent and convergent words. To be more explicit, let us explain the case $w = y_1y_2$: using the software Maple [1]

$> numer1 := asympt(eulermac(1/k^2, k = 1..n-1, 2), n);$

$$numer1 := \frac{\pi^2}{6} - \frac{1}{n} + \mathrm{O}\left(\frac{1}{n^2}\right)$$

$> numer2 := asympt(eulermac(1/k^2, k = 1..n-1, 3), n);$

$$numer2 := \frac{\pi^2}{6} - \frac{1}{n} - \frac{2}{n^2} + \mathrm{O}\left(\frac{1}{n^3}\right)$$

$> asympt(eulermac(numer1/n, n = 1..N, 2), N);$

$$\frac{1}{6}\,\pi^2 \ln(N) - \frac{1}{6}\pi^2 + \frac{1}{6}\pi^2\gamma + \frac{\frac{1}{12}\pi^2 + 1}{N} + \mathrm{O}\left(\frac{1}{N^2}\right)$$

$> asympt(eulermac(numer2/n, n = 1..N, 2), N);$

$$\frac{1}{6}\,\pi^2 \ln(N) - \frac{1}{6}\pi^2 + \frac{1}{6}\pi^2\gamma - \frac{1}{2}\zeta(3) + \frac{\frac{1}{12}\pi^2 + 1}{N} + \mathrm{O}\left(\frac{1}{N^2}\right).$$

Here we can see that expanding further the numerator modifies the final constant term. In fact, we turned this idea into an efficient (and correct) algorithm by replacing the previous wrong N-free term $-\pi^2/6 + \pi^2\gamma/6 - \zeta(3)/2$ by $\zeta_{⧢}(y_1y_2) = \gamma\zeta(2) - 2\zeta(3)$. This new algorithm, that we will call *Algo* 3 appears as the fastest of our three methods.

Here are some examples of time for computing the expansion of $\mathrm{H}_w(N)$ up to $\mathrm{O}(N^{-3})$. For *Algo* 1, the series L and $\sigma(L)Z$ are supposed to be constructed, and the computing time starts with the extraction of coefficient w in both series. The empty value on the third line means precisely that the construction of the series $\sigma(L)Z$ can not be computed in a "reasonable" time.

word w	weight	length	Algo 1	Algo 2	Algo 3
$y_1y_2y_3$	6	3	2.4	0.7	0.1
$y_1^3y_3^2$	8	5	3.0	22.4	1.2
$y_1^5y_7$	12	6	-	45.9	5.1

5. Applications to discrete probabilities

5.1. *Coupon collector's problem*

Foata et al. (2001) considered the coupon collector's problem, in the case where the collector, supposed to be the eldest, has got a brotherhood, say r brothers. When he

[1] In the following lines, to avoid some steps, unnecessary for the reader, we do not show the instructions for collecting, re-ordering, expanding etc.

gets a coupon he has already, he gives it to his oldest brother, who keeps it or gives it to the oldest of the $r-1$ remaining brothers, and so on. Defining T the (random) number of chocolate bars he must buy to complete his collection made of m different coupons, and $M_T^{(j)}$ the number of empty places in the collection of the j-th brother at this time, Foata et al. (2001) show that $\mathbb{E}[M_T^{(k)}] = \sum_{j=0}^k K_m^{(j)}$, where for all $m \geq 1$, $K_m^{(0)} = 1$ and $K_m^{(k)} = \sum_{j=2}^m K_j^{(k-1)}/j$.

Foata et al. (2001) remarked that these numbers $K_m^{(k)}$, called *hyperharmonic numbers* are variations of generalized harmonic numbers $\mathrm{H}_r(m)$. Indeed, for $k \geq 1$, $K_m^{(k)}$ can be written as $K_m^{(k)} = \underline{\mathrm{H}}_{y_1^k}(m) - \underline{\mathrm{H}}_{y_1^{k-1}}(m)$.

Proof. By induction on k. It is obvious if $k=1$ (recall that $\underline{\mathrm{H}}_\epsilon \equiv 1$). Then supposing it is true for $k > 1$, and since $K_1^{(k)} = 0$, we just apply Formula (22)

$$K_m^{(k+1)} = \sum_{j=2}^m \frac{K_j^{(k)}}{j} = \sum_{j=1}^m \frac{\underline{\mathrm{H}}_{y_1^k}(j) - \underline{\mathrm{H}}_{y_1^{k-1}}(j)}{j} = \underline{\mathrm{H}}_{y_1^{k+1}}(m) - \underline{\mathrm{H}}_{y_1^k}(m). \quad \square$$

Consequently, $\mathbb{E}[M_T^{(k)}]$ may be further simplified in $\mathbb{E}[M_T^{(k)}] = \underline{\mathrm{H}}_{y_1^k}(m)$. For instance, as m tends to $+\infty$,

$$\begin{aligned}
\mathbb{E}[M_T^{(2)}] &= \frac{1}{2}\ln^2(m) + \gamma\ln(m) + \frac{\gamma^2+\zeta(2)}{2} + \frac{\ln(m)+\gamma-1}{m} + \mathrm{O}\left(\frac{\ln(m)}{m^2}\right) \\
\mathbb{E}[M_T^{(3)}] &= \frac{1}{6}\ln^3(m) + \frac{1}{2}\ln^2(m)\,\gamma + \left(\frac{\gamma^2+\zeta(2)}{2}\right)\ln(m) \\
&\quad + \frac{\gamma\zeta(2)}{2} + \frac{\gamma^3}{6} + \frac{\zeta(3)}{3} + \mathrm{O}\left(\frac{\ln^2(m)}{m}\right) \\
\mathbb{E}[M_T^{(4)}] &= \frac{1}{24}\ln^4(m) + \frac{1}{6}\ln^3(m)\,\gamma + \left(\frac{\gamma^2+\zeta(2)}{4}\right)\ln^2(m) + \left(\frac{\gamma\zeta(2)}{2} + \frac{\gamma^3}{6} + \frac{\zeta(3)}{3}\right)\ln(m) \\
&\quad + \frac{1}{24}\gamma^4 + \frac{1}{3}\zeta(3)\,\gamma + \frac{9}{40}\zeta(2)^2 + \frac{\zeta(2)\gamma^2}{4} + \mathrm{O}\left(\frac{\ln^3(m)}{m}\right).
\end{aligned}$$

5.2. *Root-subtrees in multidimensional quadtrees*

Following the notations suggested in (Labelle and Laforest, 1995b; Labelle et al., 2006), given the hypercube $[0,1]^d$, an initial point $X_1 = (t_1, \ldots, t_d)$ divides the hypercube in 2^d hyperoctants, and we index by the binary word $\epsilon_1 \ldots \epsilon_d \in \{0,1\}^d$, the hyperoctant containing the vertex $(\epsilon_1, \ldots, \epsilon_d)$.

Let S be a set of binary words encoding hyperoctants, we denote by $J_n[S]$ the probability that n points i.i.d fall in this set. With the practical notations $t_i^{<0>} = t_i$ and $t_i^{<1>} = 1 - t_i$, probability $J_n[S]$ is given by the following multiple integral

$$J_n[S] = \int_0^1 \ldots \int_0^1 f_S(t_1, \ldots, t_d)^{n-1} dt_1 \ldots dt_d, \tag{26}$$

where $f_S(t_1, \ldots, t_d) = \sum_{\epsilon \in S} \prod_{i=1}^d t_i^{<\epsilon_i>}$.

Example 8. In dimension 2, let $X_1 = (t_1, t_2)$ and $S = \{01, 11\}$ (encoding the two quadrants "north"), or rather (identifying the binary encoding with its decimal equivalent) $S = \{1, 3\}$, then

$$J_n[S] = \int_0^1 \int_0^1 (t_1(1-t_2) + (1-t_1)(1-t_2))^{n-1} dt_1 dt_2.$$

One can remark that replacing a set S with its complement replaces, in Formula (26), the expression $f_S(t_1, \ldots, t_d)$ by its 1's complement. By consequence, $\overline{S}$ standing for the complement set of S, $J_n[S] = \sum_{k=1}^{n} (-1)^{k-1} \binom{n-1}{k-1} J_k[\overline{S}]$.

In the case where $S = \{1, 2, 3, 4, 5, 6, 7\}$, then $\overline{S} = \{0\}$, and

$$J_n[S] = \sum_{k=1}^{n} (-1)^{k-1} \binom{n-1}{k-1} \frac{1}{k^3} = -\frac{1}{n} \sum_{k=1}^{n} (-1)^k \binom{n}{k} \frac{1}{k^2}.$$

Here we are facing an Euler transform, and we can use some interesting (combinatoric) properties linking the operators ∇ and Σ, defined by - cf (Hoffman, 2005) -

$$\nabla : (a_n)_{n\geq 0} \mapsto \left(\sum_{k=0}^{n} (-1)^k \binom{n}{k} a_k \right)_{n\geq 0}, \qquad \Sigma : (a_n)_{n\geq 0} \mapsto \left(\sum_{k=0}^{n} a_k \right)_{n\geq 0}.$$

Indeed, Hoffman (2005) showed that $\Sigma\nabla\underline{\mathrm{H}}_w = -\underline{\mathrm{H}}_{w^\sharp}$, with $w^\sharp$ uniquely associated to w is constructed as follows : if $w = y_{s_1} \ldots y_{s_r}$ has for weight n, then you can consider the partial sums $\{s_1, s_1 + s_2, \ldots, s_1 + \ldots + s_{r-1}\}$ as a (maybe empty) subset of $\{1, \ldots, n-1\}$, for which you can take the complement set and construct from it the unique word $w^\sharp$ of weight n. For example, $y_2^\sharp = y_1 y_1, y_1 y_3^\sharp = y_2 y_1^2$.

Coming back to $S = \{1, 2, 3, 4, 5, 6, 7\}$, and since $\Sigma\nabla = \nabla\Sigma^{-1}$, we have

$$J_n[S] = -\frac{1}{n} \nabla\Sigma^{-1} \underline{\mathrm{H}}_{y_2}(n) = \frac{1}{n} \underline{\mathrm{H}}_{y_2^\sharp}(n) = \frac{1}{n} \underline{\mathrm{H}}_{y_1^2}(n).$$

Another example, still in dimension $d = 3$. If $S = \{2, 3, 4, 5, 6, 7\}$, then $J_k[\overline{S}] = k^{-2}$.

$$J_n[S] = \sum_{k=1}^{n} (-1)^{k-1} \binom{n-1}{k-1} \frac{1}{k^2} = -\frac{1}{n} \sum_{k=1}^{n} (-1)^k \binom{n}{k} \frac{1}{k} = -\frac{1}{n} \nabla\Sigma^{-1} \underline{\mathrm{H}}_{y_1}(n)$$
$$= \frac{1}{n} \underline{\mathrm{H}}_{y_1^\sharp}(n) = \frac{1}{n} \underline{\mathrm{H}}_{y_1}(n).$$

A last example, with $d = 3$, and $S = \{3, 4, 5, 6, 7\}$, for which $J_k[\overline{S}] = \underline{\mathrm{H}}_{y_1}(k)/k^2$.

$$J_n[S] = \sum_{k=1}^{n} (-1)^{k-1} \binom{n-1}{k-1} \frac{\underline{\mathrm{H}}_{y_1}(k)}{k^2} = -\frac{1}{n} \sum_{k=1}^{n} (-1)^k \binom{n}{k} \frac{\underline{\mathrm{H}}_{y_1}(k)}{k} = -\frac{1}{n} \nabla\Sigma^{-1} \underline{\mathrm{H}}_{y_1^2}(n)$$
$$= \frac{1}{n} \underline{\mathrm{H}}_{y_1^{2\sharp}}(n) = \frac{1}{n} \underline{\mathrm{H}}_{y_2}(n).$$

Let $Q = \{q_1, \ldots, q_n\}$ be a set of independent and identically distributed random vectors in $\mathbb{R}^d$. A point $q_i = (q_{i_1}, \ldots, q_{i_d})$ is said to be dominated by $q_j = (q_{j_1}, \ldots, q_{j_d})$ if $q_{i_k} < q_{j_k}$ for all $k \in \{1, .., d\}$ and a point q_i is called a maximum of Q if none of the other points dominates it. The number of maxima of Q is denoted by $K_{n,d}$.

Recently, Bai et al. (2005) proposed a method for computing an asymptotic expansion of the variance and the study of $\mathbb{V}ar(K_{n,d})$ for random samples from $[0,1]^d$ is precisely the goal of the present section. For that, we exploit the following result, first derived by Ivanin (1976) :

$$\mathbb{E}(K_{n,d}^2) = \sum_{1 \le i_1 \le \ldots \le i_{d-1} \le n} \frac{1}{i_1 \ldots i_{d-1}} + \sum_{1 \le t \le d-1} \binom{d}{t} \sum_{l=1}^{n-1} \frac{1}{l} \sum^{(*)} \frac{1}{i_1 \ldots i_{d-2} j_1 \ldots j_{d-1}}, \quad (27)$$

where the sum $(*)$ is taken over indices verifying $1 \le i_1 \ldots \le i_{t-1} \le l, 1 \le i_t \le \ldots \le i_{d-2} \le l$ and $l+1 \le j_1 \le \ldots \le j_{d-1} \le n$. After having given an alternative derivation for this formula, Bai et al. deduce, by analytic and combinatoric considerations, as the main result of (Bai et al., 1998), the following equivalent [2]

$$\mathbb{V}ar(K_{n,d}) \sim \left(\frac{1}{(d-1)!} + \kappa_d \right) \ln^{d-1}(n), \text{ with}$$

$$\kappa_d = \sum_{t=1}^{d-2} \frac{1}{t!(d-1-t)!} \sum_{l \ge 1} \frac{1}{l^2} \sum^{(**)} \frac{1}{i_1 \ldots i_{t-1} j_1 \ldots j_{d-2-t}} \quad (28)$$

the sum $(**)$ being calculated over all indices verifying $1 \le i_1 \le \ldots \le i_{t-1} \le l$ and $1 \le j_1 \le \ldots \le j_{d-2-t} \le l$.

We first focus on the asymptotic equivalent of $\mathbb{V}ar(K_{n,d})$ from Formula (28), κ_d being re-written, with our tools, in the following way

$$\kappa_d = \frac{1}{(d-1)!} \sum_{t=1}^{d-2} \binom{d-1}{t} \sum_{l \ge 1} \frac{1}{l^2} \underline{\mathrm{H}}_{y_1^{t-1} \sqcup\!\sqcup\, y_1^{d-2-t}}(l) \quad (29)$$

We need a last *ad hoc* notation.

Definition 5. Let S be a subset of Y, and ρ a positive integer, we define S_ρ as the set of words containing only letters in S, and of weight equal to ρ.

Example 9. Let $S = \{y_1, y_2\}$ and $\rho = 4$ then $S_\rho = \{y_1^4, y_1 y_2 y_1, y_1^2 y_2, y_2 y_1^2, y_2^2\}$.

We recall that $|w|_2$ stands for the number of occurrences of the letter y_2 in w, and so we turn Equation 29 into a closed form,

Theorem 10. $\kappa_d = \frac{1}{(d-1)!} \sum_{w \in \{y_1, y_2\}_{d-3}} (-1)^{|w|_2} \binom{2\,(d-2-|w|_2)}{d-2-|w|_2} \underline{\zeta}(y_2 w).$

[2] The value of the mean of $K_{n,d}$ is known to be $\underline{\mathrm{H}}_{y_1^{d-1}}(n)$ (Barndorff-Nielsen and Sobel, 1966)

Example 10. For $d = 7$, we get

$$6!\kappa_7 = \binom{10}{5}\underline{\zeta}(2,1,1,1,1) - \binom{8}{4}\Big(\underline{\zeta}(2,2,1,1) + \underline{\zeta}(2,1,2,1) + \underline{\zeta}(2,1,1,2)\Big) + \binom{6}{3}\underline{\zeta}(2,2,2).$$

For $d = 9$ and $d = 10$, using the reduction table, we get

$$\begin{aligned}\kappa_9 &= -\frac{17}{1920}\zeta(6,2) + \frac{11}{160}\zeta(3)\zeta(5) + \frac{1}{320}\zeta(2)\zeta(3)^2 + \frac{1891}{89600}\zeta(2)^4 \\ \kappa_{10} &= \frac{529}{75600}\zeta(2)^2\zeta(5) + \frac{33941}{6350400}\zeta(2)^3\zeta(3) + \frac{17}{3360}\zeta(2)\zeta(7) \\ &\quad + \frac{199271}{4354560}\zeta(9) + \frac{11}{12960}\zeta(3)^3.\end{aligned}$$

Now, interested in the whole expansion of the variance, we can turn Expression (27) into $\mathbb{E}(K_{n,d}^2) = \underline{\mathrm{H}}_{y_1^{d-1}}(n) + \sum_{1\le t\le d-1} \binom{d}{t} \sum_{l=1}^{n-1} \frac{1}{l} \underline{\mathrm{H}}_{y_1^{t-1}}(l)\underline{\mathrm{H}}_{y_1^{d-t-1}}(l)\underline{\mathrm{H}}_{y_1^{d-1}}(n; l+1)$, the notation $\underline{\mathrm{H}}_w(n; l+1)$ being the same than in Definition 3, but where all indices involved in the summation are bounded lowerly by $l + 1$. From this point, we get the following

Theorem 11. *For all $d \geq 2$, there exist an integer $K > 0$, some integers $(\alpha_i)_{1\le i\le K}$ and some words $w_i \in Y^*$such that $\mathbb{V}ar(K_{n,d}) = \sum_{i=1}^{K} \alpha_i \underline{\mathrm{H}}_{w_i}(n)$.*

By consequence, there exist algorithmically computable coefficients $\alpha_i, \beta_{j,k} \in \mathcal{Z}'$ such that, for any dimension d and any order M,

$$\mathbb{V}ar(K_{n,d}) = \sum_{i=0}^{2d-2} \alpha_i \ln^i(n) + \sum_{j=1}^{M} \frac{1}{n^j} \sum_{k=0}^{2d-2} \beta_{j,k} \ln^k(n) + \mathrm{o}\left(\frac{1}{n^M}\right).$$

Example 11.

$$\begin{aligned}\mathbb{V}ar(K_{\cdot,3}) &= -\underline{\mathrm{H}}_{y_2^2} + 2\underline{\mathrm{H}}_{y_1^2 y_2} - 4\underline{\mathrm{H}}_{y_1 y_2 y_1} + 2\underline{\mathrm{H}}_{y_2 y_1^2} + \underline{\mathrm{H}}_{y_1^2} \\ \mathbb{V}ar(K_{\cdot,4}) &= \underline{\mathrm{H}}_{y_2^3} + 6\,\underline{\mathrm{H}}_{y_1^3 y_2 y_1} - 14\,\underline{\mathrm{H}}_{y_1^2 y_2 y_1^2} + 6\,\underline{\mathrm{H}}_{y_1 y_2 y_1^3} + 6\,\underline{\mathrm{H}}_{y_2 y_1^4} - 2\,\underline{\mathrm{H}}_{y_2^2 y_1^2} \\ &\quad + 4\,\underline{\mathrm{H}}_{y_1^2 y_2^2} - 2\,\underline{\mathrm{H}}_{y_1 y_2 y_1 y_2} - 2\,\underline{\mathrm{H}}_{y_2 y_1^2 y_2} + \underline{\mathrm{H}}_{y_1^3} - 2\,\underline{\mathrm{H}}_{y_1 y_2^2 y_1} - 2\,\underline{\mathrm{H}}_{y_2 y_1 y_2 y_1}.\end{aligned}$$

$$\begin{aligned}\text{So,}\quad \mathbb{V}ar(K_{n,3}) &= \left(\frac{1}{2} + \kappa_3\right)\ln^2(n) + (-10\zeta(3) + 2\zeta(2)\gamma + \gamma)\ln(n) + \frac{1}{2}\gamma^2 \\ &\quad - 10\zeta(3)\gamma + \frac{83}{10}\zeta(2)^2 + \zeta(2)\gamma^2 + \frac{1}{2}\zeta(2) + \mathrm{o}(1) \\ \mathbb{V}ar(K_{n,4}) &= \left(\frac{1}{3!} + \kappa_4\right)\ln^3(n) + \left(-\frac{53}{5}\zeta(2)^2 + 6\zeta(3)\gamma + \frac{1}{2}\gamma\right)\ln^2(n) \\ &\quad + \left(97\zeta(5) - \frac{106}{5}\zeta(2)^2\gamma + 16\zeta(2)\zeta(3) + 6\zeta(3)\gamma^2 + \frac{1}{2}\zeta(2) + \frac{1}{2}\gamma^2\right)\ln(n) \\ &\quad + \frac{1}{3}\zeta(3) - \frac{53}{5}\zeta(2)^2\gamma^2 - \frac{3719}{70}\zeta(2)^3 + \frac{1}{6}\gamma^3 + \frac{1}{2}\zeta(2)\gamma \\ &\quad + 16\zeta(2)\zeta(3)\gamma - 3\zeta(3)^2 + 2\zeta(3)\gamma^3 + 97\zeta(5)\gamma + \mathrm{o}(1)\end{aligned}$$

$$\begin{aligned}\mathbb{V}ar(K_{n,5}) = &\left(\frac{1}{4!}+\kappa_5\right)\ln^4(n)+\left(\frac{1}{6}\gamma-\frac{98}{3}\zeta(5)+\frac{33}{10}\zeta(2)^2\gamma-\frac{13}{3}\zeta(2)\zeta(3)\right)\ln^3(n)\\
&+\left(\frac{10123}{140}\zeta(2)^3+\frac{47}{2}\zeta(3)^2+\frac{99}{20}\zeta(2)^2\gamma^2+\frac{1}{4}\gamma^2+\frac{1}{4}\zeta(2)-13\zeta(2)\zeta(3)\gamma\right.\\
&\left.-98\zeta(5)\gamma\right)\ln^2(n)+\left(\frac{1}{6}\gamma^3+\frac{33}{10}\zeta(2)^2\gamma^3+\frac{1}{2}\zeta(2)\gamma-950\zeta(7)\right.\\
&-13\zeta(2)\zeta(3)\gamma^2+47\zeta(3)^2\gamma+\frac{1}{3}\zeta(3)-\frac{317}{5}\zeta(3)\zeta(2)^2+\frac{10123}{70}\zeta(2)^3\gamma\\
&\left.-98\zeta(5)\gamma^2-222\zeta(2)\zeta(5)\right)\ln(n)-\frac{13}{3}\zeta(2)\zeta(3)\gamma^3+\frac{47}{2}\zeta(3)^2\gamma^2\\
&-\frac{317}{5}\zeta(3)\zeta(2)^2\gamma-\frac{98}{3}\zeta(5)\gamma^3+\frac{33}{40}\zeta(2)^2\gamma^4+\frac{32}{3}\zeta(3)\zeta(5)+\frac{10123}{140}\zeta(2)^3\gamma^2\\
&-222\zeta(2)\zeta(5)\gamma+\frac{1}{24}\gamma^4-950\zeta(7)\gamma+50\zeta(6,2)+\frac{1}{4}\zeta(2)\gamma^2+\frac{1}{3}\zeta(3)\gamma\\
&+\frac{9}{40}\zeta(2)^2+\frac{95}{6}\zeta(2)\zeta(3)^2+\frac{134739}{350}\zeta(2)^4+\mathrm{o}(1).\end{aligned}$$

Acknowledgements

We acknowledge the influence of Cartier's lectures at the GdT *Polylogarithmes et Polyzêtas*. We greatly appreciated many fruitful discussions with Boutet de Monvel, Enjalbert, Jacob, Petitot and Waldschmidt.

References

Bai, Z., Chao, C., Hwang, H., Liang, W., 1998. On the variance of the number of maxima in random vectors and its applications. Annals of Applied Probability 8, 886–895.

Bai, Z., Devroye, L., Hwang, H., Tsai, T., 2005. Maxima in hypercubes. Random structures and Algorithmes 27, 290–309.

Barndorff-Nielsen, O., Sobel, M., 1966. On the distribution of the number of admissible points in a vector random sample. Theory Proba. Appl. 11, 249–269.

Costermans, C., Enjalbert, J., Hoang Ngoc Minh, 2005a. Algorithmic and combinatoric aspects of multiple harmonic sums. DMTCS `http://www.dmtcs.org/dmtcs-ojs/index.php/proceedings/article/view/dmAD0107/`.

Costermans, C., Enjalbert, J., Hoang Ngoc Minh, Petitot, M., 2005b. Structure and asymptotic expansion of multiple harmonic sums. In: International Symposium on Symbolic and Algebraic Computation. pp. 100–107.

Costermans, C., Hoang Ngoc Minh, 2007. Variance for the number of maxima in hypercubes and generalized euler constants. In: Formal Power Series and Algebraic Combinatorics. `http://www.fpsac.cn/PDF-Proceedings/Posters/63.pdf`.

Drinfel'd, V., 1990. Quasi–hopf algebras. Len. Math. Journal 1, 1419–1457.

Flajolet, P., Sedgewick, R., 1996. An Introduction to the Analysis of Algorithms. Addison-Wesley.

Flajolet, P., Vallée, B., 2000. Continued fractions, comparison algorithms, and fine structure constants. In: Constructive, Experimental, and Nonlinear Analysis. pp. 53–82.

Foata, D., Han, G., Lass, B., 2001. Les nombres hyperharmoniques et la fratrie du collectionneur de vignettes. Jour. Élec. du Séminaire Lotharingien de Combinatoire 47.

Gessel, I., 1984. Multipartite p-partitions and inner products of skew schur functions. Combinatorics and Algebra, Contemp. Math. 34, 289–301.

Hardy, G., 2000. Divergent Series, 3rd Edition. Jacques Gabay.

Hoang Ngoc Minh, 2003a. Differential galois groups and noncommutative generating series of polylogarithms. In: 7th World Multiconference on Systemics, Cybernetics and informatics, Orlando, Floride. pp. 128–135.

Hoang Ngoc Minh, 2003b. Finite polyzêtas, poly-bernoulli numbers, identities of polyzêtas and noncommutative rational power series. In: 4th International Conference on Words. pp. 232–250.

Hoang Ngoc Minh, 2007. Algebraic combinatoric aspects of asymptotic analysis of nonlinear dynamical system with singular inputs. In: Computer Algebra and Differential Equations. Vol. 67 of B. Acta Academiae Aboensis, pp. 117–126.

Hoang Ngoc Minh, Jacob, G., Oussous, N., Petitot, M., 2001. De l'algèbre des ζ de riemann multivariées à l'algèbre des ζ de hurwitz multivariées. Journal électronique du Séminaire Lotharingien de Combinatoire 44.

Hoang Ngoc Minh, Petitot, M., Van der Hoeven, J., 1998. Polylogarithms and shuffle algebra. In: Formal Power Series and Algebraic Combinatorics.

Hoang Ngoc Minh, Petitot, M., Van der Hoeven, J., 1999. L'algèbre des polylogarithmes par les séries génératrices. In: Formal Power Series and Algebraic Combinatorics.

Hoffman, M. E., 1997. The algebra of multiple harmonic series. Jour. of Alg. 194, 477–495.

Hoffman, M. E., 2000. Quasi-shuffle products. Jour. of Alg. Comb. 11, 49–68.

Hoffman, M. E., 2004. The hopf algebra structure of multiple harmonic sums. Nuclear Phys. B. Proc Suppl. 135, 215–219.

Hoffman, M. E., 2005. Algebraic Aspects of Multiple Zeta Values. Vol. 14 of Dev. Math. Springer, New York.

Ivanin, V., 1976. Calculation of the dispersion of the number of elements of the pareto set for the choice of independent vectors with independent components. Theory of Optimal Decisions, Kiev 1, 90–100.

Knuth, D., 1997. The Art of Computer Programming, Vol. 1: Fundamental Algorithms, 3rd Edition. Addison-Wesley.

Labelle, G., Laforest, L., 1995a. Combinatorial variations on multidimensional quadtrees. Journal of Combinatorial Theory 69, 1–16.

Labelle, G., Laforest, L., 1995b. Sur la distribution de l'arité de la racine d'une arborescence hyperquaternaire. Discrete Mathematics 139, 287–302.

Labelle, G., Laforest, L., Provencal, X., 2006. Around the root of multidimensional quadtrees. DMTCS `http://www.dmtcs.org/dmtcs-ojs/index.php/proceedings/article/view/dmAG0125/1702`.

MacDonald, I., 1995. Symmetric functions and Hall polynomials, 2nd Edition. Oxford University Press, USA.

Malvenuto, C., Reutenauer, C., 1995. Duality between quasi-symmetric functions and the solomon descent algebra. Jour. of Alg. 177, 967–982.

Radford, D., 1979. A natural ring basis for shuffle algebra and an application to group schemes. Journal of Algebra 58, 432–454.

Reutenauer, C., 1993. Free Lie Algebras. Vol. 7 of London Mathematical Society Monographs, New Series. Oxford University Press, USA.