Explicit computation of the variance of the number of maxima in hypercubes

Christian Costermans and Hoang Ngoc Minh

Université Lille II, 1, Place Déliot, 59024 Lille, France

We present a combinatorial approach of the variance for the number of maxima in hypercubes. This leads to an explicit expression, in terms of Multiple Zeta Values, of the dominant term in the asymptotic expansion of this variance. Moreover, we get an algorithm to compute this expansion, and show that all coefficients occuring belong to the $\mathbb{Q}$-algebra generated by Multiple Zeta Values, and by Euler's constant γ.

Keywords: Maximal points, harmonic sums, shuffle algebra, multiple zeta values

1 Introduction

Let $\Lambda = \{x_1, \ldots, x_n\}$ be a set of independent and identically distributed random vectors in $\mathbb{R}^d$. A point $x_i = (x_{i_1}, \ldots, x_{i_d})$ is said to be dominated by $x_j = (x_{j_1}, \ldots, x_{j_d})$ if $x_{i_k} < x_{j_k}$ for all $k \in \{1, \ldots, d\}$ and a point x_i is called a maximum of Λ if none of the other points dominates it. The number of maxima of Λ is denoted by $K_{n,d}$.

Many papers were already devoted to the study of the number of maxima in a set of points, since it arises in various domains. Recently, in [2], Bai et al. proposed a method for computing an asymptotic expansion of the variance.

The study of $\mathbb{V}ar(K_{n,d})$ for random samples from $[0,1]^d$ is precisely the goal of the present paper. For that, we exploit an important result, first derived by Ivanin [5] :

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

where the sum $(*)$ is taken over indices verifying

$$1 \le i_1 \ldots \le i_{t-1} \le l, 1 \le i_t \le \ldots \le i_{d-2} \le l \quad \text{and} \quad l+1 \le j_1 \le \ldots \le j_{d-1} \le n.$$

In Formula (1), $\mu_{n,d}$ stands for the mean of $K_{n,d}$, first calculated by Barndorff-Nielsen and Sobel [3] :

$$\mu_{n,d} = \mathbb{E}(K_{n,d}) = \sum_{1 \le i_1 \le \ldots \le i_{d-1} \le n} \frac{1}{i_1 \ldots i_{d-1}}. \tag{2}$$

After having given an alternative derivation for this formula, Bai et al. deduce, by analytic and combinatoric considerations, as the main result of [1], the following equivalent

$$\mathbb{V}ar(K_{n,d}) \sim \left(\frac{1}{(d-1)!} + \kappa_d \right) \ln^{d-1}(n), \tag{3}$$

$$\text{with} \quad \kappa_d = \sum_{t=1}^{d-2} \frac{1}{t!(d-1-t)!} \sum_{l \ge 1} \frac{1}{l^2} \sum^{(**)} \frac{1}{i_1 \ldots i_{t-1} j_1 \ldots j_{d-2-t}} \tag{4}$$

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

These two formulas give rise to harmonic sums $\mathrm{A}_{\mathbf{s}}(N)$, closely related to $\mathrm{H}_{\mathbf{s}}(N)$ defined for a multi-index $\mathrm{s} = (s_1, \ldots, s_r)$ by

$$\mathrm{A}_{\mathbf{s}}(N) = \sum_{N \ge n_1 \ge \ldots \ge n_r \ge 1} \frac{1}{n_1^{s_1} \ldots n_r^{s_r}}, \qquad \mathrm{H}_{\mathbf{s}}(N) = \sum_{N \ge n_1 > \ldots > n_r \ge 1} \frac{1}{n_1^{s_1} \ldots n_r^{s_r}}.$$

We already studied in [4] the algebraic properties of $\mathrm{H}_{\mathbf{s}}(N)$; in particular, when $s_1 > 1$, $\mathrm{H}_{\mathbf{s}}(N)$ converges to the polyzêta (or MZV) $\zeta(\mathbf{s}) = \displaystyle\sum_{n_1>\ldots>n_r\geq 1} \frac{1}{n_1^{s_1}\ldots n_r^{s_r}}$. There exist explicit relations, given by Hoffman between the $\mathrm{A}_{\mathbf{s}}(N)$ and $\mathrm{H}_{\mathbf{s}}(N)$. Indeed, let $\mathrm{Comp}(n)$ be the *set of compositions* of n, i.e. sequences $(i_1,\ldots,i_r)$ of positive integers summing to n. 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. One has ($l(J)$ being the number of parts of J) :

$$\mathrm{A}_{\mathbf{s}}(N) = \sum_{J\in\mathrm{Comp}(r)} \mathrm{H}_{J\circ\mathbf{s}}(N) \quad \text{and} \quad \mathrm{H}_{\mathbf{s}}(N) = \sum_{J\in\mathrm{Comp}(r)} (-1)^{l(J)-r}\mathrm{A}_{J\circ\mathbf{s}}(N). \tag{5}$$

Here, the nature of Formulas (2) and (4) makes clear that it would be difficult to interpret both formulas in terms of $\mathrm{H}_{\mathbf{s}}(N)$. So, we prefer looking at the algebraic and combinatoric properties of $\mathrm{A}_{\mathbf{s}}(N)$, and deduce from these ones two main results, first the explicit value of κ_d in terms of Multiple Zeta Values, and then an algorithm to compute the asymptotic expansion of $\mathbb{V}ar(K_{n,d})$.

2 Combinatoric background

2.1 Combinatorics on words

To the multi-index $\mathbf{s}=(s_1,\ldots,s_r)$ we can canonically associate the word $v=y_{s_1}\ldots y_{s_r}$ over the infinite alphabet $Y=\{y_i\}_{i\geq 1}$. Its length r is denoted by $\ell(v)$, and its weight is defined as $|v|=\sum_{i=1}^r s_i$. The number of occurences of the letter y_i in the word w is denoted by $N_i(w)$. Moreover, the empty multi-index will correspond to the empty word ϵ.

Example 1 *Let* $w=y_1y_4y_1^2y_2$, *we have* $\ell(w)=5, |w|=9$ *and* $N_1(w)=3$.

Definition 1 *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 2 *Let* $S=\{y_1,y_2\}$ *and* $\rho=4$ *then* $S_\rho=\{y_1^4, y_1y_2y_1, y_1^2y_2, y_2y_1^2, y_2^2\}$.

We shall henceforth identify the multi-index $\mathbf{s}$ with its encoding by the word $v=y_{s_1}\ldots y_{s_r}$.

We denote by Y^* the free monoid generated by Y, which is the set of words over Y, and by $\mathbb{Q}\langle Y\rangle$ the algebra of non commutative polynomials with coefficients in $\mathbb{Q}$.

2.2 Shuffle product

Let $y_i,y_j\in Y$ and $u,v\in Y^*$. The *minus-stuffle* [(i)] of $u=y_iu'$ and $v=y_jv'$ is the polynomial recursively defined by

$$\begin{aligned}\epsilon \sqcup\!\!\!\sqcup\!\!\!\!-\, u &= u \sqcup\!\!\!\sqcup\!\!\!\!-\, \epsilon = u \qquad \text{and} &(6)\\ u \sqcup\!\!\!\sqcup\!\!\!\!-\, v &= y_i(u' \sqcup\!\!\!\sqcup\!\!\!\!-\, v) + y_j(u \sqcup\!\!\!\sqcup\!\!\!\!-\, v') - y_{i+j}(u' \sqcup\!\!\!\sqcup\!\!\!\!-\, v') &(7)\end{aligned}$$

For example, $y_1 \sqcup\!\!\!\sqcup\!\!\!\!-\, y_2 = y_1y_2+y_2y_1-y_3$.

Proposition 1 $y_1^r = \displaystyle\sum_{\substack{s_1,\ldots,s_r>0\\ s_1+\ldots+rs_r=r}} \frac{y_1^{\sqcup\!\!\!\sqcup\!\!\!\!-\, s_1} \sqcup\!\!\!\sqcup\!\!\!\!-\, \ldots \sqcup\!\!\!\sqcup\!\!\!\!-\, y_r^{\sqcup\!\!\!\sqcup\!\!\!\!-\, s_r}}{1^{s_1}s_1!\ldots r^{s_r}s_r!}$.

Definition 2 *Let* $w=y_{s_1}\ldots y_{s_r}\in Y^*$. *For* $N\geq k\geq 1$, *the harmonic sum* $\mathrm{A}_w(N;k)$ *is defined as* $\mathrm{A}_w(N;k) = \displaystyle\sum_{N\geq n_1\geq\ldots\geq n_r\geq k} \frac{1}{n_1^{s_1}\ldots n_r^{s_r}}$.

In particular, $\mathrm{A}_{y_{s_1}\ldots y_{s_r}}(N;1)$ stands for the multi-indexed notation $\mathrm{A}_{s_1,\ldots,s_r}(N)$.

For convenience, we will use the notation $\mathrm{A}_w(N)$ instead of $\mathrm{A}_w(N;1)$. We put $\mathrm{A}_w(0)=0$ and, for the empty word ϵ, we put $\mathrm{A}_\epsilon(N)=1$, for any $N\geq 0$. The definition is extended to $\mathbb{Q}\langle Y\rangle$ by linearity.

Proposition 2 *For any* $u,v\in Y^*$, $\mathrm{A}_{u \sqcup\!\!\!\sqcup\!\!\!\!-\, v}(N;k) = \mathrm{A}_u(N;k)\mathrm{A}_v(N;k)$.

[(i)] Note that the usual *shuffle product* and *stuffle product* of $u=y_iu'$ and $v=y_jv'$ are defined respectively by

$$u \sqcup\!\!\!\sqcup\, v = y_i(u' \sqcup\!\!\!\sqcup\, v) + y_j(u \sqcup\!\!\!\sqcup\, v'), \quad \text{and} \quad u \uplus v = y_i(u' \uplus v) + y_j(u \uplus v') + y_{i+j}(u' \uplus v').$$

Proposition 3 *For $w = y_s w' \in Y^*$, we have* $\mathrm{A}_w(N) = \sum_{N \geq l \geq 1} \frac{\mathrm{A}_{w'}(l)}{l^s}$.

For $s_1 > 1$, $A_w(N)$ converges to a limit denoted by $\Theta(w)$ and the word w is said to be convergent. By Formula (5), $\Theta(w)$ can be expressed as a linear combination of MZV.

More generally, as a consequence of the nature of coefficients occuring in the asymptotic expansion of $\mathrm{H}_w(N)$ [4], we get the following result

Proposition 4 *Let $\mathcal{Z}$ be the $\mathbb{Q}$-algebra generated by MZV, i.e. $\{\zeta(w), w \in Y^* \setminus y_1 Y^*\}$ and let $\mathcal{Z}'$ be the $\mathbb{Q}[\gamma]$-algebra generated by $\mathcal{Z}$. Then there exist algorithmically computable coefficients $b_i \in \mathcal{Z}'$, $\kappa_i \in \mathbb{N}$ and $\eta_i \in \mathbb{Z}$ such that, for any $w \in Y^*$,*

$$\mathrm{A}_w(N) \sim \sum_{i=0}^{+\infty} b_i N^{\eta_i} \log^{\kappa_i}(N), \quad \textit{for} \quad N \to +\infty.$$

3 Asymptotic equivalent for $\mathbb{V}ar(K_{n,d})$

In this section, we focus on the asymptotic equivalent of $\mathbb{V}ar(K_{n,d})$

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

κ_d given by Formula (4). This one can be re-written, with our tools, in the following way:

$$\kappa_d = \frac{1}{(d-1)!} \sum_{t=1}^{d-2} \binom{d-1}{t} \sum_{l \geq 1} \frac{1}{l^2} A_{y_1^{t-1} \sqcup\!\sqcup\, y_1^{d-2-t}}(l)$$

Remind that we denote by $N_2(w)$ the number of occurences of the letter y_2 in w.

Theorem 1 $\kappa_d = \dfrac{1}{(d-1)!} \displaystyle\sum_{w \in \{y_1, y_2\}_{d-3}} (-1)^{N_2(w)} \binom{2\,(d-2-N_2(w))}{d-2-N_2(w)} \Theta(y_2 w).$

For example, for $d = 7$, we get

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

The last step consists in reducing into polyzêtas, and then use the reduction table. The following example make explicitly appear irreducible MZV of length > 1, *which was not observed before*,

$$\begin{aligned} \kappa_{11} \;=\; & \frac{209}{302400}\,\zeta(5)\,\zeta(2)\,\zeta(3) + \frac{2893}{6048000}\,\zeta(2)^2\zeta(3)^2 + \frac{3311}{460800}\,\zeta(3)\,\zeta(7) \\ - & \frac{517}{921600}\,\zeta(8,2) + \frac{39457}{9676800}\,\zeta(5)^2 + \frac{426341}{221760000}\,\zeta(2)^5 \end{aligned}$$

4 Next terms in the Asymptotic Expansion

Let us come back to Expression (1), that we can interpret, in terms of harmonic sums, this way

$$\mathbb{E}(K_{n,d}^2) = \mathrm{A}_{y_1^{d-1}}(n) + \sum_{1 \leq t \leq d-1} \binom{d}{t} \sum_{l=1}^{n-1} \frac{1}{l} \mathrm{A}_{y_1^{t-1}}(l) \mathrm{A}_{y_1^{d-t-1}}(l) \mathrm{A}_{y_1^{d-1}}(n; l+1),$$

Proposition 5 *For any integers $n \geq l$,* $\mathrm{A}_{y_1^d}(n; l) = \displaystyle\sum_{\substack{k_1 + \ldots + d k_d = d \\ k_1, \ldots, k_d > 0}} \frac{\mathrm{A}_1^{k_1}(n; l) \ldots \mathrm{A}_d^{k_d}(n; l)}{1^{k_1} k_1! \ldots d^{k_d} k_d!}$

- Thanks to Proposition 2, we are able to turn each polynomial (in harmonic sums) into a linear combination of harmonic sums.

- Finally, there are only sums over l of type $\frac{A_w(l)}{l}$ left, but by Proposition 3, they simply reduce to $A_{y_1 w}(n-1)$.

$$\begin{aligned}
\mathbb{V}ar(K_{n,3}) &= \mathbb{E}(K_{n,3}^2) - \mu_{n,3}^2 \\
&= \mathrm{A}_{1,1}(n) + 3\,{\mathrm{A}_1}^2(n)\mathrm{A}_{1,1}(n-1) - 12\mathrm{A}_1(n)\mathrm{A}_{1,1,1}(n-1) \\
&+ 6\mathrm{A}_1(n)\mathrm{A}_{1,2}(n-1) + 18\,\mathrm{A}_{1,1,1,1}(n-1) - 12\,\mathrm{A}_{1,1,2}(n-1) \\
&- 12\mathrm{A}_{1,2,1}(n-1) + 6\mathrm{A}_{1,3}(n-1) + 3\mathrm{A}_2(n)\mathrm{A}_{1,1}(n-1) - \mathrm{A}_{1,1}^2(n).
\end{aligned}$$

- Using algorithms described in [4], we can now compute the asymptotic expansion of $\mathbb{V}ar(K_{n,d})$.

Theorem 2 *There exist algorithmically computable coefficients $\alpha_i, \beta_{j,k} \in \mathcal{Z}'$ such that, for any dimension d and any order M,*

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

For example,

$$\begin{aligned}
\mathbb{V}ar(K_{n,3}) &= \left(\frac{1}{2} + \kappa_3\right) \ln^2(n) + (-10\zeta(3) + 2\zeta(2)\gamma + \gamma)\ln(n) + \frac{1}{2}\gamma^2 \\
&- 10\zeta(3)\gamma + \frac{83}{10}\zeta(2)^2 + \zeta(2)\gamma^2 + \frac{1}{2}\zeta(2) + \mathrm{o}(1)
\end{aligned}$$

References

[1] Z.D. Bai, C.C. Chao, H.K. Hwang, W.Q. Liang.– *On the variance of the number of maxima in random vectors and its applications*, Annals of Applied Probability, 8(3), pp. 886-895, 1998. MathReview: 99f:60019

[2] Z.D. Bai, L. Devroye, H.K. Hwang, T.H. Tsai.– *Maxima in hypercubes*, Random Structures and Algorithms, 27 (3), pp. 290-309, 2005.

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

[4] C. Costermans, J.Y. Enjalbert, Hoang Ngoc Minh.– *Algorithmic and combinatoric aspects of multiple harmonic sums*, in the proceedings of Aofa, Barcelone, 06-10 June, 2005.

[5] V. M. Ivanin.– *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, Akad. Nauk. Ukrain. SSR Inst. Kibernet., Kiev., pp. 90-100, 1976.