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

Christian Costermans and Hoang Ngoc Minh

Université Lille II, 1, Place Délia, 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 occurring belong to the \( \mathbb{Q} \)-algebra generated by Multiple Zeta Values, and by Euler’s constant \( \gamma \).

**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 \( x_i = (x_{i1}, \ldots, x_{id}) \) is said to be dominated by \( x_j = (x_{j1}, \ldots, x_{jd}) \) if \( x_{ik} < x_{jk} \) for all \( k \in \{1, \ldots, d\} \) and a point \( x_i \) is called a maximum of \( \Lambda \) if none of the other points dominates it. The number of maxima of \( \Lambda \) 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 \( \text{Var}(K_{n,d}) \) 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]:

\[
\mathbb{E}(K_{n,d}^2) = \mu_{n,d} + \sum_{1 \leq i \leq d-1} \binom{d}{i} \sum_{l=1}^{n-1} \frac{1}{i_1 \cdots i_{d-2} j_1 \cdots j_{d-1}},
\]

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

\[
1 \leq i_1 \leq \ldots \leq i_{l-1} \leq l, 1 \leq i_l \leq \ldots \leq i_{d-2} \leq l \quad \text{and} \quad l+1 \leq j_1 \leq \ldots \leq j_{d-1} \leq n.
\]

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

\[
\mu_{n,d} = \mathbb{E}(K_{n,d}) = \sum_{1 \leq i_1 \leq \ldots \leq i_{d-1} \leq n} \frac{1}{i_1 \cdots i_{d-1}}.
\]

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

\[
\text{Var}(K_{n,d}) \sim \left( \frac{1}{(d-1)!} + \kappa_d \right) \ln^{d-1}(n),
\]

with \( \kappa_d = \sum_{t=1}^{d-2} \frac{1}{t!(d-1-t)!} \sum_{l \geq 1} \frac{1}{l^2} \sum_{i_1 \cdots i_{d-2} j_1 \cdots j_{d-2-t}} \frac{1}{i_1 \cdots i_{d-1} j_1 \cdots j_{d-2-t}} \)

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

These two formulas give rise to harmonic sums \( A_s(N) \), closely related to \( H_s(N) \) defined for a multi-index \( s = (s_1, \ldots, s_t) \) by

\[
A_s(N) = \sum_{N \geq n_1 \geq \ldots \geq n_r \geq 1} \frac{1}{n_1^{s_1} \cdots n_r^{s_r}}, \quad H_s(N) = \sum_{N \geq n_1 \geq \ldots \geq n_r \geq 1} \frac{1}{n_1^{s_1} \cdots n_r^{s_r}}.
\]
We already studied in [1] the algebraic properties of \( H_\alpha(N) \); in particular, when \( s_1 > 1 \), \( H_\alpha(N) \) converges to the polyzet\( \alpha \) (or MZV) \( \zeta(s) = \sum_{n_1 > \ldots > n_r > 1} \frac{1}{n_1^{s_1} \cdots n_r^{s_r}} \). There exist explicit relations, given by Hoffman between the \( A_\alpha(N) \) and \( H_\alpha(N) \). Indeed, let \( \text{Comp}(n) \) 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. \( r \)) then \( J \circ I = (i_1 + \ldots + j_1, i_1 + j_1 + 1 + \ldots + j_2, \ldots, i_k + j_k + 1 + \ldots + j_p) \) is a composition of \( n \). One has \((l(J) \text{ being the number of parts of } J) : \)

\[
A_\alpha(N) = \sum_{J \in \text{Comp}(r)} H_{J \circ \alpha}(N) \quad \text{and} \quad H_\alpha(N) = \sum_{J \in \text{Comp}(r)} (-1)^{(l(J)-r)} A_{J \circ \alpha}(N).
\]

Here, the nature of Formulas (3) and (4) makes clear that it would be difficult to interpret both formulas in terms of \( H_\alpha(N) \). So, we prefer looking at the algebraic and combinatoric properties of \( A_\alpha(N) \), and deduce from these ones two main results, first the explicit value of \( \kappa_d \) in terms of Multiple Zeta Values, and then an algorithm to compute the asymptotic expansion of \( \Psi \text{ar}(K_{n,d}) \).

## 2 Combinatorial background

### 2.1 Combinatorics on words

To the multi-index \( s = (s_1, \ldots, s_r) \) 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 \( t \) is denoted by \( \ell(v) \), and its weight is defined as \( |v| = \sum_{i=1}^{t} s_i \). The number of occurrences 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 \( \epsilon \).

**Example 1** Let \( w = y_1 y_3 y_1^2 y_2 \), we have \( \ell(w) = 5, |w| = 9 \) and \( N_1(w) = 3 \).

**Definition 1** Let \( S \) be a subset of \( Y \), and \( \rho \) a positive integer, we define \( S_\rho \) as the set of words containing only letters in \( S \), and of weight equal to \( \rho \).

**Example 2** Let \( S = \{y_1, y_2\} \) and \( \rho = 4 \) then \( S_\rho = \{y_1^4, y_1 y_2 y_1, y_2 y_1 y_2, y_2^2\} \).

We shall henceforth identify the multi-index \( s \) 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-shuffle** \(^{10}\) of \( u = y_i u' \) and \( v = y_j v' \) is the polynomial recursively defined by

\[
\epsilon \shuffle u = u \quad \text{and} \quad u \shuffle v = y_k (u \shuffle v') + y_j (u \shuffle v') - y_{i+j} (u \shuffle v')
\]

For example, \( y_1 \shuffle y_2 = y_1 y_2 + y_2 y_1 - y_3 \).

**Proposition 1** \( y_1^r = \sum_{s_1 + \ldots + s_r = r} y_1^{s_1} \ldots y_k^{s_r} / r! s_1! \ldots s_r! \).

**Definition 2** Let \( w = y_{s_1} \ldots y_{s_r} \in Y^* \). For \( N \geq k \geq 1 \), the harmonic sum \( A_w(N; k) \) is defined as

\[
A_w(N; k) = \sum_{N \geq n_1 \geq \ldots n_r \geq k} \frac{1}{n_1^{s_1} \cdots n_r^{s_r}}.
\]

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

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

**Proposition 2** For any \( u, v \in Y^* \),

\[ u \shuffle v = y_i (u \shuffle v') + y_j (u \shuffle v') \quad \text{and} \quad u \shuffle v = y_i (u \shuffle v') + y_j (u \shuffle v') + y_{i+j} (u \shuffle v'). \]

Note that the usual **shuffle product** and **stuffle product** of \( u = y_i u' \) and \( v = y_j v' \) are defined respectively by

\[
u \shuffle v = y_i (u' \shuffle v') + y_j (u \shuffle v') \quad \text{and} \quad u \shuffle v = y_i (u' \shuffle v') + y_j (u \shuffle v') + y_{i+j} (u \shuffle v'). \]
Proposition 3 For \( w = y_{s1}w' \in Y^* \), we have \( A_w(N) = \sum_{N \geq l \geq 1} \frac{A_{w'}(l)}{l^n} \).

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 \( \text{MZV} \).

More generally, as a consequence of the nature of coefficients occurring in the asymptotic expansion of \( H_w(N) \) [4], we get the following result

Proposition 4 Let \( Z \) be the \( \mathbb{Q} \)-algebra generated by \( \text{MZV} \), i.e. \( \{ \zeta(w), w \in Y^* \setminus y_1 Y^* \} \) and let \( Z' \) be the \( \mathbb{Q}[\gamma] \)-algebra generated by \( Z \). Then there exist algorithmically computable coefficients \( b_t \in Z', \text{ } \kappa_t \in \mathbb{N} \) and \( \eta_t \in \mathbb{Z} \) such that, for any \( w \in Y^* \\
A_w(N) \sim \sum_{t=0}^{+\infty} b_t N^{\eta_t} \log^{\kappa_t}(N), \text{ for } N \to +\infty.

3 Asymptotic equivalent for \( Var(K_{n,d}) \)

In this section, we focus on the asymptotic equivalent of \( Var(K_{n,d}) \)

\[
Var(K_{n,d}) \sim \left( \frac{1}{(d-1)!} + \kappa_d \right) \ln^{d-1}(n),
\]

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

\[
\kappa_d = \frac{1}{(d-1)!} \sum_{l=1}^{d-2} \binom{d-1}{l} \sum_{t \geq 1} \frac{1}{t^2} A_{y_1^{l-1}} \omega_{y_t^{d-1}}(l)
\]

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

Theorem 1 \( \kappa_d = \frac{1}{(d-1)!} \sum_{w \in \{y_1, y_2\}_{d-3}} (-1)^{N_2(w)} \binom{2(d-2-N_2(w))}{d-2-N_2(w)} \Theta(y_2w). \)

For example, for \( d = 7 \), we get

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

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

\[
\kappa_{11} = \frac{209}{302400} \zeta(5) \zeta(2) \zeta(3) + \frac{2893}{604800} \zeta(2)^2 \zeta(3)^2 + \frac{3311}{460800} \zeta(3) \zeta(7)
- \frac{426341}{221760000} \zeta(5)^2 + \frac{921600}{9676800} \zeta(8) + \frac{39457}{9676800} \zeta(5) \zeta(2)
\]

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

\[
E(K_{n,d}^2) = A_{y_{d-1}}(n) + \sum_{1 \leq l \leq d-1} \binom{d}{l} \sum_{t=1}^{n-1} \frac{1}{t^2} A_{y_{d-1}}(l) A_{y_{d-1}}(l) A_{y_{d-1}}(n; l + 1),
\]

Proposition 5 For any integers \( n \geq l \), \( A_{y_t}(n; l) = \sum_{k_1 + \ldots + k_d = d} A_{y_t}^{k_1}(n; l) \ldots A_{y_t}^{k_d}(n; l) \frac{1}{k_1! \ldots k_d!} d^{k_1 + \ldots + k_d} k_1! \ldots k_d! \)

Thanks to Proposition 5, 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 \( A_{y,w}(n-1) \).

\[
\text{Var}(K_n,3) = \mathbb{E}(K_{n,3}^2) - \mu_{n,3}^2 \\
= A_{1,1}(n) + 3A_1^2(n)A_{1,1}(n-1) - 12A_1(n)A_{1,1,1}(n-1) \\
+ 6A_1(n)A_{1,2}(n-1) + 18A_{1,1,1,1}(n-1) - 12A_{1,1,2}(n-1) \\
- 12A_{1,2,1}(n-1) + 6A_{1,3}(n-1) + 3A_2(n)A_{1,1}(n-1) - A_{1,1,1}(n).
\]

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

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

\[
\text{Var}(K_{n,d}) = \sum_{i=0}^{d-1} \alpha_i \ln^i(n) + \sum_{j=1}^{M} \frac{1}{n^j} \sum_{k=0}^{2d-2} \beta_{j,k} \ln^k(n) + o\left( \frac{1}{n^M} \right).
\]

For example,

\[
\text{Var}(K_{n,3}) = \left( \frac{1}{2} + \kappa_3 \right) \ln^2(n) + (-10\zeta(3) + 2\zeta(2)\gamma + \gamma) \ln(n) + \frac{1}{2} \gamma^2 \\
- 10\zeta(3)\gamma + \frac{83}{10} \zeta(2)^2 + \zeta(2)\gamma^2 + \frac{1}{2} \zeta(2) + o(1)
\]

**References**


