Maximal empty boxes amidst random points

Adrian Dumitrescu* Minghui Jiang†

March 19, 2013

Abstract

We show that the expected number of maximal empty axis-parallel boxes amidst n random points in the unit hypercube $[0,1]^d$ in $\mathbb{R}^d$ is $(1 \pm o(1)) \frac{(2d-2)!}{(d-1)!} n \ln^{d-1} n$, if d is fixed. This estimate is relevant for analyzing the performance of exact algorithms for computing the largest empty axis-parallel box amidst n given points in an axis-parallel box R, especially the algorithms that proceed by examining all maximal empty boxes. Our method for bounding the expected number of maximal empty boxes also shows that the expected number of maximal empty orthants determined by n random points in $\mathbb{R}^d$ is $(1 \pm o(1)) \ln^{d-1} n$, if d is fixed. This estimate is related to the expected number of maximal (or minimal) points amidst random points, and has application to algorithms for colored orthogonal range counting.

Keywords: Maximal empty box, largest empty box, lower and upper bound, random points, expected value, alternating binomial sum, algorithm, data mining.

MSC2010 Codes: 52C45 Combinatorial complexity of geometric structures; 68U05 Computer graphics; computational geometry; 68W25 Approximation algorithms.

1 Introduction

Given an axis-parallel rectangle R in the plane containing n points, the problem of computing a maximum-area empty axis-parallel sub-rectangle contained in R is one of the oldest problems in computational geometry. For instance, this problem arises when a rectangular shaped facility is to be located within a similar region which has a number of forbidden areas, or in cutting out a rectangular piece from a similarly shaped metal sheet with some defective spots to be avoided [23]. In higher dimensions, finding an empty axis-parallel box with the maximum volume has applications in data mining, in finding large gaps in a multi-dimensional data set [14].

Since the volume ratio of any box inside another box is invariant under scaling, we can assume without loss of generality that the enclosing box is a hypercube. Given a set S of n points in the unit hypercube $U_d = [0,1]^d$, $d \geq 2$, an *empty box* is an open axis-parallel hyperrectangle contained in U_d and containing no points in S (i.e., $B \subset U_d$ and $B \cap S = \emptyset$), and Maximum Empty Box is the problem of finding an empty box with the maximum volume. Some planar examples are shown in Fig. 1.

Several algorithms have been proposed over time [1, 2, 3, 9, 11, 22, 23, 24] for the Maximum Empty Box problem in the plane. The fastest one, due to Aggarwal and Suri [1], runs in $O(n \log^2 n)$

*Department of Computer Science, University of Wisconsin–Milwaukee, USA. Email: dumitres@uwm.edu. Supported in part by NSF grant DMS-1001667.

†Department of Computer Science, Utah State University, Logan, USA. Email: mjiang@cc.usu.edu.

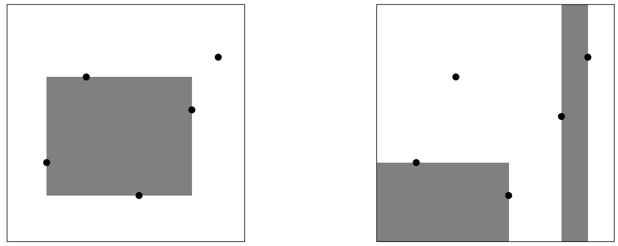

Figure 1: A maximal empty rectangle supported by one point on each side (left), and two maximal empty rectangles supported by both points and sides of $[0,1]^2$ (right).

time and $O(n)$ space. A lower bound of $\Omega(n \log n)$ in the algebraic decision tree model for the planar problem has been shown by McKenna, O'Rourke, and Suri [22]. Augustine et al. [4] and Kaplan, Mozes, Nussbaum, and Sharir [16] studied a related problem, that of finding the largest empty rectangle containing a query point.

For the Maximum Empty Box problem in higher dimensions, Backer and Keil [5, 6] proved that the problem is NP-hard when the dimension d is part of the input, and Giannopoulos, Knauer, Wahlström, and Werner [15] further showed that the problem is W[1]-hard with the dimension d as the parameter. From the positive direction, Dumitrescu and Jiang [13] presented an approximation algorithm that finds an empty box whose volume is at least $1-\varepsilon$ of the optimal in $O\big((\frac{8ed}{\varepsilon^2})^d \cdot n \cdot \log^d n\big)$ time.

Note that an empty box of maximum volume must be *maximal* with respect to inclusion. In high dimensions, i.e., for $d \geq 3$, the only approach currently known for computing the largest empty box exactly is by examining *all* candidates, i.e., all maximal empty boxes. Kaplan, Rubin, Sharir, and Verbin [17] presented an output-sensitive algorithm running in $O((n+k)\log^{d-1} n)$ time, where k is the number of maximal empty boxes. Backer and Keil [5, 6] also reported an output-sensitive algorithm running in $O(k \log^{d-2} n)$ time; in particular, the worst-case running time of their algorithm for $d = 3$ is $O(n^3 \log n)$. Previously, Datta and Soundaralakshmi [12] had reported an $O(n^3)$ time exact algorithm for the $d = 3$ case, but their analysis for the running time seems incomplete. Specifically, the $O(n^3)$ running time depends on an $O(n^3)$ upper bound on the maximum number of maximal empty boxes, but they only gave an $\Omega(n^3)$ lower bound.

Naamad, Lee, and Hsu [23] showed that in the plane, the number of maximal empty rectangles is $O(n^2)$, and that this bound is tight. It was conjectured by Datta and Soundaralakshmi [12] that the maximum number of maximal empty boxes is $O(n^d)$ for each (fixed) d. The conjecture was recently confirmed by Kaplan, Rubin, Sharir, and Verbin [17], who obtain the first tight $\Theta(n^d)$ bound on this number[1].

[1]We note that Kaplan et al. [17] seemed to be unaware of the conjecture and the earlier papers [12, 23]. Their study of maximal empty boxes is motivated by a related problem called Colored Orthogonal Range Counting, which has important database applications. In the context of this related problem, they transform each input point in $\mathbb{R}^d$ into a point in $\mathbb{R}^{2d}$ by splitting the coordinate x along the i-th axis into two coordinates x and $-x$ in the $(2i-1)$-th and $(2i)$-th axes, such that a maximal empty box in $\mathbb{R}^d$ becomes a maximal empty orthant in $\mathbb{R}^{2d}$. Then they provide an upper bound of $O(n^{\lfloor d/2 \rfloor})$ on the maximum number of maximal empty orthants in $\mathbb{R}^d$, which, by the transformation, implies an upper bound of $O(n^d)$ on the maximum number of maximal empty boxes in $\mathbb{R}^d$. They also provide matching lower bounds, $\Omega(n^{\lfloor d/2 \rfloor})$ for orthants and $\Omega(n^d)$ for boxes. Shortly after this, following the earlier work [12, 23] on the Maximum Empty Box problem, Dumitrescu and Jiang [13] obtained the lower bound $\Omega(n^d)$ independently, unaware of the recent work by Kaplan et al. [17]. Around the same time, Backer and Keil [5] also obtained the tight $\Theta(n^d)$ bound. In retrospect, we observe that the three constructions for the $\Omega(n^d)$ lower bound in [17, 13, 5] are based on essentially the same simple idea. For the $O(n^d)$ upper bound, Kaplan et al. [17] use an elegant shifting technique and cite Boissonnat, Sharir, Tagansky, and Yvinec [8] for a similar analysis, and Backer and Keil [5] use essentially the same technique (which they call deflation-inflation) in their proof and cite (the

Hence the maximum number of maximal empty boxes is $\Theta(n^d)$ for each fixed d. This means that any algorithm that computes a maximum-volume empty box by enumerating all maximal empty boxes is bound to be inefficient in the worst case. Indeed, by a standard result in parameterized complexity theory [21, Section 6.3], the aforementioned W[1]-hardness result of Giannopoulos, Knauer, Wahlström, and Werner [15, Theorem 3] implies that the existence of an exact algorithm running in $n^{o(d)}$ time is unlikely, i.e., unless the so-called Exponential Time Hypothesis (ETH) fails, i.e., unless 3-SAT can be solved in $2^{o(n)}$ time. However, as it is the case with the output-sensitive algorithm of Kaplan, Rubin, Sharir, and Verbin [17], which runs in $O((n+k)\log^{d-1} n)$ time, and the algorithm of Backer and Keil [6], which was reported to run in $O(k \log^{d-2} n)$ time, such algorithms would be much faster when there are only a few maximal empty boxes (here k denotes this number).

Let us consider the exact algorithm performing in a typical case, for instance with points randomly and uniformly distributed in a box. Naamad, Lee, and Hsu [23] obtained an $O(n \log n)$ upper bound on the expected number of maximal empty boxes amidst n random points in the plane. Datta and Soundaralakshmi claimed in [12, Lemma 2] that in 3-space the expected number of maximal empty boxes is of the order $\Theta(n \log^2 n)$. Recently, Backer and Keil [6, p. 20] acknowledged this estimate by citing it as a generalized bound of $\Theta(n \log^{d-1} n)$ for all $d \geq 3$. This bound, if true, would make their exact algorithm, which enumerates all maximal empty boxes in order to find a maximum-volume empty box, quite attractive in a typical case.

Here we show that the proof given by Datta and Soundaralakshmi [12] in support of their claim does not stand. We then provide a proof of the generalized bound of $\Theta(n \log^{d-1} n)$ for all $d \geq 1$. Our estimates are relevant for analyzing the performance of any exact algorithm for computing the largest empty axis-parallel box amidst n points in a given axis-parallel box R, that proceeds by examining all maximal empty boxes. The current most efficient algorithms for this task [17, 5, 12], are thus expected to be much faster in instances where the points are close to being randomly distributed.

Our results. Let $d \geq 1$ and $n \geq 2d$. Let $X_i = (x_{i,1}, \ldots, x_{i,d})$, $1 \leq i \leq n$, be n random points in the unit hypercube $U_d = [0,1]^d$ in $\mathbb{R}^d$, with independent coordinates sampled uniformly from the interval $[0,1]$. Note that with probability 1, the n points have distinct coordinates in the open interval $(0,1)$ along each axis. Without loss of generality, we will assume this condition in our analysis. We define $G(n,a,b)$, where $a \geq 0$, $b \geq 0$, and $a+b \leq d$, as the expected number of maximal empty boxes supported by one point in each of $2a+b$ faces: the two opposite faces orthogonal to each of the first a coordinate axes, and the upper face orthogonal to each of the next b coordinate axes. Then the expected number $F(n,d)$ of maximal empty boxes in U_d that are supported by one point in each of the $2d$ faces is

$$F(n,d) = G(n,d,0), \tag{1}$$

and the expected number $E(n,d)$ of all maximal empty boxes in U_d is

$$E(n,d) = \sum_{a=0}^{d} \sum_{b=0}^{d-a} \binom{d}{a} \binom{d-a}{b} 2^b \, G(n,a,b). \tag{2}$$

For two functions f and g, we write $f(n) \sim g(n)$ if $f(n) = g(n)(1 \pm o(1))$ as $n \to \infty$. The relation $\sim$ is clearly symmetric. Moreover, it is transitive, in the sense that for any fixed number r of functions $f_1, \ldots, f_r$, if $f_i(n) \sim f_{i+1}(n)$ for all $i = 1, \ldots, r-1$, then $f_1(n) \sim f_r(n)$.

conference version of) [17] for inspiration. This technique was subsequently used by Dumitrescu and Jiang [13] in sharpening the constant factor in the $O(n^d)$ upper bound. Unaware of the contribution of Kaplan et al. [17] to the Maximum Empty Box problem, Dumitrescu and Jiang [13] cite Backer and Keil [5] for this technique.

Our main result is a tight asymptotic estimate for $G(n, a, b)$:

Theorem 1. *For any fixed $d \geq 1$ and any fixed a and b with $0 < a + b \leq d$, $G(n, a, b) \sim \frac{(2a+b-2)!}{(a-1)!} n \ln^{a-1} n = \Theta(n \log^{a-1} n)$ if $a \geq 2$, $G(n, a, b) \sim b!\, n = \Theta(n)$ if $a = 1$, and $G(n, a, b) \sim \ln^{b-1} n = \Theta(\log^{b-1} n)$ if $a = 0$.*

By (1) and (2), we immediately have the following corollary:

Corollary 1. *For any fixed $d \geq 1$, $E(n, d) \sim F(n, d) \sim \frac{(2d-2)!}{(d-1)!} n \ln^{d-1} n = \Theta(n \log^{d-1} n)$.*

Remark. After the completion of our work, we learned that Kaplan, Rubin, Sharir, and Verbin [17] had obtained, prior to us, an upper bound of $O(n \log^{d-1} n)$ on $E(n, d)$; although many details in their proof are not spelled out and the arguments in [17] are quite sketchy, the proof seems to lead to an $O(n \log^{d-1} n)$ bound. The authors of [17] were apparently unaware of previous work on this topic in the literature, such as [12, 23]. (The focus of their paper is range counting.) Obviously, if the results claimed by Datta and Soundaralakshmi [12] were correct, then the upper bound of $O(n \log^{d-1} n)$ obtained by Kaplan et al. [17] would *not* be a new result. Here we settle this discrepancy in the literature by acknowledging that the first correct upper bound $E(n, d) = O(n \log^{d-1} n)$ was obtained by Kaplan et al. [17]. It is also worth mentioning that the method used by Kaplan et al. [17] in deriving the upper bound is completely different than ours. We don't know if their method could be adapted to obtain the sharper estimate (with the dependence on d) that we obtain.

Our estimates significantly sharpen previous estimates obtained by Kaplan et al. [17]: their upper bound $O(n \log^{d-1} n)$ had no matching lower bound; moreover our proof provides full details and our estimate is much more precise. It is interesting to note that in our general bound $G(n, a, b) \sim \frac{(2a+b-2)!}{(a-1)!} n \ln^{a-1} n = \Theta(n \log^{a-1} n)$ for $a \geq 2$, the number b of semi-bounded dimensions does not contribute logarithmic factors. Also note that the maximal empty boxes considered in the bound $G(n, a, b) \sim \ln^{b-1} n = \Theta(\log^{b-1} n)$ for $a = 0$ are in fact maximal empty orthants (for example the maximal empty rectangle at the lower-left corner in Fig. 1 right) which play a pivotal role in the algorithm of Kaplan et al. [17] for efficient colored orthogonal range counting.

Kaplan et al. [17] remarked that it would be interesting to explore the connection between the expected number of maximal empty boxes and the expected number of maximal points. We show two connections: (i) the lower bound of $\Omega(n \log^{d-1} n)$ on the expected number of direct dominance pairs [18], which is closely related to the expected number of maximal points [7], yields a lower bound of $\Omega(n \log^{d-1} n)$ on the expected number of maximal empty boxes, and (ii) the upper bound of $O(\log^{d-1} n)$ on the expected number of maximal empty orthants (the case $a = 0$ in Theorem 1) yields an upper bound of $O(\log^{d-1} n)$ on the expected number of maximal (or minimal) points [7].

The rest of the paper is organized as follows. In Section 2, we explore the connections between the results for maximal empty boxes and maximal empty orthants and earlier results [7, 18] on maximal points and direct dominance pairs, and show that the proof given by Datta and Soundaralakshmi in support of their claim [12, Lemma 2] does not stand. In Section 3, we present our proof of Theorem 1 and Corollary 1.

2 Connections between maximal empty boxes, maximal empty orthants, maximal points, and direct dominance pairs

For two points p and q in $\mathbb{R}^d$, we say that p *dominates* q if along each of the d axes the coordinate of p is larger than or equal to the coordinate of q. For a set S of points in $\mathbb{R}^d$, and a pair of points

$p, q \in S$, we say that p *directly dominates* q if (i) p dominates q, and (ii) there is no other point r in S such that p dominates r and r dominates q; then (p, q) is called a *direct dominance pair*. A point is *maximal* (respectively, *minimal*) if it is not dominated by (respectively, does not dominate) any other point in the set.

Bentley, Kung, Schkolnick, and Thompson [7] proved that the expected number of maximal points in a set of n random points in $\mathbb{R}^d$ is $\Theta(\log^{d-1} n)$ for any fixed $d \geq 2$; by symmetry, the same bound also applies to the expected number of minimal points. Klein [18] proved that the expected number of direct dominance pairs in a set of n random points in $\mathbb{R}^d$ is $\Theta(n \log^{d-1} n)$ for any fixed $d \geq 2$. In fact, one can check that if the expected number of maximal points is $O(\log^{d-1} n)$, then the expected number of direct dominance pairs is $O(n \log^{d-1} n)$. This is because the points in S that are directly dominated by any point $p \in S$ are simply the maximal points among the subset $S_p \subseteq S$ of points that are dominated by p.

By symmetry, the concept of direct dominance can be generalized to include all 2^d different types, one for each combination of preferred directions along the d axes. For example, in $\mathbb{R}^2$, each point may directly dominate other points in each of its four quadrants. The expected number of such generalized direct dominance pairs is clearly still $\Theta(n \log^{d-1} n)$ for any fixed $d \geq 2$.

Datta and Soundaralakshmi [12, Lemma 2] observed that the expected number of maximal empty boxes amidst n random points in a hypercube in $\mathbb{R}^3$ is related to the number of direct dominance pairs determined by these points. Note that in $\mathbb{R}^3$, a maximal empty box may be supported by one point on each of its six faces. They argued that "once we fix the top support as a point p_i, the other five supports should be directly dominated by p_i in its four quadrants". Then, citing the previous known results [7, 18] on the expected number of direct dominance pairs, they jumped to the conclusion that the expected number of maximal empty boxes is of the same order, i.e., $\Theta(n \log^2 n)$ in $\mathbb{R}^3$. Recently, Backer and Keil [6, p. 20] acknowledged this estimate from [12] by citing it as a generalized bound of $\Theta(n \log^{d-1} n)$ in $\mathbb{R}^d$, $d \geq 3$, without a proof.

Here we show that the argument of Datta and Soundaralakshmi [12, Lemma 2] does not stand. In relating the expected number of maximal empty boxes to the expected number of direct dominance pairs, they correctly observed that each maximal empty box is associated with only a constant number (depending on d only) of direct dominance pairs, but they failed to provide any argument in the opposite direction to show that each direct dominance pair is also associated with a constant number of maximal empty boxes. For a bipartite graph with vertex partition $V = A \cup B$, the condition that every vertex in A has constant degree, without the symmetric condition that every vertex in B also has constant degree, does not imply that the number of vertices in A is of same order as the number of vertices in B.

We also note that the argument of Datta and Soundaralakshmi does not use any special property of the random distribution of the n points. Their observation that each maximal empty box is associated with a constant number of direct dominance pairs continues to hold even for non-random point sets. As long as the points have distinct coordinates along each axis, each maximal empty box in $\mathbb{R}^3$ is supported by at most six points, one in each face. Note that the number of direct dominance pairs in any set of n points is at most the total number of pairs, i.e., $\binom{n}{2} = O(n^2)$. If their proof were sound, then following their argument, they could go further and claim that the number of maximal empty boxes amidst any n points in $\mathbb{R}^3$ is at most $O(n^2)$. But this claim is clearly false since for any fixed $d \geq 2$, there exist n-element point sets in $\mathbb{R}^d$ (or $[0, 1]^d$) with at least $\Omega(n^d)$ maximal empty boxes amidst them [5, 13, 17].

It is not difficult, however, to obtain a lower bound on the expected number of maximal empty boxes using the lower bound on on the expected number of direct dominance pairs. Consider the set of direct dominance pairs determined by n random points in $\mathbb{R}^d$ with distinct coordinates along each axis. Then each direct dominance pair (p, q) determines an empty box B with the two points

p and q at the two opposite vertices of a main diagonal. This empty box B can be expanded, in both directions along each of the $d-1$ axes except the first axis, to a maximal empty box B' with the two points p and q supporting its two faces orthogonal to the first axis. Then each direct dominance pair is associated with a distinct maximal empty box. Thus the number of maximal empty boxes is at least the number of direct dominance pairs. Since the expected number of direct dominance pairs is $\Theta(n \log^{d-1} n)$, it follows that the expected number of maximal empty boxes is $\Omega(n \log^{d-1} n)$.

Recall that the maximal empty boxes considered in the bound $G(n, 0, b) \sim \ln^{b-1} n = \Theta(\log^{b-1} n)$ in Theorem 1 are maximal empty orthants. Maximal empty orthants are closely related to minimal points. For a planar point set, the number of maximal empty orthants is exactly the number of minimal points plus one, including one orthant for each pair of consecutive points, and two orthants that degenerate to half-planes. Intuitively, these two numbers are closely related for random points in higher dimensions too. Bentley, Kung, Schkolnick, and Thompson [7] showed that the expected number of minimal points among n random points in $\mathbb{R}^d$ is $\Theta(\log^{d-1} n)$. By the case $a = 0$ in Theorem 1, the expected number of maximal empty orthants is $\Theta(\log^{d-1} n)$ too, since $G(n, 0, d)$ is the dominating term among all $G(n, 0, b)$. Thus the two numbers are indeed of the same order of magnitude for any fixed d.

Moreover, our proof of Theorem 1 gives an alternative way for deriving the $O(\log^{d-1} n)$ upper bound of [7] on the expected number of minimal points, as follows. Observe that each minimal point corresponds to an empty orthant with the point as the apex; for $d \geq 2$, this orthant is not maximal and can be expanded into one or more maximal empty orthants. For points in general position (in the sense that no two points have the same coordinate along any axis), each maximal empty orthant is supported by at most d points, and hence contains at most d such empty orthants. Hence the number of minimal points is at most d times the number of maximal empty orthants for points in general position. In particular, for random points, the expected number of minimal points is at most d times the expected number of maximal empty orthants, thus $O(\log^{d-1} n)$.

3 The proof

In this section we prove Theorem 1 and Corollary 1. Recall that $X_i = (x_{i,1}, \ldots, x_{i,d})$, $1 \leq i \leq n$, are n random points in the unit hypercube $[0, 1]^d$, and that $G(n, a, b)$ is the expected number of maximal empty boxes supported by one point in each of $2a + b$ faces: the two opposite faces orthogonal to each of the first a coordinate axes, and the upper face orthogonal to each of the next b coordinate axes.

3.1 Event $A(a, b)$

For any pair of non-negative integers a and b such that $a + b \leq d$, let $A(a, b)$ be the event that the n random points $X_1, \ldots, X_n$ satisfy the following two conditions:

C1. Along the first $a + b$ axes, and among the first a pairs of points $X_1, X_2, \ldots, X_{2a-1}, X_{2a}$ and the next b points $X_{2a+1}, \ldots, X_{2a+b}$,

(a) for $1 \leq j \leq a$, the two points X_{2j-1} and X_{2j} have the smallest and the largest coordinates, respectively, along the j-th axis:

$$x_{2j-1,j} = \min\{x_{i,j} \mid 1 \leq i \leq 2a + b\},$$
$$x_{2j,j} = \max\{x_{i,j} \mid 1 \leq i \leq 2a + b\}.$$

(b) for $1 \le j \le b$, the point X_{2a+j} has the largest coordinate along the $(a+j)$-th axis:

$$x_{2a+j,a+j} = \max\{x_{i,a+j} \mid 1 \le i \le 2a+b\}.$$

C2. The box determined by the $2a+b$ points $X_1, \dots, X_{2a+b}$ is empty of the other $n-2a-b$ points $X_{2a+b+1}, \dots, X_n$:

$$\langle X_1, \dots, X_{2a+b} \rangle \cap \{X_{2a+b+1}, \dots, X_n\} = \emptyset,$$

where $\langle X_1, \dots, X_{2a+b} \rangle$ denotes the box

$$(x_{1,1}, x_{2,1}) \times \cdots \times (x_{2a-1,a}, x_{2a,a}) \times (0, x_{2a+1,a+1}) \times \cdots \times (0, x_{2a+b,a+b}) \times (0,1)^{d-a-b}.$$

Then

$$G(n,a,b) = \binom{n}{2a+b}(2a+b)! \cdot \mathbb{P}(A(a,b)), \tag{3}$$

where the factor $\binom{n}{2a+b}(2a+b)!$ accounts for all possible permutations of the first $2a+b$ points from the n random points.

The evaluation of $G(n,a,b)$ thus reduces to the calculation of $\mathbb{P}(A(a,b))$. Our use of binomial expansion and conditional probability in the following is inspired by the technique of Klein [18] for bounding the number of directed dominance pairs among n points uniformly and randomly selected in $[0,1]^d$.

3.2 Binomial expansion

Recall that the volume of U_d is exactly 1. Thus with any fixed $X_1, \dots, X_{2a+b}$ satisfying condition C1, we have

$$\mathbb{P}(X_i \notin \langle X_1, \dots, X_{2a+b} \rangle) = 1 - \mathrm{vol}\langle X_1, \dots, X_{2a+b} \rangle$$

for each X_i with $2a+b+1 \le i \le n$. By the independence of these $n-2a-b$ random points, the probability that condition C2 is also satisfied is then

$$\mathbb{P}(\langle X_1, \dots, X_{2a+b} \rangle \cap \{X_{2a+b+1}, \dots, X_n\} = \emptyset) = \big(1 - \mathrm{vol}\langle X_1, \dots, X_{2a+b} \rangle\big)^{n-2a-b}.$$

Thus we have

$$\begin{aligned}
&\mathbb{P}(A(a,b)) \\
&= \underset{X_1,\dots,X_{2a+b} \text{ satisfying C1}}{\int \cdots \int} \big(1 - \mathrm{vol}\langle X_1, \dots, X_{2a+b} \rangle\big)^{n-2a-b} \, \mathrm{d}X_1 \dots \mathrm{d}X_{2a+b} \\
&= \sum_{m=0}^{n-2a-b} \binom{n-2a-b}{m} (-1)^m \underset{X_1,\dots,X_{2a+b} \text{ satisfying C1}}{\int \cdots \int} \big(\mathrm{vol}\langle X_1, \dots, X_{2a+b} \rangle\big)^m \, \mathrm{d}X_1 \dots \mathrm{d}X_{2a+b}.
\end{aligned} \tag{4}$$

Observe that after the binomial expansion, the integrand $\big(\mathrm{vol}\langle X_1, \dots, X_{2a+b} \rangle\big)^m$ in (4) is equal to the probability that the m points $X_{2a+b+1}, \dots, X_{2a+b+m}$ are all included in the box $\langle X_1, \dots, X_{2a+b} \rangle$. To make this integral easier to evaluate, we first transform the integrand into a simpler form by interpreting this probability in a different but equivalent way.

3.3 Conditional probability

Consider only the first $2a+b+m$ points $X_1, \ldots, X_{2a+b+m}$. Let $B_m^>$ be the event that

- for $1 \le j \le a$, X_{2j-1} has the smallest coordinate along the j-th axis among the first $2a+b+m$ points,

and let $B_m^<$ be the event that

- for $1 \le j \le a$, X_{2j} has the largest coordinate along the j-th axis among the first $2a+b+m$ points, and
- for $1 \le j \le b$, X_{2a+j} has the largest coordinate along the $(a+j)$-th axis among the first $2a+b+m$ points.

Then the probability that the m points $X_{2a+b+1}, \ldots, X_{2a+b+m}$ are all included in the box $\langle X_1, \ldots, X_{2a+b}\rangle$ is exactly $\mathbb{P}(B_m^> \cap B_m^<)$. Thus we have

$$\big(\mathrm{vol}\langle X_1, \ldots, X_{2a+b}\rangle\big)^m = \mathbb{P}(B_m^> \cap B_m^<) = \mathbb{P}(B_m^<)\,\mathbb{P}(B_m^> \mid B_m^<).$$

Now suppose that the $a+b$ maximum coordinates $x_{2,1}, \ldots, x_{2a,a}, x_{2a+1,a+1}, \ldots, x_{2a+b,a+b}$ in the event $B_m^<$ are fixed. Then for each point X_i other than X_{2j}, $1 \le j \le a$, the probability that $x_{i,j} < x_{2j,j}$ is exactly $x_{2j,j}$, and for each point X_i other than X_{2a+j}, $1 \le j \le b$, the probability that $x_{i,a+j} < x_{2a+j,a+j}$ is exactly $x_{2a+j,a+j}$. This implies that

$$\mathbb{P}(B_m^<) = (x_{2,1} \cdots x_{2a,a} x_{2a+1,a+1} \cdots x_{2a+b,a+b})^{m+2a+b-1}.$$

In addition, since along each of the first a axes, each of the $2a+b+m$ points except the point with the maximum coordinate has the same chance of being the point with the minimum coordinate, we have

$$\mathbb{P}(B_m^> \mid B_m^<) = (m+2a+b-1)^{-a}.$$

Thus the integral in (4) yields

$$\begin{aligned}
&\int\!\!\cdots\!\!\int_{X_1,\ldots,X_{2a+b}\text{ satisfying C1}} \big(\mathrm{vol}\langle X_1, \ldots, X_{2a+b}\rangle\big)^m \,\mathrm{d}X_1 \ldots \mathrm{d}X_{2a+b} \\
&= \int\!\!\cdots\!\!\int_{x_{2,1},\ldots,x_{2a,a},x_{2a+1,a+1},\ldots,x_{2a+b,a+b}} \mathbb{P}(B_m^<)\,\mathbb{P}(B_m^> \mid B_m^<)\,\mathrm{d}x_{2,1} \ldots \mathrm{d}x_{2a,a}\mathrm{d}x_{2a+1,a+1} \ldots \mathrm{d}x_{2a+b,a+b} \\
&= (m+2a+b-1)^{-a} \left(\int_0^1 x^{m+2a+b-1}\,\mathrm{d}x\right)^{a+b} \\
&= (m+2a+b-1)^{-a}(m+2a+b)^{-(a+b)},
\end{aligned}$$

and subsequently (4) yields

$$\mathbb{P}(A(a,b)) = \sum_{m=0}^{n-2a-b} \binom{n-2a-b}{m} (-1)^m (m+2a+b-1)^{-a}(m+2a+b)^{-(a+b)}.$$

Substituting $m+2a+b$ by k, we have

$$\mathbb{P}(A(a,b)) = (-1)^b \sum_{k=2a+b}^{n} \binom{n-2a-b}{k-2a-b} (-1)^k (k-1)^{-a} k^{-(a+b)}.$$

It then follows by (3) that

$$
\begin{aligned}
G(n,a,b) &= \binom{n}{2a+b}(2a+b)!\cdot \mathbb{P}(A(a,b)) \\
&= (-1)^b \sum_{k=2a+b}^{n} \frac{n!}{(n-k)!(k-2a-b)!}(-1)^k (k-1)^{-a} k^{-(a+b)} \\
&= (-1)^b \sum_{k=2a+b}^{n} \binom{n}{k}(-1)^k \frac{k!}{(k-2a-b)!(k-1)^a k^{a+b}}. \qquad (5)
\end{aligned}
$$

3.4 Partial fractional decompositions

To evaluate $G(n,a,b)$, our strategy is to decompose the complicated alternating binomial sum in (5) into a linear combination of simpler alternating binomial sums. Specifically, we need a partial fractional decomposition of

$$
f(k,a,b) := \frac{k!}{(k-2a-b)!(k-1)^a k^{a+b}}. \qquad (6)
$$

We briefly review some standard terms. A rational fraction (i.e., the quotient of two polynomials with real coefficients) is *proper* if the degree of the numerator is less than the degree of the denominator. A proper rational fraction $\Phi(x)/\Psi(x)$ is called a *partial fraction* if its denominator $\Psi(x)$ is a power of an irreducible polynomial $P(x)$, that is, $\Psi(x) = P^h(x)$, $h \ge 1$. The following fundamental theorem holds [19] [20, Ch. 5]:

> Any proper rational fraction $\Phi(x)/\Psi(x)$ has a unique decomposition into a sum of partial fractions. Moreover, if all roots of $\Psi(x)$ are real, all numerators of the partial fractions in the decomposition are constants.

For $a \ge 2$ and $b \ge 0$, $f(k,a,b)$ simplifies to

$$
f(k,a,b) = \frac{(k-2)\cdots(k-2a-b+1)}{(k-1)^{a-1}k^{a+b-1}}.
$$

Since

$$
1 - f(k,a,b) = \frac{(k-1)^{a-1}k^{a+b-1} - (k-2)\cdots(k-2a-b+1)}{(k-1)^{a-1}k^{a+b-1}}
$$

is clearly a proper rational fraction, it follows by the fundamental theorem that $f(k,a,b)$ has the following decomposition

$$
\frac{(k-2)\cdots(k-2a-b+1)}{(k-1)^{a-1}k^{a+b-1}} = 1 + \sum_{i=1}^{a-1} \frac{p_i}{(k-1)^i} + \sum_{j=1}^{a+b-1} \frac{q_j}{k^j}, \qquad (7)
$$

where the coefficients p_i and q_j depend on a and b but not on k.

For completeness, we include a self-contained proof of the decomposition (7) as follows. Observe that the numerator $(k-2)\cdots(k-2a-b+1)$ consists of exactly $2a+b-2$ factors $k-i$, $2 \le i \le 2a+b-1$. Partition the set of $2a+b-2$ factors arbitrarily into two subsets P_{k-1} and P_k of sizes $a-1$ and $a+b-1$, respectively. Consider each factor $k-i$ in P_{k-1} as the sum of two terms $k-1$ and $-i+1$, and consider each factor $k-i$ in P_k as the sum of two terms k and $-i$. Then the chain of multiplications expands the numerator into a sum of 2^{2a+b-2} terms, which leads to a decomposition in the following form after grouping terms of the same denominator:

$$\frac{(k-2)\cdots(k-2a-b+1)}{(k-1)^{a-1}k^{a+b-1}} = \sum_{i=0}^{a-1}\sum_{j=0}^{a+b-1} \frac{c_{i,j}}{(k-1)^i k^j}.$$

The coefficients p_i and q_j depend on a and b but not on k; they are constants when a and b are constants. It is clear that $c_{0,0} = 1$. Since $\frac{1}{(k-1)k} = \frac{1}{k-1} - \frac{1}{k}$, each term on the right side of the above decomposition with both $i \geq 1$ and $j \geq 1$ can be split into two terms:

$$\frac{c_{i,j}}{(k-1)^i k^j} = \frac{c_{i,j}}{(k-1)^i k^{j-1}} - \frac{c_{i,j}}{(k-1)^{i-1} k^j}.$$

Thus by repeated splitting, the decomposition can be brought into the form of (7).

For $a = 1$ and $b \geq 1$, $f(k, a, b)$ simplifies to

$$f(k, 1, b) = \frac{(k-2)\cdots(k-b-1)}{k^b}$$

and its decomposition simplifies to

$$\frac{(k-2)\cdots(k-b-1)}{k^b} = 1 + \sum_{j=1}^{b} \frac{q_j}{k^j}. \tag{8}$$

Note that the decomposition has no fractions with powers of $k-1$ in the denominator.

Similarly, for $a = 0$ and $b \geq 2$, $f(k, a, b)$ simplifies to

$$f(k, 0, b) = \frac{(k-1)\cdots(k-b+1)}{k^{b-1}}$$

and its decomposition simplifies to

$$\frac{(k-1)\cdots(k-b+1)}{k^{b-1}} = 1 + \sum_{j=1}^{b-1} \frac{q_j}{k^j}. \tag{9}$$

3.5 Alternating binomial sums

For any $n \geq 0$, the following identity is well-known:

$$\sum_{k=0}^{n} \binom{n}{k} (-1)^k = (1-1)^n = 0. \tag{10}$$

We next derive a few other alternating binomial sums that we need. For any $d \geq 2$, define

$$R(n, d) = \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{(k+1)^{d-1}}, \qquad \text{for } n \geq 0, \tag{11}$$

$$S(n, d) = \sum_{k=1}^{n} \binom{n}{k} \frac{(-1)^k}{k^{d-1}}, \qquad \text{for } n \geq 1, \tag{12}$$

$$T(n, d) = \sum_{k=2}^{n} \binom{n}{k} \frac{(-1)^k}{(k-1)^{d-1}}, \qquad \text{for } n \geq 2. \tag{13}$$

The following identity is also known; see e.g., [10, Exercise 27, p. 105]. We include here a short proof for completeness (alternative proofs using generating functions and integration, or using induction are also possible).

Lemma 1. *For any $n \geq 0$, $R(n,2) = \frac{1}{n+1}$.*

Proof. It suffices to show that $1-(n+1)R(n,2)=0$:

$$\begin{aligned}
1-(n+1)R(n,2) &= 1-\sum_{k=0}^{n}\binom{n}{k}\frac{n+1}{k+1}(-1)^k \\
&= 1+\sum_{k=0}^{n}\binom{n+1}{k+1}(-1)^{k+1} \\
&= 1+\sum_{k=1}^{n+1}\binom{n+1}{k}(-1)^k \\
&= \sum_{k=0}^{n+1}\binom{n+1}{k}(-1)^k = 0.
\end{aligned}$$

□

We also have the following three lemmas relating $R(\cdot,\cdot), S(\cdot,\cdot)$ and $T(\cdot,\cdot)$.

Lemma 2. *For any $n \geq 1$ and $d \geq 2$, $S(n,d) = -\sum_{m=0}^{n-1} R(m,d)$.*

Proof. We prove the lemma by induction on n. For $n=1$, it is easy to verify that $S(1,d)=-1=-R(0,d)$. For $n \geq 2$,

$$\begin{aligned}
S(n,d) &= \binom{n}{n}\frac{(-1)^n}{n^{d-1}}+\sum_{k=1}^{n-1}\binom{n}{k}\frac{(-1)^k}{k^{d-1}} \\
&= \binom{n-1}{n-1}\frac{(-1)^n}{n^{d-1}}+\sum_{k=1}^{n-1}\binom{n-1}{k-1}\frac{(-1)^k}{k^{d-1}}+\sum_{k=1}^{n-1}\binom{n-1}{k}\frac{(-1)^k}{k^{d-1}} \\
&= \sum_{k=1}^{n}\binom{n-1}{k-1}\frac{(-1)^k}{k^{d-1}}+\sum_{k=1}^{n-1}\binom{n-1}{k}\frac{(-1)^k}{k^{d-1}} \\
&= -\sum_{k=0}^{n-1}\binom{n-1}{k}\frac{(-1)^k}{(k+1)^{d-1}}+\sum_{k=1}^{n-1}\binom{n-1}{k}\frac{(-1)^k}{k^{d-1}} \\
&= -R(n-1,d)+S(n-1,d).
\end{aligned}$$

□

Lemma 3. *For any $n \geq 2$ and $d \geq 2$, $T(n,d) = -\sum_{m=1}^{n-1} S(m,d)$.*

Proof. We prove the lemma by induction on n. For $n=2$, it is easy to verify that $T(2,d)=1=-S(1,d)$. For $n \geq 3$,

$$\begin{aligned}
T(n,d) &= \binom{n}{n}\frac{(-1)^n}{(n-1)^{d-1}}+\sum_{k=2}^{n-1}\binom{n}{k}\frac{(-1)^k}{(k-1)^{d-1}} \\
&= \binom{n-1}{n-1}\frac{(-1)^n}{(n-1)^{d-1}}+\sum_{k=2}^{n-1}\binom{n-1}{k-1}\frac{(-1)^k}{(k-1)^{d-1}}+\sum_{k=2}^{n-1}\binom{n-1}{k}\frac{(-1)^k}{(k-1)^{d-1}} \\
&= \sum_{k=2}^{n}\binom{n-1}{k-1}\frac{(-1)^k}{(k-1)^{d-1}}+\sum_{k=2}^{n-1}\binom{n-1}{k}\frac{(-1)^k}{(k-1)^{d-1}} \\
&= -\sum_{k=1}^{n-1}\binom{n-1}{k}\frac{(-1)^k}{k^{d-1}}+\sum_{k=2}^{n-1}\binom{n-1}{k}\frac{(-1)^k}{(k-1)^{d-1}} \\
&= -S(n-1,d)+T(n-1,d).
\end{aligned}$$

□

Lemma 4. *For any $n \ge 0$ and $d \ge 3$, $R(n,d) = -S(n+1,d-1)/(n+1)$.*

Proof. We prove the lemma by induction on n. For $n = 0$, it is easy to verify that $R(0,d) = 1 = -S(1,d-1)$. For $n \ge 1$, it suffices to prove that

$$(n+1)R(n,d) - nR(n-1,d) = -S(n+1,d-1) + S(n,d-1).$$

(Then the identity in the lemma follows by addition.)

$$\begin{aligned}
(n+1)R(n,d) - nR(n-1,d) &= \sum_{k=0}^{n} \binom{n}{k} \frac{n+1}{k+1} \frac{(-1)^k}{(k+1)^{d-2}} - \sum_{k=0}^{n-1} \binom{n-1}{k} \frac{n}{k+1} \frac{(-1)^k}{(k+1)^{d-2}} \\
&= -\sum_{k=0}^{n} \binom{n+1}{k+1} \frac{(-1)^{k+1}}{(k+1)^{d-2}} + \sum_{k=0}^{n-1} \binom{n}{k+1} \frac{(-1)^{k+1}}{(k+1)^{d-2}} \\
&= -\sum_{k=1}^{n+1} \binom{n+1}{k} \frac{(-1)^k}{k^{d-2}} + \sum_{k=1}^{n} \binom{n}{k} \frac{(-1)^k}{k^{d-2}} \\
&= -S(n+1,d-1) + S(n,d-1).
\end{aligned}$$

□

For any $n \ge 1$, let $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ denote the n-th harmonic number. It is well-known that $H_n \sim \ln n$. From Lemmas 1, 2, and 3, we immediately obtain the following corollary:

Corollary 2. *For any $n \ge 1$, $S(n,2) = -H_n$.*

Corollary 3. *For any $n \ge 2$, $T(n,2) = \sum_{m=1}^{n-1} H_m$.*

By repeatedly applying Lemmas 2 and 4, we can determine the asymptotic growth rates of $R(n,d)$ and $S(n,d)$ when d is fixed:

Lemma 5. *For any fixed $d \ge 2$, $R(n,d) \sim \frac{1}{(d-2)!} n^{-1} \ln^{d-2} n = \Theta(n^{-1} \log^{d-2} n)$ and $-S(n,d) \sim \frac{1}{(d-1)!} \ln^{d-1} n = \Theta(\log^{d-1} n)$.*

Proof. We prove the lemma by induction on d. For $d = 2$, we have $R(n,2) = \frac{1}{n+1} \sim n^{-1}$ by Lemma 1, and $-S(n,2) = H_n \sim \ln n$ by Corollary 2. Next let $d \ge 3$, and assume that $-S(n,d-1) \sim \frac{1}{(d-2)!} \ln^{d-2} n = \Theta(\log^{d-2} n)$. Then, by Lemma 4, we have

$$R(n,d) = -S(n+1,d-1)/(n+1) \sim \frac{1}{(d-2)!} n^{-1} \ln^{d-2} n = \Theta(n^{-1} \log^{d-2} n).$$

Subsequently, by Lemma 2, we have

$$-S(n,d) = \sum_{m=0}^{n-1} R(m,d) = \Theta\left(\int_1^n t^{-1} \log^{d-2} t \, \mathrm{d}t\right) = \Theta(\log^{d-1} n).$$

More precisely, note that

$$-S(n,d) = \sum_{m=0}^{\lceil \ln n \rceil - 1} R(m,d) + \sum_{m=\lceil \ln n \rceil}^{n-1} R(m,d) \sim \sum_{m=\lceil \ln n \rceil}^{n-1} R(m,d)$$

since

$$\sum_{m=0}^{\lceil \ln n \rceil - 1} R(m,d) = -S(\lceil \ln n \rceil, d) = o\big(-S(n,d)\big).$$

We do not have $R(m,d) \sim \frac{1}{(d-2)!} m^{-1} \ln^{d-2} m$ when m is a constant. But after changing the lower limit of the sum from 0 to a non-constant $\lceil \ln n \rceil$, we can now apply $R(m,d) \sim \frac{1}{(d-2)!} m^{-1} \ln^{d-2} m$ to each summand and obtain:

$$- S(n,d) \sim \sum_{m=\lceil \ln n \rceil}^{n-1} R(m,d) \sim \sum_{m=\lceil \ln n \rceil}^{n-1} \frac{1}{(d-2)!} m^{-1} \ln^{d-2} m$$

$$\sim \int_{\ln n}^{n} \frac{1}{(d-2)!} t^{-1} \ln^{d-2} t \, \mathrm{d}t = \frac{1}{(d-1)!} \ln^{d-1} t \Big|_{\ln n}^{n} \sim \frac{1}{(d-1)!} \ln^{d-1} n. \quad \square$$

We can now also determine the asymptotic growth rate of $T(n,d)$:

Lemma 6. *For any fixed $d \geq 2$, $T(n,d) \sim \frac{1}{(d-1)!} n \ln^{d-1} n = \Theta(n \log^{d-1} n)$.*

Proof. We have $T(n,d) = -\sum_{m=1}^{n-1} S(m,d)$ by Lemma 3, and $-S(m,d) \sim \frac{1}{(d-1)!} \ln^{d-1} m = \Theta(\log^{d-1} m)$ by Lemma 5. Consequently,

$$T(n,d) = \sum_{m=1}^{n-1} -S(m,d) = \Theta\left(\sum_{m=1}^{n-1} \log^{d-1} m \right) = \Theta(n \log^{d-1} n).$$

More precisely,

$$T(n,d) = \sum_{m=1}^{n-1} -S(m,d) \sim \sum_{m=\lceil \ln n \rceil}^{n-1} -S(m,d) \sim \sum_{m=\lceil \ln n \rceil}^{n-1} \frac{1}{(d-1)!} \ln^{d-1} m$$

$$\sim \int_{\ln n}^{n} \frac{1}{(d-1)!} \ln^{d-1} t \, \mathrm{d}t \sim \int_{1}^{n} \frac{1}{(d-1)!} \ln^{d-1} t \, \mathrm{d}t$$

$$\sim \int_{1}^{n} \left(\frac{1}{(d-1)!} \ln^{d-1} t + \frac{1}{(d-2)!} \ln^{d-2} t \right) \mathrm{d}t$$

$$= \frac{1}{(d-1)!} t \ln^{d-1} t \Big|_{1}^{n} = \frac{1}{(d-1)!} n \ln^{d-1} n. \quad \square$$

3.6 Base cases for $G(n,a,b)$

As a warm-up exercise, we obtain the values of $G(n,a,b)$ for all combinations of a and b with $a + b \leq 2$ in order to derive the exact formulas for $F(n,2)$ and $E(n,2)$. Recall (5) that

$$G(n,a,b) = (-1)^b \sum_{k=2a+b}^{n} \binom{n}{k} (-1)^k \frac{k!}{(k-2a-b)!(k-1)^a k^{a+b}}.$$

By (10) alone, we easily obtain

$$G(n,0,0) = \sum_{k=0}^{n} \binom{n}{k} (-1)^k = 0, \tag{14}$$

$$G(n,0,1) = -\sum_{k=1}^{n} \binom{n}{k} (-1)^k = -\sum_{k=0}^{n} \binom{n}{k} (-1)^k + \sum_{k=0}^{0} \binom{n}{k} (-1)^k = 0 + 1 = 1, \tag{15}$$

$$G(n,1,0) = \sum_{k=2}^{n} \binom{n}{k} (-1)^k = \sum_{k=0}^{n} \binom{n}{k} (-1)^k - \sum_{k=0}^{1} \binom{n}{k} (-1)^k = 0 - (1-n) = n - 1. \tag{16}$$

By (10) and Corollary 2, we have

$$\begin{aligned}
G(n,0,2) &= \sum_{k=2}^{n}\binom{n}{k}(-1)^k\frac{k!}{(k-2)!k^2}\\
&= \sum_{k=2}^{n}\binom{n}{k}(-1)^k\frac{k-1}{k}\\
&= \sum_{k=2}^{n}\binom{n}{k}(-1)^k - \sum_{k=2}^{n}\binom{n}{k}\frac{(-1)^k}{k}\\
&= \left[\sum_{k=0}^{n}\binom{n}{k}(-1)^k - \sum_{k=0}^{1}\binom{n}{k}(-1)^k\right] - \left[\sum_{k=1}^{n}\binom{n}{k}\frac{(-1)^k}{k} - \sum_{k=1}^{1}\binom{n}{k}\frac{(-1)^k}{k}\right]\\
&= \big[0-(1-n)\big] - \big[S(n,2)+n\big]\\
&= -S(n,2)-1\\
&= H_n - 1, \qquad (17)
\end{aligned}$$

$$\begin{aligned}
G(n,1,1) &= -\sum_{k=3}^{n}\binom{n}{k}(-1)^k\frac{k!}{(k-3)!(k-1)^1k^2}\\
&= -\sum_{k=3}^{n}\binom{n}{k}(-1)^k\frac{k-2}{k}\\
&= -\sum_{k=2}^{n}\binom{n}{k}(-1)^k\frac{k-2}{k}\\
&= -\sum_{k=2}^{n}\binom{n}{k}(-1)^k + 2\sum_{k=2}^{n}\binom{n}{k}\frac{(-1)^k}{k}\\
&= -\left[\sum_{k=0}^{n}\binom{n}{k}(-1)^k - \sum_{k=0}^{1}\binom{n}{k}(-1)^k\right] + 2\left[\sum_{k=1}^{n}\binom{n}{k}\frac{(-1)^k}{k} - \sum_{k=1}^{1}\binom{n}{k}\frac{(-1)^k}{k}\right]\\
&= -\big[0-(1-n)\big] + 2\big[S(n,2)+n\big]\\
&= n + 2S(n,2) + 1\\
&= n - 2H_n + 1. \qquad (18)
\end{aligned}$$

To handle further values of $G(n,a,b)$ we need to obtain partial fraction decompositions of the damping factors in the alternating binomial sums. Note that

$$\frac{(k-3)(k-2)}{(k-1)k} = 1 - \frac{6}{k} + \frac{2}{k-1}.$$

Thus by (10) and Corollaries 2 and 3 we have

$$\begin{aligned}
G(n,2,0) &= \sum_{k=4}^{n} \binom{n}{k} (-1)^k \frac{k!}{(k-4)!(k-1)^2k^2} \\
&= \sum_{k=4}^{n} \binom{n}{k} (-1)^k \frac{(k-3)(k-2)}{(k-1)k} \\
&= \sum_{k=2}^{n} \binom{n}{k} (-1)^k \frac{(k-3)(k-2)}{(k-1)k} \\
&= \sum_{k=2}^{n} \binom{n}{k} (-1)^k \left(1 - \frac{6}{k} + \frac{2}{k-1}\right) \\
&= \sum_{k=2}^{n} \binom{n}{k} (-1)^k - 6\sum_{k=2}^{n} \binom{n}{k} \frac{(-1)^k}{k} + 2\sum_{k=2}^{n} \binom{n}{k} \frac{(-1)^k}{k-1} \\
&= \big[0 - (1-n)\big] - 6\big[S(n,2) + n\big] + 2\,T(n,2) \\
&= 2\,T(n,2) - 6S(n,2) - 5n - 1 \\
&= 2\sum_{m=1}^{n-1} H_m + 6H_n - 5n - 1 \\
&= 2\sum_{m=1}^{n} H_m + 4H_n - 5n - 1. \qquad (19)
\end{aligned}$$

Note that $G(n,2,0) \sim 2\ln(n!) \sim 2n\ln n$ since $H_n \sim \ln n$ and $n! \sim \sqrt{2\pi n}(n/e)^n$ (Stirling's approximation).

After obtaining exact values of $G(n,a,b)$ for all combinations of a and b with $a+b \le 2$, we can now derive exact formulas for $F(n,2)$ and $E(n,2)$ according to (1) and (2). By (19) and (1), we have

$$F(n,2) = 2\sum_{m=1}^{n} H_m + 4H_n - 5n - 1 \sim 2n\ln n. \qquad (20)$$

By (14), (15), (16), (17), (18), (19), and (2), we have

$$\begin{aligned}
E(n,2) &= \sum_{a=0}^{2}\sum_{b=0}^{2-a} \binom{2}{a}\binom{2-a}{b} 2^b\, G(n,a,b) \\
&= G(n,0,0) + 4G(n,0,1) + 2G(n,1,0) + 4G(n,0,2) + 4G(n,1,1) + G(n,2,0) \\
&= 0 + 4 + 2(n-1) + 4(H_n - 1) + 4(n - 2H_n + 1) + 2\sum_{m=1}^{n} H_m + 4H_n - 5n - 1 \\
&= 2\sum_{m=1}^{n} H_m + n + 1 \sim 2n\ln n. \qquad (21)
\end{aligned}$$

3.7 General cases for $G(n,a,b)$

By (15) and (16), we have $G(n,0,1) = 1 = \ln^{b-1} n$ for $a=0$ and $b=1$, and $G(n,1,0) = n-1 \sim b!\,n$ for $a=1$ and $b=0$, thus confirming Theorem 1 for the base cases when $a+b=1$. It remains to consider the cases when $a+b \ge 2$.

Recall (5) and (6) that

$$G(n,a,b) = (-1)^b \sum_{k=2a+b}^{n} \binom{n}{k} (-1)^k f(k,a,b),$$

where

$$f(k,a,b) := \frac{k!}{(k-2a-b)!(k-1)^a k^{a+b}}.$$

3.7.1 Case 1: $a \geq 2$ and $b \geq 0$

Recall the decomposition of $f(k,a,b)$ for $a \geq 2$ and $b \geq 0$ in (7):

$$\frac{(k-2)\cdots(k-2a-b+1)}{(k-1)^{a-1}k^{a+b-1}} = 1 + \sum_{i=1}^{a-1} \frac{p_i}{(k-1)^i} + \sum_{j=1}^{a+b-1} \frac{q_j}{k^j},$$

By multiplying both sides of the above decomposition by $(k-1)^{a-1}k^{a+b-1}$ and then setting $k = 1$, we determine the coefficient

$$p_{a-1} = (-1)^b (2a+b-2)!.$$

Expand $G(n,a,b)$ according to the decomposition:

$$\begin{aligned}
G(n,a,b) &= (-1)^b \sum_{k=2a+b}^{n} \binom{n}{k} (-1)^k \frac{(k-2)\cdots(k-2a-b+1)}{(k-1)^{a-1}k^{a+b-1}} \\
&= (-1)^b \sum_{k=2}^{n} \binom{n}{k} (-1)^k \frac{(k-2)\cdots(k-2a-b+1)}{(k-1)^{a-1}k^{a+b-1}} \\
&= (-1)^b \sum_{k=2}^{n} \binom{n}{k} (-1)^k \left(1 + \sum_{i=1}^{a-1} \frac{p_i}{(k-1)^i} + \sum_{j=1}^{a+b-1} \frac{q_j}{k^j} \right) \\
&= (-1)^b \left[\sum_{k=2}^{n} \binom{n}{k} (-1)^k + \sum_{i=1}^{a-1} p_i \left(\sum_{k=2}^{n} \binom{n}{k} \frac{(-1)^k}{(k-1)^i} \right) + \sum_{j=1}^{a+b-1} q_j \left(\sum_{k=2}^{n} \binom{n}{k} \frac{(-1)^k}{k^j} \right) \right] \\
&= (-1)^b \left[(n-1) + \sum_{i=1}^{a-1} p_i \, T(n,i+1) + \sum_{j=1}^{a+b-1} q_j \big(S(n,j+1) + n \big) \right].
\end{aligned}$$

By Lemmas 5 and 6, the dominating term in this expansion is

$$(-1)^b p_{a-1} T(n,a) = (2a+b-2)! \, T(n,a).$$

Thus

$$G(n,a,b) \sim (2a+b-2)! \, T(n,a) \sim \frac{(2a+b-2)!}{(a-1)!} n \ln^{a-1} n = \Theta(n \log^{a-1} n). \tag{22}$$

Note in particular that for $a = 2$ and $b = 0$, we have $G(n,2,0) \sim 2n \ln n$, which is consistent with (19).

3.7.2 Case 2: $a = 1$ and $b \geq 1$

Recall the decomposition of $f(k, a, b)$ for $a = 1$ and $b \geq 1$ in (8):

$$\frac{(k-2)\cdots(k-b-1)}{k^b} = 1 + \sum_{j=1}^{b} \frac{q_j}{k^j}.$$

By setting $k = 1$ on both sides of the decomposition, we determine the sum of the coefficients

$$\sum_{j=1}^{b} q_j = (-1)^b b! - 1.$$

Expand $G(n, a, b)$ according to the decomposition:

$$\begin{aligned}
G(n,1,b) &= (-1)^b \sum_{k=b+2}^{n} \binom{n}{k} (-1)^k \frac{(k-2)\cdots(k-b-1)}{k^b} \\
&= (-1)^b \sum_{k=2}^{n} \binom{n}{k} (-1)^k \frac{(k-2)\cdots(k-b-1)}{k^b} \\
&= (-1)^b \sum_{k=2}^{n} \binom{n}{k} (-1)^k \left(1 + \sum_{j=1}^{b} \frac{q_j}{k^j}\right) \\
&= (-1)^b \left[\sum_{k=2}^{n} \binom{n}{k} (-1)^k + \sum_{j=1}^{b} q_j \left(\sum_{k=2}^{n} \binom{n}{k} \frac{(-1)^k}{k^j}\right)\right] \\
&= (-1)^b \left[(n-1) + \sum_{j=1}^{b} q_j \big(S(n, j+1) + n\big)\right].
\end{aligned}$$

By Lemma 5, the dominating term in this expansion is

$$(-1)^b \left(1 + \sum_{j=1}^{b} q_j\right) n = b!\, n.$$

Thus

$$G(n,1,b) \sim b!\, n = \Theta(n). \tag{23}$$

Note in particular that for $b = 1$, we have $G(n, 1, 1) \sim n$, which is consistent with (18).

3.7.3 Case 3: $a = 0$ and $b \geq 2$

Recall the decomposition of $f(k, a, b)$ for $a = 0$ and $b \geq 2$ in (9):

$$\frac{(k-1)\cdots(k-b+1)}{k^{b-1}} = 1 + \sum_{j=1}^{b-1} \frac{q_j}{k^j}.$$

By multiplying both sides of the above decomposition by k^{b-1} and then setting $k = 0$, we determine the coefficient

$$q_{b-1} = (-1)^{b-1}(b-1)!.$$

Expand $G(n,a,b)$ according to the decomposition:

$$\begin{aligned} G(n,0,b) &= (-1)^b \sum_{k=b}^{n} \binom{n}{k} (-1)^k \frac{(k-1)\cdots(k-b+1)}{k^{b-1}} \\ &= (-1)^b \sum_{k=1}^{n} \binom{n}{k} (-1)^k \frac{(k-1)\cdots(k-b+1)}{k^{b-1}} \\ &= (-1)^b \sum_{k=1}^{n} \binom{n}{k} (-1)^k \left(1 + \sum_{j=1}^{b-1} \frac{q_j}{k^j}\right) \\ &= (-1)^b \left[\sum_{k=1}^{n} \binom{n}{k} (-1)^k + \sum_{j=1}^{b-1} q_j \left(\sum_{k=1}^{n} \binom{n}{k} \frac{(-1)^k}{k^j}\right)\right] \\ &= (-1)^b \left[-1 + \sum_{j=1}^{b-1} q_j S(n, j+1)\right]. \end{aligned}$$

By Lemma 5, the dominating term in this expansion is

$$(-1)^b q_{b-1} S(n,b) = (b-1)! \left(-S(n,b)\right) \sim \ln^{b-1} n.$$

Thus

$$G(n,0,b) \sim \ln^{b-1} n = \Theta(\log^{b-1} n). \tag{24}$$

Note in particular that for $b = 2$, we have $G(n,0,2) \sim \ln n$, which is consistent with (17). With the three equations (22), (23), and (24), the proof of Theorem 1 is now complete.

Proof of Corollary 1. After obtaining precise formulas for $G(n,a,b)$ for all combinations of a and b with $0 \le a+b \le d$, it is straightforward to derive precise formulas for $F(n,d)$ and $E(n,d)$ according to (1) and (2). Clearly, for any fixed $d \ge 1$, $G(n,d,0)$ is the dominating term in the expression of $E(n,d)$. Thus

$$E(n,d) \sim F(n,d) = G(n,d,0),$$

where $G(n,d,0) \sim \frac{(2d+0-2)!}{(d-1)!} n \ln^{d-1} n = \frac{(2d-2)!}{(d-1)!} n \ln^{d-1} n$ for $d \ge 2$, and $G(n,d,0) \sim 0!\, n = \frac{(2d-2)!}{(d-1)!} n \ln^{d-1} n$ too for $d = 1$.

Acknowledgement. The authors are grateful to an anonymous reviewer for thoughtful suggestions and careful reading.

References

[1] A. Aggarwal and S. Suri, Fast algorithms for computing the largest empty rectangle, in: *Proceedings of the 3rd Annual Symposium on Computational Geometry*, 1987, pp. 278–290.

[2] M. Atallah and G. Frederickson, A note on finding the maximum empty rectangle, *Discrete Applied Mathematics*, **13** (1986), 87–91.

[3] M. Atallah and S. R. Kosaraju, An efficient algorithm for maxdominance, with applications, *Algorithmica*, **4** (1989), 221–236.

[4] J. Augustine, S. Das, A. Maheshwari, S. C. Nandy, S. Roy, and S. Sarvattomananda, Querying for the largest empty geometric object in a desired location, 2010, `http://arxiv.org/abs/1004.0558v2`.

[5] J. Backer and M. Keil, The bichromatic rectangle problem in high dimensions, in: *Proceedings of the 21st Canadian Conference on Computational Geometry*, 2009, pp. 157–160.

[6] J. Backer and M. Keil, The mono- and bichromatic empty rectangle and square problems in all dimensions, in: *Proceedings of the 9th Latin American Symposium on Theoretical Informatics*, 2010, pp. 14–25.

[7] J. L. Bentley, H. T. Kung, M. Schkolnick, and C. D. Thompson, On the average number of maxima in a set of vectors and applications, *Journal of the ACM*, **25** (1978), 536–543.

[8] J.-D. Boissonnat, M. Sharir, B. Tagansky, and M. Yvinec, Voronoi diagrams in higher dimensions under certain polyhedral distance functions, *Discrete & Computational Geometry*, **19** (1998), 485–519.

[9] B. Chazelle, R. Drysdale and D. T. Lee, Computing the largest empty rectangle, *SIAM Journal on Computing*, **15** (1986), 300–315.

[10] C. Chuan-Chong and K. Khee-Meng, *Principles and Techniques in Combinatorics*, World Scientific, Singapore, 1996.

[11] A. Datta, Efficient algorithms for the largest empty rectangle problem, *Information Sciences*, **64** (1992), 121–141.

[12] A. Datta and S. Soundaralakshmi, An efficient algorithm for computing the maximum empty rectangle in three dimensions, *Information Sciences*, **128** (2000), 43–65.

[13] A. Dumitrescu and M. Jiang, On the largest empty axis-parallel box amidst n points, *Algorithmica*, (2012), doi:10.1007/s00453-012-9635-5. Also at `http://arxiv.org/abs/0909.3127v2`.

[14] J. Edmonds, J. Gryz, D. Liang, and R. Miller, Mining for empty spaces in large data sets, *Theoretical Computer Science*, **296** (2003), 435–452.

[15] P. Giannopoulos, C. Knauer, M. Wahlström, and D. Werner, Hardness of discrepancy computation and ε-net verification in high dimension, *Journal of Complexity*, (2011), doi:10.1016/j.jco.2011.09.001.

[16] H. Kaplan, S. Mozes, Y. Nussbaum, and M. Sharir, Submatrix maximum queries in Monge matrices and Monge partial matrices, and their applications, in *Proceedings of the 23rd ACM-SIAM Symposium on Discrete Algorithms*, 2012, pp. 338–355.

[17] H. Kaplan, N. Rubin, M. Sharir, and E. Verbin, Efficient colored orthogonal range counting, *SIAM Journal on Computing*, **38** (2008), 982–1011.

[18] R. Klein, Direct dominance of points, *International Journal of Computer Mathematics*, **19** (1986), 225–244.

[19] L. D. Kudryavtsev, The method of undetermined coefficients, in *Encyclopaedia of Mathematics* (M. Hazewinkel, editor), Springer, 2001.

[20] A. Kurosh, *Higher Algebra*, Mir Publishers, Moscow, 1975.

[21] D. Marx, Parameterized complexity and approximation algorithms, *Computer Journal*, **51** (2008), 60–78.

[22] M. McKenna, J. O'Rourke, and S. Suri, Finding the largest rectangle in an orthogonal polygon, in: *Proceedings of the 23rd Annual Allerton Conference on Communication, Control and Computing*, Urbana-Champaign, Illinois, October 1985.

[23] A. Naamad, D. T. Lee, and W.-L. Hsu, On the maximum empty rectangle problem, *Discrete Applied Mathematics*, **8** (1984), 267–277.

[24] M. Orlowski, A new algorithm for the largest empty rectangle problem, *Algorithmica*, **5** (1990), 65–73.