A NOTE ON THE EXPECTED TIME FOR FINDING MAXIMA BY LIST ALGORITHMS

Luc Devroye
School of Computer Science
McGill University

ABSTRACT. Maxima in $\mathbb{R}^d$ are found incrementally by maintaining a linked list and comparing new elements against the linked list. If the elements are independent and uniformly distributed in the unit square $[0,1]^d$, then regardless of how the list is manipulated by an adversary, the expected time is $O(n \log^{d-2} n)$. This should be contrasted with the fact that the expected number of maxima grows as $\log^{d-1} n$, so no adversary can force an expected complexity of $n \log^{d-1} n$. Note that the expected complexity is $O(n)$ for $d = 2$. Conversely, there are list-manipulating adversaries for which the given bound is attained. However, if we naively add maxima to the list without changing the order, then the expected number of element comparisons is $n + o(n)$ for any $d \geq 2$. In the paper, we also derive new tail bounds and moment inequalities for the number of maxima.

KEYWORDS AND PHRASES. Outer layers, maxima, list algorithms, expected time, randomized algorithms, probabilistic analysis.

CR CATEGORIES: 3.74, 5.25, 5.5.

Research of the author was sponsored by NSERC Grant A3456 and by FCAR Grant 90-ER-0291. Address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2A7. Email: `luc@cs.mcgill.ca`.

List algorithms and adversaries.

Given are $X_1, \ldots, X_n$, i.i.d. points uniformly distributed in $[0,1]^d$. We write $X_i = (X_{i1}, \ldots, X_{id})$. We say that X_i is a maximum if no X_j, $j \neq i$, exists for which $X_{jl} > X_{il}$ for all l. If N is the number of maxima, then it is known that

$$\mathbf{E}N \sim \frac{\log^{d-1} n}{(d-1)!}$$

(see Barndorff-Nielsen and Sobel, 1966). In 1990, Bentley, Clarkson and Levine pointed out that list algorithms may be quite efficient on the average for finding all maxima. A list algorithm is one in which a linked list of maxima is kept, and each X_i is considered in turn. The maxima are kept in the list in some order. In the worst scenario, the order may be determined by an "adversary". In the ordinary list algorithm, X_i is compared with each list element in turn until either the list is exhausted (in which case X_i is a maximum itself and is added to the list) or a list element X_j is found that dominates X_i in all components (in which case X_i is discarded). Even before handling X_i, the list may be reorganized by an adversary. It is more common though to reorganize the list after having processed X_i. For example, Bentley, Clarkson and Levine (1990, 1993) suggested the MTF heuristic: move X_j to the head of the list if X_j is the first list element that dominates X_i, and appending X_i in the rear if X_i itself is a maximum. They conjectured that this strategy would take $n + o(n)$ expected comparisons between elements where one (vector) comparison between elements may involve up to d scalar comparisons. This was later proved by Golin (1994) for $d = 2$. In this note, we extend and strengthen these results in several directions:

A. We show that the expected number of element comparisons for any list algorithm (manipulated at will by adversaries) is bounded by $3n$ when $d = 2$. For general d, the expected time is guaranteed to grow as $O(n \log^{d-2} n)$ (see Theorem 1 below). As $\mathbf{E}N$ grows as $\log^{d-1} n$, it is remarkable that no adversary can force an expected complexity of $\Omega(n \log^{d-1} n)$.

B. We show that the rates above cannot be improved.

C. Without adversaries or list manipulation of any kind, the linked list stores all current maxima in chronological order. New maxima are appended at the rear of the list. A list constructed in this manner will be called a random list. Despite the lack of any meaningful list organization, the expected number of element comparisons is $n + o(n)$ for any $d \geq 2$ (see Theorem 2).

The first result shows that for $d = 2$, all list algorithms take $O(n)$ time on the average, regardless of the list ordering strategy. For $d > 2$, the expected time linearity may be lost. Theorem 2 shows the futility of any list organization method (such as MTF) since a simple random ordering insures $n + o(n)$ element comparisons on the average. It may still be true however that MTF has a better "$o(n)$" term.

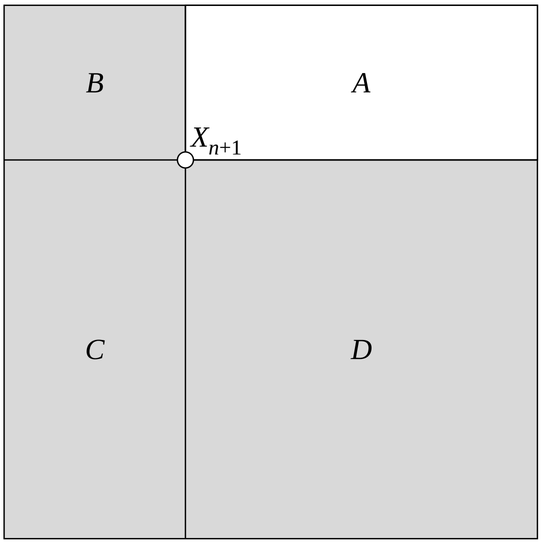
The main results.

THEOREM 1. *For any list algorithm, if* $X_1, \ldots, X_n$ *are i.i.d. random vectors uniformly drawn from* $[0,1]^d$, *the expected time is* $O(n \log^{d-2} n)$ *for all* $d \geq 2$. *Note that for* $n = 2$, *it is in fact* $O(n)$. *Conversely, there exist list adversaries (who are allowed to rearrange the lists at will, but not the order of insertion) such that the list algorithm takes expected time bounded from below by* $\Omega(n \log^{d-2} n)$.

THEOREM 2. *Let* $X_1, \ldots, X_n$ *be i.i.d. random vectors uniformly drawn from* $[0,1]^d$ *where* $d \geq 2$. *Then the expected number of comparisons between elements in the random list algorithm is* $n + o(n)$.

In addition, some interesting probability theoretical results about maxima are obtained as well. These include tail bounds for N and inequalities for the moments of N.

Proof of the upper bound when d equals 2.



Partition of the unit square by the newly inserted point X_{n+1} into four rectangles A, B, C, and D.

Assume that $X_1, \ldots, X_n$ have been processed. When a point X_{n+1} is considered, we look at the the four quadrants formed by moving the origin to X_{n+1} as in the figure above. Using the notation of that figure, we see that the number of points of A compared with X_{n+1} during the traversal of the current list is at most one. Thus, the number of comparisons during insertion is at most one plus N_n, the number of maxima in $B \cup C \cup D$. We will write N_n instead of N to denote the fact that n points have been processed before X_{n+1}. The expected time for finding the maxima of n points is thus not more than

$$n + \sum_{i=0}^{n-1} \mathbf{E} N_i \ .$$

Now, $N_i \leq N_i' + N_i''$, where N_i' is the number of maxima in $C \cup D$ and N_i'' is the number of maxima in $B \cup C$. By symmetry, we have

$$n + \sum_{i=0}^{n-1} \mathbf{E} N_i \leq n + 2 \sum_{i=0}^{n-1} \mathbf{E} N_i' \ .$$

Given X_{n+1}, we may compute $\mathbf{E} N_n'$ as follows:

$$\begin{aligned} &\mathbf{E}\{N_n' | X_{n+1}\} \\ &\quad = n\, \mathbf{P}\{X_1 \in C \cup D, X_1 \text{ is a maximum among } X_1, \ldots, X_n | X_{n+1}\} \\ &\quad = n\, \mathbf{E}\left\{ I_{X_1 \in C \cup D} \mathbf{P}\{X_1 \text{ is a maximum among } X_1, \ldots, X_n | X_1\} | X_{n+1} \right\} \end{aligned}$$

$$
\begin{aligned}
&= n\,\mathbf{E}\left\{ I_{X_1 \in C \cup D} (1 - (1 - X_{1,1})(1 - X_{1,2}))^{n-1} | X_{n+1} \right\} \\
&\quad \text{(where } X_1 = (X_{1,1}, X_{1,2})\text{)} \\
&= n \int_0^Y \int_0^1 (1 - (1-x)(1-y))^{n-1}\, dx dy \\
&= n \int_0^Y \frac{1 - (1 - (1-y))^n}{n(1-y)}\, dy \\
&= \int_0^Y \frac{1 - y^n}{1 - y}\, dy \\
&= \int_0^Y \sum_{j=0}^{n-1} y^j\, dy \\
&= \sum_{j=1}^{n} \frac{Y^j}{j} \ .
\end{aligned}
$$

If we take the expected value and note that for a uniform $[0,1]$ random variable Y, we have $\mathbf{E}Y^j = 1/(j+1)$, then

$$
\begin{aligned}
\mathbf{E}\{N_n'\} = \mathbf{E}\left\{\mathbf{E}\{N_n' | (X,Y)\}\right\} &= \mathbf{E}\left\{ \sum_{j=1}^{n} \frac{Y^j}{j} \right\} \\
&= \sum_{j=1}^{n} \frac{1}{j(j+1)} = 1 - \frac{1}{n+1} < 1 \ .
\end{aligned}
$$

Therefore, the expected number of point comparisons for finding the maxima is not more than

$$
n + 2n = 3n \ .
$$

Proof of the upper bound for d greater than 2.

Assume $d \geq 3$. In the above argument, set $X_{n+1} = Z = (Z_1, \ldots, Z_d)$, and let N'_n denote the number of maxima whose first component is less than Z_1. Then, arguing by symmetry as above, when Z is inserted, the expected time does not exceed

$$n + d \sum_{i=0}^{n-1} \mathbf{E} N'_i \ .$$

Write $X_j = (X_{j1}, \ldots, X_{jd})$. Define the set $C = \{z \in \mathbb{R}^d : z_1 \leq Z_1\}$ and the random variable $Y = \prod_{j=2}^{d} (1 - X_{1j})$. Then

$$\begin{aligned}
&\mathbf{E}\{N'_n | Z\} \\
&\quad = n \, \mathbf{P}\{X_1 \in C, X_1 \text{ is a maximum among } X_1, \ldots, X_n | Z\} \\
&\quad = n \, \mathbf{E}\left\{I_{X_1 \in C} \mathbf{P}\{X_1 \text{ is a maximum among } X_1, \ldots, X_n | X_1\} | Z\right\} \\
&\quad = n \, \mathbf{E}\left\{ I_{X_1 \in C} \left(1 - \prod_{j=1}^{d} (1 - X_{1j}) \right)^{n-1} | Z \right\} \\
&\quad = n \, \mathbf{E}\left\{ I_{X_{11} \leq Z_1} (1 - Y(1 - X_{11}))^{n-1} | Z_1 \right\} \\
&\quad = n \, \mathbf{E}\left\{ I_{U \leq Z_1} (1 - Y(1 - U))^{n-1} | Z_1 \right\}
\end{aligned}$$

where U is uniform $[0,1]$. Unconditioning, we see that

$$\begin{aligned}
\mathbf{E}\{N'_n\} &= n \, \mathbf{E}\left\{ I_{U \leq Z_1} (1 - Y(1 - U))^{n-1} \right\} \\
&= n \, \mathbf{E}\left\{ (1 - U) (1 - Y(1 - U))^{n-1} \right\} \\
&= n \, \mathbf{E}\left\{ U (1 - YU)^{n-1} \right\} \ .
\end{aligned}$$

Replace Y by VW, where V is uniform $[0,1]$ and W is the product of $d-2$ uniform $[0,1]$ random variables. It is easy to verify that the density of W is given by $f(w) = \log^{d-3}(1/w)/(d-3)!$, $0 < w < 1$ (see, e.g., Devroye, 1986). Then, taking the expectation first with respect to V and then with respect to U yields

$$\begin{aligned}
&\mathbf{E}\left\{ nU (1 - VWU)^{n-1} \right\} \\
&\quad = \mathbf{E}\left\{ nU \frac{1 - (1 - VWU)^n}{nWU} \right\} \\
&\quad = \mathbf{E}\left\{ \frac{1 - (1 - WU)^n}{W} \right\} \\
&\quad \leq \mathbf{E}\left\{ \min(1/W, nU) \right\} \\
&\quad \leq \mathbf{E}\left\{ \min(1/W, n/2) \right\} \\
&\qquad \text{(condition on } W \text{ and apply Jensen's inequality)}
\end{aligned}$$

$$
\begin{aligned}
&\le \int_0^{2/n} f(w)n/2\,dw + \int_{2/n}^1 f(w)/w\,dw \\
&\le f(2/n) + \frac{\log^{d-2}(n/2)}{(d-2)!} \\
&\sim \frac{\log^{d-2}(n/2)}{(d-2)!}\ .
\end{aligned}
$$

Thus, any list algorithm takes expected time $O(n\log^{d-2} n)$ when $d \ge 2$, even if the lists are allowed to be manipulated at will by adversaries.

Proof of the lower bound.

We finish by noting that our bound, under the adversary manipulation model, is best possible. It is trivially so for $d = 2$ as the expected time is $\Theta(n)$, but it is also true for $d > 2$. To that end, note that the expected time after manipulation of the list to make it perform at its worst, is bounded from below by

$$\sum_{i=0}^{n-1} \mathbf{E} N_i'\ .$$

Now, as noted above,

$$
\begin{aligned}
\mathbf{E}\{N_n'\} &= \mathbf{E}\left\{\frac{1-(1-WU)^n}{W}\right\} \\
&\ge \frac{1}{2}\mathbf{E}\left\{\frac{1-(1-(2/(n+1)(1/2))^n}{W} I_{W>2/(n+1)}\right\} \\
&\sim \frac{1-1/e}{2}\mathbf{E}\left\{\frac{1}{W} I_{W>2/(n+1)}\right\} \\
&\ge \int_{2/(n+1)}^1 f(w)/(4w)\,dw \\
&\ge f(1/\sqrt{(n+1)/2}) \int_{2/(n+1)}^{\sqrt{2/(n+1)}} 1/(4w)\,dw \\
&= \frac{\log^{d-3}(\sqrt{(n+1)/2}) \log\left(\sqrt{(n+1)/2}\right)}{4\,(d-3)!} \\
&= \frac{\log^{d-2}((n+1)/2)}{2^d\,(d-3)!}\ .
\end{aligned}
$$

Thus, the expected time is $\Omega(n\log^{d-2} n)$.

Random lists.

Theorem 2 relies heavily on the distributional assumption. It is true whenever the components of X_1 are independent and non-atomic. In this sense, things differ dramatically from randomized algorithms. The proof of Theorem 2 requires some basic properties of maxima, which are derived in the Appendix.

PROOF OF THEOREM 2. Let T be the expected number of list element comparisons when X_{n+1} is processed and the list contains the maxima for $X_1, \ldots, X_n$. Let N_k denote the number of maxima for $X_1, \ldots, X_k$. If the maxima are in chronological order, all permutations are equally likely. Let N be the number of maxima in set A of the figure, and let M be the number of maxima elsewhere. Recall that $\mathbf{E}\{M\} = O(\log^{d-2} n)$. In a random list with $N + M$ items, the expected number of comparisons for inserting a new element is the expected number of items encountered until one of the N items in A is seen. If the number of comparisons is T, we have

$$\mathbf{E}\{T|N, M\} = 1 + \frac{M}{N+1} \ .$$

Unconditioning, we get for any integer $m > 0$,

$$\begin{aligned}
\mathbf{E}T &\le 1 + \mathbf{E}\{M I_{N<m}\} + \mathbf{E}\left\{\frac{M}{N} I_{N \ge m}\right\} \\
&\le 1 + \sqrt{\mathbf{E}\{M^2\}\mathbf{P}\{N < m\}} + \frac{\mathbf{E}M}{m} \\
&\quad \text{(by the Cauchy-Schwarz inequality)} \\
&\le 1 + \sqrt{\mathbf{E}\{(M+N)^2\}\mathbf{P}\{N < m\}} + \frac{O(\log^{d-2} n)}{m} \\
&= 1 + \sqrt{\mathbf{E}\{N_n^2\}\mathbf{P}\{N < m\}} + \frac{O(\log^{d-2} n)}{m} \\
&\le 1 + \sqrt{(\mathbf{E}^2\{N_n\} + \mathbf{E}\{N_n\})\mathbf{P}\{N < m\}} + \frac{O(\log^{d-2} n)}{m} \\
&= 1 + O(\log^{d-1} n)\sqrt{\mathbf{P}\{N < m\}} + \frac{O(\log^{d-2} n)}{m} \ .
\end{aligned}$$

We first deal with $d = 2$, for which the short proof deserves separate treatment. Take $m = 1$. The proof is complete if we can show that

$$\mathbf{P}\{N < 1\} = o\left(\log^{-2} n\right) \ .$$

It should be noted that $N_1 \prec N_2 \prec N_3 \prec \cdots \prec N_k$, where $\prec$ denotes stochastic ordering. Given X_{n+1}, let V denote the number of X_i's with $1 \le i \le n$ that dominate X_{n+1}. Thus, these X_i's are in the quadrant A centered at X_{n+1}—see figure 1—and are uniformly distributed in A. Therefore, $N \overset{\mathcal{L}}{=} N_V$, where $\overset{\mathcal{L}}{=}$ denotes equality in distribution. By

Lemma 1 (see Appendix),

$$\begin{aligned}\mathbf{P}\{N<1\} &\le \mathbf{P}\{V<k\}+\mathbf{P}\{N_k<1\}\\ &\le \frac{k+d}{n+1}+\frac{e\log(k+1)}{k+1}\\ &= O\left(\frac{\log n}{\sqrt{n}}\right)\end{aligned}$$

if we take $k \sim \sqrt{n}$. The first part of the last step follows from the observation that if $V<k$, then one of the d coordinates of X_{n+1} must be among the $1+k/d$ largest of all similar coordinate values in $X_1,\ldots,X_{n+1}$. As the latter probability for a fixed coordinate does not exceed $(1+k/d)/(n+1))$, we have $\mathbf{P}\{V<k\} \le (k+d)/(n+1)$. This is more than is required, and the proof is complete.

For $d>2$, we must argue a bit more carefully. Define $k=\lfloor\sqrt{n}\rfloor$, and $m=\lfloor(1/2)\mathbf{E}\{N_k\}\rfloor$, and note that $m \sim \log^{d-1} n/(2^d(d-1)!)$. Observe that

$$\begin{aligned}\mathbf{E}T &\le 1+\mathbf{E}\{MI_{N=0}\}+\mathbf{E}\{MI_{0<N<m}\}+\mathbf{E}\left\{\frac{M}{N}I_{N\ge m}\right\}\\ &\le 1+\sqrt{\mathbf{E}\{M^2\}\mathbf{P}\{N=0\}}+\mathbf{E}\{MI_{0<N<m}\}+\frac{O(\log^{d-2} n)}{m}\\ &\quad \text{(by the Cauchy-Schwarz inequality)}\\ &= 1+\sqrt{\mathbf{E}\{(M+N)^2\}\mathbf{P}\{X_{n+1}\text{ is a maximum}\}}+\mathbf{E}\{MI_{0<N<m}\}+O\left(\frac{1}{\log n}\right)\\ &= 1+O(\log^{d-1} n)\sqrt{\frac{\mathbf{E}\{N_{n+1}\}}{n+1}}+\mathbf{E}\{MI_{0<N<m}\}+O\left(\frac{1}{\log n}\right)\ .\end{aligned}$$

The second term is $O(\log^{(3d-3)/2}/\sqrt{n})$ and the fourth term is $o(1)$ as well. We finish the proof by showing that the third term is $o(1)$. To bound $\mathbf{E}\{MI_{0<N<m}\}$, we note first that deterministically, $MI_{N>0} \le \sum_{j=1}^{d} N_j^*$, where N_j^* is the number of data points that are maxima in the $d-1$-dimensional subspace that does not include the j-th component. Thus, by Hölder's inequality, if we pick

$$\mu=\frac{1}{2}\left(1+\frac{d-1}{d-2}\right)\ ,\ \lambda=\frac{1}{1-\frac{1}{\mu}}$$

(so that $\mu,\lambda>1$, $1/\lambda+1/\mu=1$), then there exists a constant C only depending upon d (Lemma 2) such that

$$\begin{aligned}\mathbf{E}\{MI_{0<N<m}\} &\le \left(\mathbf{E}\left\{M^{\lambda}I_{N>0}\right\}\right)^{\frac{1}{\lambda}}\left(\mathbf{E}\left\{I^{\mu}_{N<m}\right\}\right)^{\frac{1}{\mu}}\\ &\le \left(d\sum_{j=1}^{d}\mathbf{E}\left\{N_j^{*\lambda}\right\}\right)^{\frac{1}{\lambda}}\left(\mathbf{P}\left\{N<m\right\}\right)^{\frac{1}{\mu}}\end{aligned}$$

$$
\begin{aligned}
&\leq d^{\frac{2}{\lambda}} \left(C\mathbf{E}^{\lambda}\left\{N_1^*\right\}\right)^{\frac{1}{\lambda}} \left(\mathbf{P}\{V<k\}+\mathbf{P}\{N_k<m\}\right)^{\frac{1}{\mu}} \\
&\leq (Cd^2)^{\frac{1}{\lambda}}\mathbf{E}\left\{N_1^*\right\} \left(\frac{k+d}{n+1}+\mathbf{P}\{N_k<m\}\right)^{\frac{1}{\mu}} \\
&= O\left(\log^{d-2} n\right) \left(\frac{k+d}{n+1}+\mathbf{P}\{N_k<m\}\right)^{\frac{1}{\mu}} \ . \\
&= O\left(\log^{d-2} n\right) \times \left(O\left(\frac{1}{\sqrt{n}}\right)+\mathbf{P}\{N_k-\mathbf{E}\{N_k\}<m-\mathbf{E}\{N_k\}\}\right)^{\frac{1}{\mu}} \\
&= O\left(\log^{d-2} n\right) \times \left(O\left(\frac{1}{\sqrt{n}}\right)+\mathbf{P}\{N_k-\mathbf{E}\{N_k\}<-(1/2)\mathbf{E}\{N_k\}\}\right)^{\frac{1}{\mu}} \\
&\leq O\left(\log^{d-2} n\right) \times \left(O\left(\frac{1}{\sqrt{n}}\right)+\frac{4\mathrm{Var}\{N_k\}}{\mathbf{E}^2\{N_k\}}\right)^{\frac{1}{\mu}} \\
&\quad \text{(by Chebyshev's inequality)} \\
&\leq O\left(\log^{d-2} n\right) \times \left(O\left(\frac{1}{\sqrt{n}}\right)+\frac{4C'}{\mathbf{E}\{N_k\}}\right)^{\frac{1}{\mu}} \\
&\quad \text{(by Lemma 3, where } C' \text{ depends upon } d \text{ only)} \\
&= \frac{O\left(\log^{d-2} n\right)}{\log^{\frac{d-1}{\mu}} n} \\
&= O\left(\log^{-\frac{1}{2\mu}} n\right) \\
&= O\left(\log^{-\frac{d-2}{2d-3}} n\right) \ .
\end{aligned}
$$

This concludes the proof of the theorem. □

Appendix: left tail bounds for number of extrema when d equals 2.

LEMMA 1. *Let N be the number of extrema for $X_1, \ldots, X_n$, an i.i.d. sample drawn uniformly and at random from $[0,1]^2$. Then, for integer $m > 0$,*

$$\mathbf{P}\{N < m\} \le \frac{1}{n+1} \left(\frac{e \log(n+1)}{m} \right)^m .$$

PROOF. For $d = 2$, we recall from Devroye (1988) that

$$N \overset{\mathcal{L}}{=} \sum_{i=1}^{n} Y_i ,$$

where $Y_1, \ldots, Y_n$ are i.i.d. Bernoulli random variables and $\mathbf{E}\{Y_i\} = 1/i$. This follows by first ordering the first components of the X_i's, and then noting the the corresponding second components define an extremum if and only if they correspond to records. Thus, by Chernoff's bounding method (Chernoff, 1952; Hoeffding, 1963), for $t > 0$,

$$\begin{aligned}
\mathbf{P}\{N < m\} &\le e^{tm} \mathbf{E}\left\{e^{-tN}\right\} \\
&= e^{tm} \prod_{i=1}^{n} \mathbf{E}\left\{e^{-tY_i}\right\} \\
&= e^{tm} \prod_{i=1}^{n} \left(1 - \frac{1}{i} + \frac{1}{i} e^{-t}\right) \\
&\le e^{tm} e^{-\sum_{i=1}^{n} \frac{1-e^{-t}}{i}} \\
&\le e^{tm-(1-e^{-t})\log(n+1)} \\
&= \frac{1}{n+1} \left(\frac{e \log(n+1)}{m} \right)^m \\
&\quad \text{(if we take } e^t = (\log(n+1))/m) .
\end{aligned}$$

This concludes the proof of Lemma 1. □

Appendix: auxiliary results on the number of maxima.

Let N be the number of maxima of $X_1, \ldots, X_n$, i.i.d. points uniformly distributed on $[0,1]^d$. Then

$$\mathbf{E}\{N\} = n\mathbf{E}\left\{ \left(1 - \prod_{i=1}^{d} U_i \right)^{n-1} \right\}$$

where $U_1, \ldots, U_d$ are i.i.d. uniform $[0,1]$ random variables. Setting $U = \prod_{i=1}^{d} U_i$, we note that U has density

$$f(u) = \frac{\log^{d-1}(1/u)}{(d-1)!} \ , \ 0 < u < 1 \ .$$

Thus,

$$\mathbf{E}\{N\} = n \int_0^1 (1-u)^{n-1} f(u)\, du \sim \frac{\log^{d-1} n}{(d-1)!} \ .$$

The asymptotic expression of $\mathbf{E}\{N\}$ goes back at least to Barndorff-Nielsen and Sobel (1966). One useful result is the following.

LEMMA 2 (DEVROYE, 1983). *For all $a \geq 1$, there exists a constant C depending upon a and d only such that*

$$\mathbf{E}\left\{N^a\right\} \leq C\left(\mathbf{E}\{N\}\right)^a \ .$$

LEMMA 3. *Let $X_1, \ldots, X_n$ be i.i.d. random vectors uniformly distributed on $[0,1]^d$. If N denotes the number of maxima, then there exists a universal constant C only depending upon d such that*

$$\mathrm{Var}\{N\} \leq C\mathbf{E}\{N\} \ .$$

PROOF. Let $X_1, \ldots, X_{n+1}$ be i.i.d. random vectors uniformly distributed on $[0,1]^d$. Let N_i denote the cardinality of the collection of maxima among these points if X_i is first removed. Note that $N \equiv N_{n+1}$. Let $N_{i,j}$ denote the cardinality of the collection of maxima among these points if X_i and X_j are first removed. Let $\overline{N}$ denote the average $\frac{1}{n+1}\sum_{i=1}^{n+1} N_i$. Then we have

$$\begin{aligned}
\mathrm{Var}(N_{n+1}) &\le \sum_{i=1}^{n+1} \mathbf{E}\{(N_i - \overline{N})^2\} \\
&\quad \text{(this is the Efron-Stein inequality (Efron and Stein, 1981))} \\
&= \sum_{i=1}^{n+1} \mathbf{E}\left\{\left(\frac{1}{n+1}\sum_{j\ne i}(N_j - N_i)\right)^2\right\} \\
&= \frac{1}{n+1}\mathbf{E}\left\{\left(\sum_{j=1}^{n+1}(N_j - N_1)\right)^2\right\} \\
&\le (n+1)\mathbf{E}\left\{(N_2 - N_1)^2\right\} \\
&\le 4(n+1)\mathbf{E}\left\{(N_1 - N_{1,2})^2\right\} \\
&= 4(n+1)\mathbf{E}\left\{(N_{n+1} - N_{n,n+1})^2\right\} \ .
\end{aligned}$$

Let Z_i be the indicator of the event that X_i is a maximum among $X_1, \ldots, X_{n-1}$ dominated in all components by X_n. Let W_i be an indicator of the event that X_i is a maximum. Then, for $n \ge 3$,

$$\begin{aligned}
\mathbf{E}\left\{(N_{n+1} - N_{n,n+1})^2\right\} &= \mathbf{E}\left\{W_n \times \left(1 - \sum_{i=1}^{n-1} Z_i\right)^2\right\} \\
&\le 2\mathbf{E}\{W_n\} + 2\mathbf{E}\left\{W_n\left(\sum_{i=1}^{n-1} Z_i\right)^2\right\} \\
&= 2\mathbf{E}\left\{\frac{N_{n+1}}{n}\right\} + 2\sum_{i=1}^{n-1}\mathbf{E}\{W_n Z_i\} + 2\sum_{i\ne j;1\le i,j\le n-1}\mathbf{E}\{W_n Z_i Z_j\} \\
&= 2\mathbf{E}\left\{\frac{N_{n+1}}{n}\right\} + 2(n-1)\mathbf{E}\{W_n Z_1\} + 2(n-1)(n-2)\mathbf{E}\{W_n Z_1 Z_2\} \ .
\end{aligned}$$

We compute the two expected values on the right-hand-side separately. Let X_1 have components $(1-U_1, \ldots, 1-U_d)$, where the U_i's are i.i.d. uniform $[0,1]$ random variables. Then

$$\mathbf{E}\{W_n Z_1\} = \mathbf{E}\left\{\prod_{j=1}^{d} U_j \left(1 - \prod_{j=1}^{d} U_j\right)^{n-2}\right\}$$

$$
\begin{aligned}
&= \mathbf{E}\left\{V\,(1-V)^{n-2}\right\} \\
&\quad \text{(where } V = \textstyle\prod_{j=1}^d U_j\text{)} \\
&= \int_0^1 v\,(1-v)^{n-2}\,\frac{\log^{d-1}(1/v)}{(d-1)!}\,dv \\
&= O\left(\frac{\log^{d-1} n}{n^2}\right)\ .
\end{aligned}
$$

For the product term, we also introduce the components $(1-V_1,\ldots,1-V_d)$ of X_2. It is handy to represent X_1 and X_2 slightly differently by introducing two i.i.d. uniform $[0,1]$ sequences $(m_1,\ldots,m_d)$ and $(M_1,\ldots,M_d)$, and noting that

$$
(\min(U_i,V_i),\max(U_i,V_i)) \stackrel{\mathcal{L}}{=} \left(m_i\sqrt{M_i},\sqrt{M_i}\right)\ .
$$

Then

$$
\begin{aligned}
\mathbf{E}\{W_n Z_1 Z_2\} &= \mathbf{E}\left\{\prod_{j=1}^d \min(U_j,V_j)\left(1-\prod_{j=1}^d U_j-\prod_{j=1}^d V_j+\prod_{j=1}^d \min(U_j,V_j)\right)^{n-2}\right\} \\
&\le \mathbf{E}\left\{\prod_{j=1}^d m_j\sqrt{M_j}\left(1-2\sqrt{\prod_{j=1}^d m_j\sqrt{M_j}\sqrt{M_j}}+\prod_{j=1}^d m_j\sqrt{M_j}\right)^{n-2}\right\} \\
&\quad \text{(since } a+b\ge 2\sqrt{ab} \text{ for } a,b>0\text{)} \\
&\le \mathbf{E}\left\{\prod_{j=1}^d m_j\sqrt{M_j}\left(1-\prod_{j=1}^d \sqrt{m_j}\sqrt{M_j}\right)^{n-2}\right\} \\
&= \mathbf{E}\left\{V\sqrt{W}\left(1-\sqrt{VW}\right)^{n-2}\right\} \\
&\quad \text{(where } V=\textstyle\prod_{j=1}^d m_j \text{ and } W=\prod_{j=1}^d M_j\text{)} \\
&= \mathbf{E}\left\{V\int_0^1 \sqrt{w}\left(1-\sqrt{Vw}\right)^{n-2}\frac{\log^{d-1}(1/w)}{(d-1)!}\,dw\right\} \\
&= \mathbf{E}\left\{\int_0^{\sqrt{V}} \frac{2u^2}{\sqrt{V}}\,(1-u)^{n-2}\,\frac{2\log^{d-1}(V/u^2)}{(d-1)!}\,du\right\} \\
&\le \mathbf{E}\left\{\frac{1}{\sqrt{V}}\right\}\times\int_0^1 2u^2\,(1-u)^{n-2}\,\frac{4\log^{d-1}(1/u)}{(d-1)!}\,du \\
&= O\left(\frac{\log^{d-1} n}{n^3}\right).
\end{aligned}
$$

Collecting bounds, we see that

$\mathrm{Var}\{N_{n+1}\}$

$$
\begin{aligned}
&\leq 4(n+1)\left\{\frac{2\mathbf{E}\{N_{n+1}\}}{n}+2(n-1)O\left(\frac{\log^{d-1} n}{n^2}\right)+2(n-1)(n-2)O\left(\frac{\log^{d-1} n}{n^3}\right)\right\}\\
&= O\left(\mathbf{E}\{N_{n+1}\}\right)\ .\ \square
\end{aligned}
$$

References.

O. Barndorff-Nielsen and M. Sobel, "On the distribution of the number of admissible points in a vector random sample," *Theory of Probability and its Applications*, vol. 11, pp. 249–269, 1966.

J. L. Bentley, K. L. Clarkson, and D. B. Levine, "Fast linear expected-time algorithms for computing maxima and convex hulls," in: *Proceedings of the First Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 179–187, SIAM, Philadelphia, 1990.

J. L. Bentley, K. L. Clarkson, and D. B. Levine, "Fast linear expected-time algorithms for computing maxima and convex hulls," *Algorithmica*, vol. 9, pp. 168–183, 1993.

H. Chernoff, "A measure of asymptotic efficiency of tests of a hypothesis based on the sum of observations ," *Annals of Mathematical Statistics*, vol. 23, pp. 493–507, 1952.

L. Devroye, "Moment inequalities for random variables in computational geometry," *Computing*, vol. 30, pp. 111–119, 1983.

L. Devroye, "Applications of the theory of records in the study of random trees," *Acta Informatica*, vol. 26, pp. 123–130, 1988.

L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

B. Efron and C. Stein, "The jackknife estimate of variance," *Annals of Statistics*, vol. 9, pp. 586–596, 1981.

M. Golin, "A provably-fast linear-expected-time maxima-finding algorithm," *Algorithmica*, vol. 11, pp. 501–524, 1994.

W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, pp. 13–30, 1963.