Computers Math. Applic. Vol. 25, No. 5, pp. 19–31, 1993
Printed in Great Britain.

0097-4943/93 $6.00 + 0.00

RECORDS, THE MAXIMAL LAYER, AND UNIFORM DISTRIBUTIONS IN MONOTONE SETS

LUC DEVROYE
School of Computer Science, McGill University
3480 University Street, Montreal, Canada H3A 2A7

(Received January 1992)

Abstract—Consider a nondecreasing nonnegative integrable function f on $[0,1]$. Draw an independent sample $(X_1,Y_1),\dots,(X_n,Y_n)$ of size n from the uniform distribution under f, and let N_n be the number of records in the sample, where (X_i,Y_i) is a record when, for all $j \neq i$, either $X_j > X_i$ or $Y_j < Y_i$. We study the dependence upon f of the constant C in the asymptotic formula $\mathbf{E}N_n \sim C\sqrt{n}$, and show that whenever $\int_0^1 f' > 0$, $N_n/\mathbf{E}N_n \to 1$ in probability. The results are related to the expected time analysis of algorithms for finding the collection of all records (i.e., the maximal layer).

1. INTRODUCTION

Consider a nondecreasing nonnegative function f on $[0,1]$ and let $(X_1,Y_1),\dots,(X_n,Y_n)$ be independent, identically distributed (i.i.d.) random variables uniformly distributed in the set

$$A = \{(x,y) : 0 \le x \le 1,\ 0 \le y \le f(x)\}.$$

We say that (X_i,Y_i) corresponds to a record (or is a record) if $Y_i = \max\{Y_j : X_j \le X_i\}$. Let N_n be the number of records in a sequence of length n. In this paper, we are interested in the behavior of N_n. In particular, we will obtain

- the first term in the asymptotic expansion of N_n;
- explicit inequalities related to $\operatorname{Var} N_n$ and $\mathbf{E}N_n$;
- a weak law of large numbers stating that $N_n/\mathbf{E}N_n \to 1$ in probability, as $n \to \infty$.

Some of these results require a certain smoothness on the part of f. Interestingly, the weak law of large numbers is universally valid. In the particular case $f \equiv 1$, N_n is distributed as the number of records in an i.i.d. sequence of continuous random variables, and its properties are well-known (see, e.g., [1]). Among these, we cite:

- $\mathbf{E}N_n = \sum_{i=1}^n \frac{1}{i} = \log n + \gamma + O(1)$;
- $\operatorname{Var} N_n \sim \log n$;
- $N_n/\mathbf{E}N_n \to 1$ in probability, as $n \to \infty$;
- $(N_n - \mathbf{E}N_n)/\sqrt{\operatorname{Var} N_n} \xrightarrow{\mathcal{L}} N(0,1)$ as $n \to \infty$, where $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution, and $N(0,1)$ is a standard normal random variable.

An early reference on the subject is [2]. There are also many strong convergence results available, but they do not necessarily carry over to our model because our sequence does not grow from left to right. From now on, unless mentioned otherwise, all integrals are over $[0,1]$. When f is monotonically increasing, the situation is very different, since we expect N_n to be larger than in the case $f \equiv 1$. Indeed, there is a sudden jump from $\log n$ behavior to $\sqrt{n}$ behavior as can be seen from the following result.

The author's research was sponsored by NSERC Grant A3456 and FCAR Grant 90-ER-0291.
I thank Mordecai Golin for his comments on the manuscript.

Typeset by $\mathcal{A}_\mathcal{M}\mathcal{S}$-TEX

THEOREM 1. *Let f be absolutely continuous on $[0,\alpha)$ for all $\alpha < 1$. Then*

$$\liminf_{n\to\infty} \frac{\mathbf{E}N_n}{\sqrt{n}} \geq \sqrt{\frac{\pi}{2}} \frac{\int_0^1 \sqrt{f'}}{\sqrt{\int_0^1 f}}.$$

The right-hand-side in the inequality of Theorem 1 can be ∞ for functions f with an unbounded peak as $x \uparrow 1$. Furthermore, under slight regularity conditions, equality is reached and, in fact, $N_n/\mathbf{E}N_n \to 1$ in probability. The curves for which we have no specific answer in this paper include the non-smooth curves, and those with a large infinite peak as $x \uparrow 1$. In the latter case, it is for example possible to have rates of increase that are between $\sqrt{n}$ and $O(n)$. However, we always have the following theorem.

THEOREM 2. *For any f, $\mathbf{E}N_n = O(n)$.*

The motivation for this paper is triple: first of all, the model generalizes the standard model with $f \equiv 1$ and takes a time factor into account. Second, records correspond to points on the maximal layer of the sample; the maximal layer is an object that has received some attention in computational geometry (see [3–12]). The expected size of the maximal layer in $\mathbb{R}^d$ is studied in [13–18]. Third, the properties of N_n are essential in the analysis of the expected time taken by certain algorithms for computing the maximal layer of a cloud of points.

2. TWO BASIC LEMMAS

In the remainder of the paper, we will need a fundamental fact about the differentiability of monotone functions. We have the following lemma.

LEMMA 1 (NATANSON [19]). *Let f be a bounded nondecreasing function on the real line with $\lim_{x\downarrow -\infty} f(x) = 0$, and let f' be the derivative of f wherever it exists. Then*

(a) *if S is the set of all x for which f' exists and $0 \leq f' < \infty$, then $\int_{S^c} dx = 0$, where S^c is the complement of S;*
(b) *for $x < y$, we have $\int_x^y f' \leq f(y) - f(x)$; furthermore, $\int |f'| < \infty$;*
(c) *if $f_{ac}(x) = \int_{-\infty}^x f'$ and $f_s = f - f_{ac}$, then $f'_{ac} = f'$ almost everywhere and $f'_s = 0$ almost everywhere.*

In part (c), f_{ac} represents the absolute continuous portion of f and f_s the singular portion. When f is absolutely continuous, we see that $\int_x^y f' = f(y) - f(x)$ for all $0 \leq x \leq y \leq 1$.

Let $A(x,y)$ be the set defined by $\{(u,v) : 0 \leq u < x,\ y < v \leq f(u)\}$, and $A(X,Y)$ be defined similarly provided that (X,Y) is distributed as (X_1,Y_1). Thus, (X_i,Y_i) defines a record if and only if $A(X_i,Y_i)$ contains no (X_j,Y_j) with $j \neq i$. We have the following lemma.

LEMMA 2. *N_n is distributed as*

$$\sum_{i=1}^n I_{A(X_i,Y_i) \text{ contains no data point}}.$$

Also,

$$\begin{aligned}\mathbf{E}N_n &= n\,\mathbf{P}\{A(X_1,Y_1) \text{ is empty}\}\\ &= n\,\mathbf{E}(1-\mu(A(X,Y))^{n-1})\\ &\leq n\,\mathbf{E}\,e^{-(n-1)\,\mu(A(X,Y))},\end{aligned}$$

where μ is the uniform probability measure on $A(1,0)$.

3. SOME PROOFS AND REMARKS

Note first that $\int_0^1 f(x)\,dx < \infty$, for otherwise, we would not be able to define a uniform distribution on A in view of the fact that $\int_A dx\,dy = \int_0^1 f(x)\,dx$. We will call the given area of A λ. It is possible to have $\int \sqrt{f'} = \infty$. In that case, Theorem 1 shows that

$$\frac{\mathbf{E}N_n}{\sqrt{n}} \to \infty.$$

An example is furnished by $f(x) = 1/(1-x)\log^{1+\delta}(1/(1-x))$ for $\delta > 0$. Such functions necessarily have an infinite peak at one. However, there are functions with an infinite peak for which $\int \sqrt{f'} < \infty$.

It should also be noted that for purely singular functions, we have $\int \sqrt{f'} = 0$ since $f' = 0$ almost everywhere. For such functions, $\mathbf{E}N_n$ can tend to ∞ at any rate between $\log n$ and $o(n)$. The functions with an infinite peak can attain any rate between $\sqrt{n}$ and $o(n)$.

PROOF OF THEOREM 2. From Lemma 2, we note that

$$\frac{\mathbf{E}N_n}{n} = \mathbf{E}\left\{e^{-(n-1)\,\mu(A(X,Y))}\right\},$$

and this tends to 0 by the Lebesgue dominated convergence theorem if for almost all $x(\mu)$ we have $\mu(A(x,y)) > 0$. Let S be the subset of A consisting of all (x,y) with $\mu(A(x,y)) = 0$. We will show that $\mu(S) = 0$. Observe that $S = \bigcap_n S_n$, where $S_n = \{(x,y) : \mu(A(x,y)) < 1/n\}$. Thus, $\mu(S) = \lim_{n\to\infty} \mu(S_n)$. It suffices to show that the given limit is zero. Take a constant D such that $\mu(A \cap \{y : y > D\}) < \epsilon$, for a given small ϵ. Partition the entire plane into a regular grid with grid size $1/\sqrt{n}$. Thus, each cell completely contained in A has μ measure $1/n$. In each row, mark the leftmost such cell, and in each column, mark the topmost such cell. Mark also all cells whose intersection with the complement of A and with A are both non-empty. It is easy to see that S_n is included in the marked cells. The number of marked cells with nonempty intersection with $A \cap \{y : y \le D\}$ does not exceed $3\,(1+D)\sqrt{n}$. Thus,

$$\mu(S_n) \le \epsilon + \frac{1}{n}\,(3+3D)\sqrt{n} = \epsilon + o(1).$$

This concludes the proof of Theorem 2. ∎

4. LOWER BOUNDS FOR THE EXPECTED NUMBER OF RECORDS

THEOREM 3. *For all monotone f on $[0,1]$, we have*

$$\liminf_{n\to\infty} \frac{\mathbf{E}N_n}{\sqrt{n}} \ge \sqrt{\frac{\pi}{2}}\,\frac{\int\sqrt{f'}}{\sqrt{\int f}}.$$

PROOF. From Lemma 2,

$$\mathbf{E}N_{n+1} = (n+1)\int_0^1 \int_0^{f(x)} \frac{1}{\lambda}(1-\mu(A(x,y)))^n\,dy\,dx.$$

By Fatou's lemma, we are done if we can show that for almost all x with $f'(x) > 0$,

$$\liminf_{n\to\infty} \sqrt{n+1}\int_0^{f(x)} (1-\mu(A(x,y)))^n\,dy \ge \sqrt{\frac{\pi}{2}}\,\sqrt{f'(x)}\,\sqrt{\lambda}.$$

By Lemma 1, it suffices to consider only those $x \in (0,1)$ for which f' exists and is finite and positive. Let $0 < \epsilon < f'(x)$ be arbitrary. Then, by the existence of $f'(x)$, we have for all δ smaller than some $\Delta = \Delta(x,\epsilon)$,

$$f(x) - \delta(f'(x)-\epsilon) \le f(x-\delta) \le f(x) + \delta(f'(x)+\epsilon).$$

Uniformly over all $y < f(x)$ with $f(x) - y \le \Delta\,(f'(x) - \epsilon)$, we have

$$\mu(A(x,y)) \le \frac{(f(x)-y)^2}{2\lambda\,(f'(x)-\epsilon)}.$$

So,

$$\begin{aligned}
\int_0^{f(x)} (1-\mu(A(x,y))^n\,dy &\ge \int_{\max(0,f(x)-\Delta(f'(x)-\epsilon))}^{f(x)} \left(1-\frac{(f(x)-y)^2}{2\lambda\,(f'(x)-\epsilon)}\right)^n dy \\
&= \int_0^{\Delta(x,\epsilon)\sqrt{f'(x)-\epsilon}\sqrt{n\lambda}} \left(1-\frac{z^2}{2n}\right)^n \sqrt{f'(x)-\epsilon}\,\sqrt{\lambda}\,dz/\sqrt{n} \\
&\sim \sqrt{f'(x)-\epsilon}\int_0^\infty e^{-z^2/2}\sqrt{\lambda}\,dz/\sqrt{n} \\
&= \sqrt{\frac{\pi}{2}}\sqrt{f'(x)-\epsilon}\,\sqrt{\lambda}/\sqrt{n}.
\end{aligned}$$

Since we can choose ϵ arbitrarily small, we have shown the Theorem. ∎

THEOREM 4. *For all monotone f on $[0,1]$, we have*

$$\mathbf{E}N_n \ge H_n \stackrel{\text{def}}{=} 1 + \cdots + \frac{1}{n} = \log n + \gamma + O(1),$$

where $\gamma = 0.55\ldots$ is Euler's constant.

PROOF. Let M be a very large constant to be picked later. Replace all Y_i's by

$$Y_i' = \frac{M}{f(X_i)}Y_i, \qquad 1 \le i \le n.$$

It is easy to verify that Y_i' is independent of X_i for all i. Furthermore, Y_i' is uniformly distributed on $[0,M]$. Also, if $f(X_i) \le M$ and (X_i, Y_i') is a maximal layer point in the collection $(X_1,Y_1'),\ldots,(X_n,Y_n')$, then (X_i,Y_i) is a maximal layer point in the original data set. If $f(X_i) > M$, this statement may not be true. Let N_n and N_n' be the cardinalities of the maximal layers with the original and the transformed data, respectively. Then

$$N_n \ge N_n' - \sum_{i=1}^n I_{[f(X_i)>M]}.$$

Taking expectations, and letting M tend to infinity shows that

$$\mathbf{E}N_n \ge \liminf_{M\to\infty} \mathbf{E}N_n'.$$

But when X_i is independent of Y_i', the number of maximal layer points N_n' satisfies [13]

$$\mathbf{E}N_n' = H_n.$$ ∎

THEOREM 5. *For all monotone f on $[0,1]$, we have*

$$\liminf_{n\to\infty} \frac{\mathbf{E}N_n}{\log n} \ge A_0,$$

where $A_0 = \lim_{\delta\downarrow 0} A_\delta$ and

$$A_\delta = \#x : x \in [0,1), \quad \lim_{y\downarrow x} f(y) > \lim_{y\uparrow x} f(y) + \delta.$$

(Note: A_0 is the number of points of increase of f.)

Proof. Fix $\delta > 0$. Let $0 \le x_1 < x_2 < \cdots < x_{A_\delta}$ be the points from $[0,1)$ counted in A_δ. Let

$$p_i = \int_{x_i}^{x_{i+1}} f,$$

where by definition, $x_{A_\delta+1} = 1$. On the interval $[x_i, x_{i+1})$, Define

$$b_i = \lim_{y \downarrow x_i} f(y), \quad a_i = \lim_{y \uparrow x_i} f(y).$$

For every $X_j \in [x_i, x_{i+1})$, replace Y_j by

$$Y_j' = \frac{b_i}{f(X_j)} Y_j.$$

Note that every such Y_j' is independent of X_j and uniformly distributed on $[0, b_i]$. In addition, we add A_δ new data points at the positions (x_i, a_i), $1 \le i \le A_\delta$. The crucial observation is that the number of maximal layer points in the new data set is not greater than $N_n + A_\delta$. Among (X_j, Y_j'), the number of points (say, M_i) falling in $[x_i, x_{i+1}) \times [a_i, b_i]$ is binomially distributed with parameters n and

$$q_i = p_i \times \frac{b_i - a_i}{b_i}.$$

When we condition on M_i, the expected number of maximal layer points among the M_i points just marked is at least $H_{M_i} \ge \log(1 + M_i)$ (Theorem 4). Thus,

$$\begin{aligned} \mathbf{E}N_n &\ge \sum_{i=1}^{n} \mathbf{E}\log(1 + M_i) + A_\delta - A_\delta \\ &\ge \sum_{i=1}^{A_\delta} \log(1 + \mathbf{E}M_i) \quad \text{(Jensen's inequality)} \\ &= \sum_{i=1}^{A_\delta} \log(1 + nq_i) \\ &\sim A_\delta \log n. \end{aligned}$$

Now let δ tend to zero. ∎

The lower bounds collected thus far show that there are two processes at work. The smooth increases as measured by f' contribute to the coefficient of a $\sqrt{n}$ term. Abrupt points of increase of f are counted in a coefficient of a $\log n$ term. The latter phenomenon is somehow similar to one seen in the study of the expected number of maximal layer points for uniform distributions in a staircase-shaped polygon with k steps, which grows as a constant times $k \log n$ (for a similar result for the number of convex hull points in a k-gon, see [20]). We note that in Theorem 5, A_δ can be ∞. In such cases, we can attain any rate of convergence between $\log n$ and $o(n)$. This is slightly annoying, since these counterexamples can be chosen in such a way that there are a countable number of points of increase of f, and $f' = 0$ elsewhere, as in infinite staircases. Thus, the lower bound of Theorem 3 cannot possibly have a similar-looking universal upper bound, except perhaps under smoothness conditions on f.

A Class of Counterexamples

Partition $[0,1]$ into intervals of length $1/2^i$, $i = 1, 2, \ldots$, the smaller intervals to the right of larger intervals. On the i^{th} interval, let f take the value $a_i = 2^i c_i$, where c_i is to be specified. We want the integral under f to be one, so we require

$$\sum_{i=1}^{\infty} c_i = 1.$$

We also require that $c_i \ge \frac{3}{4} c_{i-1}$ so that $a_i \uparrow$. Observe that the i^{th} interval contributes at least one maximal layer point if at least one j exists for which X_j falls in the i^{th} interval, and $Y_j \in (a_{i-1}, a_i]$ $(a_0 = 0)$. The probability of this event is

$$\begin{aligned} 1 - (1 - (a_i - a_{i-1})2^{-i})^n &= 1 - (1 - (c_i - \frac{1}{2} c_{i-1}))^n \\ &\ge 1 - (1 - \frac{1}{3} c_i)^n \\ &\ge 1 - e^{-1}, \qquad \text{if } c_i \ge \frac{3}{n}. \end{aligned}$$

Take $c_i \sim i^{-\rho}$ for some $\rho > 1$. Then the number of i's for which $c_i \ge 3/n$ grows at least as $(n/3)^{1/\rho}$. Thus,

$$\mathbf{E}N_n \ge (1 - e^{-1})(1 + o(1)) \left(\frac{n}{3}\right)^{1/\rho}.$$

This can be pushed as close to $o(n)$ as desired by decreasing ρ to 1. Observe that f' is identically zero except at the borders of the intervals. Also, f necessarily exhibits an infinite peak at one. ∎

5. UPPER BOUND AND EXACT ASYMPTOTICS

THEOREM 6. *Assume that f is monotone on $[0,1]$, and that either f is concave, convex, or Lipschitz (C), i.e., we have $|f(x) - f(y)| \le C|x - y|$ for all $x, y \in [0,1]$. Then*

$$\limsup_{n\to\infty} \frac{\mathbf{E}N_n}{\sqrt{n}} \le \sqrt{\frac{\pi}{2}} \frac{\int_0^1 \sqrt{f'}}{\sqrt{\int_0^1 f}}.$$

In fact,

$$\lim_{n\to\infty} \frac{\mathbf{E}N_n}{\sqrt{n}} = \sqrt{\frac{\pi}{2}} \frac{\int_0^1 \sqrt{f'}}{\sqrt{\int_0^1 f}}.$$

PROOF. The second part of the theorem is a corollary of Theorem 1 and the first part of this theorem.

Extend f on the real line by setting $f(x) = 0$ for $x < 0$ and $f(x) = f(1)$ for $x \ge 1$. Assume first that f is Lipschitz on the entire real line. Define $\lambda = \int_0^1 f$, let μ be the uniform probability measure defined in the proof of Theorem 1, and let $A_{x,y}$ be the collection of all (z, w) with $0 \le z < x$ and $f(z) \ge w > y$. Then

$$\begin{aligned} \mathbf{E}N_{n+1} &= (n+1) \int_0^1 \lambda^{-1} \int_0^{f(x)} (1 - \mu(A_{x,y}))^n \, dy \, dx \\ &\le (n+1) \int_0^1 \lambda^{-1} \int_0^{f(x)} \exp(-n\mu(A_{x,y})) \, dy \, dx \\ &\overset{\text{def}}{=} (n+1) \int_0^1 n^{-1/2} g_n(x) \, dx. \end{aligned}$$

If f is Lipschitz with constant C, then

$$\mu(A_{x,y}) \ge \frac{(f(x) - y)^2}{2C\lambda}.$$

Therefore,

$$\begin{aligned} g_n(x) &= \sqrt{n}\, \lambda^{-1} \int_0^{f(x)} \exp(-n\mu(A_{x,y})) \, dy \\ &\le \sqrt{n}\, \lambda^{-1} \int_0^{f(x)} \exp\left(\frac{-n(f(x) - y)^2}{2C\lambda}\right) dy \\ &\le \sqrt{\frac{\pi}{2}} \sqrt{C}. \end{aligned}$$

Thus, for all x, g_n remains bounded. Observe that

$$\frac{\mathbf{E}N_{n+1}}{\sqrt{n+1}} \le \sqrt{\frac{n+1}{n}} \int_0^1 g_n(x)\,dx$$

and that g_n is uniformly bounded over $[0,1]$. Thus, by Fatou's lemma, Theorem 6 follows for the Lipschitz functions if we can show that for almost all x, $\limsup_{n\to\infty} g_n(x) \le \sqrt{\pi f'(x)/2\lambda}$. Consider first x such that $f(x) > 0$ and $f'(x) > 0$. For fixed $\epsilon \in (0,1)$, find $\delta > 0$ such that

$$\mu(A_{x,y}) \ge (1-\epsilon)\,\frac{(f(x)-y)^2}{2f'(x)\,\lambda}$$

whenever $y > f(x) - \delta > 0$. Thus,

$$\begin{aligned} g_n(x) &\le \frac{\sqrt{n}}{\lambda}\int_0^{f(x)-\delta} \exp\left(-\frac{n\,(1-\epsilon)\,\delta^2}{2f'(x)\,\lambda}\right) dy + \frac{\sqrt{n}}{\lambda}\int_{f(x)-\delta}^{f(x)} \exp\left(-\frac{n\,(1-\epsilon)(f(x)-y)^2}{2f'(x)\,\lambda}\right) dy \\ &\le o(1) + \sqrt{\frac{\pi}{2}}\sqrt{\frac{f'(x)}{\lambda\,(1-\epsilon)}}. \end{aligned}$$

By the arbitrariness of ϵ, the desired result follows for this first case. The case $f(x) = 0$ can be discarded straight away. Finally, we consider $f(x) > 0$ and $f'(x) = 0$. For fixed $\epsilon \in (0,1)$, find $\delta > 0$ such that

$$\mu(A_{x,y}) \ge \frac{(f(x)-y)^2}{2\epsilon\,\lambda},$$

whenever $y > f(x) - \delta > 0$. Thus,

$$\begin{aligned} g_n(x) &\le \frac{\sqrt{n}}{\lambda}\int_0^{f(x)-\delta} \exp\left(-\frac{n\,\delta^2}{2\epsilon\,\lambda}\right) dy + \frac{\sqrt{n}}{\lambda}\int_{f(x)-\delta}^{f(x)} \exp\left(-\frac{n\,(f(x)-y)^2}{2\epsilon\,\lambda}\right) dy \\ &\le o(1) + \sqrt{\frac{\pi}{2}}\sqrt{\frac{\epsilon}{\lambda}}. \end{aligned}$$

This is as small as desired by our choice of ϵ, which once again establishes the sought limiting result. This concludes the proof if f is Lipschitz on the real line.

When f is merely Lipschitz on $[0,1]$, then it may have a jump at the origin. Assume thus that $f(0) > 0$. Then, partition the data points into sets I and J, where I collects all indices j between 1 and n such that $Y_j \le f(0)$. J captures the remainder of these indices. The cardinalities of I and J are denoted by $|I|$ and $|J|$. The number of maximal layer points among the points with index in $I(J)$ is denoted by $N_I(N_J)$. From the fundamental properties of records, we have

$$\mathbf{E}N_I = \mathbf{E}\left(\sum_{i=1}^{|I|} \frac{1}{i}\, I_{|I|>0}\right) \le \sum_{i=1}^{n} \frac{1}{i} \le 1 + \log n.$$

Also, we have simple subadditivity: $N_n \le N_I + N_J$. Hence,

$$\limsup_{n\to\infty} \frac{\mathbf{E}N_n}{\sqrt{n}} \le \limsup_{n\to\infty} \frac{\mathbf{E}N_J}{\sqrt{n}}.$$

From the first part of the theorem, we see that for every $\epsilon > 0$, there exists a constant M such that

$$\mathbf{E}\,(N_J \mid |J|) \le (1+\epsilon)\sqrt{\frac{\pi}{2}}\,\frac{\int\sqrt{f'}}{\sqrt{\int(f-f(0))}}\,\sqrt{|J|}\,I_{|J|\ge M} + |J|\,I_{|J|<M},$$

and thus, using

$$\mathbf{E}\sqrt{|J|} \le \sqrt{\mathbf{E}|J|} = \sqrt{\frac{n\int(f-f(0))}{\int f}},$$

we obtain

$$\mathbf{E}N_J \le (1+\epsilon)\sqrt{\frac{\pi n}{2}}\frac{\int\sqrt{f'}}{\sqrt{\int f}} + M.$$

If $\int\sqrt{f'} > 0$, the result of the theorem follows from this and the arbitrary nature of ϵ. Consider the last case: $f' = 0$ almost everywhere, and f is Lipschitz (C). Clearly, f is absolutely continuous on $[0,1]$, and therefore, by Lemma 1, f is constant on $[0,1]$. For this distribution, $\mathbf{E}N_n/\log n \to 1$, so that $\mathbf{E}N_n/\sqrt{n} \to 0$ as required.

Next, consider f concave and nondecreasing on $[0,1]$. Then we note that except at the origin, f is necessarily Lipschitz with some constant C. Assume first that $f(x) \downarrow 0$ as $x \downarrow 0$. Again, consider subsets I and J of indices between 1 and n with $i \in I$ if $X_i > x$ and $i \in J$ if $X_i \le x$, where x is picked so small that $f(x)/\int f < \delta$ for a prespecified small $\delta > 0$. By Theorem 11 below, we see that there are universal constants $c, c' > 0$ such that

$$\begin{aligned}\mathbf{E}N_J &\le c + c'\,\mathbf{E}\sqrt{\frac{f(x)\,|J|}{\int_0^x f}} \le c + c'\sqrt{\frac{f(x)\,\mathbf{E}|J|}{\int_0^x f}} \\ &= c + c'\sqrt{\frac{n\,f(x)}{\int_0^1 f}} \le c + c'\sqrt{n\,\delta}.\end{aligned}$$

By the arbitrary nature of δ and the subadditivity of N_n (i.e., $N_n \le N_I + N_J$), we see that the result of Theorem 6 follows if it is true for N_I. For this purpose, take $\epsilon > 0$. We note that it is impossible that $\int\sqrt{f'} = 0$, for this would force $f \equiv \lim_{x\downarrow 0} f(x) = 0$, thus making $\int f = 0$. By the first part of the theorem,

$$\mathbf{E}\,(N_I \mid |J|) \le (1+\epsilon)\sqrt{\frac{\pi}{2}}\frac{\int_x^1\sqrt{f'}}{\sqrt{\int_x^1 f}}\sqrt{|I|}\,I_{|I|\ge M} + |I|\,I_{|I|<M},$$

where M depends upon ϵ only. Unconditioning and using Jensen's inequality as done above shows that

$$\mathbf{E}N_I \le M + (1+\epsilon)\sqrt{\frac{\pi n}{2}}\frac{\int_x^1\sqrt{f'}}{\sqrt{\int_0^1 f}} \le M + (1+\epsilon)\sqrt{\frac{\pi n}{2}}\frac{\int_0^1\sqrt{f'}}{\sqrt{\int_0^1 f}},$$

from which the desired result follows once again. Assume next that $f(x) \to c > 0$ as $x \downarrow 0$. If $f' > 0$ on a set of positive measure, then we argue as above using the subadditivity trick again to deduce the result. Finally, if $f' = 0$ almost everywhere, then the concavity implies that $f \equiv c'$ for some constant $c' > 0$. In this case, we know that $\mathbf{E}N_n = o(\log n)$, so that the result is once again verified. This concludes the proof for concave functions f.

The theorem is also easy to verify for bounded convex functions on $[0,1]$. The proof is not given here. ∎

Remark. Theorem 6 remains valid for functions f on $[0,1]$ that are nondecreasing, and either concave or convex on each of a finite number of intervals into which $[0,1]$ can be partitioned.

6. VARIANCE

In this section, we consider general monotone f on $[0,1]$. The purpose is to prove the following inequality.

Theorem 8. *For all monotone f on $[0,1]$ and all $n > 1$,*

$$\mathrm{Var}\{N_n\} \le (5 + 2\log n)\,\mathbf{E}N_n + \frac{n}{n-1}.$$

Proof. We introduce A_i, the set of all x, y for which $y \le f(x)$, yet $x < X_i$ and $y > Y_i$. I_i is the indicator function of the event that A_i does not capture any of the (X_j, Y_j), $1 \le j \le n, j \ne i$. Define

$$p_i \stackrel{\text{def}}{=} \mathbf{E} I_i = \mathbf{E}(1-\mu(A_1))^{n-1},$$

where μ is the uniform probability measure on $\{(x,y) : 0 \le x \le 1, 0 \le y \le f(x)\}$. By the symmetry in our problem,

$$\begin{aligned}
\mathrm{Var}\{N_n\} = \mathbf{E}N_n^2 - \mathbf{E}^2 N_n &= \mathbf{E}\left(\sum_{i=1}^n I_i + \sum_{j\ne i} I_i I_j - \sum_{i=1}^n p_i^2 - \sum_{j \ne i} p_i p_j\right) \\
&= \sum_{i=1}^n p_i(1-p_i) + \sum_{j\ne i}(\mathbf{E} I_i I_j - p_i p_j) \\
&\le \mathbf{E}N_n + n(n-1)(\mathbf{E}I_1 I_2 - p_1 p_2).
\end{aligned}$$

Assume that $n \ge 2$, and let B denote the event that (X_2, Y_2) is a maximal layer point among those points with $X_2 > X_1, Y_1 < Y_2 \le f(X_1)$. Let C denote the event that (X_2, Y_2) is a maximal layer point among those points with $X_2 > X_1, Y_2 > f(X_1)$. Let S_1 be the rectangle defined by $Y_1 < y \le f(X_1)$, and $X_1 \le x \le 1$, and let $|S_1|$ be the number of (X_i, Y_i) pairs in S_1: conditional on (X_1, Y_1), this is binomially distributed with parameters $n-1$ and $\mu(S_1)$. We will also need the inequality

$$\mathbf{E}N_{n-1} \le 1 + \frac{n-1}{n}\mathbf{E}N_n,$$

which follows from

$$\mathbf{E}N_{n-1} - \mathbf{E}N_n = n\,\mathbf{E}\mu(A_1)(1-\mu(A_1))^{n-2} - \mathbf{E}(1-\mu(A_1))^{n-2} \le \frac{n}{n-1} - \frac{\mathbf{E}N_{n-1}}{n-1}.$$

By symmetry,

$$\begin{aligned}
\mathbf{E}I_1 I_2 &= 2\mathbf{E}I_1 I_2 I_{X_1<X_2} \\
&= 2\mathbf{E}I_1 I_2 I_{X_1<X_2} I_{Y_2>f(X_1)} + 2\mathbf{E}I_1 I_2 I_{X_1<X_2} I_{Y_1<Y_2\le f(X_1)} \\
&= 2\mathbf{E}(1-\mu(A_1)-\mu(A_2))^{n-2} I_{X_1<X_2} I_{Y_2>f(X_1)} + \frac{2}{n-1}\mathbf{E}\left(I_1 \sum_{j=2}^n I_j I_{(X_j,Y_j)\in S_1}\right) \\
&\le \mathbf{E}(1-\mu(A_1))^{n-2}(1-\mu(A_2))^{n-2} + \frac{2}{n-1}\mathbf{E}\left(I_1(1+\log|S_1|)\right) \\
&\le \mathbf{E}(1-\mu(A_1))^{n-2}(1-\mu(A_2))^{n-2} + \frac{2}{n-1}(1+\log n)p_1 \\
&= \left(\frac{\mathbf{E}N_{n-1}}{n-1}\right)^2 + \frac{(2+2\log n)\mathbf{E}N_n}{n(n-1)} \\
&\le \left(\frac{1}{n-1} + \frac{\mathbf{E}N_n}{n}\right)^2 + \frac{(2+2\log n)\mathbf{E}N_n}{n(n-1)} \\
&= \left(\frac{1}{n-1}\right)^2 + p_1 p_2 + \frac{(4+2\log n)\mathbf{E}N_n}{n(n-1)}.
\end{aligned}$$

We used the fact that the expected number of maximal layer points for an i.i.d. sample of size k drawn from a uniform distribution on the unit square is $\sum_{i=1}^k \frac{1}{i} \le 1 + \log k$. Combining these bounds shows that

$$\mathrm{Var}\{N_n\} \le (5 + 2\log n)\mathbf{E}N_n + \frac{n}{n-1}. \qquad \blacksquare$$

7. WEAK LAW OF LARGE NUMBERS

THEOREM 9. *For any monotone f on $[0,1]$ for which $\mathbf{E}N_n/\log n \to \infty$, we have*

$$\frac{N_n}{\mathbf{E}N_n} \to 1 \quad \textit{in probability}$$

as $n \to \infty$. In particular, this is true whenever $\int_0^1 f'(x)\,dx > 0$.

Theorem 1 follows directly from Theorems 3 and 8 and Chebyshev's inequality. In the vast majority of the cases, we have $\liminf_{n\to\infty} \mathbf{E}N_n/\sqrt{n} > 0$, so that the weak law of large numbers indeed applies. This is important to know, since it means that the expected value is a good indicator of the size of the maximal layer. It also means that the actual asymptotic value of second order quantities such as the variance of N_n is less important, except perhaps in situations in which one wants to construct some kind of statistical test or confidence interval.

8. A GENERAL UPPER BOUND

The upper bounds provided so far assumed a certain smoothness on the part of f. Basically, we have treated functions that consist of a finite number of Lipschitz, convex, or concave pieces. Without any smoothness conditions, it remains nevertheless possible to bound $\mathbf{E}N_n$ from above in useful manners. The following lemma will be useful.

LEMMA 10. *Let f be a nondecreasing function on $[0,1]$ which remains bounded, and let $p > 0$ be a given number. Let S_p be the collection of all (x,y) such that*

$$\lambda(A(x,y)) \le p,$$

where $\lambda(\cdot)$ denotes Lebesgue measure. Then

$$\lambda(S_p) \le 2p + 3\sqrt{p f(1)}.$$

PROOF. Extend f on $(1,\infty)$ by defining $f \equiv f(1)$ there. Extend the definition of $A(x,y)$ to include all x on $(0,\infty)$. Let L_p be the collection of all (x,y) for which $\lambda(A(x,y)) = p$. Construct a cover of S_p as follows. Begin with any point on L_{2p} having $y = 0$. Call this (x_1, y_1). Find any point on L_p with x-coordinate x_1, and call its y-coordinate y_2. Repeat this process by finding (x_2, y_2) on L_{2p} and so forth. Stop the process after k steps where k is the first index for which $x_k \ge 1$. (It is possible that k is infinite.) Define (x_0, y_0) as any point on L_p with $y_0 = 0$. Clearly, the sets $A(x_i, y_i)$, $1 \le i \le k$, cover S_p. Thus,

$$S_p \subseteq \bigcup_{i=1}^{k} A(x_i, y_i).$$

To describe things a bit better, we need to introduce sets $A_i = A(x_i, y_i)$, $B_i = A(x_{i-1}, y_i)$, $C_i = B_i \cap B_{i+1}$, $D_i = A_i - (B_i \cup B_{i+1})$, and $E_i = B_{i+1} - (B_i \cup B_{i+2})$. We note the following. Clearly, $B_i \cap B_{i+1} \subseteq A_i$. Also, B_i has a vertex on L_p, and C_i is not empty. If C_i were empty, then we would have

$$\lambda(A_i) \ge \lambda(B_i) + \lambda(B_{i+1}) > 2p,$$

which is a contradiction. Thus, D_i and E_i are both rectangles. Furthermore,

$$S_p \subseteq A_0 \cup A_k \cup \bigcup_{i=1}^{k-1} (D_i \cup E_i \cup C_i).$$

We will use this inclusion in a nonoptimal manner since some C_i's may overlap:

$$\lambda(S_p) \le 2p + \sum_{i=1}^{k-1} (\lambda(D_i) + \lambda(E_i) + \lambda(C_i)).$$

But since $\lambda(B_i) = \lambda(B_{i+1}) = p$, we see that $\lambda(D_i) = \lambda(C_i)$. Let D_i have dimensions (a_i, b_i). Then E_i must have dimensions (a_i, b_{i+1}). It is clear that

$$\sum_{i=1}^{k-1} a_i \le 1, \quad \sum_{i=1}^{k-1} b_i \le F \stackrel{\text{def}}{=} f(1), \quad \sum_{i=2}^{k} b_i \le F \stackrel{\text{def}}{=} f(1).$$

Furthermore, both $a_i\, b_i$ and $a_i\, b_{i+1}$ are less than or equal to p. Thus, $\lambda(S_p)$ does not exceed

$$2p + \sum_{i=1}^{k-1} a_i\,(2b_i + b_{i+1}),$$

subject to the abovementioned constraints. We will maximize

$$\sum_{i=1}^{k-1} a_i\, b_i.$$

We have

$$\begin{aligned}
\sum_{i=1}^{k-1} a_i\, b_i &= \sum_{i=1}^{k-1} \sqrt{a_i}\, b_i\, (\sqrt{a_i}) \le \sqrt{\sum_{i=1}^{k-1} a_i\, b_i^2 \sum_{i=1}^{k-1} a_i} \quad \text{(by the Cauchy-Schwarz inequality)} \\
&\le \sqrt{\sum_{i=1}^{k-1} p\, b_i \sum_{i=1}^{k-1} a_i} \quad \text{(because } a_i\, b_i \le p\text{)} \\
&\le \sqrt{F\, p}.
\end{aligned}$$

A similar argument remains valid if the b_i's are replaced by b_{i+1}. Thus,

$$\lambda(S_p) \le 2p + 3\sqrt{p\, F}. \qquad \blacksquare$$

The following theorem partially overlaps with results in [16,21].

THEOREM 11. *For any bounded nondecreasing f on $[0,1]$,*

$$\mathbf{E}N_n \le 6 + 8\sqrt{\frac{n\, f(1)}{\int f}}.$$

PROOF. We construct nested functions f_i as follows: let $f_0 = f$. In the notation of Lemma 10, we remove from the area under f the set S_p, where $p > 0$ is to be chosen further on. For every $x \in (0,1)$, define $f_1(x) = \inf\{y : y \in S_p\}$. From f_1, we construct f_2 in a similar manner, and so forth. Let M_i be the number of points (X_j, Y_j) for which $f_{i+1}(X_j) < Y_j \le f_i(X_j)$. Consider a point (X_j, Y_j) counted in M_i. Clearly, $\lambda(A(X_j, Y_j)) \ge ip$ by construction. Thus, the point is a maximal layer point with probability at most $(1 - ip/\int_0^1 f)^{n-1}$. Thus, if we define $\lambda = \int_0^1 f$ and $q = (2p + 3\sqrt{p\, f(1)})/\lambda$,

$$\begin{aligned}
\mathbf{E}N_n &\le \mathbf{E}M_0 + \left(1 - \frac{p}{\lambda}\right)^{n-1} \mathbf{E}M_1 + \left(1 - \frac{2p}{\lambda}\right)^{n-1} \mathbf{E}M_2 + \left(1 - \frac{3p}{\lambda}\right)^{n-1} \mathbf{E}M_3 + \cdots \\
&\le nq + \left(1 - \frac{p}{\lambda}\right)^{n-1} nq + \left(1 - \frac{2p}{\lambda}\right)^{n-1} nq + \cdots \le nq \sum_{i=0}^{\infty} e^{-ip/\lambda(n-1)} \\
&= \frac{nq}{1 - e^{-p/\lambda(n-1)}}.
\end{aligned}$$

Take $p = \lambda/(2(n-1))$. Then, assuming $n > 1$,

$$\begin{aligned}\mathbf{E}N_n &\le \frac{n}{(n-1)(1-\sqrt{1/e})} + \frac{3n\sqrt{f(1)/2\lambda}}{\sqrt{n-1}(1-\sqrt{1/e})} \\ &\le \frac{2}{(1-\sqrt{1/e})} + \frac{3}{(1-\sqrt{1/e})}\sqrt{n\, f(1)/\lambda}. \quad \blacksquare\end{aligned}$$

Observe that the upper bound is scale-free, and does not require any smoothness assumptions regarding f. Theorem 11 provides the main tools to the problem of bounding $\mathbf{E}N_n$ whenever f has an infinite peak at one. A typical result, one of many possible such results, is the following.

THEOREM 12. *Assume that f is nondecreasing on $[0,1]$ and that $\int f^{1+a} < \infty$ for some $a > 0$. Then*

$$\mathbf{E}N_n \le c + c'\left(\int \left(\frac{nf}{\lambda}\right)^{1+a}\right)^{1/(2a+1)}. \quad \blacksquare$$

for some universal constants c, c'.

PROOF. Again, we will use the subadditivity property of the proof of Theorem 6. Let I and J be the collections of indices of points with $Y_i \le r$ and $Y_i > r$ respectively, where $r > 0$ is a constant to be chosen further on. Clearly,

$$\mathbf{E}N_J \le \mathbf{E}|J| = \frac{n}{\lambda}\int (f - \min(f,r)) \le n\int \frac{f^{1+a}}{r^a\lambda}$$

by Chebyshev's inequality. Furthermore, by Theorem 11,

$$\begin{aligned}\mathbf{E}N_I &\le 6 + 8\,\mathbf{E}\sqrt{|I|}\sqrt{\frac{r}{\int \min(f,r)}} \le 6 + 8\sqrt{\frac{n\int \min(f,r)}{\lambda}}\sqrt{\frac{r}{\int \min(f,r)}} \\ &= 6 + 8\sqrt{\frac{nr}{\lambda}}.\end{aligned}$$

Since $\mathbf{E}N_n \le \mathbf{E}N_I + \mathbf{E}N_J$, we obtain an upper bound that is a function of r, n and f. Minimization with respect to r shows that we should take r proportional to

$$\left(\frac{n}{\lambda}\right)^{1/(2a+1)}\left(\int f^{1+a}\right)^{2/(2a+1)}. \quad \blacksquare$$

REMARK. The upper bound of Theorem 12 is again scale-free, as it should be. Note that $\int f^{1+a}$ measures the peakedness of f, while the upper bound can vary from $o(\sqrt{n})$ to $o(n)$.

REFERENCES

1. N. Glick, Breaking records and breaking boards, *American Mathematical Monthly* **85**, 2–26 (1978).
2. A. Rényi, Théorie des éléments saillants d'une suite d'observations, In *Colloquium on Combinatorial Methods in Probability Theory*, pp. 104–115, Mathematisk Institut, Aarhus Universitet, Denmark, (1962).
3. H.T. Kung, F. Luccio and F.P. Preparata, On finding the maxima of a set of vectors, *Journal of the ACM* **22**, 469–476 (1975).
4. J.L. Bentley and M.I. Shamos, Divide and conquer for linear expected time, *Information Processing Letters* **7**, 87–91 (1978).
5. 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**, 536–543 (1982).
6. H.N. Gabow, J.L. Bentley and R.E. Tarjan, Scaling and related techniques for geometry problems, In *Proceedings of the* 16th *Annual Symposium on the Theory of Computing*, pp. 135–143, (1984).
7. M. Machii and Y. Igarashi, A hashing method of finding the maxima of a set of vectors, Technical Report CS-84-2, Department of Computer Science, Gunma University, Gunma, Japan, (1984).
8. D. Kirkpatrick and R. Seidel, The ultimate planar convex hull algorithm, *SIAM Journal on Computing* **15**, 287–299 (1986).

9. S. Kapoor and P. Ramanan, Lower bounds for maximal and convex layers problems, *Algorithmica* **4**, 447–459 (1989).
10. G.N. Frederickson and S. Rodger, A new approach to the dynamic maintenance of maximal points in a plane, *Discrete Computational Geometry* **5**, 365–374 (1990).
11. J.L. Bentley, K.L. Clarkson and D.B. Levine, Fast linear expected-time alorithms for computing maxima and convex hulls, In *Proceedings of the First Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 179–187, SIAM, Philadelphia, (1990).
12. R. Janardan, On the dynamic maintenance of maximal points in the plane, *Information Processing Letters* **40**, 59–64 (1991).
13. O. Barndorff-Nielsen and M. Sobel, On the distribution of the number of admissible points in a vector, *Theory of Probability and its Applications* **11**, 249–269 (1966).
14. L. Devroye, A note on finding convex hulls via maximal vectors, *Information Processing Letters* **11**, 53–56 (1980).
15. L. Devroye, Moment inequalities for random variables in computational geometry, *Computing* **30**, 111–119 (1983).
16. L. Devroye, On the expected time required to construct the outer layer, *Information Processing Letters* **20**, 255–257 (1985).
17. C. Buchta, On the average number of maxima in a set of vectors, *Information Processing Letters* **33**, 63–65 (1989).
18. R.A. Dwyer, Kinder, gentler average-case analysis for convex hulls and maximal vectors, *SIGACT News* **21**, 64–71 (1990).
19. I.P. Natanson, *Theory of Functions of a Real Variable*, Frederick Ungar Publ. Co., New York, (1955).
20. A. Rényi and R. Sulanke, Über die konvexe Hulle von n zufällig gewahlten Punkten, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **2**, 75–84 (1963).
21. L. Devroye, *Lecture Notes on Bucket Algorithms*, Birkhäuser Verlag, Boston, (1986).