Notes on maxima in non-convex regions extended version

Mordecai Golin

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France

Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 1535

Programme 2
Calcul Symbolique, Programmation et Génie logiciel

NOTES ON MAXIMA IN NON-CONVEX REGIONS EXTENDED VERSION

Mordecai J. GOLIN

Octobre 1991

Notes on Maxima in Non-Convex Regions Extended Version

Mordecai J. Golin *
INRIA Rocquencourt
78153 Le Chesnay (France)
golin@margaux.inria.fr

October 4, 1991

Abstract: Let C be a fixed planar region. Choose n points IID uniformly from C and let X_n be the number of them which are maximal. We show that if g is a member of a large class of monotone functions then there always exists a region C such that $\mathrm{E}(X_n) = \theta(g(n))$.

Notes sur les maxima dans les régions non-convexes version longue

Résumé : Soit C une région du plan. Pour n points choisis IID à partir de la distribution uniforme sur C soit X_n le nombre de ces points qui sont maximaux. Nous montrons que si g appartient à une grande classe des functions monotones alors il existe toujours une région C telle que $\mathrm{E}(X_n) = \theta(g(n))$.

*This work was supported by a Chateaubriand fellowship from the French Ministère des Affaires Étrangères and by the European Community, Esprit II Basic Research Action Program Number 3075 (ALCOM).

Notes on Maxima in Non-Convex Regions
Extended Version

Mordecai J. Golin *
INRIA Rocquencourt
78153 Le Chesnay (France)
golin@margaux.inria.fr

October 7, 1991

Let C be a planar region. Choose n points $p_1, \ldots, p_n$ I.I.D. from the uniform distribution over C. Let M_n^C be the number of these points that are maximal. If C is convex it is known that either $\mathrm{E}\left(M_n^C\right) = \theta(\sqrt{n})$ or $\mathrm{E}\left(M_n^C\right) = O(\log n)$. In this note we will show that, for general C, there is very little that can be said, a-priori, about $\mathrm{E}\left(M_n^C\right)$. More specifically we will show that if g is a member of a large class of monotonic functions then there is a region C such that $\mathrm{E}\left(M_n^C\right) = \theta(g(n))$. This class contains all functions with regular variation and (i) exponent less than 1 or (ii) exponent equal to 1 and $n/g(n) > \ln^\beta n$ for any $\beta > 1$. For example, all functions of the form $g(n) = n^\alpha$, $0 < \alpha < 1$, or $g(n) = \ln^\beta n$, $\beta \geq 0$ satisfy condition (i) while all functions of the form $g(n) = n \ln\ln n / \ln^\beta n$, $\beta > 1$ satisfy (ii). The class also contains nondecreasing functions like $g(n) = \ln^* n$. The results in this paper remain valid in higher dimensions.

Note: Section 3 of this version of the paper, *The Second Construction*, will not be contained in the journal version. It is included here only for the sake of completeness.

1 Introduction

Identifying the maximal elements in a set of n points is a problem which arises quite naturally in certain areas of statistics, economics and of course computational geometry [1] [11]. The more efficient algorithms for finding maxima have bad worst case behavior but run quickly in a probabilistic sense, i.e. they have fast expected running times when the inputs are chosen from certain distributions [1] [2]. Since the run times of these algorithms are usually dependent upon the *number* of maxima they output there has been a great deal of study into the problem of calculating the expected number of maxima when the points are chosen from different specific distributions [1] [2] [3] [4] [6]. In this paper we will study the converse of this problem.

Suppose $g(x) \leq x$ is a monotonically increasing function. We will show that for a large class of "good" functions g it is possible to construct a planar region with the following property: when n points are chosen independently identically distributed from the uniform distribution over C then the expected number of the points that are maximal will be $\theta(g(n))$.

The class of "good" functions includes all monotonically increasing functions that are of regular variation and have (i) exponent less than 1 or (ii) exponent equal to 1 and $n/g(n) > \ln^\beta n$ for any $\beta > 1$. For example, functions of the form $g(n) = n^\alpha \ln^\beta n$ for $0 \leq \alpha < 1$ and $\beta \geq 0$ satisfy (i) and functions of the

*This work was supported by a Chateaubriand fellowship from the French Ministère des Affaires Étrangères and by the European Community, Esprit II Basic Research Action Program Number 3075 (ALCOM).

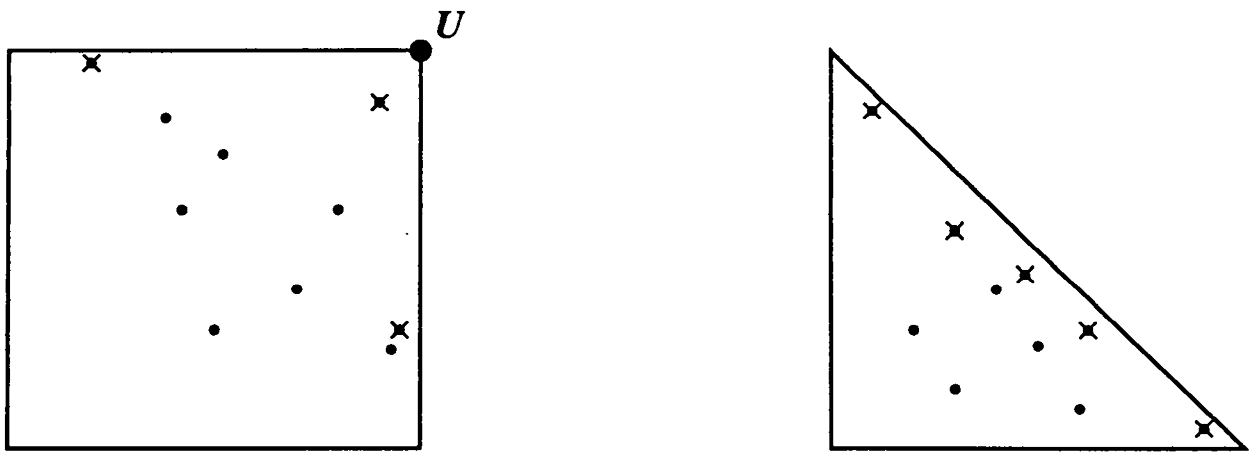

Figure 1: The square and the triangle contain ten points each. Of these, three points in the square are maximal as opposed to five in the triangle; we have marked the maximal points with x-s. The square has an upper-right-hand-corner (U) while the triangle does not have one. Therefore the expected number of maxima among n points chosen I.I.D. uniformly from the square is $O(\ln n)$ while the expected number for the triangle is $\theta(\sqrt{n})$.

---

form $g(n) = n \ln\ln n / \ln^{\beta} n$, $\beta > 1$ satisfy (ii). The class of good functions will also include nonincreasing functions such as $g(n) = \ln^{*} n$.

This result should be contrasted with previous work [9] which restricted itself to the case that C is convex. When C is convex then the expected number of maxima is either $\theta(\sqrt{n})$ or is $O(\ln n)$: nothing in between is possible.

The results in this paper remain valid in higher dimensions.

1.1 Some Background And The Main Result

Let $p = (p.x, p.y)$ and $q = (q.x, q.y)$ be points in the plane. We say that p *dominates* q if $p.x \geq q.x$ and $p.y \geq q.y$. If p does not dominate q and q does not dominate p then p and q are *incomparable*. For a set S of points in the plane the *maximal points* (*maxima*) of S are the points $p \in S$ that are not dominated by any $q \in S$. Figure 1. We denote the set of maxima in S by

$$\mathrm{MAX}(S) = \{p \in S \ : \ p \text{ is not dominated by any } q \in S,\ q \neq p.\}$$

When S is a set of n points chosen "randomly" from some distribution there has been a great deal of work done on the analysis of the expected number of maximal points, $\mathbf{E}(|\mathrm{MAX}(S)|)$ (see [6] for a survey). Much of this work defines "randomly" to mean that the points are chosen independently identically distributed (I.I.D.) from the uniform distribution over some region. For example, if the points are chosen I.I.D. uniformly from a rectangle with sides parallel to the Cartesian axes then $\mathbf{E}(|\mathrm{MAX}(S)|) = \theta(\ln n)$ [2] [3] while if they are chosen I.I.D. uniformly from a circle then $\mathbf{E}(|\mathrm{MAX}(S)|) = \theta(\sqrt{n})$ [6].

This is the problem that will interest us in this note. Let C be some fixed planar region and let $p_1, \ldots, p_n$ be chosen I.I.D. uniformly from C. We will denote the (random variable which is the) number of maxima by $M_n^C = |\mathrm{MAX}(\{p_1, \ldots, p_n\})|$ and will study the asymptotics of $\mathbf{E}\left(M_n^C\right)$, the expected number of maxima.

Quite a lot is known when C is restricted to be a convex region. There is a general upper bound, $\mathbf{E}\left(M_n^C\right) = O(\sqrt{n})$ [4] [6]. Also, obviously, C's shape must somehow influence the asymptotic behavior of $\mathbf{E}\left(M_n^C\right)$. We say that C has an *upper-right-hand-corner* if C contains a point which dominates every point in C, i.e. $|\mathrm{MAX}(C)| = 1$; the square in Figure 1 has an upper-right-hand-corner while the triangle does not. It can be shown that when C has such a corner then $\mathbf{E}\left(M_n^C\right) = O(\log n)$ while when C does not have such a corner then $\mathrm{E}\left(M_n^C\right) = \theta(\sqrt{n})$ [9], e.g. the square has $\mathrm{E}\left(M_n^C\right) = O(\log n)$ and the triangle has $\mathrm{E}\left(M_n^C\right) = \theta(\sqrt{n})$. The intuitive justification is that when C has a corner one of the p_i-s will be close to that corner and dominate most of the other p_i-s so $\mathrm{E}\left(M_n^C\right)$ will be small; when C doesn't have a corner then no one point will be able to dominate most of the others so $\mathrm{E}\left(M_n^C\right)$ will be large. A consequence of this result is that there is no convex region C such that $\mathrm{E}\left(M_n^C\right) = o(\sqrt{n})$ *and* $\mathrm{E}\left(M_n^C\right) = \omega(\log n)$.

In this paper we will show that, if we drop the requirement that C is convex, then no general statements can be made about $\mathrm{E}\left(M_n^C\right)$. Our main theorem is that if $g(x) \le x$ is a monotonically increasing function that fulfills certain conditions then there is a region C such that $\mathrm{E}\left(M_n^C\right) = \theta(g(n))$.

In order to precisely specify these conditions we introduce the concepts of regular and slow variation. Our treatment follows that of Feller [8] (section VIII.8).

Definition 1 *A positive (not necessarily) monotone function L defined on $(0,\infty)$* varies slowly *at infinity if and only if, for all $x > 0$*

$$\lim_{t\to\infty} \frac{L(xt)}{L(t)} \to 1.$$

Definition 2 *A function U defined on $(0,\infty)$* varies regularly *at infinity with exponent ρ if and only if it is of the form $x^\rho L(x)$ where L is slowly varying.*

As an example the function $\ln^2 n$ varies slowly at infinity so the function $\sqrt{n}\ln^2 n$ varies regularly at infinity with exponent 1/2. Similarly, the function $1/\ln^2 n$ varies slowly at infinity so the function $n/\ln^2 n$ varies regularly at infinity with exponent 1.

We now state our main result.

Theorem 1 *Let g be a continuous, monotonically increasing almost everywhere differentiable function from $(0,\infty)$ onto itself. Furthermore, suppose that g is regularly varying with exponent ρ and either (i) $\rho < 1$ or (ii) $\rho = 1$ and $x/g(x) \ge \ln^\beta x$ for some $\beta > 1$. Then there is some planar region C such that for n points $p_1, \ldots, p_n$ chosen I.I.D. uniformly from C the expected number of the points that are maximal is $\theta(g(n))$:*

$$\mathrm{E}\left(|\mathrm{MAX}\left(\{p_1, \ldots, p_n\}\right)|\right) = \theta(g(n)).$$

Very large classes of functions g satisfy the conditions of Theorem 1. Some examples:

1. $g(x) = x^\alpha$ where $\alpha < 1$.
2. More generally, $g(x) = x^\alpha e^{\ln^\beta x} \ln^\gamma x$ where $0 \le \alpha < 1$, $0 \le \beta < 1$ and $\gamma > 0$.
3. $g(x) = \ln^{(m)}(x)$ the m'th iterated logarithm: $\ln^{(1)}(x) = \ln x$ and $\ln^{(m+1)}(x) = \ln\left(\ln^{(m)}(x)\right)$.
4. $g(x) = \frac{x \ln\ln x}{\ln^\beta x}$ where $\beta > 1$.

Examples 1, 2, and 3 satisfy condition (i); example 4 satisfies condition (ii). The theorem therefore tells us that for each of these g-s there is some C such that $\mathrm{E}\left(M_n^C\right) = \theta(g(n))$.

We should point out that in practice the continuity, monotonicity and differentiability requirements of the theorem are not as restrictive as they might seem. This is because our problem is discrete; $\mathrm{E}\left(M_n^C\right)$ is only defined when n is integral.

Let $\{g_n\}_{n=1}^\infty$ be a nondecreasing sequence such that $\lim_{n\to\infty} g(n) = \infty$. Then it is always possible to find a monotonically increasing continuously differentiable function g such that $|g(n) - g_n| < 1$. First find a monotonically increasing sequence $\{\tilde{g}_n\}_{n=1}^\infty$ with $|g_n - \tilde{g}_n| < 1$. Next, construct a monotonically increasing continuously differentiable function g such that $g(n) = \tilde{g}_n$ for all integral n. This is the function g which is then plugged into the theorem to find a C such that $\mathrm{E}\left(M_n^C\right) = \theta(g(n)) = \theta(g_n)$.

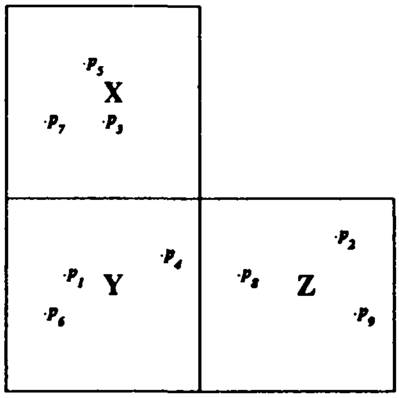

Figure 2: Let $S = \{p_1, \ldots, p_9\}$, $C' = X \cup Y$, $C'' = Y \cup Z$ and $C = C' \cup C'' = X \cup Y \cup Z$. Then $\mathrm{MAX}(S) = \{p_3, p_5, p_2, p_9\}$ and we have $\mathrm{MAX}(S) \cap C' = \{p_3, p_5\} \subseteq \{p_3, p_4, p_5\} = \mathrm{MAX}(S \cap C')$ and $\mathrm{MAX}(S) \cap C'' = \{p_2, p_9\} = \mathrm{MAX}(S \cap C'')$.

As an example suppose that $g_n = \ln^* n = \max\{i : \ln^{(i)} n \geq 1\}$. The function $\ln^*$ does vary regularly (with exponent 0) but is neither monotonic nor continuously differentiable. Following the remarks in the previous paragraph we can construct a monotonically increasing continuously differentiable function g such that $|g(n) - \ln^* n| < 1$ for all integral n. This g varies regularly because $\ln^*(x)$ varies regularly. Thus there is some region C such that $\mathbf{E}\left(M_n^C\right) = \theta(\ln^* n)$.

In sections 2 and 3 we will give two different methods for constructing C with $\mathbf{E}\left(M_n^C\right) = \theta(g(n))$. The method of Section 2 serves as the basis for the proof of Theorem 1. The region that it constructs will have a sawtooth upper boundary. The method of Section 3 only works for a small class of functions g. Its advantage is that the region that it constructs has a much "nicer" boundary than the one constructed in Section 2. In Section 4 we discuss extensions to Theorem 1. We explain how to extend the theorem to maxima in higher dimensions; to the analysis of higher moments of M_n^C; to the case where the points are chosen from a Poisson distribution. We conclude with a short discussion as to whether there are regions with $\mathbf{E}\left(M_n^C\right) = \theta(g(n))$ that are "simpler" than the regions we construct.

1.2 Notation and Observations

A short list of notation and observations that we will use.

1. Let X be a binomial random variable with parameters n and p :

$$\Pr(X = i) = \begin{cases} \binom{n}{i} p^i (1-p)^{n-i} & 0 \leq i \leq n \\ 0 & \text{otherwise.} \end{cases}$$

   We will say that X is a binomial $B(n,p)$ random variable. When X is a binomial $B(n,p)$ random variable $\mathbf{E}(X) = np$ and $\mathrm{Var}(X) = np(1-p)$.

2. Let A and B be events. We denote *the probability of A conditioned on B* by $\Pr(A \mid B)$.

3. Let $C' \subseteq C$ be two planar regions. Let $S = \{p_1, \ldots, p_n\} \subseteq C$ be a set of points. Then the points which are maximal *and* in C' are a subset of the maxima among the points *restricted to* C'. That is

$$\mathrm{MAX}(S) \cap C' \subseteq \mathrm{MAX}(S \cap C'). \tag{1}$$

   (See Figure 2.) As a consequence if C', C'' are two (not necessarily disjoint) regions such that $C' \cup C'' = C$ then

$$\mathrm{MAX}(S) \subseteq \mathrm{MAX}(S \cap C') \cup \mathrm{MAX}(S \cap C''). \tag{2}$$

4. Let g be a continuous, monotonically increasing function from $(0, \infty)$ into itself with $g(1) \leq 1$ and $\lim_{x \to \infty} g(x) = \infty$. Then the functional inverse of g is well defined and is a continuous, monotonically increasing function on $[1, \infty)$. This function, which we denote by g^{-1}, is the unique function which satisfies $g(g^{-1}(x)) = g^{-1}(g(x)) = x$. For example if $g(x) = \ln x$ then $g^{-1}(x) = e^x$.

2 The Main Construction

In this section we will prove Theorem 1. We first give sufficient conditions on g for the existence of a region C such that $\mathrm{E}\left(M_n^C\right) = \theta(g(n))$. We then show that any g that satisfies the conditions of Theorem 1 satisfies these sufficient conditions.

Theorem 2 *Let g be a continuous monotonically increasing function from $(0, \infty)$ into itself with $g(x) \leq x$ and $\lim_{x \to \infty} g(x) = \infty$. Use g^{-1} to denote the functional inverse of g : $g(g^{-1}(x)) = g^{-1}(g(x)) = x$. Suppose that g fulfills the following condition for all integers $n > 0$:*

$$\sum_{i > g(n)} \frac{1}{g^{-1}(i)} = O\left(\frac{g(n)}{n}\right). \tag{3}$$

Then there exists a connected planar region C such that, for n points $p_1, \ldots, p_n$ chosen I.I.D. uniformly from C, the expected number of the points that are maximal is $\theta(g(n))$:

$$\mathrm{E}\left(|\mathrm{MAX}\left(\{p_1, \ldots, p_n\}\right)|\right) = \theta(g(n)).$$

Proof: Let $g(x)$ be a function that satisfies the conditions of the theorem. The proof will be in two parts. In the first we will define an infinite sequence of triangles C_i of decreasing size. We will show that when n points are chosen I.I.D. uniformly from $C = \bigcup_i C_i$ then $\mathrm{E}\left(M_n^C\right) = \theta(g(n))$. In the second part we will show how to modify C to construct a new connected region C' with $\mathrm{E}\left(M_n^{C'}\right) = \theta(g(n))$. It is this C' which will satisfy the Theorem.

Part 1: The construction depends heavily upon the following:

Lemma 1 *Fix $d > 0$ and let T be the triangle with vertices $(0,0)$, $(d, 0)$ and $(2d, 2d)$. Choose n points $p_1, \ldots, p_n$ I.I.D. uniformly from T. Then*

$$\mathrm{E}\left(|\mathrm{MAX}\left(\{p_1, \ldots, p_n\}\right)|\right) \leq 2.$$

Proof: Refer to Figure 3. We may, without loss of generality, assume that $d = 1$. Otherwise we can scale the triangle by multiplying both the x and y coordinates by $1/d$ leaving $\mathrm{E}\left(M_n^C\right)$ unchanged.

Let T be the triangle with vertices $(0,0)$, $(1,0)$ and $(2,2)$. By symmetry we have that $\mathrm{E}\left(|\mathrm{MAX}\left(\{p_1, \ldots, p_n\}\right)|\right) = n \Pr(p_1 \in \mathrm{MAX}(\{p_1, \ldots, p_n\}))$. Suppose that $p_1 = (x, y)$. Then p_1 is maximal if and only if none of the points $p_2, \ldots p_n$ are in the region

$$R(x, y) = \{p \in T : p.x \geq x \text{ and } p.y \geq y\}.$$

If $(x, y) \in T$ then $y \leq x \leq y/2 + 1$. Thus

$$\begin{aligned} Area(R(x,y)) &= \frac{1}{2}[1 - y/2][2 - y] - \frac{(x-y)^2}{2} \\ &\geq \frac{1}{2}[y/2 + 1 - y][2 - (y/2 + 1)] \\ &= \frac{1}{2}[1 - y/2]^2. \end{aligned}$$

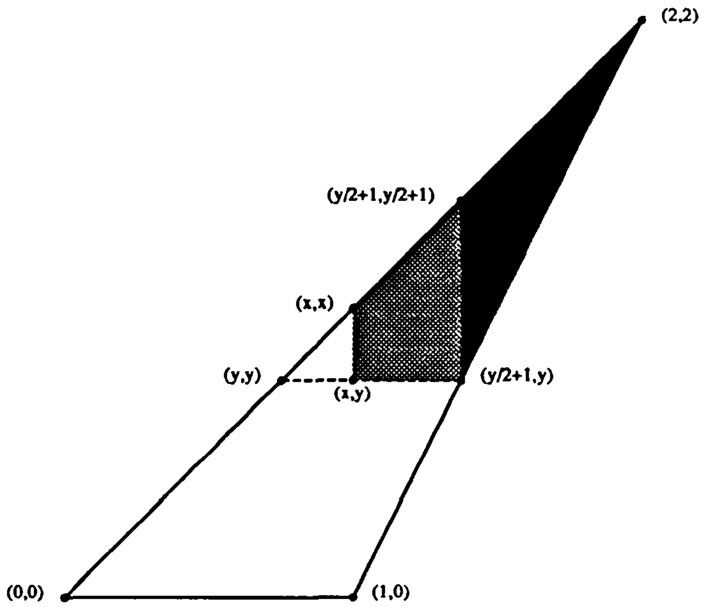

Figure 3: $R(x,y)$ is the region containing all points in T that dominate (x,y). In the diagram $R(x,y)$ is the union of the two shaded regions. The darker of the two shaded regions (the small triangle) has area $\frac{1}{2}[y/2+1-y][2-(y/2+1)] = \frac{1}{2}[1-y/2]^2$ so $Area(R(x,y)) \geq \frac{1}{2}[1-y/2]^2$.

The derivation of these last identities/inequalities can best be understood by referring to Figure 3. Now

$$\Pr\left(p_1 \in \mathrm{MAX}(\{p_1,\ldots,p_n\}) \mid p=(x,y)\right) = \left(1 - \frac{Area(R(x,y))}{Area(T)}\right)^{n-1}.$$

Note that $Area(T) = 1$ so

$$\Pr\left(p_1 \in \mathrm{MAX}(\{p_1,\ldots,p_n\}) \mid p=(x,y)\right) \leq \left(1 - \frac{1}{2}[1-y/2]^2\right)^{n-1}.$$

Integrating over all possible choices of $p=(x,y)$ we find

$$\begin{aligned} \mathrm{E}(M_n^T) &= n\int_{y=0}^{2}\int_{x=y}^{y/2+1} \Pr(p_1 \in \mathrm{MAX}(\{p_1,\ldots,p_n\}) \mid p=(x,y))\,dx\,dy \\ &\leq n\int_0^2 [1-y/2]\left[1-\frac{1}{2}(1-y/2)^2\right]^{n-1} dy. \end{aligned}$$

The change of variables $u = 1-y/2$ gives

$$\mathrm{E}(M_n^T) \leq 2n\int_0^1 u\left(1-u^2/2\right)^{n-1} du \leq 2.$$

■

We will now construct an infinite sequence C_i of smaller and smaller triangles of the type described by the lemma and define $C = \bigcup_i C_i$. Figure 4.

Construction of C :

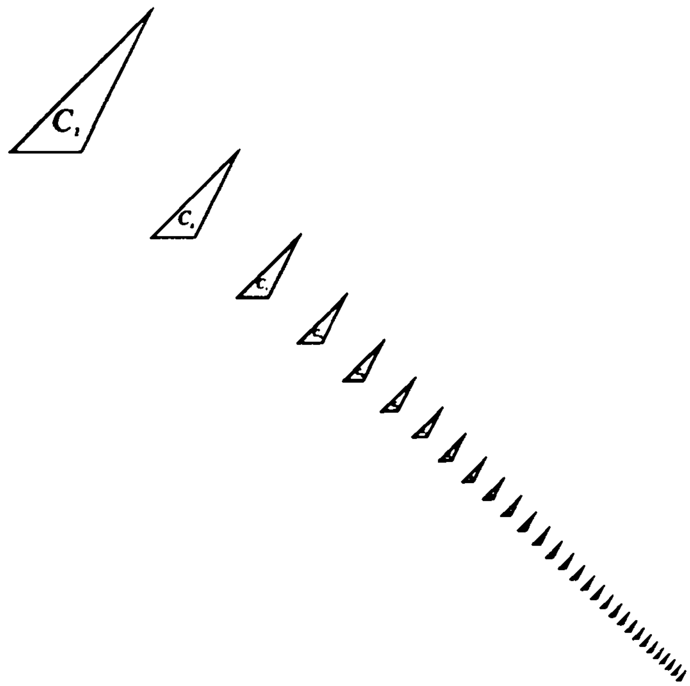

Figure 4: The region $C = \cup_i C_i$ when $g(x) = x^{5/12}$, $g^{-1}(x) = x^{12/5}$ and $f_i = i^{-6/5}$. Note that points in different triangles are incomparable; if $p \in C_i$ and $q \in C_j$ where $i \neq j$ then p and q are incomparable.

---

1. Set $f_i = 1/\sqrt{g^{-1}(i)}$.
2. Set $x_1 = y_1 = 0$. For $i > 1$ set $x_i = x_{i-1} + 2f_{i-1}$ and $y_i = y_{i-1} - 2f_i$.
3. For $i \geq 1$ define C_i to be the copy of T which has $d = f_i$ and whose lower left corner is (x_i, y_i). That is, C_i is the triangle with vertices (x_i, y_i), $(x_i + f_i, y_i)$ and $(x_i + 2f_i, y_i + 2f_i)$.
4. Set $C = \bigcup_i C_i$.

Let $c = \sum_i 1/g^{-1}(i) < \infty$. ($c$ exists because of (3)). Notice that the C_i are pairwise disjoint. Since $Area(C_i) = 1/g^{-1}(i)$ this tells us that $Area(C) = \sum_i 1/g^{-1}(i) = c$.

Not only are the C_i disjoint but points in different C_i-s are incomparable, i.e. if $p \in C_i$ and $q \in C_j$ with $i \neq j$ then p and q are incomparable. Thus for any set of points $S = \{p_1, \ldots, p_n\} \subset C$ we have

$$\mathrm{MAX}(S) = \mathrm{MAX}(\{p_1, \ldots, p_n\}) = \bigcup_i \mathrm{MAX}(S \cap C_i).$$

Because the C_i are disjoint the sets $\mathrm{MAX}(S \cap C_i)$ are also disjoint. Therefore for a set $S = \{p_1, \ldots, p_n\}$ of n points chosen I.I.D. uniformly from C

$$\mathbf{E}(|\mathrm{MAX}(\{p_1, \ldots, p_n\})|) = \sum_i \mathbf{E}(|\mathrm{MAX}(S \cap C_i)|). \tag{4}$$

Let $X_i = |S \cap C_i|$ be the number of the points in triangle C_i. If $X_i > 0$ then Lemma 1 tells us that $1 \leq \mathbf{E}(|\mathrm{MAX}(S \cap C_i)|) \leq 2$. Put another way the lemma says that

$$\mathbf{E}(|\mathrm{MAX}(S \cap C_i)|) = \theta(\Pr(X_i > 0)). \tag{5}$$

Now X_i is a binomial $B(n, Area(C_i)/Area(C)) = B(n, 1/cg^{-1}(i))$ random variable so

$$\mathrm{E}(|\mathrm{MAX}(S \cap C_i)|) = \theta\left(1 - \left[1 - \frac{1}{cg^{-1}(i)}\right]^n\right). \tag{6}$$

When $i \leq g(n)$ we have $g^{-1}(i) \leq n$ so

$$\left[1 - \frac{1}{cg^{-1}(i)}\right]^n \leq \left[1 - \frac{1}{cn}\right]^n \leq e^{-1/c}. \tag{7}$$

For general i we have

$$\left[1 - \frac{1}{cg^{-1}(i)}\right]^n = e^{-n \ln\left[1 - \frac{1}{cg^{-1}(i)}\right]} = 1 - O\left(\frac{n}{g^{-1}(i)}\right) \tag{8}$$

where we use $\ln[1-x] = -O(x)$. Therefore

$$\begin{aligned}
\sum_i \mathrm{E}(|\mathrm{MAX}(S \cap C_i)|) &= \sum_{i \leq g(n)} \mathrm{E}(|\mathrm{MAX}(S \cap C_i)|) \\
&\quad + \sum_{i > g(n)} \mathrm{E}(|\mathrm{MAX}(S \cap C_i)|) \\
&= \sum_{i \leq g(n)} \theta(1) + \sum_{i > g(n)} O\left(\frac{n}{g^{-1}(i)}\right) \\
&= \theta(g(n)) + nO\left(\sum_{i > g(n)} \frac{1}{g^{-1}(i)}\right)
\end{aligned}$$

Inserting (3), the condition of the theorem, into the last line we find that

$$\sum_i \mathrm{E}(|\mathrm{MAX}(S \cap C_i)|) = \theta(g(n)). \tag{9}$$

Referring back to (4) gives

$$\mathrm{E}(|\mathrm{MAX}(\{p_1, \ldots, p_n\})|) = \theta(g(n))$$

and Part 1 of the proof is complete.

Part 2: The C constructed in Part 1 is not really a region; it is the union of an infinite number of non-overlapping regions. In this part we will modify C to yield a *connected* region C' such that $\mathrm{E}\left(M_n^{C'}\right) = \theta(g(n))$. It is this C' that will satisfy the theorem. We will need the following lemma:

Lemma 2 *Let $d > 0$ and $d' \leq 4d$. Define T to be, as in Lemma 1, the triangle with vertices $(0,0)$, $(d,0)$ and $(2d, 2d)$. Let R be the rectangle with vertices $(0,0)$, $(d,0)$, $(0,-d')$ and $(d,-d')$. If $p_1, \ldots, p_n$ are chosen I.I.D. uniformly from $T \cup R$ then*

$$\mathrm{E}(\mathrm{MAX}(\{p_1, \ldots, p_n\})) = O(1).$$

Proof: Refer to Figure 5. From observation (1) in Section 1.2 we have

$$|\mathrm{MAX}(\{p_1, \ldots, p_n\})| \leq |\mathrm{MAX}(\{p_1, \ldots, p_n\}) \cap T| + |\mathrm{MAX}(\{p_1, \ldots, p_n\}) \cap R|.$$

From observation (2) in the same subsection we also have that

$$\mathrm{MAX}(\{p_1, \ldots, p_n\}) \cap T \subseteq \mathrm{MAX}(\{p_1, \ldots, p_n\} \cap T).$$

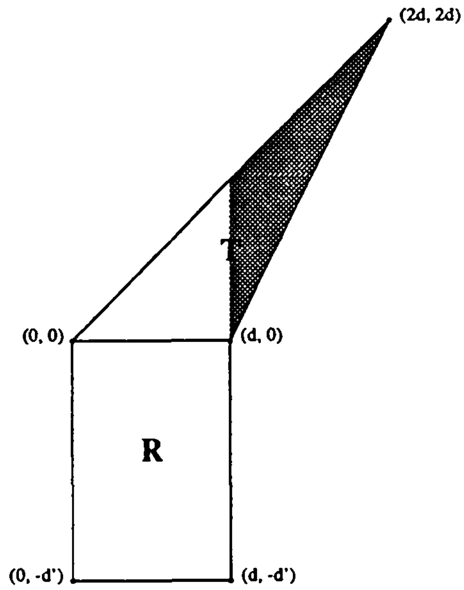

Figure 5: The region described in Lemma 2. It is the union of triangle T and rectangle R. Note that any point in R is dominated by any point in the shaded triangle. Note too that the shaded triangle has area $d^2/2$. We assume that $d' \le 4d$ so $\frac{Area(shaded\ triangle)}{Area(R \cup T)} \ge \frac{1}{10}$.

Applying Lemma 1 gives

$$\mathrm{E}(|\mathrm{MAX}(\{p_1, \ldots, p_n\}) \cap T|) = O(1). \tag{10}$$

Let T' be the triangle with vertices $(d, 0)$, (d, d) and $(2d, 2d)$. This is the shaded region in Figure 5. Suppose that $p_1 \in R$. Then $p_1 \in \mathrm{MAX}(p_1, \ldots, p_n)$ only if none of $p_2, \ldots, p_n$ are in T', an event which occurs with probability

$$\left[1 - \frac{Area(T')}{Area(T \cup R)}\right]^{n-1} \le \left(\frac{9}{10}\right)^{n-1}.$$

Thus

$$\mathrm{E}(|\mathrm{MAX}(\{p_1, \ldots, p_n\}) \cap R|) \le n\left(\frac{9}{10}\right)^{n-1}. \tag{11}$$

Combining (10) and (11) proves the lemma. ■

The new region C' will be the union of an infinite number of regions of the type defined by the lemma. Figure 6.

Construction of C' :

1. Set $f_i = 1/\sqrt{g^{-1}(i)}$.
2. Set $x_1 = y_1 = 0$ and For $i > 1$ set $x_i = x_{i-1} + 2f_{i-1}$ and $y_i = y_{i-1} - 2f_i$.
3. Define C'_1 to be the triangle with vertices $(0, 0)$, $(f_1, 0)$ and $(2f_1, 2f_1)$. For $i > 1$ define C'_i to be the triangle with vertices $(x_i + 2f_i, y_i + 2f_i)$, $(x_i - 2f_{i-1}, y_i - 2f_{i-1})$ and $(x_i + f_i - f_{i-1}, y_i - 2f_{i-1})$.
4. For $i > 0$ define R_i to be the rectangle with vertices (x_i, y_i), $(x_i + f_i, y_i)$, $(x_i + f_i, y_i - 2f_i - 2f_{i+1})$ and $(x_i, y_i - 2f_i - 2f_{i+1})$
5. Set $C' = \bigcup_i (C'_i \cup R_i)$.

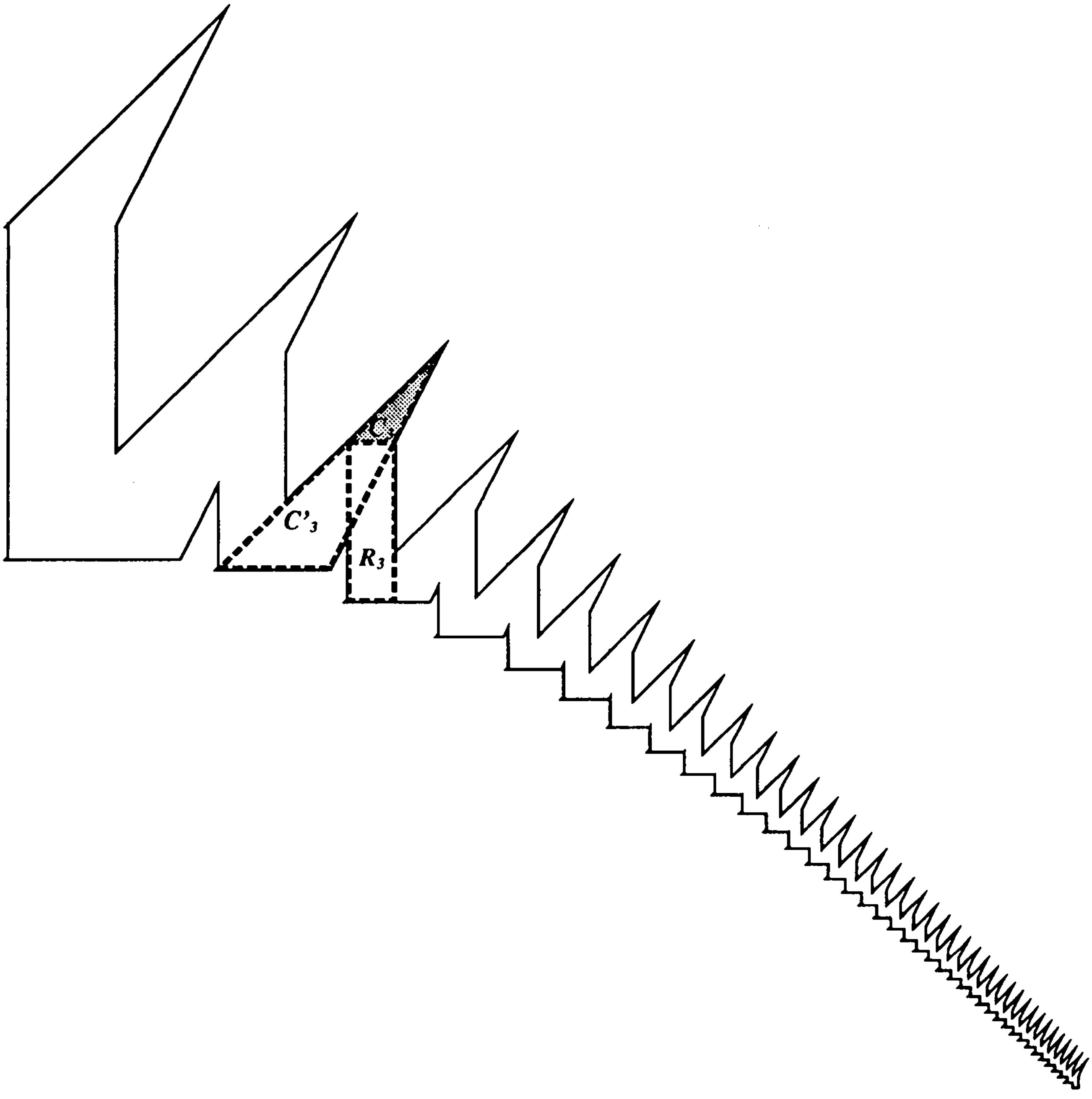

Figure 6: The region $C' = \bigcup_i (C'_i \cup R_i)$ when $g(x) = x^{5/12}$. We have emphasized C'_3 and R_3 by giving them a dashed boundary. The shaded region is the triangle C_3 defined in Part 1 of the proof. Note how $C_3 \subseteq C'_3$.

First, note that C' is connected. This follows from the fact that the C'_i were defined so that the lower left corner of R_i is also the lower left corner of C'_{i+1} (this is best understood by looking at Figure 6). Thus, for all i, the region $C'_i \cup R_i \cup C'_{i+1}$ is connected so C' is connected. Secondly, C' has bounded area. The $f_i = 1/\sqrt{g^{-1}(i)}$ are monotonically decreasing. Thus $Area(R_i) = f_i(f_i + 2f_{i+1}) \le 3f_i^2$. Similarly we find that for $i > 1$, $Area(C'_i) \le 4f_{i-1}^2$. Then

$$Area(C') \le \sum_i Area(C'_i) + \sum_i Area(R_i) = O\left(\sum_i f_i^2\right) = O\left(\sum_i 1/g^{-1}(i)\right)$$

which converges. We set $c' = Area(C')$.

The proof that $\mathbf{E}\left(M_n^{C'}\right) = \theta(g(n))$ is almost exactly like the proof of $\mathbf{E}\left(M_n^{C}\right) = \theta(g(n))$ in Part 1.

Note that the triangles C_i defined in Part 1 satisfy $C_i \subseteq C'_i$. Furthermore C' was constructed in such a way that, if $p \in C_i$ and $q \in C' \backslash C_i$ (the set C' minus the set C_i) then q cannot dominate p. Thus for a set $S = \{p_1, \ldots, p_n\}$ of n points chosen I.I.D. uniformly from C' we have $\text{MAX}(\{p_1, \ldots, p_n\}) \cap C_i = \text{MAX}(S \cap C_i)$. Taking expectations and summing over i we have

$$\sum_i \mathbf{E}(|\text{MAX}(S \cap C_i)|) \le \mathbf{E}(|\text{MAX}(S)|) = \mathbf{E}(|\text{MAX}(\{p_1, \ldots, p_n\})|). \tag{12}$$

Let $X_i = |S \cap C_i|$ be the number of points in the triangle C_i. Lemma 1 tells us that if $X_i > 0$ then $\mathbf{E}(|\text{MAX}(S \cap C_i)|) = \theta(1)$. Thus

$$\mathbf{E}(|\text{MAX}(S \cap C_i)|) = \theta(\Pr(X_i > 0)). \tag{13}$$

Now X_i is a binomial $B(n, \frac{Area(C_i)}{Area(C')}) = B(n, \frac{1}{c'g^{-1}(i)})$ random variable so

$$\mathbf{E}(|\text{MAX}(S \cap C_i)|) = \theta\left(1 - \left[1 - \frac{1}{c'g^{-1}(i)}\right]^n\right). \tag{14}$$

This is equation (6) with c replaced by c'. Repeat the analysis given in equations (7), (8) and (9) to find

$$\sum_i \mathbf{E}(|\text{MAX}(S \cap C_i)|) = \theta(g(n)). \tag{15}$$

Inserting this back into (12) gives the lower bound

$$\mathbf{E}(|\text{MAX}(p_1, \ldots, p_n)|) = \Omega(g(n)). \tag{16}$$

To find an upper bound recall that $C' = \bigcup_i (C'_i \cup R_i)$. Observation (2) of Section 1.2 tells us

$$\begin{aligned} \mathbf{E}(|\text{MAX}(p_1, \ldots, p_n)|) &\le \sum_{i \ge 1} \mathbf{E}(|\text{MAX}(S) \cap (C'_i \cup R_i)|) \\ &\le \sum_{i > 1} \mathbf{E}(|\text{MAX}(S \cap C'_i)|) \\ &\qquad + \sum_{i \ge 1} \mathbf{E}(|\text{MAX}(S \cap (C_i \cup R_i))|) \end{aligned} \tag{17}$$

Now C'_i is a triangle of the type described by Lemma 1. Its base has width $d = f_i + f_{i-1} = \theta(f_{i-1})$ so $Area(C'_i) = \theta\left(1/g^{-1}(i-1)\right)$. We can thus use the same technique that led to (15) to prove

$$\sum_i \mathbf{E}(|\text{MAX}(S \cap C'_i)|) = \theta(g(n)) \tag{18}$$

To analyze $\sum_{i\geq 1} \mathrm{E}(|\mathrm{MAX}(S \cap (C_i \cup R_i))|)$ note that $C_i \cup R_i$ is a region of the form discussed by Lemma 2 with $d = f_i$ and $d' = 2f_i + 2f_{i+1} < 4f_i$. Consequentially if $S \cap (C_i \cup R_i)$ is not empty then $\mathrm{E}(|\mathrm{MAX}(S \cap (C_i \cup R_i))|) = \theta(1)$. Note also $Area(C_i \cup R_i) = \theta(1/g^{-1}(i))$.

What follows should be monotonously familiar: we use the same type of reasoning that led to (5) to get

$$\mathrm{E}(|\mathrm{MAX}(S \cap (C_i \cup R_i))|) = \theta\left(1 - \left[1 - \frac{1}{c'\theta(g^{-1}(i))}\right]^n\right). \tag{19}$$

Following the same reasoning that led to (9), to (15) and to (18) yields

$$\mathrm{E}(|\mathrm{MAX}(\{p_j \in C_i \cup R_i\})|) = \theta(g(n)).$$

Combining this with equation (18) and inserting into (17) yields

$$\mathrm{E}(|\mathrm{MAX}(p_1, \ldots, p_n)|) = O(g(n)).$$

This upper bound matches the lower bound (16) so

$$\mathrm{E}(|\mathrm{MAX}(p_1, \ldots, p_n)|) = \theta(g(n))$$

and the proof is finished. ■

We can now prove Theorem 1.

Theorem 1 *Let g be a continuous monotonically increasing almost everywhere differentiable function from $(0,\infty)$ onto itself. Furthermore suppose that g is regularly varying with exponent ρ and either (i) $\rho < 1$ or (ii) $\rho = 1$ and $x/g(x) \geq \ln^\beta x$ for some $\beta > 1$. Then there is some planar region C such that for n points $p_1, \ldots, p_n$ chosen I.I.D. uniformly from C the expected number of the points that are maximal is $\theta(g(n))$:*

$$\mathrm{E}\left(|\mathrm{MAX}\left(\{p_1, \ldots, p_n\}\right)|\right) = \theta(g(n)).$$

Proof: We will show that if g satisfies the conditions of Theorem 1 then

$$\sum_{i>g(n)} \frac{1}{g^{-1}(i)} = O\left(\frac{g(n)}{n}\right). \tag{20}$$

The proof will follow from Theorem 2. The function g is continuous and monotonically increasing so g^{-1} is also continuous and monotonically increasing. Set $k = \lceil g(n) \rceil$. Then $1/g^{-1}(k) \leq 1/n$ and $\sum_{i>k} \frac{1}{g^{-1}(i)} \leq \int_{g(n)}^{\infty} \frac{dx}{g^{-1}(x)}$. Because g is monotonic and almost everywhere differentiable we can make the change of variable $u = g^{-1}(x)$ and then integrate by parts to find

$$\int_{g(n)}^{\infty} \frac{dx}{g^{-1}(x)} = \int_n^\infty \frac{g'(u)}{u}\, du = \left.\frac{g(u)}{u}\right|_n^\infty + \int_n^\infty \frac{g(u)}{u^2}\, du$$

If g satisfies either (i) or (ii) then $g(x)/x \to 0$ as $x \to \infty$ and

$$\sum_{i>g(n)} \frac{1}{g^{-1}(i)} = \sum_{i\geq k} \frac{1}{g^{-1}(i)} \leq \frac{1}{n} + \int_n^\infty \frac{g(u)}{u^2}\, du$$

To prove (20) it will thus suffice to prove that

$$\int_n^\infty \frac{g(u)}{u^2}\, du = O\left(\frac{g(n)}{n}\right).$$

The following theorem due to Feller ([8] page 281) gives the asymptotics of the truncated moments of regularly varying functions:

Theorem 3 *If $Z(x)$ varies regularly with exponent γ and $Z_p^*(x) = \int_x^\infty y^p Z(y)\,dy$ exists then*

$$\frac{t^{p+1}Z(t)}{Z_p^*(t)} \to \lambda \qquad as \qquad t \to \infty$$

where $\lambda = -(p+\gamma+1) \geq 0$.

We apply Feller's theorem with $Z = g$ and $p = -2$ to find that if $\int_n^\infty \frac{g(u)}{u^2}\,du$ exists then $\int_n^\infty \frac{g(u)}{u^2}\,du = O\left(\frac{g(n)}{n}\right)$. But if g satisfies either condition (i) or (ii) of our theorem then $\int_n^\infty \frac{g(u)}{u^2}\,du$ exists. The proof is complete. ∎

3 The Second Construction

In this section we will describe another method, that for certain functions g, also constructs a region C such that $\mathrm{E}\left(M_n^C\right) = \theta(g(n))$. This new construction remedies an aesthetic problem with the construction described in the proof of Theorem 2: the upper boundary of the region constructed in that theorem was a sawtooth curve; it increased, decreased and increased ad infinitum. The upper boundary of the region constructed in this section will be a nonincreasing curve. The price that we pay for getting this nicer boundary is a stricter set of conditions that g must satisfy. These are described by the next theorem.

Theorem 4 *Suppose $g(x)$ is a monotonically increasing function such that $\ln x \leq g(x) < x$. Suppose further that there exists a monotonically increasing continuous function $h(x)$ that fulfills the following two conditions: let $k = \lfloor h^{-1}(n) \rfloor$, where h^{-1} is the functional inverse of h.*

1. $\sum_{i>k} 1/h(i) = O(g(n)/n)$.
2. $k \ln n - \sum_{i\leq k} \ln h(i) = \theta(g(n))$

Then there exists a connected region C such that for n points chosen I.I.D. uniformly from C

$$\mathrm{E}(|\mathrm{MAX}(\{p_1,\ldots,p_n\})|) = \theta(g(n)).$$

We defer the proof for a moment to give an application. Suppose that $g(n) = n^\alpha$ for some $0 < \alpha < 1$. Set $h(n) = n^{1/\alpha}$. Now $k = \lfloor n^\alpha \rfloor$. Condition 1 of the theorem is satisfied because $\sum_{i>k} 1/h(i) = O\left(k^{1-1/\alpha}\right) = O\left(n^{\alpha-1}\right) = O(g(n)/n)$.

Condition 2 is satisfied because $\sum_{i\leq k} \ln h(i) = \ln[(n^\alpha)]^{1/\alpha}$. Stirling's formula tells us that $\ln[(n^\alpha)!] = n^\alpha \ln n^\alpha + n^\alpha/2 + O(1/n^\alpha)$ so we find that $k \ln n - \sum_{i\leq k} \ln h(i) = n^\alpha/2\alpha + O(1/n^\alpha) = \theta(g(n))$. Theorem 4 can therefore be applied and we find that a region C with $\mathrm{E}\left(M_n^C\right) = \theta(n^\alpha)$ exists.

Proof of Theorem 4: The proof has the same flavor as the proof of Theorem 2. It is divided into two parts. In the first part we define an infinite sequence of disjoint squares C_i of decreasing size. We show that when n points are chosen I.I.D. uniformly from $C = \bigcup_i C_i$ then $\mathrm{E}\left(M_n^C\right) = \theta(g(n))$. In the second part we will show how to modify C to construct a new connected region C' such that $\mathrm{E}(M_n^{C'}) = \theta(g(n))$.

Part 1: The proof will need the following lemma:

Lemma 3 *Let X be a $B(n,p)$ binomial random variable with $\mu = np \geq 1$. Choose X points $p_1, \ldots, p_X$ I.I.D. uniformly from a rectangle whose sides are parallel to the x and y axes. Then*

$$\mathrm{E}(|\mathrm{MAX}(p_1,\ldots,p_X)|) = \ln\mu + \gamma + O\left(\frac{\ln\mu}{\sqrt{\mu}}\right).$$

Proof: Let m be a fixed integer and choose $q_1, \ldots, q_m$ I.I.D. uniformly from the rectangle. Then it is well known [2] that $\mathrm{E}(|\mathrm{MAX}(q_1, \ldots, q_m)|) = H_m$ where $H_m = \sum_{i \le m} 1/i$ is the m'th harmonic number. Thus $\mathrm{E}(|\mathrm{MAX}(p_1, \ldots, p_X)|) = \mathrm{E}(H_X)$.

We will use the fact that X is very heavily concentrated around its mean. Mathematically this is expressed by Chernoff bounds (we use a version given in [10]). Let $0 \le \epsilon < 1$. Then

$$\Pr(X \ge (1+\epsilon)\mu) \le e^{-\epsilon^2\mu/3}, \quad \Pr(X \le (1-\epsilon)\mu) \le e^{-\epsilon^2\mu/2}.$$

Setting $\epsilon = 3\ln\mu/\sqrt{\mu}$ gives

$$\Pr(|X-\mu| \ge 3\ln\mu\sqrt{\mu}) \le \mu^{-\mu}.$$

Thus we have

$$\mathrm{E}(H_X) = \mathrm{E}(H_x \mid |X-\mu| \le \ln\mu\sqrt{\mu}) + O\left(\mu^{1-\mu}\right).$$

It is well known that $H_m = \ln m + \gamma + O(1/m)$ where $\gamma \approx 0.5772\ldots$ is Euler's constant. If $0 < k < k'$. Then $\ln(k'+k) = \ln k' + O(k/k')$. As a consequence if $|X-\mu| \le \ln\mu/\sqrt{\mu}$ then $\ln X = \ln\mu + O\left(\ln\mu/\sqrt{\mu}\right)$ and we have

$$\mathrm{E}(H_X) = \ln\mu + \gamma + O\left(\frac{\ln\mu}{\sqrt{\mu}}\right)$$

■

We will now construct an infinite sequence C_i of smaller and smaller squares and define $C = \bigcup_i C_i$. Figure 7 (a).

<u>*Construction of C :*</u>

1. Set $f_i = 1/\sqrt{h(i)}$.
2. Set $x_1 = y_1 = 0$. For $i > 1$ set $x_i = x_{i-1} + f_{i-1}$ and $y_i = y_{i-1} - f_i$.
3. For $i \ge 1$ define C_i to be the square with sides of length f_i parallel to the x and y axes and whose lower left corner is (x_i, y_i). That is, C_i is the square with vertices (x_i, y_i), $(x_i + f_i, y_i)$ and $(x_i + f_i, y_i + f_i)$ and $(x_i, y_i + f_i)$.
4. Set $C = \bigcup_i C_i$.

Let $c = Area(C) = \sum_i 1/h(i)$. Assume for the moment that $c = 1$ (we will justify this later). Let $S = \{p_1, \ldots, p_n\}$ be a set of n points chosen I.I.D. uniformly from C. Points in different C_i-s are incomparable so

$$\mathrm{E}(|\mathrm{MAX}(\{p_1, \ldots, p_n\})|) = \sum_i \mathrm{E}(|\mathrm{MAX}(S \cap C_i)|). \tag{21}$$

Let $k = \lfloor h^{-1}(n) \rfloor$. We will split this sum into two parts and evaluate them separately; the first part will include all terms with $i \le k$, the second all terms with $i > k$.

Let $X_i = |S \cap C_i|$ be the number of the points in square C_i. This X_i is a binomial $B(n, Area(C_i)/Area(C)) = B(n, 1/h(i))$ random variable. Lemma 3 then tells us that

$$E(|\mathrm{MAX}(S \cap C_i)|) = \ln\left(\frac{n}{h(i)}\right) + \gamma + O\left(\frac{\ln(n/h(i))}{\sqrt{n/h(i)}}\right)$$

Thus

$$\sum_{i \le k} \mathrm{E}(|\mathrm{MAX}(S \cap C_i)|) = \theta\left(n \ln k - \sum_{i \le k} \ln h_i\right). \tag{22}$$

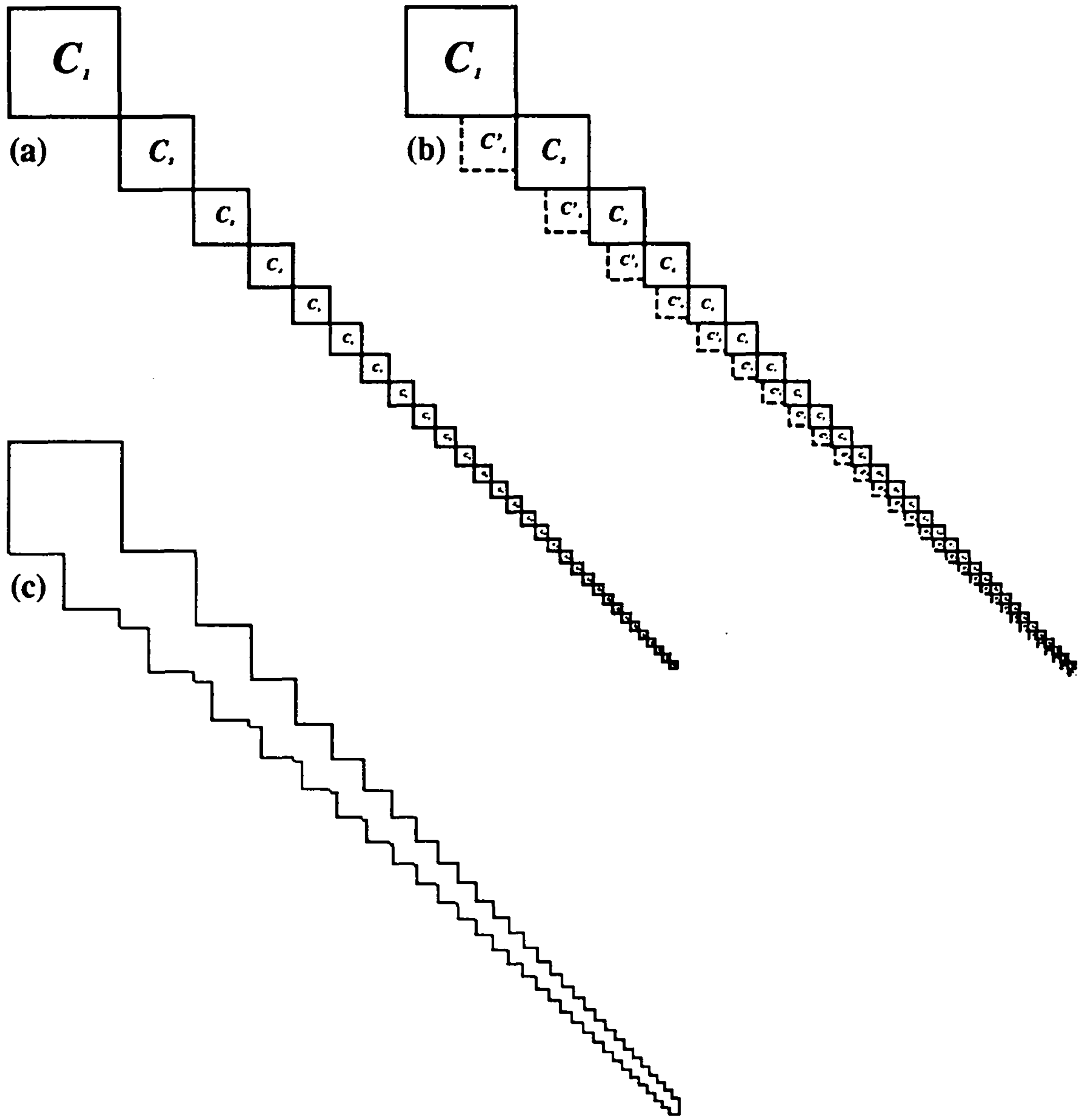

Figure 7: (a)The region $C = \cup_i C_i$ defined in Part 1 of the theorem: $g(x) = \sqrt{x}$ and $h(x) = x^2$. Note that points in different squares are incomparable; if $p \in C_i$ and $q \in C_j$ where $i \neq j$ then p and q are incomparable. (b) The region $C' = \bigcup_i (C_i \cup C'_i)$. (c) An unobstructed view of C'.

For all i the number of maxima in C_i is trivially at most the number of points in C_i so $\mathbf{E}(|\mathrm{MAX}(S \cap C_i)|) \le \mathbf{E}(X_i)$. Summing over $i > k$

$$\sum_{i>k} \mathbf{E}(|\mathrm{MAX}(S \cap C_i)|) = \sum_{i>k} \frac{n}{h_i} = O(g(n)/n) \tag{23}$$

where the last equality is a consequence of condition 1 of the theorem.

Combining (22) and (23) and using condition 1 gives

$$\sum_i \mathbf{E}(|\mathrm{MAX}(S \cap C_i)|) = \theta(g(n)). \tag{24}$$

Substituting this back into (21) yields the desired

$$E(\{p_1, \ldots, p_n\}) = \theta(g(n)).$$

We have almost finished Part 1. It remains to deal with the case that $c = Area(C) \neq 1$. We will show how to construct a new region $\hat{C}$ with $Area(\hat{C}) = 1$ and $\mathbf{E}\left(M_n^{\hat{C}}\right) = \theta(g(n))$.

First assume that $c < 1$. Let Q be the square with vertices $(0,0)$, $(0, 1-c)$, $(1, 1-c)$, and $(1,0)$. Let $\hat{C} = Q \cup C$ and choose n points $p_1, \ldots, p_n$ I.I.D. uniformly from $\hat{C}$. Points in Q and points in C are incomparable so if n points are chosen I.I.D. uniformly from C' then

$$\begin{aligned} \mathbf{E}(\mathrm{MAX}(\{p_1, \ldots, p_n\})) &= \mathbf{E}(\mathrm{MAX}(\{p_1, \ldots, p_n\}) \cap Q) \\ &\quad + \mathbf{E}(\mathrm{MAX}(\{p_1, \ldots, p_n\}) \cap C) \end{aligned}$$

A point in Q cannot dominate a point in C so

$$\mathbf{E}(\mathrm{MAX}(\{p_1, \ldots, p_n\}) \cap C) = \sum_i \mathbf{E}(|\mathrm{MAX}(S \cap C_i)|).$$

The number of points in C_i, $|S \cap C_i|$, is a $B(n, Area(C_i)/Area(\hat{C})) = B(n, 1/h(i))$ binomial random variable. But, the analysis performed above (when the points were chosen from C) that led to (24) was only dependent upon the $|S \cap C_i|$ being $B(n, 1/h(i)$ binomial random variables. Thus (24) remains valid for the n points being chosen I.I.D. uniformly from C' as well and $\mathbf{E}(\mathrm{MAX}(\{p_1, \ldots, p_n\}) \cap C) = \theta(g(n))$. For the points in Q we use Lemma 3 to show that

$$\mathbf{E}(\mathrm{MAX}(\{p_1, \ldots, p_n\}) \cap Q) = O(\ln n).$$

Recall that $g(x) \ge \ln x$. Adding the maxima in Q and the maxima in C therefore gives $\mathbf{E}(\mathrm{MAX}(\{p_1, \ldots, p_n\})) = \theta(g(n))$.

Now assume that $c > 1$. We will construct a new monotonically increasing, continuous function $\hat{h}$ with $\sum_i 1/\hat{h}(i) \le 1$. This $\hat{h}$ will have the property that for all x greater than some N, $\hat{h}(x) = h(x)$. Thus, for $k > N$,

$$\sum_{i>k} 1/\hat{h}(i) = \sum_{i>k} 1/h(i) = O(g(n)/n).$$

Furthermore, for n large enough $\hat{h}^{-1}(n) = h^{-1}(n)$ so

$$k \ln n - \sum_{i \le k} \ln \hat{h}(i) = k \ln n - \sum_{i \le k} \ln h(i) + O(\ln n) = \theta(g(n))$$

These last sentences tell us that when h satisfies conditions 1 and 2 of the theorem $\hat{h}$ will satisfy them too. Substituting $\hat{h}$ for h in the last paragraph will give us a C such that $\mathbf{E}\left(M_n^C\right) = \theta(g(n))$ and we will be finished.

To find a function $\hat{h}$ with the above properties define

$$N = \min\{n : \sum_{i>n} 1/h(i) < 1\}.$$

Set $a = \min\left[h(N+1)/2,\, 1 - \sum_{i>N} 1/h(i)\right]$ and

$$\hat{h}(x) = \begin{cases} ax/N^2 & \text{if } x \le N \\ (h(N+1) - a/N)(x-N) + a & \text{if } N < x < N+1 \\ h(x) & \text{if } N+1 \le x. \end{cases}$$

This function is the concatenation of the line segment from $(0,0)$ to $(N, a/N)$, the segment from $(N, a/N)$ to $(N+1, h(N+1))$ and the curve $(x, h(x))$ where $x \ge N+1$. Thus $\hat{h}$ is a monotonically increasing continuous function. Furthermore $\sum_i 1/\hat{h}(i) \le aN/N + \sum_{i>N} 1/h(i) \le 1$ so we are done.

Part 2: We now show how to modify C to get a connected region C' with $\mathrm{E}\left(M_n^{C'}\right) = \theta(g(n))$. The plan will be to construct a new infinite sequence C_i' of squares of decreasing size. Combining these squares with the C_i of Part 1 we will set $C' = \bigcup_i (C_i \cup C_i')$. Figures 7 (b) and (c).

Construction of C' :

1. Set $f_i = 1/\sqrt{h(i)}$.
2. Set $x_1 = y_1 = 0$. For $i > 1$ set $x_i = x_{i-1} + f_{i-1}$ and $y_i = y_{i-1} - f_i$.
3. For $i \ge 1$ define C_i to be the square with sides of length f_i parallel to the x and y axes and whose lower left corner is (x_i, y_i). That is, C_i is the square with vertices (x_i, y_i), $(x_i + f_i, y_i)$ and $(x_i + f_i, y_i + f_i)$ and $(x_i, y_i + f_i)$.
4. For $i \ge 1$ define C_i' to be the square with sides of length f_{i+2} parallel to the x and y axes whose upper right corner is (x_i, y_i). That is, C_i' is the square with vertices (x_i, y_i), $(x_i - f_{i+2}, y_i)$ and $(x_i - f_{i+2}, y_i - f_{i+2})$ and $(x_i, y_i - f_{i+2})$.
5. Set $C = \bigcup_i (C_i \cup C_i')$.

Let $c' = Area(C') = 1/h(1) + 1/h(2) + 2\sum_i 1/h(i)$. Without loss of generality we will assume that $c = 1$. If not apply the transformations described at the end of Part 1.

The C_i, C_i' were defined so that if $p \in C_i$ and $q \in C' \setminus C_i$ then p and q are incomparable. Thus if $p_1, \ldots, p_n$ are chosen I.I.D. uniformly from C' then

$$\mathrm{E}(\mathrm{MAX}(p_1, \ldots, p_n)) \ge \sum_i \mathrm{E}(\mathrm{MAX}(S \cap C_i)).$$

The square C_i' is just a translated copy of the square C_{i-2} so we also have the easy upper bound

$$\begin{aligned} \mathrm{E}(\mathrm{MAX}(p_1, \ldots, p_n)) &\le \sum_i \mathrm{E}(\mathrm{MAX}(S \cap C_i)) \\ &\quad + \sum_i \mathrm{E}(\mathrm{MAX}(\{p_j \in C_i'\})) \qquad (25) \\ &\le 2\sum_i \mathrm{E}(\mathrm{MAX}(S \cap C_i)). \end{aligned}$$

Combining the last two equations gives

$$\mathrm{E}(\mathrm{MAX}(p_1, \ldots, p_n)) = O\left(\sum_i \mathrm{E}(\mathrm{MAX}(S \cap C_i))\right).$$

In part 1 we showed (equation (24)) that if n points are chosen I.I.D. uniformly from C then $\sum_i \mathbf{E}(\mathrm{MAX}(S \cap C_i)) = \theta(g(n))$. The derivation of this fact was only dependent upon the number of points in square C_i being a binomial $B(n, 1/h(i))$ random variable. In our current construction the number of points in square C_i is also a binomial $B(n, 1/h(i))$ random variable. Thus the same proof also shows that, for n points chosen I.I.D. uniformly from C',

$$\mathbf{E}(\mathrm{MAX}(p_1, \ldots, p_n)) = \theta\left(\sum_i \mathbf{E}(\mathrm{MAX}(S \cap C_i))\right) = \theta(g(n))$$

and we are done. ∎

4 Extensions

In this section we will discuss extensions to the results derived so far. In the first subsection we will sketch how to extend Theorem 1 so that it remains valid in higher dimensions. In the second subsection we will discuss how a theorem due to Luc Devroye gives order information about the higher moments $\mathbf{E}\left((M_n^C)^p\right)$. In the third subsection we will discuss why Theorems 1 and 2 remain valid if the points are chosen from Poisson distributions in place of I.I.D. uniform ones. We conclude, in the fourth subsection, by saying a few words as to why it might be difficult to find "simpler" regions C such that $\mathbf{E}\left(M_n^C\right) = \theta(g(n))$.

4.1 Higher Dimensions

Until now we have only considered maxima in the plane. In this section we will sketch the proof that Theorem 2 (and by extension Theorem 1) is valid in higher dimensions as well.

We must first extend the idea of maxima to higher dimensions. This is a quite natural generalization of the two dimensional definition. Suppose $p = (p.1, p.2, \ldots, p.d)$ and $q = (q.1, q.2, \ldots, q.d)$ are d-dimensional points. We say that p *dominates* q if $p.i \geq q.i$ for all $1 \leq i \leq d$. If S is a set of d dimensional points then the set of *maximal points* (or *maxima*) of S is

$$\mathrm{MAX}(S) = \{p \in S\ :\ p \text{ is not dominated by any } q \in S,\ q \neq p.\}$$

Let C be a d-dimensional region. Let $S = \{p_1, \ldots, p_n\}$ be a set of n points chosen I.I.D. uniformly from C. As in the two dimensional case we set $M_n^C = |\mathrm{MAX}(S)|$; the quantity that we are interested in is $\mathbf{E}\left(M_n^C\right)$. We claim that Theorem 2 remains valid in d dimensions, i.e.

Theorem 5 *Let g be a continuous monotonically increasing function from $(0, \infty)$ into itself with $g(x) \leq x$ and $\lim_{x\to\infty} g(x) = \infty$. Use g^{-1} to denote the functional inverse of g : $g\left(g^{-1}(x)\right) = g^{-1}\left(g(x)\right) = x$. Suppose that g fulfills the following condition for all integers $n > 0$:*

$$\sum_{i > g(n)} \frac{1}{g^{-1}(i)} = O\left(\frac{g(n)}{n}\right). \tag{26}$$

Then there exists a connected d-dimensional region C such that, for n points $p_1, \ldots, p_n$ chosen I.I.D. uniformly from C the expected number of the points that are maximal is $\theta(g(n))$,

$$\mathbf{E}|(\mathrm{MAX}(\{p_1, \ldots, p_n\})) = \theta(g(n)).$$

The proof of a d-dimensional analogue to Theorem 1 will follow because the original proof of Theorem 1 never used the planarity of the chosen points.

We will only sketch the proof of Theorem 5. Looking back at the proof of Part 1 of Theorem 2 we find that the entire proof flowed from the existence of a triangle T such that $\mathbf{E}\left(M_n^T\right) = \theta(1)$ (Lemma 1). Once we had such a T we constructed scaled copies C_i such that points in different C_i-s were incomparable and $Area(C_i) = \theta(1/g^{-i}(i))$ and then set $C = \bigcup C_i$.

We will now show that for every $d \geq 2$ there is a bounded d-dimensional polytope T_d such that $\mathbf{E}\left(M_n^{T_d}\right) = \theta(1)$. Following the paradigm laid down by the two dimensional case we then construct scaled copies, C_i, $i = 1, 2, 3, \ldots$, of T_d such that points in different C_i-s are incomparable and $VOL_d(C_i) = \theta(1/g^{-i}(i))$ (we use $VOL_d()$ for d-dimensional volume). Setting $C = \bigcup C_i$ we can prove that $\mathbf{E}\left(M_n^C\right) = \theta(g(n))$ in exactly the same way that we did in two dimensions. It will then remain to show how to modify C to get a *connected* C' such that $\mathbf{E}\left(M_n^{C'}\right) = \theta(g(n))$. This is done exactly as in Part 2 of the proof of Theorem 2; by extending the C_i into C_i' and attaching a hyperrectangle to the base of C_i (the T_d that we will define have flat bases). We omit the details.

We now show that there is a region T_d that satisfies our criterion.

Lemma 4 *Let T_d be the d-dimensional polytope with vertices $q_0, q_1, \ldots, q_d$ given by*

$$q_i.j = \begin{cases} i+1-j & \text{if } j < i \\ 0 & \text{otherwise.} \end{cases}$$

Then $\mathbf{E}\left(M_n^{T_d}\right) = \theta(1)$.

Here are the vertices of the first few T_d :

T_1	T_2	T_3	T_4
(0)	(0,0)	(0,0,0)	(0,0,0,0)
(1)	(1,0)	(1,0,0)	(1,0,0,0)
	(2,1)	(2,1,0)	(2,1,0,0)
		(3,2,1)	(3,2,1,0)
			(4,3,2,1)

Proof: Let T_d' be the ($d-1$ dimensional) polytope with vertices $q_1, \ldots, q_{d-1}$. This region can be thought of as a copy of the $(d-1)$-dimensional region T_{d-1} in the hyperplane $p.d = 0$ in d-space, i.e. a point $(q.1, q.2, \ldots, q.d-1) \in T_{d-1}$ if and only if the point $(q.1, q.2, \ldots, q_{d-1}, 0) \in T_d'$.

In this notation it is not difficult to convince oneself that T_d is the polytope with base T_d' and apex q_d. Thus every point $q = (q.1, q.2, \ldots, q.d) \in T_d$ has a unique representation $q = \alpha q_d + (1-\alpha)q'$ where $q' \in T_d'$ and $0 \leq \alpha \leq 1$. Further, since T_d' lies on the hyperplane $p.d = 0$ we must have $\alpha = q.d$.

Set $F_\alpha = \{(q.1, q.2, \ldots, q.d) \in T_d \ : \ q.d = \alpha\}$. This region is $(d-1)$-dimensional and its $(d-1)$-dimensional volume is $VOL_{d-1}(F_\alpha) = (1-\alpha)^{d-1} VOL_{d-1}(T_{d-1})$.

Set $R(q)$ to be the set of all points in T_d that dominate q :

$$R(q) = \{p \in T_d, \ p \text{ dominates } q\}$$

Now q_{d-1} dominates all points in F_0 so the point $\alpha q_d + (1-\alpha)q_{d-1}$ dominates all points in F_α. Therefore, for all $q \in F_\alpha$

$$R(q) \supseteq R(\alpha q_d + (1-\alpha)q_{d-1})$$

so

$$VOL_d(R(q)) \geq VOL_d(R(\alpha q_d + (1-\alpha)q_{d-1})) = (1-\alpha)^d VOL_d(R(q_d)).$$

We can now prove the lemma. Let $p_1, \ldots, p_n$ be chosen I.I.D. uniformly from T_d. We denote $p_1 = (p.1, p.2, \ldots, p.d)$. Then, using the above observations,

$$\begin{aligned}
\mathbf{E}\left(M_n^T\right) &= n \Pr(p_1 \text{ is maximal in } \{p_1, \ldots, p_n\}) \\
&= n \int\int_{T_d} \left(1 - \frac{VOL_d\left(R(p.1, p.2, \ldots, p.d)\right)}{VOL_d(T_d)}\right)^{n-1} dp.1 \ldots dp.d \\
&\leq n \int_{\alpha=0}^{1} VOL_{d-1}(F_\alpha) \left(1 - \frac{VOL_d(R(\alpha q_d + (1-\alpha)q_{d-1}))}{VOL_d(T_d)}\right)^{n-1} d\alpha
\end{aligned}$$

$$= nVOL_{d-1}(T_{d-1}) \int_{\alpha=0}^{1} (1-\alpha)^{d-1} \left(1 - (1-\alpha)^d \frac{VOL_d(R(q_{d-1}))}{VOL_d(T_d)}\right)^{n-1} d\alpha$$

$$= a_{d-1} n \int_0^1 \alpha^{d-1} \left(1 - \frac{b_d}{a_d}\alpha^d\right) d\alpha = O(1)$$

where $a_d = VOL_d(T_d)$ and $b_d = VOL_d(R(q_{d-1}))$ are constants. ■

4.2 Higher Moments

The regions C constructed in this paper were carefully tailored so that $\mathbf{E}\left(M_n^C\right) = \theta(g(n))$ for given monotonically increasing functions g. There is a rather remarkable theorem[1] due to Devroye [5] which, for $p > 1$, gives us the higher moments $\mathbf{E}\left(\left(M_n^C\right)^p\right)$ "for free". This theorem states that if $\mathbf{E}\left(M_n^C\right) = \theta(g(n))$ where g is nondecreasing then $\mathbf{E}\left(\left(M_n^C\right)^p\right) = \theta\left(\left(\mathbf{E}\left(M_n^C\right)\right)^p\right) = \theta\left(g^p(n)\right)$. Thus we know all of the higher moments of M_n^C.

As an example suppose C was constructed so that $\mathbf{E}\left(M_n^C\right) n = \theta\left(n^{1/3}\right)$. Then $\mathbf{E}\left(\left(M_n^C\right)^p\right) = \theta\left(n^{p/3}\right)$ for all $p \geq 1$.

4.3 Poisson Distributions

In this paper we have concentrated on the case where n points are chosen I.I.D. uniformly from some region C. In this subsection we shall briefly note what occurs if the points are chosen instead using a (spatial) Poisson distribution.

A random variable X has a Poisson distribution with parameter λ if X takes only non-negative integral values and $\Pr(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$.

Let R be a region and $Area(R)$ ($Vol(R)$) be its area (volume). Let S be a finite set of points. We use $\#(R)$ to represent the number of points in $R \cap S$.

The set of points S has a spatial Poisson distribution with parameter λ if [7]

1. The probability of finding k points in R is only dependent on $Area(R)$ and not on its shape.
2. $\Pr(\#(R) > 1)$ is small compared to $Area(R)$.
3. When R and R'' are disjoint regions then $\#(R)$ and $\#(R')$ are independent random variables.

If a distribution is a spatial Poisson distribution with parameter λ then it is well known that, for every region R, $\#(R)$ is Poisson with parameter $\lambda \cdot Area(R)$.

We claim that Theorem 2 (and thus Theorem 1) are true even when the points are chosen from a spatial Poisson distribution with parameter n. That is, let g be a function that satisfies the conditions of Theorem 2 (or 1) and let C be the region that the theorem constructs such that $\mathbf{E}\left(M_n^C\right) = \theta(g(n))$. Now let S_n be a set of points chosen from C using a spatial Poisson distribution with parameter $\lambda = n$. Then

$$\mathbf{E}(\mathrm{MAX}(S_n)) = \theta(g(n)).$$

The proof of this claim is almost exactly the same as that of Theorem 2 (Theorem 1) and we will not rederive it here.

[1] We should point out that what we use is not Devroye's actual theorem, which is very general, but rather his remark 3 ([5], page 112) following the theorem, which discusses the case that $\mathbf{E}\left(M_n^C\right) = \theta(g(n))$ where g is a nondecreasing function

4.4 Overkill

We conclude with a brief discussion as to whether the regions constructed by Theorem 2 must be as complicated as the ones we have constructed. Is it possible to find much "simpler" regions, perhaps even polygons, that can be tailored to satisfy $\mathbf{E}\left(M_n^C\right) = \theta(g(n))$? Did we engage in mathematical overkill in our constructions?

We do not have a formal proof but the following argument hints that the answer is probably no. It is shown in [9] that if a region C can be decomposed into a finite number of disjoint convex regions then either (i) $\mathbf{E}\left(M_n^C\right) = \sqrt{n}$ or (ii) $\mathbf{E}\left(M_n^C\right) = O(\ln n)$. Thus if $\mathbf{E}\left(M_n^C\right) = \theta(g(n))$ and g satisfies neither (i) or (ii) and the boundary of C is piecewise linear then the boundary must have an infinite number of sides.

Acknowledgements: The author thanks Luc Devroye and Rex Dwyer for their emailed comments concerning the expected number of maxima. It was Luc Devroye who suggested that there are regions with $\mathbf{E}\left(M_n^C\right)$ between $\ln n$ and $\sqrt{n}$. The author would also like to thank Philippe Dumas, Bruno Salvy and Philippe Flajolet for discussions about classes of functions that satisfy the conditions of Theorems 2 and 4. This paper is dedicated to the memory of the late great Miles Davis, who, because he was an hour late to his *Halle That Jazz* concert, gave me the time to start thinking about this problem.

References

[1] John L. Bentley, Kenneth L. Clarkson, and David B. Levine, "Fast Linear Expected-Time Algorithms for Computing Maxima and Convex Hulls," *Symposium on Discrete Algorithms (SODA)*, (1990).

[2] 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 Association for Computing Machinery*, **25**(4) October 1978, 536-543.

[3] Christian Buchta, "On the Average Number of Maxima in a Set of Vectors," *Information Processing Letters*, **33** 10 November, 1989, 63-65.

[4] Luc Devroye, *Lecture Notes on Bucket Algorithms*, Birkhauser Verlag, Boston, 1986.

[5] Luc Devroye, "Moment Inequalities for Random Variables in Computational Geometry," *Computing*, **30** 1983, 111-119.

[6] Rex Dwyer, "Kinder, Gentler, Average-Case Analysis for Convex Hulls and Maximal Vectors," *SIGACT NEWS*, **21**(2), Spring 1990, 64-71.

[7] William Feller, *An Introduction to Probability Theory and Its Applications: Volume I (3rd edition)*, John Wiley and Sons, New York, 1968.

[8] William Feller, *An Introduction to Probability Theory and its Applications: Volume II (2nd edition)*, John Wiley and Sons, New York, 1971.

[9] Mordecai Golin, "Maxima in Convex regions," unpublished manuscript.

[10] Torben Hagerup and Christine Rub, [1990] " A Guided Tour of Chernoff Bounds," *Information Processing Letters*, **33** 1989/90, 305-308.

[11] F. P. Preparata and M. I. Shamos, *Computational Geometry: An Introduction*, Springer-Verlag, New York, 1985.

ISSN 0249-6399