GROWTH RATES OF MINIMAL SPANNING TREES
OF MULTIVARIATE SAMPLES

BY

J. MICHAEL STEELE

TECHNICAL REPORT NO. 162
OCTOBER 1980

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MCS77-16974

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
GROWTH RATES OF MINIMAL SPANNING TREES OF MULTIVARIATE SAMPLES

BY

J. MICHAEL STEELE

TECHNICAL REPORT NO. 162
OCTOBER 1980

PREPARED UNDER THE AUSPICES OF
NATIONAL SCIENCE FOUNDATION GRANT
MCS 77-16974

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ABSTRACT

The length $M_n$ of the minimal spanning tree of $n$ independent identically distributed multivariate observations is studied. In particular, it is proved that for bounded observations one has with probability one that $M_n \sim c_d n^{(d-1)/d} \int_{\mathbb{R}^d} \phi(x)^{(d-1)/d} \, dx$. Here, $c_d$ is a universal constant depending only on the dimension $d \geq 2$, and $\phi(x)$ is the density of the absolutely continuous part of the distribution of the observations.

AMS 1980 Subject Classifications: Primary 60D05, Secondary 60F15, 62H30.

Key Words And Phrases: Minimal spanning tree, subadditive Euclidean functional, Efron-Stein inequality, Jackknife.
I. Introduction

The main objective of this article is to take some first steps toward developing the probability theory for the minimal spanning tree of \( n \) independent multivariate observations. The motivation for this work comes from two sources. The first of these is the pair of articles by Friedman and Rafsky (1979,1980) on the multivariate two-sample problem. The second source is the theory of subadditive Euclidean functionals.

In Friedman and Rafsky (1979) the minimal spanning tree is used to obtain what are perhaps the most natural multivariate analogues of the Wald-Wolfowitz, Kolmogorov-Smirnov, and other tests. In the 1980 article these authors further illustrate the statistical utility of the minimal spanning tree (MST) by applying variants of the mapping technique of Lee, Slagle, and Blum (1977) to several practical problems. Here, one should note that the earlier paper of Zahn (1971) also gives some compelling evidence of the data analytic utility of minimal spanning trees.

In the following we will be concerned with \( x_i \in \mathbb{R}^d \), \( 1 \leq i \leq n \); and, in particular, with the graphs \( G \) which have vertex set \( V = \{x_1, x_2, \ldots, x_n\} \). Any graph \( G \) is determined by \( E \), its set of edges; and \( E \) is just a subset of the pairs of elements of \( V \). The length of an edge \( e = (x_i, x_j) \in E \) will be denoted \( |e| \), and it is equal to the Euclidean distance \( \|x_i - x_j\| \).
The functional of central interest is $M(x_1, x_2, \ldots, x_n)$, the length of the minimal spanning tree of $\{x_1, x_2, \ldots, x_n\}$. This quantity is defined more precisely by

$$M(x_1, x_2, \ldots, x_n) = \min_T \sum_{e \in T} |e|,$$

where the minimum is over all connected graphs $T$ with vertex set $V$. Any $T$ which attains this minimum will be called a minimal spanning tree (MST) for $V$. (The graph theoretical terminology used here is hopefully self-explanatory, but in any case it follows the conventions of Lovász (1979) and Bondy and Murty (1976).)

In subsequent sections several results will be developed which are pertinent to small samples, but the principal focus of this article is on the proof of the following limit theorem.

**Theorem 1.** Suppose $X_i$, $1 \leq i < \infty$ are independent random variables with distribution $\mu$ and values in $\mathbb{R}^d$, $d \geq 2$. If $\mu$ has bounded support, then with probability one

$$\lim_{n \to \infty} n^{-(d-1)/d} M(X_1, X_2, \ldots, X_n) = c_d \int_{\mathbb{R}^d} f(x)^{(d-1)/d} \, dx.$$

Here, $c_d$ is a constant depending only on $d$ and $f(x)dx$ represents that part of $\mu$ which is absolutely continuous.

To illustrate the theorem, one should note that if $\mu$ has bounded support and is singular with respect to Lebesgue measure, then $M(X_1, X_2, \ldots, X_n) = o(n^{-(d-1)/d})$ a.s. Part of the appeal of this observation is that it indicates that the length of the MST is a
linear measurement which in a reasonable sense can measure the
dimension of the support of a distribution. This fact is already
suggestive of the notion that the MST might prove to be a useful
adjunct to principal component analyses and other "dimension
measuring" procedures.

Theorem 1 is very closely related to the theory of subaddi-
tive Euclidean functionals (Steele (1980)), but there are some
essential differences. The most pressing of these is that
\[ M_n = M(X_1, X_2, \ldots, X_n) \]
is not an almost surely increasing sequence of random variables. This forces several subtleties on \( M_n \) which are
absent in the study of the traveling salesman problem, the Steiner
tree problem, and other monotone Euclidean functionals.

Although the proof given here of Theorem 1 cuts a general
pattern, one cannot avoid developing some preliminary inequalities
which are special to the deterministic functional \( M(x_1, x_2, \ldots, x_n) \).
These inequalities are given in Section 2 where a unified treatment
is made possible by the systematic use of a distance counting
function.

In Sections 3 and 4 a proof of Theorem 1 is given under the
stringent hypothesis that \( \mu \) is the uniform distribution in \( [0,1]^d \).
A key tool used in the proof is the recent Jackknife inequality of
Efron-Stein (1979) which states that Tukey's Jackknife estimate of
variance is conservative in expectation.

Section 4 provides a general extension principle for multi-
variate functionals. This method is based on the Dudley-Strassen
embedding theorem, and it provides a widely applicable device for extending results for multivariate Euclidean statistics from special to general distributions. This section also contains the conclusion of the proof of Theorem 1.

In the final section there are some comments on some open problems. In particular, the apparently excessive requirement that $\mu$ have bounded support is discussed at that point.

**Acknowledgement.** Persi Diaconis, Ingram Olkin, and Joe Marhoul have been kind enough to comment on an earlier draft of this paper and their advice is gratefully acknowledged.
II. Applications of a Counting Function

Let $T$ denote a minimal spanning tree of $\{x_1, x_2, \ldots, x_n\} \subset [0,1]^d$; and let $\nu_d(x)$ denote the counting function of the edge lengths of $T$, i.e.,

$$\nu_d(x) \equiv \# \{ e \in T : |e| > x \} .$$

(2.1)

The following lemma will prove useful in many subsequent computations.

**Lemma 2.1.** There is a constant $\beta_d$ depending only on $d$ such that

$$\nu_d(x) \leq \beta_d x^{-d}$$

(2.2)

for all $0 \leq x < \infty$.

**Proof.** We first note that by the pigeon-hole principle one can easily show that there is a constant $\alpha_d$ such that from any set of $n$ points in $[0,1]^d$ one can select a pair $x_i, x_j$ with $|x_i - x_j| \leq \alpha_d n^{-1/d}$. (In fact, $\alpha_d = 2^{\sqrt{d}}$ will suffice.)

Next let $\mu(n)$ denote the maximum, overall possible choices of $n$ points in $[0,1]^d$ of the length of the corresponding MST. Since any $n$ set has two points within a distance $\alpha_d n^{-1/d}$, we have

$$\mu(n) \leq \mu(n-1) + \alpha_d n^{-1/d} ,$$

and consequently

$$\mu(n) \leq 2\alpha_d n^{(d-1)/d} .$$

(2.3)
But, by the definition of \( v_d(x) \), we can deduce that if there are \( n \) points in \([0,1]^d\), then \( x v_d(x) \leq \mu(n) \), and thus
\[
v_d(x) \leq 2a_d n^{(d-1)/d}/x.
\]
For \( x \geq n^{-1/d} \), this last inequality implies
\[
(2.4) \quad v_d(x) \leq 2a_d x^{-d}.
\]
Since for \( x \leq n^{-1/d} \) it is trivial that \( v(x) \leq n \leq 1/x^d \), the theorem is proved with \( \beta_d = \max(2a_d, 1) = 2a_d \).

The next lemma illustrates a generic application of inequality (2.2).

**Lemma 2.2.** If \( T \) is any MST of \( \{x_1, x_2, \ldots, x_n\} \subset [0,1]^d \), then there exists a constant \( \beta'_d \) such that
\[
(2.5) \quad \sum_{e \in T} |e|^\alpha \leq \beta'_d n^{1-\alpha/d} \quad \text{for} \quad 1 \leq \alpha < d
\]
\[
(2.6) \quad \sum_{e \in T} |e|^d \leq \beta'_d \log n
\]
and
\[
(2.7) \quad \sum_{|e| \geq y} |e| \leq \beta'_d y^{1-d} \quad \text{for} \quad 0 \leq y < \infty.
\]

**Proof.** For any \( \lambda > 0 \) we have for \( 1 \leq \alpha < d \)
\[ \sum |e|^\alpha = \sum_{e \in \mathcal{E}} |e|^\alpha + \sum_{e \notin \mathcal{E}} |e|^\alpha \]

\[ \leq n^{1-\lambda \alpha} + \int_{n^{-\lambda}}^d x^\alpha d(n-1-v_d(x)) \]

\[ \leq n^{1-\lambda \alpha} + n^{-\alpha \lambda} v_d(n^{-\lambda}) + \alpha \int_{n^{-\lambda}}^d x^{\alpha-1} v_d(x) dx. \]

Applying (2.2) and integrating we are lead to choose \( \lambda = d^{-1} \) which will yield (2.5). By analogous arguments one can verify (2.6) and (2.7).

In the Jackknife estimates used in Section IV we will need bounds on the change in \( M(x_1, x_2, \ldots, x_n) \) as an observation is deleted.

Following the customary Jackknife notation we set \( M(x_1, x_2, \ldots, \hat{x}_i, \ldots, x_n) = M(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \). Further, we let \( N(i) = \{ j : (x_i, x_j) \in \mathcal{E} \} \).

Lemma 2.3. For all \( x_i, 1 \leq i \leq n \), we have

\[ M(x_1, x_2, \ldots, x_n) \leq M(x_1, x_2, \ldots, \hat{x}_i, \ldots, x_n) + \min_{j: j \neq i} \| x_i - x_j \| \quad (2.8) \]

\[ M(x_1, x_2, \ldots, \hat{x}_i, \ldots, x_n) \leq M(x_1, x_2, \ldots, x_n) + \sum_{j \in N(i)} \| x_i - x_j \| \quad (2.9) \]

Proof. To prove (2.8) we just note that any tree which spans \( \{ x_1, x_2, \ldots, \hat{x}_i, \ldots, x_n \} \) can be completed to a connected graph spanning \( \{ x_1, x_2, \ldots, x_n \} \) by connected \( x_i \) to the nearest \( x_j, j \neq i \).

To prove (2.9) let \( T \) be a minimal spanning tree of \( \{ x_1, x_2, \ldots, x_n \} \), and let \( x' \) be the element of \( N(i) \) nearest \( x_i \). We get a connected graph spanning \( \{ x_1, x_2, \ldots, \hat{x}_i, \ldots, x_n \} \) by taking the edges
of $T$, deleting $\{(x_i^1, x_j^1) : (x_i^1, x_j^1) \in T\}$, and adding
$A = \{(x^1, x_j^1) : j \in N(i), x_j^1 \neq x^1\}$. The total length of the edges in $A$
can be bounded by

$$\sum_{e \in A} |e| \leq \sum_{j \in N(i)} (\|x_j^1 - x_1^1\| + \|x_1^1 - x^1\|) \leq 2 \sum_{j \in N(i)} \|x_j^1 - x_1^1\|,$$

and the length of the deleted edges equals $\sum_{j \in N(i)} \|x_j^1 - x_1^1\|$. $\Box$

To bound sums like that in (2.9) one needs bounds on $|N(i)|$, the cardinality of the set $N(i)$.

**Lemma 2.4.** There is a constant $D_d$ such that for any MST in $\mathbb{R}^d$, no
vertex has degree greater than $D_d$. In particular, $D_2 = 6$ and $D_3 = 13$.

This lemma is well known so no proof needs to be given. For
the values $D_2$ and $D_3$ one can consult Roberts (1969) or the book of
Hilbert and Cohn-Vossen (1952).

The main result of this section is the following continuity
lemma which will be essential in Section 5. The proof uses all of
the preceding techniques.

**Lemma 2.5.** There is a constant $\beta''_d$ such that for any two finite subsets $\chi$ and $\chi'$ of $[0,1]^d$, we have

$$(2.10) \quad |\mathcal{M}(\chi) - \mathcal{M}(\chi')| \leq \beta''_d |\chi \Delta \chi'|^{(d-1)/d}. $$

Here, $|\chi \Delta \chi'|$ denotes the cardinality of the symmetric difference
$(\chi \cup \chi') \setminus (\chi \cap \chi')$. 

8
Proof. By the same reasoning applied in the derivation of inequality (2.9) we obtain

\[
M(\chi') \leq M(\chi \cup \chi') + \sum_{x' \in \chi' \setminus \chi} \sum_{x \in N(x')} ||x - x'|| .
\]

We also have by Lemma 2.2 (Eq. (2.5) with \( a = 1 \)) that

\[
M(\chi \cup \chi') \leq M(\chi) + M(\chi' \setminus \chi) + \sqrt{d}
\]

\[
\leq M(\chi) + \beta d' |\chi' \setminus \chi|^{(d-1)/d} + \sqrt{d} .
\]

But, by Lemma 2.2 (Eq. (2.7)) and Lemma 2.4 applied to \( T \), the MST of \( \chi \cup \chi' \), we have for any \( y \geq 0 \)

\[
\sum_{x' \in \chi' \setminus \chi} \sum_{x \in N(x')} ||x - x'|| \leq \sum_{\{e: e \in T, |e| \geq y\}} |e| + y D_d |\chi \setminus \chi'|^{1-d} + y D_d |\chi \setminus \chi'| .
\]

Taking \( y = |\chi \setminus \chi'|^{-1/d} \) and combining (2.11), (2.12), and (2.13) we obtain (2.10) by the symmetry of \( \chi \) and \( \chi' \).
III. Growth in Mean

The asymptotics of \( m_n = \text{EM}(X_1, X_2, \ldots, X_n) \) where \( X_i \) are independent and uniform on \([0,1]^d\) can be established by applying a mild variant of the technique introduced for the Traveling Salesman Problem by Beardwood, Halton, and Hammersley (1959). A different Tauberian argument is used here, and some additional care is needed due to the fact that the random variables \( M(X_1, X_2, \ldots, X_n) \) are not necessarily an increasing sequence.

By \( \pi \) we denote a Poisson point process on \( \mathbb{R}^d \) with constant intensity equal to one. For any bounded Borel \( A \), \( \pi(A) \) is almost surely a finite set of points and \( M(\pi(A)) \) equals the length of the MST of \( \pi(A) \).

The main observation is that by setting \( \phi(t) = \text{EM}(\pi([0,t]^d)) \), we get a continuous function which satisfies

\[
(3.1) \quad \phi(t) \leq m^d \phi(t/m) + 4^d t.
\]

To prove (3.1) we note that if \([0,t]^d\) is divided into \( m^d \) subcubes \( Q_i \) of edge \( t/m \), then \( \text{EM}(\pi(Q_i)) = \phi(t/m) \) for all \( 1 \leq i \leq m^d \). Moreover, by taking the MST's of all of the \( \pi(Q_i) \), and by joining these (at a cost less than \( 4^d t \)), we have

\[
(3.2) \quad M(\pi([0,t]^d)) \leq \sum_{i=1}^{m^d} M(\pi(Q_i)) + 4^d t.
\]

On taking expectations (3.2) yields (3.1).
It is now an easy consequence of (3.1) and the continuity of \( \phi \) that for some constant \( c_d \) depending only on \( d \) that

\[
(3.3) \quad \phi(t) \sim c_d t^d \quad \text{as} \quad t \to \infty.
\]

The deduction of (3.3) from (3.2) under the assumption that \( \phi(t) \) is monotone is done in detail in Steele (1979), and the changes needed to replace monotonicity by continuity are not great enough to merit repetition.

By the definition of \( \phi(t) \) and the scaling property

\[
tM(x_1, x_2, \ldots, x_n) = M(tx_1, tx_2, \ldots, tx_n)
\]

of \( M \), one can compute by conditional expectations that

\[
(3.4) \quad \phi(t) = \sum_{n=0}^{\infty} m_n e^{-t^d/n} t^{dn}/n!.
\]

Here, \( m_n = M(X_1, X_2, \ldots, X_n) \) and the \( X_i \) are independent and uniform on \([0,1]^d\).

To extract the asymptotics of \( m_n \) from (3.3) and (3.4) we first note that \( m_n \) are practically monotone, i.e.,

\[
(3.5) \quad n m_{n-1} \leq (n+2)m_n.
\]

To prove (3.5) we sum (2.9) for \( 1 \leq i \leq n \) to get

\[
(3.6) \quad \sum_{i=1}^{n} M(X_1, X_2, \ldots, \hat{X}_i, \ldots, X_n) \leq n M(X_1, X_2, \ldots, X_n)
\]

\[
+ \sum_{i=1}^{n} \sum_{j \in N(i)} \|X_i - X_j\|,
\]

11
and note that the double sum equals $2M(X_1, X_2, \ldots, X_n)$. Taking expectations in (3.6) naturally proves (3.5).

The way we use (3.5) is by observing that one can use it to easily deduce that

$$n^2 m_n \geq (n-1)^2 m_{n-1} \quad \text{for all } n \geq 1.$$  \hfill (3.7)

The last tool required is the following differentiation lemma of Hardy and Littlewood (see, e.g., Widder (1946), pp. 193-194).

**Lemma 3.1.** If $f(x) \sim Ax^\alpha$ as $x \to 0$ and $f''(x) = 0(x^{\alpha-2})$ as $x \to 0$, then $f'(x) \sim \alpha Ax^{\alpha-1}$ as $x \to 0$.

We can now assemble the pieces to prove the main result of this section.

**Lemma 3.2.** If $X_i, 1 \leq i < \infty$, are uniform on $[0,1]^d$ and $m_n = EM(X_1, X_2, \ldots, X_n)$, then for a constant $c_d$,

$$m_n \sim c_d n^{(d-1)/d} \quad \text{as } n \to \infty.$$  \hfill (3.8)

**Proof.** By (3.3), (3.4), and the change of variables $t^d = u$ we have

$$\sum_{n=0}^\infty m_n e^{-u} u^n/n! \sim c_d u^{(d-1)/d} \quad \text{as } u \to \infty.$$  \hfill (3.9)

Taking the Laplace transform of the left sided of (3.9) and applying the Abelian theorem for Laplace transforms (Widder (1946), p. 181), we obtain for $\lambda \to 0$ that
\[ \sum_{n=0}^{\infty} m_n (1+\lambda)^{-n} \sim c_d \Gamma((2d-1)/d) \lambda^{-(2d-1)/d} \quad \text{as} \quad \lambda \to 0 \quad \text{or} \]

\[ \sum_{n=0}^{\infty} m_n \lambda^{-nx} \sim c_d \Gamma((2d-1)/d) \lambda^{-(2d-1)/d} \quad \text{as} \quad \lambda \to 0 \quad \text{as} \quad x \to 0 \quad \text{(3.10)} \]

From Lemma 2.2 we know \( m_n = O(n^{(d-1)/d}) \) so by elementary estimations we also have for \( k = 0, 1, 2, \ldots \) that

\[ \sum_{n=0}^{\infty} n^k m_n e^{-nx} = O(x^{-k-2+1/d}) \quad \text{as} \quad x \to 0 \quad \text{(3.11)} \]

Finally by Lemma 3.1 applied twice with (3.10) and (3.11) we have for \( \alpha = (2d-1)/d \)

\[ \sum_{n=0}^{\infty} n^2 m_n e^{-nx} \sim \alpha(\alpha-1)c_d \Gamma(\alpha)x^{-\alpha-2} \quad \text{as} \quad x \to 0 \quad \text{(3.12)} \]

By the substitution \( e^{-x} = y \), we have as \( y \to 1 \) that

\[ \sum_{n=0}^{\infty} n^2 m_n y^n \sim \alpha(\alpha+1)c_d \Gamma(\alpha)(1-y)^{-\alpha-2} \quad \text{(3.13)} \]

By the Karamata Tauberian theorem (Feller (1971), p. 447) we now have

\[ \sum_{k=0}^{n} k^2 m_k \sim \alpha(\alpha+1)c_d \Gamma(\alpha)n^{\alpha+2}/\Gamma(\alpha+3) = c_d n^{\alpha+2}/(\alpha+2) \quad \text{(3.14)} \]

The point of our maneuvering is that the series \( \{k^2 m_k\} \) is increasing by (3.7), and by a well-known lemma (cf. Apostol (1976), p. 280) the monotonicity of the terms of a sum justifies carrying over the asymptotics to the individual terms, i.e.,

\[ n^2 m_n \sim c_d n^{\alpha-1} \quad \text{(3.15)} \]
On dividing by $n^2$ and recalling that $\alpha = 1 - 1/d$ the main lemma is proved.

One should perhaps remark here that the repeated differentiation technique applied above can be used in very many types of problems where one would like to exploit the availability of some "modest" monotonicity property. The inequality (3.5) gives one example of such a property, but it is by no means the only one which can be handled by this method.
IV. Variance Bounds

Efron and Stein (1980) established the very useful fact that Tukey's Jackknife estimate of variance is conservative in expectation. Together with the geometric lemmas of Section 2, this fact will provide very effective bounds on $\text{Var} M_n$.

Suppose now that $S(x_1, x_2, \ldots, x_{n-1})$ is any symmetric function of $n-1$ vectors $x_i \in \mathbb{R}^d$. If $X_i$, $1 \leq i \leq n-1$ are independent identically distributed random vectors in $\mathbb{R}^d$, we define the new random variables $S_i = S(x_1, x_2, \ldots, x_{i-1}, X_{i+1}, \ldots, x_n)$ and $S_0 = \frac{1}{n} \sum_{i=1}^{n} S_i$. The Efron-Stein inequality states that

$$\text{(4.1)} \quad \text{Var} S(x_1, x_2, \ldots, x_{n-1}) \leq E \sum_{i=1}^{n} (S_i - S_0)^2.$$ 

Since the right side of (4.1) is not decreased if $S_0$ is replaced by any other variable, we can apply (4.1) in the MST problem to obtain

$$\text{Var} M_{n-1} \leq E \sum_{i=1}^{n} (M(x_1, x_2, \ldots, \hat{x}_i, \ldots, x_n) - M(x_1, x_2, \ldots, x_n))^2.$$ 

Assuming now that the $X_i$ are independent and uniformly distributed on $[0,1]^d$, it will be easy to bound the preceding sum.

To avoid concern over irrelevant constants we will use the Vinogradov symbol $a_n \ll b_n$ to denote that $a_n \leq C b_n$ for some $C$ not depending on $n$. By Lemma 2.3 and $(\max(a,b))^2 \leq a^2 + b^2$ we have

$$\text{Var} M_{n-1} \leq E \sum_{i=1}^{n} \min_{j:j\neq i} \|x_i - x_j\|^2 + E \sum_{i=1}^{n} \left( \sum_{j \in \mathbb{N}(i)} \|x_i - x_j\|^2 \right)^2.$$
By elementary calculus, \( E \min_{j:j \neq 1} ||X_j - X_j||^2 \ll n^{-2/d} \), and by Lemma 2.4 plus Schwarz' inequality one gets

\[
\text{Var } M_{n-1} \ll n^{1-2/d} + E \sum_{i=1}^{n} \mathbb{E}_{j \in \mathbb{N}(i)} ||X_j - X_i||^2.
\]

By Lemma 2.2 (Eq. (2.5) and (2.6)) we can complete the proof of the following lemma.

**Lemma 4.1.** If \( X_i \), \( 1 \leq i < \infty \), are independent and uniform on \([0,1]^d\), then

(4.2) \[
\text{Var } M_n \ll \log n \quad \text{for } d = 2
\]

and

(4.3) \[
\text{Var } M_n \ll n^{1-2/d} \quad \text{for } d > 2.
\]

To show that \( n^{-(d-1)/d}(M_n - m_n) \) converges almost surely, one naturally uses a subsequence argument. For \( \lambda \geq 0 \) we set \( n_k = [k^\lambda] \) and note by Lemma (4.1) that

(4.4) \[
\sum_{k=1}^{\infty} \text{Var}(M_{n_k} / n_k^{(d-1)/d}) < \infty
\]

provided \( \lambda > d \).

To control the variability on the intervals \( [n_k, n_{k+1}] \) we consider \( V_k = \max_{n_k < m < n_{k+1}} |M_m - M_{n_k}| \). But, by Lemma 2.5, for

\[
n_k \leq n \leq n_{k+1}
\]
\[ |M_n - M_{n_k}| \leq \beta d! \| \{X_n, X_{n-1}, \ldots, X_{n_k+1} \} \|^{(d-1)/d} \ll k^{(\lambda-1)(d-1)/d} \]

and

\[(4.5) \quad \frac{V_k}{n_k^{(d-1)/d}} \ll k^{-(d-1)/d} .\]

From (4.4), (4.5), and the fact from Section 3 that
\[ E M_n \sim c_d n^{(d-1)/d} \] a standard argument suffices to complete the proof of the following theorem, which is the main result of this section.

As a summary, we give its statement in detail.

**Theorem 2.** If \( X_i, 1 \leq i < \infty \) are uniformly distributed on \([0,1]^d\) and \( M_n \) is the length of the MST of \( \{X_1, X_2, \ldots, X_n\} \), then there is a constant \( c_d \) such that with probability one

\[(4.6) \quad M_n \sim c_d n^{(d-1)/d} \quad \text{as} \quad n \to \infty .\]
V. General Extention Principle

We will now show that the results obtained so far under the strong assumption of uniform distribution can be extended to any bounded distribution. While the technique employed here does not seem to yield a simple formulation when cast as a theorem, it still seems inevitable that the method will prove useful in many related problems. For this reason some of the intermediate results are given in more generality than is required by the immediate goal.

The main idea is that a type of continuity shared the MST and related functionals seems to work naturally in concert with an almost sure embedding technique of Strassen. The only tool we need is the following.

Lemma 5.1 (Strassen (1965)). Suppose P and Q are probability measures on a bounded subset of \( \mathbb{R}^d \), and suppose also that they satisfy \( P(F) \leq Q(F) + \epsilon \) for all closed \( F \). There is then a probability measure \( \mu \) on \( \mathbb{R}^d \times \mathbb{R}^d \) such that

\[
\mu(\cdot, \mathbb{R}^d) = P(\cdot), \quad \mu(\mathbb{R}^d, \cdot) = Q(\cdot), \quad \text{and} \quad \mu\{(x,y) : x \neq y\} \leq \epsilon.
\]

For an elegant proof of this lemma one can consult Dudley (1968), or better yet, Dudley (1976).

Suppose now that \( S(x_1, x_2, \ldots, x_n) \) is any sequence \( n = 1, 2, \ldots \) of functionals on \( [0,1]^d \). Also let \( S \) denote the class of probability measures on \( [0,1]^d \) which are of the form \( \phi(x)dx + d\mu_S \) where \( \phi(x) = \sum_{i=1}^{m} \alpha_i 1_{Q_i} \). Here we require that \( m < \infty \) and the \( Q_i \) are
disjoint cubes with edges parallel to the axes. Also, \( \mu_s \) denotes a measure on \([0,1]^d\) which is purely singular, i.e., \( \mu_s([0,1]^d) = \mu_s(A) \) for some measurable \( A \) of Lebesgue measure zero.

**Theorem 3.** Suppose that there is a constant \( B \) such that 
\[
S(x_1, x_2, \ldots, x_n) \text{ satisfies the continuity condition}
\]
\[
|S(x_1, x_2, \ldots, x_n) - S(x'_1, x'_2, \ldots, x'_n)| \leq B |\{i : x_i \neq x'_i\}|^{(d-1)/d}
\]
for all \( x, x' \) in \([0,1]^d\). Suppose also that with probability one
\[
S(x_1, x_2, \ldots, x_n) \sim n^{(d-1)/d} c_d \int f(x)^{(d-1)/d} dx
\]
whenever \( \{X_i\} \) are independent and identically distributed with respect to any probability measure \( \mu = \mu_s + \phi(x)dx \) in \( S \). With probability one, one then has
\[
S(x'_1, x'_2, \ldots, x'_n) \sim n^{(d-1)/d} c_d \int f(x)^{(d-1)/d} dx
\]
whenever \( \{X'_i\} \) are independent and identically distributed with respect to any probability measure on \([0,1]^d\) with an absolutely continuous part given by \( f(x)dx \).

**Proof.** If the \( X'_i \) are distributed according to \( f(x)dx + \mu_s \) where \( \mu_s \) is singular, we take the approximation \( \phi_m(x)dx + \mu_s \) where 
\[
\phi_m(x) = \sum_{i=1}^m \alpha_i 1_{Q_i}.
\]
Here, each \( Q_i \) is one of the subcubes obtained by partitioning \([0,1]^d\) into \( m^d \) parts, and \( \alpha_i \) are defined by 
\[
\alpha_i = \int_{Q_i} f(x)dx.
\]
It is well known that in this case 
\[
\int |\phi_m(x) - f(x)| dx \to 0 \text{ as } m \to \infty.
\]
(To prove this one can first
convolve $f$ with a $C^\infty$-kernel with compact support and then use easy
calculus.) Hence, by setting

$$P(A) = \int_A f(x) \, dx + \mu_S(A) \quad \text{and} \quad Q(A) = \int_A \phi_m(x) \, dx + \mu_S(A),$$

we have $|P(A) - Q(A)| \leq \int_A |f(x) - \phi_m(x)| \, dx \leq \varepsilon$ for all $m$ sufficiently
large.

By Lemma 5.1 there is thus a probability measure $\mu$ on
$\mathbb{R}^d \times \mathbb{R}^d$ with marginals $P$ and $Q$ such that

$$\mu\{ (x,y) : x \neq y \} \leq \varepsilon.$$

We can now define random vectors $(X_i, X'_i)$, $1 \leq i < \infty$, by
taking an independent sequence generated by the measure $\mu$. By the
law of large numbers, with probability one,

$$|\{i : X_i \neq X'_i\}| \sim n \mu(\{(x,y) : x = y\}) \leq \varepsilon n.$$

By condition (5.1) we then have (5.3) which holds that

$$|S(X_1, X_2, \ldots, X_n) - S(X'_1, X'_2, \ldots, X'_n)| \leq B(\varepsilon n)^{(d-1)/d}.$$

By (5.2) this implies,

$$\limsup_{n \to \infty} S(X_1, X_2, \ldots, X_n)_{n^{-(d-1)/d}} \leq B \varepsilon^{(d-1)/d} + \limsup_{n \to \infty} S(X'_1, X'_2, \ldots, X'_n)_{n^{-(d-1)/d}} \leq B \varepsilon^{(d-1)/d} + c_d \int \phi(x)^{(d-1)/d} \, dx$$

a.s.
Since $m$ and $\epsilon > 0$ are arbitrary we have,

$$\limsup_{n \to \infty} S(X_1, X_2, \ldots, X_n)^{-(d-1)/d} \leq c_d \int f(x)^{(d-1)/d} \, dx \quad \text{a.s.}$$

Since the limit infimum can be dealt with similarly, the theorem is proved. \[ \square \]

We will now apply Theorem 3 to minimal spanning trees. By Lemma 2.5 we see that inequality (5.1) is satisfied. The remainder of this section is devoted to the verification of (5.2).

Take any $\phi(x)dx + \mu_s$ in $S$, and let $E$ denote the support of $\mu_s$. In the following, according to convenience, Lebesgue measure will be denoted by $\lambda(\cdot)$ or by $dx$. Now, since the Lebesgue measure of $E$ equals 0, and since $\phi$ is constant on a set of subcubes, we can find a finer partition $\{Q_i\}_{i \in I}$ of $[0,1]^d$ into cubes such that the following properties hold:

(5.4) On each $Q_i$, $i \in I$, $\phi(\cdot)$ is constant

(5.5) For all $i \in I$, $\lambda = \lambda(Q_i)^{1/d} < \epsilon$

and

(5.6) $E \subseteq A \cup B$ where $A$ and $B$ are disjoint, $\lambda(A) = 0$, $P(X_1 \in A) = \mu_s(A) \leq \epsilon$, and

$$B = \bigcup_{j \in J} Q_j \quad \text{where} \quad \lambda(B) = \sum_{j \in J} \lambda(Q_j) \leq \epsilon$$

We now set $C = [0,1]^d \setminus (A \cup B)$, and note

(5.7) $M(X_1, X_2, \ldots, X_n) \leq 2\sqrt{d} + M(\{X_1 : X_1 \in A\}) + M(\{X_1 : X_1 \in B\})$

$$+ M(\{X_1 : X_1 \in C\})$$

21
Since $P(X_i \in A) \leq \varepsilon$ we have

$$
\limsup_{n \to \infty} M(\{X_i : X_i \in A\}) n^{-1/(d-1)/d} \leq \beta_d' \varepsilon^{(d-1)/d}
$$

by Lemma 2.2 (Eq. (2.5)) and the law of large numbers. Similarly, on $B$ we use Lemma 2.2 with rescaling and Schwarz' inequality to obtain

(5.8) \quad $M(\{X_i : X_i \in B\}) \leq |J|^{1/d} + \sum_{j \in J} M(\{X_i : X_i \in Q_j\})$

$$
\leq |J|^{1/d} + \beta_d' \varepsilon \sum_{j \in J} |\{X_i : X_i \in Q_j\}|^{1/(d-1)/d}
$$

$$
\leq |J|^{1/d} + \beta_d' \varepsilon |J|^{1/d} n^{-1/(d-1)/d}.
$$

By (5.6), $\sum_{j \in J} \lambda(Q_j) \leq \varepsilon$ so $|J|^{d} \leq \varepsilon$ and (5.8) becomes

$$
\limsup_{n \to \infty} M(\{X_i : X_i \in B\}) n^{-1/(d-1)/d} \leq \beta_d' \varepsilon^{1/d}.
$$

Writing $J'$ for the complement of $J$, we now handle the more substantial contribution of $C$. We write $\overline{Q}_j = Q_j \setminus (A \cup B)$ and note

$$
C = \bigcup_{j \in J'} \overline{Q}_j
$$

so

(5.9) \quad $M(\{X_i : X_i \in C\}) \leq \sqrt{d} |J'| + \sum_{j \in J'} M(\{X_i : X_i \in \overline{Q}_j\})$.

Setting $\gamma_j \equiv \int_{\overline{Q}_j} \phi(x) dx = P(X_i \in \overline{Q}_j)$ we note that the constancy of $\phi$ on $Q_i$ implies $\phi(x) \equiv \gamma_j x^{-d}$ on $Q_j$.

The conditional random variable $Y_i = \{X_i | X_i \in \overline{Q}_j\}$ is uniform on $\overline{Q}_j$, so a scaled version of Theorem 2 and the law of large numbers will yield a.s. that
(5.10) \( M(\{ X_i : X_i \in \overline{Q}_j \}) \sim c_d \ell |\{ X_i : X_i \in \overline{Q}_j \}|^{(d-1)/d} \sim c_d \ell (\gamma_j n)^{(d-1)/d}. \)

Summing (5.10) and applying the previous bounds on the contributions of \( A \) and \( B \), we have

(5.11) \( \lim_{n \to \infty} \sup \frac{M(X_1, X_2, \ldots, X_n)}{n^{-(d-1)/d}} \leq \beta_d \left( \epsilon^{(d-1)/d} + \epsilon^d \right) \)

\[ + c_d \sum_{j \in J} \gamma_j^{(d-1)/d} \ell. \]

The arbitrariness of \( \epsilon > 0 \) and the bound \( \sum_{j \in J} \lambda(Q_j) \leq \epsilon \) applied in (5.11) imply that for any \( \phi(x)dx + \mu_s \) in \( S \) we have

(5.12) \( \lim_{n \to \infty} \sup \frac{M(X_1, X_2, \ldots, X_n)}{n^{-(d-1)/d}} \leq c_d \int \phi(x)^{(d-1)/d} dx. \)

The corresponding lower bound on the limit inferior is just a bit more subtle.

With \( A, B, \) and \( C \) as before we note that by Lemma 2.5 we have

(5.13) \( M(X_1, X_2, \ldots, X_n) \geq M(\{ X_i : X_i \in B \cup C \}) - \beta_d'' |\{ X_i : X_i \in A \}|^{(d-1)/d}. \)

Now for any \( T \) which is a MST of \( \{ X_i : X_i \in B \cup C \} \) we let \( I \) denote those elements of \( \{ X_i : X_i \in C \} \) which are joined by an edge of \( T \) to an element of \( \{ X_i : X_i \in B \} \).

The MST's of \( \{ X_i : X_i \in B \cup C \} \) and \( I \) will together span \( \{ X_i : X_i \in C \} \) so

(5.14) \[ M(\{ X_i : X_i \in C \}) \leq M(\{ X_i : X_i \in B \cup C \}) + M(I). \]

To bound \( |I| \) we note that each element of \( I \) is either an endpoint of an edge of length greater than \( y \) or else is a point of \( C \) within a
distance $y$ of $B$, thus

\[(5.15) \quad |X| \leq v_d(y) + \{x_1 : x_1 \in C, \min_{\omega \in B} |x_1 - \omega| \leq y\} \] .

We have by elementary probability that

\[P(X_1 \in C \text{ and } \min_{\omega \in B} |x_1 - \omega| \leq y) \leq (\max_{x} \phi(x)) \lambda^{d-1} y |J| ;\]

and by (5.6) we have $\sum_{j \in J} \lambda(Q_j) = \lambda^d |J| \leq \varepsilon$, so

\[(5.16) \quad P(X_1 \in C \text{ and } \min_{\omega \in B} |x_1 - \omega| \leq y) \leq (\max_{x} \phi(x)) \in \varepsilon \lambda^{-1} .\]

By Lemma 2.2 applied to $M(x)$ and the law of large numbers applied to (5.16) and (5.15), we have

\[(5.17) \quad \limsup_{n \to \infty} M(\mathcal{X})n^{-(d-1)/d} \leq 2\beta_d''((\max_{x} \phi(x)) \in \varepsilon \lambda^{-1})^{(d-1)/d} .\]

By (5.13), (5.14), and (5.16) we have

\[(5.18) \quad \liminf_{n \to \infty} M(X_1, X_2, \ldots, X_n)n^{-(d-1)/d} \geq \liminf_{n \to \infty} M(\{X_1 : x_1 \in C\})n^{-(d-1)/d} \]

\[- \beta_d'' \varepsilon - 2\beta_d''((\max_{x} \phi(x)) \in \varepsilon \lambda^{-1})^{(d-1)/d} .\]

By first choosing $\varepsilon$ and then choosing $y$ depending on the size of $\varepsilon \lambda^{-1}$ we see that

\[(5.19) \quad \liminf_{n \to \infty} M(X_1, X_2, \ldots, X_n)n^{-(d-1)/d} \geq \liminf_{n \to \infty} M(\{X_1 : x_1 \in C\})n^{-(d-1)/d} .\]

24
So again the problem is reduced to calculating the contribution from $C$.

We now let $Q_j = Q_j \setminus A$ and note $C = \bigcup_{j \in J'} Q_j$ where $J'$ is the complement of $J$. Take a spanning tree $T$ of $\{X_i : X_i \in C\}$ and for each $j \in J'$ let $D_j$ denote the set of edges $e \in T$ such that both end points of $e$ are in $Q_j$. Also, let $\gamma_j$ denote the set of points in $Q_j$ which are joined by an edge of $T$ to a point in $Q_j^c$. Since $D_j$ together with a MST of $\gamma_j$ will span $\{X_i : X_i \in Q_j \}$ we have

$$M(\{X_i : X_i \in Q_j \}) \leq \sum_{e \in D_j} |e| + \beta_d' \frac{|\gamma_j|^{(d-1)/d}}{d}$$

which yields after summing over $J'$ that

$$\sum_{j \in J'} M(\{X_i : X_i \in Q_j \}) \leq M(\{X_j : X_j \in C\}) + \beta_d' \sum_{j \in J'} \frac{|\gamma_j|^{(d-1)/d}}{d}.\quad (5.21)$$

To handle the last sum we note as in (5.15)

$$\sum_{j \in J'} |\gamma_j| \leq m_d(y) + \sum_{j \in J'} \frac{\min_{\omega \in Q_j^c} |X_i - \omega| \leq y|}{\omega \in Q_j^c},\quad (5.22)$$

and as in (5.16)

$$P(X_i \in Q_j^c \text{ and } \min_{\omega \in Q_j^c} |X_i - \omega| \leq y) \leq (\max \phi(x))y^{d-1}.\quad (5.23)$$

By (5.22), (5.23), Hölder's inequality and the law of large numbers

$$\limsup_{n \to \infty} \left[ \sum_{j \in J'} \frac{|\gamma_j|^{(d-1)/d}}{d} \right]^{(d-1)/d} \leq |J'|^{(d-1)/d} \left( \max \phi(x) \right)^{(d-1)/d} \frac{\kappa^d}{d}.\quad (5.24)$$
Since $y$ was arbitrary, (5.24) and (5.21) imply

\[
(5.25) \quad \liminf_{n \to \infty} M\{X_i : X_i \in C\} n^{-(d-1)/d} \geq \liminf_{n \to \infty} \sum_{j \in J'} M\{X_i : X_i \in \overline{Q}_j\},
\]

but, for each $\overline{Q}_j$, (5.10) says the limit exists so summing (5.25) we have

\[
(5.26) \quad \liminf_{n \to \infty} M\{X_i : X_i \in C\} n^{-(d-1)/d} \geq c_d \int_C \phi(x)^{(d-1)/d} \, dx.
\]

Since $C$ is all but $\epsilon$ of $[0,1]^d$, (5.26) completes the proof of the lower bound corresponding to (5.12). Moreover, this completes the verification of condition (5.2) of Theorem 3 and therefore completes the proof of Theorem 1.
VI. Concluding Remarks

There are two classes of problems which are open for almost all subadditive Euclidean functionals. These concern rates of convergence and the value of the limiting constants.

From the bounds on the variance of Section 3 one can make some progress on questions of rates of convergence, but a definative result seems very difficult. As for the constants $c_d$, some bounds can be deduced from the work of Gilbert (1965) on minimal exodic trees or from bounds on the corresponding constants in the traveling salesman problem. The exact analytical determination of $c_d$ would seem to require a miracle, but for any practical purpose $c_d$ can be computed as precisely as one likes by Monte-Carlo methods (at least for small $d$).

After the establishment of a strong law like Theorem 1, it is natural to inquire after the possibility of distributional results and an associated "central limit" theorem. As yet, there are no known asymptotic distributional results for any subadditive Euclidean functional.

Finally, there is the problem of removing the assumption that $\mu$ have bounded support. If $\mu$ fails to have a second moment, one can easily show that Theorem 1 need no longer hold. Elementary truncations seem to yield no results unless one first obtains a rate of convergence result to supplement Theorem 2; but as noted above, such rate results do not seem to be easily won.
REFERENCES


