*The Annals of Probability*
1988, Vol. 16, No. 4, 1767–1787

# GROWTH RATES OF EUCLIDEAN MINIMAL SPANNING TREES WITH POWER WEIGHTED EDGES[1]

BY J. MICHAEL STEELE

*Princeton University*

Let $X_i$, $1 \le i < \infty$, denote independent random variables with values in $\mathbb{R}^d$, $d \ge 2$, and let $M_n$ denote the cost of a minimal spanning tree of a complete graph with vertex set $\{X_1, X_2, \dots, X_n\}$, where the cost of an edge $(X_i, X_j)$ is given by $\psi(|X_i - X_j|)$. Here $|X_i - X_j|$ denotes the Euclidean distance between $X_i$ and $X_j$ and $\psi$ is a monotone function. For bounded random variables and $0 < \alpha < d$, it is proved that as $n \to \infty$ one has $M_n \sim c(\alpha, d) n^{(d-\alpha)/d} \int_{\mathbb{R}^d} f(x)^{(d-\alpha)/d}\, dx$ with probability 1, provided $\psi(x) \sim x^\alpha$ as $x \to 0$. Here $f(x)$ is the density of the absolutely continuous part of the distribution of the $\{X_i\}$.

**1. Introduction.** The main issue pursued here is the development of the probability theory for the minimal spanning tree of $n$ independent multivariate observations. For $x_i \in \mathbb{R}^d$, $1 \le i \le n$, we will be concerned with graphs $G$ which have vertex set $V = \{x_1, x_2, \dots, x_n\}$ and edge set $E = \{(x_i, x_j): 1 \le i < j \le n\}$. Here the length of an edge $e = (x_i, x_j) \in E$ will be denoted by $|e|$, where $|e| = |x_i - x_j|$ equals the Euclidean distance from $x_i$ to $x_j$.

The functional of interest is $M(x_1, x_2, \dots, x_n)$, the weight of the minimal spanning tree of $V = \{x_1, x_2, \dots, x_n\}$, where the weight assigned to edge $e$ is $\psi(|e|)$. More precisely, we focus on

$$M(x_1, x_2, \dots, x_n) = \min_T \sum_{e \in T} \psi(|e|), \tag{1.1}$$

where the minimum is over all connected graphs $T$ with vertex set $V$. The weighting function $\psi: [0, \infty) \to [0, \infty)$ which is of most interest to us is $\psi(x) = x^\alpha$, where $0 < \alpha < d$. More general $\psi$ are still of interest and some of the analysis which follows assumes only that $\psi \ge 0$. When additional properties of $\psi$ are required, those properties are made explicit.

Any tree $T$ which attains the minimum in (1.1) will be called a minimal spanning tree (MST) for $V$. In almost all of the situations considered here the minimal spanning tree is unique; but, to avoid consideration of special cases, we never rely on that uniqueness.

Subsequent sections will develop several results which pertain to finite values of $n$, but the main result is the limit theorem:

THEOREM 1. *Suppose $X_i$, $1 \le i < \infty$, are independent random variables with distribution $\mu$ having compact support in $\mathbb{R}^d$, $d \ge 2$. If the monotone*

Received September 1986; revised December 1987.
[1]Research supported in part by National Science Foundation Grant DMS-84-14069.
*AMS* 1980 *subject classifications.* Primary 60F15; secondary 05C05.
*Key words and phrases.* Minimal spanning trees, subadditive processes.

*function $\psi$ satisfies $\psi(x) \sim x^\alpha$ as $x \to 0$ for some $0 < \alpha < d$, then with probability* 1

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

*Here $f$ denotes the density of the absolutely continuous part of $\mu$ and $c(\alpha, d)$ denotes a strictly positive constant which depends only on the power $\alpha$ and the dimension $d$.*

For the case of ordinary length $\psi(x) = |x|$, Beardwood, Halton and Hammersley (1959) state that the preceding result follows from modifications of their analysis of the traveling salesman problem. The proof of Theorem 1 which is given here uses several tools which were not applied in Beardwood, Halton and Hammersley (1959). Perhaps largely for this reason, the proof given here for the general result is more direct and less taxing than the original analysis of Beardwood, Halton and Hammersley (1959) of the asymptotics of the traveling salesman problem. Still, many of the insights provided in Beardwood, Halton and Hammersley (1959) have helped guide this analysis of the minimal spanning tree.

One feature of Theorem 1 that should be noted is that if $\mu$ has bounded support and $\mu$ is singular with respect to Lebesgue measure, then we have with probability 1 that $M(X_1, X_2, \ldots, X_n) = o(n^{(d-\alpha)/d})$. Part of the appeal of this observation is the indication that the length of the minimal spanning tree is a measure of the *dimension* of the support of a distribution. This suggests that the asymptotic behavior of the minimal spanning tree might be a useful adjunct to the concept of dimension in the modeling applications and analysis of fractals; see, e.g., Mandelbrot (1977).

Theorem 1 is closely related to the theory of subadditive Euclidean functionals [Steele (1981a)], but there are some essential differences. One issue is that $M_n = M(X_1, X_2, \ldots, X_n)$ is not an almost surely increasing sequence of random variables. This fact forces subtleties on $M_n$ which are absent in the study of the traveling salesman problem, the Steiner tree problem and other monotone Euclidean functionals.

The technique applied to prove Theorem 1 was also useful in studying optimal triangulations and the directed traveling salesman problem [Steele (1982, 1986)]. Still, the minimal spanning tree function has its own unique features and, apparently, one cannot avoid developing some results which are special to the geometry of the minimal spanning tree. Many of the special features of the deterministic functional $M(x_1, x_2, \ldots, x_n)$ are expressed by the inequalities given in Section 2. The unified treatment of those inequalities is made possible by the systematic use of a distance counting function and the pattern used in the analysis of the counting function might prove useful in other geometric problems.

In Sections 3 and 4, a proof of Theorem 1 is given under the hypothesis that $\mu$ is the uniform distribution in $[0,1]^d$. A key tool used in the proof is the jackknife inequality of Efron–Stein (1981). That inequality states (approximately) that

Tukey's jackknife estimate of variance is conservative in expectation. From the point of view of combinatorial and geometric probability, the virtue of the Efron–Stein inequality is that it shows how to bound the variance of a multivariate function of independent variables when one understands how the function changes as *one* of the arguments of the function is varied.

Section 5 provides a general method for extending results for Euclidean functionals from the uniform case to arbitrary distributions. This extension method is easier and more direct than the methods used in either Beardwood, Halton and Hammersley (1959) or Steele (1981a). The idea behind the extension technique is closely linked to almost sure embedding, and an important tool in the technique is the embedding lemma of Strassen (1965).

The final section reports on those aspects of the probability theory of minimal spanning trees which could not be resolved by the methods of this paper. In particular, effort is made to lay out some open problems of interest.

**2. A counting function and its applications.** Let $T_0$ denote a minimal spanning tree of $\{x_1, x_2, \dots, x_n\} \subset [0,1]^d$, where the edge weighting function $\psi$ is taken to be $\psi_0(x) = |x|$. Also, let $\nu_d(x)$ denote the counting function of the edge lengths of $T_0$, i.e., $\nu_d(x)$ is equal to the cardinality of the set

$$\{e \in T_0 \colon |e| > x\}. \tag{2.1}$$

Lemma 2.1 does a good job of capturing the most basic information on the distribution of the edge lengths. An interesting feature of the bound is that it does not depend on $n$.

LEMMA 2.1. *There is a constant $\beta_d$ depending only on the dimension $d \geq 1$ such that*

$$\nu_d(x) \leq \beta_d x^{-d} \tag{2.2}$$

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

PROOF. We first note by the pigeonhole principle that there is a constant $\alpha_d$ such that from any set of $k$ points $\{x_1, x_2, \dots, x_k\}$ in $[0,1]^d$, one can select a pair of points $x_i$ and $x_j$ with $|x_i - x_j| \leq \alpha_d k^{-1/d}$. With a little care, one can show that $\alpha_d = 2\sqrt{d}$ will suffice, but we will not need to be concerned with precise constants either here or subsequently.

We now suppose that $\{x_1, x_2, \dots, x_n\}$ are any $n$ points in $[0,1]^d$ and let $E$ denote the set of edges of a MST for $\{x_1, x_2, \dots, x_n\}$. We recall Prim's algorithm for constructing a MST [see, e.g., Prim (1957) or Papadimitriou and Steiglitz (1982), page 273] can be expressed as (1) initially join the two nearest points and (2) iteratively join the two nearest connected components. Now if $e_j$, $1 \leq j < n$, are the edges of a MST for $\{x_1, x_2, \dots, x_n\}$ constructed by Prim's algorithm and the edges are listed in the order chosen by the algorithm, we have the bound

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

since, when $e_j$ is chosen, there are exactly $n - j + 1$ connected components of the forest produced up to that time by the algorithm.

Now if $e_{i_j}$, $1 \le j \le k$, are any $k$ edges of $E$, the fact that the edges are chosen in monotone increasing order gives us

$$\sum_{j=1}^{k} |e_{i_j}| \le \sum_{i=n-k}^{n-1} |e_i| \le \alpha_d \sum_{i=2}^{k+1} i^{-1/d} \le \tilde{\alpha}_d k^{(d-1)/d} \tag{2.3}$$

for a new constant $\tilde{\alpha}_d$. If we apply this last inequality to the set of $k = \nu_d(x)$ edges with length at least $x$, we have

$$x\nu_d(x) \le \tilde{\alpha}_d \big(\nu_d(x)\big)^{(d-1)/d}. \tag{2.4}$$

Upon division by $x\nu_d^{(d-1)/d}(x)$, inequality (2.4) establishes Lemma 2.1 with $\beta_d = (\tilde{\alpha}_d)^d$. □

Lemma 2.2 is a generic application of inequality (2.2). It suggests the correct order of magnitude for sums of powers of edge lengths.

LEMMA 2.2. *For any $\psi(x) = |x|^\gamma$, any minimal spanning tree $T$ of $\{x_1, x_2, \dots, x_n\} \subset [0,1]^d$ satisfies the inequalities*:

$$\sum_{e \in T} |e|^\gamma \le \beta'(\gamma, d) n^{(d-\gamma)/d} \quad \text{for } 0 < \gamma < d, \tag{2.5}$$

$$\sum_{e \in T} |e|^d \le \beta'(d) \log n \tag{2.6}$$

*and*

$$\sum_{|e| \ge y} |e|^\gamma \le \beta'(\gamma, d) y^{\gamma-d} \quad \text{for } 0 \le y < \infty \text{ and } 0 < \gamma < d. \tag{2.7}$$

*Here $\beta'(\gamma, d)$ is a constant depending only on the parameter $\gamma$ and the dimension $d$ and $\beta'(d)$ is a constant depending only on $d$.*

PROOF. For any $\lambda > 0$, we have for $0 < \gamma < d$,

$$\sum_{e \in T} |e|^\gamma = \sum_{|e| \le n^{-\lambda}} |e|^\gamma + \sum_{|e| > n^{-\lambda}} |e|^\gamma,$$

so, noting that $n - 1 - \nu_d(x)$ is the number of edges of the MST which are of length less than $x$, we can majorize the first sum by $n^{1-\lambda\gamma}$ and write the second sum as a Stieltjes integral to get

$$\sum_{e \in T} |e|^\gamma \le n^{1-\gamma\lambda} + \int_{n^{-\lambda}}^{\sqrt{d}} x^\gamma d\big(n - 1 - \nu_d(x)\big).$$

Integrating by parts, we find

$$\sum_{e \in T} |e|^\gamma \le n^{1-\lambda\gamma} + n^{-\gamma\lambda}\nu_d(n^{-\lambda}) + \gamma \int_{n^{-\lambda}}^{\sqrt{d}} x^{\gamma-1} \nu_d(x)\, dx.$$

Majorizing $\nu_d(x)$ by $\beta_d x^{-d}$ and integrating, we get

$$\sum_{e\in T} |e|^{\gamma} \le n^{1-\lambda\gamma} + \beta_d n^{\lambda(d-\gamma)} + \beta_d \gamma (d-\gamma)^{-1} n^{\lambda(d-\gamma)}.$$

Choosing $\lambda = 1/d$, we obtain (2.5). By analogous arguments, one can verify inequalities (2.6) and (2.7). □

REMARK. The preceding proof provides a bound on $\beta'(\gamma, d)$ which diverges to infinity as $\gamma \to d$. When $d = 2$, an explicit bound

$$\sum_{e\in T} |e|^2 \le 2\sqrt{3}$$

was proved by Gilbert and Pollak [(1968), page 16]. The method of Gilbert and Pollak uses the geometry of the plane extensively and their method appears to be difficult to generalize beyond $d = 2$.

In the jackknife estimates developed in Section 4, we need bounds on the change which takes place in the value of $M(x_1, x_2, \dots, x_n)$ as a point is deleted from $\{x_1, x_2, \dots, x_n\}$. Following the customary jackknife notation, we set $M(x_1, x_2, \dots, \hat{x}_i, \dots, x_n) = M(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$, i.e., a *hat* is used to signal a missing variable. Further, if $T$ is any graph, we let $N(i) = \{j: (x_i, x_j) \in T\}$, so $N(i)$ is the set of neighbors of $x_i$ in the graph determined by $T$. We can now bound the changes in the values of $M$ as the sample changes by adding or dropping points.

LEMMA 2.3. *For any edge weight function* $\psi$, *we have*

$$M(x_1, x_2, \dots, x_n) \le M(x_1, x_2, \dots, \hat{x}_i, \dots, x_n) + \min_{j:\, j\ne i} \psi(|x_i - x_j|) \tag{2.8}$$

*and, for any nondecreasing* $\psi$, *we also have*

$$M(x_1, x_2, \dots, \hat{x}_i, \dots, x_n) \le M(x_1, x_2, \dots, x_n) + \sum_{j\in N(i)} \psi(2|x_i - x_j|). \tag{2.9}$$

PROOF. To prove inequality (2.8), note that any tree which spans $\{x_1, x_2, \dots, \hat{x}_i, \dots, x_n\}$ can be completed to a connected graph spanning $\{x_1, x_2, \dots, x_n\}$ by connecting the point $x_i$ to the point $x_j$, $j \ne i$, for which $\psi(|x_i - x_j|)$ is minimal.

To prove inequality (2.9), let $T$ be a minimal spanning tree of $\{x_1, x_2, \dots, x_n\}$ and let $x'$ be an element $x_j$ of $N(i)$ such that $|x_j - x_i|$ is minimal. We get a connected graph spanning $\{x_1, x_2, \dots, \hat{x}_i, \dots, x_n\}$ by (1) taking the edges of $T$, (2) deleting all the edges incident to $x_i$, i.e., deleting the set $\{(x_i, x_j): (x_i, x_j) \in T\}$ and (3) adding the set of edges which join $x'$ to the other neighbors of $x_i$, i.e., adding in the set $A = \{(x', x_j): j \in N(i), x_j \ne x'\}$. This procedure implies the bound

$$M(x_1, x_2, \dots, \hat{x}_i, \dots, x_n) \le M(x_1, x_2, \dots, x_n) - \sum_{j \in N(i)} \psi(|x_i - x_j|) + \sum_{j \in N(i)} \psi(|x' - x_j|).$$

By the triangle inequality and the definition of $x'$, we have $|x' - x_j| \le |x' - x_i| + |x_i - x_j| \le 2|x_i - x_j|$, so inequality (2.9) follows by the monotonicity of $\psi$. □

To bound sums such as that in inequality (2.9), one needs a bound on $|N(i)|$, the cardinality of the set $N(i)$ of neighbors of $x_i$.

LEMMA 2.4. *If $\psi$ is strictly increasing, then there is a constant $D_d$ depending only on d such that any MST in $R^d$ has maximum vertex degree bounded by $D_d$.*

In the case of $\psi_0(x) = x$, this lemma is well known; in fact, $D_d$ is bounded by $N_d$, the number of spherical caps with angle 60° which are needed to cover the unit sphere in $R^d$. To check this bound, one can note that if a vertex had degree greater than $N_d$, then two edges of the MST would have an angle less than 60°. Simple geometry would then contradict minimality.

To prove Lemma 2.4 for general monotone $\psi$, we first note by the monotonicity that we can choose edges for a minimal spanning tree with edge weighting function $\psi$ which are also the edges of a minimal spanning tree with edge weighting function $\psi_0(x) = |x|$. To see this, just consider building the two minimal spanning trees by the algorithm of Kruskal (1956) which (1) orders the edges of the complete graph in increasing order and (2) examines each edge in order and accepts an edge into the minimal tree provided it does not create a cycle with the previously chosen edges. By the monotonicity of $\psi$, the edge ordering given by $\psi$ and $\psi_0$ coincide, so if one constructs a minimal spanning tree using Kruskal's algorithm for $\psi$ and $\psi_0$, the sets of edges can be chosen so they will also coincide. [For more detail on Kruskal's algorithm, one can consult Papadimitriou and Steiglitz (1982), page 289.]

The main result of this section is Lemma 2.5 which expresses a basic continuity of the minimal spanning tree functional.

LEMMA 2.5. *There is a constant $\beta''(\alpha, d)$ such that for any two finite subsets $\chi$ and $\chi'$ of $[0,1]^d$, we have for the weighting function $\psi(x) = x^\alpha$, $0 < \alpha < d$, that*

$$|M(\chi) - M(\chi')| \le \beta''(\alpha, d)|\chi \triangle \chi'|^{(d-\alpha)/d}. \tag{2.10}$$

*Here $\chi \triangle \chi'$ denotes the symmetric difference of the sets $\chi$ and $\chi'$, i.e., $(\chi \cup \chi') - (\chi \cap \chi')$.*

PROOF. For any $x'$ in the set $\chi' - \chi$, we let $N(x')$ denote the set of neighbors of $x'$ in the MST of $\chi \cup \chi'$. By the same considerations used in the proof of (2.9), we obtain

$$M(\chi') \le M(\chi \cup \chi') + \sum_{x' \in \chi - \chi'} \sum_{x \in N(x')} \psi(2|x' - x|).$$

We will use Lemma 2.2 to bound this double sum, but some maneuvering is required. First consider the set of edges $E' = \{(x', x): x \in N(x') \text{ and } x' \in \chi - \chi'\}$ and let $V'$ denote the set of all vertices which belong to an edge of $E'$. We know the cardinality $|V'|$ is bounded by $(D_d + 1)|\chi - \chi'|$, since no vertex $x'$ can have degree bigger than $D_d$.

A necessary and sufficient condition due to Prim (1957) for an edge $e$ to belong to some minimal spanning tree of a graph is that the vertex set of the graph can be divided into two nontrivial complementary components $A$ and $A^c$ such that $e$ is the least weight edge which meets both $A$ and $A^c$. For $e$ contained in $\mathrm{MST}\{x_1, x_2, \dots, x_n\} = E$, we can thus find $A \subset \{x_1, x_2, \dots, x_n\}$ such that $e$ meets $A$ and $A^c$ and

$$|e| = \min\{\tilde{e} \in E: \tilde{e} \cap A \ne \phi, \tilde{e} \cap A^c \ne \phi\}.$$

Now if $e \in E'$, we also have $e \in E$, and hence we have the equality

$$|e| = \min\{e \in E': e \cap A \cap V' \ne \phi, e \cap A^c \cap V' \ne \phi\}$$

and we see that $e$ must also be in $\mathrm{MST}(V')$ by Prim's sufficient condition. We therefore have the bound

$$\sum_{x' \in \chi - \chi'} \sum_{x \in N(x')} \psi(2|x - x'|) \le \sum_{e \in \mathrm{MST}(V')} \psi(2|e|),$$

so using Lemma 2.2 and the bound on the cardinality of $V'$, we have

$$M(\chi') \le M(\chi \cup \chi') + 2^\alpha \beta'(\alpha, d)(D_d + 1)^{(d-\alpha)/d}|\chi - \chi'|^{(d-\alpha)/d}. \tag{2.11}$$

To obtain an inequality in the other direction, we build a spanning tree for $\chi \cup \chi'$ by taking minimal trees for $\chi$ and $\chi' - \chi$ and joining them with a single edge. This construction gives

$$M(\chi \cup \chi') \le M(\chi) + M(\chi' - \chi) + \min_{x \in \chi', x' \in \chi' - \chi} \psi(|x - x'|). \tag{2.12}$$

Since by inequality (2.5), we have the bound $\beta'(\alpha, d)|\chi' - \chi|^{(d-\alpha)/d}$ on $M(\chi' - \chi)$ and since $\psi(|x - x'|)$ is majorized by $d^{\alpha/d}$, inequality (2.12) yields

$$M(\chi \cup \chi') \le M(\chi) + \beta'(\alpha, d)|\chi' - \chi|^{(d-\alpha)/d} + d^{\alpha/2}. \tag{2.13}$$

Now majorizing $M(\chi \cup \chi')$ in (2.11) by the bound given by (2.13) yields

$$\begin{aligned} &M(\chi') - M(\chi) \\ &\quad \le \beta'(\gamma, d)\left(1 + 2^\alpha (D_d + 1)^{(d-\alpha)/d}\right)|\chi - \chi'|^{(d-\alpha)/d} + d^{\alpha/2}. \end{aligned} \tag{2.14}$$

By the symmetry of $\chi$ and $\chi'$, we see that $|M(\chi') - M(\chi)|$ is also bounded by the right side of (2.14). Since we can either assume $|\chi - \chi'| \ge 1$ or else inequality (2.10) is trivial, we see that inequality (2.14) establishes the claimed bound (2.10) with a value of $\beta_d''$ which can be given by

$$\beta''(\alpha, d) = \beta'(\gamma, d)\left(1 + 2^\alpha (D_d + 1)^{(d-\alpha)/d}\right) + d^{\alpha/2}. \qquad \square$$

**3. Growth in mean.** The asymptotics of $m_n = EM(X_1, X_2, \dots, X_n)$, where $X_i$ are independent and uniform on $[0,1]^d$, can be established by a three step technique of Poisson smoothing, analysis of an approximate recursion and a Tauberian argument for extracting information about $m_n$ from information about averages.

By $\pi$ we denote a Poisson point process on $\mathbb{R}^d$ with constant intensity equal to 1. For any bounded Borel set $A$, the set $\pi(A)$ is almost surely a finite set of points and $M(\pi(A))$ will denote total weight of the minimal spanning tree of the finite set of points $\pi(A)$. Here, of course, the total weight of the minimal spanning tree is defined in terms of $\psi$ as expressed in (1.1).

We now let $\phi(t) = EM(\pi[0, t]^d)$; that is, $\phi(t)$ is the expected cost of the minimal spanning tree of the set of points $\pi[0, t]^d$. To make this explicit, we write $\pi[0, t]^d = \{X_1, X_2, \dots, X_N\}$, where $N$ is a Poisson random variable with mean $t^d$ and we apply (1.1) to get the representation

$$\phi(t) = E \min_T \sum_{(i,j) \in T} \psi(|X_i - X_j|),$$

where the minimum is over all spanning trees of $\{X_1, X_2, \dots, X_N\}$. The asymptotic behavior of $\phi$ will be obtained from Lemma 3.1.

LEMMA 3.1. *If $\psi(x) = x^\alpha$ with $0 < \alpha < d$, then there is a constant $C = C(\alpha, d)$ such that*

$$\phi(t) \le m^d \phi(t/m) + C t^\alpha m^{d-\alpha} \tag{3.1}$$

*for all reals $t$, $0 < t < \infty$, and integers $m \ge 1$.*

PROOF. We first partition $[0, t]^d$ into $m^d$ subcubes $Q_i$ of edge length $t/m$. Next, for each $i$ for which $\pi(Q_i) \ne 0$, we choose one representative $Y_i$ from $\pi(Q_i)$. By inequality (2.5) scaled up to $[0, t]^d$, we can obtain a minimal spanning tree $T$ of the set $\{Y_i\colon 1 \le i \le m^d,\ \pi(Q_i) \ne \phi\}$ such that $\sum_{e \in T} |e|^\alpha \le t^\alpha \beta'(\alpha, d)(m^d)^{(d-\alpha)/d} = t^\alpha \beta'(\alpha, d) m^{d-\alpha}$. But now the tree $T$, together with the minimal spanning trees of all the $\pi(Q_i)$, must form a connected graph. Such a connected graph has length no greater than that of the minimal spanning tree of $\pi([0, t]^d)$, so

$$\begin{aligned} M(\pi[0, t]^d) &\le \sum_{i=1}^{m^d} M(\pi(Q_i)) + \sum_{e \in T} |e|^\alpha \\ &\le \sum_{i=1}^{m^d} M(\pi(Q_i)) + t^\alpha \beta'(\alpha, d) m^{d-\alpha}. \end{aligned} \tag{3.2}$$

Taking expectations and using the fact that $EM(\pi(Q_i)) = \phi(t/m)$ completes the proof of the lemma with $C = \beta'(\alpha, d)$. □

As a consequence of (3.1) and the continuity of $\phi$, we can show there is a constant $c(\alpha, d) \geq 0$ depending only on $d$ and $\alpha$ such that

$$\phi(t) \sim c(\alpha, d)t^d \quad \text{as } t \to \infty. \tag{3.3}$$

To prove this, we set $\tilde{\phi}(t) = \phi(t) + 2Ct^\alpha$ and note that if $m_0$ is chosen large enough to insure $2m_0^\alpha < m_0^d$, then inequality (3.1) implies that for $m \geq m_0$, we have

$$\tilde{\phi}(tm) \leq m^d \tilde{\phi}(t).$$

For any $\varepsilon > 0$, we can let $\beta = \liminf_{t\to\infty} \tilde{\phi}(t)/t^d$ and use the continuity of $\tilde{\phi}$ to find an interval $[t_0, t_1]$ such that $\tilde{\phi}(t)/t^d \leq \beta + \varepsilon$ for all $t \in [t_0, t_1]$. Setting $A = \{s\colon \tilde{\phi}(s)/s^d \leq \beta + \varepsilon\}$, we note that the recursion for $\tilde{\phi}$ implies that $\bigcup_{m=m_0}^{\infty}[mt_0, mt_1] \subset A$. But since $mt_1 \geq (m+1)t_0$ for all $m \geq t_0(t_1 - t_0)^{-1} = t^*$, we see that for $m$ greater than $\max(t^*, m_0)$ the intervals $[mt_0, mt_1)$ are overlapping. From this we see that $A$ must contain an infinite interval $[t^{**}, \infty)$. This implies $\limsup \tilde{\phi}(t)/t^d \leq \beta + \varepsilon$, and by the arbitrariness of $\varepsilon$ we obtain $\tilde{\phi}(t) \sim \beta t^d$. This result and the fact that $\alpha < d$ implies the asymptotic relation (3.3).

By the definition of $\phi(t)$ and the scaling property $t^\alpha M(x_1, x_2, \dots, x_n) = M(tx_1, tx_2, \dots, tx_n)$ of $M$, one can compute by conditional expectations that

$$\phi(t) = t^\alpha \sum_{n=0}^{\infty} m_n e^{-t^d} t^{dn}/n!. \tag{3.4}$$

Here $m_n = EM(X_1, X_2, \dots, X_n)$, where the $X_i$ are independent and uniformly distributed on $[0,1]^d$.

To extract the asymptotics of $m_n$ from (3.3) and (3.4), we will first prove that the sequence $m_n$ has a useful monotonicity property, specifically,

$$nm_{n-1} \leq (n + 2^{1+\alpha})m_n. \tag{3.5}$$

To verify inequality (3.5), we sum inequality (2.9) over $1 \leq i \leq n$ to get

$$\begin{aligned} \sum_{i=1}^{n} M\big(X_1, X_2, \dots, \hat{X}_i, \dots, X_n\big) &\leq nM(X_1, X_2, \dots, X_n) \\ &\quad + \sum_{i=1}^{n} \sum_{j \in N(i)} 2^\alpha |X_i - X_j|^\alpha. \end{aligned} \tag{3.6}$$

Next we note that the double sum exactly equals $2^{1+\alpha}M(X_1, X_2, \dots, X_n)$, so taking expectations in inequality (3.6) completes the proof of inequality (3.5).

The way we will use inequality (3.5) is through the fact that it implies

$$n^k m_n \geq (n-1)^k m_{n-1} \quad \text{for all } n \geq 1, \tag{3.7}$$

provided $k \geq 2^{1+\alpha}$.

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

LEMMA 3.2. *If $f(x) \sim Ax^{\gamma}$ as $x \to 0$ and $f''(x) = O(x^{\gamma-2})$ as $x \to 0$, then $f'(x) \sim \gamma A x^{\gamma-1}$ as $x \to 0$.*

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

LEMMA 3.3. *If $X_i$, $1 \le i < \infty$, are independent and uniformly distributed on $[0,1]^d$ and $m_n = EM(X_1, X_2, \dots, X_n)$, then there is a constant $c(\alpha, d)$ such that*

$$m_n \sim c(\alpha, d) n^{(d-\alpha)/d} \quad \text{as } n \to \infty. \tag{3.8}$$

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

$$\sum_{n=0}^{\infty} m_n e^{-u} u^n / n! \sim c(\alpha, d) u^{(d-\alpha)/d} \quad \text{as } u \to \infty. \tag{3.9}$$

Taking the Laplace transform of the left side of (3.9) and applying the Abelian theorem for Laplace transforms [Widder (1946), page 181], we obtain for $\lambda \to 0$ that

$$\sum_{n=0}^{\infty} m_n (1+\lambda)^{-n-1} \sim c(\alpha, d)\Gamma(2 - \alpha/d)\lambda^{-2+\alpha/d}$$

or, equivalently,

$$\sum_{n=0}^{\infty} m_n e^{-nx} \sim c(\alpha, d)\Gamma(2 - \alpha/d) x^{-2+\alpha/d} \quad \text{as } x \to 0. \tag{3.10}$$

From Lemma 2.2 we know $m_n = O(n^{(d-\alpha)/d})$ and, by the Euler–Maclaurin formula, we have $\sum_{n=0}^{\infty} n^{\gamma} e^{-nx} \sim x^{-\gamma-1}\Gamma(\gamma+1)$ as $x \to 0$; so, for all $k$,

$$\sum_{n=0}^{\infty} n^k m_n e^{-nx} = O(x^{-k-2+\alpha/d}) \quad \text{as } x \to 0. \tag{3.11}$$

Finally, by the Hardy–Littlewood differentiation lemma applied successively beginning with (3.10) and using (3.11), we find for each integer $k$ as $x \to 0$ that

$$\begin{aligned}\sum_{n=0}^{\infty} n^k m_n e^{-nx} \sim{}& c(\alpha, d)\Gamma(2 - \alpha/d)(2 - \alpha/d)(3 - \alpha/d) \\ & \cdots (k + 1 - \alpha/d) x^{-k-2+\alpha/d}.\end{aligned} \tag{3.12}$$

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

$$\begin{aligned}\sum_{n=0}^{\infty} n^k m_n y^n \sim{}& c(\alpha, d)\Gamma(2 - \alpha/d)(2 - \alpha/d)(3 - \alpha/d) \\ & \cdots (k + 1 - \alpha/d)(1 - y)^{-k-2+\alpha/d}.\end{aligned} \tag{3.13}$$

The Karamata Tauberian theorem [see, e.g., Feller (1971), page 447] applied to

(3.13) now gives us information about the partial sums of the $n^k m_n$:

$$\sum_{n=1}^{N} n^k m_n \sim c(\alpha, d)\Gamma(2-\alpha/d) \times \prod_{j=1}^{k}(1+j-\alpha/d)N^{k+2-\alpha/d}/\Gamma(k+3-\alpha/d). \tag{3.14}$$

The point of this maneuvering is that the series $\{n^k m_n\}$ is increasing by (3.7) and, by a well-known lemma [cf. Apostol (1976), page 280], the monotonicity of the terms of a partial sum justifies carrying over the asymptotics to the individual terms, i.e.,

$$n^k m_n \sim c(\alpha, d) n^{k+1-\alpha/d}. \tag{3.15}$$

On dividing by $n^k$, the proof of Lemma 3.3 is complete. □

The repeated differentiation technique applied previously can be used in many problems where one needs to exploit an approximate monotonicity such as expressed by inequality (3.5). The passage from the asymptotic relation (3.3) to (3.15) can be made a bit more quickly by using a more sophisticated Tauberian theorem, but building the derivation on Karamata's theorem is simple enough. One benefit of the path chosen is that the structure of $m_n$ is drawn out a bit more fully through the development of inequality (3.7). A third method for proving (3.8) can be based on (3.7), the fact that $m_n = O(n^{(d-\alpha)/d})$ by Lemma 2.2 and direct estimation of the Borel sum in (3.9).

**4. Variance bounds.** Efron and Stein (1981) established the 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 effective bounds on $\operatorname{Var} M_n$.

Let $S(x_1, x_2, \dots, x_{n-1})$ denote any symmetric function of $n-1$ vectors $x_i \in \mathbb{R}^d$. If $X_i$, $1 \le i \le n$, are independent identically distributed random vectors in $\mathbb{R}^d$, we define $n+1$ new random variables by

$$S_i = S(X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_n) \quad \text{and} \quad \bar{S} = n^{-1}\sum_{i=1}^{n} S_i.$$

The Efron–Stein inequality says

$$\operatorname{Var} S(X_1, X_2, \dots, X_{n-1}) \le E\sum_{i=1}^{n}(S_i - \bar{S})^2. \tag{4.1}$$

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

$$\operatorname{Var} M_{n-1} \le E\sum_{i=1}^{n}(S_i - M(X_1, X_2, \dots, X_n))^2.$$

For $X_i$ that 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 use the Vinogradov symbol $a_n \ll b_n$ to denote that $a_n \le Cb_n$ for some $C$ not depending on $n$. By Lemma 2.3 and the bound $(\max(a, b))^2 \le a^2 + b^2$, we have

$$\operatorname{Var} M_{n-1} \le E\sum_{i=1}^{n} \min_{j:\, j\ne i} |X_i - X_j|^{2\alpha} + E\sum_{i=1}^{n}\Big(\sum_{j\in N(i)} 2^{\alpha}|X_i - X_j|^{\alpha}\Big)^2.$$

By elementary calculus, we can show for any $0 < \beta < \infty$ that $E\min_{j:\, j\ne i}|X_i - X_j|^{\beta} \ll n^{-\beta/d}$ so the first sum is majorized by $n^{1-2\alpha/d}$. Applying Schwarz's inequality to the second sum, yields a total bound

$$\operatorname{Var} M_{n-1} \ll n^{1-2\alpha/d} + E\sum_{i=1}^{n}\sum_{j\in N(i)} |X_j - X_i|^{2\alpha}.$$

Since the last sum equals $2E\sum_{e\in T}|e|^{2\alpha}$, where $T$ is a minimal spanning tree of $\{X_1, X_2, \dots, X_n\}$, Lemma 2.2 shows that we have completed the proof of Lemma 4.1.

LEMMA 4.1. *If $X_i$, $1 \le i < \infty$, are independent and uniform on $[0,1]^d$, then*

$$\operatorname{Var} M_n \ll n^{1-2\alpha/d} \quad \text{for } 0 < 2\alpha < d, \tag{4.2}$$

$$\operatorname{Var} M_n \ll \log n \quad \text{for } 2\alpha = d \tag{4.3}$$

*and for each $\alpha$, $d/2 < \alpha < \infty$,*

$$\operatorname{Var} M_n \text{ is uniformly bounded for all } n \ge 1. \tag{4.4}$$

To move now toward a proof that $n^{-(d-\alpha)/d}(M_n - m_n)$ converges almost surely to 0, we set up a subsequence argument. For a real number $\lambda > 1$ which will be determined later, we define a subsequence of integers $n_k$ by letting $n_k = [k^{\lambda}]$ and we note by Lemma 4.1 that we have the set of three inequalities:

$$\operatorname{Var}\big(M_n/n^{(d-\alpha)/d}\big) \ll n^{-1} \quad \text{for } 0 < \alpha < d/2, \tag{4.5a}$$

$$\operatorname{Var}\big(M_n/n^{(d-\alpha)/d}\big) \ll \log n/n \quad \text{for } \alpha = d/2, \tag{4.5b}$$

$$\operatorname{Var}\big(M_n/n^{(d-\alpha)/d}\big) \ll n^{-2+2\alpha/d} \quad \text{for } d/2 < \alpha < d. \tag{4.5c}$$

Under each of the conditions of (4.5a)–(4.5c), we see for $0 < \alpha < d$ that

$$\sum_{k=1}^{\infty} \operatorname{Var}\big\{M_{n_k}/n_k^{(d-\alpha)/d}\big\} < \infty,$$

provided also that $\lambda(-2 + 2\alpha/d) < -1$, i.e., provided $\lambda > \frac{1}{2}(1 - \alpha/d)^{-1}$. For such a value of $\lambda$, we see by the usual Borel–Cantelli argument that $M_{n_k} n_k^{-(d-\alpha)/d}$ is asymptotic to $m_{n_k}$ with probability 1 as $k \to \infty$.

To push this relation toward one valid for the full sequence of integers, we will bound the variability of $M_n$ as $n$ varies through the intervals $[n_k, n_{k+1})$, i.e., we bound $V_k = \max_{n_k \le n < n_{k+1}} |M_n - M_{n_k}|$. By Lemma 2.5, we see that for $n_k \le n \le n_{k+1}$ we have with probability 1 that

$$|M_n - M_{n_{k+1}}| \le \beta''(\alpha, d) |\{X_{n+1}, X_{n+2}, \dots, X_{n_{k+1}}\}|^{(d-\alpha)/d} \ll k^{(\lambda-1)(d-\alpha)/d}.$$

Hence we have with probability 1 that

$$V_k / n_k^{(d-\alpha)/d} \ll k^{-(d-\alpha)/d}. \tag{4.6}$$

From the identity

$$\begin{aligned} &M_n n^{-(d-\alpha)/d} - M_{n_k} n_k^{-(d-\alpha)/d} \\ &\quad = (M_n - M_{n_k}) n^{-(d-\alpha)/d} + M_{n_k}\left(n^{-(d-\alpha)/d} - n_k^{-(d-\alpha)/d}\right), \end{aligned}$$

the bound (4.6), the fact that $M_{n_k} \ll n_k^{(d-\alpha)/d}$ and $n^{-(d-\alpha)/d} - n_k^{-(d-\alpha)/d} \ll k^{-1} n_k^{-(d-\alpha)/d}$, we have the relation

$$|M_n n^{-(d-\alpha)/d} - M_{n_k} n_k^{-(d-\alpha)/d}| \ll k^{-(d-\alpha)/d} + k^{-1}. \tag{4.7}$$

Finally, inequality (4.7) and the fact that $M_{n_k} \sim c(\alpha, d) n_k^{(d-\alpha)/d}$ with probability 1, complete the proof. Our result is summarized in Theorem 2.

**Theorem 2.** *If $X_i$, $1 \le i < \infty$, are uniformly distributed on $[0,1]^d$ and $M_n$ is the length of the MST of $\{X_1, X_2, \dots, X_n\}$ using the edge weight function $\psi(x) = x^\alpha$ with $0 < \alpha < d$, then there is a constant $c(\alpha, d) > 0$ such that with probability* 1,

$$M_n \sim c(\alpha, d) n^{(d-\alpha)/d} \quad \text{as } n \to \infty. \tag{4.8}$$

This theorem has been proved except for showing $c(\alpha, d)$ is strictly positive. To see this, note that each $X_i$ is connected to some $X_j$ and hence

$$M_n \ge \frac{1}{2} \sum_{i=1}^{n} \min\{|X_j - X_i|^\alpha : 1 \le j \le n,\ j \ne i\}.$$

Since it is easy calculus to show there is a $c > 0$ such that

$$E \min\{|X_j - X_i|^\alpha : 1 \le j \le n,\ j \ne i\} \ge c n^{-\alpha/d},$$

the proof is complete.

**5. General extension principle.** We now show how the result just obtained for uniform distributions can be extended to any bounded distribution. The main observation is that the minimal spanning tree functional has enough continuity to permit approximation of samples from a general distribution to be replaced by samples from a simpler class of distributions whose asymptotic analysis follows easily from the results for uniformly distributed variables. This approximation is made easier by using an almost sure embedding technique based on the following property of marginal and joint distributions.

LEMMA 5.1 [Strassen (1965)]. *Suppose $P$ and $Q$ are probability measures on a bounded subset of $\mathbb{R}^d$ and suppose also that there is an $\varepsilon > 0$ such that $P$ and $Q$ satisfy $P(F) \le Q(F) + \varepsilon$ for all closed $F$. There is then a probability measure $\hat{\mu}$ on the product space $\mathbb{R}^d \times \mathbb{R}^d$ such that*

$$\hat{\mu}(\cdot, \mathbb{R}^d) = P(\cdot), \qquad \hat{\mu}(\mathbb{R}^d, \cdot) = Q(\cdot) \quad \textit{and} \quad \hat{\mu}\{(x, y): x \neq y\} \le \varepsilon.$$

For an elegant combinatorial proof of this lemma based on Hall's matching theorem, one can consult Dudley (1968 or 1976).

The proof of Theorem 1 begins by establishing the limit result for a special class of distributions which we call *blocked distributions*. These are probability measures on $[0,1]^d$ with the form $g(x)\,dx + d\mu_s$ where $g(x) = \sum_{i=1}^m \alpha_i 1_{Q_i}$, the measure $\mu_s$ is purely singular and the $Q_i$, $1 \le i \le m < \infty$, are disjoint cubes with edges parallel to the axes. Here we recall that a measure $\mu_s$ on $[0,1]^d$ is called a purely singular measure if $\mu_s([0,1]^d) = \mu_s(A)$ for some measurable $A$ of Lebesgue measure 0. The next result points out a continuity condition which suffices to carry asymptotic results for the class of blocked distributions to the class of bounded distributions.

THEOREM 3. *Suppose that there is a constant $B$ not depending on $n$ such that $S(x_1, x_2, \ldots, x_n)$ satisfies the continuity condition*

$$|S(x_1, x_2, \ldots, x_n) - S(x_1', x_2', \ldots, x_n')| \le B|\{i: x_i \neq x_i'\}|^{(d-\alpha)/d}. \tag{5.1}$$

*Suppose also that for every sequence of i.i.d. random variables $\{X_i\}_{1 \le i < \infty}$ distributed with a blocked distribution $\mu = \mu_s + g(x)\,dx$, we have with probability 1 that*

$$S(X_1, X_2, \ldots, X_n) n^{-(d-\alpha)/d} \sim c(\alpha, d) \int g(x)^{(d-\alpha)/d}\, dx. \tag{5.2}$$

*One then has that with probability 1,*

$$S(X_1', X_2', \ldots, X_n') \sim n^{(d-\alpha)/d} c(\alpha, d) \int f(x)^{(d-\alpha)/d}\, dx,$$

*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 an approximation $g_m(x)\,dx + \mu_s$ where $g_m(x) = \sum_{i=1}^{m^d} \alpha_i 1_{Q_i}$. Here, each $Q_i$ is one of the subcubes obtained by partitioning $[0,1]^d$ into $m^d$ parts and the $\alpha_i$ are defined by $\alpha_i = \int_{Q_i} f(x)\,dx$. It is a traditional exercise to show that for such $\alpha_i$ we have $\int |g_m(x) - f(x)|\,dx \to 0$ as $m \to \infty$. Finally, defining measures $P$ and $Q$ by

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

we have $|P(A) - Q(A)| \le \int_A |f(x) - g_m(x)|\,dx \le \varepsilon$ for all $m \ge m_0(\varepsilon)$ and $\varepsilon > 0$.

By Lemma 5.1, one can therefore define a probability measure $\hat{\mu}$ on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $P$ and $Q$ such that $\hat{\mu}\{(x, y): x \neq y\} \leq \varepsilon$.

We now define a sequence of random vectors in $\mathbb{R}^d \times \mathbb{R}^d$ by $(X_i, X_i')$, $1 \leq i < \infty$, where $(X_i, X_i')$ is the $i$th vector of an independent sequence of random vectors with distribution given by the measure $\hat{\mu}$. By the law of large numbers, we have with probability 1 that

$$|\{i: X_i \neq X_i'\}| \sim n\hat{\mu}(\{(x, y): x \neq y\}) \tag{5.3}$$

and by the bound on the off-diagonal probability mass we have $n\hat{\mu}(\{(x, y): x \neq y\}) \leq \varepsilon n$. By conditions (5.1) and (5.3), we see

$$\limsup_{n\to\infty} n^{-(d-\alpha)/d} |S(X_1, X_2, \dots, X_n) - S(X_1', X_2', \dots, X_n')| \leq B\varepsilon^{(d-\alpha)/d},$$

whence by (5.2) we have almost surely

$$\begin{aligned}
&\limsup_{n\to\infty} S(X_1, X_2, \dots, X_n) n^{-(d-\alpha)/d} \\
&\quad \leq B\varepsilon^{(d-\alpha)/d} + \limsup_{n\to\infty} S(X_1', X_2', \dots, X_n') n^{-(d-\alpha)/d} \\
&\quad \leq B\varepsilon^{(d-\alpha)/d} + c(\alpha, d) \int g_m(x)^{(d-\alpha)/d}\, dx.
\end{aligned}$$

Since $m$ and $\varepsilon > 0$ are arbitrary, we consequently have

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

Finally, since the limit infimum can be dealt with in exactly the same way as the preceding limit superior, the proof of this theorem is complete. □

We now apply Theorem 3 to minimal spanning trees. By Lemma 2.5, inequality (5.1) is satisfied by the minimal spanning tree functional, so it only remains to verify the relation (5.2).

We consider a blocked distribution $g(x)\,dx + \mu_s$ and let $E$ denote the support of $\mu_s$. We will also denote Lebesgue measure by $\lambda(\cdot)$ or by $dx$, according to convenience. Since the Lebesgue measure of $E$ equals 0 and since $g$ is constant on a set of subcubes, we can find for any $\varepsilon > 0$ a partition $\{Q_i\}_{i\in I}$ of $[0,1]^d$ into subcubes such that the following properties hold:

$$\text{on each } Q_i,\ i \in I,\ g(\cdot) \text{ is constant,} \tag{5.4}$$

$$\text{for all } i \in I,\ l = \lambda(Q_i)^{1/d} < \varepsilon, \tag{5.5}$$

$$\begin{gathered}
E \subset A \cup B, \quad \text{where } A \text{ and } B \text{ are disjoint, } \lambda(A) = 0, \\
P(X_1 \in A) = \mu_s(A) \leq \varepsilon, \\
B = \bigcup_{j\in J} Q_j, \quad \text{where } \lambda(B) = \sum_{j\in J} \lambda(Q_j) \leq \varepsilon.
\end{gathered} \tag{5.6}$$

We now set $C = [0,1]^d - (A \cup B)$ and note that by tying together the minimal spanning trees of these sample points which lie, respectively, in the sets in $A$, $B$

and $C$, we have

$$M(X_1, X_2, \ldots, X_n) \le 2d^{\alpha/2} + M(\{X_i: X_i \in A\}) + M\{X_i: X_i \in B\} + M(\{X_i: X_i \in C\}),$$

where the first term $2d^{\alpha/2}$ denotes the maximal cost needed to unite the three trees. Since $P(X_i \in A) \le \varepsilon$, the strong law of large numbers and Lemma 2.2 give the bound

$$\limsup_{n\to\infty} M(\{X_i: X_i \in A\})n^{-(d-\alpha)/d} \le \beta'(\alpha, d)\varepsilon^{(d-\alpha)/d}. \tag{5.7}$$

For the sample in the set $B$, we use a more geometric bound. Specifically, we apply Lemma 2.2, rescaling and Hölder's inequality to obtain

$$\begin{aligned} M(\{X_i: X_i \in B\}) &\le |J|d^{\alpha/2} + \sum_{j\in J} M\big(\{X_i: X_i \in Q_j\}\big) \\ &\le |J|d^{\alpha/2} + \beta'(\alpha, d)l^{\alpha} \sum_{j\in J} |\{X_i \in Q_j\}|^{(d-\alpha)/d} \\ &\le |J|d^{\alpha/2} + \beta'(\alpha, d)l^{\alpha}|J|^{\alpha/d}n^{(d-\alpha)/d}. \end{aligned} \tag{5.8}$$

By the last bound given in (5.6), we have $\sum_{j\in J}\lambda(Q_j) \le \varepsilon$; so by the definition of $l$, we have $|J|l^d \le \varepsilon$ and inequality (5.8) finally yields

$$\limsup_{n\to\infty} M(\{X_i: X_i \in B\})n^{-(d-\alpha)/d} \le \beta'(\alpha, d)\varepsilon^{\alpha/d}.$$

Writing $J'$ for the complement of $J$ in $I$, we now handle the more substantial contribution due to the points in $C$. We write $\tilde{Q}_j = Q_j - A$ and note $C = \bigcup_{j\in J'}\tilde{Q}_j$, so

$$M(\{X_i: X_i \in C\}) \le d^{\alpha/2}|J'| + \sum_{j\in J} M\big(\{X_i: X_i \in \tilde{Q}_j\}\big). \tag{5.9}$$

Setting $\gamma_j \equiv \int_{Q_i} g(x)\,dx = P(X_i \in \tilde{Q}_i)$, we note $g$ is constant on each member of the partition $\{Q_i\}_{i\in I}$ so we can write $g(x) \equiv \gamma_j l^{-d}$ for all $x$ in $Q_j$.

Conditional on the event $\{X_i \in \tilde{Q}_j\}$ the random variable $X_i$ has the uniform distribution on $\tilde{Q}_j$, so scaling, Theorem 2 and the law of large numbers tell us that with probability 1,

$$\begin{aligned} M\big(\{X_i: X_i \in \tilde{Q}_j\}\big) &\sim c(\alpha, d)l^{\alpha}|\{X_i: X_i \in \tilde{Q}_j\}|^{(d-\alpha)/d} \\ &\sim c(\alpha, d)l^{d}(\gamma_j n)^{(d-\alpha)/d}. \end{aligned} \tag{5.10}$$

When we sum the previous bounds on the contributions of $A$ and $B$ and the contribution expressed by (5.10), we have

$$\begin{aligned} &\limsup_{n\to\infty} M(X_1, X_2, \ldots, X_n)n^{-(d-\alpha)/d} \\ &\quad\le \beta'(\alpha, d)\{\varepsilon^{(d-\alpha)/d} + \varepsilon^{\alpha/d}\} + c(\alpha, d) \sum_{j\in J'} \gamma_j^{(d-\alpha)/d} l^d. \end{aligned} \tag{5.11}$$

The arbitrariness of $\varepsilon > 0$ and the bound $\sum_{j\in J}\lambda(Q_j) \le \varepsilon$ can now be applied in (5.11) permitting us to conclude that for any blocked density $g(x)\,dx + \mu_s$ we have the upper bound

$$\limsup_{n\to\infty} M(X_1, X_2, \dots, X_n) n^{-(d-\alpha)/d} \le c(\alpha, d) \int g(x)^{(d-\alpha)/d}\,dx. \tag{5.12}$$

The corresponding lower bound on the limit inferior is a bit more subtle. For $A$, $B$ and $C$ defined as before, we have by Lemma 2.5 that

$$\begin{aligned} M(X_1, X_2, \dots, X_n) \ge\ & M(\{X_i\colon X_i \in B \cup C\}) \\ & -\beta''(\alpha, d)|\{X_i\colon X_i \in A\}|^{(d-\alpha)/d}. \end{aligned} \tag{5.13}$$

Now, given a minimal spanning tree $T$ of $\{X_i\colon X_i \in B \cup C\}$, we take $\chi$ to be the set of elements of $\{X_i\colon X_i \in C\}$ which are joined by an edge of $T$ to an element of $\{X_i\colon X_i \in B\}$.

The minimal spanning tree of $\{X_i\colon X_i \in B \cup C\}$ together with a minimal spanning tree of $\chi$ will contain a spanning tree of the set $\{X_i\colon X_i \in C\}$, so we have the crude bound

$$M(\{X_i\colon X_i \in C\}) \le M\big(\{X_j\colon X_j \in B \cup C\}\big) + M(\chi). \tag{5.14}$$

To bound the cardinality $|\chi|$, we note that for each element of $\chi$ there is either (1) an endpoint of an edge of length greater than $y$ or (2) a point of $C$ within a distance $y$ of $B$. Thus

$$|\chi| \le \nu_d(y) + \left|\left\{X_i\colon X_i \in C,\ \min_{\omega\in B}|X_i - \omega| \le y\right\}\right|. \tag{5.15}$$

Since $C$ is disjoint from $B$ and the singular support of $\{X_i\}$, we have the bound

$$P\Big(X_i \in C \text{ and } \min_{\omega\in B}|X_i - \omega| \le y\Big) \le \|g\|_\infty |J| \left\{(l + 2y)^d - l^d\right\},$$

where $\|g\|_\infty = \sup|g(x)| < \infty$. Also, by (5.6) we have $\sum_{j\in J}\lambda(Q_j) = l^d|J| \le \varepsilon$, so the preceding bound gives us for $0 < y < l/2d$ that

$$P\Big(X_i \in C \text{ and } \min_{\omega\in B}|X_i - \omega| \le y\Big) \le 2e\|g\|_\infty y l^{-1}\varepsilon d, \tag{5.16}$$

where we have used $1 + x \le e^x \le 1 + ex$ for $0 \le x \le 1$.

By Lemma 2.2 applied to $M(\chi)$ and by the law of large numbers applied to inequalities (5.15) and the probability bound (5.16), we have

$$\limsup_{n\to\infty} M(\chi) n^{-(d-\alpha)/d} \le \beta''(\alpha, d)\big(2e\|g\|_\infty y l^{-1}\varepsilon d\big)^{(d-\alpha)/d}. \tag{5.17}$$

Also by (5.13) and (5.14), we have

$$\begin{aligned} M(X_1, X_2, \dots, X_n) \ge\ & M(\{X_i\colon X_i \in C\}) - M(\chi) \\ & -\beta''(\alpha, d)|\{X_i\colon X_i \in A\}|^{(d-\alpha)/d}. \end{aligned} \tag{5.18}$$

The last two terms have already been bounded; so as before, our problem is reduced to calculating the contribution from $C$.

To begin, take a spanning tree $T$ of $\{X_i\colon X_i \in C\}$ 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 $\tilde{Q}_j$. Also, let $\chi_j$ denote the set of points in $\tilde{Q}_j$ which are joined by an edge of $T$ to a point in the complement, $\tilde{Q}_j^c$. Since $D_j$ together with a MST of $\chi_j$ will span $\{X_i\colon X_i \in \tilde{Q}_j\}$, we have

$$M\Big(\big\{X_i\colon X_i \in \tilde{Q}_j\big\}\Big) \le \sum_{e \in D_j} |e| + \beta'(\alpha, d) l^{\alpha} |\chi_j|^{(d-\alpha)/d}, \tag{5.19}$$

which, after summing over $J'$, yields

$$\begin{aligned}\sum_{j \in J'} M\Big(\big\{X_i\colon X_i \in \tilde{Q}_j\big\}\Big) &\le M\big(\{X_j\colon X_j \in C\}\big)\\ &\quad + \beta'(\alpha, d) l^{\alpha} \sum_{j \in J'} |\chi_j|^{(d-\alpha)/d}.\end{aligned} \tag{5.20}$$

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

$$\sum_{j \in J'} |\chi_j| \le \nu_d(y) + \sum_{j \in J'} \left| \left\{ X_i\colon X_i \in \tilde{Q}_j,\ \min_{\omega \in \tilde{Q}_j^c} |X_i - \omega| \le y \right\} \right| \tag{5.21}$$

and, as in (5.16), we have for $0 < y < l/2d$ that

$$P\left( X_i \in \tilde{Q}_j \text{ and } \min_{\omega \in \tilde{Q}_j^c} |X_i - \omega| \le y \right) \le 2e \|g\|_{\infty} y l^{d-1} d. \tag{5.22}$$

By (5.21), (5.22), Hölder's inequality and the law of large numbers, we have

$$\begin{aligned}&\limsup_{n \to \infty} \left( \sum_{j \in J'} |\chi_j|^{(d-\alpha)/d} \right) n^{-(d-\alpha)/d}\\ &\qquad \le |J'|^{(d-\alpha)/d} \big(2e\|g\|_{\infty} y l^{d-1} d\big)^{(d-\alpha)/d}.\end{aligned} \tag{5.23}$$

Now, since $|J'| l^d \le 1$, inequality (5.23), inequality (5.20) and the fact that $y$ is arbitrary except for the constraint $0 < y < l/2d$, we can take $yl^{-1}$ to be as small as we like and conclude

$$\begin{aligned}&\liminf_{n \to \infty} M\big(\{X_i\colon X_i \in C\}\big) n^{-(d-\alpha)/d}\\ &\qquad \ge \liminf_{n \to \infty} \sum_{j \in J'} M\Big(\big\{X_i\colon X_i \in \tilde{Q}_j\big\}\Big) n^{-(d-\alpha)/d}.\end{aligned} \tag{5.24}$$

Finally, for each $\tilde{Q}_j$, relation (5.10) says the limit of $M(\{X_i\colon X_i \in \tilde{Q}_j\}) n^{-(d-\alpha)/d}$ exists and equals $c(\alpha, d)\gamma_j^{(d-\alpha)/d} l^d$, so summing (5.24), we have

$$\liminf_{n \to \infty} M\big(\{X_i\colon X_i \in C\}\big) n^{-(d-\alpha)/d} \ge c(\alpha, d) \int_C g(x)^{(d-\alpha)/d}\, dx. \tag{5.25}$$

Since the Lebesgue measure of $C$ is at least $1 - \varepsilon$, inequality (5.25) completes the proof of a lower bound on the limit infimum which complements the bound on the limit superior given in (5.12). We have therefore completed the proof of Theorem 3 under the restriction that $\psi(x) = |x|^{\alpha}$. The conclusion of the proof of

Theorem 3 under the more general assumption that $\psi(x) \sim x^{\alpha}$ as $x \to 0$ comes from Lemma 5.2.

LEMMA 5.2. *If $x_i$, $1 \le i < \infty$, is any sequence of points in $\mathbb{R}^d$ such that for $0 < \alpha < d$, we have $M'(x_1, x_2, \dots, x_n) \to \infty$ as $n \to \infty$, where*

$$M'(x_1, x_2, \dots, x_n) = \min_T \sum_{e \in T} |e|^{\alpha}.$$

*Then if $\psi$ is monotone and $\psi(x) \sim x^{\alpha}$ as $x \to 0$ and*

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

*we have*

$$M(x_1, x_2, \dots, x_n) - M'(x_1, x_2, \dots, x_n) = o(M'(x_1, x_2, \dots, x_n))$$

*as $n \to \infty$.*

PROOF. If we choose the edges of the two minimal spanning trees of $\{x_1, x_2, \dots, x_n\}$ under weight functions $\psi(x)$ and $|x|^{\alpha}$ by using Prim's algorithm, we will obtain the same tree $T$ because of the monotonicity of $\psi$. The absolute difference of the tree weights is therefore given by

$$\Delta_n = \left| \sum_{e \in T} \psi(|e|) - \sum_{e \in T} |e|^{\alpha} \right| = \sum_{e \in T} \varepsilon(|e|)|e|^{\alpha},$$

where $\varepsilon(x) = |\psi(x) - x^{\alpha}|x^{-\alpha}$ and $\varepsilon(x) \to 0$ as $x \to 0$. Since

$$\Delta_n \le \sup_{0 < x \le \delta} \varepsilon(x) \sum_{\substack{e \in T \\ |e| \le \delta}} |e|^{\alpha} + \sup_{0 < x \le d^{1/2}} \varepsilon(x) \sum_{\substack{e \in T \\ |e| > \delta}} |e|^{\alpha}, \tag{5.26}$$

we see the second sum of (5.26) is bounded independently of $n$ by inequality (2.7) of Lemma 2.2 and the first sum of (5.26) is bounded by

$$\sup_{x \le \delta} \varepsilon(x) M(x_1, x_2, \dots, x_n).$$

By our choice of $\delta > 0$, we can make $\sup_{x \le \delta} \varepsilon(x)$ as small as we like, so we see that (5.26) implies $\Delta_n = o(M'(x_1, x_2, \dots, x_n))$, as claimed. □

**6. Concluding remarks.** The results given here concern the almost sure convergence of a sequence of normalized random variables to a constant $c(\alpha, d)$. The natural questions associated with such a result are:

1. Can the constant be determined?
2. Can the basic strong law be supplemented with a result which provides information about the rate of convergence?
3. Can the strong law be supplemented with distributional results such as a central limit theorem?

Concerning the value of the constants $c(\alpha, d)$, there are some analytic bounds and Monte Carlo estimates, but there is little hope for an exact analytic determination. Gilbert (1965) proved that $c(1,2) \le 2^{-1/2} \approx 0.707$ and cited experimental evidence that suggested $c(1,2) \approx 0.68$ as a good approximation. Roberts (1968) proved that $0.5 \le c(1,2)$ and $0.554 \le c(1,3) \le 0.698$. Through a more extensive Monte Carlo analysis, Roberts (1968) estimated that $c(1,2) \approx 0.656$ with a standard deviation of 0.002 and $c(1,3) \approx 0.668$ with a standard deviation of 0.002.

There are two issues concerning the possibility of more detailed information on the rate of convergence. First, what can be said about the rate of convergence of the normalized means $m_n n^{-(d-\alpha)/d}$ to $c(\alpha, d)$? Here experience in the area of subadditive processes suggests that progress is unlikely. On the other hand, several approaches are likely to give more detailed information about the asymptotic size of $M_n - EM_n$. For example, the Efron–Stein inequality can be used iteratively as in Steele (1981b) to obtain moment bounds on $M_n - EM_n$, e.g., with $\psi(x) = |x|$ and $d = 2$, one can show that for each $1 \le p < \infty$ there is a constant $B_p < \infty$ such that

$$E(M_n - EM_n)^p \le B_p \tag{6.1}$$

for all $n \ge 1$. Even stronger bounds on the tails of $M_n - EM_n$ should be possible using the inequalities of Section 2 together with the interpolation technique of Rhee and Talagrand (1986), which has proved remarkably effective in the traveling salesman problem.

Concerning the possibility of a central limit theorem, there is encouraging progress owing to Ramey (1982), who proved that a central limit theorem would follow if one could verify a certain complicated Ansatz which expresses a type of conditional independence between distant parts of the minimal spanning tree. The required Ansatz is analogous to results in statistical mechanics which have been rigorously justified, but so far there is no complete proof of a central limit theorem for minimal spanning trees.

Outside of this set of classical questions, there is an intriguing and delicate problem raised by Robert Bland. For $X_i$, i.i.d. $U[0,1]^2$, Bland observed empirically that the sums $\sum_{e \in T}|e|^2$ seem to converge as $n \to \infty$. Bland's observations and conjecture provided basic motivation for the present analysis of power weighted minimal spanning trees, and it is intriguing that the case $\gamma = d$ seems particularly resistive. This conjecture bears an interesting relation to the result of Frieze (1985) concerning the weight $M_n$ of the MST for the graph with edge weights chosen independently from the uniform distribution on $[0,1]$. Frieze proved that $M_n$ converges in probability (and expectation) to the constant $\zeta(3) = 1.202\ldots$. An earlier result of Timofeev (1984) only established that $EM_n \le 3.29$, but the method of Timofeev is still intriguing because of its generality.

**Acknowledgments.** Persi Diaconis, Ingram Olkin, Dan Ramey, Geoffrey Grimmett and Joe Marhoul have been kind enough to comment on the Stanford Technical Report 212 which focused on $\psi(x) = |x|$ and which formed the initial basis for part of this report. The motivation to study the problem of power weighted edges was provided by Robert Bland. The present exposition has been improved by comments provided by Timothy Snyder and the referees.

## REFERENCES


APOSTOL, T. M. (1976). *Introduction to Analytic Number Theory*. Springer, New York.

BEARDWOOD, J., HALTON, H. J. and HAMMERSLEY, J. M. (1959). The shortest path through many points. *Proc. Cambridge Philos. Soc.* **55** 299–327.

DUDLEY, R. M. (1968). Distances of probability measures and random variables. *Ann. Math. Statist.* **39** 1563–1572.

DUDLEY, R. M. (1976). *Probabilities and Metrics, Convergence of Laws on Metric Spaces, with a View to Statistical Testing. Lecture Note Series* **45** 18.1–18.7. Matematisk Institut, Aarhus Universitet.

EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9** 586–596.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

FRIEZE, A. M. (1985). On the value of a random spanning tree problem. *Discrete Appl. Math.* **10** 47–56.

GILBERT, E. N. (1965). Random minimal trees. *J. Soc. Indust. Appl. Math.* **13** 376–387.

GILBERT, E. N. and POLLAK, H. O. (1968). Steiner minimal trees. *SIAM J. Appl. Math.* **16** 1–29.

KRUSKAL, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Amer. Math. Soc.* **7** 48–50.

MANDELBROT, B. B. (1977). *Fractals, Form, Chance, and Dimension*. Freeman, San Francisco.

PAPADIMITRIOU, C. H. and STEIGLITZ, K. (1982). *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, Englewood Cliffs, N.J.

PRIM, R. C. (1957). Shortest connection networks and some generalizations. *Bell System Tech. J.* **36** 1389–1401.

RAMEY, D. B. (1982). A non-parametric test of bimodality with application to cluster analysis. Thesis, Yale Univ.

RHEE, W. T. and TALAGRAND, M. (1986). Martingale inequalities, interpolation, and NP-complete problems. *Math. Oper. Res.* To appear.

ROBERTS, F. D. K. (1968). Random minimal trees. *Biometrika* **55** 255–258.

STEELE, J. M. (1981a). Subadditive Euclidean functionals and nonlinear growth in geometric probability. *Ann. Probab.* **9** 365–376.

STEELE, J. M. (1981b). Complete convergence of short paths and Karp's algorithm for the TSP. *Math. Oper. Res.* **6** 374–378.

STEELE, J. M. (1982). Optimal triangulation of random samples in the plane. *Ann. Probab.* **10** 548–553.

STEELE, J. M. (1986). Probabilistic algorithm for the directed traveling salesman problem. *Math. Oper. Res.* **11** 343–350.

STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

TIMOFEEV, E. A. (1984). Random minimal trees. *Theory Probab. Appl.* **29** 134–141.

WIDDER, D. V. (1946). *The Laplace Transform*. Princeton Univ. Press, Princeton, N.J.


PROGRAM IN OPERATIONS RESEARCH AND STATISTICS
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544