The Annals of Probability
1982, Vol. 10, No. 3, 548–553

OPTIMAL TRIANGULATION OF RANDOM SAMPLES IN THE PLANE

BY J. MICHAEL STEELE[1]

Stanford University

Let T_n denote the length of the minimal triangulation of n points chosen independently and uniformly from the unit square. It is proved that $T_n/\sqrt{n}$ converges almost surely to a positive constant. This settles a conjecture of György Turán.

1. Introduction. Triangulation of finite sets occurs naturally in several areas of mathematical analysis, and the object of this article is to take some beginning steps in the probabilistic study of the triangulation process. The main result obtained is a proof of a conjecture of György Turán (1980) on the rate of growth of the minimal triangulation of n points independently and uniformly distributed in the unit square.

Turán's conjecture was initially motivated by an envisioned analogue to Karp's probabilistic algorithm for the Traveling Salesman Problem (Karp, 1977). This in turn was motivated in part by the relatively recent development of practical fixed point algorithms. (See, for example, Karamardian, 1977).

In view of the apparently special nature of the process studied here, it seems worth noting that the method employed can be used in several problems which deal with growth rates in geometric probability. The present technique will hopefully form a useful complement to the theory of subadditive Euclidean functionals (Steele, 1980).

To state the minimization problem precisely, first let $S = \{x_1, x_2, \cdots, x_n\}$ denote a finite set of points $x_i \in R^2$. By a triangulation of S, we will mean a decomposition of the convex hull of S into triangles such that each $x_i \in S$ is a vertex of some triangle. It is *not* required that each vertex of the triangulation necessarily be an element of S.

By the length of triangulation we mean the sum of the lengths of all of the edges in the triangulation. The central quantity of interest here is $T(x_1, x_2, \cdots, x_n)$ which denotes the minimum length over all possible triangulations of S.

The following result, originally conjectured by György Turán, will be proved in the next two sections.

THEOREM 1. *If* $T_n = T(X_1, X_2, \cdots, X_n)$ *where* X_i, $1 \le i < \infty$ *are independent and uniformly distributed in* $[0, 1]^2$, *then*

$$\lim_{n\to\infty} \frac{T_n}{\sqrt{n}} = \beta$$

with probability one for some constant $\beta > 0$.

In the fourth section, the assumption of uniform distribution is relaxed, and some special results are obtained. The section also mentions several open problems.

2. Mean asymptotics. To establish a limit as in Theorem 1 it is almost essential to first show $ET_n \sim \beta\sqrt{n}$. This will be done here by means of a Poisson "smoothing" and Tauberian argument. By Π we denote a planar Poisson point process with intensity 1, so

Received June 1980; revised May 1981.

[1] Research supported in part by NSF Grant No. MCS 78-07736.

AMS 1980 *subject classifications.* Primary, 60F15, 60D05; secondary, 68C05, 68E10.

Key words and phrases. Triangulation, probabilistic algorithm, subadditive Euclidean functionals, jackknife, Efron-Stein inequality.

for each Borel $A \subset \mathbb{R}^2$, $\Pi(A)$ is a set of points uniformly distributed in A. Further, the cardinality $|\Pi(A)|$ is a Poisson random variable with parameter $\lambda(A)$, the Lebesque measure of A.

LEMMA 2.1. *The expectation* $\phi(t) = E(T(\Pi[0, t]^2))$ *satisfies*

$$\phi(t) \le m^2\phi(t/m) + Cmt \log t$$

for some constant C, *all* $m \in \mathbb{Z}^+$ *and* $t \in (3, \infty)$.

PROOF. Let $[0, 1]^2$ be decomposed into m^2 cells Q_j, $1 \le j \le m^2$, of edge length $1/m$, and set $tQ_j = \{x : x = ty,\ y \in Q_j\}$. Also, for any set of points $A \subset \mathbb{R}^2$ let $e(A)$ denote the cardinality of the set of extreme points of the convex hull of A. The main observation is that

$$T(\Pi[0, t]^2) \le \textstyle\sum_{j=1}^{m^2} T(\Pi(tQ_j)) + 4\cdot 2^{-1/2}m^{-1}t \sum_{j=1}^{m^2} e(\Pi(tQ_j)) + 2(m + 1)t. \tag{2.1}$$

To establish (2.1) we note that a triangulation of $\Pi[0, t]^2$ can be obtained by taking the minimal triangulations of each $\Pi(tQ_j)$, and then completing these to a general triangulation as follows:
(a) Adjoin each of the boundaries $\partial(tQ_j)$, $1 \le j \le m^2$, and thus add a total length of $2(m + 1)t$, and
(b) extend the triangulation of $\Pi(tQ_j)$ to a triangulation of $\Pi(tQ_j) \cup \{$extreme points of $tQ_j\}$. This second action has a total cost bounded by $4\cdot(2^{-1/2}/m)\cdot t\cdot e(tQ_j)$, since there are certainly no more than $4e(tQ_j)$ lines of length (uniformly) bounded by $t\cdot 2^{-1/2}/m$ which are necessary to complete the triangulation within tQ_j.

We now take expectations in (2.1) and note that by Euclidean translation

$$ET(\Pi(tQ_j)) = ET(\Pi[0, t/m]^2) = \phi(t/m), \quad 1 \le j \le m^2$$

If X_i are independent and uniformly distributed in $[0, 1]^2$, Renyi and Sulanke (1963) proved that

$$E(e(X_1, X_2, \cdots, X_n)) \le A \log n, \quad n \ge 3, \tag{2.2}$$

for a constant A. By elementary bounds on the Poisson distribution, this shows $Ee(\Pi[0, t]) \le A' \log t$ for all $t \ge 3$, and some constant A'. Returning to (2.1) with these bounds, one completes the proof of Lemma 2.1.

LEMMA 2.2. *If a non-decreasing function* $\phi(t)$ *satisfies*

$$\phi(t) \le m^2\phi(t/m) + Cmt \log t \tag{2.3}$$

for some $c > 0$ *and all* $m \in \mathbb{Z}^+$ *and* $t \ge t_0$, *then*

$$\lim_{t\to\infty} \phi(t)/t^2 = \liminf_{t\to\infty} \phi(t)/t^2 = \beta.$$

PROOF. From (2.3) we have

$$\frac{\phi(mt)}{(mt)^2} \le \frac{\phi(t)}{t^2} + C\frac{\log(mt)}{t}$$

and setting $\psi(t) = \phi(t) + t^{3/2}$ we have

$$\frac{\psi(mt)}{(mt)^2} \le \frac{\psi(t)}{t^2} \tag{2.4}$$

for all $m \le t^{-1} \exp(\sqrt{t}/2c)$. If we let

$$A_t = \left\{ m : \frac{\psi(mt)}{(mt)^2} \le \frac{\psi(t)}{t^2} \right\} = \{m_1 < m_2 < \cdots < m_k < \cdots\},$$

then by (2.4) applied recursively we see A_t contains all integers which are of the form 2^a3^b with $a, b \in \mathbb{Z}^+$. From this last observation and elementary number theory one can check that

$$\lim m_{k+1}/m_k = 1.$$

Now, for any $\varepsilon > 0$ we can choose a t_0 such that

$$\psi(t_0)/t_0^2 \le \beta + \varepsilon.$$

But for any s, we can find m_k and m_{k+1} in A_{t_0} such that $m_k t_0 \le s < m_{k+1} t_0$ and apply the monotonicity of ψ to obtain

$$\frac{\psi(s)}{s^2} \le \frac{\psi(m_{k+1}t_0)}{(m_k t_0)^2} \le \frac{\psi(m_{k+1}t_0)}{(m_{k+1}t_0)^2} \cdot \frac{m_{k+1}^2}{m_k^2},$$

which yields

$$\lim\sup_{s\to\infty} \frac{\psi(s)}{s^2} \le (\beta + \varepsilon).$$

This last result implies $\psi(s) \sim \beta s^2$, and since $\phi(s) = \psi(s) + o(s^2)$ the proof is complete.

All that remains is the Tauberian argument to complete the main result of this section.

LEMMA 2.3. $ET_n \sim \beta\sqrt{n}$.

PROOF. From the definition of $\phi(t)$ and the rescaling of $[0, 1]^2$ to $[0, t]^2$, one can calculate that

$$\phi(t) = t \textstyle\sum_{n=0}^{\infty} (ET_n)e^{-t^2}t^{2n}/n!.$$

The two previous lemmas can now be translated by change of variables to

$$\textstyle\sum_{n=0}^{\infty} (ET_n)e^{-u}u^n/n! \sim \beta\sqrt{u} \quad \text{as} \quad u \to \infty. \tag{2.5}$$

By the Abelian theorem for Borel summability (Doetch, 1943), (2.5) implies that as $x \uparrow 1$, one has

$$\textstyle\sum_{n=1}^{\infty} (ET_n - ET_{n-1})x^n \sim \beta\Gamma\left(\frac{3}{2}\right)(1 - x)^{-1/2}. \tag{2.6}$$

But now the (carefully chosen) definition of T_n implies $ET_{n+1} \ge ET_n$, so the Karamata Tauberian theorem (Feller, 1971) is applicable. This means that (2.6) also implies the basic result,

$$ET_n = \textstyle\sum_{k=1}^{n} (E(T_k) - E(T_{k-1})) \sim \beta\sqrt{n}.$$

3. Stabilizing variances. Rather than attempt to bound the variance of T_n directly it seems useful to introduce a close approximation whose variance is more easily bounded. We therefore consider the new random variables $T'_n = T(X_1, X_2, \cdots, X_n, (0, 0), (0, 1), (1, 0), (1, 1))$; and thus, instead of just triangulating $\{X_1, X_2, \cdots, X_n\} \subset [0, 1]^2$, one also includes the corners in the triangulation.

To check that T'_n is close to T_n, one just notes that

$$0 \le T'_n - T_n \le 4\sqrt{2}\, e(X_1, X_2, \cdots, X_n) + 4 \quad \text{for} \quad X_i \in [0, 1]^2, \quad 1 \le i \le n. \tag{3.1}$$

The first inequality in (3.1) follows from the definition of T and the second inequality follows from the same consideration applied in the proof of Lemma 2.1. Since in the case of X_i i.i.d. uniform on $[0, 1]^2$ we have $Ee(X_1, X_2, \cdots, X_n) \le A \log n$, the inequalities in (3.1) and Lemma 2.3 give in that case the asymptotic equivalence,

$$ET'_n \sim ET_n \sim \beta\sqrt{n} \quad \text{as} \quad n \to \infty. \tag{3.2}$$

The key tool in obtaining the required variance bound of this section is the Efron-Stein inequality (Efron-Stein 1981). If $S(Y_1, Y_2, \cdots, Y_{n-1})$ is any real symmetric function of the independent identically distributed random vectors Y_j, $1 \le j < \infty$, we set $S_i = S(Y_1, Y_2, \cdots, Y_{i-1}, Y_{i+1}, \cdots, Y_n)$ and $S_. = 1/n \sum_{i=1}^n S_i$. The Efron-Stein jackknife inequality then says

$$\text{(3.3)} \qquad \operatorname{Var} S_{n-1} \le E \textstyle\sum_{i=1}^n (S_i - S_.)^2.$$

This inequality will now be applied to the random variable $T'(X_1, X_2, \cdots, X_n)$.

LEMMA 3.1. *If X_i, $1 \le i < \infty$, are independent and identically distributed with support contained in* $[0, 1]^2$, *then*

$$\operatorname{Var} T'_{n-1} \le 6\sqrt{2}\, ET'_n, \quad n \ge 2.$$

PROOF. We first note that in the Efron-Stein inequality the right side is not decreased if $S_.$ is replaced by any other function of $Y_1, Y_2, \cdots, Y_n$. Applying the resulting inequality to T'_n we have

$$\text{(3.4)} \qquad \operatorname{Var} T'_{n-1} \le E \textstyle\sum_{i=1}^n (T'(X_1, X_2, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n) - T'_n)^2.$$

There are two ways to bound the differences in (3.4). If we extend a minimal triangulation of $\{(0, 0), (1, 0), (0, 1), (1, 1), X_1, X_2, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n\} \equiv A_i$ to a triangulation of all of $\{(0, 0), (1, 0), (0, 1), (1, 1), X_1, X_2, \cdots, X_n\} \equiv A$, then an increment of at most $3\sqrt{2}$ is made. This crude bound comes from the fact that X_i is contained in some triangle Δ of the first triangulation, and X_i can be connected to the three vertices of Δ at a cost less than $3\sqrt{2}$.

A second way of bounding the differences in (3.4) comes from the fact that if X_i is *removed* from a minimal triangulation $\mathscr{T}$ of A, the decrease in the cost is at most $\sum_{j \in I(i)} |X_i - X_j|$ where $I(i) = \{j : (X_i, X_j) \text{ is an edge of } \mathscr{T}\}$.

From the trivial inequality, $\min(a^2, b^2) \le ab$ applied to the two preceeding bounds and (3.4), we obtain

$$\text{(3.5)} \qquad \operatorname{Var} T'_{n-1} \le E \textstyle\sum_{i=1}^n 3\sqrt{2} \sum_{j \in I(i)} |X_i - X_j|.$$

The reason this is effective is that

$$\text{(3.6)} \qquad \textstyle\sum_{i=1}^n \sum_{j \in I(i)} |X_i - X_j| = 2T'_n,$$

because each edge of the minimal triangulation $\mathscr{T}$ of A appears exactly twice in the double sum. From (3.5) and (3.6) the lemma follows immediately.

Now it is possible to complete the proof of Theorem 1. First recall that when the X_i, $1 \le i < \infty$, are independent and uniformly distributed on $[0, 1]^2$ the relation (3.2) $ET'_n \sim \beta\sqrt{n}$ is valid. By Lemma 3.1, Chebyshev's inequality, and a 2ε argument, we then see that the probabilities $P(|T'_{n_k} - \beta\sqrt{n_k}| < \varepsilon)$ are summable for $n_k = k^3$, $1 \le k < \infty$. By Markov's inequality and the Renyi-Sulanke bound (2.2) the probabilities $P(e(X_1, X_2, \cdots, X_{n_k}) \ge \varepsilon\sqrt{n_k})$ are also summable. The inequalities (3.1) therefore imply that

$$\text{(3.7)} \qquad \lim_{k\to\infty} T_{n_k}/\sqrt{n_k} = \beta$$

with probability one. Since T_n is non-decreasing and $n_{k+1}/n_k \to 1$, the subsequence limit (3.7) suffices to prove the existence of the limit in Theorem 1.

To check that β is indeed positive we just note that

$$\text{(3.8)} \qquad 2T_n \ge \textstyle\sum_{i=1}^n \min\{|X_j - X_i| : 1 \le j \le n, \quad j \ne i\},$$

since each X_i must be joined to some other X_j by an edge of the triangulation. Now a trite calculation shows that there is a constant $\alpha > 0$ so that $E \min\{|X_j - X_i| : 1 \le j \le n, j \ne i\} \ge \alpha n^{-1/2}$ and hence that $ET_n \ge \beta\sqrt{n}$ for a $\beta > 0$.

4. General distributions and open problems. If one assumes only that the X_i, $1 \le i < \infty$ are independent and identically distributed with $E|X_i|^2 < \infty$, it still seems inevitable that $\lim_{n\to\infty} T_n/\sqrt{n}$ exist with probability one. The natural approach to such an extension, as used in the case of subadditive Euclidean functions (Steele, 1980), does not seem to help in this context even when the X_i have bounded support. The problem seems to arise from the fact that the functional $T_n = T(X_1, X_2, \cdots, X_n)$ does not have the "simple subadditivity" property.

Nevertheless, the next lemma shows that T_n does have one of the crucial properties used in extending the theory of subadditive Euclidean functionals to general distributions, the "scale boundedness" property.

LEMMA 4.1. *There is a constant C such that*

$$T'(x_1, x_2, \cdots, x_n) \le C\sqrt{n}$$

for all $n \ge 1$ *and* $x_i \in [0, 1]^2$.

PROOF. For $n \ge 1$ one sets $\psi(n) = \max T'(x_1, x_2, \cdots, x_n)$ where the maximum is over all possible choices of $x_i \in [0, 1]$. We also set $\psi(0) = 4$. The key observation is that

$$\psi(n) \le \max_{a+b+c+d=n} \tfrac{1}{2}\{\psi(a) + \psi(b) + \psi(c) + \psi(d)\}. \tag{4.1}$$

This inequality is easily proved by decomposing $[0, 1]^2$ into four quadrants, correcting for the scale of the quadrants, and being careful of the cases which leave some quadrants empty. By induction one can then prove $\psi(n) \le 30\sqrt{n}$. (In fact, $C = 2\cdot 4\cdot (2 - \sqrt{3})^{-1} \cong 29.8$ will suffice.)

As an application of the preceding lemma, one can prove the following.

THEOREM 2. *If* X_i, $1 \le i < \infty$, *are independently distributed with compact support*

$$\lim_{n\to\infty} T(X_1, X_2, \cdots, X_n)/\sqrt{n} = 0 \quad \text{a.s.,} \tag{4.2}$$

if, and only if, the support of the $\{X_i\}$ *is singular with respect to Lebesgue measure.*

PROOF. First suppose the X_i have compact support S with Lebesgue measure zero, and then choose a finite cover of S by closed squares A_j, $1 \le j \le m$, with disjoint interiors. We can further suppose that $\sum_{j=1}^m \lambda(A_j) = (16)^{-1} \sum_{j=1}^m |\partial A_j|^2 \le \varepsilon$ where $\lambda(A_j)$ denotes the Lebesgue measure for A_j and $|\partial A_j|$ denotes the length of its boundary ∂A_j.

For each j, let E_j denote the (four) extreme points of A_j and complete the set of edges $\cup_{j=1}^m \partial A_j$ to a triangulation of $\cup_{j=1}^m E_j$. Let B be the total length of the resulting triangulation.

We then have the elementary bound

$$T(X_1, X_2, \cdots, X_n) \le \sum_{j=1}^m T(E_j \cup \{X_i : X_i \in A_j\}) + B.$$

By Lemma 4.1 with a change of scale to squares of side $\lambda(A_i)^{1/2}$ we obtain

$$T(X_1, X_2, \cdots, X_n) \le C \sum_{j=1}^m \lambda(A_i)^{1/2} |E_j \cup \{X_j : X_i \in A_j\}|^{1/2} + B,$$

so by Schwarz' inequality

$$T(X_1, X_2, \cdots, X_n) \le C(\textstyle\sum_{i=1}^m \lambda(A_i))^{1/2}(4m + n)^{1/2} + B.$$

Since $\sum_{i=1}^m \lambda(A_i) < \varepsilon$ was arbitrary, the first half of the lemma is thus proved.

To prove the converse we note as before that

$$2T_n \ge \sum_{i=1}^n \min\{|X_i - X_j| : 1 \le j \le n, \quad i \ne j\}. \tag{4.3}$$

By the Lebesgue density theorem one can show (without too much difficulty) that if the

X_i do not have purely singular support that there is a constant $\delta > 0$ so that

$$E \min\{|X_1 - X_j| : 1 < j \le n\} > \delta n^{-1/2}. \tag{4.4}$$

But if (4.2) holds, the scale boundedness of Lemma 4.1 and the bounded convergence theorem imply $ET_n = o(\sqrt{n})$. This is incompatible with (4.4), so the theorem is proved.

The fact that $T_n/\sqrt{n} \to 0$ a.s. when the $\{X_i\}$ are singular with compact support suggests that more refined results might be possible in the singular case. This suspicion is confirmed by several special results of which the following is typical.

THEOREM 3. *If* X_i, $1 \le i < \infty$, *are independent and uniformly distributed on the unit circle, then*

$$T_n - ET_n = o(\log^\alpha n) \quad \text{a.s.} \tag{4.5}$$

for any constant $\alpha > ½$ *and*

$$\gamma_1 \log n \le ET_n \le \gamma_2 \log n, \tag{4.6}$$

for some constants $0 < \gamma_1 < \gamma_2 < \infty$.

The proof of (4.5) is not difficult and can be based on the Efron-Stein inequality which in this case shows Var $T_n = O(n^{-1})$. The bounds in (4.6) are not difficult if one does not aim for good values of γ_1 and γ_2. Since it is very likely that one actually has $ET_n \sim \gamma \log n$, there is little loss in omitting the proofs of the bounds (4.6).

Acknowledgment. The author wishes to thank Joe Marhoul, Larry Shepp, and a very careful referee for many useful comments. In particular, the proof given here of Lemma 2.2 corrects an earlier error and uses an idea suggested by the referee.

REFERENCES

DOETSCH, G. (1943). *Theorie und Anwendung der Laplace-Transformation*, 191–192. Dover, New York.

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, II*, 447–448. Wiley, New York.

KARAMARDIAN, S. (1977). *Fixed Points Algorithms and Applications* (in collaboration with C. B. Garcia). Academic, New York.

KARP, R. M. (1977). Probabilistic analysis of partitioning algorithms for the traveling salesman problem in the plane. *Math Oper. Res.* **2** 209–224.

RÉNYI, A. and SULANKE, R. (1963). Über die Konvexe Hülle von n zufällig gewählten Punkten. *Z. Wahrsch. Verw. Gebiete* **2** 75–85.

STEELE, J. M. (1980). Subadditive Euclidean functionals and nonlinear growth in geometric probability. *Ann. Probability.* **9** 365–376.

TURÁN, G. (1980). Informal talks, Stanford-Hungarian Academy Conference, Budapest, March 1980.

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
STANFORD, CA 94305