The expression on the right is zero for $\rho = 0$, tends to zero as ρ becomes large, and has a unique maximum for $\rho \geqslant 0$, viz.

$$(\theta-1)\{\phi/(2+\phi)\}^2\{(\epsilon-1)/\epsilon\}^2(1+\tfrac{1}{2}\theta\phi/\epsilon)^{-1} = M(\epsilon,\theta), \quad \text{say,}$$

where
$$\theta = (rk)/(\lambda t), \quad \phi = 1+(1+8\epsilon/\theta)^{\frac{1}{2}}.$$

Now $M(\epsilon,\theta)$ may be shown by elementary means to be an increasing function of θ for $\theta > 0$. Since $1 < (rk)/(\lambda t) = (1-1/t)/(1-1/k) < k/(k-1)$, an upper bound for the relative error is expressible as a function of ϵ and the block size k only; let this function be $U_k(\epsilon)$. Table 1 gives the lower and upper limits of ϵ when $U_k \leqslant 0{\cdot}01$ and $U_k \leqslant 0{\cdot}02$:

Table 1. *Lower and upper limits of ϵ for given limits for U_k*

k	2	3	4	5	8	10
$U_k \leqslant 0{\cdot}01$	0·726	0·670	0·629	0·596	0·523	0·488
	1·414	1·568	1·710	1·848	2·259	2·544
$U_k \leqslant 0{\cdot}02$	0·640	0·575	0·528	0·492	0·414	0·378
	1·646	1·921	2·194	2·473	3·408	4·147

Quite wide discrepancies in $\tilde{\rho}$ lead to only small losses of information, especially as the block size increases.

References

Kendall, M. G. & Stuart, A. (1966). *The Advanced Theory of Statistics*, vol. III. London: Griffin.

Roy, J. & Shah, K. R. (1962). Recovery of interblock information. *Sankhyā* A **24**, 269–80.

Shah, K. R. (1964). Use of interblock information to obtain uniformly better estimates. *Ann. Math. Statist.* **35**, 1064–78.

Yates, F. (1939). The recovery of inter-block information in variety trials arranged in three dimensional lattices. *Ann. Eugen. Lond.* **9**, 136–56.

[*Received July* 1967. *Revised August* 1967]

Random minimal trees

By F. D. K. ROBERTS

University of Liverpool

Summary

Random minimal trees in both 2 and 3 dimensions are discussed. The expected total length of the minimal trees connecting n points uniformly distributed in an area A, and in a volume V, are asymptotic to $\alpha_2(An)^{\frac{1}{2}}, \alpha_3(Vn)^{\frac{1}{3}}$ respectively, where α_2 and α_3 are constants. Upper and lower bounds are obtained on these constants and their values estimated by a Monte Carlo technique.

1. Introduction

Consider n points $P_1, P_2, \ldots, P_n$, which we wish to join together by straight lines to form a connected tree, or network. This tree will contain a path, directly or indirectly, between every pair of points P_i, P_j. For example, the points may be n terminals which we wish to connect together electrically. We only require $n-1$ connexions to construct a tree. According to Cayley's formula (see Riordan, 1958, p. 128) there are n^{n-2} distinct trees which can be drawn to connect n points.

The minimal tree for $P_1, P_2, \ldots P_n$ is the tree which joins these points with the minimal total length of connexions. Figure 1 shows the minimal tree connecting 25 points lying in a plane. A simple algorithm for constructing the minimal tree is given by Prim (1957). If $P_1, P_2, \ldots, P_n$ are uniformly distributed over a plane area A, then the expected total length of the connexions of the minimal tree is asymptotically proportional to $(An)^{\frac{1}{2}}$ (Beardwood, Halton & Hammersley, 1959). Hence the expected length of each connexion of the minimal tree is asymptotically proportional to $(A/n)^{\frac{1}{2}}$. Let $L_2(n)$ be the expected length of each connexion of the minimal tree joining n points uniformly distributed inside a circle of unit area, and $L_3(n)$ be the expected length of each connexion of the minimal tree joining n points uniformly distributed inside a sphere of unit volume. We have asymptotically as $n \to \infty$,

$$L_2(n) \sim \alpha_2 n^{-\frac{1}{2}}, \quad L_3(n) \sim \alpha_3 n^{-\frac{1}{3}}.$$

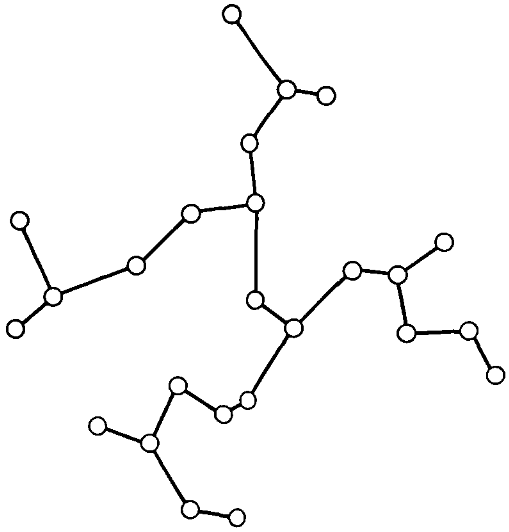

Fig. 1. Minimal tree for 25 points.

In § 2 we derive the following bounds on α_2 and α_3: $0{\cdot}500 \leqslant \alpha_2 \leqslant 0{\cdot}707$, $0{\cdot}554 \leqslant \alpha_3 \leqslant 0{\cdot}698$.

Section 3 describes a Monte Carlo technique to estimate α_2 and α_3. We estimate $\alpha_2 = 0{\cdot}656$ with standard deviation 0·002, and $\alpha_3 = 0{\cdot}668$ with standard deviation 0·002. An estimate $\alpha_2 = 0{\cdot}68$ and also the upper bound $\alpha_2 \leqslant 0{\cdot}707$ have already been obtained by Gilbert (1964).

2. Bounds on α_2 and α_3

To obtain a lower bound on α_2 we consider an infinite plane on which are scattered points according to the Poisson law with density λ. The expected size of each connexion of the minimal network joining these points is $\alpha_2/\sqrt{\lambda}$. Prim (1957) proves that every point of a minimal network is connected to its nearest neighbour. Every connexion of the minimal network is not necessarily between nearest neighbours, however, and hence the expected distance between nearest neighbours is a lower bound on the expected length of a connexion of the minimal tree. Feller (1966, p. 14) shows that the probability distribution of nearest neighbours in a two-dimensional Poisson ensemble is given by $f_2(r) = 2\lambda\pi r \exp(-\lambda\pi r^2)$. The expected distance between nearest neighbours is therefore

$$l_2 = \int_0^\infty 2\lambda\pi r^2 \exp(-\lambda\pi r^2)\, dr = 0{\cdot}5/\sqrt{\lambda}.$$

Hence we have a lower bound $0{\cdot}5 \leqslant \alpha_2$. In three dimensions, the probability distribution of nearest neighbours is given by $f_3(r) = 4\lambda\pi r^2 \exp(-\frac{4}{3}\lambda\pi r^3)$. The expected distance between nearest neighbours is therefore

$$l_3 = \int_0^\infty 4\lambda\pi r^3 \exp(-\tfrac{4}{3}\lambda\pi r^3)\, dr = 0{\cdot}554/\sqrt[3]{\lambda}.$$

Hence $0{\cdot}554 \leqslant \alpha_3$.

To obtain upper bounds on α_2 and α_3 we consider a tree which is not necessarily a minimal tree which we call an exodic tree. Gilbert (1964) has discussed this type of tree and has made an exact analysis of the mean length of an exodic tree for n points uniformly distributed in a circle. To construct an exodic tree we choose any point P_0 of the Poisson ensemble as origin, and label the remaining points $P_1, P_2, P_3, \ldots,$

the subscripts being chosen to order the points in increasing distance from the origin. For $i = 1, 2, 3, \ldots$ we connect P_i to a point P_j chosen from $P_1, P_2, \ldots, P_{i-1}$, to minimize the distance between P_i, P_j. We denote the asymptotic form of the distribution function of these connexions as $i \to \infty$ by $g_2(r)$. $g_2(r)\,dr$ is the probability that no points lie in a semicircle of radius r multiplied by the probability that a point lies between semicircles of radii r and $r+dr$. Hence

$$g_2(r)\,dr = \exp\left(-\tfrac{1}{2}\lambda\pi r^2\right)\lambda\pi r\,dr.$$

The expected length of the connexions is therefore given by

$$u_2 = \int_0^\infty \lambda\pi r^2 \exp\left(-\tfrac{1}{2}\lambda\pi r^2\right)dr = 0{\cdot}707/\surd\lambda.$$

Since the exodic tree is not necessarily the minimal tree we obtain an upper bound $\alpha_2 \leqslant 0{\cdot}707$. In the three dimensional case, the probability distribution of connexion lengths asymptotically approaches

$$g_3(r) = 2\lambda\pi r^2 \exp\left(-\tfrac{2}{3}\lambda\pi r^3\right).$$

Hence the expected length is given by

$$u_3 = \int_0^\infty 2\lambda\pi r^3 \exp\left(-\tfrac{2}{3}\lambda\pi r^3\right) = 0{\cdot}698/\sqrt[3]{\lambda}.$$

An upper bound on α_3 is therefore $\alpha_3 \leqslant 0{\cdot}698$.

3. Monte Carlo solution

The Monte Carlo technique was carried out on a KDF 9 computer using the algorithm suggested by Prim (1957). The minimal tree was constructed for n points uniformly distributed inside a circle of unit area and also a sphere of unit volume. Table 1 shows the values of n used. The random number generator used was one suggested by Downham & Roberts (1967). For each value of n, 10 runs were made and the mean length of each connexion recorded. Table 1 shows the estimates and standard deviations of α_2 and α_3. Since 750 points are not necessarily sufficient for the asymptotic formulae to hold, we use the following extrapolation technique to obtain more accurate estimates of α_2 and α_3. Let $\alpha_{2,n}$ be the estimate of α_2 obtained using n points. Then $\lim \alpha_{2,n} = \alpha_2$. Expanding $\alpha_{2,n}$ as a power series in $1/n$, we have approximately for large n, $\alpha_{2,n} = \alpha_2 + \beta_2/n$. A weighted least squares procedure was used to fit this straight line to the data for n = 100, 200, 300, 500, 750. The weights used were inversely proportional to estimates of the variance of $\alpha_{2,n}$ for each value of n. The three-dimensional case was treated similarly. The estimates obtained are $\alpha_2 = 0{\cdot}656$ with standard deviation 0·002 and $\alpha_3 = 0{\cdot}668$ with standard deviation 0·002. The time taken on KDF 9 to construct the minimal tree for 750 points is 9 min. The time for n points is approximately proportional to n^2.

Table 1. *Minimal tree connecting n random points uniformly distributed inside a circle of unit area and a sphere of unit volume. Estimates obtained are based on* 10 *computer runs for each value of n*

n	Estimate of α_2	Standard deviation of α_2	Estimate of α_3	Standard deviation of α_3
25	0·6797	0·0138	0·7411	0·0147
50	·6817	·0088	·7114	·0096
100	·6796	·0038	·6930	·0073
200	·6720	·0033	·6890	·0038
300	·6649	·0039	·6741	·0044
500	·6591	·0036	·6772	·0032
750	·6595	·0018	·6720	·0015

The author wishes to thank Dr S. H. Storey of the Data Processing Research Unit, Liverpool University, for his helpful suggestions and also the Science Research Council for their financial support.

REFERENCES

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

DOWNHAM, D. Y. & ROBERTS, F. D. K. (1967). Multiplicative congruential pseudo-random number generators. *Comput. J.* **10**, 74–7.

FELLER, W. (1966). *An Introduction to Probability Theory and its Applications*, vol. 2. New York: Wiley.

GILBERT, E. N. (1964). Random minimal trees. *J. Soc. Indust. Appl. Math.* **13**, 376–87.

PRIM, R. C. (1957). Shortest connection networks and some generalizations. *Bell System Tech. J.* **36**, 1389–401.

RIORDAN, J. (1958). *An Introduction to Combinatorial Analysis*. New York: Wiley.

[*Received October* 1967]

A three-dimensional cluster problem

BY F. D. K. ROBERTS AND S. H. STOREY

University of Liverpool

SUMMARY

The expected size of cluster in a three-dimensional Poisson ensemble is discussed. This expected size is a function of one parameter t. There exists a critical value $t = t_c$ at which the expected cluster size becomes infinite and for $t > t_c$ infinite clusters occur. A lower bound $t_c \geqslant 0{\cdot}0434$ and a Monte Carlo estimate $t_c = 0{\cdot}0889$ are obtained. An application of this problem to a percolation process is discussed.

1. INTRODUCTION

In a recent paper (Roberts, 1967) a two-dimensional cluster problem was discussed and a solution obtained by a Monte Carlo technique. In this paper we concern ourselves with the three-dimensional version of the problem.

Consider points scattered at random in three-dimensional Euclidean space according to the Poisson law with density λ. Two points are said to be connected if they lie within a distance $2R$ of each other. A cluster of size n is a set of n points each of which is connected to at least one other point in the set but none of which are connected to points not contained in the set. We consider only non-zero clusters. Although the distribution of the number of points in a cluster and the expected cluster size depend on the two parameters λ and R, dimensional analysis suggests that they are functions of one composite parameter $t = \lambda R^3$. Let $p_n(t)$ be the probability distribution of cluster sizes and $e(t)$ the expected cluster size. We assume that there exists a critical value $t = t_c$ at which $e(t)$ becomes infinite and that for values of $t > t_c$ infinite clusters occur. We obtain a graph of $e(t)$ as a function of t and estimate t_c by extrapolation.

In §2 we consider an application of this problem to a percolation process. A lower bound $t_c \geqslant 0{\cdot}0434$ is obtained in §3 and a Monte Carlo technique to obtain random samples from the distribution $p_n(t)$ is described in §4. An estimate $\hat{e}(t)$ of $e(t)$ is obtained by taking the mean value of 20 random samples from $p_n(t)$ for various values of t, see Fig. 1. In §4, an estimate of $t_c = 0{\cdot}0889$ with standard deviation $0{\cdot}0024$ is obtained by extrapolation. The Monte Carlo procedure was carried out on a KDF 9 computer.

2. A PERCOLATION PROCESS

If we consider each point of the Poisson ensemble as the centre of a spherical cavity of radius R within a solid, then for $t > t_c$ infinite clusters occur and a fluid can percolate through the solid.

Let ρ_0 be the density of the solid without the cavities and ρ the density with the cavities. We establish a relationship between ρ, ρ_0, λ, R as follows: ρ/ρ_0 represents the fraction of the volume occupied by the solid. By a theorem of Robbins (1944), this equals the probability that a point chosen at random is contained in the solid, i.e. is not contained in a cavity. This probability equals

$$\exp(-\tfrac{4}{3}\pi R^3\lambda)$$