Math. Proc. Camb. Phil. Soc. (1982), **91**, 315–322
Printed in Great Britain

Random uniform triangles and the alignment problem

By CHRISTOPHER SMALL
Statistical Laboratory, Cambridge

(*Received* 1 *April* 1981; *revised* 2 *July* 1981)

Abstract

Let n points be drawn independently and uniformly from a compact convex set K. The distribution of the shape of the resulting n-ad is determined and studied in the region corresponding to near alignment. Special attention is given to the case $n \geqslant 4$ and a table provided to help in the assessment of practical data (e.g. megalithic 'alignments'). The Broadbent factor, representing the effect of stretching the parent distribution, is computed explicitly.

1. *Introduction*

In work by S. R. Broadbent (2), and D. G. Kendall and W. S. Kendall (5), statistical techniques have been developed to determine whether a sample of points in the plane has a larger number of alignments than would be expected by chance. Kendall and Kendall (5) define three points to be ϵ-blunt (and therefore in approximate alignment) if the triangle formed by the points has a maximum internal angle within ϵ of π radians. They then extended the definition to multiple alignments of more than three points. This definition has the advantage over the one given in (2) that it is more easily computed. In this paper we shall follow the definition of (5).

At the time of writing, the most important application of these techniques is to determine whether megalithic stones, found at sites such as in Land's End in Cornwall, are significantly aligned. These alignments are known in the folklore as ley lines. One of the earliest systematic studies done of them was by A. Watkins (10). S. R. Broadbent (2) used uniform scatterings of points in rectangles to model the problem, and calculated the number of ϵ-blunt triads one would expect when ϵ is small. On the basis of his study he concluded that there was no significant alignment in the Land's End data. In (5), ϵ was treated as a nuisance parameter, and once again no significant alignment was found.

In what follows we shall see that the techniques used to study alignments can be regarded as a special case of what D. G. Kendall (4) has called the statistics of shape. We shall develop the statistics of shape for the case where the samples of points are drawn uniformly and independently from a bounded convex set, and use it to calculate probabilities of higher-order alignments.

2. *Shapes and their distributions*

D. G. Kendall (3) defines the shape of a planar sample to be the set of properties of the data which are invariant under translations, rotations and rescalings (i.e. homothetic

https://doi.org/10.1017/S0305004100059338 Published online by Cambridge University Press

expansions and contractions) of the sample. A sample in which all points are coincident will be said not to possess a shape. (See (3).) In addition, a minimum of three points is required for the sample to have a non-trivial shape.

The shape of a sample of n points, $n \geqslant 3$, can be represented as follows. Label the sample as points $A, B, C_1, C_2, \ldots, C_{n-2}$, so that A and B are not coincident. Let Q be the midpoint of AB. In triangle ABC_j, let $r_j = 2(\|QC_j\|/\|AB\|)$ and ϕ_j = angle from QB to QC_j taken in a counterclockwise direction. Then the shape of the sample can be represented by $(r_1, \phi_1, r_2, \phi_2, \ldots, r_{n-2}, \phi_{n-2})$. It should be noted that r_j can be zero or infinity, in which case ϕ_j is undefined. If $A = B$ and $n \geqslant 4$, the representation of shape breaks down. However we shall only consider cases where this occurs with probability zero.

If the samples are generated by some stochastic mechanism, then there is an induced distribution of shape on $(r_1, \phi_1, \ldots, r_{n-2}, \phi_{n-2})$. If the probability of two points being coincident is zero then A, B can be chosen arbitrarily. Suppose the points are independent and identically distributed with density $f(P)$ in the plane. In what follows, the sample shall be thought of as lying in $\mathbb{C}$, the complex plane. The sample has distribution $f(A)f(B)f(C)\ldots f(C_{n-2})\,dA\,dB\,dC_1\ldots dC_{n-2}$, where dA, dB, dC_j are differentials of area. Letting $Q = (A+B)/2$ be the midpoint of $\overline{AB}$, and $R = (B-A)/2$ be the vector from Q to B, the distribution is $4f(Q-R)f(Q+R)f(C_1)\ldots f(C_{n-2})\,dQ\,dR\,dC\ldots dC_{n-2}$. Fixing Q, R then $C_j = Q + Rr_j e^{i\phi_j}$. So $dC_j = \|R\|^2 r_j\,dr_j\,d\phi_j$. Therefore the density of the shape distribution with respect to the measure $\Pi_{j=1}^{n-2} r_j\,dr_j\,d\phi_j$ is

$$\tilde{f}(r_1, \phi_1, \ldots, r_{n-2}, \phi_{n-2}) = 4\int_{\mathbb{C}}\int_{\mathbb{C}} \|R\|^{2n-4}\left[\prod_{i=1}^{n-2} f(Q+Rr_j e^{i\phi_j})\right] f(Q-R)f(Q+R)\,dQ\,dR. \tag{2·1}$$

In the special case where f is the gaussian density, W. S. Kendall (6) has found the distribution of shapes of triplets. The calculation of the distributions of shapes for independent points in a convex set is very difficult unless the set is regular. In the case where points are independent and uniform in a circular disc, the shape density can be calculated by numerical integration. With an additional integration, the distribution of the maximum angle is found. (See Appendix for details.)

3. *The alignment problem*

We note that the alignment or ϵ-bluntness of a triad is a function of the shape of the triad. Points will be aligned if, for every j, $\phi_j = 0$ or π. Suppose that f is a uniform density on some bounded convex set K. Then $f(P) = I_K(P)/|K|$, where I_K is the indicator function of K, and $|K|$ is the area of K. The integrand of (2·1) is non-zero whenever $Q-R \in K$, $Q+R \in K$ and $Q+Rr_j e^{i\phi_j} \in K$ for all j. But if $\phi_j = 0$ or π, and $0 \leqslant r_j \leqslant 1$ for all j, then the fact that $Q-R \in K$ and $Q+R \in K$ implies that $Q+Rr_j e^{i\phi_j} \in K$ by convexity. So for these special shapes, (2·1) becomes

$$\tilde{f} = \frac{1}{|K|^n}\int_K\int_K \|R\|^{2n-4}\,dA\,dB = \frac{1}{4^{n-2}|K|^n}\int_K\int_K \|B-A\|^{2n-4}\,dA\,dB, \tag{3·1}$$

where

$$\phi_j = 0 \text{ or } \pi \quad \text{for all } j,$$
$$0 \leqslant r_j \leqslant 1 \quad \text{for all } j.$$

https://doi.org/10.1017/S0305004100059338 Published online by Cambridge University Press

This is constant over the range of shape parameters indicated. It is the simplicity of this integral in comparison with (2·1) which makes the uniform distribution on a convex set useful in modelling the alignment problem. The integral in (3·1) is related to the well-known Blaschke constants for the constant set K (see ((8), p. 46)). We write J_{2n-4} as the $(2n-4)$-th moment of the distance between two random points in K. So for aligned shapes as in (3·1) $f = 4|K|^{-(n-2)}J_{2n-4}$.

In (5), a set of n points, $n > 3$, is said to be ϵ-blunt if there exist two of these points, say A and B, such that any triangle ABC_j selected from the n points is ϵ-blunt at C_j. (In practice, ϵ will be very small.) If $\epsilon < \frac{1}{2}\pi$ then A and B will be unique. Following (5), we define the m-th order collinearity constant of K to be

$$\lambda^{(m)} = \lim_{\epsilon \to 0} \frac{P(\epsilon\text{-blunt } (m+2)\text{-ad})}{\epsilon^m}, \quad m \geqslant 1.$$

Our definition differs from (5) in that we do not condition on endpoints.

To compute $P(\epsilon$-blunt $(m+2)$-ad) fix endpoints to some preselected choice. In terms of the measure $\Pi_j r_j dr_j d\phi_j$, the measure of all ϵ-blunt $(m+2)$-ads with endpoints A, B is asymptotically $(\frac{4}{3}\epsilon)^m$. So from (3·1), $P(\epsilon$-blunt $(m+2)$-ads, endpoints $A, B)$ $\approx 4(|K|)^{-m}J_{2m}(\frac{4}{3}\epsilon)^m = \epsilon^m J_{2m}/(3|K|)^m$. There are $\binom{m+2}{2}$ ways of choosing A, B so

$$\lambda^{(m)} = \binom{m+2}{2} \frac{J_{2m}}{(3|K|)^m}. \tag{3·2}$$

For a unit circle,

$$J_{2m} = \frac{1}{m+1}\frac{1}{2m+3}\frac{1.3.5\ldots(2m+3)}{2.4.6\ldots(2m+4)}2^{2m+3},$$

((8), p. 48). Thus $J_2 = 1, J_4 = \frac{5}{3}$, giving $\lambda^{(1)} = 1/\pi$ as was found in (5) but also $\lambda^{(2)} = 10^3 \div 9\pi^2$ etc.

In order to study the significance of the alignments of data points, it seems reasonable to use the entire sample to estimate a convex set from which it might be supposed that the points are drawn uniformly and independently. Then, if a small value of ϵ is chosen, the expected number of m-th order alignments $(= \epsilon$-blunt $(m+2)$-ads) in a sample of N points is $\binom{N}{m+2}\lambda^{(m)}\epsilon^m$.

A natural estimate for K would be the maximum likelihood estimator, which is the convex hull of the data. J.-P. Rasson and B. Ripley (7) find an estimate for K which is based upon the convex hull. If we condition upon the vertices of the convex hull, then the points in its interior are uniform and independent over the convex hull. So if the vertices of the hull are discarded, the alignment problem is reduced to one in which the convex set is known. The 'price' to be paid is the loss of some data points.

S. Broadbent (2) used a rectangle fitted to the data. Similarly an ellipse can be fitted. As expected we shall see that the collinearity constant of an ellipse is an increasing function of its eccentricity. For rectangles the ratio of length to width determines the collinearity constant in a similar fashion. In view of this we investigate the change in shape distributions under a linear transformation of the generator.

https://doi.org/10.1017/S0305004100059338 Published online by Cambridge University Press

4. *Shapes and linear transformation*

Suppose that f is some density in the plane that is circularly symmetric about the origin. Let T be an area-preserving linear transformation of the plane. If $\tilde{f}_T$ is the shape distribution generated from the density fT then

$$\tilde{f}_T = 4\int_C\int_C \|R\|^{2n-4}\left[\prod_{k=2}^{n-2} f(TQ + T(Rr_k \exp(i\phi_k)))\right] f(TQ - TR) f(TQ + TR)\, dR\, dQ$$

$$= 4\int_C\int_C \|T^{-1}e^{i\alpha}\|^{2n-4}\rho^{2n-4}[\Pi f(U + V(r_k)_\alpha \exp i(\phi_k)_\alpha)] f(U - V) f(U + V)\, dU\, dV,$$

where $$U = TQ, \quad V = TR = \rho e^{i\alpha}, \quad \rho > 0$$

and $$(r_k)_\alpha = r_k \left\| \frac{T(e^{i\phi_k} T^{-1} e^{i\alpha})}{e^{i\alpha}} \right\|,$$

$$(\phi_k)_\alpha = \arg\left(\frac{T(e^{i\phi_k} T^{-1} e^{i\alpha})}{e^{i\alpha}}\right)$$

$$= \oint \|T^{-1}e^{i\alpha}\|^{2n-4}\left\{4\int_0^\infty\int_C \rho^{2n-3}[\Pi f(U + \rho e^{i\alpha}(r_k)_\alpha e^{i(\phi_k)_\alpha})] f(U - V) f(U + V)\, dU\, d\rho\right\} d\alpha.$$

Now by circular symmetry the expression

$$\{-\} \quad \text{equals} \quad [\tilde{f}((r_1)_\alpha, (\phi_1)_\alpha, \ldots, (r_{n-2})_\alpha, (\phi_{n-2})_\alpha)]/2\pi.$$

So

$$\tilde{f}_T(r_1, \phi_1, \ldots, r_{n-2}, \phi_{n-2}) = \frac{1}{2\pi}\oint \|T^{-1}e^{i\alpha}\|^{2n-4} \tilde{f}((r_1)_\alpha, (\phi_1)_\alpha, \ldots, (r_{n-2})_\alpha, (\phi_{n-2})_\alpha)\, d\alpha. \tag{4·1}$$

Once again, by restricting to aligned shapes the formula simplifies slightly. If $\phi_1, \ldots, \phi_{n-2} = 0$ or π then $(r_j)_\alpha = r_j$ and $(\phi_j)_\alpha = \phi_j$. So

$$\tilde{f}_T(r_1, \phi_1, \ldots, r_{n-2}, \phi_{n-2}) = \left\{\frac{1}{2\pi}\oint \|T^{-1}e^{i\alpha}\|^{2n-4} d\alpha\right\} \tilde{f}(r_1, \phi_1, \ldots, r_{n-2}, \phi_{n-2}), \tag{4·2}$$

where $\phi_j = 0$ or π for all j. The factor $\{(1/2\pi)\oint \|T^{-1}e^{i\alpha}\|^{2n-4} d\alpha\}$ is a generalization of the Broadbent factor in (5). Formula (4·2) will be shown later to be true under conditions more general than circular symmetry. Using (4·2) the collinearity constants for ellipses can be calculated. Suppose $f(P)$ is the uniform density on the unit disc and $T(x, y) = (s^{-\frac{1}{2}}x, s^{\frac{1}{2}}y)$ where $P = (x, y)$. Then

$$\oint \|T^{-1}e^{i\alpha}\|^{2n-4} d\alpha = \oint \left(s\cos^2\alpha + \frac{1}{s}\sin^2\alpha\right)^{n-2} d\alpha = 2\pi P_{n-2}\left(\frac{s + s^{-1}}{2}\right),$$

where P_{n-2} is the $(n-2)$-nd order Legendre polynomial. If $n = 3$ this becomes the Broadbent factor $(s + s^{-1})/2$ of (5). So if $\lambda_s^{(m)}$ is the m-th collinearity constant of the ellipse $sx^2 + s^{-1}y^2 = 1$

$$\lambda_s^{(m)} = P_m\left(\frac{s + s^{-1}}{2}\right)\lambda_1^{(m)}. \tag{4·3}$$

Therefore $\lambda_s^{(1)} = (s + s^{-1})/2\pi$, as in (5), but we also obtain $\lambda_s^{(1)} = 5(3s^2 + 2 + 3s^{-2})/36\pi^2$ etc.

https://doi.org/10.1017/S0305004100059338 Published online by Cambridge University Press

The calculation of the Broadbent factor under a stretch, s, provides a check on (3), where by a different method the Broadbent factor was calculated for the family of gaussian distributions.

The reader will find a reference table of collinearity constants at the end of the paper.

5. *A generalization*

We now generalize the conditions under which (4·2) is true. We use the following lemma.

LEMMA. *Let $L^2(X,\lambda)$ be the space of L^2 functions on the measure space (X,λ) with inner product $(f_1,f_2)=\int_X f_1 f_2\,d\lambda$. Suppose $\lambda(X)<\infty$ and $(e_\gamma)_{\gamma\in\Gamma}$ is a complete orthonormal basis for $L^2(X,\lambda)$ which contains the constant function $e_\# = [\lambda(X)]^{-\frac{1}{2}}$. Given $g,h\in L^2(X,\lambda)$ condition A implies condition B:*

(*A*) *If both $(g,e_\gamma)\neq 0$ and $(h,e_\gamma)\neq 0$ then $\gamma=\#$.*

(*B*) $$\int_X gh\,d\lambda = \frac{1}{\lambda(X)}\left(\int_X g\,d\lambda\right)\left(\int_X h\,d\lambda\right).$$

The proof is straightforward.∎

In particular, if X is the set of points $(\cos\alpha,\sin\alpha)$ in the plane, and $d\lambda = d\alpha$, the basis can be the set of normalized trigonometric functions.

THEOREM. *Let f be a density in the plane that is invariant with respect to rotations by $\pi/(2k-4)$ around the origin, for all $k=3,4,\ldots,n$. Then with T an area-preserving linear transformation, and $\phi_j = 0$ or π for all $j=1,\ldots,n-2$, $\tilde{f}_T(r_1,\phi_1,\ldots,r_{n-2},\phi_{n-2}) = (1/2\pi)\left(\int \|T^{-1}e^{i\alpha}\|^{2n-4}d\alpha\right)\tilde{f}(r_1,\phi_1,\ldots,r_{n-2},\phi_{n-2})$.*

Proof. To prove this, return to the derivation of (4·1) and consider the second last integral. We apply the Lemma to $g(\alpha)=\|T^{-1}e^{i\alpha}\|^{2n-4}$ and $h(\alpha)=\{-\}$. It can be seen that $g(\alpha)$ has, in general, non-vanishing Fourier coefficients in $\cos(2i\alpha)$ and $\sin(2j\alpha)$ only for $i,j=1,2,\ldots,n-2$, and $i=0$.

The rotational invariance condition implies that $(\{-\},\cos((2k-4)\alpha))=0$ for all $k=3,4,\ldots,n$. The result then follows immediately.∎

The conditions of the theorem for $T(x,y)=(s^{-\frac{1}{2}}x,s^{\frac{1}{2}}y)$ and $n=3$ were first developed in (6). We generalize to all T and all orders here.

6. *Tests for alignment*

The definition of approximate alignment used here is not the only one for testing the ley hypothesis. D. M. Behrend (1) has studied alignments using a different definition. In his method, n points are aligned at tolerance level c if there is a strip of width c containing all n points. The definition has the disadvantage that any sufficiently small n-ad is accepted as aligned regardless of its shape.

The definition of an aligned n-ad given in (2) is intuitively appealing for statistical tests. Unfortunately its first moment is difficult to calculate. In this definition an n-ad

https://doi.org/10.1017/S0305004100059338 Published online by Cambridge University Press

is aligned within tolerance level ϵ if every subset of three points is ϵ-blunt. We can call this ϵ-B-bluntness. It can be seen to be a more stringent definition than the one used here and is more in keeping with a 'common sense' notion of alignment. Define

$$\lambda_B^{(m)} = \lim_{\epsilon \to 0} \frac{P(\epsilon\text{-}B\text{-blunt } (m+2)\text{-ad})}{\epsilon^m}.$$

Fortunately, although $\lambda_B^{(m)}$ is difficult to calculate, (3·1) implies that $\lambda^{(m)}/\lambda_B^{(m)}$ is not a function of the convex set from which the points are drawn. Simulations for squares ((2)) show that $\lambda_B^{(2)} \approx 0{\cdot}073$. Therefore $\lambda^{(2)}/\lambda_B^{(2)} \approx 1{\cdot}7$ for any convex set. This independence is not of course true in general for non-uniform densities f in the plane. Nonetheless if f satisfies the conditions of the Theorem of section 5 and T is an area-preserving linear transformation of the plane another invariance can be demonstrated. Let $\lambda^{(m)}$ and $\lambda_B^{(m)}$ be the respective collinearity constants for $(m+2)$-ads with generator fT. Then the Theorem implies that $\lambda^{(m)}/\lambda_B^{(m)}$ is not a function of T. However, the ratio will depend upon the particular f chosen.

A difficult problem is the choice of an appropriate model for the test. The results of this paper suggest that the uniform convex model is a convenient one to use, but its suitability for the data may well vary from sample to sample. W. S. Kendall (6) has studied a gaussian model for random shapes. The test used in (5) required that the distribution of the maximum angle of a triangle be approximately uniform within 10° of 180°. In the Appendix, this is verified for K a circle. Analysis of the table at the end of this paper suggests that the Broadbent factor plays an important role in determining the alignment of points.

By fitting a convex set to the data (say by taking its convex hull) a problem of robustness arises. If the sample contains outliers then the estimate of K could be very poor. To overcome this, after discarding the vertices of the convex hull, the researcher can repeat the procedure and take the convex hull of the remaining data points. This convex hull stripping provides a more robust estimate for K at the loss of some information in the sample. For computational simplicity, the convex hull stripping can be replaced by minimal covering ellipse stripping. These techniques, currently being developed by P. Green, B. Silverman and others, are examples of the usefulness of computer algorithms in the analysis of spatial data. We note that both convex hull stripping and minimal covering ellipse stripping are consistent with affine transformations in the following sense. Let L be an affine transformation of the plane, and $K(S)$ the estimate of K from the sample S, using either convex hull or minimal covering ellipse procedures. Then $L(K(S)) = K(L(S))$. So the procedures can be used consistently with sections 4 and 5. The sample can be linearly transformed so that the minimal covering ellipse estimate for K is a circle before statistical analysis is done.

The actual distribution of the number of alignments is complicated, but an asymptotic theory for triads is available. Silverman and Brown (9) prove that the number of ϵ-blunt triads is approximately Poisson for large N, using a Poisson limit theorem for U-statistics. The simulations in (2) give some indication of the distribution of the number of higher-order alignments (in S. Broadbent's sense).

https://doi.org/10.1017/S0305004100059338 Published online by Cambridge University Press

The author would like to thank his Ph.D. supervisor, Professor D. G. Kendall, for his helpful comments on this paper.

Appendix

When K is a circular disc, the uniformity of f on K and the circular symmetry imply that (2·1) can be reduced to a twofold integration which is accessible by numerical techniques. The density of the maximum internal angle of a triad then requires one additional integration.

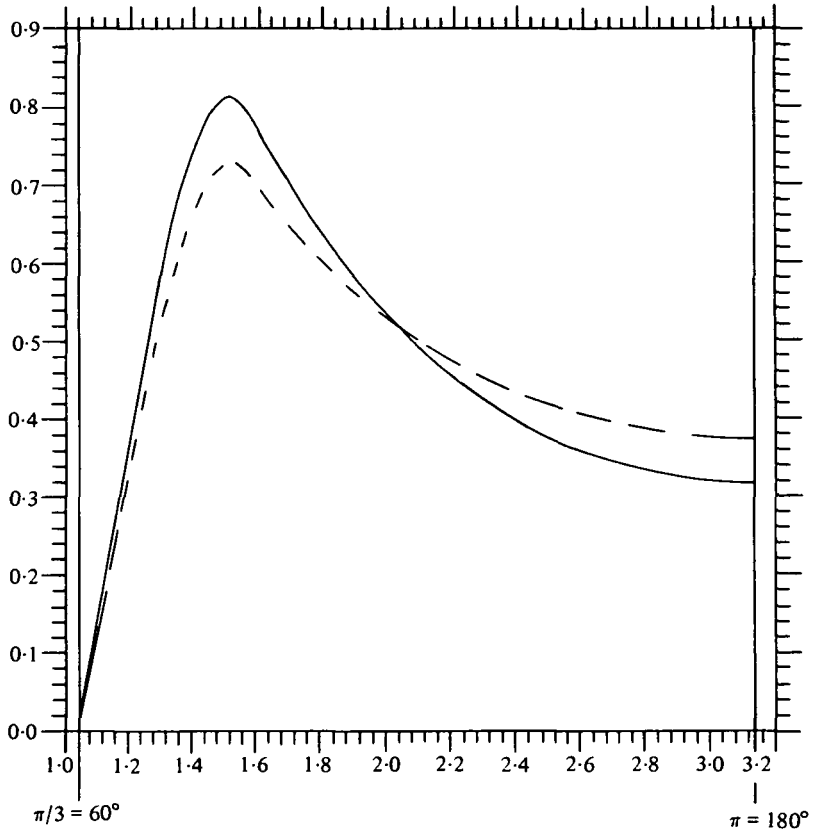


Fig. 1. Densities for the maximum internal angle (in radians). ——, Uniform triangle in a disc. – – –, Circular gaussian triangle (formula taken from (6)).

As can be seen, the density for the uniform triad (unbroken curve) is approximately constant within 10° of 180° as required for the test developed in (5). The density is also compared with the maximum angle density for a gaussian triad (broken curve). The reader is referred to (5) and (6) for more information about the gaussian maximum angle density.

The similarity of the two distributions suggests that the choice of type of distribution is not critical if circular symmetry is present. From the discussion by T. Brown at the end of (2) and the observation above, it is likely that a test using the pontogram

https://doi.org/10.1017/S0305004100059338 Published online by Cambridge University Press

method of Kendall and Kendall (5) is valid for a wide class of circularly symmetric distributions.

We finish by noting that the first collinearity constant for a circle can be read directly from the graph of the maximum angle density: it is the value of this density at π radians.

Table 1. *Collinearity constants for ellipses*

	$\frac{1}{s}x^2+sy^2=1$ for mth-order collinearity constant								
s	1·0	1·5	2·0	2·5	3·0	3·5	4·0	4·5	5·0
$m=1$	0·3183	0·3448	0·3979	0·4615	0·5305	0·6025	0·6764	0·7516	0·8276
$m=2$	0·1126	0·1419	0·2076	0·2988	0·4128	0·5488	0·7526	0·8851	1·085
$m=3$	0·4181 $\times 10^{-1}$	0·6495 $\times 10^{-1}$	0·1257	0·2277	0·3794	0·5901	0·8697	1·228	1·674
$m=4$	0·1597 $\times 10^{-1}$	0·3194 $\times 10^{-1}$	0·8299 $\times 10^{-1}$	0·1889	0·3787	0·6883	1·160	1·843	2·794
$m=5$	0·6213 $\times 10^{-2}$	0·1651 $\times 10^{-1}$	0·5769 $\times 10^{-1}$	0·1648	0·3969	0·8422	1·623	2·902	4·888

REFERENCES

(1) BEHREND, D. M. To appear.

(2) BROADBENT, S. R. Simulating the ley hunter. *J. Roy. Statist. Soc.* (*A*) **143** (1980), 109–140.

(3) KENDALL, D. G. Shape-manifolds, procrustean metrics and complex projective spaces. To be submitted to *Ann. Prob.* (1980).

(4) KENDALL, D. G. Foundations of a theory of random shape. To appear in *Bull. London Math. Soc.*

(5) KENDALL, D. G. and KENDALL, W. S. Alignments in two-dimensional random sets of points *Adv. Appl. Prob.* **12** (1980), 380–424.

(6) KENDALL, W. S. Random gaussian triangles. (1981), to appear.

(7) RIPLEY, B. D. and RASSON, J.-P. Finding the edge of a Poisson forest. *J. Appl. Prob.* **14** (1977), 483–491.

(8) SANTALO, L. A. *Integral geometry and geometric probability* (Addison-Wesley, 1976).

(9) SILVERMAN, B. W. and BROWN, T. C. Short distances, flat triangles and Poisson limits. *J. Appl. Prob.* **15** (1978), 815–825.

(10) WATKINS, A. *The Old Straight Track* (Methuen, 1925; republished Garnstone Press, 1970).

https://doi.org/10.1017/S0305004100059338 Published online by Cambridge University Press