# Intrinsic correlation in planar Poisson line processes

C.T.J. Dodson
School of Mathematics
University of Manchester
Manchester, M13 9PL, UK.

W.W. Sampson
School of Materials
University of Manchester
Manchester, M60 1QD, UK.

January 7, 2009

**Abstract**

The polygons arising from a planar Poisson line process have an exponential distribution of their side lengths and are known to be more regular as their area, perimeter or number of sides increase. Local regions with higher line density have smaller polygon side lengths and conversely. Numerical analysis of computer generated Poisson line processes shows that when pairs of adjacent polygon sides $(x, y)$ are sorted such that $x \leq y$ they are correlated with $\rho \approx 0.616$ as compared to $\rho = 1/\sqrt{5} \approx 0.447$ for independent sorted exponential $(x, y)$ pairs. This correlation is consistent with the observed regularity of polygons in realizations of planar Poisson line processes.
**Keywords:** Poisson lines, random polygons, correlation, Monte Carlo

## 1 Introduction

Random lines with uniformly distributed orientation, generated from a Poisson planar point process, partition the plane into random polygons and much is known about such processes. The lengths of polygon sides follows an exponential distribution, each line carrying a Poisson process of intersections with other lines. Miles [1, 2] showed that the expected number of sides per polygon is 4, the variance of the number of sides per polygon is $(\pi^2 - 8)/2$ and the perimeter of $n$-gons has a $\chi^2$ distribution with $2(n - 2)$ degrees of freedom. This means that perimeters of $n$-gons have equivalently a gamma distribution with dispersion parameter $(n - 2)$; in particular, for $n = 3$ the perimeter of triangles coincides with an exponential distribution.

Miles showed also that the fraction of polygons that are triangles is

$$p_3 = (2 - \frac{\pi^2}{6}) \approx 0.355.$$

Tanner [3] derived the fraction of quadrilaterals as

$$p_4 = \frac{1}{3} - \frac{7\pi^2}{36} + 4\int_0^{\pi/2} x^2 \cot x \, dx \approx 0.381.$$

The fractions $p_n$ of polygons with $n > 4$ sides is not known analytically, though Stoyan et al. [4] p325 provide numerical estimates collected from Monte Carlo methods:

$$p_5 \approx 0.192, \ p_6 \approx 0.059, \ p_7 \approx 0.013, \ p_8 \approx 0.002.$$

Realizations of isotropic random lines from simulations tend to yield random polygons that appear generally more ‘roundish’, nearly regular, rather than irregular in shape and this is reported also by Corte and Lloyd [5] for fibre networks made using laboratory and commercial filtration processes. See Miles [6] and Kovalenko [7] for proofs that this regularity is in fact a limiting property for random polygons as their area, perimeter or number of sides become large.

Such regularity suggests that the length of adjacent polygon sides is to some extent correlated, simply as a result of the random clustering of Poisson point processes. Indeed, first inspection of graphical representations of random isotropic homogeneous line processes reveals that denser regions tend to have shorter polygon sides than less dense regions. Product models for polygon areas exist, but they assume the length of adjacent polygon sides to be independent [5, 8]. Such models might be extended to account for correlation of the constituent variables, by applying a bivariate exponential or bivariate gamma distribution to represent polygon side lengths; we shall address this elsewhere. Here we report results from a Monte Carlo method to determine the correlation coefficient for lengths of adjacent polygon sides and this may be useful in guiding analytic models or in characterising other clustered spatial processes.

## 2 Equivalent ellipse models

We can exploit the result that the mean number of sides per polygon is 4 to represent the distribution of random polygons with exponentially distributed side lengths by a distribution of random rectangles. Such an approach has yielded results that agree with experiments in a range of practical cases [8, 9, 10]. Identify each random rectangle with an equivalent ellipse of the same area, with minor and major axes the random variables $x, y$. The simplest model gives $x$ an exponential distribution and $y$, which cannot be less than $x$, the distribution of $x + z$ where $z$ is an independent random variable with exponential distribution. It is easy to show that in this case we have correlation coefficient for $y$ on $x$ given in terms of the mean values $\bar{x}, \bar{z}$ by

$$\rho = \frac{1}{\sqrt{1 + (\bar{z}/\bar{x})^2}}. \tag{1}$$

In particular, we note that if $\bar{z} = \bar{x}$ then $\rho = \frac{1}{\sqrt{2}} \approx 0.707$, if $\bar{z} = 2\bar{x}$ then $\rho = \frac{1}{\sqrt{5}} \approx 0.447$ and if $\bar{z} = 3\bar{x}$ then $\rho = \frac{1}{\sqrt{10}} \approx 0.316$.

This is in fact a special case of the bivariate gamma model given by the McKay bivariate gamma distribution for correlated $0 < x < y$ which has joint probability density,

$$m(x,y) = \frac{\left(\frac{\alpha_1}{\sigma_{12}}\right)^{\frac{(\alpha_1+\alpha_2)}{2}} x^{\alpha_1-1}(y-x)^{\alpha_2-1} e^{-\sqrt{\frac{\alpha_1}{\sigma_{12}}}y}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}, \quad (2)$$

with parameters $\alpha_1, \sigma_{12}, \alpha_2 > 0$. The marginal probability densities of $x$ and $y$ are univariate gamma distributions with,

$$\bar{x} = \sqrt{\alpha_1 \, \sigma_{12}} \quad (3)$$

$$\bar{y} = (\alpha_1 + \alpha_2)\sqrt{\frac{\sigma_{12}}{\alpha_1}}. \quad (4)$$

The correlation coefficient between $x$ and $y$ is given by,

$$\rho = \sqrt{\frac{\alpha_1}{\alpha_1 + \alpha_2}}. \quad (5)$$

Since in our model we have $y - x = z$, the case of marginal exponential distributions for $x$ and $z$ is given by $\alpha_1 = \alpha_2 = 1$. Then it follows that $\rho = \frac{1}{\sqrt{2}}$, as in Equation 1 for $\bar{z} = \bar{x}$.

# 3 Sorted exponential model

We derive next the correlation of independent, but sorted, pairs of exponentially distributed variables as a reference against which the outcomes of numerical procedures in the sequel may be compared. We seek to estimate the correlation coefficient $\rho$, between the random variables $x, y$ drawn from an exponential distribution with unit mean with the ordering $x \leq y$. We start with pairs of randomly chosen numbers from the exponential distribution; hence the mean value of the product of these pairs is 1. Then convert each pair $\{x_i, y_i\}$ into an ordered pair $(x_i, y_i)$ such that $x_i \leq y_i$ and create now two distributions, one for the first member $x$ and one for the second member $y$. Intuitively, we take any $y_i < x_i$ from the source distribution of $y$ and add them to source distribution of $x$; also we take any $x_i > y_i$ from the source distribution of $x$ and add these to the source distribution of y. Note that the mean product of pairs $\overline{xy} = 1$ unaltered; however, the ordered pairs are no longer independent. This yields the probability density function for $x$

$$g_{\leq}(x) = \frac{1}{2}\, e^{-2\,x} \;\text{ so } \bar{x} = \frac{1}{2} \;\text{ and } var(x) = \frac{1}{4}. \quad (6)$$

Also, the probability density function for $y$ is

$$g_{\geq}(y) = 2\,e^{-2\,y}(1 - e^{-y}) \;\text{ so } \bar{y} = \frac{3}{2} \;\text{ and } var(y) = \frac{5}{4}. \tag{7}$$

It follows that

$$\rho = \frac{\overline{x\,y} - \bar{x}\,\bar{y}}{\sqrt{var(x)\;var(y)}} = \frac{1}{\sqrt{5}}. \tag{8}$$

# 4 Simulated random polygons

We expect a consequence of the clustering of crossings in Poisson line processes to be that pairs $(x, y)$ representing adjacent sides of polygons cannot be considered as independent. To probe the correlation between the lengths of adjacent polygon sides, we have written *Mathematica* code that extracts pairs $(x, y)$ representing the lengths of the adjacent sides of polygons arising from a Poisson line process in a unit square.

The code solves the equations of lines drawn at random within the unit square to generate the coordinates of all crossings that occur among them. Coordinates of each crossing are identified by the lines that generate it, allowing the coordinates of the adjacent crossings on these lines to be extracted; from these the lengths of adjacent pairs of polygon sides are calculated.

Note that we consider only pairs of polygon sides bounded entirely by the unit square. Where either of a pair of adjacent polygon sides cross the sides of the unit square, these are discounted from the analysis. Importantly, discarding these polygon sides from our analysis had no significant influence on the distribution of polygon sides, which was exponential, as expected. Given this, we can be confident that any difference between the correlation computed from our simulation and that calculated for independent polygon sides is an intrinsic feature of the network structure and not an artefact arising from the way the problem has been encoded within the software.

Networks with an increasing number of lines per unit area were generated using 10 random seeds, permitting the correlation to be tracked as a function of process intensity. For processes of 1000 lines in the unit square we calculate the correlation between more than a million pairs of adjacent polygon sides and observe a correlation of $\rho = 0.616 \pm 0.001$. We observe the same correlation in networks of 500 lines with a confidence interval varying only in the fourth decimal place. Note also that whereas the correlation for individual line processes may exceed this value at low process intensities, the mean correlation observed over our 10 cases was always less than 0.616 for process intensities less than 500 lines in the unit square. It is interesting that for process of 20 or more lines per unit area, the observed correlation was always greater than the value $\rho = \frac{1}{\sqrt{5}} \approx 0.447$ found in Equation 8 and $\rho$ increases rapidly towards its limiting value 0.616 with increasing intensity. For more discussion of the modelling of stochastic fibrous networks see the recent books Arwini and Dodson [9] and Sampson [10].

## References


[1] R.E. Miles. Random polygons determined by random lines in a plane. *Proc. Nat. Acad. Sci. USA* 52 901-907,1157-1160 (1964).

[2] R.E. Miles. The various aggregates of random polygons determined by random lines in a plane. *Advances in Math.* 10, 256-290 (1973).

[3] J.C. Tanner. The proportion of quadrilaterals formed by random lines in a plane. *J. Appl. Probab.* 20(2), 400-404 (1983).

[4] D. Stoyan, W.S. Kendall and J. Mecke. **Stochastic geometry and its applications** $2^{\text{nd}}$ edition, John Wiley, Chichester 1995.

[5] H. Corte and E.H. Lloyd. Fluid flow through paper and sheet structure. In **Consolidation of the Paper Web** *Trans. IIIrd Fundamental Research Symposium* (F. Bolam, ed.), pp981-1009, BPBMA, London 1966.

[6] R.E. Miles. A heuristic proof of a long-standing conjecture of D.G. Kendall concerning the shapes of certain large random polygons. *Adv. in Appl. Probab.* 27(2), 397-417 (1995).

[7] I. Kovalenko. A simplified proof of a conjecture of D.G. Kendall concerning shapes of random polygons. *J. Appl. Math. Stochastic Anal.* 12(4), 301-310 (1999).

[8] C.T.J. Dodson and W.W. Sampson. Modeling a class of stochastic porous media. *App. Math. Lett.* 10(2), 87-89 (1997).

[9] Khadiga Arwini and C.T.J. Dodson. **Information Geometry Near Randomness and Near Independence**. Lecture Notes in Mathematics, Springer-Verlag, New York, Berlin 2008.

[10] W.W. Sampson. **Modelling Stochastic Fibrous Materials with *Mathematica***. Springer-Verlag, London 2009.