Statistics of Non-Poisson Point Processes in Several Dimensions

Kenneth A. Brakke
Department of Mathematical Sciences
Susquehanna University
Selinsgrove, Pennsylvania 17870
brakke@susqu.edu

originally about 1986
current version March 16, 2006

Abstract

The Poisson point process is the most commonly studied random point process, but there are others. It is not always justifiable to assume that a random point process will have the same statistics as a Poisson point process. This paper derives the relationship between Poisson processes and some other random processes for some types of geometric statisics.

Keywords: Poisson point process, random point process, stochastic geometry, Voronoi tessellation, Dirichlet tessellation, Theissen tessellation.

AMS Classification: 60. Probability theory and stochastic processes.

1 Introduction

The phrase "a random set of points in space" is in itself ambiguous. The most popular version of randomness is a Poisson point process, but there are many other natural candidates in various circumstances. The Poisson process is often the most tractable analytically due to the independence of disjoint regions, but an actual model may have a predetermined number of points randomly scattered in a fixed area. Can the theoretical results obtained for a Poisson process be easily converted to apply to other random processes? This paper answers affirmatively for a certain class of statistic and certain other natural processes.

The motivation for this paper comes from some computer simulations of random Voronoi tessellations I have been doing. I came up with some embarassing discrepancies from theoretical values. The discrepancies were slight, but statistically significant. Eventually, it occured to me that the theoretical results were for a Poisson point process, but my simulation program was not using precisely a Poisson process. I was able to derive the adjustments necessary to have the the theoretical results apply to my program's process, and the discrepancies were explained. I present my adjustments here for the benefit of those who might be similarly involved with non-Poisson point processes.

A Voronoi tessellation is defined as follows: Let S be a set of points in a space R^n, and let each point of the space be associated with the nearest point of S. The space is thereby partitioned into convex polyhedra, or *cells*. Such a partition is called a *Voronoi tessellation*, also known as a *Dirichlet* or *Theissen tessellation*. When S is generated randomly, the result is a *random Voronoi tessellation*. Such patterns turn up in the crystallization of metals [1,2], geography [3], pattern recognition [4], numerical interpolation [5], and many other subjects. To date, the only random method for generating S that has been considered theoretically has been the Poisson point process. Theoretical results are contained in Meijering [1], Gilbert [2], and Brakke [7].

This paper considers several closely related of ways generating points: a Poisson process on a finite space, a Poisson process in addition to a fixed lattice of points, and a fixed number of random points on a finite space. Statistics for these are related to the corresponding statistics for Poisson processes on full n-space.

2 Qualifying random variables and statistics

Suppose U is a random variable that is defined to be a function of a domain W and a set S of points not necessarily contained in W. We will make the following assumptions about U:

1. U is extensive; that is, if W_1 and W_2 are disjoint domains, then $U(W_1 \cup W_2, S) = U(W_1, S) + U(W_2, S)$.

2. U is local; that is, for a given W and a given S, there is a distance R such that if any change is made to S at a distance greater than R from W, then U(W, S) is not changed.

3. U is homogeneous; that is, there is an m such that if W and S are scaled up by a factor of R, then U(W, S) is multiplied by R^m.

As the canonical example of this paper, S is the set of seeds of a Voronoi tessellation and U is the total of some cell statistic. Examples would be cell area, cell perimeter squared, the total sides of the neighboring cells, or the area of five sided cells only. A cell would be included in the tally only if its seed is in the domain.

The statistic of interest will be the expected value of U for a fixed domain W for S generated by some random process, which may depend on some parameters. The domain W will be assumed to be of unit n-dimensional measure in an n-dimensional space. The notation for the expectation of U will have the process type as a subscript and parameter values as conditions. For example, $E_P(U|D)$ would stand for the expected value of U for a Poisson process of density D.

3 Random processes

The types of random processes considered in this paper are:

P: Poisson process of density D.

F: Fixed number N of random points in the domain.

FP: A fixed number N of random points plus a Poisson process of density D.

PL: A Poisson process of density D plus a set of points mutually distant enough that their local effects do not overlap. This might rise from a relatively widely spaced lattice of points, hence the “L”. The key lattice parameter is M, the overall density of lattice points.

FL: A fixed number N of random points a lattice of density M. This is the model used in my simulations.

A Poisson point process of density D in R^n is defined by the property that the probability that a set W of n-dimensional measure A contains N generated points is given by the Poisson distribution:

$$Pr(N) = \frac{(AD)^N}{N!} e^{-AD}. \tag{3.1}$$

Due to the scale invariance of space, a local statistic u is proportional to a power of the density D, so for the sum over a fixed region we may write

$$E_P(U|D) = C_U \cdot D^q \tag{3.2}$$

for some constant C_U and some power q, not necessarily integral or positive. If the dimension of space is n and the local statistic has length dimensionality m, then q for the sum is

$$q = 1 - \frac{m}{n}. \tag{3.3}$$

4 Fixed number of random points

If there are N random points in W, it is indistinguishable whether they are due to a Poisson process of density D or to a random choice of N points. Hence a Poisson process may be considered as an ensemble of fixed number processes with Poisson probability for each number. Therefore, recalling that W is of measure one,

$$E_P(U|D) = C_U \cdot D^q = \sum_{N=0}^{\infty} E_F(U|N) \frac{D^N}{N!} e^{-D}. \tag{4.1}$$

Equation (4.1) may be solved as follows:

$$E_F(U|N) = C_U \cdot \frac{N!}{(N-q)!}. \tag{4.2}$$

This is valid for any real q with negligible error if factorial is interpreted in terms of the gamma function.

Nonintegral q are definitely of interest, so it is convenient to expand (4.2) in an asymptotic series in 1/N:

$$E_F(U|N) = C_U \cdot N^q \cdot \left[1 - \frac{q(q-1)}{2N} + \frac{q(q-1)(q-2)(3q-1)}{24N^2} + \cdots\right]. \tag{4.3}$$

Hence

$$E_F(U|N) = E_P(U|N) \cdot \left[1 - \frac{q(q-1)}{2N} + \frac{q(q-1)(q-2)(3q-1)}{24N^2} + \cdots\right], \tag{4.4}$$

and inversely,

$$E_P(U|N) = E_F(U|N) \cdot \left[1 + \frac{q(q-1)}{2N} + \frac{q(q-1)(q+1)(3q-2)}{24N^2} + \cdots\right]. \tag{4.5}$$

As a sample application, suppose one were trying to estimate the mean cell perimeter for a Poisson tessellation in the plane with simulations of 500 cells at a time on a unit torus. The proper statistic to accumulate is total cell perimeter Σp, and the scaling power is $q = 1/2$. One has an estimate of $E_F(\Sigma p|500)$ from the simulation totals, and would calculate from (4.5) and (2.1)

$$\begin{aligned} E_P(\Sigma p|500) &= E_F(\Sigma p|500) \cdot [1 - \frac{1}{4000} + \frac{1}{3200000} + \cdots], \\ C_p &= E_P(\Sigma p|500)/\sqrt{500}). \end{aligned} \tag{4.6}$$

5 Fixed number of random points plus Poisson process

Suppose M points are chosen at random in W, and a Poisson process of density D adds more. This process can be written as an ensemble of fixed number processes, with the probability of fixed number $M + N$ being the Poisson probability of N points. Letting $E_{FP}(U|M, D)$ be the statistic, we have

$$E_{FP}(U|M, D) = \sum_{N=0}^{\infty} E_F(U|M + N) \frac{D^N}{N!} e^{-D}. \tag{5.1}$$

Plugging in (3.3),

$$E_{FP}(U|M, D) = C_U \cdot \sum_{N=0}^{\infty} \frac{(M + N)!}{(M + N - q)!} \frac{D^N}{N!} e^{-D}. \tag{5.2}$$

Some index gymnastics gives

$$E_{FP}(U|M, D) = C_U \cdot \sum_{N=0}^{\infty} \sum_{i=0}^{M} \binom{M}{i} \frac{q!}{(q - i)!} \frac{D^N}{(i + N - q)!} e^{-D}, \tag{5.3}$$

where $\binom{M}{i}$ is the binomial coefficient. Interchanging the order of summation and summing over N yields with negligible error

$$E_{FP}(U|M, D) = C_U \cdot \sum_{i=0}^{M} \binom{M}{i} \frac{q!}{(q - i)!} D^{q-i}. \tag{5.4}$$

The total point density is $M + D$, so it is convenient to expand (5.4) in an asymptotic series in $M + D$, assuming M is small compared to D:

$$E_{FP}(U|M, D) = C_U \cdot (M + D)^q \cdot \left[1 - \frac{q(q - 1)M}{2(M + D)^2} + \cdots\right] \tag{5.5}$$

$$= E_P(U|M + D) \cdot \left[1 - \frac{q(q - 1)M}{2(M + D)^2} + \cdots\right]. \tag{5.6}$$

6 Lattice plus Poisson process

Suppose there are points generated by a Poisson process of density D in W, and M well separated points are added. By the nature of Poisson processes and the separatedness, the expected change caused by each added point will be the same, and will be the same as for the case $M = 1$ in section 5. Hence

$$\begin{aligned} E_{PL}(U|D,M) &= E_P(U|D) + M(E_{FP}(U|D,1) - E_P(U|D)) \\ &= C_U \cdot D^q + MC_U q \cdot D^{q-1}. \end{aligned} \tag{6.1}$$

The asymptotic expansion is

$$E_{PL}(U|D,M) = E_P(U|D,M) \cdot \left[1 - \frac{q(q-1)M^2}{2(D+M)^2} + \cdots\right]. \tag{6.2}$$

7 Lattice plus fixed number

We finally reach the case that was the original motivation for this research. In simulating random Voronoi tessellation, one method is to start with simple tessellation of a flat torus and add random points up to the desired number, updating the tessellation as points are added. A judicious starting tessellation is useful to eliminate wraparound complications and simplify the updating algorithm. Of course, one must adjust the results as shown here if one wishes to compare them to the theoretical values for a Poisson point tessellation.

Let $E_{FL}(U|M,N)$ be the expected sum for a starting lattice of M points and an additional N random points. We use the same tactic as in section 3, writing $E_{PL}(U|D,M)$ as a combination of $E_{FL}(U|M,N)$:

$$E_{PL}(U|D,M) = \sum_{N=0}^{\infty} E_{FL}(U|M,N)\frac{D^N}{N!}e^{-D}. \tag{7.1}$$

Expanding $E_{PL}(U|D,M)$ with (6.1) gives

$$C_U \cdot D^q + MC_U q \cdot D^{q-1} = \sum_{N=0}^{\infty} E_{FL}(U|M,N)\frac{D^N}{N!}e^{-D}. \tag{7.2}$$

This is solved by

$$E_{FL}(U|M,N) = C_U \cdot \frac{N!}{(N-q)!} + C_U qM \cdot \frac{N!}{(N-q+1)!}. \tag{7.3}$$

In terms of the total density $M + N$, the asymptotic expansion is

$$E_{FL}(U|N,M) = \tag{7.4}$$

$$E_P(U|M+N) \cdot \left[1 - \frac{q(q-1)}{2(M+N)} - \frac{q(q-1)M(M-1)}{2(M+N)^2} + \frac{q(q-1)(q-2)(3q-1)}{24(M+N)^2} + \cdots\right].$$

To get a feel for the significance of the various terms, consider estimating the total of the square of cell area a^2 in a plane with $M = 8$ and $M + N = 700$. Then $q = -1$, and

$$E_{FL}(\Sigma a^2|8,692) = E_P(\Sigma a^2|700) \cdot \left[1 - \frac{1}{700} - \frac{112}{980000} + \frac{12}{11760000} + \cdots\right]. \tag{7.5}$$

As another typical example, consider the total of the square of cell volume v^2 in three dimensions with $M = 32$ and $M + N = 2000$. Again $q = -1$, and

$$E_{FL}(\Sigma v^2|32,1968) = E_P(\Sigma v^2|2000) \cdot \left[1 - \frac{1}{2000} - \frac{992}{400000} + \frac{12}{96000000} + \cdots\right]. \tag{7.6}$$

The number of terms that would need to be used to adjust simulation results depends on the statistical accuracy of the simulation, which is inversely proportional to the square root of the total number of cells. For totals around $100,000,000$ cells, the second order terms definitely should be included.

8 Variances of totals

The cells in a simulation are not independent. Hence the variance of the total of a cell statistic over a simulation is not simply the sum of cell variances. This is of particular significance in estimating the accuracy of results of simulation. It will be seen that a simulation using a definite number of cells gives much more accurate results than one using a Poisson process.

For generality, the results of this section will be given for covariances. Let U and V be two qualifying random variables. Let q_1 and q_2 be the respective powers, and let $q = q_1 + q_2$. Then

$$\mathrm{Cov}(U, V) = E(U \cdot V) - E(U) \cdot E(V). \tag{8.1}$$

The formulas of the previous sections apply to $E(U)$ and $E(V)$, but not to $E(U \cdot V)$, which involves nonlocal correlations.

The starting point is the pure Poisson process of density D. Because nonneighboring parts of the domain are independent for a Poisson process, the net covariance must be the sum of local correlations, which means it is proportional to a power of the point density. Dimensional analysis gives the power as $q - 1$, so we may write

$$\mathrm{Cov}_P(U, V)(D) = C_{uv} D^{q-1} \tag{8.2}$$

for a constant C_{uv}. Thus we have

$$E_P(U \cdot V|D) = C_U D^{q_1} C_v D^{q_2} + C_{uv} D^{q-1}. \tag{8.3}$$

For fixed N random points, the arguments of section 4 can be applied to get

$$C_U C_v D^q + C_{uv} D^{q-1} = E_P(U \cdot V|D) = \sum_{N=0}^{\infty} E_F(U \cdot V|N) \frac{D^N}{N!} e^{-D}. \tag{8.4}$$

This is solved by

$$E_F(U \cdot V|N) = C_U C_v \frac{N!}{(N-q)!} + C_{uv} \frac{N!}{(N-q+1)!}. \tag{8.5}$$

Thus

$$\mathrm{Cov}_F(U, V|N) = C_U C_v \frac{N!}{(N-q)!} - C_U \frac{N!}{(N-q_1)!} . C_v \frac{N!}{(N-q_2)!} + C_{uv} \frac{N!}{(N-q+1)!}. \tag{8.6}$$

The asymptotic expansion is

$$\begin{aligned} \mathrm{Cov}_F(U, V|N) \quad = \quad & \mathrm{Cov}_F(U, V|N) \left[1 - \frac{(q-1)(q-2)}{2N} + \cdots \right] \\ & - E_P(U|N) E_P(V|N) q_1 q_2 \left[\frac{1}{N} - \frac{(q-1)^2 - q_1 q_2}{2N^2} + \cdots \right]. \end{aligned} \tag{8.7}$$

Note that variances will always be less than for the Poisson process. The effect can be quite dramatic. For example, consider the variance of the total edge length. By [7],

$$\mathrm{Var}_F(\Sigma L) = 1.0445\ldots, \tag{8.8}$$

independent of N. With $q_1 = q_2 = 1/2$ and $E_P(\Sigma L|N) = 2\sqrt{N}$, we get

$$\mathrm{Var}_F(\Sigma L|N) = 1.0445\ldots - (4N)(0.25)\left[\frac{1}{N} - \frac{1}{8N^2} + \cdots\right] \tag{8.9}$$

$$= 0.0445\ldots - \frac{1}{8N} + \cdots$$

If the $3N$ edges in N cells were independent, the variance would be

$$\mathrm{Var}(\Sigma L|N) = 3N \cdot \mathrm{Var}(L) = 0.55688\ldots, \tag{8.10}$$

with the numerical value from [7].

Covariances with a lattice of points require a bit more work since the arguments of sections 5 - 7 are not valid for $E(U \cdot V)$. First consider one fixed point with N random points. The statistics must be the same as for $N+1$ random points, so

$$E_{FL}(U \cdot V|N, 1) = E_F(U \cdot V|N+1). \tag{8.11}$$

The expectation for one lattice point plus a Poisson process of density D is a combination:

$$\begin{aligned} E_P(U \cdot V|D+1) &= \sum_{N=0}^{\infty} E_F(U \cdot V|N+1)\frac{D^N}{N!}e^{-D} \\ &= C_U C_v[D^q + qD^{q-1}] + C_{uv}[D^{q-1} + (q-1)D^{q-2}]. \end{aligned} \tag{8.12}$$

The amount of expectation added by adding one lattice point can be found by comparing (8.11) with (8.3). Since the lattice points are separated, the local changes made by adding each lattice point are independent. This can be used to find the net effect on $E(U \cdot V)$ of adding M lattice points.

The effect of adding one point is to change the configuration locally. Let U be U and V be V for the original N point configuration. Let Δu_i and Δv_i be the changes in U and V resulting from adding point i of the M lattice points. Then

$$\begin{aligned} E(\Delta u_i) &= E_P(U|N+1) - E_P(U|N) \\ E(\Delta v_i) &= E_P(V|N+1) - E_P(V|N) \end{aligned} \tag{8.13}$$

If one adds M lattice points to N random points, then

$$\begin{aligned} E_{PL}(U \cdot V|N, M) &= E((U + \sum_i \Delta u_i)(V + \sum_j \Delta v_j)) \\ &= E(UV) + \sum_i E(\Delta U_i) + \sum_j E(\Delta v_j) + \sum_i E(\Delta u_i \Delta v_i) \\ &\quad + \sum_i \sum_{j \neq i} E(\Delta \mathrm{u}_i \Delta v_j) E(\Delta \mathrm{u}_i \Delta v_j). \end{aligned} \tag{8.14}$$

For different lattice points, the local effects are independent, so for $i \neq j$,

$$E(\Delta u_i.\Delta v_j) = E(\Delta u_i)E(\Delta v_j). \tag{8.15}$$

Hence

$$E_{PL}(U \cdot V|N,M) = \tag{8.16}$$
$$E(UV) + \sum_i (E((U+\Delta u_i)(V+\Delta v_i)) - E(UV)) + M(M-1)E(\Delta u_i)E(\Delta v_i).$$

From this it follows from (8.12), (8.13), and (8.15) that

$$\mathrm{Cov}_{PL}(U,V|D,M) = C_{uv}[D^{q-1} + M(q-1)D^{q-2}] - MC_U C_V q_1 q_2 D^{q-2}. \tag{8.17}$$

Expanding in asymptotic series in terms of statistics for Poisson process of density $D+M$,

$$Cov_{PL}(U,V|D,M) = Cov_P(U,V|D+M)\left[1 - \frac{(q-1)(q-2f)^{fl}}{2(M+D)^2} + \cdots\right]$$
$$-Mq_1q_2E_P(U|D+M)E_P(V|D+M)\left[\frac{1}{(D+M)^2} + \frac{M(q-2)}{(D+M)^3} + \cdots\right]. \tag{8.18}$$

Lastly, the case of M lattice points and N random points may be derived by writing the Poisson version as a combination of the definite N version:

$$E_{PL}(V \cdot V|D,M) = \sum_{N=0}^{\infty} E_{FL}(U \cdot V|N,M)\frac{D^N}{N!}e^{-D}. \tag{8.19}$$

This implies

$$\begin{aligned} E_{FL}(U \cdot V|N,M) &= C_U C_V \frac{N!}{(N-q)!} + C_{uv}\frac{N!}{(N-q+1)!} \\ &\quad + M\left[C_U C_V q\frac{N!}{(N-q+1)!} + C_{uv}(q-1)\frac{N!}{(N-q-2)!}\right] \\ &\quad + M(M-1)C_U C_V q_1 q_2 \frac{N!}{(N-q-2)!}. \end{aligned} \tag{8.20}$$

The covariance is

$$\begin{aligned} Cov_{FL}(U,V|N,M) &= E_{FL}(U \cdot V|N,M) - E_{FL}(U|N,M) \cdot E_{FL}(V|N,M) \\ &= E_{FL}(U \cdot V|N,M) \\ &\quad - C_U\frac{N!}{(N-q_1)!}\left[1 + \frac{q_1 M}{N-q_1+1}\right] C_V\frac{N!}{(N-q_2)!}\left[1 + \frac{q_2 M}{N-q_2+1}\right]. \end{aligned} \tag{8.21}$$

The asymptotic expansion is

$$\mathrm{Cov}_{FL}(U,V|N,M) = \mathrm{Cov}_P(U,V|N+M)\left[1 - \frac{(q-1)(q-2)}{2(N+m)} + \cdots\right]$$
$$-q_1q_2E_P(U|N+M) \cdot E_P(V|N+M)\left[\frac{1}{N+M} - \frac{(q-1)^2 - q_1q_2}{(N+M)^2} + \cdots\right] \tag{8.22}$$

Note that to this order of expansion, the result depends only on the total point density.

9 Conclusion

The statistical properties of a random Voronoi tessellation depend on the exact nature of the point generating process. For generating processes that are almost Poisson, the equations derived above give asymptotic series accurate through second order terms. The statistics to which they apply directly are densities of cell statistics. The power of today's computers is making it necessary to take these slight differences into account.

References:

[1] J. L. Meijering, "Interface area, edge length, and number of vertices in crystal aggregates with random nucleation," *Phillips Res. Rep.* **8** (1953), 270–290.

[2] E. N. Gilbert, "Random subdivisions of space into crystals." *Annals of Mathematical Statistics* **33** (1962), 958–972.

[3] D. Rhynsburger, "Analytic delineation of Theissen polygons." *Geographical Analysis* **5** (1973), 133–144.

[4] R. Sibson, "The Dirichlet tessellation as an aid in data analysis." *Scandinavian J. Statistics* **7** (1980), 14–20.

[5] C. L. Lawson, "Software for C^1 surface interpolation." In *Mathematical Software, III.* Academic Press, New York, 1977, pp. 161–193.

[6] A. L. Hinde and R. E. Miles, "Monte Carlo estimates of the distributions of the random polygons of the Voronoi tessellation with respect to a Poisson process." *Journal of Statistical Computation and Simulation* **10** (1980), 205–223.

[7] K. A. Brakke, "Statistics of random plane Voronoi tessellations." (unpublished)