Statistics of Random Plane Voronoi Tessellations

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

about 1986; this version March 11, 2005

Abstract

If S is a discrete set of points in a space, and each point of the space is associated with the nearest point of S, then the resulting partition is called a *Voronoi tessellation*. This paper derives a general scheme for setting up integrals for statistics for tessellations generated from a Poisson point process. For the case of the plane, the integrals are evaluated to find the variances of cell area, edge length, perimeter, and number of sides. The distributions of several parameters, including edge length, are also found.

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

AMS Classification: 60. Probability theory and stochastic processes.

0 Introduction

If S is a set of points in a Euclidean space R^n, and each point of the space is associated with the nearest point of S, then the space is divided 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. This paper sets up a general scheme for calculating statistics of random Voronoi tessellations for sets S generated by a Poisson point process of unit density. This scheme is then applied to the particular case of the plane. Succeeding papers will deal with higher dimensions. Even for the plane, the previously known results are very few. Clearly, the mean cell area is 1. With probability 1, each vertex will have valence 3, so it follows from Euler's formula that the mean number of sides of a cell is 6. Meijering [1] derived the mean cell perimeter to be 4, whence the mean edge length is 2/3. Gilbert [2] expressed the mean square of cell area as a double integral. This paper extends Meijering's and Gilbert's methods to calculate other second order statistics and the distributions of several quantities, including edge lengths.

1 Tessellation geometry

Figure 1 shows an example of a random Voronoi tessellation. The points of the plane are of three types, depending on how many nearest neighbors in S they have. A point with exactly one nearest neighbor is in the interior of a cell, a point with two nearest neighbors is on the boundary between

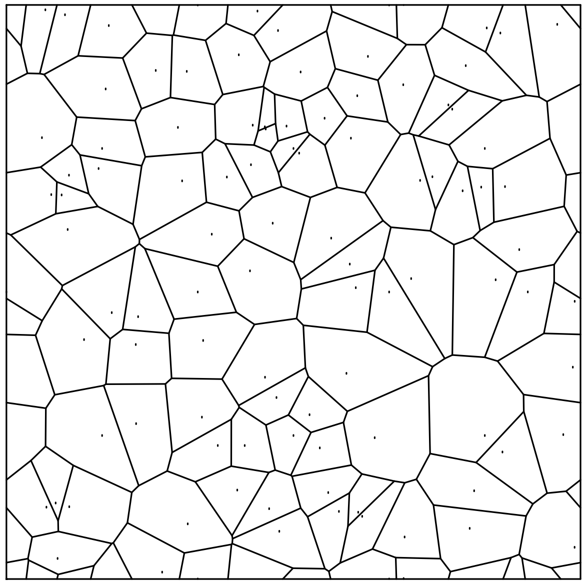

Figure 1: A 100-cell random Voronoi tessellation of a square torus.

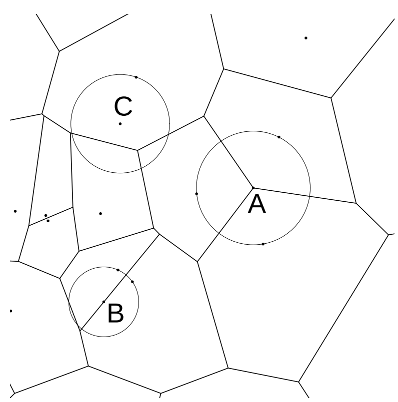

Figure 2: Types of points and their neighbor seeds and voids. Point A is a vertex with three neighbor seeds, point B is an edge point with two neighbor seeds, and point C is an interior point with one neighbor seed.

cells, and a point with three nearest neighbors is a vertex where three cells meet. Figure 2 illustrates a point of each type. There is zero probability that there will be any point with four or more nearest neighbors, so quadruple vertices do not occur in random tessellations. Any seeming quadruple vertices in diagrams are really close pairs of triple vertices. In R^n there will be $n+1$ types of points. Henceforth, the members of S will be caued *seeds* (due to their role in generating cells), and *point* will refer to a general point in the space, usually specified to be one of the types mentioned above. The open ball whose center is at a point and which has the nearest neighbor seeds of that point on its circumference will be called the *void* of the point. The central object of this paper is a configuration of seeds and points. The definition of a configuration type consists of

1. the number m of seeds involved,
2. the number k of points and their types, and
3. a specification of which seeds $S_0, \ldots, S_{m-1}$ are nearest neighbors of which points $P_1, \ldots, P_k$.

Usually there will be one or two points and their nearest seeds. An example configuration would be a vertex and its three neighbor seeds. Note that an actual instance of a configuration in a tessellation may have several possible labellings. For example, a plane vertex configuration has six possible labellings of its seeds. A *complete configuration* is a configuration that includes all the neighbor seeds of its points. Examples would be a vertex and its three neighbor seeds, or an edge point with its two neighbor seeds. An example of an incomplete configuration would be a vertex and one neighbor seed. For a complete configuration, item 3 above can be rephrased in two parts:

3a. a specification of the geometrical relation of points to seeds (i.e., on perpendicular bisector, at circumcenter), and

3b. a requirement that the voids of the points (as defined by 3a) are empty of seeds.

A set of m seeds and k (untyped) points that satisfies the geometrical relationships 3a will be called a *potential configuration*. If the condition 3b is also true, it is an *actual configuration*, and the points are necessarily of the requisite types. As an example, consider the configuration of a vertex and its three neighbor seeds. A potential configuration would be any three seeds and their circumcenter. This is an actual configuration if the void of the configuration is empty of other seeds. The method of this paper may be outlined as follows:

1. State the problem in terms of an incomplete configuration.
2. Embed the incomplete configuration in a complete configuration.
3. In the space of all potential configurations, write down the expectation measure for all potential configurations defined by the Poisson process generating the seeds.
4. Multiply by the probability the void region is empty to get the expectation measure for actual configurations.
5. Integrate over some variables to find the induced measure on the space of incomplete configurations.
6. Solve the original problem.

2 Configuration spaces and measures

Introduce a canonical parameter space W for all complete configurations of a given type as follows. Let the seeds have locations $S_0, \ldots, S_{m-1}$. The seed coordinates thus form a space $W_S = (\mathrm{R}^n)^m$. A point P_i of the k-dimensional skeleton of the tessellation is on the k-plane through the circumcenter of its neighbor seeds and perpendicular to the $(n-k)$-plane those seeds determine. Let y_i represent the coordinates of P_i in this plane. These point parameters y_i form a space $W_P = \mathrm{R}^\mathrm{q}$ for some q. Then $W = W_S \times W_P$. A tessellation T generates in W a set W_T consisting of all the instances of the configuration occurring in T. Different labelings are considered different instances. Let $d\mu_T$ be Hausdorff measure of dimension d restricted to W_T. The configuration measure $d\mu$ will be the expectation of $d\mu_T$ under the probability measure on the space of tessellations T defined by the Poisson process generating S. For potential configurations, $d\mu_T$ is the product of a sum of unit point measures on W_S (one for each ordered subset of m seeds of S) and Lebesgue measure on W_P. By the unit density of the Poisson process, the expectation of the sum of the unit point measures is Lebesgue measure on W_S. Hence the potential configuration measure $d\mu_{pot}$ is Lebesgue measure on W,

$$d\mu_{pot} = dS_0 \ldots dS_{m-1} dy_1 \ldots dy_k. \tag{2.1}$$

The actual configuration measure $d\mu$ will be nonzero only on the subdomain W_0 of W for which none of the seeds of the configuration are in any of the voids. In W_0, the probability that the voids will be empty of other seeds of S is the Poisson factor e^{-A}, where A is the n-dimensional measure of the union of the voids. Hence

$$d\mu = e^{-A} dS_0 \ldots dS_{m-1} dy_1 \ldots dy_k \text{ restricted to } W_0. \tag{2.2}$$

It will usually be convenient to change to coordinates relative to S_0 for $S_1, \ldots, S_{m-1}$. Note the Jacobian of this transformation is 1. Also, results will often be in terms of expected value per cell. For this, we may assume $S_0 = 0$ and factor S_0 out of W, leaving parameter space W', and factor off the dS_0 part of $d\mu$, which leaves the expected measure $d\sigma$ for configurations associated with a single cell (that generated by S_0):

$$d\sigma = e^{-A} dS_1 \ldots dS_{m-1} dy_1 \ldots dy_k \text{ restricted to } W_0'. \tag{2.3}$$

3 Single vertex configuration

In this section, we will use the three seed configuration around a vertex P to derive several distributions. The seed S_0 is assumed to be at the origin of coordinates. Let (R_1, θ_1) be the polar coordinates of S_1 and (R_2, θ_2) be the polar coordinates of S_2, as shown in figure 3a. The single cell configuration measure is

$$d\sigma = e^{-A} dS_1 dS_2 = e^{-A} \cdot R_1 dR_1 d\theta_1 \cdot R_2 dR_2 d\theta_2, \tag{3.1}$$

with domain W_0':

$$0 < R_1, \qquad R_2 < \infty, \qquad 0 \leq \theta_1 < 2\pi, \qquad \theta_1 < \theta_2 < \theta_1 + \pi, \tag{3.2}$$

and void area

$$A = \pi(R_1^2 - 2R_1R_2\cos(\theta_2 - \theta_1) + R_2^2)/4\sin^2(\theta_2 - \theta_1). \tag{3.3}$$

Note that this domain counts each vertex in a cell once, with S_1 and S_2 in counterclockwise order. Since there are an average of six vertices per cell,

$$\int_{W_0} d\sigma = 6. \tag{3.4}$$

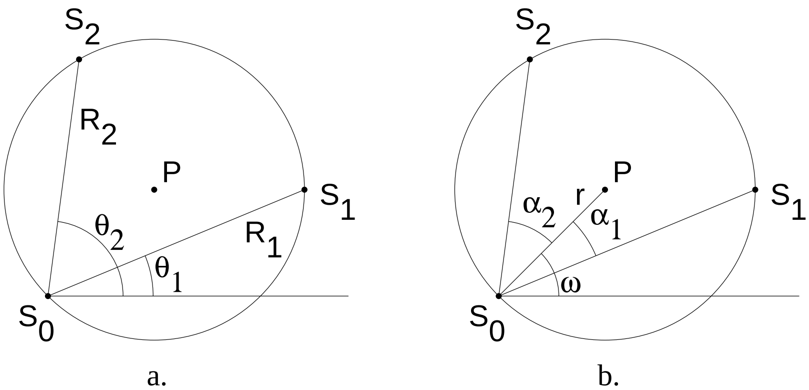

Figure 3: a. Original polar coordinates of S_1 and S_2. b. Coordinate change for integration.

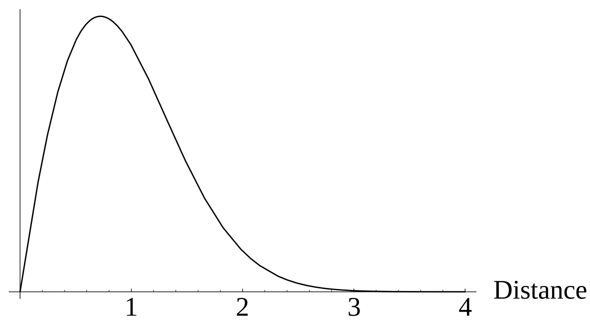

Figure 4: Density function for the distribution of the distance between neighboring seeds, from eq. 3.5.

Hence to get a probability measure, one must divide by 6.

To get the probability density function f_R for the distance R between seeds of neighboring cells, one need only take $R = R_1$ and integrate over all variables except R_1:

$$\begin{aligned} f_R(R) &= \frac{R}{6}\int_0^\infty \int_0^{2\pi} \int_{\theta_1}^{\theta_1+\pi} e^{-A} R_2 d\theta_2 d\theta_1 dR_2 \\ &= \frac{\pi R}{3}(\mathrm{erfc}(\sqrt{\pi}R/2) + Re^{-\pi R^2/4}), \quad 0 < R < \infty, \end{aligned} \tag{3.5}$$

where erfc() is the complementary error function. This distribution is plotted in figure 4.

The distribution f_θ of the angle $\theta = \theta_2 - \theta_1$ between seeds of adjacent neighbors can likewise be derived, and it turns out to be

$$f_\theta(\theta) = \frac{4}{3\pi}((\pi - \theta)\cos\theta + \sin\theta)\sin\theta, \qquad 0 < \theta < \pi. \tag{3.6}$$

This distribution is plotted in figure 5.

To get the distribution f_r of the distance r from seed to vertex, it is convenient to make a change of variables before integrating. Replace $(R_1, \theta_1, R_2, \theta_2)$ by $(r, \omega, \alpha_1, \alpha_2)$, where (r, ω) are the polar

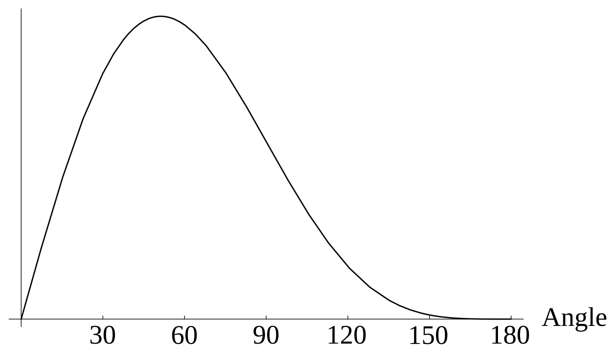

Figure 5: Density function for the distribution of the angle between adjacent neighbor seeds of a cell, subtended from the cell's seed, from eq. 3.6.

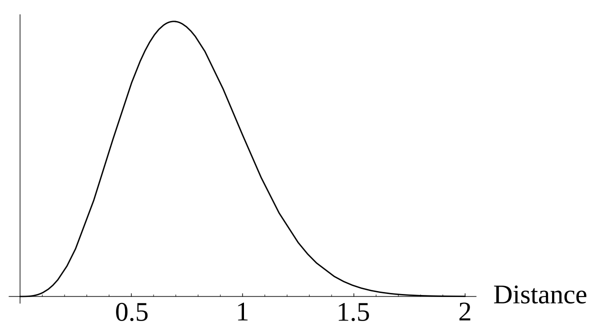

Figure 6: Density function for the distribution of distances between seeds and vertices, i.e. the vertex void radii, from eq. 3.9.

coordinates of the vertex P and α_1, α_2 are the angles from P to S_1 and S_2 respectively, as shown in figure 3b. Then

$$\begin{aligned} \theta_1 &= \omega + \alpha_1, \qquad R_1 = 2r\cos\alpha_1, \\ \theta_2 &= \omega + \alpha_2, \qquad R_2 = 2r\cos\alpha_2. \end{aligned} \tag{3.7}$$

Inserting the Jacobian, the configuration measure becomes

$$d\sigma = 16e^{-4\pi r^3/3} r^3 \cos\alpha_1 \cos\alpha_2 \sin(\alpha_2 - \alpha_1) d\alpha_1 d\alpha_2 dr d\omega,$$

$$0 < r < \infty, \qquad 0 \le \omega < 2\pi, \qquad -\pi/2 < \alpha_1 < \alpha_2 < \pi/2. \tag{3.8}$$

Integrating over all variables but r and normalizing gives

$$f_r(r) = 2\pi^2 r^3 e^{-\pi r^2}, \qquad 0 < r < \infty. \tag{3.9}$$

This distribution is plotted in figure 6.

The above probability densities may be multiplied by appropriate factors to give absolute density functions. For example, if one wanted the expected absolute radial density of neighbor seeds around a given seed, one would multiply (3.5) by 6, which is the expected number of neighbor seeds.

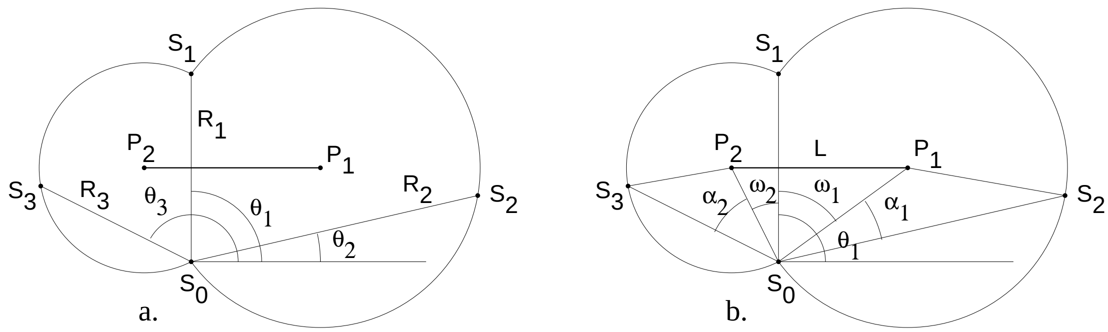

Figure 7: Edge configuration. a. Original polar coordinates of S_1, S_2, and S_3. b. Coordinate change for integration.

4 Edge length distribution

In this section, we will use the three seed configuration around a vertex P to derive several distributions. The appropriate configuration to use to find out facts about a single edge is the configuration of four seeds S_0, S_1, S_2, and S_3 such that S_0, S_1, and S_2 determine the vertex P_1 at one end of the edge and S_0, S_1, and S_3 determine the vertex P_2 at the other end. In polar coordinates for S_1, S_2, and S_3 relative to S_0, as shown in figure 7a, the configuration measure is

$$d\sigma = e^{-A} dS_1 dS_2 dS_3 = e^{-A} R_1 dR_1 d\theta_1 \cdot R_2 dR_2 d\theta_2 \cdot R_3 dR_3 d\theta_3, \tag{4.1}$$

where A is the area of the union of the void circles of the two vertices and the domain is

$$0 < R_1, R_2, R_3 < \infty, \quad 0 \le \theta_1 < 2\pi, \quad \theta_1 < \theta_2 < \theta_1 + \pi, \quad \theta_2 < \theta_3 < \theta_2 + \pi, \tag{4.2}$$

but excluding configurations wherein S_2 is in the void circle of P_2 and S_3 is in the void circle of P_1. This counts an average of six edges per cell, so the probability measure is $d\sigma/6$. For convenience, a change of variables is made to $(L,\, \theta_1, \omega_1, \omega_2,\, \alpha_1, \alpha_2)$ as shown in figure 7b, where

- L is the edge length from P_1 to P_2,
- θ_1 is the same θ_1 ,
- ω_1 is the angle $S_1 S_0 P_1$, positive clockwise,
- ω_2 is the angle $S_1 S_0 P_2$, positive counterclockwise,
- α_1 is the angle $P_1 S_0 S_2$, positive clockwise,
- α_2 is the angle $P_2 S_0 S_3$, positive counterclockwise.

Therefore

$$\begin{aligned} d\sigma = 64 e^{-A} \sec^2\omega_1 (\sin\alpha_1 + \cos\alpha_1 \tan\omega_1) \sec^2\omega_2 (\sin\alpha_2 + \cos\alpha_2 \tan\omega_2) \\ \times (\tan\omega_1 + \tan\omega_2)^{-6} \cos\alpha_1 \cos\alpha_2 d\alpha_1 d\alpha_2 d\theta_1 d\omega_1 d\omega_2 dL, \end{aligned} \tag{4.3}$$

with domain

$$0 < L < \infty, \quad 0 \le \theta_1 < 2\pi, \quad -\frac{\pi}{2} < \omega_1 < \frac{\pi}{2}, \quad -\omega_1 < \omega_2 < \frac{\pi}{2},$$

$$-\omega_1 < \alpha_1 < \frac{\pi}{2}, \quad -\omega_2 < \alpha_2 < \frac{\pi}{2}, \quad -\omega_2 < \alpha_2 < \frac{\pi}{2}. \tag{4.4}$$

Integrating over θ_1, α_1, and α_2 analytically and normalizing gives the probability density function f_L for edge length:

$$f_L(L) = \frac{16\pi}{3} L^5 \int_{-\pi/2}^{\pi/2} \int_{-w_1}^{\pi/2} e^{-L^2 B/(\tan\omega_1 + \tan\omega_2)^2} (\tan\omega_1 + \tan\omega_2)^{-6}$$
$$\times \left(\left(\frac{\pi}{2} + \omega_1 \right) \tan\omega_1 + 1 \right) \left(\left(\frac{\pi}{2} + \omega_2 \right) \tan\omega_2 + 1 \right) \sec^2\omega_1 \sec^2\omega_2 d\omega_2 d\omega_1, \tag{4.5}$$

where

$$B = \left(\frac{\pi}{2} + \omega_1 + \sin\omega_1 \cos\omega_1 \right) \sec^2\omega_1 + \left(\frac{\pi}{2} + \omega_2 + \sin\omega_2 \cos\omega_2 \right) \sec^2\omega_2. \tag{4.6}$$

The results of numerical integration are given in Table 1 and plotted in figure 8. A less singular formulation is

$$f_L(L) = \frac{16\pi}{3} L^5 \int_{-\pi/2}^{\pi/2} \int_{-w_1}^{\pi/2} e^{-L^2 \hat{B}/\sin^2(\omega_1+\omega_2)} \sin^{-6}(\omega_1 + \omega_2) \tag{4.7}$$
$$\times \left(\left(\frac{\pi}{2} + \omega_1 \right) \sin\omega_1 + \cos\omega_1 \right) \left(\left(\frac{\pi}{2} + \omega_2 \right) \sin\omega_2 + \cos\omega_2 \right) \cos^3\omega_1 \cos^3\omega_2 d\omega_2 d\omega_1,$$

where

$$\hat{B} = \left(\frac{\pi}{2} + \omega_1 + \sin\omega_1 \cos\omega_1 \right) \cos^2\omega_2 + \left(\frac{\pi}{2} + \omega_2 + \sin\omega_2 \cos\omega_2 \right) \cos^2\omega_1. \tag{4.8}$$

Exactly,

$$f_L(0) = \frac{2}{\pi}. \tag{4.9}$$

Asymptotically, for large L,

$$f_L(L) \approx \pi^2 L^2 e^{-\pi L^2/2} / 3\sqrt{2}. \tag{4.10}$$

The moments of L are

$$E[L^n] = \frac{8\pi}{3} \int_{-\pi/2}^{\pi/2} \int_{-w_1}^{\pi/2} \hat{B}^{-(n+6)/2} \sin^n(\omega_1 + \omega_2) \tag{4.11}$$
$$\times \left(\left(\frac{\pi}{2} + \omega_1 \right) \sin\omega_1 + \cos\omega_1 \right) \left(\left(\frac{\pi}{2} + \omega_2 \right) \sin\omega_2 + \cos\omega_2 \right) \cos^3\omega_1 \cos^3\omega_2 d\omega_2 d\omega_1.$$

From the same starting measure (4.3), one can also derive the distribution f_ω of the angle ω between adjacent vertices as seen from S_0. Let $\omega = \omega_1 + \omega_2$ be the angle in question. Then

$$f_w(\omega) = \frac{16\pi}{3} \int_{\omega-\pi/2}^{\pi/2} B^{-3} \left(\left(\frac{\pi}{2} + \omega_1 \right) \tan\omega_1 + 1 \right) \left(\left(\frac{\pi}{2} + \omega_2 \right) \tan\omega_2 + 1 \right) \sec^2\omega_1 \sec^2\omega_2 d\omega_2,$$
$$0 < \omega < \pi. \tag{4.12}$$

A more non-singular version is

$$f_w(\omega) = \frac{16\pi}{3} \int_{\omega-\pi/2}^{\pi/2} \hat{B}^{-3} \left(\left(\frac{\pi}{2} + \omega_1 \right) \sin\omega_1 + \cos\omega_1 \right) \left(\left(\frac{\pi}{2} + \omega_2 \right) \sin\omega_2 + \cos\omega_2 \right) \cos^3\omega_1 \cos^3\omega_2 d\omega_2. \tag{4.13}$$

The limiting values are

$$f_w(0) = \frac{20}{3\pi} - \frac{4\pi}{9}, \tag{4.14}$$

and

$$f_\omega(\omega) \approx \left(\frac{\pi}{2} - \frac{4}{3} \right) (\pi - \omega) \text{ as } \omega \to \pi. \tag{4.15}$$

The results of numerical integration are given in Table 2 and plotted in figure 9.

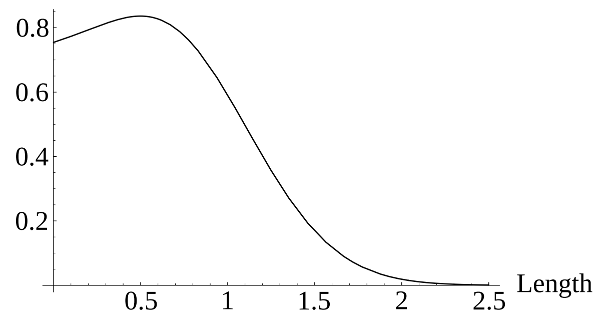

Figure 8: Density function for the distribution of edge lengths, from eq. 4.5.

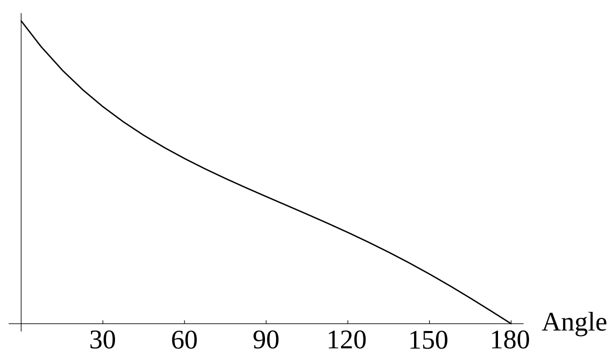

Figure 9: Density function for the distribution of angles between adjacent vertices, as subtended from seed, from eq. 4.9.

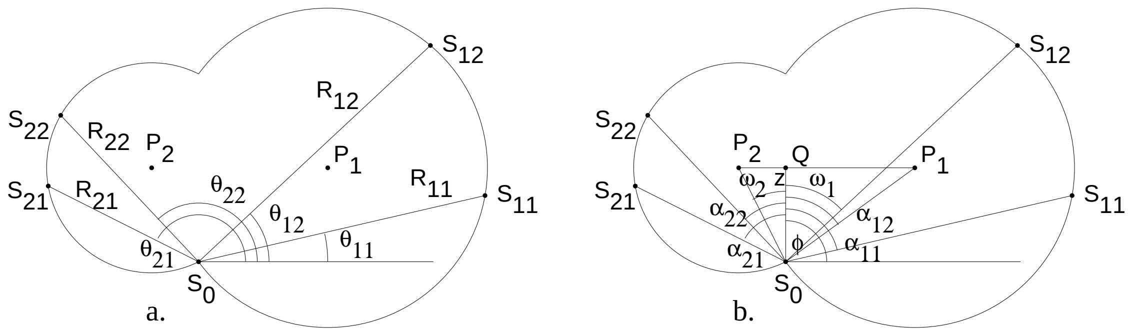

Figure 10: General point pair configuration. This shows two vertices P_1 and P_2 not on the same edge. a. Original polar coordinates. b. Variables for integration.

5 Second order statistics

Second order statistics, such as the expected square of cell area, can be calculated by finding the measure of pairs of points (P_1, P_2) associated with the cell generated by S_0. There are three types of points, and hence six types of pairs to consider. Each type of pair configuration has as its void region the union of the appropriate void circles of the two points. It will turn out below that it is only necessary to compute for configurations in which the two points have no neighbor seeds in common except S_0. With S_0 fixed, the potential configurations for P_1 and P_2 are independent. Hence, the configuration measure may be written

$$d\sigma = e^{-A} d\sigma_1 d\sigma_2, \tag{5.1}$$

where A is the void area and $d\sigma_i$ is the potential configuration measure for P_i. The configuration coordinates and measures for P_i are (see figure 10):

i. If P_i is an interior point with neighbor S_0:

$$\begin{aligned} (r_i, \theta_i) : &\ \text{polar coordinates of } P_i, \\ d\sigma_i &= r_i dr_i d\theta_i. \end{aligned} \tag{5.2}$$

ii. If P_i is a boundary point with neighbors S_0 and S_i:

$$\begin{aligned} (R_i, \theta_i) : &\quad \text{polar coordinates of } S_i, \\ y_i : &\quad \text{distance from } S_0 S_i \text{ midpoint to } P_i, \\ d\sigma_i &= R_i dR_i d\beta_i dy_i. \end{aligned} \tag{5.3}$$

iii. If P_i is a vertex with neighbors S_0, S_{i1}, and S_{i2},

$$\begin{aligned} (R_{i1}, \theta_{i1}) : &\quad \text{polar coordinates of } S_{i1}, \\ (R_{i2}, \theta_{i2}) : &\quad \text{polar coordinates of } S_{i2}, \\ d\sigma_i &= R_{i1} dR_{i1} d\theta_{i1} R_{i2} dR_{i2} d\theta_{i2}. \end{aligned} \tag{5.4}$$

To set up useful coordinates, let P_1 and P_2 be taken in counterclockwise order around S_0. To conform to the natural symmetry of the configuration, angles measured from P_1 will be positive clockwise and angles measured from P_2 will be positive counterclockwise. See figure 10b. All types of configurations will use these parameters:

(z, ϕ): polar coordinates of the point Q on the segment P_1P_2 that is closest to S_0, $0 \le z < \infty$, $0 \le \phi < 2\pi$,

ω_1: the angle QS_0P_1, $-\pi/2 < \omega_1 < \pi/2$,

ω_2: the angle QS_0P_2, $-\omega_1 < \omega_2 < \pi/2$.

In addition, if P_i is a boundary point with neighbor seed S_i, then use

α_i: the angle $P_iS_0S_i$, $-\omega_i < \alpha_i < \pi/2$.

If P_i is a vertex with neighbor seeds S_{i1} and S_{i2}, use

α_{i1}: the angle $P_iS_0S_{i1}$, $-\omega_i < \alpha_{i1} < \pi/2$.

α_{i2}: the angle $P_iS_0S_{i2}$, $\alpha_{i1} < \alpha_{i2} < \pi/2$.

Making the changes of variables, it turns out that the configuration measure can be written

$$d\sigma = e^{-Bz^2} g_1 g_2 \sin(\omega_1 + \omega_2) z dz d\omega_1 d\omega_2 d\phi, \tag{5.5}$$

where B is given by (4.6) and the factor g_i for a point P_i is

$$g_i = \begin{cases} z \sec^3 \omega_i, & P_i \text{ is an interior point;} \\ 4z^2 \sec^4 \omega_i \cos \alpha_{i1} d\alpha_{i1}, & P_i \text{ is a boundary point;} \\ 16z^3 \sec^5 \omega_i \cos \alpha_{i1} \cos \alpha_{i2} \sin(\alpha_{i2} - \alpha_{i1}) d\alpha_{i1} d\alpha_{i2}, & P_i \text{ is a vertex.} \end{cases} \tag{5.6}$$

Let $I(\cdot, \cdot)$ denote the complete integral over the appropriate configuration measure $d\sigma$ for the pair type $(\cdot, \cdot)$, where the pair elements can be a, p, or v, representing interior, boundary, or vertex points respectively. The variables z, ϕ, and all α's can be integrated analytically, leaving

$$I(\cdot,\cdot) = \pi\Gamma\left(\frac{n_1+n_2+2}{2}\right) \int_{-\pi/2}^{\pi/2} \int_{-w_1}^{\pi/2} B^{-(n_1+n_2+2)/2} G_1(\omega_1) G_2(\omega_2) \sin(\omega_1+\omega_2) d\omega_2 d\omega_1 \tag{5.7}$$

where $G_i(w_i)$ and n_i depend on the type of point P_i:

$$G_i(\omega_i) = \begin{cases} \sec^3 \omega_i, & P_i \text{ is an interior point;} \\ 4(1+\sin\omega_i)\sec^4 \omega_i, & P_i \text{ is a boundary point;} \\ 2((\pi/2+\omega_i)(1+2\sin^2\omega_i) + 3\sin\omega_i\cos\omega_i))\sec^5\omega_i, & P_i \text{ is a vertex.} \end{cases}$$

$$n_i = \begin{cases} 1, & P_i \text{ is an interior point;} \\ 2, & P_i \text{ is a boundary point;} \\ 3, & P_i \text{ is a vertex.} \end{cases} \tag{5.8}$$

In a less singular way,

$$\begin{aligned} I(\cdot,\cdot) \quad = \quad & \pi\Gamma\left(\frac{n_1+n_2+2}{2}\right) \int_{-\pi/2}^{\pi/2} \int_{-w_1}^{\pi/2} \hat{B}^{-(n_1+n_2+2)/2} \\ & \hat{G}_1(\omega_1)\hat{G}_2(\omega_2) \sin(\omega_1+\omega_2) \cos^{n_2}\omega_1 \cos^{n_1}\omega_2 d\omega_2 d\omega_1 \end{aligned} \tag{5.9}$$

where

$$\hat{G}_i(\omega_i) = \begin{cases} 1, & P_i \text{ is an interior point;} \\ 4(1+\sin\omega_i), & P_i \text{ is a boundary point;} \\ 2((\pi/2+\omega_i)(1+2\sin^2\omega_i) + 3\sin\omega_i\cos\omega_i)), & P_i \text{ is a vertex.} \end{cases} \tag{5.10}$$

$I(\cdot,\,\cdot)$ finds the expected measure of pairs of points of the types under consideration, with the second counterclockwise from the first by less than π. In the expectations and variances below, a is the area, p is the perimeter, and v is the number of vertices of a cell. $I(a,\,a)$ counts pairs of interior points in counterclockwise order, and $E[a^2]$ counts each such pair twice, so

$$E[a^2] = 2I(a,a), \qquad \mathrm{Var}[a] = E(a^2) - E(a)^2. \tag{5.11}$$

$I(a,p)$ counts only half the pairs of interior and perimeter points, and $E[ap]$ counts them all, so

$$E[ap] = 2I(a,p), \qquad \mathrm{Cov}[a,p] = E[ap] - E[a]E[p]. \tag{5.12}$$

$I(p,p)$ counts pairs of perimeter points not on the same edge, so

$$E[p^2] = 2I(p,p) + 6E[L^2], \qquad \mathrm{Var}[p] = E[p^2] - E[p]^2. \tag{5.13}$$

$I(a,\,v)$ half counts pairs of interior points and vertices, so

$$E[av] = 2I(a,v), \qquad \mathrm{Cov}[a,v] = E[av] - E[a]E[v]. \tag{5.14}$$

$I(p,v)$ half counts pairs of perimeter points and vertices not on the same edge, so

$$I(p,v) = E[p(v-2)]/2, \qquad E[pv] = 2I(p,v) + 2E[p],$$
$$\mathrm{Cov}[p,v] = E[pv] - E[p]E[v]. \tag{5.15}$$

$I(v,v)$ counts pairs of vertices not on the same edge, so

$$I(v,v) = E[v(v-3)/2], \qquad E[v^2] = 2I(v,v) + 3E[v],$$
$$\mathrm{Var}[v] = E[v^2] - E[v]^2. \tag{5.16}$$

The results of numerical integration are given in table 3.

6 Triangular cells

It is easy enough to write down the configuration measure for a k-sided cell. However, it results in integration over $2k$ variables. But with a little ingenuity, the integrals for triangles are practical. The details are messy and add no insight to what has already been presented, so I will just present the results of numerical integration:

Probability of triangle: 0.01124001348534

Mean triangle area: 0.34308914805

Mean triangle perimeter: 2.74029726648

Other information about cells with a given number of sides is so far known only through computer simulations [6].

7 Cell void distributions

There is one remarkable exception to the last statement of the last section. The *cell void* of a cell is the union of all the voids of the points of the cell, which is the same as the union of the voids of all the vertices. By a simple argument, it is possible to completely specify the distribution of measures of cell voids for k-sided cells in n dimensions. A k-sided cell has k neighbor seeds, so the configuration measure for k-sided cells is

$$d\sigma = e^{-A} R_1^{n-1} dR_1 d\theta_1 \dots R_k dR_k^{n-1} d\theta_k, \tag{7.1}$$

with an appropriate domain. The θ_i represent all angular variables. Make a change of variables to the total cell void measure A and a set of dimensionless parameters Ω. Scaling homogeneity and dimensional analysis shows that the configuration measure takes the form

$$d\sigma = g(\Omega) d\Omega e^{-A} A^{k-1} dA, \tag{7.2}$$

and the domain of Ω does not depend on A. Integrate over Ω to get the distribution function $f_k(A)$ for cell void measures of k-sided cells:

$$f_k(A) = ce^{-A} A^{k-1}. \tag{7.3}$$

Normalizing to total probability one, we get the conditional probability density function:

$$f_k(A) = A^{k-1} e^{-A}/(k-1)!, \tag{7.4}$$

which is a gamma distribution. It is interesting to note that this immediately implies that the mean cell void measure for k-sided cells is exactly k.

8 Variance of regional totals

Distant cells of a tessellation are practically independent of each other since the influence of a local configuration falls off exponentially as the square of the distance. However, nearby cells not independent. This means that the variance of the sum of a statistic in a region is not just the sum of the variances of single cases. This is relevant in estimating the accuracy of the results of simulation. For tractability, the region considered here will be a torus of area N. A torus is a square with its opposite edges identified. It is convenient here because it is a finite, flat region without any boundary. Assume the area is large enough so that wrap-around effects are negligible.

Some totals have easily found variances. For a unit density Poisson process on a torus of area N, the variance in the number of cells is just N, by the properties of the Poisson distribution. The variances of the total numbers of edges and vertices are 9N and 4N respectively, since there are three times as many edges and twice as many vertices as cells. The variance of the total edge length is not as simply found, but it can be reduced to a double integral and evaluated numerically. In what follows, L is the length of a single edge, and ΣL is the total edge length. Since

$$\mathrm{Var}[\Sigma L] = E[(\Sigma L)^2] - E[\Sigma L]^2 \tag{8.1}$$

and $E[\Sigma L] = 2N$, it remains to find $E[(\Sigma L)^2]$. Now $T = (\Sigma L)^2$ is the measure of all ordered pairs (P_1, P_2) of edge points, which can be divided into three domains:

I: pairs on the same edge,

II: pairs on different edges of the same cell, and

III: pairs not on the same cell.

The expected totals of I and II are already known:

$$E[T_I] = 3NE[L^2], \qquad E[T_{II}] = N(E[p^2] - 6E[L^2]). \tag{8.2}$$

Domain III pairs correspond to configurations of four seeds S_{11}, S_{12}, S_{21} and S_{22}, and two edge parameters y_1 and y_2, with (S_{11}, S_{12}, y_1) locating P_1 and (S_{21}, S_{22}, y_2) locating P_2. The configuration measure is

$$d\mu = \frac{1}{4}e^{-A} dS_1 dS_2 dy_1 dS_{21} dS_{22} dy_2, \tag{8.3}$$

where A is the area of the void region, which is the union of the void circles around the two edge points. The factor of $1/4$ is necessary to correct for multiple counting of the same pair due to seed labellings. Domain III can be broken up into two pieces:

IIIA: Non-overlapping voids, and

IIIB: Overlapping voids.

On IIIA, we have $A = A_1 + A_2$, where A_1 and A_2 are the areas of the void circles. Here, the configuration measure factors:

$$d\mu_{IIIA} = \frac{1}{2}e^{-A_1} dS_{11} dS_{12} dy_1 \cdot \frac{1}{2}e^{-Az} dS_{21} dS_{22} dy_2, \tag{8.4}$$

On domain IIIA, the integral of (8.4) is the product of integrals for total edge, with result $2N \cdot 2N$. From this must be subtracted the integral of (8.4) over domain IIIB. For this, change to variables (as shown in figure 11)

(x_1, x_2) :	location of first edge point,
(R, θ) :	polar coordinates of second edge point relative to the first,
r_1, r_2 :	radii of the two void circles,
$\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}$:	position angles of the seeds on the void circles.

Then (8.4) becomes

$$d\mu_{IIIB} = 4e^{-\pi r_1^2 - \pi r_2^2} r_1^2 r_2^2 R \sin\frac{\alpha_{12} - \alpha_{11}}{2} \sin\frac{\alpha_{22} - \alpha_{21}}{2} d\alpha_{11} d\alpha_{12} d\alpha_{21} d\alpha_{22} dr_1 dr_2 dR d\theta dx_1 dx_2, \tag{8.5}$$

with domain IIIB becoming

$$\begin{gathered} 0 < \alpha_{11} < \alpha_{12} < 2\pi, \qquad 0 < \alpha_{21} < \alpha_{22} < 2\pi, \\ 0 < r_1, r_2 < \infty, \qquad 0 < R < r_1 + r_2, \qquad 0 < \theta < 2\pi, \\ (x_1, x_2) \text{ in torus.} \end{gathered} \tag{8.6}$$

The range on r_1 and r_2 is not exact, but it is necessary for tractibility and the error is negligible for large N. The value of the integral is $(12 - 32/\pi)N$, so

$$E[T_{IIIA}] = 4N^2 - (12 - 32/\pi)N. \tag{8.7}$$

Finally, the integral of (8.3) over IIIB can be put into a form similar to integrals of previous sections by making the further change of variables from $(r_1,\, r_2,\, R)$ to $(\omega_1, \omega_2,\, z)$ with

$$\begin{aligned} r_1 &= z \sec\omega_1, \\ r_2 &= z \sec\omega_2, \\ R &= z(\tan\omega_1 + \tan\omega_2). \end{aligned} \tag{8.8}$$

The configuration measure becomes

$$d\mu_{IIIB} = 4e^{-Bz^2} z^7 \sec^5\omega_1 \sec^5\omega_2 \sin\frac{\alpha_{12}-\alpha_{11}}{2} \sin\frac{\alpha_{22}-\alpha_{21}}{2} \sin^2(\omega_1+\omega_2)$$
$$\times d\alpha_{11} d\alpha_{12} d\alpha_{21} d\alpha_{22} d\omega_1 d\omega_2 dz d\theta dx_1 dx_2, \tag{8.9}$$

where B is as before. Domain IIIB excluding configurations with seeds in voids is

$$\begin{gathered} 0 < z < \infty, \qquad -\frac{\pi}{2} < \omega_1 < \frac{\pi}{2}, \qquad -\omega_1 < \omega_2 < \frac{\pi}{2}, \\ \frac{\pi}{2} - \omega_1 < \alpha_{11} < \alpha_{12} < \frac{3\pi}{2} + \omega_1, \\ \frac{\pi}{2} - \omega_2 < \alpha_{21} < \alpha_{22} < \frac{3\pi}{2} + \omega_2, \\ 0 < \theta < 2\pi, \qquad (x_1, x_2) \text{ in torus.} \end{gathered} \tag{8.10}$$

Integrating over z, α_{11}, α_{12}, α_{21}, α_{22}, θ, x_1, and x_2 leaves

$$E[T_{IIIB}] = 384N \int_{-\pi/2}^{\pi/2} \int_{-\omega_1}^{\pi/2} B^{-4} F(\omega_1) F(\omega_2) \sin(\omega_1+\omega_2) d\omega_1 d\omega_2, \tag{8.11}$$

where

$$F(\omega) = (\pi/2 + \omega - \cos\omega)\sec^5\omega. \tag{8.12}$$

Numerical integration yields

$$E[TIIIB] = 8.17520721090347 \tag{8.13}$$

The net result is

$$\mathrm{Var}[\Sigma L] = 1.04456853531195N. \tag{8.14}$$

Total perimeter is twice total edge, so the variance of total perimeter is

$$\mathrm{Var}[\Sigma p] = 4.17827414124760N. \tag{8.15}$$

Note the variance is proportional to area, which is to be expected since distant regions are independent, so their variances would add.

9 Conclusion

The use of configuration measures makes it possible to write down integrals for many statistics regarding random Voronoi tessellations. The necessary condition is that the statistic be expressible in terms of the total Hausdorff measure of some type of configuration of points. The practical difficulty comes in evaluating the integrals. Future papers will report on the statistics of higher dimensional tessellations and their cross-sections.

References:

[1] J. L. Meijering, "Interface area, edge length, and number of vertices in crystal aggregates with random nucleation," *Philips 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.

Table 1. Probability density function f_L of edge lengths L of random plane Voronoi tessellations, normalized for unit seed density.

L	f_L	L	f_L	L	f_L
0.00	7.545123228060E-1	1.70	7.823754374633E-2	3.40	3.669523763846E-7
0.05	7.636277464120E-1	1.75	6.306081962957E-2	3.45	2.205022285610E-7
0.10	7.731661096492E-1	1.80	5.036087814589E-2	3.50	1.314099265451E-7
0.15	7.830985778046E-1	1.85	3.985103540144E-2	3.55	7.767123794257E-8
0.20	7.932991012282E-1	1.90	3.124793411031E-2	3.60	4.553170350443E-8
0.25	8.035106314634E-1	1.95	2.428065760305E-2	3.65	2.647242639388E-8
0.30	8.133289320658E-1	2.00	1.869727373767E-2	3.70	1.526525829628E-8
0.35	8.222060443266E-1	2.05	1.426905749239E-2	3.75	8.730705302919E-9
0.40	8.294731115367E-1	2.10	1.079272366266E-2	3.80	4.952600633009E-9
0.45	8.343800232174E-1	2.15	8.091040073822E-3	3.85	2.786506007724E-9
0.50	8.361474851396E-1	2.20	6.012195325818E-3	3.90	1.555096669929E-9
0.55	8.340258609372E-1	2.25	4.428272348244E-3	3.95	8.607087923971E-10
0.60	8.273545865994E-1	2.30	3.233138747860E-3	4.00	4.725338077814E-10
0.65	8.156161501103E-1	2.35	2.340014888064E-3	4.05	2.573157514659E-10
0.70	7.984794798673E-1	2.40	1.678927267877E-3	4.10	1.389824348877E-10
0.75	7.758289450996E-1	2.45	1.194202923612E-3	4.15	7.445890060086E-11
0.80	7.477768377829E-1	2.50	8.421135844139E-4	4.20	3.956757870629E-11
0.85	7.146589591473E-1	2.55	5.887381304827E-4	4.25	2.085601685546E-11
0.90	6.770145682878E-1	2.60	4.080793621783E-4	4.30	1.090423216158E-11
0.95	6.355532970282E-1	2.65	2.804461682718E-4	4.35	5.655006336047E-12
1.00	5.911125800252E-1	2.70	1.910942749742E-4	4.40	2.909033212120E-12
1.05	5.446096396456E-1	2.75	1.291069058891E-4	4.45	1.484375096028E-12
1.10	4.969921081934E-1	2.80	8.648977070415E-5	4.50	7.513126318300E-13
1.15	4.491910219664E-1	2.85	5.745169748060E-5	4.55	3.772090640750E-13
1.20	4.020792737766E-1	2.90	3.784188448041E-5	4.60	1.878582199449E-13
1.25	3.564377728141E-1	2.95	2.471626945074E-5	4.65	9.280408714649E-14
1.30	3.129306453972E-1	3.00	1.600816188742E-5	4.70	4.547732294753E-14
1.35	2.720899186674E-1	3.05	1.028149162106E-5	4.75	2.210627089611E-14
1.40	2.343093427142E-1	3.10	6.548386028079E-6	4.80	1.065934658233E-14
1.45	1.998463804269E-1	3.15	4.136023068908E-6	4.85	5.098489898042E-15
1.50	1.688309573588E-1	3.20	2.590650881905E-6	4.90	2.419073332676E-15
1.55	1.412793203743E-1	3.25	1.609231477082E-6	4.95	1.138551071076E-15
1.60	1.171112882989E-1	3.30	9.913293951915E-7	5.00	5.315542196010E-16
1.65	9.616926059581E-2	3.35	6.056380732764E-7		

Table 2. Probability density function f_ω of vertex-seed-vertex angles of random plane Voronoi tessellation, normalized for unit seed density.

angle	probability	angle	probability	angle	probability
0.00π	$7.25802506296473E-1$	0.34π	3.91943761215642E-1	0.68π	2.11520799538384E-1
0.01π	$7.09744517404633E-1$	0.35π	3.85945429832324E-1	0.69π	2.05992072193657E-1
0.02π	$6.94283748167271E-1$	0.36π	3.80049775560122E-1	0.70π	2.00403340159989E-1
0.03π	$6.79390752015949E-1$	0.37π	3.74205065340687E-1	0.71π	1.94751781701410E-1
0.04π	$6.65037713225037E-1$	0.38π	3.68539745691271E-1	0.72π	1.89034750303980E-1
0.05π	$6.51198337206598E-1$	0.39π	3.62912431826782E-1	0.73π	1.83249786408308E-1
0.06π	$6.37847748950422E-1$	0.40π	3.57361897529207E-1	0.74π	1.77394629837405E-1
0.07π	$6.24962398929240E-1$	0.41π	3.51882065736399E-1	0.75π	1.71467232900285E-1
0.08π	$6.12519975851668E-1$	0.42π	3.46466999825624E-1	0.76π	1.65465774146261E-1
0.09π	$6.00499325701744E-1$	0.43π	3.41110895570550E-1	0.77π	1.59388672738896E-1
0.10π	$5.88880376554612E-1$	0.44π	3.35808073753409E-1	0.78π	1.53234603412076E-1
0.11π	$5.77644068703699E-1$	0.45π	3.30552973416963E-1	0.79π	1.47002511963632E-1
0.12π	$5.66772289676010E-1$	0.46π	3.25340145743655E-1	0.80π	1.40691631234420E-1
0.13π	$5.56247813749588E-1$	0.47π	3.20164248551881E-1	0.81π	1.34301497512720E-1
0.14π	$5.46054245620922E-1$	0.48π	3.15020041401779E-1	0.82π	1.27831967295366E-1
0.15π	$5.36175967900761E-1$	0.49π	3.09902381305225E-1	0.83π	1.21283234328077E-1
0.16π	$5.26598092144542E-1$	0.50π	3.04806219036860E-1	0.84π	1.14655846838216E-1
0.17π	$5.17306413148858E-1$	0.51π	2.99726596045024E-1	0.85π	1.07950724863601E-1
0.18π	$5.08287366268282E-1$	0.52π	2.94658641963323E-1	0.86π	1.01169177571184E-1
0.19π	$4.99527987527738E-1$	0.53π	2.89597572725314E-1	0.87π	9.43129204494244E-2
0.20π	$4.91015876324549E-1$	0.54π	2.84538689286382E-1	0.88π	8.73840922481807E-2
0.21π	$4.82739160531626E-1$	0.55π	2.79477376958316E-1	0.89π	8.03852715300028E-2
0.22π	$4.74686463829037E-1$	0.56π	2.74409105363412E-1	0.90π	7.33194926869179E-2
0.23π	$4.66846875105628E-1$	0.57π	2.69329429015990E-1	0.91π	6.61902612674264E-2
0.24π	$4.59209919785597E-1$	0.58π	2.64233988540212E-1	0.92π	5.90015684494880E-2
0.25π	$4.51765532946953E-1$	0.59π	2.59118512533805E-1	0.93π	5.17579044870414E-2
0.26π	$4.44504034109980E-1$	0.60π	2.53978820087831E-1	0.94π	4.44642709502077E-2
0.27π	$4.37416103583916E-1$	0.61π	2.48810823972991E-1	0.95π	3.71261915729888E-2
0.28π	$4.30492760269500E-1$	0.62π	2.43610534503012E-1	0.96π	2.97497215171788E-2
0.29π	$4.23725340823645E-1$	0.63π	2.38374064085507E-1	0.97π	2.23414548575816E-2
0.30π	$4.17105480100433E-1$	0.64π	2.33097632470244E-1	0.98π	1.49085300916750E-2
0.31π	$4.10625092790036E-1$	0.65π	2.27777572704003E-1	0.99π	7.45863347680132E-3
0.32π	$4.04276356183924E-1$	0.66π	2.22410337800155E-1	1.00π	0.00000000000000E0
0.33π	$3.98051694001064E-1$	0.67π	2.16992508129651E-1		

Table 3. Numerical integration results for second order statistics of random plane Voronoi tessellations, normalized for unit seed density.

$E[L^2]$	0.6300717791
$E[a^2]$	1.2801760409267
$E[p^2]$	16.9454930107385
$E[v^2]$	37.7808116990122
$E[ap]$	4.4904721130071
$E[av]$	6.4008802046335
$E[pv]$	24.6505831238765
$\mathrm{Var}[L]$	0.1856273347051
$\mathrm{Var}[a]$	0.2801760409267
$\mathrm{Var}[p]$	0.9454930107385
$\mathrm{Var}[v]$	1.7808116990122
$\mathrm{Cov}[a, p]$	0.4904721130071
$\mathrm{Cov}[a, v]$	0.4008802046335
$\mathrm{Cov}[p, v]$	0.6505831238765

Legend:

L	Single edge length.
a	Cell area.
p	Cell perimeter.
v	Number of vertices of cell.