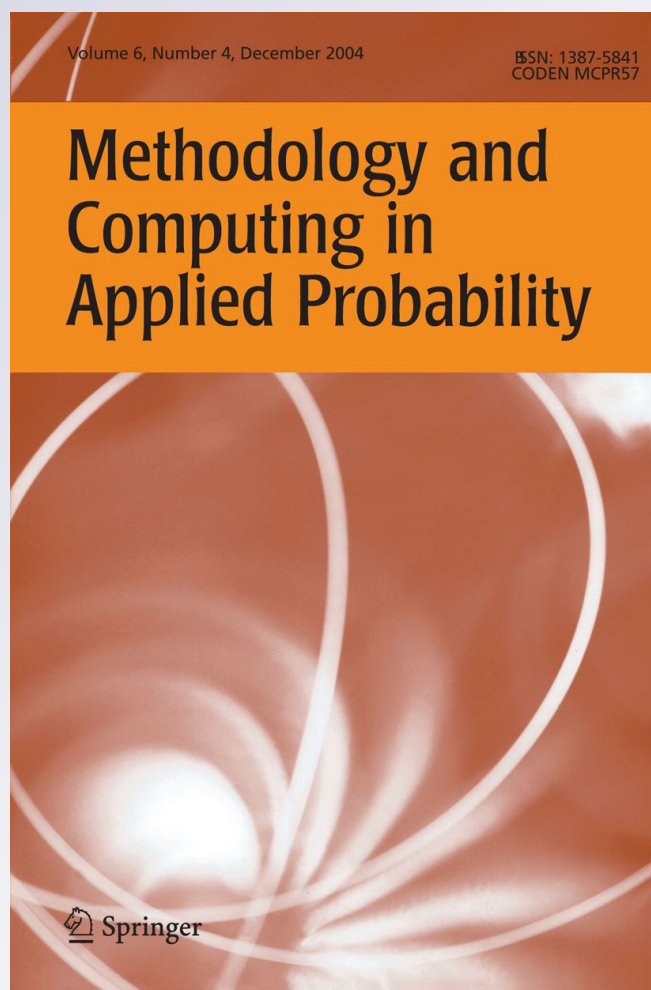# On Estimation of the Intensity Function of a Point Process

Marie-Colette N. M. van Lieshout

Methodology and Computing in Applied Probability

Springer

Springer

Springer

On Estimation of the Intensity Function of a Point Process

Marie-Colette N. M. van Lieshout

Abstract In this paper we review techniques for estimating the intensity function of a spatial point process. We present a unified framework of mass preserving general weight function estimators that encompasses both kernel and tessellation based estimators. We give explicit expressions for the first two moments of these estimators in terms of their product densities, and pay special attention to Poisson processes.

1 Introduction

Throughout this paper, consider a simple point process Φ on $\mathbb{R}^d$. We are interested in its first order moment measure defined as follows. Let $A \subset \mathbb{R}^d$ be a bounded Borel set and denote by $N(A)$ the random variable that counts the number of points of Φ that fall in A. If the expectation $M(A) = \mathbb{E}N(A)$ is finite for all bounded Borel sets A, the set function M can be extended uniquely to a σ-finite measure on the d-dimensional Borel sets, the *first order moment measure* Λ (see e.g. Daley and Vere-Jones 2003/2008 or Van Lieshout 2000 for further details). Make the further

M. N. M. van Lieshout (✉)
CWI and Eindhoven University of Technology, Science Park 123, 1098 XG,
Amsterdam, The Netherlands
e-mail: Marie-Colette.van.Lieshout@cwi.nl

assumption that Λ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$, so that

$$\Lambda(A) = \int_A \lambda(x)\, dx$$

for some measurable function $\lambda(x) \geq 0$.

Upon observation of a realisation of Φ in some bounded window, it is of interest to estimate the intensity function. Sometimes the scientific context or the data suggest a parametric form for the intensity function. In this case, likelihood based methods may be applied. For an overview of the state of the art, the user is referred to the recent handbook of spatial statistics (Gelfand et al. 2010) and the references therein. More often, non-parametric estimation is called for. In this paper, we shall review several estimators of λ given an observation of the point process Φ in some open, convex and bounded Borel set $A \neq \emptyset$ and interpret them as *general weight function estimators* of the form

$$\widehat{\lambda(x_0)} = \sum_{x \in \Phi \cap A} g(x_0; x, \Phi \cap A) \tag{1}$$

for some measurable weight function $g \geq 0$ that may depend on the point process as well as on $x_0 \in A$ and is assumed to integrate to unity, that is,

$$\int_A g(x_0; x, \Phi \cap A)\, dx_0 = 1.$$

For such estimators

$$\int_A \widehat{\lambda(x_0)}\, dx_0 = \sum_{x \in \Phi \cap A} \int_A g(x_0; x, \Phi \cap A)\, dx_0 = N(A),$$

so that $\widehat{\lambda(x_0)}$ is *mass preserving*.

The plan of this paper is as follows. First, in Section 2, we focus on kernel estimators and modify the widely used Berman–Diggle estimator so as to make it mass preserving. In Section 3 we study tessellation based estimators and show that they can be seen as adaptive kernel estimators. Then we concentrate on Poisson processes in Section 4 and present some examples. The paper closes with a brief discussion of alternative Bayesian approaches.

2 Mass Preserving Kernel Estimation

In the absence of specific information on the intensity function λ, it is natural to apply ideas from density estimation theory, as in the classic and widely used Berman–Diggle estimator (Diggle 1985; Berman and Diggle 1989)

$$\widehat{\lambda_{BD}(x_0)} = \frac{N(b(x_0, h) \cap A)}{|b(x_0, h) \cap A|} \tag{2}$$

for $x_0 \in A$. Here $b(x_0, h)$ denotes the open ball around x_0 with radius $h > 0$ and the notation $|\cdot|$ is used for Lebesgue mass (d-dimensional volume). The bandwidth h controls the amount of smoothing. Note that as A is open, one never divides by zero. In the sequel, we assume that the point locations are not subject to measurement

error. However, Eq. 2 can be adapted to noisy observations by means of deconvolution (Cucala 2008).

Another approach is based on the distance $r(x_0, k)$ of a point $x_0 \in A$ to its k-th nearest neighbour in $\Phi \cap A$, assuming there is one. Note that the closed ball $\overline{b(x_0, r(x_0, k))}$ around x_0 contains exactly k points of $\Phi \cap A$. This observation leads to the estimator

$$\widehat{\lambda_{NN}(x_0)} = \frac{k}{|b(x_0, r(x_0, k)) \cap A|}, \tag{3}$$

cf. Silverman (1986). In order to increase robustness against outliers, one may prefer to combine k-th nearest neighbour distances for several values of k (Granville 1998).

Although Eqs. 2 and 3 define natural estimators, they do not necessarily preserve the total mass in A, nor are they based on a general weight function. However, note that the Berman–Diggle estimator (2) may be written as

$$\sum_{x \in \Phi \cap A} \frac{1\{||x - x_0|| < h\}}{|b(x_0, h) \cap A|}.$$

Thus, each point x of Φ that falls in $A \cap b(x_0, h)$ is given a weight $|b(x_0, h) \cap A|^{-1}$. If x_0 is close to the boundary of A, the weight is larger than the reciprocal Lebesgue mass of a ball of radius h to compensate for the relative shortage of h-close points in $\Phi \cap A$. This form of edge correction is 'global' as $|b(x_0, h) \cap A|$ does not depend on x. Using a 'local' edge correction instead suggests the estimator

$$\widehat{\lambda_K(x_0)} = \sum_{x \in \Phi \cap A} \frac{1\{||x - x_0|| < h\}}{|b(x, h) \cap A|} \tag{4}$$

for $x_0 \in A$. Note that (4) is well-defined and coincides with (2) for x_0 such that $b(x_0, 2h) \subseteq A$. In contrast to Eq. 2, however, Eq. 4 is based on a proper weight function $g(x_0; x, \Phi \cap A) = 1\{||x - x_0|| < h\}/|b(x, h) \cap A|$, cf. Eq. 1, which does not depend on Φ. Consequently, the estimator defined by Eq. 4 is mass preserving.

In order to compute the moments of the estimator defined by Eq. 4, note that

$$\mathbb{E}\left[\sum_{x \in \Phi \cap A} \frac{1\{||x - x_0|| < h\}}{|b(x, h) \cap A|} \right] = \int_{b(x_0, h) \cap A} \frac{\lambda(x)}{|b(x, h) \cap A|} \, dx$$

by the Campbell–Mecke formula (see Daley and Vere-Jones 2003/2008 or Stoyan et al. 1995). For the second moment of $\widehat{\lambda_K(x_0)}$, write

$$\mathbb{E}\left[\widehat{\lambda_K(x_0)}^2 \right] = \mathbb{E}\left[\sum_{x, y \in \Phi \cap A}^{\neq} \left[\frac{1\{||x - x_0|| < h\}}{|b(x, h) \cap A|} \frac{1\{||y - x_0|| < h\}}{|b(y, h) \cap A|} \right] \right]$$

$$+ \mathbb{E}\left[\sum_{x \in \Phi \cap A} \frac{1\{||x - x_0|| < h\}}{|b(x, h) \cap A|^2} \right],$$

where the first term on the right hand side of the above equation is over pairs of different points, the second term over identical ones. The expectation of the second term in the right hand side can be expressed as an integral as before. To proceed, assume the second order factorial moment measure of Φ exists as a locally finite

measure that is absolutely continuous with respect to the product Lebesgue measure with Radon–Nikodym derivative $\rho^{(2)}$. Heuristically speaking, $\rho^{(2)}(x, y)$ is the joint probability of Φ placing a point in each of two infinitesimal regions centred at x and y, respectively. Then the expected value of the sum over pairs of different points in the formula for the second moment can be expressed as

$$\iint_{(b(x_0,h)\cap A)^2} \frac{\rho^{(2)}(x, y)}{|b(x, h) \cap A|\, |b(y, h) \cap A|}\, dx\, dy.$$

In summary, the first two moments of (4) are given by

$$\mathbb{E}\left[\widehat{\lambda_K(x_0)}\right] = \int_{b(x_0,h)\cap A} \frac{\lambda(x)}{|b(x, h) \cap A|}\, dx;$$

$$\mathbb{E}\left[\widehat{\lambda_K(x_0)}^2\right] = \iint_{(b(x_0,h)\cap A)^2} \frac{\rho^{(2)}(x, y)}{|b(x, h) \cap A|\, |b(y, h) \cap A|}\, dx\, dy$$
$$+ \int_{b(x_0,h)\cap A} \frac{\lambda(x)}{|b(x, h) \cap A|^2}\, dx.$$

Obviously, the spherical kernel in Eq. 4 may be replaced by any other kernel that integrates to unity.

Neither Eq. 2 nor 4 is universally better in terms of integrated mean squared error than its competitor. To see this, first consider a homogeneous Poisson process Φ with intensity $\lambda > 0$. Then, the integrated variance of the estimators defined both Eqs. 2 and 4 is equal to $\lambda \int_A |b(x, h) \cap A|^{-1}\, dx$. The bias of the Berman–Diggle estimator is zero, whereas the estimator defined by Eq. 4 is biased unless $\int_{b(x_0,h)\cap A} |b(x, h) \cap A|^{-1}\, dx = 1$. So, in general, the estimator defined by Eq. 2 will be preferred.

Next, let Φ be a Poisson process on A with intensity function $\lambda(x) = \lambda |b(x, h) \cap A|$, for some $\lambda > 0$. Then, the estimator defined by Eq. 4 is unbiased with integrated variance $\lambda |A|$. Write

$$m(x_0) = \int_{b(x_0,h)\cap A} \frac{|b(x, h) \cap A|}{|b(x_0, h) \cap A|^2}\, dx.$$

Then, $\mathbb{E}\left[\widehat{\lambda_{BD}(x_0)}\right]$ can be expressed as $\lambda(x_0)\, m(x_0)$, so its integrated squared bias is zero if and only if $m(x_0) = 1$ for almost all $x_0 \in A$. The integrated variance of the estimator defined by Eq. 2 is $\lambda \int_A m(x_0)\, dx_0$ which reduces to $\lambda |A|$ if $m(x_0) = 1$ for almost all $x_0 \in A$ so that the estimators are indistinguishable. Otherwise, the mass preserving kernel estimator should be preferred since $\int_A m(x_0)\, dx_0 \geq |A|$.

3 Tessellation Based Estimators

A potential disadvantage of kernel estimators such as Eq. 4 is that the same bandwidth is used throughout the observation window A, risking over smoothing in high intensity regions. Thus, estimators that adapt to Φ have been proposed based on tessellations.

More specifically, suppose that the realisations $\varphi = \{x_1, x_2, \dots\}$ of the point process Φ are almost surely in general quadratic position, that is, no $d + 2$ points are

located on the boundary of a sphere and no $k + 1$ points lie in a $k - 1$ dimensional affine subspace for $k = 2, \ldots, d$. Recall that the Voronoi cell of x_i in φ is the set

$$C(x_i \mid \varphi) = \{y \in \mathbb{R}^d : ||x_i - y|| \leq ||x_j - y|| \quad \forall x_j \in \varphi\}$$

that consists of all points in $\mathbb{R}^d$ that are at least as close to $x_i \in \varphi$ as to any other point of φ. Note that the Voronoi cells are closed and convex, but not necessarily bounded.

Under our assumptions, intersections between $k = 2, \ldots, d + 1$ different Voronoi cells are either empty or of dimension $d - k + 1$. In particular,

$$\bigcap_{i=1}^{d+1} C(x_i \mid \varphi) \neq \emptyset \Leftrightarrow b(x_1, \ldots, x_{d+1}) \cap \varphi = \emptyset,$$

where $b(x_1, \ldots, x_{d+1})$ is the open ball spanned by the points $x_1, \ldots, x_{d+1}$ on its boundary, and in that case the intersection is a single point, referred to as a vertex of the Voronoi tessellation.

Vertices can be used to define the second tessellation of interest to us in this paper. Suppose that φ contains at least $d + 1$ points. Each Voronoi vertex arising as the intersection of $d + 1$ cells $C(x_i \mid \varphi)$ defines a closed simplex, the convex hull of $\{x_1, \ldots, x_{d+1}\}$, which is called a Delaunay cell and denoted by $D(x_1, \ldots, x_{d+1})$. For $d = 1$, Delaunay cells are intervals, in the plane they form triangles. Alternatively, join points $x_i \neq x_j \in \varphi$ that share a common Voronoi border $C(x_i \mid \varphi) \cap C(x_j \mid \varphi) \neq \emptyset$ into a Delaunay side. Such x_i and x_j are said to be Voronoi neighbours; the set of neighbours of x_i in φ is denoted by $\mathcal{N}(x_i \mid \varphi)$. The union of Delaunay cells containing $x_i \in \varphi$ is known as the contiguous Voronoi cell, denoted $W(x_i \mid \varphi)$, of x_i in φ. For further details, see e.g. Møller (1994) or Okabe et al. (2000).

Note that the tessellation cells described above can be seen as adaptive neighbourhoods of a point of Φ. In contrast to the balls of fixed radius h used in Section 2, the sizes of the cells depend on the point process: In densely populated regions, the cells will be small, whereas they tend to be larger in regions of low intensity. This observation led Schaap and Van de Weygaert (2000), see also Schaap (2007) to introduce their Delaunay tessellation field estimator (DTFE) as follows. For $x \in \Phi \cap A$, set $\widehat{\lambda_D(x)} = (d + 1)/|W(x \mid \Phi \cap A)|$, and for any $x_0 \in A$ in the interior of some Delaunay cell, define

$$\widehat{\lambda_D(x_0)} = \frac{1}{d + 1} \sum_{x \in \Phi \cap D(x_0 \mid \Phi \cap A)} \widehat{\lambda_D(x)} \tag{5}$$

by averaging over the vertices of the Delaunay cell $D(x_0 \mid \Phi \cap A)$ that contains x_0. Note that the factor $(d + 1)^{-1}$ in Eq. 5 cancels out against the reciprocal factor in $\widehat{\lambda_D(x)}$. However, from the point of view of generalisation to interpolation schemes other than averaging over the vertices of $D(x_0 \mid \Phi \cap A)$, the formulation in Eq. 5 is useful (Schaap and Van de Weygaert 2000, Schaap 2007). An earlier variation on this theme was the Voronoi tessellation field estimator (Bernardeau and Van de Weygaert 1996; Ord 1978) that estimates $\lambda(x_0)$ by the reciprocal area of the Voronoi cell x_0 belongs to (with obvious modifications on the borders of these cells). The Voronoi approach, in contrast to that based on contiguous Voronoi cells, cannot be used in combination with linear interpolation and still be mass preserving (Schaap 2007), so we will from now on focus on Eq. 5.

In order to rephrase Eq. 5 as a general weight function estimator, write $\mathcal{D}(\Phi \cap A)$ for the family of Delaunay cells of $\Phi \cap A$, use a superscript $^\circ$ to denote topological interior, and set

$$g(x_0; x, \Phi \cap A) = \frac{\sum_{D_j \in \mathcal{D}(\Phi \cap A)} 1\{x_0 \in D_j^\circ; x \in D_j\}}{|W(x \mid \Phi \cap A)|}$$

for $x_0 \in A \setminus \Phi$, $x \in \Phi \cap A$ and $g(x; x, \Phi \cap A) = (d+1)/|W(x \mid \Phi \cap A)|$. The function g is well-defined on the event that $\Phi \cap A$ contains at least $d+1$ points. For configurations consisting of less than $d+1$ points, we may set the function g equal to $1/|A|$. Now,

$$\widehat{\lambda_D(x_0)} = \sum_{x \in \Phi \cap A} g(x_0; x, \Phi \cap A),$$

and, as $\int_A g(x_0; x, \Phi \cap A)\, dx_0 = 1$, mass preservation follows.

Arguments similar to those used in Section 2 for the derivation of the moments of the kernel estimator (4) can be used to calculate the moments of the estimator defined by Eq. 5. The main difference is that in the case of Eq. 5, $g(x_0; x, \Phi, A)$ does depend on Φ, which necessitates the use of Palm distributions. Intuitively speaking, the Palm distribution P_x of Φ at x can be seen as the conditional distribution of Φ given that a point falls at x, with similar interpretations for higher order Palm distributions. More specifically, it can be shown that the first two moments of the estimator defined by Eq. 5 are given by

$$\mathbb{E}\left[\widehat{\lambda_D(x_0)}\right] = \int_A \mathbb{E}_x\left[g(x_0; x, \Phi \cap A)\right]\lambda(x)\, dx;$$

$$\mathbb{E}\left[\widehat{\lambda_D(x_0)}^2\right] = \iint_{A^2} \mathbb{E}_{x,y}^{(2)}\left[g(x_0; x, \Phi \cap A)\, g(x_0; y, \Phi \cap A)\right]\rho^{(2)}(x, y)\, dx\, dy$$

$$+ \int_A \mathbb{E}_x\left[g^2(x_0; x, \Phi \cap A)\right]\lambda(x)\, dx, \tag{6}$$

provided the second order factorial moment measure of Φ exists as a locally finite measure that is absolutely continuous with respect to the product Lebesgue measure with Radon–Nikodym derivative $\rho^{(2)}$. Here $\mathbb{E}_x$ and $\mathbb{E}_{x,y}^{(2)}$ denote, respectively, expectation with respect to the Palm distribution of Φ at x and the two-fold Palm distribution at x, y (Hanisch 1982). Indeed, the right hand sides in Eq. 6 are valid expressions for the first and second moment of any estimator of the form (1).

If Φ is a Poisson process with intensity function $\lambda(\cdot)$, Eq. 6 can be simplified by observing that the Palm and two-fold Palm distributions coincide with the distributions of $\Phi \cup \{x\}$ and $\Phi \cup \{x, y\}$, respectively. Moreover, $\rho^{(2)}(x, y) = \lambda(x)\,\lambda(y)$.

4 Example: The Poisson Process

For stationary Poisson processes, the expressions for the moments of the estimators discussed in Sections 2 and 3 can be simplified.

Theorem 1 *Let Φ be a stationary Poisson process on $\mathbb{R}^d$ with intensity $\lambda > 0$. Then the Delaunay tessellation field estimator $\widehat{\lambda_D(0)}$ is unbiased with variance $c_d \lambda^2$, where $c_d + 1$ is given by*

$$\mathbb{E}_1 \left[\frac{1}{|W(0 \mid \Phi \cup \{0\})|} \left\{ 1 + \sum_{y \in \mathcal{N}(0 \mid \Phi \cup \{0\})} \frac{|W(0 \mid \Phi \cup \{0\}) \cap W(y \mid \Phi \cup \{0\})|}{|W(y \mid \Phi \cup \{0\})|} \right\} \right].$$

Here $\mathbb{E}_1$ denotes expectation with respect to a unit intensity Poisson process.

The proof is deferred to an Appendix. In words, the variance of the Delaunay tessellation field estimator increases quadratically in λ with a dimension dependent scalar multiplier c_d. For example if $d = 1$, $c_1 = 2 (2 - \pi^2/6) \approx 0.7$. Since the Berman–Diggle estimator is unbiased with variance $\lambda \, \omega_d^{-1} \, h^{-d}$, where ω_d is the volume of the unit ball in $\mathbb{R}^d$, it is more efficient than (5) whenever $\mathbb{E}N(b(0, h)) > c_d^{-1}$. Hence on the line, the estimator defined by Eq. 2 is the better choice if $\mathbb{E}N((-h, h)) = 2\lambda h > 1.4$. For comparison, the computation of the estimator defined by Eq. 5 in this case requires four points.

In the remainder of this section, we contrast the behaviour of the DTFE estimator on the line with that of a kernel estimator. We minimise edge effects by simulating Φ beyond the region used for estimation.

In order to choose the bandwidth in such a way that a meaningful comparison can be made, note that the computation of the right hand side of Eq. 5 involves four points. Therefore we set the global bandwidth so that an interval of length $2h$ contains on average four points.

First consider the highly fluctuating step function plotted in Fig. 1. Here the Delaunay tessellation field estimator clearly outperforms the kernel estimator which gives a misleading picture as it is large for low values of the intensity function and
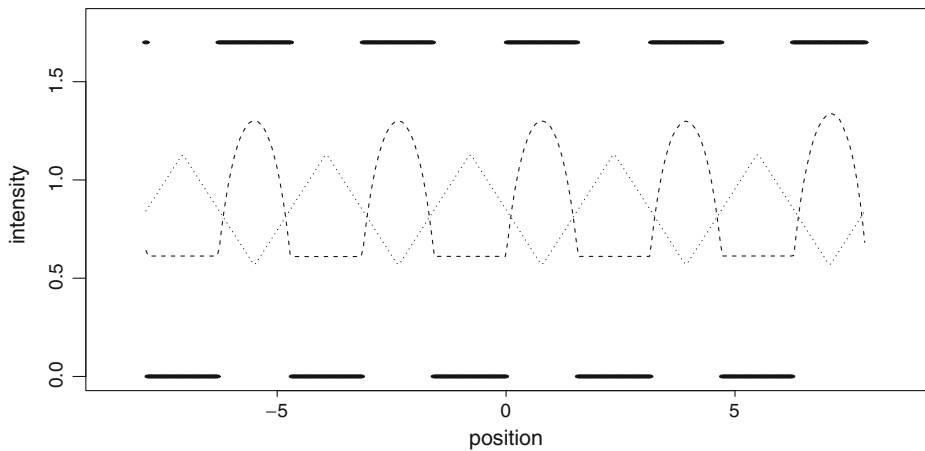

Fig. 1 Intensity function (*solid line*) with estimates of $\mathbb{E}\widehat{\lambda_D(\cdot)}$ (*coarsely dashed line*) and $\mathbb{E}\widehat{\lambda_K(\cdot)}$ (*finely dashed line*)
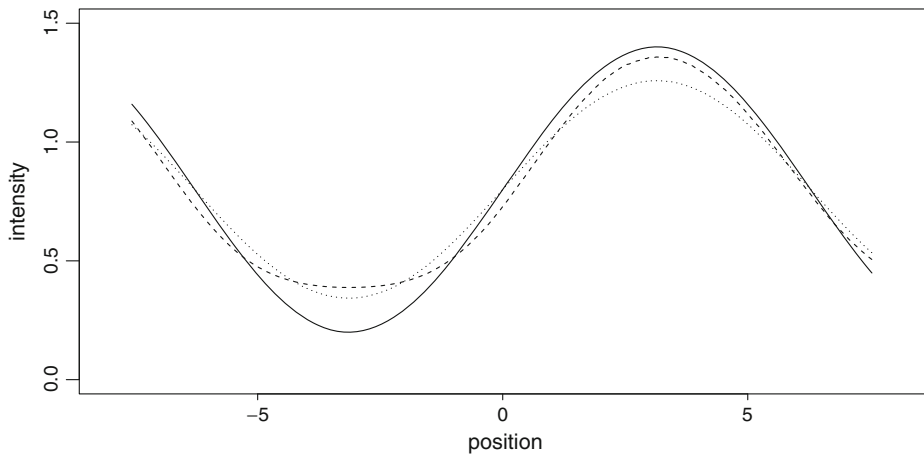
Fig. 2 Intensity function (*solid line*) with estimates of $\mathbb{E}\widehat{\lambda_D(\cdot)}$ (*coarsely dashed line*) and $\mathbb{E}\widehat{\lambda_K(\cdot)}$ (*finely dashed line*)

small when the intensity function is large. The estimated integrated mean squared error of $\widehat{\lambda_D(\cdot)}$ is 22, the integrated mean absolute error 9.6. The kernel estimator $\widehat{\lambda_K(\cdot)}$ has estimated integrated mean squared error 18 and integrated mean absolute error 16.

We also considered the mildly oscillating function $\lambda(x) = b \sin(c\,x) + d$ with $b = 0.6$, $c = 0.5$ and $d = 0.8$. The estimated expected values of the kernel and DTFE estimator are given in Fig. 2. Both follow the intensity graph, the DTFE is somewhat better able to cope with the peak, the kernel estimator with the valley. The kernel estimator $\widehat{\lambda_K(\cdot)}$ has estimated integrated mean squared error 3 and integrated mean absolute error 1.3. The estimated integrated mean squared error of $\widehat{\lambda_D(\cdot)}$ is 10, the integrated mean absolute error 1.0. Moreover, for the latter estimator, the average estimated standard deviation (0.7) is up to the first decimal equal to $\bar{\lambda}\sqrt{c_1}$ where $\bar{\lambda}$ is the average intensity. For the wilder step function considered above, the average estimated standard deviation (0.8) is larger than $\bar{\lambda}\sqrt{c_1} = 0.7$.

In summary, the Delaunay tessellation field estimator is preferred when peaks in the intensity surface are an important feature that would be flattened out by standard kernel estimation. The price to pay is a higher standard deviation as well as a higher computational cost. Therefore, in most cases, statisticians will prefer to use mass preserving kernel estimators.

5 Discussion

In this paper, we reviewed techniques for estimating the intensity function of simple point processes on Euclidean spaces. We presented mass preserving general weight function estimators and showed that both kernel and tessellation based estimators can be described as members of this class. We derived explicit expressions for the first two moments of these estimators in terms of their product densities. We then turned

to the special, computationally amenable, case of Poisson processes and presented some examples.

It should be mentioned that Bayesian methods have been proposed as well. For example, Heikkinen and Arjas' method (1998) is based on approximating the unknown intensity function by a function that is constant on the Voronoi cells of random point patterns. Smoothing between nearby intensity values is achieved by means of a Markov random field prior in the spirit of Bayesian image analysis (Winkler 2003).

Appendix: Proof of Theorem 1

Let Φ be a stationary Poisson process on $\mathbb{R}^d$ with intensity $\lambda > 0$. We begin the proof of Theorem 1 by showing the unbiasedness of $\widehat{\lambda_D(0)}$.

Let $b(x, y_1, \ldots, y_d)$ be the open ball spanned by the points $x, y_1, \ldots, y_d$ on its topological boundary, and let $D^\circ(x, y_1, \ldots, y_d)$ be the open simplex that is the interior of the convex hull of $\{x, y_1, \ldots, y_d\}$. Recall that the points $x, y_1, \ldots, y_d$ define a Voronoi vertex, or, equivalently, a Delaunay cell if and only if there are no points in $b(x, y_1, \ldots, y_d)$. Thus, $\mathbb{E}\left[\widehat{\lambda_D(0)}\right] = \lambda \int_{\mathbb{R}^d} \mathbb{E}\left[g(0; x, \Phi \cup \{x\})\right] dx$ is equal to

$$\lambda \int \mathbb{E}\left[\sum_{\{y_1, \ldots, y_d\} \subset \Phi}^{\neq} \frac{1\{0 \in D^\circ(x, y_1, \ldots, y_d); b(x, y_1, \ldots, y_d) \cap (\Phi \cup \{x\}) = \emptyset\}}{|W(x \mid \Phi \cup \{x\})|}\right] dx$$

$$= \lambda \int \mathbb{E}\left[\sum_{\{z_1, \ldots, z_d\} \subset \Phi_{-x}}^{\neq} \frac{1\{-x \in D^\circ(0, z_1, \ldots, z_d); b(0, z_1, \ldots, z_d) \cap \Phi_{-x} = \emptyset\}}{|W(0 \mid \Phi_{-x} \cup \{0\})|}\right] dx.$$

Here the notation $\sum^{\neq}$ is used to indicate that the sum is over sets of distinct points and Φ_{-x} denotes the (random) pattern obtained from Φ by translating all its points by $-x$. By stationarity, the above expression is equal to

$$\lambda \int \mathbb{E}\left[\sum_{\{z_1, \ldots, z_d\} \subset \Phi}^{\neq} \frac{1\{-x \in D^\circ(0, z_1, \ldots, z_d); b(0, z_1, \ldots, z_d) \cap \Phi = \emptyset\}}{|W(0 \mid \Phi \cup \{0\})|}\right] dx.$$

Hence, by Fubini's theorem,

$$\mathbb{E}\left[\widehat{\lambda_D(0)}\right] = \lambda \mathbb{E}\left[\frac{\sum_{\{z_1, \ldots, z_d\} \subset \Phi}^{\neq} |D^\circ(0, z_1, \ldots, z_d)| \, 1\{b(0, z_1, \ldots, z_d) \cap \Phi = \emptyset\}}{|W(0 \mid \Phi \cup \{0\})|}\right]$$

$$= \lambda \mathbb{E}\left[\frac{|W(0 \mid \Phi \cup \{0\})|}{|W(0 \mid \Phi \cup \{0\})|}\right] = \lambda,$$

which proves the claim of unbiasedness.

Next, we consider the two terms in the expression for the second moment of $\widehat{\lambda_D(0)}$, cf. Eq. 6. Our first aim is to show that

$$E(\lambda, d) := \lambda^2 \iint \mathbb{E}\left[g(0; x, \Phi \cup \{x, y\})\, g(0; y, \Phi \cup \{x, y\})\right] dx\, dy$$

$$= \lambda^2 \int \mathbb{E}_1\left[\frac{|W(0 \mid \Phi \cup \{0, x\}) \cap W(x \mid \Phi \cup \{0, x\})|}{|W(0 \mid \Phi \cup \{0, x\})|\, |W(x \mid \Phi \cup \{0, x\})|}\right] dx,$$

where $\mathbb{E}_1$ denotes expectation with respect to a unit intensity Poisson process. Consequently, by the Nguyen–Zessin formula (see e.g. Van Lieshout 2000; Stoyan et al. 1995),

$$E(\lambda, d) = \lambda^2\, \mathbb{E}_1\left[\frac{1}{|W(0 \mid \Phi \cup \{0\})|} \sum_{y \in \mathcal{N}(0 \mid \Phi \cup \{0\})} \frac{|W(0 \mid \Phi \cup \{0\}) \cap W(y \mid \Phi \cup \{0\})|}{|W(y \mid \Phi \cup \{0\})|}\right].$$

To do so, write Φ_{d-1} for sets of $d - 1$ distinct points in Φ. Then, as $g(0; x, \Phi \cup \{x, y\})\, g(0; y, \Phi \cup \{x, y\})$ vanishes whenever x and y do not belong to the same Delaunay cell containing 0 in its interior,

$E(\lambda, d)$

$$= \lambda^2 \iint \mathbb{E}\left[\sum_{z \in \Phi_{d-1}} \frac{1\{0 \in D^\circ(x, y, z); b(x, y, z) \cap \Phi = \emptyset\}}{|W(x \mid \Phi \cup \{x, y\})|\, |W(y \mid \Phi \cup \{x, y\})|}\right] dx\, dy$$

$$= \lambda^2 \iint \mathbb{E}\left[\sum_{z \in \Phi_{-x; d-1}} \frac{1\{-x \in D^\circ(0, y - x, z); b(0, y - x, z) \cap \Phi_{-x} = \emptyset\}}{|W(0 \mid \Phi_{-x} \cup \{0, y - x\})|\, |W(y - x \mid \Phi_{-x} \cup \{0, y - x\})|}\right] dx\, dy.$$

Because of stationarity, $E(\lambda, d)$ can be written as

$$\lambda^2 \iint \mathbb{E}\left[\sum_{z \in \Phi_{d-1}} \frac{1\{-x \in D^\circ(0, y - x, z); b(0, y - x, z) \cap \Phi = \emptyset\}}{|W(0 \mid \Phi \cup \{0, y - x\})|\, |W(y - x \mid \Phi \cup \{0, y - x\})|}\right] dx\, dy$$

$$= \lambda^2 \iint \mathbb{E}\left[\frac{\sum_{z \in \Phi_{d-1}} 1\{-x \in D^\circ(0, y, z); b(0, y, z) \cap \Phi = \emptyset\}}{|W(0 \mid \Phi \cup \{0, y\})|\, |W(y \mid \Phi \cup \{0, y\})|}\right] dx\, dy.$$

Scaling by $\lambda^{1/d}$ yields that $\lambda^{-2} E(\lambda, d)$ is equal to

$$\iint \mathbb{E}\left[\frac{\sum_{z \in \Phi_{d-1}} 1\{-\lambda^{1/d} x \in D^\circ(0, \lambda^{1/d} y, \lambda^{1/d} z); b(0, \lambda^{1/d} y, \lambda^{1/d} z) \cap \lambda^{1/d} \Phi = \emptyset\}}{\lambda^{-1}|W(0 \mid \lambda^{1/d} \Phi \cup \{0, \lambda^{1/d} y\})|\, \lambda^{-1}|W(\lambda^{1/d} y \mid \lambda^{1/d} \Phi \cup \{0, \lambda^{1/d} y\})|}\right] dx\, dy.$$

Since $\lambda^{1/d} \Phi$ is a unit intensity Poisson process, we obtain

$$\lambda^{-2} E(\lambda, d) = \iint \mathbb{E}_1\left[\frac{\sum_{z \in \Phi_{d-1}} 1\{-x \in D^\circ(0, y, z); b(0, y, z) \cap \Phi = \emptyset\}}{|W(0 \mid \Phi \cup \{0, y\})|\, |W(y \mid \Phi \cup \{0, y\})|}\right] dx\, dy.$$

An appeal to Fubini's theorem proves the claim.

To conclude the proof of Theorem 1, we shall show that

$$E'(\lambda, d) := \lambda \int \mathbb{E}\left[g^2(0; x, \Phi \cup \{x\})\right] dx = \lambda^2 \mathbb{E}_1\left[\frac{1}{|W(0 \mid \Phi \cup \{0\})|}\right],$$

where, as before, $\mathbb{E}_1$ denotes expectation with respect to a unit intensity Poisson process.

Arguing as in the proof of unbiasedness, $E'(\lambda, d)$ can be written as

$$\lambda \int \mathbb{E}\left[\left(\sum_{\{z_1,\ldots,z_d\}\subset\Phi}^{\neq} \frac{1\{-x \in D^\circ(0, z_1, \ldots, z_d); b(0, z_1, \ldots, z_d) \cap \Phi = \emptyset\}}{|W(0 \mid \Phi \cup \{0\})|}\right)^2\right] dx$$

$$= \lambda \int \mathbb{E}\left[\sum_{\{z_1,\ldots,z_d\}\subset\Phi}^{\neq} \frac{1\{-x \in D^\circ(0, z_1, \ldots, z_d); b(0, z_1, \ldots, z_d) \cap \Phi = \emptyset\}}{|W(0 \mid \Phi \cup \{0\})|^2}\right] dx$$

since $-x$ can belong to at most a single Delaunay interior. Write Φ_d for sets of d distinct points in Φ and scale each point by $\lambda^{1/d}$ to obtain that $E'(\lambda, d)$ is equal to

$$\lambda \int \mathbb{E}\left[\sum_{z \in \Phi_d} \frac{1\{-\lambda^{1/d}x \in D^\circ(0, \lambda^{1/d}z); b(0, \lambda^{1/d}z) \cap \lambda^{1/d}\Phi = \emptyset\}}{\lambda^{-2}|W(0 \mid \lambda^{1/d}\Phi \cup \{0\})|^2}\right] dx,$$

which, since $\lambda^{1/d}\Phi$ is a unit intensity Poisson process, reduces to

$$\lambda^2 \int \mathbb{E}_1\left[\frac{\sum_{\{z_1,\ldots,z_d\}\subset\Phi} 1\{-x \in D^\circ(0, z_1, \ldots, z_d); b(0, z_1, \ldots, z_d) \cap \Phi = \emptyset\}}{|W(0 \mid \Phi \cup \{0\})|^2}\right] dx.$$

Noting that the sum of d-volumes of Delaunay cells involving 0 is exactly equal to the volume of $W(0 \mid \Phi \cup \{0\})$ concludes the proof of the claim. Finally, an appeal to Eq. 6 and collecting all terms computed above finishes the proof of Theorem 1.

References

Berman M, Diggle PJ (1989) Estimating weighted integrals of the second-order intensity of a spatial point process. J Roy Stat Soc Ser B 51:81–92

Bernardeau F, Van de Weygaert R (1996) A new method for accurate estimation of velocity field statistics. Mon Not Roy Astron Soc 279:693–711

Cucala L (2008) Intensity estimation for spatial point processes observed with noise. Scand J Stat 35:322–334

Daley DJ, Vere-Jones D (2003/2008) An introduction to the theory of point processes. Springer, New York (First edition 1988)

Diggle PJ (1985) A kernel method for smoothing point process data. Appl Stat 34:138–147

Gelfand AE, Diggle PJ, Fuentes M, Guttorp P (2010) Handbook of spatial statistics. CRC Press, Boca Raton

Granville V (1998) Estimation of the intensity of a Poisson point process by means of nearest neighbour distances. Stat Neerl 52:112–124

Hanisch KH (1982) On inversion formulae for n-fold Palm distributions of point processes in LCS-spaces. Math Nachr 106:171–179

Heikkinen J, Arjas E (1998) Non-parametric Bayesian estimation of a spatial Poisson intensity. Scand J Stat 25:435–450

Van Lieshout MNM (2000) Markov point processes and their applications. Imperial College Press, London

Møller J (1994) Lectures on random Voronoi tessellations. Lecture notes in statistics, vol 87. Springer, New York

Okabe A, Boots B, Sugihara K, Chiu SN (2000) Spatial tessellations. Concepts and applications of Voronoi diagrams, 2nd edn. Wiley, New York

Ord JK (1978) How many trees in a forest? Math Sci 3:23–33

Schaap WE (2007) DTFE. The Delaunay tessellation field estimator. PhD thesis, University of Groningen

Schaap WE, Van de Weygaert R (2000) Letter to the editor. Continuous fields and discrete samples: reconstruction through Delaunay tessellations. Astron Astrophys 363:L29–L32

Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London

Stoyan D, Kendall WS, Mecke J (1995) Stochastic geometry and its applications, 2nd edn. Wiley, Chichester

Winkler G (2003) Image analysis, random fields and Markov chain Monte Carlo methods: a mathematical introduction. Applications of mathematics, vol 27, 2nd edn. Springer, Berlin