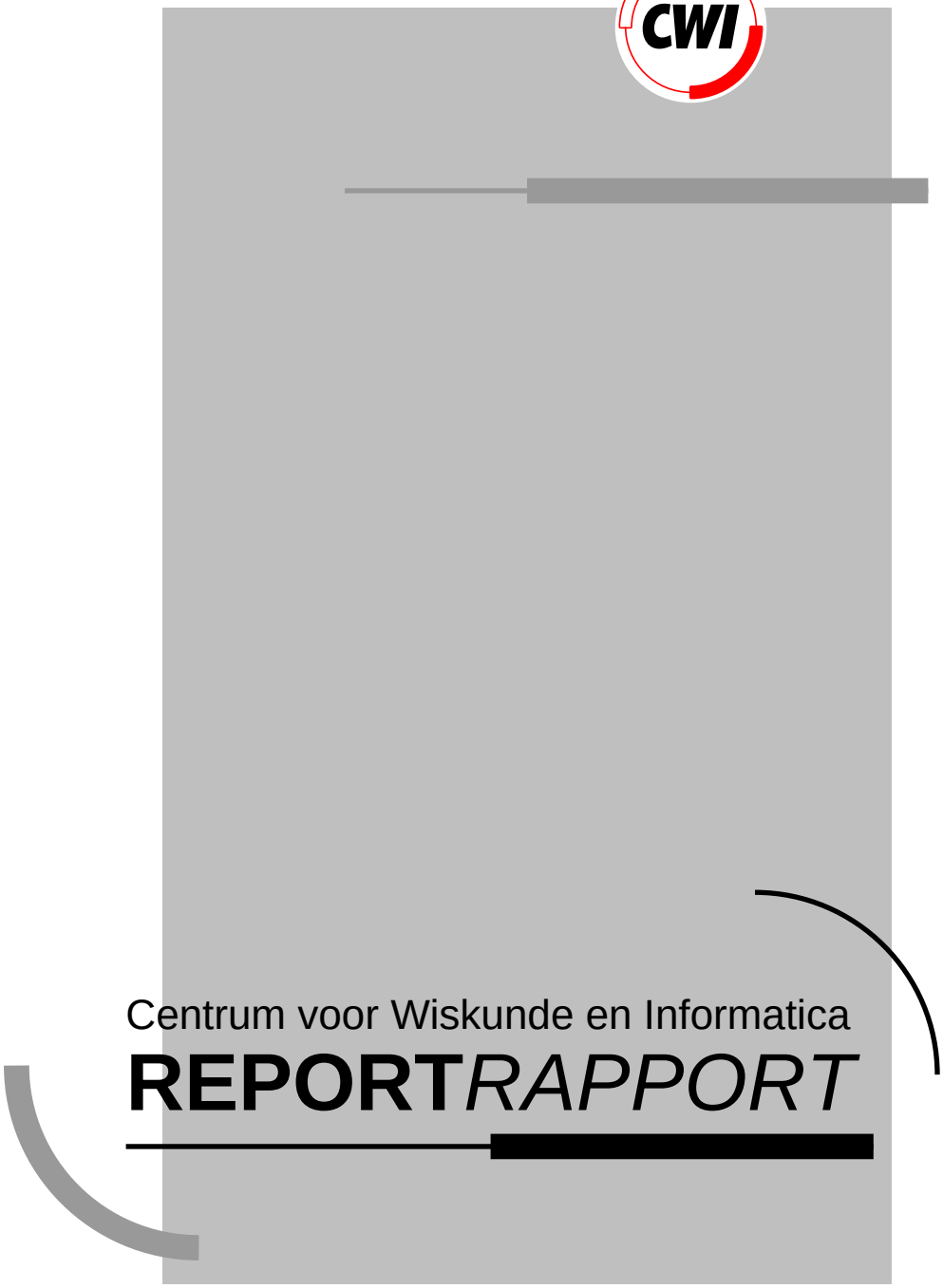

Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

Kaplan-Meier estimators for interpoint distance distributions of spatial point processes

A. Baddeley, R.D. Gill

Department of Operations Reasearch, Statistics, and System Theory

BS-R9315 1993

Kaplan-Meier Estimators of Interpoint Distance Distributions for Spatial Point Processes

Adrian Baddeley

Richard D. Gill*

CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Department of Mathematics and Computer Science
University of Leiden, The Netherlands

* *Mathematical Institute, University of Utrecht,*
Budapestlaan 6, 3584 CD Utrecht, The Netherlands

Abstract

When a spatial point process is observed through a bounded window, edge effects hamper the estimation of characteristics such as the empty space function F, the nearest neighbour distance distribution G, and the reduced second order moment function K. Here we propose and study product-limit type estimators of F, G and K based on the analogy with censored survival data: the distance from a fixed point to the nearest point of the process is right-censored by its distance to the boundary of the window. The resulting estimators have a ratio-unbiasedness property that is standard in spatial statistics. We show that the empty space function F of *any* stationary point process is absolutely continuous, and so is the product-limit estimator of F. The estimators are strongly consistent when there are independent replications or when the sampling window becomes large. We sketch a CLT for independent replications within a fixed observation window, and asymptotic theory for independent replications of sparse Poisson processes. In simulations the new estimators are generally much more efficient than the 'border method' estimator but (for estimators of K) somewhat less efficient than sophisticated edge corrections.

AMS Subject Classification (1991): 62G05; 62H11, 60D05.
Keywords & Phrases: border correction method, dilation, distance transform, edge corrections, edge effects, empty space statistic, erosion, functional delta-method, influence function, K-function, local knowledge principle, nearest-neighbour distance, product integration, reduced sample estimator, reduced second moment measure, sparse Poisson asymptotics, spatial statistics, survival data.
Note: This paper has been submitted for publication

1. Introduction

The exploratory data analysis of observations of a spatial point process often starts with the estimation of certain distance distributions: F, the distribution of the distance from an arbitrary point in space to the nearest point of the process; G, the distribution of the distance from a typical point of the process to the nearest other point of the process; and $K(t)$, the expected number of other points within distance t of a typical point of the process, divided by the intensity α. Equivalently αK is the sum over all $n = 1, 2, \ldots$ of the distribution function of the distance from a typical point of the process to the nth nearest point. Popular names for F, G and K are the empty space function, the nearest neighbour distance distribution, and Ripley's K or reduced second moment function. For a homogeneous Poisson process F, G and K take known functional forms, and deviations of estimates of F, G, K from these forms are taken as indications of 'clustered' or 'inhibited' alternatives [11, 37, 38].

However, the estimation of F, G and K is hampered by edge effects arising because the point process is observed within a bounded window W. Essentially the distance from a given reference point to the nearest point of the process is *censored* by its distance to the boundary of W. Edge effects become rapidly more severe as the dimension of space increases, or as the distance t increases.

Traditionally in spatial statistics, one uses edge-corrected estimators which are weighted empirical distributions of the observed distances. The simplest approach is the "border method" [38] in which we restrict attention (when estimating F, G or K at distance t) to those reference points lying more than t units away from the boundary of W. These are the points x for which distances up to t are observed without censoring. This approach is sometimes also justified by appealing to the "local knowledge principle" of mathematical morphology [42, pp. 49, 233]. However, the border method throws away an appreciable number of points; in three dimensions [4] it seems to be unacceptably wasteful, especially when estimating G.

In more sophisticated edge corrections, the weight $c(x, y)$ attached to the observed distance $||x-y||$ between two points x, y is the reciprocal of the probability that this distance will be observed under certain invariance assumptions (stationarity under translation and/or rotation). Corrections of this type were first suggested by Miles [33] and developed by Ripley, Lantuéjoul, Hanisch, Stoyan, Ohser and others [11, 25, 34, 36, 37], [42, p. 246]. For surveys see [38, chap. 3], [46, pp. 122–131], [8, chap. 8].

Now the estimation problem for F, G and K when observing a point process Φ through a bounded window W has a clear analogy, already implicitly drawn above, to the estimation of a survival function based on a sample of randomly censored survival times. This paper develops the analogy, and proposes Kaplan-Meier [28] or product-limit estimators for F, G and K. Since the observed, censored distances are highly interdependent, classical theory from survival analysis has little to say about statistical properties of the new estimators. One may however hope that the new estimators are better than the classical edge corrections, as in the survival analysis situation the Kaplan-Meier estimator has various large-sample optimality properties. In fact the border method for edge correction, described above, is analogous to the so-called reduced sample estimator, an inefficient competitor to the Kaplan-Meier estimator obtained using only those observations for which the censoring time is at least t when estimating the probability of survival to time t.

Surprisingly the analogy between edge effects for point processes and censoring of survival times does not seem to have been noted before. Laslett [29, 30] noted that when a spatial *line segment* process is observed within a bounded window, the resulting edge effects on the observed line segment lengths can be compared to censoring of survival times. However in that case the analogy is not especially helpful: a Kaplan-Meier type estimator for the segment length distribution is inconsistent, and the NPMLE is a different estimator which is difficult to analyse; see Wijers [49]. Zimmerman [50] proposed introducing *artificial* censoring in spatial sampling by restricting the maximum search distance from any reference point.

The estimation of F by a Kaplan-Meier type estimator poses a new (for survival analysis) problem, since one has a *continuum* of observations: for each point in the sampling window, a censored distance to the nearest point of the process. We solve this problem by using product-integration theory [23].

Together with estimators of F, G and K one would like to evaluate their accuracy. We make a start on this by using linearisation techniques (the functional delta-method, see [21]) leading to several proposals for variance estimators. This also leads to an evaluation of asymptotic efficiency in some simple, theoretical situations.

The plan of the paper is as follows: §2 recalls some definitions from spatial statistics and from the analysis of survival data; §3 introduces our Kaplan-Meier style estimator of the empty space function F; §4 discusses asymptotic properties of this estimator; §5 and §6 treat the estimation of G and K respectively in less detail. Critical comments are collected in §7.

2. Preliminaries

2.1 Spatial statistics

Let Φ be a point process in $\mathbb{R}^k$, observed through a window $W \subset \mathbb{R}^k$. We assume W is compact and inner regular (it is the closure of its interior), and denote its boundary by ∂W.

We consider Φ both as a random set in $\mathbb{R}^k$ and as a random measure. The problem is, based on the data $\Phi \cap W$ (and knowledge of W itself) to estimate the functions F, G and K defined as follows.

For $x \in \mathbb{R}^k$, $A \subset \mathbb{R}^k$ let

$$\rho(x, A) = \inf\{||x - a||_2 : a \in A\} \tag{2.1}$$

be the shortest Euclidean distance from x to A. Define

$$\begin{aligned} A_{\oplus r} &= \{x \in \mathbb{R}^k : \rho(x, A) \le r\} \\ A_{\ominus r} &= \{x \in A : \rho(x, A^c) > r\} \end{aligned}$$

where ^{c} denotes complement. For A closed, these are respectively the dilation and erosion of A by a ball of radius r:

$$\begin{aligned} A_{\oplus r} &= \bigcup_{x \in A} B(x, r) \\ A_{\ominus r} &= \left((A^c)_{\oplus r}\right)^c \end{aligned}$$

where $B(x, r)$ is the closed ball of radius r, centre x in $\mathbb{R}^k$.

Assume now that Φ is a.s. stationary under translations and has finite positive intensity α. Thus for any bounded Borel set $A \subset \mathbb{R}^k$

$$\mathbb{E}\,\Phi(A) = \alpha\,|A|_k$$

where $|\cdot|_k$ denotes k-dimensional Lebesgue volume. For $r \ge 0$ define

$$\begin{aligned} F(r) &= \mathbb{P}\{\rho(0, \Phi) \le r\} && (2.2) \\ &= \mathbb{P}\{\Phi(B(0, r)) > 0\} \\ G(r) &= \mathbb{P}\{\rho(0, \Phi \setminus \{0\}) \le r \mid 0 \in \Phi\} && (2.3) \\ &= \mathbb{P}\{\Phi(B(0, r)) > 1 \mid 0 \in \Phi\} \\ K(r) &= \alpha^{-1}\mathbb{E}\{\Phi(B(0, r) \setminus \{0\}) \mid 0 \in \Phi\} && (2.4) \end{aligned}$$

By stationarity the point 0 in these expressions may be replaced by any arbitrary point x.

The conditional expectations given $0 \in \Phi$ used above are expectations with respect to the Palm distribution of Φ at 0. Alternative definitions using the Campbell-Mecke formula [46] are

$$G(r) = \frac{\mathbb{E}\sum_{x \in \Phi \cap A} \mathbf{1}\{\rho(x, \Phi \setminus \{x\}) \le r\}}{\mathbb{E}\Phi(A)} \tag{2.5}$$

and

$$\alpha K(r) = \frac{\mathbb{E}\sum_{x \in \Phi \cap A} \Phi(B(x, r) \setminus \{x\})}{\mathbb{E}\Phi(A)} \tag{2.6}$$

holding for arbitrary measurable sets A with $|A|_k > 0$. The latter definition of K holds for any second-order stationary process [46].

Edge-corrected estimators for F, G and K based on observations of $\Phi \cap W$ are reviewed in [38, chap. 3], [46, pp. 122–131], [8, chap. 8]. For other comments see [4, 5] and for recent extensions [13, 14, 15, 16, 17, 18, 20, 43].

Most of the standard estimators in spatial statistics are not unbiased but instead are ratios of two unbiased consistent estimators

$$\widehat{\theta} = \frac{Y}{X} \quad \text{where} \quad \theta = \frac{\mathbb{E}Y}{\mathbb{E}X}$$

with

$$\begin{aligned} X, Y &\geq 0 \\ X = 0 &\Rightarrow Y = 0 \\ \mathbb{P}\{X > 0\} &> 0 \end{aligned}$$

typically arising as the mean of a weighted empirical distribution where the weights are random variables [4, 38]. We call such estimators "ratio-unbiased" and accept this property as a substitute for the generally unobtainable unbiasedness.

2.2 Survival data

Next we recall some theory of the Kaplan-Meier and reduced sample estimators. Suppose $T_1, \ldots, T_n$ are i.i.d. positive r.v.'s with distribution function F and survival function $S = 1 - F$. Let $C_1, \ldots, C_n$ be independent of the T_i's and i.i.d. with d.f. H. Let $\widetilde{T}_i = T_i \wedge C_i$, $D_i = \mathbf{1}\{T_i \leq C_i\}$ where $a \wedge b$ denotes $\min\{a, b\}$. Then $(\widetilde{T}_1, D_1), \ldots, (\widetilde{T}_n, D_n)$ is a sample of censored survival times $\widetilde{T}_i$ with censoring indicators D_i (really, *non*-censoring indicators).

The *reduced-sample estimator* of F is

$$\widehat{F}^{\mathrm{rs}}(t) = \frac{\#\{i : \tilde{T}_i \leq t \leq C_i\}}{\#\{i : C_i \geq t\}} \tag{2.7}$$

This requires that we are able to observe the censoring times C_i themselves, or at least the event $C_i \geq t$ for all t for which $F(t)$ must be estimated. This estimator is clearly pointwise unbiased for F and has values in $[0, 1]$ but in general it may not be a monotone function of t.

The *Kaplan-Meier estimator* of F is given by

$$\widehat{F}(t) = 1 - \prod_{s \leq t} \left(1 - \frac{\#\{i : \widetilde{T}_i = s,\ D_i = 1\}}{\#\{i : \widetilde{T}_i \geq s\}}\right). \tag{2.8}$$

It can be understood intuitively by considering $\#\{i : \widetilde{T}_i \in \mathbf{ds},\ D_i = 1\}/\#\{i : \widetilde{T}_i \geq s\}$, for a small interval $\mathbf{ds} = [s, s + \mathrm{d}s)$, as an estimator of $\mathbb{P}\{T_i \in \mathbf{ds} \mid T_i \geq s\}$. The complement of this probability is therefore $\mathbb{P}\{T_i \geq s + \mathrm{d}s \mid T_i \geq s\}$. Multiplying over small intervals $[s, s + \mathrm{d}s)$ partitioning $[0, t + dt)$ produces $\mathbb{P}\{T_i > t\} = 1 - F(t)$.

More formally, introduce

$$\begin{aligned} N_n(t) &= \frac{1}{n}\#\{i : \widetilde{T}_i \leq t,\ D_i = 1\} \\ Y_n(t) &= \frac{1}{n}\#\{i : \widetilde{T}_i \geq t\} \\ \widehat{\Lambda}_n(t) &= \int_0^t \frac{\mathrm{d}N_n(s)}{Y_n(s)} \\ \Lambda(t) &= \int_0^t \frac{\mathrm{d}F(s)}{1 - F(s-)}. \end{aligned}$$

Then Λ is the *cumulative hazard* belonging to F, and $\widehat{\Lambda}_n$ is the Nelson-Aalen estimator of it. One can write

$$1 - F(t) = \prod_0^t (1 - \mathrm{d}\Lambda(s)),$$

$$1 - \widehat{F}_n(t) \quad = \quad \prod_0^t \Big(1 - \mathrm{d}\widehat{\Lambda}_n(s)\Big)$$

where $\prod$ denotes **product integration:**

$$\prod_0^t (1 + \mathrm{d}A(s)) = \lim_{\max |t_i - t_{i-1}| \to 0} \prod_{i=1}^m (1 + A(t_i) - A(t_{i-1}))$$

where $0 = t_0 < \ldots < t_m = t$ forms a partition of $(0, t]$.

If F is absolutely continuous with density f then defining $\lambda(t) = f(t)/(1 - F(t))$, the hazard rate, one has $\Lambda(t) = \int_0^t \lambda(s)\,\mathrm{d}s$ and

$$1 - F(t) = \prod_0^t (1 - \mathrm{d}\Lambda(s)) = \exp\left(-\Lambda(t)\right).$$

However if F has a discrete component the relation $\Lambda = -\log(1 - F)$ no longer holds.

Gill and Johansen [23] (see also [22]) survey the general theory of product integration and show in particular how asymptotic normality of $\sqrt{n}(\widehat{F}_n - F)$ follows from joint asymptotic normality of the empirical processes $\sqrt{n}(N_n - \mathbb{E}N_n)$, $\sqrt{n}(Y_n - \mathbb{E}Y_n)$ (the Donsker theorem), and the functional delta-method applied to the sequence of mappings from (N_n, Y_n) to $(N_n, 1/Y_n)$, then to $\widehat{\Lambda}_n = \int \frac{\mathrm{d}N_n}{Y_n}$, then to $1 - \widehat{F}_n = \prod(1 - \mathrm{d}\widehat{\Lambda}_n)$. Each of these mappings is differentiable in the sense of Hadamard or compact differentiability (see [21]). Differentiability together with joint weak covergence of the empirical processes guarantees that $\sqrt{n}(\widehat{F}_n - F)$ is asymptotically equivalent (in the sense that the supremum of the difference over any bounded interval converges in probability to zero) to the linear functional of the empirical processes

$$\sqrt{n}(1 - F(t)) \int_0^t \frac{\mathrm{d}N_n(s) - Y_n(s)\,\mathrm{d}\Lambda(s)}{(1 - \Delta\Lambda(s))y(s)}, \quad 0 \le t < \tau$$

where $y(s) = \mathbb{E}Y_n(s) = (1 - F(s-))(1 - H(s-))$ and $\tau = \inf\{t : \mathbb{E}Y_n(t) = 0\}$.

3. Kaplan-Meier estimator of the empty space function

3.1 Definition of estimator

Return to the setup of §2.1. Every point x in the window W contributes one possibly censored observation of the distance from an arbitrary point in space to the point process Φ; recall that $F(r) = \mathbb{P}\{\rho(x, \Phi) \le r\}$. The analogy with survival times is to regard $\rho(x, \Phi)$ as the 'distance (time) to failure' and $\rho(x, \partial W)$ as the censoring distance. The observation is censored if $\rho(x, \partial W) < \rho(x, \Phi)$.

From the data $\Phi \cap W$ we can compute $\rho(x, \Phi \cap W)$ and $\rho(x, \partial W)$ for each $x \in W$. Note that

$$\rho(x, \Phi) \wedge \rho(x, \partial W) = \rho(x, \Phi \cap W) \wedge \rho(x, \partial W) \tag{3.9}$$

(another application of the local knowledge principle [42, pp. 49, 233]) so that we can indeed observe $\rho(x, \Phi) \wedge \rho(x, \partial W)$ and $\mathbf{1}\{\rho(x, \Phi) \le \rho(x, \partial W)\}$ for each $x \in W$. Then the set

$$\{x \in W : \rho(x, \Phi) \wedge \rho(x, \partial W) \ge r\}$$

can be thought of as the set of points '*at risk* of failure at distance r', and

$$\{x \in W : \rho(x, \Phi) = r, \quad \rho(x, \Phi) \le \rho(x, \partial W)\}$$

are the '*observed* failures at distance r'. These two sets are analogous to the points counted in the empirical functions $Y_n(s), N_n(\,\mathrm{d}s)$ respectively in the definition of the Kaplan-Meier estimator.

Figure 1: Geometry of the reduced sample (left) and Kaplan-Meier (right) estimators. Spatial process Φ indicated by filled dots. For Kaplan-Meier, points x at risk are shaded, and observed failures constitute the curved boundary of the shaded region.

Geometrically the two sets can be written as (the topological closures of)

$$W_{\ominus r} \setminus \Phi_{\oplus r}$$

and

$$\partial\left(\Phi_{\oplus r}\right) \cap W_{\ominus r};$$

that is, *within the clipped window* $W_{\ominus r}$, consider the region outside the union of balls of radius r centred at points of the process, and the surface of this union of balls. See Figure 1.

Definition 1 *Let Φ be an a.s. stationary point process and $W \subset \mathbb{R}^k$ an inner regular compact set. The Kaplan-Meier estimator $\widehat{F}$ of the empty space function F of Φ, based on data $\Phi \cap W$, is defined by:*

$$\widehat{\Lambda}(r) = \int_0^r \frac{\left|\partial\left(\Phi_{\oplus s}\right) \cap W_{\ominus s}\right|_{k-1}}{\left|W_{\ominus s} \setminus \Phi_{\oplus s}\right|_k} \mathrm{d}s \tag{3.10}$$

$$\widehat{F}(r) = 1 - \prod_0^r \left(1 - \mathrm{d}\widehat{\Lambda}(s)\right) \tag{3.11}$$

$$= 1 - \exp(-\widehat{\Lambda}(r)) \tag{3.12}$$

where $|\cdot|_{k-1}$ denotes $k-1$ dimensional surface area (Hausdorff) measure. The reduced sample estimator $\widehat{F}^{\mathrm{rs}}$ of F is

$$\widehat{F}^{\mathrm{rs}}(r) = \frac{\left|W_{\ominus r} \cap \Phi_{\oplus r}\right|_k}{\left|W_{\ominus r}\right|_k} \tag{3.13}$$

i.e. this is the border correction [38].

Here $\widehat{F}$ is the Kaplan-Meier estimator based on the continuum of observations generated by all $x \in W$. Note that the estimator is a proper distribution function and is even absolutely continuous, with hazard rate

$$\widehat{\lambda}(r) = \frac{\left|\partial\left(\Phi_{\oplus r}\right) \cap W_{\ominus r}\right|_{k-1}}{\left|W_{\ominus r} \setminus \Phi_{\oplus r}\right|_k}. \tag{3.14}$$

3.2 Unbiasedness and continuity

Our first theorem will be an 'unbiasedness' result for the hazard rate estimator (3.14).

Theorem 1 *Let Φ be any stationary point process with intensity $\alpha < \infty$. Then*

(a) the empty space function F is absolutely continuous;

(b) the hazard rate of F equals

$$\lambda(r) = \frac{\mathbb{E}\,|W \cap \partial\,(\Phi_{\oplus r})|_{k-1}}{\mathbb{E}\,|W \setminus \Phi_{\oplus r}|_k}$$

for any compact, inner regular window W;

(c) the Kaplan-Meier estimator (3.14) of λ is ratio-unbiased.

Thus our estimator $\widehat{F}(r)$ respects the smoothness of the true empty space function F. The reduced-sample estimator (3.13) is not even necessarily monotone.

To prove the Theorem we need some regularity results. The first is an example of Crofton's perturbation or 'moving manifold' formula [2, 9].

Lemma 1 *Let $Z \subset \mathbb{R}^k$ be a compact, inner regular set and $A \subset \mathbb{R}^k$ any nonempty closed set. Then for $r \geq 0$*

$$|Z \cap A_{\oplus r}|_k = |Z \cap A|_k + \int_0^r |Z \cap \partial\,(A_{\oplus s})|_{k-1}\,\mathrm{d}s;$$

the integrand is Lebesgue measurable and integrable.

Proof : The function $f(x) = \rho(x, A)$ is Lipschitz: $f(y) \leq f(x) + ||x - y||$, and hence a.e. differentiable, and its approximate Jacobian has determinant 1. Since Z is inner regular it is $(\mathcal{H}^d, d)$ rectifiable and measurable in the sense of Federer [19, p. 251]. Applying the coarea formula [19, p. 258] gives the result when we recognise $\{x : f(x) = s\} = \partial\,(A_{\oplus s})$. □

The next lemma shows that the integrand $|Z \cap \partial\,(\Phi_{\oplus r})|_{k-1}$ is uniformly bounded (over possible realisations of Φ) in such a way that dominated convergence can be used to justify interchanges of expectation and integration or differentiation with respect to r.

Lemma 2 *For any inner regular compact set Z*

$$|Z \cap \partial\,(\Phi_{\oplus r})|_{k-1} \leq \frac{k}{r}\,|Z_{\oplus r}|_k \;\wedge\; \omega_k r^{k-1}\Phi(Z_{\oplus r}) \quad a.s.$$

where $\omega_k = |\partial B(0,1)|_{k-1} = 2\pi^{k/2}/\Gamma(k/2)$.

Proof : The second term on the right is a trivial bound on the left side, since $\omega_k r^{k-1} = |\partial B(0,r)|_{k-1}$. For the first term, fix a realization of Φ and let $x_i, i = 1, \ldots, m$ be the a.s. distinct points in $\Phi \cap Z_{\oplus r}$. Thus

$$Z \cap \partial\,(\Phi_{\oplus r}) = Z \cap \bigcup_{i=1}^{m} B(x_i, r).$$

We also have

$$\bigcup_{1}^{m} B(x_i, r) = \{y \in \mathbb{R}^k : \min_i ||y - x_i|| \leq r\}.$$

Construct the Dirichlet cells formed by the x_i,

$$C_i = C(x_i|x_1,\ldots,x_m) = \{y \in \mathbb{R}^k : ||y-x_i|| = \min_j ||y-x_j||\}.$$

Split the surface $\partial(\Phi_{\oplus r})$ into pieces of surface within each cell:

$$\begin{aligned}\partial\left(\bigcup_{i=1}^m B(x_i,r)\right) &= \bigcup_{i=1}^m \{y \in \mathbb{R}^k : ||y-x_i|| = r, \min_j ||y-x_j|| = r\} \\ &= \bigcup_{i=1}^m (C_i \cap \partial B(x_i,r)) \\ &= \bigcup_{i=1}^m D_i, \text{ (say)}\end{aligned}$$

That is, within the cell defined by the point x_i the surface $\partial(\Phi_{\oplus r})$ is just the intersection of the surface of the sphere $B(x_i,r)$ with the cell. The pieces D_i are measure-disjoint since $D_i \cap D_j = \partial B(x_i,r) \cap \partial B(x_j,r)$ is a $k-2$ dimensional compact manifold (or empty), which has zero $k-1$ dimensional Hausdorff measure. Thus

$$\left|\partial\left(\bigcup_{i=1}^m B(x_i,r)\right) \cap Z\right|_{k-1} = \sum_{i=1}^m |D_i \cap Z|_{k-1}. \tag{3.15}$$

The intuitive argument is now as follows: any line segment joining x_i to a point on the corresponding surface piece $D_i \cap Z$ is contained entirely within the Dirichlet cell C_i, since this is convex. The union F_i of these segments is a solid angular cone of the sphere $B(x_i,r)$, and its outside surface equals k/r times its volume. The cones F_i are volume-disjoint since the C_i are, so the sum of the cone volumes is bounded by the volume of $Z_{\oplus r}$, yielding the result.

Formally, for each i define $f_i : D_i \times [0,1] \to \mathbb{R}^k$ by $f_i(y,t) = (1-t)x_i + ty$, so that $f_i(y,t)$ is a point on the radius joining y to the centre x_i. Define $F_i = f_i((D_i \cap Z) \times [0,1])$ and apply the coarea formula to get

$$|F_i|_k = \frac{r}{k}|D_i|_{k-1}.$$

Since C_i and $B(x_i,r)$ are convex, the range of f_i is contained in $C_i \cap B(x_i,r)$; hence the F_i are volume-disjoint subsets of Z, and

$$\sum_{i=1}^m |F_i|_k \le |Z|_k.$$

Combining these,

$$\sum_{i=1}^m |D_i \cap Z|_{k-1} \le \sum_{i=1}^m |D_i|_{k-1} = \sum_{i=1}^m |F_i|_k \le |Z_{\oplus r}|_k.$$

□

Lemma 3 *Let Φ be a point process in $\mathbb{R}^k$ and $W \subset \mathbb{R}^k$ a inner regular compact set. Then for fixed r, $|W \cap \Phi_{\oplus r}|_k$ and $|W_{\ominus r} \cap \Phi_{\oplus r}|_k$ are a.s. finite r.v.'s on the same probability space, and the following identities hold a.s.:*

$$|W \cap \Phi_{\oplus r}|_k = \int_0^r |W \cap \partial(\Phi_{\oplus s})|_{k-1}\, \mathrm{d}s \tag{3.16}$$

$$|\{x \in W : \rho(x,\Phi) \le \rho(x,\partial W),\ \rho(x,\Phi) \le r\}|_k = \int_0^r |W_{\ominus s} \cap \partial(\Phi_{\oplus s})|_{k-1}\, \mathrm{d}s \tag{3.17}$$

$$|W_{\ominus r} \setminus \Phi_{\oplus r}|_k = |W|_k - \int_0^r |\partial(W_{\ominus s} \setminus \Phi_{\oplus s})|_{k-1}\, \mathrm{d}s \tag{3.18}$$

where the integrands are well defined r.v.'s for each fixed s and are a.s. measurable and integrable functions of s.

Proof : By standard results of stochastic geometry [32, pp. 9, 19, 47], $\Phi_{\oplus r}$ is a random closed set, so that $\partial(\Phi_{\oplus r})$ is a random closed set, the intersections with W are random closed sets, and their measures are random variables. Now apply lemma 1 to each realization to get (3.16).

For the second result, we note that $\rho(x, \partial W)$ is continuous in x and $\rho(x, \Phi)$ is a random upper-semicontinuous (u.s.c.) function, so that $\rho(x, \Phi) - \rho(x, \partial W)$ is also a random u.s.c. function and $Z = \{x \in W : \rho(x, \Phi) \leq \rho(x, \partial W)\}$ is a random closed set. Recognise the left side of (3.17) as the volume of $Z \cap \Phi_{\oplus r}$, and the integrand as the surface area of $Z \cap \partial(\Phi_{\oplus s})$. Measurability arguments remain valid for the random closed set Z and we apply lemma 1 to each realization.

For the third result, observe that $W_{\ominus r} \setminus \Phi_{\oplus r} = W \setminus (\partial W \cup \Phi)_{\oplus r}$ and use the same technique as for the first identity. □

Proof of Theorem 1: By Fubini (Robbins' theorem [32, p. 47])

$$\begin{aligned} \mathbb{E}\,|W \cap \Phi_{\oplus r}|_k &= \mathbb{E} \int_W \mathbf{1}\{x \in \Phi_{\oplus r}\}\, \mathrm{d}x \\ &= \int_W \mathbb{P}\{x \in \Phi_{\oplus r}\}\, \mathrm{d}x \\ &= F(r)\ |W|_k\,. \end{aligned} \tag{3.19}$$

Since $|W \cap \Phi_{\oplus r}|_k$ is absolutely continuous with derivative given in Lemma 3, and bounded as in Lemma 2, its expectation is absolutely continuous too, with derivative

$$f(r)\,|W|_k = \mathbb{E}\,|W \cap \partial(\Phi_{\oplus r})|_{k-1}\,. \tag{3.20}$$

But complementarily to (3.19)

$$\mathbb{E}\,|W \setminus \Phi_{\oplus r}|_k = (1 - F(r))\,|W|_k\,. \tag{3.21}$$

Dividing (3.20) by (3.21) we obtain the first result of the theorem. The rest follows by replacing W with $W_{\ominus r}$. □

3.3 Discretisation and classical Kaplan-Meier estimator

In practice one would not actually compute the surface areas and volumes for each $s \in [0, r]$ in order to estimate $F(r)$. Rather one would discretize W or $[0, r]$ or both. For standard estimators of F, one typically discretises W on a regular lattice (see [12]) although Lotwick [31] showed the areas can be computed analytically.

A natural possibility here is to discretize W by superimposing a regular lattice L of points, calculating for each $x_i \in W \cap L$ the censored distance $\rho(x_i, \Phi) \wedge \rho(x_i, \partial W)$ and the indicator $\mathbf{1}\{\rho(x_i, \Phi) \leq \rho(x_i, \partial W)\}$. Then one would calculate the ordinary Kaplan-Meier estimator based on this finite dataset.

Our next result is that as the lattice becomes finer, the discrete Kaplan-Meier estimates converge to the 'theoretical' continuous estimator $\widehat{F}$.

Theorem 2 *Let $\widehat{F}_L$ be the Kaplan-Meier estimator (2.8) computed from the discrete observations at the points of $W \cap L$, where $L = \epsilon M + b$ is a rescaled, translated copy of a fixed regular lattice M. Let*

$$R = \inf\{r \geq 0 : W_{\ominus r} \cap \Phi_{\oplus r} = \emptyset\}.$$

Then as the lattice mesh ϵ converges to zero, $\widehat{F}_L(r) \to \widehat{F}(r)$ for any $r < R$. The convergence is uniform on any compact interval in $[0, R)$.

Proof of Theorem: For any inner regular compact set $A \subseteq \mathbb{R}^k$ one can easily show that

$$\epsilon^d \#(A \cap L) \to c\,|A|_k \quad \text{as } \epsilon \to 0$$

where c is a finite positive constant depending on M. Hence the functions

$$N_L(r) = \frac{\#\left(L \cap \{x \in W : \rho(x,\Phi) \le \rho(x,\partial W),\ \rho(x,\Phi) \le r\}\right)}{\#(L \cap W)}$$

and

$$Y_L(r) = \frac{\#\left(L \cap (W_{\ominus r} \setminus \Phi_{\oplus r})\right)}{\#(L \cap W)}$$

converge pointwise to

$$N(r) = \frac{|\{x \in W : \rho(x,\Phi) \le \rho(x,\partial W),\ \rho(x,\Phi) \le r\}|_k}{|W|_k} \tag{3.22}$$

and

$$Y(r) = \frac{|W_{\ominus r} \setminus \Phi_{\oplus r}|_k}{|W|_k} \tag{3.23}$$

respectively. Since $N_L(r)$ is increasing in r and the limit is continuous, $N_L \to N$ uniformly in r. Recalling (3.18) of Lemma 3, and using the argument of Lemma 2 to bound the integrand, one sees that Y_L converges uniformly in r.

Given (3.17) and by continuity of the mapping from (N, Y) to $\widehat{\Lambda}_n = \int \mathrm{d}N/Y$ (see [21, Lemma 3] which even establishes a form of differentiability) the discrete Nelson-Aalen estimators

$$\widehat{\Lambda}_L = \int \frac{\mathrm{d}N_L}{Y_L}$$

converge to $\widehat{\Lambda}$. By continuity of the product-integral mapping [23, Theorem 7] $\widehat{F}_L$ converges to $\widehat{F}$. □

A similar, simpler argument establishes that the continuous reduced-sample estimator (3.13) is the uniform limit of the discrete reduced sample estimator (2.7) under the conditions of Theorem 2.

It does not seem to be widely known in spatial statistics (cf. [8, p. 764], [12, 15]) that computation of the distances $\rho(x, \Phi \cap W), \rho(x, \partial W)$ for all points x in a fine rectangular lattice can be performed very efficiently using the *distance transform* algorithm of image processing [6, 7, 39, 40] at the price of accepting a (good) discrete approximation to the true Euclidean metric $||\cdot||$ in the definition of ρ at (2.1). Thus the reduced-sample and Kaplan-Meier estimators are equivalent in computational cost when a fine grid is used.

It is often of interest to replace Euclidean distance $||\cdot||_2$ by another metric, either for computational convenience as above, or in order to obtain different information about the process Φ [46], particularly in three dimensions [4]. It is possible to replace $||\cdot||_2$ by another vector space norm $||\cdot||$ in the above results, provided the unit ball of $||\cdot||$ is a polyhedron (in $\mathbb{R}^2$ a polygon) scaled so that

$$\sup\{\frac{||x||}{||x||_2} : ||x||_2 \le 1\} = 1. \tag{3.24}$$

Examples are $||\cdot||_\infty$, $||\cdot||_1/\sqrt{2}$ and continuous versions of the standard 'chamfer metrics' [6, 7] used in the distance transform algorithm. Redefine the ball of radius r as $B(x,r) = \{y : ||x-y|| \le r\}$ and the distance function ρ of (2.1) in terms of $||\cdot||$. Then it can be shown that

$$\frac{\mathrm{d}}{\mathrm{d}r}\,|B(x,r)|_k = |\partial B(x,r)|_{k-1}$$

and that Lemma 1 remains true when A is a finite set but not in general. Hence Theorems 1 and 2 continue to hold for the Kaplan-Meier estimator with respect to this more general metric.

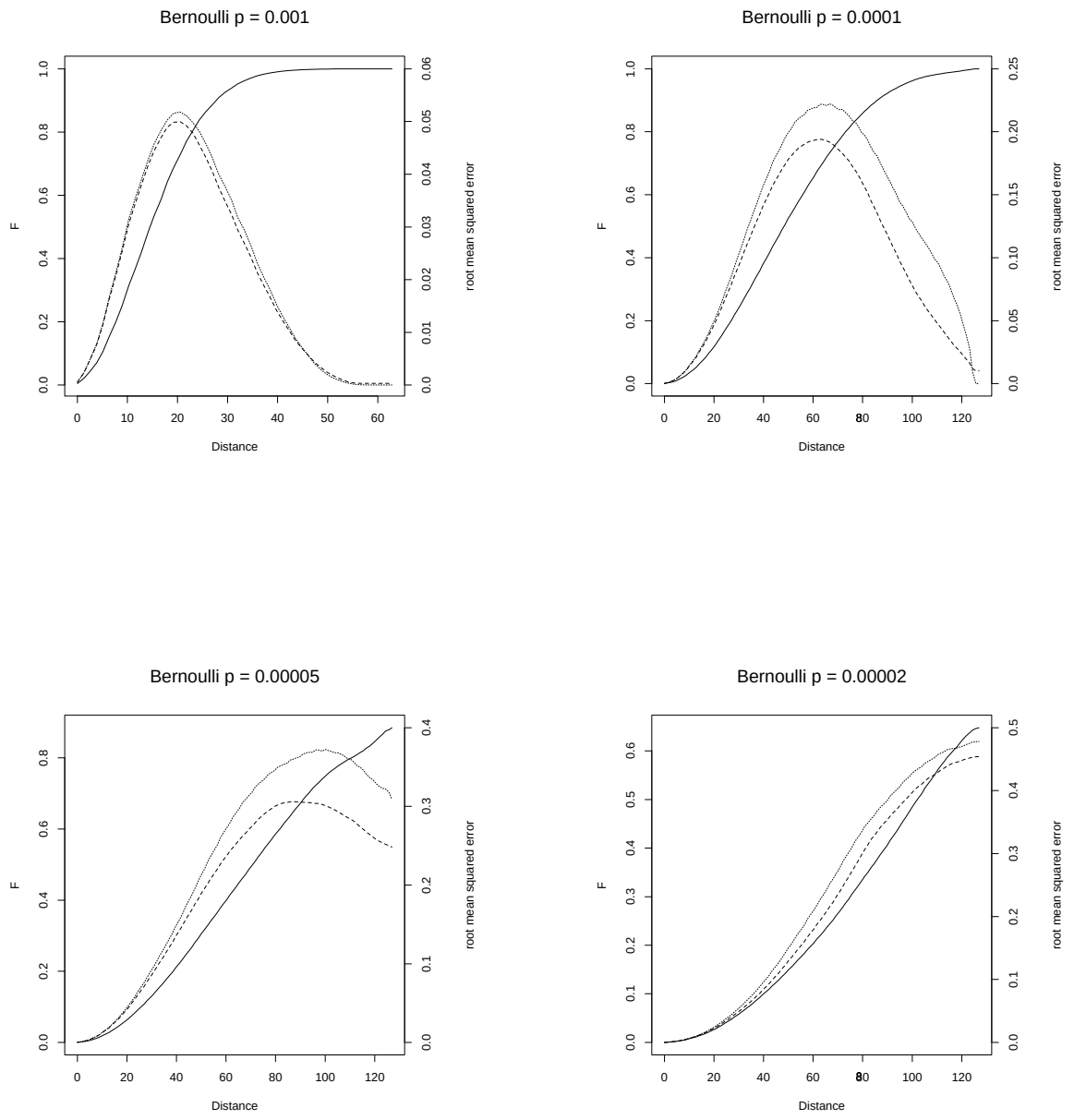

Figure 2: Root mean square error comparison for simulations of Poisson process. *Dotted lines:* reduced sample estimator; *dashed lines:* Kaplan-Meier estimator; *solid lines:* estimand F. Note use of different scales for rmse and F.

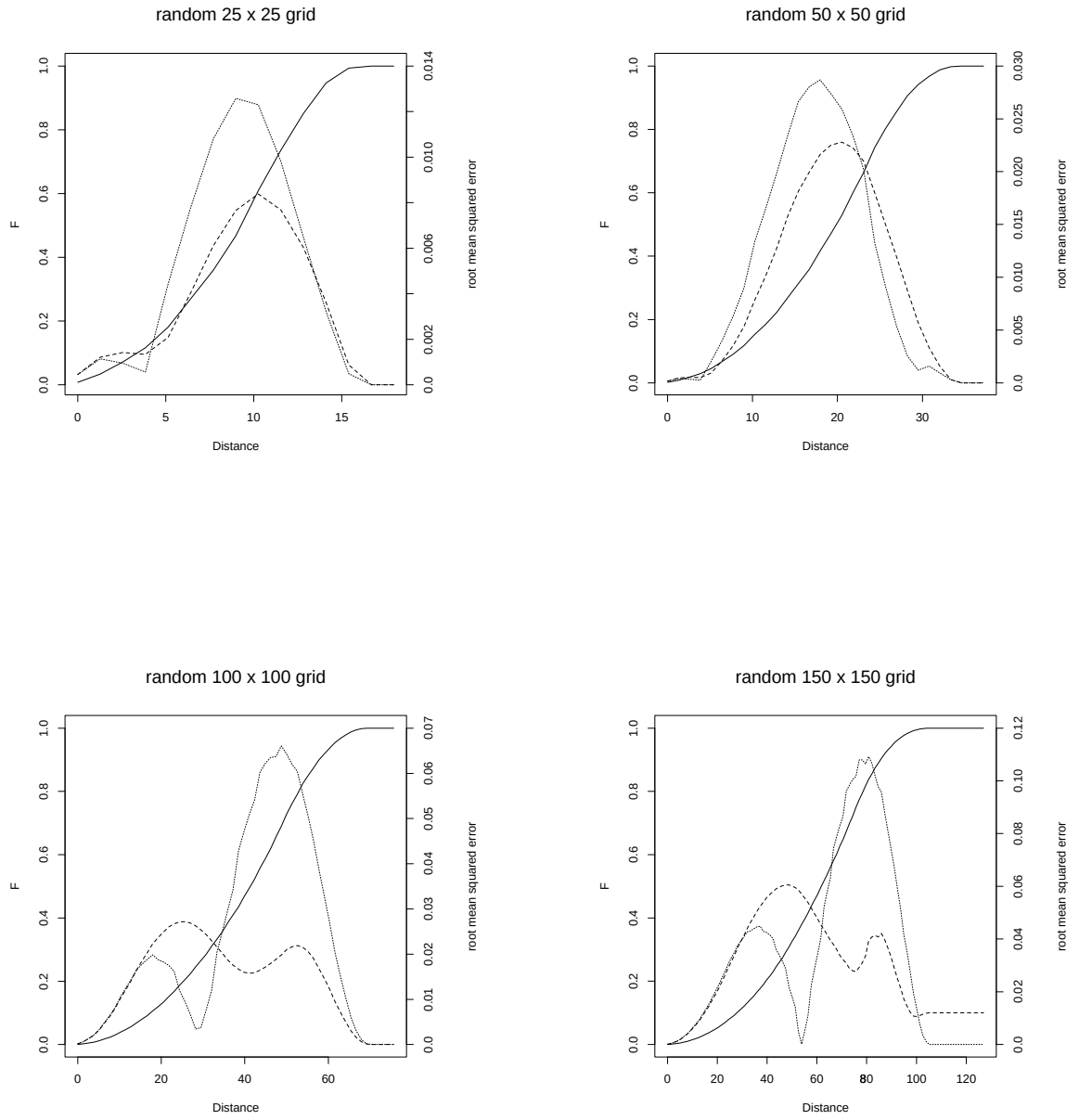

Figure 3: RMSE comparison for simulations of randomly translated grids. *Dotted lines:* reduced sample estimator; *dashed lines:* Kaplan-Meier estimator; *solid lines:* estimand F.

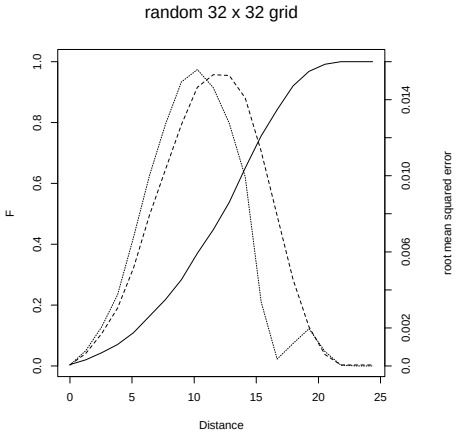

Figure 4: Extreme resonance case of previous Figure

3.4 Simulations

We have compared the performance of the Kaplan-Meier and reduced-sample estimators of F in Monte Carlo simulations of a Poisson process and of a randomly-translated square grid.

Both processes were simulated as binary images on a discrete 256×256 grid. For the Poisson process of intensity α the pixel values were i.i.d. Bernoulli variables with $p = \alpha/(256 \times 256)$. The randomly-translated square grid with spacing s is generated by choosing a starting position uniformly distributed in an $s \times s$ subrectangle and extending periodically. We generated 100 realizations of each of Bernoulli $p = 0.001$, 0.0001, 0.00005, 0.00002 (corresponding to expected number of points in the window $= \alpha\,|W|_k = 65.5$, 6.55, 3.3, 1.3) and randomly-translated grids of side $s = 25$, 32, 50, 100 and 150 (expected number of points in window = 105, 64, 26, 6.6 and 2.9). For each realization the distance transform was computed in the chamfer(5,7) metric of Borgefors [6] and the two estimators derived.

Figure 2 compares the sample standard deviation of the reduced-sample estimator with the root mean square error of the Kaplan-Meier estimator (since the reduced sample estimator is unbiased pointwise for F) for several intensities of the Poisson process. The Kaplan-Meier estimator appears to be uniformly more efficient.

Figure 3 is a similar comparison for a randomly-translated grid. Here the comparison is not uniformly favorable to the Kaplan-Meier estimator, although it is generally better. One can attribute this to periodic effects. For certain values of r the reduced-sample estimator is exact; near these values it has small variance. The mse of the Kaplan-Meier estimator oscillates for similar reasons. An extreme case is Figure 4 in which the grid dimension 32 is a divisor of the window dimension 256.

Typical two-dimensional spatial point pattern datasets in [11, 37, 38] consist of 10 to 100 points, so the behaviour of the Kaplan-Meier estimator lies somewhere between the first two cases in Figure 2 or the first three cases in Figure 3.

4. Asymptotic properties of estimators of F

4.1 Large sample theory

It is usual in spatial statistics to consider the limiting situation in which the window W is a convex set that expands to fill $\mathbb{R}^k$ (see e.g. [3, 26, 27, 44] and the discussion in [8, p. 480]). Under the additional assumption of ergodicity it is clear that the reduced-sample and Kaplan-Meier estimators of F are pointwise consistent as $W \uparrow \mathbb{R}^k$.

However, a more relevant 'large sample' situation is one in which the edge effect problem remains equally as severe as in the 'small sample' case, e.g. so that the proportion of the window within distance r from the boundary stays appreciable. We shall consider the situation when there are $n \to \infty$ independent replicated observations Φ_i of a given process Φ within a fixed window W. The pooled statistics $\widehat{F}$ and $\widehat{F}^{\mathrm{rs}}$ are obtained, not as the mean of the separate estimators for each window, but by analogues of (3.11), (3.13) in which the numerators and denominators of (3.10) and (3.13) are replaced by the sums of these expressions over all realizations Φ_i. Equivalently if Φ satisfies a mixing assumption we may consider observation of the same point process through a single window W which is the union of n small and distantly spread windows of fixed size and shape. The latter is important in applications, e.g. [4].

Asymptotics as $n \to \infty$ are now easy to derive, cf. [23, section 4.2], using a CLT of Giné and Zinn.

Lemma 4 *Let $\Phi_1, \Phi_2, \ldots$ be i.i.d. realizations of an a.s. stationary point process Φ with finite positive intensity α. Fix a* **convex** *compact inner regular set $W \subset \mathbb{R}^k$ and let $N_i(r), Y_i(r)$ for $i = 1, 2, \ldots$ be the corresponding 'fraction of failures' and 'fraction at risk' processes (3.22)-(3.23) for Φ_i in W. Then*

$$n^{1/2} \left(n^{-1} \sum_{i=1}^{n} N_i(r) - \mathbb{E}\, N(r) \right) \tag{4.25}$$

and

$$n^{1/2} \left(n^{-1} \sum_{i=1}^{n} Y_i(r) - \mathbb{E}\, Y(r) \right) \tag{4.26}$$

converge weakly in $C[0, \tau]$ to Gaussian processes, where $\tau > 0$ satisfies $F(\tau) < 1$, i.e. $\mathbb{E}Y(\tau) > 0$.

Proof : N and Y are monotone and uniformly bounded by 1. This gives a LLN uniformly on $[0, \tau]$. By a result of Giné and Zinn [24, theorem 7.4] (see [1, p. 296] and [48]) the CLT holds if we can show that the centred process $Z = Y - \mathbb{E}Y$ satisfies

$$\mathbb{E}\,|Z(t) - Z(s)| \le K\,|t - s| \tag{4.27}$$

for all $s, t \in [0, \tau]$ for some constant $K < \infty$, and similarly for $M = N - \mathbb{E}N$. Now

$$\frac{d}{\mathrm{d}r} |W_{\ominus r} \setminus \partial(\Phi_{\oplus r})|_k = |W_{\ominus r} \cap \partial(\Phi_{\oplus r})|_{k-1} + |\partial(W_{\ominus r}) \setminus \Phi_{\oplus r}|_{k-1} \quad \text{a.s.}$$

by another application of the 'moving manifold' technique [2]. The first term on the right is bounded as in Lemma 2 and the second term is dominated by $|\partial(W_{\ominus r})|_{k-1}$. Since W is convex we have [41] $|\partial(W_{\ominus r})|_{k-1} < |\partial W|_{k-1}$ and for $0 \le s < t \le \tau$

$$\begin{aligned} |W|_k\ |Y(t) - Y(s)| &\le \omega_k \Phi(W) \int_s^t r^{k-1}\,\mathrm{d}r + |t - s|\ |\partial W|_{k-1} \\ &\le \left[\omega_k \Phi(W) \tau^{k-1} + |\partial W|_{k-1}\right] |t - s| \quad \text{a.s.} \end{aligned}$$

Since $\mathbb{E}\Phi(W) = \alpha\,|W|_k$ we can apply this inequality to bound both $\mathbb{E}|Y(s) - Y(t)|$ and $|\mathbb{E}Y(s) - \mathbb{E}Y(t)|$, and it follows that Z satisfies (4.27) and hence the CLT.

Turning to the 'fraction of failures' process N we have

$$\begin{aligned} |W|_k \frac{d}{\mathrm{d}r} N(r) &= |\{x \in W : \rho(x,\Phi) \le \rho(x,\partial W),\ \rho(x,\Phi) \le r\}|_{k-1} \\ &\le |W \cap \Phi_{\oplus r}|_{k-1} \\ &\le \omega_k r^{k-1} \Phi(W_{\oplus r}) \end{aligned}$$

and again M satisfies (4.27). □

A joint CLT for (N, Y) follows immediately, since the families (4.25),(4.26) are jointly tight and the finite dimensional distributions satisfy a joint CLT. The functional delta-method [21, Theorem 3] together with differentiability of the product-integration mapping [23, Theorem 8] now imply that $\widehat{F}_n$ is consistent and converges weakly at rate $n^{1/2}$ to a Gaussian process, with linear approximation

$$\widehat{F}_n(r) - F(r) = \frac{1}{n} \sum_{i=1}^{n} I(\widehat{F}, \Phi_i, r) + \mathbf{o}_p(n^{-1/2}) \tag{4.28}$$

uniformly in $0 \le r \le \tau$, where I is the 'influence function'

$$I(\widehat{F}, \Phi, r) = (1 - F(r)) \int_0^r \frac{|W_{\ominus s} \cap \partial(\Phi_{\oplus s})|_{k-1} - |W_{\ominus s} \setminus \Phi_{\oplus s}|_k \lambda(s)}{y(s)} \, \mathrm{d}s \tag{4.29}$$

and $y(s) = \mathbb{E}|W_{\ominus s} \setminus \Phi_{\oplus s}|_k = (1 - F(s))\,|W_{\ominus s}|_k$ and τ satisfies $y(\tau) > 0$.

4.2 Calculations for the sparse Poisson limit

Here we consider asymptotic variances of the Kaplan-Meier and reduced sample *influence functions* on a fixed window W for a Poisson process whose intensity α is sent to zero. This is the asymptotic variance of the Kaplan-Meier and reduced sample *estimators* in the large-sample case of many independent replicates of observation of a fixed-intensity Poisson process through an asymptotically small window. 'Many replicates' justifies looking at the influence function, and the case of vanishing intensity with a fixed window is the same as a vanishing window, fixed intensity.

There are just two situations to consider: (i) no random point in W, with probability $e^{-\alpha|W|_k} = 1 + \mathcal{O}(\alpha)$, and (ii), one random point in W at a position x uniformly distributed over W, occurring with probability $\alpha\,|W|_k\, e^{-\alpha|W|_k} = \alpha\,|W|_k + \mathcal{O}(\alpha^2)$; the remaining possibilities have probability $\mathcal{O}(\alpha^2)$.

The influence function for Kaplan-Meier (4.29) is the difference of two terms: a part depending on surface areas at some distances from a point of Φ, and a part depending on volumes at risk, and involving the hazard rate of the empty space function. In case (i) only the second part is present and is of order α; in case (ii) the first part is also present and is of constant order.

The empty space function for the Poisson process is

$$F(r) = 1 - e^{-\alpha|B_r|_k}$$

and its hazard rate is

$$\lambda(r) = \frac{d}{dr}\left[-\log(1 - F(r))\right] = \alpha\,|\partial B_r|_{k-1}$$

where $B_r = B(0, r)$ is a ball of radius r in the Euclidean metric, so that $|B_r|_k = r^d \omega_d / d$ and $|\partial B_r|_{k-1} = r^{d-1}\omega_d$. The expected volume "at risk" is

$$y(r) = (1 - F(r))\,|W_{\ominus r}|_k\,.$$

In case (i), no random points in W, the influence function (4.29) for Kaplan-Meier is therefore

$$I(\widehat{F}, \emptyset, r) = (1 - F(r)) \left\{ - \int_0^r \frac{\alpha\,|\partial B_s|_{k-1}\,|W_{\ominus s}|_k}{|W_{\ominus s}|_k\, e^{-\alpha|B_s|_k}} \, \mathrm{d}s \right\}$$

$$
\begin{aligned}
&= (1-F(r))\left\{-\int_0^r \alpha\,|\partial B_s|_{k-1}\, e^{\alpha|B_s|_k}\,\mathrm{d}s\right\}\\
&= e^{-\alpha|B_r|_k}\left[e^{\alpha|B_s|_k}\right]_0^r\\
&= -\left(1-e^{-\alpha|B_r|_k}\right)\\
&= -\alpha\,|B_r|_k + \mathcal{O}\left(\alpha^2\right).
\end{aligned}
$$

In case (ii) the influence function is

$$
\begin{aligned}
I(\widehat{F},\{x\},r) &= (1-F(r))\left\{\int_0^r \frac{|\partial B(x,s)\cap W_{\ominus s}|_{k-1} - \alpha\,|\partial B_s|_{k-1}\,|W_{\ominus s}\setminus B(x,s)|_k}{|W_{\ominus s}|_k\, e^{-\alpha|B_s|_k}}\,\mathrm{d}s\right\}\\
&= e^{-\alpha|B_r|_k}\int_0^r \frac{|\partial B(x,s)\cap W_{\ominus s}|_{k-1}}{|W_{\ominus s}|_k\, e^{-\alpha|B_s|_k}}\,\mathrm{d}s + \mathcal{O}(\alpha)\\
&= \int_0^r \frac{|\partial B(x,s)\cap W_{\ominus s}|_{k-1}}{|W_{\ominus s}|_k}\,\mathrm{d}s + \mathcal{O}(\alpha).
\end{aligned}
$$

To check this, observe that the expected influence function is therefore, to first order in α,

$$
\begin{aligned}
\mathbb{E}\,I(\widehat{F},\Phi,r) &\sim -\alpha\,|B_r|_k + \alpha\,|W|_k\,\mathbb{E}\left(\int_0^r \frac{|\partial B(x,s)\cap W_{\ominus s}|_{k-1}}{|W_{\ominus s}|_k}\,\mathrm{d}s\right)\\
&= \alpha\left(-|B_r|_k + |W|_k\int_0^r \frac{\mathbb{E}\,|\partial B(x,s)\cap W_{\ominus s}|_{k-1}}{|W_{\ominus s}|_k}\,\mathrm{d}s\right).
\end{aligned}
$$

By a well-known result of integral geometry [41, p. 97] the expectation in the numerator is

$$
\mathbb{E}\,|\partial B(x,s)\cap W_{\ominus s}|_{k-1} = \frac{|\partial B_s|_{k-1}\;|W_{\ominus s}|_k}{|W|_k}
$$

so that the expected influence function is

$$
\begin{aligned}
\mathbb{E}\,I(\widehat{F},\Phi,r) &\sim \alpha\left(-|B_r|_k + |W|_k\int_0^r \frac{|\partial B_s|_{k-1}\;|W_{\ominus s}|_k}{|W_{\ominus s}|_k\;|W|_k}\,\mathrm{d}s\right)\\
&= \alpha\left(-|B_r|_k + \int_0^r |\partial B_s|_{k-1}\,\mathrm{d}s\right)\\
&= 0.
\end{aligned}
$$

What we are really looking for, the variance of the influence function, is to first order just the expectation of the square of the 'area of failures' term from case (ii) (since case (i) is now $\mathcal{O}\left(\alpha^2\right)$):

$$
\operatorname{var} I(\widehat{F},\Phi,r) \sim \alpha\,|W|_k\,\mathbb{E}\left(\int_0^r \frac{|\partial B(x,s)\cap W_{\ominus s}|_{k-1}}{|W_{\ominus s}|_k}\,\mathrm{d}s\right)^2 \tag{4.30}
$$

For the reduced sample estimator, the calculations are similar but easier. In case (i) the estimator is identically zero; in case (ii) it is

$$
\widehat{F}^{\mathrm{rs}}(r) = |B(x,r)\cap W_{\ominus r}|_k \,/\, |W_{\ominus r}|_k\,.
$$

Since $F(r) = 1-e^{-\alpha|B_r|_k} = \alpha\,|B_r|_k + \mathcal{O}\left(\alpha^2\right)$ the influence function (= estimator − estimand in this linear case) is in case (i)

$$
I(\widehat{F}^{\mathrm{rs}},0,r) = -\alpha\,|B_r|_k + \mathcal{O}\left(\alpha^2\right);
$$

in case (ii)

$$I(\widehat{F}^{\rm rs}, \{x\}, r) = \frac{|B(x,r) \cap W_{\ominus r}|_k}{|W_{\ominus r}|_k} + \mathcal{O}(\alpha).$$

The expectation of the influence function is, to first order,

$$\begin{aligned} \mathbb{E}\, I(\widehat{F}^{\rm rs}, \Phi, r) &\sim -\alpha\, |B_r|_k + \alpha\, |W|_k\, \mathbb{E}(\,|B(x,r) \cap W_{\ominus r}|_k)\,/\,|W_{\ominus r}|_k \\ &= \alpha \left\{ -\,|B_r|_k + |W|_k\, \frac{|B_r|_k\, |W_{\ominus r}|_k\,/\,|W|_k}{|W_{\ominus r}|_k} \right\} \\ &= 0, \end{aligned}$$

as should be the case. Here we have used another well-known integral geometric identity. The variance is

$$\text{var}\; I(\widehat{F}^{\rm rs}, \Phi, r) = \alpha\, |W|_k\, \mathbb{E}\left(\left(\frac{|B(x,r) \cap W_{\ominus r}|_k}{|W_{\ominus r}|_k} \right)^2 \right) + \mathcal{O}\left(\alpha^2\right)$$

The conclusion is that we must calculate and compare the expected squared values of

$$\int_0^r \frac{|\partial B(x,s) \cap W_{\ominus s}|_{k-1}}{|W_{\ominus s}|_k}\, \mathrm{d}s$$

and

$$\frac{|B(x,r) \cap W_{\ominus r}|_k}{|W_{\ominus r}|_k}$$

for $x \sim \text{Uniform}(W)$.

For convenience in calculation, we will take W to be the d-dimensional unit cube centred at $(\frac{1}{2}, \ldots, \frac{1}{2})$, and replace the Euclidean metric $||\cdot||$ by the L_∞ metric in the definition (2.1) of ρ and $A_{\oplus r}, A_{\ominus r}$ (see comments at the end of section 3.3). Thus F becomes the 'empty square space' function obtained by replacing $B(x,r)$ by a cube $B_\infty(x,r)$ of centre x and side length $2r$.

We need to consider all possible ways the cubes $B_\infty(x,r)$ and $W_{\ominus r}$ intersect. For fixed $x \in W$, as r increases, initially $B_\infty(x,r)$ is entirely contained in $W_{\ominus r}$, then one-by-one the faces of $B_\infty(x,r)$ pass through faces of $W_{\ominus r}$. By symmetry we may take x uniformly distributed on the simplex $\{x : x_1 < x_2 < \cdots < x_d < \frac{1}{2}\}$. The different transitions then occur as the value $2r$ passes through x_1, then x_2, ..., then x_d; and then as $(1-2r)$ passes through $x_d, x_{d-1}, \ldots, x_1$. The latter cases are only relevant when $r > 1/4$. After expressing the volume and surface area contributions in terms of the x_i in each case, we integrate over r (for Kaplan-Meier only) and then over x.

In one dimension with $W = [-1/2, 1/2]$ the variance of $n^{1/2}(\widehat{F}_W(r) - F(r))$ is approximately (ignoring terms of order $O(\alpha^2)$) equal to α times the following expression:

$$\begin{cases} 2r + (1-4r)\log(1-2r) - \frac{1}{2}(\log(1-2r))^2 & \text{for } 0 \le r \le \frac{1}{4} \\ 2r + \int_{\frac{1}{2}}^{2r} \log u \log(1-u) du - 2r \log 2r \log(1-2r) & \text{for } \frac{1}{4} \le r < \frac{1}{2} \end{cases}$$

For the reduced sample estimator $|\Phi_{\oplus r} \cap W_{\ominus r}|_k\,/\,|W_{\ominus r}|_k$, the corresponding formula is

$$\begin{cases} 4r^2(1 - \frac{8r}{3})/(1-2r)^2 & \text{for } 0 \le r \le \frac{1}{4} \\ (8r-1)/3 & \text{for } \frac{1}{4} \le r < \frac{1}{2} \end{cases}$$

These functions are plotted in Figure 5 together with the corresponding curves for two and three dimensions; the latter have been calculated (by Mathematica) with a mixture of computer algebra and numerical integration (for integrals over s) and Monte-Carlo integration (for integrals over x). The new estimator is superior over a broad range of distances r, but suprisingly deteriorates at very large

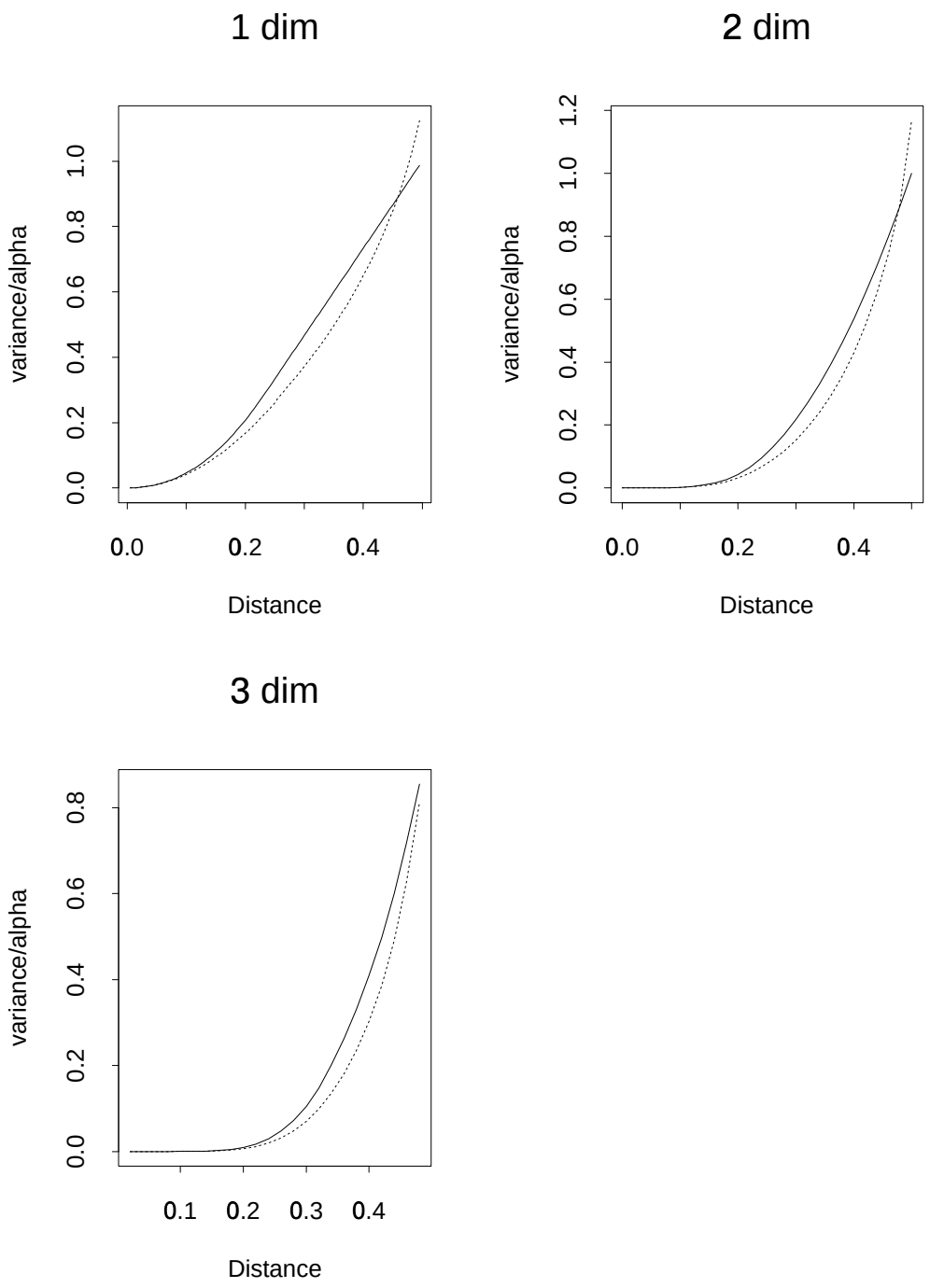

Figure 5: Sparse Poisson limit asymptotic variances (divided by α) in dimensions 1,2 and 3 using the L_∞ metric. *Solid lines:* reduced sample estimator; *dotted lines:* Kaplan-Meier estimator.

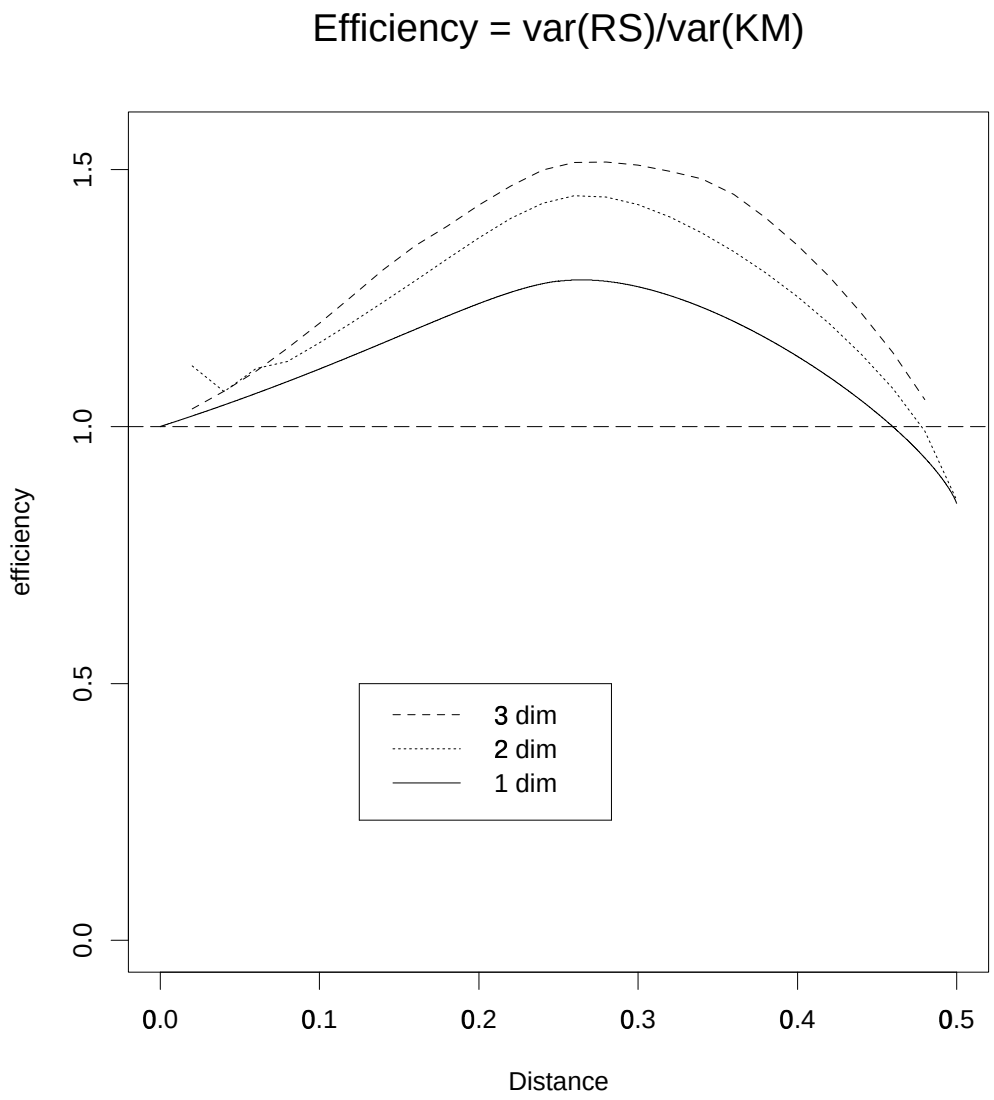

Figure 6: Asymptotic relative efficiency in 1, 2 and 3 dimensions, calculated for the L_∞ metric

distances. Apparently the kind and amount of dependence here has destroyed the uniform optimality enjoyed by the Kaplan-Meier estimator in the i.i.d. case.

Figure 6 shows the asymptotic relative efficiency (ratio of variances of reduced sample to Kaplan-Meier) in dimensions 1 to 3. The greatest gain is achieved at intermediate distances (near $\frac{1}{4}$); only for very large distances (near $\frac{1}{2}$) is there a loss in efficiency. As the dimension d increases, and hence as edge effects become more severe, Kaplan-Meier represents an ever more convincing improvement on the reduced sample estimator.

4.3 Remarks on variance estimation

The variance of $\widehat{F}(r)$ can be approximated by the sum of the squares of the summands in (4.29), in which one would have to replace $\lambda(\cdot)$ and F by their Kaplan-Meier estimates. This is similar to a jacknife or bootstrap analysis.

The computations involved in this procedure can be eased by the same discretisation procedure as was used to approximate $\widehat{F}$ itself: choose points on a regular lattice intersected with W, or many independent random points uniformly distributed over W, and average the 'influence function' for one point x:

$$(1 - F(r)) \left(\frac{\mathbf{1}\{\rho(x,\Phi) \le r,\ \rho(x,\Phi) \le \rho(x,\partial W)\}}{y(\rho(x,\Phi))} - \int_0^{r \wedge \rho(x,\Phi) \wedge \rho(x,\partial W)} \frac{\lambda(s)}{y(s)}\, \mathrm{d}s \right). \tag{4.31}$$

The expression (4.29) for $I(\widehat{F}, \Phi, r)$ can be rewritten as an integral over $x \in W$ of (4.31), with respect to Lebesgue measure, by recognising $|\cdot|_k$ and $|\cdot|_{k-1}\, \mathrm{d}s$ in (4.29) as integrals over x and then interchanging orders of integration. In order to implement this proposal one only has to numerically tabulate an estimate of the function $\int_0^r \frac{\lambda(s)}{y(s)}\, \mathrm{d}s$ together with the functions y and $1 - F$. After (4.31) has been calculated for points sampled from each subwindow W_i, one must average, square, and add over subwindows.

Alternatively one can write down the variance of $I(\widehat{F}, \Phi, r)$ in terms of the covariance structures of the random function $r(x) = \rho(x, \Phi)$ and of the window W. First of all we rewrite (4.29) as

$$I(\widehat{F}, \Phi, r) = -(1 - F(r)) \int_W \int_0^r \mathbf{1}\{\rho(x, \partial W) \ge s\} \left(\frac{\mathrm{d}^{(s)} \mathbf{1}\{x \notin \Phi_{\oplus s}\} + \mathbf{1}\{x \notin \Phi_{\oplus s}\} \lambda(s)\, \mathrm{d}s}{y(s)} \right) \mathrm{d}x$$

where $\mathrm{d}^{(s)}$ denotes Lebesgue-Stieltjes integration with respect to the s variable. After some calculation one arrives at

$$\mathbf{cov}\left(\widehat{F}(r), \widehat{F}(r') \right) \approx (1 - F(r))\ (1 - F(r')) \int_{\mathbb{R}}^k \int_0^r \int_0^r C_{W_{\ominus s}, W_{\ominus s'}}(x) \frac{h(\mathrm{d}s, \mathrm{d}s', x)}{y(s) y(s')}\, \mathrm{d}x \tag{4.32}$$

where $C_{A,B}(x)$ is the set cross-covariance function of $A, B \subset \mathbb{R}^k$

$$C_{A,B}(x) = |A \cap (B \oplus x)|_k\,, \quad x \in \mathbb{R}^k$$

with $B \oplus x$ being the translate of B by x, and

$$h(t, t', x) = \mathbf{cov}\,(M(0,t), M(x,t'))$$

where

$$M(x,t) = \mathbf{1}\{\rho(x,\Phi) \le t\} - \int_0^t \mathbf{1}\{\rho(x,\Phi) > s\} \lambda(s)\, \mathrm{d}s.$$

Note that $M(x,t)$ is a martingale in t for each $x \in \mathbb{R}^k$. Writing σ_x for the covariance function

$$\begin{aligned} \sigma_x(t,t') &= \mathbb{P}\{\Phi_{\oplus t} \not\ni 0,\ \Phi_{\oplus t'} \not\ni x\} \\ &= \mathbb{P}\{z \notin \Phi_{\oplus t},\ x + z \notin \Phi_{\oplus t'}\} \\ &= \mathbb{P}\{\rho(z,\Phi) > t,\ \rho(x+z,\Phi) > t'\} \\ &= \mathbb{P}\{\Phi(B(z,t)) = 0,\ \Phi(B(x+z,t') = 0\} \end{aligned}$$

(for arbitrary $z \in \mathbb{R}^k$) clearly $\sigma_x(t,t')$ is decreasing in t and t' for any fixed x, so we can represent

$$\begin{aligned} h(\mathrm{d}s, \mathrm{d}s', x) = \quad & \sigma_x(\mathrm{d}s, \mathrm{d}s') + \sigma_x(\mathrm{d}s, s')(x)\lambda(s')\,\mathrm{d}s' \\ & + \sigma_x(s, \mathrm{d}s')\lambda(s)\,\mathrm{d}s + \sigma_x(s,s')\lambda(s)\lambda(s')\,\mathrm{d}s\,\mathrm{d}s') \end{aligned} \tag{4.33}$$

so that (4.32) is well-defined as a sum of Stieltjes integrals.

One could try to estimate σ and plug the estimate into (4.32) using estimates of $y(s) = (1 - F(s))\,|W_{\ominus s}|_k$ and $\lambda(\cdot)$ also. Note that σ is actually a bivariate survival function so one could in principle use a Dabrowska-type estimator [10], [22, sec. 12] or just a bivariate reduced sample estimator for this purpose. However, the amount of computation needed is daunting, and the final result may be so statistically inaccurate as to be useless. Practical experience is badly needed here.

5. The nearest neighbour distance function G

The nearest neighbour distribution function G was defined in (2.3), (2.5). Note that G need not have any special continuity properties, in contrast to F; in fact G may be degenerate, as in the case of a randomly translated lattice.

5.1 Kaplan-Meier estimator

A Kaplan-Meier estimator for G is more immediate than for F: for each point x_i of the process Φ observed in the window W, one has a censored distance from x to the nearest other point of Φ, censored by its distance to ∂W. Counting 'observed failures' and 'numbers at risk' as for censored data:

$$N^G(r) = \#\{x \in \Phi \cap W : \rho(x, \Phi \setminus \{x\}) \le r,\ \rho(x, \Phi \setminus \{x\}) \le \rho(x, \partial W)\} \tag{5.34}$$

and

$$Y^G(r) = \#\{x \in \Phi \cap W : \rho(x, \Phi \setminus \{x\}) \ge r,\ \rho(x, \partial W) \ge r\} \tag{5.35}$$

define the Nelson-Aalen estimator

$$\widehat{\Lambda}^G(r) = \int_0^r \frac{dN^G(s)}{Y^G(s)} \tag{5.36}$$

and the Kaplan-Meier estimator of G

$$\widehat{G}(r) = 1 - \prod_0^r \left(1 - d\widehat{\Lambda}^G(s)\right) \tag{5.37}$$

It follows from the Campbell-Mecke formula (see (2.5)) that the numerator and denominator of (5.36) satisfy the same mean-value relation as for ordinary randomly censored data,

$$\mathbb{E}N^G(r) = \int_0^r \mathbb{E}Y^G(s)\ d\Lambda^G(s), \tag{5.38}$$

where $d\Lambda^G(s) = dG(s)/(1 - G(s-))$.

5.2 Comparison with existing estimators

To spell out (5.37) in terms of an observed point pattern $\Phi \cap W = \{x_1, \ldots, x_m\}$, writing

$$\begin{aligned} s_i &= \rho(x_i, \Phi \setminus \{x_i\}) \\ b_i &= \rho(x_i, \partial W) \end{aligned}$$

we have

$$\widehat{G}(r) = 1 - \prod_s \left(1 - \frac{\#\{i : s_i = s,\ s_i \le b_i\}}{\#\{i : s_i \ge s,\ b_i \ge s\}}\right)$$

where s in the product ranges over the finite set $\{s_i\}$. We shall compare this to the reduced-sample estimator

$$\widehat{G}_1(r) = \frac{\sum_{x \in \Phi \cap W_{\ominus r}} \mathbf{1}\{\rho(x, \Phi \setminus \{x\}) \le r\}}{\Phi(W_{\ominus r})} = \frac{\#\{i : s_i \le r, b_i \ge r\}}{\#\{i : b_i \ge r\}} \tag{5.39}$$

and the modification

$$\widehat{G}_2(r) = \frac{|W|_k}{n} \frac{\#\{i : s_i \le r, b_i \ge r\}}{|W_{\ominus r}|_k} \tag{5.40}$$

obtained by replacing $\Phi(W_{\ominus r})$ by an estimate of its expectation. Other estimators are described in [46, p. 128] [8, p. 614, 637–638] and [13, 18].

5.3 Sparse Poisson asymptotics for $\widehat{G}$

Linearization can be applied to $\widehat{G} - G$ just as well as for $\widehat{F} - F$ and the results used to motivate variance estimators through analogues to (4.29)–(4.33). We make use of the following result.

Lemma 5 *Any monotone, uniformly bounded nonnegative stochastic process on $[0,1]$ satisfies a uniform CLT in $D[0,1]$.*

Proof : We use Ossiander's [35] CLT with local bracketing, see [1]. Assume w.l.o.g. that $X(t)$ is increasing and let $\sup_t X(t) \le M$ a.s. Noting that for $s < t$

$$(X(t) - X(s))^2 \le 2M(X(t) - X(s))$$

define $h(t) := 2M\,\mathbb{E}X(t)$; then h is increasing and right-continuous. Given $\epsilon > 0$ find the points $t_i \in [0,1]$ where $\Delta h(t_i) > \epsilon^2$. There are at most M/ϵ^2 such points. For each interval $[t_{i-1}, t_i]$ find a partition $t_{i-1} = s_{i0} < s_{i1} < \ldots < s_{i,k_i} = t_i$ such that $\epsilon^2/2 \le h(s_{i,j+1}) - h(s_{i,j}) < \epsilon^2$ for all j. Then $k_i < 2M/\epsilon^2$. Then the functions δ_{t_i}, δ_{t_i-}, $\delta_{s_{ij}}$ form an L^2 net of mesh ϵ for the function class $\mathcal{F} = \{\delta_t : t \in [0,1]\}$ and have cardinality at most $2M^2/\epsilon^2 + 2M^2/\epsilon^4 = \mathcal{O}\left(\epsilon^{-4}\right)$. Hence

$$\int_0 \left[\log N^{(2)}(\epsilon, \mathcal{F}, P)\right]^{1/2} \mathrm{d}\epsilon \sim \int_0 \sqrt{-\log \epsilon}\ \mathrm{d}\epsilon < \infty$$

and the CLT follows. □

It follows that a joint LLN and CLT hold for (N, Y) uniformly on an interval $[0, \tau]$ where $\mathbb{E}Y(\tau) > 0$. Then differentiability of the product-integral mapping implies weak convergence of $\widehat{G}$ to a Gaussian process at rate $\sqrt{n}$ and the asymptotic variance is equivalent to that of the influence function.

The Kaplan-Meier influence function for the nearest-neighbour distances equals the sum over points $x \in \Phi \cap W$ of the 'usual' influence function based on a censored observation $(\rho(x, \Phi \setminus \{x\}) \wedge \rho(x, \partial W), \mathbf{1}\{\rho(x, \Phi \setminus \{x\}) \le \rho(x, \partial W)\})$. The effective censoring distribution is that of the distance to ∂W from a uniformly distributed random point in W. For example, for the L_∞ distance when W is the unit k-dimensional cube the censoring has survival function $(1-2r)^k$, $0 < r < 1/2$.

Fix the window W, an arbitrary inner regular set with Lebesgue measure 1. The information we need about W and the metric $||\cdot||$ is contained in the functions

$$\begin{aligned} b(r) &= |B_d(0,r)|_k \\ c(r) &= |\partial B_d(0,r)|_{k-1} \\ e(r) &= |W_{\ominus r}|_k \end{aligned}$$

where the erosion $W_{\ominus r}$ is defined in terms of $||\cdot||$. For the L_∞ metric and $W = [0,1]^k$ we have $e(r) = (1-2r)^k$, $b(r) = (2r)^k$ and $c(r) = 2k(2r)^{k-1}$.

The influence function for the Kaplan-Meier estimator of G is thus

$$\sum_{x\in\Phi\cap W}(1-G(r))\left[\frac{\mathbf{1}\{\rho(x,\Phi\setminus\{x\})\le r,\ \rho(x,\Phi\setminus\{x\})\le\rho(x,\partial W)\}}{y(\rho(x,\Phi\setminus\{x\}))}-\int_0^{r\wedge\rho(x,\Phi\setminus\{x\})\wedge\rho(x,\partial W)}\frac{\Lambda(\mathrm{d}s)}{y(s)}\right]$$

where Λ is the cumulative hazard function associated with G and

$$\begin{aligned}y(r) &= \mathbb{E}\{\text{“number at risk at distance } r\text{”}\}\\ &= \mathbb{E}\left\{\sum_{x\in\Phi\cap W}\mathbf{1}\{\rho(x,\Phi\setminus\{x\})\ge r,\ \rho(x,\partial W)\ge r\}\right\}\\ &= \alpha(1-G(r))e(r).\end{aligned}$$

The factor α is the expected number of points in W since $|W|_k=1$.

Suppose the process is homogeneous Poisson with intensity α; then $G(r)=\exp(-\alpha b(r))$ and $\Lambda(\mathrm{d}s)=\alpha c(s)\,\mathrm{d}s$. For α small, $1-G(r)\sim 1$ and $y(r)\sim\alpha e(s)$. The cases $\Phi(W)=0,1,2$ have probabilities $\sim 1,\alpha$ and $\frac{1}{2}\alpha^2$ and result in influence functions

$$\begin{aligned}I(\widehat{G},\emptyset,r) &= 0\\ I(\widehat{G},\{x\},r) &= -(1-G(r))\int_0^{r\wedge\rho(x,\partial W)}\frac{\Lambda(\mathrm{d}s)}{y(s)}\\ &\sim -\int_0^{r\wedge\rho(x,\partial W)}\frac{c(s)}{e(s)}\,\mathrm{d}s\\ I(\widehat{G},\{x,z\},r) &= (1-G(r))\left\{\frac{\mathbf{1}\{d(x,z)\le r;\ d(x,z)\le\rho(x,\partial W)\}+\mathbf{1}\{d(x,z)\le r;\ d(x,z)\le\rho(z,\partial W)\}}{y(d(x,z))}\right.\\ &\quad\left.-\int_0^{r\wedge d(x,z)\wedge\rho(x,\partial W)}\frac{\Lambda(\mathrm{d}s)}{y(s)}-\int_0^{r\wedge d(x,z)\wedge\rho(z,\partial W)}\frac{\Lambda(\mathrm{d}s)}{y(s)}\right\}\\ &\sim \frac{\mathbf{1}\{d(x,z)\le r;\ d(x,z)\le\rho(x,\partial W)\}+\mathbf{1}\{d(x,z)\le r;\ d(x,z)\le\rho(z,\partial W)\}}{\alpha e(d(x,z))}\\ &\quad-\int_0^{r\wedge d(x,z)\wedge\rho(x,\partial W)}\frac{c(s)}{e(s)}\,\mathrm{d}s-\int_0^{r\wedge d(x,z)\wedge\rho(z,\partial W)}\frac{c(s)}{e(s)}\,\mathrm{d}s.\end{aligned}$$

Larger values of $\Phi(W)$ have probability of order α^3 and influence functions of order α^{-1}.

The required asymptotic variance is the expectation of the square of the influence function. The leading term comes from the first part of the case $\Phi(W)=2$ and is (of constant order):

$$\begin{aligned}&\frac{1}{2}\mathbb{E}\left\{\left(\frac{\mathbf{1}\{d(x,z)\le r;\ d(x,z)\le\rho(x,\partial W)\}+\mathbf{1}\{d(x,z)\le r;\ d(x,z)\le\rho(z,\partial W)\}}{e(d(x,z))}\right)^2\right\}\\ =\ &\frac{1}{2}\mathbb{E}\left\{\frac{\mathbf{1}\{d(x,z)\le r\}\left(\mathbf{1}\{d(x,z)\le\rho(x,\partial W)\}+\mathbf{1}\{d(x,z)\le\rho(z,\partial W)\}\right)^2}{e(d(x,z))^2.}\right\}\end{aligned}$$

We now look at the reduced-sample estimator (5.39) in the same way. The expectations of numerator and denominator are $\alpha\,G(r)\,|W_{\ominus r}|_k$ and $\alpha\,|W_{\ominus r}|_k$ respectively, so that the linearized estimator minus estimand is

$$\begin{aligned}I(\widehat{G}^{\mathrm{rs}},\Phi,r) &= \frac{\sum_{x\in\Phi\cap W_{\ominus r}}\mathbf{1}\{\rho(x,\Phi\setminus\{x\})\le r\}-G(r)\Phi(W_{\ominus r})}{\alpha\,|W_{\ominus r}|_k}\\ &= \sum_{x\in\Phi\cap W}\frac{\mathbf{1}\{\rho(x,\partial W)\ge r\}\left(\mathbf{1}\{\rho(x,\Phi\setminus\{x\})\le r\}-G(r)\right)}{\alpha\,|W_{\ominus r}|_k}.\end{aligned}$$

The cases $\Phi(W) = 0, 1, 2$ give influence functions (up to higher order terms, and putting $G(r) \sim \alpha b(r)$)

$$\begin{aligned} I(\widehat{G}^{\mathrm{rs}}, \emptyset, r) &= 0 \\ I(\widehat{G}^{\mathrm{rs}}, \{x\}, r) &= -\frac{b(r)}{e(r)} \mathbf{1}\{\rho(x, \partial W) \geq r\} \\ I(\widehat{G}^{\mathrm{rs}}, \{x, z\}, r) &= \frac{\mathbf{1}\{d(x,z) \leq r\} \left(\mathbf{1}\{\rho(x, \partial W) \geq r\} + \mathbf{1}\{\rho(z, \partial W) \geq r\}\right)}{\alpha e(r)} \\ &- \frac{b(r)}{e(r)} \left(\mathbf{1}\{\rho(x, \partial W) \geq r\} + \mathbf{1}\{\rho(z, \partial W) \geq r\}\right) \end{aligned}$$

For the asymptotic variance's leading term, again only the first part of the case $\Phi(W) = 2$ contributes, giving a term (of constant order):

$$\mathrm{var}\, I(\widehat{G}^{\mathrm{rs}}, \Phi, r) \to \frac{1}{2} \mathbb{E} \left\{ \frac{\mathbf{1}\{d(U,V) \leq r\} \left(\mathbf{1}\{\rho(U, \partial W) \geq r\} + \mathbf{1}\{\rho(V, \partial W) \geq r\}\right)^2}{e(r)^2} \right\} \tag{5.41}$$

where U, V are independent uniformly distributed random points in W.

It is also easy to calculate the influence function of the estimator $\widehat{G}_2$ defined in (5.40). Its asymptotic variance turns out to be asymptotically equivalent to that of $\widehat{G}_1$ given above.

Compare (5.41) with the Kaplan-Meier result

$$\mathrm{var}\, I(\widehat{G}, \Phi, r) \to \frac{1}{2} \mathbb{E} \left\{ \frac{\mathbf{1}\{d(U,V) \leq r\} \left(\mathbf{1}\{\rho(U, \partial W) \geq d(U,V)\} + \mathbf{1}\{\rho(V, \partial W) \geq d(U,V)\}\right)^2}{e(d(U,V))^2} \right\} \tag{5.42}$$

These asymptotic variances have leading terms of *constant* order, because only a fraction α of the realizations provide any data at all; this amplifies an asymptotic variance of order α by the factor $1/\alpha$ to constant order. In the case of F asymptotic variances are of order α as we would expect.

Simple integration techniques of geometrical probability applied to (5.41)–(5.42) give, for the L_∞ metric and $W = [0,1]^k$,

$$\lim \mathrm{var}\, \widehat{G}^{\mathrm{rs}}(r) = \frac{(2r)^k}{(1-2r)^k} + \frac{v(r)^k}{(1-2r)^{2k}}$$

where

$$\begin{aligned} v(r) &= 2(1-2r)(r \wedge (1-2r)) - (r \wedge (1-2r))^2 \\ &= \begin{cases} 2r - 5r^2 & \text{for } r \leq 1/3 \\ (1-2r)^2 & \text{for } 1/3 \leq r \leq 1/2 \end{cases} \end{aligned}$$

and

$$\lim \mathrm{var}\, \widehat{G}(r) = \int_0^r \frac{2k(2s)^{k-1}}{(1-2s)^k}\, \mathrm{d}s + \int_0^{r \wedge 1/3} \frac{2k(1-3s)\left(2s - 5s^2\right)^{k-1}}{(1-2s)^{2k}}\, \mathrm{d}s$$

The results are plotted in Figure 7 for dimensions 1, 2 and 3. They show a superiority of Kaplan-Meier over the reduced sample estimator more marked than in the case of the empty space function. Moreover, the deterioration of the Kaplan-Meier estimator at large distances is not observed any more.

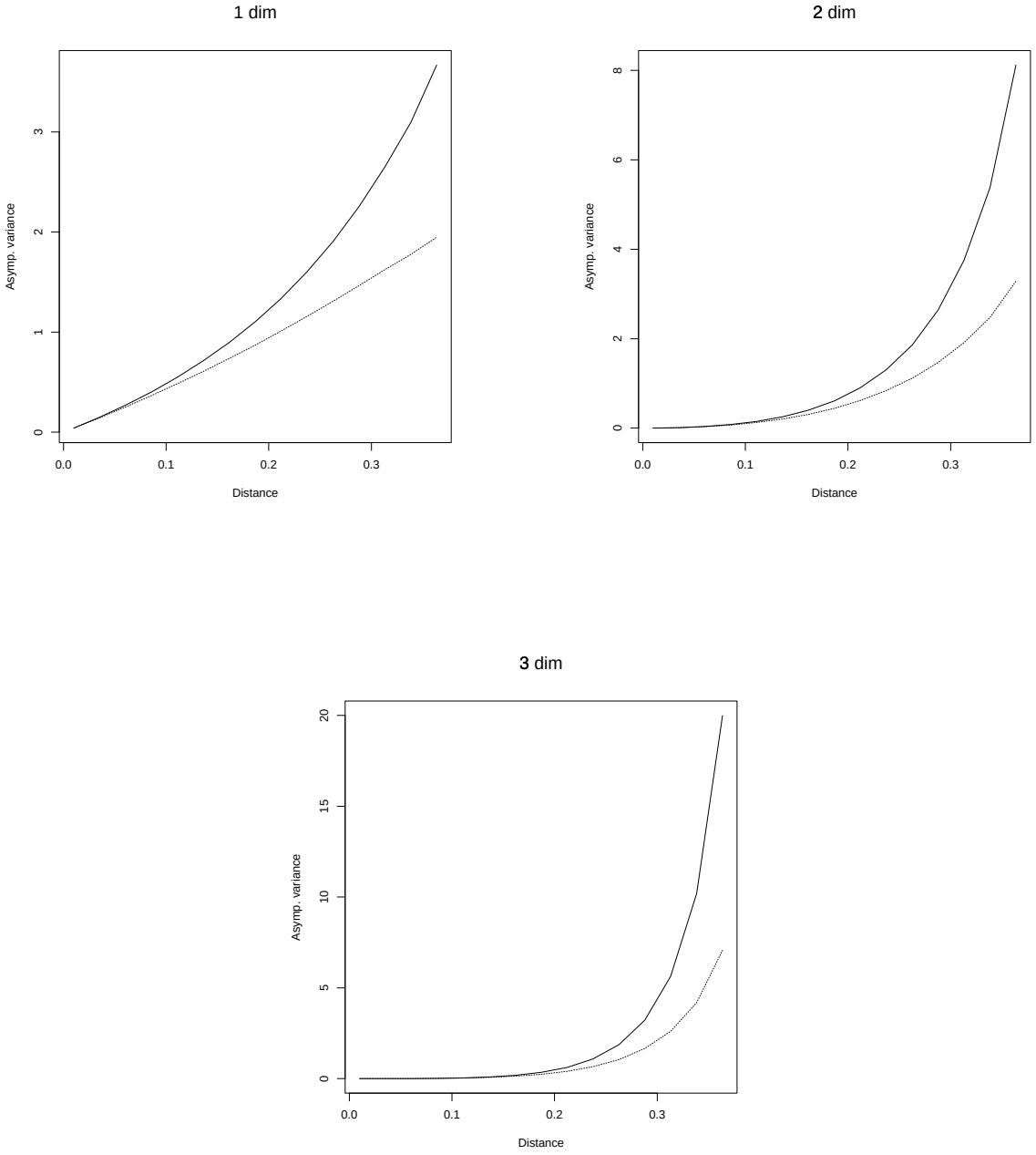

Figure 7: Asymptotic variance of estimators of G, sparse Poisson limit. *Solid lines:* reduced sample; *dotted lines:* Kaplan-Meier.

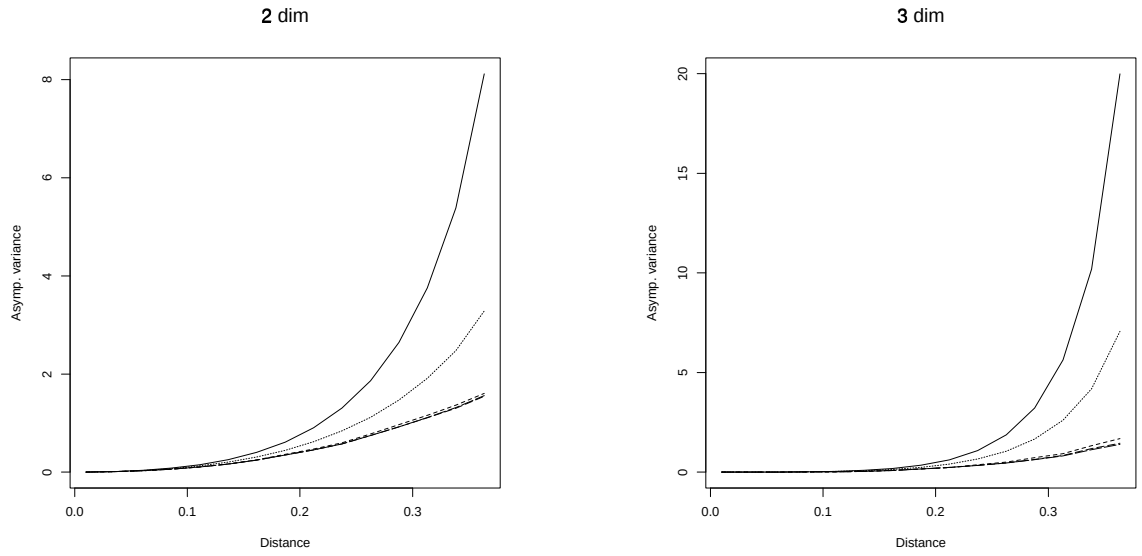

Figure 8: Asymptotic variance of estimators of K, sparse Poisson limit. *Solid lines:* reduced sample; *dotted lines:* Kaplan-Meier; *dashed lines:* weighted edge corrections, see text.

6. The K function

$K(r)$ was defined in (2.4) as $1/\alpha$ times the expected number of points within distance r of a typical point of the process. The possibility of constructing a Kaplan-Meier estimator for $K(r)$ is not so obvious until one notices that $\alpha K(r)$ equals the sum over $n = 1, 2, \ldots$ of the distribution functions G_n of the distance from a typical point to the nearest point, second nearest point, etc. For each of the distance distributions G_n one can form a Kaplan-Meier estimator, since the distance from a point $x \in \Phi$ to its kth nearest neighbour is also censored just as before by its distance to the boundary. One can check that the sequence of Kaplan-Meier estimators always satisfies the natural stochastic ordering of the distance distributions.

The theory we sketched for F and G can also be developed for K. For the estimator of G_n the influence function has a similar form to that given for G. For a Poisson process

$$G_n(r) = e^{-\alpha b(r)} \sum_{k=n}^{\infty} \frac{(\alpha b(r))^k}{k!}.$$

This means that in the previous development, terms for $\Phi(W) = 3, 4, \ldots$ remain of the same (lower) order, while those for $\Phi(W) = 0, 1, 2$ are unchanged. That is, *sparse Poisson asymptotics for* $\widehat{K}$ *coincide with those for* $\widehat{G}$. Hence our conclusions are similar to those of the previous section.

For estimating K a number of sophisticated edge corrections exist; see [11, 34, 36, 37], [38, chap. 3], [46, pp. 122–131], [8, pp. 616–619, 639–644] and for recent extensions [14, 17, 43]. The asymptotic variances of these estimators are the variances of weighted analogues of the influence function given in the previous analysis. Figure 8 shows asymptotic variances for the rigid motion correction, translation correction and isotropic correction (estimated by Monte Carlo simulation of the influence function) together with the asymptotic variances of reduced sample and Kaplan-Meier estimators carried over

from Figure 7. It turns out that as far as the sparse Poisson asymptotics are concerned, the sophisticated edge corrections are equally as good, and better than Kaplan-Meier, which itself is better than the classical border method (reduced sample) estimator.

In two-dimensional spatial statistics it is common to transform K into $L(r) = \sqrt{K(r)/\pi}$ which produces $L(t) = t$ for a Poisson process and approximately stabilises the variance. Our efficiency comparisons remain the same since all asymptotic variances are multiplied by a constant factor $(2\pi\sqrt{K(r)})^{-1}$.

7. General discussion

The Kaplan-Meier technique has been shown to provide good estimators of all three distributions F, G and K. It appears to be substantially more efficient than the simple border correction (reduced sample) estimators in most situations. However, in the case of K the Kaplan-Meier estimator is less efficient (asymptotically in the sparse Poisson limit) than the more sophisticated edge corrections currently in favour. This loss of efficiency is offset by the fact that the Kaplan-Meier estimator requires less computation and can be implemented for windows W of *arbitrary* shape, while the popular edge corrections are only easy to apply in rectangular windows.

A great deal of experimentation is needed to compare the worth of the various estimators in practical situations (see e.g. [13]). Note also that Heinrich [26] has proved large-domain limit theorems concerning the estimation of K in Poisson cluster processes, and Stoyan et al. [45] have derived approximations to the variance of kernel estimators of the pair-correlation function.

The Kaplan-Meier estimator casts new light on the 'local knowledge principle' [42, pp. 49, 233]. This states for the case of dilation by radius r that for all sets X, W

$$X_{\oplus r} \cap W_{\ominus r} = (X \cap W)_{\oplus r} \cap W_{\ominus r}$$

and that $W_{\ominus r}$ is the largest set Z satisfying

$$X_{\oplus r} \cap Z = (X \cap W)_{\oplus r} \cap Z \quad \text{for all } X.$$

In words, given observation of a set X within a window W, the dilation of X is only known within the mask $W_{\ominus r}$. While this principle has been used to justify the border method (reduced sample) estimators, we note that it is not in conflict with the construction of the Kaplan-Meier estimator since $\widehat{F}(r)$ is composed of hazard estimates for distances $s \leq r$.

The Kaplan-Meier estimators use more 'information' than the corresponding reduced sample estimators, but not all information, in the following sense. Write $C(x)$ for the censoring distance $\rho(x, \partial W)$ at a point x, and $T(x)$ for the observed failure distance $\rho(x, \Phi), \rho(x, \Phi \setminus \{x\})$ as appropriate. Then the reduced sample estimate at distance r depends only on those points x where $C(x) \geq r$, while the Kaplan-Meier estimate also involves cases where $T(x) \leq C(x)$ but $C(x) < r$. However, neither estimator makes use of cases where $C(x) < T(x)$ and it seems plausible that these may contain usable information. The sophisticated edge-correction estimators for K use information from the case $C(x) < T(x) \leq r$. Doguwa [15] argues that information should be used from all six possible orderings of $C(x), T(x), r$.

A bootstrap result for the estimators of F, G and K in independent replications case is available from the celebrated Giné-Zinn equivalence theorem that the bootstrap works if and only if the CLT holds; see, e.g. [22, sec. 11].

In applications (see, e.g. [4]) one may well have a number of replicates but typically the number n will be small (say 5 to 10) and the windows not all of the same shape and size. Consequently the formal asymptotics cannot be expected to give a very accurate picture.

One might wonder whether it is possible to improve the Kaplan-Meier estimators of F, G and K by considering the observed distances as *interval-censored* rather than just right censored. This seems possible since for a point $x \in W$, which is closer to ∂W than to other points in $\Phi \cap W$, one does know that its distance to $\Phi \setminus \{x\}$ is not greater than its distance to $(\Phi \setminus \{x\}) \cap W$; so

$$\rho(x, \partial W) \leq \rho(x, \Phi \setminus \{x\}) \leq \rho(x, (\Phi \setminus \{x\}) \cap W)$$

Similar statements can be made for the distance to the kth nearest neighbour. However treating this data as 'randomly interval-censored data' would produce asymptotically biased estimators, since the upper limit $\rho(x, (\Phi \setminus \{x\}) \cap W)$ is strongly dependent on $\rho(x, \Phi \setminus \{x\})$, unlike the lower limit $\rho(x, \partial W)$.

Acknowledgements
The simulations in §3.4 were performed using the image processing package `scilimage` [47] and a library written by Adri Steenbeek. We thank Aad van der Vaart for advice on proving the CLT, and Marie-Colette van Lieshout for reading a draft very carefully.

REFERENCES

1. N.T. Andersen, E. Giné, M. Ossiander, and J. Zinn. The central limit theorem and the law of the iterated logarithm for empirical processes under local conditions. *Probability and related fields*, 77:271–305, 1988.
2. A. J. Baddeley. Integrals on a moving manifold and geometrical probability. *Advances in Applied Probability*, 9:588–603, 1977.
3. A. J. Baddeley. A limit theorem for statistics of spatial data. *Advances in Applied Probability*, 12:447–461, 1980.
4. A. J. Baddeley, R. A. Moyeed, C. V. Howard, S. Reid, and A. Boyde. Analysis of a three-dimensional point pattern with replication. *Applied Statistics*, 42(4):641–668, 1993.
5. L. G. Barendregt and M. J. Rottschäfer. A statistical analysis of spatial point patterns. A case study. *Statistica Neerlandica*, 45:345–363, 1991.
6. G. Borgefors. Distance transformations in arbitrary dimensions. *Computer Vision, Graphics and Image Processing*, 27:321–345, 1984.
7. G. Borgefors. Distance transformations in digital images. *Computer Vision, Graphics and Image Processing*, 34:344–371, 1986.
8. N.A.C. Cressie. *Statistics for spatial data.* John Wiley and Sons, New York, 1991.
9. M. W. Crofton. Sur quelques théorèmes du calcul intégral. *Comptes Rendus de l'Académie des Sciences de Paris*, 68:1469–1470, 1869.
10. D.M. Dabrowska. Kaplan-Meier estimate on the plane. *Annals of Statistics*, 16:1475–1489, 1988.
11. P.J. Diggle. *Statistical analysis of spatial point patterns.* Academic Press, London, 1983.
12. P.J. Diggle and B. Matérn. On sampling designs for the estimation of point-event nearest neighbour distributions. *Scandinavian Journal of Statistics*, 7:80–84, 1981.
13. S.I. Doguwa. A comparative study of the edge-corrected kernel-based nearest neighbour density estimators for point processes. *J. Statist. Comp. Simul.*, 33:83–100, 1989.
14. S.I. Doguwa. On edge-corrected kernel-based pair correlation function estimators for point processes. *Biometrical Journal*, 32:95–106, 1990.
15. S.I. Doguwa. On the estimation of the point-object nearest neighbour distribution $F(y)$ for point processes. *J. Statist. Comp. Simul.*, 41:95–107, 1992.
16. S.I. Doguwa and D.N.Choji. On edge-corrected probability density function estimators for point processes. *Biometrical Journal*, 33:623–637, 1991.
17. S.I. Doguwa and G.J.G. Upton. Edge-corrected estimators for the reduced second moment measure of point processes. *Biometrical Journal*, 31:563–675, 1989.
18. S.I. Doguwa and G.J.G. Upton. On the estimation of the nearest neighbour distribution, $G(t)$, for point processes. *Biometrical Journal*, 32:863–876, 1990.

19. H. Federer. *Geometric Measure Theory*. Springer Verlag, Heidelberg, 1969.

20. Thomas Fiksel. Edge-corrected density estimators for point processes. *Statistics*, 19:67–75, 1988.

21. R.D. Gill. Non- and semiparametric maximum likelihood estimators and the von Mises method, I. (with discussion). *Scandinavian Journal of Statistics*, 16:97–128, 1989.

22. R.D. Gill. Lectures on survival analysis. In P. Bernard, editor, *22e Ecole d'Eté de Probabilités de Saint-Flour 1992*, Lecture Notes in Statistics. Springer, 1993. to appear.

23. R.D. Gill and S. Johansen. A survey of product-integration with a view toward application in survival analysis. *Annals of Statistics*, 18:1501–1555, 1990.

24. E. Giné and J. Zinn. Some limit theorems for empirical processes. *Annals of Probability*, 12:929–989, 1984.

25. K.-H. Hanisch. Some remarks on estimators of the distribution function of nearest neighbour distance in stationary spatial point patterns. *Mathematische Operationsforschung und Statistik, series Statistics*, 15:409–412, 1984.

26. L. Heinrich. Asymptotic Gaussianity of some estimators for reduced factorial moment measures and product densities of stationary Poisson cluster processes. *Statistics*, 19:87–106, 1988.

27. E. Jolivet. Central limit theorem and convergence of empirical processes for stationary point processes. In P. Bastfai and J. Tomko, editors, *Point processes and queueing problems*, pages 117–161. North-Holland, Amsterdam, 1980.

28. E.L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

29. G.M. Laslett. Censoring and edge effects in areal and line transect sampling of rock joint traces. *Mathematical Geology*, 14, 1982.

30. G.M. Laslett. The survival curve under monotone density constraints with applications to two-dimensional line segment processes. *Biometrika*, 69:153–160, 1982.

31. H. W. Lotwick. *Spatial stochastic point processes*. PhD thesis, University of Bath, 1981.

32. G. Matheron. *Random sets and integral geometry*. John Wiley and Sons, New York, 1975.

33. R.E. Miles. On the elimination of edge-effects in planar sampling. In E F Harding and D G Kendall, editors, *Stochastic geometry: a tribute to the memory of Rollo Davidson*, pages 228–247. John Wiley and Sons, London-New York-Sydney-Toronto, 1974.

34. J. Ohser. On estimators for the reduced second moment measure of point processes. *Mathematische Operationsforschung und Statistik, series Statistics*, 14:63–71, 1983.

35. M. Ossiander. A central limit theorem under metric entropy with L_2 bracketing. *Annals of Probability*, 15:897–919, 1987.

36. B.D. Ripley. Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society, Series B*, 39:172 – 212, 1977.

37. B.D. Ripley. *Spatial statistics*. John Wiley and Sons, New York, 1981.

38. B.D. Ripley. *Statistical inference for spatial processes*. Cambridge University Press, 1988.

39. A. Rosenfeld and J. L. Pfalz. Sequential operations in digital picture processing. *Journal of the Association for Computing Machinery*, 13:471, 1966.

40. A. Rosenfeld and J. L. Pfalz. Distance functions on digital pictures. *Pattern Recognition*, 1:33–61, 1968.

41. L. A. Santaló. *Integral Geometry and Geometric Probability*. Encyclopedia of Mathematics and Its Applications, vol. 1. Addison-Wesley, 1976.

42. J. Serra. *Image analysis and mathematical morphology*. Academic Press, London, 1982.

43. M.L. Stein. A new class of estimators for the reduced second moment measure of point processes. Technical Report 278, Department of Statistics, University of Chicago, Chicago, Illinois, USA, July 1990.

44. M.L. Stein. Asymptotically optimal estimation for the reduced second moment measure of point processes. Technical Report 300, Department of Statistics, University of Chicago, Chicago, Illinois, USA, April 1991.

45. D. Stoyan, U. Bertram, and H. Wendrock. Estimation variances for estimators of product densities and pair correlation functions of planar point processes. *Annals of the Institute of Statistical Mathematics*, 45:211–221, 1993.

46. D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications.* John Wiley and Sons, Chichester, 1987.

47. T. K. ten Kate, R. van Balen, A. W. M. Smeulders, F. C. A. Groen, and G. A. den Boer. SCILAIM: a multi-level interactive image processing environment. *Pattern Recognition Letters*, 11:429–441, 1990.

48. A. van der Vaart and J.A. Wellner. *Weak convergence and empirical processes.* IMS Lecture Notes–Monograph series. Institute of Mathematical Statistics, 1993.

49. B. J. Wijers. Consistent non-parametric estimation for a one-dimensional line segment process observed in an interval. Preprint 683, Department of Mathematics, University of Utrecht, September 1991.

50. D. Zimmerman. Censored distance-based intensity estimation of spatial point processes. *Biometrika*, 78:287–294, 1991.