Kaplan-Meier Estimators
of Distance Distributions
for Spatial Point Processes

Adrian Baddeley
University of Western Australia and University of Leiden

Richard D. Gill
University of Utrecht

Addresses for correspondence:
Prof. A.J. Baddeley
Department of Mathematics
University of Western Australia
Nedlands WA 6907, Australia
e-mail: adrian@maths.uwa.edu.au

Prof. R.D. Gill
Mathematical Institute
University of Utrecht
Budapestlaan 6
3584 CD Utrecht, The Netherlands
e-mail: gill@math.ruu.nl
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 Mathematics Subject Classification (1991 Revision): 62G05; 62H11, 60D05.

Keywords: 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.

RUNNING-HEAD: Kaplan-Meier estimators
1 Introduction

The exploratory data analysis of observations of a spatial point process often starts with the estimation of certain distance distributions: $F(t)$ the distribution of the distance from an arbitrary point in space to the nearest point of the process; $G(t)$ 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. 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.

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” [40] 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 [44 pp. 49233]. However the border method discards much of the data; in three dimensions [5] 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 invariance assumptions (stationarity under translation and/or rotation). Corrections of this type were first suggested by Miles [35] and developed by Ripley, Lantuéjoul, Hanisch, Stoyan, Ohser and others [11, 25, 36, 38, 39] [44, p. 246]. For surveys see [40, chap. 3] [41, pp. 122-131] [9, chap. 8] [4].

The estimation problem for $F, G$ and $K$ from data in 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 [30] 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 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 random censoring of survival times has not been explored much. Laslett [31, 32] noted that when a spatial line segment process is clipped within a bounded window the observed line segment lengths can be compared to censored survival times. However a Kaplan-Meier type estimator for the segment length distribution is inconsistent and the NPMLE is a different difficult estimator [52]. Zimmerman [53] 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 tackle this using product integration [22, 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]) and evaluate the asymptotic efficiency explicitly in a simple ‘sparse Poisson’ limiting situation. This also leads to proposals for variance estimators.

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 $\Phi$ be a simple point process in $\mathbb{R}^k$ observed through a compact window $W \subset \mathbb{R}^k$. We consider $\Phi$ 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 $x \in \mathbb{R}^k$ and any closed $A \subset \mathbb{R}^k$ let

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

be the shortest Euclidean distance from $x$ to $A$ and

$$A_{\leq r} = \{x \in \mathbb{R}^k : \rho(x, A) \leq r\}$$
\[ A_{\ominus r} = \{ x \in A : \rho(x, A^c) > r \} \]

where \( c \) denotes complement. Write \( B(x, r) \) for the closed ball of radius \( r \) centre \( x \) in \( \mathbb{R}^k \).

Assume now that \( \Phi \) is a.s. stationary under translations \( \Gamma \) with intensity \( 0 < \alpha < \infty \). Thus \( \Phi(A) = \alpha |A|_k \) for any bounded Borel \( A \subset \mathbb{R}^k \) where \( |\cdot|_k \) denotes \( k \)-dimensional Lebesgue volume. For \( r \geq 0 \) define

\[
F(r) = \{ \rho(0, \Phi) \leq r \} = \{ \Phi(B(0, r)) > 0 \}
\]

\[
G(r) = \{ \rho(0, \Phi \setminus \{0\}) \leq r \mid 0 \in \Phi \} = \{ \Phi(B(0, r)) > 1 \mid 0 \in \Phi \}
\]

\[
K(r) = \alpha^{-1} \{ \Phi(B(0, r) \setminus \{0\}) \mid 0 \in \Phi \}
\]

The conditional expectations given \( 0 \in \Phi \) used above are expectations with respect to the Palm distribution of \( \Phi \) at \( 0 \). By stationarity the point \( 0 \) here may be replaced by any arbitrary point \( x \). Using the Campbell-Mecke formula \[49\]

\[
G(r) = \frac{\sum_{x \in \Phi \cap A} 1 \{ \rho(x, \Phi \setminus \{x\}) \leq r \}}{\Phi(A)}
\]

and

\[
\alpha K(r) = \frac{\sum_{x \in \Phi \cap A} \Phi(B(x, r) \setminus \{x\})}{\Phi(A)}
\]

for arbitrary Borel \( A \) with \( 0 < |A|_k < \infty \). The latter definition of \( K \) is the original one and applies to any second-order stationary process \[49\].

Edge-corrected estimators for \( F, G \) and \( K \) based on observation of \( \Phi \) in \( W \) are reviewed in \[40\] chap. 3, \[49\] pp. 122–131, \[9\] chap. 18. See \[5\] chap. 3 and 14 of \[49\].

Many estimators in spatial statistics are not unbiased but instead are ratios of two unbiased consistent estimators

\[
\hat{\theta} = \frac{Y}{\hat{X}} \quad \text{where} \quad \theta = \frac{Y}{X}
\]
with \( X, Y \geq 0 \) \( \{ X > 0 \} > 0 \Gamma X = 0 \Rightarrow Y = 0 \) typically arising as the mean of a weighted empirical distribution where the weights are random variables \([5\Gamma40]\). 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 \( \bar{T}_i = T_i \land C_i \Gamma D_i = 1 \{ T_i \leq C_i \} \) where \( a \land b \) denotes \( \min \{ a, b \} \). Then \( (\bar{T}_1, D_1), \ldots, (\bar{T}_n, D_n) \) is a sample of censored survival times \( \bar{T}_i \) with censoring indicators \( D_i \). The **reduced-sample estimator** of \( F \) is

\[
\hat{F}^{rs}(t) = \frac{\# \{ i : \bar{T}_i \leq t \leq C_i \}}{\# \{ i : C_i \geq t \}} \tag{7}
\]

This requires that we can 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 may not be a monotone function of \( t \).

The **Kaplan-Meier estimator** of \( F \) is

\[
\hat{F}(t) = 1 - \prod_{s \leq t} \left( 1 - \frac{\# \{ i : \bar{T}_i = s, D_i = 1 \}}{\# \{ i : T_i \geq s \}} \right). \tag{8}
\]

Introduce

\[
N_n(t) = \frac{1}{n} \# \{ i : \bar{T}_i \leq t, D_i = 1 \} \tag{9}
\]

\[
Y_n(t) = \frac{1}{n} \# \{ i : \bar{T}_i \geq t \} \tag{10}
\]

\[
\hat{\Lambda}_n(t) = \int_0^t \frac{dN_n(s)}{Y_n(s)} \tag{11}
\]
\[ \Lambda(t) = \int_0^t \frac{dF(s)}{1 - F(s)}. \]  \hfill (12)

Then \( \Lambda \) is the \textit{cumulative hazard} belonging to \( F \) and \( \hat{\Lambda}_n \) is the Nelson-Aalen estimator of it. One can write
\[
1 - F(t) = \prod_0^t (1 - d\Lambda(s)),
\]
\[
1 - \hat{F}_n(t) = \prod_0^t (1 - d\hat{\Lambda}_n(s)) \tag{13}
\]
where \( \prod \) denotes \textbf{product integration}:
\[
\prod_0^t (1 + dA(s)) = \lim_{\max|t_i-t_{i-1}|\to 0} \prod_{i=1}^m (1 + A(t_i) - A(t_{i-1}))
\]
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 \( \Lambda(t) = \int_0^t \lambda(s) \, ds \) and
\[
1 - F(t) = \int_0^t (1 - d\Lambda(s)) = \exp (-\Lambda(t)).
\]

However if \( F \) has a discrete component the relation \( \Lambda = -\log(1 - F) \) no longer holds. See [22, 23] for further information on the product integral including empirical process theory.

\section{Kaplan-Meier estimator of the empty space function}

\subsection{Definition of estimator}

Return to the setup of \S 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 $\Phi$. 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 where $\partial W$ denotes the boundary of $W$. 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

$$\rho(x, \Phi) \land \rho(x, \partial W) = \rho(x, \Phi \cap W) \land \rho(x, \partial W) \tag{14}$$

(cf. [44Γ pp. 49Γ233]) so that we can indeed observe $\rho(x, \Phi) \land \rho(x, \partial W)$ and $1\{\rho(x, \Phi) \leq \rho(x, \partial W)\}$ for each $x \in W$. Then the set

$$\{x \in W : \rho(x, \Phi) \land \rho(x, \partial W) \geq r\}$$

can be thought of as the set of points ‘at risk of failure at distance $r$’ and

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

are the ‘observed failures at distance $r$’.

**Figure 1 about here**

Geometrically the two sets are the closures of $W_{\|r} \setminus \Phi_{\|r}$ and $\partial (\Phi_{\|r}) \cap W_{\|r}$ respectively. See Figure 1.

**Definition 1** Let $\Phi$ be an a.s. stationary point process and $W \subseteq k$ a fixed compact set. The Kaplan-Meier estimator $\hat{F}$ of the empty space function $F$ of $\Phi$, based on data $\Phi \cap W$, is defined by:

$$\hat{\Lambda}(r) = \int_0^r \frac{|\partial (\Phi_{\|s}) \cap W_{\|s}|_{k-1}}{|W_{\|s} \setminus \Phi_{\|s}|_{k}} ds \tag{15}$$

$$\hat{F}(r) = 1 - \prod_0^r (1 - d\hat{\Lambda}(s)) \tag{16}$$

$$= 1 - \exp(-\hat{\Lambda}(r)) \tag{17}$$
where $|\cdot|_{k-1}$ denotes $k-1$ dimensional Hausdorff measure (‘surface area’). The reduced sample estimator $\hat{F}_m$ of $F$ is

$$\hat{F}_m(r) = \frac{|W_{\partial r} \cap \Phi_{\partial r}|_k}{|W_{\partial r}|_k}$$

(18)

i.e. this is the border correction [40].

Here $\hat{F}$ 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

$$\hat{\lambda}(r) = \frac{|\partial (\Phi_{\partial r}) \cap W_{\partial r}|_{k-1}}{|W_{\partial r} \setminus \Phi_{\partial r}|_k}.$$  

(19)

**3.2 Unbiasedness and continuity**

**Theorem 1** Let $\Phi$ be any stationary point process with intensity $0 < \alpha < \infty$. Then

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

(b) the hazard rate of $F$ equals

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

for any compact window $W$ such that the denominator is positive;

(c) the Kaplan-Meier estimator (19) of $\lambda$ is ratio-unbiased.

Thus our estimator $\hat{F}(r)$ respects the smoothness of the true empty space function $F$. The reduced-sample estimator (18) is not even necessarily monotone.

To prove the Theorem we need three regularity results. The first is an example of Crofton’s perturbation or ‘moving manifold’ formula [2Γ10].
Lemma 1 Let $Z \subset k$ be compact and $A \subset k$ a finite union of compact convex sets. Then for $r \geq 0$

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

the integrand is Lebesgue measurable and integrable.

Proof: The function $f(x) = \rho(x, A)$ is Lipschitz: $f(y) \leq f(x) + ||x - y||\Gamma$ and hence a.e. differentiable with approximate Jacobian $J_1 f \leq 1$. A geometric argument shows that up to a null set $\int_1 f \equiv 1$ and $\{ x : f(x) = s, J_1 f(x) > 0 \} = \partial (A_{\varepsilon s})$. Apply the coarea formula [19p. 258] to integration of $1_Z$ over $A_{\varepsilon r}$. \(\square\)

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

Lemma 2 For any compact set $Z$

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

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 since $\omega_k r^{k-1} = |\partial B(0,r)|_{k-1}$. For the first term fix a realization of $\Phi$ and let $x_i, i = 1, \ldots, m$ be the a.s. distinct points in $\Phi \cap Z_{\varepsilon r}$. Then

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

Construct the Dirichlet cells formed by the $x_i$

$$C_i = C(x_i|x_1, \ldots, x_m) = \{ y \in k : ||y - x_i|| = \min_j ||y - x_j|| \}.$$
Split the surface $\partial(\Phi_{\mathbb{S},r})$ into pieces of surface within each cell:

$$\partial \left( \bigcup_{i=1}^{m} B(x_i, r) \right) = \bigcup_{i=1}^{m} \{ y \in k : ||y - x_i|| = r, \min_{j} ||y - x_j|| = r \}$$

$$= \bigcup_{i=1}^{m} (C_i \cap \partial B(x_i, r))$$

$$= \bigcup_{i=1}^{m} D_i, \text{ (say)}$$

The $D_i$ are measure-disjoint since $D_i \cap D_j = \partial B(x_i, r) \cap \partial B(x_j, r)$ is $k - 2$ dimensional (or empty) for $i \neq j$. Thus

$$\left| \partial \left( \bigcup_{i=1}^{m} B(x_i, r) \right) \cap Z \right|_{k-1} = \sum_{i=1}^{m} |D_i \cap Z|_{k-1}. \quad (20)$$

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 $C_i \Gamma$ since this is convex. The union $F_i$ of these segments is a solid angular cone of the sphere $B(x_i, r) \Gamma$ and its curved surface area $|D_i \cap Z|_{k-1}$ equals $k/r$ times its volume. The cones $F_i$ are volume-disjoint since the $C_i$ are $\Gamma$ and $F_i \subseteq Z_{\mathbb{S},r} \Gamma$ so the sum of the cone volumes is bounded by $|Z_{\mathbb{S},r}| \Gamma$ yielding the result. \hfill $\Box$

**Lemma 3** Let $\Phi$ be a simple point process in $^k$ and $W \subseteq ^k$ compact. Then for fixed $r$, $|W \cap \Phi_{\mathbb{S},r}|_k$ and $|W_{\mathbb{S},r} \cap \Phi_{\mathbb{S},r}|_k$ are a.s. finite r.v.'s on the same probability space, and the following identities hold a.s.:

$$|W \cap \Phi_{\mathbb{S},r}|_k = \int_{0}^{r} |W \cap \partial(\Phi_{\mathbb{S},s})|_{k-1} \, ds \quad (21)$$

$$|\{x \in W : \rho(x, \Phi) \leq \rho(x, \partial W) \wedge r\}|_k = \int_{0}^{r} |W_{\mathbb{S},s} \cap \partial(\Phi_{\mathbb{S},s})|_{k-1} \, ds \quad (22)$$

$$|W_{\mathbb{S},r} \setminus \Phi_{\mathbb{S},r}|_k = |W|_k - \int_{0}^{r} |\partial(W_{\mathbb{S},s} \setminus \Phi_{\mathbb{S},s})|_{k-1} \, ds \quad (23)$$

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

12
Proof: By [34Gamma pp. 9Gamma 19Gamma 47]Gamma, Phi_{\bar{r}} is a random closed set Gamma so that \partial (Phi_{\bar{r}}) is a random closed set Gamma the intersections with W are random compact sets Gamma and their measures are r.v.'s. Now apply lemma 1 to each realization to get (21).

For (22) we note that \rho(x, \partial W) is continuous in x and \rho(x, \Phi) is a random upper-semicontinuous (u.s.c.) function Gamma 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 (22) as the volume of Z \cap \Phi_{\bar{r}}, Gamma and the integrand as the surface area of Z \cap \partial (Phi_{\bar{r}}), Measurability arguments remain valid for the random closed set Z and we apply lemma 1 to each realization.

For (23) observe that W_{\bar{r}} \setminus \Phi_{\bar{r}} = W \setminus (\partial W \cup \Phi)_{\bar{r}} and use the same technique as for (21).

Proof of Theorem 1: By Fubini (Robbins’ theorem [34Gamma pp. 47])
\[
|W \cap \Phi_{\bar{r}}|_{k} = \int_{W} 1\{x \in \Phi_{\bar{r}}\} \, dx = \int_{W} \{x \in \Phi_{\bar{r}}\} \, dx = F(r) \, |W|_{k}, \quad (24)
\]
Since \( r \mapsto |W \cap \Phi_{\bar{r}}|_{k} \) is absolutely continuous with derivative given in Lemma 3Gamma and bounded as in Lemma 2Gamma its expectation is absolutely continuous too Gamma with derivative
\[
f(r) \, |W|_{k} = |W \cap \partial (\Phi_{\bar{r}})|_{k-1}. \quad (25)
\]
But complementarily to (24)
\[
|W \setminus \Phi_{\bar{r}}|_{k} = (1 - F(r)) \, |W|_{k}. \quad (26)
\]
Dividing (25) by (26) we obtain the first result of the theorem. The rest follows by replacing W with W_{\bar{r}}. \qed
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 discretizes \( W \) on a regular lattice (see [12]) although Lotwick [33] 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 \( 1\{\rho(x_i, \Phi) \leq \rho(x_i, \partial W)\} \). Then one would calculate the ordinary Kaplan-Meier estimator (8) 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 \( \hat{F} \).

**Theorem 2** Let \( \hat{F}_L \) be the Kaplan-Meier estimator (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

\[
R = \inf\{r \geq 0 : W_{\epsilon r} \setminus \Phi_{\epsilon r} = \emptyset\}.
\]

Then as the lattice mesh \( \epsilon \) converges to zero, \( \hat{F}_L(r) \to \hat{F}(r) \) for any \( r < R \). The convergence is uniform on any compact interval in \([0, R]\). Similarly the continuous reduced-sample estimator (18) is the uniform limit of the discrete reduced sample estimator (7).

**Proof:** For any compact set \( A \subseteq \mathbb{R}^k \) with \( |\partial A|_k = 0 \) 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 \( M \). The sets \( W_{\partial r} \Phi_{\partial r} \) and

\[
V = \{ x \in W : \rho(x, \Phi) \leq \rho(x, \partial W) \}
\]
clearly have these properties for \( r < R \). Hence the functions

\[
N_L(r) = \frac{\#(L \cap \{ x \in W : \rho(x, \Phi) \leq \rho(x, \partial W) \wedge r \})}{\#(L \cap W)}
\]
and

\[
Y_L(r) = \frac{\#(L \cap (W_{\partial r} \setminus \Phi_{\partial r}))}{\#(L \cap W)}
\]
converge pointwise to

\[
N(r) = \left[ \frac{\{ x \in W : \rho(x, \Phi) \leq \rho(x, \partial W) \wedge r \}}{[W]_k} \right]_k
\]
and

\[
Y(r) = \left[ \frac{W_{\partial r} \setminus \Phi_{\partial r}}{[W]_k} \right]_k
\]
respectively. Since \( N_L(r) \) is increasing in \( r \) and the limit is continuous, \( N_L \rightarrow N \) uniformly in \( r \). Recalling (23) and using the argument of Lemma 2 to bound the integrand, \( Y_L \) converges uniformly in \( r \).

Given (22) and by continuity of the mapping from \( (N, Y) \) to \( \hat{\Lambda}_n = \int dN/Y \) [21 Lemma 3] the discrete Nelson-Aalen estimators

\[
\hat{\Lambda}_L = \int \frac{dN_L}{Y_L}
\]
converge to \( \hat{\Lambda} \). By continuity of the product-integral mapping [23 Theorem 7] \( \hat{F}_L \) converges to \( \hat{F} \). A similar, simpler argument establishes the result for the reduced-sample estimator. \( \square \)

It does not seem to be widely known in spatial statistics (cf. [9 p. 764]) that computation of the distances \( \rho(x, \Phi \cap W), \rho(x, \partial W) \) for all
po i n ts $x$ in a fine rectangular lattice can be performed very efficiently using the distance transform algorithm of image processing [7G8F41G42] at the price of accepting a discrete approximation to the true Euclidean metric $\| \cdot \|_2$ in the definition of $\rho$ at (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 by another metric $\Gamma$ either for computational convenience as above or in order to obtain different information about the process $\Phi$ [27G19G] particularly in three dimensions [5]. 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 2 a polygon) scaled so that

$$\sup \{ \frac{\|x\|}{\|x\|_2} : \|x\|_2 \leq 1 \} = 1. \quad (29)$$

Examples are $\| \cdot \|_\infty \| \cdot \|_{1/\sqrt{2}}$ and continuous versions of the standard ‘chamfer metrics’ [7G8] used in the distance transform algorithm. Redefine the ball of radius $r$ as $B(x, r) = \{ y : \|x - y\| \leq r \}$ and the distance function $\rho$ of (1) in terms of $\| \cdot \|$. Then it can be shown that

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

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.

Estimation of $F$ for more general sets $B\Gamma$ and for more general random sets instead of the point process $\Phi\Gamma$ is treated in [26G27].

Figure 2 about here

Figure 3 about here
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 $256 \times 256$ square grid. For the Poisson process of intensity $\alpha$ the pixel values were i.i.d. Bernoulli variables with $p = \alpha/(256^2)$. We generated 100 realizations of each of Bernoulli $p = 0.001\Gamma 0.0001\Gamma 0.00005\Gamma 0.00002$ and randomly-translated grids of side $s = 25\Gamma 32\Gamma 50\Gamma 100$ and 150. For each realization the distance transform was computed in the chamfer ($5\Gamma 7$) metric of Borgefors [7] and the two estimators derived.

Figure 2 compares the sample standard deviation of the reduced-sample estimator with the sample root mean square error of the Kaplan-Meier estimator (since the reduced sample estimator is unbiased pointwise for $F$). 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.
4 Asymptotic properties of estimators of $F$

4.1 Large sample theory

In spatial statistics, many “large-sample” limiting regimes are possible (see [9], p. 480; [10]; [17], p. 224). It is common to consider the limit in which the window $W$ expands to fill $k$ [31; 28; 29; 46]. Under the additional assumption of ergodicity it is clear that the reduced-sample and Kaplan-Meier estimators of $F$ are pointwise consistent as $W \nearrow k$. Edge effects are asymptotically negligible in this limit.

However, edge effects are appreciable in practical applications, so it would be more relevant to study asymptotic regimes in which the edge effect remains equally severe for all sample sizes. One such limit is considered by Stein [47].

We shall consider the situation where there are $n$ independent replicated observations $\Phi_i$ of a process $\Phi$ within a fixed window $W$. This is becoming increasingly common in applications: the data consist of 10-100 binary images which may be treated as independent replications of the same process (e.g., [5; 27]). Equivalently, if $\Phi$ satisfies a mixing assumption we may consider observation of the same point process through $n$ distantly separated windows $W_i$ of fixed size and shape, cf. [5]. Apart from its practical relevance, study of this limiting regime ($n \to \infty$ replicates) enables qualitative comparison of different estimators and may provide suggestions for variance estimation.

Given $n$ replicated observations $\Phi_i$ in $W$, the pooled statistics $\hat{F}$ and $\hat{F}^{\text{rm}}$ are obtained not as the mean of the separate estimators for each window but by analogues of (16) and (18) in which the numerators and denominators of (15) and (18) are replaced by the sums of these expressions over all replicates $\Phi_i$. Asymptotics as $n \to \infty$ are now straightforward using empirical process theory.
Lemma 4 Let \( \Phi_1, \Phi_2, \ldots \) be i.i.d. copies of an a.s. stationary point process \( \Phi \) with finite positive intensity \( \alpha \). Fix a convex compact set \( W \subset \mathbb{R}^k \) and let \( \hat{F}_n \) be the Kaplan-Meier estimator of \( F \) obtained from \( \Phi_1, \ldots, \Phi_n \) in \( W \) by pooling as above. Let \( \tau > 0 \) satisfy \( F(\tau) < 1 \). Then \( \hat{F}_n \) is consistent and \( \sqrt{n} ( \hat{F}_n - F ) \) converges weakly in \( C[0, \tau] \) to a Gaussian process with linear approximation

\[
\hat{F}_n(r) - F(r) = \frac{1}{n} \sum_{i=1}^{n} I(\hat{F}, \Phi_i, r) + o_p(n^{-1/2})
\]

uniformly in \( 0 \leq r \leq \tau \), where \( I \) is the ‘influence function’

\[
I(\hat{F}, \Phi, r) = (1 - F(r)) \int_0^r \frac{|W_{\Phi} \cap \partial (\Phi_{s})|_{k-1} - |W_{\Phi} \setminus \Phi_{s}|_{k} \lambda(s)}{y(s)} \, ds
\]

and \( y(s) = |W_{\Phi} \setminus \Phi_{s}|_{k} = (1 - F(s)) |W_{\Phi}|_{k} \). A similar statement holds for the reduced-sample estimator \( \hat{F}_n^{rs} \) with the influence function replaced by \( I(\hat{F}_n^{rs}, \Phi, r) = \hat{F}_n^{rs}(r) - F(r) \).

Proof (sketch): Let \( N_i(r), Y_i(r) \) for \( i = 1, 2, \ldots \) be the ‘fraction of failures’ and ‘fraction at risk’ processes (27)–(28) for \( \Phi_i \) in \( W \). They are monotone and uniformly bounded by 1. This gives a LLN uniformly on \([0, \tau]\). By [23][section 4.2][24][Theorem 7.4] the CLT holds for the \( Y_i \) if \( Z = Y - Y \) satisfies \( |Z(t) - Z(s)| \leq A |t - s| \) for all \( s, t \in [0, \tau] \) for some \( A < \infty \). This can be verified by bounding the derivative of \( Y \) using the methods of Lemmas 1 and 2 and the monotonicity of boundary length for convex sets [43]. Similarly the CLT holds for the \( N_i \). A joint CLT follows immediately. Apply the functional delta-method [21][Theorem 3] to the sequence of mappings from \((N_n, Y_n)\) to \((N_n, 1/Y_n)\) then to \( \hat{A}_n = \int \frac{dN_n}{Y_n} \) then to \( 1 - \hat{F}_n = \| (1 - d\hat{A}_n) \). Each mapping is Hadamard differentiable or compactly differentiable [21][22][23][Theorem 8]. Hence \( \sqrt{n}(\hat{F}_n - F) \) 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

\[ \sqrt{n}(1 - F(t)) \int_0^t \frac{dN_n(s) - Y_n(s) \, d\Lambda(s)}{(1 - \Delta \Lambda(s))y(s)}, \quad 0 \leq t < \tau \]

where \( y(s) = Y_n(s) \).

\[ \Box \]

### 4.2 Calculations for the sparse Poisson limit

From Lemma 4 we can obtain the asymptotic variance of the Kaplan-Meier estimator as the variance of the influence function (31). However this expression is unwieldy and further simplifying assumptions are needed to obtain explicit results.

In this section we calculate variances of (31) for the extreme case of a Poisson process whose intensity \( \alpha \) is sent to zero. Edge effects become increasingly severe for small \( \alpha \).

This ‘sparse Poisson limit’ is chosen because it is mathematically tractable yet is stringent enough to reveal qualitative differences between the competing estimators. The differences emerge in the first-order approximation and not (as is usual) at higher orders. The limit also facilitates comparisons with results in survival analysis. It is of course an extreme situation which may not have direct practical impact. It may be relevant to applications where data are observed in a large number of windows each window containing relatively little information.

There are just two situations to consider as \( \alpha \to 0 \): (i) no random point in \( WT \) with probability \( e^{-\alpha |W|_k} = 1 + \mathcal{O}(\alpha) \) and (ii) one random point in \( W \) at a position \( x \) uniformly distributed over \( WT \) 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 (31) for Kaplan-Meier is the difference of two terms: a part depending on surface areas at some distances from a point of \( \Phi \Gamma \) and a part depending on volumes at risk involving the hazard rate of \( F \). In case (i) only the second part is present and is of order \( \alpha \); in case (ii) the first part is also present and is of constant order.

The empty space function for the Poisson process is

\[
F(r) = 1 - e^{-\alpha |B_r|}
\]

and its hazard rate is

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

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

\[
y(r) = (1 - F(r)) |W_{B_r}|_k.
\]

In case (i) no random points in \( W_\Gamma(31) \) is therefore

\[
I(\hat{F}, \emptyset, r) = (1 - F(r)) \left\{ - \int_0^r \frac{\alpha |\partial B_s|_{k-1} |W_{B_s}|_k}{|W_{B_s}|_k e^{-\alpha |B_s|_k}} ds \right\}
\]

\[
= e^{-\alpha |B_r|} \left\{ - \int_0^r \frac{\alpha |\partial B_s|_{k-1} e^{\alpha |B_s|_k}} {e^{-\alpha |B_s|_k}} ds \right\}
\]

\[
= - \left( 1 - e^{-\alpha |B_r|} \right)
\]

\[
= -\alpha |B_r| + O(\alpha^2).
\]

In case (ii) the influence function is

\[
I(\hat{F}, \{x\}, r) = (1 - F(r)) \left\{ \int_0^r \frac{|\partial B(x, s) \cap W_{B_s}|_{k-1} - \alpha |\partial B_s|_{k-1} |W_{B_s} \setminus B(x, s)|_k}{|W_{B_s}|_k e^{-\alpha |B_s|_k}} ds \right\}
\]

\[
= e^{-\alpha |B_r|} \int_0^r \frac{|\partial B(x, s) \cap W_{B_s}|_{k-1}}{|W_{B_s}|_k e^{-\alpha |B_s|_k}} ds + O(\alpha)
\]

\[
= \int_0^r \frac{|\partial B(x, s) \cap W_{B_s}|_{k-1}}{|W_{B_s}|_k} ds + O(\alpha).
\]
It is an interesting exercise to check this by verifying that the expected influence function is zero to first order using integral geometry [43P. 97].

Hence the variance of the influence function is to first order

$$\text{var } I(\hat{F}, \Phi, r) \sim \alpha |W|_k \left( \int_0^r \frac{|\partial B(x, s) \cap W_{\infty,r}|_{k-1}}{|W_{\infty,r}|_k} \, ds \right)^2$$

(32)

since case (i) is now $O(\alpha^2)$. For the reduced sample estimator in case (i) the estimator is identically zero; in case (ii) it is

$$\hat{F}_r^\text{rs}(r) = |B(x, r) \cap W_{\infty,r}|_k / |W_{\infty,r}|_k.$$ 

Since $F(r) = 1 - e^{-\alpha |B|_k} = \alpha |B_x|_k + O(\alpha^2)$ the influence function (= estimator – estimand in this linear case) is in case (i)

$$I(\hat{F}_r^\text{rs}, 0, r) = -\alpha |B_x|_k + O(\alpha^2);$$

in case (ii)

$$I(\hat{F}_r^\text{rs}, \{x\}, r) = \frac{|B(x, r) \cap W_{\infty,r}|_k}{|W_{\infty,r}|_k} + O(\alpha).$$

Again it can be verified using integral geometry that the expectation of the influence function is zero to first order. The variance is

$$\text{var } I(\hat{F}_r^\text{rs}, \Phi, r) \sim \alpha |W|_k \left( \left( \frac{|B(x, r) \cap W_{\infty,r}|_k}{|W_{\infty,r}|_k} \right)^2 \right)$$

(33)

For convenience in calculation of (32)–(33) 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 $||\cdot||_2$ by the $L_\infty$ metric in the definition (1) of $\rho$ and $A_{\infty,r}, A_{\infty}r$ (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 $2r$. 

22
In this case it becomes feasible to enumerate all possible ways the cubes \( B_{\infty}(x, r) \) and \( W_{\odot r} \) intersect. Expressing the volume and surface area contributions in terms of \( x \) 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 \( n^{1/2}(\hat{F}_W(r) - F(r)) \) ignoring terms of order \( O(a^2) \) equals \( a \) times the following expression:

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

For the reduced sample estimator \( |\Phi_{x,r} \cap W_{\odot r}|_{k} / |W_{\odot r}|_{k} \) the corresponding formula is

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

**Figure 5 about here**

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 surprisingly deteriorates at very large distances. Apparently the dependence here has destroyed the uniform optimality enjoyed by the Kaplan-Meier estimator in the i.i.d. case.

**Figure 6 about here**

Figure 6 shows the asymptotic relative efficiency 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.

5 The nearest neighbour function $G$

The nearest neighbour distance distribution function $G$ was defined in (3)\(\Gamma\) (5). Note that $G$ need not have any special continuity properties\(\Gamma\) in contrast to $F$; in fact $G$ may be degenerate\(\Gamma\) as in the case of a randomly translated lattice.

5.1 Kaplan-Meier estimator

Let $\Phi \cap W = \{x_1, \ldots, x_m\}$ be the observed point pattern. A Kaplan-Meier estimator for $G$ is more immediate than for $F$: for each point $x_i$ of the process $\Phi$ observed in the window $WT$ one has a censored distance from $x_i$ to the nearest other point of $\Phi$

$$s_i = \rho(x_i, \Phi \setminus \{x_i\})$$

censored by its distance to $\partial W$

$$b_i = \rho(x_i, \partial W).$$

Counting 'observed failures' and 'numbers at risk' as for censored data:

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

$$= \#\{i : s_i \leq b_i \wedge r\}$$

24
and
\[ Y^G(r) = \#\{x \in \Phi \cap W : r \leq \rho(x, \Phi \setminus \{x\}) \wedge \rho(x, \partial W)\} = \#\{i : s_i \wedge b_i \geq r\} \]
define the Nelson-Aalen estimator
\[ \hat{\Lambda}^G(r) = \int_0^r \frac{dN^G(s)}{Y^G(s)} \quad (34) \]
and the Kaplan-Meier estimator of \( G \)
\[ \hat{G}(r) = 1 - \prod_{s} \left(1 - \frac{\#\{i : s_i = s, s_i \leq b_i\}}{\#\{i : s_i \geq s, b_i \geq s\}}\right) \quad (35) \]
where \( s \) in the product ranges over the finite set \( \{s_i\} \).

It follows from the Campbell-Mecke formula (see (5)) that the numerator and denominator of (34) satisfy the same mean-value relation as for ordinary randomly censored data\( \Gamma \)
\[ N^G(r) = \int_0^r Y^G(s) \, d\Lambda^G(s), \quad (36) \]
where \( d\Lambda^G(s) = dG(s)/(1 - G(s-)) \).

Compare this to the reduced-sample estimator
\[ \hat{G}_1(r) = \sum_{x \in \Phi \cap W_{\infty}} \frac{1\{\rho(x, \Phi \setminus \{x\}) \leq r\}}{\Phi(W_{\infty})} = \frac{\#\{i : s_i \leq r, b_i \geq r\}}{\#\{i : b_i \geq r\}} \quad (37) \]
and the modification
\[ \hat{G}_2(r) = \frac{\left|W_{\infty}\right|}{n} \frac{\#\{i : s_i \leq r, b_i \geq r\}}{\left|W_{\infty}\right|_k} \quad (38) \]
obtained by replacing \( \Phi(W_{\infty}) \) by an estimate of its expectation. Other estimators are described in [19Γp. 128] [9Γp. 614Γ637–638] and [13Γ18].
5.2 Large sample theory for $\hat{G}$

Linearization can be applied to $\hat{G} - G$ just as well as for $\hat{F} - F$ and the results used to study variances.

**Lemma 5** Any monotone nondecreasing, nonnegative stochastic process $X$ on $[0, 1]$ with finite second moment satisfies a uniform CLT in $D[0, 1]$.

**Proof (sketch):** Since $\sup_t X(t)^2 = X(1)^2 < \infty$ for the bracketing CLT of [51 Thm. 2.11.9] it suffices to find for each $\epsilon > 0$ a partition of $[0, 1]$ into $N_\epsilon$ intervals $I_{j, \epsilon}$ such that $\sup_{s, t \in I_{j, \epsilon}} (X(t) - X(s))^2 < \epsilon^2$ for each $j$ and $\int_0^1 \sqrt{\log N_\epsilon} \, d\epsilon < \infty$. For $s < t \Gamma(X(t) - X(s))^2 \leq 2X(1)(X(t) - X(s))$ a.s. The function $h(t) := 2 \left[ X(1)X(t) \right]$ is finite monotone nondecreasing right-continuous. Given $\epsilon > 0$ we can find $N_\epsilon \leq 2h(1)/\epsilon^2$ intervals $[t_i, t_{i+1}]$ with $t_i \leq t_{i+1}$ such that $h(t_{i+1}) - h(t_i) \leq \epsilon^2$. The result follows. □

Since $N^G(r), Y^G(r) \leq \Phi(W)$ it follows that provided $\Phi(W)^2 < \infty$ each of $N^G, Y^G$ satisfies a LLN and CLT uniformly on an interval $[0, \tau]$ where $Y(\tau) > 0$. A joint LLN and CLT for $(N^G, Y^G)$ follow immediately. Then differentiability of the product-integral mapping implies weak convergence of $\hat{G}$ 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 $(\rho(x, \Phi \setminus \{ x \}) \wedge \rho(x, \partial W), 1\{ \rho(x, \Phi \setminus \{ x \}) \leq \rho(x, \partial W) \})$. The effective censoring distribution is that of the distance to $\partial W$ from a uniformly distributed random point in $W$.

Fix the window $W$ a arbitrary inner regular set with Lebesgue measure 1. The information we need about $W$ and the metric $\| \cdot \|$ is contained in the functions $b(r) = |B_k(0, r)|_k \Gamma e(r) = |\partial B_k(0, r)|_k^{-1} \Gamma e(r) = |W_{\geq r}|_k$ where the
erosion $W_{\Theta r}$ is defined in terms of $\| \cdot \|$. For the $L_{\infty}$ metric and $W = [0, 1]^k$ we have $c(r) = (1 - 2r)^k \Gamma b(r) = (2r)^k$ and $c(r) = 2k(2r)^{k-1}$.

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

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

where $\Lambda$ is the cumulative hazard function associated with $G$ and

$$
y(r) = \left\{ \sum_{x \in \Phi \cap W} 1 \{ \rho(x, \Phi \setminus \{ x \}) \geq r, \rho(x, \partial W) \geq r \} \right\} = \alpha(1 - G(r)) c(r).
$$

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

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

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

$$
I(\hat{G}, \emptyset, r) = 0,
$$

$$
I(\hat{G}, \{ x \}, r) = -(1 - G(r)) \int_0^{\rho(\partial W) \cap \rho(x, \partial W)} \frac{\Lambda(ds)}{y(s)}
$$

$$
\sim - \int_0^{\rho(\partial W) \cap \rho(x, \partial W)} \frac{c(s)}{c(s)} \, ds
$$

$$
I(\hat{G}, \{ x, z \}, r) = (1 - G(r)) \times
$$

$$
\left\{ \begin{align*}
1 \{ d(x, z) \leq r \} \frac{1 \{ d(x, z) \leq \rho(x, \partial W) \} + 1 \{ d(x, z) \leq \rho(z, \partial W) \}}{y(d(x, z))} \\
- \int_0^{\rho(\partial W) \cap \rho(x, \partial W)} \frac{\Lambda(ds)}{y(s)} - \int_0^{\rho(\partial W) \cap \rho(z, \partial W)} \frac{\Lambda(ds)}{y(s)}
\end{align*} \right\}
$$

$$
\sim 1 \{ d(x, z) \leq r \} \frac{1 \{ d(x, z) \leq \rho(x, \partial W) \} + 1 \{ d(x, z) \leq \rho(z, \partial W) \}}{\alpha c(d(x, z))}
$$
Larger values of $\Phi(W)$ have probability of order $\alpha^3$ and influence functions of order $\alpha^{-1}$.

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):

$$\text{var } I(\hat{G}, \Phi, r) \to \frac{1}{2} \left\{ 1 \{d(U, V) \leq r\} \frac{(1 \{\rho(U, \partial W) \geq d(U, V)\} + 1 \{\rho(V, \partial W) \geq d(U, V)\})^2}{e(d(U, V))^2} \right\}$$

(39)

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

We now look at the reduced-sample estimator (37) in the same way. The expectations of numerator and denominator are $\alpha G(r) |W_{\Xi_r}|_k$ and $\alpha |W_{\Xi_r}|_k$ respectively so that the linearized estimator minus estimand is

$$I(\hat{G}_1, \Phi, r) = \sum_{x \in \Phi \cap W_{\Xi_r}} 1 \{\rho(x, \Phi \setminus \{x\}) \leq r\} \frac{\alpha |W_{\Xi_r}|_k}{\alpha |W_{\Xi_r}|_k} - G(r) \Phi(W_{\Xi_r})$$

$$= \sum_{x \in \Phi \cap W} 1 \{\rho(x, \partial W) \geq r\} (1 \{\rho(x, \Phi \setminus \{x\}) \leq r\} - G(r))$$

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

$$I(\hat{G}_1, \emptyset, r) = 0$$

$$I(\hat{G}_1, \{x\}, r) = -\frac{b(r)}{e(r)} 1 \{\rho(x, \partial W) \geq r\}$$

$$I(\hat{G}_1, \{x, z\}, r) = \frac{1 \{d(x, z) \leq r\} (1 \{\rho(x, \partial W) \geq r\} + 1 \{\rho(z, \partial W) \geq r\})}{\alpha e(r)}$$

$$-\frac{b(r)}{e(r)} (1 \{\rho(x, \partial W) \geq r\} + 1 \{\rho(z, \partial W) \geq r\})$$

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):

$$\text{var} I(\hat{G}_1, \Phi, r) \to \frac{1}{2} \left\{ 1\{d(U, V) \leq r\} \frac{1\{\rho(U, \partial W) \geq r\} + 1\{\rho(V, \partial W) \geq r\}}{e(r)^2} \right\}$$

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

It is also easy to calculate the influence function of the estimator $\hat{G}_2$ defined in (38). Its asymptotic variance turns out to be asymptotically equivalent to that of $\hat{G}_1$ given above.

Compare (40) with the result (39) for Kaplan-Meier. These have leading terms of constant order because only a fraction $\alpha$ of the realizations provide any data at all; this amplifies an asymptotic variance of order $\alpha$ by the factor $1/\alpha$ to constant order. In the case of $F$ asymptotic variances are of order $\alpha$ as we would expect.

Integration techniques of geometrical probability applied to (39)–(40) give for the $L_\infty$ metric and $W = [0, 1]^k \Gamma$

$$\lim \text{var} \hat{G}_1(r) = \frac{(2r)^k}{(1 - 2r)^k} + \frac{v(r)^k}{(1 - 2r)^{2k}}$$

where

$$v(r) = 2(1 - 2r)(r \wedge (1 - 2r)) - (r \wedge (1 - 2r))^2$$

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

and

$$\lim \text{var} \hat{G}(r) = \int_0^r \frac{2k(2s)^{k-1}}{(1 - 2s)^k} \, ds + \int_0^{r^{1/3}} \frac{2k(1 - 3s) (2s - 5s^2)^{k-1}}{(1 - 2s)^{2k}} \, ds$$

29
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 no longer observed.

6 The $K$ function

$K(r)$ was defined in (4). Equivalently

$$aK(r) = \sum_{n=0}^{\infty} G_n(r)$$

(41)

where $G_n(r) = \{ \Phi(B(0, r)) > n \mid 0 \in \Phi \}$ is the distribution function of the distance from a typical point of $\Phi$ to the $n$th nearest point. For each $G_n$ one can form a Kaplan-Meier estimator since the distance from a point $x \in \Phi$ to its $n$th nearest neighbour is censored just as before by its distance to the boundary. The sequence of Kaplan-Meier estimators always satisfies the natural stochastic ordering of the distance distributions.

The large sample theory we sketched for $F$ and $G$ can also be developed for $K$. Again we require $\Phi(W)^2 < \infty$. For the estimator of $G_n$ the influence function has a similar form to that given for $G$. Since for a Poisson process

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

terms in the influence function 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 $\hat{K}$ coincide with those for $\hat{G}$. Hence our conclusions are similar to those of the previous section.
For estimating $K$ a number of sophisticated edge corrections exist; see [11Γ36Γ38Γ39]Γ[40Γ chap. 3]Γ[49Γpp. 122–131]Γ[9Γpp. 616–619Γ639–644] and recent investigations in [14Γ17Γ45]. 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 under sparse Poisson asymptotics, the sophisticated edge corrections are equally as good and better than Kaplan-Meier, which in turn is better than the classical border method (reduced sample) estimator.

In two-dimensional spatial statistics it is common to transform $K$ into $L(r) = \sqrt{K(r)/\pi}$. Our efficiency comparisons remain the same and 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 ease of implementing the Kaplan-Meier estimator for arbitrary windows $W T$ while the popular edge corrections are only easy to apply in rectangular windows.
Experimentation is needed to compare the worth of the various estimators in practical situations (see e.g. [13]). Heinrich [28] proved large-domain limit theorems concerning the estimation of $K$ in Poisson cluster processes and Stoyan et al. [48] 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’ [44 pp. 49-233]. This states e.g. that for all $X, W \subset k$

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

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

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

In words, given observation of a set $X$ within a window $WT$ the dilation of $X$ is known only within the mask $W_{\bar{r}}$. While this principle has been used to justify the border method (reduced sample) estimators it is not in conflict with the construction of the Kaplan-Meier estimator since $\hat{F}(r)$ is based on 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 $\rho(x, \Phi)$ or $\rho(x, \Phi \setminus \{x\})$ as appropriate. Then the reduced sample estimate at distance $r$ depends only on those points $x$ where $C(x) \leq r$ 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 $C(x), T(x), r$. 

32
A bootstrap result for the estimators of $F, G$ and $K$ in independent replications case is available from the Giné-Zinn equivalence theorem that the bootstrap works if and only if the CLT holds; see e.g. [22Γsec. 11].

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 $\partial W$ 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 $k$th 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 $\rho(x, \partial W)$.

The variance of $\hat{F}(r)$ can be approximated by the sum of the squares of the summands in (30) with $\lambda(\cdot)$ and $F$ replaced by their Kaplan-Meier estimates. The expression (31) for $I(\hat{F}, \Phi, r)$ can be rewritten as an integral over $x \in W$ of the ‘one-point influence function’

$$(1 - F(r)) \left( \frac{1\{\rho(x, \Phi) \leq r, \rho(x, \Phi) \leq \rho(x, \partial W)\}}{y(\rho(x, \Phi))} - \int_0^{\rho(x, \Phi) \wedge \rho(x, \partial W)} \frac{\lambda(s)}{y(s)} \, ds \right).$$

The integral over $x$ can be approximated by a sum over lattice points as above. In order to implement this proposal one only has to numerically tabulate an estimate of the function $\int_0^r \frac{\lambda(s)}{y(s)} \, ds$ together with the functions $y$ and $1 - F$.

Alternatively one can write down the variance of $I(\hat{F}, \Phi, r)$ in terms of the covariance structures of the random function $r(x) = \rho(x, \Phi)$ and of the
window $W$:

\[
\text{cov} \left( \hat{F}(r), \hat{F}(r') \right) \approx (1 - F(r)) (1 - F(r')) \int_0^k \int_0^t C_{W \ominus x, W \ominus x}(x) \frac{h(ds, ds', x)}{y(s)y(s')} \, dx
\]

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

\[
C_{A,B}(x) = \left| A \cap (B + x) \right|, \quad x \in k
\]

with $B + x$ being the translate of $B$ by $x$ and

\[
h(t, t', x) = \text{cov} (M(0, t), M(x, t'))
\]

where

\[
M(x, t) = 1\{\rho(x, \Phi) \leq t\} - \int_0^t 1\{\rho(x, \Phi) > s\} \lambda(s) \, ds,
\]

a martingale in $t$ for each $x \in k$.

Finally, Kaplan-Meier estimators can also be developed for contact distributions, the analogues of $F$ for random closed sets [49]. This is investigated in [26, 27].

**Acknowledgements**

The simulations in §3.4 were performed using the image processing package scilimage [50] and a library written by Adri Steenbeek. We thank Aad van der Vaart, Jon Wellner and David Pollard for advice, and Marie-Colette van Lieshout for reading a draft carefully. The first author was supported by CWI, Amsterdam.
References


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

43
Figure 4: Extreme resonance case of previous Figure
Figure 5: Sparse Poisson limit asymptotic variances divided by a) in dimensions 1, 2 and 3 using the $L^\infty$ metric. Solid lines: reduced sample estimator. Dotted lines: Kaplan-Meier estimator. Solid lines: reduced sample estimator.
Figure 6: Asymptotic relative efficiency (ratio of variances) in 1, 2 and 3 dimensions calculated for the $L_\infty$ metric.
Figure 7: Asymptotic variance of estimators of $\Gamma$-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.