MPS-RR 2000-8

ISSN 1398-599X

Nonparametric estimation of the chord length distribution

Martin B. Hansen

Department of Mathematical Sciences, Aalborg University,
Fredrik Bajers Vej 7E, DK-9220 Aalborg Ø Denmark

Erik W. van Zwet

University of California, Department of Statistics, 367 Evans Hall #3860
Berkeley, CA 94720-3860, USA

ABSTRACT
The distribution of the length of a typical chord of a stationary random set is an interesting feature of the set's whole distribution. We give a nonparametric estimator of the chord length distribution and prove its strong consistency. We report on a simulation experiment in which our estimator compared favorably to a reduced sample estimator. Both estimators are illustrated by applying them to an image sample from a yoghurt ferment. We briefly discuss the closely related problem of estimation of the linear contact distribution. We show by a simulation experiment that a transformation of our estimator of the chord length distribution is more efficient than a Kaplan–Meier type estimator of the linear contact distribution.

Mathematics Subject Classification: 62M30 (Inference from spatial processes).
Keywords and Phrases: chord length distribution, linear contact distribution, nonparametric maximum likelihood estimation, missing data.

1 Introduction

Consider a random closed set $\Xi \subset \mathbb{R}^2$ (Matheron, 1975) which we observe through a bounded 'window' B. Important characteristics of the probability distribution of a random set are the chord length distribution (CLD) and the—perhaps more familiar—linear contact distribution (LCD) (Stoyan et

al. 1995 p. 206–209). These characteristics can give us some insight in the 'architectural' features (Delfiner 1972) of the random set, such as the shape, size and dispersion of its various connected parts. Also, the CLD and LCD can help us when we want to test for stationarity (translation invariance) or for isotropy (rotation invariance).

The CLD and LCD are closely related as is established in formula (7.7) below. With this relation, estimates for the one can be transformed into estimates for the other. For instance, in Hansen et al. (1996) an estimator for the LCD is developed and it is noted that an estimator for the CLD can be obtained from it. Our primary interest is in estimation of the CLD but we shall also argue that a transformation of our estimator of the CLD is superior to the Kaplan–Meier estimator of the LCD of Hansen et al. (1996). Our argument is strongly supported by simulation experiments.

The 'classical' definition of a chord or an intercept (Weibel, 1979, p. 315) is a "line segment contained fully within an object and extending between two points on the object's surface".

We can informally define the chord length distribution in the direction e as the distribution of the length of the longest line segment through the origin in the direction e which is fully contained within Ξ, conditionally on the event that the origin lies in Ξ. If Ξ is stationary and rotation invariant then this distribution does not depend on the choice of the reference point (the origin) or on the direction e. Hence, if we should want to test whether Ξ is isotropic, we could estimate the CLD in various directions and see if the estimates differ. Note that our definition introduces a length bias; the origin is more likely to fall in a 'large' part of Ξ and hence the chord through the origin will have a tendency to be unusually long. If we want, we can correct for this to obtain an 'unweighted' CLD (see formula (2.3) below). For further discussion on chord lengths cf. Serra (1982, Chapter X), Matheron (1975, p. 53) and Stoyan et al. (1995, pp. 207–209).

The CLD has been applied in biology to study certain features of tissue, especially to characterize lung airspace dimensions, see Rosenthal (1989), Lum et al. (1990) and Oldmixon et al. (1994). Estimation of the CLD can be used also for inferential purposes in the statistical analysis of random set models (Stoyan et al. 1995, Hansen et al. 1996).

Estimation of certain characteristics of the chord length distribution can be done from stereological estimates of the surface to volume ratio and the volume fraction. Such stereological estimates are obtained by straightforward intersection and point counting methods (Stoyan et al. 1995 p. 208-209).

These methods are simple, fast and easy to deal with, but they do have their limitations. Stereological mean value formulas can only be used to obtain certain low-dimensional characteristics of the CLD, whereas knowledge of the shape of the CLD may give important additional information about the structure of the random set under consideration.

Estimation is, of course, influenced by edge effects as the random set Ξ is only observed in a bounded observation window B. More specifically, when a certain point x is used as a reference point, the chord through x could extend beyond B. A given chord is called uncensored (u.c.) when both endpoints are in B; singly censored (s.c.) when one of its endpoints is outside B; and doubly censored (d.c.) when both endpoints are outside B.

Estimation from spatial data in the presence of censoring is often dealt with by means of 'minus sampling' which is also known as the 'reduced sample method' or the 'border method'. In our situation this means that when we want to estimate the probability that the length of a chord is less than r, we restrict attention to those reference points which are further than r away from the boundary of B. The resulting estimator can be termed the 'reduced sample estimator' (Baddeley and Gill, 1997). The obvious disadvantage of minus sampling is that much information is discarded, especially for large values of r.

Oldmixon et al. (1994) suggest to make two separate estimates; one based only on the uncensored chords and another one based only on the singly censored chords. These two estimates could then be combined by taking some convex combination of them. Oldmixon et al. (1994) do not discuss how to do so optimally. Also, they discard doubly censored chords.

Estimation from censored observations belongs traditionally to the field of survival analysis, where it has—among other things—resulted in the development and study of the famous Kaplan–Meier estimator. Laslett (1982a and b) first noted the similarity of censoring in survival studies and edge effects in spatial statistics. Subsequently, this analogy has been exploited by Gill (1994), Wijers (1995), van der Laan (1995), Hansen et al. (1996,1999), Baddeley and Gill (1997) and van Zwet (1999). In particular, in Hansen et al. (1996) a Kaplan–Meier type estimator was proposed for the linear contact distribution. The unique correspondence between the LCD and the CLD suggests a transformation of the estimator of the former to obtain an estimator of the latter. Unfortunately, doing so involves a numerically unstable differentiation, which is believed to cause loss of efficiency.

We want to use modern semi-parametric methods and missing data theory

but cannot do so directly. We first consider an analogous problem, derive an NPMLE for *that* problem and then show that the estimator can also be applied in the real problem. The estimator will not be the maximum likelihood estimator in the real problem. It does—in a sense—utilize *all* the available data.

The organization of this paper is as follows. In section 2, we give a more precise definition of the chord length distribution and describe how censoring affects its estimation. In section 3 we give a general review of nonparametric maximum likelihood estimation for missing data models. In section 4 we then give an estimator for the chord length distribution. We study its asymptotics as we observe an increasing number of independent copies of the random set $\Xi \cap B$. Under this regime, we prove uniform consistency of our estimator. In section 5 we report a simulation experiment to try out our new estimator. In section 6 we apply our estimator to protein networks in a youghurt ferment. In section 7 we consider estimation of the linear contact distribution and study its relation to estimation of the chord length distribution. We note that we can transform an estimator for the chord length distibution into an estimator for the linear contact distribution. Finally, in section 8 we report on a simulation experiment to compare a transformation of our chord length estimator to an estimator for the linear conatct distribution proposed in Hansen et al. (1996). The new estimator seems best.

2 The chord length distribution

Let Ξ be a stationary random closed set in $\mathbb{R}^2$ and e a unit vector in $\mathbb{R}^2$. The *chord* through a point of Ξ in the direction e is the longest line segment through the given point with orientation e which is fully contained within Ξ. We are interested in estimation of the distribution of the length of the chord through the origin, given that the origin belongs to Ξ. Since Ξ is stationary, the particular choice of reference point (in this case the origin) is irrelevant. The 'chord length distribution' (CLD) is an interesting feature of the distribution of Ξ. The data on which we shall base our estimation will consist of n independent realizations of Ξ which are all observed through a fixed compact set B.

Let $\xi_e(s,t)$ denote the chord in the direction e through the point $(s,t) \in \Xi$. The situation is clarified in figure 1. A formal definition of the chord length distribution is as follows. Let B be some compact set of positive area. Define,

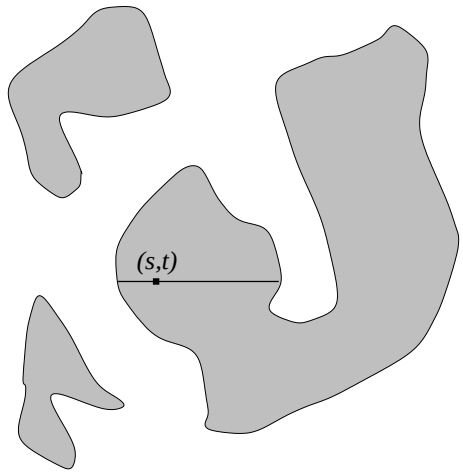

Figure 1: The chord $\xi_{(1,0)}(s,t)$. The shaded region is (part of) the random set Ξ.

for all positive x

$$W_e(x) = \{(s,t) \in \Xi \cap B : |\xi_e(s,t)| \leq x\} \tag{2.1}$$

where $|\xi_e(s,t)|$ denotes the length of the chord $\xi_e(s,t)$. Now define the (length weighted) chord length distribution function as

$$L_e(x) = \frac{E|W_e(x)|}{E|\Xi \cap B|} \tag{2.2}$$

where the expectation is with respect to the distribution of Ξ and $|.|$ denotes area. It is easily seen that this definition does not depend on the choice of B. For simplicity we fix $e = (1,0)$ and omit e from our notation.

We call L the *length weighted* chord length distribution for a reason. Recall that L may be interpreted as the distribution of the length of the chord though the origin, given that the origin lies in Ξ. Informally, the origin is more likely to fall in a big part of Ξ than in a small part. Hence L is biased towards greater chord lengths. Consider the straight horizontal line through the origin intersecting Ξ, resulting in a collection of intervals of varying lengths. Now select an interval at random in a way that does *not* depend on its length. The distribution of this interval's length is also a 'chord

length distribution', but this one is not wighted towards greater lengths. It is given by

$$L_0(x) = \frac{\int_0^x \frac{1}{t} \mathrm{d}L(t)}{\int_0^\infty \frac{1}{t} \mathrm{d}L(t)}. \tag{2.3}$$

The subscript '0' indicating 'unweighted' should not be confused with the subscript 'e' we used earlier.

Of a chord $\xi(s,t)$ through a point $(s,t) \in \Xi \cap B$ we observe only $\xi(s,t) \cap B$. Hence, from observation of $\Xi \cap B$ we can not in general infer $W(x)$. However, we do observe the length of the intersection of each chord with B and we observe how many of each chord's endpoints are outside of B. Define two functions

$$y(s,t) = |\xi(s,t) \cap B| \tag{2.4}$$

$$\delta(s,t) = \#\text{ endpoints of } \xi(s,t) \text{ outside of } B. \tag{2.5}$$

From observation of $\Xi \cap B$ we know, for all positive y and $\delta = 0, 1$ or 2

$$A(y,\delta) = \{(s,t) \in \Xi \cap B : y(s,t) \leq y,\ \delta(s,t) = \delta\}. \tag{2.6}$$

We define

$$F(y,\delta) = \frac{E|A(y,\delta)|}{E|\Xi \cap B|}, \tag{2.7}$$

where again the expectation is with respect to the distribution of Ξ. For $\delta = 0, 1, 2$, the $F(y,\delta)$ are sub-distribution functions.

We can interpret L and F in the following way which will be very useful in the next section.

Lemma 1. *Let (S,T) be uniformly distributed on B. If we condition on the event $(S,T) \in \Xi \cap B$ then*

1. *L is the distribution of $|\xi(S,T)|$;*
2. *F is the distribution of $(y(S,T), \delta(S,T))$;*
3. *conditionally on $\xi(S,T)$, (S,T) is uniformly distributed on $\xi(S,T) \cap B$.*

Proof The first two statements are immediate from the definitions of L and F. To prove the third, define, for all t, $(\Xi \cap B)_t = \{s : (s,t) \in \Xi \cap B\}$. Conditionally on $\xi(S,T)$, T is degenerate. Also, conditional on T, it is clear that S is uniformly distributed on $(\Xi \cap B)_T$. Now $(\Xi \cap B)_T$ consists of at least one line segments, $\xi(S,T) \cap B$ being one of these. Hence, conditionally on $\xi(S,T)$, (S,T) is uniformly distributed on $\xi(S,T) \cap B$. □

3 Nonparametric missing data problems

In this section we review some semi-parametric theory, especially with regards to missing data models which are also known as information loss models or mixture models. A standard reference on semi-parametrics is the monograph by Bickel et al. (1993). Our exposition here owes much to the very readable chapter 25 on semi-parametrics of a recent book by van der Vaart (1998).

Let X be a random variable on some (measurable) space $\mathcal{X}$ with distribution $Q_0 \in \mathcal{Q}$ and let C be another random variable whose conditional distribution given X is known to us. Let $Y = \Phi(X, C)$, where Φ is a known many-to-one mapping. Y takes values in a measurable space $\mathcal{Y}$. Now suppose we observe independent $Y_1, \ldots, Y_n$ which are distributed as Y and from these we want to estimate Q_0. We call the X_i the complete data and the Y_i the incomplete or observed data. Estimation of $Q_0 \in \mathcal{Q}$ (or functionals thereof) from the observations Y_i is called a missing data problem.

An equivalent way to describe the situation is as follows. Consider a *Markov kernel* $K(\mathrm{d}y; x)$. This means that $K(\mathrm{d}y; x)$ is a probability measure for every $x \in \mathcal{X}$ and $K(A; x) = \int_A K(\mathrm{d}y, x)$ is a measurable function of x when $A \subseteq \mathcal{Y}$ is a measurable set. Now for all $Q \in \mathcal{Q}$ define the 'mixture' $P = KQ$ by

$$P(\mathrm{d}y) = KQ(\mathrm{d}y) = \int_{\mathcal{X}} K(\mathrm{d}y; x)\mathrm{d}Q(x). \tag{3.1}$$

Clearly P is a probability measure on the space $\mathcal{Y}$. Usually, Q is called the 'mixing distribution'. We suppose K is known and we observe an i.i.d. sample $Y_1, \ldots, Y_n$ with common distribution $P_0 = KQ_0$, where Q_0 is known to lie in a class $\mathcal{Q}$. The model for the distribution of a generic observation Y is of course $\mathcal{P} = K\mathcal{Q} = \{KQ : Q \in \mathcal{Q}\}$. The objective, again, is to estimate Q_0. The fact that we only observe Y and not X may cause a lack of *identifiability*. We may well have that $KQ = KQ'$ while $Q \neq Q'$. We shall pay special attention to this in the next section.

Suppose we observe data $Y_1, \ldots, Y_n$ with common distribution $P_0 \in \mathcal{P}$, where $\mathcal{P}$ may be infinite dimensional. The nonparametric maximum likelihood estimator (NPMLE) in the sense of Kiefer and Wolfowitz (1956) of P_0 is any

$\hat{P}_n$ satisfying

$$\int \log \frac{\mathrm{d}\hat{P}_n}{\mathrm{d}\mu(P, \hat{P}_n)} \mathrm{d}\mathbb{P}_n \geq \int \log \frac{\mathrm{d}P}{\mathrm{d}\mu(P, \hat{P}_n)} \mathrm{d}\mathbb{P}_n, \qquad \forall P \in \mathcal{P}, \tag{3.2}$$

where $\mathbb{P}_n$ is the empirical distribution of the $Y_1, \ldots, Y_n$. If $\mathcal{P}$ is not dominated by a single sigma finite measure μ, then $\mu(P, \hat{P}_n)$ must vary with its arguments to dominate P and $\hat{P}_n$. For instance, we could take $\mu(P, \hat{P}_n) = P + \hat{P}_n$. In the special missing data setup where $P_0 = KQ_0$, we find that $\hat{Q}_n$ is the NPMLE of Q_0 (up to identifiability) if

$$\int \log \frac{\mathrm{d}K\hat{Q}_n}{\mathrm{d}\mu(KQ, K\hat{Q}_n)} \mathrm{d}\mathbb{P}_n \geq \int \log \frac{\mathrm{d}KQ}{\mathrm{d}\mu(KQ, K\hat{Q}_n)} \mathrm{d}\mathbb{P}_n, \qquad \forall Q \in \mathcal{Q}.$$

In the parametric case we usually consider the 'score function', i.e. the derivative of the log likelihood. Often the maximum likelihood estimator solves the score equations: sum of scores equals zero. The notion of a score can be extended very neatly to the infinite dimensional case. Hence we can also write down 'non-parametric' score equations. The NPMLE is of course not fully characterized as a root of the score equations. However, it is often easier to solve the score equations (for instance by applying the EM algorithm) than to solve (3.2). Also, proving desirable properties of the NPMLE such as consistency or asymptotic normality is generally done by exploiting the fact that it solves the score equations.

Scores are defined through 'differentiable submodels'. Following van der Vaart (1998), we fix a $P \in \mathcal{P}$ and consider maps $t \mapsto P_t$ from a neighborhood of 0 in $[0, \infty)$ to $\mathcal{P}$ such that $P_0 = P$ and such that there exists a measurable function $g : \mathcal{X} \to \mathbb{R}$ for which

$$\int \left[\frac{p_t^{1/2} - p^{1/2}}{t} - \frac{1}{2} g p^{1/2} \right]^2 \mathrm{d}\mu \to 0, \quad t \downarrow 0, \tag{3.3}$$

where p_t and p are Radon–Nikodym derivatives of P_t and P with respect to a sigma-finite measure $\mu(P, P_t)$, which dominates both P_t and P. We call the submodel $t \mapsto P_t$ with (3.3) differentiable at P. If the integrand converges pointwise then $g = \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0} \log \mathrm{d}P_t$, which is the familiar score function. Hence, the function g in (3.3) is called the score associated with the submodel $t \mapsto P_t$. The collection of scores associated with all differential submodels is called

the *tangent set* of the model $\mathcal{P}$ at P and denoted $\dot{\mathcal{P}}(P)$. Lemma 25.14 in van der Vaart (1998) asserts that from (3.3) it follows that $\int g \mathrm{d}P = 0$ and $\int g^2 \mathrm{d}P < \infty$. Hence, $\dot{\mathcal{P}}(P)$ can be identified (up to equivalence) with a subset of $L_2^0(P)$ (the square integrable functions which integrate to zero). We say that $\hat{P}_n$ solves the score equations (and even call it NPMLE) if

$$\int g \mathrm{d}\mathbb{P}_n = 0, \quad \forall g \in \dot{\mathcal{P}}(\hat{P}_n). \tag{3.4}$$

Again we specialize to the missing data setup $P = KQ$. We first give a few useful properties of $\mathcal{Q}$ and $\mathcal{P} = K\mathcal{Q}$.

Lemma 2. *If $\mathcal{Q}$ is convex then so is $\mathcal{P} = K\mathcal{Q}$.*

Lemma 3. *Suppose that $\mathcal{Q}$ is convex. If $Q \ll Q'$ and $\mathrm{d}Q/\mathrm{d}Q' \in L_0^2(Q')$ then $t \mapsto tQ' + (1-t)Q$ is differentiable at Q' with score*

$$\frac{\mathrm{d}Q}{\mathrm{d}Q'} - 1.$$

Lemma 4. *If $t \mapsto Q_t$ is a differentiable submodel through $Q \in \mathcal{Q}$ with score function g , then $t \mapsto P_t = KQ_t$ is a differentiable submodel through $P = KQ \in \mathcal{P}$ with score $E_Q(g(X)|Y)$.*

Proof For a proof refer to Bickel et al. (1993) proposition A.5.5 or Le Cam and Yang (1988) proposition 4. □

By lemma 4 the score equations for a missing data problem take the special form

$$\int E_{\hat{Q}_n}(g(X)|y) \mathrm{d}\mathbb{P}_n(y) = \frac{1}{n} \sum_i E_{\hat{Q}_n}(g(X_i)|Y_i) = 0, \quad \forall g \in \dot{\mathcal{Q}}(\hat{Q}_n). \tag{3.5}$$

By a completely *nonparametric* missing data problem we mean that every function g in $L_2^0(\hat{Q}_n)$ is a score. This situation typically occurs when $\mathcal{Q}$ consists of *all* probability measures on $\mathcal{X}$. In the nonparametric case we can take $g(x) = \mathbf{1}_A(x) - \hat{Q}_n(A)$ for any measurable set A to obtain the so-called *self consistency equations*

$$\hat{Q}_n(A) = \int E_{\hat{Q}_n}(\mathbf{1}_A(X)|y) \mathrm{d}\mathbb{P}_n(y)) = \frac{1}{n} \sum \hat{Q}_n(X_i \in A|Y_i). \tag{3.6}$$

Solving these equations iteratively is an instance of the EM algorithm (Dempster et al. (1977)). When we start the iterations with an initial guess $Q_n^{(0)}$ then all subsequent iterates will be dominated by it. Hence we can never do better than compute a 'sieved' NPMLE over the model $\{Q \in \mathcal{Q} : Q \ll Q_n^{(0)}\}$.

4 A consistent nonparametric estimator

We now return to the problem of estimating the chord length distribution L from n independent observations $\Xi_i \cap B$ which are all distributed as $\Xi \cap B$. We have no prior information about L so we initially choose $\mathcal{L} = \{\text{all distributions on } \mathbb{R}^+\}$ as our 'model'. Let $\Xi_1, \Xi_2, \ldots, \Xi_n$ be independent and identically distributed as Ξ. We observe the $\Xi_i \cap B$. For simplicity, let $B = [0, \tau] \times [0, 1]$ ($\tau > 0$). Note that doubly censored chords (of which neither endpoint is in B) will always have length τ. Let $\xi(i, s, t)$ be the chord through (s, t) in Ξ_i. Like before,

$$W_i(x) = \{(s,t) \in \Xi_i \cap B : |\xi(i,s,t)| \leq x\} \tag{4.1}$$

$$y(i,s,t) = |\xi(i,s,t) \cap B| \tag{4.2}$$

$$\delta(i,s,t) = \#\text{ endpoints of } \xi(i,s,t) \text{ outside of } B. \tag{4.3}$$

$$A_i(y,\delta) = \{(s,t) \in \Xi_i \cap B : y(i,s,t) \leq y,\ \delta(i,s,t) = \delta\} \tag{4.4}$$

To obtain the nonparametric maximum likelihood estimator (NPMLE) of L one should express L in terms of the law of Ξ. Then one should compute the NPMLE of the law of Ξ and insert it into our expression for L. This might be possible if we have a very simple parametric model for the distribution of Ξ, such as a Boolean model of discs. Without such a highly restrictive assumption on Ξ, maximum likelihood estimation of L appears out of the question. We take an approach which is 'inspired by maximum likelihood considerations'.

Suppose we select m points independently, uniformly in each $\Xi_i \cap B$. Thus we obtain points $(S_{ij}, T_{ij}) \in \Xi_i \cap B$ $(i = 1, \ldots, n,\ j = 1, \ldots, m)$. Associated with these points there are chords $\xi(i, S_{ij}, T_{ij})$. Of these chords we observe (possibly censored) lengths $Y_{ij} = y(i, S_{ij}, T_{ij})$ and censoring types $\Delta_{ij} = \delta(i, S_{ij}, T_{ij})$. We denote the empirical distribution of these observations by $\mathbb{P}_{nm}$.

We first consider the special case $m = 1$ because the (Y_{i1}, Δ_{i1}) are independent and identically distributed. We can describe their common distribution F (cf. (2.7)) in a convenient way.

Imagine we fix the distribution of Ξ. Then L (cf. (2.2)) and F are also fixed distribution functions. Now, let (S, T) be a uniformly distributed point in B. Associate with (S, T) a random variable X which is distributed according to L. Conditional on S and X, let R be a random variable which is uniformly distributed on the interval $[S - X, S]$. Now let ξ be the line-segment from (R, T) to $(R + X, T)$. Note that given ξ, (S, T) is uniformly distributed on $\xi \cap B$. Define $Y = |\xi \cap B| = |[R, R + X] \cap [0, \tau]|$ and Δ is the number of endpoints of ξ that fall outside of B which is the same as the number of endpoints of $[R, R+X]$ that fall outside of $[0, \tau]$. We see that our specific choices of $B = [0, \tau] \times [0, 1]$ and $e = (1, 0)$ have made T irrelevant; from (S, X, R) we can derive (Y, Δ)

Lemma 5. *The joint distribution of Y and Δ is given by F as defined in (2.7).*

Proof This is an immediate consequence of Lemma 1. □

The joint distribution of (S, X, R) is of course given by

$$\mathbf{1}_A(s, x, r)\frac{\mathrm{d}s}{\tau}L(\mathrm{d}x)\frac{\mathrm{d}r}{x} \tag{4.5}$$

where $A = \{(s, x, r) : [r, r + x] \cap [0, \tau] \neq \emptyset\}$.

From our description of the model we can explicitly derive the distribution of the observed data (Y, Δ). We introduce two very useful quantities

$$G = \int_{[\tau,\infty)} \frac{\tau}{x}\mathrm{d}L(x) \quad \text{and} \quad H = \int_{[\tau,\infty)} \frac{x - \tau}{x}\mathrm{d}L(x) \tag{4.6}$$

and note that

$$L(\tau^-) + G + H = 1 \tag{4.7}$$

where $L(\tau^-) := \int_0^{\tau^-} \mathrm{d}L(x) := \int_{(0,\tau)} \mathrm{d}L(x)$.

We can write down the distribution of (Y, Δ) in terms of L restricted to $[0, \tau)$, G and H. (Re)define functions $y(s, x, r) = |[r, r + x] \cap [0, \tau]|$ and

$\delta(s,x,r)$ = the number of endpoints of $[r, r+x]$ outside $[0,\tau]$. Now consider sets

$$\begin{aligned}
S(y,0) &= \{(s,x,r) : y(s,x,r) = y,\ \delta(s,x,r) = 0\} \\
&= \{(s,x,r) : x = y,\ 0 < r < \tau - x\} \\
S(y,1) &= \{(s,x,r) : y(s,x,r) = y,\ \delta(s,x,r) = 1\} = S(y,lc) \cup S(y,rc) \\
S(y,lc) &= \{(s,x,r) : x > y,\ s < y,\ r = y - x\} \\
S(y,rc) &= \{(s,x,r) : x > y,\ s > \tau - y,\ r = \tau - y\} \\
S(\tau,2) &= \{(s,x,r) : y(s,x,r) = \tau,\ \delta(s,x,r) = 2\} \\
&= \{(s,x,r) : x > \tau,\ \tau - x < r < 0\}.
\end{aligned}$$

Integrating the distribution of (S,X,R) given in (4.5) over these sets we find, for $0 \le y < \tau$,

$$\begin{aligned}
P(\mathrm{d}y,0) &= \int_{A\cap S(y,0)} \frac{\mathrm{d}s}{\tau}\frac{\mathrm{d}r}{x}\mathrm{d}L(x) \\
&= \int_{x=0}^{\infty}\int_{s=0}^{\tau}\int_{r=s-x}^{s} \mathbf{1}_y(x)\mathbf{1}_{\{0<r<\tau-x\}}(r)\frac{\mathrm{d}s}{\tau}\frac{\mathrm{d}r}{x}\mathrm{d}L(x) \\
&= \int_{s=0}^{\tau}\int_{r=s-y}^{s} \mathbf{1}_{\{0<r<\tau-y\}}(r)\frac{\mathrm{d}s}{\tau}\frac{\mathrm{d}r}{y}L(\mathrm{d}y) \qquad (4.8)
\end{aligned}$$

There is no need to further evaluate this expression, but it should be noted that it depends on L only through its restriction to $[0,\tau)$. For the distribution of the singly censored observations we have by symmetry of right and left censored observations

$$P(\mathrm{d}y,1) = \int_{A\cap S(y,1)} \frac{\mathrm{d}s}{\tau}\frac{\mathrm{d}r}{x}\mathrm{d}L(x) = 2\int_{A\cap S(y,rc)} \frac{\mathrm{d}s}{\tau}\frac{\mathrm{d}r}{x}\mathrm{d}L(x).$$

Hence, for $0 \le y < \tau$,

$$\begin{aligned}
P(\mathrm{d}y,1) &= 2\int_{x=0}^{\infty}\int_{s=0}^{\tau}\int_{r=s-x}^{s} \mathbf{1}_{\{x>y\}}(x)\mathbf{1}_{\{s>\tau-y\}}(s)\mathbf{1}_{\{r=\tau-y\}}(r)\frac{\mathrm{d}s}{\tau}\frac{\mathrm{d}r}{x}\mathrm{d}L(x) \\
&= 2\int_{x=y}^{\infty}\int_{s=\tau-y}^{\tau} \frac{\mathrm{d}s}{x}\frac{\mathrm{d}(\tau-y)}{\tau}\mathrm{d}L(x) \\
&= 2\int_{x=y}^{\infty} \frac{y}{x\tau}\mathrm{d}y\mathrm{d}L(x) \\
&= 2\left(\int_{x=y}^{\tau-} \frac{y}{x\tau}\mathrm{d}y\mathrm{d}L(x) + \frac{y}{\tau^2}\mathrm{d}yG\right), \qquad (4.9)
\end{aligned}$$

where we used that a random variable U (say) is uniformly distributed on $[0,\tau)$ if and only if $\tau - U$ has the same distribution.

The doubly censored observations are always of length τ.

$$\begin{aligned} P(Y=\tau,\Delta=2) &= \int_{A\cap S(y,2)} \frac{\mathrm{d}s}{\tau}\frac{\mathrm{d}r}{x}\mathrm{d}L(x) \\ &= \int_{x=0}^{\infty}\int_{s=0}^{\tau}\int_{r=s-x}^{s} \mathbf{1}_{\{x>\tau\}}(x)\mathbf{1}_{\{\tau-x<r<0\}}(r)\frac{\mathrm{d}s}{\tau}\frac{\mathrm{d}r}{x}\mathrm{d}L(x) \\ &= \int_{x=\tau-}^{\infty}\int_{s=0}^{\tau}\int_{r=\tau-x}^{0} \frac{\mathrm{d}s}{\tau}\frac{\mathrm{d}r}{x}\mathrm{d}L(x) \\ &= \int_{x=\tau-}^{\infty} \frac{x-\tau}{x}\mathrm{d}L(x) \\ &= H \end{aligned} \tag{4.10}$$

Because, by (4.7), G can be expressed in terms of $L|_{[0,\tau)}$ and H, we see that the distribution of the data is fully parameterized by $L|_{[0,\tau)}$ and H. This means that the model $\mathcal{L} = \{\text{all distributions on } \mathbb{R}^+\}$ is not identified. We could reduce $\mathcal{L}$ so that it is, but then the model will no longer be completely nonparametric. We now give a different model and a different mechanism to create (Y,Δ). This new model is both completely nonparametric and identified, while the distribution of (Y,Δ) remains the same.

Consider a *new* model $\mathcal{L}'$ of all distributions on the space $[0,\tau)\cup\{\dagger,\ddagger\}$. By (4.7) we know that $G + H < 1$ and therefore we may interpret G and H as probabilities. In fact, we identify $L(\{\dagger\}) = P(X=\dagger) = G$ and $L(\{\ddagger\}) = P(X=\ddagger) = H$. Relation (4.7) now reads

$$L(\tau^-) + L(\{\dagger\}) + L(\{\ddagger\}) = 1 \tag{4.11}$$

We have the following missing data problem which is both completely nonparametric and identified:

Draw X from $L \in \mathcal{L}'$

- if $X = x \in [0,\tau)$, then sample S uniformly on $[0,\tau]$ and given $S = s$ sample R uniformly on $[s-x, s]$. Finally, set $Y = [R, R+X]\cap[0,\tau]$ and Δ equal to the number of unobserved endpoints.
- if $X = \dagger$, Sample Y from $\mathbf{1}_{[0,\tau)}(y)2y/\tau^2$ and set $\Delta = 1$
- if $X = \ddagger$, set $(Y,\Delta) = (\tau, 2)$

It is easily verified that the distribution of the data (Y, Δ) thus obtained is indeed given by formulas (4.8) to (4.10).

Recall that we have n i.i.d. copies (Y_{i1}, Δ_{i1}) of (Y, Δ) with empirical distribution $\mathbb{P}_{n1}$. The score equations for a general nonparametric missing data problem are given by (3.6). Here, in particular, we take sets A of the form $(x, x + \mathrm{d}x)$ and $\{\ddagger\}$ and find for $x < \tau$

$$\begin{aligned}\widehat{L}_{n1}(\mathrm{d}x) &= \mathbb{P}_{n1}(\mathrm{d}x, 0) \\ &\quad + \int_{y=0}^{x} \frac{(y/x)\mathrm{d}\widehat{L}_{n1}(x)}{\int_{x=y}^{\tau}(y/x)\mathrm{d}\widehat{L}_{n1}(x) + (y/\tau)\widehat{L}_{n1}(\{\dagger\})}\mathbb{P}_{n1}(\mathrm{d}y, 1) \\ \widehat{L}_{n1}(\{\ddagger\}) &= \mathbb{P}_{n1}(\Delta = 2),\end{aligned}$$

where, using (4.11),

$$\widehat{L}_{n1}(\{\dagger\}) = 1 - \widehat{L}_{n1}(\tau^-) - \widehat{L}_{n1}(\{\ddagger\}).$$

Translating from estimation within the model $\mathcal{L}'$ to the original model $\mathcal{L}$, we identify $\widehat{H}_{n1} = \widehat{L}_{n1}(\{\ddagger\})$ and $\widehat{G}_{n1} = \widehat{L}_{n1}(\{\dagger\})$.

Let us now consider arbitrary m. We have observations (Y_{ij}, Δ_{ij}) $(i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, m)$. These observations are all distributed as (Y, Δ) but they are certainly *not* independent. In spite of this, we propose as an estimator the solution of

$$\begin{aligned}\widehat{L}_{nm}(\mathrm{d}x) &= \mathbb{P}_{nm}(\mathrm{d}x, 0) \\ &\quad + \int_{y=0}^{x} \frac{(y/x)\mathrm{d}\widehat{L}_{nm}(x)}{\int_{x=y}^{\tau}(y/x)\mathrm{d}\widehat{L}_{nm}(x) + (y/\tau)\widehat{L}_{nm}(\{\dagger\})}\mathbb{P}_{nm}(\mathrm{d}y, 1) \qquad (4.12) \\ \widehat{L}_{nm}(\{\ddagger\}) &= \mathbb{P}_{nm}(\Delta = 2), \qquad (4.13)\end{aligned}$$

where, $\widehat{L}_{nm}(\{\dagger\}) = 1 - \widehat{L}_{nm}(\tau^-) - \widehat{L}_{nm}(\{\ddagger\})$.

Proposition 1. *Let $m \geq 1$ be an arbitrary integer. As n tends to infinity, $\hat{L}_{nm}$ tends to L_0 almost surely, uniformly on sets of the form $[0, x]$ $(x < \tau)$, $\{\dagger\}$ and $\{\ddagger\}$.*

We sketch a proof, omitting technical detail. Let $\hat{P}_{nm} = K\hat{L}_{nm}$ be the common distribution of the (Y_{ij}, Δ_{ij}). The idea is to first construct a sequence

$\tilde{P}_{nm}$ which is known to converge to $P_0 = KL_0$, while $\hat{P}_{nm} \gg \tilde{P}_{nm}$. Then we show convergence of $\hat{P}_{nm}$ to $\tilde{P}_{nm}$, so that $\hat{P}_{nm}$ must also converge to P_0. Since we made sure that the model is identifiable, this will imply convergence of $\hat{L}_{nm}$ to L_0.

We start with an almost trivial, but important observation.

Lemma 6. *As n tends to infinity, $\mathbb{P}_{nm}$ converges to P_0 almost surely, uniformly over sets of the form $[0, x] \times \delta$, where $x < \tau$ and $\delta = 0, 1, 2$.*

Proof This result follows upon noting that $\mathbb{P}_{nm}$ is just the average of $\mathbb{P}_{n1}, \ldots, \mathbb{P}_{nm}$. Each of these converges to P_0 in the desired sense by the Glivenko-Cantelli Theorem. □

We now construct the comparison sequence $\tilde{P}_{nm}$. Define

$$\alpha(y) = \int_{s=0}^{\tau} \int_{r=s-y}^{s} \mathbf{1}_{\{0<r<\tau-y\}}(r) \frac{\mathrm{d}s}{\tau} \frac{\mathrm{d}r}{y},$$

and note that by (4.8) we have for the sub-distribution of the observed length of an uncensored segment

$$P(\mathrm{d}y, 0) = \alpha(y)\mathrm{d}L(y).$$

Now define, for $x < \tau$

$$\tilde{L}_{nm}(\mathrm{d}x) = \frac{1}{\alpha(x)} \mathbb{P}_{nm}(\mathrm{d}x, 0)$$
$$\tilde{L}_{nm}(\{\ddagger\}) = \mathbb{P}_{nm}(\Delta = 2).$$

Let $\tilde{P}_{nm} = K\tilde{L}_{nm}$. As a straightforward consequence of Lemma 6, we have

Lemma 7. *As n tends to infinity, $\tilde{P}_{nm}$ converges to P_0 almost surely, uniformly over sets of the form $[0, x] \times \delta$, where $x < \tau$ and $\delta = 0, 1, 2$.*

Since $\mathcal{L}'$ consists of all probability measures on a given space, it is convex. Thus the straight line $t\tilde{L}_{nm} + (1-t)\hat{L}_{nm}$ $(0 \leq t \leq 1)$ is a submodel of $\mathcal{L}'$. Since $\tilde{L}_{nm} \gg \hat{L}_{nm}$ and $\mathrm{d}\tilde{L}_{nm}/\mathrm{d}\hat{L}_{nm} \in L_0^2(\hat{L}_{nm})$, it follows by Lemma 3 that the submodel is Hellinger differentiable at $\hat{L}_{nm}$. By lemma 2 the straight line $tK\tilde{L}_{nm} + (1-t)K\hat{L}_{nm}$ is a submodel of $K\mathcal{L}'$. By Lemma 4 it follows that this submodel is Hellinger differentiable as well. Moreover, by Lemma 3 it has score $(\mathrm{d}K\tilde{L}_{nm}/\mathrm{d}K\hat{L}_{nm}) - 1$.

$K\hat{L}_{nm}$ is not the solution of score equations for a true likelihood of the data (Y_{ij}, Δ_{ij}). However, it *is* the solution of the score equations under the assumption that the (Y_{ij}, Δ_{ij}) are all independent. One of those score equations is

$$\int \left(\frac{\mathrm{d}K\tilde{L}_{nm}}{\mathrm{d}K\hat{L}_{nm}} - 1 \right) \mathrm{d}\mathbb{P}_{nm} = 0. \tag{4.14}$$

This equation we may expand into

$$\int \left(\frac{\mathrm{d}K\tilde{L}_{nm}}{\mathrm{d}K\hat{L}_{nm}} \right) \mathrm{d}(\mathbb{P}_{nm} - K\tilde{L}_{nm}) + \int \left(\frac{\mathrm{d}K\tilde{L}_{nm}}{\mathrm{d}K\hat{L}_{nm}} - 1 \right) \mathrm{d}K\tilde{L}_{nm} = 0. \tag{4.15}$$

It can be shown (cf. van de Geer (1993)) that the second summand is not less than the Hellinger distance between $K\tilde{L}_{nm}$ and $K\hat{L}_{nm}$.

$$\int \left(\frac{\mathrm{d}K\tilde{L}_{nm}}{\mathrm{d}K\hat{L}_{nm}} - 1 \right) \mathrm{d}K\tilde{L}_{nm} \geq H(K\tilde{L}_{nm}, K\hat{L}_{nm}) \geq 0.$$

If we can show convergence to zero of the first term of (4.15), we can conclude that $K\tilde{L}_{nm}$ and $K\hat{L}_{nm}$ converge in Hellinger distance, which is equivalent to convergence in total variation. This is very strong convergence indeed, and in particular it is stronger than convergence uniformly over sets of the form $[0, x] \times \delta$. Hence convergence to zero of the first term of (4.15) implies by lemma 7 convergence of $K\hat{L}_{nm}$ to KL_0 uniformly over sets of the form $[0, x] \times \delta$. This, in turn, implies convergence of $\hat{L}_{nm}$ to L_0 uniformly over sets $[0, x]$ $(x < \tau)$, $\{\dagger\}$ and $\{\ddagger\}$, as stated in Proposition 1. Now to show that the first summand of (4.15) indeed converges to zero, we split it up into the the various censoring types.

$$\begin{aligned}
&\int \left(\frac{\mathrm{d}K\tilde{L}_{nm}}{\mathrm{d}K\hat{L}_{nm}} \right) \mathrm{d}(\mathbb{P}_{nm} - K\tilde{L}_{nm}) \\
&= \int_0^{\tau-} \left(\frac{K\tilde{L}_{nm}(\mathrm{d}y, 0)}{K\hat{L}_{nm}(\mathrm{d}y, 0)} \right) (\mathbb{P}_{nm}(\mathrm{d}y, 0) - K\tilde{L}_{nm}(\mathrm{d}y, 0)) \\
&+ \int_0^{\tau-} \left(\frac{K\tilde{L}_{nm}(\mathrm{d}y, 1)}{K\hat{L}_{nm}(\mathrm{d}y, 1)} \right) (\mathbb{P}_{nm}(\mathrm{d}y, 1) - K\tilde{L}_{nm}(\mathrm{d}y, 1)) \\
&+ \int_0^{\tau-} \left(\frac{K\tilde{L}_{nm}(\mathrm{d}y, 2)}{K\hat{L}_{nm}(\mathrm{d}y, 2)} \right) (\mathbb{P}_{nm}(\mathrm{d}y, 2) - K\tilde{L}_{nm}(\mathrm{d}y, 2)).
\end{aligned}$$

Now we chose $\tilde{L}_{nm}$ in such a way that $K\tilde{L}_{nm}(\mathrm{d}y, 0) = \mathbb{P}_{nm}(\mathrm{d}y, 0)$. Hence the first term is equal to zero. Exactly the same happens with the third term. Convergence to zero of the second term is somewhat tricky. The difficulty lies at the point τ. However, restricting the integral to $[0, \sigma]$ $(\sigma < \tau)$ convergence can be obtained using Lemmas 6 and 7. Increasing σ to τ we obtain the result for $[0, \tau)$.

This concludes our sketch of the proof of Proposition 1.

The number (m) of points that we pick in each $\Xi_i \cap B$ is entirely up to ourselves. The larger m, the smaller one would expect the variance of $\hat{L}_{nm}$ to be. If (for fixed n) we ever increase m the estimator converges to

$$\widehat{L}_n(\mathrm{d}x) = \frac{\sum_i |A_i(\mathrm{d}x, 0)|}{\sum_i |\Xi_i \cap B|} + \int_{y=0}^{x} \frac{(y/x)\mathrm{d}\widehat{L}_n(x)}{\int_{x=y}^{\tau}(y/x)\mathrm{d}\widehat{L}_n(x) + (y/\tau)\widehat{L}_n(\{\dagger\})} \frac{\sum_i |A_i(\mathrm{d}y, 1)|}{\sum_i |\Xi_i \cap B|} \tag{4.16}$$

$$\widehat{L}_n(\{\ddagger\}) = \frac{\sum_i |A_i(\tau, 2)|}{\sum_i |\Xi_i \cap B|}, \tag{4.17}$$

where

$$\widehat{L}_n(\{\dagger\}) = 1 - \widehat{L}_n(\tau^-) - \widehat{L}_n(\{\ddagger\}).$$

5 Simulations of the CLD

In this section we compare our estimator to a simple alternative. Recall our definition of the chord length distribution in the direction $e = (1, 0)$. $L(x) = E|W(x)|/E|\Xi \cap B|$ where $W(x) = \{(s, t) \in \Xi \cap B : |\xi(s, t)| \leq x\}$. It is clear that this definition does not depend on B. From observing $\Xi \cap B$ we cannot in general infer $|W(x)|$.

Define C to be the closed line segment from $(-1, 0)$ to $(1, 0)$. Let xC be the closed line segment from $(-x, 0)$ to $(x, 0)$. 'Minkowski' subtraction of two sets A_1 and A_2 is defined as $A_1 \ominus A_2 = (A_1^c \oplus A_2)^c$. We shall consider $B \ominus xC = [x, \tau - x] \times [0, 1]$, for positive $x \leq \tau$. This set is called the erosion of B by xC.

Note that $W(x) \cap (B \ominus xC) = \{(s, t) \in \Xi \cap (B \ominus xC) : |\xi(s, t)| \leq x\}$ and that $|W(x) \cap (B \ominus xC)|$ can be inferred from observing $\Xi \cap B$. It is natural

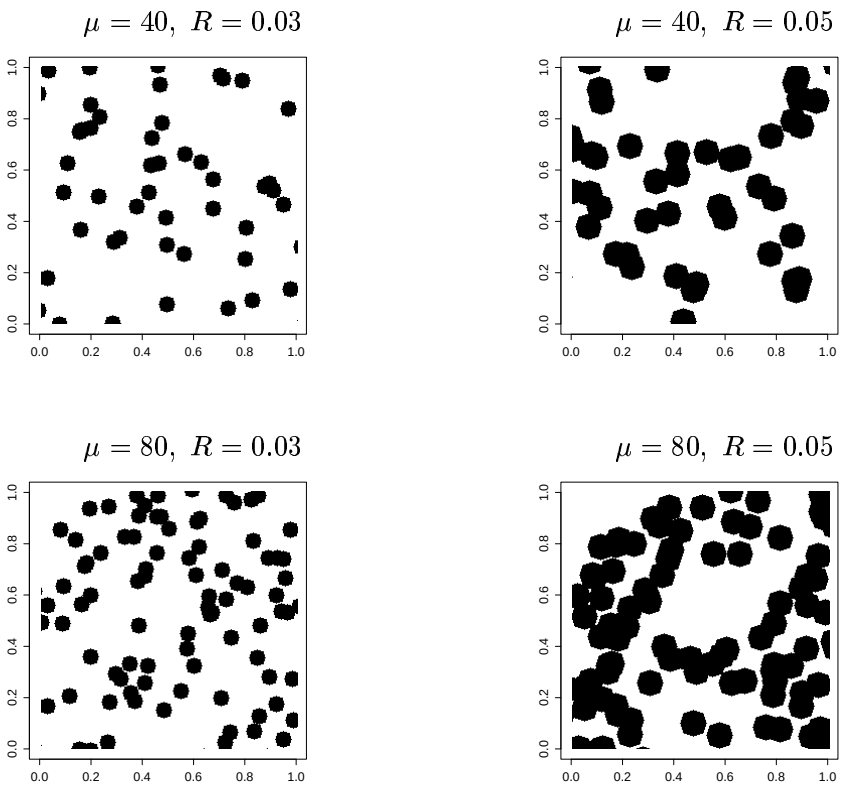

Figure 2: 'Typical' realizations of Boolean models

to define the following 'reduced sample' estimator based on an i.i.d. sample of n copies $\Xi_1, \Xi_2, \ldots, \Xi_n$ of Ξ all observed through B

$$\tilde{L}_n^*(x) = \frac{\sum_i |W_i(x) \cap (B \ominus xC)|}{\sum_i |\Xi_i \cap (B \ominus xC)|}.$$

We shall use a a similar, but more stable alternative

$$\tilde{L}_n(x) = \frac{\sum_i |W_i(x) \cap (B \ominus xC)|/|B \ominus xC|}{\sum_i |\Xi_i \cap B|/|B|}. \tag{5.1}$$

These estimators are 'ratio unbiased', meaning that they are the ratio of unbiased estimators of the numerator and denominator appearing in the definition of $L(x)$. The are clearly (uniformly, almost surely) consistent.

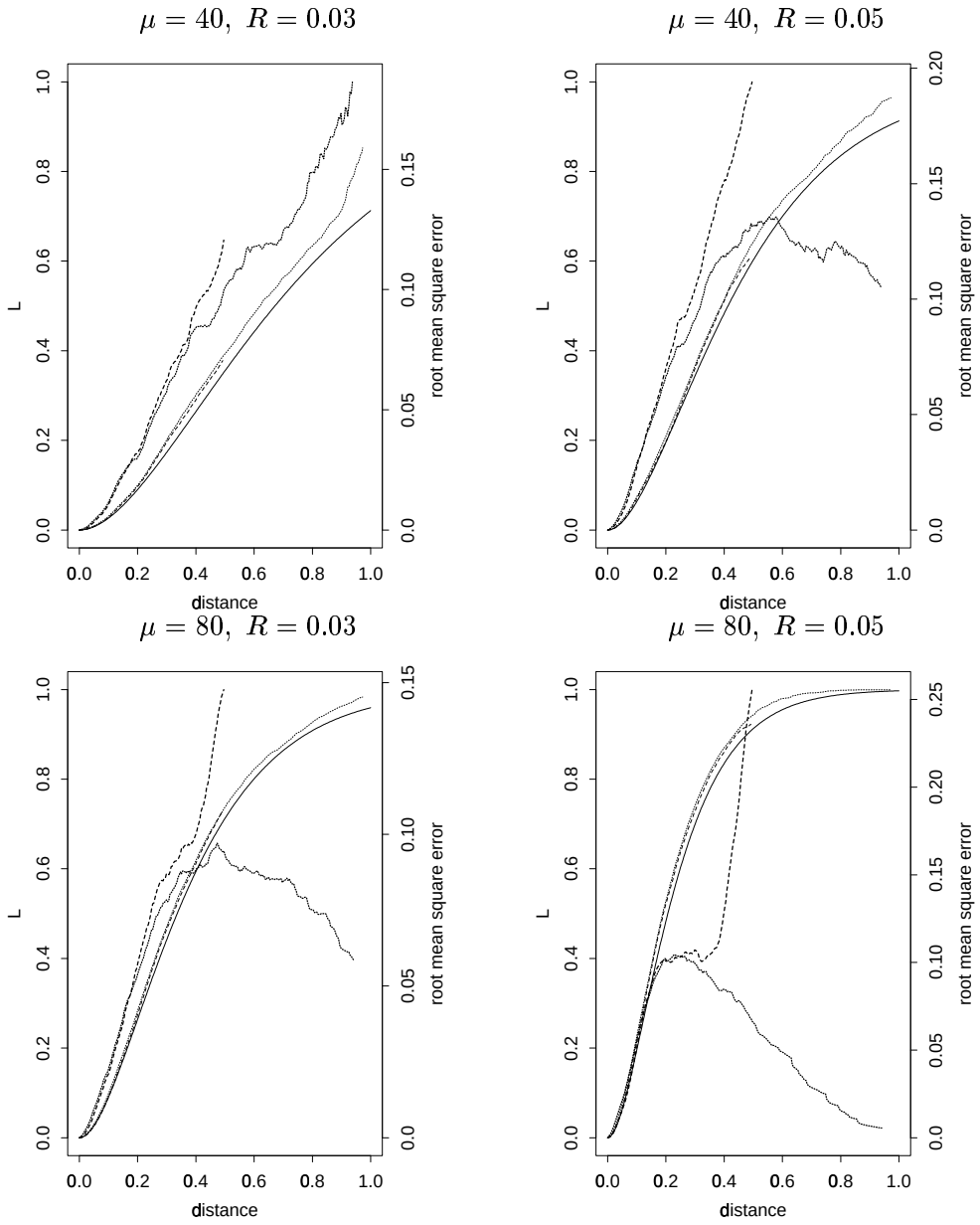

Figure 3: Root mean square error (RMSE) comparison of the NPMLE and the reduced sample estimator. The thick dashed line is the estimated RMSE of the NPMLE and the thick solid line is the estimated RMSE of the reduced sample estimator. The thin dotted line is an average of 100 NPMLE's, the thin dashed line is an average of 100 reduced sample estimators. The thin solid line is the estimand.

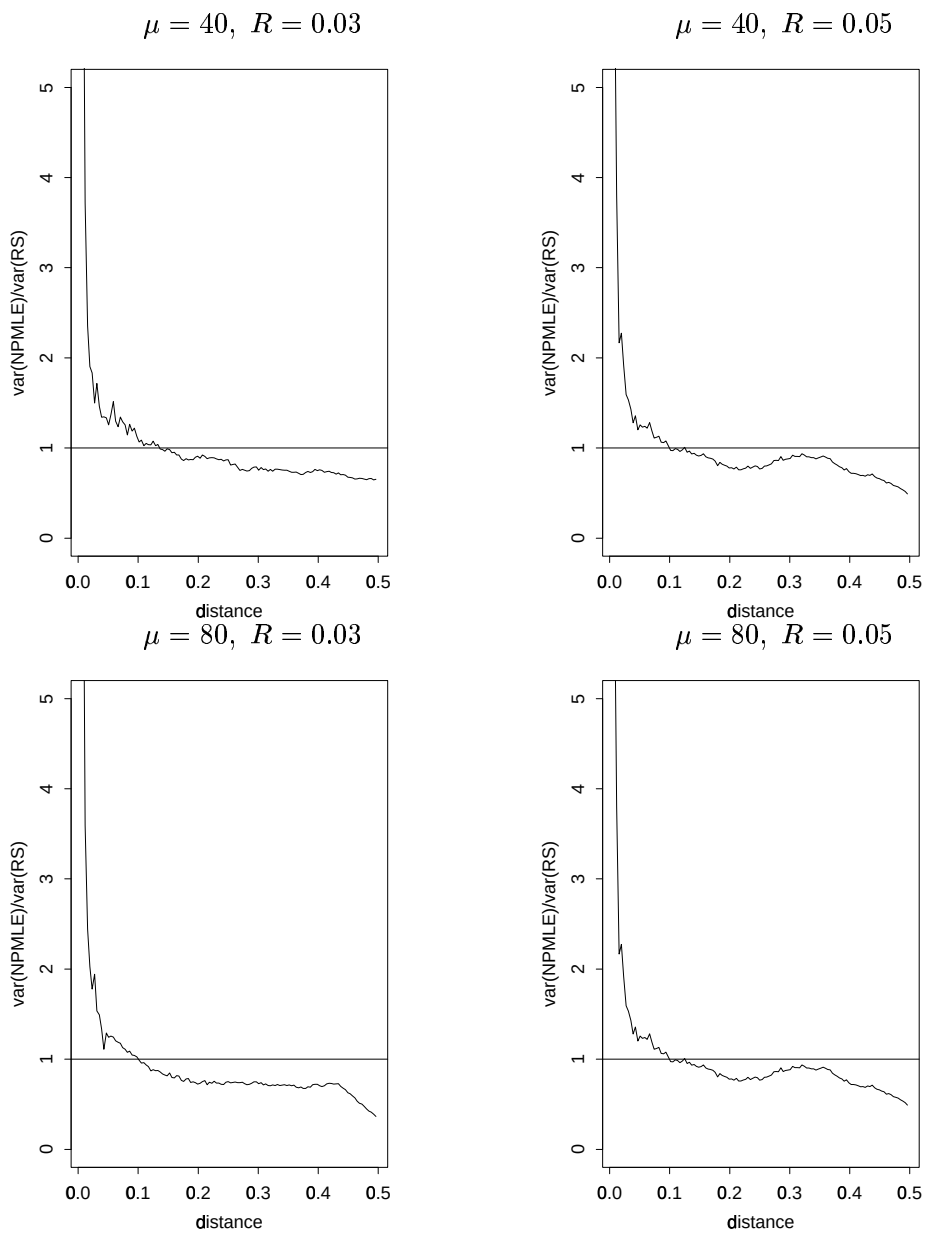

Figure 4: Estimated relative efficiency of the NPMLE and the reduced sample estimator

In Figure 2 we see realizations of Boolean models Ξ observed in the unit square, $B = [0,1]^2$. The grains are discs with a constant radius of $R = 0.03$ or 0.05. The underlying intensity of the Poisson process is $\mu = 40.0$ or 80.0.

We take the random set of interest to be the closure of the void of the Boolean model $\overline{\Xi^c}$. The reason for doing this, is that we know how to obtain a closed form expression for the chord length distribution of $\overline{\Xi^c}$. We use the so called linear contact distribution function of Ξ (Stoyan et al., 1995, p. 80)

$$H(x) = 1 - \exp(-2\mu R x),$$

where $R = 0.03,\ 0.05$ is the radius and $\mu = 40, 80$ are the intensities. The following equation relates the linear contact distribution function H of Ξ and the chord length distribution L of $\overline{\Xi}^c$ (Matheron, 1975, p. 53)

$$L(r) = H(r) - r\frac{\mathrm{d}H(r)}{\mathrm{d}r}. \tag{5.2}$$

We have generated 100 realizations $\Xi_1, \Xi_2, \ldots, \Xi_{100}$ of each Boolean model. For each Ξ_k we have derived the 'NPMLE' estimator $\hat{L}_1^{(k)}$ on the basis of (4.16) and (4.17). We also computed the reduced sample estimator $\tilde{L}_1^{(k)}$ on the basis of (5.1). The subscript "1" indicates in both cases that the estimator is based on a single sample.

In (4.16), (4.17) and (5.1) there appear areas of sets. Because our images are discretized we must make discrete approximations of these areas to approximate the estimators. This is straightforward. Let $B_N = \{1, 2, \ldots, N\}^2$ be the square grid of pixels that is the discrete analog of $B = [0,1]^2$. Suppose we observe $\Xi_1 \cap B_N$. Let (i,j) denote an element in B_N and let $y(i,j)$ denote the observed length of the chord through (i,j) and let $\delta(i,j)$ denote its censoring type. Since we measure distances in the horizontal direction only, $y(i,j)$ takes values in $\{1/N, 2/N, \ldots, 1\}$. Now for $0 \le y < 1$ and $\delta = 0, 1, 2$ define

$$\begin{aligned} |A^N(y,\delta)| &= \#\{(i,j) \in \Xi_1 \cap B_N : y(i,j) \le y,\ \delta(i,j) = \delta\}/N^2 \\ |\Xi_1 \cap B_N| &= \#\{(i,j) \in \Xi_1 \cap B_N\}/N^2. \end{aligned}$$

Of course the $|A^N(y,\delta)|$ and $|\Xi_1 \cap B_N|$ approximate the $|A_1(y,\delta)|$ and $|\Xi_1 \cap B|$

of (4.16) and (4.17). The EM algorithm was initiated with

$$\begin{aligned}
\mathrm{d}\widehat{L}^{\mathrm{init}}(i/N) &= \frac{|A_1^N((i+1)/N,0)| - |A_1^N(i/N,0)|}{|\Xi_1 \cap B_N|}, \quad i = 1,2,\ldots,N-1 \\
\widehat{L}^{\mathrm{init}}(\{\ddagger\}) &= \frac{|A_1^N(1,2)|}{|\Xi_1 \cap B_N|} \\
\widehat{L}^{\mathrm{init}}(\{\dagger\}) &= 1 - \widehat{L}^{\mathrm{init}}(1-(1/N)) - \widehat{L}^{\mathrm{init}}(\{\ddagger\}).
\end{aligned}$$

As a stop criterion of the EM-algorithm we simply chose a large number of iterations, 5000, and in all cases we experienced practically no changes in the final iterates. One could of course use more sophisticated stop criteria.

The reduced sample estimate based on $\Xi_1 \cap B_N$ is approximated by

$$\tilde{L}_1^{(1)}(x) = \frac{\#\{(i,j) \in B_N : y(i,j) \le x,\ D(i,j) > x\}}{\#\{(i,j) \in B_N : D(i,j) > x\}}, \quad 0 \le x \le 1,$$

where $D(i,j) = (N-i)/N$ is the distance from (i,j) to the right boundary of B_N.

All calculations are done by an extension of the routines presented in Hansen (1996), which are available at `www.math.auc.dk/~mbh/ficodifu`. One should notice that all distances in the grid are calculated using distance transformation algorithms from image analysis (Borgefors, 1984, 1986). These provide a good approximations of Euclidean distances and are computationally very efficient.

In Figure 3 we show by thin lines the pointwise averages of $(1/100)\sum_k \hat{L}_1^{(k)}$ and $(1/100)\sum_k \hat{L}^{(k)}$. In Figure 3 we also see a comparison of the estimated root mean square errors (RMSE) of the two estimators. That is, the figure shows

$$\left(\sum_{k=1}^{100}(\tilde{L}_1^{(k)}(x) - L(x))^2/100\right)^{1/2} \quad \text{and} \quad \left(\sum_{k=1}^{100}(\hat{L}_1^{(k)}(x) - L(x))^2/100\right)^{1/2}.$$

The bias seems to be of same order for both estimators, although we notice that the reduced sample estimator has slightly smaller bias. If we turn to the RMSE we see that the NPMLE performs uniformly better than the reduced sample estimator and one should also notice the NPMLE gives a rather confident estimate of the whole range of the image in contrast to the reduced sample estimator which is only available up to half the range of the image!

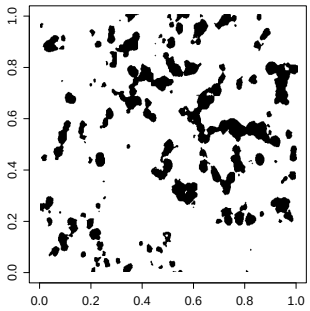

Figure 5: 'Typical' image of the protein network in youghurt. Physical scale of the image is a square with sidelengt 24.73 μm.

Since, as we observed in Figure 3, the bias in both estimators is negligible, comparison of their variances should be a good measure for their relative performance. In Figure 4 we plot the (estimated) relative efficiency. We notice that the variance of both estimators is of the same order, except for very small distances where—to our surprise—the NPMLE appears much worse. We have no explanation for the phenomenon. To the credit of the NPMLE we point out that it yields a reasonable estimate over a much wider range than does the reduced sample estimator.

We should point out that comparing our estimator to this very simple reduced sample estimator is not the last word. One could surely think of more sophisticated alternatives (such as various edge correction estimators) that would present a sterner test.

6 Example: protein network in a yoghurt ferment

The dataset to be analyzed consists of 12 microscopy images of the protein network in a yoghurt ferment magnified $\times 7500$. It is a part of a larger data set which was obtained to investigate the impact of heat treatment on the microstructure of stirred yoghurt. For details see Skriver et al. (1997). This dataset has also been analyzed in Hansen (1995), Hansen et al. (1996) and Skriver (1995). Each picture consists of 500×500 pixels and originates from a number of preprocessing steps performed to extract the relevant protein structure. A typical example of the 12 images is shown in Figure 5.

On the basis of the 12 samples we computed the NPMLE $\hat{L}_{12}$ and the

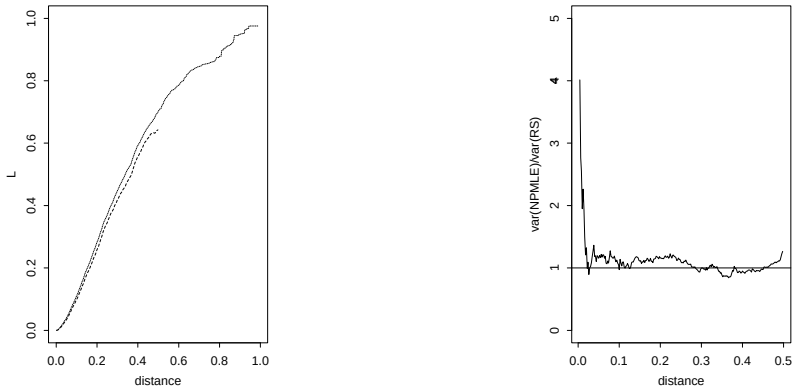

Figure 6: To the left estimates of the chord length distribution for the yoghurt example, NPMLE (dotted line) and reduced sample estimator (dashed line) are given. On the right an efficiency comparison of the NPMLE and the reduced sample estimator for the yoghurt example is performed.

reduced sample estimator $\tilde{L}_{12}$. The result is shown to the left in Figure 6. We also computed estimators $\hat{L}_1^{(k)}$ and $\hat{L}_1^{(k)}$ $(k = 1, 2, \ldots, 12)$ based on single images. On the basis of these we made pointwise estimates of the variances of the NPMLE and the reduced sample estimator. On the right of Figure 6 we plot the relative efficiency of the NPMLE to RS estimator. We notice that for very small distances the NPMLE is much worse. This behaviour was also observed in Figure 4.

In Hansen et al. (1996) a Boolean model of discs with constant intensity μ and constant disc radius R was fitted to data. This resulted in estimates $\hat{\mu} = 80$ and $\hat{R} = 0.03$. Note that these are exactly the parameter values we used to produce the lower left corner of Figure 2. We notice that Figure 6 and the lower left corner of Figure 3 look very much alike, although the appearance of Figure 5 and the lower left corner of Figure 2 is quite different.

7 The linear contact distribution

Recall that Ξ is a stationary random closed set in $\mathbb{R}^2$ and e is a unit vector in $\mathbb{R}^2$. The *contact segment* at a point outside Ξ in the direction e is the longest line-segment, starting at that point, which does not hit Ξ. The linear contact

distribution (in the direction e) is the distribution of the length of the contact segment (in the direction e) through the origin, given that the origin does not belong to Ξ. Let $\chi_e(s,t)$ denote the contact segment in the direction e originating at (s,t) and let $|\chi_e(s,t)|$ denote its length. The situation is illustrated in Figure 7.

We now formally define the linear contact distribution function. First, define

$$V_e(x) = \{(s,t) \in B \setminus \Xi : |\chi_e(s,t)| \leq x\}. \tag{7.1}$$

The linear contact distribution function is

$$H_e(x) = \frac{E|V_e(x)|}{E|B \setminus \Xi|} \tag{7.2}$$

The expectations are with respect to the distribution of Ξ. This definition does not depend on B. Again, for simplicity we take $e = (1,0)$ and drop it from our notation. Also, we take $B = [0,\tau] \times [0,1]$. We now proceed in a

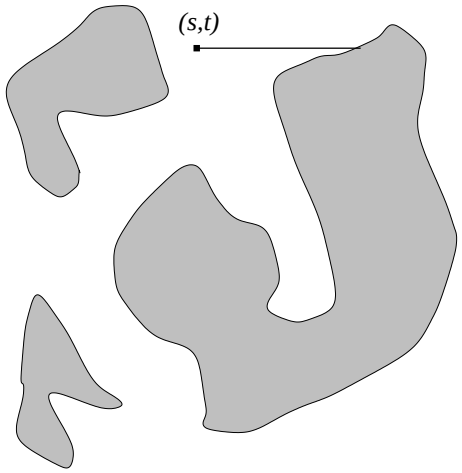

Figure 7: The contact segment $\chi_{(1,0)}(s,t)$. The shaded region is (part of) the random set Ξ.

similar way as before. In fact, as we are running out of useful letters we shall now *re-define* the functions $y(s,t)$ and $\delta(s,t)$ and the sets $A(s,t)$. Define

$$y(s,t) = |\chi(s,t) \cap B| \tag{7.3}$$

$$\delta(s,t) = \#\text{ endpoints of } \chi(s,t) \text{ outside of } B \tag{7.4}$$

to indicate a contact segment's observed length and censoring type. From observation of $\Xi \cap B$ we ascertain for all positive y and $\delta = 0$ or 1

$$A(y, \delta) = \{(s, t) \in B \setminus \Xi : y(s, t) \leq y,\ \delta(s, t) = \delta\}. \tag{7.5}$$

We define

$$G(y, \delta) = \frac{E|A(y, \delta)|}{E|B \setminus \Xi|}, \tag{7.6}$$

where the expectation is with respect to the distribution of Ξ. This definition should not be confused with (2.7). The sets $A(y, \delta)$ have a different meaning here.

Suppose that we choose a point (S, T) uniformly on B. Conditionally on the event that $(S, T) \in B \setminus \Xi$, H is the distribution function of $|\chi(S, T)|$ and G is the joint distribution function of $y(S, T) = |\chi(S, T) \cap B|$ and $\delta(S, T)$ (which is the number (0 or 1) of endpoints of $\chi(S, T)$ that fall outside of B).

Now consider the following related experiment. Let (S, T) be uniformly distributed on B and let X be independently distributed according to H. Define χ to be the line-segment between (S, T) and $(S + X, T)$. Define $Y = |\chi \cap B| = |[S, S+X] \cap [0, \tau]|$ and let Δ be the number of endpoints of χ outside of B. This set-up is the familiar random censorship model. The maximum likelihood estimator of H from observing n i.i.d. copies (Y_i, Δ_i) of (Y, Δ) is the well-known Kaplan–Meier estimator. The Kaplan–Meier estimator is a function of the empirical distribution of the (Y_i, Δ_i). In the original problem—estimation the linear contact distribution H from observing i.i.d. copies $\Xi_i \cap B$ of $\Xi \cap B$—we replace these empirical distribution functions with empirical ratios of areas of the form $\sum_i |A_i(y, \delta)| / \sum_i |B \setminus \Xi_i|$. The resulting estimator was introduced and studied in Hansen et al. (1996). There it is also compared to a reduced sample estimator.

The chord lenth distribution and the linear contact distribution are very closely related as we already noted in formula (5.2). Stoyan et al. (1987 p. 180) formulate this relation a little different. Let $\overline{\Xi^c}$ denote the closure of the complement of Ξ, which is again a stationary random closed set. Let L_0 denote the unweighted chord length distribution function of $\overline{\Xi^c}$ and let H be the linear contact distribution function of Ξ itself. Then

$$H(x) = \int_0^x (1 - L_0(t)) \mathrm{d}t / m \tag{7.7}$$

where m is the mean typical chord length

$$m = \int x \mathrm{d}L_0(x).$$

First of all we note from (7.7) that H is absolutely continuous with density $h(x) = (1 - L_0(x))/m$. We easily deduce

$$L_0(x) = 1 - h(x)/h(0). \tag{7.8}$$

Now from (7.7) it also follows that h is *monotone* or—equivalently—that H is concave. Non-parametric maximum likelihood estimation of a concave distribution function from right-censored observations is studied in Huang and Zhang (1994) and Huang and Wellner (1995). Using the NPMLE derived in these papers instead of the Kaplan–Meier estimator (and replacing empirical distribution functions by ratios of areas) we can improve the estimator of Hansen et al. (1996). Replacing the Kaplan–Meier estimator with its least concave majorant will also improve the estimator.

To estimate the chord length distribution of Ξ we can first estimate the distribution function and the density of the linear contact distribution of $\overline{\Xi^c}$. We can then apply (5.2) to obtain the weighted chord lenth distribution L or apply (7.8) to find the unweighted version L_0. However, it seems that this estimator will not improve the chord length estimator we obtained earlier, as density estimation is generally awkward. It is more fruitful to work in the other direction. First estimate the chord length distribution of $\overline{\Xi^c}$ and then use (7.7) to transform the estimate into an estimate of the linear contact distribution of Ξ. Intuitively, we expect such a transformation to perform better than direct estimates of the linear contact distribution. Namely, we can interpret a contact segment as a chord which has been subjected to multiplicative censoring. Hence, contact segments are inherently subject to added randomness.

To transform an estimate of L into one for H we note the following. By (2.3) and (4.6) we have, for $x < \tau$

$$L_0(x) = \frac{\int_0^x \frac{1}{t}\mathrm{d}L(t)}{\int_0^\infty \frac{1}{t}\mathrm{d}L(t)} = \frac{\int_0^x \frac{1}{t}\mathrm{d}L(t)}{\int_0^{\tau-} \frac{1}{t}\mathrm{d}L(t) + G/\tau}.$$

Also, note that

$$1 = \int \mathrm{d}L(t) = \int t\frac{1}{t}\mathrm{d}L(t) = \int t\mathrm{d}L_0(t) \int \frac{1}{t}\mathrm{d}L(t) = m(\int_0^{\tau-} \mathrm{d}L(t) + G/\tau)$$

hence

$$m = (\int_0^{\tau-} \mathrm{d}L(t) + G/\tau)^{-1}$$

These formulas combined with (7.7) express H on the interval $[0, \tau)$ in terms of L on the same interval and G. These we can estimate using formulae (4.16) and (4.17), where we recall that we indentified G with $L(\{\dagger\})$.

8 Simulations of the LCD

Suppose we observe n independent copies $\Xi_i \cap B$ of a random set $\Xi \cap B$. In this section we present a simulation study to compare two estimators of the linear contact distribution. To wit, $\hat{H}_n$ (the transformation of the chordlength NPMLE) and $\tilde{H}_n$ (a Kaplan–Meier estimator proposed in Hansen et al. (1996)). Similar to Section 5, we generated 100 realizations $\Xi_1, \ldots, \Xi_{100}$ of each of the Boolean models illustrated in Figure 2. On basis of, say, Ξ_k we have derived the chord length NPMLE $\hat{L}_1^{(k)}$ and transformed it into an estimate $\hat{H}_1^{(k)}$ of H. For each Ξ_k we also computed the Kaplan–Meier estimate $\tilde{H}_1^{(k)}$. The subscript "1" indicates that the estimators are based on single observations. In Figure 8 we show by thin lines the pointwise averages of $\hat{H}_1^{(k)}$ and $\tilde{H}_1^{(k)}$. In the same figure we also see a comparison of the estimated RMSE of the two estimators (obtained as described in Section 5). The picture is quite clear as the bias of the two estimators seems to be of same order but the transformed NPMLE is uniformly better than the Kaplan–Meier estimator.

Acknowledgement

We would like to express our thanks to Dr. Anne Skriver, Department of Dairy and Food Science, The Royal Veterinary and Agricultural University, Copenhagen for supplying the data set and to Richard Gill for his help with this paper from first to last. The research of the first author was supported by the European Union's research network "Statistical and Computational Methods for the Analysis of Spatial Data, ERB-FMRX-CT96-0095" and MaPhySto, Centre for Mathematical Physics and Stochastics, funded by a grant from the Danish National Research Foundation. Most of the second author's work was carried out while he was a PhD student at the University of Utrecht.

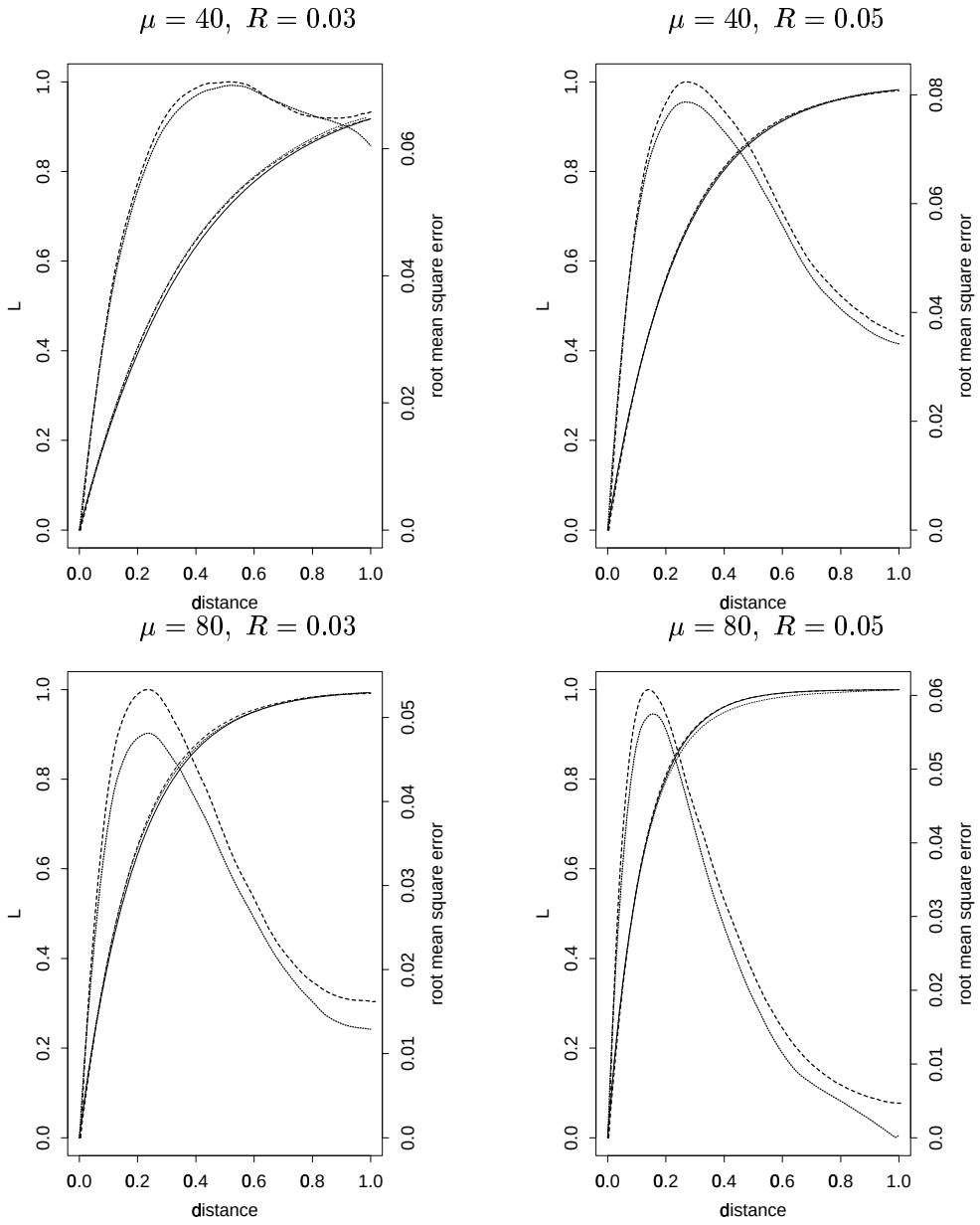

Figure 8: Root mean square error (RMSE) comparison of the transformed NPMLE and the Kaplan–Meier estimator. The thick dashed line is the estimated RMSE of the transformed NPMLE and the thick dotted line is the RMSE of the Kaplan–Meier estimator. The thin dotted line is the transformed NPMLE, the thin dashed line is the Kaplan–Meier estimator and the thin solid line is the estimand.

References

A.J. Baddeley and R.D Gill (1997). Kaplan–Meier estimators for interpoint distance distributions of spatial point processes. *Ann. Statist.* **25** 263–292.

G. Borgefors (1984). Distance transformations in arbitrary dimensions. *Comput. Vision, Graphics Image Process* **27** 321–345.

G. Borgefors (1986). Distance transformations in digital images. *Comput. Vision, Graphics Image Process* **34** 344–371.

S.N. Chiu and D. Stoyan (1998). Estimators of distance distributions for spatial patterns. *Statist. Neerlan.* **52** 239–246.

P.J. Bickel, C.A.J. Klaassen, Y. Ritov and J.A. Wellner (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore.

P. Delfiner (1972). A generalization of the concept of size. *J. Microsc.* **95** 203-216.

A.P. Dempster, N.M. Laird and D.B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

S.A. van de Geer (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21** 14–44.

R.D. Gill (1994). *Lectures on Survival Analysis.* In: D. Bakry, R.D. Gill and S. Molchanov, École d'Été de Probabilités de Saint Flour XXII-1992, P. Bernard (editor). Springer Lecture Notes in Mathematics 1581 115–241.

M.B. Hansen (1995). *Spatial Statistics for Network Structures in Processed Milk.* PhD dissertation, The Royal Veterinary and Agricultural University, Copenhagen. Dina Research Report no. 30.

M.B. Hansen (1996). Estimation of first contact distribution functions for spatial patterns in S-PLUS. In: Proceedings of COMPSTAT '96, A. Prat (editor). Physica-Verlag 295-300

M.B. Hansen, A.J. Baddeley and R.D. Gill (1996). Kaplan-Meier type estimators for linear contact distributions. *Scand. J. Statist.* **23** 129–155.

M.B. Hansen, A.J. Baddeley and R.D. Gill (1999). First contact distributions for spatial patterns: Regularity and estimation. Kaplan-Meier type estimators for linear contact distributions. *Adv. in Appl. Probab.* **31** 15–33.

Y. Huang and C.-H. Zhang (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **22** 1256–1274.

J. Huang and J.A. Wellner (1995). Estimation of a monotone density or monotone hazard under random censorship. *Scand. J. Statist.* **22** 3–33.

J. Kiefer and J. Wolfowitz (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

M.J. van der Laan (1995). Efficiency of the NPMLE in the line-segment problem. *Scand. J. Statist.* **23** 527–550.

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

G.M. Laslett (1982b). Censoring and edge effects in areal and line transect sampling of rock joint traces. *Math. Geol.* **14**, 125–140.

H. Lum, I. Huang and W. Mitzner (1994). Morphological evidence for alveolar recruitment during inflation at high transpulmonary pressure. *J. Appl. Physiol.* **68** 2280–2286.

L. Le Cam and G.L. Yang (1988). On the preservation of local asymptotic normality under infromation loss. *Ann. Statist.* **16** 483–520.

G. Matheron (1975). *Random Sets and Integral Geometry.* John Wiley & Sons, New York.

E.H. Oldmixon, J.P. Butler and F.G. Hoppin (1994). Semi-automated measurement of true chord length distributions and moments by video microscopy and image analysis. *J. Microsc.* **175** 60–69.

F.S. Rosenthal (1989). Aerosol recovery following breathholding derived from the distribution of chordlengths in pulmonary tissue. *J. Aerosol. Sci.* **20** 267–277.

D. Stoyan, W.S. Kendall and J. Mecke (1995). *Stochastic Geometry and Its Applications.* Wiley, Chichester, second edition.

A. SKRIVER (1995). *Characterization of Stirred Yoghurt by Rheology, Microscopy and sensory analysis.* PhD dissertation, Institute for Dairy Research, The Royal Veterinary and Agricultural University, Copenhagen.

A. SKRIVER, M.B. HANSEN AND K.B. QVIST (1997). Image analysis applied to electron micrographs of stirred youghurt. *J. Dairy Research*, **64**, 135-143.

A.W. VAN DER VAART (1998). *Asymptotic Statistics.* Cambridge University Press.

B.J. WIJERS (1995). Consistent non-parametric estimation for a one-dimensional line segment process observed in an interval. *Scand. J. Statist.* **22** 335–360.

E.W. VAN ZWET (1999). *Likelihood Devices in Spatial Statistics.* PhD dissertation, University of Utrecht.