Bernoulli **10**(3), 2004, 377–396

Laslett's line segment problem

ERIK W. VAN ZWET

University of Leiden, Department of Mathematics, P.O. Box 9512, 2300 RA Leiden, The Netherlands. E-mail: evanzwet@math.leidenuniv.nl

Our problem is to estimate the length distribution of fractures in a rock surface from a geological map. We do not fully observe the fractures because part of the rock surface is covered by vegetation, soil and water. The uncovered region is very irregular and, as a result, we tend to observe several pieces of a single fracture. It is quite impossible to decide from the map if two pieces belong to the same underlying fracture. Under the assumption that the observed pieces are independent, we derive the nonparametric maximum likelihood estimator of the length distribution of the underlying fractures. The assumption is clearly false, but our approach is justified by proving consistency of the estimator without appealing to the independence. We apply our estimator to the geological data.

Keywords: EM algorithm; missing data; nonparametric maximum likelihood estimation

1. Introduction

Almost two decades ago, a study was planned concerning the hazards of nuclear fuel waste disposal in underground excavations in selected plutonic rock masses of the Canadian Shield. A site within the granitic rock of the Lac du Bonnet batholith in southeastern Manitoba was selected to build an underground research laboratory. At this lab, experiments would be conducted related to thermal heating and hydraulic conductivity. For more information, we refer to a report by Stone *et al*. (1984). Figure 1 is part of a geological map in this report. It shows fractures in the rock surface of a 160 by 160 metre region at the Lac du Bonnet site. The statistical problem is to estimate the probability distribution of the lengths of these fractures, as an indication of the rock's permeability. The data from Figure 1 first entered the statistical literature with Chung (1989a; 1989b). Estimation of the length distribution of line segments observed through a bounded window is sometimes called *Laslett's line segment problem*, after Laslett (1982a; 1982b).

The one-dimensional version of Laslett's problem is also called the 'hospital problem'. The line segments represent the sojourns of patients in a hospital. We observe the presence of patients only during a fixed time interval and we are asked to estimate the distribution of the sojourn length. Note that sojourns can be 'doubly' censored, when we observe neither arrival nor departure of a given patient. Under the assumption that the patients arrive according to a homogeneous Poisson process, Laslett (1982a; 1982b) showed how the EM algorithm can be used to obtain the (sieved) nonparametric maximum likelihood estimator (NPMLE). Wijers (1995) has shown this estimator to be consistent. Gill (1994), van der Laan (1995) and Wijers (1995) have (jointly) established its asymptotic normality and efficiency.

1350–7265

Figure 1. Fractures in 160 m by 160 m granitic rock of the Lac du Bonnet batholith in Manitoba. From Stone *et al*. (1984). Digitized and post-processed by Professor A.J. Baddeley of the University of Western Australia. The irregular black region is the rock's surface. The white parts represent areas where the rock cannot be observed due to soil, vegetation or water. The white lines through the black regions indicate fractures.

With our geological map, we encounter three main difficulties. Firstly, many of the fractures are 'censored' that is, not fully observed, because the rock is only partly exposed due to vegetation, soil and water. Secondly, longer fractures have a higher probability of being (partly) observed than shorter ones. Thus, we have to account for selection bias. Finally, the area of exposed rock where we observe the cracks is not convex. This means that we might observe several pieces of a single crack. It is very difficult to assess from Figure 1 when two pieces belong to the same underlying fracture.

In Section 3 we derive the NPMLE of the fracture length distribution under the assumption that the observed *pieces* are independent. This assumption is clearly false, but our approach is justified in Section 4 by proving consistency of our estimator *without* appealing to the independence. In Section 3 we apply our estimator to the Canadian data.

2. Notation and statement of the problem

A two-dimensional line segment process is conveniently modelled by a point process $\Phi = \{(\vec{S}_i, X_i, \Theta_i)\}$ on $\mathbb{R}^2 \times \mathbb{R}^+ \times (-\pi/2, \pi/2)$. The $\vec{S}_i$ are the locations of the leftmost

endpoints (say) of the line segments. The X_i and Θ_i are the segments' lengths and orientations, respectively. We will use square brackets, writing $[\vec{s}, x, \theta]$, to denote a line segment in $\mathbb{R}^2$, rather than a point in $\mathbb{R}^2 \times \mathbb{R}^+ \times (-\pi/2, \pi/2)$.

Suppose that Φ is stationary with respect to shifts on $\mathbb{R}^2$ and that each segment's length and orientation are independent. To be precise, we assume that Φ has an intensity measure

$$\lambda \, \mathrm{d}\vec{s} \, \mathrm{d}F(x)\mathrm{d}J(\theta), \tag{1}$$

where $\lambda > 0$ and F and J are distribution functions on $\mathbb{R}^+$ and $(-\pi/2, \pi/2)$, respectively. Let μ denote the mean of the length distribution F, and suppose that μ is finite. Note that we have not completely described the distribution of Φ by giving its intensity. The distribution of a point process is only fully specified by the intensity measure if the process is Poisson.

We assume that the orientation distribution J is known and our goal is nonparametric estimation of the length distribution F. By 'nonparametric' we mean that we make no assumptions on F. Lok (1994) considers just the opposite situation: she assumes F to be known and estimates J. If both F and J are unknown we could alternate estimation of F and J.

For mathematical convenience we let the segment orientations range over $(-\pi/2, \pi/2)$, excluding vertical line segments with orientation $\pi/2$. There is no loss of generality, because we are assuming that J is known and hence we can choose the orientation of the line segment process so that J does not have a jump at $\pi/2$.

Let $\mathcal{W} \subset \mathbb{R}^2$ be a random closed set and let B be the unit square, $B = [-1/2, 1/2] \times [-1/2, 1/2]$, and suppose we can only observe the intersections of the line segments with $W = \mathcal{W} \cap B$. In the Canadian data set of Figure 1, B actually corresponds to the 160×160 m square area and W is the irregular black region through which we observe the fractures. When proving asymptotic properties of our estimator as B grows, we will assume that Φ and $\mathcal{W}$ are jointly ergodic.

3. A nonparametric estimator

In this section we derive the NPMLE of the fracture length distribution F under independence of the observed pieces. First we reparametrize the line segment process Φ into a new process Φ^* to make the problem one-dimensional. Next, we reparametrize F into a new distribution function V to account for the sampling bias due to longer fractures being more likely to be observed. In Section 3.3, we reparametrize again in order to put our problem in a form that is both identifiable and completely nonparametric. We will then have a standard nonparametric missing data problem for which we can write down the score equations. These can be solved using the EM algorithm (Dempster *et al*. 1977).

3.1. Two reparametrizations

Since our observation window W is not convex, we might observe several pieces of a single fracture. Instead of the fractures, we wish to treat these pieces as independent observations.

To this end, we reparametrize our problem. Let $\ell(r, \theta)$ denote the straight line at (signed) distance r from the origin at an angle θ with respect to the x-axis. If we define

$$r(\vec{s}, \theta) = -s_1 \sin\theta + s_2 \cos\theta,$$

then the line segment $[\vec{S}_i, X_i, \Theta_i]$ lies exactly on the line $\ell_i = \ell(r(\vec{S}_i, \Theta_i), \Theta_i)$. We assume that the intersection of W with ℓ_i is almost surely the disjoint union of a finite number, say $n(i)$, of line segments (intervals) ℓ_{ij} and we write $\ell_i \cap W = \cup_{j=1}^{n_i} \ell_{ij}$. Now let T_{ij} be the distance between $\vec{S}_i$ and the leftmost endpoint of ℓ_{ij}. We take T_{ij} to be negative when $\vec{S}_i$ is to the left of the leftmost endpoint of ℓ_{ij}. Also, let $L_{ij} = |\ell_{ij}|$ be the length of ℓ_{ij}. Thus, the intersection of Φ with W yields a new point process $\Phi^* = \{(T_{ij}, X_i, L_{ij})\}$ $(j = 1, \ldots, n(i))$ on $\mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+$. It is not difficult to show that the intensity measure of Φ^* is

$$\lambda \, \mathrm{d}t \, \mathrm{d}F(x) \mathrm{d}\nu(l) \tag{2}$$

where, for any Borel measurable set A, the measure ν is given by

$$\nu(A) = \int_r \int_\theta \#\{j : |\ell_j(r, \theta) \cap W| \in A\} \mathrm{d}r \, \mathrm{d}J(\theta), \tag{3}$$

with $\ell(r, \theta) \cap \mathcal{W} = \cup_j \ell_j(r, \theta)$ being a countable union of disjoint line segments.

Formula (2) may be understood as follows. Each fracture can be thought of as lying on an infinite straight line selected at random according to $\mathrm{d}r \, \mathrm{d}J(\theta)$. The fracture could be observed through any of the intervals that arise from this line intersecting W. The measure ν counts the expected number of such intervals. If we list the distances between the left endpoint of the fracture and the left endpoints of the intervals we have a homogeneous stationary point process on the line, as indicated by the Lebesgue measure $\mathrm{d}t$.

We have for the mean of ν,

$$\int_l l \, \mathrm{d}\nu(l) = \int_r \int_\theta |\ell(r, \theta) \cap W| \mathrm{d}r \, \mathrm{d}J(\theta) = \int_\theta |W| \mathrm{d}J(\theta) = |W|,$$

and we shall denote the total mass of ν by κ. Integrals with respect to ν can easily be estimated by Monte Carlo integration.

Our setting is now essentially one-dimensional, as all the pieces of fractures that we might observe correspond to intersections

$$[T_{ij}, T_{ij} + X_i] \cap [0, L_{ij}], \qquad j = 1, \ldots, n_i. \tag{4}$$

Of course many of these intersections are empty. To compute the expected number of non-empty intersections we define $A = \{(t, x, l) : [t, t + x] \cap [0, l] \neq \varnothing\}$ and integrate with respect to the intensity measure (2):

$$\int_A \lambda \, \mathrm{d}t \, \mathrm{d}F(x)\mathrm{d}\nu(l) = \int_l \int_x \int_{t\in[-x,l]} \lambda \, \mathrm{d}t \, \mathrm{d}F(x)\mathrm{d}\nu(l)$$

$$= \int_l \int_x (l+x)\lambda \, \mathrm{d}F(x)\mathrm{d}\nu(l)$$

$$= \int_l (l+\mu)\lambda \, \mathrm{d}\nu(l) = \lambda(|W| + \mu\kappa). \tag{5}$$

The (T_{ij}, X_i, L_{ij}) $(j = 1, \ldots, n_i)$ that fall in A are clearly not independent because for fixed i and varying j they refer to different pieces of the same fracture. However, as we noted earlier, it is very difficult to assess from Figure 1 when two pieces belong to the same underlying fracture. Hence the dependence among the (T_{ij}, X_i, L_{ij}) seems impossible to track. We will ignore this dependence and assume that the (T_{ij}, X_i, L_{ij}) are distributed on A according to a Poisson point process with intensity measure (2). Then, given their number, the (T_{ij}, X_i, L_{ij}) are distributed as the set of values in an independent and identically distributed (i.i.d.) sample from the normalized intensity

$$\mathbf{1}_A(t, x, l)\frac{\lambda \, \mathrm{d}t \, \mathrm{d}F(x)\mathrm{d}\nu(l)}{\displaystyle\int_A \lambda \, \mathrm{d}t \, \mathrm{d}F(x)\mathrm{d}\nu(l)} = \mathbf{1}_A(t, x, l)\frac{\mathrm{d}t \, \mathrm{d}F(x)\mathrm{d}\nu(l)}{|W| + \mu\kappa}. \tag{6}$$

We now address the problem of sampling bias: that longer fractures stand a better chance of being (partly) observed. The conditional distribution of X_i given $(T_{ij}, X_i, L_{ij}) \in A$, is not F but another distribution, say V, which is given by

$$V(x) = \int_{y\in(0,x]} \int_l \int_{t\in[-x,l]} \frac{\mathrm{d}t \, \mathrm{d}F(x)\mathrm{d}\nu(l)}{|W| + \mu\kappa} = \int_0^x \frac{|W| + y\kappa}{|W| + \mu\kappa} \, \mathrm{d}F(y). \tag{7}$$

Rewriting (6) in terms of V, we obtain

$$\mathbf{1}_A(t, x, l)\frac{\mathrm{d}t \, \mathrm{d}V(x)\mathrm{d}\nu(l)}{|W| + x\kappa} = \mathbf{1}_A(t, x, l)\frac{\mathrm{d}t}{l+x} \, \mathrm{d}V(x)\frac{l+x}{|W| + x\kappa} \, \mathrm{d}\nu(l). \tag{8}$$

3.2. Missing data problem

We are led to the following *missing data problem*. Suppose we have n i.i.d. copies (T_i, X_i, L_i) distributed according to (8). Suppose V is an unknown distribution function among a class $\mathcal{V}$ of all distribution functions on $\mathbb{R}^+$ with finite mean. The density (8) is easy to interpret:

1. First, select X according to V.
2. Given $X = x$, select L according to the density $(l + x)\mathrm{d}\nu(l)/(|W| + x\kappa)$.
3. Given $X = x$ and $L = l$, select T according to the uniform distribution on $(-x, l)$.

The (T_i, X_i, L_i) are the 'complete data'. The observed data are the intersections $[T_i, T_i + X_i] \cap [0, L_i]$. From these observations we wish to estimate V.

Figure 2 shows a line segment $[t, t + x]$ intersecting a line segment $[0, l]$. Regions which

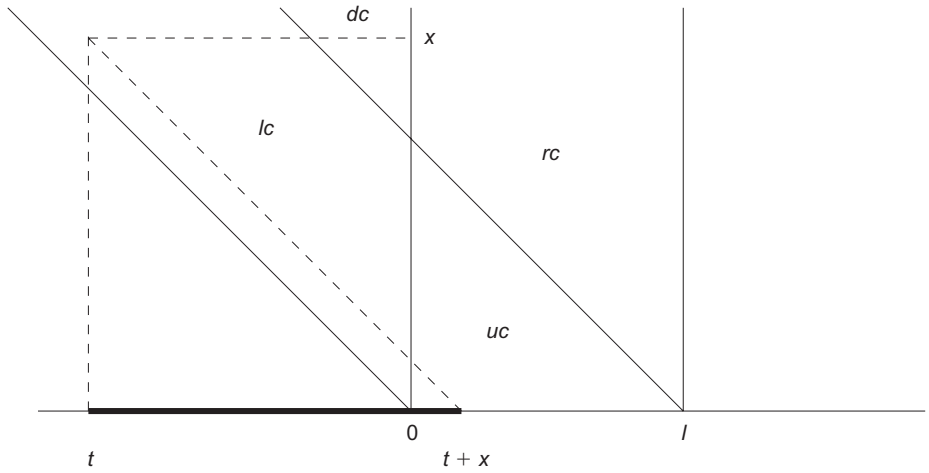

Figure 2. The different censoring types

correspond to different censoring types are indicated: uncensored (uc), left censored (lc), right censored (rc) and double censored (dc). For instance, if a point (t, x) falls in the 'left-censored' region (as illustrated), the left endpoint of the line segment $[t, t + x]$ will be not be observed through $[0, l]$.

At this point, we introduce some extra, artificial censoring. We choose a constant $\tau > 0$ and group together all observations whose length is greater than τ. We perform this grouping because observations of longer lengths become increasingly sparse, which leads to unstable estimation. Also, this fixed τ will be mathematically convenient when in Section 4 we consider increasing observation windows. Define

$$Y_i = \min(\tau, |[T_i, T_i + X_i] \cap [0, L_i]|),$$

$$\Delta_i = \begin{cases} \mathbf{1}_{\{T_i<0\}} + \mathbf{1}_{\{T_i+X_i>L_i\}} & \text{if } |[T_i, T_i + X_i] \cap [0, L_i]| \leqslant \tau, \\ 3 & \text{if } |[T_i, T_i + X_i] \cap [0, L_i]| > \tau. \end{cases}$$

Let P be the common distribution of the (Y_i, Δ_i). Because both the structure of the density (8) and the transformation from (T, X, L) to (Y, Δ) are very simple, we can explicitly evaluate P. We find P by first conditioning on L and integrating the density (8) over various regions such as indicated in Figure 3.

To describe P we introduce two useful functions:

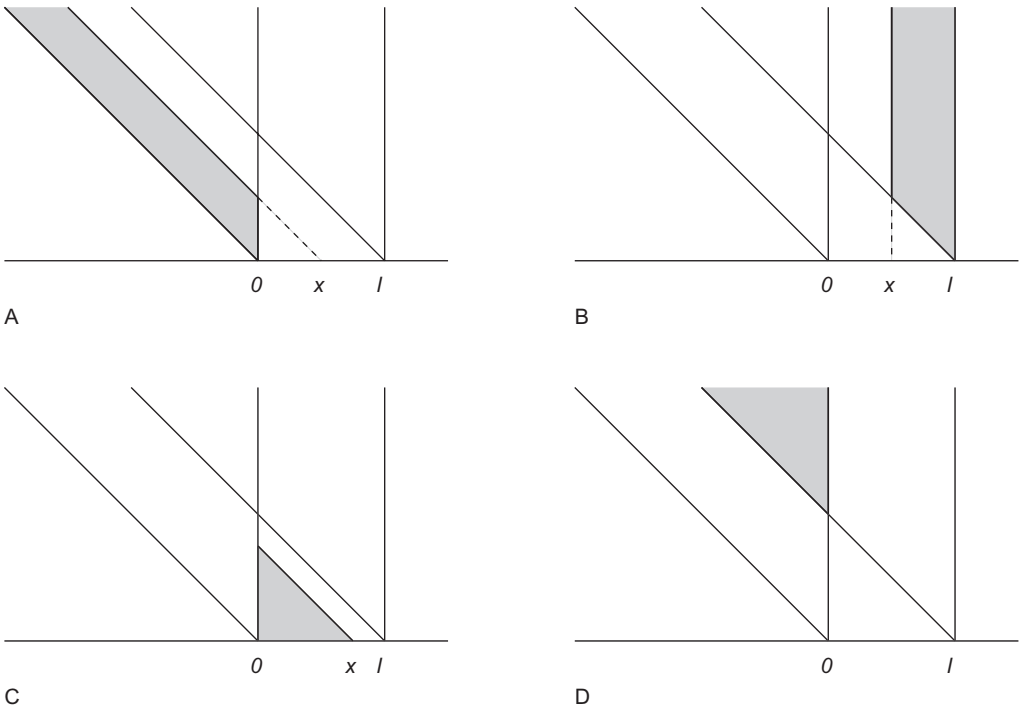

Figure 3. Integration regions for calculating P

$$g(x) = \int_x^\infty \frac{1}{|W| + \kappa y}\, \mathrm{d}V(y)$$

$$= \int_x^\tau \frac{1}{|W| + \kappa y}\, \mathrm{d}V(y) + g(\tau); \tag{9}$$

$$h(x) = \int_x^\infty \frac{y - x}{|W| + \kappa y}\, \mathrm{d}V(y)$$

$$= \int_x^\tau \frac{y - x}{|W| + \kappa y}\, \mathrm{d}V(y) + h(\tau) + (\tau - x)g(\tau). \tag{10}$$

It is easily verified that $\kappa h(x) + (|W| + \kappa x)g(x) + V(x) = 1$. We define $H = \kappa h(\tau)$ and $G = (|W| + \kappa\tau)g(\tau)$ so that

$$V(\tau) + G + H = 1. \tag{11}$$

As it turns out, P can be expressed in terms of V restricted to $[0, \tau]$, G and H. Defining

$$\alpha(x) = \frac{1}{|W| + \kappa x} \int_x^\infty (l - x)\mathrm{d}\nu(l), \tag{12}$$

we find – after much tedious manipulation – that, for all $y \leqslant \tau$,

$$P(\mathrm{d}y, 0) = \alpha(y)\mathrm{d}V(y) \tag{13}$$

$$P(\mathrm{d}y, 1) = 2\nu[y, \infty)g(y)\mathrm{d}y$$

$$= 2\nu[y, \infty)\mathrm{d}y\left(\int_y^\tau \frac{1}{|W| + \kappa x}\,\mathrm{d}V(x) + \frac{G}{|W| + \kappa\tau}\right) \tag{14}$$

$$P(\mathrm{d}y, 2) = \mathrm{d}\nu(y)h(y)$$

$$= \mathrm{d}\nu(y)\left(\int_y^\tau \frac{x - y}{|W| + \kappa x}\,\mathrm{d}V(x) + \frac{H}{\kappa} + \frac{(\tau - y)G}{|W| + \kappa\tau}\right) \tag{15}$$

and

$$P(Y = \tau, \Delta = 3) = \alpha(\tau)G + \frac{\nu(\tau, \infty)H}{\kappa}. \tag{16}$$

Because (11) expresses G in terms of $V(\tau)$ and H, the distribution P of the data is fully parametrized by V restricted to $[0, \tau]$ (we write $V|_{[0,\tau]}$) and H. $V|_{[0,\tau]}$ ranges over all (possibly defective) distribution functions on $[0, \tau]$, and H is any positive real number such that $V(\tau) + H \leqslant 1$.

We now demonstrate that there is a one-to-one correspondence between $(V|_{[0,\tau]}, H)$ and $(F|_{[0,\tau]}, \mu)$, where $F|_{[0,\tau]}$ ranges over all (possibly defective) distribution functions on $[0, \tau]$ and μ is any positive real not less than $\int_0^\tau x\mathrm{d}F(x)$. To express $F|_{[0,\tau]}$ and μ in terms of $V|_{[0,\tau]}$ and H, we note that the function $g(\cdot)$, defined in (9), is fully determined by $V|_{[0,\tau]}$ and H through (11). Also,

$$g(0) = \int_0^\infty \frac{1}{|W| + \kappa y}\,\mathrm{d}V(y) = \int_0^\infty \frac{1}{|W| + \kappa\mu}\,\mathrm{d}F(y) = \frac{1}{|W| + \kappa\mu}.$$

Hence,

$$\mu = \frac{1}{\kappa}\left(\frac{1}{g(0)} - |W|\right) \tag{17}$$

and

$$F(x) = \int_0^x \frac{|W| + \kappa\mu}{|W| + \kappa y}\,\mathrm{d}V(y) = \frac{g(0) - g(x)}{g(0)}. \tag{18}$$

3.3. Identifiability; another reparametrization

As we pointed out, the distribution of the data is fully parametrized by $V|_{[0,\tau]}$ and H. This means that our original model $\mathcal{V} = \{\text{all distributions on } \mathbb{R}^+\}$ is not identified. We could decide on an identifiable subset of $\mathcal{V}$, but that model would no longer be completely nonparametric. Why it is convenient if the model is both completely nonparametric and

identified is explained in the Appendix. We move to 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 as before.

Let us consider the – admittedly peculiar – space $[0, \tau] \cup \{\dagger, \ddagger\}$. By $\{\dagger, \ddagger\}$ we simply mean the addition of two arbitrary points to $[0, \tau]$. From (11) we know that $G + H \leqslant 1$ and therefore we may interpret G and H as probabilities. Hence, given a distribution function V on $\mathbb{R}^+$, we can define a measure Q on $[0, \tau] \cup \{\dagger, \ddagger\}$ by setting

$$Q(0, x] = V(x), \qquad \text{for all } x \in [0, \tau], \tag{19}$$

$$Q\{\dagger\} = G, \tag{20}$$

$$Q\{\ddagger\} = H. \tag{21}$$

Let $P(\mathrm{d}y, \delta|x)$ denote the conditional distribution of an observed length and censoring type, given that the underlying fracture has length x. Now, we define a Markov kernel which maps Q to P. For $y \leqslant \tau$,

$$K(\mathrm{d}y, \delta; x) = P(\mathrm{d}y, \delta|x),$$

$$K(\mathrm{d}y, \delta; \dagger) = \mathbf{1}_{\{\delta=1\}} \frac{2\nu[y, \infty)\mathrm{d}y}{|W| + \kappa\tau} + \mathbf{1}_{\{\delta=2\}} \frac{(\tau - y)\mathrm{d}\nu(y)}{|W| + \kappa\tau},$$

$$K(\mathrm{d}y, \delta; \ddagger) = \mathbf{1}_{\{\delta=2\}} \frac{\mathrm{d}\nu(y)}{\kappa}.$$

Also, set $K(Y = \tau, \Delta = 3; \dagger) = \alpha(\tau)$ and $K(Y = \tau, \Delta = 3; \ddagger) = \nu[\tau, \infty)/\kappa$, where $\alpha(\tau)$ is defined in (12). With Q defined in terms of V, one can easily verify that

$$P(\mathrm{d}y, \delta) = QK = \int K(\mathrm{d}y, \delta; x)\mathrm{d}Q(x).$$

The mapping $Q \mapsto QK$ is one-to-one on $\mathcal{Q} = \{\text{all distributions on } [0, \tau] \cup \{\dagger, \ddagger\}\}$. In other words, our new model $\mathcal{Q}$ is identified. In addition, $\mathcal{Q}$ is completely nonparametric.

3.4. EM algorithm

We are now in a position where we can apply the theory of the Appendix. We use the EM algorithm to estimate Q. Let $\mathbb{P}$ denote the empirical distribution of the observed pieces. A precise definition of $\mathbb{P}$ is given below in (30). The self-consistency equations

$$\hat{Q}(A) = \frac{1}{\text{number of observed pieces}} \sum_i \hat{Q}(A|Y_i, \Delta_i)$$

$$= \int \hat{Q}(A|y, \delta)\mathrm{d}\mathbb{P}(y, \delta)$$

become, for $x \leqslant \tau$,

$$\begin{aligned} \mathrm{d}\hat{Q}(x) &= \mathrm{d}\mathbb{P}(y,\,0) \\ &\quad + \frac{\mathrm{d}\hat{Q}(x)}{|W|+\kappa x}\int_{y=0}^{x}\frac{1}{\hat{g}(y)}\,\mathrm{d}\mathbb{P}(y,\,1) \\ &\quad + \frac{\mathrm{d}\hat{Q}(x)}{|W|+\kappa x}\int_{y=0}^{x}\frac{x-y}{\hat{h}(y)}\,\mathrm{d}\mathbb{P}(y,\,2) \end{aligned}$$

$$\hat{Q}\{\ddagger\} = \int_{y=0}^{\tau}\frac{1}{\hat{h}(y)}\,\mathrm{d}\mathbb{P}_n(y,\,2) \tag{22}$$

$$+ \frac{\nu[\tau,\,\infty)/\kappa}{\alpha(\tau)\hat{Q}\{\dagger\} + \nu[\tau,\,\infty)\hat{Q}\{\ddagger\}/\kappa}\mathbb{P}(\tau,\,3), \tag{23}$$

where

$$\hat{Q}\{\dagger\} = 1 - \hat{Q}(0,\,\tau] - \hat{Q}\{\ddagger\}$$

and (cf. (9) and (10)),

$$\hat{g}(x) = \int_{x}^{\tau}\frac{1}{|W|+\kappa y}\,\mathrm{d}\hat{Q}(y) + \frac{\hat{Q}\{\dagger\}}{|W|+\kappa\tau},$$

$$\hat{h}(x) = \int_{x}^{\tau}\frac{y-x}{|W|+\kappa y}\,\mathrm{d}Q(y) + \frac{\hat{Q}\{\ddagger\}}{\kappa} + \frac{(\tau-x)\hat{Q}\{\dagger\}}{|W|+\kappa\tau}.$$

Finally, using (17) and (18),

$$\hat{\mu} = \frac{1}{\kappa}\left(\frac{1}{\hat{g}(0)} - |W|\right) \tag{24}$$

$$\hat{F}(x) = \frac{\hat{g}(0)-\hat{g}(x)}{\hat{g}(0)}, \qquad x \leqslant \tau. \tag{25}$$

Our estimator solves score equations for a model which reflects the assumptions we have made. These assumptions include independence of the observed pieces, which certainly is not true. Also, we should be aware that solving score equations – even if they are the correct ones – does not necessarily produce a maximum likelihood estimator. Despite all this, we will refer to our estimator as an (or even 'the') NPMLE.

We have applied our estimator to the Canadian data. We assumed that the orientation distribution J is the uniform distribution on $(-\pi/2,\,\pi/2)$ and we used Monte Carlo integration to approximate the various integrals with respect to ν (these are: $|W|$, κ, $\nu[\tau,\,\infty)$ and $\alpha(\tau)$). We chose $\tau = 14\,\mathrm{m}$, which is about the length of the longest uncensored fracture we have observed. Next, we ran the EM algorithm with two different starting values. The algorithm was stopped as soon as the supremum distance between two subsequent iterates was less than 0.001. This took only a few minutes. Figure 4 shows the two resulting estimates of $F|_{[0,\tau]}$, together with our starting values. From top to bottom we see:

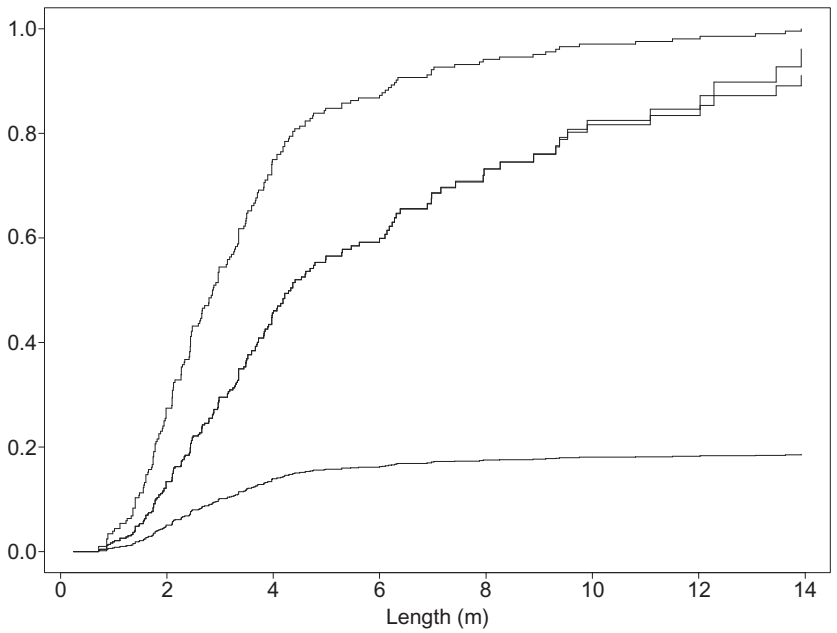

Figure 4. Some estimates of F restricted between 0 and 14 m. See description in the text.

- our first starting value, $\int_0^{\cdot} \mathbb{P}(\mathrm{d}y, 0)/\int_0^{\tau} \mathbb{P}(\mathrm{d}y, 0)$.
- the result of the EM algorithm, initialized at our first starting value. Convergence occurred after 19 iterations. The mean of F was estimated at 7.582 m.
- the result of the EM algorithm, initialized at our second starting value. Convergence occurred after 79 iterations. The mean of F was estimated at 7.565 m.
- our second starting value, $\int_0^{\cdot} \mathbb{P}(\mathrm{d}y, 0)$.

We note that our two estimates agree very well up to about 10 m. Beyond that the agreement is not as good, indicating, perhaps, that we chose τ too big or that we stopped too early.

4. Consistency

Define

$$B_n = nB = [-n/2,\ n/2] \times [-n/2,\ n/2]$$

and $B_\infty = \mathbb{R}^2$. We will now prove consistency of our estimator as we observe the line segment process Φ through an expanding sequence of observation windows $W_n = \mathcal{W} \cap B_n$, for $n = 1, 2, \ldots$.

The distribution $P = QK$ of our data depends on the observation window $\mathcal{W} \cap B_1$

through the measure ν (cf. (3)). To make the dependence on W_n explicit we will write ν_n and $P_n = Q_n K_n$. We denote our NPMLE of Q_n by $\hat{Q}_n$ and write $(\hat{F}_n|_{[0,\tau]}, \hat{\mu}_n)$ for the NPMLE of $(F|_{[0,\tau]}, \mu)$.

First, we will define the 'empirical distribution' $\mathbb{P}_n$ of the pieces we observe through the window W_n. Note that the subscript n of $\mathbb{P}_n$ does not refer to the number of observed pieces. In Lemma 2 we use an ergodic theorem to establish that $\mathbb{P}_n$ and the true $P_n = Q_n K_n$ converge. Then, we introduce a 'pilot' sequence $\tilde{Q}_n$ to 'guide' the NMPLE $\hat{Q}_n$. In Lemma 3 we show the convergence of $\tilde{Q}_n$ and the true Q_n. Next, in Lemma 4, we use the fact that the NPMLE $\hat{Q}_n$ solves certain score equations, to obtain convergence of the NPMLE $\hat{P}_n = \hat{Q}_n K_n$ and the pilot sequence $\tilde{P}_n = \tilde{Q}_n K_n$. By Lemma 3, this implies convergence of $\hat{Q}_n$ and Q_n, which is the content of Lemma 5. Finally, Theorem 2 concludes that the NPMLE $(\hat{F}_n|_{[0,\tau]}, \hat{\mu}_n)$ converges to $(F|_{[0,\tau]}, \mu)$.

The trick of using a pilot sequence to guide the NPMLE has previously been applied by Murphy (1994), Gill (1994), Wijers (1995) and Hansen and van Zwet (2001).

We shall make extensive use of an ergodic theorem for spatial processes due to Nguyen and Zessin (1979, Corollary 4.20). A spatial process on $\mathbb{R}^2$ is collection of random variables $\{X_B : B \in \mathcal{B}\}$ where $\mathcal{B}$ are the bounded Borel sets in $\mathbb{R}^2$. A spatial process is said to be *covariant* if a shift of the entire process accompanied by the same shift of the observation window B does not change the value of X_B. The process is said to be *additive* if, for disjoint $A, B \in \mathcal{B}$,

$$X_{A \cup B} = X_A + X_B.$$

We denote the collection of bounded and convex subsets of $\mathbb{R}^2$ by $\mathcal{K}$. Let $B_1 = [-1/2, 1/2]^2$ be the unit square.

Theorem 1 (Nguyen and Zessin). *If a spatial process $\{X_B : B \in \mathcal{B}\}$ is covariant and additive and if there exists a non-negative, integrable random variable Y such that*

$$|X_B| \leqslant Y, \quad a.s. \text{ for each } B \in B_1 \cap \mathcal{K},$$

then

$$\lim_{n \to \infty} \frac{1}{|B_n|} X_{B_n} = \mathrm{E}(X_{B_1}|\mathcal{J}), \quad a.s.$$

for each regular countable sequence B_n of sets in $\mathcal{K}$.

Our sequence $B_n = [-n/2, n/2]^2$ is an example of a regular countable sequence in $\mathcal{K}$. Here $\mathcal{J}$ denotes the invariant sigma-algebra. If $\{X_B : B \in \mathcal{B}\}$ is ergodic – that is, $\mathcal{J}$ is trivial – then the limit equals $\mathrm{E}(X_{B_1})$.

To begin with, we introduce some notation for Lemma 1 below. We intersect $\mathcal{W}$ with a straight line $\ell(r, \theta)$ at signed distance r from the origin and orientation θ. Suppose such intersections are always the countable union of disjoint segments (intervals): $\ell(r, \theta) \cap \mathcal{W} = \cup \ell_i(r, \theta)$. Let $\{\ell_i(r, \theta), i \in I_B\}$ be the collection of all those segments with left endpoint in a bounded Borel set B. Define measures $\nu^*(\mathrm{d}l; B)$, $B \in \mathcal{B}$, by

$$\nu^*(A;\, B) = \int_r \int_\theta \#\{i \in I_B : |\ell_i(r,\, \theta)| \in A\} \mathrm{d}r\, \mathrm{d}J(\theta).$$

Note the similarities and differences between $\nu_1(A)$ defined at (3) and $\nu^*(A;\, B_1)$.

Lemma 1. *Suppose the random closed set $\mathcal{W}$ is ergodic. Let f be a ν_n measurable function, for all n. The sequence*

$$\int_l f(l)\nu_n(\mathrm{d}l)/|B_n| \to \mathrm{E}\left(\int_l f(l)\mathrm{d}\nu^*(\mathrm{d}l;\, B_1)\right)$$

in probability.

Proof. Consider a collection of random variables $X_B = \int f(l)\nu^*(\mathrm{d}l;\, B)$, indexed by sets $B \in \mathcal{B}$. This spatial process is additive, covariant and ergodic (because $\mathcal{W}$ is ergodic). Hence, we can apply Theorem 1 to conclude that $X_{B_n}/|B_n|$ converges to $\mathrm{E}(X_{B_1})$. The difference between $\int f(l)\nu_n(\mathrm{d}l)$ and X_{B_n} comes from segments that cross the boundary of B_n. Since the expected number of such segments is $O(n)$, while $|B_n| = n^2$, it follows that

$$\left|\int f(l)\nu_n(\mathrm{d}l) - X_{B_n}\right|/|B_n| \to 0,$$

in probability and our claim follows. □

Intersecting the line segment process Φ with the random closed set $\mathcal{W}$, we obtain a collection of 'pieces' of line segments. These pieces are a new line segment process, say, $\Psi = ((\vec{S}_i,\, X_i,\, \Theta_i)\}$, of left endpoints $\vec{S}_i$, lengths X_i and orientations Θ_i. We write $[\vec{s},\, x,\, \theta]$ to indicate a line segment (subset of $\mathbb{R}^2$) rather than a point in a higher-dimensional space. Let B be a bounded Borel set. We define three functions:

$$p_B(\vec{s},\, x,\, \theta) = \text{left endpoint of } [\vec{s},\, x,\, \theta] \cap B, \tag{26}$$

$$y_B(\vec{s},\, x,\, \theta) = \min(\tau,\, |[\vec{s},\, x,\, \theta] \cap B|), \tag{27}$$

$$\delta_B(\vec{s},\, x,\, \theta) = \text{censoring type of } [\vec{s},\, x,\, \theta] \text{ relative to } \mathcal{W} \cap B. \tag{28}$$

As before, if $|[\vec{s},\, x,\, \theta] \cap B| > \tau$ we set $\delta_B(\vec{s},\, x,\, \theta) = 3$. For $0 \leqslant y \leqslant \tau$, $\delta = 0,\, 1,\, 2,\, 3$ we define

$$X_B(y,\, \delta) = \#\{(\vec{s},\, x,\, \theta) \in \Psi : [\vec{s},\, x,\, \theta] \cap B \neq \varnothing,\, y_B(\vec{s},\, x,\, \theta) \leqslant y,\, \delta_B(\vec{s},\, x,\, \theta) = \delta\}. \tag{29}$$

With $P_n = Q_n K_n$, note that

$$\int_0^y P_n(\mathrm{d}x,\, \delta) = \mathrm{E}\left(\frac{X_{B_n}(y,\, \delta)}{\sum_\delta X_{B_n}(\tau,\, \delta)}\right).$$

Observing $\Phi \cap \mathcal{W} \cap B_n$, we can evaluate $X_{B_n}(y,\, \delta)$ for all y and δ. The empirical distribution of these observations is defined by

$$\int_0^y \mathbb{P}_n(\mathrm{d}x,\,\delta) = \frac{X_{B_n}(y,\,\delta)}{\sum_{\delta} X_{B_n}(\tau,\,\delta)}. \tag{30}$$

We denote by $\|\cdot\|_A$ the uniform norm over a set A.

Lemma 2. *Suppose the random closed set $\mathcal{W}$ and the line segment process Φ are jointly ergodic. Then*

$$\left\| \int_0^{\cdot} \mathbb{P}_n(\mathrm{d}y,\,\delta) - \int_0^{\cdot} P_n(\mathrm{d}y,\,\delta) \right\|_{[0,\tau]} \to 0$$

for $\delta = 0, 1, 2$, and

$$|\mathbb{P}_n(Y = \tau,\,\Delta = 3) - P_n(Y = \tau,\,\Delta = 3)| \to 0$$

in probability.

Proof. Fix y and δ. Define, for all $B \in \mathcal{B}$,

$$X_B^*(y,\,\delta) = \#\{(\vec{s},\,x,\,\theta) \in \Psi : \vec{s} \in B,\, y_{\mathbb{R}^2}(\vec{s},\,x,\,\theta) \leqslant y,\, \delta_{\mathbb{R}^2}(\vec{s},\,x,\,\theta) = \delta\}.$$

The collection $\{X_B^*(y,\,\delta),\, B \in \mathcal{B}\}$ constitutes a covariant, additive, ergodic spatial process, so $X_{B_n}^*(y,\,\delta)/|B_n| \to \mathrm{E}X_{B_1}^*(y,\,\delta)$. The difference between $X_{B_n}(y,\,\delta)$ and $X_{B_n}^*(y,\,\delta)$ comes from segments that cross the boundary of B_n. Since the expected number of such segments is $O(n)$, while $|B_n| = n^2$, it follows that $|X_{B_n}(y,\,\delta) - X_{B_n}^*(y,\,\delta)|/|B_n| \to 0$ in probability. Hence, $X_{B_n}(y,\,\delta)/|B_n| \to \mathrm{E}X_{B_1}^*(y,\,\delta)$ in probability, and pointwise convergence of $\int_0^y \mathbb{P}_n(\mathrm{d}x,\,\delta)$ and $\int_0^y P_n(\mathrm{d}x,\,\delta)$ follows. The same argument holds if we replace $(0,\,y]$ by $(0,\,y)$. Hence, by monotonicity, the uniform convergence follows. □

We have obtained uniform convergence in probability, while below we really need almost sure convergence. The Skorohod–Dudley–Wichura almost-sure representation theorem (e.g. Shorack and Wellner, 1986) says that we can always construct an almost surely converging sequence whose marginal distributions are equal to a given weakly convergent sequence of probability measures on a metric space. Hence, we can initially pretend we have almost sure convergence and prove more almost sure convergence results. Afterwards, we must modify these results to convergence in probability.

Recalling (13), which states that

$$\mathrm{d}Q_n(x) = \frac{1}{\alpha_n(x)} P_n(\mathrm{d}x,\,0) = \frac{|W_n| + \kappa_n x}{\int_x^{\infty} (l - x)\mathrm{d}\nu_n(l)} \mathbb{P}_n(\mathrm{d}x,\,0), \qquad x \leqslant \tau,$$

we define a 'pilot' sequence $\tilde{Q}_n$ by

$$\mathrm{d}\tilde{Q}_n(x) = \frac{1}{\alpha_n(x)}\mathbb{P}_n(\mathrm{d}x,\, 0), \qquad x \leqslant \tau,$$

$$\tilde{Q}_n\{\ddagger\} = \min(\tilde{Q}_n(\tau,\, \infty),\, Q_n\{\ddagger\}). \tag{31}$$

Note that $\tilde{Q}_n$ is not a true estimator, because it cannot be computed from the data alone. Fortunately, it does not have to be. We define $\tilde{P}_n = \tilde{Q}_n K_n$ and note that $\tilde{P}_n(\cdot,\, 0) = \mathbb{P}_n(\cdot,\, 0)$.

Lemma 3. *Suppose the random closed set $\mathcal{W}$ and the line segment process Φ are jointly ergodic. Then the pilot sequence $\tilde{Q}_n$ and the true Q_n converge in the sense that*

$$\left\| \int_0^{\cdot} \mathrm{d}\tilde{Q}_n - \int_0^{\cdot} \mathrm{d}Q_n \right\|_{[0,\tau]} \to 0 \text{ and } |\tilde{Q}_n\{\ddagger\} - Q_n\{\ddagger\}| \to 0$$

in probability.

Proof. Integration by parts yields that $\tilde{Q}_n(0,\, x]$ is equal to

$$-\int_0^x \frac{\kappa_n \int_y^{\infty} (l-y)\mathrm{d}\nu_n(l) + (|W_n| + \kappa_n y)\nu_n(y,\, \infty)}{\left(\int_y^{\infty} (l-y)\mathrm{d}\nu_n(l)\right)^2} \mathbb{P}_n((0,\, y],\, 0)\mathrm{d}y$$

$$+ \mathbb{P}_n((0,\, \tau],\, 0)\frac{|W_n| + \kappa_n\tau}{\int_\tau^{\infty} (l-\tau)\mathrm{d}\nu_n(l)} + \mathbb{P}_n((0,\, x],\, 0)\frac{|W_n| + \kappa_n\tau}{\int_x^{\infty} (l-x)\mathrm{d}\nu_n(l)}.$$

A similar expression holds for Q_n, but with $\mathbb{P}_n$ replaced by P_n. If, for n large enough, $\nu_n(\tau,\, \infty)$ is positive, then the integrand will be bounded and we can use the dominated convergence theorem in combination with Lemma 2 to get the (pointwise) convergence of $\tilde{Q}_n(0,\, x]$ and $Q_n(0,\, x]$. The same argument holds if we replace $(0,\, x]$ by $(0,\, x)$. Hence, by monotonicity, the uniform convergence follows. □

We now come to the core of our consistency proof. We show that the NPMLE $\hat{P}_n$ and our pilot sequence $\tilde{P}_n = \tilde{Q}_n K_n$ converge. Together with the consistency of $\tilde{Q}_n$, this will yield consistency of the NPMLE.

Lemma 4. *Suppose the random closed set $\mathcal{W}$ and the line segment process Φ are jointly ergodic. Then*

$$\left\| \int_0^{\cdot} \hat{P}_n(\mathrm{d}y,\, \delta) - \int_0^{\cdot} \tilde{P}_n(\mathrm{d}y,\, \delta) \right\|_{[0,\tau]} \to 0$$

for $\delta = 0,\, 1,\, 2$, and

$$|\hat{P}_n(Y = \tau,\, \Delta = 3) - \tilde{P}_n(Y = \tau,\, \Delta = 3)| \to 0$$

in probability.

Proof. We prove a stronger result, namely that $\hat{P}_n$ and $\tilde{P}_n$ converge in Hellinger distance. The $\tilde{Q}_n$ only put mass on the uncensored observations. This ensures that $\hat{Q}_n \gg \tilde{Q}_n$. Moreover, the Radon–Nikodym derivative $\mathrm{d}\tilde{Q}_n/\mathrm{d}\hat{Q}_n$ is square integrable with respect to $\hat{Q}_n$. By Lemma 7, it follows that the submodel consisting of the straight line $t\tilde{Q}_n + (1-t)\hat{Q}_n$ $(0 \leqslant t \leqslant 1)$ is Hellinger differentiable. By Lemma 8, the straight line $t\tilde{Q}_nK_n + (1-t)\hat{Q}_nK_n$ $(0 \leqslant t \leqslant 1)$ is Hellinger differentiable as well. By Lemma 7 it has score $(\mathrm{d}\tilde{Q}_nK_n/\mathrm{d}\hat{Q}_nK_n) - 1$.

Our 'NPMLE' $\hat{Q}_n$ does not maximize the likelihood of our data. However, it *is* the solution of the score equations that belong to the model that says that our data (the pieces) are all independent. One of these score equations is

$$\int \left(\frac{\mathrm{d}\tilde{Q}_nK_n}{\mathrm{d}\hat{Q}_nK_n} - 1 \right) \mathrm{d}\mathbb{P}_n = 0. \tag{32}$$

We may expand this equation into

$$\int \left(\frac{\mathrm{d}\tilde{Q}_nK_n}{\mathrm{d}\hat{Q}_nK_n} \right) \mathrm{d}(\mathbb{P}_n - \tilde{Q}_nK_n) + \int \left(\frac{\mathrm{d}\tilde{Q}_nK_n}{\mathrm{d}\hat{Q}_nK_n} - 1 \right) \mathrm{d}\tilde{Q}_nK_n = 0. \tag{33}$$

It can be shown (cf. van de Geer, 1993) that the second summand is not less than the squared Hellinger distance between $\tilde{Q}_nK_n$ and $\hat{Q}_nK_n$. Hence, if we can show convergence to zero of the first summand, we can conclude that $\tilde{Q}_nK_n$ and $\hat{Q}_nK_n$ converge in Hellinger distance. So we consider the first summand

$$\begin{aligned}
\int \frac{\mathrm{d}\tilde{Q}_nK_n}{\mathrm{d}\hat{Q}_nK_n} \mathrm{d}(\mathbb{P}_n - \tilde{Q}_nK_n) &= \int \frac{\mathrm{d}\tilde{P}_n}{\mathrm{d}\hat{P}_n} \mathrm{d}(\mathbb{P}_n - \tilde{P}_n) \\
&= \int_0^\tau \frac{\mathrm{d}\tilde{P}_n(\mathrm{d}y, 0)}{\mathrm{d}\hat{P}_n(\mathrm{d}y, 0)} \mathrm{d}(\mathbb{P}_n(\mathrm{d}y, 0) - \tilde{P}_n(\mathrm{d}y, 0)) \\
&\quad + \int_0^\tau \frac{\tilde{g}_n(y)}{\hat{g}_n(y)} \mathrm{d}(\mathbb{P}_n(\mathrm{d}y, 1) - \tilde{P}_n(\mathrm{d}y, 1)) \\
&\quad + \int_0^\tau \frac{\tilde{h}_n(y)}{\hat{h}_n(y)} \mathrm{d}(\mathbb{P}_n(\mathrm{d}y, 2) - \tilde{P}_n(\mathrm{d}y, 2)) \\
&\quad + \frac{\tilde{P}_n(Y = \tau, \Delta = 3)}{\hat{P}_n(Y = \tau, \Delta = 3)} (\mathbb{P}_n(Y = \tau, \Delta = 3) - \tilde{P}_n(Y = \tau, \Delta = 3)).
\end{aligned}$$

Because $\mathbb{P}_n(\mathrm{d}y, 0) = \tilde{P}_n(\mathrm{d}y, 0)$, the first summand is zero. Convergence to zero of the fourth term is immediate from Lemma 2. Convergence of the second and third term is a matter of routine analysis. □

Lemma 5. *Suppose the random closed set $\mathcal{W}$ and the line segment process Φ are jointly ergodic. Then the NPMLE sequence $\hat{Q}_n$ and the true Q_n converge in the sense that*

$$\left\| \int_0^{\cdot} \mathrm{d}\hat{Q}_n - \int_0^{\cdot} \mathrm{d}Q_n \right\|_{[0,\tau]} \to 0 \quad \textit{and} \quad |\hat{Q}_n\{\ddagger\} - Q_n\{\ddagger\}| \to 0$$

in probability.

Proof. Because, by definition, $\hat{P}_n = \hat{Q}_n K_n$ we have

$$\mathrm{d}\hat{Q}_n(x) = \frac{1}{\alpha_n(x)} \hat{P}_n(\mathrm{d}x, 0).$$

Also, recall that $\mathbb{P}_n(\cdot, 0) = \tilde{P}_n(\cdot, 0)$, so that

$$\mathrm{d}\tilde{Q}_n(x) = \frac{1}{\alpha_n(x)} \mathbb{P}_n(\mathrm{d}x, 0) = \frac{1}{\alpha_n(x)} \tilde{P}_n(\mathrm{d}x, 0).$$

Hence we are in a very similar situation to Lemma 3. To prove our lemma we perform integration by parts and then use Lemma 4 instead of Lemma 2. □

We have arrived at our final result.

Theorem 2. *Suppose the random closed set $\mathcal{W}$ and the line segment process Φ are jointly ergodic. Then*

$$\|\hat{F}_n - F\|_{[0,\tau]} \to 0 \quad \textit{and} \quad |\hat{\mu}_n - \mu| \to 0$$

in probability.

Proof. By (17) and (18) it suffices to prove uniform convergence of

$$\hat{g}_n(x) = \int_x^{\tau} \frac{1}{|W_n| + \kappa_n y} \, \mathrm{d}\hat{Q}_n(y) + \frac{\hat{Q}_n\{\dagger\}}{|W| + \kappa\tau}$$

to $g(x)$, which is defined similarly in terms of Q_n. We have already established convergence of $\hat{Q}_n\{\dagger\}$ and $Q_n\{\dagger\}$. Integration by parts yields

$$\begin{aligned} \int_x^{\tau} \frac{1}{|W_n| + \kappa_n y} \, \mathrm{d}\hat{Q}_n(y) = &- \int_x^{\tau} \hat{Q}_n(0, y] \frac{-\kappa_n}{(|W_n| + \kappa_n y)^2} \, \mathrm{d}y \\ &+ \hat{Q}_n(0, \tau] \frac{1}{|W_n| + \kappa_n \tau} - \hat{Q}_n(0, x] \frac{1}{|W_n| + \kappa_n x}. \end{aligned}$$

We now use the previous lemma and the dominated convergence theorem (the integrand is evidently bounded) to obtain the convergence. The same argument holds if we replace the set $(x, \tau]$ by $[x, \tau]$. Hence, by monotonicity, the uniform convergence follows. □

Appendix. Nonparametric missing data problems

In this appendix we review some semi-parametric theory. See also Bickel *et al.* (1993) or van der Vaart (1998, Chapter 25).

Let X be a random variable on some (measurable) space $\mathcal{X}$ with unknown distribution $Q_0 \in \mathcal{Q}$ and let Y be a random variable in another space $\mathcal{Y}$. Using a Markov kernel K, we can define the 'mixture' $P = QK$ as

$$P(\mathrm{d}y) = QK(\mathrm{d}y) = \int_{\mathcal{X}} K(\mathrm{d}y, x)\mathrm{d}Q(x). \tag{34}$$

P is a probability measure on the space $\mathcal{Y}$. The Markov kernel K models a mechanism that causes some sort of 'missingness', for instance by means of (random) censoring. We suppose K is known and observe an i.i.d. sample $Y_1, \ldots, Y_n$ with common distribution $P_0 = Q_0K$. The model for the distribution of a generic observation Y is of course $\mathcal{P} = \mathcal{Q}K = \{QK : Q \in \mathcal{Q}\}$ and the objective 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 $QK = Q'K$ while $Q \neq Q'$. We have paid special attention to this in this paper.

If the model $\mathcal{P}$ is not dominated by a single sigma-finite measure, score functions are defined through 'differentiable submodels' of $\mathcal{P}$. The collection of all score functions is called the 'tangent set' of the model $\mathcal{P}$ at P and denoted $\dot{\mathcal{P}}(P)$. Since score functions are square integrable and integrate to zero (e.g. van der Vaart 1998, Lemma 25.14) the tangent set $\dot{\mathcal{P}}(P)$ can be identified with a subset of $L_2^0(P)$. We call a problem 'completely nonparametric' if $\dot{\mathcal{P}}(P) = L_2^0(P)$. This situation typically occurs when $\mathcal{Q}$ consists of *all* probability measures on $\mathcal{X}$.

We say that $\hat{P}_n = \hat{Q}_nK$ solves the score equations (and even call it the NPMLE) if

$$\int g \,\mathrm{d}\mathbb{P}_n = 0, \qquad \text{for all } g \in \dot{\mathcal{P}}\hat{P}_n). \tag{35}$$

We now list a few useful properties of $\mathcal{Q}$ and $\mathcal{P} = \mathcal{Q}K$ which are easy to prove.

Lemma 6. *If $\mathcal{Q}$ is convex then so is $\mathcal{P} = \mathcal{Q}K$.*

Lemma 7. *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 8. *If $t \mapsto Q_t$ is a differentiable submodel through $Q \in \mathcal{Q}$ with score function g, then $t \mapsto P_t = Q_tK$ is a differentiable submodel through $P = QK \in \mathcal{P}$ with score* $\mathrm{E}_Q(g(X)|Y)$.

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

By Lemma 8 the score equations for a missing data problem take the special form

$$\int \mathrm{E}_{\hat{Q}_n}(g(X)|y)\mathrm{d}\mathbb{P}_n(y) = \frac{1}{n}\sum_i \mathrm{E}_{\hat{Q}_n}(g(X_i)|Y_i) = 0, \qquad \text{for all } g \in \dot{\mathcal{Q}}(\hat{Q}_n). \tag{36}$$

If every square-integrable function is a score, then 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*

$$\hat{Q}_n(A) = \int \mathrm{E}_{\hat{Q}_n}(\mathbf{1}_A(X)|y)\mathrm{d}\mathbb{P}_n(y) = \frac{1}{n}\sum \hat{Q}_n(X_i \in A|Y_i). \tag{37}$$

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 $\{Q \in \mathcal{Q} : Q \ll Q_n^{(0)}\}$.

References

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

Chung, C.F. (1989a) Confidence bands for residual lifetime under random censorship model. *J. Multivariate Anal.*, **29**, 94–126.

Chung, C.F. (1989b) Confidence bands for quantile function under random censorship. *Ann. Inst. Statist. Math.*, **42**, 21–36.

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

Gill, R.D. (1994) Lectures on survival analysis. In P. Bernard (ed.), *Lectures on Probability Theory: École d'Été de Probabilités de Saint-Flour XXII – 1992*, Lecture Notes in Math. 1581, pp. 115–241. Berlin: Springer-Verlag.

Hansen, M.B. and van Zwet, E.W. (2001) Nonparametric estimation of the chord length distribution. *Adv. Appl. Probab.*, **33**, 541–558.

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

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

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

Lok, J. (1994) The NPMLE of the orientation distribution and comparison of estimators of the length distribution in the line segment problem. Master's dissertation, University of Utrecht.

Murphy, S.A. (1994) Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.*, **22**, 712–731.

Nguyen, X.X. and Zessin, H. (1979) Ergodic theorems for spatial processes. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **48**, 133–158.

Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes with Applications to Statistics.* New York: Wiley.

Stone, D., Kamineni, D.C. and Brown, A. (1984) Geology and fracture characteristics of the

underground research laboratory near Lac du Bonnet, Manitoba. Tech. Rep. 243, Atomic Energy of Canada Ltd. Research Corporation.

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

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

van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge: Cambridge University Press.

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

Received November 2001 and revised May 2003