Chapter 2

Laslett's line segments

2.1 Introduction

Almost to decades ago a study was to be made into the hazards of nuclear fuel waste disposal in underground excavations in selected plutonic rock masses of the Canadian Shield. Experiments had to be done related to thermal heating and hydraulic conductivity along fractures in the rock. As part of this research the construction was planned of an underground research laboratory within the granitic rock of the Lac du Bonnet batholith in southeastern Manitoba. For more information we refer to Stone et al. (1984). Figure 1.1 in the introduction to this thesis shows part of a map from that report of roughly 160 by 160 meters of fractures in the rock at the Lac du Bonnet site.

This particular data set enters the statistical literature with Chung (1989a and b). The statistical problem is to estimate from figure 1.1 the distribution of the lengths of the fractures. Estimation of the length distribution of line segments observed through a bounded window is sometimes called *Laslett's line segment problem* after Laslett (1982a and b).

We are not sure if the underground laboratory was ever built, but quite apart from the original motivation the estimation problem presents a very interesting statistical challenge. We encounter three main difficulties. First, we have to deal with censoring, since most of the fractures are only partly observed as the rock is only partly exposed due to vegetation, soil and water. Secondly, the sample of (partly) observed cracks is biased, because longer cracks stand a better chance of being observed than shorter ones. Thirdly, the area of exposed rock where we observe the cracks is not convex. This means that we might observe several fragments of a single crack. A single glance at Figure 1.1 will convince the reader that it would be very difficult to assess if two observed fragments belong to the same underlying fracture.

The first two problems, censoring and length bias, have been studied rather extensively Laslett (1982a and b), van der Laan (1993), Gill (1994) and Wijers (1995). The third difficulty, non-convexity of the observation window, is treated here for the first time. Also, in the above contributions the assumption is made that the locations of the fractures are distributed according to a Poisson process. We will obtain our results under far more general ergodicity assumptions.

A line segment process can be modeled as a marked point process, where the points indicate the locations of the segments and the marks their lengths and orientations. Assuming that the locations are scattered as a Poisson process greatly facilitates any statistical analysis because then, conditional on their number, the segments are independent. We will argue that an analysis under the Poisson assumption is still relevant when that assumption is not fulfilled. We feel that this is an important point and we do not want it to be snowed under by details. Therefore we give a general—albeit somewhat heuristic—discussion in the next section, which will then serve as a blueprint for the remainder of the chapter.

The reader will search in vain for the actual estimate of the length distribution of the Canadian fractures. As this thesis is on its way to the printer, we have not yet implemented our estimator and extracted the necessary data from Figure 1.1.

2.1.1 Inference for ergodic point processes

Consider a point process (see section 1.3) $\Phi = \{(T_i, X_i)\}$ on $\mathbb{R} \times \mathcal{X}$ with intensity measure

$$\mathrm{d}\Lambda_{\lambda,F}(t,x) = \lambda \mathrm{d}t \mathrm{d}F(x),$$

where F is known to belong to some class $\mathcal{F}$.

We think of Φ as a marked point process on $\mathbb{R}$ with marks X_i in a mark space $\mathcal{X}$. For instance, the T_i could be the locations of cars parked along a street and the X_i could be their make. For another example, the X_i could determine the length of a line segment starting at T_i. Evidently, this is a very general set up and in fact our restriction to T_i taking values in $\mathbb{R}$ is not at all necessary for our discussion.

We are primarily interested in estimating the mark distribution F_0 which is supposed to be in some collection $\mathcal{F}$. The 'nuisance' parameter λ is also assumed to be unknown to us.

We do not observe Φ entirely. Suppose that of points (T_i, X_i) outside a set $W \subseteq \mathbb{R} \times \mathcal{X}$ we observe nothing, while if $(T_i, X_i) \in W$ we only observe the result of some known function $Y_i = g(T_i, X_i)$. The mapping g may depend

on W. For instance, if the (T_i, X_i) represent line segments $[T_i, T_i + X_i]$ then $Y_i = g(T_i, X_i)$ could be $|[T_i, T_i + X_i] \cap [0, 1]|$, the length of the intersection with the unit interval. Then $W = \{(t, x) : [t, t + x] \cap [0, 1] \neq \emptyset\}$.

We now define a new 'point process' of observations

$$\Psi = \{Y_i\} = \{g(T_i, X_i) : (T_i, X_i) \in W\}.$$

It is convenient to think of the Y_i as a point process because the observed data will often be of similar structure as the incompletely observed model Φ.

Now suppose that Φ is *Poisson* (cf. section 1.3 on Poisson point processes). Then the intensity $\lambda \mathrm{d}t\mathrm{d}F(x)$ completely determines the distribution of Φ. The distribution of Ψ is also determined and can be described as follows.

Define, for all $F \in \mathcal{F}$

$$\mu_F = \int \int_W \mathrm{d}t\mathrm{d}F(x),$$

and assume it to be finite. The reader should mistake μ_F for the mean of F. Now,

1. let N be distributed as a Poisson random variable with mean $\lambda\mu_F$;
2. conditional on $N = n$ draw an i.i.d. sample (T_i, X_i) of size n from the 'normalized intensity' measure

$$\mathbf{1}_W(t, x)\frac{1}{\mu_F}\mathrm{d}t\mathrm{d}F(x); \tag{2.1}$$

3. compute $Y_i = g(X_i, T_i)$;
4. identify samples $Y_1, \ldots, Y_n$ that are equal up to ordering.

It is easy to write down the likelihood for the pair (λ, F) at the data $(N, Y_1, \ldots, Y_N)$.

$$\frac{e^{-\lambda\mu_F}}{N!}(\lambda\mu_F)^N \prod_{i=1}^{N} \iint_{(t,x)\in g^{-1}(Y_i)\cap W} \frac{1}{\mu_F}\mathrm{d}t\mathrm{d}F(x)N!.$$

We find the *profile likelihood* for estimating F by first fixing F and replacing λ by its maximum likelihood estimator N/μ_F. We find

$$e^{-N}N^N \prod_{i=1}^{N} \iint_{(t,x)\in g^{-1}(Y_i)\cap W} \frac{1}{\mu_F}\mathrm{d}t\mathrm{d}F(x).$$

This profile likelihood can be interpreted as the likelihood of a missing data problem (see section 1.4.4.) To make this clear we re-parameterize the problem. Define for all x

$$W(x) = \{t \in \mathbb{R} : (t, x) \in W\}.$$

Also define for all $F \in \mathcal{F}$ a transformation

$$\mathrm{d}V_F(x) = \frac{|W(x)|}{\mu_F}\mathrm{d}F(x). \tag{2.2}$$

For later use we have the inverse transformation

$$\mathrm{d}F_V(x) = \left(\int_{x'} \frac{1}{|W(x')|}\mathrm{d}V(x')\right)^{-1} \frac{1}{|W(x)|}\mathrm{d}V(x). \tag{2.3}$$

so that $F_{V_F} = F$. Define $\mathcal{V} = \{V_F,\ F \in \mathcal{F}\}$.

We can rewrite the distribution of (T, X) as

$$\mathbf{1}_W(t, x)\frac{\mathrm{d}t}{|W(x)|}\mathrm{d}V(x). \tag{2.4}$$

Our missing data problem takes the form

1. Sample X from $V \in \mathcal{V}$
2. Given $X = x$ sample T from the uniform distribution on $W(x)$
3. Set $Y = g(T, X)$

In line with section 1.4.4 we can also describe the situation as follows. Defining a Markov kernel

$$K(\mathrm{d}y; x) = \int_{W(x)} \delta_{g(t,x)}(\mathrm{d}y)\frac{\mathrm{d}t}{|W(x)|},$$

we have that

$$KV(\mathrm{d}y) = \int K(\mathrm{d}y; x)\mathrm{d}V(x)$$

is the distribution of $Y = g(T, X)$. The model for the observations Y_i is thus

$$K\mathcal{V} = \{KV : V \in \mathcal{V}\}.$$

Conditionally on $N = n$, the maximum likelihood estimator $\hat{V}_n$ satisfies

$$\int \log \frac{\mathrm{d}KV}{\mathrm{d}K\hat{V}_n}\mathrm{d}\mathbb{P}_n \geq 0 \qquad \forall V \in \mathcal{V}.$$

Suppose that F_0 is the 'true' parameter. The MLE $\hat{V}_n$ is of course supposed to estimate $V_0 = V_{F_0}$ (cf. (2.2)). We define $\hat{F}_n = F_{\hat{V}_n}$ (cf. (2.3)) as the MLE of F_0.

Now if the model $\mathcal{V}$—and hence $K\mathcal{V}$—is convex, we can attempt to use Theorem 1.2 to prove the convergence of $K\hat{V}_n$ to KV_0. To conclude convergence of $\hat{V}_n$ to V_0 we of course need to have identifiability: If $V \neq V'$ then $KV \neq KV'$. Ultimately, convergence of $\hat{V}_n$ to V_0 hopefully implies that $\hat{F}_n = F_{\hat{V}_n}$ tends to $F_{V_{F_0}} = F_0$.

To use Theorem 1.2 we need a sequence V_n such that KV_n tends to KV_0 and prove

$$\int \frac{\mathrm{d}KV_n}{\mathrm{d}K\hat{V}_n} \mathrm{d}(\mathbb{P}_n - KV_n) \to 0. \tag{2.5}$$

Such a proof, and indeed the choice of V_n, will rely on the fact that $\mathbb{P}_n$ converges to KV_0.

But what if Φ is *not* Poisson? Then we can no longer condition on the number of observations and treat them as an i.i.d. sample. We cannot write down the likelihood and maximize it. Also, the above asymptotics involving the number of observations tending to infinity become meaningless.

In spite of all this, we propose to use exactly the same estimator. Though no longer the maximum likelihood estimator, one still expects it to have nice properties. We explain.

First, some notation. We can consider Ψ as a random set of points but also as a random measure. By $\Psi(A)$ we mean the number of points of Ψ that fall in some measurable set A. In fact, we write $\Psi(A) = \int_A \mathrm{d}\Psi(y)$. Upon normalization, we obtain a random probability measure

$$\frac{\Psi(\mathrm{d}y)}{\Psi(g^{-1}(W))}.$$

Now we can define a 'maximum likelihood estimator' $\hat{V}$ as satisfying

$$\int \left(\log \frac{\mathrm{d}KV}{\mathrm{d}K\hat{V}} \right) \mathrm{d} \frac{\Psi(y)}{\Psi(g^{-1}(W))} \geq 0 \qquad \forall V \in \mathcal{V},$$

and we define $\hat{F} = F_{\hat{V}}$. Of course, all this is just notation; the estimator we just defined coincides exactly with the maximum likelihood estimator we defined earlier for the case where Φ is Poisson.

We will study the asymptotics of the present estimator as more and more of the underlying process Φ is revealed. Suppose we have a sequence $W = W_1 \subseteq W_2 \subseteq, \ldots$. Define

$$\Psi_i = \{g_i(T_i, X_i) : (T_i, X_i) \in W_i\}.$$

Note that the g_i may depend on W_i. Since we defined $W = W_1$ we have $\Psi = \Psi_1$.

Define

$$\mu_{i,F} = \int \int_{W_i} \mathrm{d}t \mathrm{d}F(x),$$

and

$$W_i(x) = \{t \in \mathbb{R} : (t, x) \in W_i\}.$$

Similarly to (2.2) and (2.3), define

$$\mathrm{d}V_{i,F}(x) = \frac{|W_i(x)|}{\mu_{i,F}} \mathrm{d}F(x) \tag{2.6}$$

and

$$\mathrm{d}F_{i,V}(x) = \left(\int_{x'} \frac{1}{|W_i(x')|} \mathrm{d}V(x') \right)^{-1} \frac{1}{|W_i(x)|} \mathrm{d}V(x). \tag{2.7}$$

Define

$$\mathcal{V}_i = \{V_{i,F},\ F \in \mathcal{F}\}.$$

Finally, define a Markov kernel

$$K_i(\mathrm{d}y; x) = \int_{W_i(x)} \delta_{g_i(t,x)}(\mathrm{d}y) \frac{\mathrm{d}t}{|W_i(x)|}.$$

We now have a sequence of statistical experiments: observation of Ψ_i. The model for the distribution of the data *under the Poisson assumption* changes with each experiment as the mapping g_i may depend on the 'window' W_i. Thus we have a sequence of models

$$K_i \mathcal{V}_i = \{K_i V : V \in \mathcal{V}_i\} = \{K_i V_{i,F} : F \in \mathcal{F}\}.$$

We can define a sequence of estimators $\hat{V}_i$ by requiring

$$\int \left(\log \frac{\mathrm{d}K_i V}{\mathrm{d}K_i \hat{V}_i} \right) \mathrm{d} \frac{\Psi_i(y)}{\Psi_i(g_i^{-1}(W_i))} \geq 0 \qquad \forall V \in \mathcal{V}_i.$$

We set $\hat{F}_i = F_{i,\hat{V}_i}$

We would like to use the ideas behind theorem 1.2 to prove that the $\hat{F}_i$ tend to F_0. However, first we must try to use ergodic properties of Φ and the sequence W_i to show that for all $F \in \mathcal{F}$ there exist $V_{\infty,F}$ and $K_\infty V_{\infty,F}$ such that

$$V_{i,F} \to V_{\infty,F} \quad \text{and} \quad K_i V_{i,F} \to K_\infty V_{\infty,F}. \tag{2.8}$$

These convergences are established for Laslett's problem in section 2.5.3. There it is also made clear in which sense they hold.

Next, we need a sequence V_i such that $K_i V_i$ tends to $K_\infty V_{\infty,F_0}$. Then we show that

$$\int \left(\frac{\mathrm{d}K_i V_i}{\mathrm{d}K_i \hat{V}_i} \right) \mathrm{d} \left(\frac{\Psi_i(y)}{\Psi_i(g^{-1}(W))} - K_i V_i \right) \to 0. \tag{2.9}$$

From this we can conclude that $K_i \hat{V}_i$ converges to $K_\infty V_{\infty,F_0}$. Identifiability will then imply that also $\hat{V}_i$ tends to V_{∞,F_0}. This will (at least for Laslett's problem) imply that $\hat{F}_i = F_{i,\hat{V}_i}$ tends to $F_{\infty,V_{\infty,F_0}} = F_0$.

For the convergence of (2.9), we will need that

$$\frac{\Psi_i(\mathrm{d}y)}{\Psi_i(g^{-1}(W))} \to K_\infty V_{\infty,F_0}(\mathrm{d}y).$$

For Laslett's problem this is shown in section 2.5.3.

2.1.2 Laslett's line segment problem

A line segment process is conveniently modeled by a marked point-process in the plane, where the points indicate the location of, say, the left endpoints and the marks indicate orientation and length. We consider a point process $\Phi = \{(\vec{S}_i, X_i, \Theta_i)\}$ on $\mathbb{R}^2 \times \mathbb{R}^+ \times (-\pi/2, \pi/2)$. We could call $\mathbb{R}^+ \times (-\pi/2, \pi/2)$ the mark-space. Take the $\vec{S}_i$ to be left endpoints of the line segments and let X_i and Θ_i be their lengths and orientations. Let us use square brackets, writing $[s, x, \vartheta]$, 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 $\mathbb{R}^2$). Also, suppose that each segment's length and orientation are independent, in the sense that Φ has an intensity measure of the form $\lambda \mathrm{d}\vec{s}\mathrm{d}F(x)\mathrm{d}K(\vartheta)$, where $\lambda > 0$ and F and K are distribution functions on $\mathbb{R}^+$ and $(-\pi/2, \pi/2)$, respectively. Let μ denote the mean of F, and suppose that it is finite. Note that by giving this intensity we have not completely described the distribution of Φ.

Now we assume that K is known. Our goal is nonparametric estimation of the length distribution F. This means that F belongs to a *model*, say $\mathcal{F}$, which consists of all probability distributions on the $\mathbb{R}^+$. Lok (1994) considers just the opposite situation: she assumes F to be known, and estimates K. If both F and K are unknown we could alternate estimation of F as presented here and estimation of K as presented by Lok (1994).

As a technical aside we point out the following. For reasons of mathematical convenience we decided to let the orientations range in $(-\pi/2, \pi/2)$,

excluding vertical line segments with orientation $\pi/2$. However, there is no loss of generality in doing so, because we can choose the orientation of the entire process so that K does not have a jump at $\pi/2$.

We have defined a stationary line segment process and stated our aim to estimate the distribution of the lengths of the line segments. We now describe the data that are available to us. Let $\mathcal{W} \subset \mathbb{R}^2$ be a random closed set and let B be the unit square; $B = [0,1] \times [0,1]$. Suppose we can only observe the intersections of the line segments with $W = \mathcal{W} \cap B$. In the Canadian dataset, c.f. Figure 1.1, B corresponds to the 160×160 meters square area and W is the irregular black region through which we observe the fractures.

After we come up with an estimator of F, shall study its asymptotic properties. In spatial statistics basically two types of asymptotics are usually considered (Cressie (1991) p. 100). The first is 'infill' asymptotics where the observation region is kept fixed, but within that region the number of observations increases. The other is 'increasing-domain' asymptotics, where the observation region is expanded. In fact, we shall define $B_n = [0,n] \times [0,n]$ and consider observation of the line segment process Φ through $\mathcal{W} \cap B_n$ as n tends to infinity. When doing so, we shall have to impose some ergodicity assumptions on Φ and $\mathcal{W}$.

2.1.3 Some history

We now present a brief overview of the work that has already been done by other authors. Much effort has been put into the one dimensional case where line segments are scattered according to a Poisson process on the real line and the segments are observed (without occlusion) through an interval. Laslett (1982 a) showed how the EM algorithm can be used to obtain the (sieved) nonparametric maximum likelihood estimator. Wijers (1995b) has shown it to be consistent. Gill (1994), van der Laan (1995) and Wijers (1995a) have (jointly) established its asymptotic normality and efficiency. Many of the methods that have been developed for the one-dimensional case carry over to the two dimensional case, as long as the observation window is convex. The trick is to subdivide the plane into parallel strips of infinitesimal width and doing so in every direction. Then through each strip we observe—as it were—a one-dimensional line segment process and then we integrate over all strips and all directions. We shall perform such calculations later on. At that time this approach will be made more precise.

Two things should be noted about assuming Φ to be Poisson and W to be convex. First, the Poisson assumption allows us to condition on the number of observations, after which we have an i.i.d. sample and we can apply standard estimation techniques. Secondly, the fact that the observation window W is

convex ensures that we can not observe more than one fragment of a single underlying line segment.

2.1.4 Scope of our contribution

We study the two dimensional line segment process observed through a non-convex window. This is prompted by the fact that the exposed rock surface of the Canadian data is obviously non-convex.

If the observation window is non-convex we might observe several fragments of a single line segment. It may occur, as it does with the Canadian data, that it is near impossible to decide if two fragments belong to the same fracture. As a consequence the observed fragments may be dependent in a way unknown to us.

In this situation it is of no use to assume that the segments are scattered as a Poisson process. Even if we could condition on the number of different segments that are observed, the fragments are not necessarily distributed as an i.i.d. sample.

However, we propose that the statistician analyze the data "as if" the observed fragments are scattered according to a Poisson process. In this simpler situation we find the non-parametric maximum likelihood estimator of the length distribution of the fractures. We shall show consistency of the estimator without using the Poisson assumption.

Under the Poisson assumption the estimator may well be efficient, though we have not tried to prove this. Without the Poisson assumption, it will certainly not be efficient because we effectively ignore dependencies among the data, thus throwing away information. However, we feel confident that this loss of information is minor.

2.2 A re-parameterization

Above we briefly explained how to extend the one-dimensional case to the two-dimensional case. We now demonstrate in detail how this works. The main step is to re-parameterize the location of each line segment relative to the unique line in the plane on which the segment lies.

Recall that we are studying a point-process $\Phi = \{(\vec{S}_i, X_i, \Theta_i)\}$ on $\mathbb{R}^2 \times \mathbb{R}^+ \times (-\pi/2, \pi/2)$, with finite intensity $\lambda \mathrm{d}\vec{s}\mathrm{d}F(x)\mathrm{d}K(\vartheta)$. The $\vec{S}_i$ denote left endpoints of line segments of length X_i and orientation Θ_i. It is straightforward to represent the locations of the left endpoints relative to rotated

coordinate axes. Define two functions

$$t_1(s_1, s_2, \vartheta) = s_1 \cos(\vartheta) + s_2 \sin(\vartheta)$$
$$t_2(s_1, s_2, \vartheta) = -s_1 \sin(\vartheta) + s_2 \cos(\vartheta).$$

For fixed ϑ this map represents $(s_1, s_2) \in \mathbb{R}^2$, with respect to the coordinate axes tilted counter-clockwise over an angle ϑ, cf. Figure 2.1. Now consider the

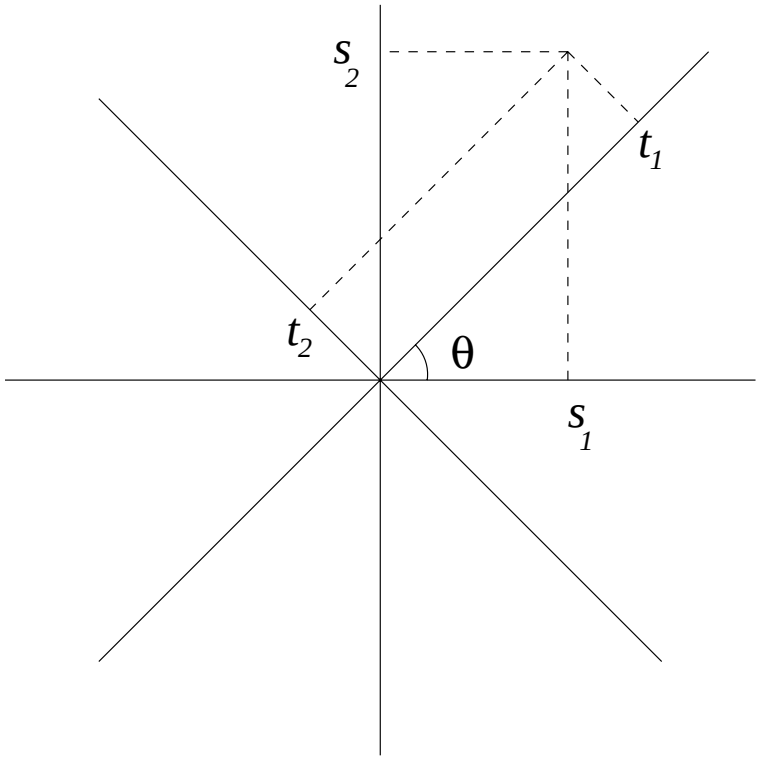

Figure 2.1: Re-parameterization of locations.

map that assigns $(t_1(s_1, s_2, \vartheta), t_2(s_1, s_2, \vartheta), x, \vartheta)$ to (s_1, s_2, x, ϑ). The image of Φ under this mapping is a new point-process $\Phi' = \{(\vec{T}_i, X_i, \Theta_i)\}$ on $\mathbb{R}^2 \times \mathbb{R}^+ \times (-\pi/2, \pi/2)$ with intensity

$$\lambda \mathrm{d}\vec{t}\mathrm{d}F(x)\mathrm{d}K(\vartheta).$$

This follows since Lebesgue measure is invariant under rotation.

Write $\ell(d, \vartheta)$ for the line with orientation ϑ at (signed) distance d from the origin. In our new parameterization, a line segment $[\vec{t}, x, \vartheta]$ lies exactly on the line $\ell(t_2, \vartheta)$.

Denote $\mathcal{W}(d,\vartheta) = \mathcal{W} \cap \ell(d,\vartheta)$. Let us assume that $\mathcal{W}(d,\vartheta)$ is always the union of at most a countable number of intervals: $\mathcal{W}(d,\vartheta) = \cup_m \mathcal{W}(d,\vartheta,m)$. The situation is depicted in Figure 2.2.

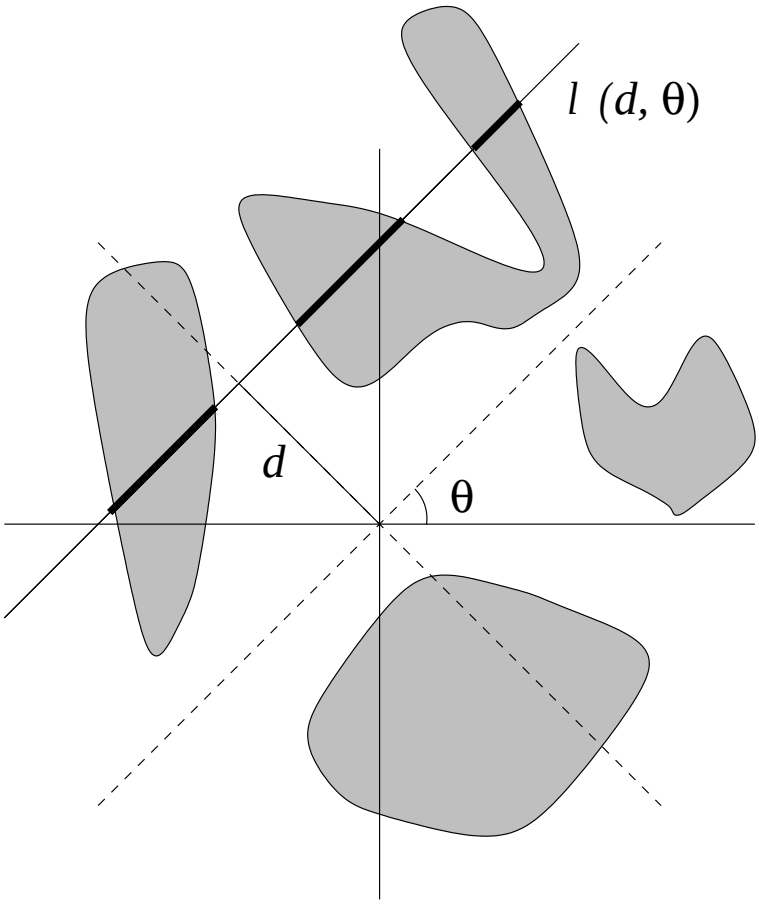

Figure 2.2: The line $\ell(d,\vartheta)$ intersecting (part of) the random set $\mathcal{W}$ (grey) producing closed intervals $\mathcal{W}(d,\vartheta,m)$ (bold).

Since $\mathcal{W}$ is non-convex, the intersection of a line segment $[\vec{t},x,\vartheta]$ with $\mathcal{W}$ might consist of several *fragments*. In other words, we may have a non-empty intersection of $[\vec{t},x,\vartheta]$ with $\mathcal{W}(t_2,\vartheta,m)$ for several values of m. Conversely, every fragment corresponds to a line segment $[\vec{t},x,\vartheta]$ and a positive number m. Our use of the words 'fragment' and 'segment' will have always precisely this meaning.

Consider countably many identical copies of $\Phi' = \{(\vec{T}_i, X_i, \Theta_i)\}$, one for each m. The result, $\Phi'' = \Phi' \times \mathbb{N}$ is a stationary point-process on $\mathbb{R}^2 \times \mathbb{R}^+ \times$

$(-\pi/2, \pi/2) \times \mathbb{N}$ with intensity

$$\lambda \mathrm{d}\vec{t}\mathrm{d}F(x)\mathrm{d}K(\vartheta)\mathrm{d}\mathcal{C}(m),$$

where $\mathcal{C}$ denotes counting measure on $\mathbb{N}$. This new point-process Φ'' can represent any fragment that could possible arise from the line segments of Φ'. A fragment $[\vec{t}, x, \vartheta, m]$ is the intersection of a line segment $[\vec{t}, x, \vartheta]$ with $\mathcal{W}(t_2, \vartheta, m)$. Of course such an intersection may very well be empty.

We must now deal with further edge effects because we observe the line segments only through the intersection of $\mathcal{W}$ with the unit square $B = [0,1] \times [0,1]$. For all $(d, \vartheta, m) \in \mathbb{R} \times (-\pi/2, \pi/2) \times \mathbb{N}$ let

$$W(d, \vartheta, m) = \mathcal{W}(d, \vartheta, m) \cap B$$

Define

$$\mathcal{D} = \{(\vec{t}, x, \vartheta, m) \in \mathbb{R}^2 \times \mathbb{R}^+ \times (-\pi/2, \pi/2) \times \mathbb{N} : W(t_2, \vartheta, m) \neq \emptyset\}.$$

Only points in $\mathcal{D}$ can correspond to fragments that could be observed through W.

We define two functions on $\mathcal{D}$, $t : \mathcal{D} \to \mathbb{R}$ and $l : \mathcal{D} \to \mathbb{R}^+$, as follows. Let $t(\vec{t}, x, \vartheta, m)$ be the distance from the left endpoint of the line segment $[\vec{t}, x, \vartheta]$ to the left endpoint of $W(t_2, \vartheta, m)$, which we take negative if the first is to the left of the latter. Let $l(\vec{t}, x, \vartheta, m)$ be the (strictly positive) length of the interval $W(t_2, \vartheta, m)$.

Now consider the map that assigns $(t(\vec{t}, x, \vartheta, m), x, l(\vec{t}, x, \vartheta, m))$ to each 'potential fragment' $(\vec{t}, x, \vartheta, m) \in \mathcal{D}$. The image under this map of $\Phi'' \cap \mathcal{D}$ is a new point-process $\Psi = \{(T_i, X_i, L_i)\}$ on $\mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+$. The points of Ψ can be interpreted as segments $[T_i, T_i + X_i]$ on the real line which can be observed through intervals $[0, L_i]$. These intersections $[T_i, T_i + X_i] \cap [0, L_i]$ are the observed fragments.

We have now accomplished what we set out to do; we are now in the one-dimensional case. Wijers (1995) studied the case where the L_i are constant almost surely and the T_i follow a homogeneous Poisson process. We shall be able to make much use of his methods. However, we wish to stress that the Poisson assumption certainly does not hold here because some of the $[T_i, T_i + X_i]$ are actually different representations of the same line segment!

We now demonstrate that the intensity of Ψ is given by

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

where

$$\mathrm{d}\nu(l) = \int_\vartheta \int_t \sum_m \delta_{l(t,\vartheta,m)}(\mathrm{d}l)\mathrm{d}t\mathrm{d}K(\vartheta).$$

Here $\delta_{l(t,\vartheta,m)}(\mathrm{d}l)$ is Dirac measure putting point mass at $l(t,\vartheta,m)$.

We must show that $\mathrm{d}t_1\mathrm{d}\nu(l)$ is the image measure of $\mathrm{d}\vec{t}\mathrm{d}K(\vartheta)\mathrm{d}\mathcal{C}(m)$ under the transformation, described above, that turned $\Phi'' \cap \mathcal{D}$ into Ψ. First we point out that $t(.)$ does not depend on x and that $l(.)$ does not depend on either t_1 or x. Abusing notation we write

$$t(\vec{t},x,\vartheta,m) = t(\vec{t},\vartheta,m).$$
$$l(\vec{t},x,\vartheta,m) = l(t_2,\vartheta,m).$$

Next, we notice that $t(\vec{t},x,\vartheta,m) = t_1 + t(0,t_2,x,\vartheta,m)$. It now follows that

$$\begin{aligned}
&\int_{\vec{t}}\int_{\vartheta}\sum_m \mathbf{1}_{\{a<t(\vec{t},x,\vartheta,m)\leq b\}}\mathbf{1}_{\{c<l(\vec{t},x,\vartheta,m)\leq d\}}\mathrm{d}\vec{t}\mathrm{d}K(\vartheta)\\
&= \int_{t_1}\int_{t_2}\int_{\vartheta}\sum_m \mathbf{1}_{\{a<t_1+t(0,t_2,\vartheta,m)\leq b\}}\mathbf{1}_{\{c<l(t_2,\vartheta,m)\leq d\}}\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}K(\vartheta)\\
&= \int_a^b \mathrm{d}t_1 \int_{t_2}\int_{\vartheta}\sum_m(\mathbf{1}_{\{c<l(t_2,\vartheta,m)\leq d\}}\mathrm{d}K(\vartheta)\mathrm{d}t_2\\
&= \int_a^b \mathrm{d}t_1 \int_c^d \mathrm{d}\nu(l).
\end{aligned}$$

The measure ν on $\mathbb{R}^+$ can be interpreted geometrically. Recall that we defined $\ell(d,\vartheta)$ to be a line with orientation ϑ at (signed) distance d from the origin. We define a random line L as follows. First, consider a circle $B(c,r)$ with center c and radius R such that $W \subset C(c,R)$. Next, draw a random variable Θ from K and a random variable D from the uniform distribution on $[-R,R]$. Finally, define L to be the line at an angle Θ with respect to the positive x-axis and at (signed) distance D from the origin. Now we condition on the event that L hits W. Intersecting L with W we obtain a collection of intervals and $\nu(a,b) = \int_a^b \mathrm{d}\nu(l)$ is the expected number of intervals of length between a and b.

We note that

$$\begin{aligned}
\int_l l\mathrm{d}\nu(l) &= \int_l\int_{\vartheta}\int_t\sum_m l\mathbf{1}_{l(t,\vartheta,m)}(\mathrm{d}l)\mathrm{d}t\mathrm{d}K(\vartheta)\\
&= \int_{\vartheta}\int_t\sum_m l(t,\vartheta,m)\mathrm{d}t\mathrm{d}K(\vartheta)\\
&= \int_{\vartheta}|W|\mathrm{d}K(\vartheta)\\
&= |W|.
\end{aligned}$$

We shall denote

$$\int_l \mathrm{d}\nu(l) = \kappa.$$

Integrals with respect to ν of such simple functions as above, are easily estimated using stereological methods. One would for instance sample random lines (or even a random grid of lines) hitting W and estimate the integrals by empirical averages. A nice introduction into the rudiments of stereology is given in Baddeley (1999a). There one will also find that if K is the uniform distribution, then $\kappa = \pi|\partial W|$, where $|\partial W|$ denotes the length of the boundary of W.

The points of $\Psi = \{(T_i, X_i, L_i)\}$ represent segments $[T_i, T_i + X_i]$ on the real line that *could* be observed through intervals $[0, L_i]$. Define

$$A = \{(t, x, l) \ : \ [t, t + x] \cap [0, l] \neq \emptyset\} = \{(t, x, l) \ : \ -x \leq t \leq l\}.$$

If a point of Ψ falls in A, the corresponding fragment is (partly) observed. Let us calculate the expected number of observed fragments, which equals the expected number of points of Ψ in A.

$$\begin{aligned} \int_A \lambda \mathrm{d}t\mathrm{d}F(x)\mathrm{d}\nu(l) &= \int_l \int_x \int_{-x}^{l} \lambda \mathrm{d}t\mathrm{d}F(x)\mathrm{d}\nu(l) \\ &= \int_l \int_x (l + x)\lambda \mathrm{d}F(x)\mathrm{d}\nu(l) \\ &= \int_l (l + \mu)\lambda \mathrm{d}\nu(l) = \lambda(|W| + \mu\kappa) \end{aligned} \tag{2.10}$$

2.3 Poisson

We have a stationary point process $\Psi = \{T_i, X_i, L_i\}$ on $\mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+$ with intensity $\lambda \mathrm{d}t\mathrm{d}F(x)\mathrm{d}\nu(l)$. These points represent line segments $[T_i, T_i + X_i]$ that might be observed through associated intervals $[0, L_i]$. The situation is very much akin to the set-up of Wijers (1995a and b). There are two differences. First he supposed the L_i to be constant. Secondly—and much more importantly—he assumed that the fragments $\{(T_i, X_i)\}$ are a Poisson process. This is certainly not the case here, even if the original line segments of Φ were Poisson. The problem is that two different fragments could have come from a single segment.

However, we may as well analyze the data *as if* Ψ were Poisson, as long as we make sure that our eventual results hold regardless of the dependencies between the points of Ψ. This is exactly what we shall do. Under the Poisson assumption we shall be able to derive the non-parametric maximum

likelihood estimator of the distribution function F. But whether or not the observations are independent, the algebraic properties of the maximum likelihood estimator remain valid. To prove its consistency when the observations are dependent we can basically copy the proof for the independent case. We only need to replace the strong law of large numbers by some ergodic theorem. We discussed these ideas in section 2.1.1.

So let us suppose Ψ is a Poisson point process. Then, given $\Psi(A) = n$, the points (T_i, X_i, L_i) in A are distributed as the set of values in an i.i.d. sample of size n from the normalized intensity, which by (2.10) equals

$$\mathbf{1}_A(t,x,l)\frac{\lambda \mathrm{d}t\mathrm{d}F(x)\mathrm{d}\nu(l)}{\int_A \lambda \mathrm{d}t\mathrm{d}F(x)\mathrm{d}\nu(l)} = \mathbf{1}_A(t,x,l)\frac{\mathrm{d}t\mathrm{d}F(x)\mathrm{d}\nu(l)}{|W|+\mu\kappa}. \tag{2.11}$$

We must now address the problem of 'length bias': Because longer line segments stand a better chance of hitting W than shorter ones, the lengths X_i of the line segments that hit W are not a sample from F. Hence, even if we observed the X_i, their empirical distribution would not estimate F. We therefore introduce the distribution function, say V, of the length X of a line segment $[T, T+X]$, given its observation through $[0, L]$. We calculate just like we did to derive (2.10)

$$\begin{aligned} V(x) &= \int_{y=0}^{x}\int_l\int_{t=-x}^{l} \frac{\mathrm{d}t\mathrm{d}F(x)\mathrm{d}\nu(l)}{|W|+\mu\kappa} \\ &= \int_{y=0}^{x} \frac{|W|+y\kappa}{|W|+\mu\kappa}\mathrm{d}F(y). \end{aligned} \tag{2.12}$$

One can show that as F ranges over all possible distributions (with finite mean μ), then V too varies over all distributions. Note that

$$\int \frac{1}{|W|+x\kappa}\mathrm{d}V(x) = \int \frac{1}{|W|+\mu\kappa}\mathrm{d}F(x) = \frac{1}{|W|+\mu\kappa}$$

so that one can recover μ from V and hence also recover F from V.

Formula (2.11) for the conditional distribution of (T, X, L) given observation becomes in terms of V

$$\mathbf{1}_A(t,x,l)\frac{\mathrm{d}t\mathrm{d}V(x)\mathrm{d}\nu(l)}{|W|+x\kappa} = \mathbf{1}_A(t,x,l)\frac{\mathrm{d}t}{l+x}\mathrm{d}V(x)\frac{l+x}{|W|+x\kappa}\mathrm{d}\nu(l). \tag{2.13}$$

Define, for all x, measures

$$\mathrm{d}\rho(l|x) = \frac{l+x}{|W|+x\kappa}\mathrm{d}\nu(l).$$

and note that these are probability measures for all x. Hence, by inspection of (2.13) we can interpret the distribution of (T, X, L) as

1. $X \sim \mathrm{d}V$
2. Given $X = x$, $L \sim \mathrm{d}\rho(l|x)$
3. Given $X = x$ and $L = l$, $T \sim \mathrm{Unif}[-x, l]$

We call the collection (T_i, X_i, L_i) the complete data. We observe only a many-to-one mapping of it when we observe the intervals $[0, L_i]$ and their intersection with the $[T_i, T_i + X_i]$. This reduction, together with the fact that the distribution of the complete data has the above cascade-like structure, casts the problem of estimating V as a non-parametric missing data problem, cf. section 1.4.4. In this special model the parameter space is convex and the distribution of a single observation is linear in the parameter.

We now derive the distribution of the (observed) data. Define $Y_i = |[T_i, T_i + X_i] \cap [0, L_i]|$, the length of the i-th fragment. Let Δ_i be the number of endpoints of $[T_i, T_i + X_i]$ that fall outside of $[0, L_i]$. Also we observe L_i, the length of the interval in which a fragment is observed. Under the assumption that the $\{T_i\}$ are Poisson it will turn out that (Y, Δ) is sufficient for estimating V. At this point we introduce two functions that will help describe the distribution of the data.

$$g(x) = \int_{[x,\infty)} \frac{1}{|W| + \kappa y} \mathrm{d}V(y) \tag{2.14}$$

$$h(x) = \int_{[x,\infty)} \frac{y - x}{|W| + \kappa y} \mathrm{d}V(y). \tag{2.15}$$

For the case $\Delta = 1$, V will enter the distribution of Y only through g, and for the case $\Delta = 2$ only through h. There is a *very* useful relation between V, g and h which is easy to check using the above two definitions. For all $x > 0$

$$1 = \kappa h(x) + (|W| + \kappa x) g(x) + V(x^-). \tag{2.16}$$

where $V(x^-) = \int_{[0,x)} \mathrm{d}V(y)$.

Suppose that $\mathrm{d}\rho(.|x)$ (the distribution of L given $X = x$) is degenerate at some fixed l. We are then *exactly* in the situation studied by Wijers (1995). We follow his derivation. First consider the region

$$A_l = \{(t, x) \ : \ [t, t + x] \cap [0, l] \neq \emptyset\} = \{(t, x) \ : \ -x \leq t \leq l\}.$$

which can be subdivided into regions *lc, rc, uc* and *dc* as in figure 2.3. For a point (t, x) in *lc* the intersection of $[t, t + x]$ with $[0, l]$ will be left censored. Similarly, we distinguish right censored (*rc*), uncensored (*uc*) and doubly

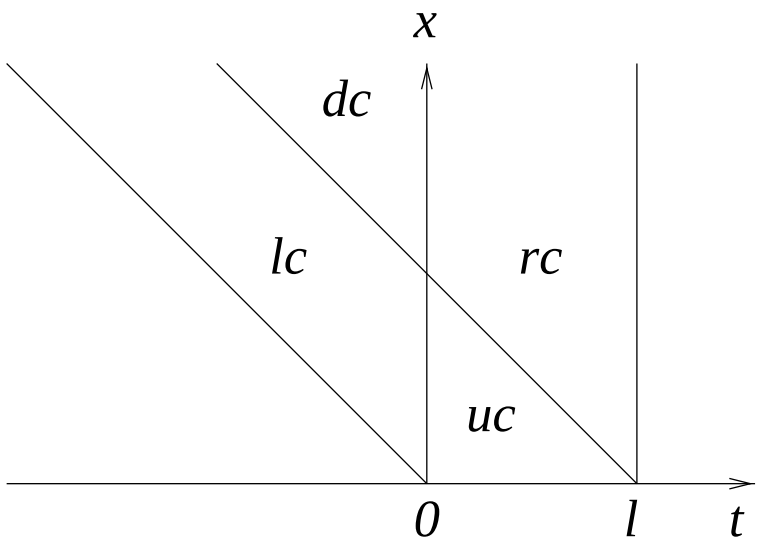

Figure 2.3: Subdividing A_l into different censoring types.

censored (dc). Left and right censored observations are taken together as censoring 'type' $\Delta = 1$; observation of one endpoint. Because $L = l$ a.s., the density of (t, x) becomes, cf. (2.13)

$$\mathbf{1}_{A_l}(t,x)\frac{\mathrm{d}t}{l+x}\mathrm{d}V(x).$$

The sub-distributions for the various censoring types can be computed by integrating, over the shaded regions in Figure 2.3. We take right and left censored observations together. We find

$$\begin{aligned} P^l(\mathrm{d}y, 0) &= \frac{l-y}{l+y}\mathrm{d}V(y) \\ P^l(\mathrm{d}y, 1) &= 2\int_y^\infty \frac{1}{l+x}\mathrm{d}V(x)\mathrm{d}y \\ P^l(\mathrm{d}y, 2) &= \int_y^\infty \frac{x-y}{l+x}\mathrm{d}V(x)\delta_l(\mathrm{d}y). \end{aligned}$$

These formulas are easily modified to accommodate general $\mathrm{d}\rho(l|x)$. Re-

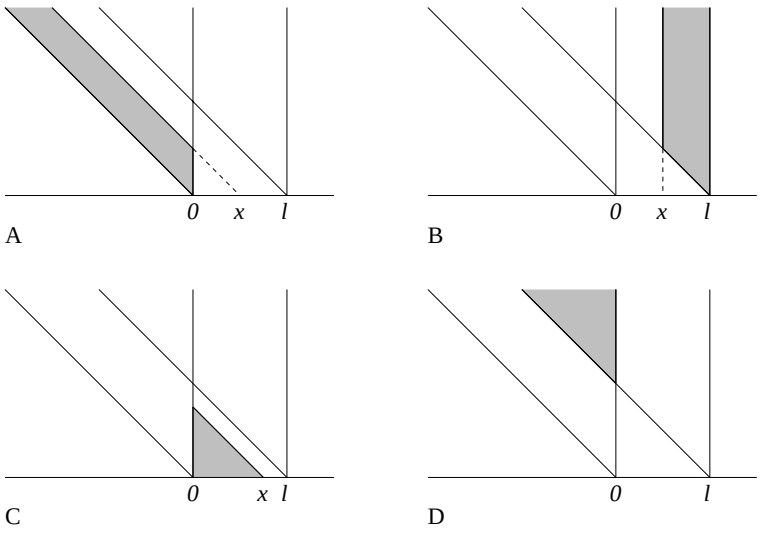

Figure 2.4: Integration regions for various sub-distribution functions.

calling that $\mathrm{d}\rho(l|x) = \frac{l+x}{|W|+x\kappa}\mathrm{d}\nu(l)$ we have

$$\begin{aligned}
P(\mathrm{d}y, \mathrm{d}l, 0) &= \mathbf{1}_{\{l\geq y\}}\frac{l-y}{l+y}\mathrm{d}\rho(l|y)\mathrm{d}V(y) = \mathbf{1}_{\{l\geq y\}}\frac{l-y}{|W|+\kappa y}\mathrm{d}\nu(l)\mathrm{d}V(y)\\
P(\mathrm{d}y, \mathrm{d}l, 1) &= 2\mathbf{1}_{\{l\geq y\}}\int_y^\infty \frac{1}{l+x}\mathrm{d}\rho(l|x)\mathrm{d}V(x)\mathrm{d}y\\
&= 2\mathbf{1}_{\{l\geq y\}}\mathrm{d}\nu(l)\int_y^\infty \frac{1}{|W|+\kappa x}\mathrm{d}V(x)\mathrm{d}y\\
&= 2\mathbf{1}_{\{l\geq y\}}\mathrm{d}\nu(l)g(y)\mathrm{d}y\\
P(\mathrm{d}y, \mathrm{d}l, 2) &= \int_y^\infty \frac{x-y}{l+x}\mathrm{d}\rho(l|x)\mathrm{d}V(x)\delta_l(\mathrm{d}y)\\
&= \mathrm{d}\nu(l)\int_y^\infty \frac{x-y}{|W|+\kappa x}\mathrm{d}V(x)\delta_l(\mathrm{d}y)\\
&= \mathrm{d}\nu(l)h(y)\delta_l(\mathrm{d}y).
\end{aligned}$$

We see that the conditional distribution of L, given Y and Δ does not depend on V. Hence (Y, Δ) is sufficient for estimation of V (at least under the present

Poisson assumption). The joint distribution of (Y, Δ) is

$$P(\mathrm{d}y, 0) = \int_l P(\mathrm{d}y, \mathrm{d}l, 0) = \frac{\int_{l \geq y} (l - y) \mathrm{d}\nu(l)}{|W| + \kappa y} \mathrm{d}V(y) \tag{2.17}$$

$$P(\mathrm{d}y, 1) = \int_l P(\mathrm{d}y, \mathrm{d}l, 1) = 2\nu([y, \infty)) g(y) \mathrm{d}y \tag{2.18}$$

$$P(\mathrm{d}y, 2) = \int_l P(\mathrm{d}y, \mathrm{d}l, 2) = \mathrm{d}\nu(y) h(y) \tag{2.19}$$

Suppose that τ is the supremum of the support of $\rho(.|x)$ over all x. Then τ is an upper bound for the length of a fragment that can be observed in W. That is, $P(Y > \tau) = 0$. Note that

$$\begin{aligned}
g(x) &= \int_{[x,\infty)} \frac{1}{|W| + \kappa y} \mathrm{d}V(y) \\
&= \int_{[x,\tau)} \frac{1}{|W| + \kappa y} \mathrm{d}V(y) + \int_{[\tau,\infty)} \frac{1}{|W| + \kappa y} \mathrm{d}V(y) \\
&= \int_{[x,\tau)} \frac{1}{|W| + \kappa y} \mathrm{d}V(y) + g(\tau) \\
h(x) &= \int_{[x,\infty)} \frac{y - x}{|W| + \kappa y} \mathrm{d}V(y) \\
&= \int_{[x,\tau)} \frac{y - x}{|W| + \kappa y} \mathrm{d}V(y) + \int_{[\tau,\infty)} \frac{y - \tau + \tau - x}{|W| + \kappa y} \mathrm{d}V(y) \\
&= \int_{[x,\tau)} \frac{y - x}{|W| + \kappa y} \mathrm{d}V(y) + h(\tau) + (\tau - x) g(\tau).
\end{aligned}$$

We define

$$H = \kappa h(\tau) \qquad G = (|W| + \kappa\tau) g(\tau).$$

and recall that relation (2.16) states that $G + H + V(\tau^-) = 1$. Later on we shall interpret G and H as probabilities. We can express the distribution of

(Y, Δ) in terms of V restricted to $[0, \tau)$, G and H. For $y < \tau$

$$
\begin{aligned}
P(\mathrm{d}y, 0) &= \frac{\int_{l \geq y} (l - y) \mathrm{d}\nu(l)}{|W| + \kappa y} \mathrm{d}V(y) && (2.20) \\
P(\mathrm{d}y, 1) &= 2\nu([y, \infty)) \mathrm{d}y g(y) = 2\nu([y, \infty)) \mathrm{d}y \int_y^\infty \frac{1}{|W| + \kappa x} \mathrm{d}V(x) \\
&= 2\nu([y, \infty)) \mathrm{d}y \Big(\int_y^{\tau-} \frac{1}{|W| + \kappa x} \mathrm{d}V(x) + \int_\tau^\infty \frac{1}{|W| + \kappa x} \mathrm{d}V(x) \Big) \\
&= 2\nu([y, \infty)) \mathrm{d}y \int_y^{\tau-} \frac{1}{|W| + \kappa x} \mathrm{d}V(x) + \frac{2\nu([y, \infty)) \mathrm{d}y}{|W| + \kappa \tau} G && (2.21) \\
P(\mathrm{d}y, 2) &= \mathrm{d}\nu(y) h(y) = \mathrm{d}\nu(y) \int_y^\infty \frac{x - y}{|W| + \kappa x} \mathrm{d}V(x) \\
&= \mathrm{d}\nu(y) \Big(\int_y^{\tau-} \frac{x - y}{|W| + \kappa x} \mathrm{d}V(x) + \int_\tau^\infty \frac{x - \tau}{|W| + \kappa x} \mathrm{d}V(x) \\
&+ \int_\tau^\infty \frac{\tau - y}{|W| + \kappa x} \mathrm{d}V(x) \Big) \\
&= \mathrm{d}\nu(y) \int_y^{\tau-} \frac{x - y}{|W| + \kappa x} \mathrm{d}V(x) + \frac{\mathrm{d}\nu(y)}{\kappa} H + \frac{(\tau - y) \mathrm{d}\nu(y)}{|W| + \kappa \tau} G. && (2.22)
\end{aligned}
$$

Because (2.16) expresses G in terms of $V(\tau^-)$ and H, the distribution of the data is fully parameterized by

$$(V|_{[0,\tau)}, H),$$

where $V|_{[0,\tau)}$ ranges over all (possibly defective) distribution functions and H is any positive real such that $V(\tau^-) + H \leq 1$. Below we demonstrate that there is a 1-1 correspondence between $(V|_{[0,\tau)}, H)$ and $(F|_{[0,\tau)}, \mu)$, where $F|_{[0,\tau)}$ ranges over all (possibly defective) distribution functions and μ is any positive real exceeding $\int_0^{\tau-} x \mathrm{d}F(x)$.

By the way, it is interesting that without any observations exceeding τ, the mean of F can be estimated.

To express $F|_{[0,\tau)}$ and μ in terms of $V|_{[0,\tau)}$ and H, we first note that $g(\tau) = G/(|W| + \kappa\tau)$ is determined by $V(\tau^-)$ and H through (2.16). We have

$$
\begin{aligned}
\int_{[0,\tau)} \frac{1}{|W| + \kappa y} \mathrm{d}V(y) + g(\tau) &= \int_0^\infty \frac{1}{|W| + \kappa y} \mathrm{d}V(y) \\
&= \int_0^\infty \frac{1}{|W| + \kappa \mu} \mathrm{d}F(y) = \frac{1}{|W| + \kappa \mu}
\end{aligned}
$$

so that

$$\mu = \frac{1}{\kappa}\left(\frac{1}{\int_{[0,\tau)} \frac{1}{|W|+\kappa y}\mathrm{d}V(y) + g(\tau)} - |W|\right) \tag{2.23}$$

and

$$\begin{aligned} F(x) &= \int_0^x \frac{|W|+\kappa\mu}{|W|+\kappa y}\mathrm{d}V(y) \\ &= \frac{1}{\int_{[0,\tau)} \frac{1}{|W|+\kappa y}\mathrm{d}V(y) + g(\tau)} \int_0^x \frac{1}{|W|+\kappa y}\mathrm{d}V(y). \end{aligned} \tag{2.24}$$

2.4 EM

Under the assumption that the fragments we observe are independent, we have a nonparametric missing data problem (see section 1.4.4).

1. $X \sim V, \;\; V \in \mathcal{V} = \{\text{all distributions on } \mathbb{R}^+\}$
2. Given $X = x, \;\; L \sim \mathrm{d}\rho(l|x)$
3. Given $X = x$ and $L = l$, $T \sim \text{Unif}[-x, l]$
4. Observe $(Y, \Delta) = g(X, T, L)$

where $Y = |[T, T+X] \cap [0, L]|$ and Δ is the number of unobserved endpoints.

As we have seen, the distribution of the data (Y, Δ) depends on V only through its restriction to $[0, \tau)$ and a functional H. This means that the model $\mathcal{V}$ is not identifiable from the data. If two distribution functions V and V' agree on $[0, \tau)$ and have the same H's, then they can never be told apart on the basis of observing (Y_i, Δ_i).

We could reduce $\mathcal{V}$ in such a way that it becomes identifiable, but then the model is no longer completely nonparametric. However, we can construct a different model consisting of all probability measures on a different space, and a new missing data problem such that model *is* identifiable. Of course we must make sure that the model for the observed data remains the same.

We now state a nonparametric, identifiable missing data problem, such that the distribution of the observed data satisfies (2.20) to (2.22). We again use the symbol $\mathcal{V}$ to denote the model for the complete data. We hope this does not confuse too much.

The (new) model $\mathcal{V}$ for the complete data, say X, consists of *all* probability distributions on the space $[0, \tau) \cup \{\dagger, \ddagger\}$. We write $V(\{\dagger\}) = P(X = \dagger) = G$ and $V(\{\ddagger\}) = P(X = \ddagger) = H$. We must have $V(\tau^-) + G + H = 1$.

Sample X from $V \in \mathcal{V}$. Next,

- If $X = x \in [0, \tau)$ we draw an L from $\mathrm{d}\rho(.|x)$ and given $L = l$ we draw a uniform T on $[-x, l]$. We set $Y = |[T, T + X] \cap [0, L]|$ and Δ is the number of endpoints of $[T, T + X]$ outside of $[0, L]$.
- If $X = \dagger$ we sample (Y, Δ) from

$$\mathbf{1}_1(\delta)\frac{2\nu([y,\infty))\mathrm{d}y}{|W| + \kappa\tau} + \mathbf{1}_2(\delta)\frac{(\tau - y)\mathrm{d}\nu(y)}{|W| + \kappa\tau}.$$

- If $X = \ddagger$ we sample (Y, Δ) from

$$\mathbf{1}_2(\delta)\frac{\mathrm{d}\nu(y)}{\kappa}.$$

It is certainly not hard to check that the distribution of (Y, Δ) under this new scheme is indeed again given by (2.20) to (2.22).

Now, because we have a completely nonparametric missing data model we can use every square integrable function which integrates to zero as a score function. In particular, we can use indicator functions of measurable sets in $[0, \tau) \cup \{\dagger, \ddagger\}$ minus their expectations. As in (1.14) we find the following system of score equations

$$\hat{V}_n(A) = \frac{1}{n}\sum_1^n P_{\hat{V}_n}(X_i \in A | Y_i, \Delta_i), \qquad A \subseteq [0, \tau) \cup \{\dagger, \ddagger\} \tag{2.25}$$

or in other words

$$\hat{V}(\mathrm{d}x) = \frac{1}{n}\sum_1^n P_{\hat{V}_n}(X_i \in \mathrm{d}x | Y_i, \Delta_i), \qquad x \in [0, \tau) \tag{2.26}$$

$$\hat{V}_n(\{\ddagger\}) = \hat{H} = \frac{1}{n}\sum_1^n P_{\hat{V}_n}(X = \ddagger | Y_i, \Delta_i) \tag{2.27}$$

and by (2.16)

$$\hat{V}_n(\{\dagger\}) = 1 - \hat{V}_n(\{\ddagger\}) - \hat{V}_n(\tau^-). \tag{2.28}$$

Solving the above equations iteratively is an instance of the EM algorithm. From one iteration step to the next the support will not increase. Hence, one should first decide on the support of $\hat{V}_n$ and then start the iterations with an initial guess which does not have smaller support. The support of the true NPMLE (as defined in (1.1)) is unknown to us. We *choose* to start the iterations with a distribution V on $[0, \tau) \cup \{\dagger, \ddagger\}$ with mass only at the

observed uncensored observations and at the points † and ‡. Thus we obtain a 'data sieved' NPMLE (see section 1.4.1).

To find explicit expressions for the above score equations, note that the distribution of the data, specified in (2.20) to (2.22), is of the form

$$P(\mathrm{d}y, \delta) = KV(\mathrm{d}y, \delta) = \int K(\mathrm{d}y, \delta|x)\mathrm{d}V(x).$$

We recognize, for $y \in [0, \tau)$ and $\delta \in \{0, 1, 2\}$,

$$\begin{aligned} K(\mathrm{d}y, \delta|x) = {} & \mathbf{1}_0(\delta)\frac{\int_{l\geq x}(l-x)\mathrm{d}\nu(l)}{|W|+\kappa x}\delta_x(\mathrm{d}y) \\ & + \mathbf{1}_1(\delta)2\nu([y,\infty))\left(\frac{\mathbf{1}_{(y,\tau)}(x)}{|W|+\kappa x} + \frac{\mathbf{1}_{\{\dagger\}}(x)}{|W|+\kappa\tau}\right)\mathrm{d}y \\ & + \mathbf{1}_2(\delta)\left(\frac{\mathbf{1}_{(y,\tau)}(x)(x-y)}{|W|+\kappa x} + \frac{\mathbf{1}_{\{\dagger\}}(x)(\tau-y)}{|W|+\kappa\tau} + \frac{\mathbf{1}_{\{\ddagger\}}(x)}{\kappa}\right)\mathrm{d}\nu(y). \end{aligned}$$

Given a sample $(y_1, \delta_1), \ldots, (y_n, \delta_n)$ suppose that $x_1, \ldots, x_m$ are the lengths of all different uncensored observations. Consider the (random) measure

$$\begin{aligned} \mu(\mathrm{d}y, \delta) = {} & \mathbf{1}_0(\delta)\sum_{i=1}^{m}\frac{\int_{l\geq x_i}(l-x_i)\mathrm{d}\nu(l)}{|W|+\kappa X_i}\delta_{x_i}(\mathrm{d}y) \\ & + \mathbf{1}_1(\delta)\mathbf{1}_{[0,\tau)}(y)\nu([y,\infty))\mathrm{d}y \\ & + \mathbf{1}_2(\delta)\mathbf{1}_{[0,\infty)}(y)\nu(\mathrm{d}y). \end{aligned}$$

This sigma-finite measure μ dominates $K(\mathrm{d}y, \delta|x)$ for all $x \in \{x_1, \ldots, x_m\} \cup \{\dagger, \ddagger\}$. The Radon–Nikodym derivative of $K(\mathrm{d}y, \delta|x)$ with respect to $\mu(\mathrm{d}y, \delta)$ is, for all $x \in \{x_1, \ldots, x_m\} \cup \{\dagger, \ddagger\}$

$$\begin{aligned} k(y, \delta; x) = {} & \mathbf{1}_0(\delta)\mathbf{1}_{[0,\tau)}(x)\mathbf{1}_x(y) \\ & + \mathbf{1}_1(\delta)\left(\frac{\mathbf{1}_{(y,\tau)}(x)}{|W|+\kappa x} + \frac{\mathbf{1}_{\{\dagger\}}(x)}{|W|+\kappa\tau}\right) \\ & + \mathbf{1}_2(\delta)\left(\frac{\mathbf{1}_{(y,\tau)}(x)(x-y)}{|W|+\kappa x} + \frac{\mathbf{1}_{\{\dagger\}}(x)(\tau-y)}{|W|+\kappa\tau} + \frac{\mathbf{1}_{\{\ddagger\}}(x)}{\kappa}\right). \end{aligned}$$

We arrange it so that the data sieved NPMLE $\hat{V}_n$ puts all its mass on $\{x_1, \ldots, x_m\} \cup \{\dagger, \ddagger\}$. Thus, for $\hat{V}_n$ almost all x the mixing kernel K is dominated by a single sigma-finite measure as was required in section 1.4.4. In that section we also specified what the score equations for a nonparametric missing data problem look like, cf. (1.17). We find

$$\hat{V}_n(A) = \int \frac{\int_A k(y, \delta; x)\mathrm{d}V(x)}{\int_{x'} k(y, \delta; x')\mathrm{d}V(x)}\mathrm{d}\mathbb{P}_n(y, \delta)$$

where $\mathbb{P}_n$ is the empirical distribution of the data putting mass $1/n$ at the observed $(y_1, \delta_1), \ldots, (y_n, \delta_n)$. The equations work out to

$$\begin{aligned}\hat{V}_n(\{x_i\}) &= \mathbb{P}_n(x_i, 0) \\ &\quad + \frac{\hat{V}_n(\{x_i\})}{|W| + \kappa x_i} \int_{y=0}^{x_i} \frac{1}{\hat{g}_n(y)} \mathbb{P}_n(\mathrm{d}y, 1) \\ &\quad + \frac{\hat{V}_n(\{x_i\})}{|W| + \kappa x_i} \int_{y=0}^{x_i} \frac{x_i - y}{\hat{h}_n(y)} \mathbb{P}_n(\mathrm{d}y, 2), \qquad i = 1, 2, \ldots, m \\ \hat{V}_n(\{\ddagger\}) = \hat{H}_n &= \hat{H}_n \int_y \frac{1}{\hat{h}_n(y)} \mathbb{P}_n(\mathrm{d}y, 2).\end{aligned}$$

where, cf. (2.16)

$$\hat{V}_n(\{\dagger\}) = \hat{G}_n = 1 - \hat{H}_n - \hat{V}_n(\tau^-)$$

$$\hat{g}_n(\tau) = \hat{G}_n \;/\; (|W| + \kappa\tau)) \quad \text{and} \quad \hat{h}_n(\tau) = \hat{H}_n/\kappa$$

and, for $x \in [0, \tau)$,

$$\begin{aligned}\hat{g}_n(x) &= \int_{[x,\tau)} \frac{1}{|W| + \kappa y} \mathrm{d}\hat{V}_n(y) + \hat{g}_n(\tau) \\ \hat{h}_n(x) &= \int_{[x,\tau)} \frac{y - x}{|W| + \kappa y} \mathrm{d}\hat{V}_n(y) + \hat{h}_n(\tau) + (\tau - x)\hat{g}_n(\tau).\end{aligned}$$

Recall that $(V_{[0,\tau)}, H)$ and $(F_{[0,\tau)}, \mu)$ stand in a one-to-one relation. In the previous section we established formulas (2.23) and (2.24), expressing $(F_{[0,\tau)}, \mu)$ in terms of $(V_{[0,\tau)}, H)$. With $g(\tau) = (1 - V(\tau^-) - H)/(|W| + \kappa\tau)$, we have

$$\mu = \frac{1}{\kappa} \left(\frac{1}{\int_{[0,\tau)} \frac{1}{|W| + \kappa y} \mathrm{d}V(y) + g(\tau)} - |W| \right)$$

and

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

It is now obvious how we define $\hat{F}_n$ and $\hat{\mu}_n$ as transformations of $\hat{V}_n$ and $\hat{H}_n$.

There really is not much point in proving desirable asymptotic properties, such as consistency, of $\hat{V}_n$ as n tends to infinity. Any result we obtain here will only hold under the Poisson assumption stated at the beginning of the

previous section. However, it will turn out that the *method* of proving consistency will—with minor modifications—work without the Poisson assumption. When we no longer have the Poisson assumption we shall have to use different asymptotics, which will lead to some complications. To give an (admittedly sketchy) proof of consistency of $\hat{V}_n$ here, allows us a more clean presentation. We shall use Theorem 1.2, but we need a *working hypothesis*:

working hypothesis
If $g(\tau) > 0$ then $\liminf \hat{g}_n(\tau) > 0$, almost surely.

Whether or not this is true, we do not know. When, in the next section, we prove consistency without the Poisson assumption this hypothesis will no longer be needed. By that time, to get around other difficulties, we shall have introduced some extra, artificial censoring. This will have the added benefit of making our present working hypothesis superfluous.

First, we note that the model $\mathcal{V}$ (of all probability measures on $[0,\tau) \cup \{\dagger, \ddagger\}$) is convex, which by Lemma 1.3 implies that also the model for the distribution of the data $K\mathcal{V} = \{KV : V \in \mathcal{V}\}$ is convex. This places us in the realm of Theorem 1.2. We can use this theorem to prove consistency (under our working hypothesis) of the NPMLE $K\hat{V}_n$ in the sense that, almost surely,

$$\|\hat{V}_n - V_0\|_{[0,\tau)} \to 0 \quad \text{and} \quad |\hat{H}_n - H_0| \to 0. \tag{2.29}$$

Here $\|.\|_{[0,\tau)}$ denotes the uniform distance on the set $[0,\tau)$. From (2.29) it is not too difficult (using empirical process theory) to conclude that also

$$\|\hat{F}_n - F_0\|_{[0,\tau)} \to 0 \quad \text{and} \quad |\hat{\mu}_n - \mu_0| \to 0, \tag{2.30}$$

almost surely.

To apply Theorem 1.2 we need to compare $K\hat{V}_n$ to a sequence KV_n for which we can easily establish consistency. We choose

$$V_n(x) = \int_0^x \frac{|W| + \kappa y}{\int_{l \geq y}(l-y)\mathrm{d}\nu(l)} \mathbb{P}_n(\mathrm{d}y, 0) \qquad x \in [0,\tau).$$

We define $V_n(\tau^-)$ in the obvious manner and note that by the law of large numbers it is consistent. Hence, for $V_n(\{\ddagger\}) = H_n$ we can take a sequence tending to H while $H_n \leq V_n(\tau)$. We define G_n, g_n and h_n in terms of $V_n|_{[0,\tau)}$ and H_n.

From (2.20) we see that $KV_n(\mathrm{d}y, 0) = \mathbb{P}_n(\mathrm{d}y, 0)$.

First we show convergence of (V_n, H_n) to satisfy the first requirement of Theorem 1.2. The collection

$$\{\mathbf{1}_{[0,x]}(y)\frac{|W|+\kappa y}{\int_{l\geq y}(l-y)\mathrm{d}\nu(l)},\ x\in[0,\tau)\}$$

is a *Glivenko–Cantelli* class of functions over which the law of large numbers holds uniformly. In other words, we have

$$\|V_n - V_0\|_{[0,\tau)} \to 0 \ \text{ and } \ |H_n - H_0| \to 0 \tag{2.31}$$

almost surely. For empirical process theory see, for instance, van der Vaart and Wellner (1996). That the above collection of functions is a Glivenko–Cantelli class follows from the fact that the functions are all the product of an indicator function (of intervals) and the same monotone function and hence they are uniformly of bounded variation: the variation is bounded uniformly over the collection of functions. Also, there is an integrable 'envelop' (majorant) for the entire collection.

Next, we need that the straight line between KV_n and $K\hat{V}_n$ is Hellinger differentiable at $K\hat{V}_n$. By Lemma 1.4 it suffices to show that the straight line between V_n and $\hat{V}_n$ is Hellinger differentiable at $\hat{V}_n$. Because $V_n \ll \hat{V}_n$ and also $\mathrm{d}V_n/\mathrm{d}\hat{V}_n \in L_2(\hat{V}_n)$ this is indeed true.

Now, if we establish the final, third condition of Theorem 1.2

$$\sum_{\delta=0}^{2}\int_{[0,\tau)}\left(\frac{\mathrm{d}KV_n}{\mathrm{d}K\hat{V}_n}\right)\mathrm{d}(\mathbb{P}_n(y,\delta) - KV_n(y,\delta)) \to 0 \tag{2.32}$$

then we can conclude that $K\hat{V}_n$ converges to KV_0 in the same sense as KV_n does. In other words, (2.32) implies (2.29), which in turn implies (2.30).

From formulas (2.17) to (2.19) we see that (2.32) reads

$$\begin{aligned}
&\sum_{\delta=0}^{2}\int_{[0,\tau)}\left(\frac{\mathrm{d}KV_n}{\mathrm{d}K\hat{V}_n}\right)\mathrm{d}(\mathbb{P}_n(y,\delta) - KV_n(y,\delta))\\
&= \int_{[0,\tau)}\left(\frac{\mathrm{d}V_n(y)}{\mathrm{d}\hat{V}_n(y)}\right)\mathrm{d}(\mathbb{P}_n(y,0) - KV_n(y,0))\\
&+ \int_{[0,\tau)}\left(\frac{g_n(y)}{\hat{g}_n(y)}\right)\mathrm{d}(\mathbb{P}_n(y,1) - KV_n(y,1))\\
&+ \int_{[0,\tau)}\left(\frac{h_n(y)}{\hat{h}_n(y)}\right)\mathrm{d}(\mathbb{P}_n(y,2) - KV_n(y,2))\\
&\to 0.
\end{aligned} \tag{2.33}$$

Since $KV_n(\mathrm{d}y, 0) = \mathbb{P}_n(\mathrm{d}y, 0)$ we get convergence of the first term of (2.33) for free. To prove convergence of the other two terms, we again apply empirical process theory to assert that not only for $\delta = 0$, but also for $\delta = 1, 2$

$$\| \int_0^{\cdot} \mathrm{d}\mathbb{P}_n(y, \delta) - \int_0^{\cdot} \mathrm{d}KV_n(y, \delta) \|_{[0,\tau)} \to 0, \tag{2.34}$$

almost surely. Now we need some elementary calculus

Lemma 2.1. *Suppose that H_n ($n = 1, 2, \dots$) and H are real valued functions on an interval $[a, b)$ which are of bounded variation. Suppose that G_n are real valued functions on $[a, b)$ which are bounded and of bounded variation uniformly in n. Suppose that the H_n tend to H uniformly on $[a, b)$. Then*

$$\int_{[a,b)} G_n \mathrm{d}(H_n - H) \to 0.$$

Proof Applying integration by parts, we rewrite the integral as

$$G_n(b^-)(H_n(b^-) - H(b^-)) - G_n(a)(H_n(a) - H(a)) + \int_{[a,b)} (H_n - H)\mathrm{d}G_n$$

The first two terms tend to zero because the G_n are uniformly bounded and the H_n converge to H. The third term is in absolute value not more than $||H_n - H||_\infty \int_{[a,b)} |\mathrm{d}G_n|$. This term tends to zero because the H_n tend uniformly to H and the G_n are uniformly of bounded variation. □

We need to show convergence of the remaining two terms of (2.33). With the above lemma and (2.34), this boils down to showing that $g_n/\hat{g}_n$ and $h_n/\hat{h}_n$ are uniformly bounded and of bounded variation (almost surely for n large enough). Under our 'working hypothesis' this can be accomplished without much difficulty.

2.5 Without the Poisson assumption

Recall that we started out with a marked point-process $\Phi = \{(\vec{S}_i, X_i, \Theta_i)\}$ on $\mathbb{R}^2 \times \mathbb{R}^+ \times (-\pi/2, \pi/2)$ to model a collection of line segments with left endpoint $\vec{S}_i$, length X_i and orientation Θ_i. We then transformed Φ into $\Phi' = \{(\vec{T}_i, X_i, \Theta_i)\}$, representing the left endpoints with respect to the unique straight line on which the associated line segment lies. Using square brackets, we write $[\vec{t}, x, \vartheta]$ to denote a line segment (i.e. a random closed subset of $\mathbb{R}^2$), and $(\vec{t}, x, \vartheta)$ to denote a point in $\mathbb{R}^2 \times \mathbb{R}^+ \times (-\pi/2, \pi/2)$. We use

$\Phi' \times \mathbb{N}$ to represent all possible fragments arising from intersecting the line segments with a random closed set $\mathcal{W}$. Recall we introduced lines $\ell(d, \vartheta)$ with orientation ϑ at (signed) distance d from the origin and defined $\mathcal{W}(d, \vartheta) = \ell(d, \vartheta) \cap \mathcal{W}$. We assumed that the $\mathcal{W}(t, \theta)$ are always the union of (at most) countably many closed intervals: $\mathcal{W}(t, \vartheta) = \cup_m \mathcal{W}(t, \vartheta, m)$. Conditionally on $\mathcal{W}$, a point $(\vec{t}, x, \vartheta, m)$ corresponds to a (possibly empty) random closed set $[\vec{t}, x, \vartheta, m] = [\vec{t}, x, \vartheta] \cap \mathcal{W}(t_2, \vartheta, m)$, which we called a fragment. Finally we represented all observed fragments as a point process $\Psi = \{(T_i, X_i, L_i)\}$ on $\mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+$ corresponding to the situation where we observe segments $[T_i, T_i + X_i]$ through intervals $[0, L_i]$.

In the preceding section we assumed that Ψ is a Poisson process. We conditioned on the number of fragments that are observed, so that the fragments could be interpreted as an independent, identically distributed sample. We were then able to interpret the observations as 'missing data' (see section 1.4.4) in a very convenient way. This led us to a set of nonparametric score equations. The nonparametric maximum likelihood estimator (see section 1.4.1) in the sense of Kiefer and Wolfowitz (1956) is a solution of these equations. We proceeded to apply general methods (Theorem 1.2) to prove consistency of the estimator as the number of observations tends to infinity.

As we explained earlier, the assumption that the fragments are Poisson is simply not justified. They are not even independent, because several fragments might belong to a single underlying segment. However, as it turns out, without the Poisson assumption it is still sensible to use the *same* estimator. Also, we can still use Theorem 1.2 to prove its consistency.

Without the Poisson assumption we no longer have i.i.d. observations, the number of which we can send to infinity. We shall have to resort to different asymptotics. Indeed, we shall consider observation of $\mathcal{W} \cap \Phi$ through expanding sets B_n. Instead of empirical distribution functions we shall have empirical averages and instead of Glivenko–Cantelli we use an ergodic theorem to obtain the necessary convergence. We discussed this approach in some generality in section 2.1.1.

2.5.1 Expanding domain asymptotics

Let

$$B_n = nB = [0, n] \times [0, n] \quad \text{and} \quad B_\infty = \mathbb{R}^2$$

and suppose we observe $\mathcal{W} \cap \Phi$ through these B_n. We must now modify some earlier definitions to accommodate the varying n. Our original set-up corresponds exactly to the case $n = 1$.

We define, for all $d \in \mathbb{R}$, $\vartheta \in (-\pi/2, \pi/2)$ and $m \in \mathbb{N}$

$$W_n(d, \vartheta, m) = \mathcal{W}(d, \vartheta, m) \cap B_n \text{ and } W_\infty(d, \vartheta, m) = \mathcal{W}(d, \vartheta, m)$$

Now consider sets

$$\mathcal{D}_n = \{(\vec{t}, x, \vartheta, m) \ : \ W_n(t_2, \vartheta, m) \neq \emptyset\}$$

and note that even the set $\mathcal{D}_\infty$ is well defined.

As before we define on $\mathcal{D}_n$ two functions: $t_n(\vec{t}, \vartheta, m)$ and $l_n(\vec{t}, \vartheta, m)$. $t_n(\vec{t}, x, \vartheta, m)$ is the distance from the left endpoint of the line segment $[\vec{t}, x, \vartheta]$ to the left-most endpoint of $W(t_2, \vartheta, m)$, which we take negative if the first is to the left of the latter. Let $l_n(\vec{t}, x, \vartheta, m)$ be the length of $W_n(t_2, \vartheta, m)$. Consider the map that assigns $(t_n(\vec{t}, x, \vartheta, m), x, l_n(\vec{t}, x, \vartheta, m))$ to $(\vec{t}, x, \vartheta, m)$ for all $(\vec{t}, x, \vartheta, m) \in \mathcal{D}_n$. The image under this map of $\Phi'' \cap \mathcal{D}$ is a point-process $\Psi_n = \{(T_i, X_i, L_i)\}$ on $\mathbb{R} \times \mathbb{R}^+ \times [0, \infty)$ with intensity

$$\lambda \mathrm{d}t \mathrm{d}F(x) \mathrm{d}\nu_n(l),$$

where

$$\mathrm{d}\nu_n(l) = \int_t \int_\vartheta \sum_m \delta_{l_n(t,\vartheta,m)}(\mathrm{d}l) \mathrm{d}t \mathrm{d}K(\vartheta).$$

The Ψ_n represent the fragments arising from intersecting $\Phi \cap \mathcal{W}$ with B_n.

We note that

$$\int_l l \mathrm{d}\nu_n(l) = |W_n|$$

and define

$$\int_l \mathrm{d}\nu_n(l) = \kappa_n.$$

As before $A = \{(t, x, l) \ : \ [t, t + x] \cap [0, l] \neq \emptyset\}$, and the expected number of fragments we observe in $W_n = \mathcal{W} \cap B_n$ is

$$\int_A \lambda \mathrm{d}t \mathrm{d}F(x) \mathrm{d}\nu_n(l) = \lambda(|W_n| + \mu\kappa_n).$$

Finally, define

$$V_n(x) = \int_0^x \frac{|W_n| + y\kappa_n}{|W_n| + \mu\kappa_n} \mathrm{d}F(y)$$

and set

$$\begin{aligned} g_n(x) &= \int_{[x,\infty)} \frac{1}{|W_n| + \kappa_n y} \mathrm{d}V_n(y) \\ h_n(x) &= \int_{[x,\infty)} \frac{y - x}{|W_n| + \kappa_n y} \mathrm{d}V_n(y). \end{aligned}$$

2.5.2 Some extra censoring

We have a sequence of statistical problems: observation of $\Phi \cap \mathcal{W}$ through sets W_n. For each fixed n we can proceed as before and, pretending that the observed fragments are independent, write down the score equations. This is exactly what we do in this section, but there is one slight complication. To overcome it, we shall introduce some artificial censoring.

The normalized version of the measure $\mathrm{d}\nu_n$ for a given x is

$$\mathrm{d}\rho_n(l|x) = \frac{l+x}{|W_n| + x\kappa_n}\mathrm{d}\nu_n(l)$$

which may be interpreted as the distribution of the length of the interval through which a fragment is observed, given that the underlying line segment has length x. Now, if we define τ_n to be the supremum of the support of $\mathrm{d}\rho_n(l|x)$ over all x, then τ_n is an upper bound for the length of a fragment that can be observed through W_n. The distribution of the data from observing $\Phi \cap W_n$ is determined by the restriction of V_n to $[0, \tau_n)$ and $H_n = \kappa_n h_n(\tau_n)$. This is changing with n in a very awkward way. We also believe that estimation of F at τ_n from fragments seen through W_n is unstable, because it is typically very unlikely to observe fragments of length near τ_n.

To regularize our problem we introduce some extra censoring. We fix a value τ such that with probability 1 (under the distribution of $\mathcal{W}$), the $\sup_x \operatorname{supp} \mathrm{d}\rho_n(l|x)$ will exceed τ. We group together all observations of length equal to, or exceeding τ, irrespective of their censoring type. In practice it would be wise to choose τ in such a way that a certain percentage of the observed fragments are longer than τ.

We write P_n for the distribution, under the parameter $(V_n|_{[0,\tau)}, H_n)$, of the fragments observed through B_n. The probability that the length of a fragment observed through W_n exceeds τ is

$$\begin{aligned}
P_n(Y > \tau) &= \int_{x>\tau}\int_{l>\tau}\int_{t=\tau-x}^{0} \frac{\mathrm{d}t\mathrm{d}\nu_n(l)\mathrm{d}V_n(x)}{|W_n| + \kappa_n\tau} \\
&+ \int_{x>\tau}\int_{l>\tau}\int_{t=0}^{l-\tau} \frac{\mathrm{d}t\mathrm{d}\nu_n(l)\mathrm{d}V_n(x)}{|W_n| + \kappa_n\tau} \\
&= \int_{l>\tau}(l-\tau)\mathrm{d}\nu_n(l)\int_{x>\tau}\frac{1}{|W_n| + \kappa_n\tau}\mathrm{d}V_n(x) \\
&+ \int_{l>\tau}\mathrm{d}\nu_n(l)\int_{x>\tau}\frac{x-\tau}{|W_n| + \kappa_n\tau}\mathrm{d}V_n(x) \\
&= \int_{l>\tau}(l-\tau)\mathrm{d}\nu_n(l)g_n(\tau) + \int_{l>\tau}\mathrm{d}\nu_n(l)h_n(\tau) \\
&= a_nG_n + b_nH_n
\end{aligned}$$

where

$$\begin{aligned} a_n &= \frac{1}{|W_n| + \kappa_n \tau} \int_{l>\tau} (l - \tau) \mathrm{d}\nu_n(l) \\ b_n &= \frac{1}{\kappa_n} \int_{l>\tau} \mathrm{d}\nu_n(l) \end{aligned}$$

The estimating equations of V_n are similar to what we found earlier, for $x < \tau$

$$\begin{aligned} \mathrm{d}\hat{V}_n(x) &= \mathrm{d}\mathbb{P}_n(x, 0) \\ &+ \frac{\mathrm{d}\hat{V}_n(x)}{|W_n| + \kappa_n x} \int_{y=0}^{x} \frac{1}{\hat{g_n}(y)} \mathrm{d}\mathbb{P}_n(y, 1) \\ &+ \frac{\mathrm{d}\hat{V}_n(x)}{|W_n| + \kappa_n x} \int_{y=0}^{x} \frac{x - y}{\hat{h_n}(y)} \mathrm{d}\mathbb{P}_n(y, 2) \end{aligned} \tag{2.35}$$

The extra censoring does yield an extra term in the estimating equation for H

$$\hat{H}_n = \hat{H}_n \int_{y=0}^{\tau-} \frac{1}{\hat{h}_n(y)} \mathrm{d}\mathbb{P}_n(y, 2) + \frac{b_n \hat{H}_n}{a_n \hat{G}_n + b_n \hat{H}_n} \mathbb{P}_n(Y > \tau)) \tag{2.36}$$

where, cf. (2.16)

$$\hat{G}_n = 1 - \hat{H}_n - \hat{V}_n(\tau^-)$$

and

$$\hat{g_n}(\tau) = \hat{G}_n \ / \ (|W_n| + \kappa_n \tau)) \quad \text{and} \quad \hat{h_n}(\tau) = \hat{H}_n / \kappa$$

and, for $x \in [0, \tau)$,

$$\begin{aligned} \hat{g_n}(x) &= \int_{[x,\tau)} \frac{1}{|W_n| + \kappa_n y} \mathrm{d}\hat{V}_n(y) + \hat{g_n}(\tau) \\ \hat{h_n}(x) &= \int_{[x,\tau)} \frac{y - x}{|W_n| + \kappa_n y} \mathrm{d}\hat{V}_n(y) + \hat{h_n}(\tau) + (\tau - x)\hat{g_n}(\tau). \end{aligned}$$

2.5.3 ergodic lemmas

In this subsection we first work through a number of results establishing that various (properly normalized) functionals of $\mathcal{W} \cap B_n$ converge as n tends to infinity. Then we consider convergence of functionals of $\Phi \cap \mathcal{W} \cap B_n$. We must assume:

Assumption 2.1. $(\Phi, \mathcal{W})$ *is jointly stationary ergodic.*

This assumption actually holds if and only if Φ is weakly mixing and $\mathcal{W}$ is ergodic or vice versa. Recall that weakly mixing implies ergodicity so that in particular we are assuming that $\mathcal{W}$ is ergodic.

Observing the line segments of Φ through $\mathcal{W} \cap B_n$ we have censoring at the boundaries of $\mathcal{W}$ and of B_n. Censoring at the boundaries of B_n is a negligible effect as n tends to infinity because the expected total number of fragments observed through $\mathcal{W} \cap B_n$ is of order n^2, while the expected number of fragments hitting the boundary of B_n is only of order n. In a nutshell, this is what we make precise in this section.

Recall our transformation $\Phi' = \{(\vec{T}_i, X_i, \Theta_i)\}$ of $\Phi = \{(\vec{S}_i, X_i, \Theta_i)\}$ to represent the location of the left endpoint of each line segment relative to the unique straight line on which it lies. Observing $\Phi \cap \mathcal{W} \cap B_n$, we see non-empty intersections $[\vec{T}_i, X_i, \Theta_i, m] = [\vec{T}_i, X_i, \Theta_i] \cap \mathcal{W}_n(T_{2,i}, X_i, \Theta_i, m) \cap B_n = [\vec{T}_i, X_i, \Theta_i] \cap W_n(T_{2,i}, X_i, \Theta_i, m)$ for all m. We named these non-empty intersections 'fragments'. Now consider a modification. Suppose we observe all non-empty intersections $[\vec{T}_i, X_i, \Theta_i, m]^* = [\vec{T}_i, X_i, \Theta_i] \cap \mathcal{W}(T_{2,i}, X_i, \Theta_i, m)$, but only if the left endpoint of $[\vec{T}_i, X_i, \Theta_i, m]^* \in B_n$. In other words, we observe all fragments with left endpoint in the interior of B_n, without censoring at the boundary of B_n. If we assume that these new fragments are a Poisson process, we can condition of their observed number and consider their lengths and censoring types as an independent sample. To describe their common distribution we need to introduce modifications ν_n^* of the measures ν_n. Define

$$l^*(t, \vartheta, m) = \begin{cases} |\mathcal{W}(t, \vartheta, m)| & \text{if the left endpoint of } \mathcal{W}(t, \vartheta, m) \text{ falls in } B_n. \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mathrm{d}\nu_n^*(l) = \int_t \int_\vartheta \sum_m \delta_{l_n^*(t,\vartheta,m)}(\mathrm{d}l)\mathrm{d}t\mathrm{d}K(\vartheta).$$

Conditionally on $\mathcal{W}$, the distribution of the lengths and censoring types of the fragments without censoring at ∂B_n is the same as that of the original fragments, but with ν replaced by ν^* throughout.

We shall now establish the convergence of various functions of ν_n. 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 a set S is a family $\{X_G : G \in \mathcal{G}\}$ where $\mathcal{G}$ is the collection of all bounded Borel sets in S. The spatial process is said to be *covariant* if for any $G \in \mathcal{G}$ and $s \in S$

$$X_{G+s}(T_s) = X_G(.)$$

It is said to be *additive* if

$$X_{H\cup G} = X_H + X_G, \quad \text{if } H, G \in \mathcal{G}, \text{ and } \quad H \cap G = \varnothing.$$

We denote by $\mathcal{K}$ the collection of bounded and convex subsets of S and by B the unit square. We have the following theorem (Nguyen and Zessin (1979),

Theorem 2.1. *If a spatial process $\{X_G : G \in \mathcal{G}\}$ is covariant and additive and if there exists a non-negative random variable $Y \in L^1(P)$ such that*

$$|X_G| \leq Y, \text{ a.s. for each } \quad G \in B \cap \mathcal{K}$$

then the limit

$$\lim_{n\to\infty} \frac{1}{|G_n|} X_{G_n} = E(X_B|\mathcal{J})$$

exists almost surely for each regular countable sequence G_n of sets in $\mathcal{K}$.

Here $\mathcal{J}$ denotes the invariant sigma-algebra. If $\{X_G : G \in \mathcal{G}\}$ is ergodic, i.e. $\mathcal{J}$ is trivial, then the limit equals $E(X_B)$.

We want to establish convergence of V_n and the sub-distribution functions $F_n(y,\delta) = \int_0^y P_n(\mathrm{d}y', \delta)$. To this end, we must first prove convergence—upon normalization—of $|W_n| = \int_l l\mathrm{d}\nu_n(l)$, $\kappa_n = \int_l \mathrm{d}\nu_n(l)$, $a_n = 1/(|W_n| + \kappa_n\tau)\int_{l>\tau}(l-\tau)\mathrm{d}\nu_n(l)$ and $b_n = (1/\kappa_n)\int_{l>\tau}\mathrm{d}\nu_n(l)$. We define $|W_n|^*$, κ_n^* a_n^* and B_n^* similarly as integrals with respect to ν_n^*. We can think of these as random quantities indexed by sets B_n. As such they are additive, covariant and ergodic. The sequence B_n is a regular, countable sequence of convex bounded sets, and $|B_n| = n^2$. By Theorem 2.1 we have $|W_n|^*/n^2 \to E_{\mathcal{W})}|W_1|^*$, $\kappa_n^*/n^2 \to E_{\mathcal{W})}\kappa_1^*$, $a_n^*/n^2 \to E_{\mathcal{W})}a_1^*$ and $b_n^*/n^2 \to E_{\mathcal{W})}\kappa_1^*$. It is not difficult to show that $(|W_n|^* - |W_n|)/n^2 \to 0$, $(\kappa_n - \kappa_n^*)/n^2 \to 0$, $(a_n - a_n^*)/n^2 \to 0$ and $(b_n - b_n^*)/n^2 \to 0$.

We can now show convergence of V_n and its relatives. Recall our definition of V_n,

$$V_n(x) = \int_0^x \frac{|W_n| + y\kappa_n}{|W_n| + \mu\kappa_n}\mathrm{d}F(y)$$

Proposition 2.1. *As n tends to infinity V_n tends almost surely, uniformly to*

$$V_\infty(x) = \int_0^x \frac{E_{\mathcal{W}}|W_1|^* + yE_{\mathcal{W}}\kappa_1^*}{E_{\mathcal{W}}|W_1|^* + \mu E_{\mathcal{W}}\kappa_1^*}\mathrm{d}F(y).$$

Proof We divide numerator and denominator by n^2 and then use the convergence of $|W_n|/n^2$ and κ_n/n^2 to obtain pointwise convergence. Uniform convergence follows from the usual arguments as V_n and V_∞ are distribution functions. □

Of course g_n, G_n, h_n and H_n converge to limits g_∞, G_∞, h_∞ and H_∞, which are defined in the obvious manner as integrals with respect to V_∞.

In section 2.3, we derived the distribution $P(\mathrm{d}y, \delta)$ of the length and censoring type of the fragments observed through $W = W_1$, under the assumption that they are independent. We noted that for a certain Markov kernel $K(\mathrm{d}y, \delta; x)$ we can write $P(\mathrm{d}y, \delta) = KV(\mathrm{d}y, \delta) = \int_x K(\mathrm{d}y, \delta; x)\mathrm{d}V(x)$. For the fragments observed through W_n we find, for $y \in [0, \tau)$,

$$\begin{aligned}
P_n(\mathrm{d}y, 0) &= \frac{\int_{l \geq y}(l - y)\mathrm{d}\nu_n(l)}{|W_n| + \kappa_n y}\mathrm{d}V_n(y) \\
P_n(\mathrm{d}y, 1) &= 2\nu_n([y, \infty))g_n(y)\mathrm{d}y \\
P_n(\mathrm{d}y, 2) &= \mathrm{d}\nu_n(y)h_n(y) \\
P_n(y \geq \tau) &= a_n G_n + b_n H_n.
\end{aligned}$$

Again, P_n can be written as $\int_x K_n(\mathrm{d}y, \delta; x)\mathrm{d}V_n(x)$, where

$$\begin{aligned}
K_n(\mathrm{d}y, \delta; x) &= \mathbf{1}_0(\delta)\mathbf{1}_{[0,\tau)}(y)\frac{\int_{l \geq x}(l - x)\mathrm{d}\nu_n(l)}{|W_n| + \kappa_n x}\delta_x(\mathrm{d}y) \\
&\quad + \mathbf{1}_1(\delta)\mathbf{1}_{[0,\tau)}(y)2\nu_n([y, \infty))\left(\frac{\mathbf{1}_{(y,\tau)}(x)}{|W_n| + \kappa_n x} + \frac{\mathbf{1}_{\{\dagger\}}(x)}{|W_n| + \kappa_n \tau}\right)\mathrm{d}y \\
&\quad + \mathbf{1}_2(\delta)\mathbf{1}_{[0,\tau)}(y)\left(\frac{\mathbf{1}_{(y,\tau)}(x)(x - y)}{|W_n| + \kappa_n x} + \frac{\mathbf{1}_{\{\dagger\}}(x)(\tau - y)}{|W_n| + \kappa_n \tau} + \frac{\mathbf{1}_{\{\ddagger\}}(x)}{\kappa_n}\right)\mathrm{d}\nu_n(y) \\
&\quad + \mathbf{1}_3(\delta)\delta_\tau(\mathrm{d}y)(a_n\mathbf{1}_{\{\dagger\}}(x) + b_n\mathbf{1}_{\{\ddagger\}}(x)).
\end{aligned}$$

Let K_∞ be defined by

$$\begin{aligned}
K_\infty(\mathrm{d}y, \delta; x) &= \mathbf{1}_0(\delta)\mathbf{1}_{[0,\tau)}(y)\frac{\int_{l \geq x}(l - x)\mathrm{d}\nu_1^*(l)}{|W_1|^* + \kappa_1^* x}\delta_x(\mathrm{d}y) \\
&\quad + \mathbf{1}_1(\delta)\mathbf{1}_{[0,\tau)}(y)2\nu_1^*([y, \infty))\left(\frac{\mathbf{1}_{(y,\tau)}(x)}{|W_1|^* + \kappa_1^* x} + \frac{\mathbf{1}_{\{\dagger\}}(x)}{|W_1|^* + \kappa_1^* \tau}\right)\mathrm{d}y \\
&\quad + \mathbf{1}_2(\delta)\mathbf{1}_{[0,\tau)}(y)\left(\frac{\mathbf{1}_{(y,\tau)}(x)(x - y)}{|W_1|^* + \kappa_1^* x} + \frac{\mathbf{1}_{\{\dagger\}}(x)(\tau - y)}{|W_1|^* + \kappa_1^* \tau} + \frac{\mathbf{1}_{\{\ddagger\}}(x)}{\kappa_1^*}\right)\mathrm{d}\nu_1^*(y) \\
&\quad + \mathbf{1}_3(\delta)(a_1^*\mathbf{1}_{\{\dagger\}}(x) + b_1^*\mathbf{1}_{\{\ddagger\}}(x)).
\end{aligned}$$

Proposition 2.2. *$K_n V_n(\mathrm{d}y, \delta)$ tends to $K_\infty V_\infty(\mathrm{d}y, \delta)$ almost surely as n tends to infinity, in the sense of uniform convergence of the associated (sub)distribution functions.*

Proof The proof is similar to that of the previous proposition. □

So far in this section we have only considered functionals of the random set $\mathcal{W}$ as observed through increasing sets B_n. We now turn to the actual data, which are fragments arising from observation of $\Phi \cap \mathcal{W}$ through the B_n. We introduce a sequence of point processes, representing the data observed through B_n.

Recall that we defined

$$\mathcal{D}_n = \{(\vec{t}, x, \vartheta, m) : \ W_n(t_2, \vartheta, m) \neq \emptyset\}.$$

Now define

$$\mathcal{E}_n = \{(\vec{t}, x, \vartheta, m) \in \mathcal{D}_n \ : \ [\vec{t}, x, \vartheta] \cap W_n(t_2, \vartheta, m) \neq \emptyset\}.$$

$\mathcal{E}_n$ is well defined for $n = \infty$, if we take $B_\infty = \mathbb{R}^2$ and $W_\infty(t_2, \vartheta, m) = \mathcal{W}(t_2, \vartheta, m)$. On the sets $\mathcal{E}_n$ we define three functions, p_n, y_n and δ_n.

$$\begin{aligned}
p_n(\vec{t}, x, \vartheta, m) &= \text{left-most endpoint of } [\vec{t}, x, \vartheta] \cap W_n(t_2, \vartheta, m) \\
y_n(\vec{t}, x, \vartheta, m) &= \min(\tau, |[\vec{t}, x, \vartheta] \cap W_n(t_2, \vartheta, m)| \\
\delta_n(\vec{t}, x, \vartheta, m) &= \#\text{ endpoints of } [\vec{t}, x, \vartheta] \text{ outside of } W_n(t_2, \vartheta, m) \\
&\qquad \text{or 3 when } y_n(\vec{t}, x, \vartheta, m) = \tau.
\end{aligned}$$

These definitions are also valid for $n = \infty$. Define, for $n = 1, 2, \ldots, \infty$, point processes Π_n by applying these functions to $\Phi'' \cap \mathcal{E}_n$. Thus, Π_n are marked point processes on B_n with mark space $\mathcal{M} = (0, \tau] \times \{0, 1, 2, 3\}$ representing the fragments observed through B_n. Π_∞ is of special interest. It is a marked point process on $\mathbb{R}^2 \times \mathcal{M}$ representing all fragments arising from intersecting Φ with $\mathcal{W}$. At the beginning of the present section, we claimed that as n tends to infinity, the censoring at the boundary of B_n becomes negligible as compared to the censoring at the boundary of $\mathcal{W}$. The next lemma makes this claim precise.

Lemma 2.2. *For any measurable set $M \subseteq \mathcal{M}$,*

$$\frac{1}{n^2} E_\Phi(|\Pi_n(B_n \times M) - \Pi_\infty(B_n \times M)||\mathcal{W}) \to 0$$

almost surely, as n tends to infinity

Proof The difference between $\Pi_n(B_n \times M)$ and $\Pi_\infty(B_n \times M)$ is less than the number of fragments that cross the boundary of B_n. This, in turn, is less than the number of line segments that cross the boundary of B_n. The expected number of line segments that cross a fixed line segment of length

n (such as one side of B_n) is no more than $n\lambda\mu$, where μ is the (finite) mean of F. This is seen by noting that for a line segment of length x and orientation ϑ to cross a given object line segment, its left-endpoint must lie in a parallelogram with the object line segment for its base, and sides of length x. The area of such a parallelogram is always less than nx. Integrating this with respect to $\lambda \mathrm{d}\vec{t}\mathrm{d}F(x)\mathrm{d}K(\vartheta)$ we find our upper bound, $n\lambda\mu$. We conclude that the expected difference between $\Pi_n(B_n \times M)$ and $\Pi_\infty(B_n \times M)$ is less than $4n\lambda\mu$ and the stated result follows. □

For a given measurable set $M \subseteq \mathcal{M}$, we have

$$\int_M P_n(\mathrm{d}y, \delta) = P_n(M) = \frac{E_\Phi(\Pi_n(B_n \times M)|\mathcal{W})}{E_\Phi(\Pi_n(B_n \times \mathcal{M})|\mathcal{W})}.$$

We define

$$P_n^*(M) = \frac{E_\Phi(\Pi_\infty(B_n \times M)|\mathcal{W})}{E_\Phi(\Pi_\infty(B_n \times \mathcal{M})|\mathcal{W})}$$

and

$$\mathbb{P}_n(M) = \frac{\Pi_n(B_n \times M)}{\Pi_n(B_n \times \mathcal{M})}.$$

We draw the reader's attention to the fact that the subscript n refers to the fact that $\Phi \cap \mathcal{W}$ is observed through B_n and *not* to the number of observed fragments. The number of observed fragments, in fact, equals $\Pi_n(B_n \times \mathcal{M})$.

Everything is now in place for the final result of this section.

Proposition 2.3. *For every measurable set $M \subseteq \mathcal{M}$,*

$$\mathbb{P}_n(M) \to P_\infty(M),$$

in probability, as n tends to infinity.

Proof First, note that $\Pi_\infty(B_n \times M)$ is an additive covariant function indexed by the sets B_n. Since we assume that Φ and $\mathcal{W}$ are jointly ergodic it follows by Theorem 2.1 that $\Pi_\infty(B_n \times M) \to E\Pi_\infty(B_n \times M)$, almost surely. Hence,

$$\frac{\Pi_\infty(B_n \times M)}{\Pi_\infty(B_n \times \mathcal{M})} \to \frac{E\Pi_\infty(B_n \times M)}{E\Pi_\infty(B_n \times \mathcal{M})}.$$

It immediately follows that $P_n^*(M)$ converges to the same limit.

By Lemma 2.2, $P_n^*(M)$ and $P_n(M)$ have the same limit. In Proposition 2.2 we found that $P_n(M)$ converges to $P_\infty(M)$, so we may now conclude that

$$\frac{\Pi_\infty(B_n \times M)}{\Pi_\infty(B_n \times \mathcal{M})} \to P_\infty(M).$$

The reasoning of Lemma 2.2 allows us also to conclude that

$$\left| \frac{\Pi_n(B_n \times M)}{\Pi_n(B_n \times \mathcal{M})} - \frac{\Pi_\infty(B_n \times M)}{\Pi_\infty(B_n \times \mathcal{M})} \right| \to 0$$

in probability. Hence, our claim follows. □

Of course the above convergence is uniform over sets M of the form $[0, y] \times \{\delta\}$ ($y < \tau$, $\delta = 0, 1, 2, 3$), corresponding to 'empirical sub-distribution functions'. Still, the convergence is weaker than when we were working under the Poisson assumption. The Glivenko–Cantelli theorem yields *almost sure* uniform convergence, and here we only have uniform convergence in probability. Fortunately, this is not much of a problem. We simply apply the Skorohod–Dudley–Wichura *almost sure representation theorem* as stated, for instance in Gill (1989). For us, this boils down to pretending that we do have almost sure convergence and going ahead and prove almost sure uniform convergence of our estimator. In the end, we must then weaken the result to uniform convergence in probability.

2.6 An estimator for the line segments

With $\mathbb{P}_n$ as defined in the previous section, we consider estimators

$$\begin{aligned} \mathrm{d}\hat{V}_n(x) &= \mathrm{d}\mathbb{P}_n(x, 0) \\ &+ \frac{\mathrm{d}\hat{V}_n(x)}{|W_n| + \kappa_n x} \int_{y=0}^{x} \frac{1}{\hat{g}_n(y)} \mathrm{d}\mathbb{P}_n(y, 1) \\ &+ \frac{\mathrm{d}\hat{V}_n(x)}{|W_n| + \kappa_n x} \int_{y=0}^{x} \frac{x - y}{\hat{h}_n(y)} \mathrm{d}\mathbb{P}_n(y, 2) \qquad (2.37) \\ \hat{H}_n &= \hat{H}_n \int_{y=0}^{\tau-} \frac{1}{\hat{h}_n(y)} \mathrm{d}\mathbb{P}_n(y, 2) + \frac{b_n \hat{H}_n}{a_n \hat{G}_n + b_n \hat{H}_n} \mathbb{P}_n(Y > \tau)) \qquad (2.38) \end{aligned}$$

where, cf. (2.16)

$$\hat{G}_n = 1 - \hat{H}_n - \hat{V}_n(\tau^-).$$

the functions $\hat{h}_n$ and $\hat{g}_n$ are defined as usual. These estimators are of course inspired by (2.35) and (2.36), but it should be noted that the $\mathbb{P}_n$ have a different meaning here. The subscript n here refers to observation through $W_n = \mathcal{W} \cap B_n$, while the subscript n in (2.35) and (2.36) referred to n observed fragments.

We estimate $(F|_{0,\tau)}, \mu)$ by

$$\hat{\mu}_n = \frac{1}{\kappa_n}\left(\frac{1}{\int_{[0,\tau)} \frac{1}{|W_n|+\kappa_n y}\mathrm{d}\hat{V}_n(y) + \hat{g}_n(\tau)} - |W_n|\right)$$

and

$$\hat{F}_n(x) = \frac{1}{\int_{[0,\tau)} \frac{1}{|W_n|+\kappa_n y}\mathrm{d}\hat{V}_n(y) + \hat{g}_n(\tau)} \int_0^x \frac{1}{|W_n|+\kappa_n y}\mathrm{d}\hat{V}_n(y).$$

We now prove the consistency of our estimator.

Theorem 2.2. *Suppose that $\mathcal{W}$ and Φ are jointly ergodic. Then*

$$\|\hat{F}_n - F_0\|_{[0,\tau)} \to 0 \quad \textit{and} \quad |\hat{\mu}_n - \mu_0| \to 0$$

in probability

Proof The proof will follow the outline we gave at the end of section 2.1.1. Also, we use elements of an earlier consistency proof we gave at the end of section 2.4. However, we no longer need the 'working hypothesis' of that section. There we had difficulty proving that the mle of $g(\tau)$ is bounded away from zero. Since the function g is non-decreasing, it suffices to show that for n large enough $\hat{G}_n(\tau)$ stays away from zero if $G_n(\tau)$ is positive. Because of the extra censoring, this is not a problem anymore.

Proposition 2.1 established that for all $F \in \mathcal{F}$ (and in particular for F_0) $V_{n,F}$ tends to $V_{\infty,F}$ almost surely, uniformly on $[0,\tau)$. Also, Proposition 2.2 yielded that $K_n V_{n,F}$ tends to $K_\infty V_{\infty,F}$ almost surely, in the sense of uniform convergence over sets of the form $[0,y] \times \{\delta\}$ $(y < \tau,\ \delta = 0,1,2,3)$, corresponding to 'empirical sub-distribution functions'.

To apply Theorem 1.2, we need to compare the sequence $K_n\hat{V}_n$ with another sequence $K_n\tilde{V}_n$, say, which converges to $K_\infty V_{\infty,V_{F_0}} = K_\infty V_{\infty,V_0}$. For $\tilde{V}_n$ we choose an appropriately scaled version of the 'empirical average' of uncensored fragments:

$$\tilde{V}_n(x) = \int_0^x \frac{|W_n| + \kappa_n y}{\int_{l\geq y}(l-y)\mathrm{d}\nu(l)}\mathbb{P}_n(\mathrm{d}y, 0), \qquad x \in [0,\tau).$$

We define $\tilde{V}_n(\tau^-)$ in the obvious way. An ergodic theorem can be used to show its convergence of $\tilde{V}_n(\tau^-)$ to $V_{\infty,F_0}(\tau^-)$. Hence, we can take a sequence $\tilde{V}_n(\{\ddagger\}) = \tilde{H}_n$ (such that $\tilde{H}_n \leq \tilde{V}_n(\tau^-)$) converging to $V_{\infty,F_0}(\{\ddagger\}) = \tilde{H}_{\infty,F_0}$. Much like we did in section 2.4, we can apply empirical process theory to

obtain almost sure uniform convergence of $\tilde{V}_n$ to V_{∞,F_0}. Again using empirical process theory, we obtain almost sure convergence of $K_n\tilde{V}_n$ to $K_\infty V_{\infty,F_0}$, uniformly over sets of the form $[0,y]\times\{\delta\}$.

Proposition 2.3 tells us that $\mathbb{P}_n$ also converges to $P_\infty = K_\infty V_{\infty,F_0}$, uniformly over sets of the form $[0,y]\times\{\delta\}$. However, this convergence is 'in probability'. For the remainder of this proof we assume that the convergence is 'almost sure'. This will lead to almost sure consistency of our estimator, which we then modify to consistency in probability, as stated in the theorem. We may do so by the Skorohod–Dudley–Wichura 'almost sure representation theorem', which we discussed above.

Using the fact that $K_n\tilde{V}_n(\mathrm{d}y,0) = \mathbb{P}_n(\mathrm{d}y,o)$ and applying Lemma 2.1, we can now conclude that

$$\int \left(\frac{\mathrm{d}K_n\tilde{V}_n}{\mathrm{d}K_n\hat{V}_n}\right) \mathrm{d}(\mathbb{P}_n - K_n\tilde{V}_n) \to 0$$

almost surely. Theorem 1.2 now allows us to conclude that $K_n\hat{V}_n$ converges to $K_\infty V_{\infty,F_0}$ in the same sense as $K_n\tilde{V}_n$ does. Since we have made sure that our parameter is identifiable it follows that $\hat{V}_n$ converges to V_{∞,F_0}, whence it follows that $\hat{F}_n = F_{n,\hat{V}_n}$ converges to $F_{\infty,V_{\infty,F_0}} = F_0$. □