CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
Telephone + 31 - 20 592 9333
Telefax + 31 - 20 592 4199
URL http://www.cwi.nl

CWI is the nationally funded Dutch institute for research in Mathematics and Computer Science.

Nonparametric estimation for a windowed line-segment process

B.J. Wijers

1991 Mathematics Subject Classification: 60XX (Statistics), 62XX (Probability theory and stochastic processes).
ISBN 90 6196 474 1
NUGI-code: 811

Printed in the Netherlands

Contents

Introduction

Imagine that you are in a tunnel of a coal-mine. On the walls around you, you see cracks of all different lengths and in all kinds of directions. One supposes that the lengths of the cracks are an indicator of rock strength and can be used to assist in deciding the sizes of pillars to support the tunnel of the mine. Therefore, we are interested in the distribution of the crack length, because the distribution function tells us with which probability we will find cracks of a certain length. The motivation for the problem comes from geologists from the CSIRO Division of Applied Geomechanics (Laslett, 1982a,b).

If we want to estimate the crack length distribution we need data. Looking at a wall of the tunnel, we will see something like Figure 0.1. We observe the cracks within some window

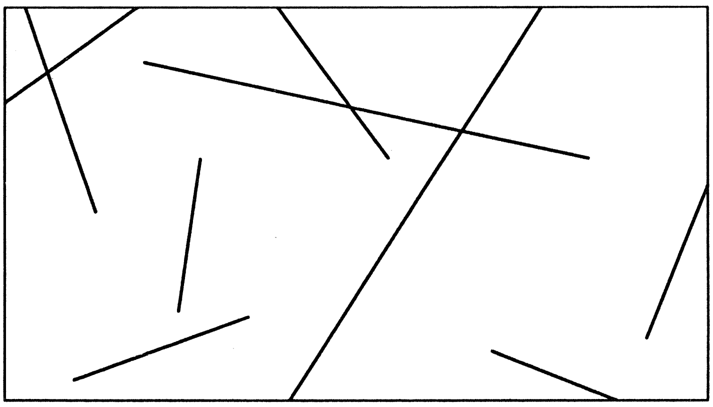

Figure 0.1: line segment process observed through **W**.

W. In this case the window is for instance one of the walls of the tunnel. Some of the lengths of the cracks are totally observed, others partially. A crack that hits the edge of the window is partially observed. The real length of the crack is beyond our visual field. The cracks in the window will be interpreted as a two-dimensional line segment process observed through a window **W**. The cracks are the line segments of the process. When we only observe part of a given crack, we say it is censored. The length of the visible portion we call a censored observation.

For each line segment in the window our data consists of: its (possibly censored) length, its direction and its position in the window. These observations will be used to construct a nonparametric maximum likelihood estimator (NPMLE) of the distribution function of the length of the line segments (the crack length distribution). We assume that lengths and directions are independent. Roughly speaking nonparametric means that we put no restrictions on the distribution function we want to estimate and the maximum likelihood estimator is that estimator that gives the highest probability of occurrence to the data.

Why do we bother about the censoring? It is possible to construct estimators of the crack length distribution based on the uncensored observations only, but it turns out that not using the censored observations is throwing away valuable information. A censored observation of length $\tilde{x}$ does not give us the true length x of the line segment, but nevertheless it tells us that $x \geq \tilde{x}$ and this means it includes information about the probability of being a line segment that is greater or equal to $\tilde{x}$. If experiments are expensive, it is not possible to generate more uncensored data if you are not satisfied with the data that is offered to you. Then, in the search for a good estimator, one is more or less forced to take the censored data into account. In the line segment processes studied in this book we show that the NPMLE of the distribution function of the length of the line segments, based on the uncensored and censored data, is in some sense the best estimator among all other estimators (efficiency).

Another application of a two-dimensional line segment process is described by Chung(1989, 1990): the storage of nuclear waste in a vault in a stable geological rock formation such as the granitic plutons in the Canadian Shield. The rock mass surrounding the vault acts as a natural barrier between the nuclear waste and the biosphere. However the ground water system contaminated with the waste migrates through the rock mass and reaches the biosphere. So the fracture system in the rock plays an important role. The fractures are approximately linear planes in nature, and we can only see linear lines on the surface of the rock mass. These lines form a two-dimensional line segment process and are possibly partially censored (by soil and vegetation). (For a picture see Gill(1994) p. 182).

In order to understand the two-dimensional case we start by analysing a simpler special case: a one-dimensional line segment process observed in an interval. If one considers the rectangular window in Figure 0.1 and all line segments are horizontal, then one obtains the one-dimensional case.

As a possible application of the one-dimensional line segment problem we consider the following so-called 'hospital model', Laslett(1982a). Suppose that in a one-dimensional line segment process we want to estimate the distribution function of the length of the line segments. Suppose we observe the process in an interval $(0, \tau)$ (this is the window $\mathbf{W}$), so there will be line segments that are censored. The edges 0 and τ can censor a line segment at the left-hand side or the right-hand side. For instance the line segments could be the periods between arrival and departure of patients in a hospital and the time interval $(0, \tau)$ is the period we are able to observe them. If the arrival time of a patient is before time 0, then the exact location of the arrival time is unknown. Also the exact locations of departure times after time τ are unknown. Now we have four kinds of observations. For a patient arriving before time 0 and departing after time τ, we observe the sojourn time from time 0 to time τ (*double censored*). The parts of the line segment (the sojourn time) from the arrival time to time 0 and from time τ to the departure time, are not observed. For a patient arriving before time 0 and departing in the time interval $(0, \tau)$, we observe the sojourn time from time 0 to the departure time (***single end censored left-hand side***). For a patient arriving and departing within the time interval $(0, \tau)$, we observe the whole sojourn time (***uncensored***). Finally for a patient arriving within the time interval $(0, \tau)$ and departing after time τ, we observe the sojourn time from the arrival time to time τ (***single end censored right-hand side***). These observations are the available data for the estimation of the distribution function of the sojourn times and form the one-dimensional line segment process observed in the interval $(0, \tau)$.

Let the line segment lengths be distributed with distribution function F; let μ denote the mean. F is the parameter of interest. Let the window $\mathbf{W}$ be convex. (Convexity ensures that two censored line segments hitting the edge of the window, do not belong to the same underlying line segment. This prevents dependence problems in the data.) Another aspect next to censoring that plays a non-trivial role in the line segment processes observed through a window is the so-called length bias: longer line segments have a bigger chance of getting (possibly partially) into the window and being observed. The observed line segments are thus not a random sample from the distribution of interest F, but from the length distribution of the *observed* line segments. The same holds for the distribution function, say K, of the angles of the line segments. The shape of the window has influence on which directions are more likely to be observed.

The length bias problem can be taken account of by estimating this length distribution of the observed line segments and because we will show that there is a 1–1 correspondence between F and this length distribution of the observed line segments, we can always turn back to our parameter of interest F.

In chapter 1 we give a formal introduction to the one- and two-dimensional line segment processes. Firstly we define the one-dimensional case: the line segment process observed in an interval $(0, \tau)$. Secondly, we formulate the two-dimensional case with convex window $\mathbf{W}$ and (un)known angle distribution K. Instead of using the parameter F we pose the problems in terms of the length distribution of the *observed* line segments V (treating the length bias). We show that there is a 1–1 correspondence between the two parametrizations. Actually we show there is a 1–1 correspondence between $(F|_{[0,\tau)}, \mu)$ and $(V|_{[0,\tau)}, h)$, because it turns out that the distribution of the data only depends on these parameters and they are identified by it. We derive the log likelihood-function. Because our model is not dominated we define the NPMLE of (V, h) following Kiefer and Wolfowitz(1956). We use the EM-algorithm to calculate the NPMLE and obtain the so-called self-consistency equations. The existence and uniqueness of the NPMLE will be shown.

There is an important reason why we will use the reparametrization (V, h) of the models. Using (V, h) we turned the cases into a special nonparametric missing data problem: missing data models where the parameter space is *convex* and the distribution of the data is *linear* in the parameter. These properties allow us to use powerful special methods to prove consistency (chapter 2) and asymptotic normality and efficiency (chapter 3).

In chapter 2 we prove consistency of the NPMLE: enlarging the sample size implies that the estimator gets closer (in supremum norm) to the underlying (V, h). We follow a general method first used by Jewell(1982) and more recently by Groeneboom and Wellner(1992). Because of the structure of the log likelihood it is essential to make some adjustments to the method. The proof of consistency will be outlined in a general setting and for the one-dimensional case and the two-dimensional case where **W** is a circle, we work out the proof in detail.

In chapter 3 we show the asymptotic normality and efficiency of the NPMLE of (V, h) (it is in some sense the best estimator), using an identity for NPMLE's in the 'convex-linear' models of Van der Laan(1993). We show that efficiency of the NPMLE of (V, h) implies efficiency of the NPMLE of (F, μ). Again, in the two-dimensional case we restrict attention to the case that **W** is a circle. The proof in the circle case relies on the assumption that the determinant Q_V of a certain 2×2 matrix is unequal to zero. So far we are not able to verify the assumption in general, though a reasonable conjecture is that it exceeds 1 for all V (see chapter 4).

It would be unfair to give the impression that all cases are covered by this book. Only for the one-dimensional case and the two-dimensional case where **W** is a circle, do we give all proofs in detail. For the circle, calculations get less complicated and it turns out that the distribution function K of the directions of the line segments plays no role in the problem. Furthermore we must admit that the efficiency results are obtained under certain conditions on the class of underlying distribution functions. In chapter 4 we give a suggestion to get rid of some of the assumptions. For the two-dimensional case where **W** is arbitrary convex and the angle distribution K is known, one can copy with more effort most of the proofs. The case that **W** is arbitrary convex and K is unknown is different. Now one has to study the joint NPMLE of (V, h) and K, because we need an estimate of K to get an estimate of our parameter of interest (V, h).

Chapter 1

The models

In this chapter we introduce the one- and two-dimensional line segment processes and explain how the data is possibly censored. After choosing a more effective parametrization, we derive the distribution functions of the data and determine the likelihood based on the data and show how to obtain the (sieved) NPMLE of the underlying distribution of the length of the line segments. Section 1.1 deals with the one-dimensional problem observed through an interval $[0, \tau]$ and in section 1.2 we find the two-dimensional case observed through a convex window $\mathbf{W}$.

In the two-dimensional case we need to introduce a distribution function K of the angles of the line segments. It turns out that in the case that $\mathbf{W}$ is a circle the distribution function K (known or unknown) does not play any role in our search for a 'nice' estimator of the distribution of the length of the line segments. We will see that with some more effort the case that $\mathbf{W}$ is arbitrary convex with known K can be treated like the 'circle-case'. In section 1.2.6 a brief remark will be made in case K is unknown.

Although the one-dimensional line segment process could be considered as a special case of the two-dimensional problem, it is certainly not redundant to pay so much attention to this case. Actually, as often happens, we reached results in the two-dimensional case through the one-dimensional problem and a good understanding of the one-dimensional problem very much helps one to foresee the difficulties in the two-dimensional case.

1.1 The one-dimensional line segment process

In this section we introduce the one-dimensional line segment process. By $\int_x^y$, where $x \leq y$ and $x, y \in [0, \infty) \setminus \{\tau\}$, we mean the integral over $(x, y]$, but by $\int_x^\tau$ and $\int_\tau^y$, where $x \leq \tau \leq y$, we mean the integrals over (x, τ) and over $[\tau, y]$ respectively. Thus we have $\int_x^\tau \mathrm{d}F(u) = F(\tau-) - F(x)$ and $\int_\tau^y \mathrm{d}F(u) = F(y) - F(\tau-)$.

1.1.1 The hospital model

Consider arrival times T_i of patients at a hospital enumerated in some way, following a homogeneous Poisson point process on $\mathbb{R}$ with rate λ. Associated with each T_i is a sojourn time X_i; and $X_1, X_2, \ldots$ are i.i.d., positive, independent of the Poisson process and have the common distribution function F. We define $\mu = \int_0^\infty x\,\mathrm{d}F(x)$. For each patient i we have a point $(T_i, X_i) \in \mathbb{R} \times \mathbb{R}^+$. All this defines a point process $\Phi = \{(T_i, X_i) : i = 1, 2, ...\}$ on $\mathbb{R} \times \mathbb{R}^+$ and one can show that Φ can also be characterized as a Poisson point process on $\mathbb{R} \times \mathbb{R}^+$ with intensity measure

$$\rho(\mathrm{d}t, \mathrm{d}x) = \lambda\,\mathrm{d}t\,\mathrm{d}F(x) \tag{1.1}$$

(Karlin(1981) p.p. 436–438, Stoyan(1987)).

Consider a time interval $(0, \tau)$. Suppose we only observe those portions of patients' sojourn times (partly or completely) overlapping $(0, \tau)$: i.e.

$$(T_i, X_i) \in A \equiv \{(t, x) \in \mathbb{R} \times \mathbb{R}^+ \mid [t, t + x] \cap (0, \tau) \neq \emptyset\}.$$

For the patients arriving at or before 0 with sojourn times passing 0, we observe pairs (W_j, E_j), where

$$W_j = \min(T_j + X_j, \tau), \quad E_j = \begin{cases} 1 & T_j + X_j \leq \tau \\ 0 & T_j + X_j > \tau \end{cases}$$

and for the patients arriving within $(0, \tau)$ we observe pairs (Z_i, D_i), where

$$Z_i = \min(X_i, \tau - T_i), \quad D_i = \begin{cases} 1 & X_i \leq \tau - T_i \\ 0 & X_i > \tau - T_i. \end{cases}$$

Together, we have four kinds of observations, respectively *single end censored (left-hand side)* (s.e.c.l.), *double censored* (d.c.), *uncensored* (u.c.) and *single end censored (right-hand side)* (s.e.c.r.) observations. The observations represent the censored line segments of the one dimensional line segment process. A typical realization is given in Figure 1.1, where an open dot means that the observation of the line segment $[T_i, T_i + X_i]$ is *censored* at that side and a closed dot means that this is not the case.

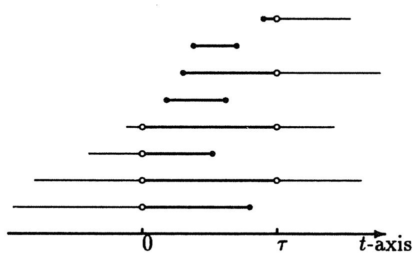

Figure 1.1: line segment process observed in an interval $(0, \tau)$.

Let N be the number of points of the Poisson point process Φ that fall in the set A. Conditioning on $N = n$, the total number of observed patients (the total number of patients with sojourn times overlapping with $(0,\tau)$), we have n independent (possibly partially observed) observations in A. N has a Poisson distribution with parameter $\lambda(\tau+\mu) = \int_A \lambda \, \mathrm{d}t \, \mathrm{d}F(x)$. So N provides information about F through the value of μ, but because λ is unknown the information about F contained in N is not useful and this justifies the presumption that conditioning on $N = n$ causes no loss of information.

In the following picture we draw the set A. Note that $A = \cup_{i=1}^{4} A_i$ and that $(T, X) \in$

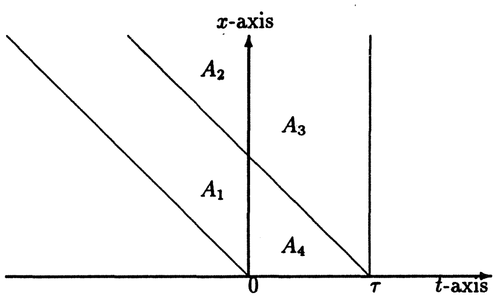

Figure 1.2: representation of set A.

A_i $(i = 1,2,3,4)$ implies that the observation associated with (T, X) is respectively ***single end censored (left-hand side)***, ***double censored***, ***single end censored (right-hand side)*** and ***uncensored.***

For each $(T, X) \in A$ we can construct the observable part of the corresponding line segment geometrically. Let $(p_1, 0)$ be the vertical projection of (T, X) on the t-axis and let $(p_2, 0)$ be the point on the t-axis such that the angle of the positive t-axis and the line through (T, X) and $(p_2, 0)$ is 135^0. In other words we have $p_1 = T$ (the arrival time) and $p_2 = T + X$ (the departure time). Now we observe the intersection of the intervals (p_1, p_2) and $(0, \tau)$. In Figure 1.3 we show this for a point (T, X) in A_2 and A_3. The corresponding line segments are drawn below the t-axis.

1.1.2 Missing data problem

We know that N, the number of points of the Poisson point process Φ that fall in the set A, has a Poisson distribution with parameter $\lambda(\tau+\mu)$. So if we condition on $N = n$, then the set of points $\Phi \cap A$ is distributed like the set of points in an i.i.d. sample of size n with probability measure $1_A(t,x)\,\lambda\,\mathrm{d}t\,\mathrm{d}F(x)/(\lambda(\tau+\mu))$ (see (1.1)). This we can write as

$$1_A(t,x) \cdot \frac{\lambda}{\lambda(\tau+\mu)} \mathrm{d}t \, \mathrm{d}F(x)$$

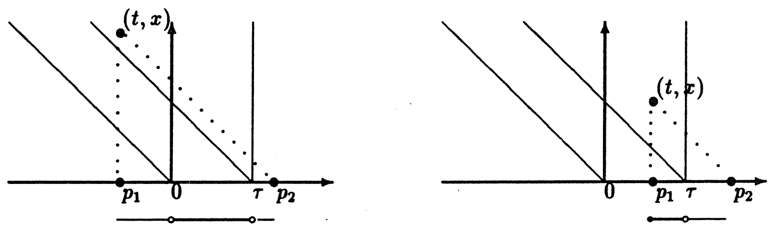

Figure 1.3: constructing observable part geometrically.

$$
\begin{aligned}
&= 1_A(t,x) \cdot \frac{\tau + x}{\tau + \mu} \, \mathrm{d}F(x) \cdot \frac{1}{\tau + x} \, \mathrm{d}t \\
&= 1_A(t,x) \cdot \mathrm{d}V(x) \cdot \frac{1}{\tau + x} \, \mathrm{d}t, \qquad (1.2)
\end{aligned}
$$

where

$$\mathrm{d}V(x) = \frac{\tau + x}{\tau + \mu} \, \mathrm{d}F(x). \qquad (1.3)$$

Note that V is a distribution function. One sees that we obtain the same probability measure if we consider the set of points in a random sample (T_i, X_i) of size n on $\mathbb{R} \times \mathbb{R}^+$, where the X_i's are i.i.d. having the common distribution function V and the T_i's, given $X_i = x_i$, are uniformly distributed on $(-x_i, \tau)$. To get the same kind of observations as in section 1.1.1, we suppose that

$$
\text{if } (T_i = t_i, X_i = x_i) \in \begin{cases} A_1 & \text{, then we observe} \quad t_i + x_i \quad \text{(s.e.c.l.).} \\ A_2 & \text{, " " "} \quad \tau \quad \text{(d.c.).} \\ A_3 & \text{, " " "} \quad \tau - t_i \quad \text{(s.e.c.r.).} \\ A_4 & \text{, " " "} \quad x_i \quad \text{(u.c.).} \end{cases}
$$

The model described in the previous two sentences is a missing data model: the observations are a function of the (T_i, X_i)'s, which are i.i.d., the X_i's having the distribution function V. Note that in the hospital model of section 1.1.1, conditioning on $N = n$, the X_i's in the set of points $\Phi \cap A = \{(T_i, X_i) \in \Phi \mid (T_i, X_i) \in A\}$ no longer have the common distribution function F, from which they are originally drawn, but the distribution function $\int_0^x ((\tau + u)/(\tau + \mu)) \, \mathrm{d}F(u)$. By (1.2) and (1.3) together with the grouping prescription we see that our hospital model can be interpreted as a missing data problem.

Let $F^{s.e.c.l.}$, $F^{d.c.}$, $F^{s.e.c.r.}$ and $F^{u.c.}$ be the subdistribution functions of the observed length of the line segments, where $F^{\cdots}(u)$ stands for the probability of obtaining a ... observation with a value $\leq u$. In Figure 1.4 we show in which set (T_i, X_i) lies, when (T_i, X_i) belongs to an ... observation with a value $\leq u$ ($u \in [0, \tau)$). The horizontal axis is the t-axis and the vertical axis is the x-axis. Now it is easy to calculate the subdistribution functions by integrating

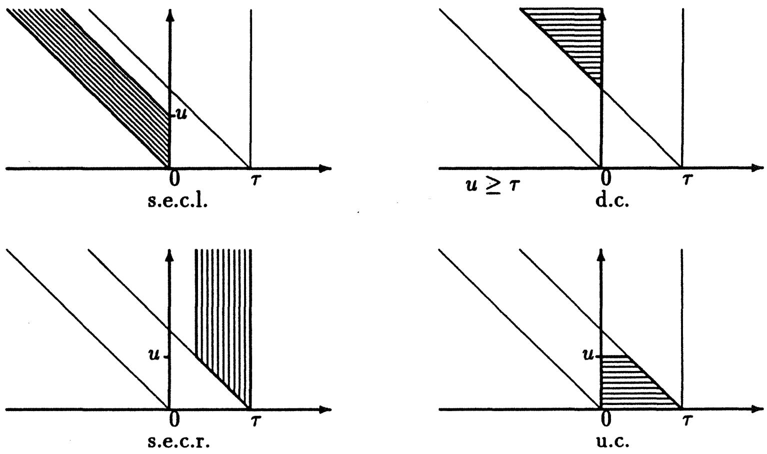

Figure 1.4: the sets corresponding to an ... observation $\leq u$.

$(1/(\tau+x))\,\mathrm{d}V(x)\,\mathrm{d}t$ over the sets drawn in Figure 1.4. For instance for $F^{s.e.c.l.}$ we find

$$\begin{aligned} F^{s.e.c.l.}(u) &= \int_{w=0}^{w=u}\int_{t=-w}^{t=0}\frac{1}{\tau+w}\,\mathrm{d}t\,\mathrm{d}V(w)+\int_{w=u}^{w=\infty}\int_{t=-w}^{t=-w+u}\frac{1}{\tau+w}\,\mathrm{d}t\,\mathrm{d}V(w)\\ &= \int_{w=0}^{w=u}\frac{w}{\tau+w}\,\mathrm{d}V(w)+u\int_{w=u}^{w=\infty}\frac{1}{\tau+w}\,\mathrm{d}V(w) \end{aligned}$$

and so

$$\mathrm{d}F^{s.e.c.l.}(u)=\frac{u}{\tau+u}\,\mathrm{d}V(u)-\frac{u}{\tau+u}\,\mathrm{d}V(u)+\int_{w=u}^{w=\infty}\frac{1}{\tau+w}\,\mathrm{d}V(w)\,\mathrm{d}u$$

($u\in[0,\tau)$). For the other subdistribution functions we can do similar calculations. We find

$$\begin{aligned} \mathrm{d}F^{s.e.c.l.}(u) &= 1_{[0,\tau)}(u)\int_u^\infty\frac{1}{\tau+w}\,\mathrm{d}V(w)\,\mathrm{d}u\\ &= 1_{[0,\tau)}(u)\cdot g(u)\,\mathrm{d}u && (1.4)\\ \mathrm{d}F^{d.c.}(u) &= \int_\tau^\infty\frac{w-\tau}{\tau+w}\,\mathrm{d}V(w)\,\mathrm{d}\delta_\tau(u)\\ &= h\ \mathrm{d}\delta_\tau(u) && (1.5)\\ \mathrm{d}F^{s.e.c.r.}(u) &= 1_{[0,\tau)}(u)\int_u^\infty\frac{1}{\tau+w}\,\mathrm{d}V(w)\,\mathrm{d}u\\ &= 1_{[0,\tau)}(u)\cdot g(u)\,\mathrm{d}u && (1.6)\\ \mathrm{d}F^{u.c.}(u) &= 1_{[0,\tau)}(u)\frac{\tau-u}{\tau+u}\,\mathrm{d}V(u), && (1.7) \end{aligned}$$

where $\delta_\tau(u) = 1$ if $u = \tau$ and 0 elsewhere and where $g(\cdot)$ and h are defined as

$$g(x) \equiv \int_x^\infty \frac{1}{\tau + w} \mathrm{d}V(w) \ , \ h \equiv \int_\tau^\infty \frac{w - \tau}{\tau + w} \mathrm{d}V(w). \tag{1.8}$$

For convenience we define

$$F^{s.e.c.}(x) \equiv F^{s.e.c.l.}(x) + F^{s.e.c.r.}(x) \tag{1.9}$$

and note that $\mathrm{d}F^{s.e.c.}(u) = 2\,g(u)\,\mathrm{d}u$. Furthermore we note that h is the probability of being double censored.

When we discuss the likelihood in section 1.1.4 we need the subdistribution functions in terms of F too. Therefore we give them here. One can find them by using (1.3) with (1.4)–(1.7). We get

$$\mathrm{d}F^{s.e.c.l.}(u) = 1_{[0,\tau)}(u) \frac{1 - F(u)}{\tau + \mu} \mathrm{d}u \tag{1.10}$$

$$\mathrm{d}F^{d.c.}(u) = \int_\tau^\infty \frac{1 - F(w)}{\tau + \mu} \mathrm{d}w \, \mathrm{d}\delta_\tau(u) \tag{1.11}$$

$$\mathrm{d}F^{s.e.c.r.}(u) = 1_{[0,\tau)}(u) \frac{1 - F(u)}{\tau + \mu} \mathrm{d}u \tag{1.12}$$

$$\mathrm{d}F^{u.c.}(u) = 1_{[0,\tau)}(u) \frac{\tau - u}{\tau + \mu} \mathrm{d}F(u). \tag{1.13}$$

Note that $F^{s.e.c.l.}$ and $F^{s.e.c.r.}$ are the same and so these observations give the same contribution to the problem. Or one can say that the edges 0 and τ of the time interval $(0, \tau)$ censor the line segments equally. In fact the departure times $T_i + X_i$ $(i = 1, ..., n)$ form a Poisson point process too, independent of the X_i's.

Before finishing this section we give some equalities, which belong to the model. We have the following equalities:

$$1 = V(\tau-) + 2\tau g(\tau) + h \tag{1.14}$$

$$\frac{\mu}{\tau + \mu} = \int_0^\tau \frac{w}{\tau + w} \mathrm{d}V(w) + \tau g(\tau) + h \tag{1.15}$$

$$\frac{\tau}{\tau + \mu} = \int_0^\tau \frac{\tau}{\tau + w} \mathrm{d}V(w) + \tau g(\tau) \tag{1.16}$$

and

$$\begin{aligned} \int_0^\tau \frac{\tau - w}{\tau + w} \mathrm{d}V(w) + 2\int_0^\tau g(w)\,\mathrm{d}w + h &= \\ -\int_0^\tau (\tau - w)\,\mathrm{d}g(w) + 2\int_0^\tau g(w)\,\mathrm{d}w + h &= 1. \end{aligned} \tag{1.17}$$

Furthermore we have

$$\frac{1}{\tau + \mu} = \frac{1}{2\tau}\,(1 - h - V(\tau-)) + \int_0^\tau \frac{1}{\tau + w} \mathrm{d}V(w) \tag{1.18}$$

and with (1.3) we can rewrite (1.18) as

$$h = 1 - \frac{2\tau}{\tau + \mu} + \int_0^\tau \frac{\tau - w}{\tau + \mu}\,\mathrm{d}F(w). \tag{1.19}$$

With (1.14) we express $g(x)$ in terms of V on $[0, \tau)$ and h:

$$\begin{aligned} g(x) &= \int_x^\tau \frac{1}{\tau + w}\,\mathrm{d}V(w) + g(\tau) \\ &= \int_x^\tau \frac{1}{\tau + w}\,\mathrm{d}V(w) + \frac{1}{2\tau}(1 - h - V(\tau-)). \end{aligned} \tag{1.20}$$

Note that '(1.14)=(1.15)+(1.16)' and (1.17) can be obtained from Figure 1.2 by integrating over A_4 which yields $\int_0^\tau (\tau - w)/(\tau + w)\,\mathrm{d}V(w)$, over A_1 and A_3 which gives $2\int_0^\tau g(w)\,\mathrm{d}w$, and over A_2 which gives h. Figure 1.2 gives a nice geometrical interpretation. One sees that h is the probability that an observation is double censored ((T_i, X_i) lies in the set A_2), $2\tau g(\tau)$ is the probability that an observation is single end censored and $X \geq \tau$ ((T_i, X_i) lies in the set $(A_1 \cup A_3) \cap \{(t, x) \in \mathbb{R} \times \mathbb{R}^+ \mid x \geq \tau\}$), $\mu/(\tau + \mu)$ is the probability that an observation is an observation of a patient arriving before time 0 ((T_i, X_i) lies in the set $A_1 \cup A_2$) and $\tau/(\tau + \mu)$ is the probability that an observation is an observation of a patient arriving within the time interval $(0, \tau)$ ((T_i, X_i) lies in the set $A_3 \cup A_4$).

1.1.3 Identification

To avoid confusion we remember that by $\int_x^y$, where $x \leq y$ and $x, y \in [0, \infty) \setminus \{\tau\}$, we mean the integral over $(x, y]$, but by $\int_x^\tau$ and $\int_\tau^y$, where $x \leq \tau \leq y$, we mean the integrals over (x, τ) and over $[\tau, y]$ respectively. Thus we have $\int_x^\tau \mathrm{d}F(u) = F(\tau-) - F(x)$ and $\int_\tau^y \mathrm{d}F(u) = F(y) - F(\tau-)$.

If we remember (1.4)–(1.7) and (1.10)–(1.13), then with (1.20) we note that the distribution of the data only depends on V (or F) through V on $[0, \tau)$ (or F on $[0, \tau)$) and h (or μ) and these two parameters are identified by it. As we already mentioned in the introduction, we want to translate the model from the parameters F and μ into the parameters V and h.

We are interested in distribution functions $F \in \mathcal{F}$, where $\mathcal{F}$ is the set

$$\mathcal{F} \equiv \{ \text{ all distribution functions on } [0, \infty) \text{ with finite mean } \}.$$

Of course the translation

$$V(x) \equiv \int_0^x \frac{\tau + w}{\tau + \mu}\,\mathrm{d}F(w) \tag{1.21}$$

does not go from $\mathcal{F}$ to $\mathcal{F}$, but from $\mathcal{F}$ to $\mathcal{F}_\infty$, where $\mathcal{F}_\infty$ is the set

$$\mathcal{F}_\infty \equiv \{ \text{ all distribution functions on } [0, \infty) \ \}.$$

One notes that $\mathcal{F} \subset \mathcal{F}_\infty$. We define the set $\mathcal{S}_\tau$ as

$$\mathcal{S}_\tau \equiv \{ \text{ all subdistribution functions on } [0, \tau) \ \}.$$

Let $P_{(F,\mu)}$ denote the distribution function of the data w.r.t. the parameters $(F,\mu) \in \mathcal{S}_\tau \times [0,\infty]$. If we define the map $m_0 : \mathcal{F}_\infty \to \mathcal{S}_\tau \times [0,\infty]$

$$m_0(G) \equiv \left(G_\tau, \int_0^\infty w \,\mathrm{d}G(w)\right) \equiv (F,\mu),$$

where $G_\tau \in \mathcal{S}_\tau$ is G restricted to $[0,\tau)$ (thus $G_\tau(x) = G(x)$ for $x \in [0,\tau)$), then our model $\mathcal{M}$, which is the set of all possible distribution functions of the data and where the set of all possible underlying distribution functions is $\mathcal{F}$, can be described as follows

$$\mathcal{M} = \left\{ P_{(F,\mu)} \,\middle|\, (F,\mu) \in m_0(\mathcal{F}) \right\}.$$

By a certain map $\varphi_\tau : \mathcal{S}_\tau \times [0,\infty] \to \mathcal{S}_\tau \times [0,\infty)$ we will pass from (F,μ) to the parametrization (V,h), where V and F satisfy (1.21) on $[0,\tau)$. If we denote by $P_{(V,h)}$ the distribution function of the data w.r.t. the parameters $(V,h) \in \mathcal{S}_\tau \times [0,\infty)$, then we can describe our model $\mathcal{M}$ as

$$\mathcal{M} = \left\{ P_{(V,h)} \,\middle|\, (V,h) \in \varphi_\tau(m_0(\mathcal{F})) \right\}.$$

Such a reparametrization is only useful if we can go back and forth between the two parametrizations in our model $\mathcal{M}$. In other words is there a 1–1 correspondence between $m_0(\mathcal{F})$ and $\varphi_\tau(m_0(\mathcal{F}))$? Because of the fact that for both parametrizations the parameters are identifiable from the data, there exists such a 1–1 correspondence.

A natural extension of $\mathcal{M}$ would be

$$\mathcal{N} = \left\{ P_{(V,h)} \,\middle|\, (V,h) \in \overline{\varphi_\tau(m_0(\mathcal{F}))} \right\},$$

where $\overline{A}$ stands for the closure of set A. By ∂A we mean the boundary of set A, thus $\partial A = \overline{A} \setminus \mathring{A}$, where $\mathring{A}$ stands for the interior of A. The model $\mathcal{N}$ is 'too big'.

The set $m_0(\mathcal{F}_\infty \setminus \mathcal{F})$ consists of elements, which do not belong to a distribution function in $\mathcal{F}$, but to a distribution function with infinite mean. If $F \in \mathcal{F}_\infty \setminus \mathcal{F}$, then all data sets contain only double censored observations a.s. and the probability of being double censored equals $h = 1$. In this case the translation from F to V gives us $V \equiv 0$. This implies

$$\varphi_\tau(m_0(\mathcal{F}_\infty \setminus \mathcal{F})) = \{(0,1)\}$$

and we will have that $\varphi_\tau(m_0(\mathcal{F}_\infty \setminus \mathcal{F})) \subset \partial\varphi_\tau(m_0(\mathcal{F}))$. Surely, there is no 1–1 correspondence between $m_0(\mathcal{F}_\infty \setminus \mathcal{F})$ and $\{(0,1)\}$.

We will see that the rest of the boundary of $\varphi_\tau(m_0(\mathcal{F}))$, which does not belong to $\varphi_\tau(m_0(\mathcal{F}))$ itself, is not empty and consists of (V,h), which can be transferred back to an $(F,\mu) \in \mathcal{S}_\tau \times [0,\infty]$, but these F on $[0,\tau)$ can not be extended to an F on $[0,\infty)$ with mean μ. Certainly, this is the case if $F(\tau-) = 1$ and $\int_0^\tau w \,\mathrm{d}F(w) < \mu$. This rest of $\partial\varphi_\tau(m_0(\mathcal{F}))$ will be denoted by $\mathcal{V}_<$. Thus we have

$$\mathcal{V}_< = \partial\varphi_\tau(m_0(\mathcal{F})) \setminus \{\{(0,1)\} \cup \varphi_\tau(m_0(\mathcal{F}))\}.$$

It will be shown that there is a 1–1 correspondence between $\mathcal{V}_<$ and $\mathcal{F}_< \equiv \varphi_\tau^{-1}(\mathcal{V}_<)$, where φ_τ^{-1} stands for the inverse image of φ_τ. We conclude that

$$\overline{\varphi_\tau(m_0(\mathcal{F}))} = \varphi_\tau(m_0(\mathcal{F})) \cup \{(0,1)\} \cup \mathcal{V}_<.$$

Let us call ψ_τ on $\varphi_\tau(m_0(\mathcal{F}))$ and on $\mathcal{V}_<$ the inverse of φ_τ. (We will show that for our choice of φ_τ this inverse exists). By φ_τ^{-1} on $\{(0,1)\}$ we denote the inverse image of φ_τ. In the following scheme we summarize the above paragraphs.

$$\begin{array}{lllllllll}
\mathcal{F}_\infty & & & = & \mathcal{F} & \cup & \mathcal{F}_\infty\setminus\mathcal{F} & & \\
\downarrow m_0 & & & & \downarrow m_0 & & \downarrow m_0 & & \\
m_0(\mathcal{F}_\infty) & \cup & \mathcal{F}_< & = & m_0(\mathcal{F}) & \cup & m_0(\mathcal{F}_\infty\setminus\mathcal{F}) & \cup & \mathcal{F}_< \\
\downarrow\varphi_\tau\ \uparrow\psi_\tau & & \downarrow\varphi_\tau\ \uparrow\psi_\tau & & \downarrow\varphi_\tau\ \uparrow\psi_\tau & & \downarrow\varphi_\tau\ \uparrow\varphi_\tau^{-1} & & \downarrow\varphi_\tau\ \uparrow\psi_\tau \\
 & \overline{\varphi_\tau(m_0(\mathcal{F}))} & & = & \varphi_\tau(m_0(\mathcal{F})) & \cup & \{(0,1)\} & \cup & \mathcal{V}_<.
\end{array}$$

Now we will fill in this scheme and describe the sets and mappings concretely. We define the sets

$$\begin{aligned}
\mathcal{F}_{\tau,\star,<} &\equiv \{(F,\mu)\,|\,F\in\mathcal{S}_\tau,\ \mu\in[0,\infty],\ \textstyle\int_0^\tau w\,\mathrm{d}F(w)+\tau(1-F(\tau-))\le\mu\} \\
\mathcal{F}_\tau &\equiv \{(F,\mu)\,|\,F\in\mathcal{S}_\tau,\ \mu\in[0,\infty),\ \textstyle\int_0^\tau w\,\mathrm{d}F(w)+\tau(1-F(\tau-))\le\mu \\
&\qquad \text{and '=' when } F(\tau-)=1\} \\
\mathcal{F}_< &\equiv \{(F,\mu)\,|\,F\in\mathcal{S}_\tau,\ \mu\in[0,\infty),\ \textstyle\int_0^\tau w\,\mathrm{d}F(w)<\mu,\ F(\tau-)=1\} \\
\mathcal{F}_\star &\equiv \{((F,\mu)\,|\,F\in\mathcal{S}_\tau,\ \mu=\infty\}\,,
\end{aligned}$$

and the sets

$$\begin{aligned}
\mathcal{V}_{\tau,\star,<} &\equiv \{(V,h)\,|\,V\in\mathcal{S}_\tau,\ h\in[0,\infty),\ h\le 1-V(\tau-)\} \\
\mathcal{V}_\tau &\equiv \{(V,h)\,|\,(V,h)\in\mathcal{V}_{\tau,\star,<},\ h\le 1-V(\tau-) \text{ and '=' when } h=0\} \\
\mathcal{V}_< &\equiv \{(V,h)\,|\,(V,h)\in\mathcal{V}_{\tau,\star,<},\ 0<h=1-V(\tau-),\ h\neq 1\} \\
\mathcal{V}_\star &\equiv \{(V,h)\,|\,(V,h)\in\mathcal{V}_{\tau,\star,<},\ h=1,\ V(\tau-)=0\} \\
&= \{(V,h)\,|\,(V,h)=(0,1)\}\,.
\end{aligned}$$

It is clear that

$$\mathcal{F}_{\tau,\star,<}=\mathcal{F}_\tau\cup\mathcal{F}_\star\cup\mathcal{F}_<\ ,\ \mathcal{V}_{\tau,\star,<}=\mathcal{V}_\tau\cup\mathcal{V}_\star\cup\mathcal{V}_<$$

and that we have

$$\overline{\mathcal{F}}_\tau=\mathcal{F}_{\tau,\star,<}\ ,\ \overline{\mathcal{V}}_\tau=\mathcal{V}_{\tau,\star,<}.$$

How are the sets $\mathcal{F}_\tau$, $\mathcal{F}_<$ and $\mathcal{F}$ related? To give an answer we define for this purpose for each $(F,\mu)\in\mathcal{S}_\tau\times[0,\infty)$ the following set:

$$\mathcal{R}_{(F,\mu)}\equiv\left\{G\in\mathcal{F}\,\middle|\,G(x)=F(x)\text{ on }[0,\tau),\ \int_0^\infty w\,\mathrm{d}G(w)=\mu\right\}\subset\mathcal{F}.$$

One sees that each $(F,\mu)\in\mathcal{S}_\tau\times[0,\infty)$ represents a set of distribution functions in $\mathcal{F}$. Actually we will see that each $(F,\mu)\in\mathcal{F}_\tau$ represents a nonempty set $\mathcal{R}_{(F,\mu)}$ in $\mathcal{F}$ and one immediately notes that for each $(F,\mu) \in \mathcal{F}_<$ we have $\mathcal{R}_{(F,\mu)} = \emptyset$. Proposition 1.1.3.1 will give us the answer to the question we posed and is formulated as follows:

$$\mathcal{F} = \bigcup_{(F,\mu)\in\mathcal{S}_\tau\times[0,\infty)} \mathcal{R}_{(F,\mu)} = \bigcup_{(F,\mu)\in\mathcal{F}_\tau} \mathcal{R}_{(F,\mu)} \tag{1.22}$$

and

$$\forall (F,\mu) \in \mathcal{F}_\tau : \mathcal{R}_{(F,\mu)} \neq \emptyset. \tag{1.23}$$

With the knowledge that $m_0(\mathcal{R}_{(F,\mu)}) = \{(F,\mu)\}$ we have that these two statements immediately imply that

$$m_0(\mathcal{F}) = \bigcup_{(F,\mu)\in\mathcal{F}_\tau} m_0(\mathcal{R}_{(F,\mu)}) = \bigcup_{(F,\mu)\in\mathcal{F}_\tau} \{(F,\mu)\} = \mathcal{F}_\tau. \tag{1.24}$$

With the definition of m_0 on $\mathcal{F}_\infty \setminus \mathcal{F}$ we may write

$$m_0(\mathcal{F}_\infty \setminus \mathcal{F}) = \{(F,\infty) \mid F \in \mathcal{S}_\tau\} = \mathcal{F}_\star. \tag{1.25}$$

Now we will define the map φ_τ on

$$\mathcal{F}_{\tau,\star,<} = m_0(\mathcal{F}_\infty) \cup \mathcal{F}_<,$$

which we use to pass from (F,μ) to the parametrization (V,h). Furthermore we define the map ψ_τ on

$$\varphi_\tau(m_0(\mathcal{F}) \cup \mathcal{F}_<),$$

which we use to pass back to the parametrization (F,μ). We define

$$\begin{aligned} \varphi_\tau(F(\cdot),\mu) &\equiv \left(\int_0^{\cdot} \frac{\tau+w}{\tau+\mu}\,\mathrm{d}F(w),\, 1 - \frac{2\tau}{\tau+\mu} + \int_0^{\tau} \frac{\tau-w}{\tau+\mu}\,\mathrm{d}F(w)\right) \\ \psi_\tau(V(\cdot),h) &\equiv \left(\frac{1}{\nu(V(\cdot),h)} \int_0^{\cdot} \frac{1}{\tau+w}\,\mathrm{d}V(w),\, \frac{1}{\nu(V(\cdot),h)} - \tau\right), \end{aligned}$$

(see also (1.19)) where

$$\nu(V(\cdot),h) \equiv \frac{1}{2\tau}(1 - h - V(\tau-)) + \int_0^{\tau} \frac{1}{\tau+w}\,\mathrm{d}V(w).$$

(We keep in mind that $\nu(V(\cdot),h)$ corresponds with $1/(\tau+\mu)$ (see (1.18)).

In proposition 1.1.3.2 we show that φ_τ and ψ_τ give a 1–1 correspondence between $\mathcal{F}_\tau$ and $\mathcal{V}_\tau$. In proposition 1.1.3.3 we prove that φ_τ and ψ_τ give a 1–1 correspondence between $\mathcal{F}_<$ and $\mathcal{V}_<$. These results with (1.24), (1.25) and the straightforward calculation of $\varphi_\tau(m_0(\mathcal{F}_\infty \setminus \mathcal{F}))$ yields

$$\begin{aligned} \varphi_\tau(m_0(\mathcal{F})) &= \varphi_\tau(\mathcal{F}_\tau) = \mathcal{V}_\tau \\ \varphi_\tau(\mathcal{F}_<) &= \mathcal{V}_< \end{aligned}$$

and

$$\varphi_\tau(m_0(\mathcal{F}_\infty \setminus \mathcal{F})) = \varphi_\tau(\{(F,\infty) \mid F \in \mathcal{S}_\tau\}) = \{(0,1)\} = \mathcal{V}_\star.$$

So we have $\varphi_\tau(m_0(\mathcal{F}_\infty) \cup \mathcal{F}_<) = \mathcal{V}_{\tau,\star,<} = \mathcal{V}_\tau = \overline{\varphi_\tau(m_0(\mathcal{F}))}$.

The scheme becomes

$$
\begin{array}{lclclclcl}
\mathcal{F}_\infty & & & = & \mathcal{F} & \cup & \mathcal{F}_\infty \setminus \mathcal{F} & & \\
\downarrow m_0 & & & & \downarrow m_0 & & \downarrow m_0 & & \\
m_0(\mathcal{F}_\infty) & \cup & \mathcal{F}_< & = & \mathcal{F}_\tau & \cup & \mathcal{F}_\star & \cup & \mathcal{F}_< \\
\downarrow \varphi_\tau \ \uparrow \psi_\tau & & \downarrow \varphi_\tau \ \uparrow \psi_\tau & & \downarrow \varphi_\tau \ \uparrow \psi_\tau & & \downarrow \varphi_\tau \ \uparrow \varphi_\tau^{-1} & & \downarrow \varphi_\tau \ \uparrow \psi_\tau \\
 & \mathcal{V}_{\tau,\star,<} & & = & \mathcal{V}_\tau & \cup & \{(0,1)\} & \cup & \mathcal{V}_<.
\end{array}
$$

From now on we describe the model $\mathcal{M}$ as

$$\mathcal{M} = \left\{ P_{(F,\mu)} \,\middle|\, (F,\mu) \in \mathcal{F}_\tau \right\} = \left\{ P_{(V,h)} \,\middle|\, (V,h) \in \mathcal{V}_\tau \right\}$$

and use the parametrization (V,h), because of its earlier mentioned advantages, and know that we can change back to $(F,\mu) \in \mathcal{F}_\tau$ without difficulties. In our search for the NPMLE we will maximize the likelihood over the set $\overline{\mathcal{V}}_\tau = \mathcal{V}_{\tau,\star,<}$ so the given specified description of the larger set could be useful.

In the remainder of this section we prove the propositions 1.1.3.1 – 1.1.3.3.

Proposition 1.1.3.1 *Using the definitions of the sets and functions from above, we have (1.22) and (1.23).*

PROOF: we begin with the first equality in (1.22). It is obvious that we have '$\supset$', because for all (F,μ) we have $\mathcal{R}_{(F,\mu)} \subset \mathcal{F}$. Each $G \in \mathcal{F}$ is an element of $\mathcal{R}_{(F_G,\mu_G)}$, where F_G equals G restricted to $[0,\tau)$ and μ_G equals the expectation of G. (Note that $G \in \mathcal{F} \Rightarrow 0 \leq \mu_G < \infty$). So we have '$\subset$' too.

If we prove $\mathcal{F} = \cup_{(F,\mu)\in\mathcal{F}_\tau} \mathcal{R}_{(F,\mu)}$, then we automatically prove the second equality in (1.22). Again, it is obvious that we have '$\supset$'. We know that each $G \in \mathcal{F}$ is an element of $\mathcal{R}_{(F_G,\mu_G)}$, where $F_G(x) = G(x)$ (on $[0,\tau)$, thus $F_G(\tau-) = G(\tau-)$) and $\mu_G = \int_0^\infty w \, \mathrm{d}G(w)$. For F_G and μ_G we note that μ_G is finite and nonnegative because $G \in \mathcal{F}$ and we have

$$
\begin{aligned}
\infty > \mu_G &= \int_0^\tau w \, \mathrm{d}G(w) + \int_\tau^\infty w \, \mathrm{d}G(w) \\
&= \int_0^\tau w \, \mathrm{d}F_G(w) + \int_\tau^\infty w \, \mathrm{d}G(w) \\
&\geq \int_0^\tau w \, \mathrm{d}F_G(w) + \tau\,(1 - G(\tau-)) \\
&= \int_0^\tau w \, \mathrm{d}F_G(w) + \tau\,(1 - F_G(\tau-))
\end{aligned}
$$

and if $F_G(\tau-) = G(\tau-) = 1$ we have

$$\int_0^\tau w \, \mathrm{d}F_G(w) = \int_0^\tau w \, \mathrm{d}G(w) = \int_0^\infty w \, \mathrm{d}G(w) = \mu_G.$$

This proves that $(F_G, \mu_G) \in \mathcal{F}_\tau$ and thus $G \in \cup_{(F,\mu)\in\mathcal{F}_\tau}\mathcal{R}_{(F,\mu)}$. So we have showed '$\subset$'. This completes the proof of (1.22).

To prove (1.23) we take a $(F,\mu) \in \mathcal{F}_\tau$. Now we look for a $G \in \mathcal{F}$ such that $G(x) = F(x)$ on $[0,\tau)$ and $\int_0^\infty w\,\mathrm{d}G(w) = \mu$. On $[0,\tau)$ we have to take $G(x) = F(x)$. If $F(\tau-) = 1$, then we are ready and choose $G(x) \equiv 1$ on $[\tau,\infty)$. (Note that $\int_0^\infty w\,\mathrm{d}G(w) = \mu$ is satisfied). If $F(\tau-) < 1$, then let us search for a G, which gives mass on $[\tau,\infty)$ only at τ and at some point $b > \tau$. By $\Delta G(\tau)$ and $\Delta G(b)$ we denote the height of the jump of G at τ respectively b.

If G is a distribution function in $\mathcal{F}$ such that it is an element of $\mathcal{R}_{(F,\mu)}$, then $\Delta G(\tau)$ and $\Delta G(b)$ have to satisfy

$$\int_0^\infty w\,\mathrm{d}G(w) = \int_0^\tau w\,\mathrm{d}F(w) + \tau\Delta G(\tau) + b\Delta G(b) = \mu \tag{1.26}$$

$$\int_0^\infty \mathrm{d}G(w) = F(\tau-) + \Delta G(\tau) + \Delta G(b) = 1 \tag{1.27}$$

under the condition that $\Delta G(\tau) \geq 0$ and $\Delta G(b) \geq 0$. Now (1.26) and (1.27) imply

$$\begin{pmatrix} \tau & b \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \Delta G(\tau) \\ \Delta G(b) \end{pmatrix} = \begin{pmatrix} \mu - \int_0^\tau w\,\mathrm{d}F(w) \\ 1 - F(\tau-) \end{pmatrix}.$$

Because $b > \tau$ the inverse of the matrix exists and we have

$$\Delta G(\tau) = \frac{1}{\tau - b}\left(\mu - \int_0^\tau w\,\mathrm{d}F(w)\right) - \frac{b}{\tau - b}(1 - F(\tau-)) \tag{1.28}$$

$$\Delta G(b) = -\frac{1}{\tau - b}\left(\mu - \int_0^\tau w\,\mathrm{d}F(w) - \tau(1 - F(\tau-))\right). \tag{1.29}$$

Now $-1/(\tau - b) > 0$ and because $(F,\mu) \in \mathcal{F}_\tau$ we have $\mu - \int_0^\tau w\,\mathrm{d}F(w) - \tau(1 - F(\tau-)) \geq 0$ and thus with (1.29) we conclude that $\Delta G(b) \geq 0$. For $b \to \infty$ we have that $-b/(\tau - b)\cdot(1 - F(\tau-))$ tends to $(1 - F(\tau-)) > 0$ and $1/(\tau - b)\cdot(\mu - \int_0^\tau w\,\mathrm{d}F(w))$ tends to 0. Thus there exists a $b \in (\tau,\infty)$ such that $\Delta G(\tau) \geq 0$. This completes the construction of a $G \in \mathcal{R}_{(F,\mu)}$ for a given $(F,\mu) \in \mathcal{F}_\tau$ and proves (1.23). □

Proposition 1.1.3.2 *If we consider the map φ_τ on $\mathcal{F}_\tau$ and the map ψ_τ on $\mathcal{V}_\tau$, then they give a 1-1 correspondence between $\mathcal{F}_\tau$ and $\mathcal{V}_\tau$.*

Proof: we have indeed that if $(V(\cdot), h) \equiv \varphi_\tau(F(\cdot),\mu)$ for an $(F,\mu) \in \mathcal{F}_\tau$, then $V \in \mathcal{S}_\tau$ because

$$V(x) \equiv \int_0^x \frac{\tau + w}{\tau + \mu}\,\mathrm{d}F(w)$$

is nondecreasing on $[0,\tau)$ and $\int_0^\tau w\,\mathrm{d}F(w) \leq \mu$ implies $0 \leq V(x) \leq 1$. For h we find with the condition in $\mathcal{F}_\tau$

$$\begin{aligned} h &\equiv 1 - \frac{2\tau}{\tau+\mu} + \int_0^\tau \frac{\tau - w}{\tau + \mu}\,\mathrm{d}F(w) \\ &= 1 - \frac{2\tau}{\tau+\mu} + \frac{1}{\tau+\mu}\left(-\tau + \tau F(\tau-) - \int_0^\tau w\,\mathrm{d}F(w) + \tau\right) \\ &\geq 1 - \frac{2\tau}{\tau+\mu} + \frac{\tau - \mu}{\tau+\mu} = 0. \end{aligned}$$

If $\int_0^\tau w\,\mathrm{d}F(w) + \tau\,(1 - F(\tau-)) < \mu$, then we have $F(\tau-) < 1$ and we may write

$$\frac{2\tau}{\tau+\mu} > \int_0^\tau \frac{2\tau}{\tau+\mu}\,\mathrm{d}F(w)$$

and this yields

$$h < 1 - \int_0^\tau \frac{2\tau}{\tau+\mu}\,\mathrm{d}F(w) + \int_0^\tau \frac{\tau - w}{\tau+\mu}\,\mathrm{d}F(w) = 1 - V(\tau-).$$

If $\int_0^\tau w\,\mathrm{d}F(w) + \tau\,(1 - F(\tau-)) = \mu$ and thus $F(\tau-) = 1$, we get that $\int_0^\tau w\,\mathrm{d}F(w) = \mu$ and so one sees immediately that $V(\tau-) = 1$ and $h = 0$. Note that we defined h in such a way that (1.19) holds. So we verified that $\varphi_\tau(\mathcal{F}_\tau) \subseteq \mathcal{V}_\tau$.

Conversely we have that if $(F(\cdot),\mu) \equiv \psi_\tau(V(\cdot),h)$ for an $(V,h) \in \mathcal{V}_\tau$, then $F \in \mathcal{S}_\tau$ because

$$F(x) \equiv \frac{1}{\nu(V(\cdot),h)} \int_0^x \frac{1}{\tau+w}\,\mathrm{d}V(w) \tag{1.30}$$

is nondecreasing on $[0,\tau)$ and

$$\begin{aligned} 0 \le F(x) &= \frac{1}{\nu(V(\cdot),h)} \int_0^x \frac{1}{\tau+w}\,\mathrm{d}V(w) \\ &\le \left(\int_0^\tau \frac{1}{\tau+w}\,\mathrm{d}V(w)\right)^{-1} \left(\int_0^x \frac{1}{\tau+w}\,\mathrm{d}V(w)\right) \le 1. \end{aligned}$$

From the definition of $\nu(V(\cdot),h)$ we get easily with the constraints in $\mathcal{V}_\tau$ that $0 < \nu(V(\cdot),h) \le (1/\tau)$ and so we have

$$0 \le \mu \equiv \frac{1}{\nu(V(\cdot),h)} - \tau < \infty.$$

(Compare the definition of μ with (1.18)). From (1.30) we get

$$\int_0^\tau (w-\tau)\,\mathrm{d}F(w) + \tau = \frac{1}{\nu(V(\cdot),h)} \int_0^\tau \frac{w-\tau}{\tau+w}\,\mathrm{d}V(w) + \tau$$

and this implies

$$\begin{aligned} &\int_0^\tau w\,\mathrm{d}F(w) + \tau(1 - F(\tau-)) \\ &= \frac{1}{\nu(V(\cdot),h)} \left(\int_0^\tau \frac{w-\tau}{\tau+w}\,\mathrm{d}V(w) + 2\tau\,\nu(V(\cdot),h)\right) - \tau. \end{aligned} \tag{1.31}$$

We have

$$\begin{aligned} &\int_0^\tau \frac{w-\tau}{\tau+w}\,\mathrm{d}V(w) + 2\tau\,\nu(V(\cdot),h) \\ &= \int_0^\tau \frac{w-\tau}{\tau+w}\,\mathrm{d}V(w) + (1 - h - V(\tau-)) + 2\tau \int_0^\tau \frac{1}{\tau+w}\,\mathrm{d}V(w) \\ &= (1-h). \end{aligned} \tag{1.32}$$

If $0 \leq h < 1 - V(\tau-) \leq 1$, then we get with (1.32) from (1.31)

$$\int_0^\tau w\,\mathrm{d}F(w) + \tau(1 - F(\tau-)) < \frac{1}{\nu(V(\cdot),h)} - \tau = \mu.$$

If $h = 1 - V(\tau-)$ and thus $h = 0$ and $V(\tau-) = 1$, then we get with (1.32) from (1.31): $\int_0^\tau w\,\mathrm{d}F(w) + \tau\,(1 - F(\tau-)) = \mu$. Furthermore with (1.30) we derive that $F(\tau-) = 1$. So we checked that $\varphi_\tau(\mathcal{V}_\tau) \subseteq \mathcal{F}_\tau$.

By straightforward calculation one shows that

$$\psi_\tau\,(\varphi_\tau(F(\cdot),\mu)) = (F(\cdot),\mu)$$

and

$$\varphi_\tau\,(\psi_\tau(V(\cdot),h)) = (V(\cdot),h).$$

This proves that there is a 1–1 correspondence between $\mathcal{F}_\tau$ and $\mathcal{V}_\tau$. □

Proposition 1.1.3.3 *If we consider the map φ_τ on $\mathcal{F}_<$ and the map ψ_τ on $\mathcal{V}_<$, then they give a 1-1 correspondence between $\mathcal{F}_<$ and $\mathcal{V}_<$.*

PROOF: (In proposition 1.1.3.2 we showed already that if $(V(\cdot),h) \equiv \varphi_\tau(F(\cdot),\mu)$ for an $(F,\mu) \in \mathcal{F}_\tau$, then $(V,h) \in \mathcal{S}_\tau \times [0,\infty)$. The same for (F,μ)). First we prove $\varphi_\tau(\mathcal{F}_<) \subseteq \mathcal{V}_<$. Let us consider a $(F,\mu) \in \mathcal{F}_<$. Because of $-\int_0^\tau w\,\mathrm{d}F(w) > -\mu$ and $F(\tau-) = 1$, we have

$$h \equiv 1 - \frac{2\tau}{\tau+\mu} + \int_0^\tau \frac{\tau - w}{\tau+\mu}\,\mathrm{d}F(w) > 1 - \frac{2\tau}{\tau+\mu} + \frac{\tau-\mu}{\tau+\mu} = 0$$

and thus $h > 0$ is satisfied. Furthermore we have

$$\begin{aligned} 1 - V(\tau-) &\equiv 1 - \int_0^\tau \frac{\tau+w}{\tau+\mu}\,\mathrm{d}F(w) \\ &= 1 - \frac{2\tau}{\tau+\mu} + \int_0^\tau \frac{\tau-w}{\tau+\mu}\,\mathrm{d}F(w) \equiv h \end{aligned}$$

and thus $1 - V(\tau-) = h$ is satisfied. We also know that if $h = 1$, then we would have

$$\frac{2\tau}{\tau+\mu} = \int_0^\tau \frac{\tau-w}{\tau+\mu}\,\mathrm{d}F(w).$$

But because we have $\int_0^\tau (\tau - w)\,\mathrm{d}F(w) \leq \tau$, this can only be true if $\mu = \infty$. This contradicts the fact that $(F,\mu) \in \mathcal{F}_<$. Thus we have $h \neq 1$. This proves $\varphi_\tau(\mathcal{F}_<) \subseteq \mathcal{V}_<$.

Secondly we prove $\psi_\tau(\mathcal{V}_<) \subseteq \mathcal{F}_<$. Let us consider a $(V,h) \in \mathcal{V}_<$. Because $h = 1 - V(\tau-)$ we have that

$$\nu(V(\cdot),h) = \int_0^\tau \frac{1}{\tau+w}\,\mathrm{d}V(w) \tag{1.33}$$

and thus

$$F(\tau-) \equiv \frac{1}{\nu(V(\cdot),h)} \int_0^\tau \frac{1}{\tau+w}\,\mathrm{d}V(w) = 1.$$

Furthermore $0 < h = 1 - V(\tau-)$ implies that $V(\tau-) < 1$ and with (1.33) we obtain

$$\begin{aligned}\int_0^\tau w\,\mathrm{d}F(w) &\equiv \frac{1}{\nu(V(\cdot),h)}\int_0^\tau \frac{w}{\tau+w}\,\mathrm{d}V(w)\\ &= \frac{1}{\nu(V(\cdot),h)}\int_0^\tau \frac{\tau+w}{\tau+w}\,\mathrm{d}V(w) - \frac{1}{\nu(V(\cdot),h)}\int_0^\tau \frac{\tau}{\tau+w}\,\mathrm{d}V(w)\\ &= \frac{1}{\nu(V(\cdot),h)}V(\tau-) - \tau\\ &< \frac{1}{\nu(V(\cdot),h)} - \tau \equiv \mu.\end{aligned}$$

We have that $F(\tau-) = 1$ and $\int_0^\tau w\,\mathrm{d}F(w) < \mu$ and this proves $\psi_\tau(\mathcal{V}_<) \subseteq \mathcal{F}_<$.

Of course we have $\psi_\tau(\varphi_\tau(F(\cdot),\mu)) = (F(\cdot),\mu)$ and $\varphi_\tau(\psi_\tau(V(\cdot),h)) = (V(\cdot),h)$ by straightforward calculation. This proves the 1–1 correspondence between $\mathcal{F}_<$ and $\mathcal{V}_<$. □

1.1.4 The likelihood and the definition of the NPMLE

We consider the hospital model of section 1.1.1. If we condition on $N = n$, the likelihood based on the probability that we have the n independent observations $(Z_i = z_i, D_i = d_i)$, $(W_j = w_j, E_j = e_j)$ $(i = 1,...,m; j = 1,...,l; n = l + m)$ is proportional to

$$\frac{1}{(\tau+\mu)^n}\prod_{i=1}^{m}(\mathrm{d}F(z_i))^{d_i}(1-F(z_i))^{1-d_i}\prod_{j=1}^{l}(1-F(w_j))^{e_j}\left(\int_\tau^\infty(1-F(u))\,\mathrm{d}u\right)^{1-e_j} \tag{1.34}$$

(use (1.10)–(1.13)). One must be aware that in the likelihood (1.34) μ in the denominator also depends on F. (Note that we have n factors $1/(\tau+\mu)$, one for each observation).

Let $x_1 < x_2 < ... < x_r$ be the ordered values of w_j and z_i for which either $e_j = 1$ or $d_i = 0, 1$ (these are the *uncensored* and *single end censored* observations) and let ϕ_i and γ_i be the number of the *uncensored* respectively *single end censored* values at x_i. Instead of using the likelihood (1.34) with the distribution function F, where we would have to deal with μ in the denominator, we regard our problem as the missing data problem described in section 1.1.2 and with (1.3) (or (1.4)–(1.7)) and (1.8) we write (1.34) in terms of V. We get

$$\begin{aligned}\mathrm{lik}(V) &\propto \prod_{i=1}^{r}(\mathrm{d}V(x_i))^{\phi_i}\left(\int_{w=x_i}^\infty \frac{1}{\tau+w}\,\mathrm{d}V(w)\right)^{\gamma_i}\cdot\left(\int_{w=\tau}^\infty \frac{w-\tau}{\tau+w}\,\mathrm{d}V(w)\right)^{n-r}\\ &= \prod_{i=1}^{r}(\mathrm{d}V(x_i))^{\phi_i}\left(\int_{x_i}^\tau \frac{1}{\tau+w}\,\mathrm{d}V(w)+g(\tau)\right)^{\gamma_i}\cdot h^{n-r}.\end{aligned} \tag{1.35}$$

Here we introduce the following empirical subdistribution functions on $[0,\tau]$:

$$F_n^{d.c.}(x) \equiv \frac{1}{n}\#\{\text{double censored observations} \le x\} \tag{1.36}$$

$$F_n^{u.c.}(x) \equiv \frac{1}{n}\#\{\text{uncensored observations} \le x\} \tag{1.37}$$

$$F_n^{s.e.c.}(x) \equiv \frac{1}{n}\#\{\text{single end censored observations} \le x\}. \tag{1.38}$$

($F^{s.e.c.}$ is defined by (1.9).) From (1.35) we write by formally taking logs the 'log likelihood-function' on $\mathcal{V}_\tau$ as

$$\Psi(\tilde{V},\tilde{h}) \equiv \int_0^\tau \log\left(\mathrm{d}\tilde{V}(x)\right)\cdot \mathrm{d}F_n^{u.c.}(x) + \int_0^\tau \log(\tilde{g}(x))\cdot \mathrm{d}F_n^{s.e.c.}(x) + \log(\tilde{h})\cdot F_n^{d.c.}(\tau), \quad (1.39)$$

where $(\tilde{V},\tilde{h}) \in \mathcal{V}_\tau$ and

$$\tilde{g}(x) \equiv \int_x^\tau \frac{1}{\tau+w}\,\mathrm{d}\tilde{V}(w) + \tilde{g}(\tau) \quad (1.40)$$

and $\tilde{g}(\tau)$ is defined by

$$\tilde{V}(\tau-) + 2\tau\tilde{g}(\tau) + \tilde{h} = 1$$

(compare with (1.14)).

Now we say that $(\hat{V}_n,\hat{h}_n)$ is a NPMLE of (V,h) in $\mathcal{V}_\tau$, if

$$\Psi(\tilde{V},\tilde{h}) - \Psi(V_0,h_0) \le \Psi(\hat{V}_n,\hat{h}_n) - \Psi(V_0,h_0)$$

for all $(\tilde{V},\tilde{h}) \in \mathcal{V}_\tau$ and $\tilde{V} \ll V_0$, $\hat{V}_n \ll V_0$ for a $(V_0,h_0) \in \mathcal{V}_\tau$ (for example $V_0 \equiv \frac{1}{2}(\tilde{V}+\hat{V}_n)$) following Kiefer and Wolfowitz(1956) (see also Scholz(1980), Gill(1989)). When subtracting log likelihoods we use the following interpretation: if G and H are measures and $G \ll H$, then we rewrite

$$\log \mathrm{d}G(x) - \log \mathrm{d}H(x) = \log\frac{\mathrm{d}G(x)}{\mathrm{d}H(x)} = \log\frac{\mathrm{d}G}{\mathrm{d}H}(x),$$

where $(\mathrm{d}G/\mathrm{d}H)(x)$ is the Radon-Nikodym derivative of G with respect to H at x.

1.1.5 Existence and uniqueness of the NPMLE

We state that there exists a nondecreasing step-function on $[0,\tau)$ with jumps at the *uncensored* and *single end censored* observations that is a NPMLE $(\hat{V}_n,\hat{h}_n)$ of (V,h). To see this we use (1.35) to note that if one puts mass between the observation points, then one can always shift this mass to the nearest observation point on the left and increase the likelihood. (Mass between 0 and the first observation point can be shifted to this observation point.) So from now we can consider discrete estimators with mass on the observation points only.

Let I_u contain all the indices of the x_i's in $x_1 < x_2 < \ldots < x_r < x_{r+1} = \tau$, which are uncensored observations and let I_s contain all the indices of the x_i's which are single end censored observations. We note that the distribution of the length of the s.e.c. observations is continuous and therefore we have $I_u \cap I_s = \emptyset$ with probability 1. In the discrete setting we denote by $\Delta\tilde{V}(x)$ the height of the jump of $\tilde{V}$ at x. If we define $v_i \equiv \Delta\tilde{V}(x_i)$, then we write the loglikelihood formally as

$$\Psi(\tilde{V},\tilde{h}) = \frac{1}{n}\sum_{i\in I_u}\phi_i\,\log(v_i) + \frac{1}{n}\sum_{i\in I_s}\gamma_i\,\log\left(\sum_{j=i}^{r}\frac{1}{\tau+x_j}v_j + \tilde{g}(\tau)\right) + \frac{n-r}{n}\cdot\log(\tilde{h}). \quad (1.41)$$

Now we define $s_j = v_j$ $(j = 1,\ldots,r)$, $s_{r+1} = \tilde{h}$ and $s_{r+2} = 2\tau\tilde{g}(\tau)$ and we write the loglikelihood formally as a function q of $\mathbf{s} = (s_1, s_2, \ldots, s_{r+2})$:

$$q(\mathbf{s}) = \sum_{i=1}^{r+1}\beta_i\,\log\left(\sum_{j=1}^{r+2}\alpha_{ij}\,s_j\right), \quad (1.42)$$

where β_i $(i = 1, ..., r+1)$ is defined as

$$\beta_i = \frac{1}{n} \cdot \phi_i^{d_i} \cdot \gamma_i^{e_i} \cdot (n-r)^{1(i=r+1)}$$

and where the $\alpha_{ij} \geq 0$ are defined as follows: for an observation i $(i = 1, ..., r+1)$ we have the following three possibilities:
1) $i \in I_u$ (there are u.c. observations x_i):

$$\alpha_{ij} = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i, \end{cases} \tag{1.43}$$

2) $i \in I_s$ (there are s.e.c. observations x_i):

$$\alpha_{ij} = \begin{cases} 0 & \text{if } j < i \\ \dfrac{1}{\tau + x_j} & \text{if } i \leq j \leq r \\ 0 & \text{if } j = r+1 \\ \dfrac{1}{2\tau} & \text{if } j = r+2 \end{cases} \tag{1.44}$$

3) $i = r+1$ (d.c. observations τ):

$$\alpha_{ij} = \begin{cases} 1 & \text{if } j = r+1 \\ 0 & \text{if } j \neq r+1. \end{cases} \tag{1.45}$$

Note that we have for all i: $\alpha_{ii} > 0$.

For a matrix M we denote by $(M)_{ij}$ the entry in M with the coordinates the i'the row and the j'th column. We use $(M)_{i.}$ and $(M)_{.j}$ to denote the i'th row respectively the j'th column and for a vector $\mathbf{a}$ we use $(\mathbf{a})_i$ (or a_i) to denote the i'th entry, thus by $(M\mathbf{a})_i$ we mean the i'th entry of the vector $M\mathbf{a}$. Now we define the $(r+2) \times (r+2)$ matrix A:

$$\begin{array}{llll} (A)_{ij} & \equiv & \alpha_{ij} & \text{for } i = 1, ..., r \text{ and } j = 1, ..., r+2 \\ (A)_{r+1,j} & \equiv & 0 & \text{for } j \neq r+1 \\ (A)_{r+1,r+1} & \equiv & 1 & \\ (A)_{r+2,j} & \equiv & 1 & \text{for } j = 1, ..., r+2. \end{array}$$

The matrix A has the following structure:

$$\begin{pmatrix} \star & \star & \star & \cdot & \cdot & \star & \star & 0 & \star \\ 0 & \star & \star & \cdot & \cdot & \star & \star & 0 & \star \\ 0 & 0 & \star & \cdot & \cdot & \star & \star & 0 & \star \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 0 & \star & 0 & \star \\ 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & \cdot & \cdot & 1 & 1 & 1 & 1 \end{pmatrix},$$

where the $\star$'s stand for some number. On the diagonal we have nonzero entries. To show that A is invertible we show that the rank is $r+2$ by using elementary row operations on A to obtain a matrix with 0's below the diagonal and nonzero elements on the diagonal. Because $(A)_{11} > 0$ we make with the first row $A_{r+2,1}$ equal to 0. Then (because $A_{22} > 0$) we make with the second row $A_{r+2,2}$ equal to 0. We repeat this until all $A_{r+2,j}$ $(j = 1, ..., r+1)$ are 0. This proves that rank A is $r+2$ and thus A^{-1} exists.

We remember that the s_i's that maximize (1.42) represent the mass of $\widehat{V}_n$ at the x_i's $(i = 1, ..., r)$ and s_{r+1} and s_{r+2} stand for the mass assigned to respectively $\widehat{h}_n$ and $2\tau\widehat{g}_n(\tau)$. Now $q(\mathbf{s})$ (see (1.42)) will be maximized over the set

$$\mathcal{S} \equiv \left\{ \mathbf{s} \in \mathbb{R}^{r+2} \,\middle|\, \sum_{j=1}^{r+2} s_j = 1, \;\; s_j \geq 0 \right\}.$$

It is obvious that q is concave on $\mathcal{S}$ (use (strict) concavity of the log to obtain $q(\lambda\,\mathbf{s}_0 + (1-\lambda)\,\mathbf{s}_1) \geq \lambda\,q(\mathbf{s}_0) + (1-\lambda)\,q(\mathbf{s}_1)$, with $\mathbf{s}_0$, $\mathbf{s}_1 \in \mathcal{S}$). It is easy to see that $q : \mathcal{S} \to \mathbb{R} \cup \{-\infty\}$ is bounded from above: $q(\mathbf{s}) \leq (r+1) \cdot \log((r+2)\max_{i,j}(\alpha_{ij}))$. Furthermore one notes that q is continuous on $\mathcal{S}$. The continuity of q on the compact set $\mathcal{S}$ guarantees the existence of a maximum in our problem.

If we define $l : \mathbb{R}^{r+2} \to \mathbb{R} \cup \{-\infty\}$ as

$$l(\mathbf{u}) = \sum_{i=1}^{r+1} \beta_i \log u_i,$$

then we note immediately $(\beta_i > 0)$ that l is strictly concave on

$$\mathcal{U}_c = \left\{ \mathbf{u} \in \mathbb{R}^{r+2} \,|\, u_i \geq 0, u_{r+2} = c \right\},$$

where $c \geq 0$ is a constant. (For each i $\log u_i$ is strictly concave on $\mathcal{U}_c$ and thus l is a sum of strictly concave functions on $\mathcal{U}_c$). Thus l is also strictly concave on

$$\mathcal{U} \equiv A(\mathcal{S}) \subset \mathcal{U}_1$$

$((A\mathbf{s})_{r+2} = \sum_{i=1}^{r+2} s_i = 1$ for all $\mathbf{s} \in \mathcal{S})$. One easily verifies that

$$q(\mathbf{s}) = l(A\mathbf{s}).$$

We know already that q has a maximum on $\mathcal{S}$. Now the fact that l is strictly concave on $\mathcal{U} = A(\mathcal{S})$ and the fact that A is invertible imply that the maximum is unique.

We know that each $\mathbf{s} \in \mathcal{S}$ corresponds with a subdistribution function $\widetilde{V}_{\mathbf{s}}$ on $[0, \tau)$, which is a step-function (with possible jumps only at the x_i's) and a $\widetilde{h}_{\mathbf{s}} \geq 0$ such that $\widetilde{h}_{\mathbf{s}} \leq 1 - \widetilde{V}_{\mathbf{s}}(\tau-)$ (and $2\tau\widetilde{g}_{\mathbf{s}}(\tau) \equiv 1 - \widetilde{h}_{\mathbf{s}} - \widetilde{V}_{\mathbf{s}}(\tau-) \geq 0$). So we have that

$$\mathcal{V}_{\mathrm{npmle}} \equiv \{(\widetilde{V}_{\mathbf{s}}, \widetilde{h}_{\mathbf{s}}) \,|\, \mathbf{s} \in \mathcal{S}\} \subset \mathcal{V}_{\tau,\star,<} = \overline{\mathcal{V}}_\tau \tag{1.46}$$

(see section 1.1.3). (Note that the set $\mathcal{V}_{\mathrm{npmle}}$ depends on the sample size). It can happen that the NPMLE $(\widehat{V}_n, \widehat{h}_n)$ which lies in $\mathcal{V}_{\mathrm{npmle}}$, lies in the boundery of $\mathcal{V}_\tau$ but not in $\mathcal{V}_\tau$ itself. For example if we have a sample of size 5 consisting of two u.c. observations $x_1, x_2 \in [0,\tau)$ and three d.c. observations τ, then we have to maximize

$$\frac{1}{5}\log\tilde{v}_1 + \frac{1}{5}\log\tilde{v}_2 + \frac{3}{5}\log\tilde{h}$$

subject to the constraints $\tilde{v}_1 + \tilde{v}_2 + \tilde{h} = 1$ and $\tilde{v}_i \geq 0$ and $\tilde{h} \geq 0$. The method of Lagrange multipliers gives us $\widehat{v}_1 = \widehat{v}_2 = 1/5$ and $\widehat{h}_n = 3/5$. Thus $0 < \widehat{h}_n = 1 - \widehat{V}_n(\tau-)$ and $\widehat{h}_n \neq 1$ and this implies that $(\widehat{V}_n, \widehat{h}_n) \in \mathcal{V}_<$. Thus the NPMLE does not have to be 'in the model'.

We have proved the following proposition:

Proposition 1.1.5.1 *The NPMLE* $(\widehat{V}_n, \widehat{h}_n)$ *exists and is unique and lies in* $\mathcal{V}_{\mathrm{npmle}} \subset \overline{\mathcal{V}}_\tau$.

Because we need some ideas from chapter 2, we will give the proof of the following proposition in section 2.2.3. The proposition says that for increasing n the probability that the NPMLE lies in the model tends to 1.

Proposition 1.1.5.2 *Suppose that we have* $g(\tau) > 0$ *for the underlying* g*, then for increasing* n *the probability that the NPMLE* $(\widehat{V}_n, \widehat{h}_n) \in \mathcal{V}_\tau$ *tends to 1.*

1.1.6 EM-algorithm and self-consistency equations

In this section we will derive the so-called self-consistency equations for the NPMLE $(\widehat{V}_n, \widehat{h}_n)$ of (V, h) and say something about the EM-algorithm for computing the NPMLE. Although most of the machinery in the next paragraph is introduced in chapter 3 ((differentiable) one-dimensional submodel through V, score operator, score equation), we illustrate here all the same how the equations (1.50)–(1.52) can be derived. If one wants to ignore this derivation at this time, then one skips the next paragraph.

Just for the moment we parametrize again with $V \in \mathcal{F}_\infty$ instead of $(V, h) \in \mathcal{V}_\tau$ and thus we parametrize the distribution of the data with P_V, $V \in \mathcal{F}_\infty$ instead of $P_{(V,h)}$ with $(V, h) \in \mathcal{V}_\tau$. We denote by P_n the empirical distribution of the data. Let us consider the following class of one-dimensional submodels through V:

$$V_{\theta,l}(x) = \frac{\int_0^x (1 + \theta\, l(u))\, \mathrm{d}V(u)}{\int_0^\infty (1 + \theta\, l(u))\, \mathrm{d}V(u)},$$

with bounded score l and where θ is sufficiently small. Now $\widehat{V}_{n\,\theta,l}$ is a dominated family of one-dimensional submodels through $\widehat{V}_n$ and because $\widehat{V}_n$ is the NPMLE we have that the log likelihood is maximal at $\theta = 0$. By differentiating the log likelihood along $P_{\widehat{V}_{n\,\theta,l}}$ (one-dimensional submodels through $P_{\widehat{V}_n}$ implied by $\widehat{V}_{n\,\theta,l}$ and with score $A_{\widehat{V}_n}(l)$, where $A_{\widehat{V}_n}$ is the score operator (see section 3.3)) and evaluating at $\theta = 0$ we obtain the score equation

$$E_{P_n}(A_{\widehat{V}_n}(l)) = 0, \tag{1.47}$$

for all bounded l (which are a score of some differentiable submodel $P_{\widehat{V}_{n\,\theta,l}}$ through $P_{\widehat{V}_n}$). For any missing data model the score operator is given by $A_{\widehat{V}_n}(l)(Y) = E_{\widehat{V}_n}(l(X) \mid Y)$, where X is the variable of interest and Y the data. The score equation becomes

$$E_{P_n}(E_{\widehat{V}_n}(l(X) \mid Y)) = 0 \tag{1.48}$$

(for heuristical argument see Gill(1989); for rigorous proof see Bickel et al.(1993) prop. A5.5).

Especially (1.48) holds for $l(X) = 1_{\{X \le x\}}(x) - \widehat{V}_n(x)$. This provides us with the so-called self consistency equation:

$$\begin{aligned} \widehat{V}_n(x) &= \int P_{\widehat{V}_n}(X \le x \mid Y)\,\mathrm{d}P_n(Y) \\ &= \frac{1}{n}\sum_{i=1}^{n} P_{\widehat{V}_n}(X_i \le x \mid \text{observation } i). \end{aligned} \tag{1.49}$$

Together with the definitions of $F_n^{d.c.}$, $F_n^{s.e.c.}$, $F_n^{u.c.}$ in (1.36)–(1.38) and $\widehat{g}_n$ we get from (1.49) that the NPMLE $(\widehat{V}_n, \widehat{h}_n)$ of (V, h) satisfies the self-consistency equations

$$\begin{aligned} \mathrm{d}\widehat{V}_n(x) &= \mathrm{d}F_n^{u.c.}(x) + \int_{v=0}^{v=x} \frac{1}{\widehat{g}_n(v)}\,\mathrm{d}F_n^{s.e.c.}(v)\cdot\frac{1}{\tau+x}\,\mathrm{d}\widehat{V}_n(x) \\ &= \mathrm{d}F_n^{u.c.}(x) \\ &\quad + \int_{v=0}^{v=x}\left(\int_{w=v}^{w=\tau}\frac{1}{\tau+w}\,\mathrm{d}\widehat{V}_n(w) + \widehat{g}_n(\tau)\right)^{-1}\mathrm{d}F_n^{s.e.c.}(v)\cdot\frac{1}{\tau+x}\,\mathrm{d}\widehat{V}_n(x) \end{aligned} \tag{1.50}$$

and

$$\widehat{h}_n = F_n^{d.c.}(\tau) = \frac{n-r}{n} \tag{1.51}$$

$$2\tau\widehat{g}_n(\tau) = 1 - \widehat{h}_n - \widehat{V}_n(\tau). \tag{1.52}$$

A solution of (1.49) and thus of (1.50)–(1.52) can be computed with the EM-algorithm. We already mentioned that we could consider discrete estimators with mass on the observation points only. In this discrete setting we mean by $\mathrm{d}\widehat{V}_n(x_i)$ the height of the jump of $\widehat{V}_n$ at x_i. Thus $\mathrm{d}\widehat{V}_n(x) = 0$ if $x \notin \{x_1, ..., x_r\}$. (Here we use this 'lazy' notation, instead of rewriting (1.50)–(1.52) in the discrete notation $\Delta\widehat{V}_n(x_i)$). If we replace in the equations (1.50)–(1.52) at the left-hand side $\widehat{V}_n$, $\widehat{h}_n$ and $\widehat{g}_n$ by $\widehat{V}_n^{\,k+1}$, $\widehat{h}_n^{\,k+1}$ and $\widehat{g}_n^{\,k+1}$ respectively and if we replace at the right-hand side $\widehat{V}_n$, $\widehat{h}_n$ and $\widehat{g}_n$ by $\widehat{V}_n^{\,k}$, $\widehat{h}_n^{\,k}$ and $\widehat{g}_n^{\,k}$ respectively, then we obtain the iterative scheme of the EM-algorithm.

We start with an initial (discrete) estimator $(\widehat{V}_n^{\,0}, \widehat{h}_n^{\,0})$ which puts positive mass at all the observation points, $\widehat{h}_n^{\,0}$ and $\widehat{g}_n(\tau)^0$. Now we evaluate for $k = 0$ the expressions at the right-hand side of (1.50)–(1.52) with $\widehat{V}_n^{\,0}$, $\widehat{h}_n^{\,0}$ and $\widehat{g}_n(\tau)^0$. This is the 'E'-step in the algorithm (where the 'E' stands for 'Expectation'; see right-hand side (1.49)). Defining $\mathrm{d}\widehat{V}_n^{\,1}(x)$, $\widehat{h}_n^{\,1}$ and $\widehat{g}_n(\tau)^1$ by (1.50)–(1.52) ($\mathrm{d}\widehat{V}_n^{\,0}(x)$, $\widehat{h}_n^{\,0}$ and $\widehat{g}_n(\tau)^0$ at the right-hand side) provides us with a new distribution function, which increases the likelihood. This is the 'M'-step in the algorithm (where the 'M' stands for 'Maximization'). Wu(1983) (see also Dempster et al.(1977) and Turnbull(1976)) shows that the likelihood increases after each iteration and converges to the maximum, the NPMLE $(\widehat{V}_n, \widehat{h}_n)$, which is in our case unique.

We make here the remark that if one does not start with an initial estimator that puts positive mass at *all* the observation points, then the EM-algorithm will converge to a solution of the self-consistency equations, but not necessarily to the NPMLE. At s.e.c. points where the initial distribution gives mass zero, all iterated distributions will give mass zero (see EM-algorithm). Thus if the NPMLE gives positive mass to a point to which the initial distribution gives mass zero, then the algorithm does not converge to the NPMLE (see also Groeneboom and Wellner(1992)).

1.1.7 The sieved NPMLE

In the previous section we already mentioned the fact that if we start in the EM-algorithm with an initial distribution that puts mass zero for instance at the s.e.c. observation points, then the algorithm will not necessarily converge to the NPMLE, which is a step-function with (possible) jumps at all observation points. But the algorithm will converge to a solution of the self-consistency equations. Actually, in the class of discrete distributions with (possible) mass at the u.c. observation points only, this solution will maximize the log likelihood. By this fact we are able to introduce a new kind of NPMLE-definition: the *sieved* NPMLE. The estimator sieves its observation points: only the u.c. observation points get mass. Of course the sieved NPMLE depends on the sieve you choose. Any subset of the observation points can be a sieve.

For the NPMLE $(\widehat{V}_n, \widehat{h}_n)$ of $(V, h) \in \mathcal{V}_r$ we know that we only have to consider discrete estimators with mass at the observation points only. Of course this is the set $\mathcal{V}_{\text{npmle}}$ defined in (1.46). This set of distributions will be dominated by $(\pi_{r+2},\ 1/(r+2))$, where π_{r+2} gives mass $1/(r+2)$ to the r observation points (in $[0,\tau)$) and the same amount of mass is given to $h = s_{r+1}$ and $2\tau g(\tau) = s_{r+2}$. Therefore if Π_{r+2} denotes the distribution function of π_{r+2} on $[0,\tau)$, we have that the definition of the NPMLE $(\widehat{V}_n, \widehat{h}_n)$ in section 1.1.4 is equivalent to the following definition:

$$(\widehat{V}_n, \widehat{h}_n) \equiv \arg \max_{(\widetilde{V}_{\mathbf{s}}, \widetilde{h}_{\mathbf{s}}) \in \mathcal{V}_{\text{npmle}}} \Big(\Psi(\widetilde{V}_{\mathbf{s}}, \widetilde{h}_{\mathbf{s}}) - \Psi(\Pi_{r+2},\ 1/(r+2)) \Big) \tag{1.53}$$

(provided the maximum exists).

Here the *sieved* NPMLE is defined as the maximizer of the log likelihood Ψ for distributions, which only give mass to the u.c. observation points (and h and $2\tau g(\tau)$). Now we define the set $\mathcal{V}_{\text{sieve}}$ analogue to the set $\mathcal{V}_{\text{npmle}}$ in (1.46). Remember the definition of the set I_s in section 1.1.5: containing all the indices of the x_i's in $x_1 < \ldots < x_r$ which are s.e.c. observation points. Now the set $\mathcal{V}_{\text{sieve}}$ is defined as

$$\mathcal{V}_{\text{sieve}} \equiv \{(\widetilde{V}_{\mathbf{s}}, \widetilde{h}_{\mathbf{s}}) \,|\, \mathbf{s} \in \mathcal{S},\ \forall i \in I_s\, s_i = 0\},$$

where $\widetilde{V}_{\mathbf{s}}$ and $\widetilde{h}_{\mathbf{s}}$ are defined as in (1.46). (Note that $\mathcal{V}_{\text{sieve}}$ depends on the sample size). Let r_0 be equal to $\#I_s$. Then this set is dominated by $(\xi_{r-r_0+2},\ 1/(r-r_0+2))$, where ξ_{r-r_0+2} gives mass $1/(r-r_0+2)$ to the $r-r_0$ u.c. observation points (in $[0,\tau)$) and the same amount of mass is given to $h = s_{r+1}$ and $2\tau g(\tau) = s_{r+2}$. If Ξ_{r-r_0+2} denotes the distribution function of ξ_{r-r_0+2} on $[0,\tau)$, we define the *sieved* NPMLE $(\widehat{V}_n \widehat{h}_n)$ as:

$$(\widehat{V}_n, \widehat{h}_n) \equiv \arg \max_{(\widetilde{V}_{\mathbf{s}}, \widetilde{h}_{\mathbf{s}}) \in \mathcal{V}_{\text{sieve}}} \Big(\Psi(\widetilde{V}_{\mathbf{s}}, \widetilde{h}_{\mathbf{s}}) - \Psi(\Xi_{r-r_0+2},\ 1/(r-r_0+2)) \Big) \tag{1.54}$$

(provided the maximum exists).

Similarly to what we did for the NPMLE, we can show that the sieved NPMLE exists and is unique in $\mathcal{V}_{\text{sieve}}$ and satisfies the self-consistency equations too. We state the next proposition:

Proposition 1.1.7.1 *The sieved NPMLE $(\widehat{V}_n, \widehat{h}_n)$ exists and is unique and lies in $\mathcal{V}_{\text{sieve}} \subset \overline{\mathcal{V}}_\tau$.*

The following statement is proved in section 2.2.3, just as proposition 1.1.5.2.

Proposition 1.1.7.2 *Suppose that we have $g(\tau) > 0$ for the underlying g, then for increasing n the probability that the sieved NPMLE $(\widehat{V}_n, \widehat{h}_n) \in \mathcal{V}_\tau$ tends to 1.*

We note that the propositions 1.1.5.2 and 1.1.7.2 do not say anything about the probability that the limit version of the (sieved) NPMLE lies in the model.

There are several reasons to consider a sieved NPMLE $(\widehat{V}_n, \widehat{h}_n)$ of (V, h). Firstly, we will see that consistency results, which are formulated in chapter 2, hold for the sieved NPMLE in all three cases distinguished there of the underlying (V, h). For the NPMLE we can prove these results just for two of the three cases.

Secondly, we will see in the sections 1.2.4 and 1.2.6 that in the two-dimensional line segment problem we are not able to show that the NPMLE lies in some discrete class. For the discrete setting we know how to use the self-consistency equations as an iterative scheme of the EM-algorithm to compute the NPMLE. Thus it might be more convenient to consider the sieved NPMLE, for which we know that it lies in a discrete class because we force it to be there.

Thirdly, instead of using the EM-algorithm to compute the sieved NPMLE $(\widehat{V}_n, \widehat{h}_n)$, we can rewrite the self-consistency equations in case of the sieved NPMLE into an iterative scheme, which works faster than the EM-algorithm. We write (1.50) as

$$\begin{aligned} \mathrm{d}\widehat{V}_n(x) &= \mathrm{d}F_n^{u.c.}(x) \\ &\quad + \int_{v=0}^{v=x} \left(\widehat{g}_n(0) - \int_{w=0}^{w=v-} \frac{1}{\tau + w}\,\mathrm{d}\widehat{V}_n(w)\right)^{-1} \mathrm{d}F_n^{s.e.c.}(v) \cdot \frac{1}{\tau + x}\,\mathrm{d}\widehat{V}_n(x), \end{aligned}$$

which gives us (because $\widehat{V}_n \ll F_n^{u.c.}$)

$$\widehat{V}_n(x) = \left(1 - \frac{1}{\tau + x} \int_{v=0}^{v=x} \left(\widehat{g}_n(0) - \int_{w=0}^{w=v-} \frac{1}{\tau + w}\,\mathrm{d}\widehat{V}_n(w)\right)_+^{-1} \mathrm{d}F_n^{s.e.c.}(v)\right)_+^{-1} \mathrm{d}F_n^{u.c.}(x). \tag{1.55}$$

Because of (1.51) we have that $1 - \widehat{h}_n = F_n^{u.c.}(\tau) + F_n^{s.e.c.}(\tau)$. This we use to write equation (1.19) as

$$\widehat{g}_n(0) = \frac{1}{2\tau}\left(F_n^{u.c.}(\tau) + F_n^{s.e.c.}(\tau) + \int_0^\tau \frac{\tau - x}{\tau + x}\,\mathrm{d}\widehat{V}_n(x)\right). \tag{1.56}$$

Now (1.55) and (1.56) imply the following iteration scheme:

$$\mathrm{d}K(x)^{\text{new}} = \left(1 - \frac{1}{\tau + x} \int_{v=0}^{v=x} \left(b^{\text{old}} - \int_{w=0}^{w=v-} \frac{1}{\tau + w}\,\mathrm{d}K(w)^{\text{new}}\right)_+^{-1} \mathrm{d}F_n^{s.e.c.}(v)\right)_+^{-1} \mathrm{d}F_n^{u.c.}(x) \tag{1.57}$$

and

$$b^{\text{new}} = R \equiv \frac{1}{2\tau}\left(F_n^{u.c.}(\tau) + F_n^{s.e.c.}(\tau) + \int_0^\tau \frac{\tau - x}{\tau + x}\,\mathrm{d}K(x)^{\text{old}}\right). \tag{1.58}$$

Note that $\widehat{V}_n$ (with $\widehat{g}_n(0)$) is the solution that satisfies these equations. The idea is to take a value for b and compute successively with (1.57) the $\mathrm{d}K(x_i)$'s at the uncensored observation points. For the computation of $\mathrm{d}K(x_i)$ we only need the $\mathrm{d}K(x_j)$'s with $j < i$. Now the equation (1.58) is used to check if the desirable accuracy ϵ is achieved. Let $\widehat{b}$ be the value to be reached. Because b^{new} is a decreasing function of b^{old} we have that $b^{\text{old}} \geq \widehat{b}$ implies $b^{\text{new}} \leq \widehat{b} \leq b^{\text{old}}$ and $b^{\text{old}} \leq \widehat{b}$ implies $b^{\text{new}} \geq \widehat{b} \geq b^{\text{old}}$. Thus $b^{\text{new}} \geq b^{\text{old}}$ implies $b^{\text{old}} \leq \widehat{b}$ and $b^{\text{new}} \leq b^{\text{old}}$ implies $b^{\text{old}} \geq \widehat{b}$. So we can see on which side of b^{old} we choose our new b. (One can use binary-search). Note that if b gets bigger at the right-hand side of (1.57), then $\mathrm{d}K(x)$ at the left-hand side gets smaller and if $\mathrm{d}K(w)$ $(w \leq x)$at the right-hand side of (1.57) gets smaller, then $\mathrm{d}K(x)$ at the left-hand side gets smaller. Furthermore if b is very big, then the $\mathrm{d}K(x_i)$'s are very small and vice versa: a solution will be reached.

1.2 The two-dimensional line segment process

In this section we introduce a two-dimensional line segment process observed through a convex window $\mathbf{W}$. A lot of the knowledge about the one-dimensional case and the techniques we used there, will be useful to get a clear insight in the two-dimensional problem.

We start in section 1.2.1 with the introduction of the model. To make it easier for understanding and because of the less complicated computations, we work out the case that $\mathbf{W}$ is a circle in the sections 1.2.2 – 1.2.4. In the sections 1.2.5 and 1.2.6 we get back to the general case: $\mathbf{W}$ is an arbitrary convex window.

1.2.1 The model

Consider the following stochastic process: to each point $\vec{T} = (T_1, T_2)$ in a homogeneous Poisson point process on $\mathbb{R} \times \mathbb{R}$ with intensity λ is assigned a line segment of length X and a direction Θ, where Θ is measured anti-clockwise relative to West–East. The $X_1, X_2, \ldots$ are i.i.d., positive and have the common distribution function $F(x)$. We define $\mu = \int_0^\infty x\,\mathrm{d}F(x)$. The $\Theta_1, \Theta_2, \ldots$ are i.i.d. and have the common distribution function $K(\theta)$ $(\theta \in [0, \pi))$. X and Θ are independent of each other, of other pairs (X^*, Θ^*) and the underlying point process. For each i we have a point $(\vec{T}_i, X_i, \Theta_i) \in \mathbb{R}^2 \times \mathbb{R}^+ \times [0, \pi)$. All this defines a point process $\mathbf{\Phi}$ on $\mathbb{R}^2 \times \mathbb{R}^+ \times [0, \pi)$ and one can show that $\mathbf{\Phi}$ can be characterized as an inhomogeneous Poisson point process on $\mathbb{R}^2 \times \mathbb{R}^+ \times [0, \pi)$ with intensity measure

$$\varrho(\mathrm{d}\vec{t}, \mathrm{d}x, \mathrm{d}\theta) = \lambda\, \mathrm{d}\vec{t}\, \mathrm{d}F(x)\, \mathrm{d}K(\theta) \tag{1.59}$$

(see Karlin(1981) p.p. 436–438, Stoyan(1987)).

Let $\mathbf{W}$ be a convex window in $\mathbb{R}^2$. We only observe those portions of the line segments intersecting $\mathbf{W}$. Again we get the same kind of observations: *uncensored* (u.c.), *single end censored* (s.e.c.) and *double censored* (d.c.) observations. The observations are the possibly censored line segments of the two-dimensional line segment process observed through $\mathbf{W}$. In Figure 1.5 we see a typical realization.

Without loss of generality, we suppose that the southern end of the segments belong to the underlying Poisson point process of the $\vec{T}_i$'s. Now consider the line segments pointing

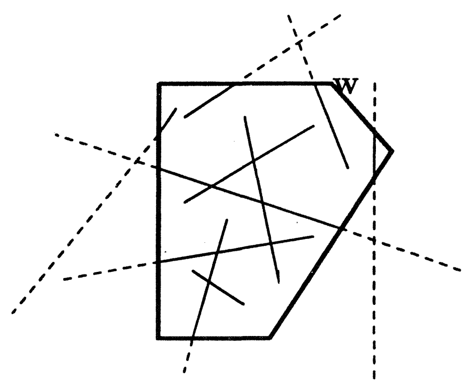

Figure 1.5: line segment process observed through **W**.

in the direction θ only in a slice of **W** parallel to θ of width dr and length τ. Considering this slice of **W** we actually have the one-dimensional case of section 1.1 and can express the observations in terms of $\vec{T}_i$, X_i and Θ_i in the same way as there.

Just as in the one-dimensional case we define the set A as the set in $\mathbb{R}^2 \times \mathbb{R}^+ \times [0,\pi)$ such that if a point of Φ say $(\vec{T}, X, \Theta) = (\vec{t}, x, \theta)$ is in A, then the corresponding line segment intersects the window **W** (in other words: is (at least partly) observed in **W**). So all the points of Φ that belong to a line segment that is observed through the window **W**, are in the set A. For each θ we define the set A_θ as the set of points of A with direction θ. Furthermore we define the set $A_{\theta,i}$ $(i = 1,2,3)$ as the subset of A_θ such that if a point of Φ say $(\vec{T}, X, \Theta) = (\vec{t}, x, \theta)$ is in $A_{\theta,i}$ then the observation is *uncensored* $(i = 1)$, *single end censored* $(i = 2)$ or *double censored* $(i = 3)$. Finally we define the set $\mathcal{A}_{\theta,x}$ as the set of points of A with a direction θ and a length x. Note that we have

$$A = \bigcup_{\theta\in[0,\pi)} A_\theta = \bigcup_{\theta\in[0,\pi)} \bigcup_{i=1,2,3} A_{\theta,i} = \bigcup_{\theta\in[0,\pi)} \bigcup_{x\in\mathbb{R}^+} \mathcal{A}_{\theta,x}.$$

In Figure 1.6 we draw one of the sets A_θ. (Window **W** as in Figure 1.5). Note that for another θ one gets a different picture. As in the one-dimensional case, for each $(\vec{T}, X, \Theta) \in A$ we can construct its observable part of the line segment geometrically (see section 1.1.1 and Figure 1.3).

In Figure 1.7 we draw a set $\mathcal{A}_{\theta,x}$. Actually, we draw the set A at level x under the angle θ. Let $z(\theta)$ be the greatest distance in **W** in the θ direction. In Figure 1.7 (a) we have a representation of set $\mathcal{A}_{\theta,x}$ if $x > z$ and in (b) if $x \leq z$. In Figure 1.7 (c) and (d) we have this for another value of θ. We see, because of the shape of the window **W**, that the shapes of A_θ and $\mathcal{A}_{\theta,x}$ heavily depend on θ. The numbers 0, 1 and 2 in Figure 1.7 refer to respectively the u.c., s.e.c. and d.c. observations. For instance if $(\vec{T}, X, \Theta) = (\vec{t}, x, \theta)$ falls in an area with number 1, then the observation belonging to this $(\vec{T}, X, \Theta)$ is s.e.c.

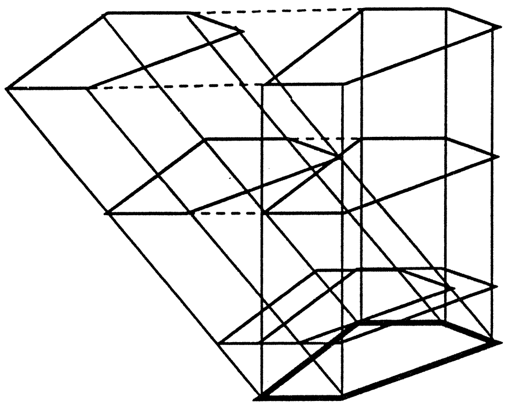

Figure 1.6: representation of a set A_θ.

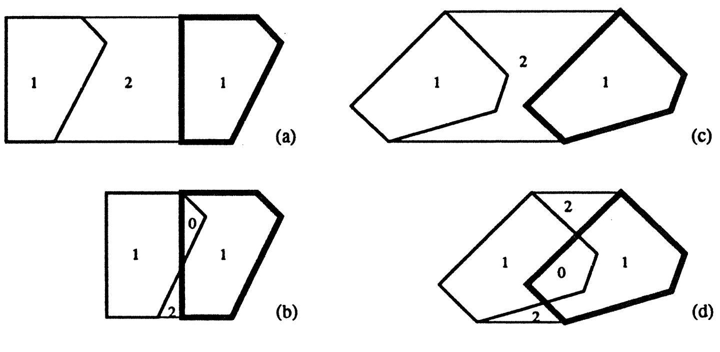

Figure 1.7: representation of $\mathcal{A}_{\theta,x}$ under different angles θ.

Let N be the number of points of the Poisson point process Φ that fall in the set A. Conditioning on $N = n$, the total number of observed line segments through window **W**, we have n independent (possibly partially observed) observations in A. N has a Poisson distribution with parameter

$$\lambda\left(|W| + \mu\, E_K \mathrm{diam}(W)\right) = \int_A \lambda\, \mathrm{d}\vec{t}\, \mathrm{d}F(x)\, \mathrm{d}K(\theta),$$

where $|W|$ is the area of **W** and $\mathrm{diam}(W, \theta)$ is the diameter of the window as seen in the θ direction and $E_K \mathrm{diam}(W)$ is the average diameter (with respect to the distribution K).

The integral is calculated as follows. For each $\Theta = \theta$ and $X = x$ we have to integrate $\lambda\, \mathrm{d}\vec{t}\, \mathrm{d}F(x)\, \mathrm{d}K(\theta)$ over the set $\mathcal{A}_{\theta,x}$. We fix an origin O and consider infinite straight lines which cross the window **W**, parametrized by the distance of the line to the origin r together with the orientation of the line θ. Let $\tau(r, \theta)$ denote the length of the intersection of the line under angle θ and at distance r with the window **W**. r varies in $[r_1(\theta), r_2(\theta)]$ (see Figure 1.8). Now if we integrate $\lambda\, \mathrm{d}\vec{t}\, \mathrm{d}F(x)\, \mathrm{d}K(\theta)$ over set $\mathcal{A}_{\theta,x}$ we get

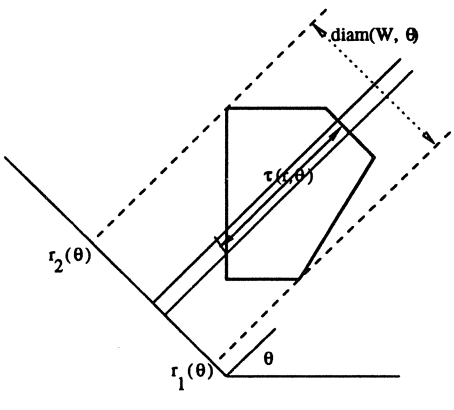

Figure 1.8: parametrization of **W** under angle θ.

$$\begin{aligned}\int_{\mathcal{A}_{\theta,x}} \lambda\, \mathrm{d}\vec{t}\, \mathrm{d}F(x)\, \mathrm{d}K(\theta) &= \lambda \int_{r_1(\theta)}^{r_2(\theta)} (\tau(r,\theta) + x)\, \mathrm{d}r\, \mathrm{d}F(x)\, \mathrm{d}K(\theta) \\ &= \lambda\left(|W| + x \mathrm{diam}(W, \theta)\right) \mathrm{d}F(x)\, \mathrm{d}K(\theta).\end{aligned}$$

If we integrate this over all values of x and θ we obtain $\lambda\left(|W| + \mu\, E_K \mathrm{diam}(W)\right)$.

Therefore if we condition on $N = n$, then the set of points $\Phi \cap A$ is distributed as the set of points in an i.i.d. sample of size n with probability measure on $\mathbb{R}^2 \times \mathbb{R}^+ \times [0, \pi)$ (see (1.59))

$$\begin{aligned}&1_A(\vec{t}, x, \theta) \frac{\lambda\, \mathrm{d}\vec{t}\, \mathrm{d}F(x)\, \mathrm{d}K(\theta)}{\lambda\left(|W| + \mu\, E_K \mathrm{diam}(W)\right)} \\ &\quad = \mathrm{d}V(x) \cdot \mathrm{d}J(\theta | X = x) \cdot \mathrm{d}A(\vec{t} | X = x, \Theta = \theta), \qquad (1.60)\end{aligned}$$

where

$$\begin{aligned} \mathrm{d}V(x) &\equiv \frac{|W| + xE_K\mathrm{diam}(W)}{|W| + \mu\, E_K\mathrm{diam}(W)}\,\mathrm{d}F(x) && (1.61)\\ \mathrm{d}J(\theta|X=x) &\equiv \frac{|W| + x\mathrm{diam}(W,\theta)}{|W| + xE_K\mathrm{diam}(W)}\,\mathrm{d}K(\theta) && (1.62)\\ \mathrm{d}A(\vec{t}|X=x,\Theta=\theta) &\equiv \frac{1}{|W| + x\mathrm{diam}(W,\theta)}\,\mathrm{d}\vec{t}\cdot 1_{\mathcal{A}_{\theta,x}}(\vec{t}). && (1.63)\end{aligned}$$

We see that we obtain the same probability measure if we consider a random sample $(\vec{T}_i, X_i, \Theta_i)$ of size n on $\mathbb{R}^2 \times \mathbb{R}^+ \times [0,\pi)$, where the X_i's are i.i.d. having the common distribution function V and the angles Θ_i given $X_i = x_i$ are drawn from the distribution $J(\cdot|X_i = x_i)$ and the $\vec{T}_i$'s given $\Theta_i = \theta_i$, $X_i = x_i$ are uniformly distributed over $\mathcal{A}_{\theta_i,x_i}$. Constructing the observable parts of the line segments as above, we get the same kind of observations. Again, as in the one-dimensional case, we have described our model as a missing data model.

1.2.2 The window W is a circle

Let the window **W** be a circle with radius R. Choosing the window **W** to be a circle, a lot of calculations will be less complicated and therefore for better understanding of the model we work out the 'circle-case' first. We have $\mathrm{diam}(W,\theta) = E_K\mathrm{diam}(W) = 2R$. The probability measure (1.60) on $\mathbb{R}^2 \times \mathbb{R}^+ \times [0,\pi)$ can be written as

$$1_A(\vec{t},x,\theta)\,\frac{\lambda\,\mathrm{d}\vec{t}\,\mathrm{d}F(x)\,\mathrm{d}K(\theta)}{\lambda\,(|W| + \mu\,2R)} = \mathrm{d}V(x)\cdot\mathrm{d}K(\theta)\cdot\frac{1}{|W| + x\,2R}\,\mathrm{d}\vec{t}\cdot 1_{\mathcal{A}_{\theta,x}}(\vec{t}), \qquad (1.64)$$

where

$$\mathrm{d}V(x) = \frac{|W| + x\,2R}{|W| + \mu\,2R}\,\mathrm{d}F(x). \qquad (1.65)$$

One notes that the distribution of $\vec{T}$ given $\Theta = \theta$ and $X = x$ has the same factor $1/(|W|+x\,2R)$ for all θ. Furthermore by (1.65) V does not depend on K (compare this with (1.61) in the general case) and we see that $\mathrm{d}J(\theta|X = x)$ in (1.62) becomes $\mathrm{d}K(\theta)$ meaning that the observed angle is distributed according to K and independent of the length X and the position $\vec{T}$. Because we are not interested in the distribution function K, this implies that it is quite irrelevant what K is. We can take w.l.o.g. for K any distribution function, for instance a degenerate distribution; all line segments have the same angle.

Let us calculate the subdistribution functions of the data. In Figure 1.9 we draw set A_θ. In the 'circle-case' we get for each θ the same picture-shape. (Compare with Figure 1.6).

In Figure 1.10 we draw set $\mathcal{A}_{\theta,x}$. Again, for each θ we have the same picture-shape. Figure 1.10 (a) is a representation of $\mathcal{A}_{\theta,x}$ if $x > 2R$ and (b) if $x \le 2R$. The length x is given in the picture. One sees (for $X = x$ and $\Theta = \theta$) in what areas $\vec{T}$ lies to obtain from $(\vec{T}, X, \Theta)$ an u.c., s.e.c. or d.c. observation (numbers in the areas are respectively 0,1 and 2).

In Figure 1.11 ((a) if $x > 2R$ and (b) if $x \le 2R$) we draw set $\mathcal{A}_{\theta,x}$ again. Now one sees (for $X = x$ and $\Theta = \theta$) in what areas $\vec{T}$ lies to obtain from $(\vec{T}, X, \Theta)$ an s.e.c. or d.c. observation

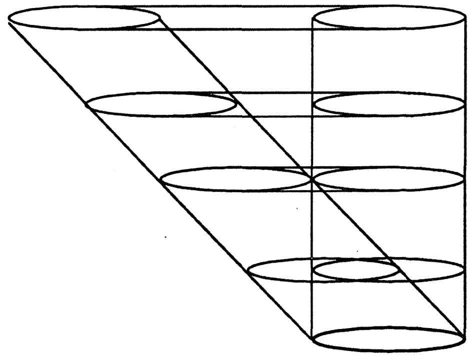

Figure 1.9: representation of A_θ (**W** being a circle).

$\widetilde{X} \le u$ or a u.c. observation $\widetilde{X} = X = x$ (respectively the dashed area, the dotted area and the lined area). x and u are given in the picture.

We define the area-integral $O(\cdot)$ as

$$O(x) \equiv \int_{v=\sqrt{R^2-\frac{1}{4}x^2}}^{v=R} \sqrt{R^2 - v^2}\,\mathrm{d}v.$$

In Figure 1.12 the area $O(x)$ is drawn (x is the length of the line segment from a to b).

If $X = x$ (and $\Theta = \theta$), the fraction of the area of $\mathcal{A}_{\theta,x}$ for which $(\vec{T}, X, \Theta)$ gives a s.e.c. observation $\widetilde{X} \le u$ $(0 \le u \le 2R)$ is given by

$$s_1(x,u) + s_2(x,u), \tag{1.66}$$

where

$$\begin{aligned} s_1(x,u) &\equiv 2\frac{4\,O(u) + 2u\sqrt{R^2 - \frac{1}{4}u^2}}{|W| + x\,2R} 1_{\{x>u\}}(u) \\ s_2(x,u) &\equiv 2\frac{4\,O(x) + 2x\sqrt{R^2 - \frac{1}{4}x^2}}{|W| + x\,2R} 1_{\{x\le u\}}(u). \end{aligned}$$

If $X = x$ (and $\Theta = \theta$), the fraction of the area of $\mathcal{A}_{\theta,x}$ for which $(\vec{T}, X, \Theta)$ gives a d.c. observation $\widetilde{X} \le u$ $(0 \le u \le 2R)$ is given by

$$s_3(x,u) + s_4(x,u), \tag{1.67}$$

where

$$\begin{aligned} s_3(x,u) &\equiv 2\frac{x(R - \sqrt{R^2 - \frac{1}{4}u^2}) - 2\,O(u)}{|W| + x\,2R} 1_{\{x>u\}}(u) \\ s_4(x,u) &\equiv 2\frac{x((R - \sqrt{R^2 - \frac{1}{4}x^2}) - 2\,O(x)}{|W| + x\,2R} 1_{\{x\le u\}}(u). \end{aligned}$$

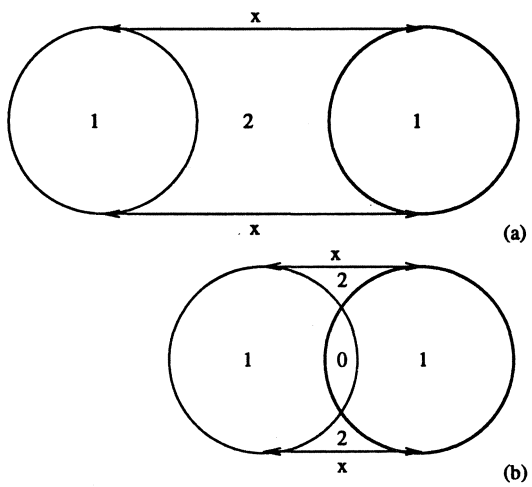

Figure 1.10: u.c., s.e.c. and d.c. 'areas' in set $\mathcal{A}_{\theta,x}$.

Finally, if $X = x$ (and $\Theta = \theta$), the fraction of the area of $\mathcal{A}_{\theta,x}$ for which $(\vec{T}, X, \Theta)$ gives an u.c. observation $\widetilde{X} \leq u$ $(0 \leq u \leq 2R)$ is given by

$$s_5(x,u) \equiv \frac{|W| - 4\,O(x) - 2x\sqrt{R^2 - \frac{1}{4}x^2}}{|W| + x\,2R}\, 1_{\{x \leq u\}}(u). \tag{1.68}$$

Note that it is obvious that the functions $s_i(\star, \cdot)$ do not depend on θ. We denote by $F^{u.c.}$, $F^{s.e.c.}$ and $F^{d.c.}$ the (conditional on $N = n$) subdistribution functions of respectively the u.c. , s.e.c. and d.c. observations $\widetilde{X}$. With (1.66), (1.67) and (1.68) we find $(0 \leq u \leq 2R)$

$$\begin{aligned}
F^{s.e.c.}(u) &= \int_{x=0}^{x=\infty} \int_{\theta=0}^{\theta=\pi} (s_1(x,u) + s_2(x,u))\,\mathrm{d}K(\theta)\,\mathrm{d}V(x) \\
&= \int_{x=u}^{x=\infty} s_1(x,u)\,\mathrm{d}V(x) + \int_{x=0}^{x=u} s_2(x,u)\,\mathrm{d}V(x) \\
F^{d.c.}(u) &= \int_{x=u}^{x=\infty} s_3(x,u)\,\mathrm{d}V(x) + \int_{x=0}^{x=u} s_4(x,u)\,\mathrm{d}V(x) \\
F^{u.c.}(u) &= \int_{x=0}^{x=u} s_5(x,u)\,\mathrm{d}V(x).
\end{aligned}$$

Computing the above integrals, provides us with the following expressions for the subdistribution functions of the data $\widetilde{X}$ $(0 \leq u \leq 2R)$

$$\mathrm{d}F^{s.e.c.}(u) = 4\sqrt{R^2 - \frac{1}{4}u^2} \int_{x=u}^{x=\infty} \frac{1}{|W| + 2xR}\,\mathrm{d}V(x)\,\mathrm{d}u$$

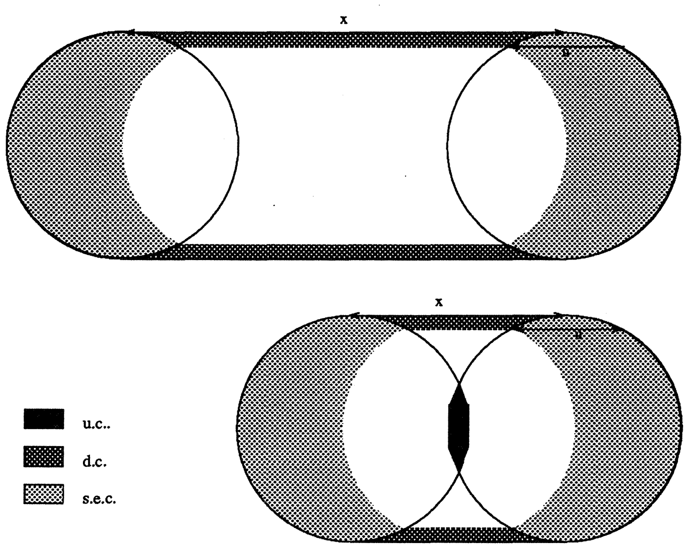

Figure 1.11: u.c., s.e.c. and d.c. 'areas' in set $\mathcal{A}_{\theta,x}$.

$$
\begin{aligned}
&= 4\sqrt{R^2 - \frac{1}{4}u^2}\, g(u)\,\mathrm{d}u && (1.69)\\
\mathrm{d}F^{d.c.}(u) &= \frac{u}{2\sqrt{R^2 - \frac{1}{4}u^2}} \int_{x=u}^{x=\infty} \frac{x-u}{|W| + 2xR}\,\mathrm{d}V(x)\,\mathrm{d}u &&\\
&= \frac{u}{2\sqrt{R^2 - \frac{1}{4}u^2}}\, d(u,u)\,\mathrm{d}u && (1.70)\\
\mathrm{d}F^{u.c.}(u) &= \frac{|W| - 4\,O(u) - 2u\sqrt{R^2 - \frac{1}{4}u^2}}{|W| + 2uR}\,\mathrm{d}V(u) &&\\
&= \frac{z(u)}{|W| + 2uR}\,\mathrm{d}V(u), && (1.71)
\end{aligned}
$$

where $g(\cdot)$ and $d(\star,\cdot)$ are defined as

$$
\begin{aligned}
g(x) &\equiv \int_x^\infty \frac{1}{|W| + 2wR}\,\mathrm{d}V(w) && (1.72)\\
d(x,y) &\equiv \int_x^\infty \frac{w-y}{|W| + 2wR}\,\mathrm{d}V(w) && (1.73)
\end{aligned}
$$

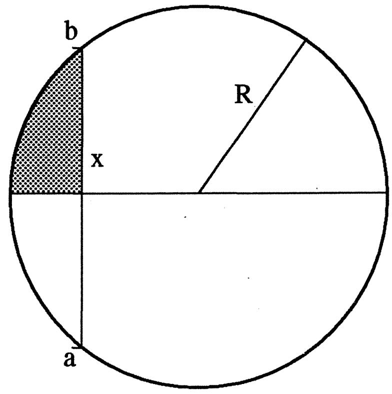

Figure 1.12: the area-integral $O(x)$.

and $z(\cdot)$ as

$$z(x) = |W| - 4\,O(x) - 2x\sqrt{R^2 - \frac{1}{4}x^2}. \qquad (1.74)$$

One sees that $g(x)\,\mathrm{d}\vec{t} = P(\vec{T} \in \mathrm{d}\vec{t}, \widetilde{X} = \text{s.e.c.}, X > x)$ and if $y \leq x$ that $d(x,y)\,\mathrm{d}t_1 = P(T_1 \in \mathrm{d}t_1, T_2 \in \bigcup_{w=x}^{\infty} \mathcal{A}_{\theta,w} \widetilde{X} = y, X > x)$ $((\vec{t}, x, \theta) \in \mathcal{A}_{\theta,x})$, where $\vec{T} = (T_1, T_2)$. In Figure 1.13 we see over what regions in A_θ we integrate (and integrating over all θ) to obtain these densities. The lines in Figure 1.13 (a) correspond with $g(x)\,\mathrm{d}\vec{t}$. Regions as in Figure 1.13 (b) correspond with $d(x,y)\,\mathrm{d}t_1$. In Figure 1.13 (b) we see the area belonging to $d(2R,0)\,\mathrm{d}t_1$.

By symmetry, the possibly censored length $\widetilde{X}$ and type (u.c., s.e.c., d.c.) of each observation is independent of its angle Θ. Moreover the distribution of Θ is the original distribution K of directions. So if we denote by $F^{u.c.}(\cdot,\star)$, $F^{s.e.c.}(\cdot,\star)$ and $F^{d.c.}(\cdot,\star)$ the (conditional on $N = n$) *joint* subdistribution functions of respectively the u.c. , s.e.c. and d.c. observations $(\widetilde{X}, \Theta)$, we have (or calculate with (1.66), (1.67) and (1.68)) for $0 \leq u \leq 2R$, $0 \leq \eta < \pi$

$$\mathrm{d}F^{s.e.c.}(u,\eta) = \mathrm{d}F^{s.e.c.}(u)\,\mathrm{d}K(\eta) \qquad (1.75)$$

$$\mathrm{d}F^{d.c.}(u,\eta) = \mathrm{d}F^{d.c.}(u)\,\mathrm{d}K(\eta) \qquad (1.76)$$

$$\mathrm{d}F^{u.c.}(u,\eta) = \mathrm{d}F^{u.c.}(u)\,\mathrm{d}K(\eta). \qquad (1.77)$$

That the joint subdistribution functions can be factorized in a part only depending on V (or u) or only depending on K (or η) is in general not the case if $\mathbf{W}$ is an arbitrary convex window (compare (1.91)–(1.93)).

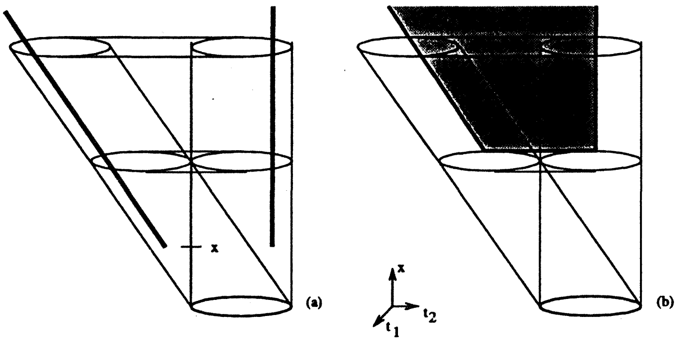

Figure 1.13: the sets in A_θ corresponding to $g(x)\,\mathrm{d}\vec{t}$ and $d(2R,0)\,\mathrm{d}t_1$.

Before we enter the next section we derive an identity for this model similar to (1.14) in the one-dimensional case. Consider in Figure 1.9 the set $\mathcal{A}_{\theta,2R}$. Obviously (since V the marginal density of X) if we integrate (1.64) over the part of set A_θ just beneath the level of set $\mathcal{A}_{\theta,2R}$ and also integrate over all θ, then one obtains the value $V(2R-)$. If we integrate (1.64) over the part of set A_θ drawn in Figure 1.14 (and also integrate over all θ), then we get the value $2\,|W|\,g(2R)$. Integrating (1.64) over the rest of set A, we find

$$c_0 \equiv 2\int_0^{2R} d(2R,u)\,\frac{u}{4\sqrt{R^2-\frac{1}{4}u^2}}\,\mathrm{d}u.$$

This can be understood as follows. If c_1 is the probability of being a double censored observation $\widetilde{X} \leq 2R$ and $X \leq 2R$ (thus that part of set A that is beneath the level $\mathcal{A}_{\theta,2R}$ (for all θ) and that belongs to the double censored region), then we have with (1.70):

$$F^{d.c.}(2R) = \int_0^{2R} d(u,u)\,\frac{u}{2\sqrt{R^2-\frac{1}{4}u^2}}\,\mathrm{d}u = c_0 + c_1.$$

Because we have integrated (1.64) over A, we derived the following equation

$$V(2R-) + 2\,|W|\,g(2R) + 2\int_0^{2R} d(2R,u)\,\frac{u}{4\sqrt{R^2-\frac{1}{4}u^2}}\,\mathrm{d}u = 1. \tag{1.78}$$

We can write

$$d(2R,x) \;=\; \int_{2R}^{\infty}\frac{u-2R}{|W|+2uR}\,\mathrm{d}V(u) + \int_{2R}^{\infty}\frac{2R-x}{|W|+2uR}\,\mathrm{d}V(u)$$

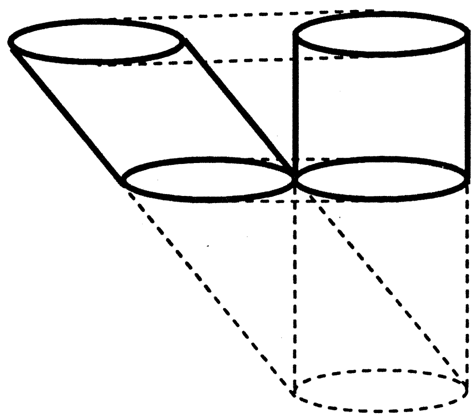

Figure 1.14: the set in A_θ corresponding to $2\,|W|\,g(2R)$

$$= \quad d(2R,2R) + (2R - x)\,g(2R).$$

So we rewrite equation (1.78) as

$$\begin{aligned} 1 &= V(2R-) + 2\,|W|\,g(2R) + 2\,d(2R,2R) \int_0^{2R} \frac{u}{4\sqrt{R^2 - \frac{1}{4}u^2}} \mathrm{d}u \\ &\quad +2\,g(2R) \int_0^{2R} \frac{u(2R-u)}{2\sqrt{R^2 - \frac{1}{4}u^2}} \mathrm{d}u \\ &= V(2R-) + 2\,|W|\,g(2R) + 2R\,d(2R,2R) + (4R^2 - |W|)\,g(2R). \end{aligned}$$

Now we define h as

$$h \equiv 2R\,d(2R,2R) \tag{1.79}$$

and finally obtain the identity

$$V(2R-) + (|W| + 4R^2)\,g(2R) + h = 1. \tag{1.80}$$

In Figure 1.15 we draw the area in set A_θ that corresponds to $h + (4R^2 - |W|)\,g(2R)$. Actually, now one reads equation (1.80) straight from the figures.

With (1.80) we express $g(x)$ and $d(x,x)$ in terms of V on $[0, 2R)$ and h

$$\begin{aligned} g(x) &= \int_x^{2R} \frac{1}{|W| + 2uR} \mathrm{d}V(u) + g(2R) \\ &= \int_x^{2R} \frac{1}{|W| + 2uR} \mathrm{d}V(u) + \frac{1}{|W| + 4R^2}(1 - V(2R-) - h) \end{aligned} \tag{1.81}$$

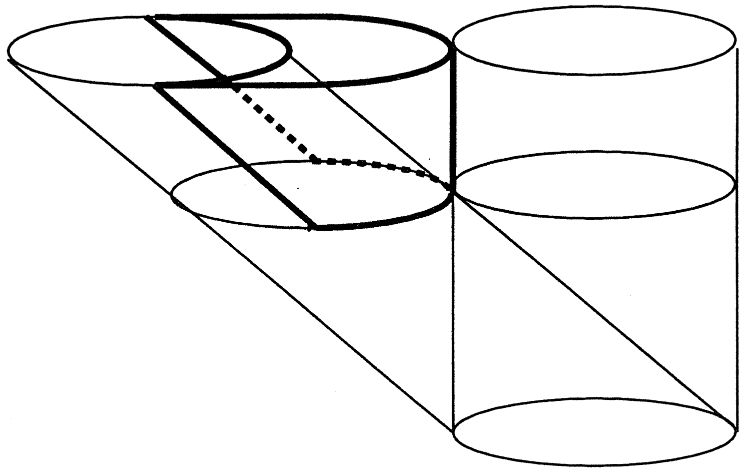

Figure 1.15: the part of A_θ that belongs to h.

$$\begin{aligned} d(x,x) &= \int_x^{2R} \frac{u-x}{|W|+2uR}\,\mathrm{d}V(u) + \frac{1}{2R}h + (2R-x)\,g(2R) \\ &= \int_x^{2R} \frac{u-x}{|W|+2uR}\,\mathrm{d}V(u) + \frac{1}{2R}h + \frac{(2R-x)}{|W|+4R^2}(1-V(2R-)-h). \quad (1.82) \end{aligned}$$

Again with (1.69)-(1.71), (1.81) and (1.82) we note that the distribution of the data only depends on V on $[0,2R)$ and h and that the parameters are identified by it.

1.2.3 Identification in the case W is a circle

To investigate the relation between the parametrizations (F,μ) and (V,h) one can almost copy section 1.1.3 of the one-dimensional case. Here we only give the definitions of φ_{2R} and ψ_{2R} (see definitions of φ_τ and ψ_τ):

$$\begin{aligned} \varphi_{2R}(F(\cdot),\mu) &\equiv \left(\int_0^{\cdot} \frac{|W|+2uR}{|W|+\mu 2R}\,\mathrm{d}F(u), 1 - \frac{|W|+4R^2}{|W|+\mu 2R} + \int_0^{2R} \frac{4R^2-2uR}{|W|+\mu 2R}\,\mathrm{d}F(u)\right) \\ \psi_{2R}(V(\cdot),h) &\equiv \left(\frac{1}{\nu(V(\cdot),h)}\int_0^{\cdot} \frac{1}{|W|+2uR}\,\mathrm{d}V(u), \frac{1}{2R}\left\{\frac{1}{\nu(V(\cdot),h)} - |W|\right\}\right), \end{aligned}$$

where $\nu(V(\cdot),h)$ is defined as

$$\nu(V(\cdot),h) \equiv \frac{1}{|W|+4R^2}(1-h-V(2R-)) + \int_0^{2R} \frac{1}{|W|+2uR}\,\mathrm{d}V(u). \quad (1.83)$$

Note that $\nu(V(\cdot),h)$ corresponds with $g(0) = 1/(|W| + \mu 2R)$. The spaces $\mathcal{F}_{2R}$ and $\mathcal{V}_{2R}$ are defined similar to $\mathcal{F}_\tau$ and $\mathcal{V}_\tau$. We define $\mathcal{S}_{2R}$ to be the set of all subdistribution functions on $[0, 2R)$. Now we define

$$\begin{aligned}\mathcal{F}_{2R} &\equiv \{(F,\mu)\,|\, F \in \mathcal{S}_{2R},\ \mu \in [0,\infty),\ \textstyle\int_0^{2R} w\,\mathrm{d}F(w) + 2R(1 - F(2R-)) \le \mu \\ &\quad \text{and '=' when } F(2R-) = 1\} \\ \mathcal{V}_{2R} &\equiv \{(V,h)\,|\, V \in \mathcal{S}_{2R},\ h \in [0,\infty),\ h \le 1 - V(2R-) \text{ and '=' when } h = 0\}.\end{aligned}$$

The maps φ_{2R} and ψ_{2R} give a 1–1 correspondence between $\mathcal{F}_{2R}$ and $\mathcal{V}_{2R}$. For the same reasons as in the one-dimensional case, we rather work in terms of V and h than in terms of F and μ.

1.2.4 The likelihood, the NPMLE, in the case W is a circle

Here we give the likelihood (conditioning on $N = n$) based on n independent observations $(\widetilde{X}_i, \Delta_i, \Theta_i) = (x_i, d_i, \theta_i)$ ($\Delta_i = d_i = 0, 1, 2$; 0 for u.c., 1 for s.e.c. and 2 for d.c.). Let $x_1 < x_2 < \ldots < x_r$ be the ordered values of the observations $\widetilde{X}_i$. Let ϕ_i, γ_i and ζ_i be the number of respectively the ***uncensored***, ***single end censored*** and ***double censored*** values at x_i. Using (1.69)–(1.71), the likelihood becomes proportional to

$$\prod_{i=1}^{r} (\mathrm{d}V(x_i))^{\phi_i} (g(x_i))^{\gamma_i} (d(x_i,x_i))^{\zeta_i} \cdot \prod_{j=1}^{n} \mathrm{d}K(\theta_j).$$

We are not interested in the (known or unknown) distribution function K. Because in the circle-case $E_K\mathrm{diam}(W) = 2R$ is known, we have that V (and therefore $g(\cdot)$ and $d(\cdot,\star)$) does not depend on K through the transformation (1.65) (compare with (1.61)) and therefore in our search for the NPMLE $(\widehat{V}_n, \widehat{h}_n)$ we use the likelihood proportional to

$$\mathrm{lik}(V,h) \propto \prod_{i=1}^{r} (\mathrm{d}V(x_i)^{\phi_i} (g(x_i))^{\gamma_i} (d(x_i,x_i))^{\zeta_i}, \tag{1.84}$$

where $g(x)$ and $d(x,x)$ can be expressed in terms of (V,h) (see (1.81) and (1.82)) in $\mathcal{V}_{2R}$.

From (1.84) we define by formally taking logs the 'log likelihood-function' on $\mathcal{V}_{2R}$ as

$$\begin{aligned}\Psi(V,h) \equiv \int_0^{2R} \log(\mathrm{d}V(x))\,\mathrm{d}F_n^{u.c.}(x) &+ \int_0^{2R} \log(g(x))\,\mathrm{d}F_n^{s.e.c.}(x) \\ &+ \int_0^{2R} \log(d(x,x))\,\mathrm{d}F_n^{d.c.}(x),\end{aligned} \tag{1.85}$$

The empirical subdistribution functions $F_n^{u.c.}$, $F_n^{s.e.c.}$ and $F_n^{d.c.}$ on $[0, 2R]$ are defined as in (1.36) – (1.38). Now with (1.85) we define the NPMLE $(\widehat{V}_n, \widehat{h}_n)$ of the underlying parameters analogue to the definition in section 1.1.4.

Similar to section 1.1.6 we obtain the self-consistency equations for the NPMLE (providing its existence) in the circle-case:

$$\begin{aligned}\mathrm{d}\widehat{V}_n(x) = \mathrm{d}F_n^{u.c.}(x) &+ \int_0^x \frac{1}{\widehat{g}_n(u)}\,\mathrm{d}F_n^{s.e.c.}(u) \cdot \frac{1}{|W| + 2xR}\,\mathrm{d}\widehat{V}_n(x) \\ &+ \int_0^x \frac{x-u}{\widehat{d}_n(u,u)}\,\mathrm{d}F_n^{d.c.}(u) \cdot \frac{1}{|W| + 2xR}\,\mathrm{d}\widehat{V}_n(x)\end{aligned} \tag{1.86}$$

and

$$\widehat{h}_n = \widehat{d}_n(2R,2R)\int_0^{2R}\frac{1}{\widehat{d}_n(u,u)}\,\mathrm{d}F_n^{d.c.}(u) \tag{1.87}$$

$$(|W|+4R^2)\,\widehat{g}_n(2R) = 1-\widehat{h}_n-\widehat{V}_n(2R). \tag{1.88}$$

(Compare these equations with (1.50)–(1.52) in the one-dimensional case). Note that in the one-dimensional case $\widehat{h}_n$ was estimated by the fraction of the double censored observations. We had only one kind of double censored observations there. In the two-dimensional problem, this is no longer the case.

If we look at the likelihood in more detail and consider the factors $g(x)$ and $d(x,x)$, which are expressed in terms of (V,h) in (1.81) and (1.82), then we see in (1.81) that $1/(|W|+2uR)$ (in the first term at the right-hand side) is decreasing in u, so it pays to move mass to the left and in (1.82) we have that $(u-x)/(|W|+2uR)$ (in the first term at the right-hand side) is increasing in u and thus for $d(x,x)$ it pays to move mass to the right. This implies that we do not have (or there is not) a simple prescription to argue that the NPMLE is a discrete estimator with mass on the observation points only.

Instead of trying to find out how the NPMLE (if it exists) distributes its mass in a nondiscrete setting, we avoid this by defining the sieved NPMLE $(\widehat{V}_n,\widehat{h}_n)$. Just as in section 1.1.7 we define $\mathcal{V}_{\text{sieve}}$ to be the set of discrete estimators in $\overline{\mathcal{V}}_{2R}$, which give (possible) mass to the u.c. observation points only (and h and $g(2R)$). Similar to the one-dimensional case one can prove the existence and uniqueness of the sieved NPMLE. Let r_0 be the number of u.c. observation points in $[0,2R)$. The measure π_{r_0+2}, which gives mass $1/(r_0+2)$ to all u.c. observation points and to h and $g(2R)$, dominates the set $\mathcal{V}_{\text{sieve}}$. If Π_{r_0+2} denotes the distribution function of π_{r_0+2} on $[0,2R)$, then we define the sieved NPMLE $(\widehat{V}_n,\widehat{h}_n)$ by

$$(\widehat{V}_n,\widehat{h}_n)\equiv\arg\max_{(\widetilde{V},\widetilde{h})\in\mathcal{V}_{\text{sieve}}}\left(\Psi(\widetilde{V},\widetilde{h})-\Psi(\Pi_{r_0+2},1/(r_0+2))\right). \tag{1.89}$$

Starting the EM-algorithm with an initial distribution like π_{r_0+2}, the algorithm will converge to a solution of the self-consistency equations (1.86)–(1.88): the sieved NPMLE (see sections 1.1.6 and 1.1.7).

1.2.5 W is an arbitrary convex window

In section 1.2.2 we derived for the two-dimensional problem the subdistribution functions of the data, in section 1.2.3 the parametrization (V,h) versus (F,μ) and in section 1.2.4 the selfconsistency equations in the case that the window $\mathbf{W}$ is a circle. Because of the completely unimportant role of θ (and the distribution function K), the calculations there are rather straightforward. In this section and the next we try to imitate the derivations to obtain similar results in the case that $\mathbf{W}$ is an arbitrary convex window. In this section we calculate the subdistribution functions of the data and in section 1.2.6 we analyse the likelihood.

We remember the probability measure (1.60). Now we calculate (conditional on $N=n$) the joint subdistribution functions of the data $(\widetilde{X}_i,\Delta_i,\Theta_i)$. We define $z(\theta)$ to be the greatest distance in $\mathbf{W}$ in the θ direction. Of course in the 'circle-case' $z(\theta)$ equals $2R$ for all θ. Furthermore we define

$$P \equiv \max_{\theta} z(\theta). \tag{1.90}$$

Let $w_1(x,u,\theta)$ be the area of the set $\mathcal{A}_{\theta,x}$ for which $(\vec{T},X,\Theta) = (\vec{t},x,\theta)$ gives a s.e.c. observation $\widetilde{X} \leq u$ $(0 \leq u \leq z(\theta))$. Because $\mathbf{W}$ is convex, there exists a continuous function $a(\cdot,\theta)$ such that

$$w_1(x,u,\theta) = \begin{cases} a(u,\theta) & \text{if } x > u \\ a(x,\theta) & \text{if } x \leq u. \end{cases}$$

One can see this from a picture. In the 'circle-case' we have

$$a(u,\theta) = 8O(u) + 4u\sqrt{R^2 - \frac{1}{4}u^2}$$

(see $s_1(\star,\cdot)$ and $s_2(\star,\cdot)$ in section 1.2.2).

Let $w_2(x,u,\theta)$ be the area of the set $\mathcal{A}_{\theta,x}$ for which $(\vec{T},X,\Theta) = (\vec{t},x,\theta)$ gives a d.c. observation $\widetilde{X} \leq u$ $(0 \leq u \leq z(\theta))$. Again (note that $\mathbf{W}$ is convex) from a picture one sees that there exist functions $p_i(\cdot,\theta)$ and $b_i(\theta)$ $(i = 1,2)$ such that

$$w_2(x,u,\theta) = \begin{cases} \displaystyle\int_{w=b_1(\theta)}^{w=u} (x-w)\,p_1(\mathrm{d}w,\theta) + \int_{w=b_2(\theta)}^{w=u} (x-w)\,p_2(\mathrm{d}w,\theta) & \text{if } x > u \\ \displaystyle\int_{w=b_1(\theta)}^{w=x} (x-w)\,p_1(\mathrm{d}w,\theta) + \int_{w=b_2(\theta)}^{w=x} (x-w)\,p_2(\mathrm{d}w,\theta) & \text{if } x \leq u. \end{cases}$$

In the 'circle-case' we have $b_i(\theta) = 0$ and

$$p_i(w,\theta) = \sqrt{R^2 - \frac{1}{4}w^2}$$

or rewritten as

$$p_i(\mathrm{d}w,\theta) = \frac{w}{4\sqrt{R^2 - \frac{1}{4}w^2}}\,\mathrm{d}w$$

for all θ and $i = 1,2$. Note that $s_3(\star,\cdot)$ and $s_4(\star,\cdot)$ in section 1.2.2 can be written like $w_2(x,u,\theta)$.

Let $w_3(x,u,\theta)$ be the area of the set $\mathcal{A}_{\theta,x}$ for which $(\vec{T},X,\Theta) = (\vec{t},x,\theta)$ gives an u.c. observation $\widetilde{X} \leq u$ $(0 \leq u \leq z(\theta))$. Note that there exists a function $q(x,\theta)$ such that $w_3(x,u,\theta)$ can be written as

$$w_3(x,u,\theta) = \begin{cases} 0 & \text{if } x > u \\ q(x,\theta) & \text{if } x \leq u. \end{cases}$$

In the 'circle-case' we have

$$q(x,\theta) = |W| - 4O(x) - 2x\sqrt{R^2 - \frac{1}{4}x^2}$$

(see $s_5(\star,\cdot)$ in section 1.2.2).

Now if $X = x$ and $\Theta = \theta$, one writes down the fractions of the areas of $\mathcal{A}_{\theta,x}$ for which $(\vec{T}, X, \Theta) = (\vec{t}, x, \theta)$ gives respectively a s.e.c., d.c. and u.c. observation $\widetilde{X} \le u$ $(0 \le u \le z(\theta))$. We will do this for the s.e.c. case and we obtain

$$\begin{aligned}\frac{w_1(x,u,\theta)}{|W| + x\text{diam}(W,\theta)} &= \frac{a(u,\theta)}{|W| + x\text{diam}(W,\theta)} 1_{\{x>u, 0\le u\le z(\theta)\}}(u) \\ &+ \frac{a(x,\theta)}{|W| + x\text{diam}(W,\theta)} 1_{\{x\le u, 0\le u\le z(\theta)\}}(u).\end{aligned}$$

Let us calculate $F^{s.e.c.}(u,\eta)$:

$$\begin{aligned}&F^{s.e.c.}(u,\eta) \\ &= \int_{x=0}^{x=\infty}\int_{\theta=0}^{\theta=\pi} \frac{w_1(x,u,\theta)}{|W| + x\text{diam}(W,\theta)} 1_{\{\theta\le\eta\}}(\theta)\, \mathrm{d}J(\theta|X=x)\, \mathrm{d}V(x) \\ &= \int_{x=u}^{x=\infty}\int_{\theta=0}^{\theta=\eta} \frac{a(u,\theta)}{|W| + x\text{diam}(W,\theta)} 1_{\{0\le u\le z(\theta)\}}(u)\, \mathrm{d}J(\theta|X=x)\, \mathrm{d}V(x) \\ &\quad + \int_{x=0}^{x=u}\int_{\theta=0}^{\theta=\eta} \frac{a(x,\theta)}{|W| + x\text{diam}(W,\theta)} 1_{\{0\le u\le z(\theta)\}}(u)\, \mathrm{d}J(\theta|X=x)\, \mathrm{d}V(x) \\ &= \int_{x=u}^{x=\infty}\int_{\theta=0}^{\theta=\eta} \frac{a(u,\theta)}{|W| + xE_K\text{diam}(W)} 1_{\{0\le u\le z(\theta)\}}(u)\, \mathrm{d}K(\theta)\, \mathrm{d}V(x) \\ &\quad + \int_{x=0}^{x=u}\int_{\theta=0}^{\theta=\eta} \frac{a(x,\theta)}{|W| + xE_K\text{diam}(W)} 1_{\{0\le u\le z(\theta)\}}(u)\, \mathrm{d}K(\theta)\, \mathrm{d}V(x).\end{aligned}$$

This yields

$$\begin{aligned}&F^{s.e.c.}(\mathrm{d}u,\mathrm{d}\eta) \\ &= a(\mathrm{d}u,\eta)\int_{x=u}^{x=\infty} \frac{1}{|W| + xE_K\text{diam}(W)}\, \mathrm{d}V(x)\, \mathrm{d}K(\eta)\, 1_{\{0\le u\le z(\eta)\}}(u) \\ &\quad - \frac{a(u,\eta)}{|W| + uE_K\text{diam}(W)}\, \mathrm{d}V(u)\, \mathrm{d}K(\eta)\, 1_{\{0\le u\le z(\eta)\}}(u) \\ &\quad + \frac{a(u,\eta)}{|W| + uE_K\text{diam}(W)}\, \mathrm{d}V(u)\, \mathrm{d}K(\eta)\, 1_{\{0\le u\le z(\eta)\}}(u).\end{aligned}$$

Doing similar computations for the other two subdistribution functions, we find

$$\begin{aligned}&F^{s.e.c.}(\mathrm{d}u,\mathrm{d}\eta) \\ &= a(\mathrm{d}u,\eta)\int_{x=u}^{x=\infty} \frac{1}{|W| + xE_K\text{diam}(W)}\, \mathrm{d}V(x)\, \mathrm{d}K(\eta)\, 1_{\{0\le u\le z(\eta)\}}(u) \\ &= a(\mathrm{d}u,\eta)\, g(u)\, \mathrm{d}K(\eta)\, 1_{\{0\le u\le z(\eta)\}}(u) \qquad (1.91) \\ &F^{d.c.}(\mathrm{d}u,\mathrm{d}\eta) \\ &= (p_1(\mathrm{d}u,\eta) + p_2(\mathrm{d}u,\eta)) \\ &\quad \times \int_{x=u}^{x=\infty} \frac{x-u}{|W| + xE_K\text{diam}(W)}\, \mathrm{d}V(x)\, \mathrm{d}K(\eta)\, 1_{\{0\le u\le z(\eta)\}}(u) \\ &= (p_1(\mathrm{d}u,\eta) + p_2(\mathrm{d}u,\eta))\, d(u,u)\, \mathrm{d}K(\eta)\, 1_{\{0\le u\le z(\eta)\}}(u) \qquad (1.92)\end{aligned}$$

$$F^{u.c.}(\mathrm{d}u,\mathrm{d}\eta) = \frac{q(u,\eta)}{|W| + uE_K\mathrm{diam}(W)}\,\mathrm{d}V(u)\,\mathrm{d}K(\eta)\,1_{\{0\le u\le z(\eta)\}}(u), \tag{1.93}$$

where $g(\cdot)$ and $d(\star,\cdot)$ are defined similar to (1.72) and (1.73):

$$g(x) = \int_x^\infty \frac{1}{|W| + wE_K\mathrm{diam}(W)}\,\mathrm{d}V(w) \tag{1.94}$$

$$d(x,y) = \int_x^\infty \frac{w-y}{|W| + wE_K\mathrm{diam}(W)}\,\mathrm{d}V(w). \tag{1.95}$$

Now we try to obtain an identity generalizing (1.80). If we integrate (1.60) over the part of set A_θ beneath the level of set $\mathcal{A}_{\theta,P}$ and integrate over all θ, then we obtain the value $V(P)$. If we integrate (1.60) over the part of set A_θ starting at level $\mathcal{A}_{\theta,P}$ and similarly drawn as in Figure 1.14 and integrating over all θ, then we get $2|W|\,g(P)$. Note that in this case the 'tubes' drawn as in Figure 1.14 do not have to hit each other at level $\mathcal{A}_{\theta,P}$ as in the 'circle-case'. Finally we integrate (1.60) over the rest of set A, which is that region of the double censored observations in A that is above the level $\mathcal{A}_{\theta,P}$ for all θ. We find

$$\begin{aligned} a_0 &\equiv \sum_{i=1,2}\int_{w=P}^{w=\infty}\int_{\theta=0}^{\theta=\pi}\int_{u=b_i(\theta)}^{u=z(\theta)} \frac{w-u}{|W| + w\mathrm{diam}(W,\theta)}\,p_i(\mathrm{d}u,\theta)\,\mathrm{d}J(\theta|X=w)\,\mathrm{d}V(w) \\ &= \sum_{i=1,2}\int_{\theta=0}^{\theta=\pi}\int_{u=b_i(\theta)}^{u=z(\theta)} d(P,u)\,p_i(\mathrm{d}u,\theta)\,\mathrm{d}K(\theta). \end{aligned}$$

We get the equality

$$V(P-) + 2\,|W|\,g(P) + a_0 = 1. \tag{1.96}$$

Again we derive $d(P,x) = d(P,P) + (P-x)\,g(P)$ and use this to rewrite (1.96) as

$$\begin{aligned} V(P-) + 2\,|W|\,g(P) &+ d(P,P)\sum_{i=1,2}\int_{\theta=0}^{\theta=\pi}\int_{u=b_i(\theta)}^{u=z(\theta)} p_i(\mathrm{d}u,\theta)\,\mathrm{d}K(\theta) \\ &+ g(P)\sum_{i=1,2}\int_{\theta=0}^{\theta=\pi}\int_{u=b_i(\theta)}^{u=z(\theta)} (P-u)\,p_i(\mathrm{d}u,\theta)\,\mathrm{d}K(\theta) = 1. \end{aligned}$$

If we define h as

$$h \equiv d(P,P)\sum_{i=1,2}\int_{\theta=0}^{\theta=\pi}\int_{u=b_i(\theta)}^{u=z(\theta)} p_i(\mathrm{d}u,\theta)\,\mathrm{d}K(\theta) \tag{1.97}$$

and

$$a_1 \equiv \sum_{i=1,2}\int_{\theta=0}^{\theta=\pi}\int_{u=b_i(\theta)}^{u=z(\theta)} (P-u)\,p_i(\mathrm{d}u,\theta)\,\mathrm{d}K(\theta),$$

we derived the identity

$$V(P-) + (2|W| + a_1)\,g(P) + h = 1. \tag{1.98}$$

(Compare (1.98) with (1.80)). One verifies that in the 'circle-case' h equals $2R\,d(2R,2R)$ and a_1 equals $4R^2 - |W|$.

For given K one can establish the 1–1 correspondence between the parametrizations (V,h) and (F,μ). One can almost copy the definitions of section 1.2.3 to get the corresponding expressions in this case and go through section 1.1.3 to convince oneself.

1.2.6 The likelihood, the NPMLE, in the case W is arbitrary convex

If we write down the likelihood (conditioning on $N = n$) based on n independent observations $(\widetilde{X}_i, \Delta_i, \Theta_i) = (x_i, d_i, \theta_i)$, then the likelihood is proportional to

$$\prod_{i=1}^{r} \left(\frac{\mathrm{d}V(x_i)}{|W| + x_i E_K \mathrm{diam}(W)} \right)^{\phi_i} (g(x_i))^{\gamma_i} (d(x_i, x_i))^{\zeta_i} \cdot \prod_{j=1}^{n} \mathrm{d}K(\theta_j), \tag{1.99}$$

where $x_1 < x_2 < \ldots < x_r$ are the ordered values of the observations $\widetilde{X}_i$ and ϕ_i, γ_i and ζ_i are the numbers of respectively the *uncensored*, *single end censored* and *double censored* values at x_i. One sees the similarity with the 'circle-case', but one must remember that through the transformation (1.61) V and therefore $g(\cdot)$ and $d(\cdot, \star)$ depend on K. If we write the (proportional) likelihood in terms of F and μ (and K) we get

$$(|W| + \mu\, E_K \mathrm{diam}(W))^{-n} \prod_{i=1}^{r} (\mathrm{d}F(x_i))^{\phi_i} (1 - F(x_i))^{\gamma_i} \left(\int_{x_i}^{\infty} (1 - F(w)) \mathrm{d}w \right)^{\zeta_i} \prod_{j=1}^{n} \mathrm{d}K(\theta_i).$$

Suppose we do not know the distribution function K of the angle Θ. So we would like to compute $\widehat{F}$ and $\widehat{K}$ by jointly maximizing the likelihood above. Unfortunately this does not decompose into separate maximization problems for F and K, but we can think of a natural iterative scheme. Firstly, we determine F given K by maximizing

$$(|W| + \mu\, E_K \mathrm{diam}(W))^{-n} \prod_{i=1}^{r} (\mathrm{d}F(x_i))^{\phi_i} (1 - F(x_i))^{\gamma_i} \left(\int_{x_i}^{\infty} (1 - F(w)) \mathrm{d}w \right)^{\zeta_i}.$$

This is something we already know how to handle (just use (1.99) with known K which is similar to the 'circle-case'). Secondly, we determine K given F by maximizing

$$(|W| + \mu\, E_K \mathrm{diam}(W))^{-n} \prod_{j=1}^{n} \mathrm{d}K(\theta_i).$$

Lok(1994) shows that the NPMLE $\widehat{K}_n$ of K for given F can be expressed as

$$\widehat{K}_n(\theta) = \frac{\int_{[0,\theta]} (|W| + \mu\, \mathrm{diam}(W, \eta))^{-1} \mathrm{d}L_n(\eta)}{\int_{[0,\pi)} (|W| + \mu\, \mathrm{diam}(W, \eta))^{-1} \mathrm{d}L_n(\eta)},$$

where $L_n(\theta) = \frac{1}{n} \sum_{i=1}^{i=n} 1_{[0,\theta)}(\Theta_i)$ is the empirical distribution of the observed angles.

Suppose we know the distribution function K of the angle Θ, then $E_K\mathrm{diam}(W)$ is known. Now the likelihood, only with factors depending on V (or F), becomes similar to (1.84) and one finds for the self-consistency equations for the NPMLE $(\widehat{V}_n, \widehat{h}_n)$ of (V, h):

$$\begin{aligned} \mathrm{d}\widehat{V}_n(x) = \mathrm{d}F_n^{u.c.}(x) \;&+\; \int_0^x \frac{1}{\widehat{g}_n(u)} \mathrm{d}F_n^{s.e.c.}(u) \cdot \frac{1}{|W| + x\, E_K \mathrm{diam}(W)} \mathrm{d}\widehat{V}_n(x) \\ &+\; \int_0^x \frac{x - u}{\widehat{d}_n(u, u)} \mathrm{d}F_n^{d.c.}(u) \cdot \frac{1}{|W| + x\, E_K \mathrm{diam}(W)} \mathrm{d}\widehat{V}_n(x) \end{aligned} \tag{1.100}$$

and

$$\widehat{h}_n = d(P,P) \sum_{i=1,2} \int_{u=0}^{u=P} \frac{1}{\widehat{d}_n(u,u)} \, \mathrm{d}F_n^{d.c.}(u) \qquad (1.101)$$

$$(2\,|W| + a_1)\,\widehat{g}_n(P) = 1 - \widehat{h}_n - \widehat{V}_n(P), \qquad (1.102)$$

where $F_n^{\cdot\cdot}(\cdot)$ are the marginal empirical subdistribution functions. The same problems arise as in the 'circle-case' in section 1.2.4, if we want to define the NPMLE. Therefore one is advised to consider here a sieved NPMLE too.

Because of the fact that in the 'circle-case' the distribution function K of the angles does not play any role in the search for an efficient NPMLE of F and the fact that information calculations (chapter 3) and consistency analysis (chapter 2) turn out to be less laborious in this case, we stick for the rest of the book to the 'circle-case'. So the two-dimensional case in the chapters 2 and 3 will be the 'circle'. We think that with more effort one can imitate the analysis in these chapters to obtain similar results for the case that $\mathbf{W}$ is arbitrary convex and K is known.

Chapter 2

Consistency

In this chapter we prove consistency results of the (sieved) NPMLE's of chapter 1. In section 2.1 the proof of consistency is outlined in a general setting and the very few conditions to be verified make this proof applicable in other linear convex models where the consistency of the (sieved) NPMLE is investigated. In section 2.2 we obtain consistency results in the one-dimensional case. In section 2.3 we do the same for the two-dimensional case. Actually, there we prove consistency in the case that **W** is a circle. The reason that we write down the 'circle-case' is that the formulas are less complicated than in the case that **W** is an arbitrary convex window and one sees immediately the similarity with the one-dimensional case. Furthermore in the case that **W** is an arbitrary convex window, we have to deal with the distribution of the angles and thus the problem arises of maximizing the likelihood jointly for V (or F) and K. If K is a known distribution function, we are convinced that section 2.3 can be more or less copied to get the consistency results for this case too.

The proof of consistency in section 2.1 is a generalization of the method described by Groeneboom(1991), following Jewell(1982). If $\phi(F)$ denotes the log likelihood function there, then the method is based on the fact that

$$\phi((1-\epsilon)\,\widehat{F}_n + \epsilon\, F_0) - \phi(\widehat{F}_n) \le 0,$$

where $\widehat{F}_n$ is a NPMLE of the underlying distribution function F_0. For our log likelihood function we could also write

$$\Psi((1-\epsilon)\,\widehat{V}_n + \epsilon\, V, (1-\epsilon)\,\widehat{h}_n + \epsilon\, h) - \Psi(\widehat{V}_n, \widehat{h}_n) \le 0,$$

where $(\widehat{V}_n, \widehat{h}_n)$ is the NPMLE of (V, h). Evaluating this difference, we encounter terms like $\log((\mathrm{d}V/\mathrm{d}\widehat{V}_n)(x))$, which equal $-\infty$ when V and $\widehat{V}_n$ are mutually singular, so the inequality becomes $-\infty \le 0$, which is completely ininformative. Instead of using the underlying distribution function V, we take some convenient empirical counterpart V_n of V (so V_n converges in supremum norm to V) and satisfying $V_n \ll \widehat{V}_n$, so that the expression $(\mathrm{d}V_n/\mathrm{d}\widehat{V}_n)(x)$ is more informative. Furthermore we will use the self-consistency equations to get an expression for the Radon-Nikodym derivative $(\mathrm{d}V_n/\mathrm{d}\widehat{V}_n)(x)$.

We think the proof of consistency, described in section 2.1, can be applied in other linear models such as the random-multiplicative censoring model of Vardi and Zhang(1992). Their approach gives not only consistency in the random-multiplicative censoring model, but the whole asymptotic characterization. Trying to use their approach in our line segment models, it unfortunately seems to fail.

For the $\widehat{T}_n$ and T in section 2.1 one reads $(\widehat{V}_n, \widehat{h}_n)$ and (V, h) respectively, if one wants the connection with our models. We note that $\mathcal{V}_T$ is convex and $(V, h) \to P_{(V,h)}$ is linear ($P_{(V,h)}$ is the distribution of the data under (V, h)). (The same for $\mathcal{V}_{2R}$). In this chapter we use sometimes $\|\cdot\|_I$ to denote the $\|\cdot\|_\infty$ (supremum norm) on the interval I.

2.1 General idea

In this section we write $\mathrm{d}P$ instead of $\mathrm{d}P(x)$ to get short notation.

Let $\widehat{T}_n$ be the NPMLE of T, T_n an ad-hoc estimator (or T itself if convenient), P_T the distribution of the data under the parameter T and P_n the empirical. Assume that $\mathcal{T}$ is a convex set.

Suppose P_T is linear in T, $T \in \mathcal{T}$. We have (because $\widehat{T}_n$ is the NPMLE)

$$\begin{aligned} 0 &\geq \int \log\left(\frac{\mathrm{d}P_{(1-\epsilon)\widehat{T}_n+\epsilon T_n}}{\mathrm{d}P_{\widehat{T}_n}}\right) \mathrm{d}P_n \\ &= \int \log\left(1+\epsilon\left(\frac{\mathrm{d}P_{T_n}}{\mathrm{d}P_{\widehat{T}_n}}-1\right)\right) \mathrm{d}P_n, \end{aligned} \tag{2.1}$$

where $\epsilon \in [0,1]$. We divide the inequality above by ϵ and let $\epsilon \downarrow 0$. We find

$$\int \left(\frac{\mathrm{d}P_{T_n}}{\mathrm{d}P_{\widehat{T}_n}}-1\right) \mathrm{d}P_n \leq 0$$

or

$$\int \frac{\mathrm{d}P_{T_n}}{\mathrm{d}P_{\widehat{T}_n}} \mathrm{d}P_n \leq 1. \tag{2.2}$$

As $n \to \infty$, we will try to arrange things that (perhaps on a subsequence only—here we will need assumptions on $\mathcal{T}$; for example $\mathcal{T}$ is a set of distribution functions) $\widehat{T}_n \to T_\infty \in \mathcal{T}$, $T_n \to T$, $P_n \to P_T$ and try to prove from this

$$\int \frac{\mathrm{d}P_{T_n}}{\mathrm{d}P_{\widehat{T}_n}} \mathrm{d}P_n \to \int \frac{\mathrm{d}P_T}{\mathrm{d}P_{T_\infty}} \mathrm{d}P_T \leq 1. \tag{2.3}$$

Showing (2.3) is the same as saying that inequality (2.2) holds in the limit. However we do have

$$\int \frac{\mathrm{d}P_T}{\mathrm{d}P_{T_\infty}} \mathrm{d}P_T \geq 1$$

for all $T_\infty \in \mathcal{T}$, with equality iff $T_\infty = T$ (we assume identifiability: $P_T = P_{T_\infty} \Longleftrightarrow T = T_\infty$). This follows because

$$\int \log\left(\frac{\mathrm{d}P_{(1-\epsilon)T_\infty+\epsilon T}}{\mathrm{d}P_{T_\infty}}\right) \mathrm{d}P_T, \quad \epsilon \in [0,1]$$

is maximal at $\epsilon = 1$. Moreover it equals

$$\int \log\left(1+\epsilon\left(\frac{\mathrm{d}P_T}{\mathrm{d}P_{T_\infty}}-1\right)\right)\,\mathrm{d}P_T$$

which is concave in ϵ (it is an average of concave functions: $\log(1+\epsilon\, c)$ is concave in ϵ, both for negative and positive c). Its derivative at $\epsilon = 0$ is therefore nonnegative: so we get

$$\int \left(\frac{\mathrm{d}P_T}{\mathrm{d}P_{T_\infty}}-1\right)\,\mathrm{d}P_T \ \geq 0$$

or

$$\int \frac{\mathrm{d}P_T}{\mathrm{d}P_{T_\infty}}\,\mathrm{d}P_T \ \geq 1. \tag{2.4}$$

We want to show that we have '>' in inequality (2.4) if $P_T \neq P_{T_\infty}$. If $(\mathrm{d}P_T\ /\mathrm{d}P_{T_\infty}) = \infty$ with positive P_T probability, then we have immediately

$$\int \frac{\mathrm{d}P_T}{\mathrm{d}P_{T_\infty}}\,\mathrm{d}P_T \ = \infty > 1$$

and $P_T \neq P_{T_\infty}$. In the other case we note that $\log(1+\epsilon\, c)$ is strictly concave in ϵ as long as

$$c = \left(\frac{\mathrm{d}P_T}{\mathrm{d}P_{T_\infty}}-1\right) \neq 0,$$

and an average of strictly concave functions is strictly concave. So we have strict concavity (and hence '>' in inequality (2.4)) unless $(\mathrm{d}P_T\ /\mathrm{d}P_{T_\infty}) = 1$ with P_T -probability 1, which implies $P_T\ = P_{T_\infty}$.

Because we have (2.3) and showed that (2.4) holds with '>' if $P_T \neq P_{T_\infty}$, we proved that we must have $P_T\ = P_{T_\infty}$, and so $T = T_\infty$. We proved the following theorem

Theorem 2.1.1 *Let $\widehat{T}_n$ be the NPMLE of T. Let P_T be the distribution of the data under the parameter T and P_n the empirical. Let T_n be an ad-hoc estimator (or T itself if convenient) such that $P_{T_n} \ll P_{\widehat{T}_n}$. We assume identifiability: $P_{T_1} = P_{T_2} \iff T_1 = T_2$. Let $\mathcal{T}$ be a convex set of distribution functions.*

If P_T is linear in T, $T \in \mathcal{T}$, and

$$\int \frac{\mathrm{d}P_{T_n}}{\mathrm{d}P_{\widehat{T}_n}}\,\mathrm{d}P_n \rightarrow \int \frac{\mathrm{d}P_T}{\mathrm{d}P_{T_\infty}}\,\mathrm{d}P_T \qquad (n\rightarrow\infty)$$

(perhaps on a subsequence only), then $T_\infty = T$. So if moreover $\widehat{T}_n \rightarrow T_\infty$, we have $\widehat{T}_n \rightarrow T$.

2.2 Consistency in the one-dimensional case

In this section we formulate consistency results for the NPMLE $(\widehat{V}_n, \widehat{h}_n)$ of (V, h) in the one-dimensional case introduced in section 1.1. Furthermore we show that these imply similar consistency results for the NPMLE $(\widehat{F}_n, \widehat{\mu}_n)$ of our original parameters (F, μ). These results are stated in the theorems 2.2.5 and 2.2.7. Theorem 2.2.5 is about the consistency of $(\widehat{V}_n, \widehat{h}_n)$ and will be proved in section 2.2.1. Theorem 2.2.7 is about the consistency of $(\widehat{F}_n, \widehat{\mu}_n)$ and follows easily from theorem 2.2.5 and the 1–1 correspondence between the parametrizations (V, h) and (F, μ).

Before we prove our consistency results, we give the following four lemmas, which will be frequently used. The first three lemmas give us some conditions, which imply the convergence of $\int_a^b G_n(x)\,\mathrm{d}H_n(x)$ to $\int_a^b G(x)\,\mathrm{d}H(x)$. Lemma 2.2.4 provides us with conditions for proving that a weakly convergent sequence of monotone functions, converges uniformly.

We note that with $\int_x^y$, where $x \leq y$ and $x, y \in [0, \tau)$, we mean the integral over $(x, y]$. Of course this implies that $\int_x^{y-}$ is the integral over (x, y). Remember that with $\int_x^\tau$ we mean the integral over (x, τ).

Lemma 2.2.1 *If G_n, G, H_n and H are measures on the interval $[a, b] \subset \mathbb{R}$ such that (i) $G_n(x) - G(x) \to 0$ for all $x \in [a, b]$ except on a negligleable set w.r.t. Lebesgue measure, (ii) $\|H_n - H\|_\infty \to 0$, (iii) G_n is bounded and of bounded variation uniformly in n and (iv) H is absolutely continuous w.r.t. Lebesgue measure and $\mathrm{d}H(x)/\mathrm{d}x$ is bounded, then*

$$\int_a^b G_n(x)\,\mathrm{d}H_n(x) - \int_a^b G(x)\,\mathrm{d}H(x) \to 0$$

for $n \to \infty$.

PROOF: the proof of this lemma is straightforward. We write

$$\begin{aligned} & \int_a^b G_n(x)\,\mathrm{d}H_n(x) - \int_a^b G(x)\,\mathrm{d}H(x) \\ = \; & G_n(b)\,(H_n(b) - H(b)) \\ & -G_n(a)\,(H_n(a) - H(a)) \\ & -\int_a^b (H_n(x) - H(x))\,\mathrm{d}G_n(x) \\ & +\int_a^b (G_n(x) - G(x))\,\mathrm{d}H(x). \end{aligned}$$

Because G_n is bounded uniformly in n (condition (iii)) and because of (ii) we see that the first and second term converge to 0. For the third term we have

$$\left| \int_a^b (H_n(x) - H(x))\,\mathrm{d}G_n(x) \right| \leq \|H_n - H\|_\infty \cdot \int_a^b |\mathrm{d}G_n(x)|\,.$$

Because of (iii) we have that $\int_a^b |\mathrm{d}G_n(x)| \leq c$ (uniformly in n) for some constant c and thus with (ii) we get that the third term converges to 0. With (i), (iii) and (iv) we apply Lebesgue's dominated convergence theorem to get the fourth term tending to 0. This proves the lemma. □

Lemma 2.2.2 *If G_n, G, H_n and H are measures on the interval $[a, b] \subset \mathbb{R}$ such that (i) $\|G_n - G\|_\infty \to 0$, (ii) $\|H_n - H\|_\infty \to 0$, (iii) G_n is bounded and of bounded variation uniformly in n and (iv) H is of bounded variation, then*

$$\int_a^b G_n(x)\,\mathrm{d}H_n(x) - \int_a^b G(x)\,\mathrm{d}H(x) \to 0$$

for $n \to \infty$.

PROOF: the proof is the same as the proof of lemma 2.2.1 except for the term $\int_a^b (G_n(x) - G(x))\,dH(x)$. For this term we have now

$$\left|\int_a^b (G_n(x) - G(x))\,dH(x)\right| \leq \|G_n - G\|_\infty \cdot \int_a^b |dH(x)|\,.$$

Because of (iv) we have that $\int_a^b |dH(x)| \leq c$ for some constant and thus with (i) we get that this term tends to 0. □

Lemma 2.2.3 *The lemmas 2.2.1 and 2.2.2 still hold if we replace* $\int_a^b$ *by* $\int_a^{b-}$.

PROOF: replace b by $b-$ in the proofs of the lemma's 2.2.1 and 2.2.2. □

Lemma 2.2.4 *If f_n and f are cadlag functions on the interval $[a,b] \subset \mathbb{R}$ such that (i) f_n and f are monotone and f is bounded, (ii) $f_n(x) \to f(x)$ for all $x \in [a,b]$ (or $x \in [a,b)$) and (iii) $f_n(x-) \to f(x-)$ for all $x \in [a,b]$, then*

$$\|f_n - f\|_\infty \to 0$$

for $n \to \infty$. ($\|f\|_\infty$ can be defined as $\sup_{x\in[a,b]} |f(x)|$ or $\sup_{x\in[a,b)} |f(x)|$).

PROOF: let $\epsilon > 0$ be given. Because f is bounded and monotone, f jumps in (a,b) at most at a finite number of points, where the height of the jump is bigger than ϵ. Let this number be m and if $m \neq 0$ let these points in (a,b) be $x_1 < x_2 < \ldots < x_m$. We define $x_0 \equiv a$ and $x_{m+1} \equiv b$. Of course x_0 and x_{m+1} can also be points where f jumps.

Without loss of generality we assume that f is monotone increasing. Because of monotonicity there are numbers $k_i = 2, 3, \ldots$ $(i = 1, \ldots, m+1)$ such that $k_i = 2, 3, \ldots$ is the smallest number for which

$$f(x_{i-1}) + (k_i - 1) \cdot \epsilon \geq f(x_i-)$$

holds. We construct the sequence $y_{ij} \in [a,b]$ $(i = 1, \ldots, m+1\,;\, j = 1, \ldots, k_i)$ such that

$$y_{i,1} = x_{i-1} \quad \text{and} \quad y_{i,k_i} = x_i$$

and for $j = 2, \ldots, k_i - 1$ we try to find that $z_j \in (x_{i-1}, x_i)$ such that

$$f(z_j) = f(x_{i-1}) + (j-1) \cdot \epsilon.$$

If such z_j exists, then we take $y_{i,j} = z_j$ and if there is not such z_j, then f makes a jump at some $s_j \in (x_{i-1}, x_i)$ and we take $y_{i,j} = s_j$. Because of monotonicity we note that the sequence $y_{i,j}$ can be constructed in such a way that $y_{i,j} < y_{i+1,j'}$ $(j = 1, \ldots, k_i\,;\, j' = 2, \ldots, k_{i+1})$ and $y_{i,k_i} = y_{i+1,1}$ and $y_{i,j} < y_{i,j+1}$. Now we have

$$|f(y_{i,j}) - f(y_{i,j+1}-)| \leq \epsilon. \tag{2.5}$$

Because of (ii) and (iii) we have that for all $i = 1, \ldots, m+1$ and for all $j = 1, \ldots, k_i - 1$ there exist $M_{i,j}$ such that for all $n \geq M_{i,j}$

$$|f_n(y_{i,j}) - f(y_{i,j})| \leq \epsilon \quad \text{and} \quad |f_n(y_{i,j+1}-) - f(y_{i,j+1}-)| \leq \epsilon. \tag{2.6}$$

Because of the monotonicity, (2.5) and (2.6) we get for all $x \in [y_{i,j}, y_{i,j+1})$ that for all $n \geq M_{i,j}$

$$|f_n(x) - f(x)| \leq 2\epsilon. \tag{2.7}$$

If we define

$$M_{i,\cdot} \equiv \max_{j=1,\ldots,k_i-1} M_{i,j} \quad \text{and} \quad M_{\cdot,\cdot} \equiv \max_{i=1,\ldots,m+1} M_{i,\cdot},$$

and note that the intervals $[y_{i,j}, y_{i,j+1})$ are a partition of $[a, b)$, then with (2.7) we have for all $n \geq M_{\cdot,\cdot}$ that for all $x \in [a, b)$: $|f_n(x) - f(x)| \leq 2\epsilon$. This yields for all $n \geq M_{\cdot,\cdot}$ that

$$\sup_{x\in[a,b)} |f_n(x) - f(x)| \leq 2\epsilon.$$

If we have that $f_n(b) \to f(b)$, then it is easy to take this into account to get the same result for $\sup_{x\in[a,b]} |f_n(x) - f(x)|$. This proves the uniform convergence of f_n to f on $[a, b)$ (or $[a, b]$). □

After this intermezzo we will formulate the consistency results for the one-dimensional line segment process of section 1.1.

Let $\tau_0 \in [0, \tau]$ be such that

$$V((\tau_0, \tau)) = 0 \tag{2.8}$$

and

$$V((\tau_0 - \epsilon, \tau_0]) > 0 \ , \quad \text{for all } \epsilon \in (0, \tau_0]. \tag{2.9}$$

We distinguish the following three cases:

Case I : for the underlying V we have $\tau_0 = \tau$,
Case II : for the underlying V we have $\tau_0 < \tau$ and $V([\tau, \infty)) > 0$,
Case III : for the underlying V we have $\tau_0 < \tau$ and $V([\tau, \infty)) = 0$.

One notes that the three cases cover all possibilities for V. To prove the consistency results we want to use theorem 2.1.1. Unfortunately, in case II we can not apply theorem 2.1.1 for the NPMLE at the moment (and indeed consistency is still an open problem), but for the *sieved* NPMLE we can. For case I and III we obtain the consistency results for the NPMLE and the *sieved* NPMLE. We state here the following theorem:

Theorem 2.2.5 *For the sieved NPMLE (and also for the NPMLE in case I and III)* $(\widehat{V}_n, \widehat{h}_n)$ *of* $(V, h) \in \mathcal{V}_\tau$ *in the one-dimensional case we have that*

$$\widehat{h}_n - h \to 0 \quad a.s.$$

and

$$\sup_{x\in[0,\tau-\epsilon]} \left|\widehat{V}_n(x) - V(x)\right| \to 0 \quad a.s.$$

for all $\epsilon \in (0, \tau]$.

The theorem is proved in section 2.2.1. Because the sieved NPMLE in case II and the (sieved) NPMLE in case III are constant on (τ_0, τ), theorem 2.2.5 implies

Corollary 2.2.6 *For the sieved NPMLE in case II and the (sieved) NPMLE in case III we have*

$$\sup_{x\in[0,\tau)} \left|\widehat{V}_n(x) - V(x)\right| \to 0 \quad a.s.$$

We remember the 1–1 correspondence between $(V, h) \in \mathcal{V}_\tau$ and $(F, \mu) \in \mathcal{F}_\tau$ in section 1.1.3. There we had

$$\begin{aligned} &\nu(V(\cdot), h) \\ \equiv\ & \frac{1}{2\tau}(1 - h - V(\tau-)) + \int_0^\tau \frac{1}{\tau + w}\, \mathrm{d}V(w) \\ =\ & \frac{1}{2\tau}(1 - h - V(\tau-)) + \frac{1}{2\tau}V(\tau-) + \int_0^\tau \frac{1}{(\tau + w)^2} V(w)\, \mathrm{d}w \\ =\ & \frac{1}{2\tau}(1 - h) + \int_0^\tau \frac{1}{(\tau + w)^2} V(w)\, \mathrm{d}w. \end{aligned}$$

Theorem 2.2.5 gives us $\widehat{V}_n(x) \to V(x)$ for all $x \in [0, \tau)$. Then by Lesbesgue's dominated convergence theorem (or use lemma 2.2.1) we obtain that

$$\int_0^\tau \frac{1}{(\tau + w)^2} \widehat{V}_n(w)\, \mathrm{d}w \to \int_0^\tau \frac{1}{(\tau + w)^2} V(w)\, \mathrm{d}w$$

and so this yields $\nu(\widehat{V}_n(\cdot), \widehat{h}_n) \to \nu(V(\cdot), h)$. Now we have proved consistency of $\widehat{\mu}_n$:

$$\widehat{\mu}_n = \left(\frac{1}{\nu(\widehat{V}_n(\cdot), \widehat{h}_n)} - \tau\right) \to \left(\frac{1}{\nu(V(\cdot), h)} - \tau\right) = \mu.$$

Together with

$$\begin{aligned} F(x) &= \frac{1}{\nu(V(\cdot), h)} \int_0^x \frac{1}{\tau + w}\, \mathrm{d}V(w) \\ &= (\tau + \mu) \left(\frac{1}{\tau + x} V(x) + \int_0^x \frac{1}{(\tau + w)^2} V(w)\, \mathrm{d}w\right) \end{aligned}$$

and theorem 2.2.5 we obtain easily the following consistency results for the parameters of interest (F, μ), which are formulated in the following theorem. (A corollary as 2.2.6 can also be obtained for F).

Theorem 2.2.7 *For the sieved NPMLE (and also for the NPMLE in case I and III) $(\widehat{F}_n, \widehat{\mu}_n)$ of $(F, \mu) \in \mathcal{F}_\tau$ in the one-dimensional case we have that*

$$\widehat{\mu}_n - \mu \to 0 \quad a.s.$$

and

$$\sup_{x\in[0,\tau-\epsilon]} \left|\widehat{F}_n(x) - F(x)\right| \to 0 \quad a.s.$$

for all $\epsilon \in (0, \tau]$.

2.2.1 Proof of consistency in the one-dimensional case

Proof of theorem 2.2.5: let τ_0 be as in (2.8) and (2.9). At the moment we do not need to discriminate between the three cases for the underlying V and between the NPMLE and the sieved NPMLE. We will point out the moments when it is necessary to distinguish between the cases.

By the Glivenko-Cantelli theorem, we have that

$$\|F_n^{d.c.} - F^{d.c.}\|_{[0,\tau]} \to 0, \ \|F_n^{s.e.c.} - F^{s.e.c.}\|_{[0,\tau]} \to 0, \ \|F_n^{u.c.} - F^{u.c.}\|_{[0,\tau]} \to 0 \qquad (2.10)$$

a.s. ($n \to \infty$) (see (1.10) - (1.13) or (1.4) - (1.7) and (1.36) - (1.38)), where $\|\cdot\|_I$ stands for the supremum norm on the interval I. With (1.51) we note that

$$F_n^{d.c.}(\tau) = \widehat{h}_n = \frac{n-r}{n} \qquad (2.11)$$

(the relative frequency of double censored observations) and we immediately see that

$$|\widehat{h}_n - h| \to 0 \text{ a.s. } (n \to \infty). \qquad (2.12)$$

In order to apply theorem 2.1.1 we start by introducing the elements that play the roles of T, T_n, $\widehat{T}_n$ and T_∞ in the theorem (or in section 2.1). Of course (V, h) plays the role of T.

- **The role of T_n:** we define

$$\mathrm{d}V_n(x) \equiv \frac{\tau + x}{\tau - x}\,\mathrm{d}F_n^{u.c.}(x) \quad , \quad h_n \equiv \widehat{h}_n \qquad (2.13)$$

and $g_n(\tau)$ is defined by $V_n(\tau-) + 2\tau g_n(\tau) + h_n = 1$ and

$$g_n(x) \equiv \int_x^\tau \frac{1}{\tau + w}\,\mathrm{d}V_n(w) + g_n(\tau) \qquad (2.14)$$

$$= \frac{1-\widehat{h}_n}{2\tau} + \frac{1}{2\tau}\left((F_n^{u.c.}(\tau) - F_n^{u.c.}(x)) - \int_0^x \frac{\tau + w}{\tau - w}\,\mathrm{d}F_n^{u.c.}(w)\right). \qquad (2.15)$$

Because of monotonicity (lemma 2.2.4) and the strong law of large numbers, one proves that

$$\left\|\int_0^x \frac{\tau + w}{\tau - w}\,\mathrm{d}F_n^{u.c.}(w) - \int_0^x \frac{\tau + w}{\tau - w}\,\mathrm{d}F^{u.c.}(w)\right\|_{[0,\tau]} \to 0$$

a.s. ($n \to \infty$). So together with (2.10) and (2.12) we conclude

$$\|g_n - g\|_{[0,\tau)} \to 0 \ , \ \|V_n - V\|_{[0,\tau)} \to 0 \quad \text{a.s. } (n \to \infty). \qquad (2.16)$$

Furthermore one notes that g_n follows from (V_n, h_n) according to (1.40).

- **The role of $\widehat{T}_n$:** again let $(\widehat{V}_n, \widehat{h}_n)$ denote the (sieved) NPMLE of (V, h) in $\mathcal{V}_\tau$. (V, h) is the image of the underlying distribution function V on $[0, \infty)$ under the map described in section 2.3. For the function $g(x) \equiv \int_x^\infty 1/(\tau + w)\,\mathrm{d}V(w)$ we obtain the estimator $\widehat{g}_n$ on $[0, \tau]$

$$\widehat{g}_n(x) = \int_x^\tau \frac{1}{\tau + w}\,\mathrm{d}\widehat{V}_n(w) + \widehat{g}_n(\tau),$$

where $\widehat{g}_n(\tau)$ is defined by $\widehat{V}_n(\tau-) + 2\tau\widehat{g}_n(\tau) + \widehat{h}_n = 1$. Here we have again that $\widehat{g}_n$ follows from $(\widehat{V}_n, \widehat{h}_n)$ according to (1.40).

• **The role of T_∞**: according to (2.10) we have

$$\|F_n^{\cdots}(\cdot,\omega) - F^{\cdots}(\cdot,\omega)\|_{[0,\tau]} \to 0$$

with probability 1 ('... = u.c., s.e.c. or d.c.'). By the Helly selection theorem we have that for some subsequence $\widehat{V}_{n_k}$ of the sequence of functions $\widehat{V}_n$ on $[0,\tau)$, there exists a nondecreasing right continuous function V_∞ on $[0,\tau)$ such that $\lim_{k\to\infty} \widehat{V}_{n_k}(x) = V_\infty(x)$ at the continuity points of V_∞. Of course we have $\widehat{h}_{n_k} \to h_\infty \equiv h$. One defines

$$g_\infty(x) \equiv \int_x^\tau \frac{1}{\tau+w}\,\mathrm{d}V_\infty(w) + g_\infty(\tau)$$

where $g_\infty(\tau)$ is defined by $V_\infty(\tau) + 2\tau g_\infty(\tau) + h_\infty = 1$. We see that

$$\begin{aligned} \widehat{g}_{n_k}(x) &= \int_x^\tau \frac{1}{\tau+w}\,\mathrm{d}\widehat{V}_{n_k}(w) + \widehat{g}_{n_k}(\tau) \\ &= \frac{1-\widehat{h}_{n_k}}{2\tau} - \frac{1}{\tau+x}\widehat{V}_{n_k}(x) + \int_x^\tau \frac{1}{(\tau+w)^2}\widehat{V}_{n_k}(w)\,\mathrm{d}w \end{aligned}$$

and by the weak convergence of $(\widehat{V}_{n_k}, \widehat{h}_{n_k})$ to $(V_\infty, h_\infty = h)$ we get $\lim_{k\to\infty} \widehat{g}_{n_k}(x) = g_\infty(x)$ at the continuity points of V_∞ (or g_∞). One verifies immediately that g_∞ follows from (V_∞, h_∞) according to (1.40).

Since $(\widehat{V}_n, \widehat{h}_n)$ is the (sieved) NPMLE we find by (2.2) and (1.4)–(1.7) that

$$\int_0^\tau \frac{\mathrm{d}V_n}{\mathrm{d}\widehat{V}_n}(x)\,\mathrm{d}F_n^{u.c.}(x) + \int_0^\tau \frac{g_n(x)}{\widehat{g}_n(x)}\,\mathrm{d}F_n^{s.e.c.}(x) + \widehat{h}_n \le 1. \tag{2.17}$$

Following the general consistency proof in section 2.1 (see theorem 2.1.1), we want to show that this holds in the limit, possibly after passing to a subsequence. The subsequence $(\widehat{V}_{n_k}, h_{n_k})$, we got by the Helly selection theorem, will do. To prove that (2.17) holds in the limit, we deal in the following three paragraphs with each case for the underlying V separately. We need to control the possible unboundedness of $1/\widehat{g}_n(x)$.

• **Case I**: $\tau_0 = \tau$. We consider the (sieved) NPMLE $(\widehat{V}_n, \widehat{h}_n)$ of (V, h). For each $a \in [0,\tau)$ we can derive the following bound for $1/\widehat{g}_n$ and $1/g_\infty$. The estimator $\widehat{V}_n$ gives at least mass $1/n$ to uncensored observations, so with (2.9) we get for n large enough

$$\frac{1}{\widehat{g}_n(x)} \le M_a\,, \quad \frac{1}{g_\infty(x)} \le M_a \tag{2.18}$$

for all points $x \in [0,a]$, where $M_a \ge 0$ is a constant.

• **Case II**: $\tau_0 < \tau$ and $V([\tau,\infty)) > 0$. We consider the *sieved* NPMLE $(\widehat{V}_n, \widehat{h}_n)$ of (V, h). We will derive a bound for $1/\widehat{g}_n$ and $1/g_\infty$ on $[0,\tau]$. The sieved estimator $\widehat{V}_n$ gives only mass to uncensored observations, so (because of (2.8)) $\widehat{g}_n(x)$ is constant on $(\tau_0, \tau]$. The same holds for V_n and g_n. Because the integrands in (2.17) are nonnegative we also have

$$\int_{\tau_0}^{\tau} \frac{g_n(x)}{\widehat{g}_n(x)} \, \mathrm{d}F_n^{s.e.c.}(x) \leq 1$$

and thus we can write

$$g_n(\tau)\,(F_n^{s.e.c.}(\tau) - F_n^{s.e.c.}(\tau_0)) \leq \widehat{g}_n(\tau).$$

We know that the left-hand side of this inequality converges to

$$g(\tau)\,(F^{s.e.c.}(\tau) - F^{s.e.c.}(\tau_0)) = g(\tau) \cdot 2 \int_{\tau_0}^{\tau} g(x)\, \mathrm{d}x.$$

Because of $V([\tau,\infty)) > 0$ this is strictly positive and thus we may conclude that there exists a constant $c > 0$ such that $g_\infty(\tau) \geq c$ and for n large enough we have $\widehat{g}_n(\tau) \geq c$. This and the fact that $g_\infty(\cdot)$ and all $\widehat{g}_n(\cdot)$ are decreasing imply that for n large enough we get

$$\frac{1}{\widehat{g}_n(x)} \leq M\,, \quad \frac{1}{g_\infty(x)} \leq M \tag{2.19}$$

for all points $x \in [0,\tau]$ and $M = (1/c)$. In this case we note that on (τ_0, τ) we have: $\mathrm{d}\widehat{V}_n(x) = \mathrm{d}V_n(x) = \mathrm{d}V_\infty(x) = \mathrm{d}V(x) = 0$ and $\mathrm{d}F_n^{u.c.}(x) = \mathrm{d}F^{u.c.}(x) = 0$.

• **Case III:** $\tau_0 < \tau$ and $V([\tau,\infty)) = 0$. We consider the (sieved) NPMLE $(\widehat{V}_n, \widehat{h}_n)$ of (V,h). For each $a \in [0,\tau_0)$ we can give the following bound on $1/\widehat{g}_n$ and $1/g_\infty$. Again the estimator $\widehat{V}_n$ gives at least mass $1/n$ to uncensored observations, so with (2.9) we get for n large enough

$$\frac{1}{\widehat{g}_n(x)} \leq M_a\,, \quad \frac{1}{g_\infty(x)} \leq M_a \tag{2.20}$$

for all points $x \in [0,a]$, where $M_a \geq 0$ is a constant. In this case we know at the moment that on (τ_0,τ) we have: $\mathrm{d}\widehat{V}_n(x) = \mathrm{d}V_n(x) = \mathrm{d}V_\infty(x) = \mathrm{d}V(x) = 0$ and $g(x) = 0$ and $\mathrm{d}F_n^{u.c.}(x) = \mathrm{d}F^{u.c.}(x) = 0$ and $\mathrm{d}F_n^{s.e.c.}(x) = \mathrm{d}F^{s.e.c.}(x) = 0$ and $\widehat{h}_n = h = 0$. So in this case we can replace the integrals over $(0,\tau)$ in (2.17) by $(0,\tau_0]$.

Using (2.18)–(2.20) and the propositions 2.2.2.1 – 2.2.2.3 we prove in lemma 2.2.2.4 in the next section that (2.17) holds in the limit for all these cases. Here we give a sketch of the proof. First of all we need the limit version of $\mathrm{d}V_n/\mathrm{d}\widehat{V}_n$. One verifies that $V_n \ll \widehat{V}_n$ and by the selfconsistency equation (1.50) we can write

$$\frac{\mathrm{d}V_n}{\mathrm{d}\widehat{V}_n}(x) = \frac{\tau + x}{\tau - x}\left(1 - \frac{1}{\tau + x}\int_{v=0}^{v=x} \frac{1}{\widehat{g}_n(v)}\, \mathrm{d}F_n^{s.e.c.}(v)\right). \tag{2.21}$$

In proposition 2.2.2.1 we present the limit version $\mathrm{d}V/\mathrm{d}V_\infty$ of $\mathrm{d}V_n/\mathrm{d}\widehat{V}_n$. Further we note that the integrands in (2.17) are nonnegative and thus the inequality still holds when we replace the τ's by an $a \in [0,\tau)$. Then in proposition 2.2.2.2 we show (along a subsequence) that

$$\int_0^a \frac{\mathrm{d}V_n}{\mathrm{d}\widehat{V}_n}(x)\, \mathrm{d}F_n^{u.c.}(x) \to \int_0^a \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x)\, \mathrm{d}F^{u.c.}(x) \tag{2.22}$$

for all a in $[0,\tau)$ in case I and II and in $[0,\tau_0)$ in case III. Thereupon we prove in proposition 2.2.2.3 (along a subsequence) that

$$\int_0^a \frac{g_n(x)}{\widehat{g}_n(x)}\,\mathrm{d}F_n^{s.e.c.}(x) \to \int_0^a \frac{g(x)}{g_\infty(x)}\,\mathrm{d}F^{s.e.c.}(x) \tag{2.23}$$

for all a in $[0,\tau)$ in case I and II and in $[0,\tau_0)$ in case III. In both proofs we need the boundedness statements of $1/\widehat{g}_n$ and $1/g_\infty$ on $[0,a]$ in (2.18)–(2.20). Because inequality (2.17) still holds, when we replace the τ's by an $a \in [0,\tau)$, we get by these convergence properties (2.22) and (2.23) a limit version (along a subsequence) for (2.17) with the τ's replaced by a. In this limit version we let a tend to τ in case I and II and to τ_0 in case III and by monotone convergence we finally prove that (2.17) holds in the limit. The details are worked out in the proof of lemma 2.2.2.4.

Having this result of lemma 2.2.2.4 (thus the condition of theorem 2.1.1 is checked), we are almost finished. For each case we only have to gather the obtained outcome and draw the corresponding conclusion.

- **Cases I and II**: by lemma 2.2.2.4 we find in case I and II that (2.17) holds in the limit:

$$\int_0^\tau \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x)\,\mathrm{d}F^{u.c.}(x) + \int_0^\tau \frac{g(x)}{g_\infty(x)}\,\mathrm{d}F^{s.e.c.}(x) + h \le 1.$$

By theorem 2.1.1 this implies that we have $V_\infty(x) = V(x)$ on $[0,\tau)$ and $h_\infty = h$ (identifiability: $P_{V_\infty} = P_V \iff V_\infty(x) = V(x)$ on $[0,\tau)$, $h_\infty = h$). This proves that outside a set of probability 0, each subsequence of the sequence of functions $\widehat{V}_n$ (or $\widehat{g}_n$) has a weakly convergent subsubsequence and all these convergent subsubsequences have the same limit V (or g). So $\widehat{V}_n$ converges weakly to $V_\infty = V$ along the whole sequence (at the continuity points).

If V is **continuous** on $[0,\tau)$, then all $x \in [0,\tau)$ are continuity points of $V_\infty = V$. In this case we have that $\widehat{V}_{n_k}(x)$ converges to $V_\infty(x) = V(x)$ for all $x \in [0,\tau)$.

If V is **discontinuous** on $[0,\tau)$, then of course we have that $\widehat{V}_{n_k}(x)$ converges to $V_\infty(x) = V(x)$ (and thus $\widehat{g}_{n_k}(x)$ converges to $g_\infty(x) = g(x)$) at all continuity points $x \in [0,\tau)$ of V. By (2.21) and (2.13) we have

$$\widehat{V}_n(x) = \int_{u=0}^{u=x} \left(1 - \frac{1}{\tau+u}\int_{v=0}^{v=u} \frac{1}{\widehat{g}_n(v)}\,\mathrm{d}F_n^{s.e.c.}(v)\right)^{-1} \mathrm{d}F_n^{u.c.}(u).$$

In the proof of proposition 2.2.2.2 we obtain (2.35) on $[0,x]$ (for all $x \in [0,\tau)$). Because we know here that we have $g_\infty = g$ at the continuity points and in S_∞ we integrate w.r.t. Lebesgue measure, we get for the S_∞ in (2.34): $S_\infty(x) = (\tau - x)/(\tau + x)$. By (2.35) we have $\|S_n^{-1} - S_\infty^{-1}\|_\infty \to 0$ on $[0,x]$. To show that S_n^{-1} is of bounded variation uniformly in n (on $[0,x]$), one uses the same argumentation as in the proof of proposition 2.2.2.2, where we do this for S_n. We write $\widehat{V}_n(x) = \int_0^x S_n^{-1}(u)\,\mathrm{d}F_n^{u.c.}(u)$ and apply lemma 2.2.3. This gives us the fact that $\widehat{V}_n(x-)$ converges to $V(x-) = \int_0^{x-} S_\infty^{-1}(u)\,\mathrm{d}F^{u.c.}(u)$ for all $x \in [0,\tau)$. Note that we did not prove that $\widehat{V}_n(\tau-)$ converges to $V(\tau-)$ in case I.

For the continuous as well for the discontinuous case we are not able to prove that $\widehat{V}_{n_k}(\tau-)$ converges to $V(\tau-)$ in case I, but for each $\epsilon \in (0,\tau]$ we have that on $[0,\tau-\epsilon]$: $\widehat{V}_n(x) \to V(x)$,

$\widehat{V}_n(x-) \to V(x-)$ and the $\widehat{V}_n$ and V are monotone and V is bounded. Now we apply lemma 2.2.4 to obtain $\|\widehat{V}_n - V\|_\infty \to 0$ on $[0, \tau - \epsilon]$, that is the consistency result in theorem 2.2.5.

• **Case III**: by lemma 2.2.2.4 we get that for case III the inequality (2.17) holds in the limit:

$$\int_0^{\tau_0} \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x)\,\mathrm{d}F^{u.c.}(x) + \int_0^{\tau_0} \frac{g(x)}{g_\infty(x)}\,\mathrm{d}F^{s.e.c.}(x) + h \le 1, \tag{2.24}$$

where we remember that with $\int_0^{\tau_0}$ we mean the integral over $(0, \tau_0]$. To get the consistency result of theorem 2.2.5, one copies the proof of the cases I and II above, but now one works on the interval $[0, \tau_0)$ (or on $[0, \tau_0]$ if τ_0 is continuity point of V). To obtain the result for $[0, \tau_0]$ for all V, we have to pay special attention to the point τ_0, if it is a discontinuity point of $V_\infty = V$. If we prove $\widehat{V}_n(\tau_0) \to V_\infty(\tau_0) = V(\tau_0)$, thus if $V_\infty(\tau_0) = V(\tau_0)$, then we are ready.

We know already that $V_\infty(x) = V(x)$ and $g_\infty(x) = g(x)$ on $[0, \tau_0)$. So on $[0, \tau_0)$ we have that $(\mathrm{d}V/\mathrm{d}V_\infty) = 1$ and $g(x)/g_\infty(x) = 1$. We also know that in this case V jumps in τ_0 to 1 and both V and V_∞ are monotone increasing. If V_∞ jumps in τ_0 differently, then the jump is smaller than the jump of V (and g_∞ does not jump to 0 in τ_0). This means that $\Delta V(\tau_0)/\Delta V_\infty(\tau_0) > 1$, where $\Delta f(x)$ stands for the height of the jump of f at x. Of course in this case we have $\Delta F^{u.c.}(\tau_0) = (\tau - \tau_0)/(\tau + \tau_0)\,\Delta V(\tau_0) > 0$. If we use these facts to calculate the left-hand side of inequality (2.24), then we get

$$\begin{aligned} &\int_0^{\tau_0-} \mathrm{d}F^{u.c.}(x) + \frac{\Delta V(\tau_0)}{\Delta V_\infty(\tau_0)} \cdot \frac{\tau - \tau_0}{\tau + \tau_0}\,\Delta V(\tau_0) + \int_0^{\tau_0} \mathrm{d}F^{s.e.c.}(x) + h \\ = \;& F^{u.c.}(\tau_0-) + \frac{\Delta V(\tau_0)}{\Delta V_\infty(\tau_0)} \cdot \frac{\tau - \tau_0}{\tau + \tau_0}\,\Delta V(\tau_0) + F^{s.e.c.}(\tau_0) + h. \end{aligned}$$

All terms are positive. If $\Delta V(\tau_0)/\Delta V_\infty(\tau_0) = 1$, then the left-hand side of (2.24) would become $F^{u.c.}(\tau_0) + F^{s.e.c.}(\tau_0) + h = 1$. If $\Delta V(\tau_0)/\Delta V_\infty(\tau_0) > 1$, then the left-hand side will be > 1. This contradicts the inequality (2.24). So $V_\infty(\tau_0)$ must equal $V(\tau_0)$ and this completes the proof for case III.

We have proved the theorem. □

2.2.2 Three propositions and a lemma

In this section one finds the propositions 2.2.2.1 – 2.2.2.3 and lemma 2.2.2.4 to which we refer in the proof of theorem 2.2.5.

Proposition 2.2.2.1 *On $[0, \tau)$ in case I and II and on $[0, \tau_0)$ in case III, we have $V \ll V_\infty$ and*

$$\frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x) = \frac{\tau + x}{\tau - x}\left(1 - \frac{1}{\tau + x}\int_{v=0}^{v=x} \frac{1}{g_\infty(v)}\,\mathrm{d}F^{s.e.c.}(v)\right),$$

where $\mathrm{d}F^{s.e.c.}(v) = 2g(v)\mathrm{d}v$ (see (1.4), (1.7)).

PROOF: here for all three cases we prove that proposition 2.2.2.1 holds. In the following derivation we get the first equality by telescoping. The second equality is obtained by integration by parts of the third term and the integrand of the second term. We start by writing

$$
\begin{aligned}
&\int_{u=0}^{u=x}\int_{v=0}^{v=u}\frac{1}{\widehat{g}_{n_k}(v)}\,\mathrm{d}F_{n_k}^{s.e.c.}(v)\cdot\frac{1}{\tau+u}\,\mathrm{d}\widehat{V}_{n_k}(u)\\
&\quad-\int_{u=0}^{u=x}\int_{v=0}^{v=u}\frac{1}{g_\infty(v)}\,\mathrm{d}F^{s.e.c.}(v)\cdot\frac{1}{\tau+u}\,\mathrm{d}V_\infty(u)\\
&=\int_{u=0}^{u=x}\int_{v=0}^{v=u}\left(\frac{1}{\widehat{g}_{n_k}(v)}-\frac{1}{g_\infty(v)}\right)\mathrm{d}F^{s.e.c.}(v)\cdot\frac{1}{\tau+u}\,\mathrm{d}V_\infty(u)\\
&\quad+\int_{u=0}^{u=x}\int_{v=0}^{v=u}\frac{1}{\widehat{g}_{n_k}(v)}\,\mathrm{d}\left(F_{n_k}^{s.e.c.}(v)-F^{s.e.c.}(v)\right)\cdot\frac{1}{\tau+u}\,\mathrm{d}V_\infty(u)\\
&\quad+\int_{u=0}^{u=x}\int_{v=0}^{v=u}\frac{1}{\widehat{g}_{n_k}(v)}\,\mathrm{d}F_{n_k}^{s.e.c.}(v)\cdot\frac{1}{\tau+u}\,\mathrm{d}\left(\widehat{V}_{n_k}(u)-V_\infty(u)\right)\\
&=\int_{u=0}^{u=x}\int_{v=0}^{v=u}\left(\frac{1}{\widehat{g}_{n_k}(v)}-\frac{1}{g_\infty(v)}\right)\mathrm{d}F^{s.e.c.}(v)\cdot\frac{1}{\tau+u}\,\mathrm{d}V_\infty(u)\\
&\quad+\int_{u=0}^{u=x}\frac{1}{\widehat{g}_{n_k}(u)}\left(F_{n_k}^{s.e.c.}(u)-F^{s.e.c.}(u)\right)\cdot\frac{1}{\tau+u}\,\mathrm{d}V_\infty(u)\\
&\quad-\int_{u=0}^{u=x}\int_{v=0}^{v=u}\left(F_{n_k}^{s.e.c.}(v)-F^{s.e.c.}(v)\right)\mathrm{d}\left(\frac{1}{\widehat{g}_{n_k}(v)}\right)\cdot\frac{1}{\tau+u}\,\mathrm{d}V_\infty(u)\\
&\quad+\frac{1}{\tau+x}\left(\widehat{V}_{n_k}(x)-V_\infty(x)\right)\int_{v=0}^{v=x}\frac{1}{\widehat{g}_{n_k}(v)}\,\mathrm{d}F_{n_k}^{s.e.c.}(v)\\
&\quad+\int_{u=0}^{u=x}\left(\widehat{V}_{n_k}(u)-V_\infty(u)\right)\frac{1}{(\tau+u)^2}\int_{v=0}^{v=u}\frac{1}{\widehat{g}_{n_k}(v)}\,\mathrm{d}F_{n_k}^{s.e.c.}(v)\,\mathrm{d}u\\
&\quad-\int_{u=0}^{u=x}\left(\widehat{V}_{n_k}(u)-V_\infty(u)\right)\frac{1}{\tau+u}\cdot\frac{1}{\widehat{g}_{n_k}(u)}\,\mathrm{d}F_{n_k}^{s.e.c.}(u).
\end{aligned}
\tag{2.25}
$$

Now we explain that all these terms tend to 0 if $k\to\infty$. Let x be a continuity point of V_∞ in $[0,\tau)$ in case I and II and in $[0,\tau_0)$ in case III.

• **The first term**: we define on $[0,x]$

$$W_{n_k}(u)\equiv\int_{v=0}^{v=u}\frac{1}{\widehat{g}_{n_k}(v)}\,\mathrm{d}F^{s.e.c.}(v)\tag{2.26}$$

and

$$W_\infty(u)\equiv\int_{v=0}^{v=u}\frac{1}{g_\infty(v)}\,\mathrm{d}F^{s.e.c.}(v).\tag{2.27}$$

To $W_{n_k}(u)-W_\infty(u)$ we apply lemma 2.2.1: $G_{n_k}=((1/\widehat{g}_{n_k})-(1/g_\infty))$, $G\equiv 0$, $H_n=H=F^{s.e.c.}$, $a=0$ an $b=u$. Now (i) follows from the fact that $\widehat{g}_{n_k}$ converges weakly to g_∞ on the continuity points of V_∞. That we have (ii) is trivial to see and (iii) follows from (2.18)–(2.20) on $[0,u]$ and the fact that the $1/\widehat{g}_n$'s and $1/g_\infty$ are monotone increasing. (iv) follows from the fact that $\mathrm{d}H(x)/\mathrm{d}x=2g(x)$ and $g(x)$ is bounded.

We can do this for each $u\in[0,x]$ and thus applying the lemma's 2.2.1 and 2.2.3 we get

$$W_{n_k}(u)-W_\infty(u)\to 0\quad\text{and}\quad W_{n_k}(u-)-W_\infty(u-)\to 0$$

for all $u \in [0,x]$. The W_{n_k}'s and W_∞ are monotone increasing and by (2.18)–(2.20) we know that W_∞ is bounded on $[0,x]$. Now lemma 2.2.4 tells us that

$$\|W_{n_k} - W_\infty\|_\infty \to 0 \quad \text{on } [0,x]. \tag{2.28}$$

Finally we conclude that the first term in (2.25) tends to 0 by applying lemma 2.2.2 (along the subsequence): $G_n = W_n - W_\infty$, $G \equiv 0$, $\mathrm{d}H_n(u) = \mathrm{d}H(u) = (1/(\tau+u))\,\mathrm{d}V_\infty(u)$, $a=0$ and $b=x$. By (2.28) we have (i) and (ii) is trivially satisfied. We use (2.18)–(2.20) and the fact that the W_n's and W_∞ are monote to get (iii) on $[0,x]$. Of course V_∞ is of bounded variation and thus we have (iv) too.

• **The second term**: for the second term we apply lemma 2.2.2: $G_n = (1/\widehat{g}_n)\cdot(F_n^{s.e.c.} - F^{s.e.c.})$, $G \equiv 0$, $\mathrm{d}H_n(u) = \mathrm{d}H(u) = (1/(\tau+u))\,\mathrm{d}V_\infty(u)$, $a=0$ and $b=x$. By (2.18)–(2.20) and (2.10) we have (i) and (ii) is again trivially satisfied. Using (2.18)–(2.20) and the fact that the $1/\widehat{g}_n$'s are monotone increasing and the fact that the $F_n^{s.e.c.}$'s and $F^{s.e.c.}$ are of bounded variation, gives us (iii) on $[0,x]$. We have (iv), because V_∞ is of bounded variation.

• **The third term**: because the $1/\widehat{g}_{n_k}$'s and V_∞ are increasing and V_∞ is bounded by 1, the absolute value of the third term can be bounded by

$$\begin{aligned} &\frac{1}{\tau}\left\|F_{n_k}^{s.e.c.} - F^{s.e.c.}\right\|_\infty \cdot \int_{v=0}^{v=x} \mathrm{d}\left(\frac{1}{\widehat{g}_{n_k}(v)}\right) \\ \le\ &\frac{1}{\tau}\left\|F_{n_k}^{s.e.c.} - F^{s.e.c.}\right\|_\infty \cdot \frac{1}{\widehat{g}_{n_k}(x)} \\ \le\ &\frac{1}{\tau} C \left\|F_{n_k}^{s.e.c.} - F^{s.e.c.}\right\|_\infty, \end{aligned}$$

where $C = M_x$ in case I and III and $C = M$ in case II ((2.18)–(2.20)). Now one uses (2.10) to see that this term tend to 0.

• **The fourth term**: because $\int_0^x (1/\widehat{g}_{n_k}(v))\,\mathrm{d}F_{n_k}^{s.e.c.}(v)$ is bounded by $C \cdot F_{n_k}^{s.e.c.}(x) \le C$, where $C = M_x$ in case I and II and $C = M$ in case III ((2.18)–(2.20)) and the fact that $\widehat{V}_{n_k}(x) - V_\infty(x)$ tends to 0 for continuity points x of V_∞, we get immediately that the fourth term tends to 0.

• **The fifth term**: in the fifth term we concentrate first on $\int_0^u (1/\widehat{g}_{n_k}(v))\,\mathrm{d}F_{n_k}^{s.e.c.}(v)$. To this integral we apply lemma 2.2.1 with $G_{n_k} = (1/\widehat{g}_{n_k})$, $G = (1/g)$, $H_n = F_n^{s.e.c.}$, $H = F^{s.e.c.}$, $a=0$ and $b=u$, where $u \in [0,x]$. Again by (2.18)–(2.20) on $[0,x]$ and the fact that the $1/\widehat{g}_n$'s are monotone increasing ,we get that G_n is of bounded variation uniformly in n. So we have that this integral converges weakly to $\int_0^u (1/g(v))\,\mathrm{d}F^{s.e.c.}(v)$ on $[0,x]$. Together with the fact that $\widehat{V}_{n_k} - V_\infty$ converges weakly to 0, we apply Lebesgue's dominated convergence theorem (or lemma 2.2.1) to the fifth term and conclude that this term tends to 0 too.

• **The sixth term**: for the sixth term we use lemma 2.2.1 again. We take $G_{n_k}(u) = (1/(\tau+u)\widehat{g}_{n_k}(u))\,(\widehat{V}_{n_k}(u) - V_\infty(u))$, $G \equiv 0$, $H_n = F_n^{s.e.c.}$ and $H = F^{s.e.c.}$, $a=0$ and $b=x$. Using the same arguments as above one sees that (i)–(iv) are satisfied.

We note here that the only place where we need that x is a continuity point of V_∞, is the sixth term. In the sixth term it is because of the factor $(\widehat{V}_{n_k}(x) - V_\infty(x))$. The fact that we only know that $\widehat{V}_{n_k}$ converges to V_∞ (and thus $\widehat{g}_{n_k}$ to g_∞) on the continuity points of V_∞, plays no role in the other terms.

Now with (2.25) we have that

$$\begin{aligned} &F^{u.c.}_{n_k}(x) - F^{u.c.}(x) \\ &+ \int_{u=0}^{u=x} \int_{v=0}^{v=u} \frac{1}{\widehat{g}_{n_k}(v)} \, \mathrm{d}F^{s.e.c.}_{n_k}(v) \cdot \frac{1}{\tau+u} \, \mathrm{d}\widehat{V}_{n_k}(u) \\ &- \int_{u=0}^{u=x} \int_{v=0}^{v=u} \frac{1}{g_\infty(v)} \, \mathrm{d}F^{s.e.c.}(v) \cdot \frac{1}{\tau+u} \, \mathrm{d}V_\infty(u) \to 0 \end{aligned} \tag{2.29}$$

for all continuity points x of V_∞ in $[0,\tau)$ in case I and II and in $[0,\tau_0)$ in case III. Furthermore by (1.50) we have

$$\widehat{V}_n(x) = F^{u.c.}_n(x) + \int_{u=0}^{u=x} \int_{v=0}^{v=u} \frac{1}{\widehat{g}_n(v)} \, \mathrm{d}F^{s.e.c.}_n(v) \cdot \frac{1}{\tau+u} \, \mathrm{d}\widehat{V}_n(u). \tag{2.30}$$

Together with (2.29) and the fact that $\widehat{V}_{n_k}$ converges weakly to V_∞ in all the continuity points x of V_∞, we obtain

$$V_\infty(x) = F^{u.c.}(x) + \int_{u=0}^{u=x} \int_{v=0}^{v=u} \frac{1}{g_\infty(v)} \, \mathrm{d}F^{s.e.c.}(v) \frac{1}{\tau+u} \, \mathrm{d}V_\infty(u) \tag{2.31}$$

at the continuity points x of V_∞. Because V_∞ is cadlag we also have the relation (2.31) for $x-$ and $x+$ and thus for all x in $[0,\tau)$ in case I and II and in $[0,\tau_0)$ in case III. Finally with (2.31) and $\mathrm{d}F^{u.c.}(x) = ((\tau-x)/(\tau+x))\,\mathrm{d}V(x)$ we find

$$\frac{\tau-x}{\tau+x} \, \mathrm{d}V(x) = \left(1 - \frac{1}{\tau+x} \int_{v=0}^{v=x} \frac{1}{g_\infty(v)} \, \mathrm{d}F^{s.e.c.}(v)\right) \mathrm{d}V_\infty.$$

This means that on $[0,\tau)$ in case I and II and on $[0,\tau_0)$ in case III we have $V \ll V_\infty$ and proposition 2.2.2.1 holds. □

Proposition 2.2.2.2

$$\int_0^a \frac{\mathrm{d}V_{n_k}}{\mathrm{d}\widehat{V}_{n_k}}(x) \cdot \mathrm{d}F^{u.c.}_{n_k}(x) \to \int_0^a \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x) \cdot \mathrm{d}F^{u.c.}(x) \qquad (k \to \infty)$$

for all a in $[0,\tau)$ in case I and II and in $[0,\tau_0)$ in case III.

Proof: let a be in $[0,\tau)$ in case I and II and in $[0,\tau_0)$ in case III. With (2.21) and proposition 2.2.2.1 we write

$$\begin{aligned} &\int_0^a \frac{\mathrm{d}V_{n_k}}{\mathrm{d}\widehat{V}_{n_k}}(x) \cdot \mathrm{d}F^{u.c.}_{n_k}(x) - \int_0^a \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x) \cdot \mathrm{d}F^{u.c.}(x) \\ &= \int_0^a \left(1 - \frac{1}{\tau+x} \int_{v=0}^{v=x} \frac{1}{\widehat{g}_{n_k}(v)} \, \mathrm{d}F^{s.e.c.}_{n_k}(v)\right) \mathrm{d}V_{n_k}(x) \\ &\quad - \int_0^a \left(1 - \frac{1}{\tau+x} \int_{v=0}^{v=x} \frac{1}{g_\infty(v)} \, \mathrm{d}F^{s.e.c.}(v)\right) \mathrm{d}V(x). \end{aligned} \tag{2.32}$$

We define on $[0,a]$ the following two functions:

$$S_{n_k}(x) \equiv 1 - \frac{1}{\tau + x} W_{n_k}(x) \tag{2.33}$$

$$S_\infty(x) \equiv 1 - \frac{1}{\tau + x} W_\infty(x), \tag{2.34}$$

where W_{n_k} and W_∞ are defined as in (2.26) and (2.27). One sees that (2.33) is the integrand w.r.t. $\mathrm{d}V_{n_k}(x)$ of the first term in (2.32) and (2.34) is the integrand w.r.t. $\mathrm{d}V(x)$ of the second term.

We apply lemma 2.2.2 to prove that (2.32) tends to 0: $G_n = S_n$, $G = S_\infty$, $\mathrm{d}H_n(x) = \mathrm{d}V_n(x)$, $\mathrm{d}H(x) = \mathrm{d}V(x)$. By (2.28) we get immediately that

$$\|S_{n_k} - S_\infty\|_\infty \to 0 \tag{2.35}$$

on $[0,a]$, thus (i) holds. By (2.16) we have that $\|V_n - V\|_\infty \to 0$ and thus we have that (ii) is satisfied. For (iii) one uses (2.18)–(2.20) and one verifies that the S_n's are monotone decreasing and thus one easily sees that S_n is of bounded variation uniformly in n. Of course we have that V is of bounded variation and so we get (iv). Using lemma 2.2.2 along the subsequence, we have proved proposition 2.2.2.2. □

Proposition 2.2.2.3

$$\int_0^a \frac{g_{n_k}(x)}{\widehat{g}_{n_k}(x)} \mathrm{d}F_{n_k}^{s.e.c.}(x) \to \int_0^a \frac{g(x)}{g_\infty(x)} \mathrm{d}F^{s.e.c.}(x) \qquad (k \to \infty)$$

for all a in $[0,\tau)$ in case I and II and in $[0,\tau_0)$ in case III.

PROOF: in order to prove proposition 2.2.2.3, we write (using telescoping)

$$\begin{aligned}&\int_0^a \frac{g_{n_k}(x)}{\widehat{g}_{n_k}(x)} \mathrm{d}F_{n_k}^{s.e.c.}(x) - \int_0^a \frac{g(x)}{g_\infty(x)} \mathrm{d}F^{s.e.c.}(x) \\ &= \int_0^a \left(\frac{g_{n_k}(x)}{\widehat{g}_{n_k}(x)} - \frac{g(x)}{g_\infty(x)}\right) \mathrm{d}F^{s.e.c.}(x) \\ &\quad + \int_0^a \frac{g_{n_k}(x)}{\widehat{g}_{n_k}(x)} \mathrm{d}\left(F_{n_k}^{s.e.c.}(x) - F^{s.e.c.}(x)\right),\end{aligned}$$

where a is in $[0,\tau)$ in case I and II and in $[0,\tau_0)$ in case III. Again, the two terms tend to 0 if $k \to \infty$. For the first term we use Lebesgue's dominated convergence theorem: $g_n(x) \to g(x)$ and $\widehat{g}_{n_k}(x) \to g_\infty(x)$ at the continuity points x of V_∞, on $[0,a]$ we have (2.18)–(2.20) and $\mathrm{d}F^{s.e.c.}(x) = 2g(x)\mathrm{d}x$ is a continuous measure.

To the second term we apply lemma 2.2.1: $G_{n_k} = (g_{n_k}/\widehat{g}_{n_k})$, $G = (g/g_\infty)$, $H_n = F_n^{s.e.c.} - F^{s.e.c.}$ and $H \equiv 0$. Just as in the proofs of the propositions 2.2.2.1 and 2.2.2.2 one verifies again that we have (i), (ii) and (iv). To (iii) we have to pay special attention. To see that G_n is of bounded variation uniformly in n (on $[0,a]$), we note that the $\widehat{g}_n$'s and g_n are positive and monotone decreasing and $1/\widehat{g}_n$ is on $[0,a]$ uniformly bounded in n by (2.18)–(2.20). There is a $D > 0$ such that $\|g_n\|_\infty \le D$ and $\|\widehat{g}_n\|_\infty \le D$ on $[0,a]$. We may write for all n:

$$\begin{aligned}
&\int_0^a \left| \mathrm{d}\left(\frac{g_n(x)}{\widehat{g}_n(x)}\right)\right| \\
&= \int_0^a \left| \frac{1}{\widehat{g}_n(x)} \mathrm{d}g_n(x) - \frac{g_n(x)}{(\widehat{g}_n(x))^2} \mathrm{d}\widehat{g}_n(x) \right| \\
&\le \int_0^a \frac{1}{\widehat{g}_n(x)} |\mathrm{d}g_n(x)| + \int_0^a \frac{g_n(x)}{(\widehat{g}_n(x))^2} |\mathrm{d}\widehat{g}_n(x)| \\
&\le -C \cdot \int_0^a \mathrm{d}g_n(x) - C^2 \cdot \int_0^a \mathrm{d}\widehat{g}_n(x) \\
&\le C\, g_n(0) + C^2\, \widehat{g}_n(0) \\
&\le (C + C^2)\, D,
\end{aligned}$$

where $C = M_a$ in case I and III and $C = M$ in case II. This proves that G_n is of bounded variation on $[0,a]$ uniformly in n. This completes the proof. □

With the propositions 2.2.2.1 – 2.2.2.3 we prove the following lemma, which is the key to the proof of theorem 2.2.5. It says that the condition in theorem 2.1.1 is satisfied.

Lemma 2.2.2.4 *The inequality (2.17) holds in the limit, possibly after passing to a subsequence.*

PROOF: we split the proof in three parts and at the beginning of each part we mention the cases, for which that part is meant.

• **For the cases I, II and III**: because the integrands in (2.17) are nonnegative we also have

$$\int_0^a \frac{\mathrm{d}V_n}{\mathrm{d}\widehat{V}_n}(x)\, \mathrm{d}F_n^{u.c.}(x) + \int_0^a \frac{g_n(x)}{\widehat{g}_n(x)}\, \mathrm{d}F_n^{s.e.c.}(x) + \widehat{h}_n \le 1 \tag{2.36}$$

for all $a \in [0,\tau)$. Together with proposition 2.2.2.2 and proposition 2.2.2.3 we obtain from (2.36), possibly after passing to a subsequence

$$\int_0^a \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x)\, \mathrm{d}F^{u.c.}(x) + \int_0^a \frac{g(x)}{g_\infty(x)}\, \mathrm{d}F^{s.e.c.}(x) + h \le 1 \tag{2.37}$$

for all $a \in [0,\tau_0)$. The integrands in (2.37) are nonnegative and so letting a converge to τ_0 we get by monotone convergence

$$\int_0^{\tau_0-} \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x)\, \mathrm{d}F^{u.c.}(x) + \int_0^{\tau_0-} \frac{g(x)}{g_\infty(x)}\, \mathrm{d}F^{s.e.c.}(x) + h \le 1. \tag{2.38}$$

• **For the cases I and II**: in case I and II the inequality (2.38) gives (2.17) in the limit, because $\tau_0 = \tau$ and because by $\int_0^\tau$ we mean the integral over $(0,\tau)$.

• **For case III**: of course in case III we also mean by $\int_0^\tau$ the integral over $(0,\tau)$, but in this case we already noted that in (2.17) we could replace $(0,\tau)$ by $(0,\tau_0]$. In (2.38) we have

(2.17) in the limit on $(0, \tau_0)$ and therefore we need an extra analysis. For convenience we define for $a \in [0, \tau_0)$

$$\begin{aligned} K_n(a) &\equiv \int_0^a \frac{\mathrm{d}V_n}{\mathrm{d}\widehat{V}_n}(x)\,\mathrm{d}F_n^{u.c.}(x) + \int_0^a \frac{g_n(x)}{\widehat{g}_n(x)}\,\mathrm{d}F_n^{s.e.c.}(x) + \widehat{h}_n \\ K(a) &\equiv \int_0^a \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x)\,\mathrm{d}F^{u.c.}(x) + \int_0^a \frac{g(x)}{g_\infty(x)}\,\mathrm{d}F^{s.e.c.}(x) + h \\ L_n(a) &\equiv \int_a^{\tau_0} \frac{\mathrm{d}V_n}{\mathrm{d}\widehat{V}_n}(x)\,\mathrm{d}F_n^{u.c.}(x) + \int_a^{\tau_0} \frac{g_n(x)}{\widehat{g}_n(x)}\,\mathrm{d}F_n^{s.e.c.}(x). \end{aligned}$$

(Thus in $K_n(a)$ and $K(a)$ we integrate over $(0, a]$ and in $L_n(a)$ over $(a, \tau_0]$). It is easy to see that in case III (2.17) can be written as

$$K_n(\tau_0-) + L_n(\tau_0-) \leq 1. \tag{2.39}$$

We know that the integrands in (2.17) are nonnegative, thus we immediately have

$$K_n(a) + L_n(\tau_0-) \leq 1 \tag{2.40}$$

and $0 \leq L_n(\tau_0-) \leq 1$. By the propositions 2.2.2.2 and 2.2.2.3 we have that $K_n(a)$ converges to $K(a)$, possibly after passing to a subsequence. Let n_k ($k \in \mathbb{N}$) be this subsequence. Because $0 \leq L_n(\tau_0-) \leq 1$ there exists a subsubsequence $p(k)$ of n_k such that $L_{p(k)}(\tau_0-)$ converges to a number $L(\tau_0-)$. For this subsubsequence we have that (2.40) converges to

$$K(a) + L(\tau_0-) \leq 1. \tag{2.41}$$

Again, the integrands in $K(a)$ are nonnegative and thus letting a converge to τ_0 from below we obtain by monotone convergence

$$K(\tau_0-) + L(\tau_0-) \leq 1. \tag{2.42}$$

One notes that if τ_0 is a continuity point of V, then $L_n(\tau_0-) = 0$ with probability 1 and thus $L(\tau_0-) = 0$. The probability of obtaining a s.e.c. observation with $\widetilde{X} = \tau_0$ is 0, and thus if τ_0 is a discontinuity point of V, we have with probability 1 that

$$\begin{aligned} L_n(\tau_0-) &= \int_{\tau_0-}^{\tau_0} \frac{\mathrm{d}V_n}{\mathrm{d}\widehat{V}_n}(x)\,\mathrm{d}F_n^{u.c.}(x) + 0 \\ &= \frac{\Delta V_n(\tau_0)}{\Delta \widehat{V}_n(\tau_0)} \cdot \Delta F_n^{u.c.}(\tau_0), \end{aligned}$$

where $\Delta f(x)$ means the height of the jump of f at x. In the case that τ_0 is a discontinuity point of V, we have for an increasing sample that the fraction of u.c. observations $\widetilde{X} = X = \tau_0$ is strictly greater than some $\delta > 0$ with probability 1. Thus for n large enough we have $\Delta\widehat{V}_n(\tau_0) > \delta > 0$. Furthermore all the points in (τ_0, τ) are continuity points of V_∞, because on (τ_0, τ) $\widehat{V}_n$ and thus V_∞ are constant. So on this interval we have $\widehat{V}_{n_k}(x) \to V_\infty(x)$ and because of the rightcontinuity we get $\widehat{V}_{n_k}(\tau_0) \to V_\infty(\tau_0)$. Furthermore we have that $V_\infty(\tau_0-)$

exists. If we combine these facts we get that $\Delta\widehat{V}_{n_k}(\tau_0) \to \Delta V_\infty(\tau_0) > \delta > 0$ with probability 1. So along the subsubsequence $p(k)$ we get that $L_{p(k)}(\tau_0-)$ converges with probability 1 to

$$L(\tau_0-) = \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(\tau_0) \equiv \frac{\Delta V(\tau_0)}{\Delta V_\infty(\tau_0)} \cdot \Delta F^{u.c.}(\tau_0) < \infty.$$

With this definition of $(\mathrm{d}V/\mathrm{d}V_\infty)(\tau_0)$ (in proposition 2.2.2.1 we had only an expression for $(\mathrm{d}V/\mathrm{d}V_\infty)(x)$ on $[0,\tau_0)$ in case III) we proved that (2.17) (thus (2.39)) holds in the limit after passing to a subsequence:

$$\begin{aligned} 1 &\geq K(\tau_0-) + L(\tau_0-) \\ &= \int_0^{\tau_0-} \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x)\,\mathrm{d}F^{u.c.}(x) + \int_0^{\tau_0-} \frac{g(x)}{g_\infty(x)}\,\mathrm{d}F^{s.e.c.}(x) + h + \int_{\tau_0-}^{\tau_0} \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x)\,\mathrm{d}F^{u.c.}(x) \\ &= \int_0^{\tau_0} \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x)\,\mathrm{d}F^{u.c.}(x) + \int_0^{\tau_0} \frac{g(x)}{g_\infty(x)}\,\mathrm{d}F^{s.e.c.}(x) + h. \end{aligned}$$

Now we have proved the lemma. □

2.2.3 Proof of two propositions of chapter 1

We promised to prove proposition 1.1.5.2 and 1.1.7.2 in this chapter. Here we give the proof. PROOF OF PROPOSITIONS 1.1.5.2 AND 1.1.7.2: for the proof of both propositions we do not have to distinguish between the NPMLE and the sieved NPMLE.

Because we have $g(\tau) > 0$ we know by (2.16) that for n large enough $g_n(\tau) \geq c > 0$ for a constant $c > 0$. Say that this holds for $n \geq N$. Now let the sample size be n with $n \geq N$ and suppose that the biggest observation point x_r in $[0,\tau)$ is s.e.c., then ($\widehat{g}_n$ and g_n are right-continuous on $[0,\tau)$) on $[x_r,\tau)$ we have that $\widehat{g}_n$ and g_n are constant and equal $\widehat{g}_n(\tau)$ and $g_n(\tau)$ respectively (because $\widehat{g}_n$ and g_n are left-continuous in τ). Furthermore we have that

$$\int_{[x_r,\tau)} \frac{g_n(x)}{\widehat{g}_n(x)}\,\mathrm{d}F_n^{s.e.c.}(x) \leq 1$$

(because the integrands in (2.17) are nonnegative). Thus we obtain

$$\widehat{g}_n(\tau) \geq g_n(\tau)\,\Delta F_n^{s.e.c.}(x_r) \geq \frac{c}{n} > 0$$

(because $g_n(\tau) \geq c > 0$ for $n \geq N$ and $\Delta F_n^{s.e.c.}(x_r)$ is at least $1/n$). This immediately implies that $0 < 2\tau\widehat{g}_n(\tau) = 1 - \widehat{h}_n - \widehat{V}_n(\tau-)$ and therefore we have $(\widehat{V}_n, \widehat{h}_n) \in \mathcal{V}_\tau$.

The only thing we still have to prove is that the probability that the biggest observation point is s.e.c. tends to 1. Because $g(\tau) > 0$ we have for each $a \in (0,\tau)$ that the probability that in the interval (a,τ) there are observation points tends to 1. Thus we are ready if we can show that the probability to have an u.c. observation point in (a,τ) given that the observation point lies in (a,τ) tends to 0 if a tends to τ. For convenience we suppose that V has a density

v (now we can apply the rule of De l'Hospital). Now we have that this limit of the conditional probability equals

$$\begin{aligned}
&\lim_{a\uparrow\tau}\left(\int_a^\tau \frac{\tau-w}{\tau+w}v(w)\,\mathrm{d}w\right)\cdot\left(\int_a^\tau \frac{\tau-w}{\tau+w}v(w)\,\mathrm{d}w+\int_a^\tau g(w)\,\mathrm{d}w\right)^{-1}\\
&= \lim_{a\uparrow\tau}\left(\frac{\tau-a}{\tau+a}v(a)\right)\cdot\left(\frac{\tau-a}{\tau+a}v(a)+g(a)\right)^{-1}\\
&= \frac{0}{0+g(\tau-)}=0.
\end{aligned}$$

(Of course $g(\tau-)>0$ because $g(\tau)>0$ and g is decreasing). This completes the proof. □

2.3 Consistency in the two-dimensional circle case

In this section we formulate consistency results for the *sieved* NPMLE $(\widehat{V}_n,\widehat{h}_n)$ of (V,h) in the two-dimensional circle case introduced in section 1.2. Again we show that these imply consistency results for $(\widehat{F}_n,\widehat{\mu}_n)$ of the original parameters (F,μ).

First we refresh our memory. We remember that the functions $g(x)$ and $d(x,x)$ are decreasing (see (1.72) and (1.73)). We have that $4\sqrt{R^2-(1/4)\,x^2}\,g(x)$ is the density of the subdistribution function of the s.e.c. observations (see (1.69)) and for the density of the subdistribution function of the d.c. observations (see (1.70)) we have that $x\,(2\sqrt{R^2-(1/4)\,x^2})^{-1}\,d(x,x)$.

Now let $2R_0\in[0,2R]$ be such that

$$V((2R_0,2R))=0 \tag{2.43}$$

and

$$V((2R_0-\epsilon,2R_0])>0 \quad \text{for all } \epsilon\in(0,2R_0]. \tag{2.44}$$

We distinguish the following four cases, which cover all (interesting) possibilities for V:

Case I : for the underlying V we have $2R_0=2R$,
Case II : for the underlying V we have $2R_0<2R$ and $V([2R,\infty))>0$ and $d(2R,2R)>0$,
Case IIa : for the underlying V we have $2R_0<2R$ and $V([2R,\infty))>0$ and $d(2R,2R)=0$,
Case III : for the underlying V we have $2R_0<2R$ and $V([2R,\infty))=0$ and $d(2R,2R)=0$,

We note that in case IIa we have $g(2R)>0$ and $d(2R,2R)=0$. If we remember the definitions of g and $d(\cdot,\cdot)$, then we know that these constraints correspond with an underlying V which gives positive mass to $2R$ but no mass to the interval $(2R,\infty)$. Of course there is no distribution function V on $[0,\infty)$ such that $V([2R,\infty))=0$ and $d(2R,2R)>0$ and thus $g(2R)=0$ and $d(2R,2R)>0$. So this case does not have to be considered.

For the sieved NPMLE $(\widehat{V}_n,\widehat{h}_n)$ we state a consistency result in the following theorem:

Theorem 2.3.1 *For the sieved NPMLE* $(\widehat{V}_n,\widehat{h}_n)$ *of* $(V,h)\in\mathcal{V}_{2R}$ *in the two-dimensional circle case we have that*

$$\widehat{h}_n-h\to 0 \quad a.s.$$

and

$$\sup_{x\in[0,2R-\epsilon]} \left|\widehat{V}_n(x) - V(x)\right| \to 0 \quad a.s.$$

for all $\epsilon \in (0, 2R]$.

The theorem is proved in section 2.3.1. (A corollary as 2.2.6 can also be formulated here).

One remembers the functions $g(x)$ and $d(x,x)$ in terms of V on $[0,2R)$ and h (see (1.81) and (1.82)). For the function g we obtain the estimator $\widehat{g}_n$ on $[0,2R]$:

$$\widehat{g}_n(x) = \int_x^{2R} \frac{1}{|W| + 2uR}\,\mathrm{d}\widehat{V}_n(u) + \widehat{g}_n(2R), \tag{2.45}$$

where $\widehat{g}_n(2R)$ is defined by $\widehat{V}_n(2R-) + (|W| + 4R^2)\widehat{g}_n(2R) + \widehat{h}_n = 1$. For $d(x,x)$ we get the estimator $\widehat{d}_n(x,x)$ on $[0,2R]$:

$$\widehat{d}_n(x,x) = \int_x^{2R} \frac{u-x}{|W| + 2uR}\,\mathrm{d}\widehat{V}_n(u) + \frac{1}{2R}\widehat{h}_n + (2R - x)\,\widehat{g}_n(2R). \tag{2.46}$$

We have that $\widehat{g}_n$ and $\widehat{d}_n(\cdot,\cdot)$ follow from $(\widehat{V}_n, \widehat{h}_n)$ according to (1.81) and (1.82).

Of course in the one-dimensional case one could prove that $\sup_{x\in[0,\tau-\epsilon]} |\widehat{g}_n(x) - g(x)|$ tends to 0 a.s. for all $\epsilon \in (0,\tau]$. The proof would be similar to the proof of the following proposition. Here we put this result for the two-dimensional circle case in a proposition together with a consistency result for $\widehat{d}_n(\cdot,\cdot)$, because we want to emphasize the difference between both consistency results; for $\widehat{g}_n$ we have strong consistency on intervals $[0, 2R-\epsilon]$ and for $\widehat{d}_n$ we have strong consistency on the whole interval $[0,2R]$.

Proposition 2.3.2 *For the functions* $\widehat{g}_n(\cdot)$ *and* $\widehat{d}_n(\cdot,\cdot)$ *both defined on* $[0,2R]$, *we have*

$$\sup_{x\in[0,2R-\epsilon]} |\widehat{g}_n(x) - g(x)| \to 0 \quad a.s.$$

for all $\epsilon \in (0, 2R]$ *and we have*

$$\sup_{x\in[0,2R]} \left|\widehat{d}_n(x,x) - d(x,x)\right| \to 0 \quad a.s..$$

Proof: using partial integration we write

$$\begin{aligned} \widehat{g}_n(x) &= -\frac{1}{|W| + 2xR}\,\widehat{V}_n(x) + \int_x^{2R} \frac{2R}{(|W| + 2uR)^2}\,\widehat{V}_n(u)\,\mathrm{d}u \\ &\quad + \frac{1}{|W| + 4R^2}\,\widehat{V}_n(2R-) + \widehat{g}_n(2R) \end{aligned} \tag{2.47}$$

and

$$\begin{aligned} \widehat{d}_n(x,x) &= \frac{1}{2R}\widehat{h}_n - (|W| + 2xR)\int_x^{2R} \frac{1}{(|W| + 2uR)^2}\,\widehat{V}_n(u)\,\mathrm{d}u \\ &\quad + (2R - x)\left(\frac{1}{|W| + 4R^2}\,\widehat{V}_n(2R-) + \widehat{g}_n(2R)\right). \end{aligned} \tag{2.48}$$

Theorem 2.3.1 gives us $\widehat{V}_n(x) \to V(x)$ for all $x \in [0, 2R)$. Then by Lesbesgue's dominated convergence theorem we obtain that

$$Y_n(x) \equiv \int_x^{2R} \frac{2R}{(|W| + 2uR)^2} \widehat{V}_n(u)\,\mathrm{d}u \to Y(x) \equiv \int_x^{2R} \frac{2R}{(|W| + 2uR)^2} V(u)\,\mathrm{d}u$$

for all $x \in [0, 2R]$. It is obvious that $Y_n(x-) \to Y(x-)$ holds. We note that the Y_n's and Y are monotone and that Y is bounded, thus we apply lemma 2.2.4 to obtain

$$\sup_{x\in[0,2R]} |Y_n(x) - Y(x)| \to 0. \tag{2.49}$$

Theorem 2.3.1 gives us $\widehat{h}_n \to h$ and because we have $(|W| + 4R^2)\,\widehat{g}_n(2R) = 1 - \widehat{h}_n - \widehat{V}_n(2R-)$ we get immediately

$$\frac{1}{|W| + 4R^2} \widehat{V}_n(2R-) + \widehat{g}_n(2R) \quad\to\quad \frac{1}{|W| + 4R^2} V(2R-) + g(2R). \tag{2.50}$$

Now with theorem 2.3.1 together with (2.47)-(2.50) one easily concludes that both statements in the proposition hold. (Because of the term $(1/(|W| + 2xR))\,\widehat{V}_n(x)$ in (2.47) we have strong consistency of $\widehat{g}_n$ only on intervals $[0, \tau - \epsilon]$). □

We formulate consistency results for the parameters (F, μ) in the next theorem:

Theorem 2.3.3 *For the sieved NPMLE* $(\widehat{F}_n, \widehat{\mu}_n)$ *of* $(F, \mu) \in \mathcal{F}_{2R}$ *in the two-dimensional circle case we have that*

$$\widehat{\mu}_n - \mu \to 0 \quad a.s.$$

and

$$\sup_{x\in[0,2R-\epsilon]} \left|\widehat{F}_n(x) - F(x)\right| \to 0 \quad a.s.$$

for all $\epsilon \in (0, 2R]$.

PROOF: the proof is identical to the proof of theorem 2.2.7 using theorem 2.3.1 and using $\nu(V(\cdot), h)$ defined in (1.83) and the transformation defined in section 1.2.3. □

2.3.1 Proof of consistency in the two-dimensional circle case

PROOF OF THEOREM 2.3.1: the proof is very similar to the proof of theorem 2.2.5. Here we only point out the essential differences.

To apply theorem 2.1.1 we give here the elements that play the roles of T, T_n, $\widehat{T}_n$ and T_∞ in the theorem. It is clear that (V, h) plays the role of T.

- **The role of T_n:** we define

$$\mathrm{d}V_n(x) \equiv \frac{|W| + 2xR}{z(x)}\,\mathrm{d}F_n^{u.c.}(x) \quad , \quad h_n \equiv h$$

and $g_n(2R)$ is defined by $V_n(2R-) + (|W| + 4R^2)g_n(2R) + h_n = 1$ and

$$\begin{aligned} g_n(x) &\equiv \int_x^{2R} \frac{1}{|W| + 2uR}\,\mathrm{d}V_n(u) + g_n(2R) \\ &= \frac{1-h_n}{|W|+4R^2} + \frac{1}{|W|+4R^2} \\ &\quad \times \left(\int_0^{2R} \frac{4R^2 - 2uR}{z(u)}\,\mathrm{d}F_n^{u.c.}(u) - (|W| + 4R^2) \int_0^x \frac{1}{z(u)}\,\mathrm{d}F_n^{u.c.}(u) \right). \end{aligned} \tag{2.51}$$

Monotonicity (lemma 2.2.4) and the strong law of large numbers provides us with the fact that

$$\left\| \int_0^x \frac{j(u)}{z(u)}\,\mathrm{d}F_n^{u.c.}(u) - \int_0^x \frac{j(u)}{z(u)}\,\mathrm{d}F^{u.c.}(u) \right\|_{[0,2R]} \to 0 \tag{2.52}$$

a.s. $(n \to \infty)$, where $\|\cdot\|_I$ stands for the supremum norm on the interval I and where $j(\cdot)$ is any bounded positive function on $[0, 2R]$ (we use $j(u) = 4R^2 - 2uR$ and $j(u) = 1$ in (2.51)). Trivially we have $h_n - h = h - h = 0 \to 0$ $(n \to \infty)$ and with (2.52) one easily verifies that

$$\|g_n - g\|_{[0,2R)} \to 0 \quad , \quad \|V_n - V\|_{[0,2R)} \to 0 \ \text{ a.s. } (n \to \infty). \tag{2.53}$$

One must be aware of the fact that, because of the choice of $h_n = h$ (compare this with the choice of h_n in the one-dimensional case), $g_n(x)$ could be negative and then $g_n(x)$ will not be a good estimate for the density $g(x)$. In fact in that case $g_n(2R)$ is negative and that would mean that negative mass is assigned to $g_n(2R)$. Then the triple $V_n(\cdot)$ (on $[0, 2R)$), h_n and $(|W| + 4R^2)g_n(2R)$ does not represent a probability measure for the model. By (2.53) we know that with probability 1 this problem will not occur for n large enough.

According to (1.82) we define $d_n(x, x)$ to be

$$\begin{aligned} d_n(x,x) &\equiv \int_x^{2R} \frac{u-x}{|W| + 2uR}\,\mathrm{d}V_n(u) + \frac{1}{2R}h_n + (2R - x)\,g_n(2R) \\ &= \frac{2R - x}{|W| + 4R^2}\,V(2R) - (|W| + 2xR) \int_x^{2R} \frac{1}{(|W| + 2uR)^2}\,V_n(u)\,\mathrm{d}u \\ &\quad + \frac{1}{2R}h_n + (2R - x)\,g_n(2R) \end{aligned}$$

and by (2.53) we get

$$\|d_n - d\|_{[0,2R]} \to 0 \ \text{ a.s. } (n \to \infty)$$

(one remembers the proof of proposition 2.3.2). We note that g_n and $d_n(\cdot,\cdot)$ follow from (V_n, h_n) according to (1.81) and (1.82).

• **The role of $\widehat{T}_n$**: let $(\widehat{V}_n, \widehat{h}_n)$ denote the sieved NPMLE of $(V, h) \in \mathcal{V}_{2R}$. The functions $\widehat{g}_n(\cdot)$ and $\widehat{d}_n(\cdot,\cdot)$ on $[0, 2R]$ are defined in (2.45) and (2.46). We know that $\widehat{g}_n$ and $\widehat{d}_n(\cdot,\cdot)$ follow from $(\widehat{V}_n, \widehat{h}_n)$ according to (1.81) and (1.82).

• **The role of T_∞**: by the Helly Selection Theorem we have that for some subsequence $(\widehat{V}_{n_k}, \widehat{h}_{n_k})$ of the sequence of $(\widehat{V}_n, \widehat{h}_n)$, there exists a nondecreasing right continuous function

V_∞ on $[0, 2R)$ and a h_∞ such that $\lim_{k\to\infty} \widehat{V}_{n_k}(x) = V_\infty(x)$ at the continuity points of V_∞ and $\widehat{h}_{n_k} \to h_\infty$. We define

$$g_\infty(x) \equiv \int_x^{2R} \frac{1}{|W| + 2uR} \, \mathrm{d}V_\infty(u) + g_\infty(2R)$$

where $g_\infty(2R)$ is defined by $V_\infty(2R-) + (|W| + 4R^2) g_\infty(2R) + h_\infty = 1$. We write

$$\begin{aligned} \widehat{g}_{n_k}(x) &= \int_x^{2R} \frac{1}{|W| + 2uR} \, \mathrm{d}\widehat{V}_{n_k}(u) + \widehat{g}_{n_k}(2R) \\ &= \frac{1 - \widehat{h}_{n_k}}{|W| + 4R^2} - \frac{1}{|W| + 2xR} \widehat{V}_{n_k}(x) + \int_x^{2R} \frac{2R}{(|W| + 2uR)^2} \widehat{V}_{n_k}(u) \, \mathrm{d}u \end{aligned}$$

and by the weak convergence of $(\widehat{V}_{n_k}, \widehat{h}_{n_k})$ to (V_∞, h_∞) we get $\lim_{k\to\infty} \widehat{g}_{n_k}(x) = g_\infty(x)$ at the continuity points of V_∞. The same result is obtained for d_∞: $\lim_{k\to\infty} d_{n_k}(x, x) = d_\infty(x, x)$, where $d_\infty(\cdot, \cdot)$ is defined as

$$d_\infty(x, x) \equiv \int_x^{2R} \frac{u - x}{|W| + 2uR} \, \mathrm{d}V_\infty(u) + \frac{1}{2R} h_\infty + (2R - x) \, g_\infty(2R).$$

Again one verifies immediately that g_∞ and $d_\infty(\cdot, \cdot)$ follow from (V_∞, h_∞) according to (1.81) and (1.82).

For the sieved NPMLE $(\widehat{V}_n, \widehat{h}_n)$ we find by (2.2)

$$\int_0^{2R} \frac{\mathrm{d}V_n}{\mathrm{d}\widehat{V}_n}(x) \, \mathrm{d}F_n^{u.c.}(x) + \int_0^{2R} \frac{g_n(x)}{\widehat{g}_n(x)} \, \mathrm{d}F_n^{s.e.c.}(x) + \int_0^{2R} \frac{d_n(x, x)}{\widehat{d}_n(x, x)} \, \mathrm{d}F_n^{d.c.}(x) \le 1. \tag{2.54}$$

(This inequality plays the role of inequality (2.17) in the proof of theorem 2.2.5). To prove that (2.54) holds in the limit, possibly after passing to a subsequence, we need not only to control the possible unboundedness of $1/\widehat{g}_n(x)$, but also the possible unboundedness of $1/\widehat{d}_n(x, x)$. In the following we find that (2.55)–(2.59) are the analogues of (2.18)–(2.20).

• **Case I**: $2R_0 = 2R$. Using the same arguments as for (2.18) we get for n large enough and for each $a \in [0, 2R)$

$$\frac{1}{\widehat{g}_n(x)} \le M_a \, , \quad \frac{1}{g_\infty(x)} \le M_a \, , \quad \frac{1}{\widehat{d}_n(x, x)} \le M_a \, , \quad \frac{1}{d_\infty(x, x)} \le M_a \tag{2.55}$$

for all points $x \in [0, a]$, where $M_a \ge 0$ is a constant.

• **Case II**: $2R_0 < 2R$ and $V([2R, \infty)) > 0$ and $d(2R, 2R) > 0$. Using the same arguments as for (2.19) we get for n large enough

$$\frac{1}{\widehat{g}_n(x)} \le M \, , \quad \frac{1}{g_\infty(x)} \le M \, , \quad \frac{1}{\widehat{d}_n(x, x)} \le M \, , \quad \frac{1}{d_\infty(x, x)} \le M \tag{2.56}$$

for all points $x \in [0, 2R]$ and M is constant. For the last two inequalties in (2.56) we use that (2.54) implies that

$$\int_{2R_0}^{2R} \frac{d_n(x,x)}{\hat{d}_n(x,x)}\, \mathrm{d}F_n^{d.c.}(x)$$

(because the integrands in (2.54) are nonnegative). We know that in this case $\hat{d}_n(x,x)$ is constant on $(2R_0, 2R]$ (see (2.43)). The same holds for $d_n(x,x)$ and thus we may write

$$d_n(2R,2R)\,(F_n^{d.c.}(2R) - F_n^{d.c.}(2R_0)) \leq \hat{d}_n(2R,2R)$$

and that the left-hand side of this inequality converges to

$$d(2R,2R)\,(F^{d.c.}(2R) - F^{d.c.}(2R_0)) = d(2R,2R) \cdot \int_{2R_0}^{2R} \frac{u}{2\sqrt{R^2 - \frac{1}{4}u^2}}\, d(u,u)\, \mathrm{d}u.$$

Because of $d(2R,2R) > 0$ this is strictly positive and therefore we conclude that there is a constant $c > 0$ such that $d_\infty(2R,2R) \geq c$ and for n large enough we have $\hat{d}_n(2R,2R) \geq c$. This and the fact that $d_\infty(\cdot,\cdot)$ and the $\hat{d}_n(\cdot,\cdot)$'s are decreasing imply the last two inequalities in (2.56).

In this case we note that on $(2R_0, 2R)$ we have: $\mathrm{d}\hat{V}_n(x) = \mathrm{d}V_n(x) = \mathrm{d}V_\infty(x) = \mathrm{d}V(x) = 0$ and $\mathrm{d}F_n^{u.c.}(x) = \mathrm{d}F^{u.c.}(x) = 0$.

• **Case IIa**: $2R_0 < 2R$ and $V([2R,\infty)) > 0$ and $d(2R,2R) = 0$. Just as in case II we find for $1/\hat{g}_n$ and $1/g_\infty$ the bound:

$$\frac{1}{\hat{g}_n(x)} \leq M\,, \quad \frac{1}{g_\infty(x)} \leq M \tag{2.57}$$

for all points $x \in [0, 2R]$ and M is constant.

The estimator $\hat{V}_n$ gives at least mass $1/n$ to uncensored observations, so with (2.44) we have for each $a \in [0, 2R_0)$ and for n large enough

$$\frac{1}{\hat{d}_n(x,x)} \leq M_a\,, \quad \frac{1}{d_\infty(x,x)} \leq M_a \tag{2.58}$$

for all points $x \in [0, a]$, where M_a is a constant.

In this case we note that on $(2R_0, 2R)$ we have: $\mathrm{d}\hat{V}_n(x) = \mathrm{d}V_n(x) = \mathrm{d}V_\infty(x) = \mathrm{d}V(x) = 0$ and $\mathrm{d}F_n^{u.c.}(x) = \mathrm{d}F^{u.c.}(x) = 0$ and $d(x,x) = 0$ and $\mathrm{d}F_n^{d.c.}(x) = \mathrm{d}F^{d.c.}(x) = 0$. In this case we can replace the integrals over $(0, 2R)$ in the first and third term of (2.54) by $(0, 2R_0]$.

• **Case III**: $2R_0 < 2R$ and $V([2R,\infty)) = 0$ and $d(2R,2R) = 0$. Similar arguments as for (2.20) give us (for each $a \in [0, 2R_0)$) for n large enough

$$\frac{1}{\hat{g}_n(x)} \leq M_a\,, \quad \frac{1}{g_\infty(x)} \leq M_a\,, \quad \frac{1}{\hat{d}_n(x,x)} \leq M_a\,, \quad \frac{1}{d_\infty(x,x)} \leq M_a \tag{2.59}$$

for all points $x \in [0, a]$, where M_a is a constant. In this case we have on $(2R_0, 2R)$: $\mathrm{d}\hat{V}_n(x) = \mathrm{d}V_n(x) = \mathrm{d}V_\infty(x) = \mathrm{d}V(x) = 0$ and $\mathrm{d}F_n^{u.c.}(x) = \mathrm{d}F^{u.c.}(x) = 0$ and $g(x) = 0$ and $d(x,x) = 0$ and $\mathrm{d}F_n^{s.e.c.}(x) = \mathrm{d}F^{s.e.c.}(x) = 0$ and $\mathrm{d}F_n^{d.c.}(x) = \mathrm{d}F^{d.c.}(x) = 0$. Just as in case III of the one-dimensional problem we can replace here the integrals over $(0, 2R)$ in (2.54) by $(0, 2R_0]$.

One verifies that $V_n \ll \widehat{V}_n$ and by (1.86) we get

$$\begin{aligned}\frac{\mathrm{d}V_n}{\mathrm{d}\widehat{V}_n}(x) &= \frac{|W|+2xR}{z(x)}\left(1-\frac{1}{|W|+2xR}\int_{v=0}^{v=x}\frac{1}{\widehat{g}_n(v)}\,\mathrm{d}F_n^{s.e.c.}(v)\right.\\ &\left.-\frac{1}{|W|+2xR}\int_{v=0}^{v=x}\frac{x-v}{\widehat{d}_n(v,v)}\,\mathrm{d}F_n^{d.c.}(v)\right) \qquad (2.60)\end{aligned}$$

(see (2.21)). From now on one imitates the proof of theorem 2.2.5. All the necessary ingredients are found in (2.55)–(2.59) and the propositions 2.3.1.1 – 2.3.1.4 and lemma 2.3.1.5. They are the analogues of the propositions 2.2.2.1 – 2.2.2.3 and lemma 2.2.2.4. In (2.54) one deals with the term

$$\int_0^{2R}\frac{d_n(x,x)}{\widehat{d}_n(x,x)}\,\mathrm{d}F_n^{d.c.}(x)$$

in the same way as the term

$$\int_0^{2R}\frac{g_n(x)}{\widehat{g}_n(x)}\,\mathrm{d}F_n^{s.e.c.}(x).$$

□

In the remainder of this section we give the propositions 2.3.1.1 – 2.3.1.4 and lemma 2.3.1.5.

Proposition 2.3.1.1 *On $[0,2R)$ in case I and II and on $[0,2R_0)$ in case IIa and III, we have $V \ll V_\infty$ and*

$$\begin{aligned}\frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x) &= \frac{|W|+2xR}{z(x)}\left(1-\frac{1}{|W|+2xR}\int_{v=0}^{v=x}\frac{1}{g_\infty(v)}\,\mathrm{d}F^{s.e.c.}(v)\right.\\ &\left.-\frac{1}{|W|+2xR}\int_{v=0}^{v=x}\frac{x-v}{d_\infty(v,v)}\,\mathrm{d}F^{d.c.}(v)\right).\end{aligned}$$

PROOF: the proof is similar to the proof of proposition 2.2.2.1. One treats the difference

$$\begin{aligned}&\int_{v=0}^{v=x}\frac{x-v}{\widehat{d}_n(v,v)}\,\mathrm{d}F_n^{d.c.}(v)\cdot\frac{1}{|W|+2xR}\,\mathrm{d}\widehat{V}_n(x)\\ &\quad-\int_{v=0}^{v=x}\frac{x-v}{d_\infty(v,v)}\,\mathrm{d}F^{d.c.}(v)\cdot\frac{1}{|W|+2xR}\,\mathrm{d}V_\infty(x)\end{aligned}$$

in the same way as the difference

$$\begin{aligned}&\int_{v=0}^{v=x}\frac{1}{\widehat{g}_n(v)}\,\mathrm{d}F_n^{s.e.c.}(v)\cdot\frac{1}{|W|+2xR}\,\mathrm{d}\widehat{V}_n(x)\\ &\quad-\int_{v=0}^{v=x}\frac{1}{g_\infty(v)}\,\mathrm{d}F^{s.e.c.}(v)\cdot\frac{1}{|W|+2xR}\,\mathrm{d}V_\infty(x),\end{aligned}$$

which is similar to the difference in (2.25). □

Proposition 2.3.1.2

$$\int_0^a \frac{\mathrm{d}V_{n_k}}{\mathrm{d}\widehat{V}_{n_k}}(x)\cdot \mathrm{d}F_{n_k}^{u.c.}(x) \to \int_0^a \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x)\cdot \mathrm{d}F^{u.c.}(x) \qquad (k\to\infty)$$

for all a in $[0,2R)$ *in case I and II and in* $[0,2R_0)$ *in case IIa and III.*

Proof: the proof is similar to the proof of proposition 2.2.2.2 and one uses the same comment as in the proof of proposition 2.3.1.1. □

Proposition 2.3.1.3

$$\int_0^a \frac{g_{n_k}(x)}{\widehat{g}_{n_k}(x)}\,\mathrm{d}F_{n_k}^{s.e.c.}(x) \to \int_0^a \frac{g(x)}{g_\infty(x)}\,\mathrm{d}F^{s.e.c.}(x) \qquad (k\to\infty)$$

for all a in $[0,2R)$ *in case I, II and IIa and in* $[0,2R_0)$ *in case III.*

Proof: the proof is identical to the proof of proposition 2.2.2.3. Just replace τ by $2R$ and τ_0 by $2R_0$ and use the corresponding definitions of $g(x)$, $g_\infty(x)$, $\widehat{g}_n(x)$, $F_n^{s.e.c.}$ and $F^{s.e.c.}$. □

Proposition 2.3.1.4

$$\int_0^a \frac{d_{n_k}(x,x)}{\widehat{d}_{n_k}(x,x)}\,\mathrm{d}F_{n_k}^{d.c.}(x) \to \int_0^a \frac{d(x,x)}{d_\infty(x,x)}\,\mathrm{d}F^{d.c.}(x) \qquad (k\to\infty)$$

for all a in $[0,2R)$ *in case I and II and in* $[0,2R_0)$ *in case IIa and III.*

Proof: the proof is the same as the proof of proposition 2.2.2.3. Just replace τ by $2R$, τ_0 by $2R_0$, $g(x)$ by $d(x,x)$, $\widehat{g}_n(x)$ by $\widehat{d}_n(x,x)$, $g_\infty(x)$ by $d_\infty(x,x)$, $g_n(x)$ by $d_n(x,x)$, $F_n^{s.e.c.}$ by $F_n^{d.c.}$ and $F^{s.e.c.}$ by $F^{d.c.}$. □

Lemma 2.3.1.5 *The inequality (2.54) holds in the limit, possibly after passing to a subsequence.*

Proof: the proof is almost the same as the proof of lemma 2.2.2.4. In case IIa we could replace in the first and third term of (2.54) the integral over $(0,2R)$ by the integral over $(0,2R_0]$. In this case and in this proof (2.38) would become

$$\int_0^{2R_0-} \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x)\,\mathrm{d}F^{u.c.}(x) + \int_0^{2R} \frac{g(x)}{g_\infty(x)}\,\mathrm{d}F^{s.e.c.}(x) + \int_0^{2R_0-} \frac{d(x,x)}{d_\infty(x,x)}\,\mathrm{d}F^{d.c.}(x) \le 1. \quad (2.61)$$

In case III we would get for (2.38):

$$\int_0^{2R_0-} \frac{\mathrm{d}V}{\mathrm{d}V_\infty}(x)\,\mathrm{d}F^{u.c.}(x) + \int_0^{2R_0-} \frac{g(x)}{g_\infty(x)}\,\mathrm{d}F^{s.e.c.}(x) + \int_0^{2R_0-} \frac{d(x,x)}{d_\infty(x,x)}\,\mathrm{d}F^{d.c.}(x) \le 1. \quad (2.62)$$

To prove that we can replace $2R_0-$ by $2R_0$ (and thus get the integral over $(0,2R_0]$) we imitate the proof of lemma 2.2.2.4 for case III. □

Chapter 3

Efficiency

In this chapter we prove asymptotic results for the (sieved) NPMLE in the two-dimensional circle-case. This we do in the sections 3.5 – 3.11. As mentioned in the introduction the results in the two-dimensional circle-case rely on the assumption that the determinant Q_V of a certain 2×2 matrix is unequal to zero. Van der Laan(1993) already studied the asymptotic behaviour in the one-dimensional case. Applying the analysis, which we used for the two-dimensional case, to the one-dimensional problem, we discover why this determinant matter did not appear here. In the one-dimensional case one can prove that this determinant $Q_V \geq 1$. This is shown in section 3.12.

To obtain the results in this chapter we have to restrict the class of underlying distribution functions V. These assumptions are formulated in section 3.4. But before we start with section 3.4, which is followed by the sections dealing with the one- and two-dimensional cases, we begin with a summary of some general efficiency theory in the sections 3.1 – 3.3. Because we estimate linear parameters in convex models, we make use of Van der Laan's(1993) identity. In section 3.3 we will show that his condition needed to obtain this identity can be changed into a condition which is easier to verify.

3.1 General notion of efficiency

3.1.1 Donsker class

Given a probability space $(\Omega, \mathcal{A}, P)$ we define the space $L^2(P)$ as

$$L^2(P) \equiv \{f : (\Omega, \mathcal{A}, P) \to \mathbb{R} : \ f \text{ measurable and } \int f^2 \, \mathrm{d}P < \infty\}$$

The set $L_0^2(P)$ will be defined as the set of all elements $f \in L^2(P)$ with $\int f \, \mathrm{d}P = 0$. We endow the spaces with the inner-product norm $\|f\|_P \equiv \sqrt{\langle f, f \rangle_P} \equiv \sqrt{\int f^2 \, \mathrm{d}P}$, which makes both spaces a Hilbert-space. Let $\mathcal{F} \subset L^2(P)$ and define

$$l^\infty(\mathcal{F}) \equiv \{H : \mathcal{F} \to \mathbb{R} : \|H\|_\mathcal{F} \equiv \sup_{f \in \mathcal{F}} |H(f)| < \infty\}.$$

We can consider function indexed empirical processes as random elements of $l^\infty(\mathcal{F})$ as follows. Let $X_1, ..., X_n$ be an i.i.d. sample of random elements in a measurable space $(\Omega, \mathcal{A}, P)$

and let P_n be the empirical measure which puts mass $1/n$ on each X_i $(i = 1, ..., n)$. Now we define the following map from a collection $\mathcal{F}$ consisting of measurable functions $f : \Omega \to \mathbb{R}$ to $\mathbb{R}$:

$$f \to P_n f = \int f \, \mathrm{d}P_n.$$

One considers $P_n = (P_n f : f \in \mathcal{F})$ as a random element of $l^\infty(\mathcal{F})$. The $\mathcal{F}$-indexed empirical process is given by

$$f \to G_n f = \sqrt{n}(P_n - P)f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(X_i) - Pf).$$

Let f be given such that Pf and Pf^2 exist, then we have the law of large numbers and the central limit theorem:

$$P_n f \to Pf \text{ a.s.} \quad , \quad G_n f \Longrightarrow^D N(0, P(f - Pf)^2).$$

If we have

$$\|P_n - P\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |P_n f - Pf| \to 0 \text{ a.s.}^*,$$

where a.s.* means outer almost surely, then we call the class $\mathcal{F}$ a P-Glivenko Cantelli class. This is the uniform version of the law of large numbers.

If we have $\sup_{f \in \mathcal{F}} |f(x) - Pf| < \infty$ for every x, then we can regard the empirical process $G_n = (G_n f : f \in \mathcal{F})$ as a random element of $l^\infty(\mathcal{F})$. Now we can talk about convergence of $G_n \Longrightarrow^D G$ in $l^\infty(\mathcal{F})$, where G is a tight measurable Gaussian law. If this is true we say that the uniform central limit theorem holds for $\mathcal{F}$ at P. We call a class $\mathcal{F}$ for which the uniform central limit theorem holds at P a P-Donsker class. For a more specific discussion we refer to Van der Vaart and Wellner(1993) and Dudley(1984, 1985). We call $\mathcal{F}$ Donsker uniformly in $P \in \mathcal{M}$ if this convergence is uniform in $P \in \mathcal{M}$ (via metrization of '$\Longrightarrow^D$'), for a certain set of probability measures $\mathcal{M}$.

Now let $I_k = \{(b_0 = 0, b_1], (b_1, b_2], ..., (b_{k-2}, b_{k-1}], (b_{k-1}, b_k = a]\}$ be a disjoint partition of the interval $I = (0, a] \subset \mathbb{R}$ consisting of k intervals. We define the variation of a function f on $(0, a]$ as follows

$$\|f\|_{v,0} = \sup_{I_k} \sum_{i=1}^{k} |f(b_i) - f(b_{i-1})|.$$

The variational norm $\|\cdot\|_v$ on $D[0, a]$ (the space of cadlag functions on $[0, a]$) is defined by

$$\|f\|_v \equiv \max\left(\|f\|_\infty, \|f\|_{v,0}\right).$$

The class of real valued functions on $(0, a] \subset \mathbb{R}$ with variational norm $\|\cdot\|_v$ smaller than some (fixed) constant $M < \infty$ is uniformly Donsker (see Dudley(1987)). Thus the set $\mathcal{F}_M \equiv \{f \in D[0, a] : \|f\|_v < M\}$ is uniformly Donsker. For monotone f one easily obtains the inequalities:

$$\|f\|_v \leq 2\|f\|_\infty \quad (3.1)$$

$$\|f\|_\infty \leq \|f\|_v. \quad (3.2)$$

3.1.2 Efficiency theory

In order to get a more or less self-contained chapter about efficiency, we give here the basic theory. This section contains all the definitions and derivations that are needed to build up the efficiency theory of section 3.2 and section 3.3. It can be found in Van der Laan(1993) sections 1.4 and 2.2.

We define a model $\mathcal{M}$ to be a set of probability measures on $(\mathcal{X}, \mathcal{B})$. Let

$$\mathcal{M}(\nu) \equiv \{P \in \mathcal{M} : P \ll \nu\}.$$

We write $p = (\mathrm{d}P/\mathrm{d}\nu)$ for the density of $P \in \mathcal{M}(\nu)$ w.r.t. ν. The collection of all these densities p corresponding with a $P \in \mathcal{M}(\nu)$ will be denoted by $\mathcal{P}(\nu)$.

We give the following definition of differentiability:

Definition 3.1.2.1 *A map $\epsilon \to p_\epsilon$ from $[0,1]$ to $\mathcal{P}(\nu)$ is called a differentiable (one- dimensional) submodel of $\mathcal{P}(\nu)$ through p if there exists an $l \in L_0^2(P)$ with*

$$\int \left(\frac{1}{\epsilon}(\sqrt{p_\epsilon} - \sqrt{p}) - \frac{1}{2}\, l \sqrt{p}\right)^2 \mathrm{d}\nu \to 0 \tag{3.3}$$

for $\epsilon \downarrow 0$.

Note that if the integrand in (3.3) converges pointwise to 0, then we would have

$$l(x) = \frac{\frac{\mathrm{d}}{\mathrm{d}\epsilon}\sqrt{p_\epsilon}|_{\epsilon=0}}{\frac{1}{2}\sqrt{p}} = \frac{\mathrm{d}}{\mathrm{d}\epsilon}\log(p_\epsilon(x))\bigg|_{\epsilon=0}.$$

Thus l can be considered as a $L^2(\nu)$ version of the score function of the one-dimensional submodel p_ϵ. Submodels P_ϵ in $\mathcal{M}$ with densities p_ϵ in $\mathcal{P}(\nu)$ for a certain measure ν satisfying (3.3) are called Hellinger differentiable.

We write $p_{\epsilon,l} \in \mathcal{P}(\nu)$ or $P_{\epsilon,l} \in \mathcal{M}(\nu)$ if we mean a one-dimensional differentiable submodel of densities (w.r.t. ν) or measures respectively, with score l as defined in definition 3.1.2.1.

Now let $\vartheta : \mathcal{M} \to \Theta \subset D$ be a parameter and let B be a collection of real valued linear mappings $b : D \to \mathbb{R}$. Given an i.i.d. sample $X_1, ..., X_n$ from an unknown $P \in \mathcal{M}$ we want to estimate the parameter $\theta = \vartheta(P)$, which is done by an estimator $\theta_n = \theta_n(X_1, X_2, ..., X_n)$. We have that $b\theta_n : (\mathcal{X}^n, \mathcal{B}^n) \to \mathbb{R}$ is a measurable map for all $b \in B$.

The Cramér-Rao lower bound, that bounds the variance of unbiased (over $P_{\epsilon,l}$) estimators of $b\vartheta(P_{\epsilon,l})$ at $\epsilon = 0$ from below, is given by

$$\frac{1}{n}\left(\frac{1}{\|l\|_P} \cdot \frac{\mathrm{d}}{\mathrm{d}\epsilon} b\vartheta(P_{\epsilon,l})\bigg|_{\epsilon=0}\right)^2 \tag{3.4}$$

(assuming $\mathrm{d}/\mathrm{d}\epsilon$ exists). If we define $\mathcal{S}(P)$ to be a class of differentiable submodels of $\mathcal{M}$ at P, then the variance of unbiased (over the whole set $\mathcal{M}$) estimators of $b\vartheta(P)$ is bounded from below by the supremum over $\mathcal{S}(P)$ of (3.4). This leads to the so called generalized Cramér-Rao lower bound. Because the bound (3.4) depends on the score l through $P_{\epsilon,l}$, the supremum is in fact a supremum over the collection of scores corresponding with $\mathcal{S}(P)$. We will define this collection more precisely

Definition 3.1.2.2 *A cone $S(P)$ in $L_0^2(P)$ is called a tangent cone at $P \in \mathcal{M}$ of $S(P)$ if for all $l \in S(P)$ there exists a differentiable one-dimensional submodel $P_{\epsilon,l} \in S(P) \subset \mathcal{M}$ through $P \in \mathcal{M}$ with score l.*

We remember that a cone C in a vector space over $\mathbb{R}$ is a subset for which the following condition holds: if $l \in C$ and $a \geq 0$, then $a\,l \in C$. Instead of taking the supremum over $S(P)$ in the generalized Cramér-Rao lower bound, we can replace it by taking the supremum over the tangent space $T(P)$, which we define by:

Definition 3.1.2.3 *For a tangent cone $S(P) \subset L_0^2(P)$ we define the tangent space $T(P) \subset L_0^2(P)$ as the closure of the linear extension of $S(P)$ within $L_0^2(P)$.*

The existence of the supremum over $S(P)$ of (3.4) will be guaranteed by the following differentiability assumption of $b\vartheta$.

Definition 3.1.2.4 *A parameter $b\vartheta : \mathcal{M} \to \mathbb{R}$ is called pathwise differentiable at $P \in \mathcal{M}$ relative to $S(P)$, if there exists a linear mapping $\dot{\vartheta} : T(P) \to (D, \|\cdot\|)$ such that $b\dot{\vartheta} : T(P) \to \mathbb{R}$ is continuous and linear and*

$$\frac{1}{\epsilon}\left(b\vartheta(P_{\epsilon,l}) - b\vartheta(P)\right) - b\dot{\vartheta}(l) \to 0$$

for all $l \in S(P)$. By the Riesz representation theorem there exists a $\tilde{l}(P, b\vartheta) \in T(P)$ such that

$$b\dot{\vartheta}(l) = \int \tilde{l}(P, b\vartheta)(x)\, l(x)\, \mathrm{d}P(x). \tag{3.5}$$

The Cramér-Rao lower bound (3.4) equals

$$\frac{1}{n}\left(\frac{b\dot{\vartheta}(l)}{\|l\|_P}\right)^2 = \frac{1}{n}\left(\frac{\int \tilde{l}(P, b\vartheta)(x)\, l(x)\, \mathrm{d}P(x)}{\|l\|_P}\right)^2.$$

With the Cauchy-Schwartz inequality this is maximized over $T(P)$ by $l = \tilde{l}(P, b\vartheta)$ and the result is exactly $(1/n)\|\tilde{l}(P, b\vartheta)\|^2$. For this reason we consider $P_{\epsilon,l}$, $l = \tilde{l}(P, b\vartheta)$ as the so called hardest one-dimensional submodel for estimating $b\vartheta(P)$. (If $\tilde{l}(P, b\vartheta) \notin S(P)$, then we still think of it as an approximate submodel). Thus $\tilde{l}(P, b\vartheta)$ is sometimes called the efficient score. The variance of $\tilde{l}(P, b\vartheta)$ is also the optimal asymptotic variance of $\sqrt{n}(\theta_n - \vartheta(P))$ for so called regular estimators (Van der Vaart, 1988).

Definition 3.1.2.5 *Let $b\theta_n$ be an estimator of $b\theta = b\vartheta(P)$ for which we have*

$$L_P\left(\sqrt{n}(b\theta_n - b\vartheta(P))\right) \to L_b.$$

We call $b\theta_n$ a $S(P)$-regular estimator of $b\theta$ if for all $l \in S(P)$ there exists a $P_{\epsilon,l} \in \mathcal{M}$ such that for $\epsilon_n = 1/\sqrt{n}$

$$L_{P_{\epsilon_n,l}}\left(\sqrt{n}(b\theta_n - b\vartheta(P_{\epsilon_n,l}))\right) \to L_b.$$

The smaller we choose $S(P)$ the larger the class of regular estimators (relative to $S(P)$) and the easier it is to verify pathwise differentiability. On the other hand the lower bound $\text{Var}(\tilde{I}(P, b\vartheta))$ represents a supremum over all Cramér-Rao lower bounds for the one- dimensional submodels $p_{\epsilon,l}$ and therefore this lower bound can only be attained if $S(P)$ is large enough and thus in order to have existence of efficient estimators one has to choose a rich enough class $S(P)$ of one-dimensional submodels $p_{\epsilon,l}$.

We consider asymptotically linear estimators:

Definition 3.1.2.6 *An estimator θ_n of $\theta = \vartheta(P)$ is called $\|\cdot\|_B$-asymptotically linear with influence curve $I(P, b\vartheta) \in L_0^2(P)$, $b \in B$, if*

$$\sqrt{n}(b\theta_n - b\theta) = \sqrt{n}(P_n - P)I(P, b\vartheta) + R_{n,b},$$

where $\|R_n\|_B \equiv \sup_{b\in B} |R_{n,b}| = o_P(1)$ and the empirical process $\int I(P, b\vartheta)\, d\sqrt{n}(P_n - P)$ indexed by $\{I(P, b\vartheta) : b \in B\}$ converges weakly.

Theorem 2.12 in Van der Vaart(1988) says that for any regular estimator $b\theta_n$ the limiting distribution L_b has a variance which is larger than $\text{Var}(\tilde{I}(P, b\vartheta))$ and that equality holds iff $b\theta_n$ is asymptotically linear with influence curve equal to $\tilde{I}(P, b\vartheta)$. One may call this result a asymptotic Cramér-Rao bound. This leads to the following definition

Definition 3.1.2.7 *$\tilde{I}(P, b\vartheta) \in T(P)$ is called the efficient influence curve w.r.t. $S(P)$ for estimating $b\vartheta(P)$ in $\mathcal{M}$.*

The convolution theorem tells us that if $S(P)$ is convex, then the limiting distribution L_b of a regular estimator $b\theta_n$ equals the sum of $N(0, \text{Var}_P(\tilde{I}(P, b\vartheta))$ and another independent random variable. Now we give the following definition of efficiency of θ_n:

Definition 3.1.2.8 *Let θ, $\theta_n \in D$ and B be a collection of real valued linear functions on D. Assume that θ_n is $\|\cdot\|_B$-asymptotically linear with efficient influence curve $\tilde{I}(P, b\vartheta)$, $b \in B$. Then we say that θ_n is $\|\cdot\|_B$-efficient.*

Later in this chapter we will prove efficiency of an estimator $(\widehat{Z}_n, \widehat{\mathcal{Z}}_n)$, which is a 1–1 function of the NPMLE $(\widehat{V}_n, \widehat{h}_n)$ of the underlying (V, h) in the two-dimensional line segment problem (see section 3.5). Furthermore in section 3.12 we prove efficiency of an estimator $(\widehat{W}_n, \widehat{\mathcal{W}}_n)$, which is a 1–1 function of the NPMLE $(\widehat{V}_n, \widehat{h}_n)$ of the underlying (V, h) in the one-dimensional case. (The reason why we do not show the efficiency of $(\widehat{V}_n, \widehat{h}_n)$ directly is the fact that in that case one has to deal with severe singularity problems. We discuss this later.) As we remember we obtained (V, h) after some reparametrization of (F, μ). Now we want to answer the question if efficiency of the NPMLE $(\widehat{Z}_n, \widehat{\mathcal{Z}}_n)$ (and $(\widehat{W}_n, \widehat{\mathcal{W}}_n)$ in the one-dimensional case) implies efficiency of the NPMLE $(\widehat{F}_n, \widehat{\mu}_n)$ of the original parameters (F, μ)? The answer to this question is formulated in theorem 3.1.1. We apply this theorem in section 3.5.3 and 3.12.1.

Let $\Phi : D_\phi \subset (D, \|\cdot\|) \to (E, \|\cdot\|_1)$, where $(D, \|\cdot\|)$ and $(E, \|\cdot\|_1)$ are normed vector spaces. Suppose that D_n, D_0, $D_\phi \subset D$ such that if G_n, $G \in D_\phi$ then $h_n \equiv \sqrt{n}(G_n - G) \in D_n$.

We say that Φ is compact differentiable if we have that if $h_n \to h$, $h \in D_0$ and D_0 separable then

$$\sqrt{n}\left(\Phi(G + (1/\sqrt{n})h_n) - \Phi(G)\right) - \mathrm{d}\Phi(G)(h) \to 0 \tag{3.6}$$

for a certain continuous linear mapping $\mathrm{d}\Phi(G) : D_0 \subset (D, \|\cdot\|) \to (E, \|\cdot\|_1)$.

The next theorem gives us that efficiency is preserved under this kind of differentiability (Van der Vaart(1991)).

Theorem 3.1.1 *Let B and B_1 be a collection of real valued linear functions on vector spaces D and E respectively, such that $(D, \|\cdot\|_B)$ and $(E, \|\cdot\|_{B_1})$ are normed vector spaces. Let $\Phi : D_\phi \subset (D, \|\cdot\|_B) \to (E, \|\cdot\|_{B_1})$ be a functional.*

If $\theta_n \in D_\phi$ is an $\|\cdot\|_B$-efficient estimator of $\theta \in D_\phi$ and Φ is compact differentiable, then $\Phi(\theta_n)$ is an $\|\cdot\|_{B_1}$-efficient estimator of $\Phi(\theta)$.

3.2 Efficiency theorem for an NPMLE

In this section and the next section we give Van der Laan's approach for linear parameters in convex models with some little improvement.

Let us assume the existence of an (NP)MLE $\mathbb{P}_n$ and let $\mathcal{S}(\mathbb{P}_n)$ be a class of one-dimensional differentiable submodels of $\mathcal{M}$ through $\mathbb{P}_n$ and let $S(\mathbb{P}_n) \subset L_0^2(\mathbb{P}_n)$ be the tangent cone corresponding to this class of submodels. Let $T(\mathbb{P}_n)$ be the tangent space at $\mathbb{P}_n$.

We suppose that $b\vartheta$ is pathwise differentiable relative to $S(\mathbb{P}_n)$ at $\mathbb{P}_n$ with efficient influence curve $\tilde{I}(\mathbb{P}_n, b\vartheta) \in T(\mathbb{P}_n)$. Let $\mathbb{P}_{n,\epsilon,l_n}$ be a one-dimensional submodel through $\mathbb{P}_n$ with score l_n and let $\mathbb{P}_{n,\epsilon,l_n} \ll \nu_n$.

If $\mathbb{P}_n$ lies in the interior of $\mathcal{M}$, then one obtains, because $\mathbb{P}_n$ is NPMLE, that the derivative w.r.t. ϵ of the loglikelihood along the submodel $\mathbb{P}_{n,\epsilon,l_n}$ evaluated in $\epsilon = 0$ equals 0. So we have

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} \int \log\left(\frac{\mathrm{d}\mathbb{P}_{n,\epsilon,l_n}}{\nu_n}(x)\right) \mathrm{d}P_n(x)\bigg|_{\epsilon=0} = 0.$$

By exchanging differentiation and integration this yields

$$\int l_n(x)\,\mathrm{d}P_n(x) = 0.$$

One notes that this holds for all $l_n \in S(\mathbb{P}_n)$ and by the linearity of $l \to \int l\,\mathrm{d}P_n$ this also hold for $\mathrm{Lin}(S(\mathbb{P}_n))$, the linear extension of $S(\mathbb{P}_n)$.

Now if we have $\tilde{I}(\mathbb{P}_n, b\vartheta) \in \mathrm{Lin}(S(\mathbb{P}_n))$, then one can write

$$\int \tilde{I}(\mathbb{P}_n, b\vartheta)(x)\,\mathrm{d}P_n(x) = P_n\tilde{I}(\mathbb{P}_n, b\vartheta) = 0. \tag{3.7}$$

If $\tilde{I}(\mathbb{P}_n, b\vartheta) \in T(\mathbb{P}_n) \setminus S(\mathbb{P}_n)$, then it might still be possible to prove (3.7) by a continuity argument. Actually for the theory below we only need

$$P_n\tilde{I}(\mathbb{P}_n, b\vartheta) = o_P(1/\sqrt{n}) \tag{3.8}$$

(Compare with the efficient score equation in theorem 3.2.1). Note that because $T(P) \subset L_0^2(P)$, we always have $P\tilde{I}(P, b\vartheta) = 0$ and thus for $P = \mathbb{P}_n$: $\mathbb{P}_n\tilde{I}(\mathbb{P}_n, b\vartheta) = 0$.

In Van der Laan(1993) we find the following argument which provides a proof of theorem 3.2.1. By definition 3.1.2.8 we know that θ_n is $\|\cdot\|_B$-efficient if and only if we have

$$b\theta_n - b\theta = \int \tilde{I}(P, b\vartheta)\,\mathrm{d}(P_n - P) + R_{n,b}, \tag{3.9}$$

where $\|R_n\|_B = o_P(1/\sqrt{n})$ and $\{\tilde{I}(P, b\vartheta) : b \in B\}$ is P-Donsker. One easily sees that $\|R_n\|_B = o_P(1/\sqrt{n})$ can be weakened to $\|R_n\|_B = o_P(\|\theta_n - \theta\|_B)$ and so we get with (3.9) that $\|\theta_n - \theta\|_B = O_P(1/\sqrt{n}) + o_P(\|\theta_n - \theta\|_B)$. This yields $\|\theta_n - \theta\|_B = O_P(1/\sqrt{n})$.

Now if we assume that $P_n\tilde{I}(\mathbb{P}_n, b\vartheta) = o_P(1/\sqrt{n}$, then (3.9) holds if (and only if)

$$\begin{aligned}&\sup_{b\in B}\left|b\theta_n - b\theta + \int \tilde{I}(\mathbb{P}_n, b\vartheta)\,\mathrm{d}P - \int \left(\tilde{I}(P, b\vartheta) - \tilde{I}(\mathbb{P}_n, b\vartheta)\right)\mathrm{d}(P_n - P)\right| \\ &\quad = o_P(\|\theta_n - \theta\|_B).\end{aligned} \tag{3.10}$$

If we suppose that there exists a P-Donsker class $\mathcal{F}$ such that $\tilde{I}(P, b\vartheta) - \tilde{I}(\mathbb{P}_n, b\vartheta) \in \mathcal{F}$ for all $b \in B$ with probability tending to 1, then it follows by the $\|\cdot\|_P$ continuity of the limiting sample paths that if $\sup_{b\in B} \rho_P(\tilde{I}(P, b\vartheta), \tilde{I}(\mathbb{P}_n, b\vartheta)) \to 0$ in probability, then we have $\sup_{b\in B} |\int (\tilde{I}(P, b\vartheta) - \tilde{I}(\mathbb{P}_n, b\vartheta))\,\mathrm{d}(P_n - P)| = o_P(1/\sqrt{n})$. (Note that

$$\rho_P(f, g)^2 \equiv \int ((f - g) - P(f - g))^2\,\mathrm{d}P\ ).$$

Showing that the sum of the other terms on the left-hand side of equation (3.10) is $o_P(\|\theta_n - \theta\|_B)$ gives us (3.9). We obtain the following theorem:

Theorem 3.2.1 *Let $X \sim P \in \mathcal{M}$ for a model $\mathcal{M}$ and let $X_1, ..., X_n$ be n i.i.d. copies of X. Let $\theta = \vartheta(P) \in D$, D a vector space and let B be a certain collection of real valued linear mappings on D. Suppose that for each $P \in \mathcal{M}$, $b\vartheta$ ($b \in B$) is pathwise differentiable at P relative to $S(P)$ with efficient influence function $\tilde{I}(P, b\vartheta)$.*

Let $\theta_n \equiv \vartheta(\mathbb{P}_n)$, $\mathbb{P}_n \in \mathcal{M}$ be an estimator of θ which satisfies the following conditions:

Efficient score equation:

$$\sup_{b\in B}\left|\int \tilde{I}(\mathbb{P}_n, b\vartheta)\,\mathrm{d}P_n\right| = o_P(1/\sqrt{n}),$$

Differentiability condition:

$$\sup_{b\in B}\left|b\vartheta(P) - b\vartheta(\mathbb{P}_n) - \int \tilde{I}(\mathbb{P}_n, b\vartheta)\,\mathrm{d}(P - \mathbb{P}_n)\right| = o_P(\|\theta_n - \theta\|_B),$$

Empirical process condition:

$$\sup_{b\in B}\left|\int \left(\tilde{I}(P, b\vartheta) - \tilde{I}(\mathbb{P}_n, b\vartheta)\right)\mathrm{d}(P_n - P)\right| = o_P(1/\sqrt{n}).$$

Then θ_n is a $\|\cdot\|_B$-asymptotically efficient estimator of θ.

We know we have

Theorem 3.2.2 *Sufficient conditions for the empirical process condition in theorem 3.2.1 are:*

P-Donsker class condition:

There exists a P-Donsker class $\mathcal{F}$ such that $\tilde{I}(P, b\vartheta) - \tilde{I}(\mathbb{P}_n, b\vartheta) \in \mathcal{F}$ for all $b \in B$ with probability tending to 1,

ρ_P-consistency:

$$\sup_{b\in B} \rho_P \left(\tilde{I}(P, b\vartheta), \tilde{I}(\mathbb{P}_n, b\vartheta)\right) \to 0 \textit{ in probability} .$$

3.3 Efficiency of NPMLE of linear parameters in convex models

In Van der Laan(1993) we find theorem 3.3.1 where an identity for linear parameters in convex models is given.

Let $\mathcal{M}$ be a convex set of probability measures and note that this implies that $\mathcal{M}(\nu)$ is convex. We recall that $\mathcal{M}(\nu) = \{P \in \mathcal{M} : P \ll \nu\}$. For every $P_1 \in \mathcal{M}(P)$ the line $\epsilon P_1 + (1-\epsilon)P$ with $\epsilon \in [0,1]$ is a submodel in $\mathcal{M}$ through P. If ν is a dominating measure for P, then the corresponding line $\epsilon p_1 + (1-\epsilon)p$ of densities w.r.t. ν can be given by ($p_1 = \mathrm{d}P_1/\mathrm{d}\nu$ and $p = \mathrm{d}P/\mathrm{d}\nu$): $p_{\epsilon,l} = (1+\epsilon l)p$, where $l = (p_1 - p)/p$. The score is given by l and if $l \in L_0^2(P)$, then $p_{\epsilon,l}$ is Hellinger differentiable because it satisfies (3.3). For efficiency calculations a natural class of one-dimensional submodels through P is

$$\mathcal{S}(P) \equiv \left\{\epsilon P_1 + (1-\epsilon)P,\ \epsilon \in [0,1] : P_1 \in \mathcal{M}(P),\ \frac{\mathrm{d}P_1}{\mathrm{d}P} \in L^2(P)\right\} \tag{3.11}$$

and in terms of densities this class (3.11) is given by

$$\left\{p_{\epsilon,l} = (1+\epsilon l)p : l = (p_1 - p)/p \in L_0^2(P),\ P_1 \in \mathcal{M}(P)\right\}.$$

The tangent cone $S(P)$ and the tangent space $T(P)$ are defined as in definition 3.1.2.2 and 3.1.2.3.

Now we give theorem 3.3.1:

Theorem 3.3.1 *Suppose that $\mathcal{M}$ is a convex model and $\vartheta : \mathcal{M} \to D$ is linear. Suppose $P,\ P_1 \in \mathcal{M}$ and that $b\vartheta$ is pathwise differentiable at P_1 relative to $S(P_1)$ with efficient influence curve $\tilde{I}(P_1, b\vartheta)$.*

Assume that the following condition holds:

Identity condition:

There exists a sequence $P_m \in \mathcal{M}(P_1)$ with $\mathrm{d}P_m/\mathrm{d}P_1 \in L^2(P_1)$ such that
$$\int \tilde{I}(P_1, b\vartheta)\,\mathrm{d}P_m \to \int \tilde{I}(P_1, b\vartheta)\,\mathrm{d}P \quad \textit{and} \quad b\vartheta(P_m) \to b\vartheta(P) \quad \textit{for } m \to \infty.$$

Then we have the following identity

$$b\vartheta(P) - b\vartheta(P_1) = \int \tilde{I}(P_1, b\vartheta)\, \mathrm{d}(P - P_1) = \int \tilde{I}(P_1, b\vartheta)\, \mathrm{d}P.$$

So for linear parameters in convex models the identity gives us the differentiability condition with remainder zero in theorem 3.2.1. One obtains immediately the following theorem:

Theorem 3.3.2 *Let $X \sim P \in \mathcal{M}$ for a convex model $\mathcal{M}$ and let $X_1, ..., X_n$ be n i.i.d. copies of X. Let $\theta = \vartheta(P) \in D$ be a linear parameter, D a vector space and let B be a certain collection of real valued linear mappings on D. Suppose that for each $P \in \mathcal{M}$, $b\vartheta$ $(b \in B)$ is pathwise differentiable at P relative to $S(P)$ with efficient influence function $\tilde{I}(P, b\vartheta)$.*

Let $\theta_n \equiv \vartheta(\mathbb{P}_n)$, $\mathbb{P}_n \in \mathcal{M}$ be an estimator of θ (typically the NPMLE) which satisfies the following conditions:

Efficient score equation in theorem 3.2.1,
Identity condition in theorem 3.3.1 with P as here and $P_1 = \mathbb{P}_n$,
P-Donsker class condition in theorem 3.2.2,
ρ_P-consistency in theorem 3.2.2.

Then θ_n is a $\|\cdot\|_B$-asymptotically efficient estimator of θ.

In the identity condition the existence of a sequence P_m had to be shown. Instead of constructing or showing the existence of such a sequence P_m, we introduce in theorem 3.3.3 a $\|\cdot\|_P$-convergence condition and show that this condition implies not only the identity of theorem 3.3.1 but even the ρ_P-consistency. The theorem is formulated as follows

Theorem 3.3.3 *Let $X \sim P \in \mathcal{M}$ for a convex model $\mathcal{M}$ and let $X_1, ..., X_n$ be n i.i.d. copies of X. Let $\theta = \vartheta(P) \in D$ be a linear parameter, D a vector space and let B be a certain collection of real valued linear mappings on D. Suppose that for each $P \in \mathcal{M}$, $b\vartheta$ $(b \in B)$ is pathwise differentiable at P relative to $S(P)$ with efficient influence function $\tilde{I}(P, b\vartheta)$.*

Let $\theta_n \equiv \vartheta(\mathbb{P}_n)$, $\mathbb{P}_n \in \mathcal{M}$ be an estimator of θ. If the following conditions hold:

$\|\cdot\|_P$-convergence conditions:

$$(1)\quad \lim_{n\to\infty} \sup_{b\in B} \|\tilde{I}(\mathbb{P}_n, b\vartheta) - \tilde{I}(P, b\vartheta)\|_P = 0 \quad \textit{(in probability)}$$

$$(2)\quad \lim_{\epsilon\downarrow 0} \|\tilde{I}((1-\epsilon)\mathbb{P}_n + \epsilon P, b\vartheta) - \tilde{I}(\mathbb{P}_n, b\vartheta)\|_P = 0 \quad \textit{(a.s.)}.$$

Then this implies the ρ_P-consistency in theorem 3.2.2 and the identity in theorem 3.3.1.

PROOF: Firstly, because

$$\rho_P(f,g) \equiv \sqrt{\int ((f-g) - P(f-g))^2\, \mathrm{d}P} \le 2\sqrt{\int (f-g)^2\, \mathrm{d}P} = 2\,\|f-g\|_P$$

we have that

$$\lim_{n\to\infty}\sup_{b\in B}\|\tilde{I}(\mathbb{P}_n, b\vartheta) - \tilde{I}(P, b\vartheta)\|_P = 0 \quad \text{(in probability)}$$

implies

$$\sup_{b\in B}\rho_P\left(\tilde{I}(P, b\vartheta), \tilde{I}(\mathbb{P}_n, b\vartheta)\right) \to 0 \quad n \to \infty \text{ (in probability)}.$$

Secondly, for convenience we define $P(\epsilon) = (1-\epsilon)\mathbb{P}_n + \epsilon P$ and let $P_{\eta,l} = \eta P + (1-\eta)P(\epsilon)$ be the differentiable submodel through $P(\epsilon)$ with score

$$l = \frac{\mathrm{d}P - \mathrm{d}P(\epsilon)}{\mathrm{d}P(\epsilon)} = \frac{\mathrm{d}P - \mathrm{d}((1-\epsilon)\mathbb{P}_n + \epsilon P)}{\mathrm{d}((1-\epsilon)\mathbb{P}_n + \epsilon P)}.$$

Note that l is well defined, $l \in S(P(\epsilon)) \subseteq L_0^2$. That $\int l\,\mathrm{d}P(\epsilon) = 0$ is obvious and from

$$\begin{aligned} |l| &= \left|\frac{\mathrm{d}P - \mathrm{d}P(\epsilon)}{\mathrm{d}P(\epsilon)}\right| \le \frac{\mathrm{d}P}{\mathrm{d}P(\epsilon)} + 1 \\ &= \frac{\mathrm{d}P}{\mathrm{d}((1-\epsilon)\mathbb{P}_n + \epsilon P)} + 1 \le \frac{\mathrm{d}P}{\epsilon \mathrm{d}P} + 1 = \frac{1}{\epsilon} + 1 \end{aligned}$$

we conclude that l is square integrable. Now by linearity of $b\vartheta$ and because of the pathwise differentiability of $b\vartheta$ at $P(\epsilon)$ relative to $S(P(\epsilon))$ we may write

$$\begin{aligned} (1-\epsilon)(b\vartheta(P) - b\vartheta(\mathbb{P}_n)) &= b\vartheta(P) - b\vartheta((1-\epsilon)\mathbb{P}_n + \epsilon P) \\ &= b\vartheta(P) - b\vartheta(P(\epsilon)) \\ &= \frac{1}{\eta}\left(b\vartheta(\eta P + (1-\eta)P(\epsilon)) - b\vartheta(P(\epsilon))\right) \\ &= \frac{1}{\eta}\left(b\vartheta(P_{\eta,l}) - b\vartheta(P(\epsilon))\right) \end{aligned}$$

and this yields

$$\begin{aligned} (1-\epsilon)(b\vartheta(P) - b\vartheta(\mathbb{P}_n)) &= \int \tilde{I}(P(\epsilon), b\vartheta)\, l\,\mathrm{d}P(\epsilon) \\ &= \int \tilde{I}(P(\epsilon), b\vartheta)\,\mathrm{d}(P - P(\epsilon)) \\ &= \int \tilde{I}(P(\epsilon), b\vartheta)\,\mathrm{d}P \\ &= \int \tilde{I}(\mathbb{P}_n, b\vartheta)\,\mathrm{d}P + \int \tilde{I}(P(\epsilon), b\vartheta)\,\mathrm{d}P - \int \tilde{I}(\mathbb{P}_n, b\vartheta)\,\mathrm{d}P \\ &= \int \tilde{I}(\mathbb{P}_n, b\vartheta)\,\mathrm{d}P + \int \left(\tilde{I}(P(\epsilon), b\vartheta) - \tilde{I}(\mathbb{P}_n, b\vartheta)\right)\mathrm{d}P. \quad (3.12) \end{aligned}$$

Now if we have that

$$\lim_{\epsilon\downarrow 0}\|\tilde{I}(P(\epsilon), b\vartheta) - \tilde{I}(\mathbb{P}_n, b\vartheta)\|_P = \lim_{\epsilon\downarrow 0}\|\tilde{I}((1-\epsilon)\mathbb{P}_n + \epsilon P, b\vartheta) - \tilde{I}(\mathbb{P}_n, b\vartheta)\|_P = 0 \quad \text{(a.s.)}$$

and thus

$$\left|\int \left(\tilde{I}(P(\epsilon), b\vartheta) - \tilde{I}(\mathbb{P}_n, b\vartheta)\right)\mathrm{d}P\right|^2 \le \|\tilde{I}(P(\epsilon), b\vartheta) - \tilde{I}(\mathbb{P}_n, b\vartheta)\|_P \to 0$$

($\epsilon \downarrow 0$), then we immediately obtain from (3.12) the identity

$$b\vartheta(P) - b\vartheta(\mathbb{P}_n) = \int \tilde{I}(\mathbb{P}_n, b\vartheta)\, \mathrm{d}P.$$

We have proved the theorem. □

The preceding argument showed that the $\|\cdot\|_P$-convergence condition (2) could be weakened to

$$\lim_{\epsilon \downarrow 0} \int \left(\tilde{I}(P(\epsilon), b\vartheta) - \tilde{I}(\mathbb{P}_n, b\vartheta)\right) \mathrm{d}P = 0.$$

In most applications if one wants to prove the ρ_P-consistency, then one proves actually the $\|\cdot\|_P$-convergence condition (1), but with a little more effort one mostly obtains (2) too. One could replace the $\|\cdot\|_P$-convergence conditions by the stronger condition:

$$\lim_{k\to\infty} \sup_{b\in B} \|\tilde{I}((1-\epsilon(k))P_1 + \epsilon(k)P_0, b\vartheta) - \tilde{I}(P_i, b\vartheta)\|_P \to 0 \tag{3.13}$$

($i = 0, 1$) for all sequences $\epsilon(k) \downarrow 0$ and $P_0, P_1 \in \mathcal{M}$. It is obvious that (3.13) implies the $\|\cdot\|_P$-convergence condition.

Often one is able to derive an inequality like

$$\sup_{b\in B} \|\tilde{I}(P_k, b\vartheta) - \tilde{I}(P_0, b\vartheta)\|_P \leq c \cdot \|F_k - F_0\|_\infty \tag{3.14}$$

(c constant), where F_k and F_0 denote the distribution functions of the measures P_k and P_0 respectively. If we take $P_k = \mathbb{P}_k$ ($n = k$) and $P_0 = P$ and if we have consistency of F_k to F_0, then this provides us with $\|\cdot\|_P$-convergence condition (1). If we take $P_k = (1-\epsilon(k))\mathbb{P}_n + \epsilon(k)P$ for $\epsilon(k) \downarrow 0$ and $P_0 = \mathbb{P}_n$ and we have consistency of F_k to F_0, then we get $\|\cdot\|_P$-convergence condition (2). In fact for the two-dimensional problem this will be done in section 3.9.

Now with theorem 3.3.3 we can give a slightly improved version of theorem 3.3.2. The improvement must be interpreted as avoiding the laborious verification of the identity condition in theorem 3.3.1 and getting through (3.13) or (3.14) the identity and the ρ_P-consistency (thus the $\|\cdot\|_P$-convergence condition) in one move. Theorem 3.3.2 and theorem 3.3.3 imply:

Theorem 3.3.4 *Let $X \sim P \in \mathcal{M}$ for a convex model $\mathcal{M}$ and let $X_1, ..., X_n$ be n i.i.d. copies of X. Let $\theta = \vartheta(P) \in D$ be a linear parameter, D a vector space and let B be a certain collection of real valued linear mappings on D. Suppose that for each $P \in \mathcal{M}$, $b\vartheta$ ($b \in B$) is pathwise differentiable at P relative to $S(P)$ with efficient influence function $\tilde{I}(P, b\vartheta)$.*

Let $\theta_n \equiv \vartheta(\mathbb{P}_n)$, $\mathbb{P}_n \in \mathcal{M}$ be an estimator of θ which satisfies the following conditions:

Efficient score equation in theorem 3.2.1,
P-Donsker class condition in theorem 3.2.2,
$\|\cdot\|_P$-convergence conditions in theorem 3.3.3 .

Then θ_n is a $\|\cdot\|_B$-asymptotically efficient estimator of θ.

Before finishing this section we want to say something about the calculation of the efficient influence function. Consider the model $\mathcal{M} = \{P_V : V \in \mathcal{V}\}$, where $\mathcal{V}$ is a convex set. Suppose that $\mathcal{V}$ is a convex set of probability measures. Let the mapping $V \to P_V$ be linear. Thus $\mathcal{M}$ is convex.

If $V_{\epsilon,l} = \epsilon V_1 + (1-\epsilon)V$ and $\epsilon \in [0,1]$ is a line from V to V_1 and $V_1 \ll V$ with score $l = (\mathrm{d}V_1 - \mathrm{d}V)/\mathrm{d}V \in L_0^2(V)$, then by linearity of $V \to P_V$ this line gives (assuming $\mathrm{d}P_{V_1}/\mathrm{d}P_V$ exists and it is square integrable) a submodel $P_{V_{\epsilon,l}} = \epsilon P_{V_1} + (1-\epsilon)P_V$ with score $(\mathrm{d}P_{V_1} - \mathrm{d}P_V)/\mathrm{d}P_V \in L_0^2(P_V)$. Note that this score is linear in l. Now we define the linear score operator A_V as follows

$$A_V(l) \equiv \frac{\mathrm{d}P_{V_1} - \mathrm{d}P_V}{\mathrm{d}P_V}$$

and denote the adjoint of A_V by A_V^T:

$$\langle A_V(l), v\rangle_{P_V} = \langle l, A_V^T(v)\rangle_V,$$

for all $l \in T(V) \subset L_0^2(V)$ and $v \in L_0^2(P_V)$. The so called information operator I_V is defined as $I_V \equiv A_V^T A_V$. If f_1 solves $I_V f_1 = f_2$, then we write $f_1 = I_V^{-1} f_2$ even if f_1 is not uniquely determined by $I_V f_1 = f_2$.

Now there is a linear mapping Γ such that $\Gamma(V) = \vartheta(P_V)$. Suppose that $b\Gamma$ is pathwise differentiable at V relative to $S(V)$ with efficient influence function $\tilde{J}(V, b\Gamma)$ and suppose that $b\vartheta$ is pathwise differentiable at P_V relative to $S(P_V)$ with efficient influence function $\tilde{I}(P_V, b\vartheta)$, then we have for all $l \in S(V)$

$$\begin{aligned}
\langle \tilde{J}(V, b\Gamma), l\rangle_V &= \frac{1}{\epsilon}(b\Gamma(V_{\epsilon,l}) - b\Gamma(V)) \\
&= \frac{1}{\epsilon}(b\vartheta(P_{V_{\epsilon,l}}) - b\vartheta(P)) \\
&= \langle \tilde{I}(P_V, b\vartheta), A_V(l)\rangle_{P_V}.
\end{aligned}$$

If $\tilde{J}(V, b\Gamma)$ lies in the range of the information operator I_V, then we obtain from this

$$\langle A_V I_V^{-1}(\tilde{J}(V, b\Gamma)), A_V(l)\rangle_{P_V} = \langle \tilde{I}(P_V, b\vartheta), A_V(l)\rangle_{P_V}$$

for all $l \in S(V)$. This implies (under the condition that $\tilde{J}(V, b\Gamma)$ lies in the range of I_V) that

$$\tilde{I}(P_V, b\vartheta) = A_V I_V^{-1}(\tilde{J}(V, b\Gamma)).$$

We see that calculation of I_V^{-1} is an important step to find an expression for the influence function $\tilde{I}(P_V, b\vartheta)$. By Gill(1989), Bickel et al.(1993) we know that in missing data models the score operator $A_V : L^2(V) \to L^2(P_V)$ is given by a conditional expectation:

$$A_V(h)(Y) = E_V(h(X) \mid Y) \tag{3.15}$$

and the adjoint $A_V^T : L^2(P_V) \to L^2(V)$ of A_V by

$$A_V^T(v)(X) = E_V(v(Y) \mid X). \tag{3.16}$$

For the information operator $I_V = A_V^T A_V$ we get then $E_V(E_V(h(X) \mid Y) \mid X)$.

3.4 The assumptions

In this section we formulate the assumptions that we use to obtain the asymptotic results in the next sections. For the two-dimensional circle-case we consider:

Assumption I: for the underlying distribution function V we have that $1/g(2R)$ and $1/d(2R, 2R)$ are bounded (or $g(2R) > 0$ and $d(2R, 2R) > 0$). Because $g(x)$ and $d(x, x)$ are decreasing on $[0, 2R]$, this is the same as saying: $1/g(x)$ and $1/d(x, x)$ are bounded on $[0, 2R]$.

Assumption II: for the underlying distribution function V we have that the determinant Q_V of the matrix N_V in (3.70) is unequal to 0. In the sections 3.10.1 and 4.4 we discuss the necessity of this assumption. We conjecture that $Q_V \geq 1$ for all V.

Assumption III: for the underlying distribution function V we have that there is a $2R_0 \in [0, 2R)$ such that V gives no mass to the interval $(2R_0, 2R)$.

Assumption IV: for the sieved NPMLE $\widehat{V}_n$ of the underlying V we have that $\widehat{V}_n(2R-) \to V(2R-)$ in probability.

Because of theorem 2.3.1 and the equation $\widehat{V}_n(2R-) + (|W| + 4R^2)\widehat{g}_n(2R) + \widehat{h}_n = 1$ this implies that $\widehat{g}_n(2R) \to g(2R)$. Of course this implies $\sup_{x\in[0,2R)} |\widehat{V}_n(x) - V(x)| \to 0$ and $\sup_{x\in[0,2R)} |\widehat{g}_n(x) - g(x)| \to 0$. One notes that if we have assumption III, then the sieved NPMLE $\widehat{V}_n$ automatically satisfies assumption IV (compare with corollary 2.2.6). Surely we do not need all the four assumptions at the same time: either we use **the assumptions I,II and III**, or we use **the assumptions I,II and IV**. The first set of assumptions restricts the class of underlying distributions V. In the second set of assumptions we must admit that we are not pleased with assumption IV, because it assumes something (maybe crucial) of the estimator, which we want to investigate. Assumption II is discussed in section 3.10.1 and 4.4. To get rid of assumption IV we give a suggestion in section 4.1 and in section 3.10.2 we look at assumption I.

Because in the one-dimensional case we can prove that the determinant Q_V is not equal to 0, we consider for this case the assumptions:

Assumption (i): for the underlying distribution function V we have that $1/g(\tau)$ is bounded.

Assumption (ii): for the underlying distribution function V we have that there is a $\tau_0 \in [0, \tau)$ such that V gives no mass to the interval (τ_0, τ).

Assumption (iii): for the (sieved) NPMLE $\widehat{V}_n$ of the underlying V we have that $\widehat{V}_n(\tau-) \to V(\tau-)$ in probability.

Assumption (iii) implies $\sup_{x\in[0,\tau)}|\widehat{V}_n(x)-V(x)|\to 0$ and $\sup_{x\in[0,\tau)}|\widehat{g}_n(x)-g(x)|\to 0$. We use the assumptions (i) and (ii) or we use the assumptions (i) and (iii) (use corollary 2.2.6). Just as for assumption IV we suggest in section 4.1 an alternative.

3.5 The parameters to be estimated

Here we consider the two-dimensional line segment process observed through a circular window. We remember the definition of the set $\mathcal{V}_{2R}$ in section 1.2.3. We know that we can parametrize the distribution of the data as $P_{V,h}$, where $(V,h)\in\mathcal{V}_{2R}$. Actually we have identifiability: $P_{V_1,h_1}=P_{V_2,h_2}\Leftrightarrow(V_1,h_1)=(V_2,h_2)$. One writes

$$P_{V,h}(\mathrm{d}\tilde{x},\{d\})=1(d=0)\cdot\mathrm{d}F^{u.c.}(\tilde{x})+1(d=1)\cdot\mathrm{d}F^{s.e.c.}(\tilde{x})+1(d=2)\cdot\mathrm{d}F^{d.c.}(\tilde{x}),$$

where the subdistribution functions $F^{u.c.}$, $F^{s.e.c.}$ and $F^{d.c.}$ are defined as in (1.69)–(1.71). Now the model is given by

$$\mathcal{M}\equiv\{P_{V,h}\,:\,(V,h)\in\mathcal{V}_{2R}\}.$$

It would be natural to think of estimating the parameter $\vartheta(P_{V,h})=(V(\cdot),h)$ where we define $b_t\vartheta_1(P_{V,h})=V(t)$ with $B=\{b_t\,:\,t\in[0,2R)\}$ and $\vartheta_2(P_{V,h})=h$. It turns out that finding the hardest submodel for $d(0,0)$ (see (1.73)) is easier than for h. For each $t\in[0,2R)$ calculations to obtain the hardest submodel for $V(t)$ are not difficult. Only for $V(2R)$ it is hard to do or maybe not possible. This causes troubles to get a uniformity result for $V(t)$, $t\in[0,2R)$. To avoid the singularity difficulties at the point $2R$ we consider the estimation of the parameter

$$\begin{aligned}\vartheta(P_{V,h}) &= (\vartheta_1(P_{V,h}),\vartheta_2(P_{V,h}))\\ &\equiv (Z(\cdot),\mathcal{Z})\\ &\equiv \left(\int_{x=0}^{x=\cdot}\frac{z(x)}{|W|+2xR}\mathrm{d}V(x)\,,\,\int_{x=0}^{x=\infty}\frac{x}{|W|+2xR}\mathrm{d}V(x)\right)\\ &= (F^{u.c.}(\cdot)\,,\,d(0,0)),\end{aligned}\tag{3.17}$$

where $z(x)$ is defined as in (1.74). We define $b_t\vartheta_1(P_{V,h})=Z(t)$ with $B=\{b_t\,:\,t\in[0,2R)\}$. We immediately see that the parameter ϑ_1 is the subdistribution function of the uncensored observations.

We take for the second parameter $\vartheta_2(P_{V,h})=\mathcal{Z}=d(0,0)$. In section 3.5.2 we will prove that the NPMLE $\widehat{\mathcal{Z}}_n$ is an efficient estimator of $\mathcal{Z}$. In section 3.5.3 we see that the relation $d(0,0)=(\mu/(|W|+2\mu R))$ gives us the possibility to express $\widehat{\mu}_n$ as a compact differentiable mapping in the efficient parameter $\widehat{\mathcal{Z}}_n$. This gives us the efficiency of $\widehat{\mu}_n$. Of course one could take $\mathcal{Z}=g(0)$. In this case one gets similar calculations to section 3.11 of the efficient influence curve and the relation $g(0)=(1/(|W|+2\mu R))$ would give us the efficiency of $\widehat{\mu}_n$. Note that $|W|g(0)$ is the probability of being a s.e.c.r. or u.c. (or a s.e.c.l. or u.c.) observation.

Now let us calculate the efficient influence curves w.r.t. $S(P_{V,h})$ for estimating $Z(t)$ and $\mathcal{Z}$ in $\mathcal{M}$. From now on instead of writing $P_{V,h}$ with $(V,h)\in\mathcal{V}_{2R}$, we write for convenience and without loss of generality the distribution function of the data as P_V with V in the set of distribution functions on $[0,\infty)$.

By (3.15) we know that the score operator $A_V : L_0^2(V) \to L_0^2(P_V)$ is given by

$$A_V(l)(\widetilde{X}, D, \Theta) = E_V(l(X) \mid \widetilde{X}, D, \Theta)$$

and its adjoint $A_V^T : L_0^2(P_V) \to L_0^2(V)$ by

$$A_V^T(\eta)(X) = E_V(\eta(\widetilde{X}, D, \Theta) \mid X).$$

The information operator $I_V : L_0^2(V) \to L_0^2(V)$ is given by

$$I_V = A_V^T A_V.$$

If f_1 solves $I_V f_1 = f_2$, then we write $f_1 = I_V^{-1} f_2$ even if f_1 is not uniquely determined by $I_V f_1 = f_2$. From the context it will be clear if this is the case.

3.5.1 The parameter $b_t\vartheta_1(P_V) = Z(t)$

We define the function χ_t to be

$$\chi_t(x) \equiv \frac{z(x)}{|W| + 2xR} \cdot 1_{(0,t]}(x) \tag{3.18}$$

and we note that

$$\chi_t - Z(t) = \chi_t - \int \chi_t(x)\, \mathrm{d}V(x) = \chi_t - E_V(\chi_t) \in L_0^2(V).$$

For the moment we assume that $I_V^{-1}(\chi_t - Z(t))$ exists. We find for $b_t\vartheta_1$

$$\begin{aligned}
&\frac{1}{\epsilon}\left(b_t\vartheta_1(P_{V_{\epsilon,l}}) - b_t\vartheta_1(P_V)\right) \\
&= \frac{1}{\epsilon}\left(Z_{\epsilon,l}(t) - Z(t)\right) \\
&= \frac{1}{\epsilon}\left(\int_{x=0}^{x=t} \frac{z(x)}{|W| + 2xR}(1 + \epsilon\, l(x))\, \mathrm{d}V(x) - \int_{x=0}^{x=t} \frac{z(x)}{|W| + 2xR}\, \mathrm{d}V(x)\right) \\
&= \int_{x=0}^{x=t} \frac{z(x)}{|W| + 2xR}\, l(x)\, \mathrm{d}V(x) \\
&= \int \frac{z(x)}{|W| + 2xR}\, 1_{(0,t]}(x)\, l(x)\, \mathrm{d}V(x) \\
&= \langle \chi_t, l \rangle_V \\
&= \langle \chi_t - Z(t), l \rangle_V \\
&= \left\langle A_V^T A_V I_V^{-1}(\chi_t - Z(t)), l \right\rangle_V \\
&= \left\langle A_V I_V^{-1}(\chi_t - Z(t)), A_V(l) \right\rangle_{P_V}.
\end{aligned} \tag{3.19}$$

Assuming that $I_V^{-1}(\chi_t - Z(t))$ exists, we have proved with (3.19) that $Z(t)$ is pathwise differentiable with efficient influence curve

$$\widetilde{I}(Z, t) = A_V I_V^{-1}(\chi_t - Z(t)). \tag{3.20}$$

Here we use the notation $\widetilde{I}(Z,t)$ instead of $\widetilde{I}(P_V, b_t\vartheta_1)$. By $\widetilde{I}(\widehat{Z}_n,t)$ we mean $\widetilde{I}(\mathbb{P}_n, b_t\vartheta_1)$, where $\mathbb{P}_n$ is the NPMLE of P_V and $\widehat{Z}_n$ is the NPMLE of Z, and we use $\widetilde{I}(Z_n,t)$ for $\widetilde{I}(P_n, b_t\vartheta_1)$, where Z_n is the estimator of Z induced by P_n.

Now we are ready to verify the conditions in theorem 3.3.4 in order to prove that $\widehat{Z}_n$ is a $\|\cdot\|_B$-asymptotically efficient estimator of Z. In section 3.7 we calculate $h_t \equiv I_V^{-1}(\chi_t - Z(t))$ and give an expression in lemma 3.7.2.1. One will notice that In section 3.8 we indirectly prove that $\|h_t\|_\infty < \infty$. In fact this holds for all distribution functions V, thus also for

$$\widehat{h}_{tn} \equiv I_{\widehat{V}_n}^{-1}(\chi_t - \widehat{Z}_n).$$

With $\|\widehat{h}_{tn}\|_\infty < \infty$ one easily proves with the calculated expression of A_V in section 3.6 that $\widetilde{I}(\widehat{Z}_n,t) = A_{\widehat{V}_n}\widehat{h}_{tn}$ has a finite supnorm and is a score in $S(\mathbb{P})$. Now we obtain from (1.47)

$$\int \widetilde{I}(\widehat{Z}_n,t)\,\mathrm{d}P_n = 0 \quad \text{for all } t \in [0,2R).$$

This provides us with the efficient score equation. In section 3.8 we prove the Donsker class condition. This is the most difficult condition to be verified. The result is stated in lemma 3.8.6.1. In section 3.9 we check the $\|\cdot\|_{P_V}$-convergence conditions. This is formulated in lemma 3.9.2.1 and (3.150) and (3.151). Applying theorem 3.3.4 we proved the following theorem:

Theorem 3.5.1.1 *Under the assumptions in section 3.4 the NPMLE $\widehat{Z}_n$ is a $\|\cdot\|_B$- asymptotically efficient estimator of Z.*

3.5.2 The parameter $\vartheta_2(P_V) = \mathcal{Z}$

Here we define the following function

$$\xi(x) \equiv \frac{x}{|W| + 2xR}. \tag{3.21}$$

One easily verifies that

$$\xi - \mathcal{Z} = \xi - \int \xi(x)\,\mathrm{d}V(x) = \xi - E_V(\xi) \in L_0^2(V).$$

If we assume for the moment that $I_V^{-1}(\xi - \mathcal{Z})$ exists, then we find for the parameter ϑ_2:

$$\frac{1}{\epsilon}\left(\vartheta_2(P_{V_{\epsilon,l}}) - \vartheta_2(P_V)\right) = \left\langle A_V\, I_V^{-1}(\xi - \mathcal{Z}), A_V(l)\right\rangle_{P_V}. \tag{3.22}$$

Assuming that $I_V^{-1}(\xi - \mathcal{Z})$ exists, we obtain with (3.22) that $\mathcal{Z}$ is pathwise differentiable with efficient influence curve

$$\widetilde{I}(\mathcal{Z}) = A_V\, I_V^{-1}(\xi - \mathcal{Z}). \tag{3.23}$$

Here we use the notation $\widetilde{I}(\mathcal{Z})$ instead of $\widetilde{I}(P_V, \vartheta_2)$. By $\widetilde{I}(\widehat{\mathcal{Z}}_n)$ we mean $\widetilde{I}(\mathbb{P}_n, \vartheta_2)$, where $\mathbb{P}_n$ is the NPMLE of P_V and $\widehat{\mathcal{Z}}_n$ is the NPMLE of $\mathcal{Z}$, and we use $\widetilde{I}(\mathcal{Z}_n)$ for $\widetilde{I}(P_n, \vartheta_2)$, where $\mathcal{Z}_n$ is the estimator of $\mathcal{Z}$ induced by P_n.

Just as we did for theorem 3.5.1.1, one verifies the conditions in theorem 3.3.4 to obtain

Theorem 3.5.2.1 *Under the assumptions in section 3.4 the NPMLE $\hat{\mathcal{Z}}_n$ is an asymptotically efficient estimator of $\mathcal{Z}$.*

In section 3.11 we calculate $h = I_V^{-1}(\xi - \mathcal{Z})$ and the result is written down in lemma 3.11.1. Compare this calculation with the calculation in section 3.7. To prove the Donsker class condition and the $\|\cdot\|_{P_V}$-convergence conditions one can almost copy the sections 3.8 and 3.9.

3.5.3 Efficiency of $(\widehat{Z}_n, \widehat{\mathcal{Z}}_n)$ implies efficiency of $(\widehat{F}_n, \widehat{\mu}_n)$

In section 3.5.1 and section 3.5.2 we showed that the NPMLE $(\hat{Z}_n, \hat{\mathcal{Z}}_n)$ is an efficient estimator of the underlying $(Z, \mathcal{Z}) \in D[0,2R) \times \mathbb{R}$. Here we will prove that this implies that the NPMLE $(\hat{F}_n, \hat{\mu}_n)$ is an efficient estimator of the underlying $(F, \mu) \in D[0, 2R - \epsilon] \times \mathbb{R}$ for every fixed $\epsilon \in (0, 2R]$.

Because of the relation

$$\mathrm{d}V(x) = \frac{|W| + 2xR}{|W| + 2\mu R}\,\mathrm{d}F(x)$$

we have

$$\mathcal{Z} = d(0,0) = \int_{x=0}^{x=\infty} \frac{x}{|W| + 2xR}\,\mathrm{d}V(x) = \frac{\mu}{|W| + 2\mu R}$$

and

$$\mathrm{d}Z(x) = \frac{z(x)}{|W| + 2xR}\,\mathrm{d}V(x) = \frac{z(x)}{|W| + 2\mu R}\,\mathrm{d}F(x).$$

From these we obtain

$$\mu = \frac{|W|\mathcal{Z}}{1 - 2R\mathcal{Z}} \tag{3.24}$$

and

$$\begin{aligned} F(t) &= (|W| + 2\mu R) \int_{x=0}^{x=t} \frac{1}{z(x)}\,\mathrm{d}Z(x) \\ &= \left(|W| + 2R\frac{|W|\mathcal{Z}}{1 - 2R\mathcal{Z}}\right)\left(\frac{Z(t)}{z(t)} + \int_{x=0}^{x=t} \frac{z'(x)}{z^2(x)}\,Z(x)\,\mathrm{d}x\right). \end{aligned} \tag{3.25}$$

We fix an $\epsilon \in (0, 2R]$. Now we define $\Phi_1 : (\mathbb{R}, |\cdot|) \to (\mathbb{R}, |\cdot|)$ as

$$\Phi_1(a) \equiv \frac{|W|\,a}{1 - 2R\,a}$$

and $\Phi_2 : (D[0,2R) \times \mathbb{R}, \|\cdot\|_0) \to (D[0, 2R-\epsilon], \|\cdot\|_\infty)$ as

$$\Phi_2(f,a)(t) \equiv \left(|W| + 2R\frac{|W|\,a}{1 - 2R\,a}\right)\left(\frac{f(t)}{z(t)} + \int_{x=0}^{x=t} \frac{z'(x)}{z^2(x)}\,f(x)\,\mathrm{d}x\right),$$

where $\|\cdot\|_0$ is defined as

$$\|(f,a)\|_0 \equiv \|f\|_\infty + |a|.$$

Note that we have $(\Phi_2(Z, \mathcal{Z}), \Phi_1(Z, \mathcal{Z})) = (F, \mu)$.

Suppose that we have $(s_n, v_n) \equiv \sqrt{n}((f_n, a_n) - (f, a)) \to (s, v)$, where $(f_n, a_n), (f, a) \in (D[0, 2R) \times \mathbb{R}, \|\cdot\|_0)$, then one easily checks that for

$$\begin{aligned} d\Phi_2(f,a)(s,v)(t) &\equiv \left(2R\,\frac{|W|}{(1-2R\,a)^2}\,v\right)\left(\frac{f(t)}{z(t)} + \int_{x=0}^{x=t} \frac{z'(x)}{z^2(x)}\,f(x)\,\mathrm{d}x\right) \\ &\quad + \left(|W| + 2R\,\frac{|W|\,a}{1-2R\,a}\right)\left(\frac{s(t)}{z(t)} + \int_{x=0}^{x=t} \frac{z'(x)}{z^2(x)}\,s(x)\,\mathrm{d}x\right) \end{aligned}$$

we have

$$\sqrt{n}(\Phi_2((f,a) + (1/\sqrt{n})(s_n, v_n))(t) - \Phi_2(f,a)(t)) - d\Phi_2(f,a)(s,v)(t) \to 0 \quad (n \to \infty)$$

uniformly in $t \in [0, 2R - \epsilon]$. Because

$$\int_{x=0}^{x=2R} \frac{z'(x)}{z^2(x)}\,\mathrm{d}x = \infty,$$

we must restrict ourselves to the interval $[0, 2R - \epsilon]$. (Note that $d\Phi_2(f,a)(s,v)$ is a continuous linear mapping in (s, v)). For Φ_1 one obtains a similar result.

Applying theorem 3.1.1 we have proved now that the NPMLE $\widehat{F}_n = \Phi_2(\widehat{Z}_n, \widehat{\mathcal{Z}}_n)$ is a $\|\cdot\|_\infty$-efficient estimator of $F \in D[0, 2R - \epsilon]$ and the NPMLE $\widehat{\mu}_n = \Phi_1(\widehat{\mathcal{Z}}_n)$ is a efficient estimator of μ.

3.6 The score operator A_V and the information operator I_V

For the two-dimensional circle problem the score operator $A_V : L_0^2(V) \to L_0^2(P_V)$ is given by

$$\begin{aligned} &A_V(h)(\tilde{x}, d, \theta) \\ &= E_V(h(X) \mid \widetilde{X} = \tilde{x}, D = d, \Theta = \theta) \\ &= h(\tilde{x}) \cdot 1(D = 0, \Theta = \theta) \\ &\quad + \frac{1}{g(\tilde{x})} \int_{x=\tilde{x}}^{x=\infty} h(x)\,\frac{1}{|W| + 2xR}\,\mathrm{d}V(x) \cdot 1(D = 1, \Theta = \theta) \\ &\quad + \frac{1}{d(\tilde{x}, \tilde{x})} \int_{x=\tilde{x}}^{x=\infty} h(x)\,\frac{x - \tilde{x}}{|W| + 2xR}\,\mathrm{d}V(x) \cdot 1(D = 2, \Theta = \theta). \end{aligned}$$

The adjoint $A_V^T : L_0^2(P_V) \to L_0^2(V)$ of A_V has the following expression

$$\begin{aligned} &A_V^T(\eta)(x) \\ &= E_V(\eta(\widetilde{X}, D, \Theta) \mid X = x) \\ &= \frac{z(x)}{|W| + 2xR} \int_{\theta=0}^{\theta=2\pi} \eta(x, 0, \theta)\,\mathrm{d}K(\theta) \cdot 1(x < 2R) \\ &\quad + \frac{1}{|W| + 2xR} \int_{\theta=0}^{\theta=2\pi} \int_{\tilde{x}=0}^{\tilde{x}=x\wedge 2R} a(\tilde{x})\,\eta(\tilde{x}, 1, \theta)\,\mathrm{d}\tilde{x}\,\mathrm{d}K(\theta) \\ &\quad + \frac{1}{|W| + 2xR} \int_{\theta=0}^{\theta=2\pi} \int_{\tilde{x}=0}^{\tilde{x}=x\wedge 2R} b(\tilde{x})\,\tilde{x}(x - \tilde{x})\,\eta(\tilde{x}, 2, \theta)\,\mathrm{d}\tilde{x}\,\mathrm{d}K(\theta), \end{aligned}$$

where $z(\cdot)$ is defined in (1.74) and $a(\cdot)$ and $b(\cdot)$ are defined as

$$\begin{aligned} a(x) &\equiv 4\sqrt{R^2 - \frac{1}{4}x^2} \\ b(x) &\equiv \frac{1}{2\sqrt{R^2 - \frac{1}{4}x^2}}. \end{aligned}$$

One checks that

$$z'(x) = -2\sqrt{R^2 - \frac{1}{4}x^2} = -\frac{1}{2}a(x) \tag{3.26}$$

and

$$\mathrm{d}F^{s.c.c.}(x) = a(x)\,g(x)\,\mathrm{d}x \ , \ \mathrm{d}F^{d.c.}(x) = x\,b(x)\,d(x,x)\,\mathrm{d}x$$

(see (1.69) and (1.70)). Now it is easy to write down the information operator $I_V = A_V^T A_V$: $L_0^2(V) \to L_0^2(V)$. We find

$$\begin{aligned} & I_V(h)(x) \\ = \ & \frac{z(x)}{|W| + 2xR} h(x) \cdot 1(x < 2R) \\ & + \frac{1}{|W| + 2xR} \int_{\tilde{x}=0}^{\tilde{x}=x \wedge 2R} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h(u) \frac{1}{|W| + 2uR} \mathrm{d}V(u)\,\mathrm{d}\tilde{x} \\ & + \frac{1}{|W| + 2xR} \int_{\tilde{x}=0}^{\tilde{x}=x \wedge 2R} \frac{b(\tilde{x})\,\tilde{x}(x - \tilde{x})}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h(u) \frac{u - \tilde{x}}{|W| + 2uR} \mathrm{d}V(u)\,\mathrm{d}\tilde{x}. \end{aligned} \tag{3.27}$$

3.7 Calculation of $I_V^{-1}(\chi_t - Z(t))$

We know that for a constant c we have $I_V c = c$ and thus $I_V^{-1} Z(t) = Z(t)$. So we only have to look at the equation

$$I_V(h_t)(x) = \chi_t(x), \tag{3.28}$$

where $\chi_t(\cdot)$ is defined in (3.18) and assuming for the moment that such a function h_t exists.

3.7.1 Invertibility of $I_V(h_t)(x) = \chi_t(x)$ for $x \geq 2R$

For $x \geq 2R$ equation (3.28) becomes

$$\begin{aligned} & \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u) \frac{1}{|W| + 2uR} \mathrm{d}V(u)\,\mathrm{d}\tilde{x} \\ & \quad + \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}(x - \tilde{x})}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u) \frac{u - \tilde{x}}{|W| + 2uR} \mathrm{d}V(u)\,\mathrm{d}\tilde{x} = 0. \end{aligned}$$

From this we yield

$$\begin{aligned} & \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u) \frac{1}{|W| + 2uR} \mathrm{d}V(u)\,\mathrm{d}\tilde{x} \\ & \quad - \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}^2}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u) \frac{u - \tilde{x}}{|W| + 2uR} \mathrm{d}V(u)\,\mathrm{d}\tilde{x} = 0 \end{aligned} \tag{3.29}$$

and

$$\int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u) \frac{u-\tilde{x}}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} = 0. \tag{3.30}$$

Before we use the equations (3.29) and (3.30), we define the following operators:

$$\begin{aligned} \Psi_V(h) &\equiv \int_{u=0}^{u=2R} \int_{\tilde{x}=0}^{\tilde{x}=u} \frac{a(\tilde{x})}{g(\tilde{x})}\,\mathrm{d}\tilde{x}\, h(u) \frac{1}{|W|+2uR}\,\mathrm{d}V(u) \\ &\quad - \int_{u=0}^{u=2R} \int_{\tilde{x}=0}^{\tilde{x}=u} \frac{b(\tilde{x})\,\tilde{x}^2(u-\tilde{x})}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\, h(u) \frac{1}{|W|+2uR}\,\mathrm{d}V(u) && (3.31) \\ \Lambda_V(h) &\equiv \int_{u=0}^{u=2R} \int_{\tilde{x}=0}^{\tilde{x}=u} \frac{b(\tilde{x})\,\tilde{x}(u-\tilde{x})}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\, h(u) \frac{1}{|W|+2uR}\,\mathrm{d}V(u) && (3.32) \end{aligned}$$

and

$$\begin{aligned} \alpha_V(h) &\equiv \int_{u=2R}^{u=\infty} h(u) \frac{1}{|W|+2uR}\,\mathrm{d}V(u) && (3.33) \\ \beta_V(h) &\equiv \int_{u=2R}^{u=\infty} h(u) \frac{u}{|W|+2uR}\,\mathrm{d}V(u) && (3.34) \end{aligned}$$

and we define the functions $(i = 1, 2, 3)$

$$\begin{aligned} c_{V,i}(x) &\equiv \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}^i}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x} \\ c_{V,4}(x) &\equiv \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{a(\tilde{x})}{g(\tilde{x})}\,\mathrm{d}\tilde{x}. \end{aligned}$$

The functions $c_{V,i}$ depend on V through the functions g and d. Now by changing the order of integration we work out the left-hand side of equation (3.29) in terms of the operators and functions above and find the next derivation

$$\begin{aligned} &\int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u) \frac{1}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} \\ &\quad - \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}^2}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u) \frac{u-\tilde{x}}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} \\ &= \int_{u=0}^{u=\infty} \int_{\tilde{x}=0}^{\tilde{x}=u\wedge 2R} \frac{a(\tilde{x})}{g(\tilde{x})}\,\mathrm{d}\tilde{x}\, h_t(u) \frac{1}{|W|+2uR}\,\mathrm{d}V(u) \\ &\quad - \int_{u=0}^{u=\infty} \int_{\tilde{x}=0}^{\tilde{x}=u\wedge 2R} \frac{b(\tilde{x})\,\tilde{x}^2(u-\tilde{x})}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\, h_t(u) \frac{1}{|W|+2uR}\,\mathrm{d}V(u) \\ &= \Psi_V(h_t) + \int_{u=2R}^{u=\infty} \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{a(\tilde{x})}{g(\tilde{x})}\,\mathrm{d}\tilde{x}\, h_t(u) \frac{1}{|W|+2uR}\,\mathrm{d}V(u) \\ &\quad - \int_{u=2R}^{u=\infty} \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}^2}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\, h_t(u) \frac{u}{|W|+2uR}\,\mathrm{d}V(u) \\ &\quad + \int_{u=2R}^{u=\infty} \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}^3}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\, h_t(u) \frac{1}{|W|+2uR}\,\mathrm{d}V(u) \\ &= \Psi_V(h_t) + (c_{V,4}(0) + c_{V,3}(0))\,\alpha_V(h_t) - c_{V,2}(0)\,\beta_V(h_t). \end{aligned}$$

We do the same for the left-hand side of equation (3.30) and we find

$$\begin{aligned}
&\int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u)\,\frac{u-\tilde{x}}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} \\
=\; & \int_{u=0}^{u=\infty} \int_{\tilde{x}=0}^{\tilde{x}=u\wedge 2R} \frac{b(\tilde{x})\,\tilde{x}(u-\tilde{x})}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\, h_t(u)\,\frac{1}{|W|+2uR}\,\mathrm{d}V(u) \\
=\; & \Lambda_V(h_t) + \int_{u=2R}^{u=\infty} \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\, h_t(u)\,\frac{u}{|W|+2uR}\,\mathrm{d}V(u) \\
& - \int_{u=2R}^{u=\infty} \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}^2}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\, h_t(u)\,\frac{1}{|W|+2uR}\,\mathrm{d}V(u) \\
=\; & \Lambda_V(h_t) - c_{V,2}(0)\,\alpha_V(h_t) + c_{V,1}(0)\,\beta_V(h_t).
\end{aligned}$$

With these calculations the equations (3.29) and (3.30) can be written as

$$\begin{pmatrix} c_{V,4}(0)+c_{V,3}(0) & -c_{V,2}(0) \\ -c_{V,2}(0) & c_{V,1}(0) \end{pmatrix} \cdot \begin{pmatrix} \alpha_V(h_t) \\ \beta_V(h_t) \end{pmatrix} = -\begin{pmatrix} \Psi_V(h_t) \\ \Lambda_V(h_t) \end{pmatrix}. \tag{3.35}$$

As one notes in (3.33) and (3.34) the operators α_V and β_V only use the values of $h(x)$ with $x \in [2R,\infty)$. Here we want to express $\alpha_V(h_t)$ and $\beta_V(h_t)$ in integrals which only use values of $h_t(x)$ with $x \in [0,2R)$. We know that Ψ_V and Λ_V are operators which only use values of $h(x)$ with $x \in [0,2R)$. So we have to solve the system of equations in (3.35).

Firstly, we note that $1/g(x)$ and $1/d(x,x)$ are bounded on $[0,2R)$ and thus we conclude that

$$c_{V,i}(0) > 0 \quad (i = 1,2,3,4).$$

This yields

$$c_{V,4}(0)\,c_{V,1}(0) > 0. \tag{3.36}$$

Secondly, one derives by the Cauchy-Schwartz inequality

$$\begin{aligned}
c_{V,2}(0)^2 &= \left(\int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}^2}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\right)^2 \\
&= \left(\int_{\tilde{x}=0}^{\tilde{x}=2R} \left(\frac{b(\tilde{x})}{d(\tilde{x},\tilde{x})}\right)^{\frac{1}{2}} \tilde{x}^{\frac{3}{2}} \cdot \left(\frac{b(\tilde{x})}{d(\tilde{x},\tilde{x})}\right)^{\frac{1}{2}} \tilde{x}^{\frac{1}{2}}\,\mathrm{d}\tilde{x}\right)^2 \\
&\leq \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})}{d(\tilde{x},\tilde{x})}\,\tilde{x}^3\,\mathrm{d}\tilde{x} \cdot \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})}{d(\tilde{x},\tilde{x})}\,\tilde{x}\,\mathrm{d}\tilde{x} \\
&= c_{V,3}(0)\,c_{V,1}(0).
\end{aligned}$$

This yields

$$c_{V,3}(0)\,c_{V,1}(0) - c_{V,2}(0)^2 \geq 0. \tag{3.37}$$

From (3.36) and (3.37) one concludes that the determinant D_V of the matrix

$$L_V \equiv \begin{pmatrix} c_{V,4}(0)+c_{V,3}(0) & -c_{V,2}(0) \\ -c_{V,2}(0) & c_{V,1}(0) \end{pmatrix}$$

is not equal to 0. Actually, one derives

$$\begin{aligned} D_V &= (c_{V,4}(0) + c_{V,3}(0))\, c_{V,1}(0) - c_{V,2}(0)^2 \\ &= c_{V,4}(0)\, c_{V,1}(0) + (c_{V,3}(0)\, c_{V,1}(0) - c_{V,2}(0)^2) > 0. \end{aligned}$$

Now we have showed that the inverse of matrix L_V exists and with (3.35) we can express $\alpha_V(h_t)$ and $\beta_V(h_t)$ as operators whose value only depends on h_t restricted to $[0, 2R)$. In other words we write them in terms of $\Psi_V(h_t)$ and $\Lambda_V(h_t)$. We get

$$\begin{pmatrix} \alpha_V(h_t) \\ \beta_V(h_t) \end{pmatrix} = -\frac{1}{D_V} \begin{pmatrix} c_{V,1}(0) & c_{V,2}(0) \\ c_{V,2}(0) & c_{V,4}(0) + c_{V,3}(0) \end{pmatrix} \cdot \begin{pmatrix} \Psi_V(h_t) \\ \Lambda_V(h_t) \end{pmatrix}. \tag{3.38}$$

From now on we may therefore regard α_V and β_V as linear operators from $D[0, 2R)$ to $\mathbb{R}$ defined by (3.38).

3.7.2 Invertibility of $I_V(h_t)(x) = \chi_t(x)$ for $x \in [0, 2R)$

One can write (3.29) and (3.30) as

$$\begin{aligned} &\int_{\tilde{x}=0}^{\tilde{x}=x} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u)\, \frac{1}{|W| + 2uR}\, \mathrm{d}V(u)\, \mathrm{d}\tilde{x} \\ &\quad - \int_{\tilde{x}=0}^{\tilde{x}=x} \frac{b(\tilde{x})\, \tilde{x}^2}{d(\tilde{x}, \tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u)\, \frac{u - \tilde{x}}{|W| + 2uR}\, \mathrm{d}V(u)\, \mathrm{d}\tilde{x} \\ &= -\int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u)\, \frac{1}{|W| + 2uR}\, \mathrm{d}V(u)\, \mathrm{d}\tilde{x} \\ &\quad + \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\, \tilde{x}^2}{d(\tilde{x}, \tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u)\, \frac{u - \tilde{x}}{|W| + 2uR}\, \mathrm{d}V(u)\, \mathrm{d}\tilde{x} \end{aligned} \tag{3.39}$$

and

$$\begin{aligned} &\int_{\tilde{x}=0}^{\tilde{x}=x} \frac{b(\tilde{x})\, \tilde{x}}{d(\tilde{x}, \tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u)\, \frac{u - \tilde{x}}{|W| + 2uR}\, \mathrm{d}V(u)\, \mathrm{d}\tilde{x} \\ &= -\int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\, \tilde{x}}{d(\tilde{x}, \tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u)\, \frac{u - \tilde{x}}{|W| + 2uR}\, \mathrm{d}V(u)\, \mathrm{d}\tilde{x}. \end{aligned} \tag{3.40}$$

We use (3.39) and (3.40) to express $I_V(h_t)(x)$ for $x \in [0, 2R)$ in terms of $\alpha_V(h_t)$, $\beta_V(h_t)$ and a Volterra integral operator (see Griffel(1981) p.p. 136) which acts on $D[0, 2R)$. For $x \in [0, 2R)$ we write

$$\begin{aligned} I_V(h_t)(x) &= \frac{z(x)}{|W| + 2xR}\, h_t(x) \\ &\quad + \frac{1}{|W| + 2xR} \int_{\tilde{x}=0}^{\tilde{x}=x\wedge 2R} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u)\, \frac{1}{|W| + 2uR}\, \mathrm{d}V(u)\, \mathrm{d}\tilde{x} \\ &\quad - \frac{1}{|W| + 2xR} \int_{\tilde{x}=0}^{\tilde{x}=x\wedge 2R} \frac{b(\tilde{x})\, \tilde{x}^2}{d(\tilde{x}, \tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u)\, \frac{u - \tilde{x}}{|W| + 2uR}\, \mathrm{d}V(u)\, \mathrm{d}\tilde{x} \end{aligned}$$

$$
\begin{aligned}
&+\frac{x}{|W|+2xR}\int_{\tilde{x}=0}^{\tilde{x}=x\wedge 2R}\frac{b(\tilde{x})\,\tilde{x}}{d(\tilde{x},\tilde{x})}\int_{u=\tilde{x}}^{u=\infty}h_t(u)\,\frac{u-\tilde{x}}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}\\
=\;& \frac{z(x)}{|W|+2xR}\,h_t(x)\\
&-\frac{1}{|W|+2xR}\int_{\tilde{x}=x}^{\tilde{x}=2R}\frac{a(\tilde{x})}{g(\tilde{x})}\int_{u=\tilde{x}}^{u=\infty}h_t(u)\,\frac{1}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}\\
&+\frac{1}{|W|+2xR}\int_{\tilde{x}=x}^{\tilde{x}=2R}\frac{b(\tilde{x})\,\tilde{x}^2}{d(\tilde{x},\tilde{x})}\int_{u=\tilde{x}}^{u=\infty}h_t(u)\,\frac{u-\tilde{x}}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}\\
&-\frac{x}{|W|+2xR}\int_{\tilde{x}=x}^{\tilde{x}=2R}\frac{b(\tilde{x})\,\tilde{x}}{d(\tilde{x},\tilde{x})}\int_{u=\tilde{x}}^{u=\infty}h_t(u)\,\frac{u-\tilde{x}}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}\\
=\;& \frac{z(x)}{|W|+2xR}\,(h_t(x)-B_V h_t(x))\\
&-\frac{z(x)}{|W|+2xR}\left(\frac{1}{z(x)}\,(c_{V,4}(x)+c_{V,3}(x)-x\,c_{V,2}(x))\;\alpha_V(h_t)\right)\\
&-\frac{z(x)}{|W|+2xR}\left(\frac{1}{z(x)}\,(x\,c_{V,1}(x)-c_{V,2}(x))\;\beta_V(h_t)\right)\\
=\;& \frac{z(x)}{|W|+2xR}\,(h_t(x)-B_V h_t(x)-r_V(x)\,\alpha_V(h_t)-s_V(x)\,\beta_V(h_t)), \qquad (3.41)
\end{aligned}
$$

where $B_V : (D[0,2R), \|\cdot\|_\infty) \to (D[0,2R), \|\cdot\|_\infty)$ is defined as

$$B_V h \equiv B_{V,1}h + B_{V,2}h. \qquad (3.42)$$

The operators $B_{V,i} : (D[0,2R), \|\cdot\|_\infty) \to (D[0,2R), \|\cdot\|_\infty)$ $(i = 1, 2)$ are given by

$$
\begin{aligned}
B_{V,1}h(x) &\equiv \frac{1}{z(x)}\int_{\tilde{x}=x}^{\tilde{x}=2R}\frac{a(\tilde{x})}{g(\tilde{x})}\int_{u=\tilde{x}}^{u=2R}h(u)\,\frac{1}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} \qquad (3.43)\\
B_{V,2}h(x) &\equiv \frac{1}{z(x)}\int_{\tilde{x}=x}^{\tilde{x}=2R}\frac{b(\tilde{x})\,\tilde{x}(x-\tilde{x})}{d(\tilde{x},\tilde{x})}\int_{u=\tilde{x}}^{u=2R}h(u)\,\frac{u-\tilde{x}}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} \qquad (3.44)
\end{aligned}
$$

and $r_V(\cdot)$ and $s_V(\cdot)$ are defined on $[0, 2R)$ as

$$
\begin{aligned}
r_V &\equiv r_{V,1} + r_{V,2} \qquad (3.45)\\
s_V(x) &\equiv \frac{1}{z(x)}\,(x\,c_{V,1}(x) - c_{V,2}(x)), \qquad (3.46)
\end{aligned}
$$

where $r_{V,i}$ $(i = 1, 2)$ on $[0, 2R)$ are defined as

$$
\begin{aligned}
r_{V,1}(x) &\equiv \frac{1}{z(x)}\,c_{V,4}(x) \qquad (3.47)\\
r_{V,2}(x) &\equiv \frac{1}{z(x)}\,(c_{V,3}(x) - x\,c_{V,2}(x)). \qquad (3.48)
\end{aligned}
$$

• **The inverse of $(I - B_V)$**

One easily shows that

$$z(x) = \mathcal{O}((R^2 - \frac{1}{4}x^2)^{\frac{3}{2}}) \tag{3.49}$$

and it is obvious that

$$\begin{aligned} a(x) &= \mathcal{O}((R^2 - \frac{1}{4}x^2)^{\frac{1}{2}}) && (3.50) \\ b(x) &= \mathcal{O}((R^2 - \frac{1}{4}x^2)^{-\frac{1}{2}}) && (3.51) \\ (\tilde{x} - x) &= \mathcal{O}((R^2 - \frac{1}{4}x^2)) \quad (0 \le x \le \tilde{x} \le 2R). && (3.52) \end{aligned}$$

We assume that $1/g(x)$ and $1/d(x,x)$ are bounded on $[0, 2R]$. Because $g(x)$ and $d(x,x)$ are decreasing functions, this is the same as saying that we assume that $1/g(2R)$ and $1/d(2R, 2R)$ are bounded. Here we define

$$M_V \equiv \max\left(\frac{1}{g(2R)}, \frac{1}{d(2R, 2R)}\right). \tag{3.53}$$

One easily sees that

$$\begin{aligned} c_{V,i}(x) &= \mathcal{O}\left(\frac{(R^2 - \frac{1}{4}x^2)^{\frac{1}{2}}}{d(2R, 2R)}\right) = \mathcal{O}(M_V) \quad (i = 1, 2, 3) && (3.54) \\ c_{V,4}(x) &= \mathcal{O}\left(\frac{(R^2 - \frac{1}{4}x^2)^{\frac{3}{2}}}{g(2R)}\right) = \mathcal{O}(M_V). && (3.55) \end{aligned}$$

So we may conclude that

$$\frac{1}{z(x)} c_{V,4}(x) = \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{a(\tilde{x})}{g(\tilde{x})} \, d\tilde{x} = \mathcal{O}\left(\frac{1}{g(2R)}\right) = \mathcal{O}(M_V) \tag{3.56}$$

and

$$\begin{aligned} \frac{1}{z(x)} c_{V,i}(x) - x\, c_{V,i-1}(x) &= \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\, \tilde{x}^{i-1}(\tilde{x} - x)}{d(\tilde{x}, \tilde{x})} \, d\tilde{x} \\ &= \mathcal{O}\left(\frac{1}{d(2R, 2R)}\right) = \mathcal{O}(M_V) \quad (i = 2, 3). && (3.57) \end{aligned}$$

This immediately implies that

$$r_{V,1}(x) = \mathcal{O}(M_V) \quad , \quad r_{V,2}(x) = \mathcal{O}(M_V) \tag{3.58}$$

and

$$r_V(x) = \mathcal{O}(M_V) \quad , \quad s_V(x) = \mathcal{O}(M_V). \tag{3.59}$$

Now we will write B_V as a Volterra integral operator (see Griffel(1981) p.p. 136). By changing the order of integration we get

$$\begin{aligned} B_{V,1}h(x) &= \int_{u=x}^{u=2R} \left(\frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=u} \frac{a(\tilde{x})}{g(\tilde{x})} \mathrm{d}\tilde{x} \right) h(u) \frac{1}{|W| + 2uR} \mathrm{d}V(u) \\ &= \int_{u=x}^{2R} K_{V,1}(x,u)\, h(u)\, \mathrm{d}V(u) \end{aligned}$$

and

$$\begin{aligned} B_{V,2}h(x) &= \int_{u=x}^{u=2R} \left(\frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=u} \frac{b(\tilde{x})\, \tilde{x}(x-\tilde{x})}{d(\tilde{x},\tilde{x})} (u-\tilde{x})\, \mathrm{d}\tilde{x} \right) h(u) \frac{1}{|W| + 2uR} \mathrm{d}V(u) \\ &= \int_{u=x}^{u=2R} K_{V,2}(x,u)\, h(u)\, \mathrm{d}V(u), \end{aligned}$$

where

$$\begin{aligned} K_{V,1}(x,u) &\equiv \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=u} \frac{a(\tilde{x})}{g(\tilde{x})} \mathrm{d}\tilde{x} \quad (0 \le x \le u \le 2R) \\ K_{V,2}(x,u) &\equiv \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=u} \frac{b(\tilde{x})\, \tilde{x}(x-\tilde{x})}{d(\tilde{x},\tilde{x})} (u-\tilde{x})\, \mathrm{d}\tilde{x} \quad (0 \le x \le u \le 2R). \end{aligned}$$

(Again we note the dependence of $K_{V,i}$ on V through g and d.) We have written $B_{V,i}$ $(i = 1,2)$ as a Volterra integral operator. Together with (3.56) and (3.57) we are able to bound $K_{V,i}(x,u)$ $(0 \le x \le u \le 2R)$ by writing

$$|K_{V,1}(x,u)| \le \frac{1}{z(x)} c_{V,4}(x) = \mathcal{O}(M_V) \tag{3.60}$$

$$|K_{V,2}(x,u)| \le \frac{2R}{z(x)} (c_{V,2}(x) - x\, c_{V,1}(x)) = \mathcal{O}(M_V). \tag{3.61}$$

If we define

$$K_V(x,u) \equiv K_{V,1}(x,u) + K_{V,2}(x,u),$$

then we have

$$B_V h(x) = \int_{u=x}^{u=2R} K_V(x,u)\, h(u)\, \mathrm{d}V(u)$$

and thus we have written B_V as a Volterra integral operator. With (3.54), (3.55), (3.58), (3.59), (3.60) and (3.61) and the definition of $K_V(x,u)$, we conclude that there exists a $C_V \ge 0$ such that

$$|K_{V,1}(x,u)| \le C_V, \quad |K_{V,2}(x,u)| \le C_V, \quad |K_V(x,u)| \le C_V \quad (0 \le x \le u \le 2R) \tag{3.62}$$

and

$$|r_{V,1}(x)| \le C_V, \quad |r_{V,2}(x)| \le C_V, \quad |s_V(x)| \le C_V \tag{3.63}$$

and

$$|c_{V,i}(x)| \le C_V \quad (i = 1,2,3,4). \tag{3.64}$$

We note that we have

$$C_V = \mathcal{O}(M_V). \tag{3.65}$$

Because B_V is a Volterra integral operator we have that

$$\|B_V^k h\|_\infty \le \frac{(C_V \cdot V(2R))^k}{k!} \|h\|_\infty. \tag{3.66}$$

This proves the existence of the inverse of $(I - B_V)$ in the space of bounded linear operators on $(D[0,2R), \|\cdot\|_\infty)$, being

$$(I - B_V)^{-1} = \sum_{k=0}^{\infty} B_V^k$$

and

$$\|(I - B_V)^{-1} h\|_\infty \le \exp(C_V \cdot V(2R)) \|h\|_\infty. \tag{3.67}$$

For $B_{V,i}$ $(i = 1,2)$ we find the same:

$$\|B_{V,i}^k h\|_\infty \le \frac{(C_V \cdot V(2R))^k}{k!} \|h\|_\infty. \tag{3.68}$$

- **The solution $h_t \equiv h_t^0 + h_t^1$ of $I_V(h_t) = \chi_t(x)$**

Now we want to solve (3.28) for $x \in [0,2R)$. Substituting (3.41) in (3.28) we get the equation

$$(I - B_V)h_t(x) = r_V(x)\,\alpha_V(h_t) + s_V(x)\,\beta_V(h_t) + 1_{(0,t]}(x).$$

We know that r_V and s_V are in $D[0,2R)$. This gives us the following equation for h_t with $x \in [0,2R)$:

$$h_t(x) = \bar{r}_V(x)\,\alpha_V(h_t) + \bar{s}_V(x)\,\beta_V(h_t) + \bar{y}_t(x), \tag{3.69}$$

where $\bar{r}_V$ and $\bar{s}_V$ and $\bar{y}_t$ in $D[0,2R)$ are defined as

$$\bar{r}_V \equiv (I - B_V)^{-1} r_V \quad , \quad \bar{s}_V \equiv (I - B_V)^{-1} s_V$$

and

$$\bar{y}_t \equiv (I - B_V)^{-1} 1_{(0,t]}.$$

Because of (3.38) we know that we can write $\alpha_V(h_t)$ and $\beta_V(h_t)$ as linear operators which only use values of $h_t(x)$ with $x \in [0,2R]$. Therefore we are allowed to apply α_V and β_V on both sides of the equation (3.69). We obtain the next system of equations

$$\begin{pmatrix} 1 - \alpha_V(\bar{r}_V) & -\alpha_V(\bar{s}_V) \\ -\beta_V(\bar{r}_V) & 1 - \beta_V(\bar{s}_V) \end{pmatrix} \cdot \begin{pmatrix} \alpha_V(h_t) \\ \beta_V(h_t) \end{pmatrix} = \begin{pmatrix} \alpha_V(\bar{y}_t) \\ \beta_V(\bar{y}_t) \end{pmatrix}.$$

If the determinant Q_V of the matrix

$$N_V \equiv \begin{pmatrix} 1 - \alpha_V(\bar{r}_V) & -\alpha_V(\bar{s}_V) \\ -\beta_V(\bar{r}_V) & 1 - \beta_V(\bar{s}_V) \end{pmatrix} \tag{3.70}$$

is not equal to 0, then the inverse ${N_V}^{-1}$ of matrix N_V exists and we are able to write

$$\begin{pmatrix} \alpha_V(h_t) \\ \beta_V(h_t) \end{pmatrix} = \frac{1}{Q_V} \begin{pmatrix} 1-\beta_V(\bar{s}_V) & \alpha_V(\bar{s}_V) \\ \beta_V(\bar{r}_V) & 1-\alpha_V(\bar{r}_V) \end{pmatrix} \cdot \begin{pmatrix} \alpha_V(\bar{y}_t) \\ \beta_V(\bar{y}_t) \end{pmatrix} \equiv \begin{pmatrix} S_V(\bar{y}_t) \\ T_V(\bar{y}_t) \end{pmatrix}. \tag{3.71}$$

Because of the result of (3.71) we have proved that, if $Q_V \neq 0$, then h_t is uniquely determined on $[0, 2R)$ by equation (3.69):

$$h_t(x) = \bar{r}_V(x)\, S_V(\bar{y}_t) + \bar{s}_V(x)\, T_V(\bar{y}_t) + \bar{y}_t(x).$$

For $x \geq 2R$ the solution h_t (there could be more possibilities) only has to satisfy (3.38). Actually, we proved the following lemma.

Lemma 3.7.2.1 *If we assume that $1/g(x)$ and $1/d(x,x)$ are bounded on $[0, 2R]$ and $Q_V \neq 0$ and let h_t^0 and h_t^1 be defined as*

$$h_t^0(x) \equiv h_t(x) \cdot 1_{[0,2R)}(x) \quad , \quad h_t^1(x) \equiv h_t(x) \cdot 1_{[2R,\infty)}(x),$$

then

$$I_V(h_t)(x) = \chi_t(x) \quad \textit{for all } x \in [0, \infty)$$

is equivalent to

$$h_t^0(x) = (\bar{r}_V(x)\, S_V(\bar{y}_t) + \bar{s}_V(x)\, T_V(\bar{y}_t) + \bar{y}_t(x)) \cdot 1_{[0,2R)}(x)$$

and h_t^1 satisfies

$$\begin{pmatrix} \alpha_V(h_t^1) \\ \beta_V(h_t^1) \end{pmatrix} = -\frac{1}{D_V} \begin{pmatrix} c_{V,1}(0) & c_{V,2}(0) \\ c_{V,2}(0) & c_{V,4}(0) + c_{V,3}(0) \end{pmatrix} \cdot \begin{pmatrix} \Psi_V(h_t^0) \\ \Lambda_V(h_t^0) \end{pmatrix}.$$

Here we see why we need assumption II of section 3.4. In the sections 3.10.1 and 4.4 we come back to it.

Something must be said about the existence of h_t^1 in lemma 3.7.2.1. Throughout the section we assumed the existence of h_t to make the analysis above. The result is formulated in lemma 3.7.2.1. There we state that h_t^1 only has to satisfy the given system of equations. Does there exist such h_t^1 that satisfies this system of equations? The answer is yes. Looking for a h_t^1 we are actually solving the system of equations

$$\begin{aligned} \int_{u=2R}^{u=\infty} h_t^1(u)\, \mathrm{d}V_1(u) &= q_1 \\ \int_{u=2R}^{u=\infty} u\, h_t^1(u)\, \mathrm{d}V_1(u) &= q_2, \end{aligned}$$

where q_1 and q_2 are the constants at the left-hand side of the system of equations in lemma 3.7.2.1 and $\mathrm{d}V_1(x) = 1/(|W| + 2xR) \cdot \mathrm{d}V(x)$. If we choose for each $x \in [2R, \infty)$: $h_t^1(u) = k \cdot 1_{[x,\infty)}(u) + l$, where k en l are constants then we get

$$\begin{aligned} k \cdot \int_{u=x}^{u=\infty} \mathrm{d}V_1(u) + l \cdot \int_{u=2R}^{u=\infty} \mathrm{d}V_1(u) = q_1 \\ k \cdot \int_{u=x}^{u=\infty} u\, \mathrm{d}V_1(u) + l \cdot \int_{u=2R}^{u=\infty} u\, \mathrm{d}V_1(u) = q_2. \end{aligned}$$

There exist a k and l iff

$$\mathrm{Det}(x) \equiv \int_{u=x}^{u=\infty} \mathrm{d}V_1(u) \cdot \int_{u=2R}^{u=\infty} u \,\mathrm{d}V_1(u) - \int_{u=x}^{u=\infty} u \,\mathrm{d}V_1(u) \cdot \int_{u=2R}^{u=\infty} \mathrm{d}V_1(u) \neq 0.$$

Now if we suppose $\mathrm{Det}(x) = 0$ for all $x \in [2R, \infty)$, then we obtain by differentiating both sides w.r.t. x

$$-\mathrm{d}V_1(x) \cdot \int_{u=2R}^{u=\infty} u \,\mathrm{d}V_1(u) + x \,\mathrm{d}V_1(x) \cdot \int_{u=2R}^{u=\infty} \mathrm{d}V_1(u) = 0 \quad \forall x \in [2R, \infty). \tag{3.72}$$

The assumption that $g(2R) > 0$ means that $\int_{u=2R}^{u=\infty} \mathrm{d}V_1(u) > 0$ and $\int_{u=2R}^{u=\infty} u \,\mathrm{d}V_1(u) > 0$. Let us define $c \equiv (\int_{u=2R}^{u=\infty} u \,\mathrm{d}V_1(u) / \int_{u=2R}^{u=\infty} \mathrm{d}V_1(u))$. We write (3.72) as

$$(-c + x) \,\mathrm{d}V_1(x) = 0 \quad \forall x \in [2R, \infty).$$

If we have that V gives positive mass to more than one point in $[2R, \infty)$ and thus V_1 gives positive mass to more than one point in $[2R, \infty)$, then we get $-c + x = 0$ for all x, to which V_1 gives positive mass . This contradicts the assumption and thus the assumption that $\mathrm{Det}(x) = 0$ for all $x \in [2R, \infty)$ is not true. Now we have proved that there exists an x such that $\mathrm{Det}(x) \neq 0$ and thus there exist constants k and l. This shows the existence of h_t^1 in the case that V gives positive mass to more than one point in $[2R, \infty)$.

In the case that $\mathrm{d}V(x) > 0$ in just one point in $[2R, \infty)$, we will not be able to observe this feature of the underlying distribution function V. So for estimating V on $[0, 2R)$ we may assume without loss of generality that V gives positive mass to more than one point in $[2R, \infty)$. Redistributing the mass in $[2R, \infty)$ (keeping the mean constant) has no effect for V on $[0, 2R)$. Estimating the redistributed distribution function on $[0, 2R)$ is the same as estimating the original underlying distribution function on $[0, 2R)$.

3.8 The Donsker class condition

To prove that there exists a P_V-Donsker class $\mathcal{K}$ (and a $m \geq 0$) such that $\tilde{I}(Z, t) - \tilde{I}(\hat{Z}_n, t) \in \mathcal{K}$ for all $n \geq m$ and uniformly in t with probability tending to 1, it is enough to show that

$$\sup_t \|\tilde{I}(Z, t)(\cdot, d) - \tilde{I}(\hat{Z}_n, t)(\cdot, d)\|_v \leq L_V, \tag{3.73}$$

(for $d = 0, 1, 2$) with probability tending to 1, where L_V is some constant (not depending on n or t). To find out what we need to obtain this result, we concentrate on $\tilde{I}(Z, t)$ where Z varies through V.

If we have that the score operator A_V is a bounded operator w.r.t. the $\|\cdot\|_v$-norm, such that

$$\|A_V(h)(\cdot, d)\|_v \leq H_V \|h\|_v \quad h \in D[0, 2R) \;, \quad \|A_V(h_t^1)(\cdot, d)\|_v \leq U_V, \tag{3.74}$$

where H_V and U_V are constants, then we can write, knowing that $Z(t) \leq m_2 V(2R)$ (with $m_2 \equiv z(0)/|W|$)

$$\begin{aligned} \|\tilde{I}(Z, t)(\cdot, d)\|_v &= \|A_V I_V^{-1}(\chi_t - Z(t))(\cdot, d)\|_v \\ &= \|A_V(h_t^0)(\cdot, d) + A_V(h_t^1)(\cdot, d) - Z(t)\|_v \\ &\leq H_V \|h_t^0\|_v + U_V + m_2 V(2R). \end{aligned} \tag{3.75}$$

We prove (3.74) in section 3.8.4.

Now we want to bound $\|h_t^0\|_v$. By lemma 3.7.2.1 we know that this can be done if we can bound $\|\bar{r}_V\|_v$, $\|\bar{s}_V\|_v$, $\|\bar{y}_t\|_v$, $|S_V(\bar{y}_t)|$ and $|T_V(\bar{y}_t)|$. In section 3.8.3 we show that S_V and T_V are bounded operators. For the others we need to show that $\|(I-B_V)^{-1}f\|_v$ is bounded for $f = r_{V,i}$, $f = s_V$ and $f = y_t$. This is done in section 3.8.1 and section 3.8.2. Together with (3.75) we find

$$\|\tilde{I}(Z,t)(\cdot,d)\|_v \le H_V^2\, C_V\,(3\,H_V+1) + U_V + m_2\,V(2R), \tag{3.76}$$

where C_V is also some constant. (One sees that the right-hand side of the inequality does not depend on t).

The inequality (3.76) holds, under some conditions, for all distribution functions V. So if we replace in (3.76) Z by $\widehat{Z}_n$ and V by $\widehat{V}_n$, we get the bound for $\|\tilde{I}(\widehat{Z}_n,t)(\cdot,d)\|_v$. We will see that C_V depends on $1/g(2R)$ and $1/d(2R,2R)$ and H_V depends on C_V and $1/|Q_V|$ and $1/D_V$ (where Q_V and D_V are the determinants defined in section 3.7.1 and section 3.7.2) and U_V depends on C_V and H_V. In section 3.8.5 we show the continuity in V of D_V and Q_V (thus we have

$$D_{\widehat{V}_n} \to D_V \quad , \quad Q_{\widehat{V}_n} \to Q_V$$

a.s.). This together with the fact that $1/\widehat{g}_n(2R)$ converges to $1/g(2R)$ and $1/\widehat{d}_n(2R,2R)$ converges to $1/d(2R,2R)$ a.s., implies that we can bound all the constants depending on V and $\widehat{V}_n$ by a constant not depending on n (and t). Finally in section 3.8.6 we show that this implies (3.73) (lemma 3.8.6.1).

3.8.1 $\|r_V\|_v$ and $\|s_V\|_v$ are bounded

Now we will bound $\|r_V\|_v$ and $\|s_V\|_v$. We remember the definitions of r_V, $r_{V,i}$ ($i=1,2$) and s_V in (3.45) – (3.48). Knowing that $z'(x) = -\frac{1}{2}\,a(x)$, one derives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x} r_{V,1}(x) &= -\frac{z'(x)}{z^2(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{a(\tilde{x})}{g(\tilde{x})}\,\mathrm{d}\tilde{x} - \frac{1}{z(x)}\frac{a(x)}{g(x)} \\ &= \frac{1}{2z^2(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \left(\frac{a(x)\,a(\tilde{x})}{g(\tilde{x})} - \frac{a(x)\,a(\tilde{x})}{g(x)}\right) \mathrm{d}\tilde{x}.\end{aligned}$$

Because g is decreasing, we find that

$$\frac{\mathrm{d}}{\mathrm{d}x} r_{V,1}(x) \ge 0 \tag{3.77}$$

and thus $r_{V,1}$ is positive and increasing on $[0,2R)$. Therefore we obtain together with (3.63)

$$\|r_{V,1}\|_v = \|r_{V,1}\|_\infty \le C_V. \tag{3.78}$$

Now we write for $i = 2,3$

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{z(x)}\left(c_{V,i}(x) - x\,c_{V,i-1}(x)\right)\right) \\ &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}^{i-1}(\tilde{x}-x)}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\right)\end{aligned}$$

$$\begin{aligned}
&= -\frac{z'(x)}{z^2(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}^{i-1}(\tilde{x}-x)}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x} \\
&\quad -\frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}^{i-1}}{d(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x} \\
&= \frac{1}{z^2(x)} \left\{-z'(x)\,c_{V,i}(x) + (x\,z'(x) - z(x))\,c_{V,i-1}(x)\right\} \\
&\leq \frac{1}{z^2(x)} \left\{-2R\,z'(x) + (x\,z'(x) - z(x))\right\}\,c_{V,i-1}(x).
\end{aligned}$$

The last inequality follows from the fact that from $c_{V,i}(x) \geq 0$ and

$$2R\,\frac{b(\tilde{x})\,\tilde{x}^{i-1}}{d(\tilde{x},\tilde{x})} \cdot \frac{\tilde{x}}{2R} \leq 2R\,\frac{b(\tilde{x})\,\tilde{x}^{i-1}}{d(\tilde{x},\tilde{x})} \cdot 1 \tag{3.79}$$

we easily get $(i = 2,3)$

$$0 \leq c_{V,i}(x) \leq 2R\,c_{V,i-1}(x).$$

One verifies that

$$0 \leq -\frac{z(x)}{2R\,z'(x)} + \frac{x}{2R} \quad (x \in [0,2R))$$

and so we get

$$0 \leq -2R\,z'(x) \leq z(x) - x\,z'(x)$$

and therefore we have

$$-2R\,z'(x) + (x\,z'(x) - z(x)) \leq 0.$$

This yields $(i = 2,3)$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{z(x)}\left(c_{V,i}(x) - x\,c_{V,i-1}(x)\right)\right) \leq 0. \tag{3.80}$$

Now we have proved for $r_{V,2}$ $(i = 3)$ that

$$\frac{\mathrm{d}}{\mathrm{d}x}\,r_{V,2}(x) \leq 0 \tag{3.81}$$

and for s_V $(i = 2)$ that

$$\frac{\mathrm{d}}{\mathrm{d}x}\,s_V(x) \geq 0. \tag{3.82}$$

So we have that $r_{V,2}$ and s_V are monotone functions and $r_{V,2}(x) \geq 0$ and $s_V(x) \leq 0$. This gives us together with (3.63) the following statement

$$\|r_{V,2}\|_v = \|r_{V,2}\|_\infty \leq C_V \tag{3.83}$$

$$\|s_V\|_v = \|s_V\|_\infty \leq C_V \tag{3.84}$$

and thus

$$\|r_V\|_v \leq \|r_{V,1}\|_\infty + \|r_{V,2}\|_\infty \leq 2\,C_V. \tag{3.85}$$

3.8.2 The boundedness of $\|(I - B_V)^{-1} f\|_v$

Before we get some results for $\|B_{V,i} f\|_v$ $(i = 1, 2)$, we introduce some short notation. We define for a function $f \in D[0, 2R)$ the operators

$$\begin{aligned} J_{V,1} f(x) &\equiv \int_{u=x}^{u=2R} f(u) \frac{1}{|W| + 2uR} \, \mathrm{d}V(u) \\ J_{V,2} f(x) &\equiv \int_{u=x}^{u=2R} f(u) \frac{u - x}{|W| + 2uR} \, \mathrm{d}V(u). \end{aligned}$$

It is easy to see that for $f \geq 0$ we have

$$0 \leq J_{V,i} f(x) \leq m_0 \, V(2R) \, \|f\|_\infty \quad (i = 1, 2), \tag{3.86}$$

where $m_0 = \max(1/|W|, 2R/|W|)$ (not depending on V) and both $J_{V,1} f$ and $J_{V,2} f$ are decreasing. Furthermore we keep in mind that we have

$$\|f\|_v \leq 2 \, \|f\|_\infty \quad , \quad \|f\|_\infty \leq \|f\|_v \quad \text{if } f \text{ is monotone} \tag{3.87}$$

and

$$\|f\|_v = \|f\|_\infty \quad \text{if } f \text{ is monotone and } f \geq 0 \text{ or } f \leq 0. \tag{3.88}$$

- **$\|B_{V,1} f\|_v$ is bounded**

Firstly, we look at the derivative of $B_{V,1} f(x)$ w.r.t. x. We obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x} B_{V,1} f(x) &= \frac{\mathrm{d}}{\mathrm{d}x} \left(\frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{a(\tilde{x})}{g(\tilde{x})} J_{V,1} f(\tilde{x}) \, \mathrm{d}\tilde{x} \right) \\ &= -\frac{z'(x)}{z^2(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{a(\tilde{x})}{g(\tilde{x})} J_{V,1} f(\tilde{x}) \, \mathrm{d}\tilde{x} - \frac{1}{z(x)} \frac{a(x)}{g(x)} J_{V,1} f(x) \\ &= \frac{1}{2z^2(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \left(\frac{a(x) \, a(\tilde{x})}{g(\tilde{x})} J_{V,1} f(\tilde{x}) - \frac{a(x) \, a(\tilde{x})}{g(x)} J_{V,1} f(x) \right) \mathrm{d}\tilde{x}. \end{aligned}$$

Now we can write $B_{V,1}$ as

$$B_{V,1} = B_{V,1}^{\mathrm{up}} + B_{V,1}^{\mathrm{down}},$$

where $B_{V,1}^{\mathrm{up}} f$ is increasing and $B_{V,1}^{\mathrm{down}} f$ is decreasing. Remember that $g(\cdot)$ is decreasing. Now we have for $f \geq 0$ (so we have that $J_{V,1} f \geq 0$ is decreasing and (3.86) holds)

$$\begin{aligned} 0 \leq \frac{\mathrm{d}}{\mathrm{d}x} B_{V,1}^{\mathrm{up}} f(x) &= \frac{1}{2z^2(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \left(\frac{a(x) \, a(\tilde{x})}{g(\tilde{x})} J_{V,1} f(\tilde{x}) - \frac{a(x) \, a(\tilde{x})}{g(x)} J_{V,1} f(x) \right) \mathrm{d}\tilde{x} \\ &\leq J_{V,1} f(x) \frac{1}{2z^2(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \left(\frac{a(x) \, a(\tilde{x})}{g(\tilde{x})} - \frac{a(x) \, a(\tilde{x})}{g(x)} \right) \mathrm{d}\tilde{x} \\ &= J_{V,1} f(x) \frac{\mathrm{d}}{\mathrm{d}x} r_{V,1}(x) \\ &\leq m_0 \, V(2R) \, \|f\|_\infty \cdot \frac{\mathrm{d}}{\mathrm{d}x} r_{V,1}(x). \end{aligned}$$

Together with (3.78) this implies that

$$\begin{aligned} \|B_{V,1}^{\mathrm{up}} f\|_v &\le m_0\, V(2R)\, \|f\|_\infty\, \|r_{V,1}\|_v \\ &\le m_0\, V(2R)\, C_V\, \|f\|_\infty . \end{aligned}$$

Because $B_{V,1}^{\mathrm{down}}$ is monotone and (3.68) and (3.87) hold, we have

$$\begin{aligned} \|B_{V,1}^{\mathrm{down}} f\|_v &\le 2\, \|B_{V,1}^{\mathrm{down}} f\|_\infty \\ &= 2\, \|B_{V,1} f - B_{V,1}^{\mathrm{up}} f\|_\infty \\ &\le 2\left(\|B_{V,1} f\|_\infty + \|B_{V,1}^{\mathrm{up}} f\|_\infty \right) \\ &\le 2\,(1+m_0)\, C_V\, V(2R)\, \|f\|_\infty . \end{aligned}$$

Now we have obtained the following result for $f \ge 0$

$$\|B_{V,1} f\|_v \le \|B_{V,1}^{\mathrm{up}} f\|_v + \|B_{V,1}^{\mathrm{down}} f\|_v \le (2+3\,m_0)\, C_V\, V(2R)\, \|f\|_\infty . \tag{3.89}$$

- **$\|B_{V,2} f\|_v$ is bounded**

Let us consider the derivative of $B_{V,2} f(x)$ w.r.t. x. We find

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}x} B_{V,2} f(x) \\ &= \frac{\mathrm{d}}{\mathrm{d}x} \left(\frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\, \tilde{x}(x-\tilde{x})}{d(\tilde{x},\tilde{x})}\, J_{V,2} f(\tilde{x})\, \mathrm{d}\tilde{x} \right) \\ &= -\frac{z'(x)}{z^2(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\, \tilde{x}(x-\tilde{x})}{d(\tilde{x},\tilde{x})}\, J_{V,2} f(\tilde{x})\, \mathrm{d}\tilde{x} + \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\, \tilde{x}}{d(\tilde{x},\tilde{x})}\, J_{V,2} f(\tilde{x})\, \mathrm{d}\tilde{x} \\ &= \frac{1}{z^2(x)} \left(z'(x) \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\, \tilde{x}^2}{d(\tilde{x},\tilde{x})}\, J_{V,2}(\tilde{x})\, \mathrm{d}\tilde{x} - (x\, z'(x) - z(x)) \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\, \tilde{x}}{d(\tilde{x},\tilde{x})}\, J_{V,2}(\tilde{x})\, \mathrm{d}\tilde{x} \right) . \end{aligned}$$

If $f \ge 0$ then (3.86) holds and we can actually use the same arguments, which lead to (3.80), to prove

$$\frac{\mathrm{d}}{\mathrm{d}x} B_{V,2} f(x) \ge 0 .$$

(Multiply in (3.79) both sides with $J_{V,2} f(x)$). It is easy to see that for $f \ge 0$ we have $B_{V,2} f \le 0$. So we have that $B_{V,2} f$ is monotone and $B_{V,2} f \le 0$ and together with (3.68) and (3.87) we obtain the following result for $f \ge 0$

$$\|B_{V,2} f\|_v = \|B_{V,2} f\|_\infty \le C_V\, V(2R)\, \|f\|_\infty . \tag{3.90}$$

- **$\|(I - B_V)^{-1} f\|_v$ is bounded**

By (3.43) and (3.44) we know that if $f \ge 0$ then $B_{V,1} f \ge 0$ and $B_{V,2} f \le 0$. If we define

$$\tilde{B}_{V,2} f \equiv -B_{V,2} f,$$

then for $f \geq 0$ we have by (3.90)

$$\|\tilde{B}_{V,2}f\|_v \leq C_V\, V(2R)\, \|f\|_\infty \tag{3.91}$$

and of course $\tilde{B}_{V,2}f \geq 0$. For notational reasons we define

$$\tilde{B}_{V,1}f \equiv B_{V,1}f.$$

Furthermore we define

$$m_1 \equiv (2 + 3\, m_0). \tag{3.92}$$

It is obvious that $m_1\, C_V\, V(2R) \geq C_V\, V(2R)$ and so we have for $f \geq 0$ by (3.89), (3.90) (or (3.91)) that

$$\|\tilde{B}_{V,i}f\|_v \leq m_1\, C_V\, V(2R)\, \|f\|_\infty \quad (i = 1, 2). \tag{3.93}$$

One easily sees that if $f \geq 0$ then

$$\tilde{B}_{V,i_1}\, \tilde{B}_{V,i_2} \cdots \tilde{B}_{V,i_k} f \geq 0 \quad (i_j = 1, 2 \;\;,\; j = 1, ..., k). \tag{3.94}$$

Because of (3.93) and (3.94) we get immediately

$$\|\tilde{B}_{V,i_1}\, \tilde{B}_{V,i_2} \cdots \tilde{B}_{V,i_k} f\|_v \leq m_1\, C_V\, V(2R) \cdot \|\tilde{B}_{V,i_2} \cdots \tilde{B}_{V,i_k} f\|_\infty \tag{3.95}$$

($i_j = 1, 2$, $j = 1, ..., k$). The Volterra structure of $\tilde{B}_{V,i}$ ($i = 1, 2$) gives us a similar result as (3.68):

$$\|\tilde{B}_{V,i_1}\, \tilde{B}_{V,i_2} \cdots \tilde{B}_{V,i_k} f\|_\infty \leq \frac{(C_V \cdot V(2R))^k}{k!}\, \|f\|_\infty \quad (i_j = 1, 2 \;\;,\; j = 1, ..., k). \tag{3.96}$$

Now we gathered all the tools to bound $\|(I - B_V)^{-1} f\|_v$. Using (3.88), (3.95) and (3.96) we find the next derivation for monotone f and $f \geq 0$:

$$\begin{aligned}
\|(I - B_V)^{-1} f\|_v &\leq \sum_{k=0}^{\infty} \|B_V^k f\|_v \\
&= \sum_{k=0}^{\infty} \|(B_{V,1}f + B_{V,2}f)^k\|_v \\
&= \sum_{k=0}^{\infty} \|(\tilde{B}_{V,1}f - \tilde{B}_{V,2}f)^k\|_v \\
&\leq \sum_{k=0}^{\infty} \sum_{l=0}^{k} \binom{k}{l} \|(-1)^l\, \tilde{B}_{V,i_1}\, \tilde{B}_{V,i_2} \cdots \tilde{B}_{V,i_k} f\|_v \\
&\qquad (l \text{ of the } i_j\text{'s are } 2\ (j = 1, ..., k), \text{ the others are } 1) \\
&\leq \|f\|_v + \sum_{k=1}^{\infty} \sum_{l=0}^{k} \binom{k}{l} \|\tilde{B}_{V,i_1}\, \tilde{B}_{V,i_2} \cdots \tilde{B}_{V,i_k} f\|_v \\
&\leq \|f\|_\infty + m_1\, C_V\, V(2R) \sum_{k=1}^{\infty} \sum_{l=0}^{k} \binom{k}{l} \|\tilde{B}_{V,i_2} \cdots \tilde{B}_{V,i_k} f\|_\infty
\end{aligned}$$

$$\begin{aligned}
&\leq \|f\|_\infty + m_1 C_V V(2R) \|f\|_\infty \sum_{k=1}^{\infty} \sum_{l=0}^{k} \binom{k}{l} \frac{(C_V V(2R))^{k-1}}{(k-1)!} \\
&\leq \|f\|_\infty + m_1 C_V V(2R) \|f\|_\infty \sum_{k=1}^{\infty} 2^k \frac{(C_V V(2R))^{k-1}}{(k-1)!} \\
&\leq (1 + 2 m_1 C_V V(2R) \exp(2 C_V V(2R)) \cdot \|f\|_\infty .
\end{aligned}$$

Here we define

$$H_{V,1} \equiv (1 + 2 m_1 C_V V(2R) \exp(2 C_V V(2R))$$

and

$$H_{V,2} \equiv 4 \frac{1}{|Q_V| D_V} m_0 C_V^2 V(2R) \left(1 + 12 \frac{1}{D_V} m_0 C_V^3 V(2R) \exp(C_V V(2R))\right)$$

and

$$H_{V,3} \equiv \max\left(1 , 2 m_0 V(2R) \frac{1}{g(2R)} , 2 m_0 V(2R) \frac{1}{d(2R, 2R)}\right) .$$

We define H_V to be

$$H_V \equiv \max (H_{V,1}, H_{V,2}, H_{V,3}) . \tag{3.97}$$

The use of $H_{V,1}$, $H_{V,2}$, $H_{V,3}$ and H_V becomes clear from (3.98), (3.104) and (3.105). We want to decrease the number of introduced constants.

Now we have for monotone f and $f \geq 0$:

$$\|(I - B_V)^{-1} f\|_v \leq H_{V,1} \|f\|_\infty \leq H_V \|f\|_\infty . \tag{3.98}$$

3.8.3 S_V and T_V are bounded operators

Here we will bound the operators S_V and T_V on $D[0, 2R)$ (to $\mathbb{R}$) defined by (3.71). Firstly, we bound the operators α_V and β_V on $D[0, 2R)$ (to $\mathbb{R}$) defined by (3.38). We remember the definitions of the operators Ψ_V and Λ_V on $D[0, 2R)$ in (3.31) and (3.32). With (3.64) one verifies that

$$\begin{aligned}
|\Psi_V(f)| &\leq m_0 c_{V,4}(0) V(2R) \|f\|_\infty + m_0 c_{V,2}(0) V(2R) \|f\|_\infty \\
&\leq 2 m_0 C_V V(2R) \|f\|_\infty .
\end{aligned}$$

For Λ_V we find

$$|\Lambda_V(f)| \leq m_0 c_{V,1}(0) V(2R) \|f\|_\infty \leq m_0 C_V V(2R) \|f\|_\infty .$$

The constant m_0 is defined as before: $m_0 \equiv \max(1/|W| , 2R/|W|)$. Now together with (3.38) and (3.64) we obtain

$$\begin{aligned}
|\alpha_V(f)| &\leq \frac{1}{D_V} (c_{V,1}(0) |\Psi_V(f)| + c_{V,2}(0) |\Lambda_V(f)|) \\
&\leq 3 \frac{1}{D_V} m_0 C_V^2 V(2R) \|f\|_\infty \qquad (3.99) \\
|\beta_V(f)| &\leq \frac{1}{D_V} (c_{V,2}(0) |\Psi_V(f)| + (c_{V,4}(0) + c_{V,3}(0)) |\Lambda_V(f)|) \\
&\leq 4 \frac{1}{D_V} m_0 C_V^2 V(2R) \|f\|_\infty . \qquad (3.100)
\end{aligned}$$

Secondly, we bound $\|\bar{r}_V\|_\infty$, $\|\bar{s}_V\|_\infty$ and $\|\bar{y}_t\|_\infty$. Because of (3.78), (3.83), (3.84) and (3.67) we immediately get

$$\begin{aligned} \|\bar{r}_V\|_\infty &= \|(I-B_V)^{-1}r_V\|_\infty \le 2\,C_V\,\exp(C_V\,V(2R)) & (3.101)\\ \|\bar{s}_V\|_\infty &= \|(I-B_V)^{-1}s_V\|_\infty \le C_V\,\exp(C_V\,V(2R)) & (3.102)\\ \|\bar{y}_t\|_\infty &= \|(I-B_V)^{-1}1_{(0,t]}\|_\infty \le \exp(C_V\,V(2R)). & (3.103)\end{aligned}$$

Now with (3.99)-(3.102) we can bound the operators S_V and T_V. We find

$$\begin{aligned} |S_V(f)| &\le \frac{1}{|Q_V|}\left((1+|\beta_V(\bar{s}_V)|)\cdot|\alpha_V(f)|+|\alpha_V(\bar{s}_V)|\cdot|\beta_V(f)|\right)\\ &\le 3\,\frac{1}{|Q_V|\,D_V}\,m_0\,C_V^2\,V(2R)\left(1+8\,\frac{1}{D_V}\,m_0\,C_V^3\,V(2R)\,\exp(C_V\,V(2R))\right)\|f\|_\infty\\ |T_V(f)| &\le \frac{1}{|Q_V|}\left(|\beta_V(\bar{r}_V)|\cdot|\alpha_V(f)|+(1+|\alpha_V(\bar{r}_V)|)\cdot|\beta_V(f)|\right)\\ &\le 4\,\frac{1}{|Q_V|\,D_V}\,m_0\,C_V^2\,V(2R)\left(1+12\,\frac{1}{D_V}\,m_0\,C_V^3\,V(2R)\,\exp(C_V\,V(2R))\right)\|f\|_\infty.\end{aligned}$$

One remembers the definition of H_V in (3.97). We have for found for f:

$$|S_V(f)|\le H_{V,2}\,\|f\|_\infty\le H_V\,\|f\|_\infty\,,\quad |T_V(f)|\le H_{V,2}\,\|f\|_\infty\le H_V\,\|f\|_\infty. \qquad (3.104)$$

3.8.4 The score operator A_V is a bounded operator w.r.t. $\|\cdot\|_v$-norm

In section 3.6 we calculated the score operator $A_V : L_0^2(V)\to L_0^2(P_V)$. One easily sees that $A_V(h)(\tilde{x},d,\theta)$ does not depend on θ at all. Therefore from now on we leave out the θ in our notation and write $A_V(h)(\tilde{x},d)$. If we restrict ourselves to functions $h(x)\,1_{[0,2R)}(x)$ ($h\in L^2(V)$), then one regards A_V as a linear operator from $D[0,2R)$ to $D[0,2R)$. Note that $h_t^0(x)=h_t(x)\cdot 1_{[0,2R)}(x)$. Now we will prove that $A_V(\star)(\cdot,d) : (D[0,2R),\|\star\|_v)\to(D[0,2R),\|\star\|_v)$ is a bounded linear operator ($d=0,1,2$).

Now let $h\in D[0,2R)$. Firstly we calculate $\mathrm{d}_{\tilde{x}}\,A_V(h)(\tilde{x},0)$:

$$\mathrm{d}_{\tilde{x}}\,A_V(h)(\tilde{x},0)=\mathrm{d}_{\tilde{x}}\,h(\tilde{x}).$$

This implies

$$\|A_V(h)(\cdot,0)\|_v=\|h\|_v.$$

Secondly we calculate $\mathrm{d}_{\tilde{x}}\,A_V(h)(\tilde{x},1)$:

$$\begin{aligned}\mathrm{d}_{\tilde{x}}\,A_V(h)(\tilde{x},1) &= \mathrm{d}_{\tilde{x}}\,\frac{1}{g(\tilde{x})}\cdot\int_{x=\tilde{x}}^{x=2R}h(x)\,\frac{1}{|W|+2xR}\,\mathrm{d}V(x)\\ &\quad-\frac{1}{g(\tilde{x})}\cdot h(\tilde{x})\,\frac{1}{|W|+2\tilde{x}R}\,\mathrm{d}V(\tilde{x}).\end{aligned}$$

We know that $1/g(\tilde{x})$ is an increasing function, so we have $\mathrm{d}_{\tilde{x}}(1/g(\tilde{x}))\ge 0$. Now one writes

$$|\mathrm{d}_{\tilde{x}}\,A_V(h)(\tilde{x},1)|\le \mathrm{d}_{\tilde{x}}\,\frac{1}{g(\tilde{x})}\cdot m_0\,V(2R)\,\|h\|_\infty+\frac{1}{g(2R)}\,m_0\,\|h\|_\infty\cdot\mathrm{d}V(\tilde{x}).$$

This implies (using the fact that $1/g(x)$ and $V(x)$ are positive and monotone (thus (3.88) holds) and the fact that $\|h\|_\infty \le \|h\|_v$)

$$\begin{aligned}\|A_V(h)(\cdot,1)\|_v &\le m_0\,V(2R)\left\|\frac{1}{g}\right\|_v \|h\|_\infty + m_0\,\|V\|_v\,\frac{1}{g(2R)}\,\|h\|_\infty \\ &\le 2\,m_0\,V(2R)\,\frac{1}{g(2R)}\,\|h\|_v.\end{aligned}$$

Thirdly one calculates $\mathrm{d}_{\tilde{x}}\,A_V(h)(\tilde{x},2)$. Similar calculations lead to the following outcome

$$\|\mathrm{d}_{\tilde{x}}\,A_V(h)(\cdot,2)\|_v \le 2\,m_0\,V(2R)\,\frac{1}{d(2R,2R)}\,\|h\|_v.$$

We have proved that $A_V(\star)(\cdot,d)$ is a linear bounded operator for $d=0,1,2$ w.r.t. the $\|\cdot\|_v$-norm and

$$\|A_V(h)(\cdot,d)\|_v \le H_{V,3}\,\|h\|_v \le H_V\,\|h\|_v, \tag{3.105}$$

where H_V is defined as in (3.97).

Now we regard A_V as an operator on $D[0,\infty)$ again. We want to investigate $A_V(h_t^1)(\tilde{x},d)$ (again we leave out the θ). If we calculate $A_V(h_t^1)(\tilde{x},0)$ we find

$$A_V(h_t^1)(\tilde{x},0) = h_t^1(\tilde{x}) = 0,$$

because $\tilde{x}\in[0,2R)$. It is obvious that

$$\|A_V(h_t^1)(\cdot,0)\|_v = 0.$$

If we calculate $A_V(h_t^1)(\tilde{x},1)$ we get (using (3.71) in the last equality)

$$\begin{aligned}A_V(h_t^1)(\tilde{x},1) &= \frac{1}{g(\tilde{x})}\int_{x=\tilde{x}}^{x=\infty} h_t^1(x)\,\frac{1}{|W|+2xR}\,\mathrm{d}V(x) \\ &= \frac{1}{g(\tilde{x})}\int_{x=2R}^{x=\infty} h_t(x)\,\frac{1}{|W|+2xR}\,\mathrm{d}V(x) \\ &= \frac{1}{g(\tilde{x})}\,\alpha_V(h_t) \\ &= \frac{1}{g(\tilde{x})}\,S_V(\bar{y}_t).\end{aligned}$$

Because $1/g(x)$ is a monotone function and because of (3.103) and (3.104) we have

$$\|A_V(h_t^1)(\cdot,1)\|_v = \left\|\frac{1}{g}\right\|_\infty \cdot |S_V(\bar{y}_t)| \le \frac{1}{g(2R)}\,H_V\,\exp(C_V\,V(2R)).$$

If we calculate $A_V(h_t^1)(\tilde{x},2)$ we obtain (using (3.71) in the last equality)

$$\begin{aligned}A_V(h_t^1)(\tilde{x},2) &= \frac{1}{d(\tilde{x},\tilde{x})}\int_{x=\tilde{x}}^{x=\infty} h_t^1(x)\,\frac{x-\tilde{x}}{|W|+2xR}\,\mathrm{d}V(x) \\ &= \frac{1}{d(\tilde{x},\tilde{x})}\int_{x=2R}^{x=\infty} h_t(x)\,\frac{x-\tilde{x}}{|W|+2xR}\,\mathrm{d}V(x) \\ &= \frac{1}{d(\tilde{x},\tilde{x})}\,\beta_V(h_t) - \frac{\tilde{x}}{d(\tilde{x},\tilde{x})}\,\alpha_V(h_t) \\ &= \frac{1}{d(\tilde{x},\tilde{x})}\,T_V(\bar{y}_t) - \frac{\tilde{x}}{d(\tilde{x},\tilde{x})}\,S_V(\bar{y}_t).\end{aligned}$$

Because $1/(d(\tilde{x},\tilde{x})$ and $\tilde{x}/d(\tilde{x},\tilde{x})$ are monotone functions and because of (3.103) and (3.104) we have

$$\|A_V(h_t^1)(\cdot,2)\|_v \leq \frac{1+2R}{d(2R,2R)} H_V \exp(C_V\, V(2R)).$$

Now we have proved for $d = 0,1,2$ that

$$\|A_V(h_t^1)(\cdot,d)\|_v \leq U_V, \tag{3.106}$$

where U_V is defined as

$$U_V \equiv \max\left(\frac{1}{g(2R)} H_V\, \exp(C_V\, V(2R))\;,\; \frac{1+2R}{d(2R,2R)} H_V \exp(C_V\, V(2R))\right). \tag{3.107}$$

3.8.5 The determinants D_V and Q_V are continuous in V

Throughout this chapter we introduced several operators and functions depending on the distribution function V. If we replace V by another distribution function V_n and V_n converges for instance in supremum norm to V, then what can we say about the convergence of the operators and functions which depend on V? Here we mean by $\|f\|_\infty$ the supremum norm on $[0,2R)$. We know that the functions $g(x)$ and $d(x,x)$ depend on V through their definitions and are decreasing. We write g_n and d_n to mark their dependence of V_n. Throughout this section we assume that we have $\|V_n - V\|_\infty \to 0$ and also $\|g_n - g\|_\infty \to 0$ and $\|d_n(\cdot,\cdot) - d(\cdot,\cdot)\|_\infty \to 0$ hold. In particular we assume that $g(2R) > 0$ and $d(2R,2R) > 0$.

- **Continuity in V of D_V**

If we consider the functions $c_{V,i}$ $(i = 1,2,3,4)$, then we can write for example for $c_{V,1}$:

$$|c_{V_n,1}(x) - c_{V,i}(x)| \leq$$
$$\int_{\tilde{x}=0}^{\tilde{x}=2R} b(\tilde{x})\,\tilde{x}\,\mathrm{d}\tilde{x} \cdot \left\| \frac{1}{d_n(\cdot,\cdot)} - \frac{1}{d(\cdot,\cdot)} \right\|_\infty \to 0.$$

One obtains similar results for the other $c_{V,i}$'s. Now we proved that $(i = 1,2,3,4)$

$$\|c_{V_n,i} - c_{V,i}\|_\infty \to 0 \quad (n \to \infty). \tag{3.108}$$

From (3.108) we get immediately the result for the determinant D_V:

$$|D_{V_n} - D_V| \to 0 \quad (n \to \infty). \tag{3.109}$$

Remember that we have $D_V > 0$ for all V and thus also for $V = V_n$. Now (3.109) implies that there exists a $D_{V,0} > 0$ such that

$$\frac{1}{D_V} \leq \frac{1}{D_{V,0}}\,,\quad \frac{1}{D_{V_n}} \leq \frac{1}{D_{V,0}} \quad \text{for all } n. \tag{3.110}$$

With (3.56), (3.57) and the proof of (3.108) one easily gets the following result for r_V and s_V:

$$\|r_{V_n} - r_V\|_\infty \to 0 \quad (n \to \infty) \tag{3.111}$$
$$\|s_{V_n} - s_V\|_\infty \to 0 \quad (n \to \infty). \tag{3.112}$$

• Continuity in V of α_V and β_V

Now let us consider the operators Ψ_V and Λ_V defined in (3.31) and (3.32). By telescoping and using integration by parts we write for Λ_V:

$$
\begin{aligned}
&\Lambda_{V_n}(h) - \Lambda_V(h) \\
&= \int_{u=0}^{u=2R} \int_{\tilde{x}=0}^{\tilde{x}=u} \frac{b(\tilde{x})\,\tilde{x}(u-\tilde{x})}{d_n(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\,h(u)\,\frac{1}{|W|+2uR}\,\mathrm{d}(V_n - V)(u) \\
&\quad + \int_{u=0}^{u=2R} \int_{\tilde{x}=0}^{\tilde{x}=u} \left(\frac{b(\tilde{x})\,\tilde{x}(u-\tilde{x})}{d_n(\tilde{x},\tilde{x})} - \frac{b(\tilde{x})\,\tilde{x}(u-\tilde{x})}{d(\tilde{x},\tilde{x})}\right)\mathrm{d}\tilde{x}\,h(u)\,\frac{1}{|W|+2uR}\,\mathrm{d}V(u) \\
&= (V_n - V)(2R) \int_{\tilde{x}=0}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}(2R-\tilde{x})}{d_n(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\,h(2R)\,\frac{1}{|W|+4R^2} \\
&\quad - \int_{u=0}^{u=2R} (V_n - V)(u) \int_{\tilde{x}=0}^{\tilde{x}=u} \frac{b(\tilde{x})\,\tilde{x}}{d_n(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\,h(u)\frac{1}{|W|+2uR}\,\mathrm{d}u \\
&\quad + \int_{u=0}^{u=2R} (V_n - V)(u) \int_{\tilde{x}=0}^{\tilde{x}=u} \frac{b(\tilde{x})\,\tilde{x}(u-\tilde{x})}{d_n(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\,h(u)\,\frac{2R}{(|W|+2uR)^2}\,\mathrm{d}u \\
&\quad - \int_{u=0}^{u=2R} (V_n - V)(u) \int_{\tilde{x}=0}^{\tilde{x}=u} \frac{b(\tilde{x})\,\tilde{x}(u-\tilde{x})}{d_n(\tilde{x},\tilde{x})}\,\mathrm{d}\tilde{x}\,\frac{1}{|W|+2uR}\,\mathrm{d}h(u) \\
&\quad + \int_{u=0}^{u=2R} \int_{\tilde{x}=0}^{\tilde{x}=u} \left(\frac{b(\tilde{x})\,\tilde{x}(u-\tilde{x})}{d_n(\tilde{x},\tilde{x})} - \frac{b(\tilde{x})\,\tilde{x}(u-\tilde{x})}{d(\tilde{x},\tilde{x})}\right)\mathrm{d}\tilde{x}\,h(u)\,\frac{1}{|W|+2uR}\,\mathrm{d}V(u).
\end{aligned}
$$

This yields (using (3.64))

$$
\begin{aligned}
&|\Lambda_{V_n}(h) - \Lambda_V(h)| \\
&\leq \|V_n - V\|_\infty \left(\frac{d(0,0)}{d_n(2R,2R)\,(|W|+4R^2)}\,(2R\,c_{V,1}(0) - c_{V,2}(0))\,\|h\|_\infty\right. \\
&\quad + \frac{2R\,d(0,0)}{d_n(2R,2R)\,|W|}\,c_{V,1}(0)\,\|h\|_\infty \\
&\quad + \frac{(2R)^2\,d(0,0)}{d_n(2R,2R)\,|W|^2}\,(2R\,c_{V,1}(0) - c_{V,2}(0))\,\|h\|_\infty \\
&\quad \left. + \frac{d(0,0)}{d_n(2R,2R)\,|W|}\,(2R\,c_{V,1}(0) - c_{V,2}(0)) \int_{u=0}^{u=2R} |\mathrm{d}h(u)|\right) \\
&\quad + \|1/d(\tilde{x},\tilde{x}) - 1/d_n(\tilde{x},\tilde{x})\|_\infty\,\frac{V(2R)}{|W|}\,(2R\,c_{V,1}(0) - c_{V,2}(0))\,\|h\|_\infty \\
&\leq \|V_n - V\|_\infty\,\frac{d(0,0)}{d_n(2R,2R)}\,\max((2R)^2/|W|^2,\,1/|W|,\,2R/|W|)\cdot 2R\,c_{V,1}(0)\,(3\,\|h\|_\infty + \|h\|_v) \\
&\quad + \|1/d(\tilde{x},\tilde{x}) - 1/d_n(\tilde{x},\tilde{x})\|_\infty\,\frac{V(2R)}{|W|}\,2R\,c_{V,1}(0)\,\|h\|_\infty \\
&\leq \|V_n - V\|_\infty\,\frac{d(0,0)}{d_n(2R,2R)}\,\max((2R)^2/|W|^2,\,1/|W|,\,2R/|W|)\cdot 2R\,C_V\,4\,\|h\|_v \\
&\quad + \|1/d(\tilde{x},\tilde{x}) - 1/d_n(\tilde{x},\tilde{x})\|_\infty\,\frac{V(2R)}{|W|}\,2R\,c_{V,1}(0)\,\|h\|_\infty.
\end{aligned}
$$

The last term tends to 0 because $\|1/d_n - 1/d\|_\infty$ tends to 0 and for the first term we note that $\|V_n - V\|_\infty$ tends to 0. Because $d_n(2R, 2R)$ converges to $d(2R, 2R) > 0$, we know that $1/d_n(2R, 2R)$ is bounded. So if $\|h\|_v$ (and thus $\|h\|_\infty$) is bounded one has proved the following result:

$$|\Lambda_{V_n}(h) - \Lambda_V(h)| \to 0 \quad (n \to \infty) \text{ if } \|h\|_v < \infty . \tag{3.113}$$

Similar arguments are used to prove this result for Ψ_V:

$$|\Psi_{V_n}(h) - \Psi_V(h)| \to 0 \quad (n \to \infty) \text{ if } \|h\|_v < \infty. \tag{3.114}$$

By (3.38), (3.108)–(3.114) we obtain easily the next statement:

$$|\alpha_{V_n}(h) - \alpha_V(h)| \to 0 \quad (n \to \infty) \text{ if } \|h\|_v < \infty \tag{3.115}$$

$$|\beta_{V_n}(h) - \beta_V(h)| \to 0 \quad (n \to \infty) \text{ if } \|h\|_v < \infty. \tag{3.116}$$

- **Continuity in V of $\bar{r}_V$ and $\bar{s}_V$**

Now let us consider the operators $B_{V,i}$ $(i = 1, 2)$ defined in (3.43) and (3.44). Again one can use the same method as for (3.113) to prove that $(i = 1, 2)$

$$\|B_{V_n,i}h - B_{V,i}h\|_\infty \to 0 \quad (n \to \infty) \text{ if } \|h\|_v < \infty$$

and this implies for our operator B_V defined in (3.42)

$$\|B_{V_n}h - B_V h\|_\infty \to 0 \quad (n \to \infty) \text{ if } \|h\|_v < \infty . \tag{3.117}$$

We know that there exist $C_{V_n} \geq 0$ such that (3.62)–(3.64) hold, where V is replaced by V_n and

$$C_{V_n} = \mathcal{O}(M_{V_n}),$$

with

$$M_{V_n} \equiv \max\left(\frac{1}{g_n(2R)}, \frac{1}{d_n(2R, 2R)}\right)$$

(compare with (3.65) and the definition of M_V in (3.53)). By the knowledge that $g_n(2R)$ converges to $g(2R) > 0$ and $d_n(2R, 2R)$ converges to $d(2R, 2R) > 0$ and the fact that $V_n(2R) \leq 1$ and $V(2R) \leq 1$ and $V_n(2R)$ converges to $V(2R)$, we conclude that there exists a $C_{V,0} \geq 0$ and $l_1 \geq 0$ such that

$$C_V \leq C_{V,0} , \quad C_{V_n} \leq C_{V,0} \quad \text{for all } n \geq l_1 \tag{3.118}$$

and

$$C_{V_n} V_n(2R) \leq C_{V,0} V(2R) \quad \text{for all } n \geq l_1 , \tag{3.119}$$

with

$$C_{V,0} = \mathcal{O}(M_V).$$

The l_1 ensures us that $g_n(2R) \neq 0$ and $d_n(2R, 2R) \neq 0$ for all $n \geq l_1$. This and the inequality (3.67) imply that we have

$$\|(I - B_{V_n})^{-1}h\|_\infty \leq \exp(C_{V,0} V(2R)) \|h\|_\infty \quad \text{for all } n \geq l_1. \tag{3.120}$$

Now one writes

$$\begin{aligned}\|(I-B_{V_n})^{-1}h-(I-B_V)^{-1}h\|_\infty &= \|(I-B_{V_n})^{-1}(B_{V_n}-B_V)(I-B_V)^{-1}h\|_\infty\\ &\le \exp(C_{V,0}V(2R))\cdot\|(B_{V_n}-B_V)f\|_\infty,\end{aligned}$$

where $f=(I-B_V)^{-1}h$. Together with (3.117) we obtain

$$\|(I-B_{V_n})^{-1}h-(I-B_V)^{-1}h\|_\infty\to 0 \quad (n\to\infty) \text{ if } \|f\|_v=\|(I-B_V)^{-1}h\|_v<\infty. \tag{3.121}$$

For $\bar{r}_V$ we derive

$$\begin{aligned}&\|\bar{r}_{V_n}-\bar{r}_V\|_\infty\\ &= \|(I-B_{V_n})^{-1}r_{V_n}-(I-B_{V_n})^{-1}r_V+(I-B_{V_n})^{-1}r_V-(I-B_V)^{-1}r_V\|_\infty\\ &\le \exp(C_{V,0}V(2R))\,\|r_{V_n}-r_V\|_\infty+\|(I-B_{V_n})^{-1}r_V-(I-B_V)^{-1}r_V\|_\infty.\end{aligned} \tag{3.122}$$

By (3.78), (3.83) and (3.98) we know that

$$\begin{aligned}\|(I-B_V)^{-1}r_V\|_v &= \|(I-B_V)^{-1}(r_{V,1}+r_{V,2})\|_v\\ &\le H_V\,\|r_{V,1}\|_\infty+H_V\,\|r_{V,2}\|_\infty\\ &\le 2\,H_V\,C_V.\end{aligned}$$

So we have

$$\|\bar{r}_V\|_v=\|(I-B_V)^{-1}r_V\|_v\le 2\,H_V\,C_V. \tag{3.123}$$

This implies that (3.121) holds for $h=r_V$ and together with (3.111) we get from (3.122) the following result:

$$\|\bar{r}_{V_n}-\bar{r}_V\|_\infty\to 0 \quad (n\to\infty). \tag{3.124}$$

A similar proof can be produced to get

$$\|\bar{s}_V\|_v=\|(I-B_V)^{-1}s_V\|_v\le H_V\,C_V \tag{3.125}$$

and

$$\|\bar{s}_{V_n}-\bar{s}_V\|_\infty\to 0 \quad (n\to\infty). \tag{3.126}$$

• **Continuity in V of Q_V**

Now we are ready to prove the continuity in V of the determinant Q_V. With (3.99), (3.110), (3.118) and (3.119) we may write

$$|\alpha_{V_n}(h)|\le 3\,\frac{1}{D_{V_n}}\,m_0\,C_{V_n}^2\,V_n(2R)\,\|h\|_\infty\le 3\,\frac{1}{D_{V,0}}\,m_0\,C_{V,0}^2\,V(2R)\,\|h\|_\infty \quad \text{for all } n\ge l_1.$$

This yields ($n\ge l_1$)

$$\begin{aligned}|\alpha_{V_n}(h_n)-\alpha_V(h)| &= |\alpha_{V_n}(h_n)-\alpha_{V_n}(h)+\alpha_{V_n}(h)-\alpha_V(h)|\\ &\le 3\,\frac{1}{D_{V,0}}\,m_0\,C_{V,0}^2\,V(2R)\,\|h_n-h\|_\infty+|\alpha_{V_n}(h)-\alpha_V(h)|\,.\end{aligned}$$

If $\|h\|_v < \infty$, then we know that (3.115) holds and together with $\|h_n - h\|_\infty \to 0$ we get

$$|\alpha_{V_n}(h_n) - \alpha_V(h)| \to 0 \quad (n \to \infty) \text{ if } \|h_n - h\|_\infty \to 0 \text{ and } \|h\|_v < \infty. \tag{3.127}$$

For β_V one derives a similar result

$$|\beta_{V_n}(h_n) - \beta_V(h)| \to 0 \quad (n \to \infty) \text{ if } \|h_n - h\|_\infty \to 0 \text{ and } \|h\|_v < \infty. \tag{3.128}$$

Applying (3.127) and (3.128) with (3.123)–(3.126) we immediately get for the determinant

$$Q_V = (1 - \alpha_V(\bar{r}_V))(1 - \beta_V(\bar{s}_V)) - \alpha_V(\bar{s}_V)\,\beta_V(\bar{r}_V) \tag{3.129}$$

the following statement:

$$|Q_{V_n} - Q_V| \to 0 \quad (n \to \infty). \tag{3.130}$$

Remember that we assume $Q_V \neq 0$. Now by (3.130) we know that there exists a $Q_{V,0} > 0$ and $l_2 \geq 0$ such that

$$\frac{1}{|Q_V|} \leq \frac{1}{Q_{V,0}}, \quad \frac{1}{|Q_{V_n}|} \leq \frac{1}{Q_{V,0}} \quad \text{for all } n \geq l_2. \tag{3.131}$$

The meaning of the l_2 is not important. It only ensures us that $Q_{V_n} \neq 0$ for all $n \geq l_2$.

3.8.6 Verification of the Donsker class condition

With all the preparations in the previous sections we will verify the Donsker class condition for a sequence V_n described in section 3.8.5 .

Firstly, we note that with (3.110), (3.118), (3.119) and (3.131) one immediately sees that there exist a $H_{V,0} \geq 0$ and $U_{V,0} \geq 0$ such that

$$H_V \leq H_{V,0}\ , \quad H_{V_n} \leq H_{V,0} \quad \text{for all } n \geq m \tag{3.132}$$

and

$$U_V \leq U_{V,0}\ , \quad U_{V_n} \leq U_{V,0} \quad \text{for all } n \geq m\ , \tag{3.133}$$

where $m = \max(l_1, l_2)$ and

$$H_{V,0} = \mathcal{O}(H_V)\ , \quad U_{V,0} = \mathcal{O}(U_V)\,.$$

Secondly, we obtain with (3.98) (knowing that $r_{V,1}$, $r_{V,2}$ and $-s_V$ are monotone and positive), (3.104), (3.78), (3.83) and (3.84) that

$$\begin{aligned}
\|h_t^0\|_v &\leq \|\bar{r}_V\|_v\, |S_V(\bar{y}_t)| + \|\bar{s}_V\|_v\, |T_V(\bar{y}_t| + \|\bar{y}_t\|_v \\
&\leq \left(\|(I - B_V)^{-1} r_{V,1}\|_v + \|(I - B_V)^{-1} r_{V,2}\|_v\right) |S_V(\bar{y}_t)| \\
&\quad + \|(I - B_V)^{-1} s_V\|_v\, |T_V(\bar{y}_t)| + \|(I - B_V)^{-1} y_t\|_v \\
&\leq H_V^2 \left(\|r_{V,1}\|_\infty + \|r_{V,2}\|_\infty\right) \|y_t\|_\infty + H_V^2\, \|s_V\|_\infty\, \|y_t\|_\infty + H_V\, \|y_t\|_\infty \\
&\leq 2\, H_V^2\, C_V + H_V^2\, C_V + H_V \\
&= H_V\,(3\, H_V\, C_V + 1).
\end{aligned}$$

Note that this holds uniformly for $t \in [0, 2R)$ ($\|y_t\|_\infty = \|1_{(0,t]}\|_\infty \leq 1$ for all $t \in [0, 2R)$). Furthermore one writes

$$0 \leq Z(t) \leq \frac{z(0)}{|W|} V(2R) = m_2 V(2R),$$

where m_2 is defined as $m_2 \equiv z(0)/|W|$. Again this holds uniformly in t.

Thirdly, we obtain with the analysis above and with (3.105) and (3.106) that

$$\begin{aligned}
\|\tilde{I}(Z,t)(\cdot,d)\|_v &= \|A_V I_V^{-1}(\chi_t - Z(t))(\cdot,d)\|_v \\
&= \|A_V I_V^{-1}(\chi_t)(\cdot,d) - Z(t)\|_v \\
&= \|A_V (h_t^0 + h_t^1)(\cdot,d) - Z(t)\|_v \\
&\leq \|A_V (h_t^0)(\cdot,d)\|_v + \|A_V (h_t^1)(\cdot,d)\|_v + Z(t) \\
&\leq H_V \|h_t^0\|_v + U_V + Z(t) \\
&\leq H_V^2 (3 H_V C_V + 1) + U_V + m_2 V(2R).
\end{aligned}$$

This yields for all $n \geq m$

$$\begin{aligned}
\|\tilde{I}(Z,t)(\cdot,d) - \tilde{I}(Z_n,t)(\cdot,d)\|_v &\leq \|\tilde{I}(Z,t)(\cdot,d)\|_v + \|\tilde{I}(Z_n,t)(\cdot,d)\|_v \\
&\leq H_V^2 (3 H_V C_V + 1) + U_V + m_2 V(2R) \\
&\quad + H_{V_n}^2 (3 H_{V_n} C_{V_n} + 1) + U_{V_n} + m_2 V_n(2R) \\
&\leq 2 (H_{V,0}^2 (3 H_{V,0} C_{V,0} + 1) + U_{V,0} + m_2).
\end{aligned}$$

Again this holds uniformly in t. If we define

$$L_V \equiv 2 (H_{V,0}^2 (3 H_{V,0} C_{V,0} + 1) + U_{V,0} + m_2)$$

(of course not depending on n and t), then we have proved the following lemma.

Lemma 3.8.6.1 *If we assume that $1/g(x)$ and $1/d(x,x)$ are bounded on $[0, 2R]$ and $Q_V \neq 0$ and if V_n is a sequence distribution functions such that $\|V_n - V\|_\infty \to 0$, $\|g_n - g\|_\infty \to 0$ and $\|d_n(\cdot,\cdot) - d(\cdot,\cdot)\|_\infty \to 0$, then we have $Q_{V_n} \to Q_V$ and thus there exists a m such that for all $n \geq m$ we have $Q_{V_n} \neq 0$ and*

$$\sup_t \|\tilde{I}(Z,t)(\cdot,d) - \tilde{I}(Z_n,t)(\cdot,d)\|_v \leq L_V$$

for $d = 0, 1, 2$.

3.9 The $\|\cdot\|_{P_V}$-convergence conditions

In this section we are going to prove the $\|\cdot\|_{P_V}$-convergence conditions in the two-dimensional 'circle'-case. We recall (3.13) and (3.14). In section 3.9.1 we write $\tilde{I}(Z,t) - \tilde{I}(Z_n,t)$ as a sum of four terms and for each term we prove that we can bound it by a constant times $K(n)$ (where $K(n)$ is defined by (3.135); compare $K(n)$ with $P_k - P$ in (3.13) or (3.14)). In section 3.9.2 we use the results in section 3.9.1 to conclude that the $\|\cdot\|_{P_V}$-convergence condition holds.

3.9.1 Calculating $\tilde{I}(Z,t) - \tilde{I}(Z_n,t)$

Let us write down the following derivation

$$\begin{aligned}
&\tilde{I}(Z,t) - \tilde{I}(Z_n,t) \\
&= A_V\, I_V^{-1}(\chi_t) - A_{V_n}\, I_{V_n}^{-1}(\chi_t) - Z(t) + Z_n(t) \\
&= A_V(h_t^0 + h_t^1) - A_{V_n}(h_{tn}^0 + h_{tn}^1) - Z(t) + Z_n(t) \\
&= (A_V - A_{V_n})(h_t^0) + A_{V_n}(h_t^0 - h_{tn}^0) + \big(A_V(h_t^1) - A_{V_n}(h_{tn}^1)\big) - (Z(t) - Z_n(t)). \quad (3.134)
\end{aligned}$$

For convenience we define

$$K(n) \equiv \max\left(\|V - V_n\|_\infty\,,\ \|1/g - 1/g_n\|_\infty\,,\ \|1/d(\cdot,\cdot) - 1/d_n(\cdot,\cdot)\|_\infty\right), \tag{3.135}$$

(thus $K(n) \to 0$) and

$$\begin{aligned}
q_V \;\equiv\; &\max\big((2\,m_0 + (m_0 + m_3 + m_4)\,C_{V,0})\,H_{V,0}\,(3\,H_{V,0}\,C_{V,0} + 1)\ , \\
&(1 + 2\,m_0\,C_{V,0})\,q_{V,1}\ ,\ q_{V,2}\ ,\ (m_2 + 2\,((2R)^2/|W|))\big)\,. \qquad (3.136)
\end{aligned}$$

We start to investigate all terms in (3.134) separately.

- **The term $(A_V - A_{V_n})(h_t^0)$**

Regard $A_V(\star)(\cdot,d)$ as a linear operator from $(D[0,2R), \|\star\|_\infty)$ to $(D[0,2R), \|\star\|_\infty)$ $(d = 0,1,2)$, then for $f \in D[0,2R)$ we write the following derivations.

If $d = 0$ it is easy. Then we have

$$\|(A_V - A_{V_n})(f)(\tilde{x}, d=0)\|_\infty = \|0\|_\infty = 0.$$

If $d = 1$ we write, using integration by parts,

$$\begin{aligned}
&(A_V - A_{V_n})(f)(\tilde{x}, d=1) \\
&= \left(\frac{1}{g(\tilde{x})} - \frac{1}{g_n(\tilde{x})}\right)\int_{x=\tilde{x}}^{x=2R} f(x)\,\frac{1}{|W| + 2xR}\,\mathrm{d}V(x) \\
&\quad + \frac{1}{g_n(\tilde{x})}\int_{x=\tilde{x}}^{x=2R} f(x)\,\frac{1}{|W| + 2xR}\,\mathrm{d}(V - V_n)(x) \\
&= \left(\frac{1}{g(\tilde{x})} - \frac{1}{g_n(\tilde{x})}\right)\int_{x=\tilde{x}}^{x=2R} f(x)\,\frac{1}{|W| + 2xR}\,\mathrm{d}V(x) \\
&\quad + \frac{1}{g_n(\tilde{x})}(V - V_n)(2R)\,f(2R)\,\frac{1}{|W| + 4R^2} \\
&\quad - \frac{1}{g_n(\tilde{x})}(V - V_n)(\tilde{x})\,f(\tilde{x})\,\frac{1}{|W| + 2\tilde{x}R} \\
&\quad - \frac{1}{g_n(\tilde{x})}\int_{x=\tilde{x}}^{x=2R}(V - V_n)(x)\,\frac{1}{|W| + 2xR}\,\mathrm{d}f(x) \\
&\quad + \frac{1}{g_n(\tilde{x})}\int_{x=\tilde{x}}^{x=2R}(V - V_n)(x)\,f(x)\,\frac{2R}{(|W| + 2xR)^2}\,\mathrm{d}x.
\end{aligned}$$

If we define

$$m_3 \equiv 2\,m_0 + \frac{(2R)^2}{|W|^2},$$

then this yields

$$\begin{aligned}
&|(A_V - A_{V_n})(f)(\tilde{x}, d=1)| \\
&\le \|1/g - 1/g_n\|_\infty \|f\|_\infty m_0\, V(2R) + \frac{2}{g_n(2R)} \|V - V_n\|_\infty m_0 \|f\|_\infty \\
&\quad + \frac{1}{g_n(2R)} \|V - V_n)\|_\infty m_0 \int_{x=\tilde{x}}^{x=2R} |\mathrm{d}f(x)| + \frac{1}{g_n(2R)} \|V - V_n\|_\infty \frac{(2R)^2}{|W|^2} \|f\|_\infty \\
&\le \|1/g - 1/g_n\|_\infty \|f\|_\infty m_0\, V(2R) \\
&\quad + \frac{1}{g_n(2R)} \|V - V_n\|_\infty \left(m_3 \|f\|_\infty + m_0 \|f\|_v\right) \\
&\le m_0 \|1/g - 1/g_n\|_\infty \|f\|_v \\
&\quad + \frac{1}{g_n(2R)} \|V - V_n\|_\infty (m_0 + m_3) \|f\|_v .
\end{aligned}$$

For $(A_V - A_{V_n})(f)(\tilde{x}, d = 2)$ one carries out similar calculations as for $d = 1$ and one finds that there exists a constant m_4 such that we have

$$\begin{aligned}
|(A_V - A_{V_n})(f)(\tilde{x}, d=2)| &\le m_0 \|1/d(\cdot,\cdot) - 1/d_n(\cdot,\cdot)\|_\infty \|f\|_v \\
&\quad + \frac{1}{d_n(2R, 2R)} \|V - V_n\|_\infty m_4 \|f\|_v .
\end{aligned}$$

Now we have by the results above, (3.118) and (3.135) that for $n \ge m$

$$\|(A_V - A_{V_n})(f)(\cdot, d)\|_\infty \le (2\,m_0 + (m_0 + m_3 + m_4)\, C_{V,0})\, \|f\|_v\, K(n) \tag{3.137}$$

Because we showed in section 3.8 that $\|h_t^0\|_v \le H_{V,0}\,(3\,H_{V,0}\,C_{V,0} + 1)$, one obtains from (3.137) with $f = h_t^0$ and (3.136) $(n \ge m)$

$$\|(A_V - A_{V_n})(h_t^0)(\cdot, d)\|_\infty \le q_V\, K(n) \tag{3.138}$$

$(d = 0, 1, 2)$.

• The term $A_{V_n}(h_t^0 - h_{tn}^0)$

Again regarding $A_V(\star)(\cdot, d)$ as a linear operator from $(D[0, 2R), \|\star\|_\infty)$ to $(D[0, 2R), \|\star\|_\infty)$ $(d = 0, 1, 2)$, one easily sees that

$$\begin{aligned}
\|A_V(f)(\cdot, d)\|_\infty &\le \|f\|_\infty + \frac{1}{g(2R)} m_0\, V(2R)\, \|f\|_\infty + \frac{1}{d(2R, 2R)} m_0\, V(2R)\, \|f\|_\infty \\
&\le \left(1 + \frac{1}{g(2R)} m_0 + \frac{1}{d(2R, 2R)} m_0\right) \cdot \|f\|_\infty .
\end{aligned} \tag{3.139}$$

(For $d = 0, 1, 2$ we get respectively the 1, the $m_0(1/g(2R))$ and the $m_0(1/d(2R, 2R))$). Now (3.139) holds for all distribution functions V, thus also for $V = V_n$. Because of (3.118) we may write for $n \geq m$

$$\|A_{V_n}(f)(\cdot, d)\|_\infty \leq (1 + 2\,m_0\,C_{V,0})\,\|f\|_\infty. \tag{3.140}$$

Furthermore we write

$$h_t^0(x) - h_{tn}^0(x) = \bar{r}_V(x)\,S_V(\bar{y}_t) - \bar{r}_{V_n}(x)\,S_{V_n}(\bar{y}_t) + \bar{s}_V(x)\,T_V(\bar{y}_t) - \bar{s}_{V_n}(x)\,T_{V_n}(\bar{y}_t).$$

With (3.115), (3.116), (3.124), (3.126) and (3.130) and telescoping we immediately get that S_V and T_V are continuous in V. We find

$$|S_{V_n}(f) - S_V(f)| \to 0 \quad (n \to \infty) \text{ if } \|f\|_v < \infty \tag{3.141}$$

$$|T_{V_n}(f) - T_V(f)| \to 0 \quad (n \to \infty) \text{ if } \|f\|_v < \infty. \tag{3.142}$$

Together with (3.124) and (3.126) we obtain

$$\|h_t^0 - h_{tn}^0\|_\infty \to 0 \quad (n \to \infty).$$

Actually, with a bit more secretarial administration, one can show that there exists a $q_{V,1} \geq 0$ such that $(n \geq m)$

$$|S_{V_n}(f) - S_V(f)| \;\leq\; q_{V,1}\,\|f\|_v\,K(n) \tag{3.143}$$

$$|T_{V_n}(f) - T_V(f)| \;\leq\; q_{V,1}\,\|f\|_v\,K(n) \tag{3.144}$$

$$\left|h_t^0(x) - h_{tn}^0(x)\right| \;\leq\; q_{V,1}\,K(n). \tag{3.145}$$

(Note that the right-hand side of inequality (3.145) does not depend on t, because $\|\bar{y}_t\|_v \leq H_V\,\|y_t\|_\infty$ and $\|y_t\|_\infty \leq 1$ for all t). Finally we obtain with (3.140), (3.145) and (3.136) for $n \geq m$

$$\|A_{V_n}(h_t^0 - h_{tn}^0)(\cdot, d)\|_\infty \leq (1 + 2\,m_0\,C_{V,0})\,q_{V,1} \cdot K(n) \leq q_V\,K(n) \tag{3.146}$$

$(d = 0, 1, 2)$.

- **The term** $A_V(h_t^1) - A_{V_n}(h_{tn}^1)$

Compare the calculations below with the calculations in section 3.8.4. If $d = 0$ we have

$$A_V(h_t^1) - A_{V_n}(h_{tn}^1)(\tilde{x}, 0) = 0 - 0 = 0.$$

If $d = 1$ we find

$$\begin{aligned} A_V(h_t^1) - A_{V_n}(h_{tn}^1)(\tilde{x}, 1) &= \frac{1}{g(\tilde{x})}\,S_V(\bar{y}_t) - \frac{1}{g_n(\tilde{x})}\,S_{V_n}(\bar{y}_t) \\ &= \left(\frac{1}{g(\tilde{x})} - \frac{1}{g_n(\tilde{x})}\right) S_V(\bar{y}_t) \\ &\quad + \frac{1}{g_n(\tilde{x})}\,(S_V(\bar{y}_t) - S_{V_n}(\bar{y}_t)). \end{aligned}$$

For $d = 2$ one get a similar expression. Using (3.118), (3.143), (3.144) and the fact that $\|\bar{y}_t\|_v \leq H_V < \infty$ (H_V of course not depending on t) one finds that there exists a $q_{V,2}$ such that for $n \geq m$

$$\|A_V(h_t^1) - A_{V_n}(h_{tn}^1)(\cdot, d)\|_\infty \leq q_{V,2}\,K(n) \leq q_V\,K(n) \tag{3.147}$$

$(d = 0, 1, 2)$.

• **The term $Z(t) - Z_n(t)$**

For the term $Z(t) - Z_n(t)$ we find (using $z(x) = -\frac{1}{2}a(x)$)

$$\begin{aligned} Z(t) - Z_n(t) &= \int_{x=0}^{x=t} \frac{z(x)}{|W| + 2xR}\,\mathrm{d}(V - V_n)(x) \\ &= (V - V_n)(t)\,\frac{z(t)}{|W| + 2tR} + \int_{x=0}^{x=t} (V - V_n)(x)\,\frac{a(x)}{2(|W| + 2xR)}\,\mathrm{d}x \\ &+ \int_{x=0}^{x=t} (V - V_n)(x)\,\frac{2R\,z(x)}{(|W| + 2xR)^2}\,\mathrm{d}x. \end{aligned}$$

Because $m_2 = z(0)/|W|$ and $a(x) \leq 4R$, this yields

$$|Z(t) - Z_n(t)| \leq \left(m_2 + 2\,\frac{(2R)^2}{|W|}\right) \|V - V_n\|_\infty \leq q_V\,K(n). \tag{3.148}$$

3.9.2 Verification of the $\|\cdot\|_{P_V}$-convergence condition

With the help of section 3.9.1 it will be easy to verify the $\|\cdot\|_{P_V}$-convergence condition. Because of (3.134), (3.138), (3.146), (3.147) and (3.148) we know now that for $n \geq m$

$$\|\tilde{I}(Z,t)(\cdot,d) - \tilde{I}(Z_n,t)(\cdot,d)\|_\infty \leq 4\,q_V\,K(n)$$

(uniformly in t). So we can write

$$\|\tilde{I}(Z,t) - \tilde{I}(Z_n,t)\|_{P_V}^2 = \int \left(\tilde{I}(Z,t)(\tilde{x},d) - \tilde{I}(Z_n,t)(\tilde{x},d)\right)^2 \mathrm{d}P_V(\tilde{x},d) \leq 3\,(4\,q_V)^2\,K^2(n) \tag{3.149}$$

and thus we have (see the definition of $K(n)$ in (3.135))

$$\begin{aligned} &\sup_t \|\tilde{I}(Z,t) - \tilde{I}(Z_n,t)\|_{P_V} \\ &\leq\ 4\sqrt{3}\,q_V\,K(n) \\ &=\ 4\sqrt{3}\,q_V\,\max\left(\|V - V_n\|_\infty\,,\ \|1/g - 1/g_n\|_\infty\,,\ \|1/d(\cdot,\cdot) - 1/d_n(\cdot,\cdot)\|_\infty\right). \end{aligned}$$

Together with the results of section 3.8 this proves the following lemma

Lemma 3.9.2.1 *If we assume that $1/g(x)$ and $1/d(x,x)$ are bounded on $[0,2R]$ and $Q_V \neq 0$ and if V_n is a sequence distribution functions such that $\|V_n - V\|_\infty \to 0$, $\|g_n - g\|_\infty \to 0$ and $\|d_n(\cdot,\cdot) - d(\cdot,\cdot)\|_\infty \to 0$, then we have $Q_{V_n} \to Q_V$ and thus there exists a m such that for all $n \geq m$ we have $Q_{V_n} \neq 0$ and*

$$\lim_{n\to\infty} \sup_t \|\tilde{I}(Z,t) - \tilde{I}(Z_n,t)\|_{P_V} = 0.$$

If we take in lemma 3.9.2.1 $V_n = \widehat{V}_n$ (and thus $g_n = \widehat{g}_n$, $d_n = \widehat{d}_n$ and $Z_n = \widehat{Z}_n$), where $\widehat{V}_n$ is the NPMLE of section 1.2.4, then we have (because we have consistency: $\|V - \widehat{V}_n\|_\infty \to 0$, $\|g - \widehat{g}_n\|_\infty \to 0$ and $\|d(\cdot,\cdot) - \widehat{d}_n(\cdot,\cdot)\|_\infty \to 0$ a.s. (see section 3.4)

$$\lim_{n\to\infty} \sup_t \|\tilde{I}(Z,t) - \tilde{I}(\widehat{Z}_n,t)\|_{P_V} = 0 \quad \text{a.s.} \tag{3.150}$$

If we replace in lemma 3.9.2.1 V (but we do not replace the V in P_V; note that we get (3.149) also with a P_{V_1} instead of P_V) by $\widehat{V}_n$ (with corresponding Z_n) and V_n by $(1-\frac{1}{k})\,\widehat{V}_n+\frac{1}{k}V$ (with corresponding Z_{nk}), then we have immediately

$$\lim_{k\to\infty} \sup_t \|\tilde{I}(Z_n,t) - \tilde{I}(Z_{nk},t)\|_{P_V} = 0 \quad \text{a.s.} \tag{3.151}$$

3.10 The determinant $Q_V \neq 0$ and assumption I

3.10.1 Some remarks about the determinant Q_V

For the existence of a hardest submodel h_t in section 3.7, we had to assume that the determinant $Q_V \neq 0$. In this case we can write down the hardest submodel as the solution h_t of the equation $I_V(h_t)(x) = \chi_t(x)$ in lemma 3.7.2.1. One has to be sure that the determinant $Q_V \neq 0$ for the underlying distribution function V. It could be possible that $Q_V = 0$ is an identity of the model for all choices of distribution functions V. Then the entire analysis in the previous sections would be worthless. Therefore we have to show that $Q_V \neq 0$ is not an empty statement.

We take a distribution function V such that $V(x) = 0$ for all $x \in [0, 2R]$. In this case we have $g(x) = g(2R)$ for all $x \in [0, 2R]$ (and $g(2R)$ is assumed to be > 0) and $d(x,x) = d(2R, 2R)$ for all $x \in [0, 2R]$ (and $d(2R, 2R)$ is assumed to be > 0). One easily checks that the operators Ψ_V and Λ_V in (3.31) and (3.32) both become the nil-operator. This means that the operators α_V and β_V as operators on $D[0, 2R)$ defined by (3.38) are nil-operator too and so we get immediately that $Q_V = 1$. This shows that there are distribution functions V for which the determinant Q_V is not equal to 0.

If one can only observe within a window with diameter $2R$, an underlying distribution function V such that $V(x) = 0$ for all $x \in [0, 2R]$, is not interesting to estimate. Are there distribution functions V with mass on $[0, 2R]$ for which $Q_V \neq 0$? Let V be a distribution function such that $V(x) = 0$ for all $x \in [0, 2R]$ and let $\tilde{V}$ be an arbitrary distribution function. Then $V_n \equiv (1-\epsilon_n)V + \epsilon_n\tilde{V}$, where $\epsilon_n \downarrow 0$, converges in supnorm to V. By section 3.8.5 we know that this implies that $Q_{V_n} \to Q_V$. So for n large enough we have that $Q_{V_n} \neq 0$. This proves that there exist distribution functions V with mass on $[0, 2R]$ for which $Q_V \neq 0$.

The assumption $Q_V \neq 0$ is not something that one can easily assume from arguments of the underlying V. On the other hand one could imagine that $Q_V \neq 0$ holds for a large class of distribution functions V. In section 3.12 we calculate the determinant Q_V in the one-dimensional line segment problem and show there that $Q_V \geq 1$. In section 4.4 we conjecture that in the two-dimensional case we also have $Q_V \geq 1$ for all V.

3.10.2 The assumption $g(2R) > 0$ and $d(2R, 2R) > 0$

Let us consider the class of distribution functions V for which $V(x) = 1$ in (e_2, ∞) for an $e_2 < 2R$. For a V in this class we can write for $g(x)$ and $d(x,x)$

$$g(x) = \int_{u=x}^{u=e_2} \frac{1}{|W| + 2uR}\, \mathrm{d}V(u) \quad , \quad d(x,x) = \int_{u=x}^{u=e_2} \frac{u-x}{|W| + 2uR}\, \mathrm{d}V(u)$$

and one notes that these are the underlying distribution functions V for which $g(2R) = 0$ and $d(2R, 2R) = 0$. For the information operator I_V we find

$$\begin{aligned} & I_V(h)(x) \\ = & \frac{z(x)}{|W| + 2xR} h(x) \cdot 1(x \le e_2) \\ + & \frac{1}{|W|+2xR} \int_{\tilde{x}=0}^{\tilde{x}=x\wedge e_2} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} h(u) \frac{1}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x} \\ + & \frac{1}{|W|+2xR} \int_{\tilde{x}=0}^{\tilde{x}=x\wedge e_2} \frac{b(\tilde{x})\, \tilde{x}(x-\tilde{x})}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} h(u) \frac{u-\tilde{x}}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x}. \qquad (3.152) \end{aligned}$$

Again we look at the equation (3.28): $I_V(h_t)(x) = \chi_t(x)$, where χ_t is defined as in (3.18). If we consider the invertibility of $I_V(h_t)(x) = \chi_t(x)$ for $x \ge 2R$, then one notes immediately that $I_V(h)(x)$ does not depend on the values of $h(x)$ for $x \in (e_2, \infty)$. So for all functions h_1 and h_2 with support on (e_2, ∞) we have $I_V(h_1) = I_V(h_2) = 0$.

For the invertibility of $I_V(h_t)(x) = \chi_t(x)$ for $x \in [0, e_2)$, we work as follows. Just as we did in section 3.7.1, we find for $x \ge 2R$ (actually, for $x \ge e_2$)

$$\begin{aligned} & \int_{\tilde{x}=0}^{\tilde{x}=e_2} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} h_t(u) \frac{1}{|W|+2xR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x} \\ & \quad + \int_{\tilde{x}=0}^{\tilde{x}=e_2} \frac{b(\tilde{x})\, \tilde{x}(x-\tilde{x})}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} h_t(u) \frac{u-\tilde{x}}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x} = 0. \end{aligned}$$

From this we obtain similar equations like (3.29) and (3.30). One derives the equations (compare with (3.39) and (3.40))

$$\begin{aligned} & \int_{\tilde{x}=0}^{\tilde{x}=x} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} h_t(u) \frac{1}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x} \\ & \quad - \int_{\tilde{x}=0}^{\tilde{x}=x} \frac{b(\tilde{x})\, \tilde{x}^2}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} h_t(u) \frac{u-\tilde{x}}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x} \\ = & \quad - \int_{\tilde{x}=x}^{\tilde{x}=e_2} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} h_t(u) \frac{1}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x} \\ & \quad + \int_{\tilde{x}=x}^{\tilde{x}=e_2} \frac{b(\tilde{x})\, \tilde{x}^2}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} h_t(u) \frac{u-\tilde{x}}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x} \qquad (3.153) \end{aligned}$$

and

$$\begin{aligned} & \int_{\tilde{x}=0}^{\tilde{x}=x} \frac{b(\tilde{x})\, \tilde{x}}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} h_t(u) \frac{u-\tilde{x}}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x} \\ = & \quad - \int_{\tilde{x}=x}^{\tilde{x}=e_2} \frac{b(\tilde{x})\, \tilde{x}}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} h_t(u) \frac{u-\tilde{x}}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x}. \qquad (3.154) \end{aligned}$$

Together with (3.153) and (3.154) we write for $x \in [0, e_2]$ (compare with (3.41))

$$I_V(h_t)(x) = \frac{z(x)}{|W|+2xR} (h_t(x) - B_V h_t(x)),$$

where $B_V : (D[0,e_2], \|\cdot\|_\infty) \to (D[0,e_2], \|\cdot\|_\infty)$ is defined as $B_V h \equiv B_{V,1}h + B_{V,2}h$ and the operators $B_{V,i} : (D[0,e_2], \|\cdot\|_\infty) \to (D[0,e_2], \|\cdot\|_\infty)$ are

$$\begin{aligned} B_{V,1}h(x) &\equiv \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=e_2} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} h(u) \frac{1}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x} \\ B_{V,2}h(x) &\equiv \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=e_2} \frac{b(\tilde{x})\,\tilde{x}(x-\tilde{x})}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} h(u) \frac{u-\tilde{x}}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x}. \end{aligned}$$

Now we want to determine the inverse of $(I - B_V)$. The method in section 3.7.2 can not be used now. Changing the order of integration will not help to find a kernel for which we can prove that it is bounded. But we use here the advantage of $e_2 < 2R$. On $[0,e_2] \times [0,e_2]$ we have that

$$0 \le \frac{a(y) + b(y)\,y(y-x)}{z(x)} \le \frac{a(0) + b(e_2)\,e_2^2}{z(e_2)} \equiv c < \infty.$$

Now we prove that

$$|B_V^k h(x)| \le \frac{(c\,(e_2 - x))^k}{k!} \|h\|_\infty \tag{3.155}$$

We derive for $k = 1$ and $x \in [0,e_2]$

$$\begin{aligned} &|B_V h(x)| \\ &\le |B_{V,1}h(x)| + |B_{V,2}h(x)| \\ &\le \|h\|_\infty \cdot \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=e_2} \frac{a(\tilde{x})}{g(\tilde{x})} g(\tilde{x}) \, \mathrm{d}\tilde{x} \\ &\quad + \|h\|_\infty \cdot \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=e_2} \frac{b(\tilde{x})\,\tilde{x}(x-\tilde{x})}{d(\tilde{x},\tilde{x})} d(\tilde{x},\tilde{x}) \, \mathrm{d}\tilde{x} \\ &\le c\,\|h\|_\infty \int_{\tilde{x}=x}^{\tilde{x}=e_2} \mathrm{d}\tilde{x} \\ &\le c\,(e_2 - x)\,\|h\|_\infty. \end{aligned}$$

Suppose (3.155) is true for a certain k. We derive for $k+1$ (and $x \in [0,e_2]$)

$$\begin{aligned} &|B_V^{k+1} h(x)| \\ &\le \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=e_2} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} |B_V^k h(u)| \frac{1}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x} \\ &\quad + \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=e_2} \frac{b(\tilde{x})\,\tilde{x}(\tilde{x}-x)}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=e_2} |B_V^k h(u)| \frac{u-\tilde{x}}{|W|+2uR} \, \mathrm{d}V(u) \, \mathrm{d}\tilde{x} \\ &\le \|h\|_\infty \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=e_2} \frac{a(\tilde{x})}{g(\tilde{x})} \frac{(c\,(e_2-\tilde{x}))^k}{k!} g(\tilde{x}) \, \mathrm{d}\tilde{x} \\ &\quad + \|h\|_\infty \frac{1}{z(x)} \int_{\tilde{x}=x}^{\tilde{x}=e_2} \frac{b(\tilde{x})\,\tilde{x}(\tilde{x}-x)}{d(\tilde{x},\tilde{x})} \frac{(c\,(e_2-\tilde{x}))^k}{k!} d(\tilde{x},\tilde{x}) \, \mathrm{d}\tilde{x} \\ &\le \|h\|_\infty \frac{c^{k+1}}{k!} \int_{\tilde{x}=x}^{\tilde{x}=e_2} (e_2 - \tilde{x})^k \, \mathrm{d}\tilde{x} \\ &= \frac{(c\,(e_2 - x))^{k+1}}{(k+1)!} \|h\|_\infty. \end{aligned}$$

Now we showed with induction that (3.155) holds for all k. This provides us with the following statement

$$\|B_V^k h\|_\infty \leq \frac{(c\,e_2)^k}{(k)!}\,\|h\|_\infty.$$

This proves the existence of the inverse of $(I - B_V)$ in the space of bounded linear operators on $D[0, e_2]$ being $(I - B_V)^{-1} = \sum_{k=0}^{k=\infty} B_V^k$ and $\|(I - B_V)^{-1}h\|_\infty \leq \exp(c\,e_2)\,\|h\|_\infty$. Now the equation $I_V(h_t)(x) = \chi_t(x)$ for $x \in [0, e_2]$ can be written as

$$h_t(x) = \bar{y}_t(x), \tag{3.156}$$

where $\bar{y}_t = (I - B_V)^{-1} 1_{(0,t]}$. Actually we proved the following lemma (compare with lemma 3.7.2.1).

Lemma 3.10.2.1 *If we consider the distribution function V for which $V(x) = 1$ on $[e_2, \infty)$ for an $e_2 < 2R$ and let h_t^0 and h_t^1 be defined as*

$$h_t^0(x) \equiv h_t(x) \cdot 1_{[0,e_2)}(x) \quad , \quad h_t^1(x) \equiv h_t(x) \cdot 1_{[e_2,\infty)}(x),$$

then

$$I_V(h_t)(x) = \chi_t(x) \quad \textit{for all } x \in [0, \infty)$$

is equivalent to

$$h_t^0(x) = \bar{y}_t(x)) \cdot 1_{[0,e_2)}(x)$$

and we can choose $h_t^1 \equiv 0$.

Actually it is not strange to find lemma 3.10.2.1 as the solution of h_t. In the case that $\mathrm{d}V(x) = 0$ for all $x \in [2R, \infty)$ we find that α_V and β_V in section 3.7.1 are the nil-operators and thus (3.156) could be expected from (3.69).

If one wants to check the Donsker class condition and the $\|\cdot\|_{P_V}$-convergence condition in this case, then one goes through the sections 3.8 and 3.9. A lot of the analysis there can be skipped. We only have to look at section 3.8.2 and the beginning of section 3.8.4 and one has to check (3.121); the continuity of $(I - B_V)^{-1}$ in V. In section 3.8.2 one must be aware that if one tries to bound $(\mathrm{d}/\mathrm{d}x)\,B_{V,1}^{\mathrm{up}} f(x)$ from above and $(\mathrm{d}/\mathrm{d}x)\,B_{V,2} f(x)$ from below, then one must use the same technique as we did above to determine the inverse of $(I - B_V)$. This provides us with similar efficiency results in the case that $g(2R) = 0$ and $d(2R, 2R) = 0$ and $V(x)$ equals 1 for an $e_2 < 2R$.

3.11 Calculation of $I_V^{-1}(\xi - \mathcal{Z})$

In this section we only give the calculation of $I_V^{-1}(\xi - \mathcal{Z})$ which we need for theorem 3.5.2.1. The calculation is at some points different to the calculation of $I_V^{-1}(\chi_t - \mathcal{Z}(t))$ in section 3.7.

We know that for a constant c we have $I_V c = c$ and thus $I_V^{-1}\mathcal{Z} = \mathcal{Z}$. Therefore we only have to concentrate on the equation

$$I_V(h)(x) = \xi,$$

assuming for the moment that such a solution h exists. (ξ as in (3.21)).

One writes ξ as

$$\xi(x) = \frac{1}{|W| + 2xR}(0 + x).$$

Now we imitate section 3.7.1. For $x \geq 2R$ we get from $I_V(h)(x) = \xi(x)$ with (3.27) the following two equations analogous to (3.35)

$$\begin{pmatrix} c_{V,4}(0) + c_{V,3}(0) & -c_{V,2}(0) \\ -c_{V,2}(0) & c_{V,1}(0) \end{pmatrix} \cdot \begin{pmatrix} \alpha_V(h) \\ \beta_V(h) \end{pmatrix} = - \begin{pmatrix} \Psi_V(h) - 0 \\ \Lambda_V(h) - 1 \end{pmatrix}. \tag{3.157}$$

(All the functionals and functions defined as in section 3.7.1). Using the same arguments as in section 3.7.1, we get the existence of the inverse of the matrix and get the system of equations analogous to (3.38)

$$\begin{aligned} \begin{pmatrix} \alpha_V(h) \\ \beta_V(h) \end{pmatrix} &= -\frac{1}{D_V} \begin{pmatrix} c_{V,1}(0) & c_{V,2}(0) \\ c_{V,2}(0) & c_{V,4}(0) + c_{V,3}(0) \end{pmatrix} \cdot \begin{pmatrix} \Psi_V(h) \\ \Lambda_V(h) - 1 \end{pmatrix} \\ &= -L_V^{-1} \begin{pmatrix} \Psi_V(h) \\ \Lambda_V(h) - 1 \end{pmatrix}. \end{aligned} \tag{3.158}$$

Again we expressed $\alpha_V(h)$ and $\beta_V(h)$ as operators which only use the values of $h(x)$ with $x \in [0, 2R)$. For each choice of h on $[0, 2R)$ one is able to find an h on $[2R, \infty)$ such that (3.158) holds. (See the remark after lemma 3.7.2.1).

If we define the operators $\tilde{\alpha}_V$ and $\tilde{\beta}_V$ for an f on $D[0, 2R)$ as

$$\begin{pmatrix} \tilde{\alpha}_V(f) \\ \tilde{\beta}_V(f) \end{pmatrix} = -L_V^{-1} \begin{pmatrix} \Psi_V(f) \\ \Lambda_V(f) \end{pmatrix} + L_V^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{3.159}$$

and we define α_V and β_V regarded as operators on $D[0, 2R)$ just the same as in (3.38) then we get by (3.158)

$$\begin{pmatrix} \tilde{\alpha}_V(h^0) \\ \tilde{\beta}_V(h^0) \end{pmatrix} = \begin{pmatrix} \alpha_V(h^0) \\ \beta_V(h^0) \end{pmatrix} + L_V^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_V(h^1) \\ \beta_V(h^1) \end{pmatrix}, \tag{3.160}$$

where $h^0(x) = h(x) \cdot 1_{[0,2R)}(x)$ and $h^1(x) = h(x) \cdot 1_{[2R,\infty)}(x)$.

For the invertibility of $I_V(h)(x) = \xi(x)$ for $x \in [0, 2R)$ we imitate section 3.7.2. One easily checks that the equations analogous to (3.39) and (3.40) (derived from (3.29) and (3.30)) are

$$\begin{aligned} &\int_{\tilde{x}=0}^{\tilde{x}=x} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h(u) \frac{1}{|W| + 2uR} \, dV(u) \, d\tilde{x} \\ &\quad - \int_{\tilde{x}=0}^{\tilde{x}=x} \frac{b(\tilde{x})\, \tilde{x}^2}{d(\tilde{x}, \tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h(u) \frac{u - \tilde{x}}{|W| + 2uR} \, dV(u) \, d\tilde{x} \\ = &\quad - \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h(u) \frac{1}{|W| + 2uR} \, dV(u) \, d\tilde{x} \\ &\quad + \int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\, \tilde{x}^2}{d(\tilde{x}, \tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h(u) \frac{u - \tilde{x}}{|W| + 2uR} \, dV(u) \, d\tilde{x} + 0 \end{aligned} \tag{3.161}$$

and

$$\int_{\tilde{x}=0}^{\tilde{x}=x} \frac{b(\tilde{x})\,\tilde{x}}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h(u)\,\frac{u-\tilde{x}}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}$$
$$= -\int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h(u)\,\frac{u-\tilde{x}}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} + 1. \tag{3.162}$$

With (3.161) and (3.162) and the definition of I_V in (3.27) we get from $I_V(h) = \xi(x)$ for $x \in [0, 2R)$

$$\begin{aligned} I_V(h)(x) &= \frac{z(x)}{|W|+2xR}\,h(x) \\ &\quad -\frac{1}{|W|+2xR}\int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{a(\tilde{x})}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h(u)\,\frac{1}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} \\ &\quad +\frac{1}{|W|+2xR}\int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}^2}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h(u)\,\frac{u-\tilde{x}}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} \\ &\quad -\frac{x}{|W|+2xR}\int_{\tilde{x}=x}^{\tilde{x}=2R} \frac{b(\tilde{x})\,\tilde{x}}{d(\tilde{x},\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h(u)\,\frac{u-\tilde{x}}{|W|+2uR}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} \\ &\quad +\frac{x}{|W|+2xR} \\ &= \frac{x}{|W|+2xR}. \end{aligned} \tag{3.163}$$

Now from (3.163) we obtain just as in (3.41)

$$I_V(h)(x) = \frac{z(x)}{|W|+2xR}\,(h(x) - B_V(h)(x) - r_V(x)\,\alpha_V(h) - s_V(x)\,\beta_V(h)) = 0$$

and this yields for $x \in [0, 2R)$

$$h(x) = \bar{r}_V(x)\,\alpha_V(h^1) + \bar{s}_V(x)\,\beta_V(h^1).$$

(See section 3.7 for all the definitions). Now we apply $\tilde{\alpha}_V$ and $\tilde{\beta}_V$ defined by (3.159) on both sides of the equation and we obtain

$$\begin{pmatrix} \alpha_V(h^1) \\ \beta_V(h^1) \end{pmatrix} = \begin{pmatrix} \alpha_V(\bar{r}_V)\,\alpha_V(h^1) + \alpha_V(\bar{s}_V)\,\beta_V(h^1) \\ \beta_V(\bar{r}_V)\,\alpha_V(h^1) + \beta_V(\bar{s}_V)\,\beta_V(h^1) \end{pmatrix} + L_V^{-1}\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

This yields

$$\begin{pmatrix} 1-\alpha_V(\bar{r}_V) & -\alpha_V(\bar{s}_V) \\ -\beta_V(\bar{r}_V) & 1-\beta_V(\bar{s}_V) \end{pmatrix} \cdot \begin{pmatrix} \alpha_V(h^1) \\ \beta_V(h^1) \end{pmatrix} = N_V \begin{pmatrix} \alpha_V(h^1) \\ \beta_V(h^1) \end{pmatrix} = L_V^{-1}\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and under the assumption that $Q_V \neq 0$ we obtain

$$\begin{pmatrix} \alpha_V(h^1) \\ \beta_V(h^1) \end{pmatrix} = N_V^{-1}\,L_V^{-1}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \equiv \begin{pmatrix} S_V \\ T_V \end{pmatrix}.$$

Now we have proved the following lemma.

Lemma 3.11.1 *If we assume that $1/g(x)$ and $1/d(x,x)$ are bounded on $[0,2R]$ and $Q_V \neq 0$ and let h^0 and h^1 be defined as*

$$h^0(x) \equiv h(x)\cdot 1_{[0,2R)}(x) \quad , \quad h^1(x) \equiv h(x)\cdot 1_{[2R,\infty)}(x),$$

then

$$I_V(h)(x) = \xi(x) \text{ for all } x \in [0,\infty)$$

is equivalent to

$$h^0(x) = \bar{r}_V(x)\, S_V + \bar{s}_V(x)\, T_V$$

and h^1 satisfies

$$\begin{pmatrix} \alpha_V(h^1) \\ \beta_V(h^1) \end{pmatrix} = -\frac{1}{D_V}\begin{pmatrix} c_{V,1}(0) & c_{V,2}(0) \\ c_{V,2}(0) & c_{V,4}(0)+c_{V,3}(0) \end{pmatrix}\cdot\begin{pmatrix} \Psi_V(h^0) \\ \Lambda_V(h^0)-1 \end{pmatrix}.$$

3.12 Efficiency proof for the one-dimensional case

The method we used to obtain the efficiency results in the 'circle'-case can also be applied to the one-dimensional case. Van der Laan(1993) studied the efficiency in the one-dimensional line segment problem, but the proof there is incorrect without the assumptions in section 3.4. If we put the proof of the one-dimensional case into the 'general' setting of the proof of the two-dimensional case, then we will see why the two-dimensional case is harder to solve.

In the one-dimensional line segment process the score operator is given by

$$\begin{aligned} & A_V(h)(\tilde{x},d) \\ &= h(\tilde{x})\cdot 1(d=0) \\ &\quad +\frac{1}{g(\tilde{x})}\int_{x=\tilde{x}}^{x=\infty} h(x)\frac{1}{\tau+x}\,\mathrm{d}V(x)\cdot 1(d=1) \\ &\quad +\frac{1}{d(\tau,\tau)}\int_{x=\tau}^{x=\infty} h(x)\frac{x-\tau}{\tau+x}\,\mathrm{d}V(x)\cdot 1(d=2). \end{aligned}$$

For the adjoint of the score operator we find

$$\begin{aligned} & A_V^T(\eta)(x) \\ &= \frac{\tau-x}{\tau+x}\eta(x,0)\cdot 1(x<\tau) \\ &\quad +\frac{2}{\tau+x}\int_{\tilde{x}=0}^{\tilde{x}=x\wedge\tau} \eta(\tilde{x},1)\,\mathrm{d}\tilde{x} \\ &\quad +\frac{x-\tau}{\tau+x}\eta(\tau,2)\cdot 1(x\geq\tau). \end{aligned}$$

Thus for the information operator $I_V = A_V^T A_V$ we obtain

$$\begin{aligned} & I_V(h)(x) \\ &= \frac{\tau-x}{\tau+x}h(x)\cdot 1(x<\tau) \end{aligned}$$

$$+\frac{2}{\tau+x}\int_{\tilde{x}=0}^{\tilde{x}=x\wedge\tau}\frac{1}{g(\tilde{x})}\int_{u=\tilde{x}}^{u=\infty}h(u)\frac{1}{\tau+u}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}$$
$$+\frac{x-\tau}{\tau+x}\cdot\frac{1}{d(\tau,\tau)}\int_{u=\tau}^{u=\infty}h(u)\frac{u-\tau}{\tau+u}\,\mathrm{d}V(u)\cdot 1(x\geq\tau).$$

Recall that the parameter h in section 1.1.2 equals $d(\tau,\tau)$ here. To avoid misunderstanding we write throughout this section $\mathcal{W}=d(\tau,\tau)$ instead of the parameter h used in chapter 1 and 2. Again for convenience and without loss of generality we write P_V (V in the set of distribution functions on $[0,\infty)$) instead of $P_{V,h}$ with $(V,h)\in\mathcal{V}_\tau$.

3.12.1 The parameters $(W,\mathcal{W})$ to be estimated

As in section 3.5, we introduce here the parameters for which we show that the NPMLE is efficient. We consider the estimation of the parameter

$$\begin{aligned}
\vartheta(P_V) &= (\vartheta_1(P_V),\vartheta_2(P_V))\\
&\equiv (W(\cdot),\mathcal{W})\\
&\equiv \left(\int_{x=0}^{x=\cdot}\frac{\tau-x}{\tau+x}\,\mathrm{d}V(x),\int_{x=\tau}^{x=\infty}\frac{x-\tau}{\tau+x}\,\mathrm{d}V(x)\right)\\
&= (F^{\mathrm{u.c.}}(\cdot),d(\tau,\tau)),
\end{aligned}$$

where we define $b_t\vartheta_1(P_V)=W(t)$ and $B=\{b_t : t\in[0,\tau)\}$. (Compare with (3.17)). Again we consider the distribution function of the uncensored observations because then we have less trouble with the singularity problems at $x=\tau$, which occur if one inverts the information operator.

- **The parameter $b_t\vartheta_1(P_V)=W(t)$**

Just as in section 3.5.1, we calculate the pathwise derivative of $W(t)$. If we define the function κ_t to be

$$\kappa_t(x)\equiv\frac{\tau-x}{\tau+x}\cdot 1_{(0,t]}(x),\tag{3.164}$$

then we note that

$$\kappa_t-W(t)=\kappa_t-\int\kappa_t(x)\,\mathrm{d}V(x)=\kappa_t-E_V(\kappa_t)\in L_0^2(V)$$

and we get the following equality analogous to (3.19)

$$\frac{1}{\epsilon}\left(W_{\epsilon,l}(t)-W(t)\right)=\left\langle A_V\,I_V^{-1}(\kappa_t-W(t)),A_V(l)\right\rangle_{P_V}.\tag{3.165}$$

Assuming that $I_V^{-1}(\kappa_t-W(t))$ exists, we have with (3.165) that $W(t)$ is pathwise differentiable with efficient influence curve

$$\tilde{I}(W,t)=A_V\,I_V^{-1}(\kappa_t-W(t)).\tag{3.166}$$

In section 3.12.2 we calculate $h_t=I_V^{-1}(\kappa_t-W(t))$. The calculation must be compared with section 3.7 of the two-dimensional case. In lemma 3.7.2.1 we have to assume that the determinant Q_V is not equal 0 to be sure of the existence of h_t. In the one-dimensional case we are able to prove that this Q_V is not equal to 0, therefore this assumption is not needed in lemma 3.12.2.1. One verifies the conditions of theorem 3.3.2 or checks the conditions of theorem 3.3.4 using section 3.8 and 3.9 (and the proof of theorem 3.5.1.1) to obtain the next theorem

Theorem 3.12.1.1 *Under the assumptions in section 3.4 the NPMLE $\widehat{W}_n$ is a $\|\cdot\|_B$- asymptotically efficient estimator of W.*

- **The parameter $\vartheta_2(P_V) = \mathcal{W}$**

We define the function γ to be

$$\gamma(x) \equiv \frac{x-\tau}{\tau+x} \cdot 1_{[\tau,\infty)}(x). \tag{3.167}$$

Again one checks that

$$\gamma - \mathcal{W} = \gamma - \int \gamma(x)\,\mathrm{d}V(x) = \gamma - E_V(\gamma) \in L_0^2(V).$$

For the parameter $\mathcal{W}$ one finds

$$\begin{aligned} \frac{1}{\epsilon}(\mathcal{W}_{\epsilon,l} - \mathcal{W}) &= \int \gamma(x)\,l(x)\,\mathrm{d}V(x) \\ &= d(\tau,\tau)\cdot A_V(l)(\tau,2) \\ &= \langle 1(d=2), A_V(l)\rangle_{P_V} \\ &= \langle 1(d=2) - \mathcal{W}, A_V(l)\rangle_{P_V}, \end{aligned}$$

showing that $\mathcal{W}$ is pathwise differentiable with efficient influence curve

$$\tilde{I}(\mathcal{W}) = 1(d=2) - \mathcal{W}.$$

We know already that the NPMLE $\widehat{\mathcal{W}}_n$ equals the fraction of double censored observations (see (1.51)). So we get

$$\begin{aligned} \widehat{\mathcal{W}}_n - \mathcal{W} &= \frac{1}{n}\sum_{i=1}^{n} 1(D_i = d_i = 2) - \mathcal{W} \\ &= \frac{1}{n}\sum_{i=1}^{n}(1(d_i=2) - \mathcal{W}) \\ &= \int \tilde{I}(\mathcal{W})\,\mathrm{d}P_n \\ &= \int \tilde{I}(\mathcal{W})\,\mathrm{d}(P_n - P_V). \end{aligned}$$

This immediately proves the efficiency of the NPMLE $\widehat{\mathcal{W}}_n$ of $\mathcal{W}$. The explicit calculation of $I_V^{-1}(\gamma - \mathcal{W})$ is not needed. Now we got without any assumption:

Theorem 3.12.1.2 *The NPMLE $\widehat{\mathcal{W}}_n$ is an asymptotically efficient estimator of $\mathcal{W}$.*

• Efficiency of $(\widehat{W}_n, \widehat{\mathcal{W}}_n)$ implies efficiency of $(\widehat{F}_n, \widehat{\mu}_n)$

Just as in section 3.5.3 one can show that the fact that the NPMLE $(\widehat{W}_n, \widehat{\mathcal{W}}_n)$ is an efficient estimator of the underlying $(W, \mathcal{W}) \in D[0,\tau) \times \mathbb{R}$ implies that the NPMLE $(\widehat{F}_n, \widehat{\mu}_n)$ is an efficient estimator of the underlying $(F, \mu) \in D[0, \tau - \epsilon) \times \mathbb{R}$ for every fixed $\epsilon \in (0, 2R)$.

Here one uses the relation

$$\mathrm{d}V(x) = \frac{\tau + x}{\tau + \mu}\,\mathrm{d}F(x)$$

to obtain

$$1 - \mathcal{W} + W(\tau) = 1 - d(\tau,\tau) + \int_{x=0}^{x=\tau} \frac{\tau - x}{\tau + x}\,\mathrm{d}V(x) = \frac{2\tau}{\tau + \mu}$$

and

$$\mathrm{d}W(x) = \frac{\tau - x}{\tau + x}\,\mathrm{d}V(x) = \frac{\tau - x}{\tau + \mu}\,\mathrm{d}F(x).$$

From these we get

$$\mu = \frac{2\tau}{1 - \mathcal{W} + W(\tau)} - \tau$$

and

$$F(t) = \frac{2\tau}{1 - \mathcal{W} + W(\tau)} \left(\frac{W(t)}{\tau - t} + \int_{x=0}^{x=t} \frac{1}{(\tau - x)^2}\,W(x)\,\mathrm{d}x \right).$$

Now one copies the proof given in section 3.5.3.

3.12.2 Calculation of $I_V^{-1}(\kappa_t - W(t))$

Again we remember that $I_V c = c$ for a constant c and thus $I_V^{-1}(W(t)) = W(t)$ and therefore we only have to concentrate on solving

$$I_V(h_t)(x) = \kappa_t(x), \tag{3.168}$$

assuming for the moment that such a h_t exists. (κ_t as in (3.164)).

• Invertibility of $I_V(h_t)(x) = \kappa_t(x)$ for $x \geq \tau$

For $x \geq \tau$ equation (3.168) becomes

$$2 \int_{\tilde{x}=0}^{\tilde{x}=\tau} \frac{1}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u)\,\frac{1}{\tau + x}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}$$
$$+(x - \tau)\,\frac{1}{d(\tau,\tau)} \int_{u=\tau}^{u=\infty} h_t(u)\,\frac{u - \tau}{\tau + u}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} = 0.$$

This yields (compare with (3.29) and (3.30))

$$2 \int_{\tilde{x}=0}^{\tilde{x}=\tau} \frac{1}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u)\,\frac{1}{\tau + x}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}$$
$$-\tau\,\frac{1}{d(\tau,\tau)} \int_{u=\tau}^{u=\infty} h_t(u)\,\frac{u - \tau}{\tau + u}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} = 0 \tag{3.169}$$

and

$$\frac{1}{d(\tau,\tau)}\int_{u=\tau}^{u=\infty} h_t(u)\,\frac{u-\tau}{\tau+u}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}=0. \tag{3.170}$$

So (3.169) with (3.170) and the assumption that $d(\tau,\tau)>0$ gives us

$$\int_{\tilde{x}=0}^{\tilde{x}=\tau}\frac{1}{g(\tilde{x})}\int_{u=\tilde{x}}^{u=\infty} h_t(u)\,\frac{1}{\tau+x}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}=0 \tag{3.171}$$

and

$$\int_{u=\tau}^{u=\infty} h_t(u)\,\frac{u-\tau}{\tau+u}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}=0. \tag{3.172}$$

Now if we define the operator Ψ_V as follows

$$\Psi_V(h)\equiv\int_{u=0}^{u=\tau}\int_{\tilde{x}=0}^{\tilde{x}=u}\frac{1}{g(\tilde{x})}\,\mathrm{d}\tilde{x}\,h(u)\,\frac{1}{\tau+u}\,\mathrm{d}V(u)$$

and we define the operators α_V and β_V as

$$\begin{aligned}\alpha_V(h) &\equiv \int_{u=\tau}^{u=\infty} h(u)\,\frac{1}{\tau+u}\,\mathrm{d}V(u)\\ \beta_V(h) &\equiv \int_{u=\tau}^{u=\infty} h(u)\,\frac{u}{\tau+u}\,\mathrm{d}V(u)\end{aligned}$$

and we define

$$c_V(x)\equiv\int_{\tilde{x}=x}^{\tilde{x}=\tau}\frac{1}{g(\tilde{x})}\,\mathrm{d}\tilde{x},$$

then the equations (3.171) and (3.172) can be written as

$$\begin{pmatrix} c_V(0) & 0\\ -\tau & 1\end{pmatrix}\cdot\begin{pmatrix}\alpha_V(h_t)\\ \beta_V(h_t)\end{pmatrix}=-\begin{pmatrix}\Psi_V(h_t)\\ 0\end{pmatrix}. \tag{3.173}$$

If we assume that $1/g(x)$ is bounded on $[0,\tau)$ we have that $0<c_V(0)<\infty$ and thus the determinant D_V of the matrix

$$\begin{pmatrix} c_V(0) & 0\\ -\tau & 1\end{pmatrix}$$

equals $c_V(0)\neq 0$ and therefore we may write

$$\begin{pmatrix}\alpha_V(h_t)\\ \beta_V(h_t)\end{pmatrix}=-\frac{1}{D_V}\begin{pmatrix} 1 & 0\\ \tau & c_V(0)\end{pmatrix}\cdot\begin{pmatrix}\Psi_V(h_t)\\ 0\end{pmatrix}. \tag{3.174}$$

(Compare (3.174) with (3.38)).

- **Invertibility of $I_V(h_t)(x)=\kappa_t(x)$ for $x\in[0,\tau)$**

Here we do the same as in section 3.7.2. One writes (3.171) as

$$\begin{aligned}&\int_{\tilde{x}=0}^{\tilde{x}=x}\frac{1}{g(\tilde{x})}\int_{u=\tilde{x}}^{u=\infty} h_t(u)\,\frac{1}{\tau+x}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}\\ &= -\int_{\tilde{x}=x}^{\tilde{x}=\tau}\frac{1}{g(\tilde{x})}\int_{u=\tilde{x}}^{u=\infty} h_t(u)\,\frac{1}{\tau+x}\,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}.\end{aligned} \tag{3.175}$$

Because of (3.170) and with (3.175) we write for $x \in [0, \tau)$

$$
\begin{aligned}
I_V(h_t)(x) &= \frac{\tau - x}{\tau + x} h_t(x) \\
&\quad + \frac{2}{\tau + x} \int_{\tilde{x}=0}^{\tilde{x}=x\wedge\tau} \frac{1}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h(u) \frac{1}{\tau + u} \,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} \\
&= \frac{\tau - x}{\tau + x} h_t(x) \\
&\quad - \frac{2}{\tau + x} \int_{\tilde{x}=x}^{\tilde{x}=\tau} \frac{1}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\infty} h_t(u) \frac{1}{\tau + x} \,\mathrm{d}V(u)\,\mathrm{d}\tilde{x} \\
&= \frac{\tau - x}{\tau + x} (h_t(x) - B_V h_t(x)) \\
&\quad - \frac{\tau - x}{\tau + x} \cdot \frac{1}{\tau - x} c_V(x)\, \alpha_V(h_t) \\
&= \frac{\tau - x}{\tau + x} (h_t(x) - B_V h_t(x) - r_V(x)\, \alpha_V(h_t)),
\end{aligned}
$$

(compare this with (3.41)) where the operator B_V is defined as

$$
B_V h(x) \equiv \frac{2}{\tau - x} \int_{\tilde{x}=x}^{\tilde{x}=\tau} \frac{1}{g(\tilde{x})} \int_{u=\tilde{x}}^{u=\tau} h_t(u) \frac{1}{\tau + u} \,\mathrm{d}V(u)\,\mathrm{d}\tilde{x}
$$

and the function r_V on $[0, \tau)$ is defined as

$$
r_V(x) \equiv \frac{1}{\tau - x} c_V(x).
$$

Now we get for (3.168) the following equation

$$
(I - B_V) h_t(x) = r_V(x)\, \alpha_V(h_t) + 1_{(0,t]}(x).
$$

Similar to section 3.7.2 one proves that $(I - B_V)^{-1}$ exists as operator on $D[0, \tau)$. So we obtain

$$
h_t(x) = \bar{r}_V(x)\, \alpha_V(h_t) + \bar{y}_t, \tag{3.176}
$$

where $\bar{r}_V$ and $\bar{y}_t$ in $D[0, \tau)$ are defined as

$$
\bar{r}_V \equiv (I - B_V)^{-1} r_V \quad , \quad \bar{y}_t \equiv (I - B_V)^{-1} 1_{(0,t]}.
$$

Just as we did in section 3.7.2 we note that we can apply α_V and β_V (as operators on $D[0, \tau)$) on both sides of the equation (3.176) and so we get the next system of equations

$$
\begin{pmatrix} 1 - \alpha_V(\bar{r}_V) & 0 \\ -\beta_V(\bar{r}_V) & 1 \end{pmatrix} \cdot \begin{pmatrix} \alpha_V(h_t) \\ \beta_V(h_t) \end{pmatrix} = \begin{pmatrix} \alpha_V(\bar{y}_t) \\ \beta_V(\bar{y}_t) \end{pmatrix}.
$$

We see immediately that the operator B_V maps positive functions to positive functions and thus $(I - B_V)^{-1} = \sum_{k=0}^{\infty} B_V^k$ maps positive functions to positive functions. Because r_V is a positive function we have now that $\bar{r}_V$ is a positive function. Furthermore we note that the operator Ψ_V maps positive functions to positive numbers and thus $\alpha_V = -(1/D_V)\,\Psi_V$ maps positive functions to negative numbers (remember that $D_V > 0$). This means that $\alpha_V(\bar{r}_V)$ is a negative number. From this we may conclude that the determinant $Q_V = 1 - \alpha_V(\bar{r}_V)$ of the matrix

$$N_V \equiv \begin{pmatrix} 1-\alpha_V(\bar{r}_V) & 0 \\ -\beta_V(\bar{r}_V) & 1 \end{pmatrix}$$

is greater or equal to 1 and thus $Q_V \neq 0$. The inverse N_V^{-1} of N_V exists and so we obtain

$$\begin{pmatrix} \alpha_V(h_t) \\ \beta_V(h_t) \end{pmatrix} = \frac{1}{Q_V}\begin{pmatrix} 1 & 0 \\ \beta_V(\bar{r}_V) & 1-\alpha_V(\bar{r}_V) \end{pmatrix}\cdot\begin{pmatrix} \alpha_V(\bar{y}_t) \\ \beta_V(\bar{y}_t) \end{pmatrix} \equiv \begin{pmatrix} S_V(\bar{y}_t) \\ T_V(\bar{y}_t) \end{pmatrix}. \tag{3.177}$$

Now we have proved that h_t is uniquely determined on $[0,\tau)$ by equation (3.176):

$$h_t(x) = \bar{r}_V(x)\,S_V(\bar{y}_t) + \bar{y}_t(x).$$

For $x \geq \tau$ the solution h_t only has to satisfy (3.174). We get the following lemma (see lemma 3.7.2.1).

Lemma 3.12.2.1 *If we assume that $1/g(x)$ is bounded on $[0,\tau]$ and let h_t^0 and h_t^1 be defined as*

$$h_t^0(x) \equiv h_t(x)\cdot 1_{[0,\tau)}(x) \quad , \quad h_t^1(x) \equiv h_t(x)\cdot 1_{[\tau,\infty)}(x),$$

then

$$I_V(h_t)(x) = \kappa_t(x) \ \ \textit{for all } x \in [0,\infty)$$

is equivalent to

$$h_t^0(x) = (\bar{r}_V(x)\,S_V(\bar{y}_t) + \bar{y}_t(x))\cdot 1_{[0,\tau)}(x)$$

and h_t^1 satisfies

$$\begin{pmatrix} \alpha_V(h_t^1) \\ \beta_V(h_t^1) \end{pmatrix} = -\frac{1}{D_V}\begin{pmatrix} 1 & 0 \\ \tau & c_V(0) \end{pmatrix}\cdot\begin{pmatrix} \Psi_V(h_t^0) \\ 0 \end{pmatrix}.$$

3.12.3 One-dimensional case versus two-dimensional case

One of the reasons why the efficiency proof in the two-dimensional case is more difficult than in the one-dimensional case, is the role of the determinant Q_V. In the one-dimensional case we could prove $Q_V \neq 0$ using simple properties of the operator B_V and the functionals Ψ_V and α_V. The structure of the determinant in the two-dimensional case is much more complex. There we can only prove the existence of a hardest submodel, the existence of h_t, if we assume that $Q_V \neq 0$.

If one considers the equations (3.69) (two-dimensional case) and (3.176) (one-dimensional case), then one actually sees why the determinant in the two-dimensional case becomes more difficult to analyse. Because we have several kinds of double censored observations in the two-dimensional case, we get the contribution of $s_V(x)$ to the equation (3.69). In the one-dimensional case, because of (3.171) and (3.172) (the double censored observations all take the same value τ), this contribution drops out. An explanation could be the fact that in the one-dimensional case the NPMLE of the distribution function of the double censored observations (equals the probability of being double censored; the distribution function is degenerate at τ and equals the parameter h) is estimated by the fraction of the double censored observations and does not depend on the NPMLE $\widehat{V}_n$ of V on $[0,\tau)$.

Chapter 4

Open problems

In this chapter we briefly discuss some open problems. In section 4.1 we give a suggestion to deal with the singularities in τ and $2R$ to obtain the efficiency results in chapter 3 for a broader class of underlying distribution functions. In section 4.2 the position points follow a non-homogeneous Poisson point process and it seems that another approach is needed. In section 4.3 we say something about the convexity condition on the window W. In section 4.4 we state the conjecture that the determinant Q_V, which comes up in the analysis in chapter 3, is greater or equal to 1 and we base this on some computations of Q_V for different choices of V.

The two-dimensional case with K unknown is still open, but we think there is a possibility to use an extension of Van der Laan's identity. The distribution function K only plays a role through $E_K\text{diam}(W)$, which is a one-dimensional parameter, thus speaking rather informally the information calculations are only $\infty + 1$ dimensional instead of ∞ dimensional. One might then try to use the identity for the NPMLE in the convex and almost linear case (Van der Laan, 1994). We have for the distribution of the data $P_{F,K} = (1/(|W| + \mu E_K\text{diam}(W))) \cdot P_F$ and $F \to |W| + E_K\text{diam}(W)$ and $F \to P_F$ are linear, which describes this case.

4.1 Suggestion to solve the τ and $2R$ singularity

In chapter 2 in the one-dimensional case we have seen that at $\tau-$ we could not prove consistency and therefore the uniform consistency results there are related to intervals $[0, \tau - \epsilon]$ for each $\epsilon \in (0, \tau]$ instead of the interval $[0, \tau)$. In the two-dimensional circle-case we had the same problem in the point $2R$. Not having the consistency results in chapter 2 uniformly on interval $[0, \tau)$ (or $[0, 2R)$) implied that we obtained the asymptotic results in chapter 3 only under the assumptions in section 3.4. Although under these conditions we could prove nice properties of the NPMLE for a broad class of the underlying distribution functions, we still want to weaken or get rid of the assumptions. In this section we give a suggestion how to solve the τ and $2R$ singularity.

The idea is to regroup the data in such a way that all the observation points in $[\tau - \epsilon, \tau)$ ($\epsilon \in (0, \tau]$) which are s.e.c.l. and all the double censored observation points together are considered to be a new kind of observations. Furthermore all the observation points in $[\tau - \epsilon, \tau)$ which are s.e.c.r. or u.c. are regarded to be a new kind of observation. We define a new h_ϵ and show that the distribution of the data only depends on V on $[0, \tau - \epsilon)$ and the new parameter h_ϵ, For $\epsilon \downarrow 0$ we will have that (V, h_ϵ) converges to the original parameters (V, h). Thus for $\epsilon = 0$ the regrouping of the data is the original grouping of the data described in chapter 1. We apply all the techniques developed in the chapters 2 and 3 on the new structure to get results for the (sieved) NPMLE $(\widehat{V}_{n,\epsilon}, \widehat{h}_{n,\epsilon})$ of (V, h_ϵ) on the whole interval $[0, \tau - \epsilon)$, because by the regrouping of the data we are able to get consistency in $\tau - \epsilon$. If we concentrate for instance on $V(\tau-)$ (or $g(\tau)$), we will have

$$\begin{aligned} \lim_{n\to\infty} \widehat{V}_{n,\epsilon}((\tau - \epsilon)-) &\to V((\tau - \epsilon)-) \quad \text{for all } \epsilon \in (0, \tau] \\ \lim_{\epsilon\downarrow 0} V((\tau - \epsilon)-) &\to V(\tau-). \end{aligned}$$

A sequence $\epsilon(n)$ that converges to 0 in such a way that

$$\widehat{V}_{n,\epsilon(n)}((\tau - \epsilon(n))-) \to V(\tau-) \quad (n \to \infty),$$

(and thus $\widehat{g}_{n,\epsilon(n)}(\tau) \to g(\tau)$) provides us with an estimator $(\widehat{V}_{n,\epsilon(n)}, \widehat{h}_{n,\epsilon(n)})$ of $(V, h) \in \mathcal{V}_\tau$ for which we have the consistency results in chapter 2 uniformly on $[0, \tau)$ and the asymptotic results in chapter 3 without the assumptions (ii) and (iii) in section 3.4.

4.1.1 Regrouping in the one-dimensional case

Let $\epsilon \in (0, \tau]$ be given. We consider the set of points (T_i, X_i) in a random sample of size n on $\mathbb{R} \times \mathbb{R}^+$, where the X_i's are i.i.d. having the common distribution function V and the T_i's given $X_i = x_i$, are uniformly distributed on $(-x_i, \tau)$. (This is the same set up as in section 1.1.2). Now the data will be grouped differently. For the $T_i = t_i \leq 0$ we observe pairs (Z_i, D_i), where

$$Z_i = \min(T_i + X_i, \tau - \epsilon), \quad D_i = \begin{cases} 1 & T_i + X_i \leq \tau - \epsilon \\ 2 & T_i + X_i > \tau - \epsilon \end{cases}$$

and for the $T_i = t_i \in (0, \tau)$ we observe pairs (Z_i, D_i), where

$$Z_i = \min(X_i, \tau - T_i, \tau - \epsilon), \quad D_i = \begin{cases} 0 & T_i + X_i \leq \tau,\ X_i \leq \tau - \epsilon \\ 1 & T_i + X_i > \tau,\ T_i > \epsilon \\ 3 & X_i > \tau - \epsilon,\ T_i \leq \epsilon. \end{cases}$$

If $D_i = 0$ we call the observation u.c. If $D_i = 1$ we call the observation s.e.c., if $D_i = 2$ d.c. and if $D_i = 3$ we call the observation 'new censored' (n.c.). We note that this grouping of the data can be obtained from the grouping in the sections 1.1.1 and 1.1.2 by labeling all the u.c. and s.e.c.r. observation points in $[\tau - \epsilon, \tau)$ with 3 and the s.e.c.l. observations in $[\tau - \epsilon, \tau)$ and the d.c. with 2 and the s.e.c. observation points in $[0, \tau - \epsilon)$ with 1 and the u.c. observation points in $[0, \tau - \epsilon)$ with 0. Instead of working with the interval $[0, \tau)$ we now deal with the interval $[0, \tau - \epsilon)$. In Figure 4.1 we see in what areas the (T_i, X_i)'s fall, which belong to a d.c. or n.c. observation.

We define h_ϵ and k_ϵ as follows:

$$h_\epsilon \equiv h + \int_0^\epsilon g(\tau - x)\,\mathrm{d}x$$

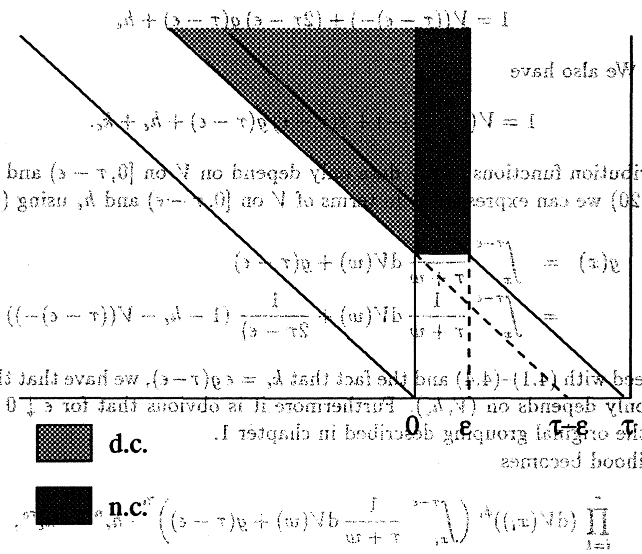

Figure 4.1: Regrouping the data.

and

$$k_\epsilon \equiv \int_{\tau-\epsilon}^{\tau} \frac{\tau - x}{\tau + x}\, \mathrm{d}V(x) + \int_0^\epsilon g(\tau - x)\, \mathrm{d}x = \epsilon\, g(\tau - \epsilon).$$

For the subdistribution functions we obtain ($u \in [0, \tau - \epsilon)$)

$$\begin{aligned} \mathrm{d}F_\epsilon^{s.e.c.}(u) &= 1_{[0,\tau-\epsilon)}(u) \cdot g(u)\, \mathrm{d}u & (4.1)\\ \mathrm{d}F_\epsilon^{d.c.}(u) &= h_\epsilon\, \mathrm{d}\delta_{\tau-\epsilon}(u) & (4.2)\\ \mathrm{d}F_\epsilon^{u.c.}(u) &= 1_{[0,\tau-\epsilon)}(u) \frac{\tau - u}{\tau + u}\, \mathrm{d}V(u), & (4.3)\\ \mathrm{d}F_\epsilon^{n.c.}(u) &= k_\epsilon\, \mathrm{d}\delta_{\tau-\epsilon}(u) & (4.4) \end{aligned}$$

where g is defined as in (1.8). Here we have the equality

$$1 = V((\tau - \epsilon)-) + (2\tau - \epsilon)\, g(\tau - \epsilon) + h_\epsilon \tag{4.5}$$

(see (1.14)). We also have

$$1 = V((\tau - \epsilon)-) + 2(\tau - \epsilon)\, g(\tau - \epsilon) + h_\epsilon + k_\epsilon. \tag{4.6}$$

The subdistribution functions of the data only depend on V on $[0, \tau - \epsilon)$ and h_ϵ. Of course, just as in (1.20) we can express $g(x)$ in terms of V on $[0, \tau - \epsilon)$ and h_ϵ using (4.5):

$$\begin{aligned} g(x) &= \int_x^{\tau-\epsilon} \frac{1}{\tau + w}\, \mathrm{d}V(w) + g(\tau - \epsilon)\\ &= \int_x^{\tau-\epsilon} \frac{1}{\tau + w}\, \mathrm{d}V(w) + \frac{1}{2\tau - \epsilon)} \left(1 - h_\epsilon - V((\tau - \epsilon)-)\right) \end{aligned}$$

and thus indeed with (4.1)–(4.4) and the fact that $k_\epsilon = \epsilon\, g(\tau - \epsilon)$, we have that the distribution of the data only depends on (V, h_ϵ). Furthermore it is obvious that for $\epsilon \downarrow 0$ we obtain the model with the original grouping described in chapter 1.

The likelihood becomes

$$\prod_{i=1}^{r} (\mathrm{d}V(x_i))^{\phi_i} \left(\int_{x_i}^{\tau-\epsilon} \frac{1}{\tau + w}\, \mathrm{d}V(w) + g(\tau - \epsilon)\right)^{\gamma_i} \cdot h_\epsilon^{\,n-r} \cdot k_\epsilon^{\,r_0},$$

where the $x_1 < x_2 < \ldots < x_r$ are the observation points (in $[0, \tau - \epsilon)$) and ϕ_i and γ_i are the number of u.c. and s.e.c. observations respectively at x_i and r_0 is the number of $n.c.$ observation points. (see (1.35)).

By $(\widehat{V}_{n,\epsilon}, \widehat{h}_{n,\epsilon})$ (and $\widehat{g}_{n,\epsilon}$ and $\widehat{k}_{n,\epsilon}$) we denote the (sieved) NPMLE of (V, h_ϵ) (and g and k_ϵ). In this case $(\widehat{V}_{n,\epsilon}, \widehat{h}_{n,\epsilon})$ satisfies the self-consistency equations: ($x \in [0, \tau - \epsilon)$)

$$\begin{aligned} \mathrm{d}\widehat{V}_{n,\epsilon}(x) &= \mathrm{d}F_{n,\epsilon}^{u.c.}(x) + \int_{v=0}^{v=x} \frac{1}{\widehat{g}_{n,\epsilon}(v)}\, \mathrm{d}F_{n,\epsilon}^{s.e.c.}(v) \cdot \frac{1}{\tau + x}\, \mathrm{d}\widehat{V}_{n,\epsilon}(x)\\ &= \mathrm{d}F_{n,\epsilon}^{u.c.}(x)\\ &\quad + \int_{v=0}^{v=x} \left(\int_{w=v}^{w=\tau-\epsilon} \frac{1}{\tau + w}\, \mathrm{d}\widehat{V}_{n,\epsilon}(w) + \widehat{g}_{n,\epsilon}(\tau - \epsilon)\right)^{-1} \mathrm{d}F_{n,\epsilon}^{s.e.c.}(v) \cdot \frac{1}{\tau + x}\, \mathrm{d}\widehat{V}_{n,\epsilon}(x) \end{aligned}$$

and

$$\begin{aligned}\widehat{h}_{n,\epsilon} &= F^{d.c.}_{n,\epsilon}(\tau-\epsilon)\\ (2\tau-\epsilon)\widehat{g}_{n,\epsilon}(\tau-\epsilon) &= F^{n.c.}_{n,\epsilon}(\tau-\epsilon)+\int_0^{\tau-\epsilon}\frac{\widehat{g}_{n,\epsilon}(\tau-\epsilon)}{\widehat{g}_{n,\epsilon}(v)}\,\mathrm{d}F^{s.e.c.}_{n,\epsilon}(v)\\ &= 1-\widehat{h}_{n,\epsilon}-\widehat{V}_{n,\epsilon}(\tau-\epsilon),\end{aligned}$$

where $F^{\cdots}_{n,\epsilon}$ stand for the empirical distribution function of $F^{\cdots}_{\epsilon}$.

We hope the reader agrees with the fact that all consistency proofs in section 2.2 can be imitated for the (sieved) NPMLE $(\widehat{V}_{n,\epsilon},\widehat{h}_{n,\epsilon})$ of (V,h_ϵ). Actually it is nothing more than replacing all the $\widehat{V}_n$'s by $\widehat{V}_{n,\epsilon}$ (all the $\widehat{g}_n$'s by ... etc.) and all the h's by h_ϵ and at several places the τ's by $\tau-\epsilon$ and in this case we have to deal with the k_ϵ. We explain now that we get consistency uniformly on the whole interval $[0,\tau-\epsilon)$. For $\widehat{V}_n$ we could not prove that $\widehat{V}_n(\tau-)\to V(\tau-)$, but for $\widehat{V}_{n,\epsilon}$ we can prove $\widehat{V}_{n,\epsilon}((\tau-\epsilon)-)\to V((\tau-\epsilon)-)$. If we write (2.2) (compare with (2.17)) for this model we find

$$\int_0^{\tau-\epsilon}\frac{\mathrm{d}V_n}{\mathrm{d}\widehat{V}_{n,\epsilon}}(x)\,\mathrm{d}F^{u.c.}_{n,\epsilon}(x)+\int_0^{\tau-\epsilon}\frac{g_n(x)}{\widehat{g}_{n,\epsilon}(x)}\,\mathrm{d}F^{s.e.c.}_{n,\epsilon}(x)+\widehat{h}_{n,\epsilon}+\frac{g_n(\tau-\epsilon)}{\widehat{g}_{n,\epsilon}(\tau-\epsilon)}F^{n.c.}_{n,\epsilon}(\tau-\epsilon)\le 1.$$

Again we note that the integrands are nonnegative and thus we also have

$$\frac{g_n(\tau-\epsilon)}{\widehat{g}_{n,\epsilon}(\tau-\epsilon)}F^{n.c.}_{n,\epsilon}(\tau-\epsilon)\le 1$$

and this implies that $\widehat{g}_{n,\epsilon}(\tau-\epsilon)>0$ (for n large enough). Now one uses (2.21) (or (2.30)) to conclude the consistency of $\widehat{V}_{n,\epsilon}(\tau-\epsilon)$ (and thus of $\widehat{g}_{n,\epsilon}(\tau-\epsilon)$).

To imitate the calculations and derivations in chapter 3, we only need the assumption that $1/g(\tau-\epsilon)$ is bounded. Again one only has to change some notation and replace several τ's by $\tau-\epsilon$'s. Together with the consistency uniformly on $[0,\tau-\epsilon)$ we obtain similar efficiency results for the parameters

$$\vartheta_1(P_{(V,h_\epsilon)})=\int_{x=0}^{x=\cdot}\frac{\tau-x}{\tau+x}\,\mathrm{d}V(x)\equiv W_\epsilon(\cdot),\quad \vartheta_2(P_{(V,h_\epsilon)})=h_\epsilon$$

(see section 3.12.1). Here we define $b_t\vartheta_1(P_{(V,h_\epsilon)})=W_\epsilon(t)$, where $B_\epsilon=\{b_t : t\in[0,\tau-\epsilon)\}$. The set $B=B_0$ will be defined as $B=\{b_t : t\in[0,\tau)\}$.

Except the condition on $g(\tau-\epsilon)$ we can drop the conditions on $\widehat{V}_n(\tau-)$ and $\widehat{g}_n(\tau)$, thus here on $\widehat{V}_{n,\epsilon}((\tau-\epsilon)-)$ and $\widehat{g}_{n,\epsilon}(\tau-\epsilon)$, because of the consistency uniformly on $[0,\tau-\epsilon)$. Of course we are only interested in small 'ϵ-models', thus it will be enough to assume that for the underlying g we have

$$g(\tau-\epsilon)>0 \text{ for all } \epsilon\in(0,a) \qquad (4.7)$$

for some a and for all these 'ϵ-models' we have the above story.

Now for a fixed ϵ we have efficiency for the (sieved) NPMLE

$$\widehat{W}_\epsilon(t)=b_t\vartheta_1(P_{(\widehat{V}_{n,\epsilon},\widehat{h}_{n,\epsilon})})$$

(uniformly on $[0, \tau - \epsilon)$) and for the (sieved) NPMLE $\hat{h}_{n,\epsilon}$. This implies (using theorem 3.1.1 or section 3.5.3) that $\hat{V}_{n,\epsilon}$ is a $\|\cdot\|_{B_\epsilon}$-asymptotically efficient estimator of V. Of course here efficient means efficient among the estimators which only make use of the information on $[0, \tau - \epsilon)$, thus efficient among the estimators, for which the regrouped data would be the 'original' data. Actually, here regrouping of the data means throwing away information, because we ignore the fact that in the original data we have u.c. and s.e.c. observation points in $[\tau - \epsilon, \tau)$. So $\hat{V}_{n,\epsilon}$ will not be efficient (for V on $[0, \tau - \epsilon)$) among the estimators which make use of all the data.

To obtain from $\widehat{W}_\epsilon(t)$ on $[0, \tau - \epsilon)$ a statement for $W(t)$ ($= W_0(t)$; see section 3.12.1) on $[0, \tau)$, we only need proposition 4.1.1.1, where we show that there exists a sequence $\epsilon(n)$ such that $\hat{V}_{n,\epsilon(n)}((\tau - \epsilon(n))-) \to V(\tau-)$. From now on we write $\hat{V}_{n,\epsilon(n)}(\tau - \epsilon(n))$ instead of $\hat{V}_{n,\epsilon(n)}((\tau - \epsilon(n))-)$.

Proposition 4.1.1.1 *There exists a sequence $\epsilon(n) \to 0$ such that in probability we have*

$$\hat{V}_{n,\epsilon(n)}((\tau - \epsilon(n))-) \to V(\tau-).$$

PROOF: for convenience we assume that the underlying V is continuous and strictly increasing on $[0, \tau)$. With more effort one proves the statement for the general case. We remember the V_n in section 2.2, for which we had consistency uniformly on $[0, \tau)$: $dV_n \equiv (\tau + x)/(\tau - x)\, dF_n^{u.c.}(x)$.

For all k there exists a $x_k \in (0, \tau]$ such that

$$0 < V_k(\tau-) - V(\tau - x_k) \leq \frac{1}{k}. \tag{4.8}$$

This implies that

$$x_k \to 0 \quad (k \to \infty). \tag{4.9}$$

(We show this later). Let $z_k \in (0, \tau]$ be such that $V(\tau - z_k) = V_k(\tau-)$ (exists because of continuity V). Because of (4.8) and V being strictly monotone increasing, there exists a $y_k \in (x_k, z_k)$ such that

$$V(\tau - x_k) < V(\tau - y_k) < V(\tau - z_k) = V_k(\tau-). \tag{4.10}$$

This easily implies that

$$0 < V_k(\tau-) - V(\tau - y_k) \leq \frac{1}{k}.$$

It follows (in the same way that (4.9) followed from (4.8)) that

$$y_k \to 0 \quad (k \to \infty).$$

Now because we (4.10) and

$$\lim_{m \to \infty} \hat{V}_{m,y_k}(\tau - y_k) = V(\tau - y_k), \tag{4.11}$$

there exists a $m(k) > m(k-1)$ such that

$$V(\tau - x_k) \leq \hat{V}_{m(k),y_k}(\tau - y_k) \leq V(\tau - z_k) = V_k(\tau-). \tag{4.12}$$

Of course by (4.11) we have for $i \geq 0$

$$V(\tau - x_k) \leq \widehat{V}_{m(k)+i,y_k}(\tau - y_k) \leq V_k(\tau-). \tag{4.13}$$

So we obtain the sequence $(m(k), y_k)_{k\geq 1}$, for which $m(1) < m(2) < m(3) < \dots$ and $y_k \to 0$. Now for $n = m(k)$ to $n = m(k+1) - 1$, we define $\epsilon(n) = \epsilon(m(k) + i) = y_k$. So we get the sequence

$$\widehat{V}_{m(1),y_1}(\tau - y_1),\ \widehat{V}_{m(1)+1,y_1}(\tau - y_1),\ \dots,\ \widehat{V}_{m(2)-1,y_1}(\tau - y_1),\ \widehat{V}_{m(2),y_2}(\tau - y_2),\ \widehat{V}_{m(2)+1,y_2}(\tau - y_2),\ \dots .$$

For this sequence we have (4.12) and (4.13) and thus immediately we get that the sequence

$$\widehat{V}_{n,\epsilon(n)}(\tau - \epsilon(n)) = \widehat{V}_{m(k)+i,y_k}(\tau - y_k)$$

lies between $V(\tau - x_k)$ and $V_k(\tau-)$, which both tend to $V(\tau-)$. This proves the proposition.

The only thing we still have to prove is (4.9). Suppose (4.9) is not true, then there is a subsequence x_{k_l} for which $x_{k_l} \to \alpha \neq \tau$ $(l \to \infty)$. Because V is strict monotone we have that $V(\tau-) - V(\tau - \alpha) > 0$. We also know that $V_k(\tau-) \to V(\tau-)$, thus from (4.8) and the continuity of V we obtain that

$$0 = \lim_{l\to\infty} (V_{k_l}(\tau-) - V(\tau - x_{k_l})) = V(\tau-) - V(\tau - \alpha).$$

We have a contradiction and thus (4.9) must be true. □

So we immediately obtain

Proposition 4.1.1.2 *Under the assumption (4.7)*

$$\widehat{W}_{\epsilon(n)} = b_t \vartheta_1 (P_{(\widehat{V}_{n,\epsilon(n)}, \widehat{h}_{n,\epsilon(n)})})$$

is a $\|\cdot\|_B$-asymptotically efficient estimator of W and $\widehat{h}_{n,\epsilon(n)}$ is a asymptotically efficient estimator of h.

Using theorem 3.1.1, this yields

Proposition 4.1.1.3 *Under the assumption (4.7) we have for each $\eta \in (0, \tau]$ that $\widehat{V}_{n,\epsilon(n)}$ is a $\|\cdot\|_\infty$-asymptotically efficient estimator of $V \in D[0, \tau - \eta]$.*

In practice one wants to have a method to construct a suitable sequence $\epsilon(n)$. A suggestion would be as follows. We consider the ad hoc estimator $\widehat{g}_{0,n}(\tau)$ of $g(\tau)$:

$$\widehat{g}_{0,n}(\tau) = \frac{F_n^{s.e.c.}(\tau) - F_n^{s.e.c.}(\tau - n^{-(1/3)})}{2\,n^{-(1/3)}},$$

which has a squared bias and variance of order $n^{-(2/3)}$. Now we choose $\epsilon(n)$ to be the smallest ϵ such that

$$\widehat{g}_{n,\epsilon}(\tau - \epsilon) > \widehat{g}_{0,n}(\tau) - n^{-(1/4)}.$$

In this way we are sure that for n large enough that $\widehat{g}_{n,\epsilon}(\tau - \epsilon) > 0$ because $\widehat{g}_{0,n}(\tau) - n^{-(1/4)} \to g(\tau)$ (in probability) (and $g(\tau) > 0$) and we are sure that such an $\epsilon(n)$ exists because we have (using Chebyshov inequality) that $P(\widehat{g}_{0,n}(\tau) - n^{-(1/4)} < g(\tau)) \to 1$. Thus we ensure that our ad hoc estimator is between 0 and $g(\tau)$. We think that in this way we have $\epsilon(n)$ converging to 0 and that the information $I(\epsilon)$ (depending on ϵ in the 'regrouped' model) will converge to the information $I(0)$ of the original model.

4.1.2 Regrouping in the two-dimensional 'circle-case'

For the two-dimensional 'circle-case' we only give the regrouping. What we did in section 4.1.1 can be similarly done for this case. Here the regrouping of the data is somehow the same as in the one-dimensional case. In Figure 1.11 we have drawn the set $\mathcal{A}_{\theta,x}$. If we take there $u = 2R - \epsilon$, than for $x > 2R$ and for $2R - \epsilon \leq x \leq 2R$ the regrouping is done by calling the observations in the white areas of the circles on the left double censored (d.c.) and on the right new censored (n.c.).

Just as in the one-dimensional regrouping, we got rid of the 'sharp' edge in the subset in A which belongs to the u.c. observations. Also in this case one shows that the distribution of the data only depends on V on $[0, 2R-\epsilon)$ and h_ϵ. Note that we do not have to regroup within the d.c. observations, because for $d(x,x)$ (density of the d.c.) we know that it is consistent uniformly on $[0, 2R)$.

4.2 Non-homogeneous Poisson point process

In the line segment processes in the previous chapters the points T_i in the one-dimensional case and the position points $\vec{T_i}$ in the two-dimensional problem, follow a homogeneous Poisson point process on respectively $\mathbb{R}$ and $\mathbb{R} \times \mathbb{R}$ with rate λ. We have seen that the number of observed line segments N has a Poisson distribution with parameter respectively $\lambda\,(\tau + \mu)$ and $\lambda\,(|W| + \mu\, E_K \text{diam}(W))$. We were not concerned about the rate λ, though its value was unknown. Fortunately by conditioning on $N = n$, we got rid of the λ, because in the probability measure it appeared as factor in the nominator and denominator and cancelled out. From that moment the (un)known rate λ did not play any role in the analysis.

What will happen if the position points follow a non-homogeneous Poisson point process with intensity measure $\rho_0(\mathrm{d}\vec{t}) = \lambda(\vec{t})\,\mathrm{d}\vec{t}$? (For convenience we assume that the intensity measure has a density w.r.t. the Lebesgue measure). Unfortunately in this case we do not get rid of $\lambda(\vec{t})$ by conditioning on $N = n$. In the nominator and denominator of the probability measure (conditioning on $N = n$) the $\lambda(\vec{t})$ does not appear as a factor that cancels out. In our search for an estimator of V (or F) we have to find an estimator for $\lambda(\cdot)$ as well.

Let us consider the one-dimensional line segment process described in section 1.1, but instead of the homogeneous Poisson point process with rate λ we assume that the position points T_i follow a non-homogeneous Poisson point process with intensity measure $\rho_0(\mathrm{d}t) = \lambda(t)\,\mathrm{d}t$. Just as in section 1.1.1 we note that the points $(T_i, X_i) \in \mathbb{R} \times \mathbb{R}^+$ follow a Poisson point process on $\mathbb{R} \times \mathbb{R}^+$ with intensity measure

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

In this case the number of observed line segments N has a Poisson distribution with parameter

$$S \equiv \int_A \lambda(t)\,\mathrm{d}t\,\mathrm{d}F(x) = \int_{x=0}^{x=\infty} \int_{t=-x}^{t=\tau} \lambda(t)\,\mathrm{d}t\,\mathrm{d}F(x)$$

(the set A is defined as in section 1.1.1). We define $S(\cdot)$ on $[-\tau, \infty)$ as

$$S(x) \equiv \int_{t=-x}^{t=\tau} \lambda(t)\,\mathrm{d}t.$$

The lengths of the observed line segments are distributed according to

$$\mathrm{d}V(x) \equiv \frac{S(x)}{S}\,\mathrm{d}F(x)$$

and the position points T_i given $X_i = x$ are distributed according to

$$\mathrm{d}A(t|X = x) \equiv 1_{(-x,\tau)}(t) \cdot \frac{\lambda(t)\,\mathrm{d}t}{S(x)}.$$

If we condition on $N = n$, then the set of points (T_i, X_i), that belong to observed line segments, are distributed as the set of points in an i.i.d. sample of size n with probability measure on $\mathbb{R} \times \mathbb{R}^+$

$$1_A(t,x) \cdot \frac{\lambda(t)\,\mathrm{d}t\,\mathrm{d}F(x)}{\int_A \lambda(t)\,\mathrm{d}t\,\mathrm{d}F(x)} = \mathrm{d}V(x)\,\mathrm{d}A(t|X = x).$$

If we write down the subdistribution function of the data, then we find for $0 \leq u \leq \tau$

$$\begin{aligned}
\mathrm{d}F^{s.e.c.l.}(u) &= \int_{x=u}^{x=\infty} \lambda(u-x)\,\frac{1}{S(x)}\,\mathrm{d}V(x)\,\mathrm{d}u \\
\mathrm{d}F^{d.c.}(u) &= \delta_\tau(u) \int_{x=\tau}^{x=\infty} \int_{t=-x+\tau}^{t=0} \lambda(t)\,\mathrm{d}t\,\frac{1}{S(x)}\,\mathrm{d}V(x) \\
\mathrm{d}F^{s.e.c.r.}(u) &= \lambda(\tau-u) \int_{x=u}^{x=\infty} \frac{1}{S(x)}\,\mathrm{d}V(x)\,\mathrm{d}u \\
\mathrm{d}F^{u.c.}(u) &= \int_{t=0}^{t=\tau-u} \lambda(t)\,\mathrm{d}t \cdot \frac{1}{S(u)}\,\mathrm{d}V(u)
\end{aligned}$$

(compare with (1.4) - (1.7)).

Let us define h similar to the definition of h in section 1.1.1:

$$h \equiv \int_{x=\tau}^{x=\infty} \int_{t=-x+\tau}^{t=0} \lambda(t)\,\mathrm{d}t\,\frac{1}{S(x)}\,\mathrm{d}V(x)$$

(the probability of being double censored). We can estimate $\lambda(t)$ only on $(0,\tau)$. Thus the $\lambda(u-x)$ in the subdistribution function of $F^{s.e.c.l.}$ can not be estimated, but the $\lambda(\tau-u)$ in $F^{s.e.c.r.}$ can. Because $\lambda(t)$ can only be estimated on $(0,\tau)$, we are not able to estimate $S(x)$ for $x \in (0,\infty)$ and thus how do we get $F(x)$ back with $\mathrm{d}F(x) = (S/S(x))\,\mathrm{d}V(x)$? The approach we used in the homogeneous case seems not to work.

In the next section we will see that if we restrict the class of possible $\lambda(t)$, we get a familiar situation.

4.2.1 $\lambda(t)$ is unknown if $t \geq 0$ and 0 otherwise

If we assume that $\lambda(t)$ is unknown if $t \geq 0$ and 0 otherwise, then we actually say that there are no s.e.c.l. and d.c. observations in the line segment process. In terms of the hospital model interpretation this would mean that in some time interval $(0,\tau)$ we set up an experiment and take only into account the patients, which arrive during the time interval $(0,\tau)$ and observe their sojourn time only during the time interval. We obtain only u.c. and s.e.c.r. observations.

We will see that in this case the rate at which patients arrive at the hospital, plays no role in estimating F on $[0, \tau)$. We have

$$S = \int_{x=0}^{x=\infty} \int_{t=0}^{t=\tau} \lambda(t)\,\mathrm{d}t\,\mathrm{d}F(x) = \int_{t=0}^{t=\tau} \lambda(t)\,\mathrm{d}t = S(x)$$

for all $x \geq 0$. So we have that the observed line segments are distributed according to $\mathrm{d}V(x) \equiv (S(x)/S)\,\mathrm{d}F(x) = \mathrm{d}F(x)$ and the arrival times T_i given $X_i = x$ are distributed according to $\mathrm{d}A(t|X = x) \equiv (\lambda(t)\,\mathrm{d}t/S) \cdot 1_{(0,\tau)}(t)$. Conditioning on $N = n$ we write down the subdistribution functions of the data $(0 \leq u \leq \tau)$:

$$\begin{aligned}
\mathrm{d}F^{s.c.c.r.}(u) &= \lambda(\tau - u) \int_{x=u}^{x=\infty} \frac{1}{S}\,\mathrm{d}F(x)\,\mathrm{d}u \\
\mathrm{d}F^{u.c.}(u) &= \frac{1}{S} \int_{t=0}^{t=\tau-u} \lambda(t)\,\mathrm{d}t \cdot \mathrm{d}F(u).
\end{aligned}$$

We know that in this case S does not depend on F. If we drop out all factors depending on $S(\cdot)$ and S, then we obtain the proportional likelihood

$$\prod_{i=1}^{r} (\mathrm{d}F(x_i))^{\phi_i} \cdot ((1 - F(x_i)))^{\gamma_i},$$

which corresponds with the Kaplan-Meier situation.

4.3 Non-convexity of the window W

In the two-dimensional case we always assumed that the observation window **W** was convex. This ensured us that two censored line segments hitting the edge of the window did not belong to the same underlying line segment. In Figure 4.2 we observe two censored line segments in the non-convex window **W**, that belong to the same underlying line segment. One can not regard the two line segments inside the window as two independent observations. One does not know whether the observed line segments belong to the same underlying line segment or not. On the other hand, in the case that the position points $\vec{T_i}$ follow a homogeneous Poisson point process on $\mathbb{R}^2$ with rate λ, the probability that two points of the underlying Poisson point process on $\mathbb{R}^2 \times \mathbb{R}^+ \times [0, \pi)$, say $(\vec{T_1}, X_1, \Theta_1) = (\vec{t_1}, x_1, \theta_1)$ and $(\vec{T_2}, X_2, \Theta_2) = (\vec{t_2}, x_2, \theta_2)$, are such that the angle of the line through the position points $\vec{T_1}$ and $\vec{T_2}$ equals $\theta_1 = \theta_2$ is 0. This means that with probability 1 in a non-convex window **W** we may say that two censored line segments, which lie on the same line through these segments, belong to the same line segment. It would be nice if we could add the observed lengths together and regard this as one observation to obtain eventually a sample of independent observations as in chapter 1 and then apply the theory we developed in the previous chapters.

Unfortunately this will not work in general. For instance it is essential that the self-consistency equations for $\hat{V}_n$ and the empirical distribution of the data obtained in this way, are also satisfied by V and the distribution of the data (which depends heavily on the shape of the window) and this is certainly not clear.

An idea is to split the non-convex window into convex pieces and each piece can be treated as in the previous chapters. Bot now the data sets belonging to the convex pieces are not independent, because a line segment can hit more pieces, therefore the proofs of the asymptotic results break down, but the results maybe not.

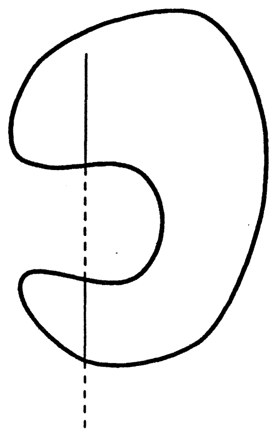

Figure 4.2: window **W** is not convex.

4.4 A conjecture: the determinant $Q_V \geq 1$

In section 3.10 we already made some remarks about the determinant Q_V. There we showed in the two-dimensional circle-case that the statement $Q_V \neq 0$ is not an empty statement. We need $Q_V \neq 0$ in the proof for the existence of a hardest submodel h_t in lemma 3.7.2.1. In the one-dimensional case one is able to show that $Q_V = 1 - \alpha_V(\bar{r}_V) \geq 1$, using simple properties of the operator B_V and the functionals Ψ_V and α_V (see section 3.12.2).

We must admit that the assumption $Q_V \neq 0$ in the two-dimensional circle-case for proving efficiency, is not satisfying. Compared with the one-dimensional case the structure of the determinant in the two-dimensional case is much more complex. In the one-dimensional case the determinant equals

$$Q_V = 1 - \alpha_V(\bar{r}_V),$$

where α_V and $\bar{r}_V$ are defined as in section 3.12.2. In the two-dimensional case the determinant equals

$$Q_V = 1 - \alpha_V(\bar{r}_V) - \beta_V(\bar{s}_V) + \alpha_V(\bar{r}_V)\,\beta_V(\bar{s}_V) - \alpha_V(\bar{s}_V)\,\beta_V(\bar{r}_V),$$

where α_V, β_V, $\bar{r}_V$ and $\bar{s}_V$ are defined as in section 3.7.2. In both cases we see immediately that if V puts all mass outside the interval $[0, \tau)$ respectively $[0, 2R)$, then the determinant equals 1. In the one-dimensional case we proved that $-\alpha_V(\bar{r}_V) \geq 0$ and thus the determinant is greater or equal to 1 for all V (satisfying: $1/g(x)$ is bounded on $[0, \tau]$). We think that in the two-dimensional case $-\alpha_V(\bar{r}_V) - \beta_V(\bar{s}_V) + \alpha_V(\bar{r}_V)\,\beta_V(\bar{s}_V) - \alpha_V(\bar{s}_V)\,\beta_V(\bar{r}_V) \geq 0$. Again this would mean that $Q_V \geq 1$ for all V (satisfying: $1/g(x)$ and $1/d(x, x)$ are bounded on $[0, 2R]$). The conjecture is based on some computations of Q_V for different choices of V, because a nice proof as in the one-dimensional case is not available yet and seems hard to find.

Suppose $R = 1$ and V puts mass on $x_1 = \sqrt{(2)}$ and $x_2 = 2\sqrt{(2)}$ only; thus on one point in the interval $[0, 2R]$ and on one point outside the interval. Now the calculations can be done by hand and one finds for the case $P(X = x_1) = 1/2$ and $P(X = x_2) = 1/2$ that $Q_V \approx 2.940329$. For the case $P(X = x_1) = 1/3$ and $P(X = x_2) = 2/3$ one finds $Q_V \approx 2.089088$. Decreasing the mass at x_1 to 0 and increasing the mass at x_2 to 1 one finds a sequence of Q_V's decreasing from above to 1.

If we take $R = 1.4$ and

$$P(X = k) = \frac{l^{k-1}}{(k-1)!} \exp(-l) \ , \quad k = 1, 2, 3, \ldots$$

(Poisson(l) on $1, 2, 3, \ldots$), then we find using the computer

l	$= 0.5$	2	2.1	5	10	15
Q_V	≈ 7.688234	2.335098	2.221792	1.125199	1.002395	1.000033.

Of course the bigger l is, the more mass of V is placed after $2R$ and thus the determinant tends to 1 (from above) if l tends to infinity. In the following tabel one sees what happens if we differ R and fix $l = 2$:

R	$= 1.01$	1.44	1.48	1.51	1.52	2.01	5.01	20.01
Q_V	≈ 3.08	2.28	2.23	6.95	6.90	17.07	1080.87	22868.13.

We also checked some continuous distribution functions V, for instance the 'Cauchy' distribution on $(0, \infty)$ with density

$$v(x) = \frac{2}{\pi\,(1 + x^2)}.$$

In this case if we take $R = 1.5$, then we find that $Q_V \geq 5.5$ and thus greater than 1. For the exponential distribution $v(x) = \lambda \exp(-\lambda\, x)$ we find for $R = 1.5$: $Q_V \geq 16$ if $\lambda = 1$, $Q_V \geq 1.9$ if $\lambda = 5$ and $Q_V \geq 1.3$ if $\lambda = 10$.

One checks that in the computations for the determinant we need to calculate certain integrals and for instance $(I - B_V)^{-1}$. To implement this in a computer program we have to discretize the problem. We will not bother the reader with the difficulties and features of the computer program we used. For the discrete distribution functions V we could approximate the determinant quite well, because we did not need to discretize. In the continuous cases we had to discretize the problem and of course the finer the grid was, the more accurate was the outcome and the more computer-time was needed. Instead of calculating the determinant as exactly as possible, we were satisfied with an underbound. Here we state the following conjecture:

Conjecture 4.4.1 *The determinant Q_V is equal or greater than 1 for all permitted V.*

Bibliography

[1] Andersen P.K., Borgan O., Gill R.D. and Keiding N. (1993), *Statistical Models based on Counting Processes*, Springer, New York.

[2] Bakry D., Gill R.D. and Molchanov S.A. (1994), *Lectures on Probability Theory*, Springer-Verlag, Berlin Heidelberg.

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

[4] Billingsley P. (1979), *Probability and Measure*, John Wiley and Sons, New York.

[5] Brookmeyer R., Gail M.H. and Polk B.F. (1987a), The prevalent cohort study and the acquired immunodeficiency syndrome, *Americ. J. Epidemiology* **126**, 14–24.

[6] Brookmeyer R. and Gail M.H. (1987b), Biases in prevalent cohorts, *Biometrics* **43**, 739–749.

[7] Chung C.F. (1989), Confidence bands for percentile residual lifetime under random censorship model, *J. Multiv. Anal.*, **29**, 94–126.

[8] Chung C.F. (1990), Confidence bands for quantile function under random censorship, *Ann. Inst. Statist. Math.* **42**, 21–36.

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

[10] Dudley R.M. (1984), A course on empirical processes, *Lecture Notes in Mathematics* 1097, 2–141, Springer-Verlag, New York.

[11] Dudley R.M. (1985), An extended Wichura theorem, definition of Donsker class, and weighted empirical distributions, *Lecture Notes in Mathematics* 1153, 141–178, Springer-Verlag, New York.

[12] Dudley R.M. (1987), Universal Donsker classes and metric entropy, *Ann. Probability* **15**, 1306–1326.

[13] Efron B. (1967), The two sample problem with censored data, *Proc. 5th Berkeley Symp. on Math. Statist. Prob.* **4**, 831–853.

[14] Gill R.D. (1989), Non- and Semi-parametric Maximum Likelihood Estimators and the Von Mises Method (Part I), *Scand. J. Statist.* **16**, 97–128.

[15] Gill R.D., Van der Laan, Wijers B.J., Laslett's line segment problem, *Bernoulli*, to appear.

[16] Giné E. and Zinn J. (1986), Lectures on the central limit theorem for empirical processes, *Lect. Notes in Math.* 1221, 50–113, Springer-Verlag, Berlin.

[17] Griffel D.H. (1981), *Applied Functional Analysis*, Ellis Horwood Series, Chichester, England.

[18] Groeneboom P. (1991), *Nonparametric maximum likelihood estimators for interval censoring and deconvolution*, Technical report no. 378, Department of Statistics Stanford University, California.

[19] Groeneboom P. and Wellner J.A. (1992), *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhäuser Verlag, Basel.

[20] Jewell N.P. (1982), Mixtures of exponential distributions, *Ann. Statist.* **10**, 479–484.

[21] Karlin S. (1981), *A Second Course in Stochastic Processes*, Academic Press.

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

[23] Koshevnik Yu. A. and Levit B. Ya. (1976), On a nonparametric analogue of the information matrix, *Theory Prob. Appl.* **21**, 738–753.

[24] Van der Laan M. J. (1993), *Efficient and Inefficient Estimation in Semiparametric Models*, thesis, Department of Mathematics, University of Utrecht.

[25] Van der Laan M.J. (1994), *Proving efficiency of NPMLE and identities*, Technical report 44, Group in Biostatistics, University of California, Berkeley.

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

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

[28] Lok J. (1994), *The NPMLE of the orientation distribution and comparison of estimators of the length distribution in the line segment problem*, master's thesis, University of Utrecht.

[29] Murphy S.A. (1993), Consistency in a proportional hazards model incorporating a random effect, *Ann. Statist.* **21**.

[30] Pollard D. (1984), *Convergence of Stochastic Processes*, Springer-Verlag, New York.

[31] Scholz F.W. (1980), Towards a unified definition of maximum likelihood, *Canad. J. Statist.* **8**, 193–203.

[32] Stoyan D., Kendall W.S. and Mecke J. (1987), *Stochastic Geometry and its Applications*, Wiley, Chichester. (Second edition 1995).

[33] Turnbull B.W. (1976), The empirical distribution function with arbitrarily grouped censored and truncated data, *J. R. Statist. Soc. B***38**, 290–5.

[34] Van der Vaart A.W. and Wellner J.A. (1993), *Weak convergence and Empirical Processes*, Springer-Verlag (to appear).

[35] Van der Vaart A.W. (1988), *Statistical Estimation in Large Parameter Spaces*, CWI tract 44, Centre for Mathematics and Computer Science, Amsterdam.

[36] Van der Vaart A.W. (1991), Efficiency and Hadamard differentiable functionals, *Scand. J. Statist*, **18**, 63–75.

[37] Vardi Y. (1982), Nonparametric estimation in renewal processes, *Ann. Statist.* **10**, 772–785.

[38] Vardi Y. and Zhang C. (1992), Large sample study of empirical distributions in a random-multiplicative censoring model, *Ann. Statist.*, **20**, 1022–1039.

[39] Wijers B.J. (1995), Consistent Non-parametric Estimation for a One-dimensional Line Segment process Observed in an Interval, *Scand. J. Statist.*, volume **22**, number 3, 335–360.

[40] Wu C.F.J. (1983), On the convergence properties of the EM algorithm, *Ann. Statist.* **11**, 95–103.

Index

N

O

P

R

S

T

U

V

W

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.
2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.
3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.
4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.
5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.
6 A.P.W. Böhm. *Dataflow computation.* 1984.
7 A. Blokhuis. *Few-distance sets.* 1984.
8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.
9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.
10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.
11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.
12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.
13 D. Grune. *On the design of ALEPH.* 1985.
14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.
15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.
16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.
17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.
18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.
19 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 1: Philosophy, framework, computer science.* 1986.
20 B.F. Schriever. *Order dependence.* 1986.
21 D.P. van der Vecht. *Inequalities for stopped Brownian motion.* 1986.
22 J.C.S.P. van der Woude. *Topological dynamix.* 1986.
23 A.F. Monna. *Methods, concepts and ideas in mathematics: aspects of an evolution.* 1986.
24 J.C.M. Baeten. *Filters and ultrafilters over definable subsets of admissible ordinals.* 1986.
25 A.W.J. Kolen. *Tree network and planar rectilinear location theory.* 1986.
26 A.H. Veen. *The misconstrued semicolon: Reconciling imperative languages and dataflow machines.* 1986.
27 A.J.M. van Engelen. *Homogeneous zero-dimensional absolute Borel sets.* 1986.
28 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 2: Applications to natural language.* 1986.
29 H.L. Trentelman. *Almost invariant subspaces and high gain feedback.* 1986.
30 A.G. de Kok. *Production-inventory control models: approximations and algorithms.* 1987.
31 E.E.M. van Berkum. *Optimal paired comparison designs for factorial experiments.* 1987.
32 J.H.J. Einmahl. *Multivariate empirical processes.* 1987.
33 O.J. Vrieze. *Stochastic games with finite state and action spaces.* 1987.
34 P.H.M. Kersten. *Infinitesimal symmetries: a computational approach.* 1987.
35 M.L. Eaton. *Lectures on topics in probability inequalities.* 1987.
36 A.H.P. van der Burgh, R.M.M. Mattheij (eds.). *Proceedings of the first international conference on industrial and applied mathematics (ICIAM 87).* 1987.
37 L. Stougie. *Design and analysis of algorithms for stochastic integer programming.* 1987.
38 J.B.G. Frenk. *On Banach algebras, renewal measures and regenerative processes.* 1987.
39 H.J.M. Peters, O.J. Vrieze (eds.). *Surveys in game theory and related topics.* 1987.
40 J.L. Geluk, L. de Haan. *Regular variation, extensions and Tauberian theorems.* 1987.
41 Sape J. Mullender (ed.). *The Amoeba distributed operating system: Selected papers 1984-1987.* 1987.
42 P.R.J. Asveld, A. Nijholt (eds.). *Essays on concepts, formalisms, and tools.* 1987.
43 H.L. Bodlaender. *Distributed computing: structure and complexity.* 1987.
44 A.W. van der Vaart. *Statistical estimation in large parameter spaces.* 1988.
45 S.A. van de Geer. *Regression analysis and empirical processes.* 1988.
46 S.P. Spekreijse. *Multigrid solution of the steady Euler equations.* 1988.
47 J.B. Dijkstra. *Analysis of means in some non-standard situations.* 1988.
48 F.C. Drost. *Asymptotics for generalized chi-square goodness-of-fit tests.* 1988.
49 F.W. Wubs. *Numerical solution of the shallow-water equations.* 1988.
50 F. de Kerf. *Asymptotic analysis of a class of perturbed Korteweg-de Vries initial value problems.* 1988.
51 P.J.M. van Laarhoven. *Theoretical and computational aspects of simulated annealing.* 1988.
52 P.M. van Loon. *Continuous decoupling transformations for linear boundary value problems.* 1988.
53 K.C.P. Machielsen. *Numerical solution of optimal control problems with state constraints by sequential quadratic programming in function space.* 1988.
54 L.C.R.J. Willenborg. *Computational aspects of survey data processing.* 1988.
55 G.J. van der Steen. *A program generator for recognition, parsing and transduction with syntactic patterns.* 1988.
56 J.C. Ebergen. *Translating programs into delay-insensitive circuits.* 1989.
57 S.M. Verduyn Lunel. *Exponential type calculus for linear delay equations.* 1989.
58 M.C.M. de Gunst. *A random model for plant cell population growth.* 1989.
59 D. van Dulst. *Characterizations of Banach spaces not containing l^1.* 1989.
60 H.E. de Swart. *Vacillation and predictability properties of low-order atmospheric spectral models.* 1989.

61 P. de Jong. *Central limit theorems for generalized multilinear forms.* 1989.

62 V.J. de Jong. *A specification system for statistical software.* 1989.

63 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part I.* 1989.

64 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part II.* 1989.

65 B.M.M. de Weger. *Algorithms for diophantine equations.* 1989.

66 A. Jung. *Cartesian closed categories of domains.* 1989.

67 J.W. Polderman. *Adaptive control & identification: Conflict or conflux?.* 1989.

68 H.J. Woerdeman. *Matrix and operator extensions.* 1989.

69 B.G. Hansen. *Monotonicity properties of infinitely divisible distributions.* 1989.

70 J.K. Lenstra, H.C. Tijms, A. Volgenant (eds.). *Twenty-five years of operations research in the Netherlands: Papers dedicated to Gijs de Leve.* 1990.

71 P.J.C. Spreij. *Counting process systems. Identification and stochastic realization.* 1990.

72 J.F. Kaashoek. *Modeling one dimensional pattern formation by anti-diffusion.* 1990.

73 A.M.H. Gerards. *Graphs and polyhedra. Binary spaces and cutting planes.* 1990.

74 B. Koren. *Multigrid and defect correction for the steady Navier-Stokes equations. Application to aerodynamics.* 1991.

75 M.W.P. Savelsbergh. *Computer aided routing.* 1992.

76 O.E. Flippo. *Stability, duality and decomposition in general mathematical programming.* 1991.

77 A.J. van Es. *Aspects of nonparametric density estimation.* 1991.

78 G.A.P. Kindervater. *Exercises in parallel combinatorial computing.* 1992.

79 J.J. Lodder. *Towards a symmetrical theory of generalized functions.* 1991.

80 S.A. Smulders. *Control of freeway traffic flow.* 1996.

81 P.H.M. America, J.J.M.M. Rutten. *A parallel object-oriented language: design and semantic foundations.* 1992.

82 F. Thuijsman. *Optimality and equilibria in stochastic games.* 1992.

83 R.J. Kooman. *Convergence properties of recurrence sequences.* 1992.

84 A.M. Cohen (ed.). *Computational aspects of Lie group representations and related topics. Proceedings of the 1990 Computational Algebra Seminar at CWI, Amsterdam.* 1991.

85 V. de Valk. *One-dependent processes.* 1994.

86 J.A. Baars, J.A.M. de Groot. *On topological and linear equivalence of certain function spaces.* 1992.

87 A.F. Monna. *The way of mathematics and mathematicians.* 1992.

88 E.D. de Goede. *Numerical methods for the three-dimensional shallow water equations.* 1993.

89 M. Zwaan. *Moment problems in Hilbert space with applications to magnetic resonance imaging.* 1993.

90 C. Vuik. *The solution of a one-dimensional Stefan problem.* 1993.

91 E.R. Verheul. *Multimedians in metric and normed spaces.* 1993.

92 J.L.M. Maubach. *Iterative methods for non-linear partial differential equations.* 1994.

93 A.W. Ambergen. *Statistical uncertainties in posterior probabilities.* 1993.

94 P.A. Zegeling. *Moving-grid methods for time-dependent partial differential equations.* 1993.

95 M.J.C. van Pul. *Statistical analysis of software reliability models.* 1993.

96 J.K. Scholma. *A Lie algebraic study of some integrable systems associated with root systems.* 1993.

97 J.L. van den Berg. *Sojourn times in feedback and processor sharing queues.* 1993.

98 A.J. Koning. *Stochastic integrals and goodness-of-fit tests.* 1993.

99 B.P. Sommeijer. *Parallelism in the numerical integration of initial value problems.* 1993.

100 J. Molenaar. *Multigrid methods for semiconductor device simulation.* 1993.

101 H.J.C. Huijberts. *Dynamic feedback in nonlinear synthesis problems.* 1994.

102 J.A.M. van der Weide. *Stochastic processes and point processes of excursions.* 1994.

103 P.W. Hemker, P. Wesseling (eds.). *Contributions to multigrid.* 1994.

104 I.J.B.F. Adan. *A compensation approach for queueing problems.* 1994.

105 O.J. Boxma, G.M. Koole (eds.). *Performance evaluation of parallel and distributed systems - solution methods. Part 1.* 1994.

106 O.J. Boxma, G.M. Koole (eds.). *Performance evaluation of parallel and distributed systems - solution methods. Part 2.* 1994.

107 R.A. Trompert. *Local uniform grid refinement for time-dependent partial differential equations.* 1995.

108 M.N.M. van Lieshout. *Stochastic geometry models in image analysis and spatial statistics.* 1995.

109 R.J. van Glabbeek. *Comparative concurrency semantics and refinement of actions.* 1996.

110 W. Vervaat, H. Holwerda (ed.). *Probability and lattices.* 1997.

111 I. Helsloot. *Covariant formal group theory and some applications.* 1995.

112 R.N. Bol. *Loop checking in logic programming.* 1995.

113 G.J.M. Koole. *Stochastic scheduling and dynamic programming.* 1995.

114 M.J. van der Laan. *Efficient and inefficient estimation in semiparametric models.* 1995.

115 S.C. Borst. *Polling models.* 1996.

116 G.D. Otten. *Statistical test limits in quality control.* 1996.

117 K.G. Langendoen. *Graph reduction on shared-memory multiprocessors.* 1996.

118 W.C.A. Maas. *Nonlinear $\mathcal{H}_\infty$ control: the singular case.* 1996.

119 A. Di Bucchianico. *Probabilistic and analytical aspects of the umbral calculus.* 1997.

120 M. van Loon. *Numerical methods in smog prediction.* 1997.

121 B.J. Wijers. *Nonparametric estimation for a windowed line-segment process.* 1997.

122 W.K. Klein Haneveld, O.J. Vrieze, L.C.M. Kallenberg (editors). *Ten years LNMB – Ph.D. research and graduate courses of the Dutch Network of Operations Research.* 1997.

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print: replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservations of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statements.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Empirical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallmann spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *GO-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewai theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequencies.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.