A STUDY OF BOOTSTRAP AND LIKELIHOOD BASED METHODS IN NON-STANDARD PROBLEMS

by

Bodhisattva Sen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
2008

Doctoral Committee:

Associate Professor Moulinath Banerjee, Co-Chair
Professor Michael B. Woodroofe, Co-Chair
Professor Mario L. Mateo
Professor George Michailidis
Assistant Professor Stilian Atanasov Stoev

In loving memory of *Amma* (my grandma)

ACKNOWLEDGEMENTS

I wish to express my profound gratitude towards my advisors – Professor Moulinath Banerjee and Professor Michael Woodroofe – for their guidance, encouragement and inspiration. I am indebted to Professor Banerjee for the time and energy he has devoted towards mentoring me. And I am privileged to have worked with Professor Woodroofe, who has helped me to improve and broaden my statistical skills, more than anyone else. I would like to thank my collaborators – Professor Mario Mateo, Matthew Walker and Javier Alonso from the Astronomy department – for truly initiating me in interdisciplinary research.

I am grateful to all my teachers, from middle school to graduate school, for their guidance and help. Special thanks go to Swapan Kumar Das, Professor Pratip Mukherjee, Professor B.V.Rao, Professor Probal Chaudhuri and Professor George Michailidis. I also thank all my friends with whom I spent the formative years of my life, especially Rupak Das, Sarang Sane and Samsiddhi Bhattacharjee.

Finally, I wish to thank my family – my wife, my parents, my brother and my late grandparents – for their encouragement, companionship and support; without them this thesis would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER 1

Introduction

In this dissertation we investigate bootstrap and likelihood based methods for constructing confidence intervals in some non-standard problems. The non-standard problems studied include problems with non root-n convergence (e.g., cube-root convergence, $\sqrt{\frac{n}{\log n}}$-rate of convergence), estimation problems where the parameter is at the boundary, and study of non-smooth/abrupt-change models.

An integral part of the statistical methodology investigated in the thesis involves inference on non-parametric function estimation that obey shape restrictions, like monotonicity/convexity. Although the estimation of such shape restricted functions has a long history in statistics, inference on these estimated functions has been theoretically and practically a challenging exercise. The pointwise limit distribution of the properly normalized (Wald-type) estimators involve non-standard asymptotics and complex limits with nuisance parameters that are difficult to estimate, thereby hindering the usefulness of these estimators. We explore two natural alternatives for inference in this situation – the use of likelihood ratio-type test statistics that give a nuisance parameter free limit and the use of bootstrap methods. Most of the methodological and theoretical contributions of the thesis has been motivated by applications in astronomy, high energy physics and epidemiology. We first give a brief preview of the motivating applications.

1.1 A preview of some of the applications

Setting confidence bounds is an essential part of the reporting of experimental results. Current physics experiments are often done to measure bounded parameters (that might be at the boundary), e.g., nonnegative parameters that are small and may be zero, and to search for small signals in the presence of backgrounds. Sometimes in such situations data is known a priori to be relatively improbable for all parameter values under consideration, and classical statistical procedures suggests a parameter estimate beyond the bound. We consider some of the typical examples that arise in high energy physics and propose methods of constructing confidence intervals for a finite-dimensional parameter of interest in presence of nuisance parameters that have the correct coverage and better finite sample properties.

In epidemiology, one often encounters data on time to infection/illness (e.g., HIV infection) gathered from a number of individuals over a period of time. Each individual is followed up at the clinic for a random number of times and the times of inspection are noted. The two successive observation times within which the individual succumbed to infection/illness is recorded. We are interested in estimating the distribution of the time to infection, which has important medical consequences. We advocate the use of a pseudo-likelihood ratio based method for constructing pointwise confidence bands around the distribution function of the time to infection with such interval censored data. Our method is computationally simple and avoids the need for estimating nuisance parameters, a major problem which earlier methods had failed to resolve satisfactorily.

A major part of the thesis has been motivated by an astronomical application - estimation of dark matter distribution in dwarf galaxies. An essential component of the application involves estimation and inference on functions that obey shape restrictions, like monotonicity/convexity. Bootstrap is probably the most commonly used inferential procedure in complex problems. We study the performance of different bootstrap methods for inference in non-parametric estimation of a monotone function.

Another feature of the astronomy application is that although our interest lies in the three-dimensional distribution of position of stars in a galaxy, we can only observe their two-dimensional projections. This gives rise to a problem in stereology – the study of three-dimensional properties of objects or matter usually observed two-dimensionally. We develop functions that capture the three-dimensional features of position under assumptions of spherical symmetry, and estimate them by utilizing natural shape constraints. We find the limit behavior of the estimators and study the consistency of bootstrap methods for constructing pointwise confidence bands.

Whether a dwarf spheroidal galaxy is in equilibrium or being tidally disrupted by the Milky Way is an important question for the study of its dark matter content and distribution. There is conjecture that in some galaxies, like Leo I, the stars in the outer halo experience streaming motion. This raises several interesting statistical questions. Is streaming motion evident in Leo I? If so, how can it be described and estimated? To what extent can it be described by a threshold model, in which streaming motion is only present for stars a sufficient distance from the center? We address these questions by modeling the effect of streaming motion using isotonic methods and change-point type models.

1.2 Summary of the thesis: a statistical perspective

On the Unified method with nuisance parameter Construction of confidence interval for a finite-dimensional parameter of interest in presence of nuisance parameters is an old problem in statistics. But the classical methods do not work well – the confidence intervals are drastically short – when the parameter of interest is bounded and the data observed is a priori known to be relatively improbable for all parameter values. Such situations occur quite often in high energy physics. Feldman and Cousins (1998), in an influential paper, showed how to construct confidence regions consisting of parameter values with high relative likelihood with exact coverage probabilities in problems with moderate sample sizes and boundary effects, like a positive normal mean or a Poisson rate that is known to exceed a background value, that are of interest in high energy physics. In Chapter 2 we discuss a generalization of the unified method by Feldman and Cousins (1998) with nuisance parameters. We demonstrate our method with several examples that arise quite frequently in high energy physics and astronomy. We also discuss the hybrid resampling method of Chuang and Lai (1998, 2000) and implement it in some of the problems.

A pseudo-likelihood method for analyzing interval censored data Interval censoring is a type of censoring that has become increasingly common in the areas that produce failure time data. In a mixed case model, an individual is observed a random number of times, and at each time it is recorded whether an "event" has happened or not. One seeks to estimate the distribution of time to event. Chapter 3 introduces a method based on a pseudo-likelihood ratio for estimating the distribution function of the survival time in a mixed-case interval censoring model. We use a Poisson process as the basis of a likelihood function to construct a pseudo-likelihood ratio statistic for testing the value of the distribution function at a fixed point. We show that the pseudo-likelihood ratio statistic converges in distribution under the null hypothesis to a nuisance parameter free limit. This family of hypotheses can be easily inverted to give pointwise confidence intervals for the failure time distribution function. The computation of the confidence sets is simple, requiring the use of the pool adjacent violators algorithm, or a standard isotonic regression algorithm. We also illustrate the superiority of the proposed method over competitors based on resampling techniques or on the limit distribution of the maximum pseudo-likelihood estimator, through simulation studies, and illustrate the different methods on a data set involving time to HIV seroconversion in a group of haemophiliacs.

Inconsistency of Bootstrap: the Grenander estimator A common method of constructing confidence intervals in complex scenarios is to resort to bootstrapping. In Chapter 4, we investigate the behavior of different bootstrap methods with the Grenander estimator, the nonparametric maximum likelihood estimator of a decreasing density on $[0, \infty)$, a prototypical example of a shape restricted estimator that exhibits cube-root asymptotics. The non-standard rate of convergence to a non-normal limit distribution makes the conventional bootstrap methods a suspect in this situation. Our main results show the inconsistency of the conventional bootstrap methods; in fact, we claim that the bootstrap estimate of the sampling distribution does not have any weak limit conditionally (given the data), in probability. We derived sufficient conditions under which different bootstrap procedures will be consistent. Our results have direct implications to estimators that exhibit cube-root asymptotics – something we plan to explore in more detail in the future.

Bootstrap in the Wicksell's problem We consider a stereological problem like that of Wicksell (1925), that arises in astronomy in connection to dark matter estimation. Let $\mathbf{X} = (X_1, X_2, X_3)$ be a spherically symmetric random vector of which only (X_1, X_2) can be observed. We focus attention on estimating F, the distribution function of the squared radius $Z := X_1^2 + X_2^2 + X_3^2$, from a random sample of (X_1, X_2). The quantity of interest can be related to functions that obey shape constraints. Using the assumption of spherical symmetry, Chapter 5 defines natural estimators of the quantity of interest – the distribution of the three-dimensional radius. We find limit distributions of the estimators, that exhibit $\sqrt{\frac{n}{\log n}}$-rate of convergence to a normal distribution with unknown variance. We propose bootstrap based confidence intervals for the estimators and prove the consistency of the procedure. Although the asymptotics involved are non-standard, but the convergence to a normal distribution plays an important role in the consistency of bootstrap methods.

Streaming motion in Leo I galaxy There is preliminary evidence that in some galaxies, like Leo I, the stars in the outer halo experience streaming motion. The main goal of Chapter 6 is to model the effect of such a streaming motion and make inference on the parameters that describe the model. We focus our attention to understanding streaming motion in Leo I. We model the effect of streaming motion, test hypothesis for significance, quantify the effect by estimating the parameters and give confidence intervals. We find that although there is evidence of streaming, the significance is not conclusive. We try to fit threshold models, in which streaming motion is only present for stars at a sufficient distance from the center, and compute the estimates of the threshold parameter and derive their limit distributions, under model misspecification. M-estimation techniques and estimation of monotone function arise naturally in this context. Key results from the empirical process literature are crucially used in deriving the limit distributions of the estimates of the change points.

CHAPTER 2

On the Unified method with nuisance parameter

In this chapter we consider the problem of constructing confidence interval for a finite-dimensional parameter of interest in presence of nuisance parameters. We discuss a generalization of the unified method by Feldman and Cousins (1998) with nuisance parameters. We demonstrate our method with several examples that arise quite frequently in high energy physics and astronomy. We also discuss the hybrid resampling method of Chuang and Lai (1998, 2000) and implement it in some of the problems.

2.1 Introduction

Confidence regions consisting of parameter values with high relative likelihood have a long tradition with Statistics and have generated a large literature, much of which emphasizes asymptotic calculations. See Reid (2003) for a recent survey article and Reid and Fraser (2003) for a relevant application. In an influential paper Feldman and Cousins (1998) showed how to implement this construction with exact coverage probabilities in problems with moderate sample sizes and boundary effects, like a positive normal mean or a Poisson rate that is known to exceed a background value, that are of interest in high energy physics. They called the construction the unified method because it makes a natural transition from an one-sided confidence bound to a two-sided confidence interval. Only problems without nuisance parameters were considered in Feldman and Cousins (1998). Here we retain the interest in problems with boundary effects and moderate sample sizes but focus on problems with nuisance parameters in addition to the parameter of primary interest. We start with describing the unified method of Feldman and Cousins (1998) applied to the signal plus noise model arising in high energy physics.

2.1.1 The Signal plus Noise problem without nuisance parameters

The KARMEN group has been searching for a neutrino oscillation signal reported by a Liquid Scintillating Neutrino Detector (LSDN) experiment. As of Summer 1998, they had expected to see 2.88 ± 0.13 background (noise) events and $1.0 - 1.5$ signal events, if the LSND results were real, but they have seen no events. From their analysis, they claimed to almost exclude the effect claimed by the LSND experiment, a claim that was later criticized by Roe and Woodroofe (1999, 2000). Attention here focusses on constructing a confidence interval for the rate of signal event, adjusting for the uncertainty in the background rate.

The background radiation is added to a signal producing a total observed count N; we assume $N \sim \text{Poisson}(b + \theta)$. Here the background and signal are assumed to be independent Poisson random variables, with mean $b \geq 0$ (assumed to be known for the time being) and $\theta \geq 0$ respectively. Feldman and Cousins (1998) proposed the *unified approach* which uses the likelihood ratio statistic (LRS) as the ordering principle (to order the probable data values) and then computes a $(1 - \alpha)$ confidence region for θ based on the exact distribution of the LRS. Mandelkern (2002) addressed the general question on setting confidence intervals for bounded parameters in the a review article, which subsequently received much attention in the statistics community. Further discussions on the unified approach and its drawbacks can be found in Roe and Woodroofe (1999). Roe and Woodroofe (2000) used bayesian methods with uniform priors in this problem, and investigated the frequentist properties of the procedures.

2.1.2 The unified method with nuisance parameters

To describe the unified method and understand the issues, suppose that a data vector X has a probability density (or mass function, in the discrete case) $f_{\theta,\eta}$ where θ is the parameter of interest and η is a nuisance parameter. For example, if a mass θ is measured with normally distributed error with an unknown standard deviation, then θ is of primary interest and the standard deviation of the measurement is a nuisance. Let L denote the likelihood function, i.e., $L(\theta,\eta|x) = f_{\theta,\eta}(x)$; further, let $\hat{\eta}_\theta = \hat{\eta}_\theta(x)$ be the value of η that maximizes $L(\theta,\eta|x)$ for a fixed θ; let $\hat{\theta} = \hat{\theta}(x)$ and $\hat{\eta} = \hat{\eta}(x)$ be the values of θ and η that maximize $L(\theta,\eta|x)$ over all allowable values; and let

$$\Lambda_\theta(x) = \frac{L(\theta, \hat{\eta}_\theta(x)|x)}{L(\hat{\theta}(x), \hat{\eta}(x)|x)}. \tag{2.1}$$

Then unified confidence intervals consist of θ for which $\Lambda_\theta(x) \geq c_\theta$, where c_θ is a value whose computation is discussed below.

For a desired level of coverage $1-\alpha$, a literal (and correct) interpretation of "confidence" requires that $P_{\theta,\eta}[\Lambda_\theta(X) \geq c_\theta] \geq 1-\alpha$ for all θ and η, where $P_{\theta,\eta}$ denotes probability computed under the assumption that the parameter values are θ and η. Equivalently it requires $\min_\eta P_{\theta,\eta}[\Lambda_\theta(X) \geq c_\theta] \geq 1-\alpha$ for each θ. Thus, c_θ should be the largest value of c for which

$$\min_{\eta} P_{\theta,\eta}\big[\Lambda_\theta(X) \geq c\big] \geq 1-\alpha. \tag{2.2}$$

For a fixed x, the confidence interval is then $\mathcal{C}(x) = \{\theta : \Lambda_\theta(x) \geq c_\theta\}$, and its coverage probability

$$P_{\theta,\eta}\big[\theta \in \mathcal{C}(X)\big] = P_{\theta,\eta}\big[\Lambda_\theta(X) \geq c_\theta\big] \geq 1-\alpha, \tag{2.3}$$

by construction. Being likelihood based, unified confidence intervals are generally reliable, even optimal, in large samples, but not necessarily so in small samples, and unified confidence intervals have been criticized in that context – e.g., Roe and Woodroofe (1999, 2000).

In some simple cases, it is possible to compute c_θ analytically. This is illustrated in Section 2.2. In other cases, one can in principle proceed by numerical calculation. This requires computing $P_{\theta,\eta}[\Lambda_\theta(X) \geq c]$ over a grid of (θ, η, c) values, either by Monte-Carlo or numerical integration, and then finding the c_θ by inspection, replacing the minimum in (2.2) by the minimum over the grid. This is feasible if η is known or absent and was done by Feldman and Cousins in two important examples. But if η is present and unknown, then numerical calculations become unwieldy, especially if η is a vector.

2.1.3 The Hybrid resampling method

One way to circumvent the unwieldy numerical problems, when η is present, is to use the chi-squared approximation to the distribution of Λ_θ, as in Rolke, W., López, A. and Conrad, J. (2005), or a chi-squared approximation supplemented by a Bartlett correction. Another is to use the hybrid resampling method of Chuang and Lai (1998, 2000). We generate random variable X^* from $P_{\theta,\hat{\eta}_\theta}$ and let $c_\theta^+ = c_\theta^+(x)$ be the largest value of c for which

$P_{\theta,\hat{\eta}_\theta}[\Lambda_\theta(X^*) \geq c] \geq 1-\alpha$. Then the hybrid confidence intervals consist of θ for which $\Lambda_\theta(x) \geq c_\theta^+$. This requires computation over a grid of θ values, but not over η for fixed θ. Unfortunately, relation (2.3) cannot be asserted for the hybrid intervals, but Chuang and Lai argue both theoretically and by example that it should be approximately true. In some cases the calculations can be done by numerical integration, but they can always be done by simulation. For a given x, generate independent $X_1^*, \cdots, X_N^*$ (pseudo) random observations from the density $f_{\theta,\hat{\eta}_\theta}$; compute $\Lambda_\theta(X_k^*)$ from (2.1) with x replaced by X_k^*; and let c_θ^* be the largest value of c for which

$$\frac{\#\{k \leq N : \Lambda_\theta(X_k^*) \geq c\}}{N} \geq 1-\alpha. \tag{2.4}$$

Here the left side of (2.4) provides a Monte Carlo estimate for $P_{\theta,\hat{\eta}_\theta}[\Lambda_\theta(X^*) \geq c]$, and c_θ^* provides an estimate of c_θ^+.

The hybrid method resembles Efron's bootstrap resampling method, but differs in one important respect. For computing (2.2) for fixed θ, θ and η are replaced by θ and $\hat{\eta}_\theta$, as opposed to $\hat{\theta}$ and $\hat{\eta}$. This is the origin of the term "hybrid". Evidence that the hybrid method is reliable – that is, that (2.3) is approximately true comes from two sources, asymptotic approximations and simulations. These are reported in Chuang and Lai (1998, 2000) and include some dramatic successes. Here the method is applied to three examples of interest to astronomers and physicists. The hybrid method has (independently) been suggested in the physics literature by Feldman (2000).

2.2 Some Examples

In this section we describe the analytic computation of c_θ based on Equation (2.2) in some problems of interest to high energy physics and astronomy.

We start with constructing a confidence interval in a normal model for mean $\theta \geq 0$ and unknown variance σ^2 (where σ^2 is the nuisance parameter). In Sub-section 2.2.2 we work out the details of the method when the parameter of interest is the angle between the mean vector of a bivariate normal population. This example has applications in astronomy. The third example we look at is a version of the "signal plus noise" problem that arises often in high energy physics. We observe $N \sim \text{Poisson}(b + \theta)$ and independently $M \sim \text{Poisson}(\gamma b)$, where γ is a known constant, θ is the signal rate (the parameter of interest) and b is the background rate (a nuisance parameter). The aim is to construct a $1 - \alpha$ confidence interval for θ. We are not able to analytically compute c_θ for this example. The details are provided in Section 2.2.3. An extension of this problem is treated in Section 2.2.4 with an application to astronomy. With every "event" we also observe a random variable with distribution depending on the type of "event" (signal event or background event). We use the EM algorithm to maximize the likelihood of this mixture model. We construct a $1 - \alpha$ confidence interval for θ using the hybrid resampling method. This generalization also arises in high energy physics.

2.2.1 The Normal Case

Suppose that $X = (Y, W)$, where Y and W are independent, Y is normally distributed with mean $\theta \geq 0$ and variance σ^2, and W/σ^2 has a chi-squared distribution with r degrees of freedom. For example, if data originally consists of a sample $Y_i = \theta + \epsilon_i,\ i = 1, \cdots, n$, where ϵ_i's are independent and identically distributed $N(0, \sigma^2)$, then one can let $Y = \bar{Y}$ and $W = (n-1)V^2/n$ where $\bar{Y}$ and V^2 denote the sample mean and variance of $Y_1, \cdots, Y_n$. The unknown parameters here are $\theta \geq 0$ and $\sigma^2 > 0$. Thus, the likelihood function is

$$L(\theta, \sigma^2 | y, w) = \frac{1}{\sqrt{2^{r+1}\pi}\Gamma(r/2)} \frac{w^{\frac{1}{2}r-1}}{\sigma^{r+1}} \exp\left\{-\frac{1}{2\sigma^2}[(y-\theta)^2 + w]\right\}.$$

For a given θ, L is maximized by

$$\hat{\sigma}^2_\theta = \frac{1}{r+1}\big[w + (y-\theta)^2\big];$$

and L is maximized with respect to θ and σ^2 jointly by $\hat{\theta} = \max[0, y] = y_+$, say, and

$$\hat{\sigma}^2 = \frac{1}{r+1}\big[w + (y_-)^2\big],$$

where $y_- = -\min[0, y]$. After some simple algebra,

$$\log[\Lambda_\theta] = -\frac{1}{2}(r+1)\log(\frac{\hat{\sigma}^2_\theta}{\hat{\sigma}^2}) = -\frac{1}{2}(r+1)\log\left[\frac{W + (Y-\theta)^2}{W + (Y_-)^2}\right].$$

Let

$$U = \frac{W}{\sigma^2} \qquad \text{and} \qquad Z = \frac{Y-\theta}{\sigma}.$$

Then U and Z are independent random variables for which $U \sim \chi^2_r$ and $Z \sim \text{Normal}(0, 1)$, and

$$\log[\Lambda_\theta] = -\frac{1}{2}(r+1)\log\left[\frac{U + Z^2}{U + [(Z + \theta/\sigma)_-]^2}\right].$$

This is an increasing function of σ for each $\theta > 0$. So, since the joint distribution of U and Z does not depend on parameters,

$$\min_{\sigma>0} P_{\theta,\sigma}\big[\Lambda_\theta \geq c\big] = \lim_{\sigma\to 0} P_{\theta,\sigma}\big[\Lambda_\theta \geq c\big] = P\left[-\frac{1}{2}(r+1)\log\big(1 + \frac{T^2}{r}\big) \geq \log(c)\right],$$

where $T = \frac{Z}{\sqrt{U/r}}$ has t-distribution with r degrees of freedom. Thus the desired c is

$$c = \exp\Big\{-\frac{1}{2}(r+1)\log\big[1 + \frac{t^2_{r,1-\frac{1}{2}\alpha}}{r}\big]\Big\},$$

where $t_{r,1-\frac{1}{2}\alpha}$ is the $1 - \frac{1}{2}\alpha$ percentile of the latter distribution and is independent of θ. To find the confidence intervals, one must solve the inequality

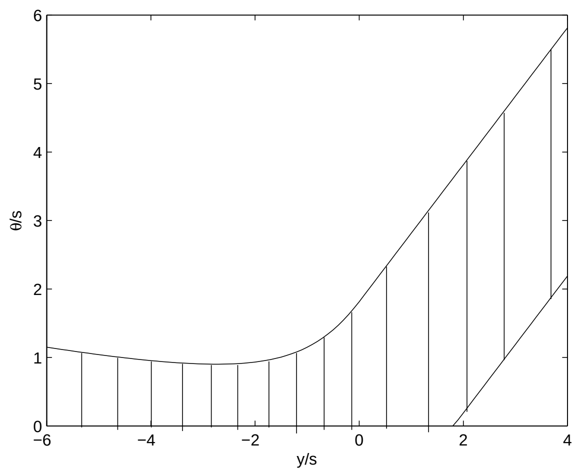

Figure 2.1: Confidence limits for θ/s as a function of y/s when $r = 10$ and $\alpha = 0.1$. Observe that the upper limit starts to increase as y decreases for $y < 0$.

$\Lambda_\theta \geq c$ for θ. Letting $s^2 = W/r$, this may be written

$$\frac{1 + (y-\theta)^2/(rs^2)}{1 + y_-^2/(rs^2)} \leq 1 + \frac{t^2_{r,1-\frac{1}{2}\alpha}}{r},$$

or

$$[y - bs]_+ \leq \theta \leq y + bs, \tag{2.5}$$

where

$$b = \sqrt{t^2_{r,1-\frac{1}{2}\alpha} + \frac{y_-^2}{s^2}\left(1 + \frac{t^2_{r,1-\frac{1}{2}\alpha}}{r}\right)}. \tag{2.6}$$

Thus, if $y > 0$, then the unified intervals are just the usual t-intervals, truncated to non-negative values; and if $y > bs$, then they are symmetric about y. This differs from the case of known σ, where the intervals are (slightly) asymmetric, even for large y. There is a more dramatic difference with the case of known σ for $y < 0$. Observe that for $y < 0$,

$$y + bs \geq s\sqrt{\frac{y^2}{s^2}\left(1 + \frac{t^2_{r,1-\frac{1}{2}\alpha}}{r}\right)} - \frac{|y|}{s} = |y|\left\{\sqrt{1 + \frac{t^2_{r,1-\frac{1}{2}\alpha}}{r}} - 1\right\}.$$

So the upper confidence limit approaches $+\infty$ as $y \to -\infty$, unlike the case of known σ where it approaches 0. Mandelkern (2002) found the latter behavior non-intuitive. If we let $r \to \infty$ and $s^2 \to \sigma^2$, then we do *not* recover the intervals of Feldman and Cousins with known σ^2. Rather, we get the interval (2.5) with the t-percentile replaced by the corresponding normal percentile.

Observe that the confidence limits for θ may be written as $[y/s - b]_+ \leq \theta/s \leq y/s + b$. Figure 2.1 shows these upper and lower confidence limits for θ/s as a function of y/s for $r = 10$ and $\alpha = .10$. For a specific example, suppose that $r = 10$, $s = 1$, $y = -.30$ and $\alpha = .10$. Then $b = \sqrt{(1.812)^2 + (.3)^2\{1 + (1.812)^2/10\}} = 1.84$, and the interval is $0 \leq \theta \leq 1.54$. The hybrid method yields $0 \leq \theta \leq 1.14$ in this example. The details are omitted here, but an example using the hybrid method is included in Section 4.

2.2.2 Angles

In Astronomy, "proper motion" refers to the angular velocity of an object in the plane perpendicular to the line of sight. An object's proper motion is given by $X = (X_1, X_2)$, where X_1 and X_2 are orthogonal components and are measured independently. In certain applications astronomers are more concerned with the direction than the magnitude of the proper motion vector. An example is the motion of a satellite galaxy whose stellar orbits may be disrupted by the tidal influence exerted by a larger parent system. Due to outward streaming of its stars, a disrupting satellite will elongate spatially and exhibit a radial velocity gradient along the direction of elongation. N-body simulations indicate that the orientations of both the elongation and velocity gradient correlate with the direction of the satellite's proper motion vector (e.g., Oh et al., 1995; Piatek & Pryor, 1995). Constraining the direction of the satellite's proper motion can therefore help determine whether or not a satellite is undergoing disruption, which in turn places constraints on applicable dynamical models.

Suppose X_1 and X_2 are normally distributed random variables with unknown means μ_1 and μ_2 and known variance σ^2. Write μ_1 and μ_2 in polar coordinates, $\mu_1 = \rho\cos(\theta)$ and $\mu_2 = \rho\sin(\theta)$, where $-\pi < \theta \leq \pi$. We consider confidence intervals for θ when ρ is the nuisance parameter.

In this example, the likelihood function,

$$L(\theta, \rho|x) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}\left[(x_1 - \rho\cos(\theta))^2 + (x_2 - \rho\sin(\theta))^2\right]\right\},$$

is maximized for a fixed θ by $\hat{\rho}_\theta = \max[0, x_1\cos(\theta) + x_2\sin(\theta)]$ and unconditionally by $\hat{\rho}$ and $\hat{\theta}$, where $x_1 = \hat{\rho}\cos(\hat{\theta})$ and $x_2 = \hat{\rho}\sin(\hat{\theta})$. Then $L(\hat{\theta}, \hat{\rho}|x) = 1/(2\pi\sigma^2)$, and

$$\Lambda_\theta = \exp\big[-\frac{1}{2\sigma^2}(\hat{\rho}^2 - \hat{\rho}_\theta^2)\big].$$

Let

$$Z_1 = \frac{1}{\sigma}\big[\cos(\theta)X_1 + \sin(\theta)X_2 - \rho\big],$$
$$Z_2 = \frac{1}{\sigma}\big[\sin(\theta)X_1 - \cos(\theta)X_2\big].$$

Then Z_1 and Z_2 are independent normal variables (both) with the same mean 0 and unit variance, and

$$\Lambda_\theta = \exp\big\{-\frac{1}{2}[(Z_1 + \rho)_-^2 + Z_2^2]\big\},$$

where (recall) $z_- = -\min[0, z]$, after some simple algebra. Thus, Λ_θ is an increasing function of ρ for fixed Z_1, Z_2, and θ. So, since the joint distribution of Z_1 and Z_1 do not depend on parameters

$$\min_{\rho} P_{\theta,\rho}\big[\Lambda_\theta \geq c\big] = \lim_{\rho\to 0} P_{\theta,\rho}\big[\Lambda_\theta \geq c\big].$$

Letting $b = -2\log(c)$, this is just

$$\begin{aligned} P\big[Z_{1,-}^2 + Z_2^2 \leq b\big] &= P\big[Z_1 \leq 0, Z_1^2 + Z_2^2 \leq b\big] + P\big[Z_1 > 0, Z_2^2 \leq b\big] \\ &= \frac{1}{2}P[\chi_1^2 \leq b] + \frac{1}{2}P[\chi_2^2 \leq b]. \end{aligned}$$

So, $c = e^{-b/2}$, where b solves $\frac{1}{2}P[\chi_1^2 \leq b] + \frac{1}{2}P[\chi_2^2 \leq b] = 1 - \alpha$. For example, when $\alpha = .90,\ b = 3.808$.

Unified confidence intervals for θ then consist of θ for which $\hat{\rho}^2 - \hat{\rho}_\theta^2 \leq b\sigma^2$, or equivalently $\hat{\rho}_\theta^2 \geq \hat{\rho}^2 - b\sigma^2$. Thus, if $\hat{\rho}^2 \leq b\sigma^2$, then the interval consists of all values $-\pi < \theta \leq \pi$. On one hand, this simply reflects the (obvious) fact that if $\hat{\rho}$ is small, then there is no reliable information for estimating θ, but it also admits the following amusing paraphrase: One is $100(1-\alpha)\%$ confident of something that is certain. If $\hat{\rho}^2 > b\sigma^2$, then the intervals consist of θ for which $\hat{\rho}\cos(\theta - \hat{\theta}) \geq \sqrt{\hat{\rho}^2 - b\sigma^2}$; that is

$$\hat{\theta} - \arccos\Big(\sqrt{1 - \frac{b\sigma^2}{\hat{\rho}^2}}\Big) \leq \theta \leq \hat{\theta} + \arccos\Big(\sqrt{1 - \frac{b\sigma^2}{\hat{\rho}^2}}\Big),$$

where $\arccos(y)$ is the unique ω for which $0 \leq \omega \leq \pi$ and $\cos(\omega) = y$ and addition is understood modulo π. Thus, there is a discontinuity in the length of the intervals as $\hat{\rho}$ passes through $b\sigma^2$: It decreases from 2π to something less than π.

Piatek et al. (2002) measured the Galactic rest-frame proper motion of the Fornax galaxy to be $(X_1, X_2) = (32, 33)$ with $\sigma = 13$ (units are in milli-arcseconds per century). Later on Dinescu et al. (2004) made a similar measurement but observed $(X_1, X_2) = (-13, 34)$ with $\sigma = 16$. We use our method to construct a 90% confidence interval for the direction θ in the two cases.

The intervals obtained are $(0.2219, 1.4119)$ for the Piatek et al. angle and $(0.9051, 2.9669)$ for the Dinescu et al. angle (where θ is measured in radians). Note that the Piatek et al. measurement places a tighter constraint on the proper motion direction, and that there is some overlap with the Dinescu et al. result.

2.2.3 Counts with Background

Suppose that $X = (N, M)$ where N and M are independent, M has the Poisson distribution with mean γb, and N has the Poisson distribution with mean $b + \theta$. It is useful to write $N = B + S$ where B and S are independent Poisson random variables with means b and θ, representing the number of background and signal events. Here b and θ are unknown; γ is assumed known and large values of γ are of interest. In this case, the likelihood function and score functions are

$$L(\theta, b|n, m) = f_{\theta,b}(n, m) = \frac{(\gamma b)^m}{m!} e^{-\gamma b} \times \frac{(\theta + b)^n}{n!} e^{-(\theta+b)},$$

$$\frac{\partial \log(L)}{\partial \theta} = \frac{n}{b+\theta} - 1,$$

and

$$\frac{\partial \log(L)}{\partial b} = \frac{m}{b} + \frac{n}{\theta + b} - (\gamma + 1).$$

Consider $\hat{b}_\theta$ for a fixed θ. If $m = 0$, then L is maximized when $b = [n/(\gamma + 1) - \theta]_+$; and if $m > 0$ it is maximized at the (positive) solution to $\partial \log(L)/\partial b = 0$, i.e.,

$$\hat{b}_\theta = \frac{[(m+n) - (\gamma+1)\theta] + \sqrt{[(\gamma+1)\theta - (m+n)]^2 + 4(\gamma+1)m\theta}}{2(\gamma+1)}; \tag{2.7}$$

and fortuitously, (2.7) also gives the correct answer when $m = 0$. The unconstrained maximum likelihood estimators may then be found as $\hat{\theta}$ and $\hat{b} = \hat{b}_{\hat{\theta}}$,

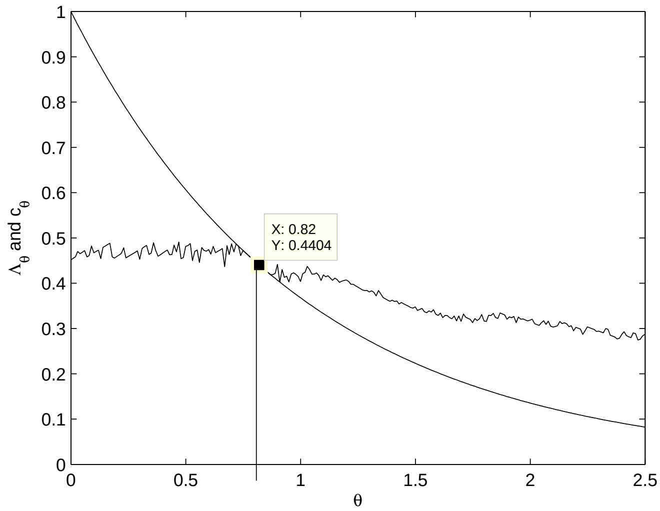

Figure 2.2: Plot of Λ_θ (smooth line) and c_θ (jagged line) against θ when $\gamma = 6, m = 23$, $n = 0$ and $\alpha = 0.10$.

where $\hat{\theta}$ maximizes the profile likelihood function $L(\theta, \hat{b}_\theta | n, m)$. Considering the cases $n \leq m/\gamma$ and $n > m/\gamma$ separately, shows that

$$\hat{\theta} = \big(n - \frac{m}{\gamma}\big)_+$$

and

$$\hat{b} = \frac{m + n - \hat{\theta}}{\gamma + 1}.$$

So,

$$\Lambda_\theta(n, m) = (\frac{\hat{b}_\theta}{\hat{b}})^m (\frac{\theta + \hat{b}_\theta}{\hat{\theta} + \hat{b}})^n \exp\big[(n + m) - (\gamma + 1)\hat{b}_\theta - \theta\big],$$

after some simple algebra.

We have been unable to find the minimizing value in (2.2) and, so, will use the Hybrid Resampling Method. This is best illustrated by an example. Figure 2.2 below shows Λ_θ and c_θ for an example in which $\gamma = 6, m = 23$, $n = 0$ and $\alpha = 0.10$. This is patterned after the original KARMEN report Eitel, K. and Zeitnitz, B. (1998), but with a larger value of $\hat{b}$ and more variability in

$\hat{b}$. The c^*_θ was computed by Monte Carlo on the grid $\theta = 0, 0.01, 0.02, \ldots, 2.50$ using $N = 10,000$ in (2.4). The right end-point of the interval is 0.82.

By construction, the hybrid-unified method always delivers non-degenerate subinterval of $[0, \infty)$, even when $n = 0$, and, thus, avoids the types of problems reported in Rolke, W., López, A. and Conrad, J. (2005). It does not avoid the problems inherent in the use of the unified method without nuisance parameters, however – for example, dependence of the interval on $\hat{b}$ when $n = 0$. We believe that the interval $[0, 2.31]$ is a more reasonable statement of the uncertainty in this example. Briefly, $[0, 2.31]$ would be the uniformly most accurate 90% confidence interval if $S = 0$ were observed; and if $N = 0$, then $B = S = 0$.

2.2.4 The star contamination problem

In studying external (to the Milky Way) galaxies, one can measure only two of the three (those orthogonal to the line of sight) components of stellar position and one (along the line of sight, from redshift of spectral features) of the three components of stellar velocity. Because the line of sight necessarily originates within the Milky Way, velocity samples for distant galaxies frequently suffer from contamination by foreground Milky Way stars. It is important to accurately identify and remove sample contamination. The most common procedure for membership determination involves fitting a normal distribution to the marginal velocity distribution of all observed stars, then iteratively rejecting outliers beyond a specified ($\sim 3\sigma$) threshold. However, this is of limited utility when the velocity distributions of target galaxy and contaminant stars overlap. Also, the trimming of outliers from an imposed distribution introduces a degree of circularity to the analysis, as it is the tar-

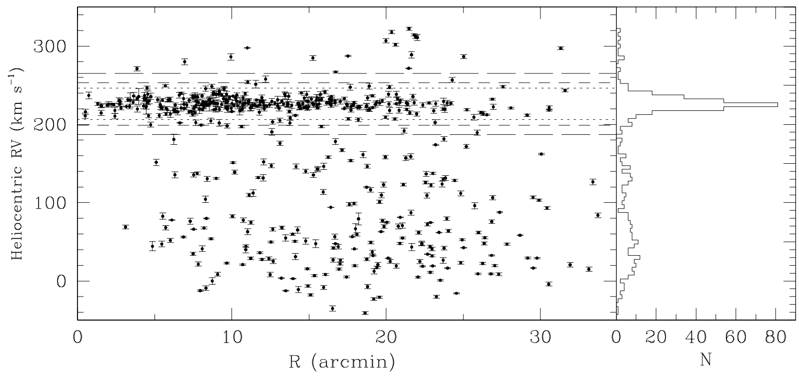

Figure 2.3: Left: Heliocentric radial velocities (RV) vs. angular distance from the Sextans center. Dotted, short-dashed, and long-dashed lines mark boundaries of 276, 294 and 303 member samples, respectively. Right: Histogram of the radial velocity of the stars.

get galaxy's velocity distribution that is under investigation. We consider results from a velocity survey of the Sextans dwarf spheroidal galaxy (see Walker et al. 2006). The unfiltered marginal velocity distribution of the 528 observed stars displays evidence of significant contamination by Milky Way foreground stars (see Figure 2.3). For the i'th star we consider the measurements $(X_{1i}, X_{2i}, U_{3i}, \sigma_i)$, where (X_{1i}, X_{2i}) is the projected position of the star, U_{3i} is the observed line-of-sight velocity, and σ_i is the error associated with the measurement of U_{3i}. In this section we develop a method of addressing sample contamination that incorporates a model of the contaminant distribution. We would like to estimate the number of "signal" (Sextans) stars and construct a $1 - \alpha$ confidence interval. Our algorithm also outputs, for each observed star, an estimate of the probability that the star belongs to the contaminant population. These probability estimates can be used as weights in subsequent analysis. See Walker et al. (2008) for applications of this algorithm on data from other dwarf galaxies.

The Statistical Model

We assign parametric distributions to the positions and velocities of the stars; the parametric models are derived from the underlying physics in most cases. The EM algorithm is then employed to find MLE's estimates of the unknown parameters. The method is described in the context of available data, but can be generalized to incorporate membership constraints provided by additional data (such as multi-color photometry data).

Suppose $N \sim \text{Poi}(b+\theta)$ is the number of stars observed using the telescope in a given amount of time. In our case we observe $N = 528$. Here θ denotes the rate for observing a signal star, i.e., a Sextans star. We assume that the foreground rate is b. We are interested in constructing a $1-\alpha$ CI for θ. The actual line of sight velocity for the i'th star will be denoted by V_{3i}. Let U_{3i} be the observed velocity; the true velocity plus a normally-distributed error. We assume that $U_{3i} = V_{3i} + \epsilon_i$, where the $\epsilon_i \sim N(0, \sigma_i^2)$ and ϵ_i's are assumed independent. Let Y_i be the indicator of a foreground star, i.e., $Y_i = 1$ if the i'th star is a foreground star, and $Y_i = 0$ otherwise. Of course, we do not observe Y_i. We need to make assumptions on the form of the joint density of $W_i = (X_{1i}, X_{2i}, U_{3i})$.

For the foreground stars (i.e., $Y_i = 1$) it might be reasonable to assume that the position (X_{1i}, X_{2i}) and velocity U_{3i} are independent. Then the joint density of W_i simplifies to $h_b(w) = f^{(b)}(x_1, x_2)g^{(b)}(u_3)$, where we take the position of the star as uniformly distributed in the field of view, i.e., $f^{(b)}(x_1, x_2) = \frac{1}{\pi M^2}$, and M is the radius of field of view (in our data set it is 35 arc min). Note that $U_{3i} \sim g^{(b)}(\cdot)$, where $g^{(b)}$ is a completely known density obtained from the Besancon Milky Way model (Robin et al. 2003), which specifies spatial and velocity distributions of Milky Way stars along a given line of sight. The density estimate $g^{(b)}(\cdot)$ was constructed using kernel density estimation techniques.

For the Sextans stars, there is a well known model in Astronomy for the distribution of the projected position of stars. The model assumes $f^{(s)}(x_1, x_2) = K(h)e^{-s/h}, 0 \leq s^2 = x_1^2 + x_2^2 \leq M^2$, where $K(h)^{-1} = 2\pi h^2\{1 - (M/h)e^{-M/h} - e^{-M/h}\}$ is the normalizing constant (M is the radius of field of view). The distribution of U_{3i} given the position is assumed to be normal with mean μ and variance $\sigma^2 + \sigma_i^2$ and its density is denoted by $g^{(s)}(\cdot)$. Thus, the joint density of W_i given that it is a signal star is $h_{s,i}(w) = f^{(s)}(x_1, x_2)g^{(s)}(u_3)$.

CI for θ: the number of "signal" stars

The likelihood for the observed data is

$$L(\theta, \eta) = e^{-(b+\theta)} \frac{(b+\theta)^N}{N!} \prod_{i=1}^{N} \left(\frac{bh_b(W_i) + \theta h_{s,i}(W_i)}{b+\theta} \right) \tag{2.8}$$

which is a essentially a mixture density problem. A simple application of the EM algorithm (details are provided in the appendix) yields the MLE's in this scenario. The hybrid resampling method can be used to construct a confidence region for θ.

The likelihood ratio statistic is defined as in (2.1) and can be computed for each θ. The hybrid resampling method was employed to find the c_θ^+ as described in the introduction. Varying θ, we get a confidence interval for θ. In our example, the 90% confidence interval turns out to be $(260.3, 318.4)$. Note that if b was known and with $\hat{\theta} \approx 290$ (the maximum likelihood estimate of θ), a 90% CI using frequentist method (obtained by intersecting uniformly most accurate 95% confidence lower and upper bounds) would be (261.7, 318.4).

This shows that the hybrid method works almost as well as the most optimal frequentist confidence region, even when b is unknown.

2.3 Appendix

We outline the implementation of the EM-algorithm described in the last section, to find the the unconstrained maximum of the observed (incomplete) data. The constrained maximization is very similar (in fact, a bit simpler). Recall our notation,

$$Y_i = \begin{cases} 1 & \text{if the } i\text{'th star is a foreground star} \\ 0 & \text{o.w.} \end{cases} \quad , \forall i = 1, 2, \ldots, N$$

Note that Y_i's are i.i.d. Bernoulli $(\frac{b}{b+\theta})$. Let $\mathbf{Z} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{U}, \mathbf{Y}, N)$ be the complete data matrix. The likelihood for the complete data can be written as

$$\tilde{L}(\theta, \eta|\mathbf{Z}) = e^{-(b+\theta)} \frac{(b+\theta)^N}{N!} \left\{ \prod_{i=1}^{N} (\frac{b}{b+\theta})^{Y_i} (\frac{\theta}{b+\theta})^{1-Y_i} h_b(W_i)^{Y_i} h_{s,i}(W_i)^{1-Y_i} \right\}.$$

The log-likelihood (up to a constant term) can be written as

$$\tilde{l}(\theta, \eta|\mathbf{Z}) = -(b+\theta) + \sum_{i=1}^{N} \left\{ Y_i \log(b h_b(W_i)) + (1 - Y_i) \log(\theta h_{s,i}(W_i)) \right\}.$$

Letting θ_n and η_n denote the parameter values obtained in the n'th step of the iteration, the E-step in the unconstrained maximization process evaluates

$$\begin{aligned} E_{\hat{\theta}_n, \hat{\eta}_n} \left(\tilde{l}(\eta|\mathbf{Z}) | \mathbf{W} \right) &= \sum_{i=1}^{N} P_{\hat{\theta}_n, \hat{\eta}_n}(Y_i = 1|\mathbf{W}) \log[b h_b(W_i)] \\ &\quad + \sum_{i=1}^{N} P_{\hat{\theta}_n, \hat{\eta}_n}(Y_i = 0|\mathbf{W}) \log[\theta h_{s,i}(W_i)] - (b+\theta) \end{aligned} \tag{2.9}$$

where $P_{\hat{\theta}_n, \hat{\eta}_n}(Y_i = 1|\mathbf{W}) = \frac{\hat{b}_n h_b(W_i)}{\hat{b}_n h_b(W_i) + \hat{\theta}_n h_{s,i}(W_i)}$ is the probability of a foreground star given the data under the current estimates of θ and η, i.e., θ_n and η_n. The

M-step maximizes (2.9), which leads to the following estimating equations:

$$
\begin{aligned}
&\frac{1}{b}\sum_{i=1}^{N} P_{\hat{\theta}_n,\hat{\eta}_n}(Y_i = 1|\mathbf{W}) - 1 = 0,\\
&\frac{1}{\theta}\sum_{i=1}^{N} P_{\hat{\theta}_n,\hat{\eta}_n}(Y_i = 0|\mathbf{W}) - 1 = 0,\\
&\sum_{i=1}^{N} P_{\hat{\theta}_n,\hat{\eta}_n}(Y_i = 0|\mathbf{W}) \left\{ \frac{1}{\sigma^2+\sigma_i^2}(U_{3i}-\mu) \right\} = 0,\\
&\sum_{i=1}^{N} P_{\hat{\theta}_n,\hat{\eta}_n}(Y_i = 0|\mathbf{W}) \left\{ \frac{(U_{3i}-\mu)^2}{2(\sigma^2+\sigma_i^2)^2} - \frac{1}{2(\sigma^2+\sigma_i^2)} \right\} = 0.
\end{aligned}
$$

The first two equations can be solved easily to give estimates $\hat{b}_{n+1} = \sum_{i=1}^N P_{\hat{\theta}_n,\hat{\eta}_n}(Y_i = 1|\mathbf{W})$ and $\hat{\theta}_{n+1} = \sum_{i=1}^N P_{\hat{\theta}_n,\hat{\eta}_n}(Y_i = 0|\mathbf{W})$. The last two equations can be slightly modified to give the following (closed form) estimates of μ and σ^2:

$$
\hat{\mu}_{n+1} = \frac{\sum_{i=1}^N \frac{P_{\hat{\theta}_n,\hat{\eta}_n}(Y_i=0|\mathbf{W})}{1+\sigma_i^2/\hat{\sigma}^2_{(n)}} U_{3i}}{\sum_{i=1}^N \frac{P_{\hat{\theta}_n,\hat{\eta}_n}(Y_i=0|\mathbf{W})}{1+\sigma_i^2/\hat{\sigma}^2_{(n)}}} \text{ and } \hat{\sigma}^2_{(n+1)} = \frac{\sum_{i=1}^N \frac{P_{\hat{\theta}_n,\hat{\eta}_n}(Y_i=0|\mathbf{W})}{(1+\sigma_i^2/\hat{\sigma}^2_{(n)})^2} (U_{3i}-\hat{\mu}_{n+1})^2}{\sum_{i=1}^N \frac{P_{\hat{\theta}_n,\hat{\eta}_n}(Y_i=0|\mathbf{W})}{1+\sigma_i^2/\hat{\sigma}^2_{(n)}}}
$$

where $\hat{\sigma}^2_{(n)}$ is the n'th step estimate of σ^2. These estimates $(\hat{\eta}_n)$ stabilize after a few iterations yielding the MLE's of η with the incomplete data. An interesting feature of this solution is that at the end of the algorithm we get estimated probabilities that the i'th star is a signal star, namely, $P_{\hat{\theta}_n,\hat{\eta}_n}(Y_i = 1|\mathbf{W})$.

CHAPTER 3

A pseudo-likelihood method for analyzing interval censored data

We introduce a method based on a pseudo-likelihood ratio for estimating the distribution function of the survival time in a mixed-case interval censoring model. In a mixed case model, an individual is observed a random number of times, and at each time it is recorded whether an event has happened or not. One seeks to estimate the distribution of time to event. We use a Poisson process as the basis of a likelihood function to construct a pseudo-likelihood ratio statistic for testing the value of the distribution function at a fixed point, and show that this converges under the null hypothesis to a known limit distribution, that can be expressed as a functional of different convex minorants of a two-sided Brownian motion process with parabolic drift. Construction of confidence sets then proceeds by standard inversion. The computation of the confidence sets is simple, requiring the use of the pool adjacent violators algorithm, or a standard isotonic regression algorithm. We also illustrate the superiority of the proposed method over competitors based on resampling techniques or on the limit distribution of the maximum pseudo-likelihood estimator, through simulation studies, and illustrate the different methods on a data set involving time to HIV seroconversion in a group of haemophiliacs.

3.1 Introduction

For an interval censored observation, one only knows a window, that is, an interval, within which the survival event (time to infection/illness) occurred. Interval-censored failure time data occur in many areas including demography, epidemiology, financial, medical and sociological studies. In the mixed–case interval censoring model each individual is followed up at the clinic for a number of times, where this number and the times of inspection themselves can vary from individual to individual. It is determined between which two successive observation times the individual succumbed to infection/illness. It is of course possible that infection/illness may not occur by the last follow up time. The term "mixed–case" is used to indicate that the number of inspection times is patient specific, and was first used by Schick and Yu (2000). Our interest lies in constructing confidence sets for F, the distribution function of time to infection/illness (failure time).

3.1.1 Current status data

The simplest form of mixed–case censoring is current status data, where the number of observation times for each patient is exactly one; see for example, Groeneboom and Wellner (1992), Jewell and van der Laan (1995), Shiboski (1998), Banerjee and Wellner (2001, 2005) and Jewell et. al. (2003). In this model, the distribution of the indicator of time to infection/illness (failure time) S, conditional on the single inspection time U, is a Bernoulli random variable. Our data consists of the pair (δ, U), where $\delta = \mathbf{1}\{S \leq U\}$ is the indicator whether the "event" occurred before or after U. Note that this is different from the right-censorship model, as we are never able to observe the exact value of the survival time. Suppose that we have data available on n independent subjects $\{(\delta_i, U_i)\}_{i=1}^n$. Let S_i's denote the unobserved survival times of interest with distribution function F and assume that U_i is independent of S_i. Then the likelihood function has the form

$$L(F) = \prod_{i=1}^{n} F(U_i)^{\delta_i}\{1 - F(U_i)\}^{1-\delta_i}.$$

The above likelihood can be maximized over all distribution functions F (identified only at the U_i's) to give the nonparametric maximum likelihood estimate (NPMLE) $\hat{F}_n$. The NPMLE is readily computable using appropriately modified versions of the pool adjacent violators algorithm (Robertson et. al., 1988). It can be shown that pointwise $\hat{F}_n$ is a strongly consistent estimator of F and that $\hat{F}_n - F$ converges to a non-normal distribution (scaled Chernoff's distribution) at rate $n^{1/3}$.

3.1.2 Mixed case interval censoring

In mixed case interval censoring, the i'th individual with failure time S_i is observed at the random (ordered) time points $0 < Y_{i,1} < Y_{i,2} < \ldots < Y_{i,n_i}$, and it is recorded in which interval the individual succumbed to illness/infection. Under the assumption of independence between the observation times and the failure time, a similar likelihood analysis can be carried out to obtain the NPMLE of F, the distribution function of time to infection/illness.

In the current status model, the computations are based on explicit representations of the maximum likelihood estimates in terms of the given data and do not involve iterative schemes. However, maximization of the likelihood function in the mixed–case setting is much more complex and requires sophisticated optimization techniques. The EM can be employed but is extremely slow (Jongbloed, 1998); a faster algorithm is the modified iterative convex minorant algorithm of Jongbloed (1998), based on the Kuhn–Tucker conditions associated with the maximization problem. However, both methods involve iterating till convergence, and can therefore be quite slow. Alternative methods for computing nonparametric maximum likelihood estimators for interval censored data have been developed by Vandal et. al. (2005) using graph theoretic representations of the unconstrained and constrained estimators. These involve reduction techniques as well as versions of the EM algorithm and the Vertex Exchange Method. It is not known how these methods compare to the modified iterative convex minorant algorithm in terms of speed.

Banerjee and Wellner (2001) showed that in the current status model the likelihood ratio statistic for testing a pointwise hypothesis of the type $H_0 : F(t_0) = \theta_0$ for some pre-specified point t_0, is asymptotically pivotal under H_0. This immediately provides a way of constructing pointwise confidence bands for F by standard inversion of the likelihood ratio statistic, with the critical values determined by the quantiles of the limiting pivotal distribution. While this result is, in principle, generalizable to mixed–case interval censoring, dealing with the likelihood function in the mixed–case model is considerably more difficult, at both a theoretical and a computational level. Only partial results, in fairly restrictive settings, exist thus far, about the limiting behavior of the nonparametric maximum likelihood estimator; consequently, the limiting behavior of the likelihood ratio statistic for testing a pointwise null hypothesis is not tractable either; see for example Groeneboom (1996), where the asymptotics of the behavior of the nonparametric maximum likelihood estimator of

F in a particular version of the Case 2 censoring model is established, and Song (2004), where estimation procedures for mixed–case censoring models and associated issues are presented.

3.1.3 Our approach

We think of mixed-case interval censored data as data on a one-jump counting process with counts available only at the inspection times and to use a pseudo-likelihood function based on the marginal likelihood of a Poisson process to construct a pseudo-likelihood ratio statistic for testing null hypotheses of the form $H_0 : F(t_0) = \theta_0$. We show that under such a null hypothesis the statistic converges to a pivotal quantity. This result can now be used to construct confidence intervals for $F(t_0)$. The pseudo–likelihood method that we adopt is based on an estimator originally proposed by Sun and Kalbfleisch (1995) whose asymptotic properties, under appropriate regularity conditions, were studied in Wellner and Zhang (2000). Indeed, our key result in Section 3.2 draws freely on the work of Wellner and Zhang (2000) and our point of view here, the fact that the interval censoring situation can be thought of as a one-jump counting process to which, consequently, the results on the pseudo-likelihood based estimators can be applied, is motivated by their work. That said, our likelihood-ratio approach for computing confidence intervals has major advantages over the Wald type intervals that can be derived from their work.

3.2 A pseudo-likelihood method for analyzing mixed–case interval censored data

3.2.1 Notation

We now introduce the stochastic processes and derived functionals that are needed to describe the asymptotic distributions. For a real–valued function f defined on $\mathbb{R}$, let slogcm(f, I) denote the left–hand slope of the greatest convex minorant of the restriction of f to the interval I. We abbreviate slogcm$(f, \mathbb{R})$ to slogcm(f). Also define

$$\mathrm{slogcm}^0(f) = \{\mathrm{slogcm}\,(f, (-\infty, 0]) \wedge 0\}\, 1_{(-\infty,0]} + [\mathrm{slogcm}\,\{f, (0, \infty)\} \vee 0] 1_{(0,\infty)}\,.$$

For positive constants c and d define the process $X_{c,d}(z) = c\,W(z) + d\,z^2$, where $W(z)$ is standard two-sided Brownian motion starting from 0. Set $g_{c,d} = \mathrm{slogcm}(X_{c,d})$ and $g^0_{c,d} = \mathrm{slogcm}^0\,(X_{c,d})$. It is known that $g_{c,d}$ is a piecewise-constant increasing function, with finitely many jumps in any compact interval. The function $g^0_{c,d}$, has the same characteristics and differs, almost surely, from $g_{c,d}$ on a finite interval containing 0. In fact, with probability 1, $g^0_{c,d}$ is identically 0 in some random neighborhood of 0, whereas $g_{c,d}$ is almost surely nonzero in some random neighborhood of 0. Also, the length of the interval $D_{c,d}$ on which $g_{c,d}$ and $g^0_{c,d}$ differ is $O_p(1)$. For more detailed descriptions of the processes $g_{c,d}$ and $g^0_{c,d}$, see Banerjee and Wellner (2001), Wellner (2003), and Banerjee (2000). Thus, $g_{1,1}$ and $g^0_{1,1}$ are the unconstrained and constrained versions of the slope processes associated with the canonical process $X_{1,1}(z)$. By Brownian scaling, the slope processes $g_{c,d}$ and $g^0_{c,d}$ can be related in distribution to the canonical slope processes $g_{1,1}$ and $g^0_{1,1}$. This leads to the following lemma.

Lemma 3.1. *For positive a and b, set*

$$\mathbb{D}_{a,b} = \int \left[\{g_{a,b}(u)\}^2 - \{g^0_{a,b}(u)\}^2 \right] du$$

and abbreviate $\mathbb{D}_{1,1}$ to $\mathbb{D}$. Then $\mathbb{D}_{a,b}$ has the same distribution as $a^2\,\mathbb{D}$.

This is proved in Chapter 3 of Banerjee (2000); alternatively, see Banerjee and Wellner (2001).

3.2.2 The pseudo-likelihood estimator

We describe our method more broadly in the context of a counting process and then specialize to the interval censoring situation. Suppose that $N = \{N(t) : t \geq 0\}$ is a counting process with mean function $E\,N(t) = \Lambda(t)$, K is an integer-valued random variable and $T = \{T_{k,j}, j = 1, \ldots, k,\ k = 1, 2, \ldots\}$ is a triangular array of potential observation times. It is assumed that N and (K, T) are independent, that K and T are independent and $T_{k,j-1} \leq T_{k,j}$ for $j = 1, \ldots, k$, for every k; we interpret $T_{k,0}$ as 0. Let $X = (N_K, T_K, K)$ be the observed random vector for an individual. Here K is the number of times that the individual was observed during a study, $T_{K,1} \leq T_{K,2} \leq \ldots \leq T_{K,K}$ are the times when they were observed and $N_K = \{N_{K,j} \equiv N(T_{K,j})\}_{j=1}^{K}$ are the observed counts at those times. The above scenario specializes easily to the mixed–case interval censoring model, when the counting process is $N(t) = 1(S \leq t)$, S being a positive random variable with distribution function F and independent of (T, K).

Suppose that we have data on n individuals; thus, we observe n independent and identically distributed copies of X, say $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ where $X_i = (N^{(i)}_{K_i}, T^{(i)}_{K_i}, K_i), i = 1, \ldots, n$. Here $(N^{(i)}, T^{(i)}, K_i), i = 1, 2, \ldots,$ are the underlying independent and identically distributed copies of (N, T, K). We are interested in estimating the mean function $\Lambda(t)$ at a pre-specified point of interest t_0. Based on our data, we can construct a pseudo-likelihood estimator, in the following manner. Pretend that the process $N(t)$ is a nonhomogeneous Poisson process. Then the marginal distribution of $N(t)$ is

$$\text{pr}\,\{N(t) = k\} = \exp\,\{-\Lambda(t)\}\,\frac{\Lambda^k(t)}{k!},\quad k = 0, 1, 2, \ldots, \,.$$

Note that, under the Poisson process assumption, the successive counts on an individual $(N_{K,1}, N_{K,2}, \ldots)$, conditional on the $T_{K,j}$'s, are actually dependent. However we choose to ignore the dependence in writing down a likelihood function for the data, conditional on the $T^{(i)}$'s and the K_i's. These do not involve Λ and hence will not contribute to the estimation procedure. Our likelihood function is

$$L_n^{\text{ps}}(\Lambda \mid \underline{X}) = \prod_{i=1}^{n}\prod_{j=1}^{K_i} \exp\{-\Lambda(T_{K_i,j}^{(i)})\}\,\frac{\Lambda(T_{K_i,j}^{(i)})^{N_{K_i,j}^{(i)}}}{N_{K_i,j}^{(i)}!}\,.$$

Thus, the log-likelihood function, up to an additive constant not depending upon the parameter, is given by

$$l_n^{\text{ps}}(\Lambda \mid \mathbf{X}) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left\{N_{K_i,j}^{(i)}\,\log\,\Lambda(T_{K_i,j}^{(i)}) - \Lambda(T_{K_i,j}^{(i)})\right\}\,.$$

The above log-likelihood can be written in a slightly neater way, as follows: Let $T_{(1)} < T_{(2)} < \ldots < T_{(M)}$ denote the ordered distinct observation times in the set of all observation time points $\{T_{K_i,j}^{(i)}, j = 1, \ldots, K_i,\;\; i = 1, \ldots, n\}$. For $1 \le l \le M$, define

$$w_l = \sum_{i=1}^{n}\sum_{j=1}^{K_i} 1\,\{T_{K_i,j}^{(i)} = T_{(l)}\}\quad,\quad \overline{N}_l = \frac{1}{w_l}\sum_{i=1}^{n}\sum_{j=1}^{K_i} N_{K_i,j}^{(i)}\,1\,\{T_{K_i,j}^{(i)} = T_{(l)}\}\,.$$

Thus w_l is the frequency of the l'th largest observation time in the sample and $w_l\,\overline{N}_l$ is the total number of events that happened by the lth largest time.

Writing $\Lambda(T_{(l)})$ as Λ_l, for convenience, we can represent the log-likelihood as

$$l_n^{\mathrm{ps}}(\Lambda \mid \mathbf{X}) = \sum_{l=1}^{M} \left(w_l \, \overline{N}_l \, \log \Lambda_l - w_l \, \Lambda_l \right) . \tag{3.1}$$

We define the nonparametric estimator $\hat{\Lambda}_n$ of Λ to be the unique nondecreasing right–continuous step-function with possible jumps only occurring at the $T_{(i)}$'s, such that the above expression is maximized. Of course, only $\Lambda_1, \ldots, \Lambda_M$ are identifiable; the choice of $\hat{\Lambda}_n$ made above is arbitrary. Other conventions are possible, but will make no difference to the asymptotics. Thus, $\hat{\Lambda}_n$, which we will subsequently refer to as $\hat{\Lambda}$ for convenience, is the unconstrained maximum pseudo-likelihood estimator. The constrained estimator $\hat{\Lambda}_n^{(0)}$, to be referred to subsequently as $\hat{\Lambda}^{(0)}$, is defined to be the unique nondecreasing step-function with possible jumps only at the $T_{(i)}$'s and at t_0, that maximizes (3.1) subject to the additional constraint that $\Lambda(t_0) = \theta_0$. Using the theory of generalized isotonic regression (Robertson et. al., 1988, Section 1.5), or by appealing to the Kuhn-Tucker theorem (Robertson et. al., 1988, Section 6.4), we can show that $\hat{\Lambda}(T_{(i)})$ is $\hat{f}_i$, where $(\hat{f}_1, \ldots, \hat{f}_M)$ minimizes $\sum_{i=1}^{M} w_i \, (g_i - f_i)^2$ over all $f_1 \leq \ldots \leq f_M$, with $g_i \equiv \overline{N}_i$ and w_i as defined above. Also, $\hat{\Lambda}^{(0)}(T_{(i)})$ is $\hat{f}_i^{(0)}$, where $(\hat{f}_1^{(0)}, \hat{f}_2^{(0)}, \ldots, \hat{f}_M^{(0)})$ solves the constrained isotonic least squares problem of minimizing $\sum_{i=1}^{M} w_i \, (g_i - f_i)^2$ over all $f_1 \leq \ldots \leq f_m \leq \theta_0 \leq f_{m+1} \leq \ldots \leq f_M$, with $T_{(m)} < t_0 < T_{(m+1)}$. The fact that none of the $T_{(i)}$'s can actually be equal to t_0, with probability 1, is guaranteed by the regularity conditions under which the asymptotic results for this model will be established; in particular, see assumptions 6 and 7 in the Appendix.

For points $\{(x_0, y_0), (x_1, y_1), \ldots, (x_k, y_k)\}$ where $x_0 = y_0 = 0$ and $x_0 < x_1 < \ldots < x_k$, consider the left-continuous function $P(x)$ such that $P(x_i) = y_i$ and such that $P(x)$ is constant on (x_{i-1}, x_i). We will denote the vector of slopes, i.e., left–derivatives, of the greatest convex minorant of $P(x)$ computed at the points $(x_1, x_2, \ldots, x_k)$ by slogcm $\{(x_i, y_i)\}_{i=0}^{k}$.

It is not difficult to see that

$$\{\hat{\Lambda}_i\}_{i=1}^{M} = \text{slogcm}\left\{\sum_{j=1}^{i} w_j, \sum_{j=1}^{i} w_j\,\overline{N}_j\right\}_{i=0}^{M}, \tag{3.2}$$

where summation over an empty set is interpreted as 0. Also,

$$\{\hat{\Lambda}_i^{(0)}\}_{i=1}^{m} = \theta_0 \wedge \text{slogcm}\left\{\sum_{j=1}^{i} w_j, \sum_{j=1}^{i} w_j\,\overline{N}_j\right\}_{i=0}^{m}, \tag{3.3}$$

where the minimum is interpreted as being taken componentwise, while

$$\{\hat{\Lambda}_i^{(0)}\}_{i=m+1}^{M} = \theta_0 \vee \text{slogcm}\left\{\sum_{j=m+1}^{i} w_j, \sum_{j=m+1}^{i} w_j\,\overline{N}_j\right\}_{i=m}^{M}, \tag{3.4}$$

where the maximum is once again interpreted as being taken componentwise.

3.2.3 Asymptotic results

Define the pseudo-likelihood ratio statistic as

$$2\,\log\,\lambda_n = 2\left\{l_n^{\text{ps}}(\hat{\Lambda} \mid \mathbf{X}) - l_n^{\text{ps}}(\hat{\Lambda}^{(0)} \mid \mathbf{X})\right\}.$$

The limit distribution of $2\,\log\,\lambda_n$ will be established under a number of regularity conditions. These are minor modifications of conditions given in Wellner and Zhang (2000), but for the sake of completeness, we state them in the Appendix and there discuss the implications of these conditions in the interval censoring framework.

Under Assumptions A1 – A4, there exist $a_0 < t_0 < b_0$ such that

$$\sup_{x\in[a_0,b_0]} \mid \hat{\Lambda}_n(x) - \Lambda(x) \mid \to 0 \ \text{ almost surely}.$$

Also, if the null hypothesis holds,

$$\sup_{x\in[a,b]} \mid \hat{\Lambda}_n^{(0)}(x) - \Lambda(x) \mid \to 0 \ \text{ almost surely}.$$

This consistency result will not be established here.

We now state the main result of this chapter, which concerns the limiting behavior of $2\,\log\,\lambda_n$.

Theorem 1. Under Assumptions A1 – A9, the pseudo-likelihood ratio statistic,

$$2\,\log\,\lambda_n \equiv 2\,\left\{ l_n^{\text{ps}}(\hat{\Lambda}_n \mid \mathbf{X}) - l_n^{\text{ps}}(\hat{\Lambda}_n^{(0)} \mid \mathbf{X}) \right\} \xrightarrow{d} \frac{\sigma^2(t_0)}{\Lambda(t_0)}\,\mathbb{D}\,,$$

when $H_0 : \Lambda(t_0) = \theta_0$ holds.

A sketch proof of this theorem is given in the Appendix and uses the following theorem on the limit distribution of the nonparametric maximum likelihood estimators of Λ. Define

$$X_n(z) = n^{1/3}\,\left\{ \hat{\Lambda}_n(t_0 + z\,n^{-1/3}) - \theta_0 \right\} \text{ and } Y_n(z) = n^{1/3}\,\left\{ \hat{\Lambda}_n^{(0)}(t_0 + z\,n^{-1/3}) - \theta_0 \right\}\,.$$

Theorem 2. Suppose that Assumptions A1 – A9 hold and set

$$a = \left(\frac{\sigma^2(t_0)}{G'(t_0)} \right)^{1/2}\,, \quad b = \frac{1}{2}\,\Lambda'(t_0)\,.$$

Then, under H_0,

$$(X_n(z), Y_n(z)) \xrightarrow{d} \left(g_{a,b}(z), g_{a,b}^0(z) \right)\,,$$

finite–dimensionally and also in the space $\mathcal{L} \times \mathcal{L}$, where $\mathcal{L}$ is the space of functions from $\mathbb{R} \to \mathbb{R}$ that are bounded on every compact set, equipped with the topology of L_2-convergence with respect to Lebesgue measure on compact sets.

3.2.4 Construction of Confidence sets

Theorem 1 gives an easy way of constructing a likelihood-ratio based confidence set for $F(t_0)$ in the mixed–case interval censoring model. This is based on the observation that under the mixed–case interval censoring framework, where the counting process $N(t)$ is $1(S \le t)$ with S following distribution F independently of (K, T), the pseudo-likelihood ratio statistic in Theorem 1 converges to $(1-\theta_0)\,\mathbb{D}$ under the null hypothesis $F(t_0) = \theta_0$. Thus, $(1-\theta_0)^{-1}\,2\,\log\,\lambda_n$ converges in distribution to $\mathbb{D}$, so that an asymptotic level-$(1-\alpha)$ confidence set for $F(t_0)$ is given by

$$\{\theta : (1-\theta)^{-1}\,2\,\log\,\lambda_n(\theta) \le q(\mathbb{D}, 1-\alpha)\}$$

, where $q(\mathbb{D}, 1-\alpha)$ is the $(1-\alpha)$'th quantile of $\mathbb{D}$ and $2\,\log\,\lambda_n(\theta)$ is the pseudo-likelihood ratio statistic computed under the null hypothesis $H_{0,\theta} : F(t_0) = \theta$. Thus, finding the confidence set amounts to computing the likelihood ratio under a family of null hypotheses. The computation is a simple affair and can be done through using elementary pool adjacent violators algorithm. Quantiles of $\mathbb{D}$ are tabulated in Banerjee and Wellner (2001).

Theorem 4.3 of Wellner and Zhang (2000) can also be derived as a special case of Theorem 2 by setting $z = 0$. Specialized to the mixed–case censoring scenario, it provides an alternative route to constructing confidence sets for $F(t_0)$. Denoting by $\hat{F}_n$ the pseudo-likelihood estimate of F, from Theorem 4.3 of Wellner and Zhang (2000), we obtain

$$n^{1/3}\,\{\hat{F}_n(t_0) - F(t_0)\} \xrightarrow{d} \left\{\frac{\theta_0\,(1-\theta_0)\,f(t_0)}{2\,G'(t_0)}\right\}^{1/3}\,2\,\mathbb{Z}\,, \tag{3.5}$$

where $\mathbb{Z} = \text{argmin}_h\,\{W(h) + h^2\}$ and $f(t)$ is the derivative of $F(t)$. An approximate level-$(1-\alpha)$ confidence interval for $F(t_0)$ is

$$\left[\hat{F}_n(t_0) - 2\,C_n\,q(\mathbb{Z}, 1-\alpha/2)\;,\;\hat{F}_n(t_0) + 2\,C_n\,q(\mathbb{Z}, 1-\alpha/2)\right]$$

where $q(\mathbb{Z}, 1-\alpha/2)$ is the $(1-\alpha/2)$'th quantile of $\mathbb{Z}$ and

$$C_n = n^{-1/3}\,\left[\frac{\hat{F}_n(t_0)\,\{1-\hat{F}_n(t_0)\}\,\hat{f}(t_0)}{2\,\hat{G}'(t_0)}\right]^{1/3}\,,$$

with $\hat{f}$ and $\hat{G}'$ denoting estimators of f and G' respectively. Quantiles of $\mathbb{Z}$ are tabulated in Groeneboom and Wellner (2001). Estimating G' involves estimating first the probability density of K and then the marginal densities of the $T_{k,j}$'s; this can be done using kernel density methods with some optimal bandwidth selection procedure like least-squares cross-validation (Loader, 1999). However, it is not difficult to see that, if K assumes a large number of values and the sample size n is moderate, there may not be sufficiently many observations to estimate the density of each $T_{k,j}$ reliably. Finally there is also the problem of estimating f, which is a trickier affair, since observations from the distribution F are not available. A discussion of the issues involved in a similar situation can be found in Banerjee and Wellner (2005). In Section 3.3, we estimate f by kernel smoothing the maximum pseudo-likelihood estimator $\hat{F}_n$, as in Banerjee and Wellner (2005), using a likelihood-based cross-validation criterion. The procedure followed is analogous to the one described in Section 3.1 of that paper, the only difference being that the likelihood used for cross-validation here is the pseudo-likelihood, as opposed to the current status likelihood used in that paper.

Thus, the estimation of nuisance parameters turns out to be the major concern in the Wald-based approach: the variability introduced through nuisance parameter estimation will tend to make the confidence intervals much more unreliable, especially at smaller sample sizes. The likelihood ratio based method, on the other hand, does not involve nuisance parameter estimation and provides an extremely clear–cut way of constructing confidence intervals for $F(t_0)$. This makes it a much more attractive option. Yet another method of obtaining confidence sets is via the use of subsampling techniques. In view of the nonstandard asymptotics involved, as manifested in the cube-root convergence of the pseudo-likelihood estimator to a non-Gaussian limit, the usual bootstrap is suspect, but subsampling without replacement works. Subsampling was implemented by drawing a large number of subsamples of size b from the original sample, without replacement, and estimating the limiting quantiles of $|n^{1/3}\{\hat{F}_n(t_0) - F(t_0)\}|$, using the empirical distribution of $|b^{1/3}\{\hat{F}_n^*(t) - \hat{F}_n(t)\}|$; here $\hat{F}_n^*(t)$ denotes the value of the maximum pseudo-likelihood estimator, based on the subsample. For consistent estimation of the quantiles, b/n should converge to 0 as n increases. In the literature, b is referred to as the block–size. For details, see the book Politis, Romano and Wolf (1999, Chapter 2). The choice of b can affect the precision of the confidence intervals in finite samples. A data-driven choice of b is often resorted to but can be computationally very intensive. For a discussion of subsampling in the context of an interval censored model, see Sections 2 and 3 of Banerjee and Wellner (2005). Since the issues in the present case are similar, we do not go into an exhaustive discussion here.

We note in closing that the pseudo-likelihood based method for constructing confidence sets at a single point can be extended to finitely many points of interest; here the relevant limit distribution is the maximum of k independent copies of $\mathbb{D}$, where k is the number of points. However, the construction of likelihood based simultaneous confidence bands for F is still an open problem.

3.3 Simulation Studies and Data Analysis

3.3.1 Simulation Studies

We present simulations from a mixed–case censoring model, in which the survival time distribution X was taken to follow the Exponential(1) distribution. The random number K of observation times for an individual was generated from the uniform distribution on the integers $\{1, 2, 3, 4\}$ and given $K = k$, the observation times $\{T_{k,i}\}_{i=1}^{k}$ were chosen as k order statistics from the uniform distribution on (0,3). We generated 1000 replicates for each sample size displayed in Table 3.1, and 95% confidence intervals for $F(\log 2) = 0.5$ were computed by the three different methods: (i) pseudo-likelihood ratio, (ii) limit distribution of the maximum pseudo-likelihood estimator with kernel based estimation of nuisance parameters, (iii) subsampling with appropriate block–size. Kernel based estimation was done in the way described in connection with the construction of confidence sets for $F(t_0)$ in Section 3.2. For the subsampling based intervals, we did not resort to a data–driven block–size selection algorithm, since this would have increased computational complexity by orders of magnitude. Since the data generating process here is known, we generated separate data sets (1000 replicates) from the mixed–case model for each sample size, and computed subsampling based intervals for $F(t_0) = 0.5$ using a selection of block–sizes. We then computed the empirical coverage of the 1000 confidence intervals produced for each block–size, and chose the optimal block–size for the simulations presented here, as the one for which the empirical coverage was closest to 0.95. Thus, block–size selection was done via pilot simulations. Of course, this is not doable in a real life setting, since the data generating process is unknown. A natural way to circumvent this problem for real data sets is using the bootstrap to generate 'pilot data' from the empirical measure of the observed data and choose the block size based on the bootstrapped samples. This idea from Delgado et.al. (2003) is used in the next subsection, where the methods are illustrated on a real data set. The results are reported in Table 3.1.

From Table 3.1 we see that the pseudo-likelihood method produces the

Table 3.1: Simulation study for mixed–case interval censoring model: Average length (AL) and empirical coverage (C) of asymptotic 95% confidence intervals using pseudo-likelihood ratio (PL), maximum pseudo-likelihood (PMLE) and subsampling based (SB) methods.

	PL		PMLE		SB	
n	AL	C	AL	C	AL	C
50	0.410	0.904	0.441	0.867	0.538	0.971
100	0.327	0.920	0.353	0.896	0.469	0.972
200	0.261	0.924	0.282	0.899	0.308	0.958
500	0.198	0.949	0.210	0.923	0.242	0.958
1000	0.157	0.938	0.167	0.914	0.174	0.945
1500	0.136	0.936	0.144	0.933	0.158	0.962
2000	0.124	0.943	0.131	0.921	0.144	0.965

narrowest confidence intervals on an average. While they tend to be anti-conservative, the coverage nevertheless is quite satisfactory, being greater than or close to 94%, provided the sample size is moderately large. The subsampling-based intervals are the widest, and not surprisingly conservative in general. The kernel based intervals perform quite poorly at lower sample sizes, being extremely anti-conservative but also giving wider confidence intervals than the likelihood ratio, and they remain anti-conservative at higher sample sizes as well. The overall picture indicates the superiority of our pseudo-likelihood–ratio method. This, added to the relative computational simplicity of our method in comparison to its competitor, where once needs to content with the choice of a smoothing parameter or block–size, makes it an attractive choice.

3.3.2 Illustration on a real data set

De Gruttola and Lagakos (1989) present an interval censored data set of the time to HIV infection in a group of haemophiliacs. Since 1978, 262 people with Type A or B haemophilia had been treated at Hôpital Kremlin Bicêtre and Hôpital Coeur des Yvelines in France. Twenty-five of them were found to be infected on their first test for infection. By August 1988, 197 had become infected and 43 of these had developed some clinical symptoms relating to their HIV infection. All the infected persons are believed to have become infected by contaminated blood factor that they received for their haemophilia.

For each patient, the only information available is that $X \in [X_L, X_R]$, where X denotes the time to infection. Here time is measured in 6-month intervals, with $X = 1$ denoting July 1, 1978. An individual was assigned $X_L = 1$ if they were found to be infected with HIV on their first test for infection. As mentioned above, there were 25 such individuals. For details see Section 6 of De Gruttola and Lagakos (1989), and their Table 1, where the (X_L, X_R) values for each patient are provided. We are interested in estimating the distribution of X, the time to infection, based on the (X_L, X_R) pairs. We do the analysis separately for the two different groups into which the patients fell: the heavily–treated group of 105 patients received at least 1000 μg/kg of blood factor for at least one year between 1982 and 1985, and the lightly–treated group of 157 patients received less than 1000 μg/kg of blood factor per year.

We model the data as Case 2 censored data. The two censoring times U and V, with $U < V$ are defined as follows. If $1 = X_L < X_R < \infty$, we set $U = X_R$ and V to be the time till the end of the study. If $1 < X_L < X_R < \infty$, we set $U = X_L$ and $V = X_R$. If $1 < X_L < X_R = \infty$, we set $U = 1$ and $V = X_L$. If $(\Delta_1, \Delta_2, \Delta_3)$ denotes the vector of indicators, with $\Delta_1 = 1(X \leq U), \Delta_2 = 1(U < X \leq V), \Delta_3 = 1(V \leq X)$, then for the first case this vector is $(1, 0, 0)$, for the second case it is $(0, 1, 0)$ and in the third case it is $(0, 0, 1)$. The given data set is really an example of mixed–case censoring in which only the relevant inspection times have been noted. The formulation of the problem as a Case 2 model is a simplification that we adopt for the purpose of illustrating our method; because of lack of information about the other inspection times, the full mixed–case model cannot be fitted to the data.

The pseudo-likelihood estimate of F, the distribution function of X, was computed for each of the two groups, and confidence intervals for the values of F at several different points were obtained using the three different methods illustrated in the simulation studies. The subsampling-based confidence intervals at any given point was computed by first determining the block–size b using the bootstrap-based block selection algorithm referred to in the previous subsection; see Banerjee and Wellner (2005) for a brief description and an application of this algorithm to current status data. Five hundred bootstrap samples were used for block-size selection, and once the optimal block size had been ascertained 1000 subsamples of that size were used to determine the confidence interval. As far as the estimation of nuisance parameters for the construction of the Wald–type confidence interval was concerned, $f(t_0)$ at a point of interest t_0 was computed by smoothing the maximum pseudo-likelihood estimator using bandwidth determined by likelihood–based cross-validation, as for the simulation experiments. However, least–squares cross-validation, for choosing the optimal bandwidths to estimate $G'(t_0)$, did not perform well, and therefore G' was estimated by differentiating the piecewise-linear modification of the empirical distribution functions of U and V.

The estimated distribution functions of the time to infection are plotted for the two different groups in Figure 3.1. The distribution function for the heavily–treated group dominates that for the lightly–treated group in the interval $[6, 14)$; between 14 and 16, the distribution function for the lightly–treated

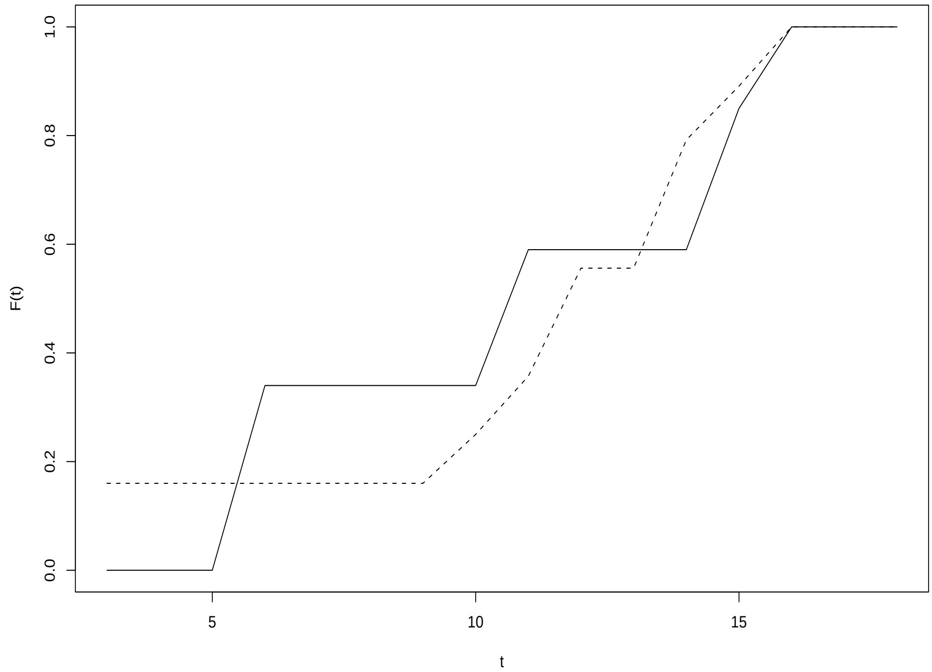

Figure 3.1: HIV infection data. The estimated distribution functions of time to HIV infection in the two different groups; heavily treated, solid line; lightly treated, dashed line.

group is higher; at 16, the two distributions coincide at the value 1. Individuals in the heavily–treated group received higher amounts of blood factor for at least a year between 1982 and 1985; the higher the amount of blood transfusion, the greater is the chance of infection through contaminated blood factor. The date of July 1, 1982 corresponds to $t = 9$, and $t = 16$, where the two distribution functions coincide, corresponds to January 1, 1986. In the range $9 - 16$, the distribution function for the heavily treated group is either equal to or almost equal to that for the lightly treated group or dominates it, except in the range $[14, 15)$; this corresponds to the year 1985.

Tables 3.2 and 3.3 give confidence intervals at different time points obtained by the three different methods. For each table, the second column gives the value of the maximum pseudo-likelihood estimator, the third gives the confidence intervals using the the pseudo-likelihood ratio, the fourth the Wald-type intervals and the fifth, the subsampling–based intervals. Note that the left extremities of the confidence intervals for the distribution function in the heavily–treated group are generally shifted to the right of those for the corresponding time points in the lightly–treated group, with violations towards the end of the table. The general shift of the left extremities to the right is predictable. The violation of this property towards the end of the table is not surprising, since there we are dealing with the time range in which the two distribution functions are essentially 'catching up' with each other, as is evident from Fig. 3.1. Also note that the likelihood ratio based confidence intervals are somewhat less erratic than the two other intervals; they exhibit monotonicity of left as well as right endpoints with increasing t. Since F is monotone in t, this is a rather nice property. On the other hand, the Wald–type or the subsampling–based intervals tend to exhibit violations of this property, though there is an overall monotonic trend.

3.4 Appendix: Technical details

We first formally state the required assumptions.

Assumption A1. The observation times $T_{k,j}$, for $j = 1, \ldots, k$ and $k = 1, 2, \ldots$, are random variables taking values in the bounded set $[0, \tau]$, where $0 < \tau < \infty$ and $E(K) < \infty$.

Assumption A2. The mean function Λ satisfies $\Lambda(\tau) \leq M$ for some $0 < M < \infty$.

Assumption A3. The random variable M_0 defined as $M_0 = \sum_{j=1}^{K} N_{K,j} \log N_{K,j}$ satisfies $E(M_0) < \infty$. Here, interpret $0 \log 0$ as 0.

Table 3.2: Confidence intervals (C.I.) of three kinds for the distribution of time to HIV infection in lightly treated group at different times: likelihood ratio based (lrt), Wald type (Wald) and subsampling based (subsampling).

t	$\hat{F}(t)$	C.I.(lrt)	C.I. (Wald)	C.I.(subsampling)
6.0	.160	0.068-0.285	0.000-0.354	0.068-0.252
7.0	.160	0.068-0.285	0.000-0.409	0.046-0.274
8.0	.160	0.068-0.298	0.000-0.410	0.042-0.278
9.0	.160	0.069-0.321	0.000-0.379	0.026-0.294
10.0	.250	0.069-0.458	0.000-0.463	0.048-0.451
11.0	.357	0.099-0.546	0.000-0.623	0.174-0.540
12.0	.556	0.187-0.660	0.381-0.730	0.396-0.716
13.0	.556	0.402-0.700	0.277-0.834	0.361-0.750
14.0	.792	0.439-0.888	0.553-1.000	0.660-0.923
15.0	. 891	0.637-0.943	0.712-1.000	0.786-0.996

Table 3.3: Confidence intervals (C.I.) of three kinds for the distribution of time to HIV infection in heavily treated group at different times: likelihood ratio based (lrt), Wald type (Wald) and subsampling based (subsampling).

t	$\hat{F}(t)$	C.I.(lrt)	C.I. (Wald)	C.I.(subsampling)
6.0	.340	0.000-0.442	0.067-0.613	0.087-0.593
7.0	.340	0.092-0.442	0.171-0.509	0.220- 0.459
8.0	.340	0.240-0.442	0.113-0.567	0.184-0.496
9.0	.340	0.240-0.442	0.179-0.501	0.213-0.467
10.0	.340	0.240-0.451	0.206-0.474	0.213-0.467
11.0	.588	0.242-0.665	0.437-.739	0.459-0.717
12.0	.588	0.472-0.665	0.451-0.725	0.490-0.686
13.0	.588	0.484-0.673	0.462-0.715	0.496-0.680
14.0	.588	0.504-0.676	0.450-0.727	0.478-0.699
15.0	. 852	0.504-0.927	0.751-0.953	0.740-0.964

For Borel subsets B of $[0,\tau]$, define the measure μ as

$$\mu(B) = E\left\{\sum_{j=1}^{K} 1\left(T_{K,j} \in B\right)\right\}.$$

Let $G(t) \equiv \mu((0,t])$ be the distribution function corresponding to the measure μ. For each k, j, denote the distribution function of the random variable $T_{k,j}$ by $G_{k,j}$. Then

$$\begin{aligned} G(t) &= E\left(\sum_{j=1}^{K} 1\left\{T_{K,j} \le t\right\}\right) \\ &= \sum_{k=1}^{\infty} \text{pr}\ (K=k) \sum_{j=1}^{k} \text{pr}\ \left(T_{k,j} \le t | K=k\right) \end{aligned}$$

$$= \sum_{k=1}^{\infty} \text{pr}\ (K=k) \sum_{j=1}^{k} G_{k,j}(t)\,. \tag{A1}$$

Call x a support point of μ if, for every $\epsilon > 0$, it is the case that $\mu(x-\epsilon, x+\epsilon) > 0$. Let S_μ denote the set of all support points of μ.

Assumption A4. The point t_0 lies in the interior of S_μ.

Assumption A5(a). The variable K has a finite moment of order greater than 2.

Assumption A5(b). There exist $\alpha > 0$ and $M_1 > 0$ such that $E\{N^{2+\alpha}(t)\} \le M_1$ for all $t \in S_\mu$.

Assumption A6. There is a neighborhood $\mathcal{U}$ of $t_0 \in S_\mu$ such that the distribution functions $G_{k,j}$ have positive continuous derivatives on $\mathcal{U}$, which are bounded by a common constant B for all k, j.

Assumption A7. There is a neighborhood $\mathcal{V}$ of $(t_0, t_0) \in \mathbb{R}^2$ such that, for all $k = 1, 2, \ldots$ and $1 \le i \le j \le k$, $G_{k,i,j}(s,t) =\text{pr}(T_{k,i} \le s, T_{k,j} \le t)$ is differentiable with respect to (s,t) and $g_{k,i,j}(s,t) = \partial^2 G_{k,i,j}(s,t)/\partial s\, \partial t$ exists. Furthermore, the functions $g_{k,i,j}$ are bounded on $\mathcal{V}$, by a common constant C, for all (k,i,j).

Assumption A8. The mean function Λ has a continuous bounded derivative on $\mathcal{U}$.

Assumption A9. The function $\sigma^2(t) \equiv \mathrm{var}\{N(t)\}$ is continuous in a neighborhood of t_0.

We discuss the implications of our assumptions in the interval censoring framework. Assumption *A2* is trivially satisfied in the interval censoring situation, since $0 \leq F(t) = \Lambda(t) \leq 1$. Assumption *A3* is also easy to check; in the interval censored situation, $N_{k,j}$ is either 1 or 0, so that $M_0 = 0$. In so far as estimation at the point t_0 is concerned, it suffices to have a positive Lebesgue density for one of the $T_{k,j}$'s in a neighborhood of the point t_0, along with $\mathrm{pr}(K = k) > 0$, for Assumption *A4* to be satisfied. Assumption *A5* is guaranteed for a K that is finitely supported, which is typically the case in applications, and for the interval censoring situation, since $N(t) \leq 1$. Assumption *A8*, in the interval censoring scenario, translates to $F(t)$ being continuously differentiable in a neighborhood of t_0 with $f(t_0) \neq 0$. Finally, Assumption *A9* is easily satisfied, since $\sigma^2(t) = F(t)\,(1 - F(t))$.

We first define the following processes:

$$\begin{aligned} V_n(t) &= \mathbb{P}_n\left(\sum_{j=1}^{K} N_{K,j}\, 1\left\{T_{K,j} \leq t\right\}\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i} N_{K_i,j}^{(i)}\, 1\left\{T_{K_i,j}^{(i)} \leq t\right\}, \\ G_n(t) &= \mathbb{P}_n\left(\sum_{j=1}^{K} 1\left\{T_{K,j} \leq t\right\}\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i} 1\left\{T_{K_i,j}^{(i)} \leq t\right\}. \end{aligned}$$

Thus, both V_n and G_n are piecewise constant right-continuous processes, with possible jumps only at the distinct observation times; the jump of G_n at the point $T_{(l)}$ is simply w_l/n, whereas the jump of V_n at the same point is $w_l\,\overline{N}_l/n$.

Also, set

$$\xi_1(X,t) = \sum_{j=1}^{K} N_{K,j}\, 1\,\{T_{K,j} \le t\} \quad \text{and} \quad \xi_0(X,t) = \sum_{j=1}^{K} 1\,\{T_{K,j} \le t\}\;.$$

Note that $G(t) = E\{\xi_0(X,t)\}$. Also, define $V(t) = E\{\xi_1(X,t)\}$. From (A1) obtain that $G^{'}(t) = \sum_{k=1}^{\infty} P(K=k)\sum_{j=1}^{k} G_{k,j}^{'}(t)\,.$ Also,

$$\begin{aligned} V(t) &= E\left(\sum_{j=1}^{K} N_{K,j}\, 1\,\{T_{K,j} \le t\}\right) \\ &= \sum_{k=1}^{\infty} P(K=k) \sum_{j=1}^{k} E\left[N_{k,j}\, 1\{T_{k,j} \le t\}\right] \\ &= \sum_{k=1}^{\infty} P(K=k) \sum_{j=1}^{k} \int_0^t \Lambda(x)\, d\,G_{k,j}(x) \\ &= \int_0^t \Lambda(x) \left(\sum_{k=1}^{\infty} P(K=k) \sum_{j=1}^{k} G_{k,j}^{'}(x)\right) dx \\ &= \int_0^t \Lambda(x)\, G'(x)\, dx\,, \end{aligned}$$

whence

$$\text{(A2)} \qquad V^{'}(t) = \Lambda(t)\, G^{'}(t)\,.$$

Proof of Theorem 1. In the following derivation, we denote by $\hat{\Lambda}_l$ the value of the unconstrained estimator $\hat{\Lambda}_n$ at the point $T_{(l)}$, and by $\hat{\Lambda}_l^{(0)}$ the value of $\hat{\Lambda}_n^{(0)}$ at the point $T_{(l)}$. The likelihood ratio statistic is then given by

$$\begin{aligned} 2\,\log\lambda_n &= 2\sum_{l=1}^{M} w_l\,(\overline{N}_l\,\log\hat{\Lambda}_l - \hat{\Lambda}_l) - 2\sum_{l=1}^{M} w_l\,(\overline{N}_l\,\log\hat{\Lambda}_l^{(0)} - \hat{\Lambda}_l^{(0)}) \\ &= 2\sum_{l=1}^{M} w_l\,\overline{N}_l\,(\log\hat{\Lambda}_l - \log\hat{\Lambda}_l^{(0)}) - 2\sum_{l=1}^{M} w_l\,(\hat{\Lambda}_l - \hat{\Lambda}_l^{(0)})\,. \end{aligned}$$

In what follows, we assume that the null hypothesis holds, so that $\Lambda(t_0) \equiv \theta_0$. We will also denote the set of indices for which $\hat{\Lambda}_l$ differs from $\hat{\Lambda}_l^{(0)}$ by D. On

Taylor expansion of $\log\hat\Lambda_l$ and $\log\hat\Lambda_l^{(0)}$ around θ_0, we obtain $2\log\lambda_n$

$$\begin{aligned} &= 2\sum_{l\in D} w_l\,\overline{N}_l\left\{\log\theta_0+\frac{1}{\theta_0}(\hat\Lambda_l-\theta_0)-\frac{1}{2\,\theta_0^2}(\hat\Lambda_l-\theta_0)^2+\frac{1}{3\,\Lambda_{l,*}^3}(\hat\Lambda_l-\theta_0)^3-\log\theta_0\right.\\ &\quad\left.-\frac{1}{\theta_0}(\hat\Lambda_l^{(0)}-\theta_0)+\frac{1}{2\,\theta_0^2}(\hat\Lambda_l^{(0)}-\theta_0)^2-\frac{1}{3\,\Lambda_{l,**}^3}(\hat\Lambda_l-\theta_0)^3\right\}-2\sum_{l\in D} w_l\,(\hat\Lambda_l-\hat\Lambda_l^{(0)})\,. \end{aligned}$$

Here $\hat\Lambda_{l,*}$ is a point intermediate between $\hat\Lambda_l$ and θ_0, and $\hat\Lambda_{l,**}$ is a point intermediate between $\hat\Lambda_l^{(0)}$ and θ_0. The above expression simplifies to $2\log\lambda_n$

$$\begin{aligned} &= 2\sum_{l\in D} w_l\,\overline{N}_l\,\frac{1}{\theta_0}\left\{(\hat\Lambda_l-\theta_0)-(\hat\Lambda_l^{(0)}-\theta_0)\right\}-2\sum_{l\in D} w_l\left\{(\hat\Lambda_l-\theta_0)-(\hat\Lambda_l^{(0)}-\theta_0)\right\}\\ &\quad-\frac{1}{\theta_0^2}\sum_{l\in D}\left\{(\hat\Lambda_l-\theta_0)^2-(\hat\Lambda_l^{(0)}-\theta_0)^2\right\}w_l\,\overline{N}_l+r_n, \end{aligned}$$

where

$$r_n=\frac{2}{3}\sum_{l\in D}\left(\frac{(\hat\Lambda_l-\theta_0)^3}{\Lambda_{l,*}^3}-\frac{(\hat\Lambda_l^{(0)}-\theta_0)^3}{\Lambda_{l,**}^3}\right)w_l\,\overline{N}_l.$$

It is not difficult to show that r_n, the remainder term arising from the third derivative of the Taylor expansion of the likelihood ratio statistic, is $o_p(1)$. Thus,

$$2\log\lambda_n=T_1-T_2+o_p(1)\,,$$

with T_1

$$=2\sum_{l\in D} w_l\,\overline{N}_l\,\frac{1}{\theta_0}\left\{(\hat\Lambda_l-\theta_0)-(\hat\Lambda_l^{(0)}-\theta_0)\right\}-2\sum_{l\in D} w_l\left\{(\hat\Lambda_l-\theta_0)-(\hat\Lambda_l^{(0)}-\theta_0)\right\},$$

$$\text{and } T_2=\frac{1}{\theta_0^2}\sum_{l\in D}\left[(\hat\Lambda_l-\theta_0)^2-(\hat\Lambda_l^{(0)}-\theta_0)^2\right]w_l\,\overline{N}_l.$$

Consider T_2. Noting that $n^{-1}\,w_l\,\overline{N}_l$ is the jump of the right-continuous process V_n at the point $T_{(l)}$, letting D_n denote the set on which $\hat\Lambda_n$ and $\hat\Lambda_n^{(0)}$ differ and setting $\tilde D_n$ to be the set $n^{1/3}(D_n-t_0)$, which is an interval and can be easily shown to be $O_p(1)$, we can write

$$T_2 \;=\; \frac{1}{\theta_0^2}\,n\int_{D_n}\left\{(\hat\Lambda_n(t)-\theta_0)^2-(\hat\Lambda_n^{(0)}(t)-\theta_0)^2\right\}d\,V_n(t)$$

$$\text{(A3)} \qquad = \frac{1}{\theta_0^2}\, n \int_{D_n} \left\{ (\hat{\Lambda}_n(t) - \theta_0)^2 - (\hat{\Lambda}_n^{(0)}(t) - \theta_0)^2 \right\} d\,V(t) + o_p(1)$$

$$= \frac{1}{\theta_0^2} \int_{\tilde{D}_n} \Big\{ n^{2/3}\, (\hat{\Lambda}_n(t_0 + n^{-1/3}\, z) - \theta_0)^2$$

$$-n^{2/3}\, (\hat{\Lambda}_n^{(0)}(t_0 + n^{-1/3}\, z) - \theta_0)^2 \Big\}\; V^{'}(t_0 + n^{-1/3}\, z)\, dz + o_p(1)$$

$$= \frac{V^{'}(t_0)}{\theta_0^2} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\} dz + o_p(1)\,,$$

where (A3) follows from the step above it, with V_n replaced by V, by a standard empirical process argument. Now, consider T_1; if we use the definitions of the processes V_n and G_n, it is straightforward to see that

$$\text{(A4)} \quad T_1 = \frac{2}{\theta_0}\, n \int_{D_n} \left\{ (\hat{\Lambda}_n(t) - \theta_0) - (\hat{\Lambda}_n^{(0)}(t) - \theta_0) \right\} d\,(V_n(t) - \theta_0\, G_n(t))$$

$$\text{(A5)} \qquad = \frac{2}{\theta_0}\, n \int_{D_n} \left\{ (\hat{\Lambda}_n(t) - \theta_0)^2 - (\hat{\Lambda}_n^{(0)}(t) - \theta_0)^2 \right\} d\,G_n(t))$$

$$\text{(A6)} \qquad = \frac{2}{\theta_0}\, n \int_{D_n} \left\{ (\hat{\Lambda}_n(t) - \theta_0)^2 - (\hat{\Lambda}_n^{(0)}(t) - \theta_0)^2 \right\} G^{'}(t)\, dt + o_p(1)$$

$$\text{(A7)} \qquad = \frac{2}{\theta_0} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\} G^{'}(t_0 + n^{-1/3}\, z)\, dz + o_p(1)$$

$$= \frac{2\, G^{'}(t_0)}{\theta_0} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\} dz + o_p(1)\,,$$

where (A5) follows from the characterization of the nonparametric maximum likelihood estimators in terms of the processes G_n and V_n and will be justified at the end, (A6) follows from (A5) with $d\,G_n(t)$ replaced by $d\,G(t) \equiv G^{'}(t)\, dt$ using standard empirical process arguments and (A7) follows if we transform to the local variable z and use the definitions of the processes X_n and Y_n. Thus, $2\,\log\,\lambda_n$

$$= \frac{2\, G^{'}(t_0)}{\theta_0} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\} dz - \frac{V^{'}(t_0)}{\theta_0^2} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\} dz + o_p(1)\,.$$

Recalling that, $a^2 = \sigma^2(t_0)/G^{'}(t_0)$ from the statement of Theorem 2 and that $V^{'}(t_0) = \Lambda(t_0)\, G^{'}(t_0)$ from equation (A2), so that

$$\frac{V^{'}(t_0)}{\theta_0^2} = \frac{G^{'}(t_0)}{\theta_0},$$

we have,

$$\begin{aligned} 2\,\log\lambda_n &= \frac{G^{'}(t_0)}{\theta_0} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\}\, dz \\ &= \frac{\sigma^2(t_0)}{\Lambda(t_0)}\, a^{-2} \int_{\tilde{D}_n} \left\{ X_n^2(z) - Y_n^2(z) \right\}\, dz \end{aligned}$$

$$\text{(A8)} \qquad \stackrel{d}{\to} \frac{\sigma^2(t_0)}{\Lambda(t_0)}\, a^{-2} \int \left[\{g_{a,b}(z)\}^2 - \left\{ g_{a,b}^0(z) \right\}^2 \right]\, dz.$$

Here (A8) follows from the previous step by applying Theorem 2 in conjunction with the continuous mapping theorem for distributional convergence and the fact that $(f,g) \mapsto \int (f^2 - g^2)\, d\,\lambda$, with λ denoting Lebesgue measure, is a continuous function from $\mathcal{L} \times \mathcal{L}$ to $\mathbb{R}$. However, $\{\sigma^2(t_0)/\Lambda(t_0)\}\, a^{-2} \int [\{g_{a,b}(z)\}^2 - \{g_{a,b}^0(z)\}^2]\, dz$ has the same distribution as $\{\sigma^2(t_0)/\Lambda(t_0)\}\, \mathbb{D}$, by Lemma 1. If, in particular, $N(t)$ is indeed a Poisson process, nonhomogeneous or otherwise, $\sigma^2(t_0) = \Lambda(t_0)$ and the limiting distribution is exactly $\mathbb{D}$.

It only remains to justify going from (A4) to (A5). It suffices to show that

$$d_1 \equiv \int_{D_n} \{\hat{\Lambda}_n(t) - \theta_0\}\, d\{V_n(t) - \theta_0\, G_n(t)\} = \int_{D_n} \{\hat{\Lambda}_n(t) - \theta_0\}^2\, d\,G_n(t),$$

$$d_2 \equiv \int_{D_n} \{\hat{\Lambda}_n^0(t) - \theta_0\}\, d\{V_n(t) - \theta_0\, G_n(t)\} = \int_{D_n} \{\hat{\Lambda}_n^0(t) - \theta_0\}^2\, d\,G_n(t)\,.$$

We will only show the latter. Let J_n denote the set of indices i such that $T_{(i)}$ belongs to D_n, ordered from smallest to largest. Partition J_n into consecutive blocks of indices $B_1, B_2, \ldots, B_k$ such that, on each B_j, we have that $\hat{\Lambda}_n^{(0)}(T_{(i)})$ is constant for all $i \in B_j$. Denote the constant value on B_j by v_j. There is potentially one block B_l on which $\hat{\Lambda}_n^{(0)}$ is equal to θ_0. On every other B_j, we have

$$v_j = \frac{n^{-1}\sum_{m\in B_j} w_m\,\overline{N_m}}{n^{-1}\sum_{m\in B_j} w_m} \equiv \frac{\sum_{m\in B_j} w_m\,\overline{N_m}}{\sum_{m\in B_j} w_m}\,.$$

This is an easy consequence of the characterization of the constrained solution. We can now write

$$\begin{aligned}
d_2 &= \sum_{j\neq l}\sum_{i\in B_j}\{\hat{\Lambda}_n^{(0)}(T_{(i)})-\theta_0\}\,(n^{-1}\,w_i\,\overline{N}_i-\theta_0\,n^{-1}\,w_i)\\
&= \sum_{j\neq l}(v_j-\theta_0)\left(\sum_{i\in B_j} n^{-1}\ w_i\,\overline{N}_i-\theta_0\sum_{i\in B_j} n^{-1}\ w_i\right)\\
&= \sum_{j\neq l}(v_j-\theta_0)\left(\sum_{i\in B_j} n^{-1}\ w_i\right)\left(\frac{\sum_{i\in B_j} w_i\,\overline{N_i}}{\sum_{i\in B_j} w_i}-\theta_0\right)\\
&= \sum_{j\neq l}(v_j-\theta_0)^2\sum_{i\in B_j} n^{-1}\ w_i\\
&= \sum_{j\neq l}\sum_{i\in B_j}\{\hat{\Lambda}_n^{(0)}(T_{(i)})-\theta_0\}^2\,n^{-1}\,w_i\\
&= \int_{D_n}\{\hat{\Lambda}_n^{0}(t)-\theta_0\}^2\,d\,G_n(t)\,.\quad\Box
\end{aligned}$$

CHAPTER 4

Inconsistency of Bootstrap: the Grenander estimator

In this chapter we investigate the (in)-consistency of different bootstrap methods for constructing confidence bands in the class of estimators that converge at rate cube-root n. The Grenander estimator (see Grenander (1956)), the nonparametric maximum likelihood estimator of an unknown non-increasing density function f on $[0, \infty)$, is a prototypical example. We focus on this example and illustrate different approaches of constructing confidence intervals for $f(t_0)$, where t_0 is an interior point, i.e., $0 < t_0 < \infty$. It is claimed that the bootstrap statistic, when generating bootstrap samples from the empirical distribution function $\mathbb{F}_n$, does not have any weak limit, conditional on the data, in probability. A similar phenomenon is shown to hold when bootstrapping from $\tilde{F}_n$, the least concave majorant of $\mathbb{F}_n$. We provide a set of sufficient conditions for the consistency of bootstrap methods in this example. A suitable version of smoothed bootstrap is proposed and shown to be strongly consistent. The m out of n bootstrap method is also proved to be consistent while generating samples from $\mathbb{F}_n$ and $\tilde{F}_n$. Although we work out the main results for the Grenander estimator, very similar techniques can be employed to draw analogous conclusions for other estimators with cube-root convergence.

4.1 Introduction

Suppose that we observe i.i.d. random variables X_1, $X_2, \ldots, X_n$ from a continuous distribution function F with non-increasing density f on $[0, \infty)$. Let $\mathbb{F}_n$ denote the empirical distribution function (e.d.f.) of the data. Grenander (1956) showed that the non-parametric maximum likelihood estimator (NPMLE) $\tilde{f}_n$ of f exists (obtained by maximizing the likelihood $\prod_{i=1}^n f(X_i)$ over all non-increasing densities) and is given by the left-derivative of $\tilde{F}_n$, the least concave majorant (LCM) of $\mathbb{F}_n$ (see Robertson, Wright and Dykstra (1988) for a derivation of this result). The main result on the distributional convergence of $\tilde{f}_n(t_0)$, for $t_0 \in (0, \infty)$, was given by Prakasa Rao (1969): If $f'(t_0) \neq 0$, then

$$n^{1/3} \left\{ \tilde{f}_n(t_0) - f(t_0) \right\} \Rightarrow \kappa Z \tag{4.1}$$

where $\kappa = 2 \left| \frac{1}{2} f(t_0) f'(t_0) \right|^{1/3}$, $Z = \arg\max_{s \in \mathbb{R}} \{\mathbb{W}(s) - s^2\}$, and $\mathbb{W}$ is a two-sided standard Brownian motion on $\mathbb{R}$ with $\mathbb{W}(0) = 0$. There are other estimators that exhibit similar asymptotic properties; for example, Chernoff's (1964) estimator of the mode, the monotone regression estimator (Brunk (1970)), Rousseeuw's (1984) least median of squares, and the estimator of the shorth (Andrews et al. (1972) and Shorack and Wellner (1986)). The seminal paper by Kim and Pollard (1990) unifies the $n^{1/3}$-rate of convergence problems in a more general M-estimation framework and provides limiting distributions of the estimators. There are further examples of shape restricted nonparametric maximum likelihood density estimators available in the literature – see, for example, estimation of convex densities (Groeneboom, Jongbloed and Wellner (2001)), estimation of k-monotone densities (Balabdaoui and Wellner (2007)) and estimation of decreasing densities that are concave in a neighborhood

(Meyer and Woodroofe (2004)) – but in this chapter we focus our attention to the Grenander estimator.

The presence of nuisance parameters in the limit distribution of the estimators complicates the construction of confidence intervals. Bootstrap intervals avoid this problem and are generally reliable and accurate in problems with $\sqrt{n}$ convergence rate (see Bickel and Freedman (1981), Singh (1981), Shao and Tu (1995) and its references). Our aim in this chapter is to study the consistency of bootstrap methods for the Grenander estimator with the goal of constructing point-wise confidence bands around $\tilde{f}_n$. The monotone density estimation problem sheds light on the behavior of bootstrap methods in other similar cube-root convergence problems discussed above.

Recently there has been considerable interest in using resampling based methods in similar $n^{1/3}$-rate convergence problems. Subsampling based confidence intervals (see Romano, Politis and Wolf (1999)) are consistent in this scenario. But subsampling requires a choice of block-size, which is quite tricky and computationally intensive. The resulting confidence intervals are also not always very accurate and can vary substantially with changing block-size. Abrevaya and Huang (2005) obtained the unconditional limit distribution for the bootstrap version of the normalized estimator in the setup of Kim and Pollard (1990) and proposed a method for constructing confidence intervals in such non-standard problems by correcting the usual bootstrap method. But as we will show in this chapter, such methods of correcting the usual bootstrap method are unlikely to work since there is extremely strong evidence to suggest that the bootstrap statistic does not have any weak limit in probability, conditional on the data. Kosorok (2007) also shows that bootstrapping from the e.d.f. is not consistent in the monotone density estimation problem. Lee and

Pun (2006) explore m out of n bootstrapping from the empirical distribution function in similar non-standard problems and prove the consistency of the method. Léger and MacGibbon (2006) describe conditions for a resampling procedure to be consistent under cube root asymptotics and assert that these conditions are generally not met while bootstrapping from the e.d.f. They propose a smoothed version of bootstrap and show its consistency for Chernoff's estimator of the mode. The authors carry out an extensive simulation study which reveals a disparity in the coverage probability of the percentile and basic bootstrap confidence intervals, also shedding doubt on the existence of a fixed conditional limit distribution for the bootstrap statistic.

In Section 4.2 we introduce notation, describe the stochastic processes of interest, and prove a uniform version of Equation (4.1) that is used later on to study the consistency of different bootstrap methods. Section 4.3 starts with a brief introduction to bootstrap procedures and formalizes the notion of consistency. We show that if the bootstrap methods (while generating bootstrap samples from either the e.d.f. $\mathbb{F}_n$ or its LCM $\tilde{F}_n$) were consistent, then two random variables would be independent, and then show by simulation that these two random variables are not independent. In fact, we show that in these two situations the bootstrap distribution of the statistic of interest does not even have any conditional weak limit, in probability. We state sufficient conditions for the consistency of any bootstrap method and propose a version of smoothed bootstrap in Section 4.4 that can be used to construct asymptotically correct confidence intervals for $f(t_0)$. Section 4.5 investigates the m out of n bootstrapping procedure, when generating bootstrap samples from $\mathbb{F}_n$ and $\tilde{F}_n$, and shows that both the methods are consistent. In Section 4.6 we discuss our findings, especially the failure of the conditional convergence of the bootstrap established in Section 4.3, which we view as one of the key contributions of our current research as it has strong implications for the behavior of the bootstrap in the broader class of cube–root estimation problems. Section 4.7, the appendix, provides the details of some arguments used in proving the main results.

4.2 Preliminaries

We begin with a uniform version of the Prakasa Rao (1969) result which will be useful later on. For the rest of the chapter we will assume that F is a distribution function with continuous non-increasing density f on $[0, \infty)$ which is continuously differentiable near $t_0 \in (0, \infty)$ with nonzero derivative. Suppose that $X_{n,1}, X_{n,2}, \ldots, X_{n,m_n}$ are i.i.d. random variables having distribution function F_n, where $m_n \leq n$ (of special interest is the case $m_n = n$). The quantity of interest to us is

$$\Delta_n := m_n^{1/3}\{\tilde{f}_{n,m_n}(t_0) - f_n(t_0)\}$$

where $\tilde{f}_{n,m_n}(t_0)$ is the Grenander estimator based on the data $X_{n,1}, X_{n,2}, \ldots,$ X_{n,m_n} and $f_n(t_0)$ can be taken as the density of F_n at t_0 (later on we allow f_n to be more flexible, and F_n need not have a density). Let $\mathbb{F}_{n,m_n}$ be the e.d.f. of the data. We study the limiting distribution of the process

$$\mathbb{Z}_n(h) := m_n^{2/3}\left\{\mathbb{F}_{n,m_n}(t_0 + hm_n^{-1/3}) - \mathbb{F}_{n,m_n}(t_0) - f_n(t_0)hm_n^{-1/3}\right\} \tag{4.2}$$

for $h \in I_{m_n} := [-t_0 m_n^{1/3}, \infty)$ and use continuous mapping arguments to deduce the limiting distribution of Δ_n, which can be expressed as the left-hand slope at 0 of the LCM of $\mathbb{Z}_n$, i.e., $\Delta_n = CM_{I_{m_n}}(\mathbb{Z}_n)'(0)$, where CM_I is the operator that maps a function $g : \mathbb{R} \to \mathbb{R}$ into the LCM of g on the interval $I \subset \mathbb{R}$

and $'$ corresponds to the left derivative. We consider all stochastic processes as random elements in $D(\mathbb{R})$, the space of cadlag function (right continuous having left limits) on $\mathbb{R}$, and equip it with the projection σ-field and the metric of uniform convergence on compacta, i.e.,

$$\rho(x,y) = \sum_{k=1}^{\infty} 2^{-k} min[1, \rho_k(x,y)]$$

where $\rho_k(x,y) = \sup_{|t|\le k} |x(t) - y(t)|$ and x and y are elements in $D(\mathbb{R})$. We say that a sequence $\{\xi_n\}$ of random elements in $D(\mathbb{R})$ converges in distribution to a random element ξ, written $\xi_n \Rightarrow \xi$, if $Eg(\xi_n) \to Eg(\xi)$ for every bounded, continuous, measurable real-valued function g. With this notion of weak convergence, the continuous mapping theorem holds (see Pollard (1984), Chapters IV and V for more details).

We decompose $\mathbb{Z}_n$ into $\mathbb{Z}_{n,1}$ and $\mathbb{Z}_{n,2}$ where

$$\mathbb{Z}_{n,1}(h) := m_n^{2/3} \left\{ (\mathbb{F}_{n,m_n} - F_n)(t_0 + hm_n^{-1/3}) - (\mathbb{F}_{n,m_n} - F_n)(t_0) \right\}$$

$$\mathbb{Z}_{n,2}(h) := m_n^{2/3} \left\{ F_n(t_0 + hm_n^{-1/3}) - F_n(t_0) - f_n(t_0) hm_n^{-1/3} \right\} \tag{4.3}$$

Now we state some conditions on the behavior of F_n and f_n (which need not be the density of F_n) to be utilized in proving the uniform version of Equation (4.1).

(a) $F_n(x) \to F(x)$ uniformly for all x in a neighborhood of t_0.

(b) $m_n^{1/3} \Big\{ F_n(t_0 + hm_n^{-1/3}) - F_n(t_0) \Big\} \to hf(t_0)$ as $n \to \infty$ uniformly on compacta.

(c) $\mathbb{Z}_{n,2}(h) \to \frac{1}{2} h^2 f'(t_0)$ as $n \to \infty$ uniformly on compacta.

(d) For each $\epsilon > 0$,

$$\left| F_n(t_0 + \beta) - F_n(t_0) - \beta f_n(t_0) - \frac{1}{2}\beta^2 f'(t_0) \right| \le \epsilon\beta^2 + o(\beta^2) + O(m_n^{-2/3})$$

for large n, uniformly in β varying over a neighborhood of zero (both n and the neighborhood can depend on ϵ).

(e) There exist a neighborhood of 0 and a constant $C > 0$ such that for all n sufficiently large,

$$|F_n(t_0 + \beta) - F_n(t_0)| \leq |\beta| C + O(m_n^{-1/3})$$

uniformly for β in the neighborhood of 0.

Letting $\mathbb{W}_1$ be a standard two-sided Brownian motion on $\mathbb{R}$ with $\mathbb{W}_1(0) = 0$, we define the following stochastic processes

$$\mathbb{Z}_1(h) = \mathbb{W}_1(f(t_0)h) \text{ and } \mathbb{Z}(h) = \mathbb{Z}_1(h) + \frac{1}{2}h^2 f'(t_0), \text{ for } h \in \mathbb{R}.$$

Proposition 1. If (b) holds then $\mathbb{Z}_{n,1} \Rightarrow \mathbb{Z}_1$. Further, if (c) holds then $\mathbb{Z}_n \Rightarrow \mathbb{Z}$.

Proof. To find the limit distribution of the process $\mathbb{Z}_n$, we make crucial use of the Hungarian embedding of Kómlos, Major and Tusnády (1975). We may suppose that $X_{n,i} = F_n^{\#}(U_i)$, where $F_n^{\#}(u) = \inf\{x : F_n(x) \geq u\}$ and $U_1, U_2, \ldots$ are i.i.d. Uniform$(0,1)$ random variables. Let $\mathbb{U}_n$ denote the empirical distribution function of $U_1\ , \ldots, U_{m_n}$, $\mathbb{E}_n(t) = \sqrt{m_n}(\mathbb{U}_n(t) - t)$, and $\mathbb{V}_n = \sqrt{m_n}(\mathbb{F}_{n,m_n} - F_n)$. Then $\mathbb{V}_n = \mathbb{E}_n \circ F_n$. We may also suppose that the probability space supports a sequence of independent Brownian Bridges $\{\mathbb{B}_n^0\}_{n\geq 1}$ for which

$$\sup_{0\leq t\leq 1} |\mathbb{E}_n(t) - \mathbb{B}_n^0(t)| = O(m_n^{-1/2} \log m_n) \text{ a.s.}$$

Let $\{\eta_n\}_{n\geq 1}$ be a sequence of N$(0,1)$ random variables independent of $\{\mathbb{B}_n^0\}_{n\geq 1}$. Define a version $\mathbb{B}_n$ of Brownian motion by $\mathbb{B}_n(t) = \mathbb{B}_n^0(t) + \eta_n t$, for $t \in [0,1]$.

Using the Hungarian construction we express $\mathbb{Z}_{n,1}$ as

$$\begin{aligned}
\mathbb{Z}_{n,1}(h) &= m_n^{1/6}\{\mathbb{V}_n(t_0+hm_n^{-1/3})-\mathbb{V}_n(t_0)\} \\
&= m_n^{1/6}\left\{\mathbb{E}_n(F_n(t_0+hm_n^{-1/3}))-\mathbb{E}_n(F_n(t_0))\right\} \\
&= m_n^{1/6}\left\{\mathbb{B}_n^0(F_n(t_0+hm_n^{-1/3}))-\mathbb{B}_n^0(F_n(t_0))\right\}+\mathbb{R}_{n,1}(h) \\
&= m_n^{1/6}\left\{\mathbb{B}_n(F_n(t_0+hm_n^{-1/3})-\mathbb{B}_n(F_n(t_0)\right\}+\mathbb{R}_n(h) \qquad (4.4)
\end{aligned}$$

where $\mathbb{R}_n = \mathbb{R}_{n,1}+\mathbb{R}_{n,2}$, $|\mathbb{R}_{n,1}(h)| \leq 2m_n^{1/6}\sup_{0\leq t\leq 1}|\mathbb{E}_n(t)-\mathbb{B}_n^0(t)| = O(m_n^{-1/3}\log m_n)$ a.s., and $|\mathbb{R}_{n,2}(h)| \leq m_n^{1/6}|\eta_n||F_n(t_0+hm_n^{-1/3})-F_n(t_0)| \to 0$, w.p.1 by condition (b). Therefore, $\mathbb{R}_n(h)\to 0$ w.p.1 as $n\to\infty$ uniformly on compacta. Letting $\mathbb{X}_n(h) := m_n^{1/6}\{\mathbb{B}_n(F_n(t_0+hm_n^{-1/3}))-\mathbb{B}_n(F_n(t_0))\}$, we observe that $\mathbb{X}_n$ is a mean zero Gaussian process defined on I_{m_n} with independent increments and covariance kernel

$$K_n(h_1,h_2) = m_n^{1/3}\{F_n(t_0+(h_1\wedge h_2)m_n^{-1/3})-F_n(t_0)\}\mathbf{1}\{sign(h_1h_2)>0\}.$$

Theorem V.19 in Pollard (1984) gives sufficient conditions for convergence of the process $\mathbb{X}_n(h)$ to $\mathbb{W}_1(f(t_0)h)$ in $D([-c,c])$ for any c that are readily verified using condition (b) in the proposition. The second part follows immediately. $\square$

We may obtain the asymptotic distribution of Δ_n from the following corollary, which is stated in a more general setup.

Corollary 1. Suppose that conditions (a), (d) and (e) hold. Let $\mathbb{Z}$ be a stochastic process on $\mathbb{R}$ such that,

(1) $\lim_{|h|\to\infty}\frac{\mathbb{Z}(h)}{|h|} = -\infty$ a.e.,

(2) $\mathbb{Z}$ is a.s. bounded above, and

(3) $CM_{[-k,k]}(\mathbb{Z})$, for $k=1,2,\ldots$, and $CM_{\mathbb{R}}(\mathbb{Z})$ are differentiable at 0 a.s.

If $\mathbb{Z}_n \Rightarrow \mathbb{Z}$ then $\Delta_n \Rightarrow CM_{\mathbb{R}}(\mathbb{Z})'(0)$.

We use the continuous mapping principle and a localization argument similar to that in Kim and Pollard (1990). The details are provided in the Appendix.

4.3 Inconsistency of the bootstrap

In this section, we show that the usual bootstrap method, generating bootstrap samples from the e.d.f. $\mathbb{F}_n$, leads to an inconsistent procedure. Not only does the bootstrap estimate fail to converge weakly to the right distribution, but there is strong evidence that it does not have any conditional limit distribution, in probability. We also consider bootstrapping from $\tilde{F}_n$, the least concave majorant of $\mathbb{F}_n$, and this procedure shows similar asymptotic behavior. We begin with a brief discussion on bootstrap.

Suppose we have i.i.d. random variables $X_1, X_2, \ldots, X_n$ having an unknown distribution function F defined on a probability space $(\Omega, \mathcal{A}, P)$ and we seek to estimate the sampling distribution of the random variable $R_n(\mathbf{X}_n, F)$, based on the observed data $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$. Let H_n be the distribution function of $R_n(\mathbf{X}_n, F)$. The bootstrap methodology can be broken into three simple steps:

Step 1: Construct an estimate $\hat{F}_n$ of F based on the data (for example, the e.d.f. $\mathbb{F}_n$).

Step 2: With $\hat{F}_n$ fixed, we draw a random sample of size m_n from $\hat{F}_n$, say $\mathbf{X}_n^* = (X_1^*, X_2^*, \ldots, X_{m_n}^*)$ (identically distributed and conditionally independent given $\mathbf{X}_n$). This is called the *bootstrap sample*.

Step 3: We approximate the sampling distribution of $R_n(\mathbf{X}_n, F)$ by the sampling distribution of $R_n^* = R_n(\mathbf{X}_n^*, \hat{F}_n)$. The sampling distribution of R_n^* can be simulated on the computer by drawing a large number of bootstrap samples and computing R_n^* for each sample.

Thus the bootstrap estimator of the sampling distribution function of $R_n(\mathbf{X}_n, F)$ is given by

$$H_n^*(x) = P^*\{R_n^* \leq x\},$$

where $P^*\{\cdot\}$ is the conditional probability given the data $\mathbf{X}_n$. Let L denote the Levy metric or any other metric metrizing weak convergence of distribution functions. We say that H_n^* is *(weakly) consistent* if $L(H_n, H_n^*) \xrightarrow{P} 0$. Similarly, H_n^* is *strongly consistent* if $L(H_n, H_n^*) \to 0$ a.s. If H_n has a weak limit H, for the bootstrap procedure to be consistent, H_n^* must converge weakly to H, in probability. In addition, if H is continuous, we must have

$$\sup_{x \in \mathbb{R}} |H_n^*(x) - H(x)| \xrightarrow{P} 0 \text{ as } n \to \infty.$$

By saying that H_n^* converges in probability to a possibly random G, in probability, we shall mean

(i) that there exists a stochastic transition function $G : \mathbb{R} \times \Omega \to [0, 1]$ such that $G(\cdot, \omega)$ is a distribution function for all $\omega \in \Omega$, and $G(x; \cdot)$ is a measurable function for every $x \in \mathbb{R}$, and

(ii) $L(H_n^*, G) \xrightarrow{P} 0$.

In fact, if $\hat{F}_n$ depends only on the order statistics of $X_1, X_2, \ldots, X_n$, the limiting G cannot depend on ω, if it exists. For if h is a bounded measurable function on $\mathbb{R}$, then any limit in probability of $\int_{\mathbb{R}} h(x) H_n^*(dx; \omega)$ must be invariant under permutations of $X_1, X_2, \ldots, X_n$ up to equivalence, and thus, must be almost surely constant by the Hewitt-Savage zero-one law (see Breiman (1968)). Let

$$\bar{G}(x) = \int_\Omega G(x;\omega)P(d\omega),$$

then $\bar{G}$ is a distribution function and $\int_{\mathbb{R}} h(x)G(dx;\omega) = \int_{\mathbb{R}} h(x)\bar{G}(dx)$ a.s. for each bounded continuous h. It follows that $G(x;\omega) = \bar{G}(x)$ a.e. ω for each x by letting h approach an indicator.

We are interested in exploring the (in)-consistency of different bootstrap procedures for the Grenander estimator. Specifically, we are interested in studying the limit behavior of

$$\Delta_n^* = m_n^{1/3}\left\{\hat{f}^*_{n,m_n}(t_0) - \hat{f}_n(t_0)\right\} \tag{4.5}$$

where $\hat{f}_n(t_0)$ is an estimate of $f(t_0)$ ($\hat{f}_n(t_0)$ can be $\tilde{f}_n(t_0)$); $\hat{f}^*_{n,m_n}(t_0)$ is the corresponding bootstrap estimate based on a bootstrap sample of size m_n.

Remark: For the rest of the chapter we make crucial use of Proposition 1 and Corollary 1. In situations where the bootstrap works, the results will be applied conditionally on the sequence $X_1, X_2, \ldots$ with $F_n = \hat{F}_n$ and $\mathbb{F}_{n,m_n} = \mathbb{F}_n^*$ (the e.d.f. of the bootstrap sample generated from $\hat{F}_n$). For scenarios where the bootstrap is inconsistent, techniques similar to that of the proof of Corollary 1 are used unconditionally to derive the unconditional limit distribution of Δ_n^*.

4.3.1 Bootstrapping from the e.d.f. $\mathbb{F}_n$

Consider now the case in which $m_n = n$ and $\hat{F}_n = \mathbb{F}_n$. The quantity of interest is $\Delta_n^* := n^{1/3}\{\tilde{f}_n^*(t_0) - \tilde{f}_n(t_0)\}$, the bootstrap analogue of $\Delta_n := n^{1/3}\{\tilde{f}_n(t_0) - f(t_0)\}$. Letting $\mathbf{X} = (X_1, X_2, \ldots)$, we define $G_n(x;\omega) = P\{\Delta_n^* \leq x|\mathbf{X}\}(\omega) = P^*\{\Delta_n^* \leq x\}(\omega)$ as the conditional distribution function of Δ_n^* given $\mathbf{X}$. We claim that G_n does not converge in P-probability.

Let us define the process

$$\mathbb{Z}_n(h) := n^{2/3}\left\{\mathbb{F}_n^*(t_0 + hn^{-1/3}) - \mathbb{F}_n^*(t_0) - \tilde{f}_n(t_0)hn^{-1/3}\right\}$$

for $h \in I_n = [-t_0 n^{1/3}, \infty)$. Then $\mathbb{Z}_n = \mathbb{Z}_{n,1} + \mathbb{Z}_{n,2}$, where

$$\mathbb{Z}_{n,1}(h) = n^{2/3}\{\mathbb{F}_n^*(t_0 + hn^{-1/3}) - \mathbb{F}_n^*(t_0) - \mathbb{F}_n(t_0 + hn^{-1/3}) + \mathbb{F}_n(t_0)\}$$

and

$$\mathbb{Z}_{n,2}(h) = n^{2/3}\{\mathbb{F}_n(t_0 + hn^{-1/3}) - \mathbb{F}_n(t_0) - \tilde{f}_n(t_0)hn^{-1/3}\}.$$

Let $\mathbb{W}_1$ and $\mathbb{W}_2$ be two independent two-sided standard Brownian motions on $\mathbb{R}$ with $\mathbb{W}_1(0) = \mathbb{W}_2(0) = 0$ and let

$$\begin{aligned}
\mathbb{Z}_1(h) &:= \mathbb{W}_1(f(t_0)h), \\
\mathbb{Z}_2^0(h) &:= \mathbb{W}_2(f(t_0)h) + \frac{1}{2}f'(t_0)h^2, \\
Z_2 &:= CM_{\mathbb{R}}[\mathbb{Z}_2^0]'(0), \\
\mathbb{Z}_2(h) &:= \mathbb{Z}_2^0(h) - hZ_2, \\
\mathbb{Z} &:= \mathbb{Z}_1 + \mathbb{Z}_2 \text{ and} \\
Z_1 &:= CM_{\mathbb{R}}[\mathbb{Z}_1 + \mathbb{Z}_2^0]'(0).
\end{aligned} \tag{4.6}$$

Note that Δ_n^* equals the left derivative at $h = 0$ of the LCM of $\mathbb{Z}_n$. We study the behavior of the process $\mathbb{Z}_n$ and then use a continuous mapping type argument to derive the behavior of Δ_n^*. It will be shown that $\mathbb{Z}_n$ does not have any weak limit conditional on $\mathbf{X}$ in P-probability. But unconditionally, $\mathbb{Z}_n$ has a limit distribution, which gives us the unconditional limit distribution of Δ_n^* that is different from the limit distribution of Δ_n.

We first state two lemmas without proof, applicable in more general scenarios, that will be used later in the chapter.

Lemma 4.1. *Let W_n and W_n^* be random vectors in $\mathbb{R}^l$ and $\mathbb{R}^k$ respectively; let Q and Q^* be distributions on the Borel sets of $\mathbb{R}^l$ and $\mathbb{R}^k$; and let $\mathcal{F}_n$ be sigma-fields for which W_n is $\mathcal{F}_n$-measurable. If the distribution of W_n converges to Q and the conditional distribution of W_n^* given $\mathcal{F}_n$ converges in probability to Q^*, then the joint distribution of (W_n, W_n^*) converges to the product measure $Q \times Q^*$.*

REMARK: The above lemma can be proved easily using characteristic functions.

Lemma 4.2. *Let $\mathbf{X}_n^*$ be a bootstrap sample generated from the data $\mathbf{X}_n$. Let $Y_n := \psi_n(\mathbf{X}_n)$ and $Z_n := \phi_n(\mathbf{X}_n, \mathbf{X}_n^*)$ where ψ_n and ϕ_n are measurable functions; and let G_n and H_n be the conditional distribution functions of $Y_n + Z_n$ and Z_n respectively. If there are distribution functions G and H for which H is non-degenerate, $L(G_n, G) \xrightarrow{P} 0$ and $L(H_n, H) \xrightarrow{P} 0$ then there is a random variable Y for which $Y_n \xrightarrow{P} Y$.*

REMARK: One proof of this lemma rests on the following idea. If $\{n_k\}$ is any subsequence for which $L(G_{n_k}, G) \to 0$ and $L(H_{n_k}, H) \to 0$ w.p.1, then $Y := \lim_{n\to\infty} Y_{n_k}$ exists by the Convergence of Types Theorems (see Loeve (1962), page 203) and Y does not depend on n_k since two subsequences can be joined. The lemma follows easily.

Proposition 2. The conditional distribution of $\mathbb{Z}_{n,1}$ given $\mathbf{X} = (X_1, X_2, \ldots)$ converges a.s. to the distribution of $\mathbb{Z}_1$. The unconditional distribution of $\mathbb{Z}_{n,2}$ converges to that of $\mathbb{Z}_2$ and the unconditional distribution of $\mathbb{Z}_n$ converges to that of $\mathbb{Z}$.

Proof. The conditional convergence of $\mathbb{Z}_{n,1}$ follows by applying Proposition 1 with $m_n = n$, $F_n = \mathbb{F}_n$, $\mathbb{F}_{n,m_n} = \mathbb{F}_n^*$. Note that as we are conditioning on $\mathbf{X}$,

$\mathbb{F}_n$, and $\tilde{f}_n$ are fixed and we can apply the proposition. Condition (b) in the Proposition is satisfied as $n^{1/3}\{\mathbb{F}_n(t_0+hn^{-1/3})-\mathbb{F}_n(t_0)\}$ can be written as

$$\begin{aligned} n^{1/3}\{F(t_0+hn^{-1/3})-F(t_0)\} \quad &+ \quad n^{1/3}\{(\mathbb{F}_n-F)(t_0+hn^{-1/3})-(\mathbb{F}_n-F)(t_0)\} \\ &= \quad hf(\alpha_n(h))+r_n(h), \end{aligned} \tag{4.7}$$

where $|r_n(h)| \leq 2n^{1/3}\sup_{s\in\mathbb{R}}|\mathbb{F}_n(s)-F(s)| \to 0$ w.p.1 (P) by the law of iterated logarithm (see Theorem 5.1.1 of Csörgő, M., and Révész, P. (1981)), and $\alpha_n(h)$ is between $t_0+hn^{-1/3}$ and t_0. Thus the conditional distribution of $\mathbb{Z}_{n,1}$ given $\mathbf{X}$ converges to that of $\mathbb{Z}_1$ a.s. As a consequence, the unconditional limit distribution of $\mathbb{Z}_{n,1}$ is the same as that of $\mathbb{Z}_1$.

To find the unconditional limit distribution of the process $\mathbb{Z}_{n,2}$ notice that $\mathbb{Z}_{n,2}$ is a function of the process

$$\mathbb{Z}^0_{n,2}(h) = n^{2/3}\{\mathbb{F}_n(t_0+hn^{-1/3})-\mathbb{F}_n(t_0)-f(t_0)hn^{-1/3}\},$$

which is quite well studied in the literature (see Kim and Pollard (1990) for more details). For $I\subset\mathbb{R}$, define the operator $G_I : f(h)\mapsto f(h)-h\cdot(CM_I f)'(0)$ for $h\in I$, $f:\mathbb{R}\to\mathbb{R}$. Observe that $\mathbb{Z}_{n,2}$ is the image of $\mathbb{Z}^0_{n,2}$ under the mapping G_{I_n}.

We apply Lemma 5.1 with $X_{n,c} = G_{[-c,c]}[\mathbb{Z}^0_{n,2}]$, $Y_n = G_{I_n}[\mathbb{Z}^0_{n,2}]$, $W_c = G_{[-c,c]}[\mathbb{Z}^0_2]$ and $Y = G_{\mathbb{R}}[\mathbb{Z}^0_2]$. For I compact, it is easy to see that $G_I : D(I)\to D(I)$ is a continuous map at all points f for which $(CM_I f)$ is differentiable at 0, i.e., both left and right derivatives exist and are equal. This shows that condition (iii) of the lemma is satisfied. Condition (ii) follows from known facts about the process $\mathbb{Z}^0_2$. Note that for any $\delta>0$, there exists $K>0$ such that for $c>K$,

$$P\{\rho(X_{n,c},W_c)>\delta\} \leq P\left\{\left|CM_{[-c,c]}[\mathbb{Z}^0_{n,2}]'(0)-CM_{I_n}[\mathbb{Z}^0_{n,2}]'(0)\right| > \frac{\delta}{2K}\right\}.$$

The Assertion in page 217 of Kim and Pollard (1990) can now be used directly to verify condition (i) of Lemma 5.1. Thus we conclude that $\mathbb{Z}_{n,2} = Y_n = G_{I_n}[\mathbb{Z}^0_{n,2}] \Rightarrow G_{\mathbb{R}}[\mathbb{Z}^0_2] = Y = \mathbb{Z}_2$.

Next we show that $\mathbb{Z}_{n,1}$ and $\mathbb{Z}^0_{n,2}$ are asymptotically independent, i.e., the joint limit distribution of $\mathbb{Z}_{n,1}$ and $\mathbb{Z}^0_{n,2}$ is the product of their marginal limit distributions. For this it suffices to show that $(\mathbb{Z}_{n,1}(t_1), \ldots, \mathbb{Z}_{n,1}(t_k))$ and $(\mathbb{Z}^0_{n,2}(s_1), \ldots, \mathbb{Z}^0_{n,2}(s_l))$ are asymptotically independent, for all choices $-\infty < t_1 < \ldots < t_k < \infty$ and $-\infty < s_1 < \ldots < s_l < \infty$. This is an easy consequence of the Lemma 4.1.

The joint unconditional distribution of $(\mathbb{Z}_{n,1}, \mathbb{Z}_{n,2})$ can be expressed as

$$\begin{pmatrix} \mathbb{Z}_{n,1}(h) \\ \mathbb{Z}_{n,2}(h) \end{pmatrix} = \begin{pmatrix} \mathbb{Z}_{n,1}(h) \\ G_{I_n}[\mathbb{Z}^0_{n,2}](h) \end{pmatrix} \Rightarrow \begin{pmatrix} \mathbb{W}_1(f(t_0)h) \\ \mathbb{Z}^0_2(h) - hZ_2 \end{pmatrix}. \tag{4.8}$$

As $\mathbb{Z}_{n,1}$ and $\mathbb{Z}^0_{n,2}$ are asymptotically independent, the process $\mathbb{Z}_n$ converges weakly to $\mathbb{Z}$. □

Corollary 2. The unconditional distribution of Δ^*_n converges to that of $CM_{\mathbb{R}}[\mathbb{Z}]'(0)$.

As in the proof of Corollary 1, we use the continuous mapping principle with a localization argument. The details are provided in the Appendix.

Proposition 3. Conditional on $\mathbf{X}$, the distribution of $\mathbb{Z}_n$ does not have a weak limit in P-probability.

Proof. We use the method of contradiction. Let $Z_n := \mathbb{Z}_{n,1}(h_0)$ and $Y_n := \mathbb{Z}_{n,2}(h_0)$ for some fixed $h_0 > 0$ (say $h_0 = 1$) and suppose that the conditional distribution of $Z_n + Y_n = \mathbb{Z}_n(h_0)$ converges in probability to the distribution function G. Observe that the distribution of Z_n converges in P-probability to a normal distribution by Proposition 1 which is obviously nondegenerate. Thus

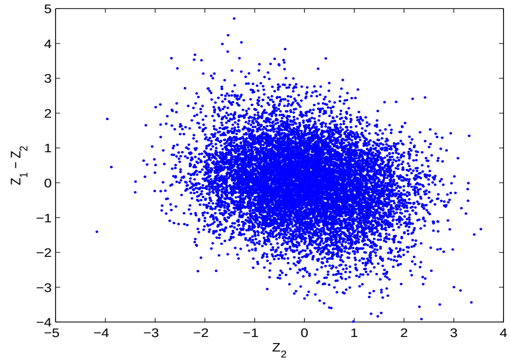

Figure 4.1: Scatter plot of 10000 random draws of $(Z_2, Z_1 - Z_2)$ when $f(t_0) = 1$ and $f'(t_0) = -2$.

the assumptions of Lemma 4.2 are satisfied and we conclude that $Y_n \xrightarrow{P} Y$, for some random variable Y. It then follows from the Hewitt-Savage zero-one law that Y is a constant, say $Y = c_0$ w.p.1. The contradiction arises since Y_n converges in distribution to $\mathbb{Z}_2^0(h_0) - h_0 Z_2$ which is not a constant a.s. $\square$

Proposition 4. If the conditional distribution function of Δ_n^* converges in P-probability, then $CM_{\mathbb{R}}[\mathbb{Z}]'(0) = Z_1 - Z_2$ must be independent of both $\mathbb{Z}_2$ and Z_2.

Proof. Note that Δ_n^* and $\mathbb{Z}_{n,2}^0$ are asymptotically independent by an application of Lemma 4.1 with $W_n = (\mathbb{Z}_{n,2}^0(t_1), \mathbb{Z}_{n,2}^0(t_2), \ldots, \mathbb{Z}_{n,2}^0(t_l))$, for $(t_1, t_2, \ldots, t_l) \in \mathbb{R}^l$, $W_n^* = \Delta_n^*$ and $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$. As $\mathbb{Z}_2$ and Z_2 are both functions of $\mathbb{Z}_2^0$, the result follows. $\square$

When combined with simulations, Proposition 4 strongly suggests that the conditional distribution of Δ_n^* does not converge in probability. The simulations clearly indicate that $Z_1 - Z_2$ and Z_2 are not independent. We have not been able to find a mathematical proof of this.

Figure 4.1 shows the scatter plot of $Z_1 - Z_2$ versus Z_2 obtained from a simulation study with 10000 samples. We took $f(t_0) = 1$ and $f'(t_0) = -2$. The correlation coefficient obtained is -0.2114 and is highly significant. This indicates that Z_2 and $Z_1 - Z_2$ are not independent.

4.3.2 Bootstrapping from $\tilde{F}_n$

One obvious problem with drawing the bootstrap samples from the e.d.f. $\mathbb{F}_n$ is that $\mathbb{F}_n$ does not have a density. In this subsection we consider bootstrapping from $\tilde{F}_n$, the LCM of $\mathbb{F}_n$, which does have a non-increasing density $\tilde{f}_n$.

Let $X^*_{n,1}, X^*_{n,2}, \ldots, X^*_{n,n}$ be a bootstrap sample generated from $\tilde{F}_n$. As before, we study the process $\mathbb{Z}_n(h) = n^{2/3}\{\mathbb{F}^*_n(t_0+hn^{-1/3}) - \mathbb{F}^*_n(t_0) - \tilde{f}_n(t_0)hn^{-1/3}\}$. We claim that $\Delta^*_n = n^{1/3}\{\tilde{f}^*_n(t_0) - \tilde{f}_n(t_0)\}$, the left derivative at $h = 0$ of the LCM of $\mathbb{Z}_n$, does not have any weak limit, conditional on $\mathbf{X}$. We show that $\mathbb{Z}_n$ does not have any limit distribution conditional on the data. But unconditionally, $\mathbb{Z}_n$ has a limit distribution which gives the unconditional limit distribution of Δ^*_n that is different from the weak limit of Δ_n, thereby illustrating that the bootstrap procedure is not consistent. We borrow the notation introduced in Equation (4.6) except that now

$$\mathbb{Z}_2(h) \;:=\; CM_{\mathbb{R}}[\mathbb{Z}^0_2](h) - CM_{\mathbb{R}}[\mathbb{Z}^0_2](0) - h \cdot CM_{\mathbb{R}}[\mathbb{Z}^0_2]'(0).$$

Theorem 3. The following hold.

(i) The conditional distribution of $\mathbb{Z}_{n,1}$, given $\mathbf{X}$, converges almost surely to the distribution of $\mathbb{Z}_1$; the unconditional distribution of $\mathbb{Z}_{n,2}$ converges to that of $\mathbb{Z}_2$; and the unconditional distribution of $\mathbb{Z}_n$ converges to that of $\mathbb{Z}$.

(ii) The unconditional distribution of Δ^*_n converges to that of $CM_{\mathbb{R}}[\mathbb{Z}]'(0)$.

(iii) Conditional on $\mathbf{X}$, $\mathbb{Z}_n$ does not have a weak limit in P-probability.

(iv) If Δ_n^* has a weak limit, conditional on $\mathbf{X}$, in P-probability, then $Z_1 - Z_2$ must be independent of the process $\mathbb{Z}_2$ and the random variable Z_2.

Proof. The proof of the result runs along similar lines as that of the propositions and corollaries in the last subsection. Using ideas similar to that in Equation (4.4) and the following discussion, the process

$$\mathbb{Z}_{n,1}(h) := n^{2/3}\{\mathbb{F}_n^*(t_0 + hn^{-1/3}) - \mathbb{F}_n^*(t_0) - \tilde{F}_n(t_0 + hn^{-1/3}) + \tilde{F}_n(t_0)\}$$

converges in distribution to $\mathbb{Z}_1(h) = \mathbb{W}_1(f(t_0)h)$ conditional on $\mathbf{X}$, a.s. We express $\mathbb{Z}_{n,2}$ as a function of the process $\mathbb{Z}_{n,2}^0$ and apply a continuous mapping type argument to find its limiting distribution. Note that $\mathbb{Z}_{n,2}(h)$ can be expressed as

$$\begin{aligned}
& n^{2/3}\left\{\tilde{F}_n(t_0 + hn^{-1/3}) - \tilde{F}_n(t_0)\right\} - \tilde{f}_n(t_0)hn^{1/3} \\
= \;& n^{2/3}\left\{\tilde{F}_n(t_0 + hn^{-1/3}) - \mathbb{F}_n(t_0) - f(t_0)hn^{-1/3}\right\} \\
- \;& n^{2/3}\left\{\tilde{F}_n(t_0) - \mathbb{F}_n(t_0)\right\} - n^{1/3}h\left\{\tilde{f}_n(t_0) - f(t_0)\right\} \\
= \;& CM_{I_n}[\mathbb{Z}_{n,2}^0](h) - CM_{I_n}[\mathbb{Z}_{n,2}^0](0) - h \cdot CM_{I_n}[\mathbb{Z}_{n,2}^0]'(0)
\end{aligned}$$

An application of the continuous mapping principle (with a localization argument) yields the unconditional convergence of $\mathbb{Z}_{n,2} \Rightarrow \mathbb{Z}_2$. The proof of part (ii) uses similar techniques as that in the proof of Corollary 2 and is given in the appendix. Using Proposition 3 we can argue that $\mathbb{Z}_n$ does not converge to any weak limit, conditional on $\mathbf{X}$, in P-probability. Proposition 4 can be employed to complete the proof of (iv) of the theorem. □

As before, extensive simulations show that $Z_1 - Z_2$ and Z_2 are not independent, which suggests that Δ_n^* does not have a conditional weak limit in probability.

4.4 Bootstrapping from a smoothed version of $\tilde{F}_n$

One of the major reasons for the inconsistency of bootstrap methods discussed in the previous section is the lack of smoothness of the distribution from which the bootstrap samples are generated. The e.d.f. $\mathbb{F}_n$ does not have a density, and $\tilde{F}_n$ does not have a differentiable density, whereas F is assumed to have a nonzero differentiable density at t_0. The results from Section 4.2 are directly applied to derive sufficient conditions on the smoothness of the distribution from which the bootstrap samples are generated.

Theorem 4. Suppose that we generate a bootstrap sample $X_{n,1}^*, X_{n,2}^*, \ldots, X_{n,m_n}^*$ from a distribution function $\hat{F}_n$ constructed from the data $X_1, X_2, \ldots, X_n$. Let $\hat{f}_n$ be an estimate of the density of $\hat{F}_n$. Let $\tilde{f}_n^*$ be the NPMLE based on the bootstrap sample. Also suppose that conditions (a)-(e) used in Proposition 1 hold a.s. with $F_n = \hat{F}_n$ and $f_n = \hat{f}_n$. Then the bootstrap distribution is strongly consistent, i.e., for almost all $\mathbf{X}$, the conditional limit distribution of $\Delta_n^* = m_n^{1/3}\left\{\tilde{f}_n^*(t_0) - \hat{f}_n(t_0)\right\}$ is the same the unconditional limit distribution of $\Delta_n = n^{1/3}\left\{\tilde{f}_n(t_0) - f(t_0)\right\}$. Equivalently,

$$\sup_{x\in\mathbb{R}} \left|P^*\left\{\Delta_n^* \le x\right\} - P\left\{\Delta_n \le x\right\}\right| \xrightarrow{a.s.} 0 \tag{4.9}$$

Proof. Conditional on $\mathbf{X}$, $\hat{F}_n$ and $\hat{f}_n$ are fixed, and we can apply Proposition 1 with $F_n = \hat{F}_n$ and $f_n = \hat{f}_n$ to obtain the limit distribution of the process $\mathbb{Z}_n$ (defined in Equation (4.2)). Equation (4.9) follows directly from an application of Corollary 1 (as the conditions (1)-(3) on the limit process $\mathbb{Z}$ are satisfied) and Polya's theorem, noticing that the conditional limit distribution of Δ_n^* is continuous. $\square$

As an example, we construct a kernel smoothed version of $\tilde{F}_n$ and show that it leads to a consistent bootstrap procedure. The usual kernel smoothing of the Grenander estimator would give rise to a boundary effect at 0, as f is supported on $[0, \infty)$, and might violate the assumption of monotonicity. To avoid these difficulties, we transform the observations by taking logarithms, kernel smooth the transformed data points, which are now supported on $\mathbb{R}$, and back transform the smoothed density to obtain an estimate of f. The result is

$$\begin{aligned} \check{f}_n(x) &:= \frac{1}{xh_n} \int_0^\infty K\left(\frac{\log u - \log x}{h_n}\right) \tilde{f}_n(u) du \\ &= \frac{1}{h_n} \int_0^\infty K\left(\frac{\log v}{h_n}\right) \tilde{f}_n(vx) dv \end{aligned}$$

for $x \in [0, \infty)$, where h_n is the smoothing bandwidth, and $K(\cdot)$ is a symmetric (around 0) density function on $\mathbb{R}$ satisfying the following conditions:

(i) K' exists and is bounded on $\mathbb{R}$.

(ii) K'' exists and is continuous on $\mathbb{R}$.

(iii) $\int_{-\infty}^{\infty} |K^{(i)}(u)| \max\{1, e^{u\epsilon}\} du < \infty$ for some $\epsilon > 0$, and $i = 0, 1, 2$.

It is easy to see that $\check{f}_n$ is a non-increasing density function supported on $[0, \infty)$. We generate bootstrap samples from $\check{F}_n$, the distribution function having density $\check{f}_n$. To simplify notation, let $K_{h_n}(u, x) := \frac{1}{xh_n} K\left(\frac{\log u - \log x}{h_n}\right)$. The following display gives an alternative expression for $\check{f}_n$ which directly follows from integration by parts and noticing that $\lim_{u \to \infty} K_h(u, x) = 0$ for every $x \in (0, \infty)$, $h_n > 0$,

$$\check{f}_n(x) = \int_0^\infty K_{h_n}(u, x) \tilde{f}_n(u) du = -\int_0^\infty \frac{\partial}{\partial u} \left[K_{h_n}(u, x)\right] \tilde{F}_n(u) du.$$

The next theorem shows the consistency of the bootstrap procedure when generating n data points $X^*_{n,1}, X^*_{n,2}, \ldots, X^*_{n,n}$ from $\check{F}_n$.

Theorem 5. Assume that $h_n \to 0$ and $h_n^2(n/\log\log n)^{1/2} \to \infty$ as $n \to \infty$. Then the bootstrap method is strongly consistent, i.e., Equation (4.9) holds with $\Delta_n^* = n^{1/3}\left\{\tilde{f}_n^*(t_0) - \check{f}_n(t_0)\right\}$.

Proof. Let $\mathbb{F}_n^*$ be the e.d.f. of $X_{n,1}^*, X_{n,2}^*, \ldots, X_{n,n}^*$. We define $\mathbb{Z}_n(z) := n^{2/3}\{\mathbb{F}_n^*(t_0 + zn^{-1/3}) - \mathbb{F}_n^*(t_0) - \check{f}_n(t_0)zn^{-1/3}\}$ for $z \in [-t_0 n^{1/3}, \infty]$. We show that the conditions (a)-(e) hold a.s. and use Theorem 4 to get the desired result.

As before, let $\mathbb{Z}_n(z) = \mathbb{Z}_{n,1}(z) + \mathbb{Z}_{n,2}(z)$, where $\mathbb{Z}_{n,1}(z) = n^{2/3}[\{\mathbb{F}_n^*(t_0 + zn^{-1/3}) - \mathbb{F}_n^*(t_0)\} - \{\check{F}_n(t_0 + zn^{-1/3}) - \check{F}_n(t_0)\}]$, and $\mathbb{Z}_{n,2}(z) = n^{2/3}[\{\check{F}_n(t_0 + zn^{-1/3}) - \check{F}_n(t_0)\} - \check{f}_n(t_0)zn^{-1/3}]$.

As a first step, we establish (c), i.e., $\mathbb{Z}_{n,2}(z) \stackrel{a.s.}{\to} \frac{z^2}{2}f'(t_0)$ uniformly on compacta. Fix a compact set $[-M, M] \subset \mathbb{R}$. As $\check{F}_n$ is twice continuously differentiable, we can use Taylor expansion to simplify $\mathbb{Z}_{n,2}(z)$ to $\frac{z^2}{2}\check{f}_n'(t_n(z))$ where $t_n(z)$ is an intermediate point between t_0 and $t_0 + zn^{-1/3}$. We now show that $\check{f}_n'(t_n(z)) \stackrel{a.s.}{\to} f'(t_0)$ uniformly for $z \in [-M, M]$. Towards this end, let us define

$$\bar{f}_n(x) = \int_0^\infty K_{h_n}(u,x)f(u)du = -\int_0^\infty \frac{\partial}{\partial u}\left[K_{h_n}(u,x)\right]F(u)du; \tag{4.10}$$

$\bar{f}_n$ is just a smoothed version of the original density function f. We first show that $\check{f}_n'(t) - \bar{f}_n'(t) \stackrel{a.s.}{\to} 0$ uniformly on $[t_0 - \delta, t_0 + \delta]$ where $\delta > 0$ is such that $t_0 - \delta > 0$ and f is continuously differentiable in the interval. For $t \in [t_0 - \delta, t_0 + \delta]$,

$$\begin{aligned}
|\check{f}_n'(t) - \bar{f}_n'(t)| &= \left|\int_0^\infty \frac{\partial^2}{\partial t \partial u}\left[K_{h_n}(u,t)\right]\left\{\tilde{F}_n(u) - F(u)\right\}\right| du \\
&\leq \int_0^\infty \left|\frac{\partial^2}{\partial t \partial u}\left[K_{h_n}(u,t)\right]\right| \left|\tilde{F}_n(u) - F(u)\right| du \\
&\leq D_n \int_0^\infty \left|\frac{\partial^2}{\partial t \partial u}\left[K_{h_n}(u,t)\right]\right| du \text{ where } D_n = \|\tilde{F}_n - F\|_\infty \\
&= \frac{D_n}{h_n^2}\left\{h_n^2 \int_0^\infty \left|\frac{\partial^2}{\partial t \partial u}\left[K_{h_n}(u,t)\right]\right| du\right\} \stackrel{a.s.}{\to} 0 \qquad (4.11)
\end{aligned}$$

uniformly as $h_n^2(n/\log\log n)^{1/2} \to \infty$ ($(n/\log\log n)^{1/2} D_n = O(1)$ w.p.1 from Theorem 7.2.1 in Robertson, Wright and Dykstra (1988)) and the fact that $h_n^2 \int_0^\infty |\frac{\partial^2}{\partial t \partial u}[K_{h_n}(u,t)]| du$ is uniformly bounded (as a consequence of assumption (iii) about the kernel K). To show that $\bar{f}_n'(t) \to f(t)$ uniformly on $I := [t_0 - \delta/2, t_0 + \delta/2]$, we express

$$\bar{f}_n(t) = \int_{C_n} K(v) f(te^{vh_n}) e^{vh_n} dv + \int_{I^c} K_{h_n}(u,t) f(u) du$$

where $C_n := \left[\frac{\log(t_0-\delta/2)-\log t}{h_n}, \frac{\log(t_0+\delta/2)-\log t}{h_n}\right]$. On differentiating and some simplification we have

$$|\bar{f}_n'(t) - f'(t)| \leq \int_{C_n} K(v) e^{2vh_n} \left| f'(te^{vh_n}) - f'(t) \right| dv +$$

$$\int_{I^c} \left| \frac{\partial}{\partial t} K_{h_n}(u,t) \right| f(u) du + |f'(t)| \left| \int_{C_n} K(v) e^{2vh_n} dv - 1 \right|. \tag{4.12}$$

By uniform continuity of f' on $[t_0 - \delta, t_0 + \delta]$, the first term can be made uniformly small. It is easy to see that the third term goes to zero. The second term can be shown to vanish by using properties (i) and (iii) about the kernel and an application of Cauchy-Schwarz inequality. From Equations (4.11) and (4.12) we see that $\mathbb{Z}_{2,n}(z) \stackrel{a.s.}{\to} \frac{z^2}{2} f'(t_0)$ uniformly on $[-M, M]$. Notice that,

$$\int_0^\infty |\check{f}_n(t) - \bar{f}_n(t)| dt \leq \int_0^\infty \int_0^\infty K_{h_n}(u,t) |\tilde{f}_n(u) - f(u)| du\, dt$$

$$= \int_0^\infty |\tilde{f}_n(u) - f(u)| \int_0^\infty K_{h_n}(u,t) dt\, du = \int_0^\infty |\tilde{f}_n(u) - f(u)| du \to 0 \text{ a.s.}$$

by interchanging the order of integration (and noticing that the inner integral evaluates to 1) and using Theorem 8.3 of Devroye (1987). Also note that $\bar{f}_n(t) \to f(t)$ for all $t > 0$, by an application of the dominated convergence theorem. By Scheffé's theorem, $\int_0^\infty |\bar{f}_n(t) - f(t)| dt \to 0$. Thus, we conclude that

$$\int_0^\infty |\check{f}_n(t) - f(t)| dt \to 0 \text{ a.s.}$$

Therefore, $\check{F}_n$ converges uniformly on $(0, \infty)$ to F a.s., which shows that (a) holds. Also as F has a continuous density f, $\check{f}_n(t) \to f(t)$ a.s. for every $t > 0$ by the lemma in page 330 of Robertson, Wright and Dykstra (1988). As $\check{f}_n$'s are monotonically decreasing functions converging pointwise to a continuous f, the convergence is uniform on the compact neighborhood $[t_0 - \delta, t_0 + \delta]$. Now, to show that condition (b) holds, for $z \in [-M, M]$, we use a one term Taylor series expansion to bound

$$\begin{aligned} & |n^{1/3}\{\check{F}_n(t_0 + zn^{-1/3}) - \check{F}_n(t_0)\} - zf(t_0)| \\ \leq\ & M\left\{\max_{|s| \leq Mn^{-1/3}} |\check{f}_n(t_0 + s) - f(t_0 + s)| + \max_{|s| \leq Mn^{-1/3}} |f(t_0 + s) - f(t_0)|\right\} \end{aligned}$$

which converges to 0 a.s. by the above discussion and the continuity of f. A similar argument also shows that (e) holds, with the $O(m_n^{-1/3})$ term identically 0.

To prove condition (d), let $\epsilon > 0$ be given. We use a two term Taylor expansion to bound the right-hand side of (d) as

$$\begin{aligned} & |\check{F}_n(t_0 + \beta) - \check{F}_n(t_0) - \check{f}_n(t_0)\beta - \frac{1}{2}\beta^2 f'(t_0)| \\ \leq\ & \frac{1}{2}\beta^2 \max_{|s| \leq |\beta|} |\check{f}_n'(t_0 + s) - f'(t_0)| \\ \leq\ & \frac{1}{2}\beta^2 \left\{\max_{|s| \leq |\beta|} |\check{f}_n'(t_0 + s) - f'(t_0 + s)| + \max_{|s| \leq |\beta|} |f'(t_0 + s) - f'(t_0)|\right\} \\ \leq\ & \epsilon\beta^2 + o(\beta^2). \end{aligned}$$

The last inequality follows from the uniform convergence of $\check{f}_n'(s)$ to $f'(s)$ in a neighborhood of t_0 (which is proved in Equations (4.11) and (4.12)) and the continuity of f' at t_0, by choosing a sufficiently large n and a sufficiently small neighborhood for β around 0. $\square$

4.5 m out of n Bootstrap

In Section 4.3 we showed that the two most intuitive methods of bootstrapping are inconsistent. In this section we show that the corresponding m out of n bootstrap procedures are weakly consistent. The following theorem considers generating bootstrap samples $X_{n,1}^*, X_{n,2}^*, \ldots, X_{n,m_n}^*$ from $\mathbb{F}_n$, where m_n is strictly less than n. The quantity of interest is $\Delta_n^* = m_n^{1/3}\left\{\tilde{f}_{m_n}^*(t_0) - \tilde{f}_n(t_0)\right\}$.

Theorem 6. If $m_n = o(n)$ then the bootstrap procedure is weakly consistent, i.e.,

$$\sup_{x\in\mathbb{R}} \left|P^*\left\{\Delta_n^* \leq x\right\} - P\left\{\Delta_n \leq x\right\}\right| \xrightarrow{P} 0. \tag{4.13}$$

Proof. We verify conditions (a)-(e) (with some modification) as in Theorem 4 with $F_n = \mathbb{F}_n$ and $f_n = \tilde{f}_n$ to establish the desired result. Conditions (a), (b) and (e) hold a.s. and are easy to establish.

Fix a compact set $[-M, M] \subset \mathbb{R}$. We show that (c) holds in probability, i.e., $\mathbb{Z}_{n,2}(z) \xrightarrow{P} \frac{z^2}{2} f'(t_0)$ uniformly on $[-M, M]$. Towards this end, we simplify $\mathbb{Z}_{n,2}(z)$, for $z \in [-M, M]$, in the following way

$$\begin{aligned}
& m_n^{2/3}\left\{\mathbb{F}_n(t_0 + z m_n^{-1/3}) - \mathbb{F}_n(t_0)\right\} - m_n^{1/3} z \tilde{f}_n(t_0) \\
= \ & m_n^{2/3}\left\{(\mathbb{F}_n - F)(t_0 + z m_n^{-1/3}) - (\mathbb{F}_n - F)(t_0)\right\} \\
+ \ & \left\{m_n^{1/3} z f(t_0) + \frac{z^2}{2} f'(t_0 + \alpha_n(z))\right\} - m_n^{1/3} z \tilde{f}_n(t_0) \\
& \qquad \text{where } \alpha_n(z) \text{ is between } t_0 \text{ and } t_0 + z m_n^{-1/3} \\
= \ & o_P(1) - m_n^{1/3} z \left\{\tilde{f}_n(t_0) - f(t_0)\right\} + \frac{z^2}{2} f'(t_0 + \alpha_n(z)) \\
\xrightarrow{P} \ & \frac{z^2}{2} f'(t_0) \text{ as } n \to \infty
\end{aligned} \tag{4.14}$$

as $\sup_{z\in[-M,M]}\left|(\mathbb{F}_n - F)(t_0 + z m_n^{-1/3}) - (\mathbb{F}_n - F)(t_0)\right| = O_P(n^{-1/2} m_n^{-1/6}) = o_P(m_n^{-2/3})$.

To verify condition (d), let $\epsilon > 0$ be given. By Equation (4.21) we can choose a small enough neighborhood of 0 for β and n large so that the righthand-side of (d) can be bounded by $o_P(m_n^{-2/3}) + \epsilon\beta^2 + o(\beta^2)$.

Given any subsequence $\{n_k\} \subset \mathbb{N}$, there exists a further subsequence $\{n_{k_l}\}$ such that conditions (c) and (d) hold a.s. and Theorem 4 is applicable. Thus Equation (4.9) holds for the subsequence $\{n_{k_l}\}$ which proves Equation (4.13). $\square$

The next theorem shows that the m out of n bootstrap method is also weakly consistent when we generate bootstrap samples from $\tilde{F}_n$. We will assume slightly stronger conditions on F, namely, conditions (a)-(d) mentioned in Theorem 7.2.3 of Robertson, Wright and Dykstra (1988).

Theorem 7. If $m_n = O(n(\log n)^{-3/2})$ then Equation (4.13) holds.

Proof. The proof is similar to that of Theorem 6. We only show that condition (c) holds. Letting $z \in [-M, M] \subset \mathbb{R}$, we add and subtract the term $m_n^{2/3}\left\{\mathbb{F}_n(t_0 + zm_n^{-1/3}) - \mathbb{F}_n(t_0)\right\}$ from

$$\mathbb{Z}_{n,2}(z) = m_n^{2/3}\left\{\tilde{F}_n(t_0 + zm_n^{-1/3}) - \tilde{F}_n(t_0)\right\} - m_n^{1/3} z \tilde{f}_n(t_0)$$

and then use the following result due to Kiefer and Wolfowitz (1976)

$$\begin{aligned} & \left|\left\{\tilde{F}_n(t_0 + zm_n^{-1/3}) - \tilde{F}_n(t_0)\right\} - \{\mathbb{F}_n(t_0 + zm_n^{-1/3}) - \mathbb{F}_n(t_0)\}\right| \\ \leq \;& 2\|\tilde{F}_n - \mathbb{F}_n\| = o_P(n^{-2/3}\log n) = o_P(m_n^{2/3}). \end{aligned}$$

This, coupled with the convergence of

$$m_n^{2/3}\{\mathbb{F}_n(t_0 + zm_n^{-1/3}) - \mathbb{F}_n(t_0)\} - zm_n^{1/3}\tilde{f}_n(t_0) \xrightarrow{P} \frac{z^2}{2} f'(t_0)$$

uniformly on $[-M, M]$ (see Equation (4.14)) establishes (c). $\square$

4.6 Discussion

We worked with the Grenander estimator as a prototypical example of cube-root asymptotics, but believe that our results have broader implications for the (in)-consistency of the bootstrap methods in problems with an $n^{1/3}$ convergence rate. We consider in this connection the work of Abrevaya and Huang (2005).

The setup is similar to that of Kim and Pollard (1990), where a general M-estimation framework is considered. For mathematical simplicity, we use the same notation as in Abrevaya and Huang (2005). Let $W_n := r_n(\theta_n - \theta_0)$ and $\hat{W}_n := r_n(\hat{\theta}_n - \theta_n)$ be the sample and bootstrap statistic of interest. In our case $r_n = n^{1/3}$, $\theta_0 = f(t_0)$, $\theta_n = \tilde{f}_n(t_0)$ and $\hat{\theta}_n = \tilde{f}_n^*(t_0)$. Theorem 2 of Abrevaya and Huang (2005) claims that

$$\hat{W}_n \Rightarrow \arg\max \hat{Z}(t) - \arg\max Z(t)$$

conditional on the original sample, in P^∞-probability, where $Z(t) = -\frac{1}{2}t'Vt + W(t)$ and $\hat{Z}(t) = -\frac{1}{2}t'Vt + W(t) + \hat{W}(t)$, W and $\hat{W}$ are two independent Gaussian processes, both with continuous sample paths and mean zero (see Abrevaya and Huang (2005) for more details). We also know that $W_n \Rightarrow \arg\max Z(t)$. An application of Lemma 4.1 with W_n and $\hat{W}_n$, shows that $\arg\max Z(t)$ and $\arg\max \hat{Z}(t) - \arg\max Z(t)$ should be independent. Now, if we specialize to cube-root asymptotics, we can take $Z(t) = W(t) - t^2$ and $\hat{Z}(t) = W(t) + \hat{W}(t) - t^2$, where $W(t)$ and $\hat{W}(t)$ are two independent two sided standard Brownian motions on $\mathbb{R}$ with $W(0) = \hat{W}(0) = 0$. There is abundant numerical evidence to suggest that $\arg\max Z(t)$ and $\arg\max \hat{Z}(t) - \arg\max Z(t)$ are not independent in this situation, contradicting Abrevaya and Huang's claim.

Section 4 of Abrevaya and Huang (2005) gives a method for correcting the bootstrap confidence interval. In light of the above discussion the construction of asymptotically correct bootstrap confidence intervals in this situation is suspect.

In case of the Grenander estimator, the LCM of the e.d.f. is another obvious choice for generating the bootstrap samples, as it is a concave distribution function. It is probably more natural to expect that bootstrapping from the LCM of the e.d.f. would work, as it has a well-defined probability density, while the e.d.f. does not have a density. But this bootstrap procedure is also inconsistent, and we claim that the bootstrap statistic does not have any conditional weak limit, in probability.

We have derived sufficient conditions for the consistency of bootstrap methods for this problem. Using these conditions we have shown the strong consistency of a smoothed version of bootstrap, and weak consistency of the m out of n bootstrap procedure when generating bootstrap samples from $\mathbb{F}_n$ and $\tilde{F}_n$.

4.7 Appendix section

We will use the following lemma to prove Corollary 1.

Lemma 4.3. *Let $\Psi : \mathbb{R} \to \mathbb{R}$ be a function such that $\Psi(h) \leq M$ for all $h \in \mathbb{R}$, for some $M > 0$, and*

$$\lim_{|h| \to \infty} \frac{\Psi(h)}{|h|} = -\infty. \tag{4.15}$$

Then there exists $c_0 > 0$ such that for any $c \geq c_0$, $CM_{\mathbb{R}}[\Psi](h) = CM_{[-c,c]}[\Psi](h)$ for all $|h| \leq 1$.

Proof. Note that for any $c > 0$, $CM_{\mathbb{R}}[\Psi](h) \geq CM_{[-c,c]}[\Psi](h)$ for all $h \in [-c, c]$. Let us define $\Phi_c : \mathbb{R} \to \mathbb{R}$ such that $\Phi_c(h) = CM_{[-c,c]}[\Psi](h)$ for $h \in [-1, 1]$, and Φ_c is the linear extension of $CM_{[-c,c]}[\Psi]\big|_{[-1,1]}$ outside $[-1,1]$.

We will show that there exists $c_0 > 2$ such that $\Phi_{c_0} \geq \Psi$. Then Φ_{c_0} will be a concave function everywhere greater than Ψ, and thus $\Phi_{c_0} \geq CM_{\mathbb{R}}[\Psi]$. Hence, $CM_{\mathbb{R}}[\Psi](h) \leq \Phi_{c_0}(h) = CM_{[-c_0,c_0]}[\Psi](h)$ for $h \in [-1,1]$, yielding the desired result.

For any $c > 2$, let $\Phi_c(h) = a_c + \Phi_c'(1)h$ for $h \geq 1$. Using the min-max formula, we can bound $\Phi_c'(1)$ as

$$\begin{aligned}
\Phi_c'(1) &= \min_{-c\leq s\leq 1}\max_{1\leq t\leq c}\frac{\Psi(t)-\Psi(s)}{t-s} \geq \min_{-c\leq s\leq 1}\frac{\Psi(2)-\Psi(s)}{2-s} \\
&\geq \min_{-c\leq s\leq 1}\frac{\Psi(2)-M}{2-s} = \Psi(2)-M =: B_0 \leq 0.
\end{aligned}$$

We can also bound a_c by using the inequality $\Psi(1) \leq \Phi_c(1) = a_c + \Phi_c'(1)$. Thus for $h \geq 1$,

$$\begin{aligned}
\Phi_c(h) &= a_c + \Phi_c'(1)h \geq \{\Psi(1) - \Phi_c'(1)\} + \Phi_c'(1)h \\
&\geq \Psi(1) + (h-1)B_0 \geq -K_1 h
\end{aligned} \tag{4.16}$$

for some suitably chosen $K_1 > 0$.

Similarly, for any $c > 2$, let $\Phi_c(h) = b_c + \Phi_c'(-1)h$ for $h \leq -1$. We can bound $\Phi_c'(-1)$ as

$$\begin{aligned}
\Phi_c'(-1) &= \min_{-c\leq s\leq -1}\max_{-1\leq t\leq c}\frac{\Psi(t)-\Psi(s)}{t-s} \leq \max_{-1\leq t\leq c}\frac{\Psi(t)-\Psi(-2)}{t+2} \\
&\leq \max_{-1\leq t\leq c}\frac{M-\Psi(-2)}{t+2} = M-\Psi(-2) =: B_1 \geq 0.
\end{aligned}$$

We can also bound b_c by noticing that $\Psi(-1) \leq \Phi_c(-1) = b_c - \Phi_c'(-1)$. Thus for $h \leq -1$,

$$\begin{aligned}
\Phi_c(h) &= b_c + \Phi_c'(-1)h \geq \Psi(-1) + \Phi_c'(-1)(h+1) \\
&\geq \{\Psi(-1) + B_1\} + hB_1 \geq K_2 h = -K_2|h|
\end{aligned} \tag{4.17}$$

for some suitably chosen $K_2 > 0$. Note that K_1 and K_2 do not depend on the choice of c. Given $K = \max\{K_1, K_2\}$, there exists $c_0 > 2$ such that $\Psi(h) \leq -K|h|$ for all $|h| \geq c_0$ from Equation (4.15). But from Equations (4.16) and (4.17) $\Phi_{c_0}(h) \geq -K|h|$ for all $|h| \geq 1$. Combining, we get $\Psi(h) \leq -K|h| \leq \Phi_{c_0}(h)$ for all $|h| \geq c_0 > 1$. Further, we know that $\Phi_{c_0}(h) \geq CM_{[-c_0,c_0]}[\Psi](h) \geq \Psi(h)$ for $|h| \leq c_0$. Thus we have been able to show that there exists $c_0 > 2$ such that $\Phi_{c_0} \geq \Psi$. $\square$

We will use the following easily verified fact (see Pollard (1984), page 70).

Lemma 4.4. *If $\{X_{n,c}\}, \{Y_n\}, \{W_c\}$ and Y are sets of random elements taking values in a metric space $(\mathcal{X}$,$d)$, $n = 0, 1, \ldots,$ and $c \in \mathbb{R}$ such that for any $\delta > 0$,*

(i) $\lim_{c\to\infty} \limsup_{n\to\infty} P\{d(X_{n,c}, Y_n) > \delta\} = 0,$

(ii) $\lim_{c\to\infty} P\{d(W_c, Y) > \delta\} = 0,$

(iii) $X_{n,c} \Rightarrow W_c$ *as* $n \to \infty$ *for every* $c \in \mathbb{R}$.

Then $Y_n \Rightarrow Y$ *as* $n \to \infty$.

Proof of Corollary 1. For the proof of the corollary we appeal to Lemma 5.1. We take $X_{n,c} = m_n^{1/3}\{\tilde{f}_{n,m_n,c}(t_0) - f_n(t_0)\}$ where $\tilde{f}_{n,m_n,c}(t_0)$ is the slope at t_0 of the LCM of $\mathbb{F}_{n,m_n}$ restricted to $[t_0 - cm_n^{-1/3}, t_0 + cm_n^{-1/3}]$, and $Y_n = m_n^{1/3}\{\tilde{f}_{n,m_n}(t_0) - f_n(t_0)\}$. Let us denote by $\mathbb{C}_{n,c}$ the LCM of the restriction of $\mathbb{Z}_n$ to $[-c, c]$. Also, we take W_c as the left-hand slope at 0 of $\mathbb{C}_c$, the LCM of the restriction of $\mathbb{Z}$ to $[-c, c]$, and Y as the slope at 0 of $\mathbb{C}$, the LCM of $\mathbb{Z}$.

Note that as $X_{n,c} = \mathbb{C}'_{n,c}(0) = CM_{[-c,c]}[\mathbb{Z}](0)$, an application of the usual continuous mapping theorem (see lemma on page 330 of Robertson, Wright and Dykstra (1988)) and the uniform convergence of $\mathbb{Z}_n$ to $\mathbb{Z}$ on $[-c, c]$ with condition (3) of the corollary yields $X_{n,c} \Rightarrow W_c = \mathbb{C}'_c(0)$, for every c. This shows that condition (iii) of the lemma holds.

To verify condition (ii) of the lemma we will make use of Lemma 4.3. For a.e. ω, let $c_0(\omega)$ be the smallest positive integer such that for any $c \geq c_0$, $CM_{\mathbb{R}}[\mathbb{Z}](h) = CM_{[-c,c]}[\mathbb{Z}](h)$ for all $|h| \leq 1$. Note that such a c_0 exists and is finite w.p.1. Then the event $\{W_c \neq Y\} \subset \{c_o > c\}$ and thus for any $\delta > 0$,

$$P\{d(W_c, Y) > \delta\} \leq P\{c_o > c\} \to 0 \text{ as } c \to \infty.$$

Next we show that condition (i) holds and apply Lemma 5.1 to conclude that Y_n converges to Y, thereby completing the proof of the corollary. The following series of claims are adopted from the assertion in page 217 of Kim and Pollard (1990).

Claim 1. Condition (i) of Lemma 5.1 follows if we can show the existence of random variables $\{\tau_n\}$ and $\{\sigma_n\}$ of order $O_P(1)$ such that $\tau_n < 0 \leq \sigma_n$ and $\mathbb{C}_n(\tau_n) = \mathbb{Z}_n(\tau_n)$ and $\mathbb{C}_n(\sigma_n) = \mathbb{Z}_n(\sigma_n)$.

Proof of Claim 1. Let $\epsilon > 0$ be given. As $\{\tau_n\}$ and $\{\sigma_n\}$ are of order $O_P(1)$, we can get $M_\epsilon > 0$ such that $\limsup_{n\to\infty} P\{A_\epsilon\} < \epsilon$, where $A_\epsilon = \{\tau_n < -M_\epsilon, \sigma_n > M_\epsilon\}$. Take $\omega \in A_\epsilon^c$. Then $-M_\epsilon \leq \tau_n(\omega) < 0$ and $0 \leq \sigma_n(\omega) \leq M_\epsilon$. Note that

$$\begin{aligned} \mathbb{Z}_n(\tau_n(\omega)) &\leq \mathbb{C}_{n,c}(\tau_n(\omega)) \leq \mathbb{C}_n(\tau_n(\omega)) \text{ and} \\ \mathbb{Z}_n(\sigma_n(\omega)) &\leq \mathbb{C}_{n,c}(\sigma_n(\omega)) \leq \mathbb{C}_n(\sigma_n(\omega)) \end{aligned} \tag{4.18}$$

for $c > M_\epsilon$. From the given condition in the claim we have equality in Equation (4.18) and by using a property (noted as as remark below) of concave majorants it follows that $\mathbb{C}_{n,c}(h)(\omega) = \mathbb{C}_n(h)(\omega)$ for all $h \in [\tau_n, \sigma_n]$.

REMARK. Let $[a,b] \subset B \subset \mathbb{R}$ and suppose that $CM_{[a,b]}(g)(x_1) = CM_B(g)(x_1)$ and $CM_{[a,b]}(g)(x_2) = CM_B(g)(x_2)$, for $x_1 < x_2$ in $[a,b]$. Then $CM_{[a,b]}(g)(t) = CM_B(g)(t)$ for all t in $[x_1, x_2]$.

Thus, $X_{n,c}(\omega) = Y_n(\omega)$. Therefore, $A_\epsilon^c \subset \{X_{n,c} = Y_n\}$ which implies that for any $\delta > 0$, $\limsup_{n\to\infty} P\{d(X_{n,c}, Y_n) > \delta\} \leq \limsup_{n\to\infty} P\{A_\epsilon\} < \epsilon$, for $c > M_\epsilon$. □

Therefore it suffices to show that we can construct random variables τ_n and σ_n of order $O_P(1)$ so that $\mathbb{C}_n(\tau_n) = \mathbb{Z}_n(\tau_n)$ and $\mathbb{C}_n(\sigma_n) = \mathbb{Z}_n(\sigma_n)$ for $\tau_n < 0 \leq \sigma_n$.

CLAIM 2. There exist random variables $\{\tau_n\}$ and $\{\sigma_n\}$ of order $O_P(1)$ such that $\tau_n < 0$, $\sigma_n \geq 0$ and $\mathbb{C}_n(\tau_n) = \mathbb{Z}_n(\tau_n)$ and $\mathbb{C}_n(\sigma_n) = \mathbb{Z}_n(\sigma_n)$.

PROOF OF CLAIM 2. Let K_n denote the LCM of $\mathbb{F}_{n,m_n}$. The line through $(t_0, K_n(t_0))$ with slope $\tilde{f}_{n,m_n}(t_0)$ must lie above $\mathbb{F}_{n,m_n}$ touching it at the two points $t_0 - L_n$ and $t_0 + R_n$, where $L_n > 0$ and $R_n \geq 0$. Note that $t_0 - L_n$ and $t_0 + R_n$ are the nearest points to t_0 such that K_n and $\mathbb{F}_{n,m_n}$ coincide. The line segment from $(t_0 - L_n, \mathbb{F}_{n,m_n}(t_0 - L_n))$ to $(t_0 + R_n, \mathbb{F}_{n,m_n}(t_0 + R_n))$ makes up part of K_n. It will suffice to show that $L_n = O_P(m_n^{-1/3})$, as then $\tau_n := -m_n^{1/3} L_n = O_P(1)$. The argument depends on the inequality

$$K_n(t_0) + \tilde{f}_{n,m_n}(t_0)\beta \geq \mathbb{F}_{n,m_n}(t_0 + \beta) \text{ for all } \beta,$$

with equality at $\beta = -L_n$ and $\beta = R_n$.

Let $\Gamma_n(\beta) = \mathbb{F}_{n,m_n}(t_0 + \beta) - \mathbb{F}_{n,m_n}(t_0) - \beta \tilde{f}_{n,m_n}(t_0)$. Γ_n is the distance between $\mathbb{F}_{n,m_n}(t_0+\beta)$ and $\mathbb{F}_{n,m_n}(t_0)+\beta\tilde{f}_{n,m_n}(t_0)$. It follows that $\Gamma_n(\beta)$ achieves its maximum at $\beta = -L_n$ and $\beta = R_n$ and $\Gamma_n(-L_n) = \Gamma_n(R_n)$. We can easily show using condition (a) that L_n, R_n and $\gamma_n := \tilde{f}_{n,m_n}(t_0) - f_n(t_0)$ are of order

$o_P(1)$. That lets us argue locally. Let

$$g_n(y,\beta) := \mathbf{1}\{y \le t_0+\beta\} - \mathbf{1}\{y \le t_0\} - f_n(t_0)\beta.$$

CLAIM 3. For any $\epsilon > 0$, we have

$$\frac{1}{m_n}\left|\sum_{i=1}^{m_n}\{g_n(X_{n,i},\beta) - Eg_n(X_{n,i},\beta)\}\right| \le \epsilon\beta^2 + O_P(m_n^{-2/3})$$

uniformly over β in a neighborhood of zero.

For the time being, we assume Claim 3, which is proved below. From condition (d), $|Eg_n(\cdot,\beta) - \frac{1}{2}\beta^2 f'(t_0)| \le \epsilon\beta^2 + o(\beta^2) + O(m_n^{-2/3})$ for sufficiently large n. Thus

$$\begin{aligned} & |\Gamma_n(\beta) + \beta\gamma_n - \frac{1}{2}\beta^2 f'(t_0)| \\ = \ & |\mathbb{F}_{n,m_n}(t_0+\beta) - \mathbb{F}_{n,m_n}(t_0) - \beta f_n(t_0) - \frac{1}{2}\beta^2 f'(t_0)| \\ \le \ & 2\epsilon\beta^2 + o(\beta^2) + O_P(m_n^{-2/3}) \end{aligned} \tag{4.19}$$

uniformly for β in a neighborhood of 0 by Claim 3 and the triangle inequality. As $f'(t_0) < 0$, for $n \to \infty$, there exist constants $c_1, c_2 > 0$ such that, with probability tending to 1, for β in a small neighborhood of 0,

$$-\frac{1}{2}c_2\beta^2 - \beta\gamma_n - O_P(m_n^{-2/3}) \le \Gamma_n(\beta) \le -\frac{1}{2}c_1\beta^2 - \beta\gamma_n + O_P(m_n^{-2/3}).$$

The quadratic $-\frac{1}{2}c_1\beta^2 - \beta\gamma_n$ assumes its maximum of $\frac{1}{2}\gamma_n^2/c_1$ at $-\gamma_n/c_1$, and takes negative values for those β with the same sign of γ_n. It follows that with probability tending to 1,

$$\max_\beta \Gamma_n(\beta) = \min(\Gamma_n(-L_n), \Gamma_n(R_n)) \le O_P(m_n^{-2/3}).$$

We also have $\max_\beta \Gamma_n(\beta) \ge \Gamma_n(-\gamma_n/c_2) \ge \frac{1}{2}\gamma_n^2/c_2 - O_P(m_n^{-2/3})$.

These two bounds imply that $\gamma_n = O_P(m_n^{-1/3})$. With this rate for convergence for $\{\gamma_n\}$ we can now deduce from the inequalities

$$0 = \Gamma_n(0) \le \Gamma_n(-L_n) \le \frac{1}{2}c_1(L_n - \gamma_n/c_1)^2 + \frac{1}{2}\gamma_n^2/c_1 + O_P(m_n^{-2/3})$$

that $L_n = O_P(m_n^{-1/3})$, as required. Similarly, we can show that $R_n = O_P(m_n^{-1/3})$.

Proof of Claim 3. Let us define $\mathbb{G}_n(\beta)$ as

$$\frac{1}{m_n}\sum_{i=1}^{m_n}\{g_n(X_{n,i},\beta) - Eg_n(\cdot,\beta)\} = (\mathbb{F}_{n,m_n} - F_n)(t_0+\beta) - (\mathbb{F}_{n,m_n} - F_n)(t_0).$$

We will show that $|\mathbb{G}_n(\beta)| \le \epsilon\beta^2 + m_n^{-2/3}M_n^2$ uniformly over a neighborhood of 0, for M_n of order $O_P(1)$. We fix a neighborhood $[-b,b]$ for β obtained from condition (e). We define $M_n(\omega)$ as the infimum (possibly $+\infty$) of those values for which the asserted uniform inequality holds. Let us define $A(n,j)$ to be the set of those β in $[-b,b]$ for which $(j-1)m_n^{-1/3} \le |\beta| < jm_n^{-1/3}$. Then for m constant,

$$\begin{aligned}
P\{M_n > m\} &\le P\{\exists\beta\in[-b,b] : |\mathbb{G}_n(\beta)| > \epsilon\beta^2 + m_n^{-2/3}m^2\}\\
&\le \sum_{j:jm_n^{-1/3}\le b} P\{\exists\beta\in A(n,j) : m_n^{2/3}|\mathbb{G}_n(\beta)| > \epsilon(j-1)^2 + m^2\}\\
&\le \sum_{j:jm_n^{-1/3}\le b} \frac{E\left(\sup_{|\beta|<jm_n^{-1/3}} m_n^{4/3}|\mathbb{G}_n(\beta)|^2\right)}{\{\epsilon(j-1)^2+m^2\}^2}\\
&\le \sum_{j:jm_n^{-1/3}\le b} \frac{C'j}{\{\epsilon(j-1)^2+m^2\}^2} \qquad (4.20)
\end{aligned}$$

for m_n sufficiently large. The last inequality follows from a maximal inequality as in part (ii) of Result 3.1 of Kim and Pollard (1990) and using condition (e). To be more precise, fix $j \ge 1$ such that $jm_n^{-1/3} \le b$ and let $\mathcal{F} := \{h_\beta : |\beta| < jm_n^{-1/3}\}$ be a collection of functions where $h_\beta(x) = \mathbf{1}\{x \le t_0+\beta\} - \mathbf{1}\{x \le t_0\}$. Note that $\mathcal{F}$ is a class of functions with envelope function $H(x) = \mathbf{1}\{x \le t_0 + jm_n^{-1/3}\} - \mathbf{1}\{x \le t_0 - jm_n^{-1/3}\}$. From the maximal inequality in 3.1 of Kim and Pollard (1990) we can bound $m_n^{4/3}E(\sup_{\mathcal{F}}|\mathbb{G}_n(\beta)|^2)$ by

$$J^2(1)m_n^{1/3}\{F_n(t_0+jm_n^{-1/3}) - F_n(t_0 - jm_n^{-1/3})\} \le C'j$$

for n sufficiently large, by adding and subtracting $F_n(t_0)$ and using conditions (e), where J is a continuous and increasing function with $J(0) = 0$ and $J(1) < \infty$, not depending on n and C is a constant. We can therefore ensure that the sum in Equation (5.37) is suitably small for large m_n by choosing m large enough. This proves the claim. $\square$

Proof of Corollary 2. To prove the corollary we appeal to Lemma 5.1 by establishing conditions (i)-(iii) (in the lemma) with $X_{n,c} = CM_{[-c,c]}[\mathbb{Z}_n]'(0)$, $Y_n = CM_{I_n}[\mathbb{Z}_n]'(0)$, $W_c = CM_{[-c,c]}[\mathbb{Z}]'(0)$ and $Y = CM_{\mathbb{R}}[\mathbb{Z}]'(0)$. Note that the process $\mathbb{Z}$ satisfies conditions (1)-(3) of Corollary 1 and so condition (ii) of the lemma holds. An application of the continuous mapping theorem and the uniform convergence of $\mathbb{Z}_n$ to $\mathbb{Z}$ on $[-c, c]$ yields condition (iii).

If we can show that (τ_n, σ_n), defined as in the proof of Claim 2 of Corollary 1 with $m_n = n$, $F_n = \mathbb{F}_n$, and $\mathbb{F}_{n,m_n} = \mathbb{F}_n^*$, are of order $O_P(1)$, then using Claim 1 in the proof of Corollary 1 we can establish condition (i). But this step requires a bit of work. Although the argument is similar to that of the proof of Claim 2 of Corollary 1, there are some subtle differences. Note that here we want to study the unconditional behavior of (τ_n, σ_n), and so $F_n = \mathbb{F}_n$ cannot be treated as fixed.

As a first step we show that slightly modified versions of conditions (a), (d) and (e), to be used later in the proof, are satisfied. Condition (a) trivially holds a.s. Condition (e) also holds a.s. and can be verified using Equation (4.7). Note that the neighborhood for β around 0 in condition (e) can be chosen to be a fixed interval a.s. (not depending on $\mathbf{X}$, but possibly on F). Let $\epsilon > 0$ be given. We show that condition (d) holds with $O(m_n^{-2/3})$ replaced by

$O_P(n^{-2/3})$. The term of interest can be grouped as

$$\begin{aligned}
& \left|\mathbb{F}_n(t_0+\beta) - \mathbb{F}_n(t_0) - \beta\tilde{f}_n(t_0) - \frac{1}{2}\beta^2 f'(t_0)\right| \\
\leq\ & |(\mathbb{F}_n - F)(t_0+\beta) - (\mathbb{F}_n - F)(t_0)| + |\beta|\left|\tilde{f}_n(t_0) - f(t_0)\right| \\
+\ & \left|F(t_0+\beta) - F(t_0) - \beta f(t_0) - \frac{1}{2}\beta^2 f'(t_0)\right| \\
\leq\ & \epsilon\beta^2 + o(\beta^2) + O_P(n^{-2/3}). \qquad (4.21)
\end{aligned}$$

The first term can be bounded by $\left\{O_P(n^{-2/3}) + \frac{1}{2}\epsilon\beta^2\right\}$ uniformly for β in a small neighborhood of 0, using Claim 3 in the proof of Corollary 1 with $\mathbb{F}_{n,m_n} = \mathbb{F}_n$ and $F_n = F$. The second term $|\beta|\left|\tilde{f}_n(t_0) - f(t_0)\right|$ can be bounded by

$$\frac{1}{2}\epsilon\beta^2 + \frac{1}{2\epsilon}\left|\tilde{f}_n(t_0) - f(t_0)\right|^2 = \frac{1}{2}\epsilon\beta^2 + O_P(n^{-2/3}). \qquad (4.22)$$

By Taylor expansion it is easy to see that the third term is of order $o(\beta^2)$.

Next we define L_n, R_n and γ_n as in Claim 2. It is easy to show that L_n, R_n and γ_n are of order $o_P(1)$, using condition (a). The main crux of the argument in the proof of Claim 2 of Corollary 1 is establishing Equation (5.36) uniformly for β in a neighborhood of 0. We show that Equation (5.36) still holds unconditionally in our context, thereby yielding $(\tau_n, \sigma_n) = O_P(1)$, from the discussion succeeding the equation. Observe that,

$$\left|\Gamma_n(\beta) + \beta\gamma_n - \frac{1}{2}\beta^2 f'(t_0)\right| \quad = \quad \left|\mathbb{F}_n^*(t_0+\beta) - \mathbb{F}_n^*(t_0) - \beta\tilde{f}_n(t_0) - \frac{1}{2}\beta^2 f'(t_0)\right|$$

can be bounded by the sum of $\left|\mathbb{F}_n(t_0+\beta) - \mathbb{F}_n(t_0) - \beta\tilde{f}_n(t_0) - \frac{1}{2}\beta^2 f'(t_0)\right|$ and $|(\mathbb{F}_n^* - \mathbb{F}_n)(t_0+\beta) - (\mathbb{F}_n^* - \mathbb{F}_n)(t_0)|$. Equation (4.21) is employed to bound the first term, whereas the following result

$$|(\mathbb{F}_n^* - \mathbb{F}_n)(t_0+\beta) - (\mathbb{F}_n^* - \mathbb{F}_n)(t_0)| \leq \epsilon\beta^2 + O_P(n^{-2/3})$$

bounds the second. Combining, we have

$$\left|\Gamma_n(\beta) + \beta\gamma_n - \frac{1}{2}\beta^2 f'(t_0)\right| \;\leq\; 2\epsilon\beta^2 + o(\beta^2) + O_P(n^{-2/3})$$

for β in a neighborhood of 0. Note that an application of the maximal inequality as in the proof of Claim 3, conditional on $\mathbf{X}$, gives us the bound

$$|(\mathbb{F}_n^* - \mathbb{F}_n)(t_0 + \beta) - (\mathbb{F}_n^* - \mathbb{F}_n)(t_0)| \leq \epsilon\beta^2 + T_n$$

uniformly for β in a neighborhood of 0, not depending on $\mathbf{X}$, where $T_n = O_{P^*}(n^{-2/3})$ a.s. From the following series of inequalities it follows that $T_n = O_P(n^{-2/3})$. Suppose that $\{S_n\}$ is a sequence of random variables that are $O_{P^*}(1)$ a.s., i.e,

$$\begin{aligned}
\lim_{T\to\infty} \limsup_{n\to\infty} P^*\{|S_n| \geq T\} \;&\to\; 0 \text{ a.s., then} \\
\lim_{T\to\infty} \limsup_{n\to\infty} P\{|S_n| \geq T\} \;&=\; \lim_{T\to\infty} \limsup_{n\to\infty} E[P^*\{|S_n| \geq T\}] \\
\leq \lim_{T\to\infty} E\left[\limsup_{n\to\infty} P^*\{|S_n| \geq T\}\right] \;&=\; E\left[\lim_{T\to\infty} \limsup_{n\to\infty} P^*\{|S_n| \geq T\}\right] = 0
\end{aligned}$$

by an application of Fatou's lemma and the dominated convergence theorem. □

Proof of Theorem 3 (iii). We use Lemma 5.1 to prove the result. Note that here (τ_n, σ_n) are defined as in the proof of Claim 2 of Corollary 1 with $m_n = n$, $F_n = \tilde{F}_n$, and $\mathbb{F}_{n,m_n} = \mathbb{F}_n^*$. The proof is very similar to that of Corollary 2. We only need to show that (τ_n, σ_n) are of order $O_P(1)$. Conditions (a) and (e) hold a.s. It is enough to show that condition (d) holds with the $O(m_n^{-2/3})$ term replaced by $O_P(n^{-2/3})$, with probability increasing to 1; as then Equation (5.36) holds, and from the discussion succeeding the equation it follows that (τ_n, σ_n) are of order $O_P(1)$.

Let $\epsilon > 0$ be given. Without loss of generality we can assume that $f'(t_0) < -4\epsilon$. It is enough to show that

$$\left|\tilde{F}_n(t_0+\beta) - \mathbb{F}_n(t_0) - \beta f(t_0) - \frac{1}{2}\beta^2 f'(t_0)\right| \leq 2\epsilon\beta^2 + O_P(n^{-2/3}) \tag{4.23}$$

uniformly in a neighborhood of 0, as we can bound the left hand-side of (d) by $\left|\tilde{F}_n(t_0+\beta) - \mathbb{F}_n(t_0) - \beta f(t_0) - \frac{1}{2}\beta^2 f'(t_0)\right| + \left|\mathbb{F}_n(t_0) - \tilde{F}_n(t_0)\right| + |\beta||\tilde{f}_n(t_0) - f(t_0)|$, where the second term is $O_P(n^{-2/3})$ (by Theorem 1 of Wang (1994)) and the third term can be bounded by $\epsilon\beta^2 + O_P(n^{-2/3})$ (see Equation (4.22)).

Given ϵ, there exists a neighborhood of 0 for β such that

$$\left|F(t_0+\beta) - F(t_0) - \beta f(t_0) - \frac{1}{2}\beta^2 f'(t_0)\right| \;\leq\; \epsilon\beta^2$$

by the twice differentiability of F at t_0. Thus, there exists $\delta > 0$, such that

$$\left|\mathbb{F}_n(t_0+\beta) - \mathbb{F}_n(t_0) - \beta f(t_0) - \frac{1}{2}\beta^2 f'(t_0)\right| \;\leq\; 2\epsilon\beta^2 + O_P(n^{-2/3}) \tag{4.24}$$

uniformly for $\beta \in [-2\delta, 2\delta]$, by the discussion following Equation (4.21). Therefore, for $\beta \in [-2\delta, 2\delta]$,

$$\mathbb{F}_n(t_0+\beta) - \mathbb{F}_n(t_0) \leq 2\epsilon\beta^2 + \beta f(t_0) + \frac{1}{2}\beta^2 f'(t_0) + O_P(n^{-2/3}) \tag{4.25}$$

Letting $\tilde{F}_n^\delta$ be the LCM of the restriction of $\mathbb{F}_n$ on $[-2\delta, 2\delta]$, we have,

$$\tilde{F}_n^\delta(t_0+\beta) - \mathbb{F}_n(t_0) \;\leq\; 2\epsilon\beta^2 + \beta f(t_0) + \frac{\beta^2}{2} f'(t_0) + O_P(n^{-2/3})$$

for $\beta \in [-2\delta, 2\delta]$, by taking concave majorants on both sides of Equation (4.25) and realizing that the $O_P(n^{-2/3})$ is uniform in β. Since $\tilde{F}_n \geq \mathbb{F}_n$, it is immediate from Equation (4.24) that

$$\tilde{F}_n(t_0+\beta) - \mathbb{F}_n(t_0) \;\geq\; -2\epsilon\beta^2 + \beta f(t_0) + \frac{\beta^2}{2} f'(t_0) - O_P(n^{-2/3}).$$

Letting

$$A_n := \left\{\tilde{F}_n^\delta(t_0+\beta) = \tilde{F}_n(t_0+\beta) \text{ for all } \beta \in [-\delta, \delta]\right\}$$

it is easy to show from the strict concavity of F around t_0 that $\lim_{n\to\infty} P\{A_n\} = 1$ (for a complete proof of this see Proposition 6.1 of Wang and Woodroofe (2007)). Thus Equation (4.23) holds with probability tending to 1 on $[-\delta, \delta]$. This completes the argument. $\square$

CHAPTER 5

Bootstrap in the Wicksell's problem

Let $\mathbf{X} = (X_1, X_2, X_3)$ be a spherically symmetric random vector of which only (X_1, X_2) can be observed. We focus attention on estimating F, the distribution function of the squared radius $Z := X_1^2 + X_2^2 + X_3^2$, from a random sample of (X_1, X_2). We relate F to a function V which is decreasing and can be estimated from observed data. We define three estimators of F and derive their limit distributions. The non-standard asymptotics involved manifests itself with a nonstandard rate of convergence $\sqrt{\frac{n}{\log n}}$. We show that the isotonized estimator of V and F have exactly half the limiting variance when compared to the naive estimators, which does not incorporate the shape constraint. We also state sufficient conditions for the consistency of any bootstrap procedure in constructing confidence intervals for V and F and show that the conditions are met by the conventional bootstrap method (while generating samples from the empirical distribution function).

5.1 Introduction

Stereology is the study of three-dimensional properties of objects or matter usually observed two-dimensionally. We consider such a problem, which arises in Astronomy, where the quantity of interest can be related to functions that obey shape restrictions. Our treatment is similar in flavor to Groeneboom and Jongbloed's (1995) study of the Wicksell's (1925) "Corpuscle problem".

Let $\mathbf{X} = (X_1, X_2, X_3)$ be a spherically symmetric random vector denoting the three dimensional position of a star in a galaxy. But we can only observe the projected stellar positions, i.e., (X_1, X_2) (with a proper choice of co-ordinates). We are interested in estimating F, the distribution function of the squared radius $Z := X_1^2 + X_2^2 + X_3^2$ from a random sample of (X_1, X_2). Suppose that $\mathbf{X}$ has density $\rho(z)$, $z = x_1^2 + x_2^2 + x_3^2$ and $Y := X_1^2 + X_2^2$ has density g. Then

$$g(y) = \pi \int_y^\infty \frac{\rho(z)}{\sqrt{z-y}}\, dz. \tag{5.1}$$

The reader may recognize Equation (5.1) as Abel's transformation. This may be inverted as follows. Let

$$V(y) := \int_y^\infty \frac{g(u)}{\sqrt{u-y}}\, du,$$

then we see that

$$V(y) = \pi \int_y^\infty \int_u^\infty \frac{\rho(z)\, dz\, du}{\sqrt{z-u}\sqrt{u-y}} = \pi^2 \int_y^\infty \rho(z)\, dz, \tag{5.2}$$

which shows that V is a non-increasing function. A natural (unbiased) "naive" estimator of V is

$$V_n(y) = \int_y^\infty \frac{d\mathbb{G}_n(u)}{\sqrt{u-y}}$$

where $\mathbb{G}_n$ is the e.d.f. of a sample of squared circle radii. This naive estimator can be improved by imposing the shape constraint. If V_n were square integrable, this could be accomplished by minimizing the integral of $(W - V_n)^2$ over all non-increasing functions W, or equivalently, by minimizing

$$\int_0^\infty W^2(y)\, dy - 2\int_0^\infty W(y)V_n(y)\, dy. \tag{5.3}$$

The function V_n is not square integrable, but it is integrable, so Equation (5.3) is well defined. Let $\tilde{V}_n$ be the non-increasing function W that minimizes Equation (5.3). Existence and uniqueness can be shown along the lines of Theorem 1.2.1 of Robertson, Wright, and Dykstra (1988), replacing the sums by integrals.

Groeneboom and Jongbloed (1995) derived the limit distributions of V_n and $\tilde{V}_n$: Let $x_0 > 0$ and $\epsilon_n = \sqrt{n^{-1}\log n}$, then under appropriate conditions,

$$\epsilon_n^{-1}\{V_n(x_0) - V(x_0)\} \Rightarrow N(0, g(x_0)) \text{ and} \tag{5.4}$$

$$\epsilon_n^{-1}\{\tilde{V}_n(x_0) - V(x_0)\} \Rightarrow N\left(0, \frac{1}{2}g(x_0)\right). \tag{5.5}$$

The quantity of interest, F, can be related to V as

$$F(x) = \int_0^x 2\pi\sqrt{u}\rho(u)\,du = 1 + \frac{2}{\pi}\int_x^\infty \sqrt{z}\,dV(z) \tag{5.6}$$

where the last equality follows from noting that $2\pi\int_0^\infty \sqrt{u}\,\rho(u)\,du = 1$. Using V_n and $\tilde{V}_n$ we can define two estimators of F as

$$F_n(x) = 1 + \frac{2}{\pi}\int_x^\infty \sqrt{z}\,dV_n(z) \text{ and} \tag{5.7}$$

$$\tilde{F}_n(x) = 1 + \frac{2}{\pi}\int_x^\infty \sqrt{z}\,d\tilde{V}_n(z). \tag{5.8}$$

Note that F_n is not even non-decreasing. The restriction of $\tilde{F}_n$ to $[0,1]$, i.e., $\max\{\tilde{F}_n, 0\}$ (as $\tilde{F}_n \leq 1$), is a valid distribution function and a much more appealing estimator of F. Yet another estimator of F can be gotten by isotonizing F_n over all non-decreasing functions. Let $\check{F}_n$ be the non-decreasing function that is closest to F_n, in the sense that it minimizes Equation (5.3) with V_n replaced by F_n. It is not difficult to see that then $\max\{0, \min(\check{F}_n, 1)\}$

is a valid distribution function. It will be shown later that for $x_0 > 0$,

$$
\begin{aligned}
&(5.9) \qquad & \epsilon_n^{-1}\{F_n(x_0) - F(x_0)\} \;&\Rightarrow\; N\left(0, \frac{4}{\pi^2} x_0 g(x_0)\right), \\
&(5.10) \qquad & \epsilon_n^{-1}\{\tilde{F}_n(x_0) - F(x_0)\} \;&\Rightarrow\; N\left(0, \frac{2}{\pi^2} x_0 g(x_0)\right) \text{ and} \\
&(5.11) \qquad & \epsilon_n^{-1}\{\check{F}_n(x_0) - F(x_0)\} \;&\Rightarrow\; N\left(0, \frac{2}{\pi^2} x_0 g(x_0)\right).
\end{aligned}
$$

Notice that the isotonized estimators have exactly half limiting variances when compared to the corresponding naive estimators. Construction of confidence intervals for $F(x_0)$ using these limiting distributions is still complicated as it requires the estimation of the nuisance parameter $g(x_0)$. Bootstrap intervals avoid this problem and are generally reliable and accurate in problems with $\sqrt{n}$ convergence rate (see Bickel and Freedman (1981), Singh (1981), Shao and Tu (1995) and its references). In this chapter we also investigate the consistency of bootstrap procedures for constructing pointwise confidence intervals around these shape constrained functions.

In Section 5.2 we prove uniform versions of Equations (5.4), (5.5), (5.9), (5.10) and (5.11) that are utilized in the later sections. Section 5.3 establishes the consistency of bootstrap methods in approximating the sampling distribution of the various estimators of V and F while generating samples from the e.d.f. Section 5.4, the Appendix, gives the details of some of the arguments in the proofs of the main results.

5.2 Preliminaries

Let Y be a random variable with c.d.f. G and density g where g is related to ρ according to Equation (5.1). Assume that g is continuous on $[0, \infty)$. We also define the following functions

$$V(y) := \int_y^\infty \frac{g(u)}{\sqrt{u-y}}\,du \text{ and } U(x) := \int_0^x V(t)\,dt.$$

From Equation (5.2) we see that V is non-increasing. We can simplify U to obtain

$$U(x) = \int_0^\infty \int_0^{u\wedge x} \frac{dt}{\sqrt{u-t}}\,g(u)\,du \quad = \quad 2\int_0^\infty \{\sqrt{u} - \sqrt{(u-x)_+}\}\,g(u)\,du$$

where $y_+ = \max\{y,0\}$. Letting $J(t) = \int_t^\infty \sqrt{z-t}\,dV(z)$, we express G as

$$\begin{aligned} G(t) &= \int_0^t g(y)\,dy = \pi\int_0^\infty \int_0^{t\wedge z} \frac{\rho(z)}{\sqrt{z-y}}\,dy\,dz \\ &= 2\pi\int_0^\infty \{\sqrt{z} - \sqrt{(z-t)_+}\}\rho(z)\,dz = 1 + \frac{2}{\pi}J(t). \end{aligned} \tag{5.12}$$

Suppose that we have i.i.d. triangular data $\{Y_{n,i}\}_{i=1}^n$ having distribution function G_n. We consider a special construction of $Y_{n,i}$, namely, let $Y_{n,i} = G_n^{-1}(T_i)$, where $G_n^{-1}(u) = \inf\{x : G_n(x) \geq u\}$ and $T_1, T_2, \ldots$ are i.i.d. Uniform$(0,1)$ random variables. Let V_n and U_n be defined as

$$V_n(y) := \int_y^\infty \frac{dG_n(u)}{\sqrt{u-y}} \text{ and } U_n(x) := \int_0^x V_n(y)\,dy. \tag{5.13}$$

Let LCM_I be the operator that maps a function $h : \mathbb{R} \to \mathbb{R}$ into the least concave majorant (LCM) of h on the interval $I \subset \mathbb{R}$. Define $\tilde{V}_n := LCM_{[0,\infty)}[U_n]'$ where $'$ denotes the right derivative. Note that

$$V_n^\#(y) := \int_y^\infty \frac{d\mathbb{G}_n(u)}{\sqrt{u-y}} = \frac{1}{n}\sum_{i:Y_{n,i}>y} \frac{1}{\sqrt{Y_{n,i}-y}},$$

where $\mathbb{G}_n$ is the e.d.f. of $Y_{n,1}, Y_{n,2}, \ldots, Y_{n,n}$, is an unbiased estimate of $V_n(y)$, but not monotonic when viewed as a function of y; $V_n^\#$ has an infinite jump at each observation $Y_{n,i}$. We will call $V_n^\#$ as the *naive* estimator. This naive estimator can be improved by imposing the shape constraint as in Equation (5.3) with V_n replaced by $V_n^{\#}$. Let $\tilde{V}_n^{\#}$ be the non-increasing function W that minimizes Equation (5.3).

Observe that

$$U_n^{\#}(x) := \frac{2}{n}\sum_{i=1}^{n}\left\{\sqrt{Y_{n,i}} - \sqrt{(Y_{n,i}-x)_+}\right\}$$

is an unbiased estimate of $U_n(x)$ for all $x \in [0,\infty)$; $U_n^{\#}$ is a non-decreasing function; $V_n^{\#}$ is the derivative of $U_n^{\#}$ a.e. Let $\tilde{U}_n^{\#}$ be the LCM of $U_n^{\#}$. Then $\tilde{V}_n^{\#}$ is the right-derivative of $\tilde{U}_n^{\#}$. Let us also define F_n and $F_n^{\#}$ as

$$\begin{aligned}
F_n(z) &:= 1 + \frac{2}{\pi}\int_z^\infty \sqrt{x}\, dV_n(x) \\
&= 1 + \frac{2}{\pi}\left\{-\sqrt{z}V_n(z) - \int_z^\infty \frac{V_n(u)}{2\sqrt{u}}\, du\right\} \text{ and} \\
F_n^{\#}(z) &:= 1 + \frac{2}{\pi}\int_z^\infty \sqrt{x}\, dV_n^{\#}(x) \\
&= 1 + \frac{2}{\pi}\left\{-\sqrt{z}V_n^{\#}(z) - \int_z^\infty \frac{V_n^{\#}(u)}{2\sqrt{u}}\, du\right\}. \qquad (5.14)
\end{aligned}$$

5.2.1 Uniform CLT for estimates of V

Fix $x_0 \in (0,\infty)$. We consider two estimates of V, namely $V_n^{\#}$ and $\tilde{V}_n^{\#}$. The limit distribution of $V_n^{\#}$ is easily obtainable from the following proposition.

Proposition 5. If $g(x_0) > 0$, then $\epsilon_n^{-1}\{V_n^{\#}(x_0) - V_n(x_0)\} \Rightarrow N\left(0, g(x_0)\right)$.

Proof of Proposition 5. Applying the triangular central limit theorem for sums of independent random variables with infinite variances (similar to Theorem 4 of Chow and Teicher (1988), page 305) to the random variables $\frac{\mathbf{1}\{Y_{n,i}>z\}}{\sqrt{Y_{n,i}-z}}$, we obtain the desired result. □

Next we study the limit distribution of

$$\Delta_n := \epsilon_n^{-1}\{\tilde{V}_n^{\#}(x_0) - \hat{V}_n(x_0)\}$$

where $\hat{V}_n(x_0)$ can be $V_n(x_0)$ or $\tilde{V}_n(x_0)$. Note that Δ_n is the right-hand slope at 0 of the LCM of the process

$$\mathbb{Z}_n(t) = \epsilon_n^{-2}\{U_n^{\#}(x_0 + \epsilon_n t) - U_n^{\#}(x_0) - \hat{V}_n(x_0)\epsilon_n t\}$$

for $t \in I_n := [-\epsilon_n^{-1}x_0, \infty)$. We will study the limiting behavior of the process $\mathbb{Z}_n$ and use continuous mapping arguments to derive the limiting distribution of Δ_n. We consider all stochastic processes as random elements in $C(\mathbb{R})$, the space of continuous functions on $\mathbb{R}$, and equip it with the Borel σ-field and the metric of uniform convergence on compacta.

To better understand the limiting behavior of $\mathbb{Z}_n$, we decompose $\mathbb{Z}_n$ into $\mathbb{Z}_{n,1}$ and $\mathbb{Z}_{n,2}$ where

$$\begin{aligned} \mathbb{Z}_{n,1}(t) &= \epsilon_n^{-2}\{(U_n^{\#} - U_n)(x_0 + \epsilon_n t) - (U_n^{\#} - U_n)(x_0)\} \text{ and} \\ \mathbb{Z}_{n,2}(t) &= \epsilon_n^{-2}\{U_n(x_0 + \epsilon_n t) - U_n(x_0) - \hat{V}_n(x_0)\epsilon_n t\} \end{aligned} \tag{5.15}$$

Note that $\Delta_n = LCM_{I_n}[\mathbb{Z}_n]'(0)$. We define the processes

$$\mathbb{Z}_1(t) = tW \text{ and } \mathbb{Z}(t) = \mathbb{Z}_1(t) + \frac{1}{2}t^2 V'(x_0),$$

for $t \in \mathbb{R}$, where W is a normal random variable having mean 0 and variance $\frac{1}{2}g(x_0)$. We state some conditions on the behavior of G_n, $\hat{V}_n$ and U_n used to obtain the limiting distribution of Δ_n.

(a) $D_n := \|G_n - G\| = O(\epsilon_n)$.

(b) $\mathbb{Z}_{n,2}(t) \to \frac{1}{2}t^2 V'(x_0)$ as $n \to \infty$ uniformly on compacta.

(c) for each $\epsilon > 0$,

$$\left| U_n(x_0 + \beta) - U_n(x_0) - \beta\hat{V}_n(x_0) - \frac{1}{2}\beta^2 V'(x_0) \right| \le \epsilon\beta^2 + o(\beta^2) + O(\epsilon_n^2)$$

for large n, uniformly in β varying over a neighborhood of zero.

Theorem 8. Under condition (a) the distribution of $\mathbb{Z}_{n,1}$ converges to that of $\mathbb{Z}_1$. Further, if (b) holds, then the distribution of $\mathbb{Z}_n$ converges to that of $\mathbb{Z}$.

Proof of Theorem 8. Fix a compact set $K = [-M, M]$, $M > 0$. We will show that $\mathbb{Z}_{n,1}$ converges weakly to $\mathbb{Z}_1$ in the metric of uniform convergence on K. Note that $\mathbb{Z}_{n,1}(t)$ has mean 0 for all $t \in I_n$. To compute the covariance of $\mathbb{Z}_{n,1}(s)$ and $\mathbb{Z}_{n,1}(t)$, for $s \leq t \in K$, we define the function

$$\phi(y, \eta) = \sqrt{(y - x_0)_+} - \sqrt{(y - x_0 - \eta)_+}$$

for $y, \eta \in \mathbb{R}$. The two following properties of $\phi(y, \eta)$ will be used in the sequel.

(P1) $|\phi(\cdot, \eta)| \leq \sqrt{|\eta|}$.

(P2) $\int_0^\infty |\phi'(y, \eta)|\, dy = 2\sqrt{|\eta|}$. The result follows as, for $\eta > 0$,

$$\begin{aligned} \int_0^\infty |\phi'(y, \eta)|\, dy &= \int_{x_0}^{x_0+\eta} \frac{dy}{2\sqrt{y - x_0}} + \int_{x_0+\eta}^\infty \left\{ \frac{1}{2\sqrt{y - x_0 - \eta}} - \frac{1}{2\sqrt{y - x_0}} \right\} dy \\ &= \{\sqrt{\eta} + \sqrt{\eta}\} = 2\sqrt{\eta}. \end{aligned}$$

Observe that

$$\mathbb{Z}_{n,1}(t) = \frac{2}{\epsilon_n^2} \int \phi(u, \epsilon_n t)\, d(\mathbb{G}_n - G_n)(u) \text{ and}$$

$$Cov(\mathbb{Z}_{n,1}(s), \mathbb{Z}_{n,1}(t)) = \frac{4}{n\epsilon_n^4} Cov(\phi(Y_{n,1}, \epsilon_n s), \phi(Y_{n,1}, \epsilon_n t))$$

where $Y_{n,1} \sim G_n$. Note that $E[\phi(Y_{n,1}, \epsilon_n t)]$ can be simplified as

$$\begin{aligned} U_n(x_0 + t\epsilon_n) - U_n(x_0) &= 2\int_0^\infty \left\{ \sqrt{(u - x_0)_+} - \sqrt{(u - x_0 - t\epsilon_n)_+} \right\} dG_n(u) \\ &= 2\int \phi(u, \epsilon_n t)\, d(G_n - G)(u) + \{U(x_0 + t\epsilon_n) - U(x_0)\}. \end{aligned} \tag{5.16}$$

The first term can be bounded using integration by parts as

$$\begin{aligned} &= 2\left| 0 - \int_0^\infty (G_n - G)(u)\phi'(u, \epsilon_n t) du \right| \leq 2D_n \int_0^\infty |\phi'(u, \epsilon_n t)|\, du \\ &= 2O(\epsilon_n) 2\sqrt{|\epsilon_n t|} = O(\epsilon_n^{3/2}). \end{aligned}$$

The second term in Equation (5.16) can be shown to be of order $O(\epsilon_n)$ by using a one term Taylor expansion. Thus, $E[\phi(Y_{n,1}, \epsilon_n t)] = O(\epsilon_n)$ which shows that the product of the expectations

$$E[\phi(Y_{n,1}, \epsilon_n s)]E[\phi(Y_{n,1}, \epsilon_n t)] = O(\epsilon_n^2).$$

Decomposing $E[\phi(Y_{n,1}, \epsilon_n s)\phi(Y_{n,1}, \epsilon_n t)]$ as

$$\int \phi(u, \epsilon_n s)\phi(u, \epsilon_n t) d(G_n - G)(u) + \int \phi(u, \epsilon_n s)\phi(u, \epsilon_n t) dG(u)$$

we know that

$$\int \phi(u, \epsilon_n s)\phi(u, \epsilon_n t) dG(u) = -\frac{1}{4} g(x_0) st \epsilon_n^2 \log \epsilon_n + O(\epsilon_n^2) \tag{5.17}$$

from the proof of Lemma 3 of Groeneboom and Jongbloed (1995), page 1539. Using integration by parts we can write

$$\begin{aligned} & \int \phi(u, \epsilon_n s)\phi(u, \epsilon_n t) d(G_n - G)(u) \\ = \quad & -\int_0^\infty \left\{\phi'(u, \epsilon_n s)\phi(u, \epsilon_n t) + \phi(u, \epsilon_n s)\phi'(u, \epsilon_n t)\right\}(G_n - G)(u) du. \end{aligned} \tag{5.18}$$

Now,

$$\begin{aligned} & \left| \int_0^\infty \{\phi'(u, \epsilon_n s)\phi(u, \epsilon_n t)(G_n - G)(u)\, du \right| \\ \leq \quad & \|G_n - G\| \int_0^\infty |\phi'(u, \epsilon_n s)|\, |\phi(u, \epsilon_n t)|\, du \\ \leq \quad & D_n \sqrt{|\epsilon_n t|}\ \{2\sqrt{|\epsilon_n s|}\} = O(\epsilon_n^2) \end{aligned} \tag{5.19}$$

using properties (P1) and (P2). Similarly the other term in Equation (5.18) can be shown to be $O(\epsilon_n^2)$, and thus,

$$\begin{aligned} Cov(\mathbb{Z}_{n,1}(s), \mathbb{Z}_{n,1}(t)) &= \frac{4}{n\epsilon_n^4}\left\{O(\epsilon_n^2) - \frac{1}{4} st g(x_0)\epsilon_n^2 \log \epsilon_n + O(\epsilon_n^2) + O(\epsilon_n^2)\right\} \\ &= \frac{1}{2} g(x_0) st \left\{1 - \frac{\log\log n}{\log n}\right\} + O\left(\frac{1}{\log n}\right). \end{aligned}$$

Using the Lindeberg-Feller central limit theorem for triangular arrays it is easy to show that $\mathbb{Z}_{n,1}(1) \Rightarrow N(0, \frac{1}{2}g(x_0))$. An application of Chebyshev's inequality implies that for all fixed $s, t \in K$, $|s\mathbb{Z}_{n,1}(t) - t\mathbb{Z}_{n,1}(s)| = o_P(1)$ as $n \to \infty$. Therefore the finite dimensional distributions of $\mathbb{Z}_{n,1}$ converges weakly to the finite dimensional distributions of $\mathbb{Z}_1$. To verify the stochastic equicontinuity condition we apply the maximal inequality given in Kim and Pollard (1990) (Section 3.1, page 199) to the function class

$$\mathcal{F}_{n,\delta}^M := \left\{\epsilon_n^{-2}\left[\sqrt{(y - x_0 - \epsilon_n t)_+} - \sqrt{(y - x_0 - \epsilon_n s)_+}\right] : |s - t| < \delta, \max(|s|, |t|) \leq M\right\}$$

with the envelope $F_n^M = \epsilon_n^{-2}\left[\sqrt{(y - x_0 + \epsilon_n M)_+} - \sqrt{(y - x_0 - \epsilon_n M)_+}\right]$. Note that $\mathcal{F}_{n,\delta}^M$ is a manageable class of functions and so the maximal inequality can be applied. Appealing to Theorem 2.3 of Kim and Pollard (1990) we obtain the convergence in distribution of $\mathbb{Z}_{n,1}$ to $\mathbb{Z}_1$. Using condition (b), it immediately follows that $\mathbb{Z}_n$ converges in distribution to $\mathbb{Z}$. $\square$

A rigorous proof for the convergence of Δ_n involves a little more than an application of a continuous mapping theorem. The convergence $\mathbb{Z}_n \Rightarrow \mathbb{Z}$ is only in the sense of the metric of uniform convergence on compacta. A concave majorant near the origin might be determined by values of the process long way from the origin; the convergence $\mathbb{Z}_n \Rightarrow \mathbb{Z}$ by itself does not imply the convergence $LCM_{I_n}[\mathbb{Z}_n] \Rightarrow LCM_{\mathbb{R}}[\mathbb{Z}]$. We need to show that $LCM_{I_n}[\mathbb{Z}_n]$ is determined by values of $\mathbb{Z}_n$ for t in an $O_P(1)$ neighborhood of the origin. Corollary 3 shows the convergence of Δ_n, and its proof is given in the Appendix.

Corollary 3. Under conditions (a)-(c), the distribution of Δ_n converges to that of $W \stackrel{d}{=} LCM_{\mathbb{R}}[\mathbb{Z}]'(0)$.

5.2.2 Uniform CLT for estimates of F

We consider three estimates of F, namely $F_n^{\#}$, $\tilde{F}_n^{\#}$ and $\check{F}_n^{\#}$ where

$$\begin{aligned}
F_n^{\#}(x_0) &= 1 + \frac{2}{\pi}\int_{x_0}^{\infty} \sqrt{z}\, dV_n^{\#}(z) \\
\tilde{F}_n^{\#}(x_0) &= 1 + \frac{2}{\pi}\int_{x_0}^{\infty} \sqrt{z}\, d\tilde{V}_n^{\#}(z)
\end{aligned}$$

and $\check{F}_n^{\#}$ is the closest (in the sense of minimizing Equation (5.3) with V_n replaced with $F_n^{\#}$) non-decreasing function to $F_n^{\#}$. We start by deriving the limit distribution of $F_n^{\#}$. Let $\sigma^2 := Var\left[\sin^{-1}\sqrt{1 \wedge \frac{x_0}{Y}}\right]$ where $Y \sim G$.

Proposition 6. If $g(x_0) > 0$ and $\|G_n - G\| \to 0$ as $n \to \infty$, then

$$\sqrt{n}\int_{x_0}^{\infty} \frac{V_n^{\#}(u) - V_n(u)}{2\sqrt{u}}\, dz \Rightarrow N(0, \sigma^2) \tag{5.20}$$

As a consequence,

$$\epsilon_n^{-1}\{F_n^{\#}(x_0) - F_n(x_0)\} \;\Rightarrow\; N\left(0, \frac{4}{\pi^2} x_0 g(x_0)\right). \tag{5.21}$$

Proof of Proposition 6. Using Equation (5.14), we have $F_n^{\#}(x_0) - F_n(x_0)$

$$= -\frac{2}{\pi}\sqrt{x_0}\{V_n^{\#}(x_0) - V_n(x_0)\} - \frac{2}{\pi}\int_{x_0}^{\infty} \frac{V_n^{\#}(u) - V_n(u)}{2\sqrt{u}}\, du$$

$$\begin{aligned}
\text{Notice that, } \int_{x_0}^{\infty} \frac{V_n^{\#}(u)}{2\sqrt{u}}\, du &= \frac{1}{n}\sum_{i=1}^{n}\int_{x_0}^{\infty} \frac{\mathbf{1}\{Y_{n,i} > u\}}{2\sqrt{u}\sqrt{Y_{n,i} - u}}\, du \\
&= \frac{\pi}{2} - \frac{1}{n}\sum_{i=1}^{n} \sin^{-1}\sqrt{1 \wedge \frac{x_0}{Y_{n,i}}}
\end{aligned}$$

after some simplification. Similarly,

$$\begin{aligned}
\int_{x_0}^{\infty} \frac{V_n(u)}{2\sqrt{u}}\, du &= \int_{x_0}^{\infty}\int_{u}^{\infty} \frac{dG_n(y)}{2\sqrt{u}\sqrt{y - u}}\, du \\
&= \frac{\pi}{2} - \int_{0}^{\infty} \sin^{-1}\sqrt{1 \wedge \frac{x_0}{y}}\, dG_n(y).
\end{aligned}$$

Equation (5.20) now follows from the Lindeberg-Feller CLT. From Proposition 5 we know that

$$\epsilon_n^{-1}\{V_n^{\#}(x_0) - V_n(x_0)\} \Rightarrow N(0, g(x_0)).$$

Combining, $\epsilon_n^{-1}\{F_n^{\#}(x_0) - F_n(x_0)\}$

$$= \frac{2\sqrt{x_0}}{\pi}\epsilon_n^{-1}\{V_n^{\#}(x_0) - V_n(x_0)\} + o_P(1) \Rightarrow N\left(0, \frac{4}{\pi^2}x_0 g(x_0)\right).$$

This completes the proof. □

Applying the proposition with $G_n = G$ verifies Equation (5.9). Next we derive the limiting distribution of $\tilde{F}_n^{\#}$. Let $\tilde{F}_n$ be as in Equation (5.8).

Proposition 7. Suppose that (a)-(c) hold with $\hat{V}_n = \tilde{V}_n$, then,

$$\epsilon_n^{-1}\{\tilde{F}_n^{\#}(x_0) - \tilde{F}_n(x_0)\} \quad \Rightarrow \quad N\left(0, \frac{2}{\pi^2}x_0 g(x_0)\right). \tag{5.22}$$

Proof of Proposition 7. We simplify $\tilde{F}_n^{\#}(x_0) - \tilde{F}_n(x_0)$ using integration by parts (see Equation (5.14)) as

$$\frac{2}{\pi}\sqrt{x_0}\left\{\tilde{V}_n(x_0) - \tilde{V}_n^{\#}(x_0)\right\} + \frac{2}{\pi}\int_{x_0}^{\infty}\frac{\tilde{V}_n(u) - \tilde{V}_n^{\#}(u)}{2\sqrt{u}}\,du$$

From Corollary 3, we know that $\epsilon_n^{-1}\left\{\tilde{V}_n(x_0) - \tilde{V}_n^{\#}(x_0)\right\} \Rightarrow N\left(0, \frac{g(x_0)}{2}\right)$. We bound $\left|\int_{x_0}^{\infty}\frac{\tilde{V}_n(u)-\tilde{V}_n^{\#}(u)}{2\sqrt{u}}\,du\right|$ using integration by parts as

$$\begin{aligned}
& 0 + \frac{|\tilde{U}_n(x_0) - \tilde{U}_n^{\#}(x_0)|}{2\sqrt{x_0}} + \frac{1}{4}\left|\int_{x_0}^{\infty}\frac{\tilde{U}_n(u) - \tilde{U}_n^{\#}(u)}{u^{3/2}}\,du\right| \\
\leq\ & \frac{\|\tilde{U}_n - \tilde{U}_n^{\#}\|}{2\sqrt{x_0}} + \|\tilde{U}_n - \tilde{U}_n^{\#}\|\frac{1}{2\sqrt{x_0}} \\
=\ & \frac{\|\tilde{U}_n - \tilde{U}_n^{\#}\|}{\sqrt{x_0}} \leq \frac{\|U_n - U_n^{\#}\|}{\sqrt{x_0}} = O_P(n^{-1/2}) = o_P(\epsilon_n)
\end{aligned}$$

by Marshall's lemma and using maximal inequality 3.1 of Kim and Pollard (1990) to bound $\|U_n - U_n^{\#}\|$. Therefore, $\epsilon_n^{-1}\{\tilde{F}_n^{\#}(x_0) - \tilde{F}_n(x_0)\}$

$$= \frac{2\sqrt{x_0}}{\pi}\epsilon_n^{-1}\{\tilde{V}_n(x_0) - \tilde{V}_n^{\#}(x_0)\} + o_P(1) \Rightarrow N\left(0, \frac{2}{\pi^2}x_0 g(x_0)\right)$$

which completes the proof. $\square$

Let $H_n(x) := \int_0^x F_n(z)\ dz$ and $H_n^{\#}(x) := \int_0^x F_n^{\#}(z)\ dz$. Note that $F_n^{\#}$ is the derivative of $H_n^{\#}$ a.e. Let $\check{H}_n^{\#}$ be the greatest convex minorant (GCM) of $H_n^{\#}$. Then $\check{F}_n^{\#}$ is the right-derivative of $\check{H}_n^{\#}$. We want to study the limit distribution of

$$\Lambda_n := \epsilon_n^{-1}\{\check{F}_n^{\#}(x_0) - \hat{F}_n(x_0)\}$$

where $\hat{F}_n$ can be F_n or $\tilde{F}_n$. Note that Λ_n is the right-hand slope at 0 of the GCM of the process

$$\mathbb{X}_n(t) := \epsilon_n^{-2}\{H_n^{\#}(x_0 + \epsilon_n t) - H_n^{\#}(x_0) - \hat{F}_n(x_0)\epsilon_n t\},$$

for $t \in I_n := [-\epsilon_n^{-1}x_0, \infty)$. As before, we will study the limiting behavior of the process $\mathbb{X}_n$ and use continuous mapping arguments to derive the limiting distribution of Λ_n. We decompose $\mathbb{X}_n$ into $\mathbb{X}_{n,1}$ and $\mathbb{X}_{n,2}$ where

$$\begin{aligned} \mathbb{X}_{n,1}(t) &:= \epsilon_n^{-2}\{(H_n^{\#} - H_n)(x_0 + \epsilon_n t) - (H_n^{\#} - H_n)(x_0)\} \text{ and} \\ \mathbb{X}_{n,2}(t) &:= \epsilon_n^{-2}\{H_n(x_0 + \epsilon_n t) - H_n(x_0) - \hat{F}_n(x_0)\epsilon_n t\} \end{aligned} \tag{5.23}$$

Let GCM_I be the operator that maps a function $h : \mathbb{R} \to \mathbb{R}$ into the GCM of h on the interval $I \subset \mathbb{R}$. Note that $\Lambda_n = GCM_{I_n}[\mathbb{X}_n]'(0)$. We define the processes

$$\mathbb{X}_1(t) = tW \text{ and } \mathbb{X}(t) = \mathbb{X}_1(t) + \frac{1}{2}t^2 f(x_0),$$

for $t \in \mathbb{R}$, where W is a normal random variable having mean 0 and variance $\frac{2}{\pi^2}x_0 g(x_0)$ and f is the density of $Z = X_1^2 + X_2^2 + X_3^2$, i.e., $f = F'$. We state some conditions on the behavior of G_n, $\hat{F}_n$ and H_n used to obtain the limiting distribution of Λ_n.

(a') $D_n := \|G_n - G\| = O(\epsilon_n)$.

(b') $\mathbb{X}_{n,2}(t) \to \frac{1}{2}t^2 f(x_0)$ as $n \to \infty$ uniformly on compacta.

(c') for each $\epsilon > 0$,

$$\left| H_n(x_0+\beta) - H_n(x_0) - \beta \hat{F}_n(x_0) - \frac{1}{2}\beta^2 f(x_0) \right| \leq \epsilon\beta^2 + o(\beta^2) + O(\epsilon_n^2)$$

for large n, uniformly in β varying over a neighborhood of zero.

Theorem 9. Under condition (a') the distribution of $\mathbb{X}_{n,1}$ converges to that of $\mathbb{X}_1$. Further, if (b') holds, then the distribution of $\mathbb{X}_n$ converges to that of $\mathbb{X}$.

Proof of Theorem 9. Using Equation (5.14) and the definition of H_n and $H_n^{\#}$, we get $H_n^{\#}(x) - H_n(x)$

$$\begin{aligned} &= -\frac{2}{\pi}\left[\int_0^x \sqrt{z}\{V_n^{\#}(z) - V_n(z)\}\,dz + \int_0^x \int_z^\infty \frac{(V_n^{\#} - V_n)(u)}{2\sqrt{u}}\,du\,dz\right] \\ &= -\frac{2}{\pi}\left[\int_0^\infty \int_0^{x\wedge y} \sqrt{\frac{z}{y-z}}\,dz\,d(\mathbb{G}_n - G_n)(y) + \int_0^x \int_z^\infty \frac{(V_n^{\#} - V_n)(u)}{2\sqrt{u}}\,du\,dz\right] \end{aligned}$$

We can uniformly bound $\left|\int_z^\infty \frac{(V_n^{\#}-V_n)(u)}{2\sqrt{u}}\,du\right|$ as

$$\begin{aligned} &\left|\int_z^\infty \int_z^y \frac{du}{2\sqrt{y-u}\sqrt{u}}\,d(\mathbb{G}_n - G_n)(y)\right| \\ &= \left|\int_z^\infty \left\{\frac{\pi}{2} - \sin^{-1}\sqrt{\frac{z}{y}}\right\} d(\mathbb{G}_n - G_n)(y)\right| \\ &= 0 + \left|\int_z^\infty \frac{d}{dy}\left[\sin^{-1}\sqrt{\frac{z}{y}}\right](\mathbb{G}_n - G)(y)\,dy\right| \\ &\leq \|\mathbb{G}_n - G\| \left|\left[\sin^{-1}\sqrt{\frac{z}{y}}\right]_z^\infty\right| = \frac{\pi}{2}\|\mathbb{G}_n - G_n\| = o(\epsilon_n) \text{ a.s.} \end{aligned} \tag{5.24}$$

by using the law of iterated logarithms $\|\mathbb{G}_n - G_n\| = o(\epsilon_n)$ a.s. Fix a compact set $K = [-M, M]$. We will show that $\mathbb{X}_{n,1}$ converges weakly to $\mathbb{X}_1$ with the metric of uniform convergence on K. Letting

$$\theta(y,\eta) = \int_0^{(x_0+\eta)\wedge y} \sqrt{\frac{z}{y-z}}\,dz - \int_0^{x_0\wedge y} \sqrt{\frac{z}{y-z}}\,dz, \eta \in \mathbb{R}, \tag{5.25}$$

we decompose $\mathbb{X}_{n,1}$ further into $\mathbb{X}_{n,3}$ and $\mathbb{X}_{n,4}$ where

$$
\begin{aligned}
(5.26) \qquad -\frac{\pi\epsilon_n^2}{2}\mathbb{X}_{n,3}(t) &= \int_{x_0}^{x_0+t\epsilon_n} \sqrt{z}\{V_n^{\#}(z) - V_n(z)\}\, dz \\
&= \int_0^{\infty} \theta(y, t\epsilon_n)\, d(\mathbb{G}_n - G_n)(y), \text{ and} \\
-\frac{\pi\epsilon_n^2}{2}\mathbb{X}_{n,4}(t) &= \int_{x_0}^{x_0+t\epsilon_n} \int_z^{\infty} \frac{(V_n^{\#} - V_n)(u)}{2\sqrt{u}}\, du\, dz.
\end{aligned}
$$

From the uniform bound on $\int_z^{\infty} \frac{(V_n^{\#}-V_n)(u)}{2\sqrt{u}}\, du$ in Equation (5.24) it is easy to see that $\mathbb{X}_{n,4}(t) = o(1)$ a.s. uniformly on K. We proceed to prove the limit distribution of $\mathbb{X}_{n,3}$. Observe that for $\eta \in \mathbb{R}$,

$$
(5.27) \qquad 2\sqrt{x_0 - |\eta|}\, \phi(y, \eta) \le \theta(y, \eta) \le 2\sqrt{x_0 + |\eta|}\, \phi(y, \eta).
$$

$$
\begin{aligned}
\text{Therefore,} \qquad & \epsilon_n^{-2} \int_0^{\infty} \theta(y, t\epsilon_n)\, d(\mathbb{G}_n - G_n)(y) \\
&= \frac{2\sqrt{x_0}}{\epsilon_n^2} \int_0^{\infty} \phi(y, t\epsilon_n)\, d(\mathbb{G}_n - G_n)(y) + R_n(t),
\end{aligned}
$$

where $|R_n(t)|$ is bounded by

$$
\le \{\sqrt{x_0 + M\epsilon_n} - \sqrt{x_0}\} 2\epsilon_n^{-2} \int_0^{\infty} \phi(y, t\epsilon_n)\, d(\mathbb{G}_n - G_n)(y) \xrightarrow{P} 0
$$

uniformly for $t \in K$ as the process $\mathbb{Z}_{n,1}(t) = 2\epsilon_n^{-2} \int_0^{\infty} \phi(y, t\epsilon_n)\, d(\mathbb{G}_n - G_n)(y)$ converges weakly to a tight measure on $C(K)$ by Theorem 8. Therefore,

$$
\mathbb{X}_{n,1}(t) = -\frac{2}{\pi}\left\{\epsilon_n^{-2} \int_0^{\infty} \theta(y, t\epsilon_n)\, d(\mathbb{G}_n - G_n)(y) + o(1)\right\} \Rightarrow \mathbb{X}_1(t)
$$

The other part of the theorem follows immediately. $\square$

Corollary 4. Under conditions (a')-(c'), the distribution of Λ_n converges to that of $W \stackrel{d}{=} GCM_{\mathbb{R}}[\mathbb{X}]'(0)$.

Proof of Corollary 4. The proof is very similar to that Corollary 3 with the LCMs changed to GCMs. We appeal to Lemma 5.1 with $X_{n,c} =$

$\epsilon_n^{-1}\{\check{F}_{n,c}^{\#}(x_0) - \hat{F}_n(x_0)\}, Y_n = \epsilon_n^{-1}\{\check{F}_n^{\#}(x_0) - \hat{F}_n(x_0)\}$ where $\check{F}_{n,c}^{\#}(x_0)$ is the slope at x_0 of the GCM of $H_n^{\#}$ restricted to $[x_0 - c\epsilon_n^{-1}, x_0 + c\epsilon_n^{-1}]$. Let us denote by $\mathbb{C}_{n,c}$ the GCM of $\mathbb{X}_n$ restricted to $[-c, c]$. Also, we take W_c as the right-hand slope at 0 of $\mathbb{C}_c$, the GCM of $\mathbb{X}$ restricted to $[-c, c]$, and Y as the slope at 0 of $\mathbb{C}$, the GCM of $\mathbb{X}$. Note that as $\mathbb{X}$ is itself convex, $\mathbb{C}_c = \mathbb{C} = \mathbb{X}$, for all $c > 0$ and thus $W_c = \mathbb{C}_c' = Y$.

We have to show that a result similar to Claim 3 of Corollary 3 holds in our case, i.e., for every $\epsilon > 0$, we have

$$\left|(H_n^{\#} - H_n)(x_0 + \beta) - (H_n^{\#} - H_n)(x_0)\right| \leq \epsilon\beta^2 + O_P(\epsilon_n^2) \tag{5.28}$$

uniformly over a neighborhood of zero. By Equation (5.26) and the following discussion,

$$\begin{aligned} & |(H_n^{\#} - H_n)(x_0 + t\epsilon_n) - (H_n^{\#} - H_n)(x_0)| \\ \leq \; & \frac{2}{\pi}\left\{\left|\int_0^\infty \theta(y, \beta)\, d(\mathbb{G}_n - G_n)(y)\right| + |\beta| o(\epsilon_n)\right\}. \end{aligned} \tag{5.29}$$

For $\epsilon > 0$, from Equation (5.27) we see that

$$\begin{aligned} \left|\int_0^\infty \theta(y, \beta)\, d(\mathbb{G}_n - G_n)(y)\right| & \leq 2(x_0 + |\beta|)\left|\int_0^\infty \phi(y, \beta)\, d(\mathbb{G}_n - G_n)(y)\right| \\ & \leq (x_0 + |\beta|)\{\epsilon\beta^2 + O_P(\epsilon_n)\} \end{aligned} \tag{5.30}$$

by Equation (5.35) (proved in the appendix). Noting that $|\beta| o(\epsilon_n) \leq \epsilon\beta^2 + o(\epsilon_n^2)$ we can show that Equation (5.28) holds for suitably chosen ϵ.

5.3 Consistency of Bootstrap methods

5.3.1 Bootstrapping $\tilde{V}_n$

The results of Chapter 4 casts serious doubt of the use of bootstrap methods in isotonic problems. Given data $Y_1, Y_2, \ldots, Y_n \sim G$ let $\mathbb{G}_n$ denote its e.d.f.

Suppose that we draw conditionally independent and identically distributed random variables $Y_{n,1}^*, Y_{n,2}^*, \ldots, Y_{n,n}^*$ having distribution function $\mathbb{G}_n$. Let $\mathbb{G}_n^*$ be the e.d.f. of the bootstrap sample. Letting

$$\begin{aligned} V_n^*(y) &:= \frac{1}{n} \sum_{i:Y_{n,i}^*>y} \frac{1}{\sqrt{Y_{n,i}^* - y}} = \int \frac{\mathbf{1}_{[y,\infty)}}{\sqrt{u-y}}\, d\mathbb{G}_n^*(u) \text{ and} \\ U_n^*(x) &:= \frac{2}{n} \sum_{i=1}^{n} \left\{ \sqrt{Y_{n,i}^*} - \sqrt{(Y_{n,i}^* - x)_+} \right\} \\ &= 2 \int \{ \sqrt{u} - \sqrt{(u-x)_+} \}\, d\mathbb{G}_n^*(u), \end{aligned} \tag{5.31}$$

the isotonic estimate of V based on the bootstrap sample is $\tilde{V}_n^* := \tilde{U}_n^{*\prime} = LCM_{[0,\infty)}[U_n^*]'$.

The bootstrap estimate of $\Delta_n = \epsilon_n^{-1}\{\tilde{V}_n(x_0) - V(x_0)\}$ is

$$\Delta_n^* := \epsilon_n^{-1}\{\tilde{V}_n^*(x_0) - \tilde{V}_n(x_0)\}.$$

To find the limit distribution of Δ_n^* we define the process

$$\mathbb{Z}_n^*(t) = \epsilon_n^{-2}\{U_n^*(x_0 + \epsilon_n t) - U_n^*(x_0) - \tilde{V}_n(x_0)\epsilon_n t\},\ t \in I_n := [-\epsilon_n^{-1} x_0, \infty).$$

We decompose $\mathbb{Z}_n^*$ into $\mathbb{Z}_{n,1}^*$ and $\mathbb{Z}_{n,2}^*$ where

$$\begin{aligned} \mathbb{Z}_{n,1}^*(t) &= \epsilon_n^{-2}\{(U_n^* - U_n)(x_0 + \epsilon_n t) - (U_n^* - U_n)(x_0)\} \\ \mathbb{Z}_{n,2}^*(t) &= \epsilon_n^{-2}\{U_n(x_0 + \epsilon_n t) - U_n(x_0) - \tilde{V}_n(x_0)\epsilon_n t\} \end{aligned} \tag{5.32}$$

Recall that $\mathbb{Z}_1(t) = tW$ and $\mathbb{Z}(t) = \mathbb{Z}_1(t) + \frac{1}{2}t^2 V'(x_0)$ are two processes defined for $t \in \mathbb{R}$, where W is a normal random variable having mean 0 and variance $\frac{1}{2}g(x_0)$. Let $\mathbf{Y} = (Y_1, Y_2, \ldots)$. The following theorem shows that bootstrapping from the e.d.f. $\mathbb{G}_n$ is weakly consistent.

Theorem 10. Suppose that V is continuously differentiable around x_0, and $g(x_0) \neq 0$. Then we have the following results:

(i) The conditional distribution of the process $\mathbb{Z}^*_{n,1}$, given $\mathbf{Y}$, converges to that of $\mathbb{Z}_1$ a.s.

(ii) Unconditionally, $\mathbb{Z}^*_{n,2}(t)$ converges in probability to $\frac{1}{2}t^2V'(x_0)$, uniformly on compacta.

(iii) The conditional distribution of the process $\mathbb{Z}^*_n$, given $\mathbf{Y}$, converges to that of $\mathbb{Z}$, in probability.

(iv) The bootstrap procedure is weakly consistent, i.e., the conditional distribution of Δ^*_n, given $\mathbf{Y}$, converges to that of W, in probability.

Proof of Theorem 10. To find the conditional distribution of $\mathbb{Z}^*_{n,1}$ given $\mathbf{Y}$ we appeal to Theorem 8 with $G_n = \mathbb{G}_n$, $\mathbb{G}_n = \mathbb{G}^*_n$ and $P\{\cdot\} = P^*\{\cdot\} = P\{\cdot|\mathbb{G}_n\}$. Note that condition (a) required for Theorem 8 holds a.s.

Let us define the process

$$\mathbb{Z}^0_n(t) = \epsilon_n^{-2}\{U_n(x_0 + t\epsilon_n) - U_n(x_0) - \epsilon_n tV(x_0)\},\ t \in I_n.$$

Using Theorem 8 with $G_n = G$, $V_n = V$ and $U_n = U$ for all n, we can show that unconditionally $\mathbb{Z}^0_n$ converges in distribution to $\mathbb{Z}$. To prove (ii) note that

$$\mathbb{Z}^*_{n,2}(t) = \mathbb{Z}^0_n(t) - t \cdot LCM_{I_n}[\mathbb{Z}^0_n]'(0).$$

Unconditionally, using the continuous mapping theorem along with a localization argument as in Corollary 3, we obtain $\mathbb{Z}^*_{n,2}(t) \Rightarrow \mathbb{Z}(t) - t \cdot LCM_{\mathbb{R}}[\mathbb{Z}]'(0) = \frac{1}{2}t^2V'(x_0)$. As the limiting process is a constant, $\mathbb{Z}^*_{n,2}(t) \xrightarrow{P} \frac{1}{2}t^2V'(x_0)$.

Let $\{n_k\}$ be a subsequence of $\mathbb{N}$. We will show that there exists a further subsequence such that conditional on $\mathbf{Y}$, $\mathbb{Z}_n \Rightarrow \mathbb{Z}$ a.s. along the subsequence. Now, given $\{n_k\}$, there exists a further subsequence $\{n_{k_l}\}$ such that $\mathbb{Z}^*_{n_{k_l},2}(t) \rightarrow \frac{1}{2}t^2V'(x_0)$ uniformly on compacta a.s. Thus the conditional distribution of $\mathbb{Z}^*_{n_{k_l}}$ given $\mathbf{Y}$, converges to that of $\mathbb{Z}$, for a.e. $\mathbf{Y}$. This completes the proof of (iii).

To prove (iv) we use Corollary 3. Although conditions (a) and (b) hold in probability, condition (c) holds with $\hat{V}_n = \tilde{V}_n$ and the $O(\epsilon_n^2)$ term replaced by $O_P(\epsilon_n^2)$. Thus we cannot appeal directly to Corollary 3. Let $\xi > 0$ and $\eta > 0$ be given. We will show that there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $P\{L(H, H_n^*) > \xi\} < \eta$, where L is the Levy metric (see Chapter 4.3), H is the distribution function of $W \sim N(0, \frac{1}{2}g(x_0))$ and H_n^* is the distribution function of Δ_n^*, conditional on the data. For $\epsilon > 0$, sufficiently small, let us define the set

$$A_n := \left\{ \left| U_n(x_0 + \beta) - U_n(x_0) - \beta \tilde{V}_n(x_0) - \frac{1}{2}\beta^2 V'(x_0) \right| < C\epsilon_n^2 + \epsilon\beta^2 \right\}$$

where $C > 0$ is chosen such that $P\{A_n^c\} < \frac{\eta}{2}$. This can be done since (c) holds with $O(\epsilon_n^2)$ term replaced by $O_P(\epsilon_n^2)$. Further, let H_n^0 be the distribution function of Δ_n^* under the probability measure P_n^0, where

$$P_n^0\{E\}(\omega) = \begin{cases} P\{E|\mathbb{G}_n\}(\omega) = P^*\{E\} & \text{if } \omega \in A_n, \\ P\{E|G\}(\omega) & \text{if } \omega \in A_n^c. \end{cases} \tag{5.33}$$

Note that under P_n^0, $L(H, H_n^0) \xrightarrow{P} 0$, as Corollary 3 can be applied. Therefore, for all sufficiently large n, $P\{L(H, H_n^*) > \xi\}$

$$\begin{aligned} &\leq P\left\{L(H, H_n^0) > \frac{\xi}{2}\right\} + P\left\{L(H_n^0, H_n^*) > \frac{\xi}{2}\right\} \\ &\leq \frac{\eta}{2} + P\left\{L(H_n^0, H_n^*) > \frac{\xi}{2}, A_n\right\} + P\left\{L(H_n^0, H_n^*) > \frac{\xi}{2}, A_n^c\right\} \\ &\leq \frac{\eta}{2} + 0 + \frac{\eta}{2} \end{aligned}$$

as when A_n occurs, $L(H_n^0, H_n^*) = 0$. This completes the proof of (iv). $\square$

Remark 1. Let $J_n(t) = \int_t^\infty \sqrt{z - t}\, dV_n(z)$ for $t \geq 0$. We could have generated the bootstrap samples from $\hat{G}_n$ where

$$G_n^\#(t) = 1 + \frac{2}{\pi} J_n(t).$$

It is interesting to note that a simplification yields $G_n^{\#} = \mathbb{G}_n$, the e.d.f. of the data. So, drawing bootstrap samples from the e.d.f. is equivalent to generating samples from $G_n^{\#}$, a model based estimate of G.

5.3.2 Bootstrapping F_n, $\tilde{F}_n$ and $\check{F}_n$

The three estimators of F under study based on the bootstrap sample are F_n^*, $\tilde{F}_n^*$ and $\check{F}_n^*$ defined analogously as in Section 5.2.2; e.g., $F_n^*(x) = 1 + \frac{2}{\pi}\int_x^\infty \sqrt{z}\, dV_n^*(z)$. We approximate the sampling distribution of $\epsilon_n^{-1}\{F_n(x_0) - F(x_0)\}$ by the bootstrap distribution of $\epsilon_n^{-1}\{F_n^*(x_0) - F_n(x_0)\}$. The bootstrap samples are generated from $\mathbb{G}_n$, the e.d.f. of the Y_i's. By appealing to Proposition 6 with $G_n = \mathbb{G}_n$, it is easy to see that the bootstrap method is consistent.

The sampling distribution of $\epsilon_n^{-1}\{\tilde{F}_n(x_0) - F(x_0)\}$ is approximated by that of $\epsilon_n^{-1}\{\tilde{F}_n^*(x_0) - \tilde{F}_n(x_0)\}$. Using Proposition 7, we can establish the consistency of the method. Note that the proof of Theorem 10 shows how conditions (a)-(c) are satisfied with $G_n = \mathbb{G}_n, \hat{V}_n = \tilde{V}_n$ required to apply Proposition 7.

Recall that $\check{F}_n^*$ is the non-increasing function closest to F_n^*. Let $H_n(x) := \int_0^x F_n(z)\, dz$ and $H_n^*(x) := \int_0^x F_n^*(z)\, dz$. Next we show that approximating the distribution of $\Lambda_n = \epsilon_n^{-1}\{\check{F}_n(x_0) - F(x_0)\}$ by the bootstrap distribution of $\Lambda_n^* := \epsilon_n^{-1}\{\check{F}_n^*(x_0) - \check{F}_n(x_0)\}$ is consistent. To find the limit distribution of Λ_n^* we define the process

$$\mathbb{X}_n^*(t) = \epsilon_n^{-2}\{H_n^*(x_0 + \epsilon_n t) - H_n^*(x_0) - \check{F}_n(x_0)\epsilon_n t\},\ t \in I_n := [-\epsilon_n^{-1} x_0, \infty)$$

and decompose it into $\mathbb{X}_{n,1}^*$ and $\mathbb{X}_{n,2}^*$, where

$$\begin{aligned} \mathbb{X}_{n,1}^*(t) &= \epsilon_n^{-2}\{(H_n^* - H_n)(x_0 + \epsilon_n t) - (H_n^* - H_n)(x_0)\} \\ \mathbb{X}_{n,2}^*(t) &= \epsilon_n^{-2}\{H_n(x_0 + \epsilon_n t) - H_n(x_0) - \check{F}_n(x_0)\epsilon_n t\} \end{aligned} \tag{5.34}$$

Recall that $\mathbb{X}_1(t) = tW$ and $\mathbb{X}(t) = \mathbb{X}_1(t) + \frac{1}{2}t^2 f(x_0)$ are two processes defined for $t \in \mathbb{R}$, where W is a normal random variable having mean 0 and variance $\frac{2}{\pi^2} x_0 g(x_0)$.

Theorem 11. Suppose that F is continuously differentiable around x_0, and $g(x_0) \neq 0$. Then we have the following results:

(i) The conditional distribution of the process $\mathbb{X}^*_{n,1}$, given $\mathbf{Y}$, converges to that of $\mathbb{X}_1$ a.s.

(ii) Unconditionally, $\mathbb{X}^*_{n,2}(t)$ converges in probability to $\frac{1}{2}t^2 f(x_0)$, uniformly on compacta.

(iii) The conditional distribution of the process $\mathbb{X}^*_n$, given $\mathbf{Y}$, converges to that of $\mathbb{X}$, in probability.

(iv) The bootstrap procedure is weakly consistent, i.e., the conditional distribution of Λ^*_n, given $\mathbf{Y}$, converges to that of W, in probability.

Proof of Theorem 11. The proof is very similar to that of Theorem 10. To find the conditional distribution of $\mathbb{X}^*_{n,1}$ given $\mathbf{Y}$ we appeal to Theorem 9 with $G_n = \mathbb{G}_n$, $\mathbb{G}_n = \mathbb{G}^*_n$ and $P\{\cdot\} = P^*\{\cdot\} = P\{\cdot|\mathbb{G}_n\}$. Note that condition (a') required for Theorem 9 holds a.s.

We express $\mathbb{X}^*_{n,2}(t)$ as $\mathbb{X}^0_n(t) - t \cdot GCM_{I_n}[\mathbb{X}^0_n]'(0)$ where

$$\mathbb{X}^0_n(t) = \epsilon_n^{-2}\{H_n(x_0 + t\epsilon_n) - H_n(x_0) - F(x_0)\epsilon_n t\}.$$

Note that unconditionally $\mathbb{X}^0_n$ converges in distribution to $\mathbb{X}$ by an application of Theorem 9 with $G_n = G$, $\hat{F}_n = F$ and $H_n = H$ for all n.

Unconditionally, using the continuous mapping theorem along with a localization argument as in Corollary 3, we obtain $\mathbb{X}^*_{n,2}(t) \Rightarrow \mathbb{X}(t) - t \cdot GCM_{\mathbb{R}}[\mathbb{X}]'(0) = \frac{1}{2}t^2 f(x_0)$. As the limiting process is a constant, $\mathbb{X}^*_{n,2}(t) \xrightarrow{P} \frac{1}{2}t^2 f(x_0)$.

An argument using subsequences as in the proof of (iii) of Theorem 10 shows that the conditional distribution of the process $\mathbb{X}_n^*$, given $\mathbf{Y}$, converges to that of $\mathbb{X}$, in probability. The last part of the theorem follows along similar lines as in the proof of (iv) of Theorem 10.

5.4 Appendix

We will use the following lemma which can be proved easily (see Pollard (1984), page 70).

Lemma 5.1. *If $\{X_{n,c}\}, \{Y_n\}, \{W_c\}, Y$ are sets of random elements taking values in a metric space $(\mathcal{X}, d)$, $n = 0, 1, \ldots,$ and $c \in \mathbb{R}$ such that for any $\delta > 0$,*

(i) $\lim_{c\to\infty} \limsup_{n\to\infty} P\{d(X_{n,c}, Y_n) > \delta\} = 0$*,*

(ii) $\lim_{c\to\infty} P\{d(W_c, Y) > \delta\} = 0$*,*

(iii) $X_{n,c} \Rightarrow W_c$ *as* $n \to \infty$ *for every* $c \in \mathbb{R}$*.*

Then $Y_n \Rightarrow Y$ *as* $n \to \infty$*.*

Proof of Corollary 3. For the proof of this corollary, we appeal to Lemma 5.1 with $X_{n,c} = \epsilon_n^{-1}\{\tilde{V}_{n,c}^{\#}(x_0) - \hat{V}_n(x_0)\}$ and $Y_n = \epsilon_n^{-1}\{\tilde{V}_n^{\#}(x_0) - \hat{V}_n(x_0)\}$ where $\tilde{V}_{n,c}^{\#}(x_0)$ is the slope at x_0 of the LCM of $U_n^{\#}$ restricted to $[x_0 - c\epsilon_n^{-1}, x_0 + c\epsilon_n^{-1}]$. Let us denote by $\mathbb{C}_{n,c}$ the LCM of $\mathbb{Z}_n$ restricted to $[-c, c]$. Also, we take W_c as the right-hand slope at 0 of $\mathbb{C}_c$, the LCM of $\mathbb{Z}$ restricted to $[-c, c]$, and Y as the slope at 0 of $\mathbb{C}$, the LCM of $\mathbb{Z}$. Note that as $\mathbb{Z}$ is itself concave, $\mathbb{C}_c = \mathbb{C} = \mathbb{Z}$, for all $c > 0$ and thus $W_c = \mathbb{C}_c' = Y$.

Note that as $X_{n,c} = \mathbb{C}_{n,c}'(0)$, an application of the usual continuous mapping theorem (see lemma on page 330 of Robertson, Wright and Dykstra (1988)) and the uniform convergence of $\mathbb{Z}_n$ to $\mathbb{Z}$ on $[-c, c]$ yields $X_{n,c} \Rightarrow W_c = \mathbb{C}_c'(0)$, for every c. This shows that condition (iii) of the lemma holds. Condition (ii) of lemma holds trivially as $\mathbb{Z}$ is itself concave. We will only need to show that condition (i) holds to apply the lemma and conclude that Y_n converges to Y, thereby completing the proof of the theorem. The following series of claims are adopted from the assertion in page 217 of Kim and Pollard (1990).

Claim 1: Condition (i) of the lemma follows if we can show the existence of random variables $\{\tau_n\}$ and $\{\sigma_n\}$ of order $O_P(1)$ such that $\tau_n \geq 0$, $\sigma_n > 0$ and $\mathbb{C}_n(\tau_n) = \mathbb{Z}_n(\tau_n)$ and $\mathbb{C}_n(\sigma_n) = \mathbb{Z}_n(\sigma_n)$.

Proof of Claim 1: Let $\epsilon > 0$ be given. As $\{\tau_n\}$ and $\{\sigma_n\}$ are of order $O_P(1)$, we can get $M_\epsilon > 0$ such that $\limsup_{n\to\infty} P\{A_\epsilon\} < \epsilon$, where $A_\epsilon = \{\tau_n < -M_\epsilon, \sigma_n > M_\epsilon\}$. Take $\omega \in A_\epsilon^c$. Then $-M_\epsilon \leq \tau_n(\omega) \leq 0$ and $0 < \sigma_n(\omega) \leq M_\epsilon$. Note that

$$\mathbb{Z}_n(\tau_n(\omega)) \leq \mathbb{C}_{n,c}(\tau_n(\omega)) \leq \mathbb{C}_n(\tau_n(\omega)) \text{ and } \mathbb{Z}_n(\sigma_n(\omega)) \leq \mathbb{C}_{n,c}(\sigma_n(\omega)) \leq \mathbb{C}_n(\sigma_n(\omega))$$

for $c > M_\epsilon$. From the hypothesis and using properties of concave majorants it follows that $\mathbb{C}_{n,c}(h)(\omega) = \mathbb{C}_n(h)(\omega)$ for all $h \in [\tau_n, \sigma_n]$. Thus, $X_{n,c}(\omega) = Y_n(\omega)$. Therefore, $A_\epsilon^c \subset \{X_{n,c} = Y_n\}$ which implies $\limsup_{n\to\infty} P\{(X_{n,c}, Y_n) > \delta\} \leq \limsup_{n\to\infty} P\{A_\epsilon\} < \epsilon$, for $c > M_\epsilon$. $\square$

Therefore it suffices to show that we can construct random variables τ_n and σ_n of order $O_P(1)$ so that $\mathbb{C}_n(\tau_n) = \mathbb{Z}_n(\tau_n)$ and $\mathbb{C}_n(\sigma_n) = \mathbb{Z}_n(\sigma_n)$ for $\tau_n \leq 0 < \sigma_n$.

Claim 2: There exist random variables $\{\tau_n\}$ and $\{\sigma_n\}$ of order $O_P(1)$ such that $\tau_n \leq 0$, $\sigma_n > 0$ and $\mathbb{C}_n(\tau_n) = \mathbb{Z}_n(\tau_n)$ and $\mathbb{C}_n(\sigma_n) = \mathbb{Z}_n(\sigma_n)$.

Proof of Claim 2: Let K_n denote the LCM of $U_n^{\#}$. The line through $(x_0, K_n(x_0))$ with slope $\tilde{V}_n(x_0)$ must lie above $U_n^{\#}$ touching it at the two points $x_0 - L_n$ and $x_0 + R_n$, where $L_n \leq 0$ and $R_n > 0$. Note that $x_0 - L_n$ and $x_0 + R_n$

are the nearest points to x_0 such that K_n and $U_n^\#$ coincide. The line segment from $(x_0 - L_n, U_n^\#(x_0 - L_n))$ to $(x_0 + R_n, U_n^\#(x_0 + R_n))$ makes up part of K_n. It will suffice to show that $L_n = O_P(\epsilon_n)$, as then $\tau_n := -\epsilon_n^{-1} L_n = O_P(1)$. The argument depends on the inequality

$$K_n(x_0) + \tilde{V}_n(x_0)\beta \geq U_n^\#(x_0 + \beta) \text{ for all } \beta,$$

with equality at $\beta = -L_n$ and $\beta = R_n$.

Let $\Gamma_n(\beta) = U_n^\#(x_0 + \beta) - U_n^\#(x_0) - \beta \tilde{V}_n^\#(x_0)$. Γ_n is the distance between $U_n^\#(x_0+\beta)$ and $U_n^\#(x_0)+\beta\tilde{V}_n^\#(x_0)$. It follows that $\Gamma_n(\beta)$ achieves its maximum at $\beta = -L_n$ and $\beta = R_n$ and $\Gamma_n(-L_n) = \Gamma_n(R_n)$. We can easily show using condition (a) that L_n, R_n and $\gamma_n := \tilde{V}_n^\#(x_0) - \hat{V}_n(x_0)$ are of order $o_P(1)$. That lets us argue locally. Let

$$g_n(y, \beta) := \sqrt{(y - x_0)_+} - \sqrt{(y - x_0 - \beta)_+} - \beta \hat{V}_n(x_0).$$

Claim 3**:** Then for any $\epsilon > 0$, we have

$$\frac{1}{n} \left| \sum_{i=1}^{n} \{ g_n(X_{n,i}, \beta) - E g_n(X_{n,i}, \beta) \} \right| \leq \epsilon \beta^2 + O_P(\epsilon_n^2) \tag{5.35}$$

uniformly over a neighborhood of zero.

For the time being, we assume the claim, which is proved later in the Appendix. From condition (d), we get $\left| E g_n(\cdot, \beta) - \frac{1}{2}\beta^2 V'(x_0) \right| \leq \epsilon\beta^2 + o(\beta^2) + O(\epsilon_n^2)$ for sufficiently large n. Thus

$$\begin{aligned} & |\Gamma_n(\beta) + \beta\gamma_n - \frac{1}{2}\beta^2 V'(x_0)| \\ = \; & |U_n^\#(x_0 + \beta) - U_n^\#(x_0) - \beta \hat{V}_n(x_0) - \frac{1}{2}\beta^2 V'(x_0)| \\ \leq \; & \epsilon\beta^2 + o(\beta^2) + O_P(\epsilon_n^2) \qquad (5.36) \end{aligned}$$

uniformly for β over a neighborhood of 0. As $V'(x_0) < 0$, for $n \to \infty$, there exist constant $c_1, c_2 > 0$ such that with probability tending to 1 for β in a small neighborhood of 0,

$$-\frac{1}{2}c_2\beta^2 - \beta\gamma_n - O_P(\epsilon_n^2) \leq \Gamma_n(\beta) \leq -\frac{1}{2}c_1\beta^2 - \beta\gamma_n + O_P(\epsilon_n^2).$$

The quadratic $-\frac{1}{2}c_1\beta^2 - \beta\gamma_n$ has its maximum of $\frac{1}{2}\gamma_n^2/c_1$ at $-\gamma_n/c_1$, and takes negative values for those β with the same sign of γ_n. It follows that with probability tending to 1,

$$\max_\beta \Gamma_n(\beta) = \min(\Gamma_n(-L_n), \Gamma_n(R_n)) \leq O_P(\epsilon_n).$$

We also have

$$\max_\beta \Gamma_n(\beta) \geq \Gamma_n(-\gamma_n/c_2) \geq \frac{1}{2}\gamma_n^2/c_2 - O_P(\epsilon_n).$$

These two bounds imply that $\gamma_n = O_P(\epsilon_n)$. With this rate for convergence for $\{\gamma_n\}$ we can now deduce from the inequalities

$$0 = \Gamma_n(0) \leq \Gamma(-L_n) \leq \frac{1}{2}c_1(L_n - \gamma_n/c_1)^2 + \frac{1}{2}\gamma_n^2/c_1 + O_P(\epsilon_n^2)$$

that $L_n = O_P(\epsilon_n)$, as required. Similarly, we can show that $R_n = O_P(\epsilon_n)$. $\square$

Proof of Claim 3: Note that $\frac{1}{n}\sum_{i=1}^n \{g_n(X_{n,i},\beta) - Eg_n(\cdot,\beta)\} = U_n^{\#}(x_0 + \beta) - U_n^{\#}(x_0) - U_n(x_0+\beta) + U_n(x_0) =: \mathbb{H}_n(\beta)$. Let $\eta > 0$ be given. We will show that $|\mathbb{H}_n(\beta)| \leq \eta\beta^2 + \epsilon_n^2 M_n^2$ uniformly over a neighborhood of 0, for M_n of order $O_P(1)$. We introduce the function class $\mathcal{H} := \{h_\beta : [0,\infty) \to \mathbb{R} | h_\beta(y) = 2\sqrt{(y-x_0)_+} - 2\sqrt{(y-x_0-\beta)_+}, \beta \in \mathbb{R}\}$ and for $R > 0$, its subclass $\mathcal{H}_R = \{h_\beta \in \mathcal{H} : |\beta| \leq R\}$. It can be shown that there is a positive integer R_0 such that the envelope $H_R(y) = 2\{\sqrt{(y-x_0-R)_+} - \sqrt{(y-x_0)_+}\}$ of $\mathcal{H}_R$ satisfies

$$\int H_R^2(y)g(y)dy \leq -2g(x_0)R^2 \log R,$$

for all $R \leq R_0$ (see Lemma 3 and 4 of Groeneboom and Jongbloed (1995)). Now define $M_n(\omega)$ as the infimum (possibly $+\infty$) of those values for which

$$|\mathbb{H}_n(\beta)| = \left| \int h_\beta(y) d(\mathbb{G}_n - G)(y) \right| \leq \eta\beta^2 + \epsilon_n^2 M_n^2$$

holds for all $|\beta| \leq R_0$. Let us also define $A(n,j)$ to be the set of those β in $\mathbb{R}$ for which $(j-1)\epsilon_n \leq |\beta| < j\epsilon_n$. Then for ν constant,

$$\begin{aligned} P\{M_n > \nu\} &\leq P\{\exists \beta : |\mathbb{H}_n(\beta)| > \eta\beta^2 + \epsilon_n^2 \nu^2\} \\ &\leq \sum_{j=1}^{R_0/\epsilon_n} P\{\exists \beta \in A(n,j) : \epsilon_n^{-2} |\mathbb{H}_n(\beta)| > \eta(j-1)^2 + \nu^2\} \\ &\leq \sum_{j=1}^{R_0/\epsilon_n} \frac{E\left(\sup_{|\beta|<j\epsilon_n} \epsilon_n^{-4} |\mathbb{H}_n(\beta)|^2\right)}{\{\eta(j-1)^2 + \nu^2\}^2} \\ &\leq \sum_{j=1}^{\infty} \frac{Cj^2}{\{\eta(j-1)^2 + \nu^2\}^2} \qquad (5.37) \end{aligned}$$

for n sufficiently large. The last inequality follows from a maximal inequality as in part (ii) of Result 3.1 of Kim and Pollard (1990). To be more precise, fix $j \geq 1$ and consider the class $\mathcal{H}_{j\epsilon_n}$ with envelope function $H_{j\epsilon_n}$. From the maximal inequality in 3.1 of Kim and Pollard (1990), we have

$$\epsilon_n^{-4} E(\sup_{\mathcal{H}_{j\epsilon_n}} |\mathbb{H}_n(\beta)|^2) \leq J^2(1) \frac{\epsilon_n^{-4}}{n} E H_{j\epsilon_n}^2 \leq J^2(1) C j^2$$

where J is a continuous and increasing function with $J(0) = 0$ and $J(1) < \infty$, not depending on n and C is a constant. We can therefore ensure that the sum is suitably small by choosing ν large enough. This proves the claim. $\square$

CHAPTER 6

Streaming motion in Leo I galaxy

Whether a dwarf spheroidal galaxy is in equilibrium or being tidally disrupted by the Milky Way is an important question for the study of its dark matter content and distribution. This question is investigated using 328 recent observations from the dwarf spheroidal Leo I [published in Mario, Olszewski, and Walker (2008)]. For Leo I, tidal disruption is detected, at least for stars sufficiently far from the center, but the effect appears to be quite modest. Statistical tools include isotonic and split point estimators, asymptotic theory, and resampling methods.

6.1 Introduction

The dwarf spheroidal galaxies near the Milky Way are among the least luminous galaxies in the night sky. While they have stellar populations similar to those of globular clusters, approximately $10^6 - 10^7$ stars, they are considerably larger systems, typically hundreds, even thousands of parsecs in size compared to radii of tens of parsecs characteristic of clusters. They are excellent candidates for the study of dark matter because they are nearby and they generally have extremely low stellar densities. Moreover, due to their proximity to larger galaxies such as the Milky Way, many of these dwarf galaxies are also potentially strongly affected by disruptive tidal effects. The mere existence of dwarf spheroidal galaxies suggests these systems contain much dark matter since it is unclear how they could have avoided the effects of tidal disruption without the added gravitational force from a considerable reservoir of unseen matter [see Muñoz, Majewski, and Johnston (2008) and Mario, Olszewski, and Walker (2008)].

Detailed kinematic studies, Wang et al. (2005), Walker et al. (2006), and Wang et al. (2008a), confirm the widespread belief that the dwarf spheroidal galaxies are dominated by dark matter, in many cases finding that the dark matter densities exceed that of visible matter by a few orders of magnitude. These latest studies differed from their predecessors, for example Mario et al. (1993), Mario (1998a), Mario et al. (1998b) (and references therein), and Kleyna (2003), statistically by using a non-parametric analysis to estimate the distribution of dark matter. While the non-parametric analysis did not require a specific form for this distribution, it did assume that the galaxies are in equilibrium and isotropic. The purpose of the present chapter is to explore the gravitational effects of the Milky Way on the dwarf spheroidals and, implicitly, to probe the underlying assumptions used in Wang et al. (2005), Walker et al. (2006), and Wang et al. (2008a).

To this end, our approach is to address these issues using recent data, Mateo, Olszewski and Walker (2008), for the dwarf spheroidal galaxy Leo I from which we obtain kinematic observations of 328 stars. Among the Milky Way's dSph satellites, Leo I is perhaps the most distant, at 255 ± 3 kpc, and is receding from the Milky Way at a relatively large velocity of 179.5 ± 0.5 km/s. The combination of the large distance and high velocity lead Byrd et. al. (1994) to suggest that Leo I may not be bound to the Milky Way. Bound or unbound, the large outward velocity means that Leo I passed much closer to the Galactic Center in the past. In the preferred model of Byrd et. al. (1994), Leo I passed within 70 kpc of the Galactic Center, a distance similar to the closest present day dwarf spheroidal galaxies. Recent papers [Sohn et al. (2007), Mateo, Olszewski, and Walker (2008)] suggest more specific models in which Leo I passed within 10-20 kpc from the center of the Milky Way some 1-2 Gyr ago.

So, the question becomes: What effect, if any, did this close encounter have on Leo I? In some cases, a close encounter with the Galactic Center can change the shape of a dwarf spheroidal by producing tidal arms, [Oh, Lin and Aarseth (1995), Piatek and Pryor, C. (1995)]. Prominent tidal arms are not observed in Leo I, but this may reflect our unfavorable viewing angle rather than the actual lack of such features [Mateo, Olszewski and Walker (2008)]. A more subtle but related effect is streaming motion. The practical observational signal of this process arises from the fact that both leading and trailing stars move away from the center of the main body of the perturbed systems in the reference frame of that galaxy. The magnitude of the streaming motion is likely to increase beyond a threshold radius in the perturbed galaxy and be aligned with the apparent major axis of the system. A more detailed account of streaming motion may be found in Section 4.3 of Mateo, Olszewski and Walker (2008).

There are several interesting statistical questions here. Is streaming motion evident in Leo I? If so, how can it be described and estimated? To what extent can it be described by a threshold model, in which streaming motion is only present for stars at a sufficient distance from the center? We answer these questions within the context of a model, called the cosine model below, that incorporates the qualitative features of streaming motion described above (increasing with distance from the center and largest along the major axis). The answers may be summarized: The magnitude of streaming motion appears to be modest, at most 6.19 km/s, but is (nearly) significant at the 5% level. The streaming motion does appear to be consistent with a threshold model, but it is difficult to constrain the threshold. This may reflect the inherent "fuzziness" of such a threshold radius, plus the fact that, due to projection effects, stars associated with streaming motions can be superposed on the sky with regions of stars that do not show any streaming.

The data are described in Section 6.2. In Section 6.3, we review the bisector test used in Mateo, Olszewski and Walker (2008). The cosine model is introduced in Section 6.4 and used to estimate the magnitude of streaming motion and motivate a test for significance. Threshold models are considered in Section 6.5. Section 6.6 contains remarks, outlining possible extensions. The Appendix provides the technical arguments for proving some of the asymptotic results used in the chapter.

6.2 The Data

The Data. The data used here consist of position and velocity measurements for candidate member stars from Leo I. These were derived from observations using the multi-fiber Hectochelle spectrograph on the MMT telescope at Las Campanas Observatory during March and April of 2005, 2006, and 2007. The raw spectra were converted to velocity measurements using **fxcor** in IRAF (the Image Reconstruction and Analysis Facility), which returns a velocity measurement and an estimate of the standard deviation of measurement error for each star. A detailed description of the observation and reduction processes

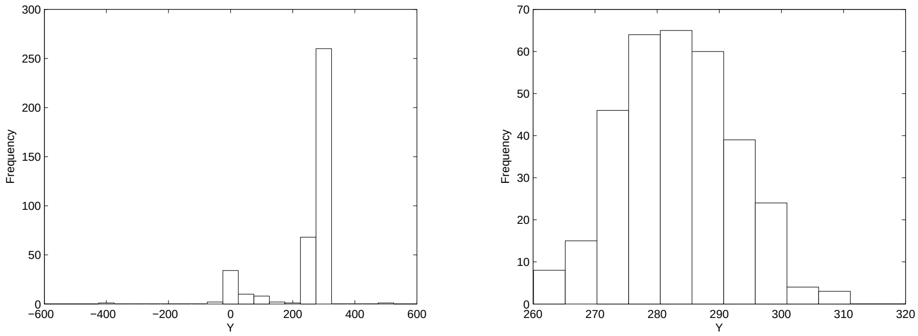

Figure 6.1: Histogram of velocities for Leo I before and after trimming (note the change in range in the two plots)

is included in Mateo, Olszewski and Walker (2008). For each star the four variables of primary interest here were line of sight velocity, position projected on the plane orthogonal to the line of sight, and the standard deviation of measurement error for the velocity. Velocities Y and the standard deviations Σ are expressed in km/sec. Position is expressed in polar coordinates (R, Θ) with R measured in arc seconds and Θ in degrees, so defined that $\Theta = 0$ along the major axis. For Leo I, 400 arc seconds are roughly 500 parsecs.

Trimming. A complicating feature of the data is that not all stars in the sample are really members of the galaxy. Some are foreground stars, located along our line of sight toward Leo I. Fortunately, due to Leo I's large systemic velocity, the radial velocities of non-members are quite distinct from those of the galaxy itself. Velocities of the galaxy members are fairly tightly clustered around a well-defined and, in this case, very large positive velocity, while velocities of non-members have a much broader distribution centered much closer to zero heliocentric velocity. See Figure 6.1. To eliminate the foreground stars, we computed (an estimate of) the probability that each star is a galaxy member, using the method of Sen, Walker and Woodroofe (2008). For the Leo

Table 6.1: Descriptive statistics for Leo I

	R	Θ	$\cos(\Theta)$	Y	Σ
min	2.3000	-256.3000	-1.0000	260.1000	1.6000
max	848.5000	99.8000	1.0000	311.1000	7.6000
med	259.8000	-74.2000	0.0932	282.6500	2.0000
mean	283.3213	-77.6591	0.0932	283.0927	2.1302
stdev	171.9573	100.4565	0.7619	9.4144	0.6470

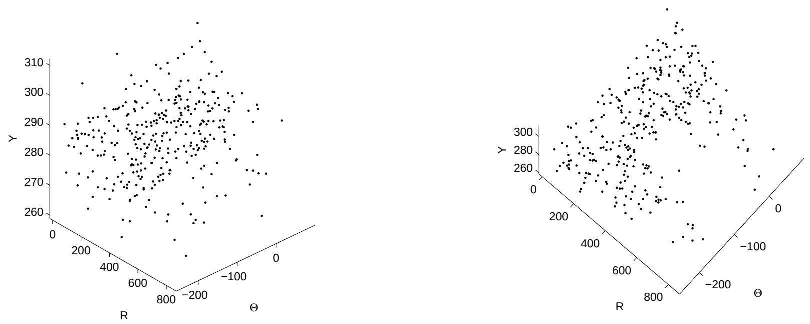

Figure 6.2: Scatterplot of (R, Θ, Y) for the Leo I data from two different perspectives

I, these probabilities were either at least .99 or at most .01. We eliminated the stars with low probabilities and kept 328 others. Some descriptive statistics of the trimmed sample are presented in Table 1. Observe that the trimmed sample consists of stars whose velocities are within three standard deviations of their mean. Figure 6.2 shows a scatter plot of positions and velocities for the trimmed sample.

Selection. The data are regarded as a random sample from Leo I, but not a simple random sample, since some regions were sampled more extensively than others. Thus, the joint density of (R, Θ) is of the form

$$f(r, \theta) \propto u(r, \theta) g(r, \theta),$$

where g is the density of R and Θ within the population and u is the selection function. Figure 6.3 presents a scatter plot of (R, Θ) for the trimmed sample

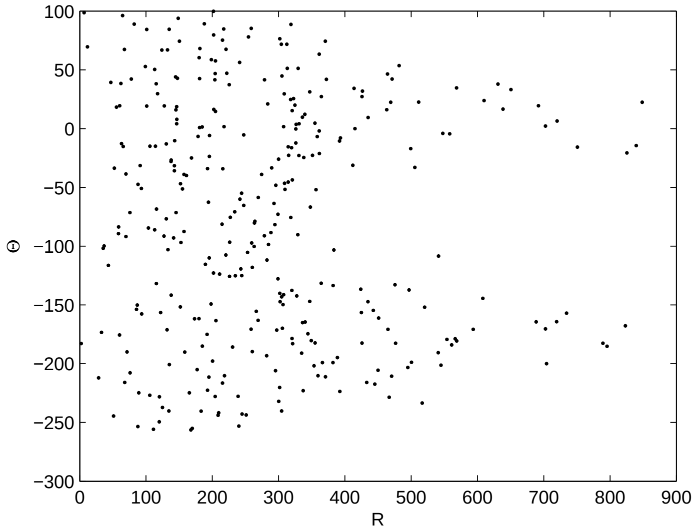

Figure 6.3: Scatterplot of (R, Θ) for the Leo I data

from which some effects of selection may be seen: Within the population of Leo I stars, it is not unreasonable to assume that R and Θ are independent and that Θ has a uniform distribution [Mateo, Olszewski and Walker (2008)]. The data in Figure 6.3 are clearly not consistent with this assumption, though this was not intentional since candidate members were selected as uniformly as feasible in Θ over the full range of R shown in Figure 6.3.

If we do suppose that R and Θ are independent and that Θ has a uniform distribution, within the population, then it is possible to estimate the selection function. The marginal density of R within the population of Leo I stars can be estimated with some precision from the large sample of positions reported in Irwin and Hatzidimitriou (1995). Thus the joint density g of R and Θ can be estimated with some precision. It is also possible to estimate f from our selected sample, using a kernel estimate, for example, and then u can be recovered from (6.1). Calculations of this nature are reported in Wang et al. (2005). We do not pursue this here because most of our analysis is conditional on position and, so, unaffected by the selection.

A Model. To describe the effects of streaming motion let V denote the line of sight velocity of a star and suppose that within a galaxy: R and Θ have a joint density g and $V = \nu(R, \Theta) + \epsilon$, where ϵ is a random fluctuation with mean 0 and variance σ^2, and ϵ is independent of (R, Θ). Thus, $\nu(r, \theta)$ is the expected velocity, given $R = r$ and $\Theta = \theta$. Velocity is measured with some error. We observe (Y, Σ), where $Y = V + \delta$ and the conditional distribution of δ given $(R, \Theta, \epsilon, \Sigma)$ is normal with mean 0 and standard deviation Σ. Thus, for the selected sample, $(R_i, \Theta_i, Y_i, \Sigma_i),\ i = 1, \cdots, n = 328$, are independent and identically distributed random vectors for which (R_i, Θ_i) have density f,

$$Y_i = \nu(R_i, \Theta_i) + \epsilon_i + \delta_i,$$

where $(R_i, \Theta_i),\ \epsilon_i$, and Σ_i are independent, and the conditional distribution of δ_i given $(R_i, \Theta_i, \epsilon_i, \Sigma_i)$ is normal with mean 0 and variance Σ_i^2.

For the remainder of the chapter, let $r_1 \le r_2 \le \cdots \le r_n$ denote the ordered values of $R_1, \cdots, R_n$, and let $\theta_1, \cdots, \theta_n,\ \sigma_1, \cdots, \sigma_n$, and $y_1, \cdots, y_n$ the concomitant order statistics of Θ_i, Σ_i, and Y_i. To avoid selection effects, we condition on the position variables $r_1, \cdots, r_n, \theta_1, \cdots, \theta_n$ in subsequent analysis. Probability and expectation mean conditional probability and expectation, unless otherwise noted.

6.3 The Bisector Test

An intuitive test for the presence of streaming motion was developed in Mario, Olszewski, and Walker (2008). To begin, stars distant from the center, say $r \ge r_0$, were selected. There were two reasons for this selection: The effect of streaming motion is expected to be small for stars close to the center and increase with distance from it, and the sample divides into two quite distinct branches for stars at least 400 arc sec from the center. See Figure 6.3. This resulted in reduced samples which were then divided into two groups by passing bisectors through the data set, and the difference in average velocities for stars in the two groups were computed. The bisectors were of the form $\cos(\theta - \omega) = 0$, where ω was allowed to vary.

In more detail, let

$$\Delta V(\omega) = \frac{\sum_{r_i > r_0, \cos(\theta_i - \omega) > 0} y_i/\sigma_i^2}{\sum_{r_i > r_0, \cos(\theta_i - \omega) > 0} 1/\sigma_i^2} - \frac{\sum_{r_i > r_0, \cos(\theta_i - \omega) \le 0} y_i/\sigma_i^2}{\sum_{r_i > r_0, \cos(\theta_i - \omega) \le 0} 1/\sigma_i^2},$$

and consider the test statistic $B = \max_\omega \Delta V(\omega)$. The attained significance levels for the reduced samples $r > 400$, $r > 455$, $r > 600$, and $r < 400$, were .030, .006, .014, and .101, using a permutation test.

The idea is sound, but there are details. Supposing that $\nu(r, \theta) = \nu$ is constant, so that there is no streaming motion, let

$$\hat{\nu}_0 = \frac{\sum_{i=1}^n y_i/(\hat{\sigma}_0^2 + \sigma_i^2)}{\sum_{i=1}^n 1/(\hat{\sigma}_0^2 + \sigma_i^2)} \quad \text{and} \quad \hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n [(y_i - \hat{\nu}_0)^2 - \sigma_i^2].$$

Then $\hat{\nu}_0$ and $\hat{\sigma}_0^2$ are $\sqrt{n}$-consistent estimators of ν and σ^2. Using the weights $1/(\hat{\sigma}_0^2 + \sigma_i^2)$ in place of $1/\sigma_i^2$ in (6.1) and permuting $(y_1, \sigma_1), \cdots, (y_n, \sigma_n)$ in the permutation test, we obtained somewhat higher significance levels. Plots of $\Delta V(\omega)$ for the reduced samples $r < 400$ and $r > 500$, are shown in Figure 6.4. One expects the effect of streaming motion to be large along the major axis of Leo I, and this is the case in Figure 6.4 and others like it (not included). Given that the sample divides into two distinct branches, corresponding to stars on the two sides of the galaxy along the major axis, the test that rejects for large values of $|\Delta V(0)|$ is also considered.

Table 2 shows (estimated) significance levels for B and $|\Delta V(0)|$ for several values of r_0. While the significance levels are higher than reported in Mario,

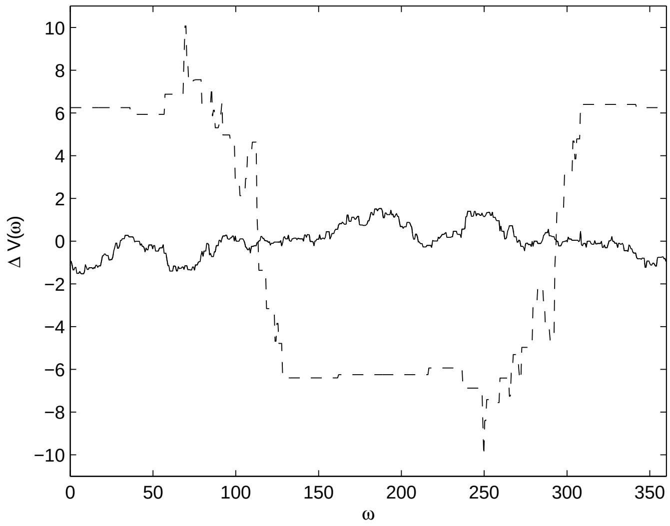

Figure 6.4: ΔV for $r < 400$ (solid) and $r > 500$ (dashed)

Table 6.2: Test statistics and significance levels

	B	p_B	$\Delta V(0)$	p_0
$r < 400$	1.5264	.809	0.9485	.401
$r > 400$	5.9065	.210	4.2182	.108
$r > 450$	7.5152	.106	4.4935	.132
$r > 500$	10.0682	.032	6.2498	.070
$r > 600$	12.2756	.129	9.9049	.052
$r > 700$	15.0053	.080	11.2687	.064
$r > 750$	19.4157	.047	2.8974	.688

Olszewski, and Walker (2008), they still suggest that streaming motion in present for stars sufficiently far from the center. The dependence on r_0 is troubling, however, and the results are far from conclusive. In the next section, we present another test which avoids the arbitrary choice of r_0, at the expense of setting $\omega = 0$.

The details of the permutation test are as follows. Consider a test statistic $T = T(\mathbf{r}, \theta, \mathbf{y}, \sigma)$, where $\mathbf{r} = (r_1, \cdots, r_n)$, $\theta = (\theta_1, \cdots, \theta_n)$, $\mathbf{y} = (y_1, \cdots, y_n)$, and $\sigma = (\sigma_1, \cdots, \sigma_n)$. If there is no streaming motion, then (R, Θ) and (Y, Σ)

are independent. In this case $(y_1, \sigma_1), \cdots, (y_n, \sigma_n)$ are conditionally i.i.d. given $(r_1, \theta_1), \cdots, (r_n, \theta_n)$, and the conditional probability that $T > t$ given $(\mathbf{r}, \theta)$ and the unordered values $\{(y_1, \sigma_1), \cdots, (y_n,\ \sigma_n)\}$ is $\#\{\pi : T(\mathbf{r}, \theta, \pi\mathbf{y}, \pi\sigma) > t\}/n!$, where π denotes a permutation of $\{1, \cdots, n\}$ and $\pi\mathbf{y}$ and $\pi\sigma$ denote permuted versions of $(y_1, \cdots, y_n)$ and $(\sigma_1, \cdots, \sigma_n)$. Of course, it is not possible to examine all $n!$ permutations, but it is possible to estimate the conditional probability by sampling permutations. The significance levels listed in Table 2, were obtained from $10,000$ permutations of the reduced samples. Observe that we permute the pairs (y_i, σ_i).

6.4 The Cosine Model

We now suppose that $\nu(r, \theta) = E(Y|R = r, \Theta = \theta)$ is of the form

$$\nu(r, \theta) = \nu + \lambda(r)\cos(\theta),$$

where ν is a constant and λ is a non-negative, non-decreasing function. Thus, $|\nu(r, \theta) - \nu|$ is assumed to be non-decreasing in r and largest along the major axis ($\theta = 0$).

Estimation. Assuming σ^2 to be known, the (weighted) conditional least squares estimators $\hat{\nu}$ and $\hat{\lambda}$, given $(\mathbf{r}, \theta, \sigma)$, minimize

$$\sum_{i=1}^{n} \frac{[y_i - v - u(r_i)\cos(\theta_i)]^2}{\sigma^2 + \sigma_i^2},$$

with respect to $v \in I\!R$ and non-negative, non-decreasing functions u. Differentiation then gives the following conditions for the least squares estimators

$$\sum_{i=1}^{n} \frac{y_i - \hat{\nu} - \hat{\lambda}(r_i)\cos(\theta_i)}{\sigma^2 + \sigma_i^2} = 0$$

and

$$\sum_{i=1}^{n} \cos(\theta_i) \left(\frac{y_i - \hat{\nu} - \hat{\lambda}(r_i)\cos(\theta_i)}{\sigma^2 + \sigma_i^2} \right) \xi_i \le 0$$

for all non-negative, non-decreasing $0 \leq \xi_1 \leq \cdots \leq \xi_n$. We will use these conditions with σ^2 replaced by

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left\{ [y_i - \hat{\nu} - \hat{\lambda}(r_i)\cos(\theta_i)]^2 - \sigma_i^2 \right\}$$

Thus, letting $\hat{w}_i = 1/(\hat{\sigma}^2 + \sigma_i^2)$ and and $\hat{W}_n = \hat{w}_1 + \cdots + \hat{w}_n$,

$$\hat{\nu} = \frac{1}{\hat{W}_n} \sum_{i=1}^{n} \hat{w}_i [y_i - \hat{\lambda}(r_i)\cos(\theta_i)]$$

and

$$\hat{\lambda}(r_k) = \max[0, \tilde{\lambda}(r_k)]$$

where

$$\tilde{\lambda}(r_k) = \max_{i \leq k} \min_{j \geq k} \frac{\hat{w}_i \cos(\theta_i)(y_i - \hat{\nu}) + \cdots + \hat{w}_j \cos(\theta_j)(y_j - \hat{\nu})}{\hat{w}_i \cos^2(\theta_i) + \cdots + \hat{w}_j \cos^2(\theta_j)}.$$

Alternatively, letting $\hat{T}_k = \hat{w}_1 \cos^2(\theta_1) + \cdots + \hat{w}_k \cos^2(\theta_k)$,

$$\Lambda^{\#}(t) = \sum_{i:\hat{T}_i \leq t} \hat{w}_i \cos(\theta_i)[y_i - \hat{\nu}],$$

and $\tilde{\Lambda} = \mathrm{GCM}(\Lambda^{\#})$, the greatest convex minorant of $\Lambda^{\#}$, $\tilde{\lambda}(r_k) = \tilde{\Lambda}'(\hat{T}_k)$, the left hand derivative. See Robertson, Wright and Dykstra (1988), Chapter 1 for background on isotonic estimation.

For Leo I, iterating (6.1), (6.1), and (6.1) leads to convergence to three decimal places after four iterations. For this data set, $\hat{\sigma} = 9.0107$ and $\hat{\nu} = 283.1040$. The function $\hat{\lambda}$ is graphed in left panel of Figure 6.5. The large value at the right end point is almost certainly due to the spiking problem Woodroofe and Sun (1993). To eliminate spiking we replace $\hat{\lambda}(r_n)$ by 6.193, the average of the last thirteen values of $\hat{\lambda}(r_k)$, adapting the suggestion of Kulikov and Lopuhaä (2006) to our context where data are much sparser.

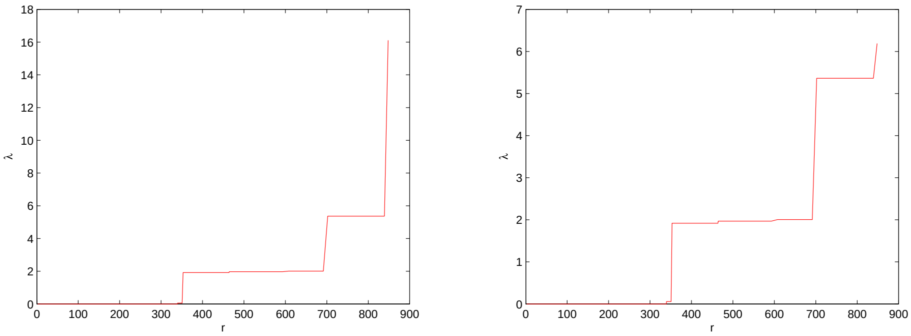

Figure 6.5: Left: $\hat{\lambda}$ before truncation. Right: $\hat{\lambda}$ after truncation.

This limits the effect of the last observation in the subsequent calculations. The truncated $\hat{\lambda}$ appears in the right panel of Figure 6.5.

Confidence Intervals. The asymptotic distribution of $\hat{\lambda}(r)$ can be derived under modest conditions. Suppose that $\lambda'(r) > 0$, and let $C = E[1/(\sigma^2 + \Sigma^2)]$ and

$$\gamma_r = 2\left|\frac{\lambda'(r)}{2C\int f(r,\theta)\cos^2(\theta)d\theta}\right|^{\frac{1}{3}}.$$

Then, a (fairly) straightforward application of the Argmax Theorem, van der Vaart and Wellner (2000), shows that the asymptotic unconditional distribution of $\mathbb{C}_n = n^{\frac{1}{3}}[\hat{\lambda}(r) - \lambda(r)]/\gamma_r$ is Chernoff's distribution, Groeneboom and Wellner (2001), the distribution of $\arg\min_t \mathbb{W}(t) + t^2$, where $\mathbb{W}$ denotes a standard two sided Brownian motion. Thus, $\mathbb{C}_n$ is an asymptotic pivot, but it is difficult to use this result to set confidence intervals for $\lambda(r)$, because it is difficult to estimate $\lambda'(r)$ and hence the normalizing constant γ_r. Moreover, even the condition $\lambda'(r) > 0$ is suspect on the interval where $\hat{\lambda} = 0$.

It is possible to avoid the problem of estimating γ_r, though not the condition $\lambda'(r) > 0$, by adapting the likelihood based confidence intervals of Banerjee and Wellner (2001) to the present problem. For fixed $r_0, \xi_0 > 0$, and $\sigma > 0$,

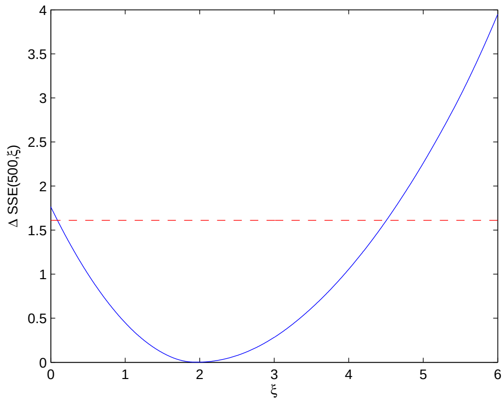

Figure 6.6: $\Delta SSE(500, \xi)$ as ξ varies from 0 to 6, with the 90% cut-off mark

let

(6.1)

$$\Delta SSE(r_0, \xi_0) = \min_{u(r_0)=\xi_0} \sum_{i=1}^{n} \frac{[y_i - v - u(r_i)\cos(\theta_i)]^2}{\sigma^2 + \sigma_i^2} - \min \sum_{i=1}^{n} \frac{[y_i - v - u(r_i)\cos(\theta_i)]^2}{\sigma^2 + \sigma_i^2},$$

where implicitly both minimizations are over $v \in I\!R$ and non-negative, non-decreasing functions u. If $\lambda(r_0) = \xi_0$ and $\lambda'(r_0) > 0$, then $\Delta SSE(r_0, \xi_0)$ has a limiting distribution that does not depend on any unknown parameters, under regularity conditions; and the same asymptotic distribution is obtained with σ replaced by $\hat{\sigma}$. See the Appendix. A description of the asymptotic distribution, including graphs and percentiles, may be found in Banerjee and Wellner (2005). In particular, (Monte Carlo estimates of) the 90^{th} and 95^{th} percentile are 1.61 and 2.29 and, for example, $\{\xi : \Delta SSE(r_0, \xi) \leq 1.61\}$ is an approximate 90% confidence set for $\lambda(r_0)$. A plot of $\Delta SSE(500, \xi)$ is shown in Figure 6.6, and selected confidence intervals are listed in Table 3.

Resampling provides still another way to set confidence intervals. Recent results, Kosorok (2007), Lee and Pun (2006), Léger, C. and MacGibbon (2006), Sen, Banerjee and Wellner (2008), some obtained for related problems, suggest that direct use of the bootstrap will not provide consistent estimators of the

Table 6.3: 90% and 95% confidence intervals for $\lambda(r)$

	ΔSSE						Bootstrap				
	0.90			0.95			0.90		0.95		
r_0	L	U	CP	L	U	CP	L	U	L	U	$\hat{\lambda}$
400	0	3.54	.901	0	3.86	.952	0	3.57	0	3.57	1.92
450	0	3.63	.887	0	4.01	.943	0	3.74	0	3.85	1.92
500	0.10	4.50	.882	0	5.02	.936	0	3.58	0	3.90	1.98
550	0.26	4.65	.852	0	5.19	.916	0	3.44	0	3.76	1.98
600	0.26	6.66	.827	0	7.30	.897	0	3.30	0	3.64	1.98
650	0.36	6.70	.865	0.05	7.39	.922	0	3.58	0	3.97	1.99
700	0.36	8.88	.913	0.05	9.56	.961	0	4.26	0	4.69	1.99
750	1.85	8.88	.906	1.37	9.56	.952	0.44	7.86	0	8.37	5.37

Notes: The leftmost column shows the radial distance. The next two columns are lower and upper endpoints of an approximate 90% confidence interval computed from ΔSSE; fourth column is a bootstrap estimate of the coverage probability; the fifth, sixth and seventh columns provide the same information for 95% confidence intervals. The next four columns are lower and upper endpoints of approximate 90% and 95% confidence intervals computed from the bootstrap. The last column provides the value of $\hat{\lambda}$.

distribution of sampling error but that use of a smoothed bootstrap or m out of n bootstrap will. Thus let

$$\hat{\lambda}_s(e^t) = \int_{-\infty}^{\infty} \hat{\lambda}(e^u)\frac{1}{b}K\left(\frac{u-t}{b}\right) du,$$

where K is a kernel and b a bandwidth. We used the standard normal density for K and chose the bandwidth $b = .1$ (subjectively) to compromise between smoothness and fit. The result is shown in left panel of Figure 6.7. The right panel shows the derivative of the smoothed estimator, illustrating the difficulty in estimating $\lambda'(r)$.

Now, let e_i denote the residuals, $e_i = y_i - \hat{\nu} - \hat{\lambda}_s(r_i)\cos(\theta_i),\ i = 1, \cdots, n$. Let $x_i = (\sigma_i, e_i)$, and let $F^{\#}$ denote the empirical distribution of $x_1, \cdots, x_n$. Further, let $(S_1, Z_1), \cdots, (S_n, Z_n) \sim F^{\#}$ be conditionally independent given $(\mathbf{r}, \theta, \mathbf{y}, \sigma)$; let

$$y_i^* = \hat{\nu} + \hat{\lambda}_s(r_i)\cos(\theta_i) + Z_i \quad \text{and} \quad \sigma_i^* = S_i;$$

and let $\hat{\lambda}^*$ denote the (truncated) isotonic estimator (6.1) computed from $y_1^*, \cdots, y_n^*$ and $\sigma_1^*, \cdots, \sigma_n^*$ with $r_1, \cdots, r_n$ and $\theta_1, \cdots, \theta_n$ held fixed. To set confidence intervals for $\lambda(r_0)$, we estimate the distribution of $\hat{\lambda}(r_0) - \lambda(r_0)$ by the conditional distribution of $\hat{\lambda}^*(r_0) - \hat{\lambda}_s(r_0)$, which may be computed from simulation. The left panel in Figure 6.8 shows a histogram of $10,000$ values of $\hat{\lambda}^*(500) - \hat{\lambda}_s(500)$. Bootstrap confidence intervals for selected r_0 are listed in Table 3. Similarly, to set confidence bands for λ, we approximate the distribution of $D = \max_r |\hat{\lambda}(r) - \lambda(r)|$ by that of $D^* = \max_r |\hat{\lambda}^*(r) - \hat{\lambda}_s(r)|$. The right panel in Figure 6.8 shows a histogram of $10,000$ values of D^*. The 90^{th} and 95^{th} percentiles of this distribution are 5.194 and 6.074. We have not standardized these variables before bootstrapping, because it is difficult to estimate λ'.

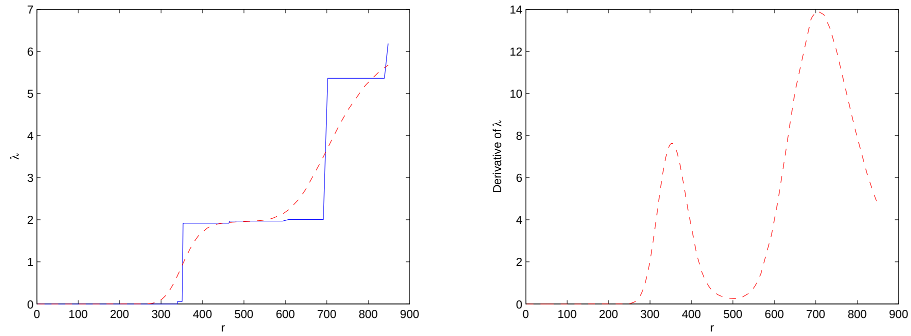

Figure 6.7: Left: The smoothed (dashed) and unsmoothed (solid) estimators with $b = 0.1$. Right: The derivative of the smoothed estimator.

Unfortunately, there are major differences between the two methods for setting confidence intervals. To some extent these can be explained by the constructions: The bootstrap intervals attempt to balance the error probabilities equally, left and right; the intervals derived from ΔSSE make no such attempt. There are more serious differences, however, between the asymptotic values and the bootstrap estimates. The difference can be seen in the left panel of Figure 6.8: The histogram is asymmetric, whereas Chernoff's distribution is symmetric. The ΔSSE method depends on the approximations $P\{\Delta SSE[r_0, \lambda(r_0)] \leq 1.61\} \approx .90$ and $P\{\Delta SSE[r_0, \lambda(r_0)] \leq 2.29\} \approx .95$ where now P denotes the unconditional probability. The bootstrap estimates of these probabilities, $P^*\{\Delta SSE^*[r_0, \hat{\lambda}_s(r_0)] \leq 1.61\}$ and $P^*\{\Delta SSE^*[r_0, \hat{\lambda}_s(r_0)] \leq 2.29\}$ are reported in columns three and six of Table 3. There is good agreement for $r_0 \leq 450$ and $r_0 \geq 700$. This is important, because the positive lower confidence bounds on the last two lines of Table 3 reinforce the conclusions in Section 6.3 that there is streaming motion. But the bootstrap estimates are substantially less than the nominal values for $550 \leq r \leq 650$, an interval that includes values for which $\hat{\lambda}_s'(r_0)$ is very small, and the positive lower confidence limits in this region should not be trusted. The disagreement between

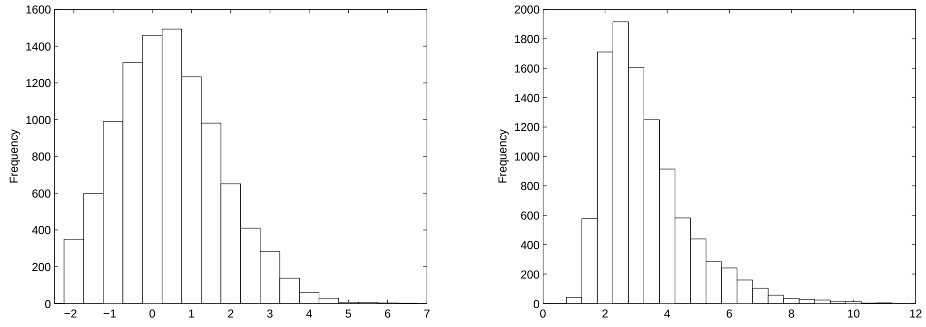

Figure 6.8: Histograms of 10,000 realizations of $\hat{\lambda}^*(500) - \hat{\lambda}_s(500)$ (left) and D^* (right)

the bootstrap intervals and ones derived from asymptotic distributions is difficult to resolve, in part, because the justification for the bootstrap is itself asymptotic and requires the condition $\lambda'(r_0) > 0$.

Testing. Within the cosine model (6.1), testing for steaming motion means testing the null hypothesis $\lambda = 0$. The positive lower confidence limits on the last lines of Table 3 suggest that this hypothesis can be rejected. This point can be made in another way that does not depend on asymptotics or even the validity of (6.1). Consider the F-like statistic

$$\mathcal{F} = \sum_{i=1}^{n} \hat{w}_i \cos^2(\theta_i)\hat{\lambda}^2(r_i)$$

which suggests itself for this problem. If ν and σ^2 were known, $\epsilon_1, \cdots, \epsilon_n$ were normal, $\hat{w}_i$ were replaced by $w_i = 1/(\sigma^2 + \sigma_i^2)$, and $\hat{\lambda}$ were replaced by the isotonic estimator for known ν and σ^2, then $-2\mathcal{F}$ would be the log-likelihood ratio statistic for testing $\lambda = 0$. See, Chapter 2 of Robertson, Wright and Dykstra (1988).

For the Leo I data set, the observed value of $\mathcal{F}$ was 6.69. We again assess significance from the permutation distribution of $\mathcal{F}$, but computed from the full sample $(r_i, \theta_i, y_i, \sigma_i)$, $i = 1, \cdots, n = 328$. In a sample of 10,000 permutations, the permuted value of $\mathcal{F}$ exceeded the observed value 553 times, roughly confirming the conclusion based on confidence intervals and asymptotic calculations. The effect of truncation on the test statistic is amusing. Without the truncation the value of $\mathcal{F}$ would have been substantially larger 8.72, but the significance level would have been essential unchanged .0543.

Observe that (6.1) was used only to motivate the form of the test statistic. The F-like test statistic serves also as a test of the hypothesis $\nu(r,\theta) = \nu$, in (6.1).

6.5 Thresholds and the Break Point

By a threshold or breakpoint we mean a distance from the center of Leo I below which there is no streaming motion, or very little, and above which streaming motion is appreciable. We consider two approaches to defining and estimating such a point, change point models and split points, as in Banerjee and McKeague (2007).

Change Point Models. Let τ denote an upper limit for r. In the change point model it is assumed that there is a $\rho > 0$ for which $\lambda(r) = 0$ for $r \le \rho$ and $\lambda(r) > 0$ for $\rho < r \le \tau$, in which case we call ρ the *threshold.*

We may obtain an upper confidence bound by modifying the F-like statistic (6.2). For a given ρ_0 consider the hypothesis $\rho \ge \rho_0$. Let

$$\mathcal{F}(\rho_0) = \sum_{i:r_i \le \rho_0} \hat{w}_i \cos^2(\theta_i)\hat{\lambda}^2(r_i)$$

and let m be the largest integer for which $r_m \le \rho_0$. Then the conditional null distribution of $\mathcal{F}(\rho_0)$ is invariant under permutations of $(y_1, \sigma_1), \cdots, (y_m, \sigma_m)$, so that significance can again be assessed from a permutation distribution of $\mathcal{F}(\rho_0)$. Of course, the set of ρ_0 for which the hypothesis is accepted at level α is a level $1-\alpha$ confidence set for ρ. For the Leo I data set with $1-\alpha = .9$,

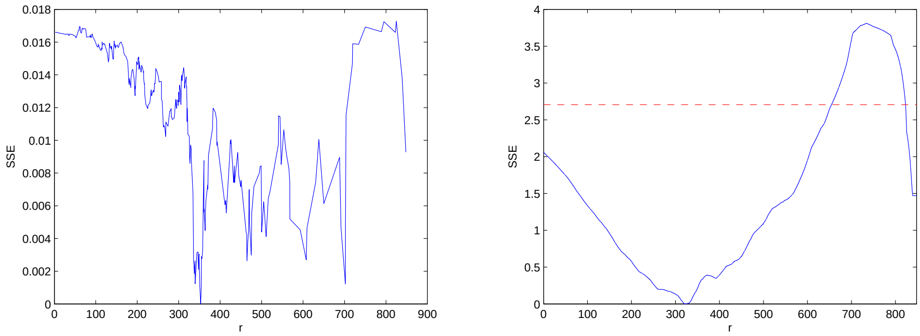

Figure 6.9: Graphs of $SSE_r - SSE_{\hat{\rho}}$ for $\psi(x) = \max(0, x)$ (left) and $\psi(x) = \mathbf{1}_{(0,\infty)}(x)$ (right)

this hypothesis is rejected when $\rho_0 = 720$ which, therefore, serves as an upper confidence bound. Again, the cosine model (6.1) was used only to motivate the form of the test statistic. The test just described also serves as a test of $\nu(r, \theta) \equiv \nu$ for all θ and all $r \leq \rho_0$.

Unsurprisingly, adopting an even more structured model suggests a lower bound. Suppose that $\lambda(r) = \beta\psi(r - \rho)$, where $\beta > 0$ is an unknown parameter and ψ is a known function for which $\psi(x) = 0$ for $x < 0$ and $\psi(x) > 0$ for $x > 0$. Thus,

$$y_i = \nu + \beta\psi(r_i - \rho)\cos(\theta_i) + \epsilon_i + \delta_i$$

for $i = 1, \cdots, n$. For a fixed ρ this is a simple linear regression model. Let $\hat{\beta}_r$ and $\hat{\nu}_r$ denote the weighted least squares estimators, using the weights $\hat{w}_i = 1/(\hat{\sigma}^2 + \sigma_i^2)$, derived from (6.2) assuming $\rho = r$, and let SSE_r denote the residual sum of squares,

$$SSE_r = \sum_{i=1}^{n} \hat{w}_i[y_i - \hat{\nu}_r - \hat{\beta}_r\psi(r_i - r)\cos(\theta_i)]^2.$$

Then the LSE $\hat{\rho}$ of ρ minimizes SSE_r with respect to r. Figure 6.9 shows graphs of $SSE_r - SSE_{\hat{\rho}}$ for two choices of ψ, $\psi(x) = \mathbf{1}_{(0,\infty)}(x)$ and $\psi(x) =$

$\max(0, x)$. The latter choice leads to the the segmented regression model considered in Hinkley (1971), Feder (1975), and recently in Hušková (1998). For this choice it may be shown that $SSE_\rho - SSE_{\hat\rho}$ has a limiting χ_1^2 distribution, assuming (6.2). So, $\{r : SSE_r - SSE_{\hat\rho} \le c\}$ is an asymptotic level $P[\chi_1^2 \le c]$ confidence set for ρ. For Leo I, the 90% asymptotic confidence set $[0, 654.5] \cup [823.1, 848.5]$ so obtained is disconnected, but there are only four stars for which $823.1 \le r \le 848.5$, and this interval is of little interest. Letting $\psi(x) = \mathbf{1}_{(0,\infty)}(x)$ leads to (a minor variation on) the classical change point problem. The asymptotic distribution of $SSE_\rho - SSE_{\hat\rho}$ may be obtained for this case too; but it is complicated and unnecessary in the sense that $SSE_r - SSE_{\hat\rho}$ rises and falls so sharply near $r = 333.5$ and $r = 702.7$.

Split Points. It is possible to define and estimate a breakpoint without assuming that $\lambda(r)$ is actually equal to 0 for small r, by fitting a stump model $\beta\mathbf{1}_{(\gamma,\tau]}$ to λ, as in Banerjee and McKeague (2007). This is accomplished by minimizing an expression of the form

$$\kappa(b, r) = \int_0^r \lambda^2(s)h(s)dr + \int_r^\tau [\lambda(s) - b]^2 h(s)ds,$$

with respect to b and r. Here h is a positive weight function. We used $h = 1$ in Figure 6.10. Another possibility is to let h be the marginal density of R, which is known from Irwin and Hatzidimitriou (1995). The minimization with respect to b, of course, is simple, and

$$\kappa_0(r) := \min_b \kappa(b, r) = \int_0^\tau \lambda^2(s)ds - \frac{[\Lambda(\tau) - \Lambda(r)]^2}{\tau - r},$$

where $\Lambda(r) = \int_0^r \lambda(s)ds$. We define the *break point* γ to be the minimizing value of r, assuming that the minimum is attained at a unique point. If λ is

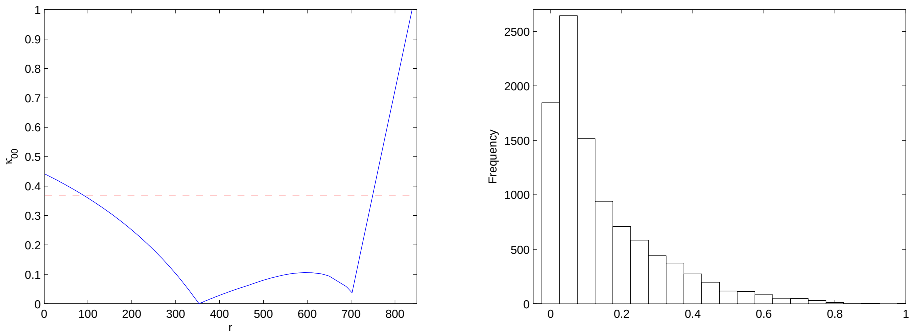

Figure 6.10: Left: $\hat{\kappa}_{00}(r)$. Right: Bootstrap estimate of the distribution of $\hat{\kappa}_{00}(\gamma)$.

continuous and $\lambda(0) < \Lambda(\tau)/(2\tau)$, then

$$\lambda(\gamma) = \frac{[\Lambda(\tau) - \Lambda(\gamma)]}{2(\tau - \gamma)}.$$

Observe that if there is a threshold ρ (for which $\lambda(r) = 0$ for $r \leq \rho$), then $\gamma \geq \rho$, because $\Lambda(r)$ is then constant for $r \leq \gamma$.

The simplest way to estimate γ is to replace λ by $\hat{\lambda}$ in the definition. The left panel of Figure 6.10 shows a graph of $\hat{\kappa}_{00}(r)$, where $\hat{\kappa}_{00}$ denotes κ_0 with λ replaced by $\hat{\lambda}$ and rescaled to take values between 0 and 1. That is, letting $\hat{\kappa}_0$ denote κ_0 with λ replaced by $\hat{\lambda}$ and $\hat{\gamma}$ a minimizing value of $\hat{\kappa}_0(r)$,

$$\hat{\kappa}_{00}(r) = \frac{\hat{\kappa}_0(r) - \hat{\kappa}_0(\hat{\gamma})}{\max_s \hat{\kappa}_0(s) - \hat{\kappa}_0(\hat{\gamma})}.$$

The estimated break point is 353.5 arc sec, but there is a near minimum at about 700.

Asymptotic theory provides little useful guidance here. The asymptotic unconditional distribution of $\hat{\gamma}$ can be obtained along the lines outlined in the Appendix, but depend on $\lambda'(\gamma)$ and would have to be approximated by simulation in any case. A bootstrap procedure does provide some guidance, however. Define y_i^* and σ_i^* as in (6.2); let $\hat{\gamma}_s$ denote the value of γ obtained by replacing λ by $\hat{\lambda}_s$; and let $\hat{\gamma}^*$ and $\hat{\kappa}_{00}^*$ denote the values of $\hat{\gamma}$ and $\hat{\kappa}_{00}$

respectively, computed from $y_1^*, \cdots, y_n^*, \sigma_1^*, \cdots, \sigma_n^*$, with $r_1, \cdots, r_n, \theta_1, \cdots, \theta_n$ held fixed. Then the conditional distribution of $\hat{\kappa}_{00}^*(\hat{\gamma}_s)$ provides an estimate of the distribution of $\hat{\kappa}_{00}(\gamma)$. A histogram of $10,000$ values of $\hat{\kappa}_{00}^*(\hat{\gamma}_s)$ is shown in the right panel of Figure 6.10. The 90^{th} and 95^{th} percentiles of this distribution are .3694 and .4690. So, for example, the set of r for which $\hat{\kappa}_{00}(r) \leq .3694$ is a bootstrap confidence set for γ. Unfortunately, this is a large interval, $[91.1, 734.3]$.

6.6 Some Remarks

1. The conclusions regarding the presence of streaming motion have to be tentative, because of the large significance levels. One of the key factors behind this is the comparatively small sample size ($n = 328$) and, in particular, the very few observations far out from the center of the galaxy, the region of interest. In fact, we just have 64 data points above 400 arc seconds. We hope that with more data in the future our methods can be used more effectively to draw stronger conclusions. We also expect to obtain data on other dwarf spheroidal galaxies, e.g., Draco, Fornax, etc. and will be applying variants of our methods to analyze the samples.

2. With more data we can resort to more flexible modeling. For example, instead of simply using $\cos\theta$, we could model the effect of angle θ, by a function $h(\cos\theta)$ or $h(\cos(\theta - \omega))$, where h is non–decreasing and ω is an unknown parameter representing the direction of tidal streaming. The velocity dispersion parameter, σ^2, a quantity of independent interest to astronomers, has been assumed to be constant in our approach. For Leo I, there is evidence for such an assumption [see Mario, Olszewski, and

Walker (2008)]. For other galaxies, it is conceivable that σ may depend on r, the radial distance from the center of the galaxy, in which case we ought to incorporate it in the analysis. Our methods should be adaptable to this heteroscedasticity, but may require non–trivial extensions.

3. Monotone regression splines provide a method for combining monotonicity constraint and smoothness. These were used effectively in Wang et al. (2005), (2008a), (2008b) and could be investigated in the present context.

4. The smoothed bootstrap was invoked at several places in this chapter for purposes of uncertainty assessment – for example, in constructing pointwise and simultaneous confidence bands for λ, and confidence sets for the split point γ. While there is evidence [see Kosorok (2007), Sen, Banerjee and Wellner (2008) and Léger and McGibbon (2006)] that the smoothed bootstrap provides consistent estimates of pointwise confidence sets for λ, the use of this method for approximating the distribution of D and that of $\hat{\kappa}_{00}(\gamma)$ remains to be vindicated. In particular, nothing is known about the limiting behavior of D or $\hat{\kappa}_{00}(\gamma)$. An alternative to the smoothed bootstrap would have been to use subsampling or the m out of n bootstrap.

5. The asymptotic distribution of the least squares estimate of γ in the split point model, derived in the Appendix, is curious in the sense that (a multiple of) Chernoff's distribution no longer arises and is replaced by a non-standard limit. We know of no other situations in the published literature where this limit distribution has been encountered. Furthermore, it does not seem possible to represent the limit as a multiple of a universal distribution, which renders the computation of quantiles diffi-

cult. We have also not been able to find a complete proof of the rate of convergence of the least squares estimators.

6. An interesting but difficult problem is to estimate the threshold parameter ρ nonparametrically, assuming only that $\lambda(r) = 0$ for $r \leq \rho$, $\lambda(r) > 0$ for $r > \rho$ and that λ is increasing. We expect the rate at which ρ can be estimated to depend crucially on the smoothness of the join between the two segments of the function at ρ; the smoother the join, the slower the convergence. The intuitive estimator $\inf\{t : \hat{\lambda}(t) > 0\}$ under-estimates ρ heavily. One suggested modification is to replace 0 by a positive threshold that decreases to 0 at an appropriate rate. Yet another approach would be to construct a penalized least squares estimate of λ under monotonicity constraints, where one penalizes monotone functions with low values of the threshold parameter.

7. Yet another way of estimating γ is to observe that $\gamma = d_0$ where $(v_0, \beta_0, d_0) = \arg\min_{(v,\beta,d)} \mathbb{M}(v, \beta, d)$ and

$$\mathbb{M}(v, \beta, d) = E\left[\{Y_i - v - \beta \mathbf{1}(R > d) \cos\Theta\}^2 \frac{1}{\phi(R, \Theta, \Sigma)}\right],$$

with $\phi(R, \Theta, \Sigma) = f(r, \theta)(\sigma^2 + \Sigma^2)/h(r)$. We can approximate this criterion function $\mathbb{M}$, by the empirical expectation and construct an estimate of γ as the threshold that minimizes the sample analogue. This method avoids the estimation of λ. However, it needs knowledge of ϕ, which in turn, involves estimation of f. This is not feasible with the currently available sample size, so we have not explored this approach in the chapter.

8. The derivations of unconditional asymptotic distributions of $\Delta SSE(r_0, \xi_0)$

and $\hat{\gamma}$ are outlined in the Appendix. We believe that the conditional distributions given $\mathbf{r}$ and θ may have the same limits, but do not have a complete proof.

Appendix

We start with deriving the limit distribution of $\Delta SSE(r_0, \xi_0)$ (see (6.1)) and $\hat{\lambda}$. The asymptotic distribution of $\hat{\gamma}$, in the split point model, is derived in the second part of the Appendix.

The residual sum of squares statistic. The main goal in this section is to analyze the (unconditional) limit behavior of the residual sum of squares statistic $\Delta SSE(r_0, \xi_0)$ introduced in (6.1) of Section 4. We study this quantity but with two simplifications – we assume that σ and ν are known. This simplification is justified because the estimates of ν and σ used in the chapter converge at a faster ($\sqrt{n}$) rate than the isotonic estimators of λ which drive the asymptotics of $\Delta SSE(r_0, \xi_0)$. See Huang (2002) for a discussion of this issue in the context of a semi-linear monotone regression model.

Define $\tilde{V}_i = (Y_i - \nu)/\cos(\Theta_i)$ (scaled response) and $W_i = \cos^2(\Theta_i)/(\sigma^2 + \Sigma_i^2)$. The unconstrained and constrained estimators of λ, $\hat{\lambda}_n$ and $\hat{\lambda}_n^0$ are characterized as:

$$\hat{\lambda}_n = \arg\min_{\lambda\uparrow} \sum_{i=1}^{n} [\tilde{V}_i - \lambda(R_i)]^2 \, W_i \quad \text{and} \quad \hat{\lambda}_n^0 = \arg \min_{\lambda\uparrow;\, \lambda(r_0)=\xi_0} \sum_{i=1}^{n} [\tilde{V}_i - \lambda(R_i)]^2 \, W_i .$$

The residual sum of squares is then given by:

$$\Delta SSE(r_0, \xi_0) \;\; = \;\; \sum_{i=1}^{n} [\tilde{V}_i - \hat{\lambda}_n^0(R_i)]^2 \, W_i - \sum_{i=1}^{n} [\tilde{V}_i - \hat{\lambda}_n(R_i)]^2 \, W_i \, .$$

Some notation is necessary. For a real–valued function f defined on $\mathbb{R}$, let $\text{slogcm}(f, I)$ denote the left–hand slope of the greatest convex minorant (GCM) of the restriction of f to the interval I. We abbreviate slogcm$(f, \mathbb{R})$ to slogcm(f). Take

$$\text{slogcm}^0(f) = (\text{slogcm}\,(f, (-\infty, 0]) \wedge 0)\, 1_{(-\infty,0]} + (\text{slogcm}\,(f, (0, \infty)) \vee 0)\, 1_{(0,\infty)}\,.$$

For positive constants c and d define the process $X_{c,d}(z) = c\,W(z) + d\,z^2$, where $W(z)$ is standard two-sided Brownian motion starting from 0. Set $g_{c,d} = \text{slogcm}(X_{c,d})$ and $g^0_{c,d} = \text{slogcm}^0\,(X_{c,d})$. For details about the processes $g_{c,d}$ and $g^0_{c,d}$, see Banerjee and Wellner (2001) and Banerjee (2007). Thus, $g_{1,1}$ and $g^0_{1,1}$ are the unconstrained and constrained versions of the slope processes associated with the "canonical" process $X_{1,1}(z)$. By Brownian scaling, the slope processes $g_{c,d}$ and $g^0_{c,d}$ can be related in distribution to the canonical slope processes $g_{1,1}$ and $g^0_{1,1}$. Set $\mathbb{D}_{c,d} = \int \left\{(g_{c,d}(u))^2 - (g^0_{c,d}(u))^2\right\} du$ and abbreviate $\mathbb{D}_{1,1}$ to $\mathbb{D}$. The following lemma holds [see, for example, Banerjee and Wellner (2001)].

Lemma 1. The random variable $\mathbb{D}_{c,d}$ has the same distribution as $c^2\,\mathbb{D}$.

To describe the asymptotic properties of the least squares estimates $\hat{\lambda}_n$ and $\hat{\lambda}^0_n$, define processes $\mathbb{G}_n$ and $\mathbb{V}_n$ as:

$$\mathbb{G}_n(t) = \mathbb{P}_n\,[W\,(1(R \leq t)] \;\text{ and }\; \mathbb{V}_n(t) = \mathbb{P}_n\,[\tilde{V} W\,1(R \leq t)]$$

where $\mathbb{P}_n$ is the empirical distribution of $X_i = (R_i, \Theta_i, Y_i, \Sigma_i), i = 1, \cdots, n$. Then $\hat{\lambda}_n(t) = \text{slogcm}[\mathbb{V}_n \circ \mathbb{G}_n^{-1}](\mathbb{G}_n(t))$, and $\hat{\lambda}^0_n$ has a similar characterization in terms of slopes of greatest convex minorants of the same processes restricted to the intervals $(-\infty, r_0]$ and $[r_0, \infty)$. Set $\mathbb{V}^c_n(t) = \mathbb{P}_n\,[(\tilde{V} - \lambda(r_0))W\,1(R \leq t)]$. Letting $B := E[1/(\sigma^2 + \Sigma^2)] \int \cos^2(\theta) f(r_0, \theta) d\theta$, define localized versions of the processes $\mathbb{V}_n^c$ and $\mathbb{G}_n$ respectively, as:

$$\begin{aligned}\tilde{G}_n(z) &= n^{1/3}B^{-1}[\mathbb{G}_n(r_0 + z\,n^{-1/3}) - \mathbb{G}_n(r_0)] \text{ and} \\ \tilde{V}_n^c(z) &= n^{2/3}B^{-1}[\mathbb{V}_n^c(r_0 + z\,n^{-1/3}) - \mathbb{V}_n^c(r_0)].\end{aligned} \tag{6.2}$$

Define the localized LSE processes X_n and Y_n as:

$$\begin{aligned}X_n(z) &= n^{1/3}\{\hat{\lambda}_n(r_0 + z\,n^{-1/3}) - \lambda(r_0)\}, \\ Y_n(z) &= n^{1/3}\{\hat{\lambda}_n^0(r_0 + z\,n^{-1/3}) - \lambda(r_0)\}.\end{aligned}$$

Then,

$$(X_n(z), Y_n(z)) = (\text{slogcm}[\tilde{V}_n^c \circ \tilde{G}_n^{-1}](\tilde{G}_n(z)), \text{slogcm}^0[\tilde{V}_n^c \circ \tilde{G}_n^{-1}](\tilde{G}_n(z))).$$

Let $a = B^{-1/2}$ and $b = \lambda'(r_0)/2$. Then standard calculations show that the process $\tilde{G}_n(z)$ converges in probability to the deterministic function z, uniformly on compacta, and the process $\tilde{V}_n^c$ converges weakly, under the topology of uniform convergence on compacta to $X_{a,b}(z)$. Invoking continuous mapping arguments for slopes-of-greatest-convex-minorant estimators (see, for example, the proof of Theorem 2.1 in Banerjee (2007) and the companion technical report), we conclude that:

[A]: The processes $(X_n(z), Y_n(z))$ converge to the processes $(g_{a,b}(z), g_{a,b}^0(z))$ finite-dimensionally, and also in the space $L_2[-K, K] \times L_2[-K, K]$, for every $K > 0$. (The space $L_2[-K, K]$ is the space of real measurable functions defined on $[-K, K]$ equipped with the topology of L_2 convergence with respect to Lebesgue measure. The cartesian product carries the usual meaning, as in product space.)

[B]: Let D_n denote the interval around r_0 on which $\hat{\lambda}_n$ and $\hat{\lambda}_n^0$ differ. Then, $\tilde{D}_n := n^{1/3}(D_n - r_0)$ can be eventually trapped inside a compact interval around 0, with arbitrarily high probability.

Let $R_{(1)} < R_{(2)} < \ldots < R_{(n)}$ denote the ordered values of the radii and $\tilde{V}_{(i)}$ and $W_{(i)}$, the scaled response and weight corresponding to the i'th largest radius. Also, let J_n denote the set of indices such that $\hat{\lambda}_n(R_{(i)}) \neq \hat{\lambda}_n^0(R_{(i)})$. Then, the following hold (as easy consequences of the characterization of isotonic regression estimates as block-wise averages) with probability increasing to 1:

[C]: The set of indices J_n can be split into a set of ordered blocks of indices $B_1, B_2, \ldots, B_k$, such that: for each B_j, $\hat{\lambda}_n(R_{(i)})$ assumes the same value, say v_j, whenever $i \in B_j$, and this common value is characterized as $v_j = \sum_{i \in B_j} \tilde{V}_{(i)} W_{(i)} / \sum_{i \in B_j} W_{(i)}$. Also, $v_1 < v_2 < \ldots < v_k$.

[D]: The set of indices J_n can also be split into a set of ordered blocks of indices $B_1^0, B_2^0, \ldots, B_l^0$, such that: for each B_j^0, $\hat{\lambda}_n^0(R_{(i)})$ assumes the same value, say v_j^0, whenever $i \in B_j^0$, and this common value, so long as it does not equal ξ_0, is characterized as $v_j^0 = \sum_{i \in B_j^0} \tilde{V}_{(i)} W_{(i)} / \sum_{i \in B_j^0} W_{(i)}$. Also, $v_1^0 < v_2^0 < \ldots < v_l^0$.

Some algebra shows that $\Delta SSE(r_0, \xi_0) = I_n - II_n$, where

$$I_n = \sum_{i \in J_n} \{\hat{\lambda}_n^0(R_{(i)}) - \lambda(r_0)\}^2 W_{(i)} - 2 \left[\sum_{i \in J_n} \{\tilde{V}_{(i)} - \lambda(r_0)\}\{\hat{\lambda}_n^0(R_{(i)}) - \lambda(r_0)\} W_{(i)} \right]$$

and II_n has the form as I_n but with $\hat{\lambda}_n^0$ replaced by $\hat{\lambda}_n$. It is easy to see that the first term in the display defining I_n equals the sum within the square brackets in the second term, by breaking the latter into sums over the blocks

B_i^0 and using [D]. The term II_n can be simplified similarly using [C], yielding:

$$\begin{aligned}
\Delta SSE(r_0, \xi_0) &= \sum_{i \in J_n} \{\hat{\lambda}_n(R_{(i)}) - \lambda(r_0)\}^2 W_{(i)} - \sum_{i \in J_n} \{\hat{\lambda}_n^0(R_{(i)}) - \lambda(r_0)\}^2 W_{(i)} \\
&= n\mathbb{P}_n[\{(\hat{\lambda}_n(R) - \lambda(r_0))^2 - (\hat{\lambda}_n^0(R) - \lambda(r_0))^2\} W\, 1(R \in D_n)] \\
&= n\,(\mathbb{P}_n - P)\,[\{(\hat{\lambda}_n(R) - \lambda(r_0))^2 - (\hat{\lambda}_n^0(R) - \lambda(r_0))^2\} W\, 1(R \in D_n)] \\
&\qquad + n\,P[\{(\hat{\lambda}_n(R) - \lambda(r_0))^2 - (\hat{\lambda}_n^0(R) - \lambda(r_0))^2\} W\, 1(R \in D_n)] \\
&\equiv A_n + B_n.
\end{aligned}$$

Using arguments from empirical process theory in conjunction with [A] and [B], it is readily deduced, as in [4], that A_n is $o_P(1)$. Setting $I(r) := E(W|R = r)$, we have,

$$\begin{aligned}
\Delta SSE(r_0, \xi_0) &= n\,P[\{(\hat{\lambda}_n(R) - \lambda(r_0))^2 - (\hat{\lambda}_n^0(R) - \lambda(r_0))^2\} W\, 1(R \in D_n)] + o_P(1) \\
&= n \int_{D_n} \{(\hat{\lambda}_n(r) - \lambda(r_0))^2 - (\hat{\lambda}_n^0(r) - \lambda(r_0))^2\}\, I(r) f(r) dr + o_P(1) \\
&= \int_{\tilde{D}_n} \{X_n^2(z) - Y_n^2(z)\} I(r_0 + z\, n^{-1/3}) f(r_0 + z\, n^{-1/3}) dz + o_P(1) \\
&\xrightarrow{d} I(r_0)\, f(r_0) \int \{(g_{a,b}(z))^2 - (g_{a,b}^0(z))^2\} dz \stackrel{d}{=} I(r_0)\, f(r_0)\, a^2\, \mathbb{D} = \mathbb{D},
\end{aligned}$$

using the fact that $I(r_0) f(r_0) = a^{-2}$ (as can be verified directly) and Lemma 1 above. □

Remark: Setting $z = 0$ in [A], we obtain: $n^{1/3}\,(\hat{\lambda}_n(r_0) - \lambda(r_0)) \xrightarrow{d} g_{a,b}(0)$. Using Brownian scaling, which allows us to relate $g_{a,b}$ to $g_{1,1}$ [see Banerjee and Wellner (2001)], and the switching relationship on the process $X_{1,1}$ which shows that $g_{1,1}(0) \stackrel{d}{=} 2\mathbb{C}$ where $\mathbb{C}$ has Chernoff's distribution, we can deduce that $n^{1/3}\,(\hat{\lambda}_n(r_0) - \lambda(r_0)) \xrightarrow{d} (8a^2 b)^{1/3}\, \mathbb{C}$, thus verifying the claim made in Section 4 before the discussion on the residual sum of squares statistic. As can be noted, the calculation of confidence intervals using this result is problematic owing to the presence of several nuisance parameters.

The split point estimation procedure. Recall, that in this procedure, the goal is to find the stump model that best approximates λ; the main objective being to estimate the point of discontinuity (of the best–fitting stump) which defines the breakpoint. The approximation is defined in terms of the L_2 metric with respect to a measure induced by a weight function h. We make a slight change of notation from the body of the chapter. We denote the fitted stump by $\beta\, 1(r \in (d, \tau])$. The criterion to be optimized is given by:

$$\kappa(\beta, d) = \int_0^d \lambda^2(r)h(r)dr + \int_d^\tau [\lambda(r) - \beta]^2 h(r)dr,$$

with respect to β and d. It follows from simple algebra that the minimizer (β_0, d_0) also minimizes

$$\mathbb{M}(\beta, d) \equiv \beta^2\{H(\tau) - H(d)\} - 2\beta\{\Lambda(\tau) - \Lambda(d)\},$$

$\Lambda(u) = \int_0^u \lambda(s)h(s)ds$ and $H(u) = \int_0^u h(s)ds$. Setting the partial derivatives of $\mathbb{M}$ to 0 gives us the normal equations that characterize the parameters (β_0, d_0). We have

$$\beta_0 = 2\lambda(d_0) \text{ and } \beta_0 = \frac{\Lambda(\tau) - \Lambda(d_0)}{H(\tau) - H(d_0)}.$$

We point out that d_0 is of primary interest and was referred to as γ in Section 5. Consistent estimates $(\hat{\beta}_n, \hat{d}_n)$ of (β_0, d_0) are obtained by minimizing $\mathbb{M}_n(\beta, d)$ where

$$\mathbb{M}_n(\beta, d) = \beta^2\{H(\tau) - H(d)\} - 2\beta\{\hat{\Lambda}_n(\tau) - \hat{\Lambda}_n(d)\},$$

and $\hat{\Lambda}_n$ is a consistent estimate of Λ and is defined as $\hat{\Lambda}_n(u) = \int_0^u \hat{\lambda}_n(s)h(s)ds$; $\hat{\lambda}_n$ being the isotonic regression estimate of λ considered in the previous section.

The following assumptions are crucial to the subsequent development: [a] The parameter (β_0, d_0) exists and is unique. [b] The function λ is continuously differentiable in a neighborhood of 0, with $\lambda'(d_0) \neq 0$. [c] The estimate $(\hat{\beta}_n, \hat{d}_n)$ obtained by minimizing the criterion $\mathbb{M}_n(\beta, d)$ converges to the true value (β_0, d_0) at rate $n^{-1/3}$, i.e., $n^{1/3}(\hat{\beta}_n - \beta_0, \hat{d}_n - d_0)$ is $O_P(1)$. Here [c] is strongly suggested by published work in closely related models but we do not yet have a complete proof. We now derive the (unconditional) asymptotic distributions of our estimates. Define a normalized version of the process $\mathbb{M}_n$ as:

$$\mathbb{Q}_n(t_1, t_2) = n^{2/3}[\mathbb{M}_n(\beta_0 + t_1 n^{-1/3}, d_0 + t_2 n^{-1/3}) - \mathbb{M}_n(\beta_0, d_0)], t_1, t_2 \in \mathbb{R}.$$

The minimizer $(\hat{t}_{1,n}, \hat{t}_{2,n})$ of $\mathbb{Q}_n$ is precisely $n^{1/3}(\hat{\beta}_n - \beta_0, \hat{d}_n - d_0)$. We can decompose $\mathbb{Q}_n$ as $\mathbb{Q}_{n,1} + \mathbb{Q}_{n,2}$, where

$$\mathbb{Q}_{n,1}(t_1, t_2) = n^{2/3}[(\mathbb{M}_n - \mathbb{M})(\beta_0 + t_1 n^{-1/3}, d_0 + t_2 n^{-1/3}) - (\mathbb{M}_n - \mathbb{M})(\beta_0, d_0)],$$

and

$$\mathbb{Q}_{n,2}(t_1, t_2) = n^{2/3}[\mathbb{M}(\beta_0 + t_1 n^{-1/3}, d_0 + t_2 n^{-1/3}) - \mathbb{M}(\beta_0, d_0)].$$

Routine calculus yields that $\mathbb{Q}_{n,2}(t_1, t_2)$ converges uniformly on compact sets to $t^T V t/2$ where $t = (t_1, t_2)$ and V is the Hessian of $\mathbb{M}$ at the point (β_0, d_0), i.e.,

$$V = \begin{pmatrix} 2\{H(\tau) - H(d_0)\} & -2\Lambda'(d_0) \\ -2\Lambda'(d_0) & 4\lambda(d_0)\lambda'(d_0)h(d_0) \end{pmatrix}.$$

Now, $(\mathbb{M}_n - \mathbb{M})(\beta, d) = -2\beta[\{\hat{\Lambda}_n(\tau) - \Lambda(\tau)\} - \{\hat{\Lambda}_n(d) - \Lambda(d)\}]$ and $\mathbb{Q}_{n,1}(t_1, t_2)$ simplifies to

$$\begin{aligned} & 2\beta_0[n^{2/3}\{\hat{\Lambda}_n(d_0 + t_2 n^{-1/3}) - \hat{\Lambda}_n(d_0)\} - n^{2/3}\{\Lambda(d_0 + t_2 n^{-1/3}) - \Lambda(d_0)\}] \\ & +2t_1 n^{1/3}(\hat{\Lambda}_n - \Lambda)(d_0 + t_2 n^{-1/3}) - 2t_1 n^{1/3}\{\hat{\Lambda}_n(\tau) - \Lambda(\tau)\}. \end{aligned} \tag{6.3}$$

The term in the second line is $o_P(1)$ using the fact that (a) $\sup_{d \in \mathrm{nbhd}(d_0)} |\hat{\Lambda}_n(d) - \Lambda(d)|$ and (b) $|\hat{\Lambda}_n(\tau) - \Lambda(\tau)|$ are both $o_P(n^{-1/3})$, so that asymptotically only the first term, say I_n, contributes. Consider, $I_n(t_2)$ for $0 \leq t_2 \leq K$, for some $K > 0$. We have:

$$\begin{aligned} I_n(t_2) &= 2\beta_0 \left[n^{2/3} \int_{d_0}^{d_0+t_2 n^{-1/3}} \{\hat{\lambda}_n(u) - \lambda(u)\} h(u) du \right] \\ &= 2\beta_0 \left[n^{1/3} \int_0^{t_2} \{\hat{\lambda}_n(d_0 + vn^{-1/3}) - \lambda(d_0 + vn^{-1/3})\} h(d_0 + vn^{-1/3}) dv \right] \\ &= 2\beta_0 h(d_0) \left[\int_0^{t_2} X_n(v) dv - \int_0^{t_2} n^{1/3} \{\lambda(d_0 + vn^{-1/3}) - \lambda(d_0)\} dv \right] + o_P(1). \end{aligned}$$

The fact that $h(d_0 + vn^{-1/3})$ can be replaced by $h(d_0)$ at the expense of an $o_P(1)$ term is not difficult to justify. Since $X_n(z)$ converges in distribution to $g_{a,b}(z)$ on the set $[0, K]$ in the L_2 sense (see [A] in the previous section), it follows (using continuous mapping) that $\int_0^{t_2} X_n(v) dv$ converges in distribution to $\int_0^{t_2} g_{a,b}(v) dv = G_{a,b}(t_2) - G_{a,b}(0)$ under the topology of uniform convergence on $[0, K]$ (here, $G_{a,b}$ is the greatest convex minorant of $X_{a,b}$). The term $\int_0^{t_2} n^{1/3} \{\lambda(d_0 + vn^{-1/3}) - \lambda(d_0)\} dv$ converges to $(1/2)\lambda'(d_0) t_2^2$, uniformly on $[0, K]$. It follows that $I_n(t_2)$ converges in distribution, uniformly on $[0, K]$ to the process $G_{a,b}(t_2) - G_{a,b}(0) - (1/2)\lambda'(d_0) t_2^2$. This result can be readily strengthened to convergence on $[-K, K]$ by considering t_2's less than 0.

We conclude that the process $\mathbb{Q}_n(t_1, t_2)$ converges in distribution, under the topology of uniform convergence on compacts, to the process

$$\mathbb{Q}(t_1, t_2) = 2\beta_0 h(d_0)[G_{a,b}(t_2) - G_{a,b}(0) - bt_2^2] + \frac{1}{2} t^T V t.$$

The limit process is a.s. in $C(\mathbb{R}^2)$ with an a.s. unique minimizer. Conclude that $(\hat{t}_{1,n}, \hat{t}_{2,n})$ converges in distribution to $(\mathbf{t_1}, \mathbf{t_2})$, the almost surely unique minimizer of $\mathbb{Q}$. Note that $(\mathbf{t_1}, \mathbf{t_2})$ is also the minimizer of the process $\tilde{\mathbb{Q}}(t_1, t_2) := 2\beta_0 h(d_0) P_{a,b}(t_2) + t^T V t/2$ where $P_{a,b}(t_2) = G_{a,b}(t_2) - bt_2^2$.

The process $\tilde{\mathbb{Q}}(t_1, t_2)$ can be written out, in expanded form, as

$$\tilde{\mathbb{Q}}(t_1, t_2) = 4\lambda(d_0) h(d_0) \left[P_{a,b}(t_2) + \frac{1}{4} \frac{H(\tau) - H(d_0)}{\lambda(d_0) h(d_0)} t_1^2 + \frac{1}{2} \lambda'(d_0) t_2^2 - \frac{1}{2} t_1 t_2 \right].$$

It follows that:

$$(\mathbf{t_1}, \mathbf{t_2}) = \arg\min_{t_1,t_2} \left[P_{a,b}(t_2) + \frac{1}{4} \frac{H(\tau) - H(d_0)}{\lambda(d_0)h(d_0)} t_1^2 + \frac{1}{2}\lambda'(d_0)t_2^2 - \frac{1}{2}t_1 t_2 \right].$$

For fixed t_2, the process inside square brackets in the above display is minimized at a unique point, $\mathbf{t_1}(t_2)$, which is obtained by setting the partial derivative with respect to t_1 to 0 (this, indeed, provides a minimum, since for fixed t_2, the process is a deterministic quadratic polynomial with the co-efficient of the quadratic term being positive). We get:

$$\mathbf{t_1}(t_2) = \frac{\Lambda'(d_0)}{H(\tau) - H(d_0)} t_2.$$

Plugging this back for t_1 in the expression for the process inside square brackets and simplifying, we obtain: $\mathbf{t_2} = \arg\min_{t_2} [P_{a,b}(t_2) + c t_2^2]$, where $c = \frac{1}{2}\left\{\lambda'(d_0) - \frac{1}{2}\frac{\lambda(d_0)h(d_0)}{(H(\tau)-H(d_0))}\right\}$ and this is greater than 0, by virtue of the fact that V is p.d. $\square$

CHAPTER 7

Future Directions

The dissertation explores a variety of statistical methodologies – parametric finite-dimensional models with likelihood based inference, likelihood based methods in non-parametric scenarios, least squares estimators under shape constraints, and different bootstrap methods and their consistency in non-regular problems. Each of the diverse methodologies investigated in the chapters has its own direction; each raise new and challenging questions and opens up possibilities for further exploration and analysis. In the following we briefly discuss some of these problems that we plan to explore in greater depth in future.

Bootstrap in non-standard problems Chapters 4 and 5 are devoted to understanding the behavior of bootstrap methods for constructing pointwise confidence bands around the Grenander estimator and in Wicksell's problem, respectively. Both problems exhibit non-standard asymptotics and a non-standard rate of convergence, namely, $n^{1/3}$ and $\sqrt{\frac{n}{\log n}}$. But the usual bootstrap method (generating samples with replacement from the empirical distribution function) is inconsistent for the Grenander estimator, while it is consistent in Wicksell's problem. Both problems shed light on the behavior of bootstrap in shape restricted scenarios.

We claim that in the setting of cube-root asymptotics, the conventional bootstrap estimate of the sampling distribution of the statistic of interest does not have any limit, in probability. But, we have not been able to find a complete proof of this result. One step in the sequence of arguments depends on the fact that two specified functionals of Brownian motion with quadratic drifts are dependent, a fact that is easily verified through simulation, but we have not been able to prove it analytically. We plan to study the related functionals in more detail in future, and if possible, find a formal proof of the dependence.

We now present two related projects that have interesting theoretical and applied implications.

- **Bootstrap in convex function estimation:** The Grenander estimator is a prototypical example of a monotone non-parametric estimator. More complicated shape restrictions like convexity/concavity also arise commonly in applications, e.g., in econometrics, epidemiology and astronomy. Groeneboom, Jongbloed and Wellner (2001) studied the estimation and asymptotic theory of the nonparametric least squares estimators of convex regression and density functions. The asymptotic distribution theory relies on the existence of an "invelope function" for integrated two-sided Brownian motion plus a fourth power drift term (compare this with the quadratic drift that arises in monotone function estimation). The estimated convex function converges at $n^{2/5}$-rate to a multiple of the second derivative of the 'invelope function" at 0 – a complicated distribution with nuisance parameters. Thus, bootstrap methods arise naturally in the construction of confidence intervals using these shape-restricted estimators. But the non-standard asymptotics involved with the slower rate of convergence shed doubt on the consistency of bootstrap methods. We plan to investigate the performance of bootstrap methods in this set-up.

- **Smoothed-bootstrap in cube-root asymptotics:** Our work on different bootstrap procedures with the Grenander estimator has implications in general cube-root convergence problems. Smoothed bootstrap methods (like kernel smoothing) generally yield valid bootstrap methods in this situation, but with regression-type estimators – e.g., the maximum score estimator of Manski, least median of squares estimator, maximum likelihood estimator of failure time in current status model and so on – there are different ways of smoothing giving rise to different procedures. A natural question that arises is: "what is the minimal amount of smoothing required to make the bootstrap consistent?". This is an important question because with an increase in the dimension of data, the performance and implementation of smoothing methods are drastically affected. We want to address these issues with special attention to bootstrapping the maximum score estimator.

Likelihood ratio based methods under monotonicity constraints To find the distribution function of survival time in the mixed case interval censoring method we worked with a pseudo-likelihood (see Chapter 3). Working with the actual likelihood would yield more efficient estimators, but is difficult to study, as the estimators no longer have closed form solutions. Although a heuristic argument still suggests that the limit distribution of the likelihood ratio statistic would be nuisance parameter free, some of the intermediate steps are technically challenging and require deeper analysis. Keeping the technical issues aside, such a result on the convergence of the likelihood ratio statistic would have immense practical implications; it would readily yield a convenient method for constructing pointwise confidence bands around the survival function in the most general case of interval censoring with minimal assumptions.

Estimating the distribution function of survival time is a special case of estimating the mean function of a counting process, namely, a one-jump counting process. Another project that would be a direct fallout of our procedure is the extension of our results to general counting processes with covariates. Considering that counting processes arise naturally in demographic studies, clinical trials, etc., the extension of our method could prove important.

Abrupt change models for threshold detection In Chapter 6, we took a more applied approach, analyzing data on stars from Leo I galaxy with emphasis on detection of streaming motion. Due to time constraint and other considerations, we also avoided providing complete proofs of the main results.

In the present analysis we assumed that streaming, if it exists, must have its maximal effect along the position angle of the galaxy (known a priori), i.e., ω (see Chapter 6) was held fixed and known. Although this assumption can be argued for Leo I galaxy, in general ω may not be completely known. Methods for estimating ω from the data are called for, and we plan to explore this issue in future. We also plan to investigate more general models (beyond the "cosine model") to allow for flexible modeling of the variation in streaming motion along different angular positions. With just 328 member stars we refrained from using more complicated models. As more measurements on stars from Leo I and other galaxies expected to be taken in the near future, we plan to continue exploring related statistical methodologies. We expect to fill in the technical details in the forthcoming papers on this application. In Chapter 6.6 we discuss other possible extensions of our work on threshold models.

The signal plus noise model in high energy physics As discussed in Chapter 2, the signal plus noise model arises quite often in particle physics. An important question in this situation is to be able to test the presence of any signal. A related question, probably more intuitive, is to ask whether we have "seen a signal yet?". We are working on a bayesian approach to derive an upper bound for the probability that a signal event has been observed that is independent of prior distribution within broad limits.

BIBLIOGRAPHY

[1] Abrevaya, J. and Huang, J. (2005). On the Bootstrap of the Maximum Score Estimator. *Econometrica*, **73** 1175–1204.

[2] Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers,W. H. and Tukey, J. W. (1972). *Robust Estimates of Location.* Princeton Univ. Press, Princeton, N.J.

[3] Balabdaoui, F. and Wellner, J.A. (2007). Estimation of a k-monotone density: Limit distribution theory and the spline connection. *Ann. Statist.*, **35**, 2536–2564.

[4] Banerjee, M. (2000). Likelihood Ratio Inference in Regular and Nonregular Problems. Ph.D. dissertation, Univ. Washington.

[5] Banerjee, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.*, **35**, 931–956.

[6] Banerjee, M. and McKeague, I.W. (2007). Confidence sets for split points in decision trees. *Ann. Statist.*, **35**, 543–574.

[7] Banerjee, M. and Wellner, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.*, **29**, 1699–1731.

[8] Banerjee, M. and Wellner, J.A. (2005). Score statistics for current status data: comparison with likelihood ratio and Wald statistics. *Int. J. Biostat.*, **1**, Iss. 1, Article 3.

[9] Banerjee, M. and Wellner, J. A. (2005). Confidence intervals for current status data. *Scand. J. Statist.*, **32**, 405–424.

[10] Berk, R.H. and Jones, D.H. (1979). Goodness of fit statistics that dominate the Kolmogorov statistics. *Z. Wahr. verw. Geb.*, **47**, 47–59.

[11] Bickel, P. and Freedman, D. (1981). Some Asymptotic Theory for the Bootstrap. *Ann. Statis.*, **9**, 1196–1217.

[12] Breiman, L. (1968). *Probability.* Addison-Wesley Series in Statistics.

[13] Brunk, H. D. (1968). Estimation of Isotonic Regression. *Nonparametric Techniques in Statistical Inference*, 177–195. Cambridge Univ. Press.

[14] Byrd, G., Valtonen, M., McCall, M. and Innanen, K. (1994). Orbits of the Magellanic Clouds and Leo I in Local Group History, *Astrophysical Journal*, **107**, 2055.

[15] Chuang, C. and Lai, T.L. (1998). Resampling methods for confidence intervals in group sequential trials. *Biometrika*, **85**, 317–332.

[16] Chuang, C. and Lai, T.L. (2000). Hybrid resampling methods for confidence intervals. *Statist. Sinica*, **10**, 1–50.

[17] Chernoff, H.(1964). Estimation of the mode. *Ann. Inst. Statis. Math.*, **16**, 31–41.

[18] Chow, Y.S. and Teicher, H. (1988). *Probability: Independence, Interchangeability and Martingales*. Springer, New York.

[19] Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York-Akadémiai Kiadó, Budapest.

[20] De Gruttola, V. and Lagakos, S.W. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics*, **45**, 1–11.

[21] Delgado, M. A., Rodriguez-Poo, J. and Wolf, M. (2001). Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator. *Econ. Lett.*, **73**, 241–50.

[22] Dempster, A., Laird, N. and Rubin, D. (1977). Maximum Likelihood for incomplete data via the EM algorithm. *J. R. Stat. Soc. B*, **39**, 1–38.

[23] Devroye, L. (1987). *A course in Density Estimation*. Birkhäuser, Boston.

[24] Dinescu, D. I., Keeney, B. A., Majewski, S. R. and Girard, T. M. (2004). *Astronomical Journal*, **128**, 687–699.

[25] Eitel, K. and Zeitnitz, B. (1998). The search for neutrino oscillations $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$ with KARMEN. Available at arXiv/hepex/9809007.

[26] Feder, P.I. (1975). On asymptotic distribution theory in segmented regression problems – identified case. *Ann. Statist.*, **3**, 49–83.

[27] Feldman, G. and Cousins, R. (1998). Unified approach to the classical statistical analysis of small signal. *Phys. Rev. D*, **57**, 3873-3889.

[28] Feldman, G. (2000). Multiple Measurements and Parameters in the Unified Approach. Talk at the FermiLab Workshop on Confidence Limits.

[29] Fraser, D. A. S, and Reid, N. (1993). Third Order Asymptotic Models: Likelihood Functions Leading to Accurate Approximations for Distribution Functions. *Statist. Sinica*, **3**, 67–82.

[30] Grenander, U. (1956). On the theory of mortality measurement, Part II. *Skand. Akt.*, **39**, 125–153.

[31] Groeneboom, P. (1996). Lectures on Inverse Problems (École d'Été de Probabilités de Saint-Flour XXIV–1994). Ed. P. Bernard, 67–164. *Lecture Notes in Mathematics*, **1648**, New York: Springer Verlag.

[32] Groeneboom, P. and Jongbloed, G. (1995). Isotonic Estimation and Rates of Convergence in Wicksell's Problem. *Ann. Statis.*, **23**, 1518–1542.

[33] Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001). Estimation of a Convex Function: Characterizations and Asymptotic Theory. *Ann. Statis.*, **29**, 1653–1698.

[34] Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Boston: Birkhäuser.

[35] Groeneboom, P. and Wellner, J. A. (2001). Computing Chernoff's distribution. *J. Comp. Graph. Statist.*, **10**, 388–400.

[36] Hinkley, D. (1971). Inference in two-phase regression. *J. Amer. Statist. Assoc.*, **66**, 736–743.

[37] Huang, J. (2002). A Note on Estimating a Partly Linear Model under Monotonicity Constraints. *J. Statist. Plan. Inf.*, **107**, 343–351.

[38] Huang, Y. and Zhang, C. (1994). Estimating a monotone density from censored observations. *Ann. Statist.*, **22**, 1256–1274.

[39] Hušková, M. (1998). Estimators in the location model with gradual changes. *Comment. Math. Univ. Carolinae*, **39**, 147–157.

[40] Irwin, M. and Hatzidimitriou, D. (1995). Structural Parameters for the Galactic Dwarf Spheroidals. *Monthly Notices of the Royal Astronomical Society*, **277**, 1354–1378.

[41] Jewell, N. P. and Van der Laan, M. J. (1995). Generalizations of current status data with applications. *Lifetime Data Anal.*, **1**, 101–109.

[42] Jewell, N. P., Van der Laan, M. J. and Henneman, T. (2003). Nonparametric estimation from current status data with competing risks. *Biometrika*, **90**, 183–197.

[43] Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comp. Graph. Statist.*, **7**, 310–321.

[44] Kiefer, J. and Wolfowitz, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrsch. Verw. Gebiete.*, **34**, 73–85.

[45] Kim, J. and Pollard, D. (1990). Cube-root Asymptotics. *Ann. Statis.*, **18**, 191–219.

[46] Kleyna, J., Wilkinson, M.I., Evans, N.W., Gilmore, G., and Frayn, C. (2002). Dark matter in dwarf spheroidals - II. Observations and modelling of Draco. *Monthly Notices of the Royal Astronomical Society*, **330**, 792–806.

[47] Kómlos, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF.I. *Z. Wahrsch. Verw. Gebiete.*, **32**, 111-131.

[48] Kosorok, M. (2007). Bootstrapping the Grenander estimator. *Beyond Parametrics in Interdisciplinary Research: Festschrift in honour of Professor Pranab K. Sen.* IMS Lecture Notes and Monograph Series. Eds.: N. Balakrishnan, E. Pena and M. Silvapulle.

[49] Kulikov, V. N. and Lopuhaä, H. P. (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *Ann. Statist.*, **34**, 742–768.

[50] Lee, S. M. S. and Pun, M. C. (2006). On m out of n Bootstrapping for Nonstandard M-Estimation With Nuisance Parameters. *J. Amer. Statis. Assoc.*, **101**, 1185–1197.

[51] Léger, C. and MacGibbon, B. (2006). On the bootstrap in cube root asymptotics. *Can. J. of Statis.*, **34**, 29–44.

[52] Loader, C. (1999). Bandwidth selection: classical or plug-in ? *Ann. Statist.*, **27**, 415–438.

[53] Loeve, M. (1962). *Probability Theory*, 3rd Edition. New York: Van Nostrand.

[54] Mandelkern, M. (2002). Setting Confidence Intervals for Bounded Parameters. *Stat. Sci.*, **17**, 149–172.

[55] Mateo, M.L. (1998a). *Ann. Reviews of Astronomy and Astrophysics*, **36**, 435.

[56] Mateo, M., Olszewski, E.W., Pryor, C., Welch, D.L., and Fischer, P. (1993). *Astronomical Journal*, **105**, 510.

[57] Mateo, M., Olszewski, E.W., Vogt, S.S., and Keane, M.J. (1998b). *Astronomical Journal*, **116**, 231.

[58] Mateo, M., Olszewski, E.W., and Walker, M.G. (2008). *The Velocity Dispersion Profile of the Remote Dwarf Spheroidal Galaxy Leo I: A Tidal Hit and Run?.* *Astrophysical Journal*, **675**, 201.

[59] Meyer, M. and Woodroofe, M. (2004). Consistent maximum likelihood estimation of a unimodal density using shape restrictions. *Can. J. of Statis.*, **32**, 85–100.

[60] Muñoz, R.R., Majewski, S.R. and Johnston, K.V. (2008). Modelling the structure and synamics of dwarf spheroidal galaxies with dark matter and tides. *Astrophysical Journal*, **679**, 346.

[61] Oh, K. S., Lin, D. N. C., and Aarseth, S. J. (1995). *Astrophysical Journal*, **442**, 142–158.

[62] Owen, A, B. (1995). Nonparametric likelihood confidence bands for a distribution function. *J. Am. Statist. Assoc.*, **90**, 516–521.

[63] Pan, W. (2000). Smooth estimation of the survival function for interval censored data. *Statist. Med.*, **19**, 2611–2624.

[64] Piatek, S. and Pryor, C. (1995). *Astronomical Journal*, **109**, 1071–1085.

[65] Piatek, S., Pryor, C., Olszewski, E. W., Harris, H. C., Mateo, M., Minniti, D., Monet, D. G., Morrison, H., and Tinney, C. G. (2002). *Astronomical Journal*, **124**, 3198–3221.

[66] Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling.* Springer-Verlag, New York.

[67] Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York. Available at http://www.stat.yale.edu/∼pollard/1984book/pollard1984.pdf

[68] Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling.* New York: Springer Verlag.

[69] Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhȳa Ser. A*, **31**, 23–36.

[70] Reid, N. (2003). Asymptotics and the theory of inference. *Ann. Statis.*, **31**, 1695-2095.

[71] Reid, N. and Fraser, D. A. S. (2003). Likelihood Inference in the Presence of Nuisance Parameters. *PHYSTAT 2003, SLAC.*, 265–271 (arXiv: physics/0312079).

[72] Robertson,T., Wright, F. T. and Dykstra, R.L. (1988). *Order restricted statistical inference.* Wiley, New York.

[73] Robin, A. C., Reylé, C., Derrière, S. and Picaud, S. (2003). *Astronomy & Astrophysics*, **409**, 523–540.

[74] Roe, B. and Woodroofe, M. (1999). Improved probability method for estimation the signal in the presence of background. *Phys. Rev. D*, **60**, 053009.

[75] Roe, B. and Woodroofe, M. (2000). Setting confidence belts. *Physical Review D*, **63**, 013009.

[76] Rolke, W., López, A. and Conrad, J. (2005). Limits and Confidence Intervals in the Presence of Muisance Parameters. *Nucl.Instrum.Meth.*, A551, 493–503.

[77] Rousseeuw, P. J. (1984). Least median of squares regression. *J. Amer. Statis. Assoc.*, **79**, 871–880.

[78] Schick, A. and Yu, Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scand. J. Statist.*, **27**, 45–55.

[79] Sen, B., Banerjee, M. and Woodroofe, M. (2008). Inconsistency of Bootstrap: the Grenander estimator. Submitted.

[80] Sen, B., Walker, M.G. and Woodroofe, M. (2008). On the Unified method with nuisance parameters. To appear in *Statistica Sinica*.

[81] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York.

[82] Shiboski, S. (1998). Generalized ddditive models for current status data. *Lifetime Data Anal.*, **4**, 29–50.

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

[84] Singh, K. (1981). On asymptotic accuracy of Efron's bootstrap. *Ann. Statist.*, **9**, 1187–1195.

[85] Sohn, S.T. et al. (2007). *Astrophysical Journal*, **663**, 96.

[86] Song, S. (2004). Estimation with univariate "mixed case" interval censored data. *Statist. Sinica*, **14**, 269–82.

[87] Sun, J. and KALBFLEISCH, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica*, **5**, 279–290.

[88] Vandal, A. C., Gentleman, R. and LIU, X. (2005). Constrained estimation and likelihood intervals for censored data. *Can. J. Statist.*, **33**, 71–84.

[89] van der Vaart, A. W. and Wellner, J. A. (2000). *Weak Convergence and Empirical Processes*. Springer, New York.

[90] Walker, M.G., Mateo, M., Olszewski, E.W., Bernstein, R., Wang, X. and Woodroofe, M. (2006). Internal Kinematics of the Fornax Dwarf Spheroidal Galaxy. *The Astronomical Journal*, **131**, 2114–2139.

[91] Walker, M., Mateo, M., Olszewski, E., Pal, J., Sen, B. and Woodroofe, M. (2006). On Kinematic Substructure in the Sextans Dwarf Galaxy. *Astrophysical Journal*, **642**, L41–L44.

[92] Walker, M., Mateo, M., Olszewski, E., Sen, B. and Woodroofe, M. (2006). Clean Kinematic Samples in Dwarf Spheroidals: An Algorithm for Evaluating Membership and Estimating Distribution Parameters When Contamination is Present. (Submitted)

[93] Wang, X., Walker, M.G., Pal, J., Woodroofe, W. and Mateo, M. (2008a). Model independent estimates of dark matter distributions. To appear in *J. Amer. Statist. Assoc.*

[94] Wang, X., Woodroofe, M., Pal, J., Walker, M.G. and Mateo, M. (2008b). Non-parametric estimation of dark matter distributions. *Statistical Challenges in Modern Astronomy, IV* (Jagdish Babu and Eric Feigelson, eds).

[95] Wang, X. and Woodroofe, M. (2007). A Kiefer Wolfowitz Comparison Theorem for Wicksell's Problem, *Ann. Statis.*, **35**, 1559–1575.

[96] Wang, X., Woodroofe, M., Walker, M.G., Mateo, M. and Olszewski, E.W. (2005). Estimating Dark Matter Distributions. *The Astrophysical Journal*, **626**, 145–158.

[97] Wang, Y. (1994). The limit distribution of the concave majorant of an empirical distribution function. *Statist. Probab. Lett.* **20**, 81-84.

[98] Wellner, J.A. (2003). Gaussian white noise models: some results for monotone functions. In *Crossing Boundaries: Statistical Essays in Honor of Jack Hall*, Ed. J.E. Kolassa and D.Oakes, 87–104. IMS Lecture Notes-Monograph Series, **43**.

[99] Wellner, J. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.*, **28**, 779–814.

[100] Wicksell, S.D. (1925). The corpuscle problem. *Biometrika*, **17**, 84–99.

[101] Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimator of $f(0+)$ when f in non-increasing. *Statistica Sinica*, **3**, 505-515.