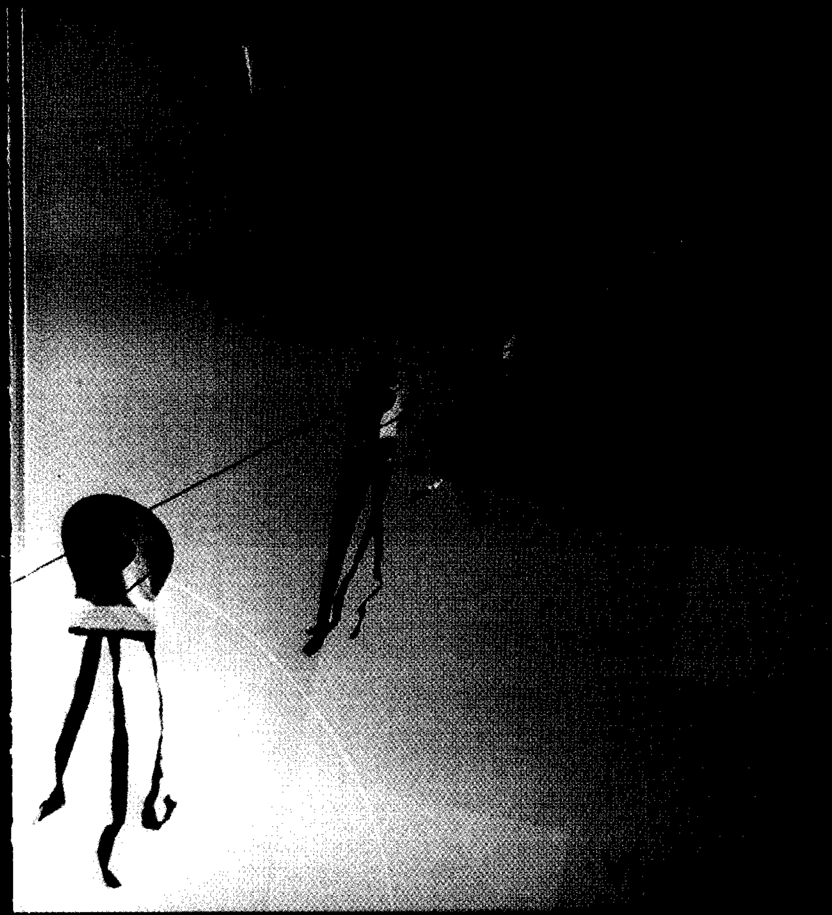
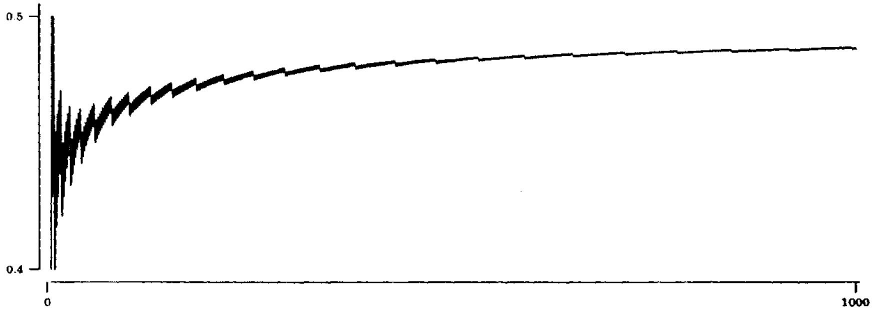


Direct and Indirect Use of Maximum Likelihood

5. Stel een vaas bevat 2 gouden en 3 zilveren ballen. Beschouw het volgende spel: voor elke gouden bal die u uit de vaas trekt, ontvangt u één euro, maar voor elke zilveren bal moet u één euro betalen. Een getrokken bal wordt niet in de vaas teruggelegd. Men kan op elk moment stoppen met spelen. De vraag is: is het verstandig om sowieso te beginnen met dit spel?

   Dit probleem is bekend en vrij eenvoudig, maar we kunnen dezelfde vraag stellen met k goudenen $n-k$ zilveren ballen. Verder kunnen we ons afvragen wat een optimale stoptijd zou zijn. Hoewel het wat lastig te beschrijven is in woorden, kan het algoritme gemakkelijk grafisch gerepresenteerd worden. Het volgende plaatje laat het minimale percentage gouden ballen zien waarvoor het winstgevend is om het spel te spelen.



   Een computer berekening laat zien dat $n = 5$ en $n = 10$ in zekere zin optimaal zijn (beiden hebben een 40% minimale ratio) en dat als $n \to \infty$, de minimale ratio naar $1/2$ gaat, waarschijnlijk met logaritmische snelheid. De optimale strategie is: speel totdat de verkregen ratio onder de grafiek valt.

6. Van de boeken van Gabriel Garcia Márquez worden "Honderd jaar eenzaamheid" en "De herfst van de patriarch" meestal gezien als zijn beste. Zij zijn echter vrij lang en somber en niet iedereen die deze boeken begint te lezen, zal ze ook uitlezen. Ik denk dat een ander, eerder boek van Márquez, "De kolonel krijgt geen post", meer aandacht en erkenning verdient. Het bevat alle opmerkelijke aspecten van Márquez' ideeën en stijl in een notendop en zou waarschijnlijk meer mensen helpen om zichzelf bekend te maken met de werken van deze auteur en om te wennen aan hun sfeer.

7. De Duivel die een rol speelt in Boelgakov's "De Meester en Margharita" is nogal atypisch. Door consistent slechte daden af te straffen en goede daden te belonen (zij het niet op een algemeen aanvaarde manier), speelt hij in feite de traditionele rol van God. Het citaat uit "Faustus", dat in het boek is opgenomen, is gebaseerd op het gebruikelijke geloof dat de duivel alleen een instrument in de handen van God is. Desondanks kan dit citaat niet het gedrag van deze Duivel van Boelgakov verklaren, omdat hij omschreven wordt als veel machtiger dan Christus. Deze interpretatie van de duivel is misschien uniek in de wereldliteratuur.

8. Ondanks mijn waardering voor de "The Lord of the Rings" trilogie, met name voor het eerste boek, denk ik dat de trilogie het stempel van de Koude Oorlog draagt. De wereld verdelend in "Noord", "West", "Zuid" en "Oost", suggereert het boek dat mensen uit het Westen half-elfen zijn, en daarom het beste van het beste (naast de elfen), mensen uit het Noorden zijn sterk en fatsoenlijk, in feite nakomelingen van Westerlingen. Mensen uit het Zuiden zijn gecorrumpeerd, maar nog niet zonder hoop, terwijl in het Oosten het Kwaad huist, dat volledig vernietigd moet worden en wel zo snel mogelijk.

9. Als men de kwaliteit van de werken van Tolstoj en Leskov vergelijkt, zou het een verrassing kunnen zijn dat Tolstoj veel populairder is. Een mogelijke reden van Tolstoj's populariteit is dat de lange novellen van Tolstoj de enige Russische tegenhangers zijn van de lange Europese novellen (Stendhal, Flaubert, Dickens) die zo populair waren in die tijd. Een andere reden zou kunnen zijn dat Leskov, een zeer traditioneel man, niet de revolutionaire veranderingen kon accepteren, noch de mensen die achter deze veranderingen stonden. Misschien dat zij om deze reden omgekeerd geen waardering voor zijn schrijven konden opbrengen.

DEZE STELLINGEN WORDEN VERDEDIGBAAR GEACHT EN ZIJN ALS ZODANIG GOEDGEKEURD DOOR DE PROMOTOR, PROF.DR. P.GROENEBOOM.

Propositions accompanying the thesis
Direct and Indirect Use of Maximum Likelihood
by V.N.Kulikov

1. Consider the supremum distance between the maximum likelihood estimator $\hat{F}_n$ of the current status censored distribution function and the distribution function F itself. It is proved in [1], §5.4, that, if the support S of the distribution F is compact and the infima of the densities of the censored distribution F and the censoring distribution G are positive on S, then

$$P\left\{\sup_{t\in S}\left|\hat{F}_n(t)-F(t)\right|>n^{-1/3}\log n\right\}\to 0,\ n\to\infty. \tag{2}$$

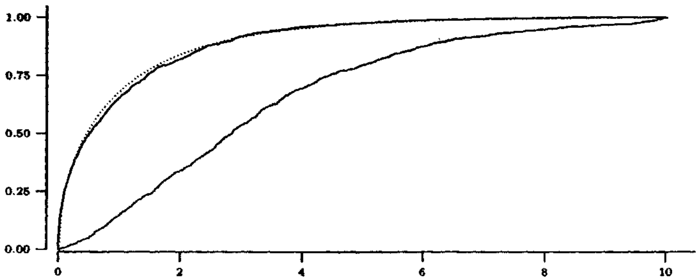
Using techniques, similar to those applied in §2.2 and §3.2 of this dissertation, the proof of this fact can be extended to a proof of the following statement:
as soon as

$$0<\inf_{t\in S}f(t)/g(t)\le\sup_{t\in S}f(t)/g(t)<\infty,$$

where S can be unbounded, (2) holds.

[1] Groeneboom, P. and Wellner, J.A. (1992) *Information bounds and nonparametric maximum likelihood estimation*, Birkhäuser Verlag.

2. In the case of testing against alternatives other than Lehmann (Proportional hazard) alternatives, the statement of Theorem 3.1.1 of this dissertation does not have to hold. For example, when considering location shift alternatives $F_1(t)=F_0(t-\theta)$, $\theta\in\mathbf{R}$, we would have to deal with score functions, depending on the density as well as the distribution function. A simulation study shows that in this case the distribution of twice the logarithm of the likelihood ratio is unlikely to be approximately χ_1^2, but it is much heavier tailed.



Here the two solid lines are empirical distribution functions of $2\log\mathbf{T}_n$ under the null hypothesis for a Lehmann alternative and for a location shift alternative, and the dotted line corresponds to the distribution function of χ_1^2.

3. The statement *(ii)* of Theorem 5.3.1 of this dissertation suggests that

$$\hat{f}_n(\hat{B}_{02}n^{-1/3})-n^{-1/3}EV/\hat{A}_{02},$$

where parameters A_{02} and B_{02} are estimates, based on $\hat{f}_n$, is a much better estimator of $f(0)$ than $\hat{f}_n(n^{-1/3})$. Unlike $\hat{f}_n(n^{-1/3})$, it is asymptotically unbiased and its limiting distribution will probably not depend on the underlying distribution. Nevertheless, computer simulations show that for $n=10000$ the estimation of $f'(0)$, playing a role in A_{02} and B_{02}, is still far too inaccurate, which leads to a much worse estimate of $f(0)$. Simulations indicate that $EV\approx-0.3115$.

4. It follows from Theorem 2.1 in [2] that

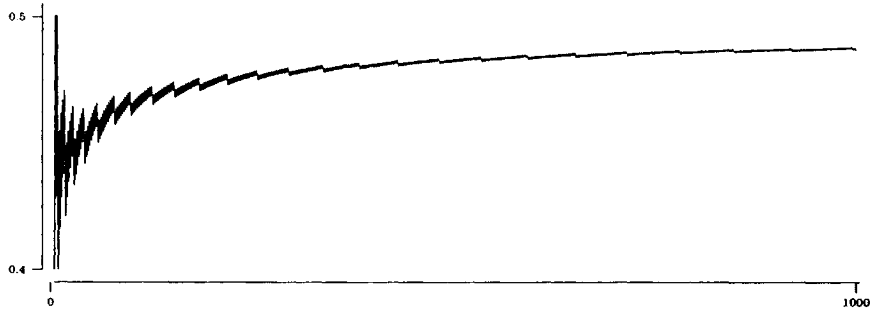
$$\sup_{t\in[f(1),f(0)]}|U_n(t)-g(t)|=O_p\left(n^{-1/3}(\log n)^{1/3}\right).$$

This result does not provide any information on the length of the first interval, where the Grenander estimator $\hat{f}_n$ is constant. We can prove that the length of this interval is $O_p\left(n^{-1}(\log n)^2\right)$. To obtain this result, one can follow a line of argument that is similar to arguments, used in Chapter 5 of this dissertation.

[2] GROENEBOOM, P., HOOGHIEMSTRA, G. AND LOPUHAÄ, H.P. (1999). Asymptotic normality of the L_1-error of the Grenander estimator, *Ann. Statist.* **27**, No.4, p.1316–1347.

5. Suppose a vase contains 2 golden and 3 silver balls. Consider the following game: for each golden ball taken from the vase one obtains one euro, whereas for a silver ball one must pay one euro. A ball taken from the vase is not put back. One can stop playing at any moment. The question is: does it make sense to play at all?

   This exercise is well-known and rather simple but we can ask the same question for k golden balls and $n-k$ silver balls. Furthermore, what would be an optimal stopping time? Although it is difficult to describe in words, the algorithm is easily represented graphically. The following picture shows the minimal percentage of golden balls for which it makes sense to play the game.



   A computer calculations show that $n = 5$ and $n = 10$ are the most favorable numbers of balls in the vase (both having 40% minimal ratio) and that, as $n \to \infty$, the minimal percentage tends to 1/2, probably at a logarithmic rate. Then the optimal strategy is: play until the obtained ratio crosses the graph.

6. Among the books of Gabriel Garcia Márquez, "One hundred years of solitude" and "The autumn of the patriarch" are usually considered as the most outstanding. Nevertheless they are quite long and grim and not everyone who starts reading them will finish even one book. On the other hand, I believe that another, earlier book of Márquez, "No one writes to the colonel", deserves more attention and recognition. It contains all remarkable aspects of Márquez' ideas and style in a nutshell and would probably help more people to familiarize themselves with the works of this author and to get used to their atmosphere.

7. The Devil acting in Bulgakov's "The Master and Margarita" is rather atypical. Consistently punishing bad actions and rewarding good actions (though not doing this in a commonly accepted way) he is, in fact, playing the traditional role of God. The quotation from "Faustus", accompanying the book, is based on the usual belief that after all the devil is just an instrument in the hands of God. Nevertheless, this quotation fails to explain the behavior of this Devil of Bulgakov, since he is described as much more powerful than Christ. This interpretation of the devil may be unique in world literature.

8. In spite of my admiration for the "The Lord of the Rings" trilogy, in particular for the first volume, I think that it bears the stamp of the Cold War. Dividing the population of the world into "North", "West", "South" and "East", the book suggests that people of the West are half-elfs, and therefore the best of the best (next to elfs), people of the North are strong and decent, in fact being offsprings of the Western folk. The people of the South are very spoilt, but still not beyond hope, whereas the East is the residence of evil, which must be completely destroyed as soon as possible.

9. If one compares the quality of the works of Tolstoy and Leskov, it may come as a surprise that Tolstoy is much more popular. A reason might be that the long novels of Tolstoy were the only Russian match to the long European novels (Stendhal, Flaubert, Dickens) so popular in his time. Another reason might be that Leskov, being very traditionalistic, could not accept revolutionary changes and people supporting these changes. Perhaps for this reason they, in turn, could not have appreciation for his writing.

THESE PROPOSITIONS ARE CONSIDERED DEFENDABLE AND AS SUCH HAVE BEEN APPROVED BY THE SUPERVISOR, PROF.DR. P.GROENEBOOM.

TR 3995 S

STELLINGEN BIJ HET PROEFSCHRIFT
DIRECT AND INDIRECT USE OF MAXIMUM LIKELIHOOD
DOOR V.N.KULIKOV

1. Beschouw de supremum afstand tussen de maximum likelihood schatter $\hat{F}_n$ van de current status gecensureerde verdelingsfunctie en de verdelingsfunctie F zelf. In [1] wordt bewezen dat als de drager S van de verdeling F compact is en de infima van de dichtheden van de gecensureerde verdeling F en de censureringsverdeling G positief zijn op S, dan

$$P\left\{\sup_{t\in S}\left|\hat{F}_n(t)-F(t)\right| > n^{-1/3}\log n\right\} \to 0,\ n\to\infty. \quad (1)$$

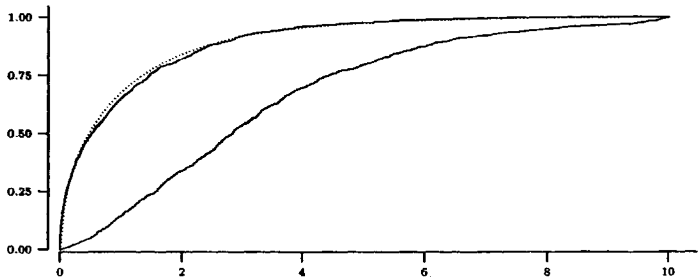
Gebruik makend van technieken analoog aan die toegepast in §2.2 en §2.3 van dit proefschrift, kan het bewijs van dit feit worden uitgebreid naar een bewijs van de volgende stelling:
Als geldt

$$0 < \inf_{t\in S} f(t)/g(t) \le \sup_{t\in S} f(t)/g(t) < \infty,$$

waarbij S onbegrensd kan zijn, dan geldt (1).

[1] GROENEBOOM, P. EN WELLNER, J.A. (1992) *Information bounds and nonparametric maximum likelihood estimation*, Birkhäuser Verlag.

2. In het geval van het toetsen tegen alternatieven anders dan de Lehmann (Proportional hazard) alternatieven, hoeft Stelling 3.1.1 van dit proefschrift niet te gelden. Bijvoorbeeld, als we translatie alternatieven beschouwen, krijgen we te maken met score functies die afhangen van zowel de dichtheid als de verdelingsfunctie. Een simulatie studie laat zien dat in deze situatie de verdeling van twee maal de logaritme van de likelihood ratio niet goed de verdeling van χ_1^2 benadert, maar een veel dikkere staart heeft.



De twee doorgetrokken lijnen zijn de empirische verdelingsfunctie van $2\log \mathbf{T}_n$ onder de nulhypothese voor een Lehmann alternatief en voor een translatie alternatief, en de stippellijn komt overeen met de verdelingsfunctie van χ_1^2.

3. Stelling 5.3.1 van dit proefschrift suggereert dat

$$\hat{f}_n(\hat{B}_{02}n^{-1/3}) - n^{-1/3}EV/\hat{A}_{02},$$

waar de parameters A_{02} en B_{02} schattingen zijn, gebaseerd op $\hat{f}_n$, een veel betere schatter is van $f(0)$ dan $\hat{f}_n(n^{-1/3})$. Anders dan $\hat{f}_n(n^{-1/3})$ is deze schatter consistent en wij vermoeden dat bewezen kan worden dat de limietverdeling niet afhangt van de onderliggende verdeling. Desalniettemin laten computer simulaties zien dat voor $n = 10000$ de schatter voor $f'(0)$, die een rol speelt in A_{02} en B_{02}, zo onnauwkeurig is dat dit leidt tot een veel slechtere schatting van $f(0)$. Deze simulaties suggereren dat $EV \approx -0.3115$.

4. Het volgt uit Theorem 2.1 in [2] dat

$$\sup_{t\in[f(1),f(0)]} |U_n(t) - g(t)| = O_p\left(n^{-1/3}(\log n)^{1/3}\right).$$

Dit resultaat geeft geen informatie over de lengte van het eerste interval waar de Grenander schatter $\hat{f}_n$ constant is. Wij kunnen bewijzen dat de lengte van dit interval $O_p\left(n^{-1}(\log n)^2\right)$ is. Om dit te bewijzen kan men gelijksoortige argumenten gebruiken als in Hoofdstuk 5 van dit proefschrift.

[2] GROENEBOOM, P., HOOGHIEMSTRA, G. EN LOPUHAÄ, H.P. (1999). Asymptotic normality of the L_1-error of the Grenander estimator, *Ann. Statist.* 27, No.4, p.1316–1347.

Direct and Indirect Use of Maximum Likelihood

Vladimir Nikolaevich Kulikov

Direct and Indirect Use of Maximum Likelihood

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.dr.ir. J.T.Fokkema,
voorzitter van het College voor Promoties,
in openbaar te verdedigen op dinsdag 21 januari 2003 om 13.30 uur

door VLADIMIR NIKOLAEVICH KULIKOV

Master of Science in Mathematics,
Moscow Lomonosov State University in Moscow, Russia

geboren te Moskou, Rusland

Dit proefschrift is goedgekeurd door de promotor

Prof.dr. P.Groeneboom

Toegevoegd promotor: Dr. H.P.Lopuhaä.

Samenstelling promotiecomissie:

Rector Magnificus, voorzitter
Prof.dr. P.Groeneboom, Technische Universiteit Delft en Vrije Universiteit Amsterdam, promotor
Dr. H.P.Lopuhaä, Technische Universiteit Delft, toegevoegd promotor
Prof.dr. R.D.Gill, Universiteit Utrecht
Prof.dr. F.M.Dekking, Technische Universiteit Delft
Prof.dr. S.A. van de Geer, Universiteit Leiden
Prof.dr. J.A.Wellner, University of Washington, Seattle
Dr.ir. G.Jongbloed, Vrije Universiteit Amsterdam

THOMAS STIELTJES INSTITUTE FOR MATHEMATICS

The research in this thesis has been carried under the auspices of the Thomas Stieltjes Institute for Mathematics, at the University of Technology in Delft.

ISBN 90-6734-105-3

Cover photo by Olena Kyryliuk

Printed by [OPTIMA] Grafische Communicatie, Rotterdam

Acknowledgments

Now when I am looking back to almost four years of my Ph.D. study in Delft, I feel I need to thank so many people who have dedicated their time and attention to me. Although I don't have enough place to thank them all here, let me acknowledge some of them in chronological order. Those who are nevertheless not mentioned here, let me assure them that my gratitude and warm feelings to them are simply remained on another branch of the chronological tree.

The first person I met in the Netherlands, after the unknown Schiphol worker returned me my hat from the puddle under the plane, where it was thrown by the Dutch wind, was Piet. I must thank him for all he has done for me, for our discussions about mathematics, where I really must say - discussions, since at the beginning Piet had to look for grains between all chaffs I produced and push me in the right direction. Furthermore I want to thank him for teaching me to find a right direction myself where I hope he has at least partly succeeded. Also Piet always remained an unending source of books and publications that I must read and, I have to admit, I still have to read more carefully. Speaking about books, our talks about books and music have always been a pleasure. But the first thing that comes to mind when I think for what I must thank Piet is his care about my settlement in Delft at the very beginning and the first part of it: carrying part of my luggage at Schiphol on the very first day.

Another person I met this day and who played a most important role in my Ph.D. study, is Rik. Rik always attended both mathematical and comprehensive aspects of my writings and I hope I have learned from him not only mathematics, and even not only reaching the same results in shorter and more elegant ways, but also explaining what has crystallized in my mind in a way that does not remind the reader of porridge as it was some three years ago. I must also thank Rik for teaching me to teach. I hope that someday I will be able to follow his style of explaining things to the audience more fully.

Next I want to acknowledge my roommate Eric, together with whom we invented some ideas later realized in this thesis. Furthermore I want to thank him for involving me in his ideas and explaining to me things that have unfortunately escaped my mind before. Eric was a perfect roommate all this time and the atmosphere he was a part of, greatly helped me in my studies.

Thinking of people from the workgroup who have participated in my Ph.D. study in general, I certainly must think of them as the audience of the Probability and Statistics seminar. Talking first about things that concern me mostly, I want to thank Michel, Gerard, Cor, Ludolf, Hans, Svetlana and Andre for not laughing during my first talk in Delft, not complaining at the second and for not falling asleep at the third one. I must also thank them for their very interesting and informative talks they have given in the last years and for inviting such interesting guests to the seminar. Certainly they have also taken care of my study and education in one way or another which I simply have no place to describe here properly.

I also want to thank my parents who always supported me in all possible ways and who have laid the foundation of my personality and my education. I think that this dissertation is their merit more than mine. My girlfriend Ksenia has also supported me greatly and her patience, understanding and love have stimulated me to continue my work properly.

I cannot stand the expression "last but not least I thank" and therefore at last I wish to thank the one who cannot even thought of as least. Thanks God we all have done it!

Contents

Chapter 1

Introduction

1.1 Scope of the dissertation

In this dissertation we consider estimators that satisfy certain order restrictions. These estimators can arise in different ways. For example, if we want to estimate a distribution function, there are natural monotonicity constraints. But also in the context of density estimation, as well as estimation of regression curves, monotonicity constraints can arise naturally. For these situations certain isotonic estimators have been in use for considerable time. Often these estimators can be seen as maximum likelihood estimators in a semi-parametric setting. Although conceptually these estimators have a great appeal and are easy to formulate, their distributional properties are usually of a very complicated nature.

PRAKASA RAO (1969) obtained the earliest result on the asymptotic pointwise behavior of the Grenander estimator, which is the maximum likelihood estimator of a decreasing density. One immediately striking feature of this result is that the rate of convergence is of the same order as the rate of convergence of histogram estimators, and that the asymptotic distribution is not normal. It took much longer to develop distributional theory for global measures of performance for this estimator. The first distributional result for a global measure of deviation was the convergence of the L_1-distance in GROENEBOOM (1985). Moreover, only a sketch of proof was given there and a rigorous proof was given fairly recently in GROENEBOOM, HOOGHIEMSTRA AND LOPUHAÄ (1999). A similar result in the regression setting has been obtained even more recently by DUROT (2000).

In this dissertation we will extend the result for the L_1-distance to the L_k-distance. Immediately, new interesting features emerge. In our study of the behavior of the L_k-distance it becomes clear that there is a kind of transition point at $k = 2.5$ in the sense that for the value of $k > 2.5$ we need a modification of the L_k-distance in order to obtain the analogous limiting result. The major reason for this is the behavior of the estimator near zero. As a spin-off of our study of the behavior near zero we also develop a $n^{1/3}$-consistent estimator for $f(0+)$.

A pointwise result also exists in the case of estimating a concave distribution function, studied in KIEFER AND WOLFOWITZ (1976). They showed that the convergence rate of the supremum distance between the empirical distribution function F_n and its concave majorant $\hat{F}_n$ is faster than $\sqrt{n}$, whereas pointwise convergence of the difference $\hat{F}_n - F_n$ was obtained by WANG (1994). We extend this result to process convergence of this difference and in addition obtain a distributional result for the L_k-distance between $\hat{F}_n$ and F_n. Recently, similar results for the pointwise behavior and the L_k-distance were obtained in a regression setting by DUROT

AND TOCQUET (2002).

The maximum likelihood estimator (MLE) of the distribution function in a situation where only interval censored observations are available, turns out to have some properties which are similar to those of the Grenander estimator. For example, the rate of pointwise convergence in the simplest case of interval censoring, case I (also called current status), is the same as that of the pointwise convergence of the Grenander estimator (see GROENEBOOM AND WELLNER (1992)). On the other hand, certain smooth functionals of the distribution function can be estimated at the usual $n^{1/2}$-rate by estimators that are based on the MLE. These estimators also have the usual asymptotic normal behavior (see GESKUS AND GROENEBOOM (1996), GESKUS AND GROENEBOOM (1997) and GESKUS AND GROENEBOOM (1999)). These results suggested that in the situation of interval censored data, a $n^{1/2}$-convergent test based on the MLE can be developed. One of the first results of this type was obtained in HUANG (1996), where the Cox model for the interval censored data is considered and where the finite-dimensional parameter is estimated at $n^{1/2}$-rate (test for the finite-dimensional parameter). In this dissertation we develop several tests of different nature in the context of two-samples testing. One type of test that we develop is a score-type test, which is easy to apply and has certain optimality properties in a restricted setting. The other type of test is the likelihood ratio test, which is harder to implement, but has a somewhat wider range of applicability. We also prove efficiency of both tests.

In the remainder of this introductory chapter we will discuss the above matters in more detail and also provide some background information. Full proofs of the results are given in Chapters 2 to 6.

1.2 Maximum likelihood and likelihood ratio

Consider a class $\mathcal{P}$ of distributions and for each $P \in \mathcal{P}$, let f_P be the density with respect to some dominating measure. The maximum likelihood estimator of the distribution P_0 based on the sample $\{X_i\}$ from P_0, is the element $\hat{P}_n$ of the class that maximizes the likelihood

$$L_n(P) = \prod_{i=1}^{n} f_P(X_i),$$

when it is well-defined. Sometimes it is more convenient to use the loglikelihood

$$l_n(P) = \log L_n(P),$$

because logarithms transform products into sums, and maximizing $L_n(P)$ is equivalent to maximizing $l_n(P)$. Therefore,

$$\hat{P}_n = \operatorname*{argmax}_{P\in\mathcal{P}} L_n(P) = \operatorname*{argmax}_{P\in\mathcal{P}} l_n(P).$$

Often the MLE has the fastest possible rate of convergence and a normal limiting distribution. This fact is explained by the theory of smooth functionals, which we will shortly discuss in Section 1.7. However, the functional of P that is of interest may not always be a smooth functional. For instance, the value of the distribution function of interval censored observations at a fixed point, is a non-smooth functional. For this reason, instead of converging at rate $n^{1/2}$

and being asymptotically normal, the MLE converges at rate $n^{1/3}$ and has a limiting distribution that is the same as the last point where a two-sided Brownian motion with parabolic drift attains its maximum (see Section 1.3).

The convergence in Hellinger distance is also of interest. If f_1 and f_2 are two densities with respect to a measure μ, then the Hellinger distance between the densities is defined as

$$h(f_1, f_2) = \left(\frac{1}{2}\int \left(\sqrt{f_1} - \sqrt{f_2}\right)^2 d\mu\right)^{1/2}.$$

In VAN DE GEER (2000) it is shown that, under certain entropy conditions, the maximum likelihood estimator is Hellinger consistent:

$$h(\hat{f}_n, f_0) \to 0, \quad \text{almost surely.}$$

The notion of entropy is discussed in Section 1.8. This result applies to many settings such as estimation of smooth densities and estimation of a monotone density. The last case will be discussed in Section 1.6, whereas in Section 1.3 another example of a Hellinger consistent MLE is given. The rate of convergence of the maximum likelihood estimator is also well-studied. According to VAN DE GEER (2000),

$$P\left\{h(\hat{f}_n, f_0) > \delta\right\} \le C \exp\left\{-\frac{n\delta^2}{C^2}\right\},$$

for any $\delta > \delta_n$, where (δ_n) is a sequence satisfying $\sqrt{n}\delta_n^2 \ge C\Upsilon(\delta_n)$, for a universal constant C and a suitably chosen function Υ. In the case of estimating smooth densities, which have at least m derivatives, the rate will not be slower than $n^{-m/(2m+1)}$. The case of estimating a monotone density is discussed in the Section 1.6 and Section 1.3 is devoted to the model for current status data.

In the context of testing hypotheses a likelihood based test statistic is the likelihood ratio

$$\mathbf{T}_n = \frac{\sup_{P\in H_1} L_n(P)}{\sup_{P\in H_0} L_n(P)},$$

where H_0 represents the null hypothesis and H_1 the alternative. The better the underlying distribution fits the null hypothesis, the smaller is the likelihood ratio $\mathbf{T}_n$. This immediately implies that the critical region corresponding to testing at level α is of the form $[t_\alpha, \infty)$, where t_α is the $(1-\alpha)$-quantile of the distribution of $\mathbf{T}_n$.

Often it is possible to show that $2\log\mathbf{T}_n$ has a limiting χ_1^2 distribution. To illustrate this, first consider the simple case of testing the null hypothesis that parameter θ of an exponential distribution is equal to one, against a two-sided alternative. In this case f_θ is sufficiently smooth with respect to θ. Writing $l_n(\theta)$ instead of $l_n(P_\theta)$ and using that $l_n'(\hat{\theta}_n) = 0$, by definition of $\hat{\theta}_n$, one can expand $2\log\mathbf{T}_n$ in terms of $\sqrt{n}(\hat{\theta}_n - 1)$ as follows,

$$2\log\mathbf{T}_n = 2l_n(\hat{\theta}_n) - 2l_n(1) \sim l_n''(\hat{\theta}_n)\left(\hat{\theta}_n - 1\right)^2 \sim \frac{1}{n}l_n''(1)\left(\sqrt{n}(\hat{\theta}_n - 1)\right)^2.$$

Next, a limiting χ_1^2 distribution is obtained if $\sqrt{n}(\hat{\theta}_n - 1)$ can be shown to be asymptotically normal. For this, use that

$$0 = l_n'(\hat{\theta}_n) \sim l_n'(1) + l_n''(1)(\hat{\theta}_n - 1),$$

which means that the MLE $\hat{\theta}_n$ behaves as

$$\sqrt{n}(\hat{\theta}_n - 1) \sim -\frac{\frac{1}{\sqrt{n}} l_n'(1)}{\frac{1}{n} l_n''(1)}.$$

Since $E_\theta \left[\frac{\partial}{\partial\theta} \log f_\theta(X)\right] = 0$, asymptotic normality of $\sqrt{n}(\hat{\theta}_n - 1)$ now follows from an application of the central limit theorem and the fact that

$$\frac{1}{n} l_n''(1) \to E_\theta \left[\frac{\partial^2}{\partial\theta^2} \log f_\theta(X)\right].$$

The χ_1^2 distribution is obtained as limiting distribution by using that

$$E_\theta \left(\frac{\partial}{\partial\theta} \log f_\theta(X)\right)^2 = -E_\theta \left[\frac{\partial^2}{\partial\theta^2} \log f_\theta(X)\right].$$

A brief sketch of a similar proof for models with finite dimensional parameters can be found in Murphy and van der Vaart (1997). Other examples of more complicated setups in which $2 \log \mathbf{T}_n$ has a limiting χ_1^2 distribution are given in Murphy and van der Vaart (1997), Murphy and van der Vaart (2000) and Huang (1996). Among these is the Cox regression model for current status data, which is the most interesting for our purposes. This model is discussed in Section 1.4. The general scheme of proof in Murphy and van der Vaart (1997) is the same as we have described above for the one-dimensional model. First $2 \log \mathbf{T}_n$ is expanded in terms of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ and then one uses the following result obtained in Huang (1996) (writing Pg for $\int g \, dP$)

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = I_0^{-1} \sqrt{n} P_n l^* + o_p(1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_0^{-1}),$$

where P_n denotes the empirical measure, l^* is the efficient score function, and $I_0 = P_0(l^*)^2$ is the Fisher information. The main difference between the Cox model and the model we will work with (see Section 1.4) is that we have to deal with logarithmic terms, which introduces additional difficulties near zero, whereas in the Cox model one deals with exponential terms having negative powers.

The model we will consider in Chapter 3, also described in the Section 1.3, is a semi-parametric model. Therefore, it contains a one-dimensional parameter θ, which will be shown to be a smooth functional. This makes it possible to construct a test for $\theta = 0$ against contiguous alternatives $\theta_n = \theta_0/\sqrt{n}$. Furthermore, the theory of smooth functionals allows us to investigate whether the test is asymptotically (locally) most efficient. The test based on the score statistic studied in Chapter 2 is most efficient, as well as the test based on the likelihood ratio statistic. Another advantage of the likelihood ratio statistic is the following. Usually its limiting distribution does not depend on initial parameters. Other statistics, like the score statistic, although easier to calculate (methods of simulation will be discussed in Section 1.5), may depend on initial parameters in a quite complicated way.

1.3 Current status data

Sometimes one cannot observe the variable of interest directly. Instead, another variable (or several other variables) is observed, that carries some indirect information on the distribution of the variable of interest. One such example is interval censoring. As an illustration, consider the situation where one is interested in the distribution of the time of onset X of a certain disease. One screens a patient at censoring time T and is only able to observe whether the patient already has the disease, i.e., $X \leq T$, or not. Hence, instead of the variable of interest X, we observe another random variable T, independent of X, and an indicator $\Delta = 1_{\{X \leq T\}}$. The space of all possible realizations of (X, T) is usually called the hidden space, and the space of all possible realizations of (T, Δ) is called the observation space. Interval censoring can then be considered as a mapping from the hidden space into the observation space (with loss of information). In the model described above, we observe the status of a patient only at a single time point. In a more complicated setup one would observe the status at more than one time point. We will confine ourselves to the simplest model, often called interval censoring: case I, also known as the current status model.

The maximum likelihood estimator of the distribution function of the random variable X is defined as follows. Since the distribution G of the censoring variable T is not an object of interest here, we consider the density of the pair (T, Δ) with respect to the measure $\mu = G \times$ (counting measure on $\{0, 1\}$). The density $f_{(T,\Delta)}$ is given by

$$f_{(T,\Delta)}(t, \delta) = \delta F(t) + (1 - \delta)(1 - F(t)),$$

where F is the distribution function of X. The likelihood of a sample $\{(T_i, \Delta_i)\}_{i=1}^n$ is then given by

$$L_n(F) = \prod_{i=1}^{n} \{\Delta_i F(T_i) + (1 - \Delta_i)(1 - F(T_i))\},$$

and the maximizer $\hat{F}_n$ of $L_n(F)$ in the class of all (possibly degenerate) distribution functions is well defined. Usually, the likelihood is maximized over the set of piecewise-constant distribution functions with jumps only at points $T_1, \ldots, T_n$. In this case it is also unique (see e.g. GESKUS AND GROENEBOOM (1999)). Later we will consider the loglikelihood

$$l_n(F) = \sum_{i=1}^{n} \{\Delta_i \log F(T_i) + (1 - \Delta_i) \log(1 - F(T_i))\}.$$

An important property of the MLE, which can be found e.g. in GROENEBOOM AND WELLNER (1992) is that for any $a_n(t) \geq 0$, that is piecewise constant on the same intervals as $\hat{F}_n(t)$, it holds that

$$\int_{\hat{F}_n(t) \in (0,1),\ t \leq t_0} a_n(t) \left(\frac{\delta}{\hat{F}_n(t)} - \frac{1 - \delta}{1 - \hat{F}_n(t)} \right) dP_n(t, \delta) \leq 0,$$

$$\int_{\hat{F}_n(t) \in (0,1)} a_n(t) \left(\frac{\delta}{\hat{F}_n(t)} - \frac{1 - \delta}{1 - \hat{F}_n(t)} \right) dP_n(t, \delta) = 0.$$

Both relations play an important role in finding the MLE, as well as in the study of its asymptotic behavior and in fact they have a close connection to the theory of smooth functionals (see Section 1.7). The second equation is often called the score rule.

Consistency and the rate of convergence of $\hat{F}_n$ have been established. GROENEBOOM AND WELLNER (1992) have shown that, if the distribution functions F_0 and G of X and T have densities, and the measure P_{F_0} is dominated by P_G, then the MLE is strongly consistent:

$$P\left\{\lim_{n\to\infty}\sup_{t\in\mathbb{R}}\left|\hat{F}_n(t)-F_0(t)\right|=0\right\}=1.$$

Another convenient measure of the accuracy of the estimator is the Hellinger distance. In VAN DE GEER (2000) it is shown that the Hellinger distance between the underlying density $f_{(T,\Delta)}$ and its maximum likelihood estimator $\hat{f}_{(T,\Delta)}$ converges to zero at rate $n^{1/3}$:

$$h(\hat{f}_{(T,\Delta)}, f_{(T,\Delta)}) = O_p(n^{-1/3}).$$

By using boundedness of the distribution function, it follows that

$$\sqrt{\int\left(\hat{F}_n(t)-F_0(t)\right)^2 dG(t)} = O_p(n^{-1/3}).$$

Extra conditions on the distribution G may provide a more informative bound. The maximum likelihood estimator in the smaller class of possible distributions can have a faster rate of convergence. For example, in the case of compact support, the MLE of a concave density of interval censored observations can be estimated at rate $n^{2/5}$ (for more details see VAN DE GEER (2000) and DÜMBGEN, FREITAG AND JONGBLOED (2002)). Further results on the supremum distance can be found in GROENEBOOM AND WELLNER (1992). For instance, if F_0 has bounded support, on which densities f_0 and g are bounded and bounded away from zero, the MLE converges at a rate faster than $n^{1/3}/\log n$:

$$P\left\{\sup_{t\in\operatorname{supp}F_0}\left|\hat{F}_n(t)-F_0(t)\right| > n^{-1/3}\log n\right\}\to 0,\ n\to\infty.$$

Some distribution theory has also been obtained for the MLE. GROENEBOOM AND WELLNER (1992) have shown that for a point t_0 in the interior of the supports of F_0 and G, it holds that

$$\left\{\frac{g(t_0)}{4f_0(t_0)F_0(t_0)(1-F_0(t_0))}\right\}^{1/3} n^{1/3}\left(\hat{F}_n(t_0)-F_0(t_0)\right)\xrightarrow{\mathcal{D}} Z,$$

where Z denotes the last time where standard two-sided Brownian motion minus the parabola $y(t)=t^2$ reaches its maximum. The rate $n^{1/3}$ is characteristic for non-smooth functionals, such as $F_0(t_0)$. Functionals, such as $\int t\,dF_0(t)=\int(1-F_0(t))\,dt$ are smooth, and therefore the corresponding estimator $\int(1-\hat{F}_n(t))\,dt$ converges at rate $\sqrt{n}$. Moreover, under certain conditions, the estimator is asymptotically normal:

$$\sqrt{n}\int_{\operatorname{supp}F_0}(\hat{F}_n(t)-F_0(t))\,dt\xrightarrow{\mathcal{D}}\mathcal{N}(0,\sigma^2),$$

with limiting variance σ^2 depending on F_0 and G (see e.g. GROENEBOOM AND WELLNER (1992)).

1.4 Current status and two samples testing

The ordinary two-samples test is concerned with testing the null hypothesis whether two independent samples (possibly of different size) are from the same distribution. For further notational convenience, instead of considering two samples $\xi_1, \ldots, \xi_{n_1}$ and $\eta_1, \ldots, \eta_{n_2}$ separately, we will join them and consider a single sample of pairs

$$(X_1, Z_1), \ldots, (X_n, Z_n), \quad n = n_1 + n_2,$$

where $\{X_1, \ldots, X_n\} = \{\xi_1, \ldots, \xi_{n_1}, \eta_1, \ldots, \eta_{n_2}\}$ and Z_i is an indicator, specifying to which original sample X_i belongs. Hence, $Z_i = 1_{X_i \in \{\eta_1, \ldots, \eta_{n_2}\}}$ is a Bernoulli random variable with parameter $p = n_2/(n_1 + n_2)$. We consider the two-sample problem in the context of interval censoring, case I (see Section 1.3). Suppose that X_i is censored by some random variable T_i. In practice, the possibly different nature of the two original samples often implies a different distribution for the censoring variable. Therefore T_i is assumed only to be conditionally independent of X_i given $Z_i = z$. For $z = 0, 1$, define

$$\begin{aligned} F_z(x) &= P\{X_i \le x \mid Z_i = z\}, \\ G_z(t) &= P\{T_i \le t \mid Z_i = z\}. \end{aligned}$$

Then the null hypothesis we need to test is $H_0 : F_0 = F_1$.

Next, we will make some assumptions about the alternative H_1, since for the general alternative $H_1 : F_0 \ne F_1$, the construction of a test and the investigation of its distribution becomes much more complicated. Interval censoring has a lot of applications in biostatistics and medicine, where, as well as in many other fields of science, the *proportional hazard model* plays an important role. An important concept in this model is the cumulative hazard function corresponding to distribution function F, which is defined by $\Lambda_F(x) = -\log(1 - F(x))$. Its derivative represents the instantaneous death rate of individuals who have survived up to time x. The alternative hypothesis in the two-sample problem for the proportional hazard model is: $\Lambda_{F_1} = \theta \Lambda_{F_0}$, for some $\theta > 0$, i.e., the death rate corresponding to F_1 is proportional to the death rate corresponding to F_0. In view of this, we will consider alternatives of the form

$$1 - F_1 = (1 - F_0)^{1+\theta}, \quad \text{for } \theta > -1.$$

Another possibility are Lehmann alternatives

$$F_1 = F_0^{1+\theta}, \quad \text{for } \theta > -1.$$

In both cases we test the null hypothesis $H_0 : \theta = 0$ against $H_1 : \theta \ne 0$.

Note that we only need to consider Lehmann alternatives, since the proportional hazard alternatives can be related to Lehmann alternatives as follows. Suppose that $1 - F_1 = (1 - F_0)^{1+\theta}$. Define new random variables $\tilde{X}_i = -X_i$, $\tilde{T}_i = -T_i$, $\tilde{\Delta}_i = 1_{\{\tilde{X}_i \le \tilde{T}_i\}}$ and $\tilde{Z}_i = Z_i$. Then $\tilde{\Delta}_i = 1 - \Delta_i$ and

$$\begin{aligned} \tilde{F}_1(x) &= P\{\tilde{X}_i \le x \mid \tilde{Z}_i = 1\} = P\{X_i \ge -x \mid Z_i = 1\} \\ &= 1 - F_1(-x) = (1 - F_0(-x))^{1+\theta} \\ &= P\{X_i \ge -x \mid Z_i = 0\}^{1+\theta} = P\{\tilde{X}_i \le x \mid \tilde{Z}_i = 0\}^{1+\theta} = \tilde{F}_0(x)^{1+\theta}. \end{aligned}$$

This means that we are back with Lehmann alternatives, and all results obtained for Lehmann alternatives can immediately be transferred to proportional hazard alternatives.

To apply likelihood based methods we consider the density $f_{(T,\Delta,Z)}$ of the triple (T, Δ, Z) under a Lehmann alternative. Similar to Section 1.3 we will consider the density with respect to a measure μ, such that $f_{(T,\Delta,Z)}$ will not depend on G_0 and G_1. Let μ_0 and μ_1 be defined on $\mathbb{R} \times \{0, 1\}$ by

$$\mu_0 = G_0 \times (\text{counting measure on } \{0,1\}),$$
$$\mu_1 = G_1 \times (\text{counting measure on } \{0,1\}),$$

and for $A \subseteq \mathbb{R} \times \{0,1\}^2$ define

$$\mu(A) = \mu_0\{(t,\delta) : (t,\delta,0) \in A\} + \mu_1\{(t,\delta) : (t,\delta,1) \in A\}.$$

Then the density $f_{(T,\Delta,Z)}$ with respect to μ is given by

$$f_{(T,\Delta,Z)}(t,\delta,z) = \delta F(t)^{1+\theta z} + (1-\delta)\left(1 - F(t)^{1+\theta z}\right),$$

where F is the distribution function of X, and the loglikelihood is given by

$$l_n(\theta, F) = \sum_{i=1}^{n} \left\{\Delta_i(1+\theta Z_i)\log F(T_i) + (1-\Delta_i)\log(1 - F(T_i)^{1+\theta Z_i})\right\}.$$

Calculation of the maximum likelihood estimator $(\hat{\theta}_n, \hat{F}_n) = \operatorname{argmax} l_n(\theta, F)$, where the maximum is taken over all pairs (θ, F), with $\theta > -1$ and F a piecewise-constant (possibly degenerate) distribution function with jumps only at points $T_1, \ldots, T_n$, will be discussed in Section 1.5. This estimator is always well-defined and is studied in Chapter 3. Its basic properties turn out to be similar to those of the MLE for F_0 in the interval censoring model for a single sample.

In Chapters 2 and 3, two different tests of the null hypothesis against Lehmann alternatives are derived. In Chapter 2 we investigate a score test, which is obtained as follows. Define

$$\hat{F}_n^0 = \operatorname*{argmax}_F l_n(0, F)$$

as the MLE under the null hypothesis. Next, consider the loglikelihood $l_n(\theta, \hat{F}_n^0)$ as function of θ, and define $\hat{\theta}_n^0 = \operatorname{argmax} l_n(\theta, \hat{F}_n^0)$. This means that $\partial l_n(\theta, \hat{F}_n^0)/\partial\theta = 0$ at $\theta = \hat{\theta}_n^0$. Hence, under the null hypothesis, it seems reasonable that $\hat{\theta}_n^0 \approx 0$, so that $\partial l_n(\theta, \hat{F}_n^0)/\partial\theta$ is small at $\theta = 0$. This leads to the following statistic, which we will call the score statistic (the term 'score' refers to the fact that it has a close relation to the theory of smooth functionals, see Section 1.7):

$$S_n = \frac{1}{\sqrt{n}}\frac{\partial l_n(0, \hat{F}_n^0)}{\partial\theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i \hat{F}_n^0(T_i)\log \hat{F}_n^0(T_i)\left(\frac{\Delta_i}{\hat{F}_n^0(T_i)} - \frac{1-\Delta_i}{1-\hat{F}_n^0(T_i)}\right).$$

In Chapter 2, asymptotic normality of S_n is established, and it is shown that the test based on S_n is asymptotically most efficient (see Section 1.7 for the definition of efficiency). The proof of this result uses appropriate versions of the Chaining lemma and stochastic equicontinuity (see Section 1.8) and methods similar to those from Geskus and Groeneboom (1999). The score test is easy to implement, since it only depends on the observations and the MLE $\hat{F}_n^0$ of a single sample, which can be computed fast. Its major disadvantage is that the limiting distribution depends on the underlying distributions F_0, G_0 and G_1. Estimation of these quantities leads to additional computations and makes the test less accurate.

A traditional statistic, whose limiting distribution is independent of the initial parameters, is the likelihood ratio statistic $\mathbf{T}_n = L_n(\hat{\theta}_n, \hat{F}_n)/L_n(0, \hat{F}_n^0)$. In Chapter 3 we show that $2\log\mathbf{T}_n$ has a limiting χ_1^2 distribution. The general scheme of the proof is similar as that for the one-dimensional model considered in Section 1.2. The main difference is, that in the case of Lehmann alternatives, the loglikelihood $l_n(\hat{\theta}_n, \hat{F}_n)$ must be expanded with respect to both arguments. To overcome this difficulty, analogously to MURPHY AND VAN DER VAART (1997), we use optimality of the MLE together with the following bounds

$$2l_n(\hat{\theta}_n, \hat{F}_n^0(t) - \hat{\theta}_n h_n^0(t)) - 2l_n(0, \hat{F}_n^0) \leq 2\log\mathbf{T}_n \leq 2l_n(\hat{\theta}_n, \hat{F}_n) - 2l_n(0, \hat{F}_n(t) + \hat{\theta}_n h_n(t)),$$

where $h_n(t)$ and $h_n^0(t)$ are piecewise constant approximations of the least favorable subdirection. Both bounds can be expanded as functions of $\hat{\theta}_n$, and can be shown to have a limiting χ_1^2 distribution.

As one may expect, the likelihood ratio statistic is asymptotically most efficient as well. Computer simulations provided in Chapters 2 and 3 suggest that the distributions of the score statistic and likelihood ratio statistic are close to the limiting distribution already for sample size $n = 50$. The likelihood ratio statistic, which has a limiting distribution independent of the underlying distributions, is preferable for practical applications with samples of sizes $n < 1000$. For the larger n, faster computation of the score statistic may become crucial.

1.5 Implementation

In this section we discuss the calculation of the non-parametric MLE $\hat{F}_n^0$ and the semiparametric MLE $(\hat{\theta}_n, \hat{F}_n)$ introduced in Section 1.4. One of the methods to compute $\hat{F}_n^0$ is the EM algorithm. That is: first take the set $t_1, \ldots, t_n, t_{n+1}$, where all t's are ordered and t_{n+1} is an arbitrary point greater than all t's in the sample. Furthermore, take an arbitrary starting distribution $F^{(0)}$ with positive masses at all points t_i:

$$p_i^{(0)} = P_{F_0}\{X = t_i\},$$

for example the discrete uniform distribution $p_i^{(0)} = 1/(n+1)$.

An "Expectation step" consists of calculating the conditional expectation of the log likelihood

$$E^{(0)}\left\{\sum_{i=1}^{n}\log f(X_i)\middle|\ \delta_1, \ldots, \delta_n, t_1, \ldots, t_n\right\},$$

where $f(x) = P_F\{X = x\}$, and $E^{(0)}$ is an expectation under the probability measure $P_{F^{(0)}}$. In the next "Maximization step" we maximize the above expectation over all discrete distributions with probability density f with respect to the counting measure on the set $\{t_1, \ldots, t_{i+1}\}$. This yields new probability masses $p_i^{(1)}$. These steps are repeated until certain solution criteria (for example the Fenchel duality conditions, see below) are satisfied. The number of iterations needed to reach the solution with an accuracy of, say, two decimals, will increase with the sample size n (more on this and on why the method works can be found in GROENEBOOM AND WELLNER (1992), Section 3.1).

The EM algorithm can be applied in various settings but for the case of interval censoring, case I, there exist one step algorithm letting to find MLE $\hat{F}_n^0$ exactly. It is based on Theorem 1.5.1 of ROBERTSON, WRIGHT AND DYKSTRA (1988), according to which maximization of

$$\Phi(x) = \sum_{i=1}^{n} (\delta_i \log x_i + (1-\delta_i)\log(1-x_i)), \ 0 \le x_1 \le \ldots \le x_n \le 1$$

is equivalent to minimization of

$$\phi(x) = \sum_{i=1}^{n} (x_i - \delta_i)^2,$$

where δ_i are ordered with respect to the t-component of the pair.

The minimizing $\hat{x}$ can be represented as $\hat{x} = \sum_{i=1}^{n} \hat{\alpha}_i z^{(i)}$, where $z^{(i)} = \sum_{j=i}^{n} e_j$ and $e_j = (0,\ldots,0,1,0,\ldots,0)$ (1 at position j). Furthermore, $\hat{x}$ is a minimizer if and only if

$$\sum_{j=i}^{n} \hat{x}_j \ge \sum_{j=i}^{n} \delta_j, \ i = 1,\ldots,n,$$

$$\sum_{j=i}^{n} \hat{x}_j = \sum_{j=i}^{n} \delta_j, \ \text{if } \hat{\alpha}_i > 0 \text{ or } i = 1.$$

Let $P_0 = (0,0)$ and $P_i = (i, \sum_{j=i}^{n} \delta_j)$, $i = 1,\ldots,n$. Moreover, let C be the pointwise largest convex function on $[0,1]$, lying below (or touching) the points of P_i. The set of points P_i is called cusum (cumulative sum) diagram and the function C - the (greatest) convex minorant of this cusum diagram. Then $\hat{x}_i$ is the left derivative of C at i. Executing the one step convex minorant algorithm takes approximately Cn^2 operations.

Finding the semi-parametric MLE $(\hat{\theta}_n, \hat{F}_n)$ consists of maximizing

$$\Psi(\theta, x) = \sum_{i=1}^{n} \left\{\delta_i(1+\theta z_i)\log x_i + (1-\delta_i)\log\left(1 - x_i^{1+\theta z_i}\right)\right\}, \ 0 \le x_1 \le \ldots \le x_n \le 1,$$

which is neither convex nor concave in x for all θ, $-1 < \theta < \infty$. The ICM (iterative convex minorant) algorithm, applicable for finding non-parametric MLE in the case of interval censoring, case II, can be applied here as well. The method is based on approximation of a non-convex loglikelihood by a parabola (Taylor expansion up to the second term). At the first step we choose one of the interior points x^0 of the cone $0 \le x_1 \le \ldots \le x_n \le 1$ and a diagonal matrix D with positive diagonal elements. Approximate $\Psi(x, \theta)$ by

$$\tilde{\Psi}(\theta, x) = \Psi(\theta, x^0) + \langle x - x^0, \nabla\Psi(x^0)\rangle - \frac{1}{2}(x - x^0)^T D (x - x^0),$$

and take x^1 to be a minimizer of the convex function

$$x \to \left(x - x^0 - D^{-1}\nabla\Psi\left(x^0\right)\right)^T D \left(x - x^0 - D^{-1}\nabla\Psi\left(x^0\right)\right)$$

over the convex cone. Then x_i^1 is the left derivative of the convex minorant of the cusum diagram

$$P_i = \left(\sum_{j=1}^{i} d_j, \sum_{j=1}^{i} \left\{d_j x_j^0 - \frac{\partial}{\partial x_j^0}\Psi(x^0)\right\}\right), \ i = 1,2,\ldots,n.$$

Usually one can get arbitrarily close to the solution (as far as numerical accuracy allows) by repeating these iterations.

The minimizer of the convex approximation at each step does not necessary lie in the convex cone but to avoid this difficulty one can use Jongbloed's modification of the ICM algorithm. This means that instead of the minimizer of the approximation we choose the point of the segment between each x^i and the minimizer of the quadratic form such that Ψ decreases sufficiently. For more details on the ICM algorithm and Jongbloed's modification, see GROENEBOOM (1996) and JONGBLOED (1998).

The above method allows one to find the maximizer $\hat{F}_n(\theta)$ of $L_n(\theta, F)$ for any $\theta > -1$, and, by further maximizing over the grid, to find the absolute maximizer $(\hat{\theta}_n, \hat{F}_n)$. This is possibly not the fastest method, since it takes about Cn^4 operations to find the estimator, which differs considerably from Cn^2 for finding $\hat{F}_n^0$.

1.6 Grenander estimator

The Grenander estimator is another name of the maximum likelihood estimator of a monotone density. In this class of densities the MLE is well-defined and as GRENANDER (1956) has shown, it is equal to the left derivative of concave majorant of the empirical distribution function F_n (the smallest concave function greater than or equal to F_n).

PRAKASA RAO (1969) has shown that if $f'(t) < 0$, for some point t in the interior of the support of f, then the Grenander estimator $\hat{f}_n(t)$ converges at rate $n^{1/3}$:

$$n^{1/3}\left|\frac{1}{2}f(t)f'(t)\right|^{-1/3}\left(\hat{f}_n(t) - f(t)\right) \xrightarrow{\mathcal{D}} 2Z,$$

where Z is distributed as the location of maximum of the process $\{W(u) - u^2,\ u \in \mathbb{R}\}$ and W is standard two-sided Brownian motion on $\mathbb{R}$ originating from zero. Later a more elegant proof of this result was given in GROENEBOOM (1985). In this paper also the limit behavior of the L_1-error of the Grenander estimator was formulated for the first time. It was proved rigorously in GROENEBOOM, HOOGHIEMSTRA AND LOPUHAÄ (1999). Assume that f is a strictly decreasing density with compact support. Let its first derivative be bounded away from zero and let the second derivative be bounded. Then

$$n^{1/6}\left\{n^{1/3}\int_0^1 \left|\hat{f}_n(t) - f(t)\right| dt - \mu\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2).$$

The Grenander estimator and related objects are difficult to study since they depend (in principle) on the whole empirical distribution. In GROENEBOOM, HOOGHIEMSTRA AND LOPUHAÄ (1999) the natural interpretation of the integral as an area leads to considering the L_1-error of the inverse process $\{U_n(a) : a \in [f(1), f(0)]\}$ as an estimator of $g(t) = f^{-1}(t)$, where $U_n(a)$ is defined as the last time when the process $F_n(t) - at$ attains its maximum:

$$U_n(a) = \sup\{t \in [0,1] : F_n(t) - at \text{ is maximal}\}.$$

Nevertheless the process $U_n(a)$ still depends on the whole process $\{F_n(x) : x \in [0,1]\}$, although there are reasons to believe that asymptotically there will only be local dependence in shrinking neighborhoods. The latter phenomenon is perhaps easiest to see by approximating the process

F_n locally by Brownian motion. To this end, the so-called Hungarian embedding (see KOMLOS, MAJOR AND TUSNADY (1975)) is used. It establishes that a version of the Brownian Bridge can be defined on the same probability space as F_n in such a way that for any x

$$P\left\{ \sup_{y\in\text{support}(F)} \left|F_n(y) - F(y) - n^{-1/2}B_n(F(y))\right| > \frac{1}{n}(C\log n + x)\right\} < Ke^{-\lambda x},$$

where C, K and λ are some positive absolute constants.

In Chapter 6 we will generalize the result for L_1-error of the Grenander estimator to the L_k-error. It will be shown that $k = 2.5$ is a kind of transition point in the sense that for $1 \le k < 2.5$

$$n^{1/6}\left\{n^{1/3}\left(\int_0^1 \left|\hat{f}_n(x) - f(x)\right|^k dx\right)^{1/k} - \mu_k\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_k^2),$$

whereas for $k > 2.5$ there is no converge at all. To get convergence for $k \ge 2.5$ we need to consider a modified L_k-error by taking the integral over the interval $[n^{-\epsilon}, 1 - n^{-\epsilon}]$. Then for any $1/6 < \epsilon < \frac{k-1}{3(k-2)}$,

$$n^{1/6}\left\{n^{1/3}\left(\int_{n^{-\epsilon}}^{1-n^{-\epsilon}} \left|\hat{f}_n(x) - f(x)\right|^k dx\right)^{1/k} - \mu_k\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_k^2).$$

The main differences with the L_1 case are that, firstly, the interpretation of the integral as an area will be lost, and, secondly, the inconsistency of the Grenander estimator at the boundary of the support of the underlying density is going to play a role for $k \ge 2.5$.

Pointwise convergence of the Grenander estimator takes place only at interior points of the interval of support. The limiting distribution of $\hat{f}_n(0+)$ is given in WOODROOFE AND SUN (1993):

$$\hat{f}_n(0+) \xrightarrow{\mathcal{D}} f(0+) \sup_{1\le k<\infty} \frac{k}{\Gamma_k},$$

where Γ_n are the partial sums of i.i.d. standard exponential random values. Similar kind of inconsistency occurs at the other end of interval of support (if the support is bounded).

There are different ways to construct an estimator that has the same properties as the Grenander estimator in the interior of the interval of support and also estimates $f(0)$ consistently. In WOODROOFE AND SUN (1993) a penalized maximum likelihood estimator is proposed. The penalized maximum likelihood estimator in this case is the maximizer $\hat{f}_n(\alpha, x)$ of the penalized likelihood

$$l_\alpha(f) = \sum_{i=1}^n \log f(x_i) - n\alpha f(0+),$$

where f is continuous non-increasing density with $f(t) = 0$ if $t < 0$ and where $\alpha > 0$ is a smoothing parameter. The exact formula for the solution of this maximization problem can be found in SUN AND WOODROOFE (1996), where consistency of the penalized estimator $\hat{f}_n(\alpha, 0+)$ is proved as $\alpha \to 0$ and $n\alpha \to \infty$. For $x > 0$, the difference between the penalized MLE $\hat{f}_n(\alpha, x)$ and the Grenander estimator is negligible.

In Chapter 5 we introduce another approach to this problem. Extending the result for the pointwise convergence of the Grenander estimator to values $t_n = n^{-\alpha}$ tending to zero, and using an inverse functions technique proposed in GROENEBOOM (1985), we prove convergence in distribution of $n^{\beta}(\hat{f}_n(n^{-\alpha}) - f(n^{-\alpha}))$ to

$$\sqrt{f(0) \operatorname*{argmax}_{t\in[0,\infty)}\{W(t)-t^2\}} \qquad \text{for } \beta = \frac{1-\alpha}{2} \text{ if } 1/3 < \alpha < 1,$$
$$(2f(0)|f'(0)|)^{1/3} \operatorname*{argmax}_{t\in(-\infty,\infty)}\{W(t)-t^2\} \text{ for } \beta = 1/3 \quad \text{if } 0 < \alpha < 1/3.$$

For $\alpha = 1/3$ we prove convergence in distribution to a non-degenerate random variable Z for $\beta = 1/3$. Based on this result we introduce an estimator $\hat{f}_n(n^{-1/3})$ of $f(0)$ which has a smaller mean squared error than that of $\hat{f}(\alpha_n, 0)$ proposed in WOODROOFE AND SUN (1993).

We also introduce a procedure for testing whether the underlying distribution function is concave. The testing procedure we propose is to compare two estimators of the underlying distribution function: the empirical distribution function F_n and the integrated Grenander estimator. The latter represents the concave majorant $\hat{F}_n$ of F_n. If the underlying distribution function is concave, one would expect that the process $\hat{F}_n(t) - F_n(t)$ is asymptotically small in a certain sense. For convenience of notation, we define the process $\{A_n(t) : t \in \operatorname{supp} F\}$, where

$$A_n(t) = n^{2/3}(\hat{F}_n(t) - F(t)).$$

The process $n^{-2/3}A_n$ was studied by KIEFER AND WOLFOWITZ (1976), where the upper bound of the supremum distance was established. They showed that if the underlying decreasing density f has a compact support and continuous derivative, then for sufficiently large n

$$P\left\{\sup_{x\in\operatorname{supp} F}\left|\hat{F}_n(x) - F_n(x)\right| > n^{-2/3}(\log n)^{5/6}\right\} < 2n^{-2}.$$

The pointwise convergence of A_n was obtained in WANG (1994). Under the condition that the density is strictly decreasing and differentiable at t_0:

$$\left|\frac{f'(t_0)}{2f(t_0)^2}\right|^{1/3} A_n(t_0) \xrightarrow{\mathcal{D}} [CM_{\mathbb{R}}\{Z\}](0),$$

where $CM_{\mathbb{R}}\{Z\}$ denotes the concave majorant of the process $Z(t) = W(t) - t^2$, which is two-sided Brownian motion on $\mathbb{R}$ originating from zero with a parabolic drift. A similar result was established in DUROT AND TOCQUET (2002) for monotone regression.

In Chapter 4 we study further properties of the process $\{A_n(t) : t \in \operatorname{supp} F\}$. First we extend the pointwise convergence, established in WANG (1994), to the local convergence in distribution of the whole process. The rescaled process

$$\left\{c_1(t_0)A_n\left(t_0 + n^{-1/3}c_2(t_0)s\right) : s \in \mathbb{R}\right\},$$

where outside of the interval of support the difference is taken to be zero, converges in distribution to the process

$$\{[CM_{\mathbb{R}}Z](s) - Z(s) : s \in \mathbb{R}\}$$

in the space $D(-\infty, \infty)$ of cadlag functions. For criteria of convergence in distribution in $D(-\infty, \infty)$, see e.g. Section 1.9.

Although this result may be of interest in its own right, it does not provide a test statistic. A possible test statistic is provided by the integral $\int |\hat{F}_n(t) - F_n(t)|^k dt$. In Chapter 4 we prove that if the underlying decreasing density f has compact support and a continuous second derivative on the interval of support, than for any $k \geq 1$ and any continuous function g

$$n^{1/6}\left\{\int A_n(t)^k g(t)dt - \mu_k\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_k^2).$$

The proof is based on techniques developed in GROENEBOOM, HOOGHIEMSTRA AND LOPUHAÄ (1999). A similar result was obtained by DUROT AND TOCQUET (2002) for monotone regression.

1.7 Smooth functionals and efficiency of testing

We now briefly discuss the theory of smooth functionals and its application to proving the asymptotic efficiency of the test. Suppose the unknown distribution P_0 is contained in a family of distributions $\mathcal{P}$, which is dominated by a σ-finite measure μ. We want to estimate some function of this underlying distribution P_0 (that may be P_0 as well), which we denote as $\Psi(P_0)$. The family $\mathcal{P}$ need not be one-dimensional, but we nevertheless consider one-dimensional submodels $\{P_s,\ s \in [0,\delta)\} \subseteq \mathcal{P}$, with P_0 the same as before. Such a submodel is called Hellinger differentiable if it is smooth in the following sense:

$$\int \left[\frac{1}{s}\left(\sqrt{p_s} - \sqrt{p_0}\right) - \frac{1}{2}a\sqrt{p_0}\right]^2 d\mu \to 0,\ \ s \downarrow 0,\ \text{ for some } a \in L_2(P_0).$$

Here p_s denotes the density P_s with respect to the dominating measure μ. Hellinger differentiability can be seen as an L_2-version of the pointwise differentiability of $\log p_s(x)$ at $s = 0$, since

$$\lim_{s\downarrow 0} \frac{\sqrt{p_s} - \sqrt{p_0}}{s} = \frac{1}{2}\sqrt{p_0}\left(\frac{\partial}{\partial s}\log p_s\bigg|_{s=0}\right).$$

The function a in the last definition is called a tangent or score function. The linear space T_{P_0} of all tangents a corresponding to all Hellinger differentiable submodels of $\mathcal{P}$, having P_0 the endpoint is called the tangent space. As proved in, e.g., GROENEBOOM (1996), one has $T_{P_0} \subseteq L_2^0(P_0)$.

The functional $\Psi : \mathcal{P} \to \mathbb{R}$ is pathwise differentiable at P_0 if for each Hellinger differentiable one-dimensional submodel $\{P_s,\ s \in [0,\delta)\}$ of $\mathcal{P}$ with tangent $a \in T_{P_0}$,

$$\lim_{s\downarrow 0} \frac{1}{s}\left(\Psi(P_s) - \Psi(P_0)\right) = \Psi'_{P_0}(a),$$

where $\Psi'_{P_0} : T_{P_0} \to \mathbb{R}$ is continuous and linear. Each functional of this kind can be represented as an inner product $\langle a, b\rangle_{P_0} = \int ab\,dP_0$. Consider an extension to a continuous linear functional $\Psi'_{P_0} : L_2(P_0) \to \mathbb{R}$. This extension (which we denote by $\bar{\Psi}'_{P_0}$) is not unique, but by the Riesz representation theorem for each extension we can find a unique $\theta_{P_0} \in L_2(P_0)$ such that

$$\bar{\Psi}'_{P_0}(h) = \langle \theta_{P_0}, h\rangle_{P_0} \ \text{ for all } h \in L_2(P_0).$$

Such a θ_{P_0} is called a *gradient.* Note that the orthogonal projection $\mathbf{P}_T(\theta_{P_0})$ onto $\bar{T}$, the closure of the subspace T, is uniquely defined and also represents a gradient. This projection (denoted by $\tilde{k} = \mathbf{P}_T(\theta_{P_0})$) we call the canonical gradient or efficient score.

The following theorem shows that smooth functional theory provides a criterion for the asymptotic efficiency of an estimator of $\Psi(P_0)$. It is Theorem 3.11.2 of VAN DER VAART AND WELLNER (1996), but here we state it as formulated in GROENEBOOM (1996). Let $T_n = t_n(X_1, \ldots, X_n)$ be a real valued measurable function of the sample. We consider a collection of sequences $\{P_{c/\sqrt{n}}\}$, where $c > 0$.

Theorem 1.7.1 *(Convolution Theorem)*
Suppose that

1. *Ψ is pathwise differentiable at $P_0 \in \mathcal{P}$ along a Hellinger differentiable path.*

2. *T_n is a regular estimator, meaning that $\sqrt{n}(T_n - \Psi(P_{c/\sqrt{n}}))$ converges, under $P_{c/\sqrt{n}}$, in distribution to a random variable Z, which does not depend on the direction, i.e., the tangent of the path $\{P_{c/\sqrt{n}}\}$ to P_0.*

3. *The set of all directions is a linear space.*

Then there exist random variables Z_0 and Δ_0 such that

A. *Z has the same distribution as $Z_0 + \Delta_0$.*

B. *Z_0 and Δ_0 are independent.*

C. *$Z_0 \sim \mathcal{N}\left(0, ||\tilde{k}_{P_0}||^2_{P_0}\right)$.*

An efficient estimator of the parameter $\Psi(P_0)$ is a regular estimator with limiting distribution equal to the distribution of Z_0 in the theorem. As one can see from condition 2 (excluding superefficiency), Z_0 has minimal variance within the set of all regular estimators.

We will apply the theory of smooth functionals to testing. Suppose we want to test the null hypothesis $H_0 : \Psi(P) \leq 0$ against the one-sided alternative $H_1 : \Psi(P) > 0$.

Theorem 1.7.2 (VAN DER VAART (1998), Theorem 25.44)
Let the functional $\Psi : \mathcal{P} \to \mathbb{R}$ be differentiable at P relative to the tangent space T_P with the canonical gradient $\tilde{k}_P$. Suppose that $\Psi(P) = 0$. Then for every sequence of power functions $P \to \pi_n(P)$ of level-α test for $H_0 : \Psi(P) \leq 0$, and for every $a \in T_P$ with $\langle \tilde{k}_P, a\rangle_P > 0$ and every $c > 0$,

$$\limsup_{n\to\infty} \pi_n(P_{c/\sqrt{n},a}) \leq 1 - \Phi\left(z_{1-\alpha} - c\frac{\langle \tilde{k}_P, a\rangle}{\sqrt{||\tilde{k}_P||^2_P}}\right),$$

where Φ is the normal $\mathcal{N}(0,1)$ distribution function and z_α denotes a corresponding quantile.

Note that if $\langle \tilde{k}_P, a\rangle_P < 0$, the subfamily $P_{c/\sqrt{n},a}$ lies in H_0 for n sufficiently large, and then the power of the test converges to zero. The theorem provides an upper bound for the asymptotic power function and if this bound is attained by some test, it is asymptotically (locally) the most efficient one. The analogous result can be proved for the simple null hypothesis.

Theorem 1.7.3 *Let the functional $\Psi : \mathcal{P} \to \mathbb{R}$ be differentiable at P relative to the tangent space T_P with the canonical gradient $\tilde{k}_P$. Suppose that $\Psi(P) = 0$. Then for every sequence of power functions $P \to \pi_n(P)$ of a level-α test for $H_0 : \Psi(P) = 0$, and for every $a \in T_P$ and every $c > 0$,*

$$\limsup_{n\to\infty} \pi_n(P_{c/\sqrt{n},a}) \le 1 - \Phi\left(z_{1-\alpha/2} - c\frac{\langle \tilde{k}_P, a\rangle}{\sqrt{||\tilde{k}_P||_P^2}} \right) + \Phi\left(z_{\alpha/2} - c\frac{\langle \tilde{k}_P, a\rangle}{\sqrt{||\tilde{k}_P||_P^2}} \right).$$

1.8 Empirical CLT and stochastic equicontinuity

In this section we discuss results which play an important role in many of our proofs. Let S be a set, endowed with a metric ρ. The covering number $N(\delta, S, \rho)$ is defined by

$$N(\delta, S, \rho) = \inf_{S_\delta} \#\{S_\delta \subseteq S : \text{for every } s \in S \text{ there is } s' \in S_\delta \text{ such that } \rho(s, s') \le \delta\}$$

or in words, the number of elements in the minimal δ-net in S. If there is no such finite δ-net we assume $N(\delta, S, \rho) = \infty$. The set S will often be a set of functions with metric ρ, defined by

$$\rho_{L_p(Q)}(s_1, s_2) = ||s_1 - s_2||_{p,Q} = \left(\int |s_1(t) - s_2(t)|^p dQ(t) \right)^{\frac{1}{p}}.$$

For a set of functions S the covering number with bracketing is defined by

$$N_B(\delta, S, \rho) = \inf_{S_\delta} \#\{S_\delta \subseteq S : \text{for any } s \in S \text{ there are } s', s'' \in S_\delta,$$
$$\text{such that } \rho(s', s'') \le \delta \text{ and for any } t \;\; s'(t) \le s(t) \le s''(t)\}.$$

The entropy $H(\delta, S, \rho)$ is the logarithm of the covering number $H(\delta, S, \rho) = \log N(\delta, S, \rho)$, and the entropy integral is defined by

$$J(\delta, S, \rho) = \int_0^\delta \sqrt{\log\left(\frac{N(t, S, \rho)^2}{t} \right)} dt.$$

Let S be a subset of $L_2(P_0)$. The uniform law of large numbers (ULLN), which can be found e.g. in van de Geer (2000) implies that if S has a P-integrable envelope and $\frac{1}{n} H\left(\delta, S, \rho_{L_1(P_n)}\right)$ converges to zero in probability, then the ULLN holds, i.e., $\sup_{s\in S} \left| \int s d(P_n - P_0) \right|$ converges to zero almost surely.

The empirical central limit theorem can be formulated in a similar way. Let us denote the empirical process, indexed by the functions s, as $\{\nu_n(s),\ s \in S\}$, where ν_n is defined by

$$\nu_n(s) = \sqrt{n} \int s d(P_n - P_0).$$

Assume that the sample $X_1, \dots, X_n$ is defined on the probability space $(\Omega, \mathcal{F}, P_0)$. Let $\mathcal{K}$ be the set of all bounded, real-valued functions on S, equipped with the supremum norm and let S be equipped with the $L_2(P_0)$-norm. Moreover, let $\overline{\mathcal{B}}$ be the σ-algebra generated by the set of all (open or closed) balls in $\mathcal{K}$. Suppose ν_n to be $(\mathcal{F}/\overline{\mathcal{B}})$-measurable. If $\{\nu(s),\ s \in S\}$ is a Gaussian process with zero mean and

$$\operatorname{cov}(\nu(s_1), \nu(s_2)) = \int s_1 s_2 dP_0 - \int s_1 dP_0 \int s_2 dP_0,\ s_1, s_2 \in S$$

(the proof of existence of such a process can be found in POLLARD (1984)), we call the class S P_0-Donsker if

$$Ef(\nu_n) \to Ef(\nu),$$

for every bounded, continuous function $f : \mathcal{K} \to \mathbb{R}$ that is $\overline{\mathcal{B}}/\mathcal{B}$-measurable, where $\mathcal{B}$ is the collection of Borel sets on $\mathbb{R}$. This means that the empirical process converges in distribution to the Gaussian process, defined above, and therefore the two statements: "the empirical central limit theorem holds" and "S is a P_0-Donsker class" are equivalent.

DUDLEY (1984) has given a sufficient condition for the class S to be P_0-Donsker class. It can be expressed in terms of stochastic equicontinuity. Processes Z_n indexed by elements of the set S are called stochastic equicontinuous (with respect to the metric ρ), if for all $\eta > 0$ there exists a $\delta > 0$ such that

$$\limsup_{n\to\infty} P\left\{\sup_{s_1, s_2 \in S,\ \rho(s_1, s_2)\le\delta} |Z_n(s_1) - Z_n(s_2)| > \eta\right\} < \eta.$$

The sufficient condition for S to be P_0-Donsker class is total boundedness of the class S (which means that for any $\delta > 0$, $N(\delta, S, \rho_{L_2(P_0)}) < \infty$) together with the property that the empirical process ν_n is stochastic equicontinuous with respect to the $L_2(P_0)$-metric.

A sufficient condition for the class S to be stochastic equicontinuous, which can be applied in proving P_0-Donsker property of this class, is given in, e.g., POLLARD (1984). It characterizes the class S as stochastically equicontinuous with respect to $\rho_{L_2(P_0)}$, if it has an envelope $G \in L_2(P_0)$ and if for each $\eta > 0$ and $\epsilon > 0$ there exists $\gamma > 0$ such that

$$\limsup_{n\to\infty} P\left\{J(\gamma, S, \rho_{L_2(P_n)}) > \eta\right\} < \epsilon. \tag{1.8.1}$$

As further detailed in POLLARD (1984), the given condition is a necessary condition as well for a large set of classes. In Chapter 2 an analogue of the criteria for the class S to be P_0-Donsker is established for the case when $S = S_n$ depends on n and when (1.8.1) fails.

For proving various modifications of the Donsker property, we often use Hoeffding's inequality and the chaining lemma. Hoeffding's inequality is used for obtaining a probability bound for a sum of zero mean independent random variables Y_i. If $a_i \le Y_i \le b_i$, for any i, then, for any $\eta > 0$:

$$P\left\{\sum_{i=1}^n Y_i \ge \eta\right\} \le \exp\left[-2\eta^2 \Big/ \sum_{i=1}^n (b_i - a_i)^2\right].$$

The chaining lemma holds under the exponential type bound,

$$P\left\{|Z(t) - Z(s)| > \eta\rho(s,t)\right\} \le 2\exp\left(-\frac{\eta^2}{2D^2}\right) \quad \text{for any } s, t \in T \text{ and } \eta > 0,$$

for a stochastic process Z with index set T. Often, a countable dense subset T^* of the set T can be found, such that for any ϵ, $0 < \epsilon < 1$,

$$P\left\{|Z(s) - Z(t)| > 26DJ(\rho(s,t), T, \rho) \text{ for some } s, t \in T^* \text{ with } \rho(s,t) \le \epsilon\right\} \le 2\epsilon.$$

The same holds for T^* replaced by T, if the process Z has continuous sample paths.

1.9 Convergence of stochastic processes

In this section we discuss the notation for convergence of stochastic processes in distribution. Consider the process $\{Z(t),\ t \in T\}$, where the parameter t does not have to be real. This process is considered to be a random element of the space $\mathcal{X}$ of functions on T, assumed to be large enough to contain all possible realizations of the process. The process Z_n converges to the process Z in distribution if

$$\lim_{n\to\infty} Ef(Z_n) = Ef(Z)$$

for every real-valued continuous bounded function f, $f : \mathcal{X} \to \mathbb{R}$, measurable with respect to the σ-algebra on $\mathcal{X}$ and the Borel σ-algebra on $\mathbb{R}$. Therefore the definition of the continuous bounded function depends on the choice of metric on $\mathcal{X}$.

Consider first the case $T = [0,1]$ and $\mathcal{X} = D[0,1]$, the space of *cadlag* functions, i.e., right continuous functions having left limits in each point of T, provided with the uniform metric

$$d(z_1, z_2) = \sup_{t\in[0,1]} |z_1(t) - z_2(t)|.$$

The name *cadlag* comes from the French expressions "continue á droite" and "limites á gauche". In most applications, in particular those where Brownian motion arises as a limit, more complicated settings are not required. For a sequence of random elements X_i, $i = 1, 2, \ldots$ and X in the cadlag space $D[0,1]$, assumed to be measurable under the uniform metric and projection σ-field, suppose that there is some separable subset C of $D[0,1]$, such that $P\{X \in C\} = 1$. Then a necessary and sufficient condition for X_i to converge to X in distribution, given in, e.g., POLLARD (1984), is that firstly, any finite dimensional projection $(X_n(t_1), \ldots, X_n(t_k))$, for $0 \le t_i \le \ldots \le t_k \le 1$, must converges in distribution to $(X(t_1), \ldots, X(t_k))$, and, secondly, for each $\epsilon > 0$ and $\delta > 0$ there is a grid $0 = t_0 < t_1 < \ldots < t_k = 1$ such that

$$\limsup P\left\{\max_i \sup_{J_i} |X_n(t_i) - X_n(t)| > \delta\right\} < \epsilon,$$

where J_i denotes the interval $[t_i, t_{i+1})$, for $i = 0, \ldots, k-1$. The projection σ-fields are introduced to avoid certain measurability difficulties.

The space $D[0,1]$ has the same structure as any *cadlag* space, indexed by a closed interval $[a,b]$ of the extended real line (possibly $[-\infty,\infty]$). This gives that the empirical process $\nu_n = \sqrt{n}(P_n - P_0)$, based on independent sampling from a distribution function F_0, converges as a random element of $D[a,b]$, to a Gaussian process ν, which has zero mean and covariance given by

$$\mathrm{cov}(\nu(r), \nu(s)) = F_0(r)(1 - F_0(s)) \quad \text{for } r \le s.$$

The proof of this result can be found in, e.g., POLLARD (1984), where also the existence of the Gaussian process is shown. In this setting, convergence of the empirical process ν_n to the Gaussian process ν mentioned in Section 1.8 can be considered to be a generalization of the central limit theorem for processes, mentioned above (we can take the subset $S_0 \subset S$, $S_0 = \{1_{(-\infty,t]},\ t \in \mathbb{R}\}$).

In Section 1.8 we looked at process convergence in another setting. The index space S was supposed to be totally bounded with respect to $\rho_{L_2(P_0)}$ class of functions, having an $L_2(P_0)$-integrable envelope. Furthermore, $\mathcal{K}$ was assumed to be the set of all bounded, real-valued functions on S, equipped with the supremum norm.

If the space of functions $\mathcal{X} = D[0,\infty)$ (therefore limits in $+\infty$ does not have to exist), a necessary and sufficient condition for the process to convergence can be expressed in terms of convergence of restrictions of the process to finite closed intervals. Let the processes $X_1, X_2 \ldots,$ and X be random elements of $D[0,\infty)$ and assume that there is a separable subset C of $D[0,\infty)$ such that $P\{X \in C\} = 1$. Then (see again POLLARD (1984)) a necessary and sufficient condition for X_n to converge to X is that, for any $k \in \mathbb{N}$, the restriction $L_k X_n$ of the process $\{X_n(t) : t \in [0,\infty)\}$ to the interval $[0,k]$ converges in distribution in $D[0,k]$ to the restriction $L_k X$ of the process X to $[0,k]$. Similar result can be established for the space $\mathcal{X} = D(-\infty,\infty)$.

1.10 Open problems

We think that the results we have obtained for two sample tests under Lehmann alternatives are rather general (in this setting) and that our conditions are close to the weakest possible. The case of different supports of the censored and censoring distributions follows immediately from our results. But how to test in the case of interval censoring, case II or higher, remains an open problem. We hope that this can be solved in the near future. Another challenging problem is to consider the covariate Z to be not simply Bernoulli but, as in MURPHY AND VAN DER VAART (1997), to be distributed according one or another continuous distribution. Both of these extensions of the theory have many practical applications, so it is unfortunate that solutions are not yet available.

Concerning practical application we did not give any advice on the best way of estimating the parameters of the limiting distribution of the score statistics. Furthermore, to make the likelihood ratio test easier to apply in practice, faster algorithms for computing the maximum likelihood estimator in the case of Lehmann alternative are needed. Combination of the EM algorithm with some other existing algorithm may be the way to go here.

Next, in connection with the Grenander estimator, one can think of the supremum distance, for which the exact rate of convergence and asymptotic distribution have not been established yet. The supremum distance between the empirical distribution function and its concave majorant is probably easier to study than the supremum distance for the Grenander estimator itself, but for the supremum distance between the empirical distribution function and its concave majorant there are no (sharp) results available either.

In the monotone regression setup (considered by DUROT AND TOCQUET (2002)) the study of the behavior of the supremum distance may be somewhat easier, since then the process can be directly embedded into Brownian motion. Convergence as a process in $D(-\infty,\infty)$ may be a stepping stone in such a proof, which will be analogous to what we have done in Chapter 4.

Chapter 2

Two samples score test

We propose a two-sample test for testing that the distribution functions F_0 and F_1, generating the two samples, are equal, in the case that the samples are subject to current status censoring (also called "interval censoring, case 1"). The proposed test is a score test and tests the null hypothesis $F_0 = F_1$ against Lehmann alternatives $F_1 = F_0^{1+\theta}$, for some $\theta > -1$. The test statistic is shown to converge at rate $\sqrt{n}$ and to be asymptotically normal. Moreover, we show that the test is asymptotically efficient for testing against the alternative $F_1 = F_0^{1+\theta}$.

2.1 Introduction

We say that a random variable X is subject to interval censoring, case 1, if instead of the (unobservable) X, we observe a pair (T, Δ), where T is a random variable independent of X, and distributed according to an unknown distribution G, and where Δ is defined by

$$\Delta = 1_{\{X \le T\}}. \tag{2.1.1}$$

Data of this type are also referred to as "current status data".

To facilitate treatment and also to put everything into the more general context of linear models, we consider instead of two samples just one joint sample where each element has a covariate Z equal to zero or to one. Formally the situation considered in the present paper is as follows:

Given the i.i.d. sample of triples $\{(X_i, T_i, Z_i)\}_{i=1}^n$ where $X_i, T_i \in \mathbb{R}$ and $Z_i \in \{0,1\}$, X_i and T_i are independent and distributed in the following way:

$$\begin{aligned}
&Z_i \sim Bernoulli(p)\\
&P\{X_i \le t|\ Z_i = 0\} = F_0(t)\\
&P\{X_i \le t|\ Z_i = 1\} = F_1(t)\\
&P\{T_i \le t|\ Z_i = 0\} = G_0(t)\\
&P\{T_i \le t|\ Z_i = 1\} = G_1(t)
\end{aligned}$$

We want to test the null hypothesis $H_0: \ F_0 = F_1$ on the basis of the sample of triples

$$\{(T_i, \Delta_i, Z_i)\}_{i=1}^n \quad \textit{(interval censoring, case I)}$$

Note that we allow the distribution of the observation times T_i to be different for the two samples. Two sample permutation tests for interval censored data have been considered by PETO AND PETO (1972). Since they rely on the permutation distribution, such tests can only be used when the censoring mechanism is the same in both samples.

Up to now we did not specify the alternative H_1. All results given below are proved for Lehmann alternatives. That is: the null hypothesis is given by:

$$H_0: \ F_0 = F_1$$

and the alternative by:

$$H_1: \ F_1(t) = F_0(t)^{1+\theta}, \ \ \theta \in (-1, \infty) \setminus \{0\}.$$

The alternative in terms of survival functions:

$$H_1^{'}: \ 1 - F_1(t) = (1 - F_0(t))^{1+\theta}, \ \theta \in (-1, \infty) \setminus \{0\},$$

is treated similarly.

To explain the construction of the test and the intuition behind it, we first have to introduce some notation. Fixing the measure μ on $\mathbb{R} \times \{0,1\}^2$ by

$$\mu = \mu_{\text{Lebesgue}} \times (\text{counting measure on } \{0,1\}^2),$$

the density of the triple (T, Δ, Z) under the Lehmann alternative is given by

$$[L(\theta, F)]\,(t, \delta, z)\,((1-z)(1-p)g_0(t) + pzg_1(t)),$$

where

$$[L(F,\theta)]\,(t,\delta,z) = \delta F(t)^{1+z\theta} + (1-\delta)\left\{1 - F(t)^{1+z\theta}\right\},$$

and where F and θ are the underlying distribution and parameter in the Lehmann alternative model, respectively. The likelihood $L_n(\theta, F)$ of our sample of triples (T_i, Δ_i, Z_i) is then given by:

$$L_n(\theta, F) \prod_{i=1}^{n} ((1-z_i)(1-p)g_0(t_i) + pz_i g_1(t_i)),$$

where

$$L_n(\theta, F) = \prod_{i=1}^{n} [L(\theta, F)]\,(T_i, \Delta_i, Z_i)$$

and the maximum likelihood estimator $(\hat{\theta}_n, \hat{F}_n)$ of the underlying pair (θ_0, F_0) is defined by

$$(\hat{\theta}_n, \hat{F}_n) = \underset{\theta, F}{\operatorname{argmax}}\, L_n(\theta, F), \tag{2.1.2}$$

where the argmax is taken over all possible distribution functions F and parameters $\theta \in (-1, \infty)$.

To test the null hypothesis, we also need the restricted maximum likelihood estimator $\hat{F}_n^0$ of the underlying distribution function F_0, under the null hypothesis:

$$\hat{F}_n^0 = \operatorname*{argmax}_F L_n(0, F). \tag{2.1.3}$$

We now show how this leads to a score test. Define the functions

$$\begin{aligned} l(\theta, x, \delta, z) &= \delta(1 + z\theta)\log x + (1-\delta)\log\left(1 - x^{1+z\theta}\right), \\ l_1(\theta, x, \delta, z) &= \frac{\partial l(\theta, x, \delta, z)}{\partial\theta} = \delta z \log x - (1-\delta) z \frac{x^{1+z\theta}\log x}{1 - x^{1+z\theta}}, \\ l_2(\theta, x, \delta, z) &= \frac{\partial l(\theta, x, \delta, z)}{\partial x} = \frac{\delta(1+z\theta)}{x} - \frac{(1-\delta)(1+z\theta)x^{z\theta}}{1 - x^{1+z\theta}}. \end{aligned}$$

Using this notation, we can rewrite (2.1.2) and (2.1.3) as

$$(\hat{\theta}_n, \hat{F}_n) = \operatorname*{argmax}_{\theta, F} \int l(\theta, F(t), \delta, z)\, dP_n(t, \delta, z) \tag{2.1.4}$$

and

$$\hat{F}_n^0 = \operatorname*{argmax}_F \int l(0, F(t), \delta, z)\, dP_n(t, \delta, z). \tag{2.1.5}$$

Consider $\int l(\theta, \hat{F}_n^0(t), \delta, z) dP_n(t, \delta, z)$ as a function of θ under the null hypothesis $\theta_0 = 0$. The function $L_n(\theta, \hat{F}_n^0)$ is approximately maximized by $\theta = 0$. This means that, at the derivative level, we get

$$\int l_1(0, \hat{F}_n^0(t), \delta, z)\, dP_n(t, \delta, z) \approx 0, \tag{2.1.6}$$

while this will not be true for (local) alternatives.

The statistic on the left-hand side of (2.1.6) can also be written as

$$\int z\hat{F}_n^0(t) \log \hat{F}_n^0(t) \left(\frac{\delta}{\hat{F}_n^0(t)} - \frac{1-\delta}{1 - \hat{F}_n^0(t)} \right) dP_n(t, \delta, z).$$

Now let the functional $[S(F)](t, \delta, z)$ be defined by

$$[S(F)](t, \delta, z) = zw(F(t)) \left(\frac{\delta}{F(t)} - \frac{1-\delta}{1 - F(t)} \right),$$

where w is a (weight) function $w: [0,1] \to \mathbb{R}$.

The main result of this paper is:

Theorem 2.1.1 *(Main Theorem)*
Consider alternatives of the form: $\theta_n = \theta_0/\sqrt{n}$, where $\theta_0 = 0$ is allowed. Suppose that F_0, G_0 and G_1 satisfy

(i) $\text{support}(F_0) = \text{support}(G_0) = \text{support}(G_1)$,

(ii) F_0, G_0 and G_1 have densities satisfying

$$0 < \inf_{t \in \text{support}(F_0)^0} \frac{f_0(t)}{(1-p)g_0(t) + pg_1(t)} \leq \sup_{t \in \text{support}(F_0)^0} \frac{f_0(t)}{(1-p)g_0(t) + pg_1(t)} < \infty, \tag{2.1.7}$$

where support$(F_0)^0$ *denotes the interior of the support of the density of F_0, and*

$$r(t) = \frac{g_0(F_0^{-1}(t))}{g_1(F_0^{-1}(t))} \text{ is differentiable and } \sup_{t \in [0,1]} |r'(t)| \left(1 \wedge r(t)^{-2}\right) < \infty. \tag{2.1.8}$$

Finally, suppose the weight function w to be

- *either a Lipschitz function satisfying $w(0) = w(1) = 0$*
- *or the function $w(x) = x|\log x|^m$, $m \geq 1$.*

Then

$$S_n = \sqrt{n} \int \left[S(\hat{F}_n^0)\right](t, \delta, z) dP_n(t, \delta, z) \xrightarrow{\mathcal{D}} \mathcal{N}(\mu, \sigma^2)$$

where the mean and variance of the asymptotic distribution are given by

$$\mu = p(1-p)\,\theta_0 \int \frac{w\left(F_0(t)\right) \log F_0(t)}{1 - F_0(t)} \cdot \frac{g_0(t) g_1(t)}{(1-p)g_0(t) + pg_1(t)}\, dt, \tag{2.1.9}$$

and

$$\sigma^2 = p(1-p) \int \frac{w\left(F_0(t)\right)^2}{F_0(t)\left(1 - F_0(t)\right)} \cdot \frac{g_0(t) g_1(t)}{(1-p)g_0(t) + pg_1(t)}\, dt \tag{2.1.10}$$

and where $p \in (0,1)$ is the Bernoulli parameter.

We need condition (2.1.8) in certain entropy calculations below. Among the cases where this condition is satisfied one may think of

- The censoring distributions G_0 and G_1 are the same.
- $$\inf_{t \in \text{support}(F_0)} f_0(t) > 0, \quad \sup_{t \in \text{support}(F_0)} \left|g_i'(t)\right| < \infty, \; i = 0, 1$$

  This can only occur if the support of F_0 is compact.
- $F_0 \sim \exp(\lambda_1)$, $G_0 \sim \exp(\lambda_2)$, $G_1 \sim \exp(\lambda_3)$ distributions (thus $F_0(x) = (1 - \exp(-\lambda_1 x))\, 1_{x \geq 0}$ etc.). Suppose $\lambda_2 \leq \lambda_3$.

For the third example with the exponential distributions, (2.1.7) is equivalent to $\lambda_1 = \lambda_2$. The condition (2.1.8) then boils down to $\lambda_3 \geq \lambda_1 + \lambda_2$.

The requirements (2.1.8) can be replaced by the following conditions. Let A_n and B_n be defined by

$$A_u = \left\{t \in [0,1] : \; |r'(t)| \left(1 \wedge r(t)^{-2}\right) \geq u\right\} \tag{2.1.11}$$

$$B_u = \left\{t \in [0,1] : \; \exists s \in A_n, \; |t - s| \leq u^{-2}\right\}. \tag{2.1.12}$$

Then there must be

$$\lim_{u\to\infty} u(\log u)^{2(m+2)} G_i \circ F_0^{-1}(B_u) = 0, \quad i = 1, 2, \tag{2.1.13}$$

and

$$\frac{g_0(t)}{g_1(t)} \text{ -- monotone on support}(F_0).$$

In case if w is a Lipschitz function we assume here and later $m = 0$.

For the exponential distributions condition (2.1.13) is always satisfied. In relation (2.1.7) we suppose that support(F_0) is an interval (possibly infinite). If somewhere inside this interval f_0 is equal to zero, condition is satisfied if g_0 and g_1 are also equal to zero in this region.

The proof of this theorem is given in the next section. In that section we give the main line of argument, which relies on certain technical lemmas the proof of which is postponed to later sections.

The third section is contributed to the proof of the asymptotic efficiency of the testing procedure based on the statistic S_n and the Theorem above. This relies on the theory of smooth functionals (in the present case the smooth functional is given by $\theta(P_0)$).

The fourth section describes properties of the maximum likelihood estimator we will use. This is mostly a generalization of the results of VAN DE GEER (2000) and GROENEBOOM AND WELLNER (1992) to the present situation. In sections 5 and 6 we analyze the Donsker terms in the main representation separately and thereby complete the proof of our main result. Section 7 gives some results of the computer simulations showing convergence to the limiting distribution in the Main Theorem.

Finally, we want to make some comments on the merits of the proposed test. The score test is asymptotically efficient. The likelihood ratio test, to be discussed in separate paper, is also asymptotically efficient. But in choosing the test in practical applications we have to think of (at least) the following three things:

- How large should the sample size be for the distribution of the test statistic to be sufficiently close to the theoretical?
- How accurate can we estimate the parameters of the theoretical distribution?
- How difficult is it to compute the test statistic?

Comparing the likelihood ratio test and the score test, we would recommend to use the score test when the sample size is rather large (say, larger than 1000), whereas for the smaller sample size it is more advisable to use the likelihood ratio test. Computer simulations indicate that the likelihood ratio statistics converges somewhat faster to the asymptotical distribution both under the null hypothesis and under alternatives.

The score test, however, has the advantage of being easier to compute. For the score statistic S_n we only have to compute the maximum likelihood estimator under the null hypothesis. Under the null hypothesis we can apply a one step algorithm that uses a number of operations of order n^2. On the other hand, for the likelihood ratio test we must (in principle) compute the maximum likelihood estimator for all possible θ's. For this we need to use an iterative algorithm, and maximize the likelihood over a grid of θ's, since $L_n(\theta, \hat{F}_{n,\theta})$ is not a simple explicit function. This will amount to a number of operations of order n^4.

2.2 Proof of the main theorem

The maximum likelihood estimator for interval censored data is usually studied under the assumption of compact support of the distribution of the censored variables X_i and positivity of the density (respectively marginal densities in the case of interval censoring with more observation times per unobservable) of the observation distribution on the support of the distribution of the censored variables. But generally speaking we would expect that a condition, stipulating that the censored and censoring values are "comparably densely" distributed would be sufficient.

Note that the statistic S_n evaluates $\hat{F}_n^0$ only at the observation times T_i and that $\hat{F}_n^0$ depends on $\{T_i\}$ only via the ordering of $\{\Delta_i\}$ (see (2.1.3)). This suggests making a transformation in the hidden space (the space corresponding to the unobservable triples (X_i, T_i, Z_i)) that would change the distributions of X_i and T_i without changing the indicators $\Delta_i = 1_{\{X_i \leq T_i\}}$ or the ordering w.r.t. T_i. The solution of the maximization problem (2.1.3) will change, but not the statistic S_n. The transformation is quite simple and given in the next lemma. The construction is somewhat similar to the quantile transformation.

Lemma 2.2.1 *The statement of Theorem 2.1.1 is implied by the corresponding statement where F_0 is the Uniform(0,1) distribution function.*

Proof: Suppose that for the parameter θ, for the censoring distributions G_0 and G_1, and for the censored distribution F_0, the conditions of the main theorem are satisfied. Consider the sequence of i.i.d. random triples $\{(X_i, T_i, Z_i)\}_{i=1}^n$ in the hidden space, where the distribution of the triples satisfies the conditions of Theorem 2.1.1. Now take another sequence of triples $\{(U_i, \tilde{T}_i, \tilde{Z}_i)\}_{i=1}^n$ defined by

$$U_i = F_0(X_i),\ \tilde{T}_i = F_0(T_i),\ \tilde{Z}_i = Z_i.$$

Their distributions are given by

$$\begin{aligned} P\{U_i \leq u |\ Z_i = z\} &= P\{F_0(X_i) \leq u |\ Z_i = z\} = P\left\{X_i \leq F_0^{-1}(u) \middle|\ Z_i = z\right\} \\ &= F_0\left(F_0^{-1}(u)\right)^{1+Z\theta_n} = \left(F_{Un[0,1]}\right)^{1+Z\theta_n} \end{aligned}$$

and

$$\begin{aligned} P\left\{\tilde{T}_i \leq t \middle|\ Z_i = z\right\} &= P\{F_0(T_i) \leq t |\ Z_i = z\} \\ &= P\left\{T_i \leq F_0^{-1}(t) \middle|\ Z_i = z\right\} = G_z\left(F_0^{-1}(t)\right), \end{aligned}$$

where F_0^{-1} is defined by

$$F_0^{-1}(x) = \sup\{t :\ F_0(t) \leq x\}.$$

Note that, by continuity, $F_0\left(F_0^{-1}(x)\right) = x$.

Now consider the corresponding sequences of triples in the observation space:

$$(X_i, T_i, Z_i) \longrightarrow (T_i, \Delta_i = 1_{\{X_i \leq T_i\}}, Z_i), \tag{2.2.1}$$

$$(U_i, \tilde{T}_i, \tilde{Z}_i) \longrightarrow (\tilde{T}_i, \tilde{\Delta}_i = 1_{\{U_i \leq \tilde{T}_i\}}, \tilde{Z}_i). \tag{2.2.2}$$

Note that since F_0 is monotone, and since almost surely there are no points X_i and T_i on the intervals where it is constant, we have:

$$\begin{aligned} F_0(X_i) \le F_0(T_i) &\iff X_i \le T_i, \\ F_0(T_i) \le F_0(T_j) &\iff T_i \le T_j. \end{aligned} \tag{2.2.3}$$

Due to (2.2.3) the Δ_i's in the observed triples (2.2.1) and (2.2.2) have the same values.

Define

$$x^n = \left(\hat{F}_n^0(T_1), \dots, \hat{F}_n^0(T_n)\right).$$

Then x^n satisfies

$$x^n = \operatorname*{argmax}_{x\in\mathbb{R}^n,\ 0\le x_i\le x_{i+1}\le 1} \sum_{i=1}^{n} l(0, x_i, \Delta_i, Z_i).$$

Similarly, let

$$\tilde{x}^n = \left(\tilde{F}_n^0(\tilde{T}_1), \dots, \tilde{F}_n^0(\tilde{T}_n)\right),$$

where $\tilde{F}_n$ is defined as

$$\tilde{F}_n = \operatorname*{argmax}_F \sum_{i=1}^{n} l\left(0, F(\tilde{T}_i), \tilde{\Delta}_i, \tilde{Z}_i\right).$$

Then we get:

$$\tilde{x}^n = \operatorname*{argmax}_{x\in\mathbb{R}^n,\ 0\le x_i\le x_{i+1}\le 1} \sum_{i=1}^{n} l(0, x_i, \tilde{\Delta}_i, \tilde{Z}_i) = \operatorname*{argmax}_{x\in\mathbb{R}^n,\ 0\le x_i\le x_{i+1}\le 1} \sum_{i=1}^{n} l(0, x_i, \Delta_i, Z_i) = x^n.$$

Hence the statistics S_n and $\tilde{S}_n$, corresponding to the samples (2.2.1) and (2.2.2), are equal since they can be represented as

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i w(x_i^n) \left(\frac{\Delta_i}{x_i^n} - \frac{1-\Delta_i}{1-x_i^n}\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Z}_i w(\tilde{x}_i^n) \left(\frac{\tilde{\Delta}_i}{\tilde{x}_i^n} - \frac{1-\tilde{\Delta}_i}{1-\tilde{x}_i^n}\right) = \tilde{S}_n.$$

■

The above lemma shows that the proof of the Theorem 2.1.1 will follow from

Theorem 2.2.1 *Suppose that $\theta_n = \theta_0/\sqrt{n}$ where $\theta_0 = 0$ is allowed. Suppose*

(i) $\operatorname{supp} G_0 = \operatorname{supp} G_1 = [0, 1]$

(ii) G_0 and G_1 satisfies (2.1.7) and (2.1.8) with $F_0 = F_{Un(0,1)}$

Moreover, suppose the weight function w to be

- *either the Lipschitz function satisfying $w(0) = w(1) = 0$*
- *or $w(x) = x|\log x|^m$, $m \ge 1$.*

Then

$$S_n = \sqrt{n} \int \left[S(\hat{F}_n^0)\right](t, \delta, z)dP_n(t, \delta, z) \xrightarrow{\mathcal{D}} \mathcal{N}(\mu, \sigma^2)$$

where the parameters μ and σ^2 are given by (2.1.9) and (2.1.10).

We need to pay special attention to the behavior of our test statistic near the boundary of the support of F_0 (from now supposed to be the Uniform(0,1) distribution function). To this end we define

$$I_n = \left[n^{-1/3}(\log n)^2, 1 - n^{-1/3}(\log n)^2\right], \; I_n^c = [0,1] \setminus I_n,$$
$$I_n^\ell = \left[0, n^{-1/3}(\log n)^2\right], \; I_n^r = \left[1 - n^{-1/3}(\log n)^2, 1\right].$$

In the proof of the following lemma and also in the rest of this section we will use that for F belonging to the set:

$$\mathcal{F}_n = \left\{ F : \text{ distributions on } [0,1], \sup_{t\in[0,1]} |F(t) - F_0(t)| \le n^{-1/3} \log n \right\} \tag{2.2.4}$$

we have:

$$\frac{1}{2} \le \inf_{t\in I_n, F\in\mathcal{F}_n} \frac{F(t)}{F_0(t)} \le \sup_{t\in I_n, F\in\mathcal{F}_n} \frac{F(t)}{F_0(t)} \le 2, \tag{2.2.5}$$

for n sufficiently large. It will be shown in Lemma 2.4.6 that

$$P\{\hat{F}_n^0 \in \mathcal{F}_n\} \to 1, \; n \to \infty.$$

Lemma 2.2.2 *Under the conditions of Theorem 2.2.1:*

$$S_n = \sqrt{n} \int_{I_n} \left[S(\hat{F}_n^0)\right](t,\delta,z) dP_n(t,\delta,z) + o_p(1)$$

Proof: We have:

$$\left| S_n - \sqrt{n} \int_{I_n} \left[S(\hat{F}_n^0)\right](t,\delta,z) dP_n(t,\delta,z) \right| \le \sqrt{n} \left| \int_{I_n^c} \left[S(\hat{F}_n^0)\right](t,\delta,z) dP_n(t,\delta,z) \right|$$
$$\le \sqrt{n} \int_{I_n^\ell} \frac{\left|w\left(\hat{F}_n^0(t)\right)\right| \delta}{\hat{F}_n^0(t)} dP_n(t,\delta,z) + \sqrt{n} \int_{I_n^\ell} \frac{\left|w\left(\hat{F}_n^0(t)\right)\right| (1-\delta)}{1 - \hat{F}_n^0(t)} dP_n(t,\delta,z)$$
$$+ \sqrt{n} \int_{I_n^r} \frac{\left|w\left(\hat{F}_n^0(t)\right)\right| \delta}{\hat{F}_n^0(t)} dP_n(t,\delta,z) + \sqrt{n} \int_{I_n^r} \frac{\left|w\left(\hat{F}_n^0(t)\right)\right| (1-\delta)}{1 - \hat{F}_n^0(t)} dP_n(t,\delta,z).$$

Denote the last four terms by I, II, III and IV. Before we show them all to be negligible we note that

$$\sup_{t\in[0,1]} \frac{|w(t)|}{1-t} \le C_1 < \infty \;\text{ and }\; \sup_{t\in[0,1]} \frac{|w(t)|/t}{(1 + |\log t|^m)} \le C_2 < \infty \tag{2.2.6}$$

as follows from the condition on w. Using this we obtain

$$I \le \sqrt{n} C_2 \int_{I_n^\ell} \left(1 + \left|\log \hat{F}_n^0(t)\right|^m\right) \delta dP_n(t,\delta,z) \le \sqrt{n} (\log n)^m C_3 \int_{I_n^\ell} \delta dP_n(t,\delta,z) = o_p(1)$$

since, by Lemma 2.4.4, $\hat{F}_n^0(T_i) \ge \frac{\Delta_i}{n}$ and since, by the Markov inequality and (2.1.7),

$$P\left\{ \int_{I_n^\ell} \delta dP_n(t,\delta,z) \ge n^{-1/2} (\log n)^{-(m+1)} \right\} \le \sqrt{n} (\log n)^{(m+1)} \int_{I_n^\ell} \delta dP_0(t,\delta,z)$$
$$= \sqrt{n} (\log n)^{(m+1)} \int_{I_n^\ell} \left((1-p)F_0(t) g_0(t) + p F_0(t)^{1+\theta_n} g_1(t)\right) dt \to 0, \; n \to \infty.$$

For the second term we have:

$$
\begin{aligned}
II &\le \frac{\sqrt{n}}{1-\hat F_n^0\left(n^{-1/3}(\log n)^2\right)} \int_{I_n^\ell} w\left(\hat F_n^0(t)\right) dP_n(t,\delta,z)\\
&\le C_5\sqrt{n}\int_{I_n^\ell}\left(\hat F_n^0(t)+\hat F_n^0(t)\left|\log \hat F_n^0(t)\right|^m\right) dP_n(t,\delta,z)+o_p(1)\\
&\le C_6\sqrt{n}(\log n)^m\int_{I_n^\ell}\hat F_n^0(t)dP_n(t,\delta,z)+o_p(1)\\
&\le C_6\sqrt{n}(\log n)^m \hat F_n^0\left(n^{-1/3}(\log n)^2\right)\int_{I_n^\ell} dP_n(t,\delta,z)+o_p(1)=o_p(1),
\end{aligned}
$$

by an application of (2.2.5), Lemmas 2.4.4 and 2.4.6, and the Markov inequality.

For the terms III and IV the proof is similar to the proof for the terms I and II:

$$
\begin{aligned}
III &\le C_7\frac{\sqrt{n}}{\hat F_n^0\left(1-n^{-1/3}(\log n)^2\right)}\int_{I_n^r}\left(1-\hat F_n^0(t)\right)dP_n(t,\delta,z)\\
&\le C_8\frac{\sqrt{n}\left\{1-\hat F_n^0(1-n^{-1/3}(\log n)^2)\right\}}{\hat F_n^0\left\{1-n^{-1/3}(\log n)^2\right\}}\int_{I_n^r} dP_n(t,\delta,z)=o_p(1)
\end{aligned}
$$

and

$$
IV \le C_9\sqrt{n}\int_{I_n^r}(1-\delta)dP_n(t,\delta,z)=o_p(1).
$$

■

We are now ready to give the outlined proof of Theorem 2.1.1. Let the functional $[R(F)]$ be defined by

$$
[R(F)](t,\delta,z)=pw(F(t))\left(\frac{\delta}{F(t)}-\frac{1-\delta}{1-F(t)}\right).
$$

We can then write the statistic S_n as

$$
\begin{aligned}
S_n &= \sqrt{n}\int_{I_n}\left[S(\hat F_n^0)\right](t,\delta,z)dP_n(t,\delta,z)+o_p(1)\\
&= \sqrt{n}\int_{I_n}\left(\left[S\left(\hat F_n^0\right)\right](t,\delta,z)-\left[S\left(F_0\right)\right](t,\delta,z)\right)d(P_n-P_0)(t,\delta,z)1_{\hat F_n^0\in\mathcal{F}_n} \quad (2.2.7)\\
&\quad -\sqrt{n}\int_{I_n}\left(\left[R\left(\hat F_n^0\right)\right](t,\delta,z)-\left[R\left(F_0\right)\right](t,\delta,z)\right)d(P_n-P_0)(t,\delta,z)1_{\hat F_n^0\in\mathcal{F}_n} \quad (2.2.8)\\
&\quad +\sqrt{n}\int_{I_n}\left(\left[S\left(\hat F_n^0\right)\right](t,\delta,z)-\left[R\left(\hat F_n^0\right)\right](t,\delta,z)\right)dP_0(t,\delta,z)1_{\hat F_n^0\in\mathcal{F}_n} \quad (2.2.9)\\
&\quad +\sqrt{n}\int_{I_n}\left([S(F_0)](t,\delta,z)-[R(F_0)](t,\delta,z)\right)d(P_n-P_0)(t,\delta,z)1_{\hat F_n^0\in\mathcal{F}_n} \quad (2.2.10)\\
&\quad +o_p(1),
\end{aligned}
$$

where we used that, by Lemma 2.4.3 in section 2.4,

$$
\int\left[R\left(\hat F_n^0\right)\right](t,\delta,z)dP_n(t,\delta,z)=0
$$

and that, as in Lemma 2.2.2,

$$\sqrt{n}\int_{I_n^c}\left[R\left(\hat{F}_n^0\right)\right](t,\delta,z)dP_n(t,\delta,z)=o_p(1).$$

Now convergence $S_n \xrightarrow{\mathcal{D}} \mathcal{N}(\mu,\sigma^2)$ follows from the Lemmas 2.5.1, 2.5.2 and 2.6.2. Note that the first two terms in the final representation are so-called Donsker terms that are shown to be asymptotically small, using the analog of the equicontinuity lemma in POLLARD (1984). The last two terms (say N_n) are the non-negligible Central Limit terms and will also be treated in Lemma 2.5.2. It is perhaps not immediately clear why the (2.2.9) term is a Central Limit term (if $\theta_0 = 0$ and $G_0 = G_1$ it is simply equal to zero by the independence of Z and (T,δ)), but this will be explained in Lemma 2.5.1. ∎

2.3 Asymptotic efficiency of the score test

In this section we will show that the score test is asymptotically efficient. The proof relies on the theory of smooth functionals.

The model we consider has the following form:

$$\mathcal{P}=\{(\tilde{F}_0,\tilde{G}_0,\tilde{G}_1,\tilde{\theta},\tilde{p}),\ \text{where } \tilde{F}_0,\tilde{G}_0,\tilde{G}_1 \text{ are distribution functions},\ \tilde{\theta}\in(-1,\infty),\ \tilde{p}\in(0,1)\}.$$

The functional we have to estimate (and for which we have to show "smoothness") is:

$$\psi(\tilde{F}_0,\tilde{G}_0,\tilde{G}_1,\tilde{\theta},\tilde{p})=\tilde{\theta},$$

and we have to study its behavior in the neighborhood of the fixed point:

$$P^0=(F_0,G_0,G_1,0,p).$$

Take the dominating measure μ to be:

$$\mu=(\text{Lebesgue measure on }\mathbb{R})\times(\text{Counting measure on }\{0,1\})^2.$$

For $P=(\tilde{F}_0,\tilde{G}_0,\tilde{G}_1,\tilde{p},\tilde{\theta})$ the density of the triple (T,δ,Z) w.r.t. μ is given by

$$f_P(t,\delta,z)=\left\{\delta\tilde{F}_0(t)^{1+z\tilde{\theta}}+(1-\delta)\left(1-\tilde{F}_0(t)^{1+z\tilde{\theta}}\right)\right\}\{(1-z)(1-\tilde{p})\tilde{g}_0(t)+z\tilde{p}\tilde{g}_1(t)\}.$$

We assume here that the densities $\tilde{g}_0$ and $\tilde{g}_1$ exist.

Our first step is to find the tangent space corresponding to the model. It is given in the following lemma.

Lemma 2.3.1 *Each element of the tangent space of the family $\mathcal{P}$ at the point P^0 can be represented as*

$$\begin{aligned}a(t,\delta,z)&=(1-z)a_0(t)+za_1(t)+\left(\frac{z}{p}-\frac{1-z}{1-p}\right)p'\\&+\left(z\theta' F_0(t)\log F_0(t)+h(t)\right)\left(\frac{\delta}{F_0(t)}-\frac{1-\delta}{1-F_0(t)}\right),\end{aligned}$$

where

- $a_0 \in L_0^2(G_0)$
- $a_1 \in L_0^2(G_1)$
- $\theta', p' \in \mathbb{R}$
- $\sup_{t \in \text{support}(F_0)} |h(t)| < \infty$, $h|_{\partial \text{support}(F_0)} = 0$

  and $h^2/\{F_0(1-F_0)\} \in L^1(G_0) \bigcap L^1(G_1)$.

The lemma gives us that if the path f^s is differentiable, then the one-dimensional parameters θ^s and p^s are differentiable in the usual sense, as real-valued functions of s, and the paths g_0^s and g_1^s are Hellinger differentiable.

The proof of this lemma uses standard techniques and is therefore omitted. One can check that the usual interpretation of the tangent – the logarithmic derivative of the family of densities along the path – remains true here.

Lemma 2.3.2 *The canonical gradient (or "efficient influence function") $\tilde{k}$, corresponding to the model $\mathcal{P}$ at the point P^0 is equal to*

$$\tilde{k}(t,\delta,z) = C\left(\frac{\delta}{F_0(t)} - \frac{1-\delta}{1-F_0(t)}\right)\left(z - \frac{pg_1(t)}{pg_1(t)+(1-p)g_0(t)}\right)F_0(t)\log F_0(t),$$

where

$$C = \left(p(1-p)\int_{\text{support}(F_0)} \frac{F_0(t)(\log F_0(t))^2}{1-F_0(t)} \cdot \frac{g_0(t)g_1(t)}{pg_1(t)+(1-p)g_0(t)}dt\right)^{-1}$$

Proof:

Since the canonical gradient is a unique element of the tangent space we write

$$\begin{aligned}\tilde{k}(t,\delta,z) &= (1-z)\tilde{a}_0(t) + z\tilde{a}_1(t) + C_1\left(\frac{z}{p} - \frac{1-z}{1-p}\right) \\ &\quad + \left(zC_2F_0(t)\log F_0(t) + \tilde{h}(t)\right)\left(\frac{\delta}{F_0(t)} - \frac{1-\delta}{1-F_0(t)}\right),\end{aligned}$$

where $\tilde{a}_0$, $\tilde{a}_1$ and $\tilde{h}$ are as in Lemma 2.3.1. The derivative $\left.\frac{\partial \psi(P^s)}{\partial s}\right|_{s=0}$ in the "direction" a, has the representation as an inner product $\langle a, \tilde{k}\rangle_{P^0}$. We write this in the following form:

$$\begin{aligned}&\left.\frac{\partial}{\partial s}\theta^s\right|_{s=0} = \\ &\left\langle (1-z)\tilde{a}_0 + z\tilde{a}_1 + C_1\left(\frac{z}{p} - \frac{1-z}{1-p}\right), (1-z)a_0 + za_1 + \left.\frac{\partial}{\partial s}p^s\right|_{s=0}\left(\frac{z}{p} - \frac{1-z}{1-p}\right)\right\rangle_{P^0} \\ &+ \left\langle \left(zC_2F_0\log F_0 + \tilde{h}\right)\frac{\delta - F_0}{F_0\{1-F_0\}}, \left(z\left.\frac{\partial}{\partial s}\theta^s\right|_{s=0}F_0\log F_0 + h\right)\frac{\delta - F_0}{F_0\{1-F_0\}}\right\rangle_{P^0},\end{aligned}$$

since the integral of $\{\delta - F_0\}/\{F_0(1-F_0)\}$, multiplied by a function that does not depend on δ, is zero when integrated w.r.t. the measure dP^0.

The first inner product is independent of $\frac{\partial}{\partial s}\theta^s\big|_{s=0}$ and must therefore be equal to zero. The choice $a_0 = \tilde{a}_0$, $a_1 = 0$, $\frac{\partial}{\partial s}p^s\big|_{s=0} = 0$ gives

$$\left\langle (1-z)\tilde{a}_0 + z\tilde{a}_1 + C_1\left(\frac{z}{p} - \frac{1-z}{1-p}\right), (1-z)\tilde{a}_0 \right\rangle_{P^0} =$$
$$\int \left\{(1-z)\tilde{a}_0^2 - C_1\tilde{a}_0\frac{1-z}{1-p}\right\} dP^0(t,\delta,z) = (1-p)\int \tilde{a}_0(t)^2 g_0(t)dt = 0,$$

and hence $\tilde{a}_0 = 0$. Similarly, $\tilde{a}_1 = 0$ and $C_1 = 0$.

The second inner product must be equal to $\frac{\partial}{\partial s}\theta^s\big|_{s=0}$, and hence we should have:

$$\frac{\partial}{\partial s}\theta^s\Big|_{s=0}\left\langle \left\{C_2 z\, F_0\log F_0 + \tilde{h}\right\}\frac{\delta - F_0}{F_0\{1-F_0\}}, \frac{z\{\delta - F_0\}F_0\log F_0}{F_0\{1-F_0\}} \right\rangle_{P^0}$$
$$+\left\langle \left\{C_2 z\, F_0\log F_0 + \tilde{h}\right\}\frac{\delta - F_0}{F_0\{1-F_0\}}, \frac{\{\delta - F_0\}h}{F_0\{1-F_0\}} \right\rangle_{P^0}$$
$$= \frac{\partial}{\partial s}\theta^s\Big|_{s=0}$$

and hence for any h as in Lemma 2.3.1 the second inner product should be zero. This means

$$\int \left\{C_2 zF_0(t)\log F_0(t) + \tilde{h}(t)\right\}\frac{\{\delta - F_0\}^2}{F_0^2\{1-F_0\}^2}h(t)\, dP^0(t,\delta,z)$$
$$= \int \frac{h(t)}{F_0(t)(1-F_0(t))}\left\{\tilde{h}(t)((1-p)g_0(t) + pg_1(t)) + C_2 pg_1(t)F_0(t)\log F_0(t)\right\} dt = 0.$$

Taking

$$h(t) = \tilde{h}(t) + C_2\frac{pg_1(t)}{(1-p)g_0(t) + pg_1(t)}F_0(t)\log F_0(t),$$

the above can only be satisfied if

$$\tilde{h}(t) = -C_2\frac{pg_1(t)}{(1-p)g_0(t) + pg_1(t)}F_0(t)\log F_0(t).$$

To find the constant C_2 we use condition on the first inner product above:

$$\left\langle \left(zC_2F_0\log F_0 + \tilde{h}\right)\frac{\delta - F_0}{F_0\{1-F_0\}}, zF_0\log F_0\frac{\delta - F_0}{F_0\{1-F_0\}} \right\rangle_{P^0}$$
$$= \left\langle \left(z - \frac{pg_1}{(1-p)g_0 + pg_1}\right)\frac{\delta - F_0}{F_0\{1-F_0\}}C_2F_0\log F_0, zF_0\log F_0\frac{\delta - F_0}{F_0\{1-F_0\}} \right\rangle_{P^0}$$
$$= 1.$$

That is

$$C_2\int z\left(1 - \frac{pg_1(t)}{(1-p)g_0(t) + pg_1(t)}\right)\frac{\{\delta - F_0(t)\}^2}{\{1-F_0(t)\}^2}\log^2 F_0(t)\, dP^0(t,\delta,z)$$
$$= C_2 p(1-p)\int \frac{F_0(t)\log^2 F_0(t)}{1-F_0(t)}\frac{g_0(t)g_1(t)}{(1-p)g_0(t) + pg_1(t)}dt = 1,$$

which finishes the proof of Lemma 2.3.2. ∎

To prove the next lemma we need the Theorem 25.44 of VAN DER VAART (1998)

Theorem 2.3.1 *(van der Vaart)*
Let the functional $\psi : \mathcal{P} \to \mathbb{R}$ be differentiable at P^0 relative to the tangent space T with the canonical gradient $\tilde{k}$. Suppose $\psi(P^0) = 0$. Than for every sequence of power functions $P \to \pi_n(P)$ of level-α tests for $H_0 : \psi(P) \le 0$, and every Hellinger differentiable path P^s with the corresponding tangent $a \in T$, $\langle \tilde{k}, a\rangle_{P^0} > 0$ and every $h > 0$,

$$\limsup_{n\to\infty} \pi_n\left(P^{h/\sqrt{n}}\right) \le 1 - \Phi\left(z_{1-\alpha} - h\frac{\langle \tilde{k}, a\rangle_{P^0}}{\sqrt{\langle \tilde{k}, \tilde{k}\rangle_{P^0}}}\right), \tag{2.3.1}$$

where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution.

By the same arguments the asymptotic power for the two-sided alternative $H_1 : \psi(P) \ne 0$ satisfies

$$\limsup_{n\to\infty} \pi_n\left(P^{h/\sqrt{n}}\right) \le 1 - \Phi\left(z_{1-\frac{\alpha}{2}} - h\frac{\langle \tilde{k}, a\rangle_{P^0}}{\sqrt{\langle \tilde{k}, \tilde{k}\rangle_{P^0}}}\right) + \Phi\left(z_{\frac{\alpha}{2}} - h\frac{\langle \tilde{k}, a\rangle_{P^0}}{\sqrt{\langle \tilde{k}, \tilde{k}\rangle_{P^0}}}\right). \tag{2.3.2}$$

Lemma 2.3.3 *For $w(t) = t\log t$ the score test (based on the statistic S_n) of H_0 against H_1 is (locally) asymptotically efficient.*

Proof: We have to prove that the asymptotical (local) power function of the score test attains the upper bound, given by (2.3.2). The alternative distribution $P^{1/\sqrt{n}}$ is in our case given by

$$P^{1/\sqrt{n}} = \left(F_0, G_0, G_1, \theta_n = \frac{\theta_0}{\sqrt{n}}, p\right),$$

which can be considered as the point of the Hellinger differentiable path

$$P^s = (F_0, G_0, G_1, \theta_n = \theta_0 s, p).$$

taking the value P^0 at $s = 0$. According to Lemma 2.3.1 the corresponding tangent is:

$$a(t, \delta, z) = z\theta_0\left(\frac{\delta}{F_0(t)} - \frac{1-\delta}{1-F_0(t)}\right) F_0(t)\log F_0(t).$$

Calculating the value of $\langle \tilde{k}, a\rangle_{P^0}/\sqrt{\langle \tilde{k}, \tilde{k}\rangle_{P^0}}$ in (2.3.2), we obtain

$$\begin{aligned}
&\langle \tilde{k}, a\rangle_{P^0} = \\
&C\theta_0 \int_{\text{support}(F_0)} zF_0(t)^2 \log^2 F_0(t)^2 \frac{\{\delta - F_0(t)\}^2}{F_0(t)^2\{1-F_0(t)\}^2} \frac{(1-p)g_0(t)}{(1-p)g_0(t) + pg_1(t)}\, dP_0(t,\delta,z) = \\
&C\theta_0 p(1-p) \int_{\text{support}(F_0)} \frac{F_0(t)(\log F_0(t))^2}{1-F_0(t)} \frac{g_0(t)g_1(t)}{(1-p)g_0(t)+pg_1(t)} dt = \theta_0,
\end{aligned}$$

where the constant C is the same as in the Lemma 2.3.2. It also follows from this lemma that

$$\langle \tilde{k}, \tilde{k}\rangle_{P^0} = p(1-p)\int_{\text{support}(F_0)} \frac{F_0(t)(\log F_0(t))^2}{1-F_0(t)} \frac{g_0(t)g_1(t)}{(1-p)g_0(t)+pg_1(t)} dt,$$

and hence

$$\frac{\langle \tilde{k}, a\rangle_{P^0}}{\sqrt{\langle \tilde{k}, \tilde{k}\rangle_{P^0}}} = \theta_0 \left\{ p(1-p) \int_{\text{support}(F_0)} \frac{F_0(t)(\log F_0(t))^2}{1-F_0(t)} \frac{g_0(t)g_1(t)}{(1-p)g_0(t)+pg_1(t)} dt \right\}^{-1/2}.$$

If $w(t) = t\log t$, the ratio μ/σ of the parameters of the limiting distribution of the statistic S_n, given in the main theorem, is equal to this value and hence the (local) asymptotic power of the score test attains the upper bound given by (2.3.2). ■

2.4 Properties of the maximum likelihood estimator

In this section we discuss properties of the maximum likelihood estimator $\hat{F}_n^0$ defined by (2.1.3) under conditions of the Main Theorem. As mentioned in the introduction, the results presented here are mostly generalizations of results in VAN DE GEER (2000) and GROENEBOOM AND WELLNER (1992).

First of all we note that the relation (2.1.3) only defines $\hat{F}_n^0$ in the points T_i and for this reason the statistic S_n depends on $\hat{F}_n^0$ only via its values $\hat{F}_n^0(T_i)$, $i = 1, \ldots, n$. We define $\hat{F}_n^0$ to be the piecewise constant right continuous function, with jumps only at the points T_i, which maximizes $L_n(F, 0)$. We put $\hat{F}_n^0(t) = 0$ for any $t < T_{(1)}$. This may sometimes result in a defective distribution function, but this does not cause a real problem, as is clear from the following lemma.

Lemma 2.4.1 *Under the conditions of Theorem 2.1.1:*

$$P\left\{\hat{F}_n^0 \text{ is defective}\right\} \to 0, \quad n \to \infty.$$

Proof:

$$P\left\{\lim_{x\to\infty} \hat{F}_n^0(x) = 1\right\} = P\left\{\hat{F}_n(T_{(n)}) = 1\right\} = P\left\{\Delta_{(n)} = 1\right\},$$

since, if $\hat{F}_n(T_{(n)}) = 1$ and $\Delta_{(n)} = 0$, the log likelihood would be equal to minus infinity. Here and in the sequel the lower index (i) in $\Delta_{(i)}$ indicates that the triples are ordered w.r.t. the order statistics $T_{(i)}$, following a convention, introduced in GROENEBOOM AND WELLNER (1992).

To show that $P\left\{\Delta_{(n)} = 1\right\}$ tends to one, we note that, from the conditions on the intervals of support in Theorem 2.1.1, there exists an increasing sequence of numbers α_n, satisfying: $P\left\{T_{(n)} < \alpha_n\right\} \to 0$ and $F_0(\alpha_n) \to 1,\ n \to \infty$. This implies:

$$\begin{aligned} P\left\{\Delta_{(n)} = 0\right\} &\le P\left\{\Delta_{(n)} = 0 \mid T_{(n)} \ge \alpha_n\right\} P\left\{T_{(n)} \ge \alpha_n\right\} + P\left\{T_{(n)} < \alpha_n\right\} \\ &\le 2 - \left(F_0(\alpha_n) + F_0(\alpha_n)^{1+\theta_n}\right) + P\left\{T_{(n)} < \alpha_n\right\} \to 0,\ n \to \infty. \end{aligned}$$

■

The next lemma gives a unicity property of the MLE.

Lemma 2.4.2 *There exists only one piecewise constant right continuous function with jumps at the points T_i satisfying (2.1.3).*

Proof: We use the alternative definition (2.1.5). The proof of the lemma will then follow from the strict concavity of $l(0, x, \delta, z)$ w.r.t. x.

Assume that there are two different solutions of (2.1.5): $F_{1,n}$ and $F_{2,n}$, take $F_n(t) = \frac{1}{2}(F_{1,n}(t) + F_{2,n}(t))$. This will mean

$$\int l(0, F_n(t), \delta, z) dP_n(t, \delta, z)$$
$$> \frac{1}{2} \int l(0, F_{1,n}(t), \delta, z) dP_n(t, \delta, z) + \frac{1}{2} \int l(0, F_{2,n}(t), \delta, z) dP_n(t, \delta, z).$$

But the last two terms must be equal. This leads to a contradiction, since then $F_{1,n}$ and $F_{2,n}$ cannot be solutions of (2.1.5). ■

The next lemma provides the necessary condition for $\hat{F}_n^0$ to be the solution of (2.1.5).

Lemma 2.4.3 *For any function $a : I\!R \to I\!R$, constant on the same intervals as $\hat{F}_n^0$,*

$$\int a(t) l_2(0, \hat{F}_n^0(t), \delta, z) 1_{\hat{F}_n^0(t) \in (0,1)} dP_n(t, \delta, z) = 0. \tag{2.4.1}$$

Proof:
Let $\{[\tau_i, \tau_{i+1})\}_{i=0}^N$ be the system of intervals where $\hat{F}_n^0$ is piecewise constant, where $\tau_0 = 0$ and $\tau_{N+1} = 1$. Note that $\{\tau_i\}_{i=1}^N \subset \{T_i\}_{i=1}^n$.

Assume that $\hat{F}_n^0$ is non-degenerate. Consider $F_i(t, u) = \hat{F}_n^0(t) + u 1_{t \in [\tau_i, \tau_{i+1})}$, $i = 1, \ldots, N-1$. There exists an $\epsilon > 0$ such that for any u, satisfying $|u| \le \epsilon$, $t \mapsto F_i(t, u)$ is a distribution function. Hence

$$\int l(0, F_i(t, u), \delta, z) dP_n(t, \delta, z) \le \int l(0, \hat{F}_n^0(t), \delta, z) dP_n(t, \delta, z).$$

We now obtain (2.4.1) by taking derivative w.r.t. u. ■

Another important property of $\hat{F}_n^0$ we will need is that it can, for fixed n, not be arbitrarily close to zero or one, without actually becoming equal to zero or one, respectively. This property is somewhat analogous to a corresponding property of a 1-dimensional empirical distribution function.

Lemma 2.4.4

$$\hat{F}_n^0(T) \ge \frac{1}{n} 1_{\{\hat{F}_n^0(T) > 0\}} \text{ and } 1 - \hat{F}_n^0(T) \ge \frac{1}{n} 1_{\{\hat{F}_n^0(T) < 1\}}.$$

Remark 2.4.1 Note that analogously to Lemma 2.4.1, Lemma 2.4.4 can also be formulated as

$$\hat{F}_n^0(T) \ge \frac{\Delta}{n},\ 1 - \hat{F}_n^0(T) \ge \frac{1 - \Delta}{n}.$$

Proof of Lemma 2.4.4: The two statements of Remark 2.4.1 are symmetric and we therefore only prove $\hat{F}_n^0(T_j) > \Delta_j / n$. If $\Delta_j = 0$ or $\hat{F}_n^0(T_j) = 1$ the statement is trivial. Otherwise $0 < \hat{F}_n^0(T_j) \le 1$. Using Lemma 2.4.3, with $a(t) = 1_{\{t = \hat{F}_n^0(T_j)\}}$, we obtain

$$0 = \sum_{i:\ \hat{F}_n^0(T_i) = \hat{F}_n^0(T_j)} \left(\frac{\Delta_i}{\hat{F}_n^0(T_i)} - \frac{1 - \Delta_i}{1 - \hat{F}_n^0(T_i)} \right) \ge \frac{1}{\hat{F}_n^0(T_j)} - \frac{n-1}{1 - \hat{F}_n^0(T_j)},$$

and hence $\hat{F}_n^0(T_j) \geq 1/n$. ∎

The following lemma relates the maximum likelihood estimator $\hat{F}_n^0$ to the maximum likelihood estimators of F_0 and $F_0^{1+\theta_n}$, respectively, not using the restriction of the null hypothesis for these distribution functions.

Lemma 2.4.5 *Let the triples $\{(T_i, \Delta_i, Z_i)\}_{i=1}^n$ be defined as before. Then we can define on the same probability space i.i.d. random pairs $\{(T_i, \Delta_i^0)\}_{i=1}^n$ and $\{(T_i, \Delta_i^1)\}_{i=1}^n$ such that the T_i's are the same as the T_i in the triples and such that Δ_i^j, $j = 0, 1$, are indicators satisfying*

$$P\{\Delta_i^j = 1 | \; T_i = t\} = F_0(t)^{1+j\theta_n}.$$

Let $\hat{F}_n^0$ be as in (2.1.3) and let $\hat{F}_{0,n}$, $\hat{F}_{1,n}$ be functions satisfying

$$\hat{F}_{j,n} = \underset{F}{\operatorname{argmax}} \prod_{i=1}^n L(0, F)(T_i, \Delta_i^j, 0), \; j = 0, 1. \tag{2.4.2}$$

Then, if $\theta_0 \geq 0$, we have for all t

$$\hat{F}_{1,n}(t) \leq \hat{F}_n^0(t) \leq \hat{F}_{0,n}(t)$$

with reversed inequalities if $\theta_0 < 0$.

The usefulness of this lemma is that $\hat{F}_{0,n}$ and $\hat{F}_{1,n}$ are the maximum likelihood estimators of F_0 and $F_0^{1+\theta_n}$ in the usual current status problem. These maximum likelihood estimators are well studied and we are going to show uniformity of the results w.r.t. the parametric class $\{F_0^{1+\theta};\ \theta \in (-1, \infty)\}$. We will then use the fact that $F_0(t)^{1+\theta_n}$ is sufficiently close to F_0 for obtaining a bound on the supremum distance between $\hat{F}_n^0$ and F_0.

Proof: Let $\{(T_i', U_i', Z_i')\}_{i=1}^n$ be i.i.d. triples such that U_i' is a Uniform(0,1) random variable, independent of the other two elements of the triple, Z_i' a Bernoulli(p) random variable, and such that T_i' is distributed according to

$$P\{T_i' \leq t | \; Z_i' = z\} = G_z(t).$$

We consider the sequence of triples $\{(T_i, \Delta_i, Z_i)\}$ as the image of the sequence above

$$T_i = T_i'; \; \Delta_i = \{U_i \leq F_0(T_i')^{1+Z_i'\theta_n}\}; \; Z_i = Z_i'; \; i = 1, \ldots, n,$$

where we write $\{\cdot\}$ instead of $1_{\{\cdot\}}$. Now take

$$T_i^0 = T_i'; \; \Delta_i^0 = \{U_i \leq F_0(T_i')\}; \; i = 1, \ldots, n,$$

and

$$T_i^1 = T_i'; \; \Delta_i^1 = \{U_i \leq F_0(T_i')^{1+\theta_n}\}; \; i = 1, \ldots, n.$$

Note that if $\theta_n \leq 0$, we have $\Delta_i^0 \leq \Delta_i \leq \Delta_i^1$, with reversed inequalities if $\theta_n > 0$. It is therefore enough to prove that if there are two sets of pairs $\{(T_i, \Delta_i^0)\}_{i=1}^n$ and $\{(T_i, \Delta_i^1)\}_{i=1}^n$ for which there exists J, $1 \leq J \leq n$, satisfying

$$\begin{aligned} &\delta_i^0 = \delta_i^1, && i \neq J, \\ &\delta_i^0 = 0, \; \delta_i^1 = 1, && i = J, \end{aligned}$$

then the functions $\hat{F}_n^j$, defined as in (2.4.2), satisfy:

$$\hat{F}_{0,n}(t) \le \hat{F}_{1,n}(t), \forall t \ge 0. \tag{2.4.3}$$

We first prove that in the point $t = t_J$ the inequality (2.4.3) holds. This is trivial if $\hat{F}_{0,n}(t_J) = 0$. To obtain it for $\hat{F}_{0,n}(T_J) > 0$, use the definitions of $\hat{F}_{0,n}$ and $\hat{F}_{1,n}$:

$$\begin{cases} \prod_{i=1}^n \left[L(0,\hat{F}_{1,n})\right](t_i,\delta_i^1,0) \ge \prod_{i=1}^n \left[L(0,\hat{F}_{0,n})\right](t_i,\delta_i^1,0), \\ \prod_{i=1}^n \left[L(0,\hat{F}_{1,n})\right](t_i,\delta_i^0,0) \le \prod_{i=1}^n \left[L(0,\hat{F}_{0,n})\right](t_i,\delta_i^0,0). \end{cases}$$

Dividing the first inequality by the second one we obtain

$$\frac{\hat{F}_{1,n}(t_J)}{1-\hat{F}_{1,n}(t_J)} \ge \frac{\hat{F}_{0,n}(t_J)}{1-\hat{F}_{0,n}(t_J)}.$$

Thus $\hat{F}_{0,n}(t_J) \le \hat{F}_{1,n}(t_J)$, since $x \mapsto x/(1-x)$ is an increasing function on $(0,1)$.

Now take $F_n(t) = \hat{F}_{0,n}(t) \bigvee \hat{F}_{1,n}(t)$. By definition

$$\prod_{i=1}^n \left[L(0,\hat{F}_{1,n})\right](t_i,\delta_i^1,0) \ge \prod_{i=1}^n \left[L(0,F_n)\right](t_i,\delta_i^1,0).$$

Divide the last inequality by

$$\prod_{\{i:\hat{F}_{1,n}(t_i)\ge\hat{F}_{0,n}(t_i)\}} \left(\left[L(0,\hat{F}_{1,n})\right](t_i,\delta_i^1,0) \Big/ \left[L(0,\hat{F}_{0,n})\right](t_i,\delta_i^0,0)\right).$$

Note that except for $i = J$ all delta's are the same. Hence we obtain the inequality

$$\prod_{i=1}^n \left[L(0,\hat{F}_{1,n} \bigwedge \hat{F}_{0,n})\right](t_i,\delta_i^0,0) \ge \prod_{i=1}^n \left[L(0,\hat{F}_{0,n})\right](t_i,\delta_i^0,0),$$

and thus, by Lemma 2.4.2, $\hat{F}_{0,n} = \hat{F}_{0,n} \bigwedge \hat{F}_{1,n}$. ∎

The next lemma shows that the supremum distance between $\hat{F}_n^0$ and the underlying F_0 is $O_p(n^{-1/3}\log n)$.

Lemma 2.4.6 *Under conditions of the Theorem 2.2.1*

$$P\left\{\sup_{t\in[0,1]} \left|\hat{F}_n^0(t) - F_0(t)\right| \ge n^{-1/3}\log n\right\} \to 0,\ n \to \infty. \tag{2.4.4}$$

Proof: Using Lemma 2.4.5 we only have to obtain the supremum distance for $\hat{F}_{0,n}$ and $\hat{F}_{1,n}$. But since these are the nonparametric maximum likelihood estimators (MLE's) for interval censoring, case I, we can apply results from Groeneboom and Wellner (1992). $\hat{F}_{0,n}$ is the

MLE of the distribution F_0 censored by the distribution $(1-p)G_0(t)+pG_1(t)$. Thus by Lemma 5.9, GROENEBOOM AND WELLNER (1992), p. 116,

$$P\left\{\sup_{t\in[0,1]}\left|\hat{F}_{0,n}(t)-F_0(t)\right|\geq n^{-1/3}\log n\right\}\to 0,\ n\to\infty. \tag{2.4.5}$$

To obtain the analogous result for $\hat{F}_{1,n}$ we have to take a closer look at the proof of Lemma 5.9 in GROENEBOOM AND WELLNER (1992). The estimated distribution $F_0^{1+\theta_n}$ in that case can have a density that is not bounded from above (if $\theta_0<0$) or a density that is not bounded away from zero (if $\theta_0>0$). But to obtain the inequality analogous to (2.4.5), it is sufficient to have

$$0<\inf_{n;t\geq n^{-1/3}\log n} f_n(t)\leq \sup_{n;t\geq n^{-1/3}\log n} f_n(t)<\infty, \tag{2.4.6}$$

where $F_n(t)=F_0(t)^{1+\theta_n}$ and f_n is the corresponding density $f_n(t)=(1+\theta_n)f_0(t)F_0(t)^{\theta_n}$. Relation (2.4.6) follows from:

$$\inf_{t\geq n^{-1/3}\log n} f_n(t)\geq (1+\theta_n)\left(\inf f_0\right)\left(n^{-1/3}\log n(\inf f_0)\right)^{|\theta_n|}\geq C_1e^{-C_2n^{-1/2}\log n}>0,$$

and

$$\sup_{t\geq n^{-1/3}\log n} f_n(t)\leq (1+\theta_n)\left(\sup f_0\right)\left(n^{-1/3}\log n(\inf f_0)\right)^{-|\theta_n|}\leq C_3e^{C_4n^{-1/2}\log n}<\infty,$$

uniformly in n.

In the proof of Lemma 5.9 in GROENEBOOM AND WELLNER (1992) the following inequality is obtained:

$$P\left\{\left|\hat{F}_n^{-1}(a_i)-F^{-1}(a_i)\right|>n^{-1/3}\log n\right\}\leq C_5\exp\left(-C_6(\log n)^2\right),\ i=1,\ldots,m_n-1,$$

where F is a distribution on $[0,1]$ with $\inf f>0$, censored by the distribution G with density g, satisfying $0<\inf g\leq\sup g<\infty$, $\hat{F}_n$ is the corresponding MLE and where $\hat{F}_n^{-1}$ is the pseudo-inverse

$$\hat{F}_n^{-1}(a)=\sup\left\{t\in[0,1]:\ \hat{F}_n(t)\leq a\right\},$$

and where the sequence a_i is defined as

$$a_i=in^{-1/3}\log n,\ i=1,\ldots,m_n-1;\quad m_n=[n^{1/3}/\log n]\,(=\text{the integer part of } n^{1/3}/\log n).$$

The constants C_5 and C_6 are proved to be absolute constants only depending on the infimum and supremum of densities.

More precisely the constants C_5 and C_6 are dependent on the supremum and infimum on the interval $\left[a_1-n^{-1/3}\log n;a_{m_n-1}+n^{-1/3}\log n\right]$ and hence, using (2.4.6), we can similarly prove

$$P\left\{\left|\hat{F}_{1,n}^{-1}(a_i)-F_n^{-1}(a_i)\right|>n^{-1/3}\log n\right\}\leq C_5\exp\left(-C_6(\log n)^2\right),\ i=2,\ldots,m_n-2.$$

Now fix $C_7 = 2\sup_{a>n^{-1/3}\log n} \left(F_n^{-1}\right)'(a) + 1 < \infty$. Then by monotonicity

$$P\left\{\sup_{a\in[a_i,a_{i+1})}\left|\hat{F}_{1,n}^{-1}(a) - F_n^{-1}(a)\right| > C_7 n^{-1/3}\log n\right\}$$
$$\leq 2C_5\exp(-C_6(\log n)^2), \quad i = 2,\ldots,m_n-2,$$

and

$$P\left\{\sup_{a\in[a_{m_n-1},1]}\left|\hat{F}_{1,n}^{-1}(a) - F_n^{-1}(a)\right| > C_7 n^{-1/3}\log n\right\} \leq 2C_5\exp(-C_6(\log n)^2),$$

which means

$$P\left\{\sup_{a\geq a_2}\left|\hat{F}_{1,n}^{-1}(a) - F_n^{-1}(a)\right| > C_7 n^{-1/3}\log n\right\} \leq 2C_5 n^{1/3}\exp(-C_6(\log n)^2). \tag{2.4.7}$$

Let

$$C_8 = C_7 \sup_{t\geq n^{-1/3}\log n} f_n(t),$$

and suppose that for $t > \alpha_n \stackrel{\text{def}}{=} F_n^{-1}(a_2) + (2 + C_7)n^{-1/3}\log n$:

$$\left|\hat{F}_{1,n}(t) - F_n(t)\right| > C_8 n^{-1/3}\log n.$$

We will then find an $a > a_2$ such that $\left|\hat{F}_{1,n}^{-1}(a) - F_n^{-1}(a)\right| > C_7 n^{-1/3}\log n$, implying, by (2.4.7),

$$P\left\{\sup_{t\geq\alpha_n}\left|\hat{F}_{1,n}(t) - F_n(t)\right| > C_8 n^{-1/3}\log n\right\} \to 0,\ n\to\infty. \tag{2.4.8}$$

The choice of a depends on which is larger: $\hat{F}_{1,n}(t)$ or $F_n(t)$. Suppose

$$\hat{F}_{1,n}(t) - F_n(t) > C_8 n^{-1/3}\log n$$

take then

$$\tilde{a} = \hat{F}_{1,n}(t) > F_n(t) > F_n(F_n^{-1}(a_2)) = a_2.$$

By Taylor expansion for $\xi_1 \in \left[F_n(t), \hat{F}_{1,n}(t)\right]$

$$\begin{aligned} F_n^{-1}(\tilde{a}) - t &= F_n^{-1}\left(F_n(t) + \left(\hat{F}_{1,n}(t) - F_n(t)\right)\right) - t = \left(F_n^{-1}\right)'(\xi_1)\left(\hat{F}_{1,n}(t) - F_n(t)\right) \\ &\geq \frac{1}{f_n(F_n^{-1}(\xi_1))}\left(\hat{F}_{1,n}(t) - F_n(t)\right) > \frac{C_8 n^{-1/3}\log n}{\sup_{u\geq n^{-1/3}\log n} f_n(u)} = C_7 n^{-1/3}\log n. \end{aligned}$$

With the present choice of $\tilde{a}$, we get $\hat{F}_{1,n}^{-1}(\tilde{a}) \geq t$, but by the continuity of F_n we can choose a such that $a_2 < a < \tilde{a}$, $\hat{F}_{1,n}^{-1}(a) \leq t$ and $F_n^{-1}(a) - t > C_7 n^{-1/3}\log n$.

For the case

$$\hat{F}_{1,n}(t) - F_n(t) < -C_8 n^{-1/3}\log n, \tag{2.4.9}$$

we take

$$a = F_n\left(t - C_7 n^{-1/3}\log n\right) > F_n(F_n^{-1}(a_2)) = a_2.$$

This a also satisfies

$$\begin{aligned} a &= F_n(t - C_7 n^{-1/3}\log n) = F_n(t) - f_n(\xi_2) C_7 n^{-1/3}\log n \\ &\geq F_n(t) - \left(\sup_{u \geq n^{-1/3}\log n} f_n(u)\right) C_7 n^{-1/3}\log n \geq F_n(t) - C_8 n^{-1/3}\log n > \hat{F}_{1,n}(t), \end{aligned}$$

due to (2.4.9), since $\xi_2 \in \left[t - C_7 n^{-1/3}\log n, t\right]$. But this gives us $\hat{F}_{1,n}^{-1}(a) > t$ (by right continuity) and thus

$$\hat{F}_{1,n}^{-1}(a) - F_n^{-1}(a) > C_7 n^{-1/3}\log n.$$

This proves (2.4.8). So the only thing left to prove is:

$$P\left\{\sup_{0\leq t\leq \alpha_n}\left|\hat{F}_{1,n}(t) - F_n(t)\right| > C_8 n^{-1/3}\log n\right\} \to 0, \; n \to \infty.$$

But we have:

$$\begin{aligned} \sup_{t\leq\alpha_n}\left|\hat{F}_{1,n}(t) - F_n(t)\right| &\leq \hat{F}_{1,n}(\alpha_n) + F_n(\alpha_n) \leq 2F_n(\alpha_n) + \left|\hat{F}_{1,n}(\alpha_n) - F_n(\alpha_n)\right| \\ &\leq 2F_n(\alpha_n) + \sup_{t\geq\alpha_n}\left|\hat{F}_{1,n}(t) - F_0(t)\right|, \end{aligned} \tag{2.4.10}$$

and

$$F_n(\alpha_n) = F_n\left(F_n^{-1}(a_2) + (2 + C_7)n^{-1/3}\log n\right) = a_2 + f_n(\xi_3)(2 + C_7)n^{-1/3}\log n,$$

where $F_n^{-1}(a_2) \leq \xi_3 \leq F_n^{-1}(a_2) + (2 + C_7)n^{-1/3}\log n$.
Using the explicit form of f_n and the fact that $F_n^{-1}(t) = F_0^{-1}\left(t^{\frac{1}{1+\theta_n}}\right)$ we get

$$\begin{aligned} f_n(\xi_3) &= (1 + \theta_n) f_0(\xi_3) F_0(\xi_3)^{\theta_n} \leq (1 + |\theta_0|)(\sup f_0) F_0\left(F_n^{-1}(a_2)\right)^{-|\theta_n|} \\ &\leq C_9 F_0\left(F_0^{-1}\left(a_2^{\frac{1}{1+\theta_n}}\right)\right)^{-|\theta_n|} \leq C_9 a_2^{\frac{-|\theta_n|}{1-|\theta_n|}} \leq C_9 \exp\left(\frac{1}{2}\frac{|\theta_0|\log n}{\sqrt{n} - |\theta_0|}\right) \to 1, \quad n \to \infty \end{aligned}$$

and finally

$$F_n(\alpha_n) \leq C_{10} n^{-1/3}\log n,$$

for finite positive constants C_9 and C_{10}. Relations (2.4.8) and (2.4.10) now imply

$$P\left\{\sup_{t\in[0,1]}\left|\hat{F}_{1,n}(t) - F_n(t)\right| \geq C_{11} n^{-1/3}\log n\right\} \to 0, \quad n \to \infty.$$

The statement of Lemma 2.4.6 now follows from the above and from

$$\sup_{t\in[0,1]}|F_n(t) - F_0(t)| = \sup_{t\in[0,1]}\left|t^{1+\theta_n} - t\right| = O(n^{-1/2})$$

since the monotonicity of the derivative

$$\frac{d}{dt}\left(t^{1+\theta_n} - t\right) = (1+\theta_n)t^{\theta_n} - 1$$

shows that the supremum is reached at $x = (1+\theta_n)^{-1/\theta_n}$ and since we have:

$$\sqrt{n}\left|x^{1+\theta_n} - x\right| = \sqrt{n}\frac{|\theta_n|}{1+\theta_n}(1+\theta_n)^{-1/\theta_n} \to |\theta_0|/e, \quad n \to \infty.$$

∎

Remark 2.4.2 The supremum distance evaluates $\hat{F}_n^0$ only in the points of jump and thus, by arguments similar to those given in Lemma 2.2.1, the result of Lemma 2.4.6 can also be proved under the conditions of Theorem 2.1.1.

Remark 2.4.3 For τ_i defined as in Lemma 2.4.3

$$P\left\{\sup_i |\tau_{i+1} - \tau_i| > n^{-1/3}\log n\right\} \to 0, \ n \to \infty. \tag{2.4.11}$$

Proof: Note that

$$\sup_i |\tau_{i+1} - \tau_i| \le 2 \sup_{a\in[0,1]}\left|\left(\hat{F}_n^0\right)^{-1}(a) - F_0^{-1}(a)\right|,$$

due to the continuity of F_0^{-1}. Furthermore, using that both F_0 and $\hat{F}_n^0$ are monotone, we write

$$\inf_{\{t:f_0(t)>0\}} f_0(t)\cdot \sup_{a\in[0,1]}\left|\left(\hat{F}_n^0\right)^{-1}(a) - F_0^{-1}(a)\right| \le \sup_{t\in[0,1]}\left|\hat{F}_n^0(t) - F_0(t)\right|$$

to obtain the statement of the remark. ∎

2.5 The normal components of the representation

In this section we consider terms (2.2.9) and (2.2.10). Since (2.2.10) already has the Central Limit form, we only have to derive the limit behavior of (2.2.9). The term (2.2.9) will be proved to be asymptotically equivalent to an integral w.r.t. the measure $\sqrt{n}(P_n - P_0)$, and next we will consider the joint limiting distribution of (2.2.9) and (2.2.10). We first have the following representation of (2.2.9).

Lemma 2.5.1 *The expression (2.2.9) is equivalent to*

$$\mu - p(1-p)\sqrt{n}\int_{I_n}\frac{w\left(F_0(t)\right)}{F_0(t)\left(1-F_0(t)\right)}\left(\delta - F_0(t)\right)d(t)d(P_n - P_0)(t,\delta,z)1_{\hat{F}_n^0\in\mathcal{F}_n} + o_p(1),$$

where μ is given by (2.1.9) and $d(t)$ by (2.5.7) below.

Proof. The proof of this lemma is mostly based on Lemma 2.4.3 and the lemmas for the Donsker-type components proved in the section 2.6. We have the following representation:

$$\sqrt{n}\int_{I_n}\left(\left[S\left(\hat{F}_n^0\right)\right](t,\delta,z)-\left[R\left(\hat{F}_n^0\right)\right](t,\delta,z)\right)dP_0(t,\delta,z)\cdot 1_{\{\hat{F}_n^0\in\mathcal{F}_n\}}$$

$$= p(1-p)\sqrt{n}\int_{I_n}\frac{w\left(\hat{F}_n^0(t)\right)}{\hat{F}_n^0(t)\left(1-\hat{F}_n^0(t)\right)}\left[\left(pF_0(t)^{1+\theta_n}g_1(t)+(1-p)F_0(t)g_0(t)\right)\cdot\right. \tag{2.5.1}$$

$$\left.\cdot\frac{g_1(t)-g_0(t)}{(1-p)g_0(t)+pg_1(t)}-\hat{F}_n^0(t)(g_1(t)-g_0(t))\right]dt\cdot 1_{\{\hat{F}_n^0\in\mathcal{F}_n\}}$$

$$+p(1-p)\sqrt{n}\int_{I_n}\frac{w\left(\hat{F}_n^0(t)\right)}{\hat{F}_n^0(t)\left(1-\hat{F}_n^0(t)\right)}\left[F_0(t)^{1+\theta_n}g_1(t)-F_0(t)g_0(t)\right. \tag{2.5.2}$$

$$\left.-\left(pF_0(t)^{1+\theta_n}g_1(t)+(1-p)F_0(t)g_0(t)\right)\frac{g_1(t)-g_0(t)}{(1-p)g_0(t)+pg_1(t)}\right]dt\cdot 1_{\{\hat{F}_n^0\in\mathcal{F}_n\}}.$$

The last term will be shown to converge to the constant μ. We rewrite it as

$$p(1-p)\sqrt{n}\int_{I_n}\frac{w\left(\hat{F}_n^0(t)\right)}{\hat{F}_n^0(t)\left(1-\hat{F}_n^0(t)\right)}\cdot\frac{F_0(t)g_0(t)g_1(t)}{(1-p)g_0(t)+pg_1(t)}\left(F_0(t)^{\theta_n}-1\right)dt\cdot 1_{\{\hat{F}_n^0\in\mathcal{F}_n\}}. \tag{2.5.3}$$

Taylor expansion shows that the last factor of the integrand, multiplied by $\sqrt{n}$, is equal to

$$\sqrt{n}\left(F_0(t)^{\theta_n}-1\right)=\sqrt{n}\left(e^{\theta_n\log F_0(t)}-1\right)=\theta_0\log F_0(t)+\frac{\theta_0^2\left(\log F_0(t)\right)^2e^{\xi_t}}{\sqrt{n}},$$

where for any $t\in I_n$ $|\xi_t|\le|\theta_n\log F_0(t)|\le C_1|\theta_0|\frac{\log n}{\sqrt{n}}$. Hence:

$$\sqrt{n}\left(F_0(t)^{\theta_n}-1\right)=\theta_0\log F_0(t)+r_n(t),$$

$$\sup_{t\in I_n}|r_n(t)|\le C_2\frac{\theta_0^2\left(\log n\right)^2\exp\left(C_1|\theta_0|\frac{\log n}{\sqrt{n}}\right)}{\sqrt{n}}=o(1).$$

The term (2.5.2) is therefore equal to:

$$p(1-p)\theta_0\int_{I_n}\frac{w\left(\hat{F}_n^0(t)\right)}{\hat{F}_n^0(t)\left(1-\hat{F}_n^0(t)\right)}\cdot\frac{g_0(t)g_1(t)F_0(t)\log F_0(t)}{(1-p)g_0(t)+pg_1(t)}dt\cdot 1_{\{\hat{F}_n^0\in\mathcal{F}_n\}} \tag{2.5.4}$$

$$+r_n\int_{I_n}\frac{w\left(\hat{F}_n^0(t)\right)}{\hat{F}_n^0(t)\left(1-\hat{F}_n^0(t)\right)}\cdot\frac{F_0(t)g_0(t)g_1(t)}{(1-p)g_0(t)+pg_1(t)}dt\cdot 1_{\{\hat{F}_n^0\in\mathcal{F}_n\}}. \tag{2.5.5}$$

For the class $\mathcal{F}_n$ and the set I_n we have that $\sup_{F\in\mathcal{F}_n}\frac{|w(F(t))|}{F(t)(1-F(t))}1_{\{t\in I_n\}}$ is bounded by

$$\begin{cases} C_3 & \text{, if } w \text{ is a Lipschitz function,}\\ \sup_{F\in\mathcal{F}_n}\frac{|\log(F(t))|^m}{1-F(t)}1_{\{t\in I_n\}} & \text{, if } w(x)=x|\log x|^m.\end{cases}$$

In the second case this is bounded by

$$\sup_{F\in\mathcal{F}_n} \tfrac{|\log(F(t))|^m}{1-F(t)} \cdot 1_{\{t\in I_n\}} \le \sup_{\frac12\le x\le 1} \frac{|\log x|^m}{1-x} + \sup_{F\in\mathcal{F}_n,\ F(t)<\frac12} \frac{|\log(F(t))|^m}{1-F(t)} 1_{t\in I_n}$$
$$\le C_4 + 2\left|\log\left(\frac12 F_0(t)\right)\right|^m \le C_5\left(1+|\log F_0(t)|^m\right)$$

for n sufficiently large. Using the upper bound, obtained for (2.5.5), we get

$$(2.5.5) \le r_n \int_0^1 \frac{(1+|\log F_0(t)|^m)\, F_0(t) g_0(t) g_1(t)}{(1-p)g_0(T)+pg_1(T)} dt = o(1),$$

since, under the conditions of Theorem 2.2.1, the integrand is bounded by a constant (if w is a Lipschitz function, we take $m = 0$).

To treat the term (2.5.4) we define

$$A_n^+ = p(1-p)\theta_0 \int_{I_n} \sup_{F\in\mathcal{F}_n} \frac{w(F(t))}{F(t)(1-F(t))} \cdot \frac{g_0(t)g_1(t)F_0(t)\log F_0(t)}{(1-p)g_0(t)+pg_1(t)} dt$$

and

$$A_n^- = p(1-p)\theta_0 \int_{I_n} \inf_{F\in\mathcal{F}_n} \frac{w(F(t))}{F(t)(1-F(t))} \cdot \frac{g_0(t)g_1(t)F_0(t)\log F_0(t)}{(1-p)g_0(t)+pg_1(t)} dt$$

and show that A_n^+ and A_n^- converge to the same constant μ. Then we are done due to the fact that (2.5.4) belongs to the interval $[A_n^-, A_n^+]$ with a probability given by $P\{\hat F_n^0 \in \mathcal{F}_n\}$. Convergence of both A_n^- and A_n^+ to

$$\mu = p(1-p)\theta_0 \int_0^1 \frac{w(F_0(t))}{F_0(t)(1-F_0(t))} \cdot \frac{g_0(t)g_1(t)F_0(t)\log F_0(t)}{(1-p)g_0(t)+pg_1(t)} dt$$

follows from Lebesgue's dominated convergence theorem, since we have found an integrable majorant, and since for any $t \in (0,1)$:

$$\lim_{n\to\infty} \inf_{F\in\mathcal{F}_n} \frac{w(F(t))}{F(t)(1-F(t))} = \lim_{n\to\infty} \sup_{F\in\mathcal{F}_n} \frac{w(F(t))}{F(t)(1-F(t))} = \frac{w(F_0(t))}{F_0(t)(1-F_0(t))},$$

using the continuity of the function $x \mapsto w(x)/\{x(1-x)\}$ at the point $F_0(t)$.

Denote (2.5.1) by $\tilde N_n$. Then $\frac{\tilde N_n}{p(1-p)}$ equals to

$$\sqrt{n} \int_{I_n} \frac{w\left(\hat F_n^0(t)\right)}{\hat F_n^0(t)\left(1-\hat F_n^0(t)\right)} \left(\delta - \hat F_n^0(t)\right) \frac{g_1(t)-g_0(t)}{(1-p)g_0(t)+pg_1(t)} dP_0(t,\delta,z) \cdot 1_{\{\hat F_n^0 \in \mathcal{F}_n\}}. \qquad (2.5.6)$$

We can now use Lemma 2.4.3, after replacing the function d, defined by

$$d(t) = \frac{g_1(t)-g_0(t)}{(1-p)g_0(t)+pg_1(t)}, \qquad (2.5.7)$$

by a piecewise constant function d_n, defined by

$$d_n(t) = d(\tau_i),\ \tau_i \le t < \tau_{i+1},$$

where the τ_i are the same as in the section 2.4. The function d_n is close to d, since by Remark 2.4.3:

$$P\left\{\sup_i |\tau_{i+1}-\tau_i| > n^{-1/3}\log n\right\} \to 0,\ n\to\infty,$$

and since, under the conditions of Theorem 2.2.1, d is differentiable, implying:

$$|d'(t)| = \left|\left(\frac{1-r(t)}{(1-p)r(t)+p}\right)'\right| = \left|\frac{r'(t)}{((1-p)r(t)+p)^2}\right| \le \frac{1}{p^2(1-p)^2}|r'(t)|(1\wedge r(t)^{-2}).$$

Hence $\sup_{t\in[0,1} |d_n(t)-d(t)| = O_p(n^{-1/3}\log n)$.

This allows us to rewrite (2.5.6). First of all, Lemma 2.4.3 implies that it is equal to

$$\sqrt{n}\int_{I_n}\frac{w\left(\hat F_n^0(t)\right)}{\hat F_n^0(t)\left(1-\hat F_n^0(t)\right)}\left(\delta-\hat F_n^0(t)\right)(d(t)-d_n(t))dP_0(t,\delta,z)\cdot 1_{\{\hat F_n^0\in\mathcal{F}_n\}} \tag{2.5.8}$$

$$-\sqrt{n}\int_{I_n}\frac{w\left(\hat F_n^0(t)\right)}{\hat F_n^0(t)\left(1-\hat F_n^0(t)\right)}\left(\delta-\hat F_n^0(t)\right)d_n(t)d(P_n-P_0)(t,\delta,z)\cdot 1_{\{\hat F_n^0\in\mathcal{F}_n\}}. \tag{2.5.9}$$

Applying Lemma 2.4.6, we bound the absolute value of (2.5.8) by

$$\begin{aligned}
&\sqrt{n}\int_{I_n}\frac{w\left(\hat F_n^0(t)\right)}{\hat F_n^0(t)\left(1-\hat F_n^0(t)\right)}\left[(1-p)g_0(t)\left|F_0(t)-\hat F_n^0(t)\right|+pg_1(t)\left|F_0(t)^{1+\theta_n}-\hat F_n^0(t)\right|\right]\cdot\\
&\qquad\qquad\cdot|d(t)-d_n(t)|\,dt\cdot 1_{\{\hat F_n^0\in\mathcal{F}_n\}}\\
&\le C_6\sqrt{n}(\log n)^m\sup_{t\in[0,1]}|d_n(t)-d(t)|\left\{\sup_{t\in[0,1]}\left|\hat F_n^0(t)-F_0(t)\right|+\sup_{t\in[0,1]}\left|F_0(t)^{1+\theta_n}-F_0(t)\right|\right\}\\
&\qquad\qquad\cdot\int_{I_n}[(1-p)g_0(t)+pg_1(t)]dt\\
&= o_p(1).
\end{aligned}$$

The term (2.5.9) can be rewritten as a sum

$$-\sqrt{n}\int_{I_n}\frac{w(F_0(t))}{F_0(t)(1-F_0(t))}(\delta-F_0(t))\,d(t)d(P_n-P_0)(t,\delta,z)1_{\hat F_n^0\in\mathcal{F}_n} \tag{2.5.10}$$

$$-\sqrt{n}\int_{I_n}\left[\frac{w\left(\hat F_n^0(t)\right)d_n(t)}{\hat F_n^0(t)}-\frac{w(F_0(t))\,d(t)}{F_0(t)}\right]\delta d(P_n-P_0)(t,\delta,z)1_{\hat F_n^0\in\mathcal{F}_n} \tag{2.5.11}$$

$$+\sqrt{n}\int_{I_n}\left[\frac{w\left(\hat F_n^0(t)\right)d_n(t)}{1-\hat F_n^0(t)}-\frac{w(F_0(t))\,d(t)}{1-F_0(t)}\right](1-\delta)d(P_n-P_0)(t,\delta,z)\cdot 1_{\{\hat F_n^0\in\mathcal{F}_n\}}, \tag{2.5.12}$$

where (2.5.10) is what we intended to obtain in the current lemma, while (2.5.11) and (2.5.12) will be shown to be asymptotically negligible in Lemma 2.6.2. ■

The last lemma of this section is an application of the Central Limit theorem to the sum of (2.5.10) and (2.2.10).

Lemma 2.5.2

$$N_n \xrightarrow{\mathcal{D}} \mathcal{N}(\mu, \sigma^2).$$

Proof: Let μ be defined by (2.1.9). It was shown in Lemma 2.5.1 that N_n is equivalent to

$$\mu + \sqrt{n}\int_{I_n} \frac{w\left(F_0(t)\right)}{F_0(t)\left(1-F_0(t)\right)}\left(\delta - F_0(t)\right)\left(z-p-p(1-p)d(t)\right)d(P_n - P_0)(t,\delta,z) + o_p(1),$$

since $P\{\hat{F}_n^0 \in \mathcal{F}_n\} \to 1,\ n \to \infty$. Let

$$d_0(t) = p + p(1-p)d(t) = \frac{pg_1(t)}{(1-p)g_0(t)+pg_1(t)}$$

and let

$$\tilde{\xi}_i^n = \frac{w\left(F_0(T_i)\right)}{F_0(T_i)\left(1-F_0(T_i)\right)}\left(\Delta_i - F_0(T_i)\right)\left(Z_i - d_0(T_i)\right)1_{T_i\in I_n};\quad \xi_i^n = \tilde{\xi}_i^n - E\tilde{\xi}_i^n,$$

where $\xi_1^n,\dots,\xi_n^n$ are i.i.d. random values with a distribution dependent on n (which is the reason that we have to check the Lindeberg condition). The variance of ξ_i^n is

$$\begin{aligned}\operatorname{var}(\xi_i^n) &= \int_{I_n}\frac{w\left(F_0(t)\right)^2}{F_0(t)^2\left(1-F_0(t)\right)^2}\left(\delta-F_0(t)\right)^2\left(z-d_0(t)\right)^2 dP_0(t,\delta,z)\\ &= \int_0^1 \frac{w\left(F_0(t)\right)^2}{F_0(t)^2\left(1-F_0(t)\right)^2}\left[\left\{F_0(t)-F_0(t)^2\right\}(1-p)g_0(t)d_0(t)^2\right.\\ &\qquad \left.+\left\{F_0(t)^{1+\theta_n}(1-2F_0(t))+F_0(t)^2\right\}pg_1(t)(1-d_0(t))^2\right]\cdot 1_{\{t\in I_n\}}dt.\end{aligned}$$

For n sufficiently large there exists an integrable majorant $C_1\{g_0(T)+g_1(T)\}$ for the absolute value of the integrand, where C_1 is a positive finite constant. This follows from the definition of d_0, convergence of θ_n to zero, and boundednes of $w(F_0(T))/\{F_0(T)^{\frac{3}{4}}(1-F_0(T))\}$ (under conditions of the Theorem 2.2.1). Hence, using the pointwise convergence of the integrand we obtain

$$\begin{aligned}\operatorname{var}(\xi_i^n) &= \int_0^1 \frac{w\left(F_0(t)\right)^2}{F_0(t)\left(1-F_0(t)\right)}\left[(1-p)g_0(t)d_0(t)^2+pg_1(t)(1-d_0(t))^2\right]dt + o(1)\\ &= p(1-p)\int_0^1 \frac{w\left(F_0(t)\right)^2}{F_0(t)\left(1-F_0(t)\right)}\cdot\frac{g_0(t)g_1(t)}{(1-p)g_0(t)+pg_1(t)}dt + o(1).\end{aligned}$$

To finish the proof we have to check the Lindeberg condition, namely that, $\forall\epsilon > 0$,

$$\frac{1}{n}\sum_{i=1}^n E\left[(\xi_i^n)^2 1_{|\xi_i^n|\ge\sqrt{n}\epsilon}\right] \to 0,\ n\to\infty.$$

But this follows from

$$\left|\tilde{\xi}_i^n\right| \le C_2\frac{|w(F_0(t))|}{F_0(t)(1-F_0(t))}1_{t\in I_n} \le \begin{cases} C_3(\log n)^m & ,\ \text{if } w(x) = x|\log x|^m\\ C_3 & ,\ \text{if } w \text{ is Lipschitz.}\end{cases}$$

■

2.6 The Donsker components of the representation

In this section we show the terms (2.2.7), (2.2.8), (2.5.11) and (2.5.12) to be asymptotically negligible. To explain the idea, let us, as an example, consider the term

$$\sqrt{n}\int_{I_n}\left(\frac{w\left(\hat{F}_n^0(t)\right)d_n(t)}{\hat{F}_n^0(t)}-\frac{w\left(F_0(t)\right)d(t)}{F_0(t)}\right)\delta d\left(P_n-P_0\right).$$

If the expression within the brackets were a fixed function, the integral would be asymptotically normally distributed according to the Central Limit theorem. But in our case the pair $\left(\hat{F}_n^0, d_n\right)$ is getting closer to (F_0, d) as n goes to infinity. If the set of functions to which $\left(\hat{F}_n^0, d_n\right)$ belongs is not too rich this will guarantee that our expression is asymptotically small. This result is given in the next lemma which is essentially the generalization of the equicontinuity lemma of POLLARD (1984), p 150 and uses the same chaining technique. It is close in spirit to Theorem 3.3 of GESKUS AND GROENEBOOM (1999).

We now first introduce some notation and summarize relevant results. Following POLLARD (1984), we denote the empirical measure $\sqrt{n}(P_n - P_0)$ by E_n and the symmetrized empirical measure $\sqrt{n}P_n^\circ$ by E_n°. It is known that for any fixed function f

$$P\left\{\left|\int f dE_n\right| > \epsilon\right\} \le 4P\left\{\left|\int f dE_n^\circ\right| > \frac{1}{4}\epsilon\right\}, \tag{2.6.1}$$

see (11), p. 15, POLLARD (1984), where it is called the "second symmetrization lemma". Note that $\int f dE_n^\circ$ is a function of the random sample X_n and the Rademacher symmetrization sequence (call it ξ_n). Taking $X_n = x$ fixed, we can apply Hoeffding's exponential inequality (see for example POLLARD (1984), p 191), yielding

$$P\left\{\left|\int f dE_n^\circ\right| \ge \epsilon \mid X_n = x\right\} \le \exp\left(-\frac{\epsilon^2}{2\int f^2 dP_n}\right). \tag{2.6.2}$$

We also need the covering integral $J(\delta, \mathcal{A}, \rho)$. Suppose $\mathcal{A}$ is the relevant set of functions and ρ a pseudo-metric on $\mathcal{A}$. Let $N(u, \mathcal{A}, \rho)$ be the number of elements of the minimal u-net in $\mathcal{A}$ w.r.t. the metric ρ. Then

$$J(\delta, \mathcal{A}, \rho) = \int_0^\delta \sqrt{\log\left(\frac{N(u, \mathcal{A}, \rho)^2}{u}\right)}du.$$

Note that $\forall \rho > 0:\ J\left(\frac{1}{2}, \mathcal{A}, \rho\right) \ge \frac{1}{2}\sqrt{\log 2} > \frac{1}{2}$, since each u-net consists of at least one element. This means that independently of the chosen metric the inverse function $J(\cdot, \mathcal{A}, \rho)^{-1}(\epsilon)$ is well-defined for any $\epsilon \le \frac{1}{2}$ and that

$$\sup_\rho J(\cdot, \mathcal{A}, \rho)^{-1}(\epsilon) \le \frac{1}{2}. \tag{2.6.3}$$

Finally, we define

$$P^{x_n} = P^n\{\cdot \mid X_n = x_n\},$$

where x_n is a realization of the sample X_n, with empirical measure P_n. We have the following lemma:

Lemma 2.6.1 *Let P^n be a sequence of the probability measures on the set X with a sequence of subsets $\{A_n\}_{n=1}^{\infty}$. Let $\mathcal{B}$ be the set of all functions from X to $\mathbb{R}$ and let $\mathcal{A}$ be some fixed set with a sequence of subsets $\{\mathcal{A}_n\}_{n=1}^{\infty}$. Suppose $\exists a_0 \in \mathcal{A}: \ a_0 \in \bigcap_{n=1}^{\infty} \mathcal{A}_n$. Furthermore, let $\{\alpha_n\}_{n=1}^{\infty}$ be a sequence of functions $\alpha_n : \mathcal{A} \to \mathcal{B}$. Suppose that for some sequence ρ_{x_n}, $n = 1, 2, \dots$, $x_n \in X^n$, of metrics on $\mathcal{A}$, depending on the sample realizations $X_n = x_n$ the following is satisfied:*

1. $P^n\{A_n\} \to 1, \ n \to \infty,$

2. *For any $x_n \in A_n$ and for any $a_1, a_2 \in \mathcal{A}_n$*

$$\int (\alpha_n(a_1) - \alpha_n(a_2))^2 \, dP^{x_n} \le \rho_{x_n}(a_1, a_2)^2,$$

3. *For any ϵ, $0 < \epsilon < \frac{1}{2}$*

$$\lim_{n\to\infty} \sup_{x_n \in A_n} \frac{\sup_{a\in\mathcal{A}_n} \rho_{x_n}(a, a_0)}{J_{x_n}^{-1}(\epsilon)} = 0,$$

where $J_{x_n}^{-1}(\epsilon)$ is the inverse function of $J(\cdot, \mathcal{A}, \rho_{x_n})$ (which we will further denote as $J_{x_n}(\delta)$).

Then

$$\sup_{a\in\mathcal{A}_n} \left| \int (\alpha_n(a) - \alpha_n(a_0)) \, dE_n \right| = o_p(1).$$

Proof: Let $0 < \epsilon \le 1$ be fixed. We intend to show that

$$\lim_{n\to\infty} P^n \left\{ \sup_{a\in\mathcal{A}_n} \left| \int (\alpha_n(a) - \alpha_n(a_0)) \, dE_n \right| \ge 12\epsilon \right\} = 0.$$

Symmetrizing (see (2.6.1)) gives

$$\begin{aligned}
&P^n \left\{ \sup_{a\in\mathcal{A}_n} \left| \int (\alpha_n(a) - \alpha_n(a_0)) \, dE_n \right| \ge 12\epsilon \right\} \\
&\le 4P^n \left\{ \sup_{a\in\mathcal{A}_n} \left| \int (\alpha_n(a) - \alpha_n(a_0)) \, dE_n^{\circ} \right| \ge 3\epsilon \right\} \\
&\le 4P^n \left\{ \sup_{a\in\mathcal{A}_n} \left| \int (\alpha_n(a) - \alpha_n(a_0)) \, dE_n^{\circ} \right| \ge 3\epsilon \,\middle|\, X_n \in A_n \right\} + P^n\{A_n^c\}
\end{aligned}$$

and

$$\begin{aligned}
&P^n \left\{ \sup_{a\in\mathcal{A}_n} \left| \int (\alpha_n(a) - \alpha_n(a_0)) \, dE_n^{\circ} \right| \ge 3\epsilon \,\middle|\, X_n \in A_n \right\} \\
&\le \sup_{x_n \in A_n} P^n \left\{ \sup_{a\in\mathcal{A}_n} \left| \int (\alpha_n(a) - \alpha_n(a_0)) \, dE_n^{\circ} \right| \ge 3\epsilon \,\middle|\, X_n = x_n \right\}.
\end{aligned}$$

We now temporarily fix an arbitrary $x_n \in A_n$ and bound the probability from above. For simplicity we will write P for $P^n\{\cdot \mid X_n = x_n\}$ and ρ_n for ρ_{x_n}.

Define

$$\beta_i^n = 2^{1-i} J_n^{-1}(\epsilon/8),\ i = 0, 1, \ldots,\quad n = 1, 2, \ldots$$

and fix

$$k_n = \inf\left\{ i = 0, 1, \ldots :\ \beta_i^n \le \frac{\epsilon}{\sqrt{n}} \right\}.$$

For each n and i choose the minimal β_i^n-net $\mathcal{P}_i^n$ in $\mathcal{A}_n$ w.r.t. ρ_n metric and denote $|\mathcal{P}_i^n| = N_i^n$.

One of the possible ways to choose a (possibly non-minimal) β_i^n-net is to take the minimal $\frac{1}{2}\beta_i^n$-net in $\mathcal{A}$ and take $p' \in B_{\frac{1}{2}\beta_i^n}(p) \bigcap \mathcal{A}_n$ for each element p of this net when it is not empty. If p is an element of the first net and, for some $a \in \mathcal{A}_n$, $\rho_n(a,p) < \frac{1}{2}\beta_i^n$, we get by the triangle inequality

$$\rho_n(a,p') \le \rho_n(a,p) + \rho_n(p,p') \le \beta_i^n.$$

The new net consists of no more than $N(\frac{1}{2}\beta_i^n, \mathcal{A}, \rho_n)$ points. Let $a_i^n(a) \in \mathcal{P}_i^n$ be a point, minimizing the function $a' \mapsto \rho_n(a', a)$. For any fixed $a \in \mathcal{A}_n$ we get:

$$\begin{aligned}
&\left| \int (\alpha_n(a) - \alpha_n(a_0))\, dE_n^\circ \right| \\
&\le \left| \int \left(\alpha_n(a) - \alpha_n(a_{k_n}^n(a))\right) dE_n^\circ \right| + \left| \int \left(\alpha_n(a_{k_n}^n(a)) - \alpha_n(a_0)\right) dE_n^\circ \right| \\
&\le \left| \int \left(\alpha_n(a_{k_n}^n(a)) - \alpha_n(a_0)\right) dE_n^\circ \right| + \sqrt{n}\rho(a, a_{k_n}^n(a)),
\end{aligned}$$

where, by the definition of k_n, the last term is less than or equal to ϵ. Therefore we need to bound

$$P\left\{ \sup_{a \in \mathcal{P}_{n_k}^n} \left| \int (\alpha_n(a) - \alpha_n(a_0))\, dE_n^\circ \right| \ge 2\epsilon \right\}.$$

At this stage $\mathcal{P}_{n_k}^n$ is still too large to get a useful bound from Hoeffding's inequality and therefore we have to tighten the set one more time.

Define

$$\eta_i^n = 4\beta_{i+1}^n \sqrt{\log\left(\frac{(N_i^n)^2}{\beta_{i+1}^n}\right)}.$$

Then $\sum_{i=1}^{k_n} \eta_i^n$ is bounded from above by

$$8\sum_{i=1}^{\infty} (\beta_{i+1}^n - \beta_{i+2}^n) \inf_{\beta \in [\beta_{i+2}^n, \beta_{i+1}^n]} \sqrt{\log\left(\frac{(N(\beta, \mathcal{A}, \rho_n))^2}{\beta}\right)} \le 8J_n(\beta_2^n) \le 8J_n(J_n^{-1}(\epsilon/8)) = \epsilon,$$

and:

$$P\left\{\sup_{a\in\mathcal{P}^n_{n_k}}\left|\int(\alpha_n(a)-\alpha_n(a_0))\,dE^\circ_n\right|\geq 2\epsilon\right\}$$

$$\leq P\left\{\sup_{a\in\mathcal{P}^n_0}\left|\int(\alpha_n(a)-\alpha_n(a_0))\,dE^\circ_n\right|+\sum_{i=1}^{k_n}\sup_{a\in\mathcal{P}^n_i}\left|\int\left(\alpha_n(a)-\alpha_n(a^n_{i-1}(a))\right)dE^\circ_n\right|\geq 2\epsilon\right\}$$

$$\leq N^n_0\sup_{a\in\mathcal{P}^n_0}P\left\{\left|\int(\alpha_n(a)-\alpha_n(a_0))\,dE^\circ_n\right|\geq\epsilon\right\} \tag{2.6.4}$$

$$+\sum_{i=1}^{k_n}N^n_i\sup_{a\in\mathcal{P}^n_i}P\left\{\left|\int\left(\alpha_n(a)-\alpha_n(a^n_{i-1}(a))\right)dE^\circ_n\right|\geq\eta^n_i\right\}. \tag{2.6.5}$$

Applying (2.6.2) to bound (2.6.4) we get

$$N^n_0\exp\left(-\frac{\epsilon^2}{2\sup_{a\in\mathcal{P}^n_0}\int(\alpha_n(a)-\alpha_n(a_0))^2\,dP_n}\right)\leq N^n_0\exp\left(-\frac{\epsilon^2}{2\sup_{a\in\mathcal{A}_n}\rho_n(a,a_0)^2}\right).$$

Moreover, $\log N^n_0$ can be estimated from above in the following way:

$$\log N^n_0\leq\log\left(\frac{N(J_n^{-1}(\epsilon/8),\mathcal{A},\rho_n)^2}{J_n^{-1}(\epsilon/8)}\right)\leq\left(\frac{J_n(J_n^{-1}(\epsilon/8))}{J_n^{-1}(\epsilon/8)}\right)^2=\left(\frac{\epsilon}{8J_n^{-1}(\epsilon/8)}\right)^2,$$

since $J_n^{-1}(\epsilon/8)\leq\frac{1}{2}$ and (2.6.4) is bounded by

$$\exp\left(\frac{\epsilon^2}{64\,(J_n^{-1}(\epsilon/8))^2}-\frac{\epsilon^2}{2\sup_{a\in\mathcal{A}_n}\rho_n(a,a_0)^2}\right). \tag{2.6.6}$$

This takes care of (2.6.4). For (2.6.5) an application of Hoeffding's inequality (see (2.6.2)) provides the upper bound

$$\sum_{i=1}^{k_n}N^n_i\exp\left\{-\frac{(\eta^n_i)^2}{2\sup_{a\in\mathcal{P}^n_i}\rho_n(a,a^n_{i-1}(a))^2}\right\}=\sum_{i=1}^{k_n}N^n_i\exp\left\{-\frac{(4\beta^n_{i+1})^2\log\left(\frac{(N^n_i)^2}{\beta^n_{i+1}}\right)}{2\sup_{a\in\mathcal{P}^n_i}\rho_n(a,a^n_{i-1}(a))^2}\right\}$$

$$\leq\sum_{i=1}^{k_n}\exp\left\{\frac{1}{2}\log\beta^n_{i+1}\left(\frac{\beta^n_{i-1}}{\sup_{a\in\mathcal{P}^n_i}\rho_n(a,a^n_{i-1}(a))}\right)^2\right\}. \tag{2.6.7}$$

Let $\gamma_n=\sup_{a\in\mathcal{A}_n}\rho_n(a,a_0)$. We divide the sum (2.6.7) into two parts, a summation over indices i such that $\beta^n_{i-1}\geq\sqrt{\gamma_n}$ and a summation over the rest of the indices. The first set of indices satisfies

$$0<i\leq i_n=\frac{1}{2\log 2}\log\left(\frac{4J_n^{-1}(\epsilon/8)^2}{\gamma_n}\right)\vee 0\leq\frac{1}{2\log 2}|\log\gamma_n|.$$

This yields the estimate:

$$\sum_{0<i\leq i_n}(\beta^n_{i+1})^{\frac{1}{2}\left(\frac{\beta^n_{i-1}}{\gamma_n}\right)^2}+\sum_{i>i_n}\sqrt{\beta^n_{i+1}}\leq i_n2^{-\frac{1}{2\gamma_n}}+\sum_{i>i_n}\sqrt{\beta^n_{i+1}}$$

$$\leq\frac{|\log\gamma_n|2^{-1/(2\gamma_n)}}{2\log 2}+\frac{\gamma_n^{1/4}\sqrt{2}}{\sqrt{2}-1}. \tag{2.6.8}$$

Combining (2.6.6) and (2.6.8) we now have obtained the following exact inequality, valid for any n:

$$P^n\left\{\sup_{a\in\mathcal{A}_n}\left|\int(\alpha_n(a)-\alpha_n(a_0))\,dE_n^\circ\right|\geq 3\epsilon\,\middle|\, X_n\in A_n\right\}$$
$$\leq C_1\sup_{x_n\in A_n}\left\{\exp\left(\frac{\epsilon^2}{64\left(J_{x_n}^{-1}(\epsilon/8)\right)^2}-\frac{\epsilon^2}{2\sup_{a\in\mathcal{A}_n}\rho_{x_n}(a,a_0)^2}\right)+\right.$$
$$\left.+\left|\log\sup_{a\in\mathcal{A}_n}\rho_n(a,a_0)\right|2^{-\frac{1}{2\sup_{a\in\mathcal{A}_n}\rho_n(a,a_0)}}+\left(\sup_{a\in\mathcal{A}_n}\rho_n(a,a_0)\right)^{1/4}\right\}.$$

The right-hand side converges to zero, as n tends to infinity. Note that it follows from condition 3 of Lemma 2.6.1 that

$$\lim_{n\to\infty}\sup_{x_n\in A_n}\sup_{a\in\mathcal{A}_n}\rho_{x_n}(a,a_0)=0.$$

∎

Now we can apply Lemma 2.6.1 to prove the asymptotically negligibility of the terms considered in this section. First consider (2.5.11). Taking the set X and the sequence of the probability measures corresponding to the random triple (T,Δ,Z) we define the set $\mathcal{A}=\mathcal{F}\times\mathcal{D}$ and $\mathcal{A}_n=\mathcal{F}_n\times\mathcal{D}_n$, where

$$\mathcal{F}=\{\text{all df's on }[0,1]\},$$
$$\mathcal{F}_n=\{\text{all df's on }[0,1];\ \sup_{t\in[0,1]}|F_0(t)-F(t)|<n^{-1/3}\log n\},$$
$$\mathcal{D}=\left\{\tilde d:\tilde d(t)=d(\tau_i)\text{ if }t\in[\tau_i,\tau_{i+1}),\ 0=\tau_0<\tau_1<\ldots<\tau_k=1\right\}\bigcup\{d\},$$
$$\mathcal{D}_n=\left\{\tilde d:\tilde d(t)=d(\tau_i)\text{ if }t\in[\tau_i,\tau_{i+1}),\right.$$
$$\left.0=\tau_0<\tau_1<\ldots<\tau_k=1,\quad\sup_{i=0,\ldots,k-1}|\tau_{i+1}-\tau_i|<n^{-1/3}\log n\right\}\bigcup\{d\}.$$

Taking $a_0=(F_0,d)$ and

$$\alpha_n(a)(t,\delta,z)=\frac{w(a^1(t))}{a^1(t)}a^2(t)\delta 1_{\{t\in I_n\}},$$

we note that (2.5.11) can be bounded in the following way:

$$\left|\sqrt{n}\int_{I_n}\left[\frac{w\left(\hat F_n^0(t)\right)d_n(t)}{\hat F_n^0(t)}-\frac{w(F_0(t))\,d(t)}{F_0(t)}\right]\delta d(P_n-P_0)(t,\delta,z)1_{\hat F_n^0\in\mathcal{F}_n}\right|$$
$$\leq\sup_{a\in\mathcal{A}_n}\left|\int(\alpha_n(a)-\alpha_n(a_0))\,dE_n\right|+o_p(1),\tag{2.6.9}$$

since, as we have seen in Remark 2.4.3, $P\{d_n\in\mathcal{D}_n\}\to 1,\ n\to\infty$. In order to show that this is asymptotically negligible we have to find the proper metric ρ_{x_n} on $\mathcal{A}$ and sequence $\{A_n\}_{n=1}^{\infty}$ such that conditions 1, 2 and 3 of Lemma 2.6.1 are satisfied.

Using (2.2.5), uniform boundednes of the set $\mathcal{D}$, and

$$\sup_{t\in I_n,\ F\in\mathcal{F}_n} \frac{|w(F(t))|}{F(t)} \le C_1(\log n)^m; \quad \sup_{u,v\in I_n,\ F\in\mathcal{F}_n} \frac{|w(F(u))-w(F(v))|}{|F(u)-F(v)|} \le C_2(\log n)^m,$$

one can write, for any $a_1, a_2 \in \mathcal{A}_n$ and for any $t\in I_n$,

$$|\alpha_n(a_1)-\alpha_n(a_2)| \le \frac{|w(a_2^1(t))|}{a_2^1(t)}|a_1^2(t)-a_2^2(t)|\delta + |a_1^2(t)|\cdot\left|\frac{w(a_1^1(t))}{a_1^1(t)} - \frac{w(a_2^1(t))}{a_2^1(t)}\right|\delta$$
$$\le \frac{|w(a_2^1(t))|}{a_2^1(t)}|a_1^2(t)-a_2^2(t)|\delta$$
$$+2|a_1^2(t)|\left\{\frac{|w(a_1^1(t))|}{a_1^1(t)} + \frac{|w(a_1^1(t))-w(a_2^1(t))|}{|a_1^1(t)-a_2^1(t)|}\right\}\cdot\frac{|a_1^1(t)-a_2^1(t)|}{F_0(t)}\delta$$
$$\le C_3\delta(\log n)^m\left\{\frac{|a_1^1(t)-a_2^1(t)|}{F_0(t)} + |a_1^2(t)-a_2^2(t)|\right\},$$

where C_3 does not depend on a_1 and a_2. Moreover, the metric $\rho_{x_n}(a_1,a_2)$ defined by

$$\rho_{x_n}(a_1,a_2) = \rho_{x_n}^1(a_1^1,a_2^1) + \rho_{x_n}^2(a_1^2,a_2^2),$$

where

$$\rho_{x_n}^1(F_1,F_2) = (\log n)^{m+1}\left(\int_{I_n}\frac{(F_1(t)-F_2(t))^2}{F_0(t)^2}\delta dP_n(t,\delta,z)\right)^{\frac{1}{2}},$$
$$\rho_{x_n}^2(d_1,d_2) = (\log n)^{m+1}\left(\int_0^1 (d_1(t)-d_2(t))^2 dP_n(t,\delta,z)\right)^{\frac{1}{2}},$$

satisfies condition 2 of Lemma 2.6.1 independently of A_n. To satisfy condition 3, we choose a specific A_n:

$$A_n = \left\{X_n : \int_{I_n}\frac{\delta}{F_0(t)^2}dP_n(t,\delta,z) \le (\log n)^2\right\}.$$

By the Markov inequality $P\{A_n\}$ tends to 1, since

$$P\{A_n^c\} \le (\log n)^{-2}\int_{I_n}\frac{\delta}{F_0(t)^2}dP_0(t,\delta,z) \le 2(\log n)^{-2}\int_{I_n}\frac{(1-p)g_0(t)+pg_1(t)}{t}dt \to 0,$$

as $n\to\infty$.

To check condition 3 of the Lemma 2.6.1, we will estimate the covering number from above, providing us with a lower bound for the inverse. One way to construct an appropriate $(\delta,\mathcal{A},\rho_{x_n})$-net is to construct a $(\delta/2,\mathcal{F},\rho_{x_n}^1)$-net as well as a $(\delta/2,\mathcal{D},\rho_{x_n}^2)$-net and take their product. The covering number will be bounded by the product of these two covering numbers.

The result for the $(\delta/2,\mathcal{F},\rho_{x_n}^1)$-net follows from the estimation obtained by Birman and Solomjak (see for example VAN DE GEER (2000), p. 18). The supremum over all $L_2(P)$-metrics for uniformly bounded measures P is bounded as

$$\sup_{\rho_{L_2(P)}} \log N(\delta,\mathcal{F},\rho_{L_2(P)}) \le \frac{A}{\delta},$$

where A only depends on the uniform bound of the measures. Our measure $\rho^1_{x_n}$ is the L_2-measure and the measure

$$\tilde{\rho}_{x_n} = (\log n)^{-(m+2)} \rho^1_{x_n}$$

is bounded by one for any $x_n \in A_n$. Thus

$$\sup_{x_n \in A_n} \log N(\delta/2, \mathcal{F}, \rho^1_{x_n}) \leq \frac{A(\log n)^{m+2}}{\delta}. \tag{2.6.10}$$

The function d has a bounded variation on the interval $[0,1]$. Therefore, it can be represented as a difference between two monotone increasing, bounded functions d_1 and d_2. Then defining $\mathcal{D}_1$ and $\mathcal{D}_2$ similar to $\mathcal{D}$ and using the triangle inequality we conclude that

$$\log N(\delta, \mathcal{D}, \rho) \leq \log N(\delta/2, \mathcal{D}_1, \rho) + \log N(\delta/2, \mathcal{D}_2, \rho) \leq 2 \log N(\delta/2, \mathcal{F}, \rho).$$

The result of Birman and Solomjak above implies

$$\sup_{x_n \in A_n} \log N(\delta/2, \mathcal{D}, \rho^2_{x_n}) \leq \frac{B(\log n)^{m+1}}{\delta}. \tag{2.6.11}$$

Collecting (2.6.10) and (2.6.11) we get

$$\sup_{x_n \in A_n} J_{x_n}(\delta, \mathcal{A}, \rho_{x_n}) \leq C_1 (\log n)^{\frac{m+2}{2}} \int_0^\delta \frac{du}{\sqrt{u}} = C_2 (\log n)^{\frac{m+2}{2}} \sqrt{\delta},$$

and

$$\sup_{x_n \in A_n} \frac{1}{J^{-1}_{x_n}(\epsilon)} \leq C_3 \epsilon^{-2} (\log n)^{m+2}.$$

Furthermore $\sup_{x_n \in A_n} \sup_{a \in \mathcal{A}_n} \rho_{x_n}(a, a_0)$ is less than

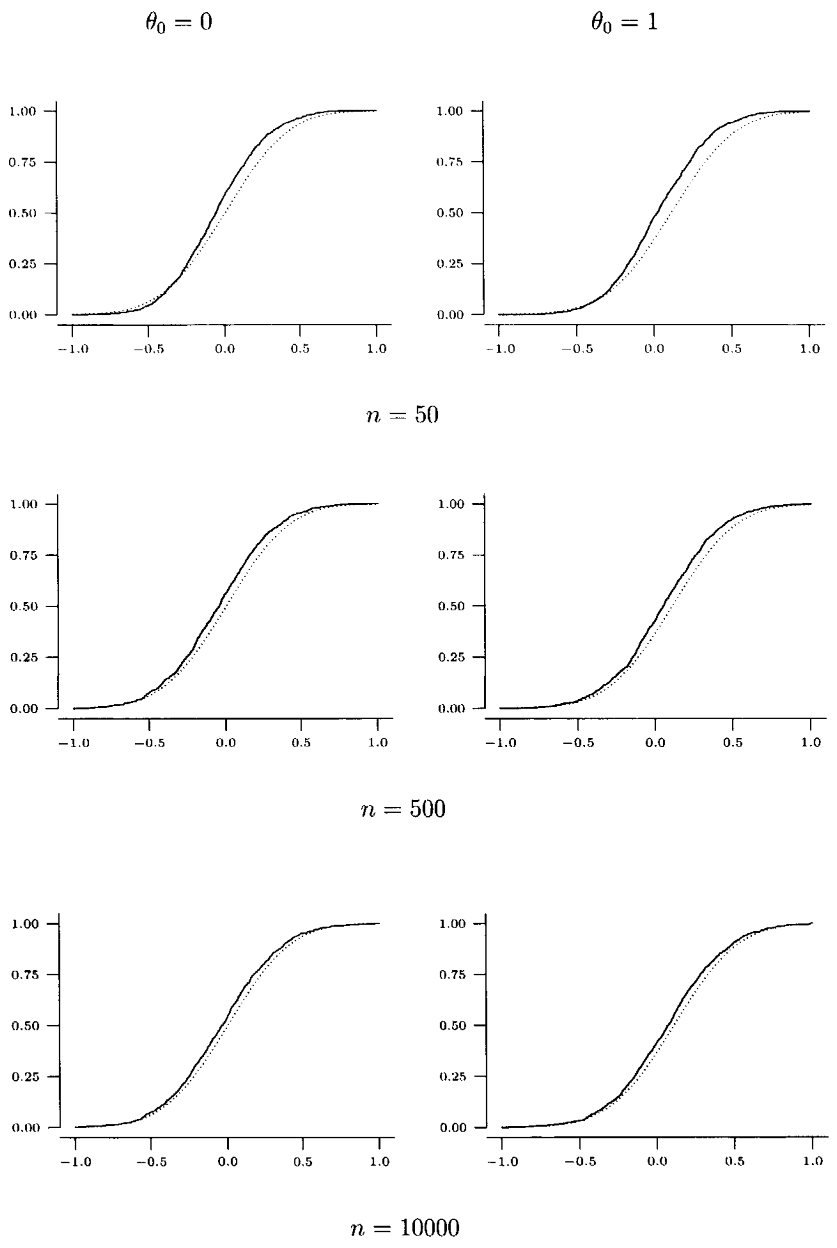
$$C_4 (\log n)^{m+2} \left(\sup_{F \in \mathcal{F}_n} |F_0(t) - F(t)| + \sup_{\tilde{d} \in \mathcal{D}_n} |d(t) - \tilde{d}(t)| \right) \leq C_5 n^{-1/3} (\log n)^{m+3}.$$

Application of Lemmas 2.6.1 to 2.6.9 finishes the proof. The terms (2.2.7), (2.2.8) and (2.5.12) can be treated similarly.

Lemma 2.6.2 *Under conditions of the Theorem 2.2.1 terms (2.2.7), (2.2.8), (2.5.11) and (2.5.12) are of asymptotically small order in probability.*

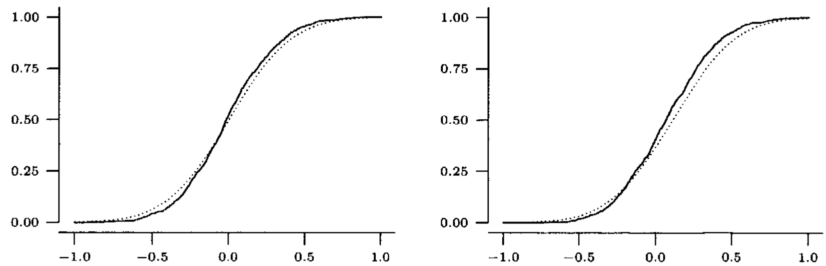
2.7 Results of simulations

This section is contributed to the results of the computer simulations of the statistic S_n for different sample sizes n. All twelve pictures below represent the empirical distribution functions of S_n (solid lines) compared to the proved above limiting normal distribution (dotted lines); the sample size everywhere equals to 1000. The first series of pictures corresponds to the underlying distributions $Z \sim \text{Bernoulli}(0.5)$, $F_0, G_0 \sim \exp(1)$, $G_1 \sim \exp(2)$.

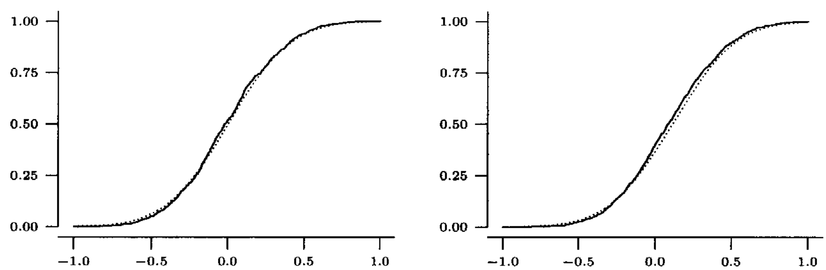


Another series of pictures corresponds to the underlying distributions $Z \sim$ Bernoulli(0.5), $F_0, G_0, G_1 \sim \mathrm{Un}[0, 1]$.

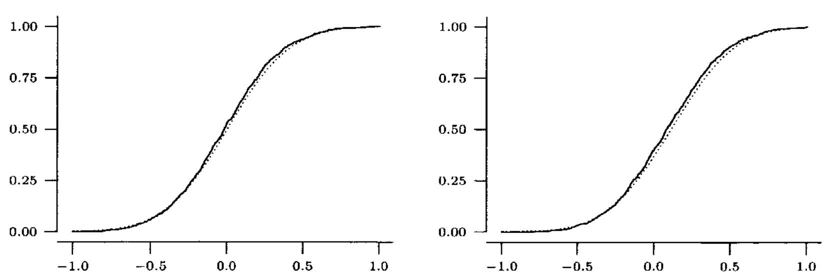
$\theta_0 = 0$ $\theta_0 = 1$



$n = 50$



$n = 500$



$n = 10000$

Chapter 3

Two samples likelihood ratio test

We propose a two-sample test for testing that the distribution functions F_0 and F_1, generating the two samples, are equal, in the case that the samples are subject to current status censoring (also called "interval censoring, case 1"). The proposed test is a likelihood ratio test and tests the null hypothesis $F_0 = F_1$ against Lehmann alternatives $F_1 = F_0^{1+\theta}$, for some $\theta > -1$. Two logarithms of the likelihood ratio statistic is shown to be asymptotically distributed as a normal random variable squared. Moreover, we show that the test is asymptotically efficient for testing against the alternative $F_1 = F_0^{1+\theta}$.

3.1 Introduction

In this paper we study the two sample likelihood ratio test for current status data. We want to test whether two samples $\xi_1, \ldots, \xi_{n_1}$ and $\eta_1, \ldots, \eta_{n_2}$ are generated from the same distribution, where both samples are subject to current status censoring. To facilitate notation we will join both samples in $X_1, \ldots, X_n$, where $n = n_1 + n_2$, and introduce indicators $Z_i = 1_{X_i \in \{\eta_1, \ldots, \eta_{n_2}\}}$, describing to which of the two samples an element X_i belongs. All elements are censored, i.e., we do not observe X_i directly, but instead we observe triples (T_i, Δ_i, Z_i), where T_i is the censoring time and $\Delta_i = 1_{X_i \le T_i}$ is an indicator specifying whether X_i lies before or after T_i. The (T_i, Δ_i, Z_i) are assumed to be independent and identically distributed, with T_i being independent of X_i, conditionally on the indicator Z_i, which is assumed to have a Bernoulli distribution with parameter $0 < p < 1$.

Denote the marginal distributions of X_i, conditionally on Z_i, by

$$F_z(x) = P\{X_i \le x | Z_i = z\}, \quad z = 0, 1,$$

and similarly for T_i

$$G_z(t) = P\{T_i \le t | Z_i = z\}, \quad z = 0, 1.$$

Hence, $\xi_1, \ldots, \xi_{n_1}$ is a random sample from F_0 and $\eta_1, \ldots, \eta_{n_2}$ is a sample from F_1.

Our null hypothesis is $H_0 : F_0 = F_1$. We will consider Lehmann alternatives

$$F_1(t) = F_0(t)^{1+\theta}, \quad \theta \in (-1, \infty).$$

Lehmann alternatives are often applied in biostatistics and are closely connected to the proportional hazard model $1 - F_1(t) = (1 - F_0(t))^{1+\theta}$, $\theta \in (-1, \infty)$. The proportional hazard model can be transformed into a Lehmann alternative by considering $X_i' = -X_i$, $T_i' = -T_i$ and $\Delta_i' = 1 - \Delta_i$.

Denote by $\mathcal{F}$ the family of all distribution functions and $\Theta = (-1, \infty)$. Then, as explained in more detail in Section 3.3, the likelihood ratio test statistic for testing the null hypothesis against Lehmann alternatives is given by

$$\mathbf{T}_n = \frac{\sup_{F\in\mathcal{F},\theta\in\Theta} \prod_{i=1}^{n} [L(\theta, F)](T_i, \Delta_i, Z_i)}{\sup_{F\in\mathcal{F}} \prod_{i=1}^{n} [L(0, F)](T_i, \Delta_i, Z_i)}, \tag{3.1.1}$$

where

$$[L(\theta, F)](t, \delta, z) = \delta F(t)^{1+z\theta} + (1-\delta)\left(1 - F(t)^{1+z\theta}\right).$$

We will investigate the asymptotic behavior of $\mathbf{T}_n$ under contiguous Lehmann alternatives, i.e., we let θ depend on n: $\theta_n = \theta_0/\sqrt{n}$, and consider a sequence of alternatives $F_0(t)^{1+z\theta_0/\sqrt{n}}$.

The purpose of this paper is to show that in this case, $\mathbf{T}_n$ behaves similar to the likelihood ratio statistic in the uncensored case, i.e., $2\log\mathbf{T}_n$ converges in distribution to a χ_1^2 random variable. We will assume that

(C1) the distributions F_0, G_0 and G_1 have the same support,

and that F_0, G_0 and G_1 have continue densities f_0, f_1 and g_1 satisfying

(C2) $$0 < \inf_{t\in\text{support}(F_0)} \frac{f_0(t)}{(1-p)g_0(t) + pg_1(t)} \le \sup_{t\in\text{support}(F_0)} \frac{f_0(t)}{(1-p)g_0(t) + pg_1(t)} < \infty,$$

and

(C3) $r(t) = \dfrac{g_0(F_0^{-1}(t))}{g_1(F_0^{-1}(t))}$ is differentiable with $\sup_{t\in[0,1]} |r'(t)|\left(1 \wedge r(t)^{-2}\right) < \infty$.

Conditions (C2) and (C3) are somewhat technical. However, note that if the censoring distributions are the same, i.e. $G_0 = G_1$, then (C3) is automatically satisfied. Moreover, if f_0 and g_i have compact support, such that $\inf f_0 > 0$ and $\sup |g_i'| < \infty$, then (C3) is satisfied as well. Finally, if all three distributions are exponential with parameters $\lambda_{F_0}, \lambda_{G_0}$ and λ_{G_1}, assuming $\lambda_{G_0} \le \lambda_{G_1}$, then it can be shown that the statement of Theorem 3.1.1 below (our main theorem) holds if $\lambda_{G_0} = \lambda_{F_0}$.

Theorem 3.1.1 *(Main Theorem)*
Suppose that F_0, G_0 and G_1 satisfy conditions (C1)-(C3) and that θ depends on n in the following way: $\theta_n = \theta_0/\sqrt{n}$. Let $\mathbf{T}_n$ be the likelihood ratio statistic defined by (3.1.1). Then $2\log\mathbf{T}_n$ converges in distribution to the random variable Y^2, where $Y \sim \mathcal{N}\left(\theta_0/\sqrt{I_{F_0}}, 1\right)$, and

$$I_{F_0} = \left(p(1-p)\int_{\text{support}(F_0)} \frac{F_0(t)(\log F_0(t))^2}{1-F_0(t)} \frac{g_0(t)g_1(t)}{(1-p)g_0(t) + pg_1(t)} dt\right)^{-1}.$$

In GROENEBOOM, KULIKOV AND LOPUHAÄ (2002) it was shown that the power $\pi_n(\theta_0)$ of testing H_0 against H_1 is bounded by:

$$\limsup_{n\to\infty} \pi_n(\theta_0) \leq 1 - \Phi\left(z_{1-\alpha/2} - \theta_0/\sqrt{I_{F_0}}\right) + \Phi\left(z_{\alpha/2} - \theta_0/\sqrt{I_{F_0}}\right).$$

Hence, as a corollary of Theorem 3.1.1, we have that the test based on the likelihood ratio statistic is also asymptotically (locally) most efficient.

The method used to prove Theorem 3.1.1 is similar to methods used in MURPHY AND VAN DER VAART (1997), and uses optimality of the maximum likelihood estimator $\hat{\theta}_n$ for θ_0 together with the fact that $\hat{\theta}_n$ is asymptotically normal. In Section 3.3 we derive some properties of maximum likelihood estimators for the parameters in the current semi-parametric setup. Asymptotic normality of $\hat{\theta}_n$ is obtained by analogous means as used in HUANG (1996). For this we need stochastic equicontinuity of empirical processes corresponding to the derivatives of the log likelihood, which is obtained in Section 3.4, and a suitable Taylor expansion, to be discussed in Section 3.5. Consequences of the optimality of the maximum likelihood estimators are studied in Section 3.6, whereas the proof of Theorem 3.1.1 is presented in Section 3.7. We conclude the paper by a small simulation study in Section 3.8.

3.2 Some notes on the proof

We can, at the cost of some extra technicalities, prove the Theorem 3.1.1 under conditions weaker than (C3).

Remark 3.2.1 Define

$$A_u = \left\{t \in [0,1]:\ |r'(t)|\left(1 \wedge r(t)^{-2}\right) > u\right\},$$
$$B_u = \left\{t \in [0,1]:\ \text{there is } s \in A_u \text{ such that } |t-s| \leq u^{-2}\right\}.$$

If $r(t)$ is a monotone function on support(F_0), condition (C3) can be replaced by

$$\lim_{u\to\infty} u(\log u)^6 G_i \circ F_0^{-1}(B_u) = 0,\ i = 0, 1.$$

This remark takes care of the important case of exponential-type distributions. Monotonicity of $r(t)$ makes entropy calculations for the quite special class $\mathcal{Q}$ (see section 3.6) unnecessary, since then this class is included into well-studied class of monotone bounded functions.

The general scheme of proof of Theorem 3.1.1 is discussed in section 3.1. But before following this scheme we need to show that the proof of Theorem 3.1.1 for the bounded support case implies the proof for more general distributions.

Consider the sequence of triples $\{(X_i, T_i, Z_i)\}$ in the hidden space and define

$$\tilde{X}_i = F_0(X_i),$$
$$\tilde{T}_i = F_0(T_i),$$
$$\tilde{Z}_i = Z_i.$$

They are distributed as

$$P\{\tilde{X}_i \leq x | \tilde{Z}_i = z\} = \left(F_{Un[0,1]}(x)\right)^{1+\theta z},\ z = 0, 1,$$
$$P\{\tilde{T}_i \leq t | \tilde{Z}_i = z\} = G_z\left(F_0^{-1}(t)\right),\ z = 0, 1,$$
$$\tilde{Z}_i \sim \text{Bernoulli}(p),$$

where F_0^{-1} is defined by

$$F_0^{-1}(x) = \sup\{t : \ F_0(t) \le x\}.$$

Notice that almost surely

$$\tilde{X}_i \le \tilde{X}_j \iff X_i \le X_j,$$
$$\tilde{T}_i \le \tilde{T}_j \iff T_i \le T_j.$$

Therefore for the observation space sequences of triples $\{(T_i, \Delta_i, Z_i)\}$ and $\{(\tilde{T}_i, \tilde{\Delta}_i, \tilde{Z}_i)\}$ (assume they are both ordered w.r.t. the first element) we have $\tilde{\Delta}_i = \Delta_i$ and

$$\begin{aligned}\mathbf{T}_n &= \frac{\sup_{F\in\mathcal{F},\theta\in\Theta}\prod_{i=1}^n\left(\Delta_i F(T_i)^{1+\theta Z_i} + (1-\Delta_i)\left(1-F(T_i)^{1+\theta Z_i}\right)\right)}{\sup_{F\in\mathcal{F}}\prod_{i=1}^n\left(\Delta_i F(T_i) + (1-\Delta_i)(1-F(T_i))\right)} \\ &= \frac{\sup_{0\le y_1\le\ldots\le y_n\le 1,\theta\in\Theta}\prod_{i=1}^n\left(\Delta_i y_i^{1+\theta Z_i} + (1-\Delta_i)\left(1-y_i^{1+\theta Z_i}\right)\right)}{\sup_{0\le y_1\le\ldots\le y_n\le 1}\prod_{i=1}^n\left(\Delta_i y_i + (1-\Delta_i)(1-y_i)\right)} \\ &= \frac{\sup_{F\in\mathcal{F},\theta\in\Theta}\prod_{i=1}^n\left(\tilde{\Delta}_i F(\tilde{T}_i)^{1+\theta\tilde{Z}_i} + (1-\tilde{\Delta}_i)\left(1-F(\tilde{T}_i)^{1+\theta\tilde{Z}_i}\right)\right)}{\sup_{F\in\mathcal{F}}\prod_{i=1}^n\left(\tilde{\Delta}_i F(\tilde{T}_i) + (1-\tilde{\Delta}_i)\left(1-F(\tilde{T}_i)\right)\right)} = \tilde{\mathbf{T}}_n.\end{aligned}$$

Making a substitution in the formula of I_{F_0} we conclude that as soon as we have proved Theorem 3.1.1 for the case of $F_0 = F_{Un[0,1]}$, which we will assume from now on, the proof for all F_0, G_0 and G_1 will follow automatically.

3.3 Properties of the maximum likelihood estimator

In this section we discuss properties of the maximum likelihood estimator in the current semiparametric setup. Define measures μ_0 and μ_1 on $\mathbb{R} \times \{0,1\}$ by

$$\mu_0 = G_0 \times (\text{counting measure on } \{0,1\}), \tag{3.3.1}$$
$$\mu_1 = G_1 \times (\text{counting measure on } \{0,1\}). \tag{3.3.2}$$

Then, if μ is a measure on $\{(t,\delta,z) : t\in\mathbb{R}, \delta\in\{0,1\}, z\in\{0,1\}\}$ defined by

$$\mu(A) = \mu_0\{(t,\delta) : (t,\delta,0)\in A\} + \mu_1\{(t,\delta) : (t,\delta,1)\in A\},$$

the density of the triple (T,Δ,Z) under the Lehmann alternative $F_0(t)^{1+z\theta}$ is

$$[L(\theta,F_0)](t,\delta,z) = \delta F_0(t)^{1+z\theta} + (1-\delta)(1-F_0(t)^{1+z\theta}).$$

The log likelihood can be written as

$$l_n(F_0,\theta) = \sum_{i=1}^n l(\theta, F_0(T_i), \Delta_i, Z_i),$$

where

$$l(\theta,x,\delta,z) = \delta(1+z\theta)\log x + (1-\delta)\log\left(1-x^{1+\theta z}\right).$$

Denote by $(\hat{\theta}_n, \hat{F}_n)$ the maximum likelihood estimator of (θ_0, F_0).

The next lemma shows that the maximum likelihood estimator $\hat{\theta}_n$ of the one-dimensional contiguous parameter $\theta_n = \theta_0/\sqrt{n}$ is of order $O_p(n^{-1/3})$ in probability. In fact, the quality of the estimator is better than $O_p(n^{-1/3})$ but to prove this we will need the preliminary estimate below.

Lemma 3.3.1 *Suppose that conditions of Theorem 3.1.1 are satisfied. Then* $\hat{\theta}_n = O_p(n^{-1/3})$.

Proof: Define the family

$$\mathcal{P}^{1/2} = \left\{ \sqrt{L(\theta, F)},\ \theta \in (-1, \infty),\ F \text{ is a distribution function} \right\}.$$

Estimate the entropy with bracketing $H_B(\epsilon, \mathcal{P}^{1/2}, \rho_{L_2(\mu)})$, where $\rho_{L_2(\mu)}$ is L_2-distance w.r.t. measure μ using

$$\mathcal{P}_1^{1/2} = \left\{ \sqrt{L_1(F_0, F_1)},\ F_0, F_1 \text{ are distribution functions} \right\},$$

where

$$L_1(F_0, F_1)(t, \delta, z) = (1-z)\left(\delta F_0(t) + (1-\delta)(1-F_0(t))\right) + z\left(\delta F_1(t) + (1-\delta)(1-F_1(t))\right).$$

Since $\mathcal{P}^{1/2} \subseteq \mathcal{P}_1^{1/2}$, by the triangle inequality

$$H_B(\epsilon, \mathcal{P}^{1/2}, \rho_{L_2(\mu)}) \le H_B(\epsilon/2, \mathcal{P}_1^{1/2}, \rho_{L_2(\mu)}).$$

Furthermore for the family

$$\mathcal{P}_2^{1/2} = \left\{ \sqrt{L_1(F, F)},\ F \text{ is a distribtuion function} \right\}$$

and measures μ_0 and μ_1 defined by (3.3.1) and (3.3.2),

$$H_B(\epsilon/2, \mathcal{P}_1^{1/2}, \rho_{L_2(\mu)}) \le H_B(\epsilon/4, \mathcal{P}_2^{1/2}, \rho_{L_2(\mu_0)}) + H_B(\epsilon/4, \mathcal{P}_2^{1/2}, \rho_{L_2(\mu_1)}) < \frac{A}{\epsilon},$$

where A is a positive universal constant (see VAN DE GEER (2000)). By Theorem 7.4 in VAN DE GEER (2000), the Hellinger distance between estimated and underlying densities satisfies

$$h_\mu(L(\hat{\theta}_n, \hat{F}_n), L(\theta_n, F_0)) = O_p(n^{-1/3}),$$

uniformly in F_0 and θ_n. Uniform boundednes of $L(\theta, F)$ implies

$$\left|\left| L(\hat{\theta}_n, \hat{F}_n) - L(\theta_n, F_0) \right|\right|^2_{L_2(\mu)} \le 8 \int \frac{1}{2} \left(\sqrt{L(\hat{\theta}_n, \hat{F}_n)} - \sqrt{L(\theta_n, F_0)} \right)^2 d\mu.$$

Continuity of g_0 and g_1 together with support$(G_i) = [0,1]$, $i = 0, 1$ implies the existence of an interval $[a, b]$ such that $0 < a < b < 1$ and $\inf_{t \in [a,b]} g_i(t) > 0$, $i = 0, 1$. Therefore

$$\int_a^b \left(\hat{F}_n(t) - F_0(t) \right)^2 dt = O_p(n^{-2/3}), \tag{3.3.3}$$

$$\int_a^b \left(\hat{F}_n(t)^{1+\hat{\theta}_n} - F_0(t)^{1+\theta_n} \right)^2 dt = O_p(n^{-2/3}).$$

The maximum of the function $|t - t^{1+\theta_n}|$ on $[0,1]$ is given by

$$\sup_{t \in [0,1]} |t - t^{1+\theta_n}| = |\theta_n| (1+\theta_n)^{-\left(1 + \frac{1}{\theta_n}\right)} = o(n^{-1/3}). \tag{3.3.4}$$

Using the triangle inequality one more time we get

$$\int_a^b \left(\hat{F}_n(t)^{1+\hat{\theta}_n} - F_0(t)\right)^2 dt = O_p(n^{-2/3}). \tag{3.3.5}$$

Let $\alpha_n(t) = \hat{F}_n(t) - F_0(t)$, $\beta_n(t) = \hat{F}_n(t)^{\hat{\theta}_n} - 1$ and subtract (3.3.3) from (3.3.5). Then

$$2\int_a^b \alpha_n(t)\beta_n(t)\hat{F}_n(t)dt + \int_a^b \hat{F}_n(t)^2\beta_n(t)^2 dt = O_p(n^{-2/3}). \tag{3.3.6}$$

Next consider the integrals

$$A_n = \left(\int_a^b \alpha_n(t)^2 dt\right)^{1/2},$$
$$B_n = \left(\int_a^b \beta_n(t)^2 dt\right)^{1/2}.$$

By the Cauchy-Schwarz inequality, the left hand side of (3.3.6) is bounded from below by $B_n\left(\hat{F}_n(a)^2 B_n - 2A_n\right)$, which is less than t, $t \geq 0$ if and only if

$$0 \leq B_n < \frac{2}{\hat{F}_n(a)^2}\left(A_n + \sqrt{A_n^2 + t\hat{F}_n(a)^2}\right).$$

Using (3.3.3) we conclude $B_n = O_p(n^{-1/3})$. It also follows from (3.3.3) that for some C_1 and C_2: $0 < C_1 < C_2 < 1$

$$\lim_{n\to\infty} P\left\{C_1 \leq \inf_{t\in[a,b]} \hat{F}_n(t) \leq \sup_{t\in[a,b]} \hat{F}_n(t) \leq C_2\right\} = 1.$$

Since $\hat{F}_n(t)^{\hat{\theta}_n}$ is monotone in t, we get that with probability tending to one

$$B_n/(b-a) \geq \left|1 - C_2^{\hat{\theta}_n}\right|,$$

and therefore

$$|\hat{\theta}_n| \leq -C_3 \log(1 - |B_n|/(b-a)) = O_p(n^{-1/3}).$$

■

The next result is trivial to prove, but we give it as a lemma for easier references. It prevents us from using the same argument twice in the sequel.

Lemma 3.3.2 *Let $\mathcal{F}$ and $\mathcal{G}$ be two sets and $H : \mathcal{F} \to \mathcal{G}$, so that there exists $H^{-1} : \mathcal{G} \to \mathcal{F}$ satisfying $H^{-1}(H(F)) = F$ for all $F \in \mathcal{F}$. Then for any $W : \mathcal{F} \to \mathbb{R}$:*

$$\operatorname*{argmax}_{F\in\mathcal{F}} W(F) = H^{-1}\left(\operatorname*{argmax}_{G\in H(\mathcal{F})} W(H^{-1}(G))\right).$$

To show that the supremum distance $\sup|\hat{F}_n(t) - F_0(t)|$ is small, we will use the following construction closely connected to a construction in GROENEBOOM, KULIKOV AND LOPUHAÄ (2002). Define for the sample of triples $\{(t_i, \delta_i, z_i)\}$ and $\theta \in (-1, \infty)$

$$\hat{F}^\theta_{0,n} = \underset{F}{\operatorname{argmax}} \sum_{i=1}^n l(\theta, F(t_i), \delta_i, 0), \tag{3.3.7}$$

$$\hat{F}^\theta_{1,n} = \underset{F}{\operatorname{argmax}} \sum_{i=1}^n l(\theta, F(t_i), \delta_i, 1), \tag{3.3.8}$$

$$\hat{F}^\theta_{n} = \underset{F}{\operatorname{argmax}} \sum_{i=1}^n l(\theta, F(t_i), \delta_i, z_i). \tag{3.3.9}$$

The derivatives of l_n can be written in terms of the functions

$$\begin{aligned} l_1(\theta, x, \delta, z) &= \frac{\partial l(\theta, x, \delta, z)}{\partial \theta} = z \log x \left(\delta - \frac{(1-\delta)x^{1+\theta}}{1 - x^{1+\theta}}\right), \\ l_2(\theta, x, \delta, z) &= \frac{\partial l(\theta, x, \delta, z)}{\partial x} = (1 + z\theta)\left(\frac{\delta}{x} - \frac{(1-\delta)x^{\theta z}}{1 - x^{1+\theta z}}\right). \end{aligned}$$

Lemma 3.3.3 *Suppose $|\theta| < 1/2$. Then $\hat{F}_n$ lies between $\hat{F}_{0,n}$ and $\hat{F}_{1,n}$.*

Proof: Take t_i and t_j such that $\hat{F}^\theta_n(t_{i-1}) < \hat{F}^\theta_n(t_i) = \hat{F}^\theta_n(t_j) < \hat{F}^\theta_n(t_{j+1})$. Then

$$\sum_{k=i}^j l_2(\theta, \hat{F}^\theta_n(t_k), \delta_k, z_k) = 0, \tag{3.3.10}$$

since otherwise for some small ϵ we can define $F_{n,\epsilon}(t) = \hat{F}^\theta_n(t) + \epsilon 1_{\hat{F}^\theta_n(t) = \hat{F}^\theta_n(t_i)}$ satisfying $l_n(\hat{F}^\theta_n, \theta) < l_n(F_{n,\epsilon}, \theta)$. Analogously, for any $l \geq i$

$$\sum_{k=i}^l l_2(\theta, \hat{F}^\theta_n(t_k), \delta_k, z_k) \geq 0. \tag{3.3.11}$$

Consider the case $\theta > 0$ first. Then

$$\frac{\delta(1+\theta z)}{x} - (1-\delta)\frac{(1+\theta z)x^{\theta z}}{1 - x^{1+\theta z}} \geq \frac{\delta}{x} - \frac{1-\delta}{1-x}. \tag{3.3.12}$$

Suppose now there is a minimal t_l such that $\hat{F}^\theta_n(t_l) < \hat{F}^\theta_{0,n}(t_l)$. Consider the interval $[t_l, t_j]$, where $\hat{F}^\theta_n(t_l) = \hat{F}^\theta_n(t_j) < \hat{F}^\theta_n(t_{j+1})$. Using (3.3.10), (3.3.11), (3.3.12) and that interval where $\hat{F}^\theta_{0,n}$ is constant starts at t_l we obtain

$$\begin{aligned} 0 \geq \textstyle\sum_{k=l}^j l_2(\theta, \hat{F}^\theta_n(t_k), \delta_k, z_k) &\geq \sum_{k=l}^j l_2(\theta, \hat{F}^\theta_n(t_k), \delta_k, 0) \\ &> \sum_{k=l}^j l_2(\theta, \hat{F}^\theta_{0,n}(t_l), \delta_k, 0) \geq \sum_{k=l}^j l_2(\theta, \hat{F}^\theta_{0,n}(t_k), \delta_k, 0) \geq 0. \end{aligned}$$

This is a contradiction.

Assume there is a maximal t_l such that $\hat{F}_n^\theta(t_l) > \hat{F}_{1,n}^\theta(t_l)$ and consider the sum over $i \le k \le l$, where $\hat{F}_n^\theta(t_{i-1}) < \hat{F}_n^\theta(t_i) = \hat{F}_n^\theta(t_l)$. This again leads to a contradiction. For the case $\theta < 0$ apply Lemma 3.3.2 with $\mathcal{F} = \{\text{all distributions}\}$ and $H(F) = F^{1+\theta}$. Rewrite

$$\begin{aligned}
\hat{F}_n^\theta &= \underset{F\in\mathcal{F}}{\operatorname{argmax}} \sum_{i=1}^n \left(\delta_i(1+\theta z_i)\log F(t_i) + (1-\delta_i)\log\left(1 - F(t_i)^{1+\theta z_i}\right)\right) \\
&= \left\{ \underset{F\in\mathcal{F}}{\operatorname{argmax}} \sum_{i=1}^n \left(\delta_i \frac{1+\theta z_i}{1+\theta}\log F(t_i) + (1-\delta_i)\log\left(1 - F(t_i)^{\frac{1+\theta z_i}{1+\theta}}\right)\right)\right\}^{\frac{1}{1+\theta}} \\
&= \left\{ \underset{F\in\mathcal{F}}{\operatorname{argmax}} \sum_{i=1}^n \left(\delta_i \left(1+\theta' z_i'\right)\log F(t_i) + (1-\delta_i)\log\left(1 - F(t_i)^{1+\theta' z_i'}\right)\right)\right\}^{\frac{1}{1+\theta}},
\end{aligned}$$

where $\theta' = -\frac{\theta}{1+\theta}$ and $z_i' = 1 - z_i$. The proof of Lemma 3.3.3 for $\theta < 0$ follows from the statement of Lemma 3.3.3, proved for positive $\theta' = -\frac{\theta}{1+\theta}$, since for the functions $\tilde{F}_n^{\theta'}$, $\tilde{F}_{0,n}^{\theta'}$ and $\tilde{F}_{1,n}^{\theta'}$, defined by (3.3.7), (3.3.8) and (3.3.9), we have for any $t \in [0,1]$:

$$\hat{F}_n^\theta(t) = \left(\tilde{F}_n^{\theta'}(t)\right)^{\frac{1}{1+\theta}} \in \left[\left(\tilde{F}_{0,n}^{\theta'}(t)\right)^{\frac{1}{1+\theta}}, \left(\tilde{F}_{1,n}^{\theta'}(t)\right)^{\frac{1}{1+\theta}}\right] = \left[\hat{F}_{1,n}^\theta(t), \hat{F}_{0,n}^\theta(t)\right].$$

■

Lemma 3.3.4 *Suppose that conditions of Theorem 3.1.1 are satisfied. Then*

$$\sup_{t\in[0,1]} \left|\hat{F}_n(t) - F_0(t)\right| = O_p(n^{-1/3}\log n).$$

Proof: Note that $\hat{F}_{0,n}^\theta$ in fact does not depend on θ and represents the estimator under the null hypothesis. For the case of contiguous alternatives it is already studied in Groeneboom, Kulikov and Lopuhaä (2002) and from this we know $\sup_{t\in[0,1]}\left|\hat{F}_{0,n}^\theta(t) - F_0(t)\right| = O_p(n^{-1/3}\log n)$. By an argument similar to one used in Lemma 3.3.3, $\hat{F}_{1,n}^\theta(t) = \left(\hat{F}_{0,n}^\theta(t)\right)^{\frac{1}{1+\theta}}$. Using Lemma 3.3.1, (3.3.4) and that $\hat{F}_n(t) = \hat{F}_n^{\hat{\theta}_n}(t)$, we obtain Lemma 3.3.4. ■

Closing this section let us prove a fact we later need in upper bound estimates:

Lemma 3.3.5 *Suppose $|\hat{\theta}_n| \le 1/2$ and $n > 1$. Then*

$$\inf\left\{\hat{F}_n(t_i) :\ \hat{F}_n(t_i) > 0\right\} > \frac{1}{9n^2},$$
$$\inf\left\{1 - \hat{F}_n(t_i) :\ \hat{F}_n(t_i) < 1\right\} > \frac{1}{5n}.$$

Proof: If $\hat{F}_n$ does not take values in the interval $(0,1)$, both statements are trivial. Otherwise, to obtain the first inequality take t_i and t_j such that $\inf\left\{\hat{F}_n(t_k) :\ \hat{F}_n(t_k) > 0\right\} = \hat{F}_n(t_i) = \hat{F}_n(t_j) < \hat{F}_n(t_{j+1})$ and $\hat{F}_n(t_{i-1}) = 0$ if any. Using (3.3.10)

$$\sum_{k=i}^j \left(\frac{\delta_k(1+\hat{\theta}_n z_k)}{x} - \frac{(1-\delta_k)(1+\hat{\theta}_n z_k)x^{\hat{\theta}_n z_k}}{1 - x^{1+\hat{\theta}_n z_k}}\right) = 0,$$

where $x = \hat{F}_n(t_i)$. At least one of δ_k, $k = i, \dots, j$ is equal to one and using $|\hat{\theta}_n| \leq 1/2$ we conclude

$$\frac{1}{2x} - \frac{3(j-i)x^{-1/2}}{2\,(1-x^{1/2})} \leq 0.$$

Next we bound $j - i \leq n - 1$ and obtain $x \geq 1/(9n^2)$.

To get the second inequality, take t_i and t_j such that $\hat{F}_n(t_{i-1}) < \hat{F}_n(t_i) = \hat{F}_n(t_j) = \sup\left\{\hat{F}_n(t_i) : \hat{F}_n(t_i) < 1\right\}$ and $\hat{F}_n(t_{j+1}) = 1$ if any. By the same argument we obtain an inequality

$$\frac{3(n-1)}{2x} - \frac{x^{1/2}}{2\,(1-x^{3/2})} \geq 0,$$

where as before $x = \hat{F}_n(t_i)$. The second statement of Lemma follows from the inequality $x \leq \left(1 - \frac{1}{3n-2}\right)^{2/3} < 1 - \frac{1}{5n}$. ∎

3.4 Stochastic equicontinuity

In this section we show stochastic equicontinuity of the "derivatives" of the loglikelihood. The method of proving the asymptotic normality of the maximum likelihood estimator $\hat{\theta}_n$ is analogous to the treatment in Huang (1996). We will consider both stochastic equicontinuity expressions from Lemma 3.4.4 and use consequences of the optimality of $(\hat{\theta}_n, \hat{F}_n)$, considered in section 3.6. Moreover we use exact calculations for the underlying model applying an expansion in terms linear in $\hat{\theta}_n$ and $\hat{F}_n$. The proof of this expansion can be found in the section 3.5. Application of the Central limit theorem and properties of the efficient score function discussed later in this section will complete the proof.

First we need some notation. For the interval $I_n = [n^{-1/3}(\log n)^3, 1 - n^{-1/3}(\log n)^3]$ let

$$\begin{aligned}
S_{1,n}(\theta, F) &= \int_{t\in I_n} l_1(\theta, F(t), \delta, z)\,dP_n,\\
S_{2,n}(\theta, F)[h] &= \int_{t\in I_n} l_2(\theta, F(t), \delta, z)h(t)\,dP_n,\\
S_1(\theta, F) &= \int_{t\in I_n} l_1(\theta, F(t), \delta, z)\,dP_{\theta_n},\\
S_2(\theta, F)[h] &= \int_{t\in I_n} l_2(\theta, F(t), \delta, z)h(t)\,dP_{\theta_n};
\end{aligned}$$

where h is a function on $[0, 1]$ and under P_{θ_n} we understand the probability measure corresponding to the underlying distribution (thus $dP_{\theta_n} = L(\theta_n, F_0)d\mu$) and under P_n - the empirical measure. Define further

$$\begin{aligned}
S_{1,1}(\theta, F) &= \int_{t\in I_n} l_1(\theta, F(t), \delta, z)^2\,dP_{\theta_n},\\
S_{1,2}(\theta, F)[h] &= \int_{t\in I_n} l_1(\theta, F(t), \delta, z)l_2(\theta, F(t), \delta, z)h(t)\,dP_{\theta_n},\\
S_{2,2}(\theta, F)[h_1, h_2] &= \int_{t\in I_n} l_2(\theta, F(t), \delta, z)^2 h_1(t)h_2(t)\,dP_{\theta_n}.
\end{aligned}$$

As we will see later this functions play role of the first and second derivatives in expansion of likelihood w.r.t. F and θ.

As we have seen in Lemma 3.2 of GROENEBOOM, KULIKOV AND LOPUHAÄ (2002), the canonical gradient $\tilde{k}$ for this model is

$$\tilde{k}(t,\delta,z) = I_{F_0}\left(\frac{\delta}{F_0(t)} - \frac{1-\delta}{1-F_0(t)}\right)\left(z - \frac{pg_1(t)}{pg_1(t)+(1-p)g_0(t)}\right)F_0(t)\log F_0(t),$$

where

$$I_{F_0} = \left(p(1-p)\int \frac{F_0(t)(\log F_0(t))^2}{1-F_0(t)}\cdot\frac{g_0(t)g_1(t)}{pg_1(t)+(1-p)g_0(t)}dt\right)^{-1}.$$

In fact, this is

$$\tilde{k}(t,\delta,z) = I_{F_0}\left(l_1(0,F_0(t),\delta,z) - l_2(0,F_0(t),\delta,z)[\mathbf{h}^*]\right)$$

with

$$\mathbf{h}^*(t) = \frac{pg_1(t)}{pg_1(t)+(1-p)g_0(t)}F_0(t)\log F_0(t).$$

Lemma 3.4.1 *Suppose that conditions of Theorem 3.1.1 are satisfied. Then*

$$\sqrt{n}\left(S_{1,n}(\theta_n,F_0) - S_{2,n}(\theta_n,F_0)[\mathbf{h}^*]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1/I_{F_0}).$$

Proof: We need to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(l_1(\theta_n,F_0(T_i),\Delta_i,Z_i) - l_2(\theta_n,F_0(T_i),\Delta_i,Z_i)\mathbf{h}^*(T_i)\right)1_{T_i\in I_n}$$

is asymptotically normal. Taking expectations, conditional on $T=t$, we show that

$$\int_{t\in I_n} l_1(\theta_n,F_0(t),\delta,z)dP_{\theta_n} = 0,$$
$$\int_{t\in I_n} l_2(\theta_n,F_0(t),\delta,z)\mathbf{h}^*(t)dP_{\theta_n} = 0.$$

The next thing is to consider the variance of Y_i defined as

$$Y_i = \left(l_1(\theta_n,F_0(T_i),\Delta_i,Z_i) - l_2(\theta_n,F_0(T_i),\Delta_i,Z_i)\mathbf{h}^*(T_i)\right)1_{T_i\in I_n}.$$

$$\begin{aligned}EY_i^2 &= \int_{t\in I_n}\frac{z\left((1-p)g_0(t)+\theta_n pg_1(t)\right)^2 + (1-z)(pg_1(t))^2}{(pg_1(t)+(1-p)g_0(t))^2}\cdot\\ &\qquad\qquad \left(\frac{\delta}{t^2} + \frac{(1-\delta)t^{2z\theta_n}}{(1-t^{1+z\theta_n})^2}\right)t^2(\log t)^2 dP_{\theta_n}\\ &= \int_{t\in I_n}\left(\frac{p(1-p)^2g_0(t)^2g_1(t)}{(pg_1(t)+(1-p)g_0(t))^2} + r_1(t)\right)\cdot\frac{t^{1+\theta_n}(\log t)^2}{1-t^{1+\theta_n}}dt\\ &\qquad + \int_{t\in I_n}\frac{p^2(1-p)g_0(t)g_1(t)^2}{(pg_1(t)+(1-p)g_0(t))^2}\frac{t(\log t)^2}{1-t}dt,\end{aligned}$$

where $r_1(t) = \frac{2\theta_n p^2(1-p)g_0(t)g_1(t)^2+\theta_n^2 p^3 g_1(t)^3}{(pg_1(t)+(1-p)g_0(t))^2}$, satisfying $|r_1(t)| \le C_1|\theta_n|(g_0(t)+g_1(t))$ for any t. Define next $r_2(t)$ as

$$r_2(t) = \left|\frac{t^{1+\theta_n}(\log t)^2}{1-t^{1+\theta_n}} - \frac{t(\log t)^2}{1-t}\right|.$$

For n sufficiently large and $t \in I_n$

$$\left|\frac{t^{1+\theta_n}(\log t)^2}{1-t^{1+\theta_n}} - \frac{t(\log t)^2}{1-t}\right| \le \frac{\left|t^{\theta_n}-1\right| t(\log t)^2}{(1-t)\,(1-t^{1+\theta_n})} \le C_2\frac{|\theta_n| t(\log t)^3}{(1-t)^2} \le C_3|\theta_n| \tag{3.4.1}$$

and

$$|EY_i^2 - 1/I_{F_0}| \le C_4 \int_{t\in I_n} (|r_1(t)| + |r_2(t)|(g_1(t)+g_2(t)))\,dt = O(\theta_n).$$

Finally, the Lindeberg condition is satisfied, since $|Y_i| \le n^{1/3}$. ∎

Lemma 3.4.2 *Suppose conditions of Theorem 3.1.1 are satisfied. Then*

$$S_{1,1}(\theta_n, F_0) - S_{1,2}(\theta_n, F_0)[\mathbf{h}^*] \to 1/I_{F_0},\ \ n \to \infty, \tag{3.4.2}$$

$$|S_{1,2}(\theta_n, F_0)[h] - S_{2,2}(\theta_n, F_0)[\mathbf{h}^*, h]| \le C_1|\theta_n| \sup_{t\in[0,1]} |h(t)|. \tag{3.4.3}$$

Proof: Rewrite (3.4.2)

$$\int_{I_n} (l_1(\theta_n, F_0(t), \delta, z) - l_2(\theta_n, F_0(t), \delta, z)\mathbf{h}^*(t))\, l_1(\theta_n, F_0(t), \delta, z) dP_{\theta_n} =$$
$$p(1-p)\int_{I_n} \frac{t^{1+\theta_n}(\log t)^2}{1-t^{1+\theta_n}} \cdot \frac{g_0(t)g_1(t)}{pg_1(t)+(1-p)g_0(t)}\,dt.$$

Use of the dominated convergence theorem finishes the proof.

On the other hand, for any function h, (3.4.3) can be written as

$$\int_{I_n} (l_1(\theta_n, F_0(t), \delta, z) - l_2(\theta_n, F_0(t), \delta, z)[\mathbf{h}^*])\, l_2(\theta_n, F_0(t), \delta, z)h(t) dP_{\theta_n} =$$
$$\int_{I_n} \left(\frac{p(1-p)g_1(t)g_2(t)}{(1-p)g_0(t)+pg_1(t)} + \theta_n \frac{p^2 g_1(t)^2}{(1-p)g_0(t)+pg_1(t)}\right) \frac{t^{\theta_n}\log t}{1-t^{1+\theta_n}} h(t)dt$$
$$- \int_{I_n} \frac{p(1-p)g_1(t)g_2(t)}{(1-p)g_0(t)+pg_1(t)} \frac{\log t}{1-t} h(t)dt.$$

Using (3.4.1), and the fact that for any $t \in (0,1)$, $|\log t|/(1-t) \le C_2(1+|\log t|)$, the absolute value of the right hand side of the above equality is bounded by

$$C_3|\theta_n| \sup_{t\in[0,1]} |h(t)| \int_{t\in I_n} \left(1+(\log t)^2\right) dt.$$

∎

To prove stochastic equicontinuity of $S_{1,n}$ and $S_{2,n}$ on sets

$$\mathcal{F}_n = \left\{F:\ F \text{ is a distribution function on } [0,1];\ \sup|F(t)-F_0(t)| \le n^{-1/3}(\log n)^2\right\}$$
$$\Theta_n = [-n^{-1/3}\log n, n^{-1/3}\log n]$$

we will use Lemma 6.1 in GROENEBOOM, KULIKOV AND LOPUHAÄ (2002):

Lemma 3.4.3 *Let P^n be a sequence of the probability measures on the set X with a sequence of subsets $\{A_n\}_{n=1}^{\infty}$. Let $\mathcal{B}$ be the set of all functions from X to $\mathbb{R}$ and let $\mathcal{A}$ be some fixed set with a sequence of subsets $\{\mathcal{A}_n\}_{n=1}^{\infty}$. Suppose $\exists a_0 \in \mathcal{A}:\ a_0 \in \bigcap_{n=1}^{\infty} \mathcal{A}_n$. Furthermore, let $\{\alpha_n\}_{n=1}^{\infty}$ be a sequence of functions $\alpha_n:\ \mathcal{A} \to \mathcal{B}$. Suppose that for some sequence ρ_{x_n}, $n = 1, 2, \dots,\ x_n \in X^n$, of metrics on $\mathcal{A}$, depending on the sample realizations $X_n = x_n$ the following is satisfied:*

1. $P^n\{A_n\} \to 1,\ n \to \infty,$
2. *For any $x_n \in A_n$ and for any $a_1, a_2 \in \mathcal{A}_n$*
$$\int (\alpha_n(a_1) - \alpha_n(a_2))^2\, dP^{x_n} \le \rho_{x_n}(a_1, a_2)^2,$$
3. *For any ϵ, $0 < \epsilon < \frac{1}{2}$*
$$\lim_{n\to\infty} \sup_{x_n \in A_n} \frac{\sup_{a\in\mathcal{A}_n} \rho_{x_n}(a, a_0)}{J_{x_n}^{-1}(\epsilon)} = 0,$$
where $J_{x_n}^{-1}(\epsilon)$ is the inverse function of $J(\cdot, \mathcal{A}, \rho_{x_n})$ (which we will further denote as $J_{x_n}(\delta)$).

Then
$$\sup_{a\in\mathcal{A}_n} \left| \int (\alpha_n(a) - \alpha_n(a_0))\, dE_n \right| = o_p(1).$$

Lemma 3.4.4 *Suppose conditions of Theorem 3.1.1 are satisfied. Then*
$$\sqrt{n}\left\{ (S_{1,n} - S_1)(\hat\theta_n, \hat F_n) - (S_{1,n} - S_1)(\theta_n, F_0) \right\} = o_p(1),$$
$$\sqrt{n}\left\{ (S_{2,n} - S_2)(\hat\theta_n, \hat F_n)[\mathbf{h}^*] - (S_{2,n} - S_2)(\theta_n, F_0)[\mathbf{h}^*] \right\} = o_p(1).$$

Proof: Using Lemmas 3.3.1 and 3.3.4 and separating terms, corresponding to δ and $(1-\delta)$ it suffices to show that, for $q(t) = pg_1(t)/\{pg_1(t) + (1-p)g_0(t)\}$,

$$A.\ \sup_{F\in\mathcal{F}_n} \left| \int_{t\in I_n} z\delta \left(\log F(t) - \log F_0(t)\right) dE_n \right| = o_p(1),$$
$$B.\ \sup_{F\in\mathcal{F}_n, \theta\in\Theta_n} \left| \int_{t\in I_n} z(1-\delta) \left(\frac{F(t)^{1+\theta}\log F(t)}{1-F(t)^{1+\theta}} - \frac{F_0(t)\log F_0(t)}{1-F_0(t)} \right) dE_n \right| = o_p(1),$$
$$C.\ \sup_{F\in\mathcal{F}_n, \theta\in\Theta_n} \left| \int_{t\in I_n} \delta \left(\frac{1+z\theta}{F(t)} - \frac{1}{F_0(t)} \right) F_0(t)\log F_0(t) q(t) dE_n \right| = o_p(1),$$
$$D.\ \sup_{F\in\mathcal{F}_n, \theta\in\Theta_n} \left| \int_{t\in I_n} (1-\delta) \left(\frac{(1+z\theta)F(t)^{z\theta}}{1-F(t)^{1+z\theta}} - \frac{1}{1-F_0(t)} \right) F_0(t)\log F_0(t) q(t) dE_n \right| = o_p(1).$$

To prove it we apply the same technique to all terms A-D. As an example we treat term C. Apply Lemma 3.4.3 with $X = \mathbb{R} \times \{0,1\}^2$, a sequence of probability measures P_{θ_n}, $\mathcal{A} = \Theta \times \mathcal{F}$ and $\mathcal{A}_n = \Theta_n \times \mathcal{F}_n$. The intersection $\bigcap \mathcal{A}_n = (0, F_0)$ will be denoted by a_0. It will follow from

Lemma 3.4.3 that C is asymptotically of smaller order in probability, if we show that for the function $\alpha_n : \mathcal{A} \to \mathcal{B}$ defined by

$$\alpha_n(a) = \delta \frac{1+z\theta}{F(t)} q(t) F_0(t) \log F_0(t) 1_{\{t\in I_n\}},$$

where $\mathcal{B}$ is the set of all functions from X to $\mathbb{R}$, conditions of Lemma 3.4.3 are satisfied.

For any functions $F_1, F_2 \in \mathcal{F}_n$, any $\theta \in \Theta_n$ and $t \in I_n$ we have:

$$\begin{aligned} |\alpha_n(a_1) - \alpha_n(a_2)| &\le z\delta|\theta_1 - \theta_2| \frac{t\log t}{F_1(t)} + 2\frac{|F_1(t) - F_2(t)|}{F_1(t)F_2(t)} t \log t \delta 1_{t\in I_n} \\ &\le C_1 \left(|\theta_1 - \theta_2| + \frac{|F_1(t) - F_2(t)|}{t} \right) \delta \log t 1_{t\in I_n} \end{aligned}$$

for n sufficiently large, since for $n > 2$, we have

$$\sup_{F\in\mathcal{F}_n, t\in I_n} \frac{t}{F(t)} \le \frac{n^{-1/3}(\log n)^3}{n^{-1/3}(\log n)^2(\log n - 1)}.$$

Therefore, for the metric ρ_{x_n}, defined by $\rho_{x_n}(a_1, a_2) = \rho'_{x_n}(\theta_1, \theta_2) + \rho''_{x_n}(F_1, F_2)$, where

$$\rho'_{x_n}(\theta_1, \theta_2) = |\theta_1 - \theta_2| C_2 \left(\int_{t\in I_n} (\log t)^2 \delta dP_n \right)^{1/2},$$

$$\rho''_{x_n}(F_1, F_2) = C_2 \left(\int_{t\in I_n} \frac{(\log t)^2}{t^2} (F_1(t) - F_2(t))^2 \delta dP_n \right)^{1/2},$$

it holds that, for any $a_1, a_2 \in \mathcal{A}_n$ and n sufficiently large,

$$\int (\alpha_n(a_1) - \alpha_n(a_2))^2 dP_{x_n} \le \rho_{x_n}(a_1, a_2)^2.$$

Define the set A_n by

$$A_n = \left\{ x_n \in X^n : \int_{t\in I_n} \delta(\log t)^2 dP_n \le \log n \text{ and } \int_{t\in I_n} \frac{(\log t)^2}{t^2} \delta dP_n \le (\log n)^4 \right\}.$$

By the Markov inequality $P\{A_n^c\} \to 0,\ n \to \infty$, since, for large n, expectations are bounded as

$$\int_{t\in I_n} \delta(\log t)^2 dP_{\theta_n} \le C_3 \int_0^1 \sqrt{t}(\log t)^2 dt < \infty$$

and

$$\int_{t\in I_n} \frac{(\log t)^2}{t^2} \delta P_{\theta_n} = \int_{I_n} (\log t)^2 \left(t^{\theta_n - 1} p g_1(t) + t^{-1}(1-p) g_0(t) \right) dt = o\left((\log n)^4 \right).$$

Therefore for any $x_n \in A_n$ holds that $\sup_{a\in\mathcal{A}_n} \rho_{x_n}(a, a_0) \le C_4 n^{-1/3}(\log n)^4$. To bound $J_{x_n}^{-1}(\epsilon)$ from below we need to bound the entropy from above. Since a δ-net in $\mathcal{A}$ can be chosen as a product of $\delta/2$-nets in Θ and $\mathcal{F}$,

$$\log N(\delta, \mathcal{A}, \rho_{x_n}) \le \log N(\delta/2, \Theta, \rho'_{x_n}) + \log N(\delta/2, \mathcal{F}, \rho''_{x_n}).$$

Uniform bounds on the entropy of the set $\mathcal{F}$ w.r.t. the $L_2(\mu)$-norm for a bounded positive measure μ were studied in BIRMAN AND SOLOMJAK (1967). This gives, for any $x_n \in A_n$,

$$\log N(\delta/2, \mathcal{F}, \rho''_{x_n}) \leq \frac{C_5(\log n)^2}{\delta}.$$

Boundary $N(\delta/2, \Theta, \rho'_{x_n}) \leq \frac{4\sqrt{\log n}}{\delta}$ is trivial and therefore

$$J(\delta, \mathcal{A}, \rho_{x_n}) = \int_0^\delta \sqrt{\log\left(\frac{N(t, \mathcal{A}, \rho_{x_n})^2}{t}\right)} dt \leq C_6\sqrt{\delta}\log n.$$

Condition 3 of Lemma 3.4.3 is therefore satisfied with

$$\sup_{x_n \in A_n} \frac{\sup_{a\in\mathcal{A}_n} \rho_{x_n}(a, a_0)}{J_{x_n}^{-1}(\epsilon)} \leq C_7\epsilon^{-2}n^{-1/3}(\log n)^6.$$

■

3.5 Smoothness

We first obtain an approximation of $S_1(\theta, F)$ and $S_2(\theta, F)[\mathbf{h}^*]$ in a neighborhood of $(0, F_0)$ by terms linear in θ and $(F - F_0)$. To prove asymptotic normality of $\sqrt{n}\hat{\theta}_n$ we will need the error to be $o_p(n^{-1/2})$.

Lemma 3.5.1 *Suppose conditions of Theorem 3.1.1 are satisfied. Then*

$$\sup_{F\in\mathcal{F}_n, \theta\in\Theta_n} \{|S_1(\theta, F) - S_1(\theta_n, F_0) + S_{1,1}(\theta_n, F_0)(\theta - \theta_n) + S_{1,2}(\theta_n, F_0)[F - F_0]|\} = o(n^{-1/2}),$$

$$\sup_{F\in\mathcal{F}_n, \theta\in\Theta_n} \{|S_2(\theta, F)[\mathbf{h}^*] - S_2(\theta_n, F_0)[\mathbf{h}^*] + S_{1,2}(\theta_n, F_0)[\mathbf{h}^*](\theta - \theta_n) + S_{2,2}(\theta_n, F_0)[\mathbf{h}^*, F - F_0]|\} = o\left(n^{-1/2}\right).$$

Proof: The argument is analogous for both expressions. Therefore we will only prove the first statement of the lemma. We have:

$$\begin{aligned} S_1(\theta, F) &= \int_{t\in I_n} z \log F(t) \left(\delta - \frac{(1-\delta)F(t)^{1+\theta}}{1 - F(t)^{1+\theta}}\right) dP_{\theta_n} \\ &= p\int_{t\in I_n} \log F(t) \left(t^{1+\theta_n} - \frac{(1 - t^{1+\theta_n})F(t)^{1+\theta}}{1 - F(t)^{1+\theta}}\right) g_1(t)dt. \end{aligned}$$

Expand this, for $t \in I_n$, $\theta \in \Theta_n$ and $F \in \mathcal{F}_n$,

$$\log F(t)\left(1 - \frac{1-t^{1+\theta_n}}{1-F(t)^{1+\theta}}\right) = -\left\{\frac{1 - t^{1+\theta_n}}{(1 - u^{1+v})^2}u^{1+v}(\log u)^2\right\}(\theta - \theta_n) \tag{3.5.1}$$

$$+\left\{\frac{1}{u}\left(1 - \frac{1 - t^{1+\theta_n}}{1 - u^{1+v}}\right) - \frac{1 - t^{1+\theta_n}}{(1 - u^{1+v})^2}(1 + v)u^\theta \log u\right\} \cdot \tag{3.5.2}$$

$$(F(t) - t),$$

where u lies between t and $F(t)$ and v between θ_n and θ. We will also use that $S_1(\theta_n, F_0) = 0$ and

$$S_{1,1}(\theta_n, F_0)(\theta-\theta_0) = p(\theta-\theta_0)\int_{t\in I_n}(\log t)^2\frac{t^{1+\theta_n}}{1-t^{1+\theta_n}}g_1(t)dt, \tag{3.5.3}$$

$$S_{1,2}(\theta_n, F_0)[F-F_0] = p\int_{t\in I_n}\log t\frac{t^{\theta_n}(1+\theta_n)}{1-t^{1+\theta_n}}(F(t)-t)g_1(t)dt. \tag{3.5.4}$$

Consider the difference between (3.5.1) and integrand of (3.5.3). Let $w_1 = u^{1+v}$ and $w_2 = t^{1+\theta_n}$. Further define

$$\mathcal{F}_n' = \left\{F:\ F \text{ is a distribution function, } \sup_{t\in[0,1]}|t-F(t)| \le 2n^{-1/3}(\log n)^2\right\}.$$

From (3.3.4) we have that both $w_1(t)$ and $w_2(t)$ belong to $\mathcal{F}'$ for large n. For the same reason,

$$\frac{1}{2} \le \inf_{t\in I_n}\frac{1-w_i(t)}{1-t} \le \sup_{t\in I_n}\frac{1-w_i(t)}{1-t} \le 2, \quad i = 1,2.$$

Then we get for the difference of integrands:

$$(1-w_2)\left|\frac{1}{(1+v)^2}\frac{w_1(\log w_1)^2}{(1-w_1)^2} - \frac{1}{(1+\theta_n)^2}\frac{w_2(\log w_2)^2}{(1-w_2)^2}\right| \le C_1\left(1+(\log t)^2\right)n^{-1/3}(\log n)^2.$$

To obtain the last inequality we have used the boundednes of the function $t(\log t)^2/(1-t)^2$ on the interval $(0,1)$, and the bound

$$\left|\left(\frac{t(\log t)^2}{(1-t)^2}\right)'\right| \le C_2\frac{1+(\log t)^2}{1-t} \quad \text{for any } t\in(0,1).$$

Integrating we find that the difference of θ-terms is bounded by

$$pC_1 n^{-2/3}(\log n)^3\int_{I_n}(1+(\log t)^2)g_1(t)dt = o\left(n^{-1/2}\right).$$

Next consider the difference between (3.5.2) and integrand of (3.5.4). First we bound

$$\frac{1}{u}\left|1-\frac{1-t^{1+\theta_n}}{1-u^{1+v}}\right| \le \frac{|w_1-w_2|}{(1-w_1)u} \le \frac{C_3 n^{-1/3}(\log n)^2}{t(1-t)}.$$

Using the inequality

$$\sup_{F\in\mathcal{F}_n,\theta\in\Theta_n,t\in I_n}\left|F(t)^\theta - 1\right| \le C_4 n^{-1/3}(\log n)^3$$

we conclude that

$$\left|\frac{1-w_2}{(1-w_1)^2}u^\theta\log w_1 - \frac{1}{1-w_2}t^{\theta_n}\log w_2\right| \le$$
$$(1-w_2)\left|\frac{\log w_1}{(1-w_1)^2} - \frac{\log w_2}{(1-w_2)^2}\right| + C_5\frac{n^{-1/3}(\log n)^4}{1-t} \le C_6\frac{n^{-1/3}(\log n)^4}{t(1-t)},$$

using the inequality

$$\left|\left(\frac{\log t}{(1-t)^2}\right)'\right| \le \frac{C_7}{t(1-t)^2}.$$

Integrating we get

$$C_7 n^{-2/3}(\log n)^6 \int_{t\in I_n} \frac{g_1(t)}{t(1-t)}\,dt = o\left(n^{-1/2}\right).$$

■

3.6 Two important lemmas

In this section we discuss consequences of the optimality of the maximum likelihood pair $(\hat{\theta}_n, \hat{F}_n)$ for $S_{1,n}$ and $S_{2,n}[\mathbf{h}^*]$. The next two lemmas together with the results of sections 3.4 and 3.5 will be applied in showing that $\hat{\theta}_n$ is asymptotically normal.

Lemma 3.6.1 *Suppose that conditions of Theorem 3.1.1 are satisfied. Then*

$$S_{1,n}(\hat{\theta}_n, \hat{F}_n) = o_p\left(n^{-1/2}\right).$$

Proof: The maximizer of

$$\theta \mapsto \int l(\theta, \hat{F}_n(t), \delta, z)\,dP_n$$

(seen as a differentiable function of θ) is $\hat{\theta}_n$. This implies

$$\int l_1(\hat{\theta}_n, \hat{F}_n(t), \delta, z)\,dP_n = 0$$

and to prove Lemma 3.6.1 we need to show that

$$\int_{t\in[0,1]\setminus I_n} l_1(\hat{\theta}_n, \hat{F}_n(t), \delta, z)\,dP_n = o_p\left(n^{-1/2}\right).$$

Denote the intervals

$$I_n^l = \left[0, n^{-1/3}(\log n)^3\right] \quad \text{and} \quad I_n^r = \left[1 - n^{-1/3}(\log n)^3, 1\right].$$

Separating integrals of δ- and $(1-\delta)$-terms over I_n^l and I_n^r it suffices to prove

$$\int_{t\in I_n^l} z\delta\left|\log \hat{F}_n(t)\right| dP_n = o_p\left(n^{-1/2}\right),$$

$$\int_{t\in I_n^r} z\delta\left|\log \hat{F}_n(t)\right| dP_n = o_p\left(n^{-1/2}\right),$$

$$\int_{t\in I_n^l} z(1-\delta)\left|\log \hat{F}_n(t)\right| \frac{\hat{F}_n(t)^{1+\hat{\theta}_n}}{1-\hat{F}_n(t)^{1+\hat{\theta}_n}}\,dP_n = o_p\left(n^{-1/2}\right),$$

$$\int_{t\in I_n^r} z(1-\delta)\left|\log \hat{F}_n(t)\right| \frac{\hat{F}_n(t)^{1+\hat{\theta}_n}}{1-\hat{F}_n(t)^{1+\hat{\theta}_n}}\,dP_n = o_p\left(n^{-1/2}\right).$$

Here we only consider the first and the third expressions since the rest is treated analogously.

$$\int_{t\in I_n^l} z\delta \left|\log \hat{F}_n(t)\right| dP_n \leq C_1 \log n \int_{I_n^l} z\delta dP_n = o_p\left(n^{-1/2}\right)$$

by an application of Lemma 3.3.5 and the Markov inequality.

Using boundednes of the function $t^{1/3}\log t/(1-t)$ on the interval $(0,1)$ we estimate

$$\int_{t\in I_n^l} z(1-\delta)\left|\log \hat{F}_n(t)\right| \frac{\hat{F}_n(t)^{1+\hat{\theta}_n}}{1-\hat{F}_n(t)^{1+\hat{\theta}_n}} dP_n \leq \frac{C_2}{1+\hat{\theta}_n}\hat{F}_n\left(n^{-1/3}(\log n)^3\right)^{2(1+\hat{\theta}_n)/3}\int_{t\in I_n^l} dP_n$$

which is $o_p\left(n^{-1/2}\right)$ due to the Lemmas 3.3.1, 3.3.4 and the Markov inequality. ■

Lemma 3.6.2 *Suppose that conditions of Theorem 3.1.1 are satisfied. Then*

$$S_{2,n}(\hat{\theta}_n, \hat{F}_n)[\mathbf{h}^*] = o_p\left(n^{-1/2}\right).$$

Proof: The function $\hat{F}_n(t) + \epsilon\mathbf{h}^*(t)$ can not be a distribution function for both positive and negative ϵ. To use optimality of $\hat{F}_n$ we consider first $h_n(t) = q_n(t)\hat{F}_n(t)\log \hat{F}_n(t)$, where q_n is a piecewise constant version of q, defined by $q_n(t) = q\left(F_0^{-1}\left(\hat{F}_n(t)\right)\right)$. We also use the notation

$$\epsilon_n^{'} = \min\left\{\left|\hat{F}_n(u) - \hat{F}_n(v)\right| : \ \hat{F}_n(u) \neq \hat{F}_n(v)\right\}$$

and

$$\epsilon_n = \min\left\{\epsilon_n^{'}, \hat{F}_n(k_n), 1 - \hat{F}_n(1-k_n)\right\},$$

where $k_n = n^{-1/3}(\log n)^3$. For τ_n^l and τ_n^r defined by

$$\tau_n^l = \min\{\tau_i : \ \tau_i \geq k_n\},$$
$$\tau_n^r = \max\{\tau_i : \ \tau_i \leq 1 - k_n\},$$

where the τ_i are the points of jump of $\hat{F}_n(t)$, it follows from Lemma 3.3.4 that intervals $J_n^l = [k_n, 2k_n]$ and $J_n^r = [1-2k_n, 1-k_n]$ usually contain τ_n^l and τ_n^r:

$$P\{\tau_n > 2k_n\} \leq P\left\{\sup_{t\in[0,1]}\left|\hat{F}_n(t) - F_0(t)\right| > \frac{1}{2}\left(F_0(2k_n) - F_0(k_n)\right)\right\} \to 0, \quad n\to\infty$$

and, for the same reason, $P\{\tau_n^r < 1 - 2k_n\} \to 0, \ n \to \infty$.

Finally define $F_{n,\epsilon}(t) = \hat{F}_n(t) + \epsilon h_n(t) 1_{t\in[\tau_n^l,\tau_n^r]}$, for any ϵ, $|\epsilon| \leq \epsilon_n/2\sup_{t\in I_n}|h_n(t)|$. For such ϵ the function $F_{n,\epsilon}$ remains a distribution function with probability tending to one and therefore

$$\int l(\hat{\theta}_n, F_{n,\epsilon}(t), \delta, z)dP_n \leq \int l(\hat{\theta}_n, \hat{F}_n(t), \delta, z)dP_n.$$

Differentiating w.r.t. ϵ we conclude

$$\int_{t\in[\tau_n^l,\tau_n^r]} l_2(\hat{\theta}_n, \hat{F}_n(t), \delta, z)h_n(t)dP_n = 0.$$

Therefore we need to show

$$\int_{t\in J_n^l \bigcup J_n^r} \left| l_2(\hat{\theta}_n, \hat{F}_n(t), \delta, z) h_n(t) \right| dP_n = o_p\left(n^{-1/2}\right).$$

Separating as in Lemma 3.6.1 and using boundednes of q it suffices to prove:

$$\sup_{F\in\mathcal{F}_n,\theta\in\Theta_n} \int_{t\in J_n^l} (1+z\theta)\delta|\log F(t)|dP_n = o_p\left(n^{-1/2}\right),$$

$$\sup_{F\in\mathcal{F}_n,\theta\in\Theta_n} \int_{t\in J_n^r} (1+z\theta)\delta|\log F(t)|dP_n = o_p\left(n^{-1/2}\right),$$

$$\sup_{F\in\mathcal{F}_n,\theta\in\Theta_n} \int_{t\in J_n^l} \frac{(1+z\theta)F(t)^{1+z\theta}|\log F(t)|}{1-F(t)^{1+z\theta}}(1-\delta)dP_n = o_p\left(n^{-1/2}\right),$$

$$\sup_{F\in\mathcal{F}_n,\theta\in\Theta_n} \int_{t\in J_n^r} \frac{(1+z\theta)F(t)^{1+z\theta}|\log F(t)|}{1-F(t)^{1+z\theta}}(1-\delta)dP_n = o_p\left(n^{-1/2}\right).$$

We are integrating over subsets of I_n and therefore inequalities, used in Lemmas 3.4.4 and 3.5.1, can be applied here as well. Consider, for example, the third expression:

$$\sup_{F\in\mathcal{F}_n,\theta\in\Theta_n} \int_{t\in J_n^l} \frac{(1+z\theta)F(t)^{1+z\theta}|\log F(t)|}{1-F(t)^{1+z\theta}}(1-\delta)dP_n \leq C_1 \int_{t\in J_n^l} t|\log t|dP_n$$

which is $o_p\left(n^{-1/2}\right)$ using the Markov inequality and the fact that, for n sufficiently large, $\sup_{F\in\mathcal{F}_n,\theta\in\Theta_n,t\in J_n^l} F(t)^\theta / \left\{1-F(t)^{1+\theta}\right\} \leq 2$.

Next we will prove

$$\int_{t\in I_n} l_2(\theta_n, F_0(t), \delta, z)\left(\mathbf{h}^*(t) - h_n(t)\right) dP_n = o_p\left(n^{-1/2}\right).$$

Considering expectations w.r.t. P_{θ_n}, conditional on $T=t$, we obtain

$$\int_{t\in I_n} l_2(\theta_n, F_0(t), \delta, z)\left(\mathbf{h}^*(t) - h_n(t)\right) dP_{\theta_n} = 0.$$

Therefore we must show

$$\sup_{F\in\mathcal{F}_n,\tilde{q}\in\mathcal{Q}_n} \left| \int_{t\in I_n} l_2(\theta_n, F_0(t), \delta, z)\left(\tilde{q}(t)F(t)\log F(t) - q(t)F_0(t)\log F_0(t)\right) dE_n \right| = o_p(1), \quad (3.6.1)$$

where

$$\mathcal{Q} = \left\{\tilde{q}(t):\ \tilde{q}(t) = q(s_i),\ s_i \leq t < s_{i+1}\right\},$$
$$\mathcal{Q}_n = \left\{\tilde{q}(t):\ \tilde{q}(t) = q(s_i),\ s_i \leq t < s_{i+1},\ \sup|s_{i+1}-s_i| \leq n^{-1/3}(\log n)^2\right\}.$$

As we have seen in the current proof, $P\{q_n \in \mathcal{Q}_n\} \to 1,\ n\to\infty$. To prove (3.6.1) it is sufficient to show that

$$A.\quad \sup_{F\in\mathcal{F}_n,\tilde{q}\in\mathcal{Q}_n} \left| \int_{t\in I_n} \frac{\delta(1+z\theta_n)}{F_0(t)}\left(\tilde{q}(t)F(t)\log F(t) - q(t)F_0(t)\log F_0(t)\right) dE_n \right| = o_p(1),$$

$$B.\quad \sup_{F\in\mathcal{F}_n,\tilde{q}\in\mathcal{Q}_n} \left| \int_{t\in I_n} \frac{(1-\delta)(1+z\theta_n)F_0(t)^{\theta_n z}}{1-F_0(t)^{1+\theta_n z}} \right.$$
$$\left. \cdot\left(\tilde{q}(t)F(t)\log F(t) - q(t)F_0(t)\log F_0(t)\right) dE_n \right| = o_p(1).$$

The proofs of A and B are similar and here we will consider only A. Apply Lemma 3.4.3 with $X = \mathbb{R} \times \{0,1\}^2$, $\mathcal{A} = \mathcal{F} \times \mathcal{Q}$, $\mathcal{A}_n = \mathcal{F}_n \times \mathcal{Q}_n$ and $a_0 = (F_0, q)$. Define $\alpha_n(a)$ as

$$\alpha_n(a) = \frac{\delta(1+z\theta_n)}{t}\tilde{q}(t)F(t)\log F(t)1_{t\in I_n}.$$

Using boundednes of q and the fact that, for large n, any $F_1, F_2 \in \mathcal{F}_n$ and any $t \in I_n$,

$$|F_1(t)\log F_1(t) - F_2(t)\log F_2(t)| \le C_1\left(|\log t| + 1\right)|F_1(t) - F_2(t)|,$$

uniformly in t, we get the bound

$$|\alpha_n(a_1) - \alpha_n(a_2)| \le C_2\left(\frac{|F_1(t) - F_2(t)|}{t} + |q_1(t) - q_2(t)|\right)\delta(|\log t| + 1)1_{t\in I_n}.$$

If we define metrics ρ'_{x_n} on $\mathcal{F}$ and ρ''_{x_n} on $\mathcal{Q}$

$$\rho'_{x_n}(F_1, F_2) = C_3\left(\int_{t\in I_n}\frac{\delta((\log t)^2+1)}{t^2}(F_1(t) - F_2(t))^2 dP_n\right)^{1/2},$$

$$\rho''_{x_n}(\tilde{q}_1, \tilde{q}_2) = C_3\left(\int_{t\in I_n}\delta((\log t)^2+1)(\tilde{q}_1(t) - \tilde{q}_2(t))^2 dP_n\right)^{1/2},$$

then for $\rho_{x_n}(a_1, a_2) = \rho'_{x_n}(F_1, F_2) + \rho''_{x_n}(\tilde{q}_1, \tilde{q}_2)$:

$$\int(\alpha_n(a_1) - \alpha_n(a_2))^2 dP_n \le \rho_{x_n}(a_1, a_2)^2.$$

For A_n given by

$$\left\{x_n \in X^n : \int_{t\in I_n}\frac{\delta((\log t)^2+1)}{t^2}dP_n \le (\log n)^4 \text{ and } \int_{t\in I_n}\delta((\log t)^2+1)dP_n \le \log n\right\}$$

the Markov inequality implies $P\{A_n\} \to 1,\ n \to \infty$.

It is proved in GROENEBOOM, KULIKOV AND LOPUHAÄ (2002) that, under condition (C3),

$$\log N(\delta, \mathcal{Q}, \rho''_{x_n}) \le C_4\delta^{-1}\sqrt{\log n},$$

uniformly in $x_n \in A_n$ and $\sup_{t\in[0,1]}|q'(t)| < \infty$. Together with $\log N(\delta, \mathcal{F}, \rho'_{x_n}) \le C_4\delta^{-1}(\log n)^2$, which is obtained in BIRMAN AND SOLOMJAK (1967), and arguing similarly as in Lemma 3.4.4, we obtain

$$\sup_{x_n\in A_n, a\in\mathcal{A}_n}\rho_{x_n}(a, a_0) \le n^{-1/3}(\log n)^4,$$

$$\inf_{x_n\in A_n} J^{-1}(\epsilon, \mathcal{A}, \rho_{x_n}) \ge \frac{C_5}{(\log n)^2\epsilon^2}.$$

Hence the conditions of Lemma 3.4.3 are satisfied, which finishes the proof of A.

Therefore we now need to show

$$\int_{t\in I_n}\left(l_2(\hat{\theta}_n, \hat{F}_n(t), \delta, z) - l_2(\theta_n, F_0(t), \delta, z)\right)(\mathbf{h}^*(t) - h_n(t))\,dP_n = o_p\left(n^{-1/2}\right).$$

It suffices to prove that

$$\sup_{F\in\mathcal{F}_n,\tilde{q}\in\mathcal{Q}_n,\theta\in\Theta_n}\left|\int_{t\in I_n}(l_2(\theta,F(t),\delta,z)-l_2(\theta_n,F_0(t),\delta,z))\right.$$
$$\left.(\tilde{q}(t)F(t)\log F(t)-q(t)F_0(t)\log F_0(t))\,dP_n\right| = o_p\left(n^{-1/2}\right).$$

Recalling the argument above this can be reduced to

$$\sup_{F\in\mathcal{F}_n,\tilde{q}\in\mathcal{Q}_n,\theta\in\Theta_n}\left|\int_{t\in I_n}\left(\frac{\delta}{t^2}+\frac{1-\delta}{t(1-t)^2}\right)(1+(\log t)^2)dP_n\right|$$
$$\cdot\left(\sup_{t\in[0,1]}|F(t)-F_0(t)|+\sup_{t\in[0,1]}|\tilde{q}(t)-q(t)|+|\theta-\theta_n|\right)^2.$$

Application of the Markov inequality finishes the proof of the lemma. ■

3.7 Proof of Theorem 3.1.1

Using Lemmas 3.4.1, 3.6.1 and 3.6.2, the statement of the Lemma 3.4.4 can be rewritten as

$$\begin{cases}\sqrt{n}\left(S_1(\hat{\theta}_n,\hat{F}_n)+S_{1,n}(\theta_n,F_0)\right)=o_p(1)\\ \sqrt{n}\left(S_2(\hat{\theta}_n,\hat{F}_n)[\mathbf{h}^*]+S_{2,n}(\theta_n,F_0)[\mathbf{h}^*]\right)=o_p(1).\end{cases}$$

Next we apply Lemma 3.5.1 to expand:

$$\begin{cases}\sqrt{n}\left(S_{1,n}(\theta_n,F_0)-S_{1,1}(\theta_n,F_0)(\hat{\theta}_n-\theta_n)-S_{1,2}(\theta_n,F_0)[\hat{F}_n-F_0]\right)=o_p(1)\\ \sqrt{n}\left(S_{2,n}(\theta_n,F_0)[\mathbf{h}^*]-S_{1,2}(\theta_n,F_0)[\mathbf{h}^*](\hat{\theta}_n-\theta_n)-S_{2,2}(\theta_n,F_0)[\mathbf{h}^*,\hat{F}_n-F_0]\right)=o_p(1).\end{cases}$$

Hence, using Lemmas 3.3.1, 3.3.4 and 3.4.2 we get

$$\sqrt{n}(\hat{\theta}_n-\theta_n)=I_{F_0}\sqrt{n}\left(S_{1,n}(\theta_n,F_0)-S_{2,n}(\theta_n,F_0)[\mathbf{h}^*]\right)+o_p(1). \tag{3.7.1}$$

Application of Lemma 3.4.1 now leads to a proof of the theorem:

Theorem 3.7.1 *Suppose conditions of Theorem 3.1.1 are satisfied. Then*

$$\sqrt{n}\hat{\theta}_n\xrightarrow{\mathcal{D}}\mathcal{N}\left(\theta_0,I_{F_0}\right),\ n\to\infty.$$

Let us define the second derivatives of $l(\theta,x,\delta,z)$:

$$l_{1,1}(\theta,x,\delta,z)=\frac{\partial l_1(\theta,x,\delta,z)}{\partial\theta}=-\frac{(1-\delta)z(\log x)^2x^{1+\theta}}{(1-x^{1+\theta})^2},$$
$$l_{1,2}(\theta,x,\delta,z)=\frac{\partial l_1(\theta,x,\delta,z)}{\partial x}=\frac{z\delta}{x}-\frac{(1-\delta)zx^{\theta}}{1-x^{1+\theta}}+\frac{(1-\delta)z(1+\theta)x^{\theta}\log x}{(1-x^{1+\theta})^2},$$

$$
\begin{aligned}
l_{2,2}(\theta, x, \delta, z) &= \frac{\partial l_2(\theta, x, \delta, z)}{\partial x} \\
&= -\frac{\delta(1+z\theta)}{x^2} - \frac{(1-\delta)z(1+\theta)\theta x^{1-\theta}}{(1-x^{1+\theta})^2} - \frac{(1-\delta)(1+\theta z)x^{2\theta z}}{(1-x^{1+\theta z})^2}.
\end{aligned}
$$

We will also need the piecewise constant versions

$$
\begin{aligned}
q_n(t) &= q\left(F_0^{-1}\left(\hat{F}_n(t)\right)\right), \\
q_n^0(t) &= q\left(F_0^{-1}\left(\hat{F}_n^0(t)\right)\right),
\end{aligned}
$$

where $\hat{F}_n^0 = \operatorname{argmax}_F l_n(0, F)$. Therefore the piecewise constant versions of the efficient score function will be

$$
\begin{aligned}
h_n(t) &= q_n(t)\hat{F}_n(t)\log \hat{F}_n(t), \\
h_n^0(t) &= q_n^0(t)\hat{F}_n^0(t)\log \hat{F}_n^0(t).
\end{aligned}
$$

The next three lemmas are important in proving Theorem 3.1.1. They allow us to apply the Taylor expansion up to the second power of the argument deviation. The following lemma was proved in GROENEBOOM, KULIKOV AND LOPUHAÄ (2002):

Lemma 3.7.1 *Suppose conditions of Theorem 3.1.1 are satisfied. Then*

$$
\int \left(l_1(0, \hat{F}_n^0(t), \delta, z) - l_2(0, \hat{F}_n^0(t), \delta, z)h_n^0(t)\right) dP_n = \\
\theta_n / I_{F_0} + ((S_{1,n}(0, F_0) - S_{2,n}(0, F_0)[\mathbf{h}^*]) - (S_1(0, F_0) - S_2(0, F_0)[\mathbf{h}^*])) + o_p\left(n^{-1/2}\right).
$$

Proof: As we have seen in (3.3.10)

$$
\int l_2(0, \hat{F}_n^0(t), \delta, z)h_n^0(t) dP_n = 0.
$$

In Lemma 5.2 of GROENEBOOM, KULIKOV AND LOPUHAÄ (2002) we obtained:

$$
\int l_1(0, \hat{F}_n^0(t), \delta, z) dP_n = \\
\theta_n / I_{F_0} + \int_{I_n} \left(l_1(0, F_0(t), \delta, z) - l_2(0, F_0(t), \delta, z)\mathbf{h}^*(t)\right) d(P_n - P_{\theta_n}) + o_p\left(n^{-1/2}\right).
$$

■

But for the proof we need another representation of this expression, given in the next lemma.

Lemma 3.7.2 *Suppose the conditions of Theorem 3.1.1 are satisfied. Then*

$$
\int \left(l_1(0, \hat{F}_n^0(t), \delta, z) - l_2(0, \hat{F}_n^0(t), \delta, z)h_n^0(t)\right) dP_n = \\
\theta_n / I_{F_0} + (S_{1,n}(\theta_n, F_0) - S_{2,n}(\theta_n, F_0)[\mathbf{h}^*]) + o_p\left(n^{-1/2}\right).
$$

Proof: Using Lemmas 3.4.1 and 3.7.1, it is seen that the statement of Lemma 3.7.2 is equivalent to

$$((S_{1,n}(0,F_0) - S_{2,n}(0,F_0)[\mathbf{h}^*]) - (S_1(0,F_0) - S_2(0,F_0)[\mathbf{h}^*])) = \\ ((S_{1,n}(\theta_n,F_0) - S_{2,n}(\theta_n,F_0)[\mathbf{h}^*]) - (S_1(\theta_n,F_0) - S_2(\theta_n,F_0)[\mathbf{h}^*])) + o_p\left(n^{-1/2}\right).$$

Now proof follows by the application of Lemma 3.4.4. ∎

Continuity of the second order derivatives is stated in the next Lemma. Its proof is in fact a series of applications of Taylor expansion, as in the proof of Lemma 3.5.1. Here we will state it without proof:

Lemma 3.7.3 *Suppose the conditions of Theorem 3.1.1 are satisfied. Let* $\left|\tilde{\theta}_n^0\right| < \left|\hat{\theta}_n\right|$ *and* $\tilde{F}_n^0(t)$ *be a function on* $[0,1]$*, such that for any* $t \in [0,1]$ $\tilde{F}_n^0(t)$ *lies between* $\hat{F}_n^0(t)$ *and* $\hat{F}_n^0(t) - \hat{\theta}_n h_n^0(t)$*. Then*

$$\sup_{\tilde{F}_n^0,\tilde{\theta}_n^0}\left|\int \left(l_{1,1}(\tilde{\theta}_n^0,\tilde{F}_n^0(t),\delta,z) - l_{1,1}(0,F_0(t),\delta,z)\right) dP_n\right| = o_p(1),$$

$$\sup_{\tilde{F}_n^0,\tilde{\theta}_n^0}\left|\int \left(l_{1,2}(\tilde{\theta}_n^0,\tilde{F}_n^0(t),\delta,z)h_n^0(t) - l_{1,2}(0,F_0(t),\delta,z)\mathbf{h}^*(t)\right) dP_n\right| = o_p(1),$$

$$\sup_{\tilde{F}_n^0,\tilde{\theta}_n^0}\left|\int \left(l_{2,2}(\tilde{\theta}_n^0,\tilde{F}_n^0(t),\delta,z)h_n^0(t)^2 - l_{2,2}(0,F_0(t),\delta,z)\mathbf{h}^*(t)^2\right) dP_n\right| = o_p(1).$$

The last lemma we will need to prove Theorem 3.1.1 gives convergence of the second order derivatives in probability:

Lemma 3.7.4 *Suppose conditions of Theorem 3.1.1 are satisfied. Then*

$$\int \left(-l_{1,1}(0,F_0(t),\delta,z) + 2l_{1,2}(0,F_0(t),\delta,z)\mathbf{h}^*(t) - l_{2,2}(0,F_0(t),\delta,z)\mathbf{h}^*(t)^2\right) dP_n \\ = 1/I_{F_0} + o_p(1).$$

Proof: First we define

$$Y_i = \left(-l_{1,1}(0,F_0(T),\Delta,Z) + 2l_{1,2}(0,F_0(T),\Delta,Z)\mathbf{h}^*(T) - l_{2,2}(0,F_0(T),\Delta,Z)\right)\mathbf{h}^*(t)^2.$$

By straightforward calculations we get $EY_i = 1/I_{F_0} + o(1)$, as $n \to \infty$. On the other hand $Y_i \le C_1\left(1 + (\log T)^2\right)$, which assures that all moments of Y_i are uniformly bounded. Application of the Markov inequality finishes the proof. ∎

Now we are ready to prove Theorem 3.1.1. The statistic $2\log \mathbf{T}_n$ can be written as

$$2\log \mathbf{T}_n = 2n\int \left(l\left(\hat{\theta}_n,\hat{F}_n(t),\delta,z\right) - l\left(0,\hat{F}_n^0(t),\delta,z\right)\right) dP_n.$$

If we can show that both $\hat{F}_n(t) + \hat{\theta}_n h_n(t)$ and $\hat{F}_n^0(t) - \hat{\theta}_n h_n^0(t)$ are distribution functions (with probability tending to one) it will follow from the optimality of the maximum likelihood estimator that, for L_n and R_n defined by

$$L_n = 2n\int \left(l\left(\hat{\theta}_n,\hat{F}_n^0(t) - \hat{\theta}_n h_n^0(t),\delta,z\right) - l\left(0,\hat{F}_n^0(t),\delta,z\right)\right) dP_n,$$

$$R_n = 2n\int \left(l\left(\hat{\theta}_n,\hat{F}_n(t),\delta,z\right) - l\left(0,\hat{F}_n(t) + \hat{\theta}_n h_n(t),\delta,z\right)\right) dP_n,$$

we have:

$$L_n \le 2\log \mathbf{T}_n \le R_n \tag{3.7.2}$$

with probability tending to one. Therefore we need to study the function $u_\theta(t) = t + \theta q(t) t \log t$ on the interval $[0, 1]$. We have: $\lim_{t\to 0} u_\theta(t) = 0$, $\lim_{t\to 1} u_\theta(t) = 1$ and for $|\theta|$ small enough we get, by comparing the derivative u_θ' with zero, that for $\exp(-C_1/|\theta|) \le t \le 1$, the function $t \mapsto u_\theta(t)$ is increasing. The statement now follows from Lemma 3.3.5.

Expanding integrand of L_n around $\left(0, \hat{F}_n^0\right)$ we get:

$$\begin{aligned} L_n = \;& 2n\hat{\theta}_n \int l_1\left(0, \hat{F}_n^0(t), \delta, z\right) dP_n - 2n\hat{\theta}_n \int l_2\left(0, \hat{F}_n^0(t), \delta, z\right) h_n^0(t) dP_n \\ & + n\hat{\theta}_n^2 \int l_{1,1}\left(\tilde{\theta}_n^0, \tilde{F}_n^0(t), \delta, z\right) dP_n - 2n\hat{\theta}_n^2 \int l_{1,2}\left(\tilde{\theta}_n^0, \tilde{F}_n^0(t), \delta, z\right) h_n^0(t) dP_n \\ & + n\hat{\theta}_n^2 \int l_{2,2}\left(\tilde{\theta}_n^0, \tilde{F}_n^0(t), \delta, z\right) h_n^0(t)^2 dP_n, \end{aligned}$$

where $\left|\tilde{\theta}_n^0\right| < \left|\hat{\theta}_n\right|$ and $\tilde{F}_n^0(t)$ is a function on $[0, 1]$, such that, for any $t \in [0, 1]$, $\tilde{F}_n^0(t)$ lies between $\hat{F}_n^0(t)$ and $\hat{F}_n^0(t) - \hat{\theta}_n h_n^0(t)$. Using Lemma 3.7.2 and (3.7.1), we obtain that the first two integrals are equal to $2n\hat{\theta}_n^2/I_{F_0} + o_p(1)$, and using Lemmas 3.7.3 and 3.7.4, we get that the last three integrals are equal to $-n\hat{\theta}_n^2/I_{F_0} + o_p(1)$. Applying all this together with Theorem 3.7.1 we obtain

$$L_n \xrightarrow{\mathcal{D}} Y^2, \; n \to \infty, \tag{3.7.3}$$

where $Y \sim \mathcal{N}\left(\theta_0/\sqrt{I_{F_0}}, 1\right)$.

On the other hand, by expanding the integrand of R_n around $\left(\hat{\theta}_n, \hat{F}_n\right)$ we obtain

$$\begin{aligned} R_n = \;& 2n\hat{\theta}_n \int l_1\left(\hat{\theta}_n, \hat{F}_n(t), \delta, z\right) dP_n - 2n\hat{\theta}_n \int l_2\left(\hat{\theta}_n, \hat{F}_n(t), \delta, z\right) h_n(t) dP_n \\ & - n\hat{\theta}_n^2 \int l_{1,1}\left(\tilde{\theta}_n, \tilde{F}_n(t), \delta, z\right) dP_n + 2n\hat{\theta}_n^2 \int l_{1,2}\left(\tilde{\theta}_n, \tilde{F}_n(t), \delta, z\right) h_n(t) dP_n \\ & - n\hat{\theta}_n^2 \int l_{2,2}\left(\tilde{\theta}_n, \tilde{F}_n(t), \delta, z\right) h_n(t)^2 dP_n, \end{aligned}$$

where, as before, $\left|\tilde{\theta}_n\right| < \left|\hat{\theta}_n\right|$, and $\tilde{F}_n(t)$ is a function on $[0, 1]$ such that, for any $t \in [0, 1]$, $\tilde{F}_n(t)$ lies between $\hat{F}_n(t)$ and $\hat{F}_n(t) + \hat{\theta}_n h_n(t)$. Now we have that

$$\int l_1\left(\hat{\theta}_n, \hat{F}_n(t), \delta, z\right) dP_n = 0,$$

$$\int l_2\left(\hat{\theta}_n, \hat{F}_n(t), \delta, z\right) h_n(t) dP_n = 0,$$

due to the optimality of the pair $\left(\hat{\theta}_n, \hat{F}_n\right)$ and since the deviation lies in the accepted area, as we saw above. Now, by application of the Lemmas 3.7.3 and 3.7.4 together with Theorem 3.7.1, we obtain

$$R_n \xrightarrow{\mathcal{D}} Y^2, \; n \to \infty, \tag{3.7.4}$$

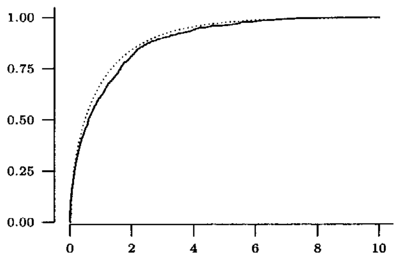
where $Y \sim \mathcal{N}\left(\theta_0/\sqrt{I_{F_0}}, 1\right)$. Combination of (3.7.2), (3.7.3) and (3.7.4) completes the proof of Theorem 3.1.1.

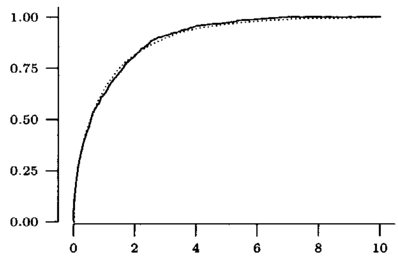
3.8 Results of simulations

This section shows results of the computer simulations of the likelihood ratio statistic for different sample sizes n. All twelve pictures below represent the empirical distribution functions of the likelihood ratio (solid lines) compared to the theoretical limiting distributions (doted lines); the sample size everywhere equals to 1000.

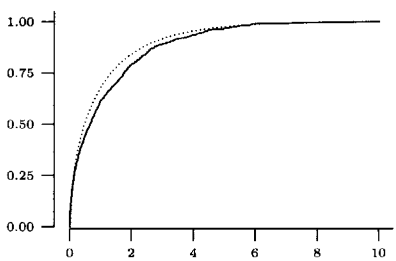
The first series of pictures corresponds to the underlying distributions $Z \sim$ Bernoulli(0.5), $F_0, G_0 \sim \exp(1)$, $G_1 \sim \exp(2)$.

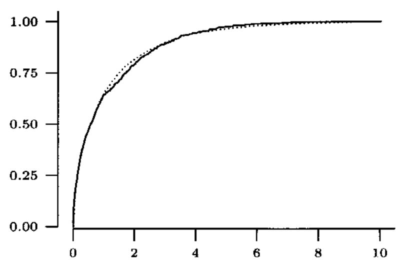
$\boldsymbol{\theta_0 = 0}$ $\boldsymbol{\theta_0 = 1}$



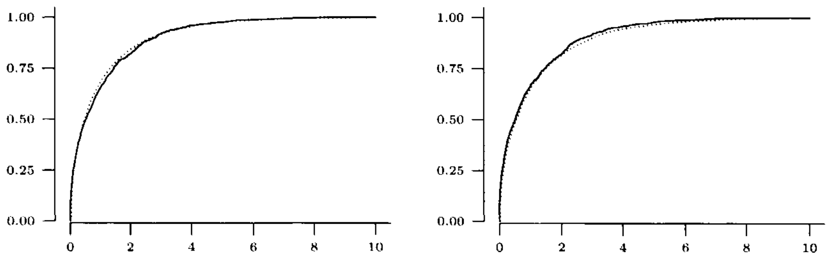


$n = 50$





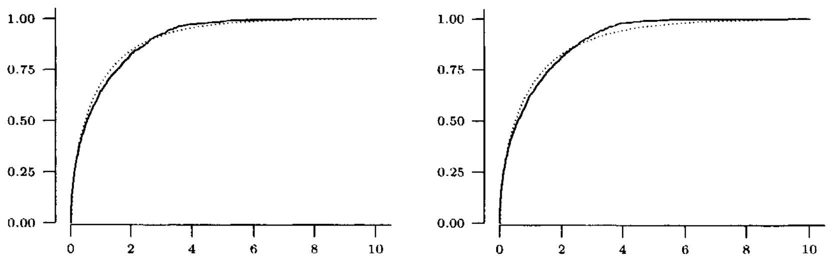
$n = 100$



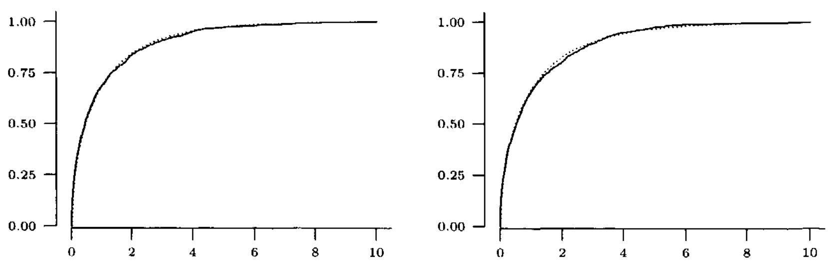
$n = 500$

Another series of pictures corresponds to the underlying distributions $Z \sim$ Bernoulli(0.5), $F_0, G_0, G_1 \sim$ Un[0, 1].

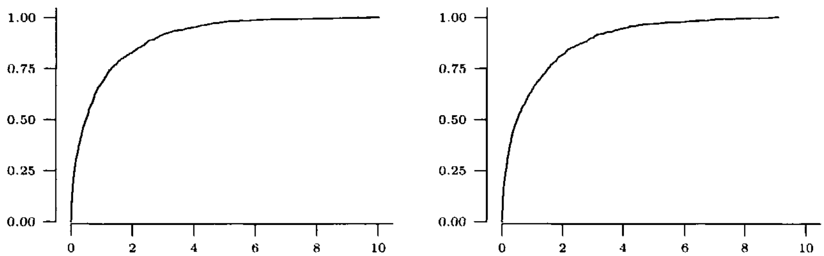
$\boldsymbol{\theta_0 = 0}$ $\boldsymbol{\theta_0 = 1}$



$n = 50$



$n = 100$



$n = 500$

Chapter 4

Test of concavity

We consider the process $\hat{F}_n - F_n$, being the difference between the empirical distribution function F_n and its least concave majorant $\hat{F}_n$. We prove that this difference process converges in distribution to the corresponding difference process for two-sided Brownian motion with parabolic drift. As a consequence we also derive asymptotic normality for L_k-functionals $\int |\hat{F}_n - F_n|^k g(t)\,dt$.

4.1 Introduction

Let $X_1, X_2, \ldots, X_n$ be a sample from a decreasing density f. Suppose that f has bounded support, which then without loss of generality may be taken to be the interval $[0,1]$. Let $\hat{F}_n$ be the least concave majorant of the empirical distribution function F_n on $[0,1]$, by which we mean the smallest concave function that lies above F_n. In this paper we study the difference between the processes $\hat{F}_n$ and F_n. Apart from being of interest in its own right, the distribution of this difference is needed if one wants to construct a statistical test for concavity based on the difference between $\hat{F}_n$ and F_n. For instance a test based on the L_k-distance $\int_0^1 |\hat{F}_n(t) - F_n(t)|^k\,dt$.

Define the process

$$A_n(t) = n^{2/3}\left\{\hat{F}_n(t) - F_n(t)\right\}, \quad t \in [0,1]. \tag{4.1.1}$$

In Kiefer and Wolfowitz (1976), it was shown that $(\log n)^{-1} \sup_t |A_n(t)|$ converges to zero with probability one, but the precise rate of convergence or limiting distribution was not given. Wang (1994) investigated the asymptotic behavior of $A_n(t)$, for t being a fixed point in $(0,1)$. The limiting distribution can be described in terms of the operator CM_I that maps a function $h : \mathbb{R} \to \mathbb{R}$ into the least concave majorant of h on the interval $I \subset \mathbb{R}$. If we define the process

$$Z(t) = W(t) - t^2, \tag{4.1.2}$$

where W denotes standard two-sided Brownian motion originating from zero, then it is shown in Wang (1994) that, for $t \in (0,1)$ fixed, $A_n(t)$ converges in distribution to $c_1(t)\zeta(0)$, where $c_1(t)$ is some constant depending on f and t, and

$$\zeta(t) = (\mathrm{CM}_{\mathbb{R}} Z)(t) - Z(t). \tag{4.1.3}$$

Recently, Durot and Tocquet (2002) obtained the same result in a regression setting.

In the present paper we will extend the pointwise result of WANG (1994) by proving that for any $t \in (0,1)$ fixed, a properly scaled version $\zeta_{nt}(s) = c_1(t)A_n(t + c_2(t)sn^{-1/3})$ converges in distribution to the process $\zeta(s)$ in the space $D(\mathbb{R})$ of cadlag functions on $\mathbb{R}$. Moreover we show that for any function g that is continuous on $[0,1]$, the L_k-functional $\int_0^1 A_n(t)^k g(t)\,dt$ is asymptotically normal. This result is similar to the one in DUROT AND TOCQUET (2002) who, independently of our efforts, proved asymptotic normality of L_k distances in the regression setting.

One of the main tools in proving process convergence is the continuous mapping theorem. Observe that A_n is the image of F_n under the mapping $h \mapsto \text{CM}_I h - h$, which is a continuous mapping from the space $D(I)$ into itself. This is one of basic properties of concave majorants that are described in Lemma 4.2.1. In Section 4.3, we use the Hungarian embedding and the representation $B(F(s)) = W(F(s)) - W(1)F(s)$ for the Brownian bridge, to approximate the empirical process by the process

$$s \mapsto W\left(n^{1/3}\big(F(t + n^{-1/3}s) - F(t)\big)\right). \tag{4.1.4}$$

As a consequence, a properly scaled version of the process F_n converges to Brownian motion with parabolic drift. After establishing some preliminary results for this process in Section 4.3, application of the continuous mapping theorem yields convergence of the process A_n. The limit process is obtained in Section 4.4.

In Section 4.5 the L_k-functionals are shown to be asymptotically normal. One of the main differences between the regression setting and our setup is the embedding of the empirical process. In the regression setting it can be embedded directly into Brownian motion itself, whereas in our setup it can only be embedded in the process (4.1.4). This introduces an additional difficulty of approximating the value of the concave majorant of the process (4.1.4) at zero by the corresponding value of the process $s \mapsto W\big(f(t)s\big)$. Although the maximum difference between the two processes is too large, the key observation that makes things work is that the values of the concave majorants at zero are sufficiently close, as is shown in Lemma 4.5.3. Next, the proof of asymptotic normality is along the lines of the proof of Theorem 1.1 in GROENEBOOM, HOOGHIEMSTRA AND LOPUHAÄ (1999). We first approximate the process A_n by a Brownian motion version A_n^W. Then we use that Brownian motion has independent increments to obtain strong mixing for the process A_n^W and prove asymptotic normality by using the method of big blocks and small blocks.

4.2 Basic properties of concave majorants

We start by proving a number of properties of the operator CM. By h'_r and h'_l, we will denote the right- and left-derivative of a function h. The following lemma lists a number of general basic properties of the operator CM.

Lemma 4.2.1 *Let g and h be functions on an interval $B \subset \mathbb{R}$. Then the following properties hold.*

1. $[\text{CM}_B g](t) \leq \sup_B g$, *for all* $t \in B$.

2. *For any linear function* $l(t) = at + b$ *on* B, *we have* $[\text{CM}_B(g + l)](t) = [\text{CM}_B g](t) + l(t)$ *for all* $t \in B$.

3. *For any interval A, such that $A \subset B$, we have $[\mathrm{CM}_A g](t) \le [\mathrm{CM}_B g](t)$, for all $t \in A$.*

4. *If $g \le h$ on B, then $\mathrm{CM}_B g \le \mathrm{CM}_B h$ on B.*

5. $(\mathrm{CM}_B g) + \inf_B h \le \mathrm{CM}_B(g+h) \le (\mathrm{CM}_B g) + \sup_B h$ *on B.*

6. *Let $a, b, t \in B$, such that $a < t < b$ and suppose that*

$$[\mathrm{CM}_B g]'_r(a) > [\mathrm{CM}_B g]'_l(t) \ge [\mathrm{CM}_B g]'_r(t) > [\mathrm{CM}_B g]'_l(b).$$

Then $[\mathrm{CM}_B g](t) = \left[\mathrm{CM}_{[a,b]} g\right](t)$.

7. *Suppose $[x-1, x+1] \subset B$. Then*

$$\left|[\mathrm{CM}_B g]'(x)\right| \le \max\left\{\sup_B g - g(x-1), \sup_B g - g(x+1)\right\}.$$

8. *Let $[a,b] \subset B \subset I\!R$ and suppose that $[\mathrm{CM}_{[a,b]}g](a) = [\mathrm{CM}_B g](a)$ and $[\mathrm{CM}_{[a,b]}g](b) = [\mathrm{CM}_B g](b)$. Then $[\mathrm{CM}_{[a,b]}g](t) = [\mathrm{CM}_B g](t)$, for all $t \in [a,b]$.*

Proof: ad 1. The statement is trivial in the case $\sup_B g = \infty$. In the case $\sup_B g < \infty$, $\mathrm{CM}_B g$ attains its maximum at some vertex τ. At each vertex τ we must have $g(\tau) = [\mathrm{CM}_B g](\tau)$, which proves 1.

ad 2. Let $t \in B$ and let $\tau_1 < \tau_2$ be two consecutive vertices of $\mathrm{CM}_B(g+l)$, such that $\tau_1 \le t \le \tau_2$. By definition of $\mathrm{CM}_B(g+l)$, on the interval B, the function $g+l$ is below the line trough the points $(\tau_i, g(\tau_i) + a\tau_i + b)$, with $i = 1, 2$. This means that for any $s \in B$,

$$\begin{aligned} g(s) &\le [\mathrm{CM}_B(g+l)](s) - l(s) \\ &\le g(\tau_2) + a\tau_2 + b + \left(\frac{g(\tau_2) - g(\tau_1)}{\tau_2 - \tau_1} + a\right)(s - \tau_2) - (as + b) \\ &= g(\tau_2) + \frac{g(\tau_2) - g(\tau_1)}{\tau_2 - \tau_1}(s - \tau_2). \end{aligned}$$

Hence, on the interval B, the function g is below the line through the points $(\tau_1, g(\tau_1))$ and $(\tau_2, g(\tau_2))$. This means that the line segment between $(\tau_1, g(\tau_1))$ and $(\tau_2, g(\tau_2))$ is part of a segment of $\mathrm{CM}_B g$. This implies that

$$\begin{aligned} [\mathrm{CM}_B(g+l)](t) &= g(\tau_2) + a\tau_2 + b + \left(\frac{g(\tau_2) - g(\tau_1)}{\tau_2 - \tau_1} + a\right)(t - \tau_2) \\ &= g(\tau_2) + \frac{g(\tau_2) - g(\tau_1)}{\tau_2 - \tau_1}(t - \tau_2) + at + b \\ &= [\mathrm{CM}_B g](t) + l(t). \end{aligned}$$

ad 3. Suppose that for some $t \in A$ we have $[\mathrm{CM}_A g](t) > [\mathrm{CM}_B g](t)$. Let $[\tau_1, \tau_2] \subset A$ be the segment of $\mathrm{CM}_A g$ that contains t. Then $\tau_1, \tau_2 \in B$ and by concavity of $\mathrm{CM}_B g$ it follows that either $g(\tau_1) = [\mathrm{CM}_A g](\tau_1) > [\mathrm{CM}_B g](\tau_1)$ or $g(\tau_2) = [\mathrm{CM}_A g](\tau_2) > [\mathrm{CM}_B g](\tau_2)$. This is in contradiction with $\mathrm{CM}_B g \ge g$ on B.

ad 4. Suppose $[\mathrm{CM}_B g](t) > [\mathrm{CM}_B h](t)$ for some $t \in B$ and let $[\tau_1, \tau_2] \subset B$ be the segment of $\mathrm{CM}_B g$ that contains t. Then by concavity of $\mathrm{CM}_B g$ we either have $[\mathrm{CM}_B g](\tau_1) > [\mathrm{CM}_B h](\tau_1)$ or $[\mathrm{CM}_B g](\tau_2) > [\mathrm{CM}_B h](\tau_2)$. This implies that

$$g(s) = [\mathrm{CM}_B g]\,(s) > [\mathrm{CM}_B h]\,(s) \geq h(s),$$

either for $s = \tau_1$ or $s = \tau_2$, which is in contradiction with $g \leq h$ on B.

ad 5. By property 2, we have $\mathrm{CM}_B(g + c) = (\mathrm{CM}_B g) + c$, for any constant $c \in \mathbb{R}$. Hence, by applying property 4, we get

$$(\mathrm{CM}_B g) + \inf h = \mathrm{CM}_B(g + \inf h) \leq \mathrm{CM}_B(g + h) \leq \mathrm{CM}_B(g + \sup h) = (\mathrm{CM}_B g) + \sup h.$$

ad 6. Let $[\tau_1, \tau_2] \subset B$ be the segment of $\mathrm{CM}_B g$ that contains t. Consider the line through the points $(\tau_1, g(\tau_1))$ and $(\tau_2, g(\tau_2))$:

$$s \mapsto g(\tau_2) + \frac{g(\tau_1) - g(\tau_2)}{\tau_1 - \tau_2}(s - \tau_2). \tag{4.2.1}$$

It remains to show that the line segment between through $(\tau_1, g(\tau_1))$ and $(\tau_2, g(\tau_2))$ is a part of $\mathrm{CM}_{[a,b]} g$. For this, it is sufficient to show that for any $s \in [a, \tau_1)$ and any $s \in (\tau_2, b]$, the value $g(s)$ is below the line (4.2.1). By definition of $\mathrm{CM}_B g$, we have for any $s \in B$ such that $s < \tau_1$ or $s > \tau_2$, that $g(s)$ is below the line (4.2.1). Since $[\mathrm{CM}_B g]_r'(a) > [\mathrm{CM}_B g]_r'(t) = [\mathrm{CM}_B g]_r'(\tau_1)$, by concavity of $\mathrm{CM}_B g$, we must have $a < \tau_1$. Hence for any $s \in [a, \tau_1)$, the value $g(s)$ is below the line (4.2.1). Similarly for $s \in (\tau_2, b]$, since $[\mathrm{CM}_B g]_l'(\tau_2) = [\mathrm{CM}_B g]_l'(t) > [\mathrm{CM}_B g]_l'(b)$ implies that $\tau_2 < b$.

ad 7. Suppose $[\mathrm{CM}_B g]'(x) > 0$. Since $[x - 1, x] \subset B$, $g(x - 1)$ is below the line through the points $(x, [\mathrm{CM}_B g](x))$ and $(x - 1, [\mathrm{CM}_B g](x - 1))$. This means that the line through the points $(x, [\mathrm{CM}_B g](x))$ and $(x - 1, g(x - 1))$ has a slope that is greater than $[\mathrm{CM}_B g]'(x)$. Hence by property 1 we find

$$0 < [\mathrm{CM}_B g]'(x) \leq \frac{[\mathrm{CM}_B g](x) - g(x - 1)}{x - (x - 1)} \leq \left(\sup_B g\right) - g(x - 1).$$

Similarly, if $[\mathrm{CM}_B g]'(x) \leq 0$,

$$0 \geq [\mathrm{CM}_B g]'(x) \geq \frac{[\mathrm{CM}_B g](x) - g(x + 1)}{x - (x + 1)} \geq -\left(\sup_B g\right) + g(x + 1).$$

ad 8. In view of property 3, suppose that there exists a $t \in (a, b)$, such that $\left[\mathrm{CM}_{[a,b]} g\right](t) < [\mathrm{CM}_B g]\,(t)$. Let $[\tau_1, \tau_2] \subset [a, b]$ be the segment of $\mathrm{CM}_{[a,b]} g$ that contains t. Since $\mathrm{CM}_{[a,b]} g = \mathrm{CM}_B g$ at the endpoints of the interval $[a, b]$, there must be a vertex τ_0 of $\mathrm{CM}_B g$ between τ_1 and τ_2, for which $\left[\mathrm{CM}_{[a,b]} g\right](\tau_0) < [\mathrm{CM}_B g]\,(\tau_0)$. But then $g(\tau_0) = [\mathrm{CM}_B g](\tau_0) > [\mathrm{CM}_{[a,b]}](\tau_0)$, which is in contradiction with $\mathrm{CM}_B g \geq g$ on B. ∎

4.3 Brownian approximation

In this section we show that by means of a local scaling the empirical process can be approximated by a Brownian motion, and we prove some preliminary results for Brownian motion with drift. Let E_n denote the empirical process $\sqrt{n}(F_n - F)$ and let B_n be a Brownian bridge constructed on the same probability space as the uniform empirical process $E_n \circ F^{-1}$ via the Hungarian embedding of KOMLOS, MAJOR AND TUSNADY (1975). Then

$$\sup_{t\in[0,1]} |E_n(t) - B_n(F(t))| = \mathcal{O}_p(n^{-1/2}\log n). \tag{4.3.1}$$

Define versions W_n of Brownian motion by

$$W_n(t) = B_n(t) + \xi_n t, \quad t \in [0,1]. \tag{4.3.2}$$

For $t \in (0,1)$ fixed, define the process

$$X_{nt}(s) = n^{2/3}\left(F_n(t+sn^{-1/3}) - F_n(t) - \left(F(t+sn^{-1/3}) - F(t)\right)\right). \tag{4.3.3}$$

Let F have a continuously differentiable density f on $[0,1]$ that satisfies

(A) $0 < \inf_{t\in[0,1]} |f'(t)| < \sup_{t\in[0,1]} |f'(t)| < \infty.$

Then on compact sets the process X_{nt} converges to a time-scaled Brownian motion.

Lemma 4.3.1 *Let F satisfy condition (A). Fix $t \in (0,1)$ and let X_{nt} be defined by (4.3.3). Then the process $\{X_{nt}(s) : s \in \mathbb{R}\}$ converges in distribution to the process $\{W(f(t)s) : s \in \mathbb{R}\}$ in $D(\mathbb{R})$, the space of cadlag functions on $\mathbb{R}$.*

Proof: All trajectories of the limiting process belong to $C(\mathbb{R})$, the separable subset of continuous functions on $\mathbb{R}$. This means that similar to Theorem V.23 in POLLARD (1984), it suffices to show that for any compact set $I \subset \mathbb{R}$ the process $\{X_{nt}(s) : s \in I\}$ converges in distribution to the process $\{W(f(t)s) : s \in I\}$ in $D(I)$, the space of cadlag functions on I. We will apply Theorem V.3 in POLLARD (1984), which is stated for $D[0,1]$, but the same result holds for $D(I)$.

By applying (4.3.1), we can write

$$\begin{aligned} X_{nt}(s) &= n^{1/6}\left\{E_n(t+sn^{-1/3}) - E_n(t)\right\} \\ &= n^{1/6}\left\{B_n(F(t+sn^{-1/3})) - B_n(F(t))\right\} + \mathcal{O}_p(n^{-1/3}\log n) \\ &= n^{1/6}\left\{W_n(F(t+sn^{-1/3})) - W_n(F(t))\right\} + \mathcal{O}_p(n^{-1/6}\log n), \end{aligned}$$

where the big $\mathcal{O}$-term is uniform for $s \in I$. By using Brownian scaling, a simple Taylor expansion, and the uniform continuity of Brownian motion on compacta, we find that

$$X_{nt}(s) \stackrel{d}{=} W(f(t)s) + R_n(s),$$

where $\sup_{s\in I} |R_n(s)| \to 0$ in probability. From this representation it follows immediately that the process $\{X_{nt}(s) : s \in I\}$ satisfies the conditions of Theorem V.3 in POLLARD (1984). This proves the lemma. ■

Let D_I be the operator that maps a function $h : \mathbb{R} \to \mathbb{R}$ into the difference between the least concave majorant of h on the interval I and h itself

$$\mathrm{D}_I h = \mathrm{CM}_I h - h.$$

Then the process A_n can be written as a functional of F_n only: $A_n = n^{2/3}[D_{[0,1]}F_n]$. This means that in order to obtain the limiting behavior of A_n we must investigate the limiting behavior of F_n itself. Note that

$$n^{2/3}\left(F(t+sn^{-1/3}) - F(t)\right) \approx n^{1/3}f(t)s + \frac{1}{2}f'(t)s^2.$$

By property 2 of Lemma 4.2.1 the operator $D_{[0,1]}$ will be invariant under addition of linear functions. Hence the term $n^{1/3}f(t)s$ will have no effect on the limiting behavior of A_n. In view of Lemma 4.3.1 this means that limiting behavior of A_n will be determined by the concave majorant of Brownian motion with a parabolic drift.

The following two lemmas are concerned with the tail behavior of Brownian motion with polynomial drift. The first lemma ensures that the probability that the process $W(t) - K|t|^\alpha$ is still positive for $|t| > a$, decreases exponentially as $a \to \infty$. The second lemma states that the distribution function of $\sup_{t\in\mathbb{R}}(W(t) - K|t|^\alpha)$ has exponential tails.

Lemma 4.3.2 *For all $K > 0$ and $\alpha > \frac{1}{2}$, there exists a $C > 0$, such that for every $a \geq 0$,*

$$P\left\{\sup_{|t|>a}\left(W(t) - K|t|^\alpha\right) \geq 0\right\} \leq C\exp\left(-\frac{K^2}{4^{\alpha+1}}a^{2\alpha-1}\right).$$

Proof: The statement is trivial for $0 \leq a < 2$, since in that case we can take $C = \exp(K^2/8)$. For $a \geq 2$, we have

$$\begin{aligned}
&P\left\{\sup_{|t|>a}\left(W(t) - K|t|^\alpha\right) \geq 0\right\} \\
&\leq 2\sum_{i=[a]}^{\infty} P\left\{\sup_{t\in[i,i+1]} W(t) \geq Ki^\alpha\right\} \\
&\leq 2\sum_{i=[a]}^{\infty} P\left\{\sup_{t\in[i,i+1]}\left(W(t) - W(i)\right) \geq \frac{1}{2}Ki^\alpha\right\} + 2\sum_{i=[a]}^{\infty} P\left\{W(i) \geq \frac{1}{2}Ki^\alpha\right\}. \qquad (4.3.4)
\end{aligned}$$

Note that for any $i \geq 1$ we have that

$$P\{W(i) \geq x\} = P\{W(1) \geq x/\sqrt{i}\} \leq e^{-\frac{x^2}{2i}}\int_0^\infty \phi(u)\,du = \frac{1}{2}e^{-\frac{x^2}{2i}}, \qquad (4.3.5)$$

and that for all $x \geq 0$ and $\beta \geq 0$,

$$(1+x)^\beta \geq 1 + x^\beta. \qquad (4.3.6)$$

Then, by using (4.3.5) and (4.3.6), we can bound the first term in (4.3.4) as follows.

$$\begin{aligned}
&2\sum_{i=[a]}^{\infty} P\left\{\sup_{t\in[i,i+1]}\left(W(t) - W(i)\right) \geq \frac{1}{2}Ki^\alpha\right\} = 2\sum_{i=[a]}^{\infty} P\left\{\sup_{t\in[0,1]} W(t) \geq \frac{1}{2}Ki^\alpha\right\} \\
&= 4\sum_{i=[a]}^{\infty} P\left\{W(1) \geq \frac{1}{2}Ki^\alpha\right\} \leq 2\sum_{i=0}^{\infty} e^{-\frac{1}{8}K^2[a]^{2\alpha}(1+i/[a])^{2\alpha}} \\
&\leq 2e^{-\frac{1}{8}K^2[a]^{2\alpha}}\sum_{i=0}^{\infty} e^{-\frac{1}{8}K^2 i^{2\alpha}} \leq C_1 e^{-\frac{1}{8}K^2[a]^{2\alpha}} \leq C_1\exp\left(-\frac{K^2}{2^{2\alpha+3}}a^{2\alpha}\right),
\end{aligned}$$

where $C_1 = 2\sum_{i=0}^{\infty} e^{-\frac{1}{8}K^2 i^{2\alpha}} < \infty$, and where we used that $[a] \geq a - 1 \geq a/2$, for $a \geq 2$. Similarly, the second term in (4.3.4) can be bounded by

$$2\sum_{i=[a]}^{\infty} P\left\{W(i) \geq \frac{1}{2}Ki^{\alpha}\right\} \leq C_2 \exp\left(-\frac{K^2}{2^{2\alpha+2}}a^{2\alpha-1}\right),$$

where $C_2 = 2\sum_{i=0}^{\infty} e^{-\frac{1}{8}K^2 i^{2\alpha-1}} < \infty$. It follows that

$$\begin{aligned} &P\left\{\sup_{|t|>a}\left(W(t) - K|t|^{\alpha}\right) \geq 0\right\} \\ &\leq \exp\left(-\frac{K^2}{2^{2\alpha+2}}a^{2\alpha-1}\right)\left\{C_1 \exp\left(-\frac{K^2}{2^{2\alpha+3}}a^{2\alpha-1}(a-2)\right) + C_2\right\} \\ &\leq C \exp\left(-\frac{K^2}{4^{\alpha+1}}a^{2\alpha-1}\right), \end{aligned}$$

where $C > 0$ only depends on K and α, since $a^{2\alpha-1}(a-2) \geq 0$, for $a \geq 2$. ■

Lemma 4.3.3 *For all $K > 0$ and $\alpha > \frac{1}{2}$, there exists a $C > 0$, such that for every $x \geq 0$,*

$$P\left\{\sup_{\mathbb{R}}\left(W(t) - K|t|^{\alpha}\right) > x\right\} \leq C(1 + x^{1/\alpha})\exp\left(-\frac{1}{8}K^{1/\alpha}x^{2-\frac{1}{\alpha}}\right).$$

Proof: We have that

$$\begin{aligned} &P\left\{\sup_{\mathbb{R}}\left(W(t) - K|t|^{\alpha}\right) > x\right\} \\ &\leq 2\sum_{i=0}^{\infty} P\left\{\sup_{t\in[i,i+1]} W(t) > x + Ki^{\alpha}\right\} \\ &\leq 2\sum_{i=0}^{\infty} P\left\{\sup_{t\in[0,1]} W(t) > (x + Ki^{\alpha})/2\right\} + 2\sum_{i=1}^{\infty} P\left\{W(i) > (x + Ki^{\alpha})/2\right\}. \end{aligned} \quad (4.3.7)$$

By application of (4.3.5) and (4.3.6), similar to the proof of Lemma 4.3.2, we can bound the first term in (4.3.7) by

$$2\sum_{i=0}^{\infty} e^{-\frac{1}{8}(Ki^{\alpha}+x)^2} \leq C_1 e^{-\frac{1}{8}x^2},$$

where $C_1 = \sum_{i=0}^{\infty} e^{-\frac{1}{8}K^2 i^{2\alpha}} < \infty$. Similarly, the second term in (4.3.7) can be bounded by

$$\sum_{i=1}^{\infty} e^{-(Ki^{\alpha}+x)^2/(8i)} = \sum_{1\leq i\leq(x/K)^{1/\alpha}} e^{-(Ki^{\alpha}+x)^2/(8i)} + \sum_{i>(x/K)^{1/\alpha}} e^{-(Ki^{\alpha}+x)^2/(8i)}.$$

By using $(Ki^{\alpha} + x)^2 \geq x^2$, the first term on the right hand side can be bounded by

$$(x/K)^{1/\alpha}e^{-\frac{1}{8}K^{1/\alpha}x^{2-1/\alpha}},$$

and by using $(Ki^\alpha + x)^2 \geq K^2 i^{2\alpha}$, the second term on the right hand side can be bounded by

$$\sum_{i>(x/K)^{1/\alpha}} e^{-\frac{1}{8}K^2 i^{2\alpha-1}} = \sum_{i>0} e^{-\frac{1}{8}K^2(i+(x/K)^{1/\alpha})^{2\alpha-1}} \leq C_2 e^{-\frac{1}{8}K^{1/\alpha}x^{2-1/\alpha}},$$

where $C_2 = \sum_{i>0} e^{-\frac{1}{8}K^2 i^{2\alpha-1}} < \infty$, again using (4.3.5) and (4.3.6). Similar to the proof of Lemma 4.3.2, it follows that

$$P\left\{\sup_{\mathbb{R}}\left(W(t) - K|t|^\alpha\right) > x\right\} \leq C(1 + x^{1/\alpha})e^{-\frac{1}{8}K^{1/\alpha}x^{2-1/\alpha}},$$

for $C > 0$ only depending on K and α. ∎

With suitable standardization, the limiting Brownian motion with parabolic drift can be transformed to the process Z defined in (4.1.2). We will be dealing with the concave majorants of this process on large bounded intervals and on the whole real line. Property 8 of Lemma 4.2.1 guarantees that both concave majorants are the same on an interval as soon as their values coincide on the boundary of the interval. The next lemma states that for large intervals this happens with large probability. For $d > 0$, consider the event

$$N(d) = \left\{ [\mathrm{CM}_{\mathbb{R}} Z](s) = \left[\mathrm{CM}_{[-d,d]} Z\right](s), \text{ for } s = \pm d/2 \right\} \tag{4.3.8}$$

Lemma 4.3.4 *There exist constants $C_1 > 0$ and $C_2 > 0$, such that for all $d \geq 0$*

$$P(N(d)^c) \leq C_1(1 + d^{1/2}) \exp\left(-C_2 d^{3/2}\right).$$

Proof: Let $\bar{Z}$ be the process

$$\bar{Z}(s) = Z(s + d/2).$$

Then, by symmetry and property 3 of Lemma 4.2.1, we have

$$\begin{aligned} P\{N(d)^c\} &\leq 2P\left\{[\mathrm{CM}_{\mathbb{R}} Z](d/2) > \left[\mathrm{CM}_{[-d,d]} Z\right](d/2)\right\} \\ &= 2P\left\{\left[\mathrm{CM}_{\mathbb{R}} \bar{Z}\right](0) > \left[\mathrm{CM}_{[-d/2,3d/2]} \bar{Z}\right](0)\right\} \\ &\leq 2P\left\{\left[\mathrm{CM}_{\mathbb{R}} \bar{Z}\right](0) > \left[\mathrm{CM}_{[-d/2,d/2]} \bar{Z}\right](0)\right\}. \end{aligned}$$

Note that

$$\bar{Z}(s) = W(s + d/2) - (s + d/2)^2 \stackrel{d}{=} Z(s) + W(d/2) - d^2/4 - sd.$$

Hence by property 2 of Lemma 4.2.1, we find that

$$\begin{aligned} P\{N(d)^c\} &\leq 2P\left\{\left[\mathrm{D}_{\mathbb{R}} \bar{Z}\right](0) > \left[\mathrm{D}_{[-d/2,d/2]} \bar{Z}\right](0)\right\} \\ &= 2P\left\{[\mathrm{D}_{\mathbb{R}} Z](0) > \left[\mathrm{D}_{[-d/2,d/2]} Z\right](0)\right\} \\ &= 2P\left\{[\mathrm{CM}_{\mathbb{R}} Z](0) > \left[\mathrm{CM}_{[-d/2,d/2]} Z\right](0)\right\}. \end{aligned}$$

Distinguish between

1. $Z(s) \leq -\frac{1}{2}s^2$, for all $|s| > d/2$,
2. $Z(s) > -\frac{1}{2}s^2$, for some $|s| > d/2$.

Consider the first case. Let $[\tau_1, \tau_2] \subset [-d/2, d/2]$ be the segment of $\mathrm{CM}_{[-d/2,d/2]}$ that contains zero. If $[\mathrm{CM}_{\mathbb{R}} Z](0) > [\mathrm{CM}_{[-d/2,d/2]} Z](0)$, then the line through the points $(\tau_1, Z(\tau_1))$ and $(\tau_2, Z(\tau_2))$ must intersect the process Z outside the interval $[-d/2, d/2]$. In case 1, this can only happen if this line intersects the parabola $p(s) = -\frac{1}{2}s^2$ outside the interval $[-d/2, d/2]$. This is only possible if the slope of this line is greater than the tangent of $p(s)$ in the point $s = -d/2$ or smaller than the tangent of $p(s)$ in the point $s = d/2$:

$$\left| \left[\mathrm{CM}_{[-d/2,d/2]} Z\right]'(0) \right| \geq d/2.$$

We find that

$$P\{N(d)^c\} \leq 2P\left\{ \sup_{|s|>d/2} \left(Z(s) + \tfrac{1}{2}s^2\right) > 0 \right\} + 2P\left\{ \left| \left[\mathrm{CM}_{[-d/2,d/2]} Z\right]'(0) \right| \geq d/2 \right\}. \quad (4.3.9)$$

By Lemma 4.3.2, the first term on the right hand side of (4.3.9) is equal to

$$2P\left\{ \sup_{|s|>d/2} \left(W(s) - \tfrac{1}{2}s^2\right) > 0 \right\} \leq 2Ce^{-d^3/2048}.$$

By property 7 of Lemma 4.2.1 the second term on the right hand side of (4.3.9) is bounded by

$$\begin{aligned} &2P\left\{ \sup_{s\in\mathbb{R}} \left(W(s) - s^2\right) + 1 - \min\{W(1), W(-1)\} \geq d/2 \right\} \\ &\leq 2P\left\{ \sup_{s\in\mathbb{R}} \left(W(s) - s^2\right) \geq d/4 \right\} + 2P\left\{1 - \min\{W(1), W(-1)\} \geq d/4\right\}. \end{aligned}$$

According to Lemma 4.3.3, the first term is bounded by $C(1+d^{1/2}/2)e^{-d^{3/2}/64}$. Property (4.3.5) implies that the second term on the right hand side is bounded by

$$2P\{W(1) \leq 1 - d/4\} + 2P\{W(-1) \leq 1 - d/4\} \leq e^{-\frac{1}{2}(1-d/4)^2}.$$

This proves the lemma. ∎

Likewise we will have to deal with the difference between concave majorants of F_n on intervals $[t - dn^{-1/3}, t + dn^{-1/3}]$ and on $[0, 1]$, as well as with the difference between concave majorants of the Brownian approximation of F_n. To this end define $F_n^E = F_n$ and let F_n^W be its Brownian approximation defined by

$$F_n^W(t) = F(t) + n^{-1/2} W_n(F(t)), \quad t \in [0, 1], \quad (4.3.10)$$

where W_n is defined in (4.3.2). For $t \in [0, 1]$ and $d > 0$ let $I_{nt}(d) = [t - dn^{-1/3}, t + dn^{-1/3}]$, and for $J = E, W$ define the event

$$N_{nt}^J(d) = \left\{ [\mathrm{CM}_{[0,1]} F_n^J](t + sn^{-1/3}) = [\mathrm{CM}_{I_{nt}(d)} F_n^J](t + sn^{-1/3}), \text{ for } s = \pm d/2 \right\}. \quad (4.3.11)$$

The following lemma ensures that the value of the two concave majorants of F_n^J coincide at $t \pm \frac{1}{2}dn^{-1/3}$ with high probability.

Lemma 4.3.5 *Let $t \in (0,1)$ and $d > 0$ be such that $0 < t - n^{-1/3}d < t + n^{-1/3}d < 1$. Moreover, suppose that for $\epsilon_n = \frac{1}{4}\inf|f'|n^{-1/3}d$,*

$$f(1) < f(t + n^{-1/3}d) - \epsilon_n < f(t - n^{-1/3}d) + \epsilon_n < f(0).$$

Then, for any distribution function F satisfying condition (A), there exist a constant $C_2 > 0$, such that

$$P\left\{N^J_{nt}(d)^c\right\} \leq 8\exp\left(-C_2 d^3\right),$$

where C_2 does not depend on d, t and n.

Proof: For $J = E, W$ define $\hat{f}^J_n$ as the left-derivative of $\mathrm{CM}_{[0,1]}F^J_n$. Define

$$U^J_n(a) = \underset{t\in[0,1]}{\operatorname{argmax}}\left\{F^J_n(t) - at\right\} \quad \text{and} \quad V^J_n(a) = n^{1/3}\left(U^J_n(a) - g(a)\right),$$

where g denotes the inverse of f. The process U^J_n is related to $\hat{f}^J_n$ by the relation

$$\hat{f}^J_n(t) \leq a \Longleftrightarrow U^J_n(a) \leq t \text{ with probability one.} \tag{4.3.12}$$

Suppose that the concave majorants of F^J_n on the intervals $[0,1]$ and $[t - dn^{-1/3}, t + dn^{-1/3}]$ differ at $t - dn^{-1/3}/2$. A simple picture shows that in that case there cannot be point of jump of $\hat{f}^J_n$ between $t - dn^{-1/3}$ and $t - dn^{-1/3}/2$, which implies that $\hat{f}^J_n(t - dn^{-1/3}) \leq \hat{f}^J_n(t - dn^{-1/3}/2)$. Similarly, if the concave majorants of F^J_n on the intervals $[0,1]$ and $[t - dn^{-1/3}, t + dn^{-1/3}]$ differ at $t + dn^{-1/3}/2$, then $\hat{f}^J_n(t + dn^{-1/3}) \leq \hat{f}^J_n(t + dn^{-1/3}/2)$. Hence

$$\begin{aligned} P\left\{N^J_{nt}(d)^c\right\} \;\leq\; & P\left\{\hat{f}^J_n(t - n^{-1/3}d) \leq \hat{f}^J_n(t - n^{-1/3}d/2)\right\} \\ & + P\left\{\hat{f}^J_n(t + n^{-1/3}d) \geq \hat{f}^J_n(t + n^{-1/3}d/2)\right\}. \end{aligned} \tag{4.3.13}$$

Consider the first probability on the right hand side of (4.3.13). Then with $s = t - n^{-1/3}d$ and $x = d/2$, we have

$$\begin{aligned} & P\left\{\hat{f}^J_n(t - n^{-1/3}d) \leq \hat{f}^J_n(t - n^{-1/3}d/2)\right\} \\ & = P\left\{\hat{f}^J_n(s + n^{-1/3}x) \geq \hat{f}^J_n(s)\right\} \\ & = P\left\{\left(\hat{f}^J_n(s + n^{-1/3}x) - f(s + n^{-1/3}x)\right) - \left(\hat{f}^J_n(s) - f(s)\right) \geq f(s) - f(s + n^{-1/3}x)\right\} \\ & \leq P\left\{\left(\hat{f}^J_n(s + n^{-1/3}x) - f(s + n^{-1/3}x)\right) - \left(\hat{f}^J_n(s) - f(s)\right) \geq n^{-1/3}x\inf|f'|\right\} \\ & \leq P\left\{\hat{f}^J_n(s + n^{-1/3}x) - f(s + n^{-1/3}x) \geq \epsilon_n\right\} + P\left\{\hat{f}^J_n(s) - f(s) \leq -\epsilon_n\right\}. \end{aligned} \tag{4.3.14}$$

By using (4.3.12), the first probability on the right hand side of (4.3.14) is equal to

$$\begin{aligned} & P\left\{U^J_n\left(f(s + xn^{-1/3}) + \epsilon_n\right) \geq s + n^{-1/3}x\right\} \\ & = P\left\{V^J_n\left(f(s + xn^{-1/3}) + \epsilon_n\right) \geq n^{1/3}\left(s + n^{-1/3}x - g(f(s + xn^{-1/3}) + \epsilon_n)\right)\right\} \\ & \geq P\left\{V^J_n\left(f(s + xn^{-1/3}) + \epsilon_n\right) \geq n^{1/3}\epsilon_n\inf|g'|\right\} \\ & = P\left\{V^J_n\left(f(s + xn^{-1/3}) + \epsilon_n\right) \geq \frac{\inf|f'|d}{4\sup|f'|}\right\}. \end{aligned}$$

Since $f(s+xn^{-1/3})+\epsilon_n = f(t-dn^{-1/3}/2)+\epsilon_n \in [f(1), f(0)]$, it follows from Theorems 2.1 and 3.1 in GROENEBOOM, HOOGHIEMSTRA AND LOPUHAÄ (1999) that

$$P\left\{V_n^J\left(f(s+xn^{-1/3})+\epsilon_n\right) \geq \frac{\inf|f'|d}{4\sup|f'|}\right\} \leq 2e^{-C_2d^3}, \tag{4.3.15}$$

for some constant $C_2 > 0$, not depending on n, t and d. The second probability on the right hand side of (4.3.14) can be bounded similarly,

$$P\left\{\hat{f}_n^J(s) - f(s) \leq -\epsilon_n\right\} \leq 2e^{-C_2d^3}.$$

Together with (4.3.15) we conclude that the probability of the first event on the right hand side of (4.3.13) can be bounded as follows

$$P\left\{\hat{f}_n^J(t-n^{-1/3}d) \geq \hat{f}_n^J(t-n^{-1/3}d/2)\right\} \leq 4e^{-C_2d^3}.$$

The probability of the second event on the right hand side of (4.3.13) can be bounded similarly, by taking $s = t + n^{-1/3}d/2$ and $x = d/2$ and using the same argument as above. This proves the lemma. ■

4.4 Process convergence

For $t \in (0,1)$ fixed and $t + c_2(t)sn^{-1/3} \in (0,1)$, define

$$\zeta_{nt}(s) = c_1(t)A_n(t + c_2(t)sn^{-1/3}), \tag{4.4.1}$$

where

$$c_1(t) = \left(\frac{|f'(t)|}{2f^2(t)}\right)^{1/3} \quad \text{and} \quad c_2(t) = \left(\frac{4f(t)}{|f'(t)|^2}\right)^{1/3}. \tag{4.4.2}$$

Define $\zeta_{nt}(s) = 0$ for $t + c_2(t)sn^{-1/3} \notin (0,1)$. The following theorem states that the process ζ_{nt} converges to the process ζ on $D(\mathbb{R})$.

Theorem 4.4.1 *Let f be decreasing with support on $[0,1]$. Suppose that f is continuous differentiable on $[0,1]$ and satisfies*

(A) $0 < \inf_{t\in[0,1]} |f'(t)| < \sup_{t\in[0,1]} |f'(t)| < \infty.$

Let the processes ζ and ζ_{nt} be defined as in (4.1.3) and (4.4.1). Then the process $\{\zeta_{nt}(s) : s \in \mathbb{R}\}$ converges in distribution to the process $\{\zeta(s) : s \in \mathbb{R}\}$ in $D(\mathbb{R})$, the space of cadlag functions of $\mathbb{R}$.

Proof: Similar to the proof of Theorem 4.3.1 it is enough to show that for any compact set $K \subset \mathbb{R}$, the process $\{\zeta_{nt}(s) : s \in K\}$ converges in distribution to the process $\{\zeta(s) : s \in K\}$ on $D(K)$. Note that for this, it suffices to show that the process $\{A_n(t + sn^{-1/3}) : s \in K\}$ converges in distribution to the process $\{[D_{\mathbb{R}}Z_t](s) : s \in K\}$, where

$$Z_t(s) = W(f(t)s) + \frac{1}{2}f'(t)s^2. \tag{4.4.3}$$

This follows from the fact that by Brownian scaling $c_1(t)Z_t(c_2(t)s) \stackrel{d}{=} Z(s) = W(s) - s^2$.

Let $t \in (0,1)$ fixed, and let $I_{nt} = [-tn^{1/3}, (1-t)n^{1/3}]$. Write $E_{nt}(s) = n^{2/3}F_n(t + sn^{-1/3})$, for $s \in I_{nt}$. Then by definition

$$A_n(t + sn^{-1/3}) = [\mathrm{D}_{I_{nt}} E_{nt}](s) \quad \text{for } s \in I_{nt}.$$

Now take K fixed. For the processes $\{[\mathrm{D}_{I_{nt}} E_{nt}](s) : s \in K\}$ and $\{[\mathrm{D}_{\mathbb{R}} Z_t](s) : s \in K\}$, we must show that for any $g : D(K) \to \mathbb{R}$ bounded and continuous:

$$|Eg(\mathrm{D}_{I_{nt}} E_{nt}) - Eg(\mathrm{D}_{\mathbb{R}} Z_t)| \to 0.$$

Let $\epsilon > 0$ and let $I = [-d, d]$ be an interval, where according to Lemmas 4.3.4 and 4.3.5, $d > 0$ is chosen sufficiently large such that $K \subset [-d/2, d/2]$, and such that

$$P(N(d)^c) < \epsilon \quad \text{and} \quad P(N_{nt}^E(d)^c) < \epsilon, \tag{4.4.4}$$

where $N(d)$ and $N_{nt}^E(d)$ are defined in (4.3.8) and (4.3.11). Let n be sufficiently large, such that $K \subset [-d/2, d/2] \subset I \subset I_{nt}$. For $g : D(K) \to \mathbb{R}$ bounded and continuous, and processes $\{[\mathrm{D}_{I_{nt}} E_{nt}](s) : s \in K\}$, $\{[\mathrm{D}_{\mathbb{R}} Z_t](s) : s \in K\}$, and $\{[\mathrm{D}_I Z_t](s) : s \in K\}$, we have

$$\begin{aligned} |Eg(\mathrm{D}_{I_{nt}} E_{nt}) - Eg(\mathrm{D}_{\mathbb{R}} Z_t)| \leq\ & |Eg(\mathrm{D}_{I_{nt}} E_{nt}) - Eg(\mathrm{D}_I E_{nt})| \\ & + |Eg(\mathrm{D}_I E_{nt}) - Eg(\mathrm{D}_I Z_t)| \\ & + |Eg(\mathrm{D}_I Z_t) - Eg(\mathrm{D}_{\mathbb{R}} Z_t)|. \end{aligned} \tag{4.4.5}$$

For the first term on the right hand side of (4.4.5) we have that

$$\begin{aligned} & |Eg(\mathrm{D}_{I_{nt}} E_{nt}) - Eg(\mathrm{D}_I E_{nt})| \\ & \leq 2 \sup |g| \cdot P\{\mathrm{D}_{I_{nt}} E_{nt} \neq \mathrm{D}_I E_{nt} \text{ on } [-d/2, d/2]\} \\ & \leq 2 \sup |g| \cdot P\{[\mathrm{D}_{I_{nt}} E_{nt}](s) \neq [\mathrm{D}_I E_{nt}](s) \text{ for } s = -d/2 \text{ or } s = d/2\} \\ & = 2 \sup |g| \cdot P\{[\mathrm{CM}_{I_{nt}} E_{nt}](s) \neq [\mathrm{CM}_I E_{nt}](s) \text{ for } s = -d/2 \text{ or } s = d/2\}. \end{aligned}$$

Suppose that $[\mathrm{CM}_{I_{nt}} E_{nt}](s) \neq [\mathrm{CM}_I E_{nt}](s)$. This means that the concave majorants of F_n itself on the intervals $[0,1]$ and $I_{nt}(d) = [t - dn^{-1/3}, t + dn^{-1/3}]$ differ at $t + sn^{-1/3}$. Hence the probability on the right hand side above can be bounded as follows:

$$P\{[\mathrm{CM}_{I_{nt}} E_{nt}](s) \neq [\mathrm{CM}_I E_{nt}](s) \text{ for } s = -d/2 \text{ or } s = d/2\} \leq P(N_{nt}^E(d)^c).$$

According to (4.4.4), this yields

$$|Eg(\mathrm{D}_{I_{nt}} E_{nt}) - Eg(\mathrm{D}_I E_{nt})| \leq 2 \sup |g| \cdot \epsilon. \tag{4.4.6}$$

Similarly, application of property 8 of Lemma 4.2.1 gives

$$\begin{aligned} |Eg(\mathrm{D}_I Z_t) - Eg(\mathrm{D}_{\mathbb{R}} Z_t)| & \leq 2 \sup |g| \cdot P\{\mathrm{D}_I Z_t \neq \mathrm{D}_{\mathbb{R}} Z_t \text{ on } K\} \\ & \leq 2 \sup |g| \cdot P\{\mathrm{D}_I Z_t \neq \mathrm{D}_{\mathbb{R}} Z_t \text{ on } [-d/2, d/2]\} \\ & \leq 2 \sup |g| \cdot P(N(d)^c). \end{aligned}$$

Once more (4.4.4) yields

$$|Eg(\mathrm{D}_I Z_t) - Eg(\mathrm{D}_{\mathbb{R}} Z_t)| \leq 2 \sup |g| \cdot \epsilon. \tag{4.4.7}$$

In order to bound the second term on the right hand side of (4.4.5) define

$$Z_{nt}(s) = n^{2/3}\left(F_n(t+sn^{-1/3}) - F_n(t) - \left(F(t+sn^{-1/3}) - F(t)\right)\right) + \frac{1}{2}f'(t)s^2.$$

It follows from Lemma 4.3.1, that the process $\{Z_{nt}(s) : s \in I\}$ converges in distribution to the process $\{Z_t(s) : s \in I\}$. Because according to property 5 of Lemma 4.2.1, the mapping $\mathrm{D}_I : D(I) \to D(I)$ is continuous, this means that

$$|Eh(\mathrm{D}_I Z_{nt}) - Eh(\mathrm{D}_I Z_t)| \to 0,$$

for any $h : D(I) \to \mathbb{R}$ bounded and continuous. Note that we can also write

$$E_{nt}(s) = Z_{nt}(s) + n^{2/3}F_n(t) + f(t)sn^{1/3} + R_{nt}(s),$$

where

$$R_{nt}(s) = n^{2/3}\left[F(t+sn^{-1/3}) - F(t) - f(t)sn^{-1/3} - \frac{1}{2}f'(t)s^2n^{-2/3}\right].$$

Note that for some $|\theta - t| \le n^{-1/3}|s|$, with $s \in I$, we have

$$R_{nt}(s) = \frac{1}{2}|f'(\theta) - f'(t)|s^2 \to 0,$$

uniformly for $s \in I$, using that f' is continuous. By continuity of the mapping D_I together with property 2 of Lemma 4.2.1, it then follows that on I:

$$\mathrm{D}_I Z_{nt} = \mathrm{D}_I(E_{nt} - R_{nt}) = \mathrm{D}_I E_{nt} + o(1),$$

where the $o(1)$-term is uniform for $s \in I$. We conclude that for any $h : D(I) \to \mathbb{R}$ bounded and continuous, and processes $\{[\mathrm{D}_{I_{nt}} E_{nt}](s) : s \in I\}$ and $\{[\mathrm{D}_I Z_t](s) : s \in I\}$,

$$|Eh(\mathrm{D}_I E_{nt}) - Eh(\mathrm{D}_I Z_t)| \to 0. \tag{4.4.8}$$

Now let $\pi_K : D(I) \to D(K)$ be defined as the restriction of an element of $D(I)$ to the set K. Since for any $g : D(K) \to \mathbb{R}$ bounded and continuous the composition $h = g \circ \pi_K$ is also bounded and continuous, (4.4.8) implies that for $g : D(K) \to \mathbb{R}$ bounded and continuous, and processes $\{[\mathrm{D}_{I_{nt}} E_{nt}](s) : s \in K\}$ and $\{[\mathrm{D}_I Z_t](s) : s \in K\}$,

$$|Eg(\mathrm{D}_I E_{nt}) - Eg(\mathrm{D}_I Z_t)| \to 0. \tag{4.4.9}$$

Putting together (4.4.6), (4.4.7), (4.4.9) and (4.4.5) proves the theorem. ∎

4.5 Convergence of L_k-functionals

To obtain asymptotic normality of $\int_0^1 A_n(t)^k g(t)\,dt$, we will approximate the process $A_n = n^{2/3}[D_{[0,1]}F_n]$ by a Brownian version. We will need slightly stronger conditions of f. In addition to the conditions imposed in Theorem 4.4.1 we will assume that f is twice continuously differentiable satisfying

(B) $\sup_{t\in[0,1]} |f''(t)| < \infty$.

Let F_n^W be defined as in (4.3.10), and let

$$A_n^W(t) = n^{2/3}[D_{[0,1]}F_n^W](t). \tag{4.5.1}$$

The next lemma shows that for $J = E, W$, a properly scaled version of F_n^J can be approximated by the process

$$Y_{nt}(s) = W_n\left(n^{1/3}\left(F(t+n^{-1/3}s) - F(t)\right)\right) + \frac{1}{2}f'(t)s^2, \quad \text{for } -\infty < s < \infty. \tag{4.5.2}$$

plus linear term, where W_n is defined in (4.3.2).

Lemma 4.5.1 *Suppose that f satisfies conditions (A) and (B). Let $F_n^E = F_n$ and let F_n^W be defined as in (4.3.10). Then for $t \in (0,1)$ fixed, $J = E, W$ and $s \in (-tn^{1/3}, (1-t)n^{1/3})$:*

$$n^{2/3}F_n^J(t+n^{-1/3}s) = Y_{nt}(s) + L_{nt}^J(s) + R_{nt}^J(s),$$

where Y_{nt} is defined in (4.5.2), $L_{nt}^J(s)$ is linear in s, and where for all $k \geq 1$,

$$E \sup_{|s|\leq \log n} \left|R_{nt}^J(s)\right|^k = \mathcal{O}(n^{-k/3}(\log n)^{3k}),$$

uniformly in $t \in (0,1)$.

Proof: Taylor expansion together with (4.3.10) and (4.3.2) yields that

$$n^{2/3}F_n^W(t+n^{-1/3}s) = Y_{nt}(s) + L_{nt}^W(s) + R_{nt}^W(s),$$

with Y_{nt} as defined in (4.5.2), $L_{nt}^W(s)$ is linear in s:

$$L_{nt}^W(s) = n^{2/3}F(t) + n^{1/6}W_n(F(t)) + n^{1/3}f(t)s,$$

and $R_{nt}^W(s) = \frac{1}{6}n^{-1/3}f''(\theta_1)s^3$, for some $|\theta_1 - t| \leq n^{-1/3}|s|$. Similarly

$$\begin{aligned} n^{2/3}F_n^E(t+n^{-1/3}s) &= n^{2/3}F_n^W(t+n^{-1/3}s) \\ &\quad + n^{1/6}\left\{E_n(t+n^{-1/3}s) - B_n(F(t+n^{-1/3}s))\right\} \\ &\quad - n^{1/6}\xi_n\left\{F(t) + f(t)n^{-1/3}s + \frac{1}{2}f'(\theta_2)n^{-2/3}s^2\right\} \\ &= Y_{nt}(s) + L_{nt}^E(s) + R_{nt}^E(s), \end{aligned}$$

where $L_{nt}^E(s) = L_{nt}^W(s) - n^{1/6}\xi_nF(t) - n^{-1/6}\xi_nf(t)s$ is linear in s, and

$$R_{nt}^E(s) = R_{nt}^W(s) - n^{1/6}\left\{E_n(t+n^{-1/3}s) - B_n(F(t+n^{-1/3}s))\right\} - \frac{1}{2}n^{-1/2}\xi_nf'(\theta_2)s^2,$$

for some $|\theta_2 - t| \leq n^{-1/3}|s|$. It follows immediately from condition (B) that:

$$\sup_{|s|\leq \log n} \left|R_{nt}^W(s)\right|^k \leq C_1n^{-k/3}(\log n)^{3k}. \tag{4.5.3}$$

Note that

$$\sup_{|s|\le\log n}\left|R^E_{nt}(s)\right| \le \sup_{|s|\le\log n}\left|R^W_{nt}(s)\right| + n^{1/6}S_n + \frac{1}{2}\sup|f'|n^{-1/2}(\log n)^2|\xi_n|,$$

where $S_n = \sup_{s\in\mathbb{R}}|E_n(s) - B_n(F(s))|$. From KOMLOS, MAJOR AND TUSNADY (1975) we have that

$$P\left\{S_n \ge n^{-1/2}(C\log n + x)\right\} \le Ke^{-\lambda x},$$

for positive constants C, K, and λ. This implies that for all $k \ge 1$,

$$ES_n^k = \mathcal{O}(n^{-k/2}(\log n)^k). \tag{4.5.4}$$

Next use that for all $a, b > 0$ and $k \ge 1$

$$(a+b)^k \le 2^k(a^k + b^k). \tag{4.5.5}$$

Then from condition (A) together with (4.5.4) and (4.5.3) we find that

$$\begin{aligned} E\sup_{|s|\le\log n}\left|R^E_{nt}(s)\right|^k &= \mathcal{O}\left(n^{-k/3}(\log n)^{3k}\right) + \mathcal{O}\left(n^{-k/3}(\log n)^k\right) + \mathcal{O}\left(n^{-k/2}(\log n)^{2k}\right) \\ &= \mathcal{O}\left(n^{-k/3}(\log n)^{3k}\right). \end{aligned}$$

This proves the lemma. ∎

The next step is to approximate the moments of $A^J_n(t)$ by corresponding moments of the concave majorant of the process

$$Y_t(s) = W\left(n^{1/3}\left(F(t+n^{-1/3}s) - F(t)\right)\right) + \frac{1}{2}f'(t)s^2, \quad \text{for } -\infty < s < \infty. \tag{4.5.6}$$

Lemma 4.5.2 *Suppose that f satisfies conditions (A) and (B). For $t \in (0,1)$ fixed let Y_t be defined as in (4.5.6). Let $A^E_n(t) = A_n(t)$ and $A^W_n(t)$ be defined in (4.1.1) and (4.5.1). Then for all $k \ge 1$, and for $J = E, W$*

$$EA^J_n(t)^k = E\left[\mathrm{D}_{[-\log n,\log n]}Y_t\right](0)^k + o(n^{-1/6}),$$

uniformly for $t \in (0,1)$.

Proof: Let $I_{nt} = [t - n^{-1/3}\log n, t + n^{-1/3}\log n]$ and note that for $J = E, W$ on the event $N^J_{nt}(\log n)$, as defined in (4.3.11), we have

$$\begin{aligned} \left[\mathrm{CM}_{\mathbb{R}}F^J_n\right]'_r(t - n^{-1/3}\log n) &> \left[\mathrm{CM}_{\mathbb{R}}F^J_n\right]'_l(t) \\ &\ge \left[\mathrm{CM}_{\mathbb{R}}F^J_n\right]'_r(t) \\ &> \left[\mathrm{CM}_{\mathbb{R}}F^J_n\right]'_l(t + n^{-1/3}\log n). \end{aligned}$$

Hence by property 8 of Lemma 4.2.1 we have

$$A^J_n(t)1_{N^J_{nt}(\log n)} = n^{2/3}[\mathrm{D}_{I_{nt}}F^J_n](t)1_{N^J_{nt}(\log n)} \quad \text{for } J = E, W. \tag{4.5.7}$$

By definition $|A_n^E(t)| \le 2n^{2/3}$ and

$$\sup_{s\in(0,1)} |A_n^W(s)| \le 2n^{2/3}\left(1+n^{-1/2}\sup_{s\in[0,1]}|W_n(s)|\right),$$

so that

$$\begin{aligned} &E\left|A_n^J(t)^k - n^{2k/3}[\mathrm{D}_{I_{nt}}F_n^J](t)^k\right| 1_{N_{nt}^J(\log n)^c} \\ &\le 4n^{2k/3}E\left(1+n^{-1/2}\sup_{s\in[0,1]}|W_n(s)|\right)^k 1_{N_{nt}^J(\log n)^c} \\ &\le 4n^{2k/3}\left\{E\left(1+n^{-1/2}\sup_{s\in[0,1]}|W_n(s)|\right)^{2k}\right\}^{1/2}\left\{P\left(N_{nt}^J(\log n)^c\right)\right\}^{1/2}. \end{aligned}$$

Next use (4.5.5) together with the fact that all moments of $\sup_{s\in[0,1]}|W_n(s)|$ are finite. Then it follows from Lemma 4.3.5 that

$$\begin{aligned} EA_n^J(t)^k &= n^{2k/3}E[\mathrm{D}_{I_{nt}}F_n^J](t)^k + E\left(A_n^J(t)^k - n^{2k/3}[\mathrm{D}_{I_{nt}}F_n^J](t)^k\right)1_{N_{nt}^J(\log n)^c} \\ &= n^{2k/3}E[\mathrm{D}_{I_{nt}}F_n^J](t)^k + n^{2k/3}\mathcal{O}(e^{-\frac{1}{2}C_2(\log n)^3}), \end{aligned}$$

uniformly for $t\in(0,1)$. According to Lemma 4.5.1, for $s\in[-\log n,\log n]$:

$$n^{2/3}F_n^J(t+n^{-1/3}s) = Y_{nt}(s) + L_{nt}^J(s) + R_{nt}^J(s),$$

where Y_{nt} has the same distribution as the process Y_t defined in (4.5.6) and $L_{nt}^J(s)$ is linear in s. Hence by property 2 of Lemma 4.2.1

$$\begin{aligned} n^{2/3}\left[\mathrm{D}_{I_{nt}}F_n^J\right](t) &= \left[\mathrm{D}_{[-\log n,\log n]}(Y_{nt}+L_{nt}^J+R_{nt}^J)\right](0) \\ &= \left[\mathrm{D}_{[-\log n,\log n]}(Y_{nt}+R_{nt}^J)\right](0) \\ &= \left[\mathrm{D}_{[-\log n,\log n]}Y_{nt}\right](0) + \Delta_{nt}, \end{aligned}$$

where

$$\Delta_{nt} = \left[\mathrm{D}_{[-\log n,\log n]}(Y_{nt}+R_{nt}^J)\right](0) - \left[\mathrm{D}_{[-\log n,\log n]}Y_{nt}\right](0).$$

We find that

$$EA_n^J(t)^k = E\left[\mathrm{D}_{[-\log n,\log n]}Y_{nt}\right](0)^k + \epsilon_{nt} + n^{2k/3}\mathcal{O}\left(e^{-\frac{1}{2}C_2(\log n)^3}\right), \tag{4.5.8}$$

where, by application of the mean value theorem,

$$\begin{aligned} |\epsilon_{nt}| &\le E\left|\left(\left[\mathrm{D}_{[-\log n,\log n]}Y_{nt}\right](0)+\Delta_{nt}\right)^k - \left[\mathrm{D}_{[-\log n,\log n]}Y_{nt}\right](0)^k\right| \\ &= kE|\theta_{nt}|^{k-1}|\Delta_{nt}| \\ &\le k\left\{E|\theta_{nt}|^{2k-2}\right\}^{1/2}\left\{E|\Delta_{nt}|^2\right\}^{1/2}, \end{aligned} \tag{4.5.9}$$

with $|\theta_{nt} - \left[\mathrm{D}_{[-\log n,\log n]}Y_{nt}\right](0)| \le |\Delta_{nt}|$. Since $Y_{nt} \stackrel{d}{=} Y_t$, by application of (4.5.5)

$$E|\theta_{nt}|^{2k-2} \le 4^{2k-2}\left(E\sup_{|s|\le\log n}|Y_t(s)|^{2k-2} + E|\Delta_{nt}|^{2k-2}\right), \tag{4.5.10}$$

where according to property 5 of Lemma 4.2.1 together with Lemma 4.5.1, for all $k \geq 1$

$$E|\Delta_{nt}|^k \leq 2^k E \sup_{|s|\leq \log n} |R^J_{nt}(s)|^k = \mathcal{O}\left(n^{-k/3}(\log n)^{3k}\right), \qquad (4.5.11)$$

uniformly in $t \in (0,1)$. On the other hand, for $|s| \leq \log n$, there exist constants $C_3, C_4 > 0$ that only depend on f, such that

$$\sup_{|s|\leq \log n} |Y_t(s)| \leq \sup_{|s|\leq C_3 \log n} |W(s)| + C_4(\log n)^2 \stackrel{d}{=} (C_3 \log n)^{1/2} \sup_{|s|\leq 1} |W(s)| + C_4(\log n)^2.$$

Because all moments of $\sup_{|s|\leq 1} |W(s)|$ are finite, from (4.5.10) and (4.5.11) we conclude that

$$E|\theta_{nt}|^{2k-2} = \mathcal{O}((\log n)^{4k-4}) + \mathcal{O}(n^{-(2k-2)/3}(\log n)^{6k-6}).$$

Hence from (4.5.9) and (4.5.11), we find that $\epsilon_{nt} = \mathcal{O}(n^{-1/3}(\log n)^{2k+1})$. Together with (4.5.8) this implies that

$$\begin{aligned} EA^J_n(t)^k &= E\left[\mathrm{D}_{[-\log n, \log n]} Y_{nt}\right](0)^k + \mathcal{O}(n^{-1/3}(\log n)^{2k+1}) + n^{2k/3}\mathcal{O}\left(e^{-\frac{1}{2}C_2(\log n)^3}\right) \\ &= E\left[\mathrm{D}_{[-\log n, \log n]} Y_t\right](0)^k + o(n^{-1/6}), \end{aligned}$$

uniformly for $t \in (0,1)$. ■

By uniform continuity of Brownian motion on compacta, the process Y_t is close to the process $Z_t(s) = W(f(t)s) + \frac{1}{2}f'(t)s^2$. Hence the next step is to approximate the value of the concave majorant of Y_t at zero by the corresponding value of Z_t. In view of Theorem 4.5.1 this difference must be of smaller order than $n^{-1/6}$. Unfortunately, it does not suffice to bound the difference of the concave majorants by

$$\sup_{|s|\leq \log n} \left| W\left(n^{1/3}\left(F(t+n^{-1/3}s) - F(t)\right)\right) - W(f(t)s) \right|,$$

which is of order $\mathcal{O}(n^{-1/6}\log n)$ according to the properties of the modulus of continuity for Brownian motion. However, as a consequence of Lemma 4.5.3 the two concave majorants at zero are sufficiently close, as is shown in the next lemma.

Lemma 4.5.3 *Let g be a function on an interval $B \subset \mathbb{R}$. Let $0 \in B^\circ$ and let $\phi : \mathbb{R} \to \mathbb{R}$ be invertible with $\phi(0) = 0$. Let $\sup_B g < \infty$ and suppose there exists an $\alpha \in [0, 1/2]$ such that*

$$1 - \alpha \leq \frac{\phi(t)}{t} \leq 1 + \alpha, \qquad (4.5.12)$$

for all $t \in B$. Then

$$\left| \left[\mathrm{CM}_{\phi^{-1}(B)}(g \circ \phi)\right](0) - \left[\mathrm{CM}_B g\right](0) \right| \leq 4\alpha \left\{ \sup_B g - \left[\mathrm{CM}_B g\right](0) \right\}.$$

Proof: Consider the function $h(t) = g(t) - \sup_B g$, so that $h(t) \leq 0$. Let $a \leq 0 < b$, then with property (4.5.12), t and $\phi(t)$ have the same sign. Hence, $\phi^{-1}(a) \leq 0 < \phi^{-1}(b)$. This yields the following inequality

$$\frac{1+\alpha}{1-\alpha}\cdot\frac{h(a)b-h(b)a}{b-a} \leq \frac{h(a)\phi^{-1}(b)-h(b)\phi^{-1}(a)}{\phi^{-1}(b)-\phi^{-1}(a)} \leq \frac{1-\alpha}{1+\alpha}\cdot\frac{h(a)b-h(b)a}{b-a}. \tag{4.5.13}$$

Let $[\tau_1,\tau_2] \subset B$ be the segment of $\mathrm{CM}_B h$ that contains zero, and denote $t_i = \phi^{-1}(\tau_i)$, for $i = 1,2$. Similarly, let $[\xi_1,\xi_2] \subset \phi^{-1}(B)$ be the segment of $\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)$ that contains zero, and denote $x_i = \phi(\xi_i)$, for $i = 1,2$. Consider the line between $(x_1, h(x_1))$ and $(x_2, h(x_2))$. Since $[x_1,x_2] \subset B$, the intercept at zero of this line must be below $[\mathrm{CM}_B h](0)$:

$$\frac{h(x_1)x_2-h(x_2)x_1}{x_2-x_1} \leq [\mathrm{CM}_B h](0) = \frac{h(\tau_1)\tau_2-h(\tau_2)\tau_1}{\tau_2-\tau_1}. \tag{4.5.14}$$

Similarly, consider the line between $\big(t_1,(h\circ\phi)(t_1)\big)$ and $\big(t_2,(h\circ\phi)(t_2)\big)$. Since $[t_1,t_2] \subset \phi^{-1}(B)$, the intercept at zero of this line must be below $[\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)](0)$:

$$\frac{(h\circ\phi)(t_1)t_2-(h\circ\phi)(t_2)t_1}{t_2-t_1} \leq [\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)](0) = \frac{(h\circ\phi)(\xi_1)\xi_2-(h\circ\phi)(\xi_2)\xi_1}{\xi_2-\xi_1},$$

or equivalently,

$$\frac{h(\tau_1)\phi^{-1}(\tau_2)-h(\tau_2)\phi^{-1}(\tau_1)}{\phi^{-1}(\tau_2)-\phi^{-1}(\tau_2)} \leq [\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)](0) = \frac{h(x_1)\phi^{-1}(x_2)-h(x_2)\phi^{-1}(x_1)}{\phi^{-1}(x_2)-\phi^{-1}(x_1)}.$$

Together with (4.5.14) and (4.5.13), this implies that

$$\begin{aligned}\frac{1+\alpha}{1-\alpha}[\mathrm{CM}_B h](0) &\leq \frac{h(\tau_1)\phi^{-1}(\tau_2)-h(\tau_2)\phi^{-1}(\tau_1)}{\phi^{-1}(\tau_2)-\phi^{-1}(\tau_1)}\\ &\leq [\mathrm{CM}_{\phi^{-1}(B)}(h\circ\phi)](0)\\ &\leq \frac{1-\alpha}{1+\alpha}\cdot\frac{h(x_1)x_2-h(x_2)x_1}{x_2-x_1} \leq \frac{1-\alpha}{1+\alpha}[\mathrm{CM}_B h](0).\end{aligned}$$

Now use that $(1-\alpha)/(1+\alpha) \geq 1-4\alpha$ and $(1+\alpha)/(1-\alpha) \leq 1+4\alpha$, for $\alpha \in [0,1/2]$ and apply property 2 of Lemma 4.2.1 to the function $g(t) = h(t) + \sup_B g$. ∎

Lemma 4.5.4 *Suppose that f satisfies conditions (A) and (B). Let $t \in (0,1)$ and let ζ be defined as in (4.1.3). Let $A_n^E(t) = A_n(t)$ and $A_n^W(t)$ be defined in (4.1.1) and (4.5.1). Then for all $k \geq 1$, and for $J = E, W$,*

$$EA_n^J(t)^k = \left(\frac{2f(t)^2}{|f'(t)|}\right)^{k/3} E\zeta(0)^k + o(n^{-1/6}),$$

uniformly in $t \in (0,1)$.

Proof: For $t \in (0,1)$ fixed let Y_t and Z_t be defined as in (4.5.6) and (4.4.3). Define the interval

$$J_{nt} = \left[\frac{n^{1/3}\left(F(t-n^{-1/3}\log n)-F(t)\right)}{f(t)}, \frac{n^{1/3}\left(F(t+n^{-1/3}\log n)-F(t)\right)}{f(t)}\right],$$

and the mapping

$$\phi_{nt}(s) = \frac{n^{1/3}\big(F(t+n^{-1/3}s) - F(t)\big)}{f(t)}.$$

Then $[-\log n, \log n] = \phi_{nt}^{-1}(J_{nt})$, and there exists constant $C_1 > 0$ only depending on f, such that for all $s \in [-\log n, \log n]$,

$$1 - \alpha_n \leq \frac{\phi_{nt}(s)}{s} \leq 1 + \alpha_n,$$

where $\alpha_n = C_1 n^{-1/3} \log n$. By definition of Z_t and Y_t, we have

$$(Z_t \circ \phi_{nt})(s) = Y_t(s) + \frac{1}{2} f'(t) s^2 \left(\frac{\phi_{nt}(s)^2}{s^2} - 1 \right).$$

According to property 5 of Lemma 4.2.1, there exists constant $C_2 > 0$ only depending on f, such that

$$\big|[\mathrm{D}_{[-\log n, \log n]} Y_t](0) - [\mathrm{D}_{[-\log n, \log n]}(Z_t \circ \phi_{nt})](0)\big| \leq C_2 n^{-1/3} (\log n)^3. \tag{4.5.15}$$

Now apply Lemma 4.5.3 with $g = Z_t$, $\phi = \phi_{nt}$, $\alpha = \alpha_n$ and $B = J_{nt}$. This yields that

$$\big|[\mathrm{D}_{[-\log n, \log n]}(Z_t \circ \phi_{nt})](0) - [\mathrm{D}_{J_{nt}} Z_t](0)\big| \leq 8\alpha_n \sup_{s \in \mathbb{R}} |Z_t(s)|.$$

Together with (4.5.15) we conclude that there exists a constant $C > 0$ only depending on f, such that

$$\big|[\mathrm{D}_{[-\log n, \log n]} Y_t](0) - [\mathrm{D}_{J_{nt}} Z_t](0)\big| \leq C n^{-1/3} \log n \left((\log n)^2 + \sup_{s \in \mathbb{R}} |Z_t(s)| \right). \tag{4.5.16}$$

Similar to the proof of Lemma 4.5.2, this implies that

$$E[\mathrm{D}_{[-\log n, \log n]} Y_t](0)^k = E[\mathrm{D}_{J_{nt}} Z_t](0)^k + \epsilon_{nt}, \tag{4.5.17}$$

where $|\epsilon_{nt}| \leq k \left\{ E|\theta_{nt}|^{2k-2} \right\}^{1/2} \left\{ E|\Delta_{nt}|^2 \right\}^{1/2}$, with

$$\Delta_{nt} = [\mathrm{D}_{[-\log n, \log n]} Y_t](0) - [\mathrm{D}_{J_{nt}} Z_t](0),$$

and $|\theta_{nt} - [\mathrm{D}_{[-\log n, \log n]} Y_t](0)| \leq |\Delta_{nt}|$. Note that with $c_1(t)$ and $c_2(t)$ as defined in (4.4.2), by Brownian scaling

$$c_1(t) Z_t(c_2(t)s) \stackrel{d}{=} Z(s). \tag{4.5.18}$$

Hence, by means of Lemma 4.3.3 it follows that for all $k \geq 1$,

$$E\left(\sup_{s \in \mathbb{R}} |Z_t(s)| \right)^k \leq C_1 E\left(\sup_{s \in \mathbb{R}} |Z(s)| \right)^k < \infty,$$

for a constant $C_1 > 0$ only depending on f. From (4.5.16) we conclude that for all $k \geq 1$

$$E|\Delta_{nt}|^k = \mathcal{O}(n^{-k/3} (\log n)^{3k}). \tag{4.5.19}$$

Similar to the proof of Lemma 4.5.2, using an inequality similar to (4.5.10) together with (4.5.19), we find that

$$E|\theta_{nt}|^{2k-2} = \mathcal{O}((\log n)^{4k-4}) + \mathcal{O}(n^{-(2k-2)/3}(\log n)^{6k-6}).$$

Together with (4.5.19) this implies that $\epsilon_{nt} = \mathcal{O}(n^{-1/3}(\log n)^{2k+1})$, so that from (4.5.17)

$$E[\mathrm{D}_{[-\log n, \log n]} Y_t](0)^k = E[\mathrm{D}_{J_{nt}} Z_t](0)^k + \mathcal{O}(n^{-1/3}(\log n)^{2k+1}).$$

Together with Lemma 4.5.2 and scaling property (4.5.18), we find that

$$EA_n^J(t)^k = c_1(t)^{-k} E\left[\mathrm{D}_{\mathbb{R}} Z\right](0)^k + c_1(t)^{-k} E\left(\left[\mathrm{D}_{I_{nt}} Z\right](0)^k - \left[\mathrm{D}_{\mathbb{R}} Z\right](0)^k\right) + o(n^{-1/6}), \quad (4.5.20)$$

where $I_{nt} = c_2(t)^{-1} J_{nt}$. Let $M > 0$ be a constant, only depending on f, such that $I_{nt} \subset [-M\log n, M\log n]$. According to property 8 of Lemma 4.2.1, on the event $N(2M\log n)$, as defined in (4.3.8), we have $\left[\mathrm{D}_{I_{nt}} Z\right](0)^k = \left[\mathrm{D}_{\mathbb{R}} Z\right](0)^k$. Hence

$$\begin{aligned}
\left|E\left(\left[\mathrm{D}_{I_{nt}} Z\right](0)^k - \left[\mathrm{D}_{\mathbb{R}} Z\right](0)^k\right)\right| &\leq E\left|\left[\mathrm{D}_{I_{nt}} Z\right](0)^k - \left[\mathrm{D}_{\mathbb{R}} Z\right](0)^k\right| 1_{N(2M\log n)^c} \\
&\leq 2^{k+1} E\left(\sup_{s\in\mathbb{R}} |Z(s)|\right)^k 1_{N(2M\log n)^c} \\
&\leq 2^{k+1}\left\{E\left(\sup_{s\in\mathbb{R}}|Z(s)|\right)^k\right\}^{1/2}\left\{P(N(2M\log n)^c)\right\}^{1/2}.
\end{aligned}$$

Since Lemma 4.3.3 yields that $E(\sup|Z|)^k < \infty$, it follows from Lemma 4.3.4 that

$$E\left(\left[\mathrm{D}_{I_{nt}} Z\right](0)^k - \left[\mathrm{D}_{\mathbb{R}} Z\right](0)^k\right) = \mathcal{O}((\log n)^{1/4} e^{-\frac{1}{2}C_2(\log n)^{3/2}} = o(n^{-1/6}).$$

Together with (4.5.20) and the fact that $\zeta = \mathrm{D}_{\mathbb{R}} Z$ this proves the lemma. ∎

A direct consequence of Lemma 4.5.4 is that for all $k \geq 1$, the difference between the processes $A_n(t)^k$ and $A_n^W(t)^k$ is of smaller order than $n^{-1/6}$.

Lemma 4.5.5 *Suppose that f satisfies conditions (A) and (B). Let $A_n^E = A_n$ and A_n^W be defined by (4.1.1) and (4.5.1). Then for all $k \geq 1$*

$$\int_0^1 \left|A_n^E(t)^k - A_n^W(t)^k\right| dt = o_p(n^{-1/6}).$$

Proof: By Markov's inequality it suffices to prove that $E\left|A_n^E(t)^k - A_n^W(t)^k\right| = o(n^{-1/6})$ uniformly in $t \in (0,1)$. Let $I_{nt} = [t - n^{-1/3}\log n, t + n^{-1/3}\log n]$ and let $K_{nt} = N_{nt}^E(\log n) \cap N_{nt}^W(\log n)$, where for $J = E, W$, the event $N_{nt}^J(\log n)$ is defined in (4.3.11). Then according to (4.5.7):

$$\begin{aligned}
E\left|A_n^E(t)^k - A_n^W(t)^k\right| &= n^{2k/3} E\left|[D_{I_{nt}} F_n^E](t)^k - [D_{I_{nt}} F_n^W](t)^k\right| 1_{K_{nt}} \qquad (4.5.21)\\
&\quad + E\left|A_n^E(t)^k - A_n^W(t)^k\right| 1_{K_{nt}^c}.
\end{aligned}$$

We first bound the second term on the right hand side of (4.5.21). We have that

$$\begin{aligned}
E\left|A_n^E(t)^k - A_n^W(t)^k\right| 1_{K_{nt}^c} &\leq EA_n^E(t)^k 1_{K_{nt}^c} + EA_n^W(t)^k 1_{K_{nt}^c} \\
&\leq \left\{EA_n^E(t)^{2k}\right\}^{1/2}\left\{P(K_{nt}^c)\right\}^{1/2} + \left\{EA_n^W(t)^{2k}\right\}^{1/2}\left\{P(K_{nt}^c)\right\}^{1/2},
\end{aligned}$$

where $P(K_{nt}^c) \le 16e^{-C_2(\log n)^3}$ uniformly in $t \in (0,1)$, according to Lemma 4.3.5. Since from Lemma 4.5.4 we know that $EA_n^J(t)^{2k}$ are bounded uniformly in n and $t \in (0,1)$, we conclude that

$$E\left|A_n^E(t)^k - A_n^W(t)^k\right| 1_{K_{nt}^c} = \mathcal{O}(e^{-\frac{1}{2}C_2 \log n^3}), \tag{4.5.22}$$

uniformly in $t \in (0,1)$.

To bound the first term in (4.5.21), apply the mean value theorem to write

$$\begin{aligned} & n^{2k/3}\left|[D_{I_{nt}}F_n^E](t)^k - [D_{I_{nt}}F_n^W](t)^k\right| 1_{K_{nt}} \\ & \le k|\theta_{nt}|^{k-1} n^{2/3}\left|[D_{I_{nt}}F_n^E](t) - [D_{I_{nt}}F_n^W](t)\right| 1_{K_{nt}} \\ & \le k\left(A_n^E(t)^{k-1} + A_n^W(t)^{k-1}\right) n^{2/3}\left|[D_{I_{nt}}F_n^E](t) - [D_{I_{nt}}F_n^W](t)\right|. \end{aligned} \tag{4.5.23}$$

Note that

$$[D_{I_{nt}}F_n^J](t) = [D_{[-\log n, \log n]}\tilde{F}_{nt}^J](0) \quad \text{for } J = E, W,$$

where $\tilde{F}_{nt}^J(s) = F_n^J(t + n^{-1/3}s)$, so that

$$n^{2/3}\left|[D_{I_{nt}}F_n^E](t) - [D_{I_{nt}}F_n^W](t)\right| = n^{2/3}\left|[D_{[-\log n,\log n]}\tilde{F}_{nt}^E](0) - [D_{[-\log n,\log n]}\tilde{F}_{nt}^W](0)\right|.$$

Taylor expansion together with (4.3.10) and (4.3.2) yields that

$$\begin{aligned} F_n^E(t + n^{-1/3}s) &= F_n^W(t + n^{-1/3}s) + n^{-1/2}\left\{E_n(t + n^{-1/3}s) - B_n(F(t + n^{-1/3}s))\right\} \\ & \quad - n^{-1/2}\xi_n\left\{F(t) + f(t)n^{-1/3}s + \frac{1}{2}f'(\theta)n^{-2/3}s^2\right\}. \end{aligned}$$

Since $F(t)+f(t)n^{-1/3}s$ is linear in s, first applying property 2 of Lemma 4.2.1 and then property 5 of Lemma 4.2.1 yields that the right hand side of (4.5.23) can be bounded by

$$\begin{aligned} & kE\left(A_n^E(t)^{k-1} + A_n^W(t)^{k-1}\right)\left(n^{1/6}S_n + \frac{1}{2}\sup|f'|n^{-1/2}(\log n)^2|\xi_n|\right) \\ & \le k\left\{E\left(A_n^E(t)^{k-1} + A_n^W(t)^{k-1}\right)^2\right\}^{1/2}\left\{E\left(n^{1/6}S_n + \frac{1}{2}\sup|f'|n^{-1/2}(\log n)^2|\xi_n|\right)^2\right\}^{1/2}, \end{aligned}$$

where $S_n = \sup_{s\in\mathbb{R}}|E_n(s) - B_n(F(s))|$. From Lemma 4.5.4 together with (4.5.5), it follows that the first expectation is bounded uniformly for $t \in (0,1)$. Similarly, condition (A) together with (4.5.4) and (4.5.5), yields that the second expectation is of the order $\mathcal{O}(n^{-1/3}\log n)$. Together with (4.5.21) and (4.5.22) this proves the lemma. ∎

We will need some independence structure for the process $\{A_n^W(t) : t \in [0,1]\}$. The fact that Brownian motion has independent increments will ensure that the process A_n^W is mixing, as is stated in the next lemma.

Lemma 4.5.6 *Suppose that f satisfies conditions (A) and (B). The process $\{A_n^W(t) : t \in [0,1]\}$ is strong mixing process with mixing function*

$$\alpha_n(d) = C_1 e^{-C_2 n d^3},$$

where $C_1 > 0$ and $C_2 > 0$ only depend on f. More specifically, for $d > 0$,

$$\sup|P(A \cap B) - P(A)P(B)| \le \alpha_n(d),$$

where the supremum is taken over all sets $A \in \sigma\{A_n^W(s) : 0 < s \le t\}$ and $B \in \sigma\{A_n^W(u) : t + d \le u < 1\}$.

Proof: Let $t \in (0,1)$ arbitrary and take $0 < s_1 \leq s_2 \leq \cdots \leq s_k = t < t + d = u_1 \leq u_2 \leq \cdots \leq u_l < 1$. Consider events

$$\begin{aligned} E_1 &= \left\{A_n^W(s_1) \in B_1, \ldots, A_n^W(s_k) \in B_k\right\}, \\ E_2 &= \left\{A_n^W(u_1) \in C_1, \ldots, A_n^W(u_l) \in C_l\right\}, \end{aligned}$$

for Borel sets $B_1, \ldots, B_k$ and $C_1, \ldots, C_l$ of $\mathbb{R}$. Note that cylinder sets of the form E_1 and E_2 generate the σ-algebras $\sigma\{A_n^W(s) : 0 < s \leq t\}$ and $\sigma\{A_n^W(u) : t + d \leq u < 1\}$, respectively. Let U_n^W be defined as in the proof of Lemma 4.3.5. Take $M_n = \frac{1}{4} d n^{1/3}$ and define the events

$$\begin{aligned} E_1' &= E_1 \cap \left\{U_{n,M_n}^W(f(t)) = U_n^W(f(t))\right\}, \\ E_2' &= E_2 \cap \left\{U_{n,M_n}^W(f(t+d)) = U_n^W(f(t+d))\right\}, \end{aligned}$$

where similar to the definition of U_n^W,

$$U_{n,M_n}(a) = \underset{n^{1/3}|s-g(a)| \leq M_n}{\operatorname{argmax}} \left\{F_n^W(s) - as\right\}.$$

By monotonicity of U_n^W, it follows that the event E_1' depends only on the increments of Brownian motion before time $F(t+n^{-1/3}M_n)$ and the event E_2' depends only on the increments of Brownian motion beyond time $F(t + d - n^{-1/3}M_n)$. By definition of M_n, we have $F(t + n^{-1/3}M_n) < F(t + d - n^{-1/3}M_n)$, so that E_1' and E_2' are independent. Therefore by means of Theorem 3.1 from Groeneboom, Hooghiemstra and Lopuhaä (1999),

$$\begin{aligned} &|P(E_1 \cap E_2) - P(E_1)P(E_2)| \\ &\quad \leq 3P\left(U_{n,M_n}^W(f(t)) \neq U_n^W(f(t))\right) + 3P\left(U_{n,M_n}^W(f(t+d)) \neq U_n^W(f(t+d))\right) \\ &\quad \leq 3P\left(n^{1/3}|U_n^W(f(t)) - t| > M_n\right) + 3P\left(n^{1/3}|U_n^W(f(t+d)) - (t+d)| > M_n\right) \\ &\quad \leq 12e^{-CM_n^3}, \end{aligned}$$

for some constant $C > 0$ that only depends on f. This proves the lemma. ∎

From Lemmas 4.5.5 and 4.5.4 it follows immediately that for proving asymptotic normality of $n^{1/6}\int_0^1 \left(A_n(t)^k - EA_n(t)^k\right) g(t)\, dt$, it suffices to prove that its Brownian version

$$T_n^W = n^{1/6} \int_0^1 \left(A_n^W(t)^k - EA_n^W(t)^k\right) g(t)\, dt, \tag{4.5.24}$$

is asymptotically normal. The proof runs along the lines of the proof of Theorem 4.1 in Groeneboom, Hooghiemstra and Lopuhaä (1999) and needs two lemmas that bound covariances by the mixing coefficient. The lemmas are analogous to Theorems 17.2.1 and 17.2.2 in Ibragimov and Linnik (1971) and can be proven similarly, since stationarity is not essential in these theorems.

Lemma 4.5.7 *If X is measurable with respect to $\{\sigma\{A_n^W(s) : 0 < s \leq t\}$ and Y is measurable with respect to $\{\sigma\{A_n^W(u) : t + d \leq u < 1\}$ $(d > 0)$, and if $|X| \leq C_1$ and $|Y| \leq C_2$ a.s., then*

$$|E(XY) - E(X)E(Y)| \leq 4C_1C_2\alpha_n(d).$$

Lemma 4.5.8 *If X is measurable with respect to $\{\sigma\{A_n^W(s) : 0 < s \le t\}$ and Y is measurable with respect to $\{\sigma\{A_n^W(u) : t + d \le u < 1\}$ $(d > 0)$, and if for some $\delta > 0$,*

$$E|X|^{2+\delta} \le C_3 \quad \text{and} \quad E|Y|^{2+\delta} \le C_4,$$

then

$$|E(XY) - E(X)E(Y)| \le C_5 \left(\alpha_n(d)\right)^{\delta/(2+\delta)},$$

where $C_5 > 0$ only depends on C_3 and C_4.

We first derive the asymptotic variance of T_n^W. To this end we introduce the Brownian version of the process ζ_{nt} defined in (4.4.1): $t \in (0,1)$ fixed and $t + c_2(t)sn^{-1/3} \in (0,1)$,

$$\zeta_{nt}^W(s) = c_1(t)A_n^W(t + c_2(t)sn^{-1/3}), \tag{4.5.25}$$

where A_n^W is defined in (4.5.1) and $c_1(t)$ and $c_2(t)$ are defined in (4.4.2). From Theorem 4.4.1 and Lemma 4.5.5 it follows immediately that the process

$$\{\zeta_{nt}^W(s) : s \in I\!R\} \to \{\zeta(s) : s \in I\!R\} \text{ in distribution.} \tag{4.5.26}$$

Furthermore, note that Lemma 4.5.4 implies that for every $m = 1, 2, \ldots$ there exists a constant $M > 0$ such that $EA_n^W(t)^{km} < M$, uniformly in $n = 1, 2 \ldots$ and $t \in (0,1)$. Hence it follows from Markov's inequality, that for all $m = 1, 2, \ldots$ there exists a constant $M' > 0$

$$P\{|\zeta_{nt}^W(s)|^k > y\} \le \frac{M'}{y^m},$$

uniformly in $n = 1, 2 \ldots$, $t \in (0,1)$ and $t + c_2(t)sn^{-1/3} \in (0,1)$. This guarantees uniform integrability of the sequence $\zeta_{nt}^W(s)^k$ for s, t and k fixed, so that together with (4.5.26) it implies convergence of moments of $(\zeta_{nt}^W(0)^k, \zeta_{nt}^W(s)^k)$ to the corresponding moments of $(\zeta(0)^k, \zeta(s)^k)$. This leads to the following lemma.

Lemma 4.5.9 *Suppose that f satisfies conditions (A) and (B). Then for any function g that is continuous on $[0,1]$, and any $k \ge 1$,*

$$\text{var}\left(n^{1/6}\int_0^1 A_n^W(t)^k g(t)dx\right) \to \int_0^1 \frac{2^{(2k+5)/3} f(t)^{(4k+1)/3}}{|f'(t)|^{(2k+2)/3}} g(t)^2\,dt \int_0^\infty \text{cov}(\zeta(0), \zeta(s))\,ds.$$

Proof: We have with ζ_{nt}^W as defined in (4.5.25),

$$\begin{aligned}
&\text{var}\left(n^{1/6}\int_0^1 A_n^W(t)^k g(t)dx\right)\\
&= 2n^{1/3}\int_0^1\int_u^1 \text{cov}\left(A_n^W(t)^k, A_n^W(u)^k\right) g(t)g(u)\,dt\,du\\
&= 2\int_0^1 \frac{c_2(t)}{c_1(t)^{2k}}\int_0^{n^{1/3}(1-t)/c_2(t)} \text{cov}(\zeta_{nt}^W(0)^k, \zeta_{nt}^W(s)^k) g(t) g(t + c_2(t)sn^{-1/3})\,dt\,ds,
\end{aligned}$$

by change of variables of integration $u = t + c_2(t)sn^{-1/3}$. As noted above for s and t fixed,

$$\text{cov}(\zeta_{nt}^W(0)^k, \zeta_{nt}^W(s)^k) \to \text{cov}(\zeta(0)^k, \zeta(s)^k).$$

Lemma 4.5.4 implies that uniformly in $n = 1, 2, \ldots$, s and t, we have $E|\zeta_{nt}^W(0)|^{3k} \leq C_3$ and $E|\zeta_{nt}^W(s)|^{3k} \leq C_4$. Hence by Lemma 4.5.8, it follows that

$$\operatorname{cov}(\zeta_{nt}^W(0)^k, \zeta_{nt}^W(s)^k) \leq C_5\alpha_n \left(n^{-1/3}c_2(t)s\right)^{1/3} \leq D_1 \exp(-D_2|s|^3),$$

where $D_1, D_2 > 0$ do not depend on n, s and t. Substituting $c_1(t)$, $c_2(t)$ as defined in (4.4.2), and using that g is uniformly bounded on $[0, 1]$, it follows by dominated convergence that

$$\operatorname{var}\left(n^{1/6}\int_0^1 A_n^W(t)^k g(t)dx\right) \to \int_0^1 \frac{2^{(2k+5)/3}f(t)^{(4k+1)/3}}{|f'(t)|^{(2k+2)/3}}g(t)^2\,dt \int_0^\infty \operatorname{cov}(\zeta(0)^k, \zeta(s)^k\,ds. \quad \blacksquare$$

Theorem 4.5.1 *Let f be decreasing with support on $[0, 1]$. Suppose that f is twice continuously differentiable satisfying conditions (A) and (B). Let g be a continuous function on $[0, 1]$ and let A_n by defined by (4.1.1). Then for all $k \geq 1$, with*

$$\mu = E\zeta(0)^k \int_0^1 \frac{2^{k/3}f(t)^{2k/3}}{|f'(t)|^{k/3}}g(t)\,dt,$$

$n^{1/6}\left(\int_0^1 A_n(t)^k g(t)\,dt - \mu\right)$ *converges in distribution to a normal random variable with mean zero and variance*

$$\sigma^2 = \int_0^1 \frac{2^{(2k+5)/3}f(t)^{(4k+1)/3}}{|f'(t)|^{(2k+2)/3}}g(t)^2\,dt \int_0^\infty \operatorname{cov}(\zeta(0), \zeta(s))\,ds.$$

Proof: It suffices to prove the statement for T_n^W as defined in (4.5.24). Define

$$X_n(t) \stackrel{def}{=} \left(A_n^W(t)^k - EA_n^W(t)^k\right)g(t).$$

Let

$$L_n = n^{-1/3}\log^3 n, \quad M_n = n^{-1/3}\log n, \quad N_n = \left[\frac{1}{L_n + M_n}\right],$$

where $[x]$ denotes the integer part of x. We divide interval $[0, 1]$ into blocks of alternating length

$$\begin{aligned} A_j &= [(j-1)(L_n + M_n), (j-1)(L_n + M_n) + L_n], \\ B_j &= [(j-1)(L_n + M_n) + L_n, j(L_n + M_n)], \end{aligned}$$

where $1 \leq j \leq N_n$. Now write $T_n^W = S_n' + S_n'' + R_n$, where

$$\begin{aligned} S_n' &= n^{1/6}\sum_{j=1}^{N_n}\int_{A_j} X_n(t)\,dt, \\ S_n'' &= n^{1/6}\sum_{j=1}^{N_n}\int_{B_j} X_n(t)\,dt, \\ R_n &= n^{1/6}\int_{N_n(L_n+M_n)}^1 X_n(t)\,dt. \end{aligned}$$

According to Lemma 4.5.4 and the Cauchy-Schwarz inequality, for all $s, t \in (0, 1)$,

$$E|X_n(s)X_n(t)| \leq C, \tag{4.5.27}$$

where C is uniform with respect to s, t and n. Together with a fact that length of the interval of integration for R_n is $O(n^{-1/3}(\log n)^3)$ this shows $E|R_n| \to 0$ and hence $R_n = o_p(1)$.

Next we show that contribution of integrals over small blocks is negligible. To this end consider

$$E(S_n'')^2 = n^{1/3} \sum_{j=1}^{N_n} E\left(\int_{B_j} X_n(t)\,dt\right)^2 + n^{1/3} \sum_{i \neq j} \int_{B_i} \int_{B_j} EX_n(s)X_n(t)\,ds\,dt.$$

We have that

$$|EX_n(s)X_n(t)| = |g(s)g(t)||\mathrm{cov}(A_n^W(s)^k, A_n^W(t)^k)| \leq D_1 e^{-D_2 n|s-t|^3},$$

where $D_1, D_2 > 0$ do not depend s, t and n, using the fact that g in uniformly bounded on $[0,1]$ together with Lemma 4.5.8. Moreover, for $s \in B_i$ and $t \in B_j$, we have $|s-t| \geq n^{-1/3}(\log n)^3$. Since $N_n = \mathcal{O}(n^{-1/3}(\log n)^3)$ this implies that

$$\left| n^{1/3} \sum_{i \neq j} \int_{B_i} \int_{B_j} EX_n(s)X_n(t)\,ds\,dt \right| \leq n^{1/3} N_n^2 M_n^2 D_1 e^{-D_2(\log n)^9} \to 0.$$

Hence, using (4.5.27) we obtain

$$E(S_n'')^2 = \mathcal{O}(n^{1/3} N_n M_n^2) + o(1) \to 0,$$

so that the contribution of the small blocks is negligible.

Define

$$Y_j = n^{1/6} \int_{A_j} X_n(t)\,dt \quad \text{and} \quad \sigma_n^2 = \mathrm{var}\left(\sum_{j=1}^{N_n} Y_j\right),$$

so that $S_n' = \sum_{j=1}^{N_n} Y_j$ and $\sigma_n^2 = \mathrm{var}(S_n')$. We have

$$\begin{aligned}
&\left| E\exp\left\{\frac{iu}{\sigma_n}\sum_{j=1}^{N_n} Y_j\right\} - \prod_{j=1}^{N_n} \exp\left\{\frac{iu}{\sigma_n} Y_j\right\} \right| \\
&\leq \sum_{k=2}^{N_n} \left| E\exp\left\{\frac{iu}{\sigma_n}\sum_{j=1}^{k} Y_j\right\} - E\exp\left\{\frac{iu}{\sigma_n}\sum_{j=1}^{k-1} Y_j\right\} E\exp\left\{\frac{iu}{\sigma_n} Y_k\right\} \right| \\
&\leq 4(N_n - 1)\alpha_n(M_n),
\end{aligned}$$

where the last inequality follows from Lemma 4.5.7. Observe that $(N_n - 1)\alpha_n(M_n) \to 0$, which means that we can apply the Central Limit Theorem to independent copies of Y_j. Asymptotic normality of S_n' follows if we can show that the independent copies of the Y_j's satisfy the Lindeberg condition, i.e., for all $\epsilon > 0$,

$$\frac{1}{\sigma_n^2} \sum_{j=1}^{N_n} EY_j^2 1_{\{|Y_j| > \epsilon\sigma_n\}} \to 0,$$

as $n \to \infty$. Note that by the Markov inequality $EY_j^2 1_{\{|Y_j| > \epsilon\sigma_n\}} \leq E|Y_j|^3/(\epsilon\sigma_n)$. Again using the Cauchy-Schwarz inequality and uniform boundedness of the moments of $|X_n(t)|$ we obtain

$$\sup_{1 \leq j \leq N_n} E|Y_j|^3 = n^{1/2}\mathcal{O}(|A_j|^3) = \mathcal{O}(n^{-1/2}(\log n)^9).$$

Hence

$$\frac{1}{\sigma_n^2}\sum_{j=1}^{N_n} EY_j^2 1_{|Y_j|>\epsilon\sigma_n} \le \frac{1}{\epsilon\sigma_n^3} N_n \sup_{1\le j\le N_n} E|Y_j|^3 = \mathcal{O}(\sigma_n^{-3} n^{-1/6}(\log n)^6).$$

Note that

$$\sigma_n^2 = \mathrm{var}(S_n^{'}) = \mathrm{var}\left(T_n^W\right) + \mathrm{var}(S_n^{''} + R_n) - 2ET_n^W(S_n^{''} + R_n).$$

Using the already obtained results $E(S_n^{''})^2 = o(1)$ and $ER_n^2 = o(1)$, together with the Cauchy-Schwarz inequality, we conclude that

$$\mathrm{var}(S_n^{''} + R_n) = E(S_n^{''})^2 + ER_n^2 + 2E(S_n^{''}R_n) \to 0,$$

and that according to the Lemma 4.5.9

$$ET_n^W(S_n^{''} + R_n) \le \sqrt{E\left(T_n^W\right)^2 \mathrm{var}(S_n^{''} + R_n)} \to 0.$$

So we find that $\sigma_n^2 = \mathrm{var}(S_n^{'}) = \sigma^2 + o(1)$, which implies

$$\frac{1}{\sigma_n^2}\sum_{j=1}^{N_n} EY_j^2 1_{\{|Y_j|>\epsilon\sigma_n\}} = o(n^{-1/6}(\log n)^6) \to 0. \quad \blacksquare$$

Chapter 5

Grenander estimator at the boundaries of support

We investigate the behavior of the nonparametric maximum likelihood estimator $\hat{f}_n$ for a decreasing density f near the boundaries of the support of f. It is shown that $\hat{f}_n(n^{-\alpha})$ is consistent for $f(0)$, and its limiting distribution is obtained, where we need to distinguish between different values of $0 < \alpha < 1$. Similar results are obtained for the upper endpoint of the support, in the case it is finite. This yields consistent estimators for the values of f at the boundaries of the support. The limit distribution of these estimators is established and their performance is compared with the adjusted NPMLE of WOODROOFE AND SUN (1993).

5.1 Introduction

Let f be a non-increasing density on $[0, \infty)$. The non-parametric maximum likelihood estimator $\hat{f}_n$ for f has been discovered by GRENANDER (1956). It is defined as the left derivative of the concave majorant of the empirical distribution function F_n constructed from a sample $X_1, \ldots, X_n$ from f.

PRAKASA RAO (1969) obtained the asymptotic pointwise behavior of $\hat{f}_n$. GROENEBOOM (1985) provided an elegant proof of the same result, which can be formulated as follows,

$$|4f(x_0)f'(x_0)|^{-1/3}n^{1/3}\left\{\hat{f}_n(x_0) - f(x_0)\right\} \to \underset{t\in\mathbb{R}}{\operatorname{argmax}}\left\{W(t) - t^2\right\},$$

for each $x_0 > 0$, where W denotes standard two sided Brownian motion originating from zero. In contrast, WOODROOFE AND SUN (1993) showed that $\hat{f}_n$ is not consistent at zero. They proposed a penalized maximum likelihood estimator $\hat{f}_n^{\mathrm{P}}(0)$ and in SUN AND WOODROOFE (1996) it was shown that

$$n^{1/3}\left\{\hat{f}_n^{\mathrm{P}}(0) - f(0)\right\} \to \sup_{t>0}\frac{W(t) - (c - \frac{1}{2}f(0)f'(0)t^2)}{t},$$

where c depends on the penalization.

The first distributional result for a global measure of deviation for $\hat{f}_n$ was the convergence of the L_1-distance $\|\hat{f}_n - f\|_1$ in GROENEBOOM (1985) (see GROENEBOOM, HOOGHIEMSTRA AND LOPUHAÄ (1999) for a rigorous proof). Here, the density f was assumed to have compact support, and surprisingly, the inconsistency of $\hat{f}_n$ at the boundaries of the support did not influence the behavior of $\|\hat{f}_n - f\|_1$. Nevertheless, the inconsistency at the boundaries will have an affect if one studies other global measures of deviation, such as the L_k-distance, for k larger than one, or the supremum distance.

In this paper we study the behavior of the Grenander estimator at the boundaries of the support of f. We first consider a non-increasing density f on $[0, \infty)$ and investigate the behavior of

$$n^{\beta}\left\{\hat{f}_n(n^{-\alpha}) - f(n^{-\alpha})\right\}, \tag{5.1.1}$$

where $0 < \alpha < 1$, and $\beta > 0$ is chosen suitably in order to make (5.1.1) converge in distribution. This will imply that $\hat{f}_n(n^{-1/3})$ is a consistent estimator for $f(0)$ at rate $n^{1/3}$ with a limiting distribution that is some complicated functional of W. This estimator will be compared with the penalized maximum likelihood estimator from Sun and Woodroofe (1996). For non-increasing f with compact support, say $[0,1]$, we also investigate the behavior near one. Similarly, this will lead to a consistent estimator for $f(1)$. Moreover, the results on the behavior of $\hat{f}_n$ at the boundaries of $[0,1]$ allows an adequate treatment of the L_k-distance between $\hat{f}_n$ and f. It turns out that for $k > 2.5$, the inconsistency of $\hat{f}_n$ starts to have an affect on the behavior of $\|\hat{f}_n - f\|_k$ (see Kulikov and Lopuhaä (2002)).

In Section 5.2 we give a brief outline of our approach for studying differences such as (5.1.1), and prove some preliminary results for the argmax functional. Section 5.3 is devoted to the behavior of $\hat{f}_n$ near zero. Section 5.4 deals with the behavior of $\hat{f}_n$ near the boundary at the other end of the support for a density f on $[0,1]$. In Section 5.5 we compare our estimator $\hat{f}_n(n^{-1/3})$ with the penalized maximum likelihood estimator from Sun and Woodroofe (1996).

5.2 Preliminaries

Instead of studying the process $\{\hat{f}_n(t) : t \in [0,1]\}$ itself, we will use the more tractable inverse process $\{U_n(a) : a \in [0, f(0)]\}$, where $U_n(a)$ is defined as the last time that the process $F_n(t) - at$ attains its maximum:

$$U_n(a) = \operatorname*{argmax}_{t\in[0,\infty)} \{F_n(t) - at\}.$$

Its relation with $\hat{f}_n$ is as follows: with probability one

$$\hat{f}_n(x) \le a \Leftrightarrow U_n(a) \le x. \tag{5.2.1}$$

Let us first describe the line of reasoning used to prove convergence in distribution of (5.1.1). We illustrate things for the case $0 < \alpha < 1/3$. It turns out that in this case the proper choice for β is $1/3$. Hence, we will consider events of the following type

$$n^{1/3}\left\{\hat{f}_n(n^{-\alpha}) - f(n^{-\alpha})\right\} \le x.$$

According to relation (5.2.1), this event is equivalent with

$$U_n\left(f(n^{-\alpha}) + xn^{-1/3}\right) - n^{-\alpha} \le 0.$$

The left hand is the argmax of the process

$$Z_n(t) = F_n(t + n^{-\alpha}) - f(n^{-\alpha})t - xtn^{-1/3}.$$

With suitable scaling, the process Z_n converges in distribution to some Gaussian process Z. The next step is to use an argmax version of the continuous mapping theorem from KIM AND POLLARD (1990). The version that suffices for our purposes, is stated below for further reference.

Theorem 5.2.1 *Let $\{Z(t) : t \in \mathbb{R}\}$ be a continuous random process satisfying*

(i) Z has a unique maximum with probability one,

(ii) $Z(t) \to -\infty$, as $|t| \to \infty$, with probability one.

Let $\{Z_n(t) : t \in \mathbb{R}\}$ be a sequence of random processes satisfying

(iii) $\operatorname{argmax}_{t\in\mathbb{R}} Z_n(t) = O_p(1)$, as $n \to \infty$.

If Z_n converges in distribution to Z, in the topology of uniform convergence on compacta, then $\operatorname{argmax}_{t\in\mathbb{R}} Z_n(t)$ converges in distribution to $\operatorname{argmax}_{t\in\mathbb{R}} Z(t)$.

Application of this theorem yields that $U_n\left(f(n^{-\alpha}) + xn^{-1/3}\right)$, properly scaled, convergences in distribution to the argmax of a Gaussian process. Convergence of (5.1.1) then follows from another application of (5.2.1).

The main difficulty in verifying the conditions of Theorem 5.2.1, is showing that (iii) holds. In the process of proving condition (iii) we will frequently use the following lemma, which enables us to suitably bound the argmax from above.

Lemma 5.2.1 *Let f and g be continuous functions on $K \subset \mathbb{R}$.*

(i) Suppose that g is non-increasing. Then $\operatorname{argmax}_{x\in K}\{f(x)+g(x)\} \leq \operatorname{argmax}_{x\in K} f(x)$.

(ii) Let $C > 0$ and suppose that for all $s, t \in K$, such that $t - s \geq C$, we have that $g(t) \leq g(s)$. Then $\operatorname{argmax}_{x\in K}\{f(x)+g(x)\} \leq C + \operatorname{argmax}_{x\in K} f(x)$.

Proof: Let $x_0 = \operatorname{argmax}_{x\in K} f(x) < \infty$. If $x_0 = \infty$, there is nothing left to prove, therefore assume that $x_0 < \infty$.

(i) By definition of x_0 and the fact that g is non-increasing, for $x \geq x_0$, we must have $f(x) + g(x) \leq f(x_0) + g(x_0)$. Hence, we must have

$$\operatorname*{argmax}_{x\in K}\{f(x)+g(x)\} \leq x_0 = \operatorname*{argmax}_{x\in K} f(x).$$

This proves (i).

(ii) If $(C + x_0, \infty) \cap K = \emptyset$, the statement is trivially true, so only consider the case $(C + x_0, \infty) \cap K \neq \emptyset$. Then by defintion $f(x) \leq f(x_0)$, for all $x \in (C + x_0, \infty) \cap K$, and by the property of g, we also have $g(x) \leq g(x_0)$, for $x \in (C + x_0, \infty) \cap K$. This implies $f(x) + g(x) \leq f(x_0) + g(x_0)$, for all $x \in (C + x_0, \infty) \cap K$. Hence, we must have

$$\operatorname*{argmax}_{x\in K}\{f(x)+g(x)\} \leq C + x_0 = C + \operatorname*{argmax}_{x\in K} f(x).$$

This proves the lemma. ■

In studying processes like Z_n we will use a Brownian approximation similar to one used in in GROENEBOOM, HOOGHIEMSTRA AND LOPUHAÄ (1999). Let E_n denote the empirical process

$\sqrt{n}(F_n - F)$. For $n \geq 1$, let B_n be versions of the Brownian bridge constructed on the same probability space as the uniform empirical process $E_n \circ F^{-1}$ via the Hungarian embedding, where

$$\sup_{t\in[0,1]} |E_n(t) - B_n(F(t))| = O_p(n^{-1/2}\log n) \tag{5.2.2}$$

(see Komlos, Major and Tusnady (1975)). Define versions W_n of Brownian motion by

$$W_n(t) = B_n(t) + \xi_n t, \quad t \in [0,1],$$

where ξ_n is standard normal random variable independent of B_n. This means that we can represent B_n by $B_n(t) \stackrel{d}{=} W_n(t) - tW_n(1)$.

5.3 Behavior near zero

We first consider the case that f is a non-increasing density on $[0, \infty)$ satisfying

(C1) $0 < f(0) = \lim_{x\downarrow 0} f(x) < \infty$.

(C2) f' is right continuous at zero, such that $0 < |f'(0)| < \sup_{s\geq 0} |f'(s)| < \infty$.

The latter condition will ensure that $F(t) - f(0)t$ is suitably bounded, as is shown in the following lemma.

Lemma 5.3.1 *Suppose that f satisfies (C2). Then there exists a value $t_0 > 0$, such that $\inf_{0\leq s\leq t_0} |f'(s)| > 0$, and for any $0 \leq c \leq \frac{1}{2}t_0$,*

$$F(c+t) - F(c) - f(c)t \leq \begin{cases} -\frac{1}{2}\inf_{0\leq s\leq t_0}|f'(s)|t^2, & 0 \leq t \leq \frac{1}{2}t_0, \\ -\frac{1}{4}t_0 \inf_{0\leq s\leq t_0}|f'(s)|t, & t > \frac{1}{2}t_0. \end{cases}$$

Proof: The existence of $t_0 > 0$ follows directly from condition (C2). For $0 \leq t \leq \frac{1}{2}t_0$, the inequality is a direct consequence of a Taylor expansion. For $t > \frac{1}{2}t_0$,

$$\begin{aligned} F(c+t) - F(c) - f(c)t &= F(c+\tfrac{1}{2}t_0) - F(c) - \tfrac{1}{2}f(c)t_0 \\ &\quad + F(c+t) - F(c+\tfrac{1}{2}t_0) - f(c+\tfrac{1}{2}t_0)(t-\tfrac{1}{2}t_0) \\ &\quad + \left(f(c+\tfrac{1}{2}t_0) - f(c)\right)(t-\tfrac{1}{2}t_0) \\ &= \tfrac{1}{8}f'(\theta_1)t_0^2 + \tfrac{1}{2}f'(\theta_2)(t-\tfrac{1}{2}t_0)^2 + \tfrac{1}{2}f'(\theta_3)(t-\tfrac{1}{2}t_0)t_0 \\ &\leq -\tfrac{1}{8}\inf_{0\leq s\leq t_0}|f'(s)|t_0^2 - \tfrac{1}{2}\inf_{0\leq s\leq t_0}|f'(s)|t_0(t-\tfrac{1}{2}t_0) \\ &\leq -\tfrac{1}{4}t_0\inf_{0\leq s\leq t_0}|f'(s)|t. \quad \blacksquare \end{aligned}$$

In studying the behavior of (5.1.1), we follow the line of reasoning described in Section 5.2. We start by establishing convergence in distribution of the relevant processes. It turns out that we have to distinguish between three cases concerning the rate at which $n^{-\alpha}$ tends to zero.

Lemma 5.3.2 *Let f satisfy (C1)-(C2) and let W denote standard two-sided Brownian motion on $\mathbb{R}$. For $1/3 \leq \alpha < 1$, $t \geq 0$ and $x \in \mathbb{R}$, define*

$$Z_{n1}(x,t) = n^{(1+\alpha)/2}\left(F_n(tn^{-\alpha}) - f(0)tn^{-\alpha}\right) - xt.$$

(i) *For $1/3 < \alpha < 1$, the process $\{Z_{n1}(x,t) : t \in [0,\infty)\}$ converges in distribution, in the uniform topology on compacta, to the process $\{W(f(0)t) - xt : t \in [0,\infty)\}$.*

(ii) *For $\alpha = 1/3$, the process $\{Z_{n1}(x,t) : t \in [0,\infty)\}$ converges in distribution, in the uniform topology on compacta, to the process $\{W(f(0)t) - xt + \frac{1}{2}f'(0)t^2 : t \in [0,\infty)\}$.*

(iii) *For $0 < \alpha < 1/3$, $t \geq -n^{1/3-\alpha}$ and $x \in \mathbb{R}$, define*

$$Z_{n2}(x,t) = n^{2/3}\left(F_n(n^{-\alpha} + tn^{-1/3}) - F_n(n^{-\alpha}) - f(n^{-\alpha})tn^{-1/3}\right) - xt.$$

Then the process $\{Z_{n2}(x,t) : t \in [-n^{1/3-\alpha},\infty)\}$ converges in distribution, in the uniform topology on compacta, to the process $\{W(f(0)t) - xt + \frac{1}{2}f'(0)t^2 : t \in \mathbb{R}\}$.

Proof: (i) Decompose the process Z_{n1} as follows,

$$\begin{aligned} Z_{n1}(x,t) &= n^{\alpha/2}W_n\left(F(tn^{-\alpha})\right) + n^{(1+\alpha)/2}\left\{F(tn^{-\alpha}) - f(0)tn^{-\alpha}\right\} - xt \\ &\quad -n^{\alpha/2}F(tn^{-\alpha})W_n(1) + n^{\alpha/2}H_n(tn^{-\alpha}), \end{aligned}$$

where $H_n(t) = E_n(t) - B_n(F(t))$. By Brownian scaling, the process $n^{\alpha/2}W_n(F(tn^{-\alpha}))$ has the same distribution as the process $W(n^\alpha F(tn^{-\alpha}))$, and by uniform continuity of Brownian motion on compacta,

$$W\left(n^\alpha F(tn^{-\alpha})\right) - W(f(0)t) \to 0,$$

uniformly for t in compact sets. Since $\alpha > 1/3$ we have that

$$n^{(1+\alpha)/2}\left\{F(tn^{-\alpha}) - f(0)tn^{-\alpha}\right\} \to 0,$$

uniformly for t in compact sets. Because $n^{\alpha/2}F(tn^{-\alpha})W_n(1) = \mathcal{O}_p(n^{-\alpha/2})$, together with (5.2.2) this proves (i). In case (ii), where $\alpha = 1/3$, the only difference is the behavior of the deterministic term

$$n^{2/3}\left\{F(tn^{-1/3}) - f(0)tn^{-1/3}\right\} \to \tfrac{1}{2}f'(0)t^2,$$

uniformly for t in compact sets. Similar to the proof of (i), using Brownian scaling and uniform continuity of Brownian motion on compacta this proves (ii).

For case (iii) the process Z_{n2} can be written as

$$\begin{aligned} & n^{1/6}\left\{W_n\big(F(n^{-\alpha} + tn^{-1/3})\big) - W_n\big(F(n^{-\alpha})\big)\right\} \\ & +n^{2/3}\left\{F(n^{-\alpha} + tn^{-1/3}) - F(n^{-\alpha}) - f(n^{-\alpha})tn^{-1/3}\right\} - xt \\ & -n^{1/6}\left\{F(n^{-\alpha} + tn^{-1/3}) - F(n^{-\alpha})\right\}W_n(1) + n^{1/6}H_n(n^{-\alpha} + tn^{-1/3}) - n^{1/6}H_n(n^{-\alpha}). \end{aligned}$$

The process $n^{1/6}\left\{W_n\big(F(n^{-\alpha} + tn^{-1/3})\big) - W_n\big(F(n^{-\alpha})\big)\right\}$ has the same distribution as the process $W\big(n^{1/3}(F(n^{-\alpha} + tn^{-1/3}) - F(n^{-\alpha}))\big)$, and by uniform continuity of Brownian motion on compacta,

$$W\big(n^{1/3}(F(n^{-\alpha} + tn^{-1/3}) - F(n^{-\alpha}))\big) - W\big(f(0)t\big) \to 0,$$

uniformly for t in compact sets. Finally,

$$n^{2/3}\left\{F(n^{-\alpha} + tn^{-1/3}) - F(n^{-\alpha}) - f(n^{-\alpha})tn^{-1/3}\right\} \to \tfrac{1}{2}f'(0)t^2,$$

uniformly for t in compact sets. Since $n^{1/6}\left\{F(n^{-\alpha} + tn^{-1/3}) - F(n^{-\alpha})\right\}W_n(1) = \mathcal{O}_p(n^{-1/6})$, together with (5.2.2) this proves (iii). ■

The next step is to use Theorem 5.2.1. The following lemma will ensure that condition (iii) of this theorem is satisfied.

Lemma 5.3.3 *Suppose that f satisfies (C1)-(C2). Let Z_{n1} and Z_{n2} be defined as in Lemma 5.3.2.*

(i) For $1/3 < \alpha < 1$ and $x > 0$, $\operatorname{argmax}_{t\in[0,\infty)} Z_{n1}(x,t) = \mathcal{O}_p(1)$.

(ii) For $\alpha = 1/3$ and $x \in \mathbb{R}$, $\operatorname{argmax}_{t\in[0,\infty)} Z_{n1}(x,t) = \mathcal{O}_p(1)$.

(iii) For $0 < \alpha < 1/3$ and $x \in \mathbb{R}$, $\operatorname{argmax}_{t\in[-n^{1/3-\alpha},\infty)} Z_{n2}(x,t) = \mathcal{O}_p(1)$.

Proof: (i) First note that $\operatorname{argmax}_t Z_{n1}(x,t)$ has the same distribution as $\operatorname{argmax}_t M_{n1}(t)$, where

$$\begin{aligned} M_{n1}(t) &= W_n\left(n^{\alpha}F(tn^{-\alpha})\right) + n^{(1+\alpha)/2}\left(F(tn^{-\alpha}) - f(0)tn^{-\alpha}\right) - xt \\ &\quad -n^{\alpha/2}F(tn^{-\alpha})W_n(1) + n^{\alpha/2}H_n(tn^{-\alpha}). \end{aligned}$$

Let $0 < \epsilon < x$ and define $X_{n1}(t) = n^{\alpha/2}H_n(tn^{-\alpha}) - \frac{1}{2}\epsilon t$. Next, consider the event

$$A_{n1} = \{X_{n1}(s) \geq X_{n1}(t), \text{ for all } s,t \in [0,\infty), \text{ such that } t - s \geq \delta_n\}. \tag{5.3.1}$$

Then with $\delta_n = n^{-(1-\alpha)/2}(\log n)^2$, by using (5.2.2) we have that

$$P(A_{n1}) \geq P\left\{\sup_{t\in[0,\infty)} |H_n(t)| \leq \frac{\epsilon}{4}n^{-1/2}(\log n)^2\right\} \to 1.$$

Also define the process $X_{n2}(t) = -n^{\alpha/2}F(tn^{-\alpha})W_n(1) - \frac{1}{2}\epsilon t$, and consider the event

$$A_{n2} = \{X_{n2}(s) \geq X_{n2}(t), \text{ for all } 0 \leq s \leq t < \infty\}. \tag{5.3.2}$$

Then, since every sample path of the process X_{n2} is differentiable, we have

$$P(A_{n2}) \geq P\left\{-f(tn^{-\alpha})W_n(1) - \frac{\epsilon}{2}n^{\alpha/2} \leq 0, \text{ for all } t \in [0,\infty)\right\} \to 1.$$

Hence, if $A_n = A_{n1} \cap A_{n2}$, then $P(A_n) \to 1$. Since for any $\eta > 0$,

$$P\left\{\operatorname*{argmax}_{t\in[0,\infty)} M_{n1}(t)1_{A_n^c} > \eta\right\} \leq P(A_n^c) \to 0,$$

we conclude that $(\operatorname{argmax}_t M_{n1}(t))\, 1_{A_n^c} = \mathcal{O}_p(1)$. This means that we only have to consider $(\operatorname{argmax}_t M_{n1}(t))\, 1_{A_n}$. From Lemma 5.2.1, we have

$$\left(\operatorname*{argmax}_{t\in[0,\infty)} M_{n1}(t)\right) 1_{A_n} \leq \operatorname*{argmax}_{t\in[0,\infty)} S_{n1}(t) + \delta_n, \tag{5.3.3}$$

where

$$S_{n1}(t) = W_n\left(n^{-\alpha}F(tn^{-\alpha})\right) - (x-\epsilon)t + n^{(1+\alpha)/2}\left(F(tn^{-\alpha}) - f(0)tn^{-\alpha}\right).$$

Since $F(tn^{-\alpha}) - f(0)tn^{-\alpha}$ is non-increasing for $t \geq 0$, according to Lemma 5.2.1,

$$\operatorname*{argmax}_{t\in[0,\infty)} S_{n1}(t) \leq \operatorname*{argmax}_{t\in[0,\infty)} \left\{W_n\big(n^{\alpha}F(tn^{-\alpha})\big) - (x-\epsilon)t\right\}. \tag{5.3.4}$$

By change of variables $u = G(t) = n^{\alpha}F(tn^{-\alpha})$, and using that for $u \in [0, n^{\alpha}]$,

$$\frac{u}{f(0)} \leq G^{-1}(u) \leq \frac{u}{f\big(F^{-1}(un^{-\alpha})\big)}, \tag{5.3.5}$$

we find that

$$\begin{aligned}\operatorname*{argmax}_{t\in[0,\infty)} \left\{W_n\big(n^{\alpha}F(tn^{-\alpha})\big) - (x-\epsilon)t\right\} &\leq \sup_{t\in[0,\infty)} \left\{W_n\big(n^{\alpha}F(tn^{-\alpha})\big) - (x-\epsilon)t \geq 0\right\} \\ &= G^{-1}\left(\sup\left\{u \in [0, n^{\alpha}] : W_n(u) - (x-\epsilon)G^{-1}(u) \geq 0\right\}\right) \\ &\leq G^{-1}\left(\sup\left\{u \in [0,\infty) : W_n(u) - \frac{x-\epsilon}{f(0)}u \geq 0\right\}\right).\end{aligned}$$

By Brownian scaling, $\sup\{u \in [0,\infty) : W_n(u) - f(0)^{-1}(x-\epsilon)u \geq 0\}$ has the same distribution as

$$\frac{f(0)^2}{(x-\epsilon)^2} \sup\{u \in [0,\infty) : W(u) - u \geq 0\},$$

which is of order $\mathcal{O}_p(1)$. The latter can be seen for instance from the law of iterated logarithm for Brownian motion:

$$P\left\{\limsup_{|u|\to\infty} \frac{W(u)}{\sqrt{2|u|\log\log|u|}} = 1\right\} = 1. \tag{5.3.6}$$

Because $\delta_n = n^{-(1-\alpha)/2}(\log n)^2 = o(1)$, together with (5.3.3), (5.3.4) and (5.3.5), it follows that

$$0 \leq \operatorname*{argmax}_{t\in[0,\infty)} M_{n1}(t) \leq \left(\operatorname*{argmax}_{t\in[0,\infty)} M_{n1}(t)\right) 1_{A_n} + \mathcal{O}_p(1) \leq \frac{\mathcal{O}_p(1)}{f(F^{-1}(\mathcal{O}_p(n^{-\alpha})))} + \mathcal{O}_p(1),$$

which proves (i).

(ii) In this case $\alpha = 1/3$, so that the argument up to (5.3.3) is the same. Let $\epsilon > 0$ and $A_n = A_{n1} \cap A_{n2}$, where A_{n1} as defined in (5.3.1) with $\delta_n = n^{-1/3}(\log n)^2$ and A_{n2} as defined in (5.3.2). We now find that

$$\left(\operatorname*{argmax}_{t\in[0,\infty)} M_{n1}(t)\right) 1_{A_n} \leq \operatorname*{argmax}_{t\in[0,\infty)} S_{n1}(t) + \delta_n \leq \sup\{t \geq 0 : S_{n1}(t) \geq 0\} + \delta_n, \tag{5.3.7}$$

where

$$S_{n1}(t) = W_n\big(n^{1/3}F(tn^{-1/3})\big) - (x-\epsilon)t + n^{2/3}\left(F(tn^{-1/3}) - f(0)tn^{-1/3}\right).$$

Let t_0 be the value from Lemma 5.3.1 and consider the event

$$D_{n1} = \left\{n^{-1/3}\sup\left\{t \geq 0 : S_{n1}(t) \geq 0\right\} \leq \frac{1}{2}t_0\right\}.$$

If $S_{n1}(t) \geq 0$, then according to Lemma 5.3.1, writing $\inf|f'|$ instead of $\inf_{0\leq s\leq t_0}|f'(s)|$, for $tn^{-1/3} > \frac{1}{2}t_0$ and n sufficiently large, we find that

$$\begin{aligned}
0 &\leq W_n\big(n^{1/3}F(tn^{-1/3})\big) - (x-\epsilon)t + n^{2/3}\left(F(tn^{-1/3}) - f(0)tn^{-1/3}\right) \\
&\leq n^{1/6}\left\{\sup_{0\leq u\leq 1}|W_n(u)| - (x-\epsilon)tn^{-1/6} - \frac{1}{4}t_0\inf|f'|tn^{1/6}\right\} \\
&\leq n^{1/6}\left\{\sup_{0\leq u\leq 1}|W_n(u)| - \frac{1}{4}t_0\inf|f'|tn^{1/6}\left(1+\frac{(x-\epsilon)n^{-1/3}}{\frac{1}{4}t_0\inf|f'|}\right)\right\} \\
&\leq n^{1/6}\left\{\sup_{0\leq u\leq 1}|W_n(u)| - \frac{1}{16}t_0^2\inf|f'|n^{1/2}\right\}.
\end{aligned}$$

Therefore

$$P\left(D_{n1}^c\right) \leq P\left(\sup_{0\leq u\leq 1}|W(u)| \geq \frac{1}{16}t_0^2\inf|f'|n^{1/2}\right) \to 0.$$

This means we can restrict ourselves to the event $A_n \cap D_{n1}$, so that by analogous reasoning as before, from (5.3.7) we get

$$\begin{aligned}
\left(\operatorname*{argmax}_{t\in[0,\infty)} M_{n1}(t)\right)1_{A_n\cap D_{n1}} &\leq \sup\left\{t\geq 0 : S_{n1}(t)\geq 0\right\}1_{D_{n1}} + \delta_n \\
&\leq \sup\left\{0\leq t\leq \frac{1}{2}t_0n^{1/3} : S_{n1}(t)\geq 0\right\} + \delta_n.
\end{aligned}$$

According to Lemma 5.3.1, for $0 \leq tn^{-1/3} \leq \frac{1}{2}t_0$, we have $n^{2/3}\left(F(tn^{-1/3}) - f(0)tn^{-1/3}\right) \leq -\frac{1}{2}\inf|f'|t^2$, so that

$$\begin{aligned}
0 &\leq \left(\operatorname*{argmax}_{t\in[0,\infty)} M_{n1}(t)\right)1_{A_n\cap D_{n1}} \qquad (5.3.8)\\
&\leq \sup\left\{0\leq tn^{-1/3}\leq \frac{1}{2}t_0 : W_n\big(n^{1/3}F(tn^{-1/3})\big) - (x-\epsilon)t - \tfrac{1}{2}t^2\inf|f'|\geq 0\right\} + \delta_n.
\end{aligned}$$

Next, distinguish between

(A) $-(x-\epsilon)t - \frac{1}{4}t^2\inf|f'| \geq 0$,

(B) $-(x-\epsilon)t - \frac{1}{4}t^2\inf|f'| < 0$,

Since $t \geq 0$, case (A) can only occur when $x - \epsilon < 0$, in which case we have $0 \leq t \leq 4(\epsilon - x)/\inf|f'|$, which is of order $\mathcal{O}(1)$. In case (B), it follows that $W_n\big(F(t)\big) - \frac{1}{4}t^2\inf|f'| \geq 0$. We conclude from (5.3.8) that

$$\begin{aligned}
0 &\leq \left(\operatorname*{argmax}_{t\in[0,\infty)} M_{n1}(t)\right)1_{A_n\cap D_{n1}} \\
&\leq \sup\left\{0\leq tn^{-1/3}\leq \frac{1}{2}t_0 : W_n\big(n^{1/3}F(tn^{-1/3})\big) - \tfrac{1}{4}t^2\inf|f'|\geq 0\right\} + \mathcal{O}_p(1) + \delta_n \\
&\leq \sup\left\{t\in[0,\infty) : W_n\big(n^{1/3}F(tn^{-1/3})\big) - \tfrac{1}{4}t^2\inf|f'|\geq 0\right\} + \mathcal{O}_p(1). \qquad (5.3.9)
\end{aligned}$$

Similar to the proof of (i), by change of variables $u = G(t) = n^{1/3}F(tn^{-1/3})$ and using (5.3.5) with $\alpha = 1/3$, we find that the argmax on the right hand side of (5.3.9) is bounded from above by

$$G^{-1}\left(\sup\left\{u \in [0,\infty) : W_n(u) - \frac{\inf|f'|}{4f(0)^2}u^2 \geq 0\right\}\right) + \mathcal{O}_p(1).$$

By Brownian scaling, $\sup\{u \in [0,\infty) : W_n(u) - (4f(0)^2)^{-1}\inf|f'|u^2 \geq 0\}$ has the same distribution as

$$\left(\frac{4f(0)^2}{\inf|f'|}\right)^{2/3}\sup\left\{u \in [0,\infty) : W(u) - u^2 \geq 0\right\}.$$

Again by using (5.3.6), this is of order $\mathcal{O}_p(1)$. Similar to the proof of (i), from (5.3.9) we find that

$$0 \leq \operatorname*{argmax}_{t\in[0,\infty)} M_{n1}(t) \leq \left(\operatorname*{argmax}_{t\in[0,\infty)} M_{n1}(t)\right) 1_{A_n\cap D_{n1}} + \mathcal{O}_p(1) \leq \frac{\mathcal{O}_p(1)}{f\left(F^{-1}(\mathcal{O}_p(n^{-1/3}))\right)} + \mathcal{O}_p(1),$$

which proves (ii).

(iii) Because adding constant terms to the process Z_{n2} does not change the location of the maximum, we have that $\operatorname{argmax}_t Z_{n2}(x,t)$ has the same distribution as $\operatorname{argmax}_t M_{n2}(t)$, where for $t \in [-n^{1/3-\alpha},\infty)$,

$$\begin{aligned} M_{n2}(t) &= W_n\left(n^{1/3}(F(n^{-\alpha}+tn^{-1/3}) - F(n^{-\alpha}))\right) \\ &\quad + n^{2/3}\left(F(n^{-\alpha}+tn^{-1/3}) - F(n^{-\alpha}) - f(n^{-\alpha})tn^{-1/3}\right) - xt \\ &\quad - n^{1/6}F(n^{-\alpha}+tn^{-1/3})W_n(1) + n^{1/6}H_n(n^{-\alpha}+tn^{-1/3}). \end{aligned}$$

Let $\epsilon > 0$ and $A_n = A_{n1} \cap A_{n2}$, with A_{n1} defined similar to (5.3.1) with $\delta_n = n^{-1/3}(\log n)^2$, and A_{n2} defined similar to (5.3.2). By the same argument as in the proof of (i) and (ii), it suffices to consider $(\operatorname{argmax}_t M_{n2}(t))\,1_{A_n}$. We find

$$\left(\operatorname*{argmax}_{t\in[-n^{1/3-\alpha},\infty)} M_{n2}(t)\right) 1_{A_n} \leq \operatorname*{argmax}_{t\in[-n^{1/3-\alpha},\infty)} S_{n2}(t) + \delta_n \leq \sup\{t \geq 0 : S_{n2}(t) \geq 0\} + \delta_n,$$

where

$$\begin{aligned} S_{n2}(t) &= W_n\left(n^{1/3}(F(n^{-\alpha}+tn^{-1/3}) - F(n^{-\alpha}))\right) \\ &\quad + n^{2/3}\left(F(n^{-\alpha}+tn^{-1/3}) - F(n^{-\alpha}) - f(n^{-\alpha})tn^{-1/3}\right) - (x-\epsilon)t. \end{aligned}$$

As in the proof of (ii), consider $D_{n2} = \left\{n^{-1/3}\sup\{t \geq 0 : S_{n2}(t) \geq 0\} \leq \frac{1}{2}t_0.\right\}$. By the same reasoning as used in the proof of (ii), it again follows from Lemma 5.3.1 that $P(D_{n2}^c) \to 0$, so that we only have to consider $(\operatorname{argmax}_t S_{n2}(t))\,1_{D_{n2}}$. Hence, similar to the proof of (ii) we get

$$\left(\operatorname*{argmax}_{t\in[-n^{1/3-\alpha},\infty)} M_{n2}(t)\right) 1_{A_n\cap D_{n2}} \leq \sup\left\{0 \leq tn^{-1/3} \leq \frac{1}{2}t_0 : S_{n2}(t) \geq 0\right\} + \delta_n.$$

According to Lemma 5.3.1, $n^{2/3}\left(F(n^{-\alpha}+tn^{-1/3}) - F(n^{-\alpha}) - f(n^{-\alpha})tn^{-1/3}\right) \leq -\frac{1}{2}t^2\inf|f'|$, for $0 \leq tn^{-1/3} \leq \frac{1}{2}t_0$, so that $\sup\{0 \leq tn^{-1/3} \leq \frac{1}{2}t_0 : S_{n2}(t) \geq 0\}$ is bounded from above by

$$\begin{aligned} \sup\Big\{0 \leq tn^{-1/3} \leq \tfrac{1}{2}t_0 &: W_n\left(n^{1/3}(F(n^{-\alpha}+tn^{-1/3}) - F(n^{-\alpha}))\right) \\ &\qquad - (x-\epsilon)t - \tfrac{1}{2}t^2\inf|f'| \geq 0\Big\}. \end{aligned}$$

Similar to (5.3.9), we conclude that $(\mathrm{argmax}_t\, M_{n2}(t))\, 1_{A_n \cap D_{n2}}$ is bounded from above by

$$\sup\left\{t \in [0,\infty) : W_n\left(n^{1/3}\left(F(n^{-\alpha}+tn^{-1/3}) - F(n^{-\alpha})\right)\right) - \tfrac{1}{4}t^2 \inf|f'| \geq 0\right\} + \mathcal{O}_p(1). \quad (5.3.10)$$

Next, change variables $u = G(t) = n^{1/3}\left(F(n^{-\alpha}+tn^{-1/3}) - F(n^{-\alpha})\right)$. Then for any $u \in [0, n^{1/3}(1-F(n^{-\alpha}))]$,

$$\frac{u}{f(0)} \leq G^{-1}(u) \leq \frac{u}{f\left(F^{-1}(un^{-1/3}+F(n^{-\alpha}))\right)}, \quad (5.3.11)$$

so that (5.3.10) is bounded from above by

$$G^{-1}\left(\sup\left\{u \geq 0 : W_n(u) - \frac{\inf|f'|}{4f(0)^2}u^2 \geq 0\right\}\right) + \mathcal{O}_p(1).$$

As in the proof of (ii), by Brownian scaling together with (5.3.11), we find that

$$\begin{aligned} \underset{t\in[-n^{1/3-\alpha},\infty)}{\mathrm{argmax}}\ M_{n2}(t) &\leq \left(\underset{t\in[-n^{1/3-\alpha},\infty)}{\mathrm{argmax}}\ M_{n2}(t)\right) 1_{A_n\cap D_{n2}} + \mathcal{O}_p(1) \quad (5.3.12)\\ &\leq \frac{\mathcal{O}_p(1)}{f\left(F^{-1}(\mathcal{O}_p(n^{-1/3})+F(n^{-\alpha}))\right)} + \mathcal{O}_p(1) = \mathcal{O}_p(1). \end{aligned}$$

To obtain a lower bound for the left hand side of (5.3.12), first note that

$$\underset{t\in[-n^{1/3-\alpha},\infty)}{\mathrm{argmax}}\ M_{n2}(t) \geq \underset{t\in[-n^{1/3-\alpha},0]}{\mathrm{argmax}}\ M_{n2}(t) = -\underset{t\in[0,n^{1/3-\alpha}]}{\mathrm{argmax}}\ M_{n2}(-t). \quad (5.3.13)$$

From here, the argument runs along the same lines as for the upper bound. Let $\epsilon > 0$ and, similar to (5.3.1) and (5.3.2), define the events A_{n1} and A_{n2} with

$$\begin{aligned} X_{n1}(t) &= n^{1/6}H_n(n^{-\alpha}-tn^{-1/3}) - \tfrac{1}{2}\epsilon t,\\ X_{n2}(t) &= -n^{1/6}F(n^{-\alpha}-tn^{-1/3}) - \tfrac{1}{2}\epsilon t. \end{aligned}$$

With $A_n = A_{n1}\cap A_{n2}$, as before we get $(\mathrm{argmax}_t\, M_{n2}(-t))\, 1^c_{A_n} = \mathcal{O}_p(1)$. Moreover, using that $F(n^{-\alpha}-tn^{-1/3}) - F(n^{-\alpha}) + f(n^{-\alpha})tn^{-1/3} + \frac{1}{2}t^2\inf|f'|n^{-2/3}$ is non-increasing for $t \in [0, n^{1/3-\alpha}]$, similar to proof of the upper bound, we find that $-(\mathrm{argmax}_t\, M_{n2}(-t))\, 1_{A_n}$ is bounded from below by

$$-\sup\left\{t \in [0,n^{1/3-\alpha}] : W_n\left(n^{1/3}\left(F(n^{-\alpha}-tn^{-1/3}) - F(n^{-\alpha})\right)\right) - \tfrac{1}{4}t^2\inf|f'| \geq 0\right\} + \mathcal{O}_p(1).$$

After change of variables $u = G(t) = n^{1/3}\left(F(n^{-\alpha}-tn^{-1/3}) - F(n^{-\alpha})\right)$, and using that for $u \in [-n^{1/3}F(n^{-\alpha}), 0]$, one has

$$-\frac{u}{f(0)} \leq G^{-1}(u) \leq -\frac{u}{f(n^{-\alpha})},$$

we now find that

$$-\underset{t\in[0,n^{1/3-\alpha}]}{\mathrm{argmax}}\ M_{n2}(-t) \geq \frac{1}{f(n^{-\alpha})}\sup\left\{u \leq 0 : W_n(u) - \frac{\inf|f'|}{4f(0)^2}u^2 \geq 0\right\} + \mathcal{O}_p(1).$$

As above, by Brownian scaling together with (5.3.13), it follows that

$$\operatorname*{argmax}_{t\in[-n^{1/3-\alpha},\infty)} M_{n2}(t) \geq \frac{\mathcal{O}_p(1)}{f(n^{-\alpha})} + \mathcal{O}_p(1) = \mathcal{O}_p(1).$$

Together with (5.3.12) this proves the lemma. ■

We are now able to determine the behavior of the Grenander estimator at zero. With the proper normalizing constants the limit distribution of $n^{-\beta}\big(\hat{f}_n(n^{-\alpha}) - f(n^{-\alpha})\big)$ is independent of f. In the case $\alpha = 1/3$, we are only able to specify the distribution function of the limiting random variable.

Theorem 5.3.1 *Let f satisfy conditions (C1)-(C2). Then*

(i) For $1/3 < \alpha < 1$ and $A_{01} = f(0)^{-1/2}$, we have that

$$A_{01}n^{(1-\alpha)/2}\left(\hat{f}_n(n^{-\alpha}) - f(n^{-\alpha})\right) \to \sqrt{\operatorname*{argmax}_{t\in[0,\infty)} \{W(t) - t\}}$$

in distribution, as $n \to \infty$.

(ii) For $A_{02} = |4f(0)f'(0)|^{-1/3}$ and $B_{02} = 4^{1/3}f(0)^{1/3}|f'(0)|^{-2/3}$, we have that

$$A_{02}\left\{n^{1/3}\left(\hat{f}_n(B_{02}n^{-1/3}) - f(n^{-1/3})\right) + f'(0)\right\} \to V$$

in distribution, as $n \to \infty$, where V has distribution function

$$\Psi(x) = P\left\{\operatorname*{argmax}_{t\geq 0}\{W(t) - (t+x)^2\} \leq 1\right\}.$$

(iii) For $0 < \alpha < 1/3$ and $A_{03} = |4f(0)f'(0)|^{-1/3}$, we have that

$$A_{03}n^{1/3}\left(\hat{f}_n(n^{-\alpha}) - f(n^{-\alpha})\right) \to \operatorname*{argmax}_{t\in\mathbb{R}} \{W(t) - t^2\}$$

in distribution, as $n \to \infty$.

Proof: (i) First note that by condition (C1),

$$n^{(1-\alpha)/2}\left(\hat{f}_n(n^{-\alpha}) - f(n^{-\alpha})\right) = n^{(1-\alpha)/2}\left(\hat{f}_n(n^{-\alpha}) - f(0)\right) + o(n^{(1-3\alpha)/2}),$$

where $(1-3\alpha)/2 < 0$. For $x > 0$, according to (5.2.1),

$$P\left\{n^{(1-\alpha)/2}\left(\hat{f}_n(n^{-\alpha}) - f(0)\right) \leq x\right\} = P\left\{n^{\alpha}U_n(f(0) + xn^{-(1-\alpha)/2}) \leq 1\right\}. \tag{5.3.14}$$

If Z_{n1} is the process defined in Lemma 5.3.2(i), then

$$0 \leq n^{\alpha}U_n(f(0) + xn^{-(1-\alpha)/2}) = \operatorname*{argmax}_{t\in[0,\infty)} Z_{n1}(x,t) = \mathcal{O}_p(1), \tag{5.3.15}$$

where, according to Lemma 5.3.2, the process $\{Z_{n1}(x,t) : t \in [0,\infty)\}$ converges in distribution to the process $\{W(f(0)t) - xt : t \in [0,\infty)\}$. To apply Theorem 5.2.1, we have to extend the above processes to the whole real line. Therefore define

$$\tilde{Z}_{n1}(t) = \begin{cases} Z_{n1}(x,t) & ,t \geq 0, \\ t & ,t \leq 0. \end{cases}$$

Then, for x fixed, $\tilde{Z}_{n1}$ converges in distribution to the process Z_1, where

$$Z_1(t) = \begin{cases} W(f(0)t) - xt & ,t \geq 0, \\ t & ,t \leq 0. \end{cases}$$

Moreover, since $Z_{n1}(x,0) = 0$, together with (5.3.15), it follows that

$$\operatorname*{argmax}_{t\in\mathbb{R}} \tilde{Z}_{n1}(t) = \operatorname*{argmax}_{t\in[0,\infty)} \tilde{Z}_{n1}(t) = n^\alpha U_n(f(0) + xtn^{-(1-\alpha)/2}) = \mathcal{O}_p(1).$$

The process Z_1 is continuous and since $\mathrm{Var}(Z_1(s) - Z_1(t)) \neq 0$, for $s,t > 0$ with $s \neq t$, it follows from Lemma 2.6 in Kim and Pollard (1990) that Z_1 has a unique maximum with probability one. By an application of (5.3.6) it can be seen that $Z_1(t) \to -\infty$, as $|t| \to \infty$. Theorem 5.2.1 now yields that

$$\operatorname*{argmax}_{t\in\mathbb{R}} \tilde{Z}_{n1}(t) \to \operatorname*{argmax}_{t\in\mathbb{R}} Z_1(t)$$

in distribution. Using (5.3.14), this implies that

$$\begin{aligned} P\left\{n^{(1-\alpha)/2}\left(\hat{f}_n(n^{-\alpha}) - f(0)\right) \leq x\right\} &= P\left\{\operatorname*{argmax}_{t\in\mathbb{R}} \tilde{Z}_{n1}(t) \leq 1\right\} \\ &\to P\left\{\operatorname*{argmax}_{t\in\mathbb{R}} Z_1(t) \leq 1\right\} \\ &= P\left\{\operatorname*{argmax}_{t\geq 0} \{W(f(0)t) - xt\} \leq 1\right\} \\ &= P\left\{\sqrt{f(0)\operatorname*{argmax}_{t\geq 0}\{W(t) - t\}} \leq x\right\}. \end{aligned}$$

It remains to show that

$$P\left\{n^{(1-\alpha)/2}\left(\hat{f}_n(n^{-\alpha}) - f(0)\right) \leq 0\right\} \to 0.$$

But this is evident as for any $\epsilon > 0$,

$$\begin{aligned} P\left\{n^{(1-\alpha)/2}\left(\hat{f}_n(n^{-\alpha}) - f(0)\right) \leq 0\right\} &\leq P\left\{n^{(1-\alpha)/2}\left(\hat{f}_n(n^{-\alpha}) - f(0)\right) \leq \epsilon\right\} \\ &\to P\left\{\sqrt{f(0)\mathrm{argmax}_{t\in[0,\infty)}\{W(t) - t\}} \leq \epsilon\right\}. \end{aligned}$$

When $\epsilon \downarrow 0$, the right hand side tends to zero, which can be seen from

$$P\left\{\limsup_{u\downarrow 0} \frac{W(u)}{\sqrt{2u\log\log(1/u)}} = 1\right\} = 1.$$

This proves (i).

(ii) First note that by condition (C2),

$$n^{1/3}\left(\hat{f}_n(B_{02}n^{-1/3}) - f(n^{-1/3})\right) + f'(0) = n^{1/3}\left(\hat{f}_n(B_{02}n^{-1/3}) - f(0)\right) + o(1).$$

According to (5.2.1), we have

$$P\left\{n^{1/3}\left(\hat{f}_n(B_{02}n^{-1/3}) - f(0)\right) \le x\right\} = P\left\{B_{02}^{-1}n^{1/3}U_n(f(0) + xn^{-1/3}) \le 1\right\}. \tag{5.3.16}$$

With Z_{n1} being the process defined in Lemma 5.3.2 with $\alpha = 1/3$, we get

$$B_{02}^{-1}n^{1/3}U_n(f(0) + xn^{-1/3}) = \operatorname*{argmax}_{t\in[0,\infty)}\{Z_{n1}(x, B_{02}t)\} = \mathcal{O}_p(1).$$

Again, we first extend the above process to the whole real line:

$$\tilde{Z}_{n1}(t) = \begin{cases} Z_{n1}(x, B_{02}t) & ,t \ge 0, \\ t & ,t \le 0. \end{cases}$$

Then, according to Lemma 5.3.2, $\tilde{Z}_{n1}$ converges in distribution to the process

$$Z_2(t) = \begin{cases} W(f(0)B_{02}t) - B_{02}xt + \frac{1}{2}f'(0)B_{02}^2t^2 & ,t \ge 0, \\ t & ,t \le 0. \end{cases}$$

Similar to the proof of (i), it follows from Theorem 5.2.1 that $\operatorname{argmax}_t \tilde{Z}_{n1}(t)$ converges in distribution to $\operatorname{argmax}_t Z_2(t)$, which implies that

$$P\left\{A_{02}n^{1/3}\left(\hat{f}_n(B_{02}n^{-1/3}) - f(0)\right) \le x\right\} \to \Psi(x),$$

where

$$\begin{aligned}\Psi(x) &= P\left\{\operatorname*{argmax}_{t\ge 0}\left\{W(f(0)B_{02}t) - A_{02}^{-1}B_{02}xt + \tfrac{1}{2}f'(0)B_{02}^2t^2\right\} \le 1\right\} \\ &= P\left\{\operatorname*{argmax}_{t\ge 0}\left\{W(t) - 2xt - t^2\right\} \le 1\right\} = P\left\{\operatorname*{argmax}_{t\ge 0}\left\{W(t) - (t+x)^2\right\} \le 1\right\},\end{aligned}$$

by means of Brownian scaling.

(iii) According to (5.2.1), we have

$$P\left\{n^{1/3}\left(\hat{f}_n(n^{-\alpha}) - f(n^{-\alpha})\right) \le x\right\} = P\left\{n^{1/3}\left(U_n(f(n^{-\alpha}) + xn^{-1/3}) - n^{-\alpha}\right) \le 0\right\}, \tag{5.3.17}$$

and with Z_{n2} as defined in Lemma 5.3.3(iii), we get

$$n^{1/3}\left(U_n(f(n^{-\alpha}) + xn^{-1/3}) - n^{-\alpha}\right) = \operatorname*{argmax}_{t\in[-n^{1/3-\alpha},\infty)} Z_{n2}(x,t) = \mathcal{O}_p(1).$$

As in the proof of (i) and (ii), we first extend the above process to the whole real line:

$$\tilde{Z}_{n2}(t) = \begin{cases} Z_{n2}(x,t), & ,t \ge -n^{1/3-\alpha}, \\ Z_{n2}(x, -n^{1/3-\alpha}) + (t + n^{1/3-\alpha}) & ,t < -n^{1/3-\alpha}. \end{cases}$$

Then, by Lemma 5.3.2, Z_{n2} converges in distribution to the process Z_3, where

$$Z_3(t) = W(f(0)t) - xt + \tfrac{1}{2}f'(0)t^2, \quad t \in \mathbb{R}.$$

Similar to the proof of (i) and (ii), it follows from Theorem 5.2.1 that $\text{argmax}_t\, Z_{n2}(t)$ converges in distribution to $\text{argmax}_t\, Z_3(t)$. Together with (5.3.17), this implies that

$$\begin{aligned}
&P\left\{n^{1/3}A_{03}\left(\hat{f}_n(n^{-\alpha}) - f(n^{-\alpha})\right) \le x\right\} \\
&\to P\left\{\underset{t\in\mathbb{R}}{\text{argmax}}\,\{W(f(0)t) - A_{03}^{-1}xt + \tfrac{1}{2}f'(0)t^2\} \le 0\right\} \\
&= P\left\{\underset{t\in\mathbb{R}}{\text{argmax}}\left\{W(t) - \frac{|f'(0)|}{2\sqrt{f(0)}}\left(t + \frac{x}{A_{03}|f'(0)|}\right)^2\right\} \le 0\right\} \\
&= P\left\{\underset{t\in\mathbb{R}}{\text{argmax}}\left\{W(t) - \frac{|f'(0)|}{2\sqrt{f(0)}}t^2\right\} \le \frac{x}{A_{03}|f'(0)|}\right\} \\
&= P\left\{\underset{t\in\mathbb{R}}{\text{argmax}}\,\{W(t) - t^2\} \le x\right\},
\end{aligned}$$

by means of Brownian scaling. This proves the theorem. ∎

5.4 Behavior near the end of the support

Suppose that f has compact support and, without loss of generality, assume this to be the interval $[0,1]$. In this section we investigate the behavior of $\hat{f}_n$ near one. Although there seems to be no simple symmetry argument to derive the behavior near one from the results in Section 5.3, the arguments to obtain the behavior of

$$n^{\beta}\left\{f(1-n^{-\alpha})) - \hat{f}_n(1-n^{-\alpha})\right\},$$

are similar to the ones used in studying (5.1.1). If $f(1) > 0$, then $\hat{f}_n(1)$ will always under estimate $f(1)$, since by definition $\hat{f}_n(1) = 0$. Nevertheless, the behavior near the end of the support is similar to the behavior near zero. We will assume that

(C3) $0 < f(1) = \lim_{x\uparrow 1} f(x) < \infty$.

(C4) f' is left continuous at one, such that $0 < |f'(1)| < \sup_{s\in[0,1]} |f'(s)| < \infty$.

Similar to Section 5.3 we need suitable bounds for $F(1-t) - F(1) + f(1)t$. This is guaranteed by the following lemma. The lemma is the complete analogue of Lemma 5.3.1, so that the proof is left to the reader.

Lemma 5.4.1 *Suppose that f satisfies (C4). Then there exists a value $0 < t_0 \le 1$, such that $\inf_{1-t_0\le s\le 1} |f'(s)| > 0$, and for any $1 - \frac{1}{2}t_0 \le c \le 1$,*

$$F(c-t) - F(c) + f(c)t \le \begin{cases} -\frac{1}{2}\inf_{1-t_0\le s\le 1}|f'(s)|t^2, & 0 \le t \le \frac{1}{2}t_0, \\ -\frac{1}{4}t_0\inf_{1-t_0\le s\le 1}|f'(s)|t, & t > \frac{1}{2}t_0. \end{cases}$$

The next lemma is the analogue of Lemma 5.3.2 and states that, with a suitable normalization, the processes corresponding to the argmax's in Lemma 5.4.3 converge in distribution.

Lemma 5.4.2 *Let W denote standard two-sided Brownian motion on $\mathbb{R}$. For $x \in \mathbb{R}$ and $1/3 \le \alpha < 1$ define $Y_{n1}(x,t)$ by*

$$Y_{n1}(x,t) = n^{(1+\alpha)/2}\left(F_n(1-tn^{-\alpha}) - F_n(1) + f(1)tn^{-\alpha}\right) - xt.$$

(i) For $1/3 < \alpha < 1$, the process $\{Y_{n1}(x,t) : t \in [0, n^{\alpha}]\}$ converges in distribution, in the uniform topology on compacta, to the process $\{W(f(1)t) - xt : t \in [0,\infty)\}$.

(ii) For $\alpha = 1/3$ and $x \in \mathbb{R}$, the process $\{Y_{n1}(x,t) : t \in [0, n^{1/3}]\}$ converges in distribution, in the uniform topology on compacta, to the process $\{W(f(1)t) - xt + \frac{1}{2}f'(1)t^2 : t \in [0,\infty)\}$.

(iii) For $0 < \alpha < \frac{1}{3}$ and $x \in \mathbb{R}$, define $Y_{n2}(x,t)$ by

$$Y_{n2}(x,t) = n^{2/3}\left(F_n(1-n^{-\alpha}-tn^{-1/3}) - F_n(1-n^{-\alpha}) + f(1-n^{-\alpha})tn^{-1/3}\right) - xt.$$

Then the process $\{Y_{n2}(x,t) : t \in [-n^{1/3-\alpha}, n^{1/3}(1-n^{-\alpha})]\}$ converges in distribution, in the uniform topology on compacta, to the process $\{W(f(1)t) - xt + \frac{1}{2}f'(1)t^2 : t \in \mathbb{R}\}$.

Proof: (i) Similar to the proof of Lemma 5.3.2, the process $Y_{n1}(x,t)$ can be written as

$$\begin{aligned}
&n^{\alpha/2}\left\{W_n(F(1-tn^{-\alpha})) - W_n(1)\right\}\\
&+n^{(1+\alpha)/2}\left\{F(1-tn^{-\alpha}) - 1 + f(1)tn^{-\alpha}\right\} - xt\\
&-n^{\alpha/2}\left\{F(1-tn^{-\alpha}) - 1\right\}W_n(1) + n^{\alpha/2}H_n(1-tn^{-\alpha}).
\end{aligned}$$

First note that the process $n^{\alpha/2}\left\{W_n(F(1-tn^{-\alpha})) - W_n(1)\right\}$ has the same distribution as the process $W(n^{\alpha}(1-F(1-tn^{-\alpha})))$, which can be approximated by the process $W(f(1)t)$, using uniform continuity of Brownian motion on compacta. Since $\alpha > 1/3$,

$$n^{(1+\alpha)/2}\left\{F(1-tn^{-\alpha}) - 1 + f(1)tn^{-\alpha}\right\} \to 0,$$

uniformly for t in compact sets. As in the proof Lemma 5.3.2, the remainder terms vanish, which proves (i). In case (ii), where $\alpha = 1/3$, the only difference is the behavior of the deterministic term

$$n^{2/3}\left\{F(1-tn^{-1/3}) - 1 + f(1)tn^{-1/3}\right\} \to \tfrac{1}{2}f'(1)t^2,$$

uniformly for t in compact sets. Similar to the proof of (i), using Brownian scaling and uniform continuity of Brownian motion on compacta this proves (ii). For case (iii), the process Y_{n2} can be written as

$$\begin{aligned}
&n^{1/6}\left\{W_n(F(1-n^{-\alpha}-tn^{-1/3})) - W_n(F(1-n^{-\alpha}))\right\}\\
&+n^{2/3}\left\{F(1-n^{-\alpha}-tn^{-1/3}) - F(1-n^{-\alpha}) + f(1-n^{-\alpha})tn^{-1/3}\right\} - xt\\
&-n^{1/6}\left\{F(1-n^{-\alpha}-tn^{-1/3}) - F(1-n^{-\alpha})\right\}W_n(1)\\
&+n^{1/6}H_n(1-n^{-\alpha}-tn^{-1/3}) - n^{1/6}H_n(1-n^{-\alpha}).
\end{aligned}$$

The process $n^{1/6}\left\{W_n(F(1-n^{-\alpha}-tn^{-1/3})) - W_n(F(1-n^{-\alpha}))\right\}$ has the same distribution as the process $W(n^{1/3}(F(1-n^{-\alpha}) - F(1-n^{-\alpha}-tn^{-1/3})))$, which can be approximated by the process $W(f(1)t)$, again by using uniform continuity of Brownian motion on compacta. Because $\alpha < 1/3$,

$$n^{2/3}\left\{F(1-n^{-\alpha}-tn^{-1/3})-F(1-n^{-\alpha})+f(1-n^{-\alpha})tn^{-1/3}\right\} \to \tfrac{1}{2}f'(1)t^2,$$

uniformly for t in compact sets. As before, the remainder terms vanish, which proves (iii). ■

Lemma 5.4.3 *Let $x \in I\!R$ and Y_{n1} and Y_{n2}, as defined in Lemma 5.4.2.*

(i) For $1/3 < \alpha < 1$ and $x > 0$, $\operatorname{argmax}_{t\in[0,n^\alpha]} Y_{n1}(x,t) = O_p(1)$.

(ii) For $\alpha = 1/3$, $\operatorname{argmax}_{t\in[0,n^{1/3}]} Y_{n1}(x,t) = O_p(1)$.

(iii) For $0 < \alpha < 1/3$, $\operatorname{argmax}_{t\in[-n^{1/3-\alpha},n^{1/3}(1-n^{-\alpha})]} Y_{n2}(x,t) = O_p(1)$.

Proof: The proof mimics the proof of Lemma 5.3.3. In case (i), first note that $\operatorname{argmax}_t Y_{n1}(x,t)$ has the same distribution as $\operatorname{argmax}_t N_{n1}(t)$, where

$$\begin{aligned} N_{n1}(t) &= W_n\left(n^\alpha(F(1-tn^{-\alpha})-1)\right)+n^{(1+\alpha)/2}\left\{F(1-tn^{-\alpha})-1+f(1)tn^{-\alpha}\right\}-xt \\ &\quad -n^{\alpha/2}\left\{F(1-tn^{-\alpha})-1\right\}W_n(1)+n^{\alpha/2}H_n(1-tn^{-\alpha}). \end{aligned}$$

Let $0 < \epsilon < x$. Define processes

$$\begin{aligned} X_{n1}(t) &= n^{\alpha/2}H_n(1-tn^{-\alpha})-\frac{1}{2}\epsilon t, \\ X_{n2}(t) &= -n^{\alpha/2}\left\{F(1-tn^{-\alpha})-1\right\}W_n(1)-\frac{1}{2}\epsilon t, \end{aligned}$$

and define the event A_n as in the proof of Lemma 5.3.3(i). It follows that $(\operatorname{argmax}_t N_{n1}(t))1_{A_n^c} = O_p(1)$, so that we only have to deal with $(\operatorname{argmax}_t N_{n1}(t))1_{A_n}$. Proceeding as in the proof of Lemma 5.3.3(i), using that the function $F(1-tn^{-\alpha})-1+f(1)tn^{-\alpha}$ is non-increasing, we find that

$$0 \le \left(\operatorname*{argmax}_{t\in[0,n^\alpha]} N_{n1}(t)\right)1_{A_n} \le \operatorname*{argmax}_{t\in[0,n^\alpha]}\left\{W_n\left(n^\alpha(F(1-tn^{-\alpha})-1)\right)-(x-\epsilon)t\right\}+\delta_n,$$

where $\delta_n = n^{-(1-\alpha)/2}(\log n)^2$. Finally, by change of variables $u = H(t) = n^\alpha(1-F(1-tn^{-\alpha}))$, and the fact that for any $u \in [0,n^\alpha]$,

$$\frac{u}{f(0)} \le H^{-1}(u) \le \frac{u}{f(1)}, \tag{5.4.1}$$

we find that

$$\begin{aligned} &\operatorname*{argmax}_{t\in[0,n^\alpha]}\left\{W_n\left(n^\alpha(F(1-tn^{-\alpha})-1)\right)-(x-\epsilon)t\right\} \\ &\le \sup\left\{t\in[0,n^\alpha]: W_n\left(n^\alpha(F(1-tn^{-\alpha})-1)\right)-(x-\epsilon)t \ge 0\right\} \\ &\le \frac{1}{f(1)}\sup\left\{u\in[0,\infty): W(-u)-\frac{x-\epsilon}{f(0)}u \ge 0\right\} = O_p(1), \end{aligned}$$

which proves (i).

(ii) As in the proof of Lemma 5.3.3(ii), similar to (5.3.7) we obtain

$$\left(\operatorname*{argmax}_{t\in[0,n^{1/3}]} N_{n1}(t)\right) 1_{A_n} \leq \operatorname*{argmax}_{t\in[0,n^{1/3}]} R_{n1}(t) + \delta_n \leq \sup\{t \geq 0 : R_{n1}(t) \geq 0\} + \delta_n,$$

where

$$R_{n1}(t) = W_n\left(n^{1/3}(F(1-tn^{-1/3}) - 1)\right) - (x-\epsilon)t + n^{2/3}\left(F(1-tn^{-1/3}) - 1 + f(1)tn^{-1/3}\right).$$

As in the proof of Lemma 5.3.3(ii), restrict to $D_{n1} = \left\{n^{-1/3}\sup\{t \geq 0 : R_{n1}(t) \geq 0\} \leq \frac{1}{2}t_0\right\}$, for which $P(D_{n1}^c) \to 0$. Then by application of Lemma 5.3.1, similar to (5.3.8) we find that $(\operatorname{argmax}_t N_{n1}(t))\, 1_{A_n \cap D_{n1}}$ is bounded from above by

$$\sup\left\{t \in [0,\infty) : W_n\left(n^{1/3}(F(1-tn^{-1/3}) - 1)\right) - (x-\epsilon)t - \tfrac{1}{2}t^2 \inf|f'| \geq 0\right\} + \delta_n.$$

Proceeding as in the proof of Lemma 5.3.3(ii), similar to (5.3.9) this supremum is bounded by

$$\sup\left\{t \in [0,\infty) : W_n\left(n^{1/3}(F(1-tn^{-1/3}) - 1)\right) - \tfrac{1}{4}t^2 \inf|f'| \geq 0\right\} + \mathcal{O}_p(1).$$

By change of variables $u = H(t) = n^{1/3}(1 - F(1-tn^{-1/3}))$, and using (5.4.1), we find that this argmax is bounded by

$$\frac{1}{f(1)}\sup\left\{u \in [0,\infty) : W(u) - \frac{\inf|f'|}{4f(0)^2}u^2 \geq 0\right\} = \mathcal{O}_p(1),$$

which proves (ii).

For case (iii), first note that $\operatorname{argmax}_t Y_{n2}(x,t)$ has the same distribution as $\operatorname{argmax}_t N_{n2}(t)$, where

$$\begin{aligned} N_{n2}(t) &= W_n\left(n^{1/3}(F(1-n^{-\alpha}-tn^{-1/3}) - F(1-n^{-\alpha}))\right) \\ &\quad + n^{2/3}\left(F(1-n^{-\alpha}-tn^{-1/3}) - F(1-n^{-\alpha}) + f(1-n^{-\alpha})tn^{-1/3}\right) - xt \\ &\quad - n^{1/6}F(1-n^{-\alpha}-tn^{-1/3})W_n(1) + n^{1/6}H_n(1-n^{-\alpha}-tn^{-1/3}). \end{aligned}$$

Let $\epsilon > 0$ and let A_n be same event as in the proof of (i) and (ii) with $\delta_n = n^{-1/3}(\log n)^2$. Write $I_n = [-n^{1/3-\alpha}, n^{1/3}(1-n^{-\alpha})]$, then by the same argument as in the proof of (i) and (ii), we find that

$$\left(\operatorname*{argmax}_{t\in I_n} N_{n2}(t)\right) 1_{A_n} \leq \operatorname*{argmax}_{t\in I_n} R_{n2}(t) + \delta_n \leq \sup\{t \geq 0 : R_{n2}(t) \geq 0\} + \delta_n,$$

where

$$\begin{aligned} R_{n2}(t) &= W_n\left(n^{1/3}(F(1-n^{-\alpha}-tn^{-1/3}) - F(1-n^{-\alpha}))\right) \\ &\quad + n^{2/3}\left(F(1-n^{-\alpha}-tn^{-1/3}) - F(1-n^{-\alpha}) + f(1-n^{-\alpha})tn^{-1/3}\right) - (x-\epsilon)t. \end{aligned}$$

According to Lemma 5.4.1,

$$n^{2/3}\left(F(1-n^{-\alpha}-tn^{-1/3}) - F(1-n^{-\alpha}) + f(1-n^{-\alpha})tn^{-1/3}\right) \leq -\frac{1}{2}t^2 \inf|f'|,$$

for $0 \leq tn^{-1/3} \leq \frac{1}{2}t_0$, so that $\sup\{0 \leq tn^{-1/3} \leq \frac{1}{2}t_0 : R_{n2}(t) \geq 0\}$ is bounded from above by

$$\sup\left\{t \in [0,\infty) : W_n\left(n^{1/3}(F(1-n^{-\alpha}-tn^{-1/3}) - F(1-n^{-\alpha}))\right) - \tfrac{1}{4}t^2 \inf|f'| \geq 0\right\}.$$

Then by change of variables $u = H(t) = n^{1/3}\big(F(1-n^{-\alpha}) - F(1-n^{-\alpha}-tn^{-1/3})\big)$, and using that for any $u \in [0, n^{1/3}F(1-n^{-\alpha})]$,

$$\frac{u}{f(0)} \leq H^{-1}(u) \leq \frac{u}{f(1)},$$

it follows that this argmax is bounded by

$$\frac{1}{f(1)} \sup\left\{u \in [0,\infty) : W(u) - \frac{\inf|f'|}{4f(0)^2}u^2 \geq 0\right\} = \mathcal{O}_p(1).$$

The lower bound for $\operatorname{argmax}_t N_{n2}(t)$ is obtained by the same type of argument as for the lower bound in the proof of Lemma 5.3.3(iii). This proves the lemma. ∎

We are now able to determine the behavior of the Grenander estimator at the end of the support. Similar to Theorem 5.3.1, in the case $\alpha = 1/3$, we are only able to give the distribution function of the limiting distribution of $\hat{f}_n(1-n^{-\alpha})$.

Theorem 5.4.1 *Let f satisfy conditions (C3)-(C4). Then*

(i) For $1/3 < \alpha < 1$ and $A_{11} = f(1)^{-1/2}$, we have that

$$A_{11}n^{(1-\alpha)/2}\left(f(1-n^{-\alpha}) - \hat{f}_n(1-n^{-\alpha})\right) \to \sqrt{\operatorname*{argmax}_{t\in[0,\infty)}\{W(t)-t\}}$$

in distribution, as $n \to \infty$.

(ii) For $A_{12} = |4f(1)f'(1)|^{-1/3}$ and $B_{12} = 4^{1/3}f(1)^{1/3}|f'(1)|^{-2/3}$, we have that

$$A_{12}\left\{n^{1/3}\left(f(1-n^{-1/3}) - \hat{f}_n(B_{12}(1-n^{-1/3}))\right) + f'(1)\right\} \to V$$

in distribution, as $n \to \infty$, where V has distribution function

$$\Psi(x) = P\left\{\operatorname*{argmax}_{t\geq 0}\{W(t)-(t+x)^2\} \leq 1\right\}.$$

(iii) For $0 < \alpha < 1/3$ and $A_{13} = |4f(1)f'(1)|^{-1/3}$, we have that

$$A_{13}n^{1/3}\left(f(1-n^{-\alpha}) - \hat{f}_n(1-n^{-\alpha})\right) \to \operatorname*{argmax}_{t\in\mathbb{R}}\{W(t)-t^2\}$$

in distribution, as $n \to \infty$.

Proof: To prove case (i), similar to the proof of Theorem 5.3.1(i), it suffices to consider $n^{(1-\alpha)/2}(f(1) - \hat{f}_n(1 - n^{-\alpha}))$. For $x > 0$, according to (5.2.1), we have

$$P\left\{n^{(1-\alpha)/2}\left(f(1) - \hat{f}_n(1 - n^{-\alpha})\right) \leq x\right\} = P\left\{n^{\alpha}\left(1 - U_n(f(1) - xn^{-(1-\alpha)/2})\right) \leq 1\right\},$$

where according to Lemma 5.4.3(i),

$$n^{\alpha}\left(1 - U_n(f(1) - xn^{-(1-\alpha)/2})\right) = \underset{t\in[0,n^{\alpha}]}{\operatorname{argmax}} Y_{n1}(x,t) = \mathcal{O}_p(1).$$

From here on, the proof proceeds in completely the same manner as that of Theorem 5.3.1(i). We conclude that for $x > 0$,

$$\begin{aligned} P\left\{n^{(1-\alpha)/2}\left(f(1) - \hat{f}_n(1 - n^{-\alpha})\right) \leq x\right\} &\to P\left\{\underset{t\geq 0}{\operatorname{argmax}}\{W(f(1)t) - xt\} \leq 1\right\} \\ &= P\left\{\sqrt{f(1)\underset{t\geq 0}{\operatorname{argmax}}\{W(t) - t\}} \leq x\right\}. \end{aligned}$$

Furthermore, similar to the proof of Theorem 5.3.1(i) it follows that

$$P\left\{n^{(1-\alpha)/2}\left(f(1) - \hat{f}_n(1 - n^{-\alpha})\right) \leq 0\right\} \to 0.$$

This proves (i). For (ii), first note that

$$\begin{aligned} &n^{1/3}\left(f(1 - n^{-1/3}) - \hat{f}_n\left(B_{12}(1 - n^{-1/3})\right)\right) + f'(1) \\ &\qquad = n^{1/3}\left(f(1) - \hat{f}_n\left(B_{12}(1 - n^{-1/3})\right)\right) + o(1). \end{aligned}$$

According to (5.2.1), we have

$$P\left\{n^{1/3}\left(f(1) - \hat{f}_n\left(B_{12}(1 - n^{-1/3})\right)\right) \leq x\right\} = P\left\{B_{12}^{-1}n^{1/3}\left(1 - U_n(f(1) - xn^{-1/3})\right) \leq 1\right\},$$

where, according to Lemma 5.4.3(ii),

$$B_{12}^{-1}n^{1/3}\left(1 - U_n(f(1) - xn^{-1/3})\right) = \underset{t\in[0,n^{1/3}]}{\operatorname{argmax}} Y_{n1}(x, B_{12}t) = \mathcal{O}_p(1).$$

The rest of the proof is completely similar to that of Theorem 5.3.1(ii). For (iii), note that according to (5.2.1), $P\left\{n^{1/3}\left(f(1 - n^{-\alpha}) - \hat{f}_n(1 - n^{-\alpha})\right) \leq x\right\}$ is equal to

$$P\left\{n^{\alpha}\left(1 - n^{-\alpha} - U_n(f(1 - n^{-\alpha}) - xn^{-1/3})\right) \leq 0\right\},$$

where, according to Lemma 5.4.3(iii),

$$n^{\alpha}\left(1 - n^{-\alpha} - U_n(f(1 - n^{-\alpha}) - xn^{-1/3})\right) = \underset{t\in[0,n^{1/3}]}{\operatorname{argmax}} Y_{n2}(x,t) = O_p(1).$$

From here the proof is completely similar to that of Theorem 5.3.1(iii). ■

5.5 A comparison with the penalized NPMLE

Consider a decreasing density f on $[0,\infty)$. As pointed out in WOODROOFE AND SUN (1993), the NPMLE $\hat{f}_n$ for f is not consistent at zero. They proposed a penalized NPMLE $\hat{f}_n^{\mathrm{P}}(\alpha,0)$, and in SUN AND WOODROOFE (1996) it was shown that

$$n^{1/3}\left\{\hat{f}_n^{\mathrm{P}}(\alpha_n,0)-f(0)\right\}\to\sup_{t>0}\frac{W(t)-(c-\frac{1}{2}f(0)f'(0)t^2)}{t},$$

where c is related to the smoothing parameter $\alpha_n=cn^{-2/3}$. SUN AND WOODROOFE (1996)also provide (to some extent) an adaptive choice for c that leads to an estimate $\hat{\alpha}_n$ of the smoothing parameter, and report some results of a simulation experiment for $\hat{f}_n^{\mathrm{P}}(\hat{\alpha}_n,0)$.

We propose $\hat{f}_n(n^{-1/3})$ as an estimate for $f(0)$. This estimator is straightforward and does not have any additional smoothing parameters. As a consequence of Theorem 5.3.1, this estimator is consistent for $f(0)$ with rate $n^{1/3}$, and has a limiting distribution that is some complicated functional of W:

$$n^{1/3}\left\{\hat{f}_n(n^{-1/3})-f(0)\right\}\to S_0,$$

where S_0 has distribution function

$$F_0(s)=P\left\{\operatorname*{argmax}_{t\geq 0}\left\{W(f(0)t)-st+\tfrac{1}{2}f'(0)t^2\right\}\leq 1\right\}.$$

Similarly, in the case of a density f with compact support $[0,1]$, we propose $\hat{f}_n(1-n^{-1/3})$ as an estimate for $f(1)$. As a consequence of Theorem 5.4.1,

$$n^{1/3}\left\{f(1)-\hat{f}_n(1-n^{-1/3})\right\}\to S_1,$$

where S_1 has distribution function

$$F_1(s)=P\left\{\operatorname*{argmax}_{t\geq 0}\left\{W(f(1)t)-st+\tfrac{1}{2}f'(1)t^2\right\}\leq 1\right\}.$$

In this section we will compare the estimator $\hat{f}_n(n^{-1/3})$ for $f(0)$ with the penalized NPMLE by means of a simulation study.

We simulated 10000 samples of sizes $n=50,100,200$, and 10000 from a standard exponential distribution with mean one. For each sample the values of $n^{1/3}\{\hat{f}_n(n^{-1/3})-f(0)\}$ and $n^{1/3}\{\hat{f}_n^{\mathrm{P}}(\hat{\alpha}_n,0)-f(0)\}$ were computed. The value of $\hat{\alpha}_n$ was computed as proposed in SUN AND WOODROOFE (1996),

$$\hat{\alpha}_n=0.649\cdot\hat{\beta}_n^{-1/3}n^{-2/3},$$

where

$$\hat{\beta}_n=\max\left\{\hat{f}_n^{\mathrm{P}}(\alpha_0,0)\frac{\hat{f}_n^{\mathrm{P}}(\alpha_0,0)-\hat{f}_n^{\mathrm{P}}(\alpha_0,x_m)}{2x_m},n^{-q}\right\},$$

is an estimate for $\beta=-\frac{1}{2}f(0)f'(0)$. Here, x_m denotes the second point of jump of the penalized NPMLE $\hat{f}_n^{\mathrm{P}}(\alpha_0,\cdot)$ computed with smoothing parameter α_0. The parameter $\alpha_0=c_0n^{-2/3}$, and q should be taken between 0 and 0.5. However, SUN AND WOODROOFE (1996) do not specify how to choose q and c_0 in general. We took $q=1/3$, and for α_0 the values as listed in their

Table 2: $\alpha_0 = 0.0526, 0.0325$ and 0.0205 for sample sizes $n = 50, 100$ and 200. For sample size $n = 10000$ we took the theoretical optimal value $\alpha_0 = 0.649\,\beta^{-1/3}n^{-2/3}$, with $\beta = 0.5$.

In Table 5.1 we listed simulated values for the mean, variance and mean squared error of both estimators. The penalized NPMLE is less biased, but has a larger variance. Estimator

n	$n^{1/3}\{\hat{f}_n(n^{-1/3}) - f(0)\}$			$n^{1/3}\{\hat{f}^{\mathrm{P}}_n(\hat{\alpha}_n, 0) - f(0)\}$		
	Mean	Variance	MSE	Mean	Variance	MSE
50	−0.8471	0.4392	1.1569	−0.0721	1.2955	1.3007
100	−0.8531	0.4835	1.2114	−0.0785	1.5304	1.5366
200	−0.8677	0.5363	1.2893	−0.0747	1.7319	1.7375
10000	−0.9169	0.7003	1.5410	−0.1950	1.9130	1.9510

Table 5.1: Simulated mean, variances and mean squared error for both estimators.

$\hat{f}_n(n^{-1/3})$ performs better in the sense of mean squared error. We complete our comparison by displaying the densities of both estimators. In Figure 5.1, kernel estimates are plotted of the 10000 simulated values of $n^{1/3}\{\hat{f}_n(n^{-1/3}) - f(0)\}$ (solid line) and $n^{1/3}\{\hat{f}^{\mathrm{P}}_n(\hat{\alpha}_n, 0) - f(0)\}$ (dotted line) for samples sizes $n = 50, 100, 200$ and 10000.

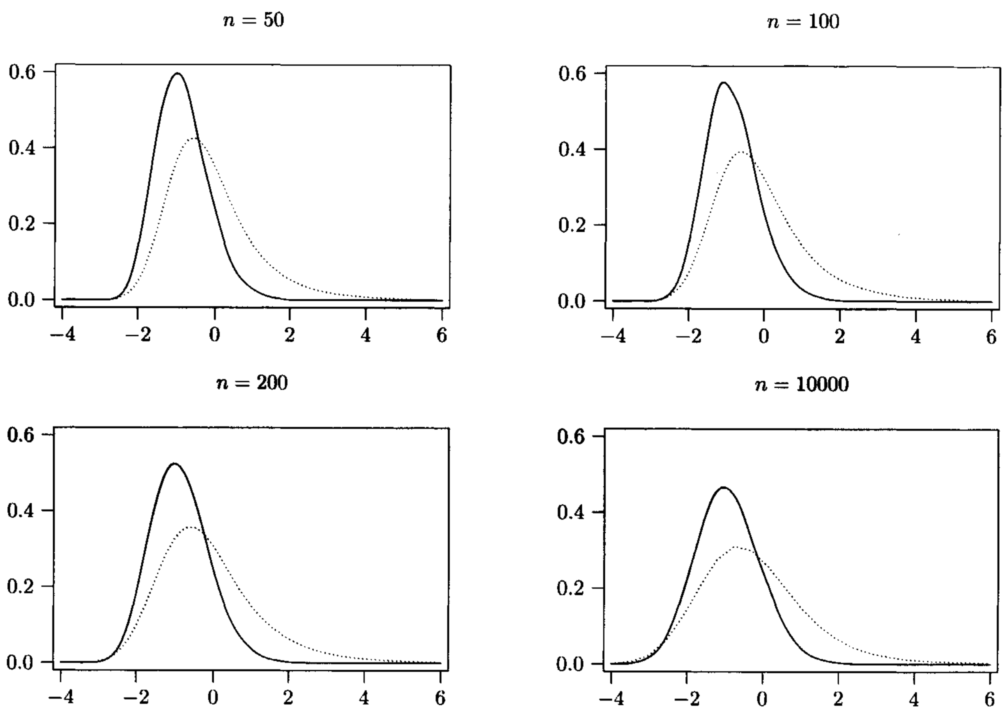


Figure 5.1: Simulated densities of both estimators for sample sizes $n = 50, 100, 200$ and 10000.

Chapter 6

The L_k-error of the Grenander estimator

Asymptotic normality of L_1-distance between a decreasing density f and its nonparametric maximum likelihood estimator $\hat{f}_n$ is based on interpreting this distance as the area between the graphs of f and $\hat{f}_n$. Since this area is also the area between the graphs of the inverse g of f and the more tractable inverse U_n of $\hat{f}_n$, the problem can be reduced to deriving asymptotic normality of the L_1-distance between U_n and g. In this chapter we investigate the limit behavior of the L_k-distance between f and $\hat{f}_n$ for $k \geq 1$. Due to the inconsistency of $\hat{f}_n$ at zero, the case $k = 2.5$ turns out to be some kind of transition point. We extent asymptotic normality of the L_1-distance to the L_k-distance for $1 \leq k < 2.5$, and obtain the analogous limiting result for a modification of the L_k-distance for $k \geq 2.5$.

6.1 Introduction

Let f be a non-increasing density with compact support. Without loss of generality, assume this to be the interval $[0, 1]$. The non-parametric maximum likelihood estimator $\hat{f}_n$ for f has been discovered by GRENANDER (1956). It is defined as the left derivative of the concave majorant of the empirical distribution function F_n constructed from a sample $X_1, \dots, X_n$ from f. PRAKASA RAO (1969) obtained the earliest result on the asymptotic pointwise behavior of the Grenander estimator. One immediately striking feature of this result is that the rate of convergence is of the same order as the rate of convergence of histogram estimators, and that the asymptotic distribution is not normal. It took much longer to develop distributional theory for global measures of performance for this estimator. The first distributional result for a global measure of deviation was the convergence to a normal distribution of the L_1-error mentioned in GROENEBOOM (1985) (see GROENEBOOM, HOOGHIEMSTRA AND LOPUHAÄ (1999) for a rigorous proof). A similar result in the regression setting has been obtained by DUROT (2000).

In this paper we will extend the result for the L_1-distance to the L_k-distance, for $k \geq 1$. We will follow the same approach as in GROENEBOOM, HOOGHIEMSTRA AND LOPUHAÄ (1999), where, instead of comparing $\hat{f}_n$ to f, the more tractable inverse process U_n was compared to the inverse g of f, where for $a \in [f(1), f(0)]$:

$$U_n(a) = \underset{x \in [0,1]}{\operatorname{argmax}} \{F_n(x) - ax\}. \qquad (6.1.1)$$

Asymptotic normality of the L_1-error

$$\|U_n - g\|_1 = \int_{f(1)}^{f(0)} |U_n(a) - g(a)|\, da$$

was obtained by approximating $F_n(t) - at$ by a Gaussian process and approximating U_n by the argmax of this process. This immediately yields asymptotic normality of L_1-error

$$\|\hat{f}_n - f\|_1 = \int_0^1 |\hat{f}_n(x) - f(x)|\, dx$$

between $\hat{f}_n$ and f, since $\|\hat{f}_n - f\|_1$ represents the area between the graphs of $\hat{f}_n$ and f, which is also the area between the graphs of U_n and g. Clearly, for $k > 1$ we no longer have such an easy correspondence between the two L_k-errors. Nevertheless, we will show that the L_k-distance between $\hat{f}_n$ and f can still be approximated by a scaled version of the L_k-distance between U_n and g, and that this scaled version is asymptotically normal.

Another important difference between the case $k > 1$ and the case $k = 1$, is the fact that for large k, the inconsistency of $\hat{f}_n$ at zero, as shown by Woodroofe and Sun (1993), starts to dominate the behavior of the L_k-distance. By using results from Kulikov and Lopuhaä (2002) on the behavior of $\hat{f}_n$ near the boundaries of the support of f, we will show that for $1 \leq k < 2.5$ the L_k-distance between $\hat{f}_n$ and f is asymptotically normal. This result can be formulated as follows. Define for $c \in \mathbb{R}$,

$$V(c) = \sup\{t : W(t) - (t-c)^2 \text{ is maximal}\}, \tag{6.1.2}$$

with $\{W(t) : -\infty < t < \infty\}$ denoting standard two-sided Brownian motion on $\mathbb{R}$ originating from zero (i.e. $W(0) = 0$).

Theorem 6.1.1 (Main theorem). *Let f be a twice differentiable decreasing density on $[0,1]$, satisfying:*

(A1) $0 < f(1) \leq f(y) \leq f(x) \leq f(0) < \infty$, *for* $0 \leq x \leq y \leq 1$;

(A2) $0 < \inf_{x\in(0,1)} |f'(x)| \leq \sup_{x\in(0,1)} |f'(x)| < \infty$;

(A3) $\sup_{x\in(0,1)} |f''(x)| < \infty$.

Then for $1 \leq k < 2.5$, with $\mu_k = E|V(0)| \left\{\int_0^1 (4f(x)|f'(x)|)^{k/3}\, dx\right\}^{1/k}$,

$$n^{1/6}\left\{n^{1/3}\left(\int_0^1 |\hat{f}_n(x) - f(x)|\, dx\right)^{1/k} - \mu_k\right\}$$

converges in distribution to a normal random variable with zero mean and variance

$$\frac{\int_0^1 f(x)^{(2k+1)/3}|f'(x)|^{(2k-2)/3}\, dx}{k^2\left(E|V(0)|^k \int_0^1 (f(x)|f'(x)|)^{k/3}\, dx\right)^{(2k-2)/k}} \cdot 8\int_0^\infty \mathrm{cov}(|\xi(0)|^k, |\xi(c)|^k)dc.$$

Note that the theorem holds under the same conditions as in GROENEBOOM ET AL. (1999). For $k \geq 2.5$, Theorem 6.1.1 is no longer true. However, the results from KULIKOV AND LOPUHAÄ (2002) enable us to show that an analogous limiting result still holds for a modification of the L_k-error.

In Section 6.2 we introduce a Brownian approximation of U_n and prove asymptotic normality of a scaled version of the L_k-error between U_n and g. The proof relies heavily on results obtained in GROENEBOOM ET AL. (1999). In Section 6.3 we show that on segments $[s,t]$, where the graph of $\hat{f}_n$ does not cross the graph of f, the difference

$$\left| \int_s^t |\hat{f}_n(x) - f(x)|^k \, dx - \int_{f(t)}^{f(s)} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}} \, da \right|$$

is of negligible order. Together with the behavior near the boundaries of the support of f, we establish asymptotic normality of the L_k-error for $1 \leq k < 2.5$ in Section 6.4. In Section 6.5 we investigate the case $k > 2.5$, and prove a result analogous to Theorem 6.1.1 for a modified L_k-error.

6.2 Brownian approximation

In this section we will prove the asymptotic normality of the L_k-error of the inverse process of the Grenander estimator. The proof will follow the same line of reasoning as Sections 3 and 4 in GROENEBOOM ET AL.(1999).

Let E_n denote the empirical process $\sqrt{n}(F_n - F)$. For $n \geq 1$, let B_n be versions of the Brownian bridge constructed on the same probability space as the uniform empirical process $E_n \circ F^{-1}$ via the Hungarian embedding, and define versions W_n of Brownian motion by

$$W_n(t) = B_n(t) + \xi_n t, \quad t \in [0,1], \tag{6.2.1}$$

where ξ_n is a standard normal random variable, independent of B_n. Then for fixed $a \in (f(1), f(0))$ and $J = E, B, W$ define

$$V_n^J(a) = \underset{t}{\operatorname{argmax}} \left\{ X_n^J(a,t) + n^{2/3} \left[F(g(a) + n^{-1/3}t) - F(g(a)) - n^{-1/3}at \right] \right\}, \tag{6.2.2}$$

where

$$\begin{aligned} X_n^E(a,t) &= n^{1/6} \left\{ E_n(g(a) + n^{-1/3}t) - E_n(g(a)) \right\}, \\ X_n^B(a,t) &= n^{1/6} \left\{ B_n(F(g(a) + n^{-1/3}t)) - B_n(F(g(a))) \right\}, \\ X_n^W(a,t) &= n^{1/6} \left\{ W_n(F(g(a) + n^{-1/3}t)) - W_n(F(g(a))) \right\}. \end{aligned} \tag{6.2.3}$$

One can easily check that $V_n^E(a) = n^{1/3}\{U_n(a) - g(a)\}$, where $U_n(a)$ is defined in (6.1.1). The random variable $V_n^B(a)$ is a Brownian *bridge* approximation of $V_n^E(a)$ based on the Hungarian embedding, whereas $V_n^W(a)$ can be seen as a Brownian *motion* approximation. A graphical interpretation and basic properties of V_n^J can be found in GROENEBOOM ET AL.(1999). Since we will use these frequently, we state them for easy reference.

First, the tail probabilities of V_n^J have a uniform exponential upper bound.

Lemma 6.2.1 *For $J = E, B, W$, let V_n^J be defined by (6.2.2). Then there exist constants $C_1, C_2 > 0$ only depending on f, such that for all $n \geq 1$, $a \in (f(1), f(0))$ and $x > 0$,*

$$P\left\{|V_n^J(a)| \geq x\right\} \leq C_1 \exp(-C_2 x^3).$$

Properly normalized versions of $V_n^J(a)$ converge in distribution to

$$\xi(c) = V(c) - c, \tag{6.2.4}$$

where $V(c)$ is defined in (6.1.2). To be more precise, for $a \in (f(1), f(0))$, let

$$J_n(a) = \left\{c : a - \phi_2(a)cn^{-1/3} \in (f(1), f(0))\right\},$$

and for $J = E, B, W$ and $c \in J_n(a)$, define,

$$V_{n,a}^J(c) = \phi_1(a)V_n^J(a - \phi_2(a)cn^{-1/3}), \tag{6.2.5}$$

where

$$\phi_1(a) = \frac{|f'(g(a))|^{2/3}}{(4a)^{1/3}} > 0, \tag{6.2.6}$$

$$\phi_2(a) = (4a)^{1/3}|f'(g(a))|^{1/3} > 0. \tag{6.2.7}$$

Then we have the following property.

Lemma 6.2.2 *For $J = E, B, W$, integer $d \geq 1$, $a \in (f(1), f(0))$ and $c \in J_n(a)^d$, we have joint distributional convergence of $(V_{n,a}^J(c_1), \ldots, V_{n,a}^J(c_d))$ to the random vector $(\xi(c_1), \ldots, \xi(c_d))$.*

Due to the fact that Brownian motion has independent increments, the process V_n^W is mixing.

Lemma 6.2.3 *The process $\{V_n^W(a)) : a \in (f(1), f(0))\}$ is strong mixing with mixing function: $\alpha_n(d) = 12e^{-C_3 n d^3}$, where the constant $C_3 > 0$ only depends on f.*

As a direct consequence we have the following lemma, which is a slight extension of Lemma 4.1 in GROENEBOOM ET AL.(1990).

Lemma 6.2.4 *Let l and m be fixed such that $l + m > 0$ and let h be a continuous function. Define*

$$c_h = 2\int_0^1 (4f(x))^{\frac{2(l+m)+1}{3}} |f'(x)|^{\frac{4-4(l+m)}{3}} h(f(x))^2 \, dx.$$

Then,

$$\mathrm{var}\left(n^{1/6}\int_{f(1)}^{f(0)} V_n^W(a)^l |V_n^W(a)|^m h(a)\, da\right) \to c_h \int_0^\infty \mathrm{cov}(\xi(0)^l|\xi(0)|^m, \xi(c)^l|\xi(c)|^m)\, dc,$$

as $n \to \infty$.

Proof: The proof runs along the lines of the proof of Lemma 4.1 in GROENEBOOM ET AL.(1999). We first have that

$$\operatorname{var}\left(n^{1/6}\int_{f(1)}^{f(0)} V_n^W(a)^l|V_n^W(a)|^m h(a)\,da\right)$$
$$= -2\int_{f(1)}^{f(0)}\int_0^{n^{1/3}\phi_2(a)^{-1}(a-f(0))}(4a)^{\frac{2(l+m)+1}{3}}|g'(a)|^{\frac{4(l+m)-1}{3}}h(a)h(a-\phi_2(a)n^{-1/3}c)$$
$$\cdot\operatorname{cov}\left(V_{n,a}^W(0)^l|V_{n,a}^W(0)|^m, V_{n,a}^W(c)^l|V_{n,a}^W(c)|^m\right)\,dc\,da.$$

According to Lemma 6.2.1, for a and c fixed, the sequence $V_{n,a}^W(c)^l|V_{n,a}^W(c)|^m$ is uniformly integrable. Hence by Lemma 6.2.2 the moments of $\left(V_{n,a}^W(0)^l|V_{n,a}^W(0)|^m, V_{n,a}^W(c)^l|V_{n,a}^W(c)|^m\right)$ converge to corresponding moments of $(\xi(0)^l|\xi(0)|^m, \xi(c)^l|\xi(c)|^m)$. Again Lemma 6.2.1 and the fact that $l+m>0$, yields that

$$E|V_{n,a}^W(0)|^{3(l+m)} < C \quad \text{and} \quad E|V_{n,a}^W(c)|^{3(l+m)} < C,$$

where $C>0$ does not depend on n, a and c. Together with Lemma 6.2.3 and Lemma 3.2 in GROENEBOOM ET AL.(1999) this yields that

$$\left|\operatorname{cov}\left(V_{n,a}^W(0)^l|V_{n,a}^W(0)|^m, V_{n,a}^W(0)^l|V_{n,a}^W(c)|^m\right)\right| \le D_1 e^{-D_2|c|^3},$$

where D_1 and D_2 do not depend on n, a and c. It follows by dominated convergence that

$$\operatorname{var}\left(n^{1/6}\int_{f(1)}^{f(0)} V_{n,a}^W(0)^l|V_n^W(a)|^m h(a)\,da\right)\to$$
$$-c_h\int_0^{-\infty}\operatorname{cov}\left(\xi(0)^l\,|\xi(0)|^m, \xi(c)^l\,|\xi(c)|^m\right)\,dc$$
$$= c_h\int_0^{\infty}\operatorname{cov}\left(\xi(0)^l\,|\xi(0)|^m, \xi(c)^l\,|\xi(c)|^m\right)\,dc,$$

using that the process ξ is stationary, where

$$\begin{aligned} c_h &= 2\int_{f(1)}^{f(0)}(4a)^{\frac{2(l+m)+1}{3}}|g'(a)|^{\frac{4(l+m)-1}{3}}h(a)^2\,da \\ &= 2\int_0^1(4f(x))^{\frac{2(l+m)+1}{3}}|f'(x)|^{\frac{4-4(l+m)}{3}}h(f(x))^2\,dx. \end{aligned}$$

This proves the lemma. ■

We are now able to prove asymptotic normality for a Brownian version of the L_k-error between U_n and g.

Theorem 6.2.1 *Let V_n^W be defined as in (6.2.2) and ξ by (6.2.4). Then*

$$n^{1/6}\int_{f(1)}^{f(0)}\frac{|V_n^W(a)|^k - E|V_n^W(a)|^k}{|g'(a)|^{k-1}}\,da$$

converges in distribution to a normal random variable with zero mean and variance

$$\sigma^2 = 2\int_0^1(4f(x))^{\frac{2k+1}{3}}|f'(x)|^{\frac{2k-2}{3}}\,dx\int_0^{\infty}\operatorname{cov}(|\xi(0)|^k, |\xi(c)|^k)\,dc.$$

Proof: Write

$$W_n^k(a) = \frac{|V_n^W(a)|^k - E|V_n^W(a)|^k}{|g'(a)|^{k-1}},$$

and define

$$\begin{aligned} L_n &= (f(0)-f(1))n^{-1/3}(\log n)^3, \\ M_n &= (f(0)-f(1))n^{-1/3}\log n, \\ N_n &= \left[\frac{(f(0)-f(1))}{L_n+M_n}\right] = \left[\frac{n^{1/3}}{\log n + (\log n)^3}\right], \end{aligned}$$

where $[x]$ denotes the integer part of x. We divide the interval $(f(1), f(0))$ into $2N_n+1$ blocks of alternating length

$$\begin{aligned} A_j &= (f(1)+(j-1)(L_n+M_n), f(1)+(j-1)(L_n+M_n)+L_n], \\ B_j &= (f(1)+(j-1)(L_n+M_n)+L_n, f(1)+j(L_n+M_n)], \end{aligned}$$

where $j = 1, \ldots, N_n$. Now write

$$T_{n,k} = S'_{n,k} + S''_{n,k} + R_{n,k},$$

where

$$\begin{aligned} S'_{n,k} &= n^{1/6}\sum_{j=1}^{N_n}\int_{A_j} W_n^k(a)\,da, \\ S''_{n,k} &= n^{1/6}\sum_{j=1}^{N_n}\int_{B_j} W_n^k(a)\,da, \\ R_{n,k} &= n^{1/6}\int_{f(1)+N_n(L_n+M_n)}^{f(0)} W_n^k(a)\,da. \end{aligned}$$

From here on the proof is completely the same as the proof of Theorem 4.1 in GROENEBOOM ET AL. (1999). Therefore we omit all specific details and only give a brief outline of the argument. Lemmas 6.2.1 and 6.2.3 imply that all moments of $W_n^k(a)$ are bounded uniformly in a, and that $E|W_n^k(a)W_n^k(b)| \leq D_1 \exp(-D_2 n|b-a|^3)$. This is used to ensure that $ER_n^2 \to 0$ and that the contribution of the small blocks is negligible: $E(S''_{n,k})^2 \to 0$. We then only have to consider the contribution over the big blocks. When

$$Y_j = n^{1/6}\int_{A_j} W_n^k(a)\,da \quad \text{and} \quad \sigma_n^2 = \text{var}\left(\sum_{j=1}^{N_n} Y_j\right),$$

then one finds that

$$\left| E\exp\left\{\frac{iu}{\sigma_n}\sum_{j=1}^{N_n} Y_j\right\} - \prod_{j=1}^{N_n} E\exp\left\{\frac{iu}{\sigma_n}Y_j\right\}\right| \leq 4(N_n-1)\exp(-C_3 n M_n^3) \to 0,$$

where $C_3 > 0$ only depends on f. This means that we can apply the central limit theorem to independent copies of Y_j. Hence, asymptotic normality of $S'_{n,k}$ follows if we show that the contribution of the big blocks satisfies the Lindeberg condition, i.e., for each $\varepsilon > 0$,

$$\frac{1}{\sigma_n^2}\sum_{j=1}^{N_n} EY_j^2 1_{\{|Y_j|>\varepsilon\sigma_n\}} \to 0, \quad n\to\infty. \tag{6.2.8}$$

By using the uniform boundednes of the moments of $|W_n^k(a)|$, we have that

$$\frac{1}{\sigma_n^2}\sum_{j=1}^{N_n} EY_j^2 1_{\{|Y_j|>\epsilon\sigma_n\}} \le \frac{1}{\epsilon\sigma_n^3} N_n \sup_{1\le k\le N_n} E|Y_j|^3 = \mathcal{O}(\sigma_n^{-3} n^{-1/6}(\log n)^6)).$$

By similar computations as in the proof of Theorem 4.1 in GROENEBOOM ET AL.(1999), we find that

$$\sigma_n^2 = \mathrm{var}(T_{n,k}) + \mathcal{O}(1).$$

By application of Lemma 6.2.4, with $l = 0$, $m = k$ and $h(a) = 1/|g'(a)|^{k-1}$, it follows that $\sigma_n^2 \to \sigma^2$, which implies (6.2.8). ■

The next lemma shows that the limiting expectation in Theorem 6.2.1 is equal to

$$\mu_k = E|V(0)| \left\{\int_0^1 (4f(x)|f'(x)|)^{k/3}\, dx\right\}^{1/k}. \tag{6.2.9}$$

Lemma 6.2.5 *Let V_n^W be defined by (6.2.2) and let μ_k be defined by (6.2.9). Moreover let $V(0)$ be defined by (6.1.2). Then for $k \ge 1$*

(i) For all a such that

$$n^{1/3}\{F(g(a)) \wedge (1 - F(g(a)))\} \ge \log n, \tag{6.2.10}$$

we have

$$E|V_n^W(a)|^k = E|V(0)|^k \frac{(4a)^{k/3}}{|f'(g(a))|^{2k/3}} + \mathcal{O}(n^{-1/3}(\log n)^{k+3}),$$

where the term $\mathcal{O}(n^{-1/3}(\log n)^{k+3})$ is uniform in all a satisfying (6.2.10).

(ii)

$$\lim_{n\to\infty} n^{1/6}\left\{\int_{f(1)}^{f(0)} \frac{E|V_n^W(a)|^k}{|g'(a)|^{k-1}}\, da - \mu_k^k\right\} = 0.$$

Proof: The proof relies on the proof of Corollary 3.2 in GROENEBOOM ET AL.(1999). There it is shown that, if we define

$$H_n(y) = n^{1/3}\left\{H\left(F(g(a)) + n^{-1/3}y\right) - g(a)\right\},$$

with H being the inverse of F, and

$$V_{n,b} = \sup\left\{y \in [-n^{1/3}F(g(a)), n^{1/3}(1 - F(g(a)))] : W(y) - by^2 \text{ is maximal}\right\},$$

with $b = |f'(g(a))|/(2a^2)$, then for the event $A_n = \{|V_n^W(a)| \leq \log n, |H_n(V_{n,b})| \leq \log n\}$, one has that $P\{A_n^c\}$ is of the order $\mathcal{O}(e^{-C(\log n)^3})$, which then implies that

$$\sup_{a\in(f(1),f(0))} E|V_n^W(a) - H_n(V_{n,b})| = \mathcal{O}(n^{-1/3}(\log n)^4).$$

Similarly, together with an application of the mean value theorem, this yields

$$\sup_{a\in(f(1),f(0))} E\left||V_n^W(a)|^k - |H_n(V_{n,b})|^k\right| = \mathcal{O}(n^{-1/3}(\log n)^{3+k}). \tag{6.2.11}$$

Note that by definition, the argmax $V_{n,b}$ closely resembles the argmax $V_b(0)$, where

$$V_b(c) = \underset{t\in\mathbb{R}}{\operatorname{argmax}}\{W(t) - b(t-c)^2\}. \tag{6.2.12}$$

Therefore we write

$$E\,|H_n(V_{n,b})|^k = E\,|H_n(V_b(0))|^k + E\left(|H_n(V_{n,b})|^k - |H_n(V_b(0))|^k\right). \tag{6.2.13}$$

Since by Brownian scaling $V_b(c)$ has the same distribution as $b^{-2/3}V(cb^{2/3})$, where V is defined in (6.1.2), together with the conditions on f, we find that

$$E\,|H_n(V_b(0))|^k = a^{-k}E\,|V_b(0)|^k + \mathcal{O}(n^{-1/3}) = \frac{(4a)^{k/3}}{|f'(g(a))|^{2k/3}}E|V(0)|^k + \mathcal{O}(n^{-1/3}).$$

As in the proof of Corollary 3.2 in Groeneboom et al.(1999), $V_{n,b}$ can only be different from $V_b(0)$ with probability of order $e^{-\frac{2}{3}(\log n)^3}$. Hence, from (6.2.13) we conclude

$$E\,|H_n(V_{n,b})|^k = \frac{(4a)^{k/3}}{|f'(g(a))|^{2k/3}}E|V(0)|^k + \mathcal{O}(n^{-1/3}).$$

Together with (6.2.11) this proves (i).

(ii) This immediately follows from (i). The values of a for which condition (6.2.10) does not hold, gives a contribution of order $n^{-1/3}\log n$ to the integral $\int E|V_n^W(a)|^k\,da$, and finally,

$$\int_{f(1)}^{f(0)} \frac{(4a)^{k/3}}{|f'(g(a))|^{2k/3}|g'(a)|^{k-1}}\,da = \int_0^1 (4f(x))^{k/3}|f'(x)|^{k/3}dx. \qquad \blacksquare$$

The next step is to transfer the result of Theorem 6.2.1 to the L_k-error of V_n^E. This is done in the next two lemmas. The first lemma shows that the difference between the L_k-errors of V_n^W and V_n^B is of negligible order. The second lemma does the same for the L_k-errors of V_n^B and V_n^E.

Lemma 6.2.6 *Let V_n^W and V_n^B be defined as in (6.2.2). Then for $k \geq 1$, we have*

$$n^{1/6}\int_{f(1)}^{f(0)} \left(|V_n^B(a)|^k - |V_n^W(a)|^k\right)da = o_p(1).$$

Proof: The proof relies on the proof of the Corollary 3.3 in GROENEBOOM ET AL.(1990). Here it is shown, that if for a belonging to the set

$$J_n = \{a : \text{both } a \text{ and } a(1 - \xi_n n^{-1/2}) \in (f(1), f(0))\},$$

we define

$$V_n^B(a, \xi_n) = V_n^B(a(1 - n^{-1/2}\xi_n)) + n^{1/3}\left\{g(a(1 - n^{-1/2}\xi_n)) - g(a)\right\},$$

then for the event $A_n = \{|\xi_n| \le n^{1/6}, |V_n^W(a)| \le \log n, |V_n^B(a, \xi_n)| \le \log n\}$, one has that $P\{A_n^c\}$ is of the order $\mathcal{O}(e^{-C(\log n)^3})$, which then implies that

$$\int_{a\in J_n} E\left|V_n^B(a, \xi_n) - V_n^W(a)\right| da = \mathcal{O}(n^{-1/3}(\log n)^3).$$

Hence, by using the same method as in proof of Lemma 6.2.7, we obtain:

$$\int_{a\in J_n} E\left||V_n^B(a, \xi_n)|^k - |V_n^W(a)|^k\right| da = \mathcal{O}(n^{-1/3}(\log n)^{k+2}).$$

From Lemma 6.2.1 it also follows that $E|V_n^B(a)|^k = \mathcal{O}(1)$ and $E|V_n^W(a)|^k = \mathcal{O}(1)$, uniformly with respect to n and $a \in (f(1), f(0))$. Hence the contribution of the integrals over $[f(1), f(0)] \setminus J_n$ is negligible, and it remains to show that

$$n^{1/6} \int_{a\in J_n} \left\{|V_n^B(a, \xi_n)|^k - |V_n^B(a)|^k\right\} da = o_p(1). \tag{6.2.14}$$

For $k = 1$, this is shown in the proof of Corollary 3.3 in GROENEBOOM ET AL.(1999), so we may assume that $k > 1$. Completely similar to the proof in the case $k = 1$, we first obtain

$$\begin{aligned} &n^{1/6} \int_{a\in J_n} \left\{|V_n^B(a, \xi_n)|^k - |V_n^B(a)|^k\right\} da \\ &= n^{1/6} \int_{f(1)}^{f(0)} \left\{\left|V_n^B(a) - ag'(a)\xi_n n^{-1/6}\right|^k - |V_n^B(a)|^k\right\} da + \mathcal{O}_p(n^{-1/3}). \end{aligned}$$

Let $\epsilon > 0$ and write $\Delta_n(a) = ag'(a)\xi_n n^{-1/6}$. Then we can write

$$n^{1/6} \int_{f(1)}^{f(0)} \left\{\left|V_n^B(a) - \Delta_n(a)\right|^k - |V_n^B(a)|^k\right\} da$$

$$= n^{1/6} \int_{f(1)}^{f(0)} \left\{\left|V_n^B(a) - \Delta_n(a)\right|^k - |V_n^B(a)|^k\right\} 1_{[0,\epsilon]}(|V_n^B(a)|)\, da \tag{6.2.15}$$

$$+ n^{1/6} \int_{f(1)}^{f(0)} \left\{\left|V_n^B(a) - \Delta_n(a)\right|^k - |V_n^B(a)|^k\right\} 1_{(\epsilon,\infty)}(|V_n^B(a)|)\, da. \tag{6.2.16}$$

First consider the term (6.2.15) and distinguish between

1. $|V_n^B(a)| < 2|\Delta_n(a)|$,
2. $|V_n^B(a)| \geq 2|\Delta_n(a)|$.

In case 1,

$$\left||V_n^B(a) - \Delta_n(a)|^k - |V_n^B(a)|^k\right| \leq 3^k|\Delta_n(a)|^k + 2^k|\Delta_n(a)|^k \leq (3^k + 2^k)|ag'(a)\xi_n|^k n^{-k/6}.$$

In case 2, note that

$$\left||V_n^B(a) - \Delta_n(a)|^k - |V_n^B(a)|^k\right| = k|\theta|^{k-1}|\Delta_n(a)|,$$

where θ is between $|V_n^B(a)| \leq \epsilon$ and $|V_n^B(a) - \Delta_n(a)| \leq \frac{3}{2}\epsilon$. Using that ξ_n and V_n^B are independent, the expectation of (6.2.15) is bounded from above by

$$C_1\epsilon^{k-1}E|\xi_n| \int_{f(1)}^{f(0)} |ag'(a)|P\{|V_n^B(a)| \leq \epsilon\}\,da + \mathcal{O}_p(n^{-(k-1)/6}),$$

where $C_1 > 0$ only depends on f and k. Hence, since $k > 1$, we find that

$$\limsup_{n\to\infty} n^{1/6} \int_{f(1)}^{f(0)} \left\{\left|V_n^B(a) - ag'(a)\xi_n n^{-1/6}\right|^k - |V_n^B(a)|^k\right\} 1_{[0,\epsilon]}(|V_n^B(a)|)\,da \tag{6.2.17}$$

is bounded from above by $C_2\epsilon^{k-1}$, where $C_2 > 0$ only depends on f and k. Letting $\epsilon \downarrow 0$ and using that $k > 1$, then yields that (6.2.15) tends to zero.

The term (6.2.16) is equal to

$$\int_{f(1)}^{f(0)} \frac{-2\xi_n ag'(a)V_n^B(a) + (ag'(a)\xi_n)^2 n^{-1/6}}{|V_n^B(a) - \Delta_n(a)| + |V_n^B(a)|} \cdot k\theta(a)^{k-1}1_{(\epsilon,\infty)}(|V_n^B(a)|)\,da, \tag{6.2.18}$$

where $\theta(a)$ is between $|V_n^B(a) - \Delta_n(a)|$ and $|V_n^B(a)|$. Note that for $|V_n^B(a)| > \epsilon$,

$$\left|\frac{2V_n^B(a)}{|V_n^B(a) - \Delta_n(a)| + |V_n^B(a)|} - \frac{V_n^B(a)}{|V_n^B(a)|}\right| \leq \frac{|ag'(a)n^{-1/6}\xi_n|}{\epsilon} = \mathcal{O}_p(n^{-1/6}),$$

uniformly in $a \in (f(1), f(0))$, so that (6.2.18) is equal to

$$-k\xi_n \int_{f(1)}^{f(0)} ag'(a)V_n^B(a)|V_n^B(a)|^{k-2}1_{(\epsilon,\infty)}(|V_n^B(a)|)\,da$$
$$+k\xi_n \int_{f(1)}^{f(0)} ag'(a)\frac{V_n^B(a)}{|V_n^B(a)|}\left(|V_n^B(a)|^{k-1} - \theta(a)^{k-1}\right)1_{(\epsilon,\infty)}(|V_n^B(a)|)\,da + \mathcal{O}_p(n^{-1/6}).$$

We have that

$$\left||V_n^B(a)|^{k-1} - \theta(a)^{k-1}\right| \leq |V_n^B(a)|^{k-1}\left|\left|1 - \frac{\Delta_n(a)}{V_n^B(a)}\right|^{k-1} - 1\right| = \mathcal{O}_p(n^{-1/6}),$$

where the big $\mathcal{O}$-term in uniform in a. This means that (6.2.18) is equal to

$$-k\xi_n \int_{f(1)}^{f(0)} ag'(a)V_n^B(a)|V_n^B(a)|^{k-2}\,da \tag{6.2.19}$$

$$+k\xi_n \int_{f(1)}^{f(0)} ag'(a)\text{sign}(V_n^B(a))|V_n^B(a)|^{k-1}1_{[0,\epsilon)}(|V_n^B(a)|)\,da + \mathcal{O}_p(n^{-1/6}). \tag{6.2.20}$$

The integral in (6.2.20) is of the order $\mathcal{O}(\epsilon^{k-1})$, whereas $E\xi_n^2 = 1$. Since $k > 1$, this means that after letting $\epsilon \downarrow 0$, (6.2.20) tends to zero. Finally, let $S_n^B(a) = ag'(a)V_n^B(a)\left|V_n^B(a)\right|^{k-2}$ and write

$$E\left(\xi_n \int_{f(1)}^{f(0)} S_n^B(a)\,da\right)^2 = \operatorname{var}\left(\int_{f(1)}^{f(0)} S_n^B(a)\,da\right) + \left(E\int_{f(1)}^{f(0)} S_n^B(a)\,da\right)^2.$$

Then, since according to Lemma 6.2.1, all moments of $|S_n^B(a)|$ are bounded uniformly in a, we find by dominated convergence and Lemma 6.2.2 that

$$\lim_{n\to\infty} E\int_{f(1)}^{f(0)} S_n^B(a)\,da = \int_{f(1)}^{f(0)} \frac{a|g'(a)|}{(\phi_1(a))^k}\left(E\xi(0)\,|\xi(0)|^{k-2}\right)da = 0,$$

because the distribution of $\xi(0)$ is symmetric. Applying Lemma 6.2.4 with $l = 1$, $m = k-2$ and $h(a) = ag'(a)$ we obtain

$$\operatorname{var}\left(\int_{f(1)}^{f(0)} ag'(a)V_n^B(a)\left|V_n^B(a)\right|^{k-2} da\right) = O(n^{-1/3}).$$

We conclude that (6.2.18) tends to zero in probability. This proves the lemma. ■

Lemma 6.2.7 *Let V_n^E and V_n^B be defined as in (6.2.2). Then for $k \geq 1$, we have*

$$\int_{f(1)}^{f(0)} \left||V_n^E(a)|^k - |V_n^B(a)|^k\right| da = \mathcal{O}_p(n^{-1/3}(\log n)^{k+2}).$$

Proof: The proof relies on the proof of Corollary 3.1 in GROENEBOOM ET AL.(1999). There it is shown that for the event $A_n = \left\{|V_n^B(a)| < \log n\,, |V_n^E(a)| < \log n\right\}$ one has that $P\{A_n^c\}$ is of the order $e^{-C(\log n)^3}$. Furthermore, if $K_n = \{\sup_t |E_n(t) - B_n(F(t))| \leq n^{-1/2}(\log n)^2\}$, then $P(K_n) \to 1$ and

$$E\left||V_n^E(a)| - |V_n^B(a)|\right| 1_{A_n\cap K_n} = \mathcal{O}(n^{-1/3}(\log n)^3), \tag{6.2.21}$$

uniformly in $a \in (f(1), f(0))$. By the mean value theorem, together with (6.2.21), we now have that

$$\begin{aligned} E\left||V_n^E(a)|^k - |V_n^B(a)|^k\right| 1_{K_n} &\leq k(\log n)^{k-1}E\left||V_n^E(a)| - |V_n^B(a)|\right| 1_{A_n\cap K_n} + 2n^{k/3}P\{A_n^c\} \\ &= \mathcal{O}(n^{-1/3}(\log n)^{k+2}) + \mathcal{O}(n^{k/3}e^{-C(\log n)^3}). \end{aligned}$$

This proves the lemma. ■

Corollary 6.2.1 *Let U_n be defined by (6.1.1) and let μ_k be defined by (6.2.9). Then for any $k \geq 1$,*

$$n^{1/6}\left(n^{k/3}\int_{f(1)}^{f(0)} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}}\,da - \mu_k^k\right) \to N(0,\sigma^2)$$

in distribution, as $n \to \infty$, where σ^2 is defined in Theorem 6.2.1.

6.3 Relating both L_k-errors

When $k = 1$, the L_k-error has an easy interpretation as the area between two graphs. In that case $\int |U_n(a) - g(a)|\, da$ is the same as $\int |\hat{f}_n(x) - f(x)|\, dx$, up to some boundaries effects. This is precisely Corollary 2.1 in GROENEBOOM ET AL.(1999). In this section we show that a similar approximation holds for $\int_s^t |\hat{f}_n(x) - f(x)|^k\, dx$ on segments $[s, t]$, where the graphs of $\hat{f}_n$ and f do not intersect. In order to avoid boundary problems, we will apply this approximation in subsequent sections to a suitable cut-off version $\tilde{f}_n$ of $\hat{f}_n$.

Lemma 6.3.1 *Let $\tilde{f}_n$ be a piecewise constant left-continuous non-increasing function on $[0, 1]$ with a finite number of jumps. Suppose that $f(1) \leq \tilde{f}_n \leq f(0)$, and define its inverse function by*

$$\tilde{U}_n(a) = \sup\left\{x \in [0,1] : \tilde{f}_n(x) \geq a\right\},$$

for $a \in [f(1), f(0)]$. Suppose that $[s,t] \subseteq [0,1]$, such that one of the following situations applies:

1. *$\tilde{f}_n(x) \geq f(x)$, for $x \in (s,t)$, such that $\tilde{f}_n(s) = f(s)$ and $\tilde{f}_n(t+) \leq f(t)$,*
2. *$\tilde{f}_n(x) \leq f(x)$, for $x \in (s,t)$, such that $\tilde{f}_n(t) = f(t)$ and $\tilde{f}_n(s) \geq f(s)$.*

If

$$\sup_{x\in[s,t]} \left|\tilde{f}_n(x) - f(x)\right| < \frac{(\inf_{x\in[0,1]} |f'(x|)^2}{2\sup_{x\in[0,1]} |f''(x)|}, \tag{6.3.1}$$

then for $k \geq 1$,

$$\left|\int_s^t \left|\tilde{f}_n(x) - f(x)\right|^k dx - \int_{f(t)}^{f(s)} \frac{\left|\tilde{U}_n(a) - g(a)\right|^k}{|g'(a)|^{k-1}}\, da\right| \leq C \int_{f(t)}^{f(s)} \frac{\left|\tilde{U}_n(a) - g(a)\right|^{k+1}}{|g'(a)|^k}\, da,$$

where $C > 0$ only depends on f and k.

Proof: Let us first consider case 1. Let $\tilde{f}_n$ have m points of jump on (s,t). Denote them in increasing order by $\xi_1 < \cdots < \xi_m$, and write $s = \xi_0$ and $\xi_{m+1} = t$. Denote by $\alpha_1 > \cdots > \alpha_m$ the points of jump of $\tilde{U}_n$ on the interval $(f(t), f(s))$ in decreasing order, and write $f(s) = \alpha_0$ and $\alpha_{m+1} = f(t)$ (see Figure 6.1). We then have

$$\int_s^t \left|\tilde{f}_n(x) - f(x)\right|^k dx = \sum_{i=0}^m \int_{\xi_i}^{\xi_{i+1}} \left|\tilde{f}_n(\xi_{i+1}) - f(x)\right|^k dx.$$

Apply Taylor expansion to f in the point $g(\alpha_i)$ for each term, and note that $\tilde{f}_n(\xi_{i+1}) = \alpha_i$. Then, if we abbreviate $g_i = g(\alpha_i)$, for $i = 0, 1, \ldots, m$, we can write the right hand side as

$$\sum_{i=0}^m \int_{\xi_i}^{\xi_{i+1}} |f'(g_i)|^k (x - g_i)^k \left|1 + \frac{f''(\theta_i)}{2f'(g_i)}(x - g_i)\right|^k dx,$$

for some θ_i between x and g_i, also using the fact that $g_i < \xi_i < x \leq \xi_{i+1}$. Due to condition (6.3.1) and the fact that $\tilde{f}_n(\xi_{i+1}) = \tilde{f}_n(x)$, for $x \in (\xi_i, \xi_{i+1}]$, we have that

$$\left|\frac{f''(\theta_i)}{f'(g_i)}(x - g_i)\right| \leq \frac{\sup|f''|}{\inf|f'|}\frac{|f(x) - f(g_i)|}{\inf|f'|} \leq \frac{\sup|f''|}{(\inf|f'|)^2}\left|f(x) - \tilde{f}_n(x)\right| \leq \frac{1}{2}. \tag{6.3.2}$$

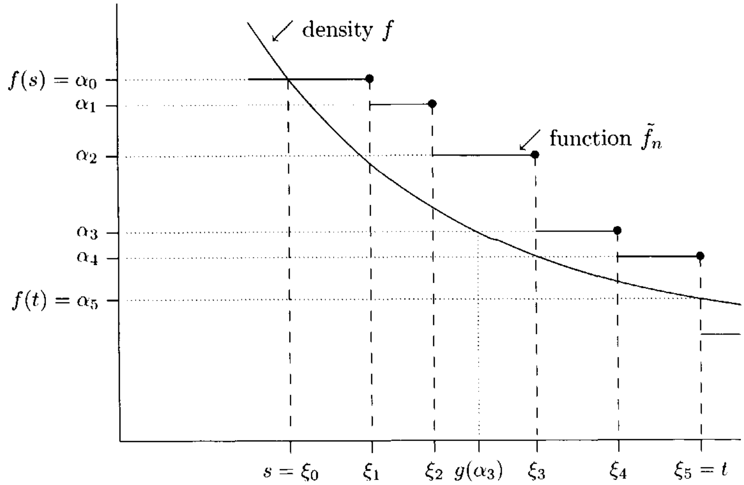


Figure 6.1: Segment $[s,t]$ where $\tilde{f}_n \geq f$.

Hence for $x \in (\xi_i, \xi_{i+1}]$,

$$\left|1 + \frac{f''(\theta_i)(x-g_i)}{2f'(g_i)}\right|^k \leq 1 + \frac{|f''(\theta_i)|(x-g_i)}{2|f'(g_i)|} \sup_{z \in [\frac{1}{2},\frac{3}{2}]} kz^{k-1} \leq 1 + C_1(x-g_i),$$

where $C_1 = \frac{\sup|f''|}{2\inf|f'|} k \left(\frac{3}{2}\right)^{k-1}$. Similarly,

$$\left|1 + \frac{f''(\theta_i)(x-g_i)}{2f'(g_i)}\right|^k \geq 1 - C_1(x-g_i).$$

Therefore we obtain the following inequality

$$\left|\int_s^t \left|\tilde{f}_n(x) - f(x)\right|^k dx - \sum_{i=0}^m \int_{\xi_i}^{\xi_{i+1}} |f'(g_i)|^k (x-g_i)^k \, dx\right| \leq C_1 \sum_{i=0}^m \int_{\xi_i}^{\xi_{i+1}} (x-g_i)^{k+1} \, dx.$$

After integration, we can rewrite this inequality in the following way:

$$\left|\int_s^t \left|\tilde{f}_n(x) - f(x)\right|^k dx - \frac{1}{k+1} \sum_{i=0}^m |f'(g_i)|^k \left\{(\xi_{i+1} - g_i)^{k+1} - (\xi_i - g_i)^{k+1}\right\}\right|$$
$$\leq \frac{C_1}{k+2} \sum_{i=0}^m \left\{(\xi_{i+1} - g_i)^{k+2} - (\xi_i - g_i)^{k+2}\right\}. \tag{6.3.3}$$

Next, let us consider the second integral in the statement of the lemma:

$$\int_{f(t)}^{f(s)} \frac{\left|\tilde{U}_n(a) - g(a)\right|^k}{|g'(a)|^{k-1}} \, da = \sum_{i=0}^m \int_{\alpha_{i+1}}^{\alpha_i} \frac{|\xi_{i+1} - g(a)|^k}{|g'(a)|^{k-1}} \, da = \sum_{i=0}^m \int_{g_i}^{g_{i+1}} (\xi_{i+1} - x)^k |f'(x)|^k \, dx,$$

now using that $g_i < x < g_{i+1} < \xi_{i+1}$. Apply Taylor expansion to f' in the point g_i. For the right hand side, we then obtain

$$\sum_{i=0}^{m} \int_{g_i}^{g_{i+1}} (\xi_{i+1} - x)^k |f'(g_i) + f''(\theta_i)(x - g_i)|^k \, dx,$$

for some θ_i between x and g_i. Using (6.3.2), by means of the same arguments as above we get the following inequality:

$$\left| \int_{f(t)}^{f(s)} \frac{\left|\tilde{U}_n(a) - g(a)\right|^k}{|g'(a)|^{k-1}} \, da - \sum_{i=0}^{m} \int_{g_i}^{g_{i+1}} |f'(g_i)|^k (\xi_{i+1} - x)^k \, dx \right|$$
$$\leq C_1 \sum_{i=0}^{m} \int_{g_i}^{g_{i+1}} (\xi_{i+1} - x)^k (x - g_i) \, dx. \tag{6.3.4}$$

Since $g_i < x < g_{i+1} < \xi_{i+1}$, for each term on the right hand side of (6.3.4), we have that

$$\begin{aligned} \int_{g_i}^{g_{i+1}} (\xi_{i+1} - x)^k (x - g_i) \, dx &\leq (\xi_{i+1} - g_i) \int_{g_i}^{g_{i+1}} (\xi_{i+1} - x)^k \, dx \\ &= \frac{1}{k+1} \left\{ (\xi_{i+1} - g_i)^{k+2} - (\xi_{i+1} - g_{i+1})^{k+1} (\xi_{i+1} - g_i) \right\} \\ &\leq \frac{1}{k+1} \left\{ (\xi_{i+1} - g_i)^{k+2} - (\xi_{i+1} - g_{i+1})^{k+2} \right\}. \end{aligned}$$

Hence from (6.3.4) we find that

$$\left| \int_{f(t)}^{f(s)} \frac{\left|\tilde{U}_n(a) - g(a)\right|^k}{|g'(a)|^{k-1}} \, da - \frac{1}{k+1} \sum_{i=0}^{m} |f'(g_i)|^k \left\{ (\xi_{i+1} - g_i)^{k+1} - (\xi_{i+1} - g_{i+1})^{k+1} \right\} \right|$$
$$\leq \frac{C_1}{k+1} \sum_{i=0}^{m} \left\{ (\xi_{i+1} - g_i)^{k+2} - (\xi_{i+1} - g_{i+1})^{k+2} \right\}. \tag{6.3.5}$$

For the third integral in the statement of the lemma, similarly as before, we can write

$$\int_{f(t)}^{f(s)} \frac{\left|\tilde{U}_n(a) - g(a)\right|^{k+1}}{|g'(a)|^k} \, da = \sum_{i=0}^{m} \int_{g_i}^{g_{i+1}} |f'(g_i)|^{k+1} (\xi_{i+1} - x)^{k+1} \left| 1 + \frac{f''(\theta)}{f'(g_i)} (x - g_i) \right|^{k+1}.$$

According to (6.3.2) we have that for $x \in (g_i, g_{i+1})$,

$$\left| 1 + \frac{f''(\theta)}{f'(g_i)} (x - g_i) \right| \geq \frac{1}{2},$$

so that after integration we obtain

$$\int_{f(t)}^{f(s)} \frac{\left|\tilde{U}_n(a) - g(a)\right|^{k+1}}{|g'(a)|^k} \, da \geq \frac{C_2}{k+2} \sum_{i=0}^{m} \left\{ (\xi_{i+1} - g_i)^{k+2} - (\xi_{i+1} - g_{i+1})^{k+2} \right\}, \tag{6.3.6}$$

where $C_2 = \left(\frac{1}{2}\right)^{k+1} \inf |f'|^{k+1}$.

Now, let us define Δ as the difference between the first two integrals:

$$\Delta \stackrel{\text{def}}{=} \int_s^t \left|\tilde{f}_n(x) - f(x)\right|^k dx - \int_{f(t)}^{f(s)} \frac{\left|\tilde{U}_n(a) - g(a)\right|^k}{|g'(a)|^{k-1}} da.$$

By (6.3.3) and (6.3.5) and the fact that $\xi_0 = g_0$ and $\xi_{m+1} = g_{m+1}$, we find that

$$\begin{aligned} |\Delta| \leq\ & D\sum_{i=0}^{m}(\xi_{i+1} - g_{i+1})^{k+1} \left||f'(g_i)|^k - |f'(g_{i+1})|^k\right| \\ & + D\sum_{i=0}^{m}\left\{(\xi_{i+1} - g_i)^{k+2} - (\xi_i - g_i)^{k+2}\right\} \\ & + D\sum_{i=0}^{m}\left\{(\xi_{i+1} - g_i)^{k+2} - (\xi_{i+1} - g_{i+1})^{k+2}\right\}, \end{aligned} \tag{6.3.7}$$

where D is some positive constant that depends only on the function f and k. By a Taylor expansion, the first term on the right hand side of (6.3.7) can be bounded by

$$\begin{aligned} & D\sum_{i=0}^{m}(\xi_{i+1} - g_{i+1})^{k+1} \left||f'(g_i)|^k - |f'(g_i) + f''(\theta_i)(g_{i+1} - g_i)|^k\right| \\ & \leq D\sum_{i=0}^{m}(\xi_{i+1} - g_{i+1})^{k+1}|f'(g_i)|^k \left|1 - \left|1 + \frac{f''(\theta_i)(g_{i+1} - g_i)}{f'(g_i)}\right|^k\right| \\ & \leq D\sum_{i=0}^{m}(\xi_{i+1} - g_{i+1})^{k+1}(g_{i+1} - g_i) \sup|f'|^k \frac{\sup|f''|}{\inf|f'|} \sup_{x\in[\frac{1}{2},\frac{3}{2}]} kz^{k-1} \\ & \leq C_3\sum_{i=0}^{m}(\xi_{i+1} - g_{i+1})^{k+1}(g_{i+1} - g_i), \end{aligned}$$

with C_3 only depending on f and k, where we also use (6.3.2) and the fact that according to (6.3.1), we have that $(g_{i+1} - g_i)\sup|f''|/\inf|f'| < \frac{1}{2}$. Since $g_i < g_{i+1} < \xi_{i+1}$, this means that the first term on the right hand side of (6.3.7) can be bounded by

$$\begin{aligned} C_3\sum_{i=0}^{m}(\xi_{i+1} - g_{i+1})^{k+1}(g_{i+1} - g_i) & \leq C_3\sum_{i=0}^{m}\left\{(\xi_{i+1} - g_i) - (\xi_{i+1} - g_{i+1})\right\}(\xi_{i+1} - g_{i+1})^{k+1} \\ & \leq C_3\sum_{i=0}^{m}\left\{(\xi_{i+1} - g_i)^{k+2} - (\xi_{i+1} - g_{i+1})^{k+2}\right\}. \end{aligned}$$

Because $\xi_0 = g_0$ and $\xi_{m+1} = g_{m+1}$, for the second term on the right hand side of (6.3.7), we have that

$$\sum_{i=0}^{m}\left\{(\xi_{i+1} - g_i)^{k+2} - (\xi_i - g_i)^{k+2}\right\} = \sum_{i=0}^{m}\left\{(\xi_{i+1} - g_i)^{k+2} - (\xi_{i+1} - g_{i+1})^{k+2}\right\}.$$

Putting things together and using (6.3.6) we find that

$$|\Delta| \leq C_4 \sum_{i=0}^{m} \left\{ (\xi_{i+1} - g_i)^{k+2} - (\xi_{i+1} - g_{i+1})^{k+2} \right\} \leq C_5 \int_{f(t)}^{f(s)} \frac{\left|\tilde{U}_n(a) - g(a)\right|^{k+1}}{|g'(a)|^k} \, da,$$

where C_5 only depends on f and k. This proves the lemma for case 1.

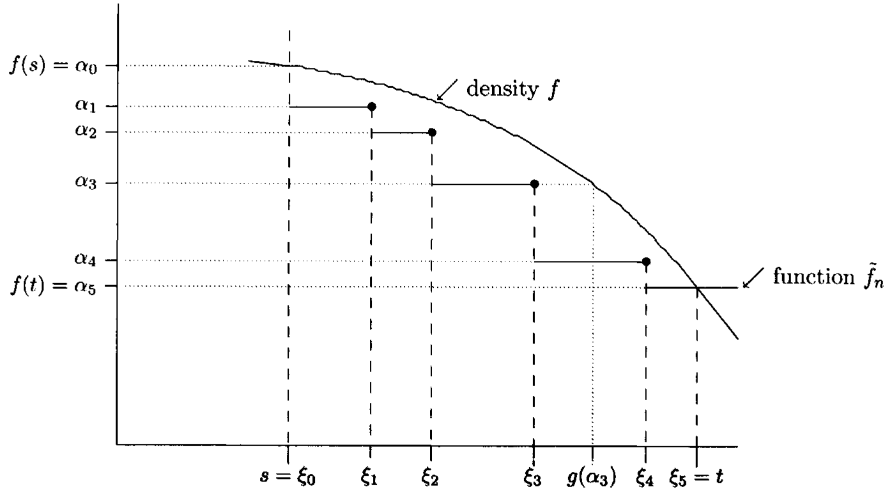


Figure 6.2: Segment $[s, t]$ where $\tilde{f}_n \leq f$.

For case 2, the proof is completely similar. The main difference are that $\tilde{f}_n(\xi_i) = \alpha_i$ (see Figure 6.2) and that the Taylor expansions are applied to f in g_{i+1} instead of g_i. Similar to (6.3.3), now using that $g_{i+1} \geq \xi_{i+1} \geq \xi_i$, we now obtain

$$\left| \int_s^t \left|\tilde{f}_n(x) - f(x)\right|^k dx - \frac{1}{k+1} \sum_{i=0}^{m} |f'(g_{i+1})|^k \left\{ (g_{i+1} - \xi_i)^{k+1} - (g_{i+1} - \xi_{i+1})^{k+1} \right\} \right|$$
$$\leq \frac{C_1}{k+2} \sum_{i=0}^{m} \left\{ (g_{i+1} - \xi_i)^{k+2} - (g_{i+1} - \xi_{i+1})^{k+2} \right\}. \tag{6.3.8}$$

Similar to (6.3.5), now using that $g_{i+1} \geq g_i \geq \xi_i$, we now obtain

$$\left| \int_{f(t)}^{f(s)} \frac{\left|\tilde{U}_n(a) - g(a)\right|^k}{|g'(a)|^{k-1}} \, da - \frac{1}{k+1} \sum_{i=0}^{m} |f'(g_{i+1})|^k \left\{ (g_{i+1} - \xi_i)^{k+1} - (g_i - \xi_i)^{k+1} \right\} \right|$$
$$\leq \frac{C_1}{k+1} \sum_{i=0}^{m} \left\{ (g_{i+1} - \xi_i)^{k+2} - (g_i - \xi_i)^{k+2} \right\}, \tag{6.3.9}$$

and similar to (6.3.6) we find

$$\int_{f(t)}^{f(s)} \frac{\left|\tilde{U}_n(a) - g(a)\right|^{k+1}}{|g'(a)|^k} \, da \geq \frac{C_2}{k+2} \sum_{i=0}^{m} \left\{ (g_{i+1} - \xi_i)^{k+2} - (g_i - \xi_i)^{k+2} \right\}. \tag{6.3.10}$$

For the difference between the two integrals, again using that $\xi_0 = g_0$ and $\xi_{m+1} = g_{m+1}$, we now find

$$\begin{aligned} |\Delta| \;\leq\; & D \sum_{i=0}^{m} (g_i - \xi_i)^{k+1} \left| |f'(g_i)|^k - |f'(g_{i+1})|^k \right| \\ & + D \sum_{i=0}^{m} \left\{ (g_{i+1} - \xi_i)^{k+2} - (g_i - \xi_i)^{k+2} \right\} \\ & + D \sum_{i=0}^{m} \left\{ (g_{i+1} - \xi_i)^{k+2} - (g_{i+1} - \xi_{i+1})^{k+2} \right\} \end{aligned} \tag{6.3.11}$$

where D is some positive constant that depends only on the function f and k. The first two terms on the right hand side of (6.3.11) can be bounded similar to the first two terms on the right hand side of (6.3.7), which results in

$$|\Delta| \leq C_4 \sum_{i=0}^{m} \left\{ (g_{i+1} - \xi_i)^{k+2} - (g_i - \xi_i)^{k+2} \right\} \leq C_5 \int_{f(t)}^{f(s)} \frac{\left|\tilde{U}_n(a) - g(a)\right|^{k+1}}{|g'(a)|^k} \, da,$$

where C_5 only depends on f and k. This proves the lemma for case 2. ■

6.4 Asymptotic normality of the L_k-error

We will apply Lemma 6.3.1, to the following cut-off version of $\hat{f}_n$:

$$\tilde{f}_n(t) = \begin{cases} f(0) & \text{if } \hat{f}_n(x) \geq f(0), \\ \hat{f}_n(x) & \text{if } f(1) \leq \hat{f}_n(x) < f(0), \\ f(1) & \text{if } \hat{f}_n(x) < f(1). \end{cases} \tag{6.4.1}$$

The next lemma shows that $\tilde{f}_n$ satisfies condition (6.3.1) with probability tending to one.

Lemma 6.4.1 *Define the event*

$$A_n = \left\{ \sup_{x \in [0,1]} \left| \tilde{f}_n(x) - f(x) \right| \leq \frac{\inf_{x \in [0,1]} |f'(x)|^2}{2 \sup_{t \in [0,1]} |f''(x)|} \right\}.$$

Then $P\{A_n^c\} \to 0$.

Proof: It is sufficient to show that $\sup |\tilde{f}_n(x) - f(x)|$ tends to zero. For this we can follow the line of reasoning in Section 5.4 in Groeneboom and Wellner (1992). Similar to their Lemma 5.9 we derive from our Lemma 6.2.1 that for each $a \in (f(1), f(0))$,

$$P(|U_n(a) - g(a)| \geq n^{-1/3} \log n) \leq C_1 \exp\{-C_2 (\log n)^3\}.$$

This, in turn implies, by monotonicity of U_n and the conditions of f that there exists a constant $C_3 > 0$ such that

$$P\left(\sup_{a\in(f(1),f(0))} |U_n(a) - g(a)| \geq C_3 n^{-1/3}\log n\right) \leq C_1 \exp\{-\tfrac{1}{2}C_2(\log n)^3\}.$$

This implies that the maximum distance between successive points of jump of $\hat{f}_n$ is of the order $\mathcal{O}(n^{-1/3}\log n)$. Since both $\tilde{f}_n$ and f are monotone and bounded by $f(0)$, this also means that the maximum distance between $\tilde{f}_n$ and f is of the order $\mathcal{O}(n^{-1/3}\log n)$. ■

The difference between the L_k-errors for $\hat{f}_n$ and $\tilde{f}_n$ is bounded as follows

$$\begin{aligned}\left|\int_0^1 |\hat{f}_n(x) - f(x)|^k\,dx - \int_0^1 |\tilde{f}_n(x) - f(x)|^k\,dx\right| \\ \leq \int_0^{U_n(f(0))} |\hat{f}_n(x) - f(x)|^k\,dx + \int_{U_n(f(1))}^1 |\hat{f}_n(x) - f(x)|^k\,dx.\end{aligned} \qquad (6.4.2)$$

The next lemma shows that the integrals on the right hand side are of negligible order.

Lemma 6.4.2 *Let U_n be defined in (6.1.1). Then*

$$\int_0^{U_n(f(0))} |\hat{f}_n(x) - f(x)|^k\,dx = o_p(n^{-\frac{2k+1}{6}}),$$

and

$$\int_{U_n(f(1))}^1 |\hat{f}_n(x) - f(x)|^k\,dx = o_p(n^{-\frac{2k+1}{6}}).$$

Proof: Consider the first integral, which can be bounded by

$$\begin{aligned}2^k\int_0^{U_n(f(0))} |\hat{f}_n(x) - f(0)|^k\,dx + 2^k\int_0^{U_n(f(0))} |f(x) - f(0)|^k\,dx \\ \leq 2^k\int_0^{U_n(f(0))} |\hat{f}_n(x) - f(0)|^k\,dx + \frac{2^k}{k+1}\sup|f'|^k U_n(f(0))^{k+1}.\end{aligned} \qquad (6.4.3)$$

Define the event $B_n = \{U_n(f(0)) \leq n^{-1/3}\log n\}$. Then $U_n(f(0))^{k+1}1_{B_n} = o_p(n^{-(2k+1)/6})$. Moreover, according to the Theorem 2.1 in GROENEBOOM ET AL.(1999) it follows that $P\{B_n^c\} \to 0$. Since for any $\eta > 0$,

$$P\left(n^{\frac{2k+1}{6}}|U_n(f(0))|^{k+1}1_{B_n^c} > \eta\right) \leq P(B_n^c) \to 0,$$

this implies that the second term in (6.4.3) is of the order $o_p(n^{-(2k+1)/6})$. The first term in (6.4.3) can be written as

$$2^k\left(\int_0^{U_n(f(0))} |\hat{f}_n(x) - f(0)|^k\,dx\right)1_{B_n} + 2^k\left(\int_0^{U_n(f(0))} |\hat{f}_n(x) - f(0)|^k\,dx\right)1_{B_n^c}, \qquad (6.4.4)$$

where the second integral is of the order $o_p(n^{-(2k+1)/6})$ by the same reasoning as before. To bound the first integral in (6.4.4), we will construct a suitable sequence $(a_i)_{i=1}^m$, such that the intervals $(0, n^{-a_1}]$ and $(n^{-a_i}, n^{-a_{i+1}}]$, for $i = 1, 2, \ldots, m-1$, cover the interval $(0, U_n(f(0))]$, and such that the integrals over these intervals can be bounded appropriately. First of all let

$$1 > a_1 > a_2 > \cdots > a_{m-1} \geq 1/3 > a_m, \tag{6.4.5}$$

and let $z_0 = 0$ and $z_i = n^{-a_i}$, $i = 1, \ldots, m$, so that $0 < z_1 < \cdots < z_{m-1} \leq n^{-1/3} < z_m$. On the event B_n, for n sufficiently large, the intervals $(0, n^{-a_1}]$ and $(n^{-a_i}, n^{-a_{i+1}}]$ cover $(0, U_n(f(0))]$. Hence, when we denote $J_i = [z_i \wedge U_n(f(0)),\ z_{i+1} \wedge U_n(f(0))]$, the first integral in (6.4.4) can be bounded by

$$\sum_{i=0}^{m-1} \left(\int_{J_i} \left(\hat{f}_n(x) - f(0) \right)^k dx \right) 1_{B_n} \leq \sum_{i=0}^{m-1} (z_{i+1} - z_i) |\hat{f}_n(z_i) - f(0)|^k,$$

using that $\hat{f}_n$ is decreasing and the fact that $J_i \subset (0, U_n(f(0))]$, so that $\hat{f}_n(z_i) - f(0) \geq \hat{f}_n(x) - f(0) \geq 0$ for $x \in J_i$. It remains to show that

$$\sum_{i=0}^{m-1} (z_{i+1} - z_i) |\hat{f}_n(z_i) - f(0)|^k = o_p(n^{-\frac{2k+1}{6}}). \tag{6.4.6}$$

From Woodroofe and Sun (1993), we have that

$$\hat{f}_n(0) \xrightarrow{\mathcal{D}} f(0) \sup_{1 \leq j < \infty} \frac{j}{\Gamma_j}, \tag{6.4.7}$$

where Γ_j are partial sums of standard exponential random variables. Therefore

$$z_1 |\hat{f}_n(0) - f(0)|^k = O_p(n^{-a_1}). \tag{6.4.8}$$

Since for any $i = 1, \ldots, m-1$ we have that $a_i \geq 1/3$, from Theorem 3.1 in Kulikov and Lopuhaä (2002), it follows that

$$|\hat{f}_n(z_i) - f(0)| \leq |\hat{f}_n(z_i) - f(z_i)| + \sup |f'| z_i = \mathcal{O}_p(n^{-(1-a_i)/2}) + \mathcal{O}_p(n^{-a_i}) = \mathcal{O}_p(n^{-(1-a_i)/2}).$$

This implies that for $i = 1, \ldots, m-1$,

$$(z_{i+1} - z_i) |\hat{f}_n(z_i) - f(0)|^k = \mathcal{O}_p(n^{-a_{i+1} - k(1-a_i)/2}). \tag{6.4.9}$$

Therefore, if we can construct a sequence (a_i) satisfying (6.4.5), as well as

$$a_1 > \frac{2k+1}{6}, \tag{6.4.10}$$

$$a_{i+1} + \frac{k(1-a_i)}{2} > \frac{2k+1}{6}, \quad \text{for all } i = 1, \ldots, m-1, \tag{6.4.11}$$

then (6.4.6) follows from (6.4.8) and (6.4.9). One may take

$$\begin{aligned} a_1 &= \frac{2k+7}{12} \\ a_{i+1} &= \frac{k(a_i - 1)}{2} + \frac{2k+3}{8}, \quad \text{for } i = 1, \ldots, m-1. \end{aligned}$$

Note that the left hand side of (6.4.11) is larger than $(2k+3)/8$. Therefore, since $k < 2.5$, it immediately follows that (6.4.10) and (6.4.11) are satisfied. To show that (6.4.5) holds, first note that $1 > a_1 > 1/3$, because $k < 2.5$. It remains to show that the described sequence strictly decreases and reaches $1/3$ in finitely many steps. As long as $a_i > 1/3$, it follows that

$$a_i - a_{i+1} = \frac{2-k}{2}a_i + \frac{2k-3}{8}.$$

When $k = 2$, this equals $1/8$. For $1 \leq k < 2$, use that $a_i > 1/3$, to find that $a_i - a_{i+1} > 1/24$, and for $2 \leq k < 2.5$, use that $a_i \leq a_1 = (2k+1)/7$, to find that $a_i - a_{i+1} > (k+1)(2.5-k)/12$. This means that the sequence (a_i) also satisfies (6.4.5), which proves (6.4.6).

Similar to (6.4.3), the second integral can be bounded by

$$2^k \int_{U_n(f(1))}^1 |f(1) - \hat{f}_n(x)|^k \, dx + \frac{2^k}{k+1} \sup |f'|^k (1 - U_n(f(1))^{k+1}.$$

From here the proof is similar as above. We can use the same sequence (a_i) as before, and take $B_n = \{1 - U_n(f(1)) \leq n^{-1/3} \log n\}$. If we now define $z_0 = 1$, $z_i = 1 - n^{-a_i}$, for $i = 1, 2, \ldots, m-1$, then similar to the argument above, we are left with considering

$$\left(\int_{U_n(f(1))}^1 |f(1) - \hat{f}_n(x)|^k \, dx \right) 1_{B_n} \leq \sum_{i=0}^{m-1} (z_i - z_{i+1}) |f(1) - \hat{f}_n(z_i)|^k. \tag{6.4.12}$$

The first term is

$$\left(1 - (1 - n^{-a_1})\right) |f(1) - \hat{f}_n(1)|^k = n^{-a_1} f(1)^k,$$

and according to Theorem 4.1 in KULIKOV AND LOPUHAÄ (2002), it follows that for $i = 1, 2, \ldots, m-1$ each term is $\mathcal{O}_p(n^{-a_{i+1} - k(1-a_i)/2})$. As before the sequence (a_i) chosen above satisfies (6.4.10) and (6.4.11), which implies that (6.4.12) is $o_p(n^{-(2k+1)/6})$. This proves the lemma.

■

We are now able to prove our main result concerning the asymptotic normality of the L_k-error, for $1 \leq k < 2.5$.

Proof of Theorem 6.1.1: First consider the difference

$$\left| \int_0^1 |\hat{f}_n(x) - f(x)|^k \, dx - \int_{f(1)}^{f(0)} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}} \, da \right|, \tag{6.4.13}$$

which can be bounded by

$$\left| \int_0^1 |\hat{f}_n(x) - f(x)|^k \, dx - \int_0^1 |\tilde{f}_n(x) - f(x)|^k \, dx \right| + R_n, \tag{6.4.14}$$

where

$$R_n = \left| \int_0^1 |\tilde{f}_n(x) - f(x)|^k \, dx - \int_{f(1)}^{f(0)} \frac{|U_n(a) - g(a)|^k}{|g'(t)|^{k-1}} \, da \right|.$$

Let A_n be the event defined in Lemma 6.4.1, so that $P\{A_n^c\} \to 0$. As in the proof of Lemma 6.4.2, this means that $R_n 1_{A_n^c} = o_p(n^{-(2k+1)/6})$. Note that on the event A_n, the function $\tilde{f}_n$ satisfies the conditions of Lemma 6.3.1, and that for any $a \in [f(1), f(0)]$,

$$U_n(a) = \sup\{t \in [0,1] : \hat{f}_n(t) > a\} = \sup\{t \in [0,1] : \tilde{f}_n(t) > a\} = \tilde{U}_n(a).$$

Moreover, we can construct a partition $[0, s_1], (s_1, s_2], \ldots, (s_l, 1]$ of $[0,1]$ in such a way that on each element of the partition, $\tilde{f}_n$ satisfies either condition 1 or condition 2 of Lemma 6.3.1. This means that we can apply Lemma 6.3.1 to each element of the partition. Putting things together, it follows that $R_n 1_{A_n}$ is bounded from above by

$$C \int_{f(1)}^{f(0)} \frac{|U_n(a) - g(a)|^{k+1}}{|g'(a)|^k} \, da.$$

Corollary 6.2.1 implies that this integral is of order $\mathcal{O}_p(n^{-(k+1)/3})$, so that $R_n 1_{A_n} = o_p(n^{-(2k+1)/6})$. Finally, the first difference in (6.4.14) can be bounded as in (6.4.2), which means that according to Lemma 6.4.2 it is of the order $o_p(n^{-(2k+1)/6})$. Together with Corollary 6.2.1, this implies that

$$n^{1/6}\left(n^{k/3}\int_0^1 |\hat{f}_n(x) - f(x)|^k \, dx - \mu_k^k\right) \to N(0, \sigma^2),$$

where σ^2 is defined in Theorem 6.2.1. An application of the δ-method then yields that

$$n^{1/6}\left(n^{1/3}\left(\int_0^1 |\hat{f}_n(x) - f(x)|^k dx\right)^{1/k} - \mu_k\right)$$

converges to a normal random variable with mean zero and variance

$$\left\{\frac{1}{k}\left(\mu_k^k\right)^{1/k-1}\right\}^2 \sigma^2 = \frac{\sigma^2}{k^2 \mu_k^{2k-2}} = \sigma_k^2. \quad \blacksquare$$

6.5 Asymptotic normality of a modified L_k-error for a large k

For $k \geq 2.5$ the result of Theorem 6.1.1 does not hold. We will use the next lemma to show that the order of variance of the standard L_k integral is too big then.

Lemma 6.5.1 *Let $k \geq 2.5$ and $z_n = 1/(2nf(0))$. Then there are a_1, b_1, a_2 and b_2 such that $0 < a_1 < b_1 < a_2 < b_2 < \infty$ and for $i = 1, 2$*

$$\liminf_{n\to\infty} P\left\{n \int_0^{z_n} \left|\hat{f}_n(x) - f(x)\right|^k dx \in [a_i, b_i]\right\} > 0.$$

Proof: Let the event $A_n = \{X_{(1)} \geq z_n\}$, so that $P\{A_n\} \to 1/\sqrt{e} > 1/2$. Then, since $a_i > 0$ and since under A_n the function $\hat{f}_n(x)$ is constant on the interval $[0, z_n]$,

$$\begin{aligned} &P\left\{n \int_0^{z_n} \left|\hat{f}_n(x) - f(x)\right|^k dx \in [a_i, b_i]\right\} \\ &\qquad \geq P\left\{\left(n \int_0^{z_n} \left|\hat{f}_n(0) - f(x)\right|^k dx\right) 1_{A_n} \in [a_i, b_i]\right\} \\ &\qquad = P\left\{\left(\frac{1}{2f(0)} \left|\hat{f}_n(0) - f(0)\right|^k + R_n\right) 1_{A_n} \in [a_i, b_i]\right\}, \end{aligned} \tag{6.5.1}$$

where

$$R_n = kn \int_0^{z_n} \theta^{k-1} \left(\left|\hat{f}_n(0) - f(x)\right| - \left|\hat{f}_n(0) - f(0)\right|\right) dx \tag{6.5.2}$$

and θ_n lies between $\left|\hat{f}_n(0) - f(x)\right|$ and $\left|\hat{f}_n(0) - f(0)\right|$. Using (6.4.7) we obtain that (6.5.2) is of the order $\mathcal{O}_p(n^{-1})$ and therefore

$$\frac{1}{2f(0)} \left|\hat{f}_n(0) - f(0)\right|^k + R_n \to \frac{1}{2} f(0)^{k-1} \left|\sup_{j \geq 1} \frac{j}{\Gamma_j} - 1\right|^k$$

in distribution.

Let us choose a_i, b_i, $i = 1, 2$ so that

$$P\left\{\frac{1}{2} f(0)^{k-1} \left|\sup_{j \geq 1} \frac{j}{\Gamma_j} - 1\right|^k \in [a_i, b_i]\right\} > (1 - 1/\sqrt{e}).$$

Then (6.5.1) is bounded from below by

$$P\left\{\left(\frac{1}{2f(0)} \left|\hat{f}_n(0) - f(0)\right|^k + R_n\right) \in [a_i, b_i]\right\} - P\{A_n^c\}$$

which converges to a positive value. ∎

Application of this lemma implies that

$$\operatorname{var}\left(n \int_0^{z_n} \left|\hat{f}_n(x) - f(x)\right|^k\right) \geq \frac{1}{4}(a_2 - b_1)^2 \inf_{i=1,2} P\left\{n \int_0^{z_n} \left|\hat{f}_n(x) - f(x)\right|^k \in [a_i, b_i]\right\}$$

asymptotically will be bounded away from zero. Therefore to have a finite variance of the L_k-error, other scaling is required. The usual L_k-error can be modified so that a result similar to Theorem 6.1.1 will hold. For $k \geq 2.5$ we will consider a modified L_k-error of the form

$$\left(\int_{n^{-\epsilon}}^{1-n^{-\epsilon}} \left|\hat{f}_n(x) - f(x)\right|^k dx\right)^{1/k},$$

where $1/6 < \epsilon < (k-1)/(3(k-2))$. In this way we avoid a region where the Grenander estimator is inconsistent and we are still able to determine its global performance. In order to prove a result analogous to Theorem 6.1.1, we define another cut-off version of the Grenander estimator:

$$f_n^\epsilon(x) = \begin{cases} f(n^{-\epsilon}) & \text{if } \hat{f}_n(x) \geq f(n^{-\epsilon}), \\ \hat{f}_n(x) & \text{if } f(1-n^{-\epsilon}) \leq \hat{f}_n(x) < f(n^{-\epsilon}), \\ f(1-n^{-\epsilon}) & \text{if } \hat{f}_n(x) < f(1-n^{-\epsilon}), \end{cases}$$

and its inverse function

$$U_n^\epsilon(a) = \sup\left\{x \in [n^{-\epsilon}, 1-n^{-\epsilon}] : \hat{f}_n(x) \geq a\right\}, \tag{6.5.3}$$

for $a \in [f(1-n^{-\epsilon}), f(n^{-\epsilon})]$. The next lemma is the analogue of Lemma 6.4.1.

Lemma 6.5.2 *Define the event*

$$A_n^\epsilon = \left\{\sup_{x\in[0,1]} |f_n^\epsilon(x) - f(x)| \leq \frac{\inf_{x\in[0,1]} |f'(x)|^2}{2\sup_{t\in[0,1]} |f''(x)|}\right\}.$$

Then $P(A_n^\epsilon) \to 1$.

Proof: Its suffices to show that $\sup_{x\in[0,1]} |f_n^\epsilon(x) - f(x)| \to 0$. Using the definition of f_n^ϵ we can bound

$$\sup_{x\in[0,1]} |f_n^\epsilon(x) - f(x)| \leq \sup_{x\in[0,1]} \left|f_n^\epsilon(x) - \tilde{f}_n(x)\right| + \sup_{x\in[0,1]} \left|\tilde{f}_n(x) - f(x)\right|, \tag{6.5.4}$$

where $\tilde{f}_n$ is defined in (6.4.1). The first term on the right hand side of (6.5.4) is smaller than $\sup|f'|n^{-\epsilon}$ which, together with Lemma 6.4.1, implies that $\sup_{x\in[0,1]} |f_n^\epsilon(x) - f(x)| = o_p(n^{-1/6})$. ∎

Similar to (6.4.2), the difference between the modified L_k-errors for $\hat{f}_n$ and f_n^ϵ is bounded as follows

$$\begin{aligned} &\left|\int_{n^{-\epsilon}}^{1-n^{-\epsilon}} |\hat{f}_n(x) - f(x)|^k\,dx - \int_{n^{-\epsilon}}^{1-n^{-\epsilon}} |f_n^\epsilon(x) - f(x)|^k\,dx\right| \\ &\qquad \leq \int_{n^{-\epsilon}}^{U_n^\epsilon(f(n^{-\epsilon}))} |\hat{f}_n(x) - f(x)|^k\,dx + \int_{U_n^\epsilon(f(1-n^{-\epsilon}))}^{1-n^{-\epsilon}} |\hat{f}_n(x) - f(x)|^k\,dx. \end{aligned} \tag{6.5.5}$$

The next lemma is the analogue of Lemma 6.4.2 and shows that both integrals on the right hand side are of negligible order.

Lemma 6.5.3 *Let U_n^ϵ be defined in (6.5.3). Then*

$$\int_{n^{-\epsilon}}^{U_n^\epsilon(f(n^{-\epsilon}))} |\hat{f}_n(x) - f(x)|^k\,dx = o_p(n^{-\frac{2k+1}{6}}),$$

and

$$\int_{U_n^\epsilon(f(1-n^{-\epsilon}))}^{1-n^{-\epsilon}} |\hat{f}_n(x) - f(x)|^k\,dx = o_p(n^{-\frac{2k+1}{6}}).$$

Proof: Consider the first integral, then similar to (6.4.3) we have that

$$\begin{aligned} &2^k \int_{n^{-\epsilon}}^{U_n^\epsilon(f(n^{-\epsilon}))} |\hat{f}_n(x) - f(n^{-\epsilon})|^k \, dx + 2^k \int_{n^{-\epsilon}}^{U_n^\epsilon(f(n^{-\epsilon}))} |f(n^{-\epsilon}) - f(x)|^k \, dx \\ &\leq 2^k \int_{n^{-\epsilon}}^{U_n^\epsilon(f(n^{-\epsilon}))} |\hat{f}_n(x) - f(n^{-\epsilon})|^k \, dx + \frac{2^k}{k+1} \sup |f'|^k (U_n^\epsilon(f(n^{-\epsilon})) - n^{-\epsilon})^{k+1}. \end{aligned} \tag{6.5.6}$$

If we define the event $B_n^\epsilon = \{U_n^\epsilon(f(n^{-\epsilon})) - n^{-\epsilon} \leq n^{-1/3} \log n\}$, then by a similar reasoning as in the proof of Lemma 6.4.2, it follows that $(U_n^\epsilon(f(n^{-\epsilon})) - n^{-\epsilon})^{k+1} = o_p(n^{-(2k+1)/6})$. The first integral on the right hand side of (6.5.6) can be written as

$$\left(\int_{n^{-\epsilon}}^{U_n^\epsilon(f(n^{-\epsilon}))} |\hat{f}_n(x) - f(n^{-\epsilon})|^k \, dx \right) 1_{B_n} + \left(\int_{n^{-\epsilon}}^{U_n^\epsilon(f(n^{-\epsilon}))} |\hat{f}_n(x) - f(n^{-\epsilon})|^k \, dx \right) 1_{B_n^c},$$

where the second term is of the order $o_p(n^{-\frac{2k+1}{6}})$ by the same reasoning as before. To bound

$$\left(\int_{n^{-\epsilon}}^{U_n^\epsilon(f(n^{-\epsilon}))} |\hat{f}_n(x) - f(n^{-\epsilon})|^k \, dx \right) 1_{B_n}, \tag{6.5.7}$$

we distinguish between two cases:

(i) $1/6 < \epsilon \leq 1/3$,

(ii) $1/3 < \epsilon < (k-1)/(3k-6)$.

In case (i), the integral (6.5.7) can be bounded by $|\hat{f}_n(n^{-\epsilon}) - f(n^{-\epsilon})|^k n^{-1/3} \log n$. Since $1/6 < \epsilon \leq 1/3$, according to Theorem 3.1 in Kulikov and Lopuhaä (2002), it follows that $|\hat{f}_n(n^{-\epsilon}) - f(n^{-\epsilon})| = \mathcal{O}_p(n^{-1/3})$ and therefore (6.5.7) is of the order $o_p(n^{-(2k+1)/6})$.

In case (ii), similar to Lemma 6.4.2, we will construct a suitable sequence $(a_i)_{i=1}^m$, such that the intervals $(n^{-a_i}, n^{-a_{i+1}}]$, for $i = 1, 2, \ldots, m-1$ cover the interval $(n^{-\epsilon}, U_n(f(n^{-\epsilon}))]$, and such that the integrals over these intervals can be bounded appropriately. First of all let

$$\epsilon = a_1 > a_2 > \cdots > a_{m-1} \geq 1/3 > a_m, \tag{6.5.8}$$

and let $z_i = n^{-a_i}$, $i = 1, \ldots, m$, so that $0 < z_1 < \cdots < z_{m-1} \leq n^{-1/3} < z_m$. Then, similar to the proof of Lemma 6.4.2, we can bound (6.5.7) as follows

$$\left(\int_{n^{-\epsilon}}^{U_n^\epsilon(f(n^{-\epsilon}))} |\hat{f}_n(x) - f(n^{-\epsilon})|^k \, dx \right) 1_{B_n} \leq \sum_{i=1}^{m-1} (z_{i+1} - z_i) |\hat{f}_n(z_i) - f(n^{-\epsilon})|^k.$$

For $k > 2.5$, we have $1/3 \leq a_i \leq \epsilon < 1$, for $i = 1, \ldots, m-1$. This means that we can apply Theorem 3.1 in Kulikov and Lopuhaä (2002) to each term and conclude each term is of order $\mathcal{O}_p(n^{-(1-a_i)/2})$. Therefore, it suffices to construct a sequence (a_i) satisfying (6.5.8), as well as

$$a_{i+1} + \frac{k(1-a_i)}{2} > \frac{2k+1}{6}, \quad \text{for all } i = 1, \ldots, m-1. \tag{6.5.9}$$

One may take

$$\begin{aligned} a_1 &= \epsilon \\ a_{i+1} &= \frac{k(a_i-1)}{2} + \frac{2k+1}{6} + \frac{1}{8}\left(\frac{k-1}{3(k-2)} - \epsilon\right), \qquad \text{for } i = 1, \dots, m-1. \end{aligned}$$

Then (6.5.9) is satisfied and it remains to show that the described sequence strictly decreases and reaches 1/3 in finitely many steps. This follows from the fact that as soon as $a_i \leq \epsilon$, for $k > 2.5$ we have:

$$a_i - a_{i+1} = \frac{k-2}{2}\left(\frac{k-1}{3(k-2)} - a_i\right) - \frac{1}{8}\left(\frac{k-1}{3(k-2)} - \epsilon\right) \geq \frac{4k-9}{8}\left(\frac{k-1}{3(k-2)} - \epsilon\right) > 0.$$

As in the proof of Lemma 6.4.2 the argument for the second integral is similar. Now take $B_n^\epsilon = \{1 - n^{-\epsilon} - U_n^\epsilon(f(1-n^{-\epsilon})) \leq n^{-1/3}\log n\}$. The case $1/6 < \epsilon \leq 1/3$ can be treated in the same way as before. For the case $1/3 < \epsilon < \frac{k-1}{3(k-2)}$, we can use the same sequence (a_i) as above, but now define $z_i = 1 - n^{-a_i}$, $i = 1, \dots, m$, so that $1 > z_1 > \dots > z_{m-1} \geq 1 - n^{-1/3} > z_m$. Then we are left with considering

$$\left(\int_{U_n^\epsilon(f(1-n^{-\epsilon}))}^{1-n^{-\epsilon}} |f(1-n^{-\epsilon}) - \hat{f}_n(x)|^k \, dx\right) 1_{B_n} \leq \sum_{i=1}^{m-1} (z_i - z_{i+1}) |f(1-n^{-\epsilon}) - \hat{f}_n(z_i)|^k.$$

As before, each term in the sum is of the order $\mathcal{O}_p\left(n^{-a_{i+1} - k(1-a_i)/2}\right)$, for $i = 1, \dots, m-1$. The sequence chosen above satisfies (6.5.9) and (6.5.8), which implies that the sum above is of order $o_p(n^{-(2k+1)/6})$. ■

Apart from (6.5.5) we also need to bound the difference between integrals for U_n and its cut-off version U_n^ϵ:

$$\begin{aligned} &\left| \int_{f(1)}^{f(0)} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}} \, da - \int_{f(1-n^{-\epsilon})}^{f(n^{-\epsilon})} \frac{|U_n^\epsilon(a) - g(a)|^k}{|g'(a)|^{k-1}} \, da \right| \qquad (6.5.10) \\ &\quad \leq \int_{\hat{f}_n(n^{-\epsilon})}^{f(0)} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}} \, da + \int_{f(1)}^{\hat{f}_n(1-n^{-\epsilon})} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}} \, da. \end{aligned}$$

The next lemma shows that both integrals on the right hand side are of negligible order.

Lemma 6.5.4 *Let $1/6 < \epsilon < \frac{k-1}{3(k-2)}$. Furthermore let U_n be defined in (6.1.1) and let $\tilde{f}_n$ be defined in (6.4.1). Then*

$$\int_{\tilde{f}_n(n^{-\epsilon})}^{f(0)} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}} da = o_p(n^{-\frac{2k+1}{6}}),$$

and

$$\int_{f(1)}^{\tilde{f}_n(1-n^{-\epsilon})} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}} da = o_p(n^{-\frac{2k+1}{6}}).$$

Proof: Consider the first integral and define the event $A_n = \{f(0) - \tilde{f}_n(n^{-\epsilon}) < n^{-1/6}/\log n\}$. For $1/6 < \epsilon \leq 1/3$, according to Theorem 3.1 we have that

$$\begin{aligned} f(0) - \tilde{f}_n(n^{-\epsilon}) \leq |\hat{f}_n(n^{-\epsilon}) - f(0)| &\leq |\hat{f}_n(n^{-\epsilon}) - f(n^{-\epsilon})| + \sup|f'|\, n^{-\epsilon} \\ &= \mathcal{O}_p(n^{-1/3}) + \mathcal{O}(n^{-\epsilon}) = o_p(n^{-1/6}/\log n). \end{aligned}$$

This means that if $1/6 < \epsilon \leq 1/3$, the probability $P\{A_n^c\} \to 0$. For $1/3 < \epsilon < 1$,

$$P\{A_n^c\} \leq P(f(0) - \tilde{f}_n(n^{-\epsilon}) > 0) \leq P\left(\hat{f}_n(n^{-\epsilon}) - f(n^{-\epsilon}) < n^{-\epsilon}\sup|f'|\right) \to 0,$$

since according to Theorem 3.1 in Kulikov and Lopuhaä (2002), $\hat{f}_n(n^{-\epsilon}) - f(n^{-\epsilon})$ is of order $n^{-(1-\epsilon)/2}$. Next, write the first integral as

$$\left(\int_{\tilde{f}_n(n^{-\epsilon})}^{f(0)} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}}\, da\right) 1_{A_n} + \left(\int_{\tilde{f}_n(n^{-\epsilon})}^{f(0)} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}}\, da\right) 1_{A_n^c}. \tag{6.5.11}$$

Similar to the argument used in Lemma 6.4.2, the second integral in (6.5.11) is of the order $o_p(n^{-\frac{2k+1}{6}})$. The expectation of the first integral is bounded by

$$\begin{aligned} E\int_{f(0)-n^{-1/6}/\log n}^{f(0)} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}}\, da &\leq n^{-k/3} C_1 \int_{f(0)-n^{-1/6}/\log n}^{f(0)} E\left|V_n^E(a)\right|^k da \\ &\leq C_2 n^{-\frac{2k+1}{6}}/\log n, \end{aligned}$$

using Lemma 6.2.1. The Markov inequality implies that the first term in (6.5.11) is of the order $o_p(n^{-\frac{2k+1}{6}})$. For the second integral the proof is similar. ■

Theorem 6.5.1 *Suppose conditions (A1) - (A3) of Theorem 6.1.1 are satisfied. Then for* $k \geq 2.5$ *and for any* ϵ*, such that* $1/6 < \epsilon < \frac{k-1}{3(k-2)}$,

$$n^{1/6}\left\{n^{1/3}\left(\int_{n^{-\epsilon}}^{1-n^{-\epsilon}} |\hat{f}_n(x) - f(x)|^k\, dx\right)^{1/k} - \mu_k\right\}$$

converges in distribution to a normal random variable with zero mean and variance σ_k^2*, where* μ_k *and* σ_k^2 *are defined in Theorem 6.1.1.*

Proof: As in the proof of Theorem 6.1.1, it suffices to show that the difference

$$\left|\int_{n^{-\epsilon}}^{1-n^{-\epsilon}} |\hat{f}_n(x) - f(x)|^k\, dx - \int_{f(1)}^{f(0)} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}}\, da\right|$$

is of the order $o_p(n^{-(2k+1)/6})$. We can bound this difference by

$$\left|\int_{n^{-\epsilon}}^{1-n^{-\epsilon}} |\hat{f}_n(x) - f(x)|^k\, dx - \int_{n^{-\epsilon}}^{1-n^{-\epsilon}} |f_n^\epsilon(x) - f(x)|^k\, dx\right| \tag{6.5.12}$$

$$+\left|\int_{f(1)}^{f(0)} \frac{|U_n(a) - g(a)|^k}{|g'(a)|^{k-1}}\, da - \int_{f(1-n^{-\epsilon})}^{f(n^{-\epsilon})} \frac{|U_n^\epsilon(a) - g(a)|^k}{|g'(a)|^{k-1}}\, da\right| \tag{6.5.13}$$

$$+\left|\int_{n^{-\epsilon}}^{1-n^{-\epsilon}} |f_n^\epsilon(x) - f(x)|^k\, dx - \int_{f(1-n^{-\epsilon})}^{f(n^{-\epsilon})} \frac{|U_n^\epsilon(a) - g(t)|^k}{|g'(a)|^{k-1}}\, da\right|. \tag{6.5.14}$$

Differences (6.5.12) and (6.5.13) can be bounded as in (6.5.5) and (6.5.10), so that Lemmas 6.5.3 and 6.5.4 imply that these terms are of the order $o_p(n^{-\frac{2k+1}{6}})$. Finally, Lemma 6.3.1 implies that (6.5.14) is bounded by

$$\int_{f(1-n^{-\epsilon})}^{f(n^{-\epsilon})} \frac{|U_n^\epsilon(a) - g(a)|^{k+1}}{|g'(a)|^k} \, da.$$

Write the integral as

$$\int_{f(1)}^{f(0)} \frac{|U_n(a) - g(a)|^{k+1}}{|g'(a)|^k} \, da + \left(\int_{f(1)}^{f(0)} \frac{|U_n(a) - g(a)|^{k+1}}{|g'(a)|^k} da - \int_{f(1-n^{-\epsilon})}^{f(n^{-\epsilon})} \frac{|U_n^\epsilon(a) - g(a)|^{k+1}}{|g'(a)|^k} da \right).$$

Then Corollary 6.2.1 and Lemma 6.5.4 imply that both terms are of the order $o_p(n^{-\frac{2k+1}{6}})$. This proves the theorem. ■

Bibliography

BILLINGSLEY, P. (1968) *Weak convergence of probablity measures*, Wiley, New York.

BICKEL, P.J., KLAASSEN, C.A.J., RITOV, Y., WELLNER, J.A. (1993) *Efficient and adaptive estimation for semiparametric models*, Springer-Verlag.

BIRMAN, M.Š. AND SOLOMJAK, M.Z. (1967) Piecewise-polynomial approximations of functions of the classes W_p^α, *Math.USSR-Sbornik* **2**, No. 3.

DUDLEY, R.M. (1984) A course on empirical processes, *Lecture notes in mathematics*, **1097**, p.2–141, Springer-Verlag.

DÜMBGEN, L., FREITAG, S., AND JONGBLOED, G. (2002) Consistency of concave regression with an application to current-status data, *Report of the Department of Stochastics*, Vrije Universiteit Amsterdam.

DUROT, C. (2000) Sharp asymptotics for isotonic regression, *Probab. Theory and Related Fields* **122**, no. 2, p.222–240.

DUROT, C. AND TOCQUET, A.-S. (2002) On the distance between the empirical process and its concave majorant in a monotone regression framework, submitted to *Annales de l'I.H.P (serie B)*.

VAN DE GEER, S. (2000) *Empirical processes in M-estimation*, The Cambridge University Press.

GESKUS, R.B. AND GROENEBOOM, P. (1996) Asymptotically optimal estimation of smooth functionals for interval censoring, part 1, *Statistica Neerlandica* **50**, No.1, p.69–88.

GESKUS, R.B. AND GROENEBOOM, P. (1997) Asymptotically optimal estimation of smooth functionals for interval censoring, part 2, *Statistica Neerlandica* **51**, No.2, p.201–219.

GESKUS, R.B. AND GROENEBOOM, P. (1999) Asymptotically optimal estimation of smooth functionals for interval censoring, case 2, *Ann. Statist.* **27**, No. 2, p.627–674.

GRENANDER, U. (1956) On the theory of mortality measurements, Part II, *Skand. Akt. Tid.* **39**, p.125–153.

GROENEBOOM, P. (1985) Estimating a monotone density, In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (Edited by L.LeCam and R.Olshen) **2**, p.539–555.

GROENEBOOM, P. (1989) Brownian motion with a parabolic drift and Airy functions, *Z. Wahrsch. Verw. Gebiete* **81**, p.79–109.

GROENEBOOM, P. AND WELLNER, J.A. (1992) *Information bounds and nonparametric maximum likelihood estimation*, Birkhäuser Verlag.

GROENEBOOM, P. (1996) Lectures on probability theory and statistics, *Lecture Notes in Mathematics*, Springer-Verlag, Berlin **1648**, p.67–164.

GROENEBOOM, P., HOOGHIEMSTRA, G. AND LOPUHAÄ, H.P. (1999) Asymptotic normality of the L_1-error of the Grenander estimator, *Ann. Statist.* **27**, No.4, p.1316–1347.

GROENEBOOM, P., KULIKOV, V.N. AND LOPUHAÄ, H.P. (2002) A two sample test for current status data. *About to be submitted.*

GROENEBOOM, P., KULIKOV, V.N. AND LOPUHAÄ, H.P. (2002) A two sample likelihood ratio test for current status data. *About to be submitted.*

HAJEK, J., SIDAK, Z. AND SEN, P.K. (1999) *Theory of rank tests*, Academic Press.

HUANG, J. (1996) Efficient estimation for the proportional hazards model with interval censoring, *Ann. Statist.* **24**, No. 2, p.540–568.

IBRAGIMOV, I.A. AND LINNIK, Y.V. (1971) *Independent and stationary sequences of random variables*, Wolters-Nordhoff, Groningen.

JONGBLOED, G. (1995) The iterative convex minorant algorithm for non-parametric estimation, *J. Comput. Graph. Statist.* **7**, No. 3, p.310–321.

KIEFER, J. AND WOLFOWITZ, J. (1976) Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrsch. verw. Gebiete* **34**, p.73-85.

KIM, J. AND POLLARD, D. (1990) Cube root asymptotics, *Ann. Statist.* **18**, No.1, p.191–219.

KOMLOS, J., MAJOR, P. AND TUSNADY, G. (1975) An approximation of partial sums of independent rv'-s, and the sample DF. I, *Z. Wahrsch. Verw. Gebiete* **32**, p.111–131.

KULIKOV, V.N. AND LOPUHAÄ, H.P. (2002) The limit process of the difference between the empirical distribution function and its concave majorant. Submitted to *Ann. Statist.*

KULIKOV, V.N. AND LOPUHAÄ, H.P. (2002) The behavior of the NPMLE of a decreasing density near the boundaries of the support. Submitted to *Ann. Statist.*

KULIKOV, V.N. AND LOPUHAÄ, H.P. (2002) Asymptotic normality of the L_k-error of the Grenander estimator. *About to be submitted.*

MURPHY, S.A. AND VAN DER VAART, A.W. (1997) Semiparametric likelihood ratio inference, *Ann. Statist.* **25**, No. 4, p.1471–1509.

MURPHY, S.A. AND VAN DER VAART, A.W. (2000) On profile likelihood, *Journal of the American Statistical Association* **95**, No. 450, Theory and Methods, p.449–485.

PETO, R. AND PETO, J. (1972) Asymptotically efficient rank invariant test procedures (with discussion). *J. R. Stat. Soc. A.* **135**, p.185–207.

POLLARD, D. (1984) *Convergence of stochastic processes*, Springer-Verlag.

PRAKASA RAO, B.L.S. (1969) Estimation of a unimodal density, *Sankhya Ser.* A 31, p.23–36.

ROBERTSON, T., WRIGHT, F.T. AND DYKSTRA, R.L. (1988) *Order restricted statistical inference*, Wiley, New York.

SUN, J. AND WOODROOFE, M. (1996) Adaptive smoothing for a penalized NPMLE of a non-increasing density, *J. Statist. Planning and Inference* **52**, p.143–159.

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

van der Vaart, A.W. and Wellner, J.A. (1996) *Weak convergence and empirical process*, Springer-Verlag.

Wang, Y. (1994) The limit distribution of the concave majorant of an empirical distribution function. *Statistics and Probability Letters* **20**, p.81–84.

Woodroofe, M. and Sun, J. (1993) A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. *Statistica Sinica* **3**, p. 501–515.

Summary

Direct and Indirect Use of Maximum Likelihood

Maximum likelihood methods certainly belong to the most known and commonly used instruments for estimation and testing in statistics. But in contrast with one-parametric problems, where application of these methods is well-studied and usually does not cause any difficulties, in case of non-parametric or semi-parametric models it becomes more complicated. It may happen that the rate of convergence will differ from the usual $\sqrt{n}$ and the limiting distribution will differ from the normal one. Furthermore, when models with possible loss of data are considered, even applicability of the standard likelihood ratio testing procedures may depend on the family of alternatives in a quite sophisticated way. Finally, standard testing procedures based on maximum likelihood ideology may become computationally too complicated and other, modified procedures, combining advantages of the maximum likelihood and simplicity of other known methods should be established for the practical application.

In the first chapter of this thesis we review some known results on the application of maximum likelihood methods in case of a decreasing density model and in case of interval censored Cox model. Besides these particular results we also recall the usual instruments applied to studying the limiting distributions of certain functionals of the maximum likelihood estimator and to establish its asymptotic efficiency. Some notes on the implementation of the maximum likelihood in these settings are also provided.

In chapters 2 and 3 we will construct two-sample tests for testing that the distribution functions F_0 and F_1, generating the two samples, are equal, in the case that the samples are subject to current status censoring. We will test the null hypothesis $F_0 = F_1$ against Lehmann alternatives $F_1 = F_0^{1+\theta}$, for some $\theta > -1$ or against the proportional hazard alternatives. This problem has many applications, for example in medicine and biostatistics. Suppose we want to find out if some working or living conditions cause certain disease when this disease can only be found at later medical examination. Comparing two groups of people where members of the second group have encountered the factor of interest during the known period of time, the hypothesis of equal distributions of the time until a person from the first and from the second groups becomes ill can be tested using the established methods.

The test proposed in chapter 2 is based on the score function and the NPMLE under the null hypothesis. The test statistic is shown to converge at rate $\sqrt{n}$ and to be asymptotically normal both under the null hypothesis and under a contiguous alternative. Moreover, we show that the test is asymptotically efficient for testing against the alternative $F_1 = F_0^{1+\theta}$. This testing procedure is easy to implement, but the parameters of the limiting normal distribution depend on the underlying distributions. They must be estimated too, which certainly diminishes the accuracy of testing.

Therefore we also propose a test based on the likelihood ratio statistic. Under the null hypothesis this statistic is shown to be asymptotically chi-squared with one degree of freedom

and under a contiguous alternative the limiting distribution is that of a squared normally distributed random variable with non-zero mean. This test is also asymptotically efficient for testing against the alternative $F_1 = F_0^{1+\theta}$. Its implementation is more complicated since it includes a semi-parametric maximum likelihood estimation which takes much more calculations than non-parametric estimation under the null hypothesis. Pictures showing results of a short simulation study can be found at the end of chapters 2 and 3.

The most important instrument used in chapters 2 and 3 is a generalization of the chaining lemma and, in fact, of the uniform central limit theorem. Another important intermediate result that should be mentioned here is that although the semi-parametric MLE under the Lehmann alternative has a rate of convergence slower than $\sqrt{n}$, under a contiguous Lehmann alternative the MLE of the one-dimensional parameter θ converges at rate $\sqrt{n}$ and has an asymptotically normal distribution.

In the last three chapters of this dissertation we concentrate on another application of the maximum likelihood estimation. Properties of the NPMLE of an isotonic density, also known as the Grenander estimator, such as rate of convergence and pointwise convergence in distribution are well studied. We tried to answer some questions about applicability and performance of the method. First of all the Grenander estimator should be applied only if the underlying density is isotonic, therefore we need a testing procedure to test this null hypothesis against any non-isotonic alternative. In chapter 4 we consider the difference between the integrated Grenander estimator and the empirical distribution function under the hypothesis of a decreasing density. After re-scaling this process converges locally to a limiting process and furthermore the integral of this difference is shown to converge at rate $n^{2/3}$ and to be asymptotically normal after appropriate re-scaling. Choosing this as a test statistic we therefore can test the hypothesis that the underlying density is monotone decreasing.

Furthermore we study the behavior of the Grenander estimator near the boundary of support (suppose without loss of generality it is $[0, 1]$). Pointwise inconsistency of the Grenander estimator at these points is well known and there are some uniformly consistent modifications of the estimator proposed by other authors. In chapter 5 we show that the difference $(\hat{f}_n(n^{-\alpha}) - f(n^{-\alpha}))$ converges at rate $n^{(1-\alpha)/2}$ for $1/3 \leq \alpha < 1$ and at rate $n^{1/3}$ for $0 < \alpha < 1/3$. We also obtain the limiting distribution of this difference. That result is close in spirit and method of proof to pointwise convergence results for the Grenander estimator known before. Based on it we propose another uniformly consistent modification of the Grenander estimator, which combines small mean squared error and easy implementation. Result of the simulation study and comparing with the modification known before can be found at the end of chapter 5.

Another consequence of the properties established in chapter 5 is that the L_k-error of the Grenander estimator has $k = 2.5$ as a kind of transition point. For $k < 2.5$ this global measure of performance converges at rate $n^{1/3}$ and after certain re-scaling is asymptotically normal. On the other hand, for $k \geq 2.5$ the rate of convergence of the standard L_k-error is slower than $n^{1/3}$ and the limiting distribution differs from a normal distribution. This is explained by a greater contribution of integration near the boundaries of support and by the inconsistency of the Grenander estimator in these points. Nevertheless, we can modify the L_k-error in a way that has a close connection to the standard L_k-error of the modified estimator, which modification converges at rate $n^{1/3}$ and after certain re-scaling has normal limiting distribution. This generalizes the L_1-error result known before.

BY V.N.KULIKOV

Samenvatting

DIRECT EN INDIRECT GEBRUIK VAN MAXIMUM LIKELIHOOD

Ongetwijfeld behoren de maximum likelihood methodes tot de meest bekende en algemeen aanvaarde instrumenten voor het schatten en toetsen in de statistiek. Maar in tegenstelling tot één-parameter problemen, waarbij deze methodes volledig bestudeerd zijn en meestal geen moeilijkheden veroorzaken, is het geval van niet-parametrische of semi-parametrische modellen meer gecompliceerd. Het kan gebeuren dat de snelheid van convergentie van de gebruikelijke $\sqrt{n}$ zal verschillen en dat de limietverdeling zal verschillen van de normale verdeling. Bovendien, als modellen met eventueel verlies van de data worden beschouwd, kan zelfs toepasbaarheid van de standaard likelihood ratio toetsingsprocedure op een ingewikkelde manier van het alternatief afhangen. Ten slotte, standaard toetsingsprocedures, die op de maximum likelihood ideologie gebaseerd zijn, kunnen ook veel te moeilijk om uit te rekenen worden. Voor praktische toepassingen moeten andere, gemodificeerde procedures, die zoveel mogelijk voordelen van de maximum likelihood en eenvoud van andere bekende methodes combineren, ontwikkeld worden.

In het eerste hoofdstuk van dit proefschrift beschouwen we bekende resultaten van de maximum likelihood methode toegepast op modellen met een dalende dichtheid en ook op interval gecensureerde Cox modellen. Behalve deze resultaten beschouwen we ook gebruikelijke instrumenten, die toepassing hebben op het verkrijgen van de limietverdeling van verschillende functionalen van de maximum likelihood schatter en voor bewijs van zijn asymptotische efficiëntie. Er zijn ook enkele opmerkingen te vinden betreffende de implementatie van de maximum likelihood schatter in de voornoemde modellen.

In hoofdstukken 2 en 3 bestuderen wij het toetsen of twee verdelingsfuncties F_0 en F_1, waaruit twee steekproeven getrokken werden, gelijk zijn aan elkaar, in het geval dat deze steekproeven interval gecensureerd zijn. We zouden de nulhypothese $F_0 = F_1$ willen toetsen tegen het Lehmann alternatief $F_1 = F_0^{1+\theta}$, voor een $\theta > -1$, of tegen het proportional hazard model. Dit toetsingsprobleem heeft veel toepassingen, onder andere in de geneeskunde en de biostatistiek. Bijvoorbeeld, om te bepalen of een of andere ziekte beïnvloed wordt door bepaalde woon- of werkomstandigheden, als deze ziekte pas tijdens later medisch onderzoek wordt geconstateerd, kan de twee-steekproeven toets voor interval gecensureerde data gebruikt worden. In dit geval zullen we twee groepen vergelijken, waar mensen uit de tweede groep gedurende een bekende termijn met de factor, die we willen bestuderen, zijn geconfronteerd, en waar de hypothese van gelijke verdelingsfuncties van het tijdstrip waarop een persoon uit de eerste en uit de tweede groep ziek is geworden wordt getoetst door middel van de ontwikkelde methodes.

De in hoofdstuk 2 voorgestelde toets is gebaseerd op de score functies en op de NPMLE onder de nulhypothese. Wij tonen aan dat de toetsingsgrootheid met snelheid $\sqrt{n}$ convergeert en asymptotisch normaal is onder zowel de nulhypothese als een contigu alternatief. Bovendien bewijzen wij dat deze toets asymptotisch efficiënt is tegen het alternatief $F_1 = F_0^{1+\theta}$. Implementatie van deze toetsingsprocedure is eenvoudig, maar parameters van de normale limietverdeling hangen af van de onderliggende verdelingen. Wanneer men de toets uitvoert, moeten de parameters geschat worden. Dat zal ongetwijfeld de nauwkeurigheid van de toets verminderen.

Daarom stellen we tevens een op de likelihood ratio gebaseerde toets voor. We bewijzen dat de limietverdeling van deze toetsingsgrootheid onder de nulhypothese chi-kwadraat is met één vrijheidsgraad, terwijl onder een contigu alternatief de limietverdeling die van een gekwadrateerde, niet gecentreerde normaal verdeeld stochast is. Deze toets is ook asymptotisch efficiënt tegen het alternatieve $F_1 = F_0^{1+\theta}$. Implementatie van deze toestsingsprocedure is meer gecompliceerd omdat het de semi-parametrische maximum likelihood schatter bevat die veel meer rekentijd kost in verhouding tot de niet-parametrische schatter onder de nulhypothese gebruikt in de eerste methode.

Het belangrijkste instrument gebruikt in hoofdstukken 2 en 3 is een generalisatie van het stochastic equicontinuity lemma dat, in feite, neerkomt op een generalisatie van de uniforme centrale limietstelling. Een andere belangrijke tussenstap die we hier willen vermelden is dat ondanks dat de semi-parametrische MLE onder het Lehmann alternatief convergeert met een snelheid lager dan $\sqrt{n}$, onder een contigu Lehmann alternatief convergeert de MLE van de één-dimensionale parameter θ wel met snelheid $\sqrt{n}$ en zijn asymptotische verdeling is normaal.

In de laatste drie hoofdstukken van dit proefschrift concentreren wij ons op een andere toepassing van de maximum likelihood schatter. Eigenschappen van de NPMLE van een monotone dichtheid, ook bekend als de Grenander schatter, zoals snelheid van convergentie en puntsgewijze convergentie in verdeling, zijn bekend. Wij hebben een studie verricht naar de toepasbaarheid en prestatie van deze methode. Ten eerste, de Grenander schatter kan alleen toegepast worden als de onderliggende dichtheid monotoon is. Daarom willen we een toets om tegen alle niet-monotone alternatieven te toetsten. In hoofdstuk 4 beschouwen we het verschil tussen de geïntegreerde Grenander schatter en de empirische verdelingsfunctie, aangenomen dat de onderliggende dichtheid dalend is. Na bepaalde herschaling convergeert dit proces lokaal naar een limietproces. Bovendien, de integraal van dit verschil is aangetoond met snelheid $n^{2/3}$ te convergeren en na een herschaling asymptotisch normaal te zijn. Deze integraal kan als een toetsingsgrootheid gekozen worden om te toetsen of de onderliggende verdeling dalend is.

Daarna bestuderen we het gedrag van de Grenander schatter nabij de grenzen van de drager van de dichtheid (zonder verlies van algemeenheid stel dit is $[0, 1]$). Het is wel bekend dat in deze punten de Grenander schatter niet consistent is en er zijn verschillende modificaties van uniform consistente schatters door anderen voorgesteld. In hoofdstuk 5 laten we zien dat het verschil $(\hat{f}_n(n^{-\alpha}) - f(n^{-\alpha}))$ convergeert met snelheid $n^{(1-\alpha)/2}$ voor $1/3 \leq \alpha < 1$ en met snelheid $n^{1/3}$ voor $0 < \alpha \leq 1/3$. Ook hebben we de limietverdeling van dit verschil afgeleid. Ons resultaat en het bewijs hiervan zijn vergelijkbaar met resultaten die bekend zijn voor puntsgewijze convergentie van de Grenander schatter. Bovendien stellen we een andere uniform consistente modificatie van de Grenander schatter gebaseerd op dit resultaat voor, die makkelijk te implementeren is en een kleine verwachte kwadratische fout heeft. Resultaten van computer simulaties en vergelijking van onze modificatie met een eerder bekende modificatie zijn aan het einde van hoofdstuk 5 te vinden.

Wanneer we de L_k-fout van de Grenander schatter bestuderen, blijkt uit de resultaten van hoofdstuk 5 dat $k = 2.5$ een soort overgangspunt is. Voor $k < 2.5$ convergeert deze globale maat van nauwkeurigheid van de schatter met snelheid $n^{1/3}$ en na een geschikte herschaling is hij asymptotisch normaal. Maar voor $k \geq 2.5$ is de snelheid van convergentie van de standaard L_k-fout lager dan $n^{1/3}$ en zijn limietverdeling verschilt van de normale verdeling. Dit is een gevolg van een grotere bijdrage aan de integraal door de intervallen nabij de grenzen van de drager en van inconsistentie van de Grenander schatter in deze punten. Niettemin, in hoofdstuk 6 er is ook een modificatie van de L_k-fout voorgesteld, die verband heeft met een standaard L_k-fout van de gemodificeerde Grenander schatter, deze convergeert met snelheid $n^{1/3}$ en na een geschikte herschaling een normale limietverdeling heeft.

DOOR V.N.KULIKOV

Curriculum Vitae

De schrijver van dit proefschrift werd geboren op 27 januari 1976 te Moskou, Rusland. Van 1983 tot 1990 bezocht hij de openbare school #135 van Moskou en van 1990 tot 1993 - de openbare school #820, gespecialiseerd in wiskunde, natuurkunde en informatica, waar hij in 1993 cum laude zijn diploma haalde.

Daarna vervolgde hij zijn studie bij de Faculteit der Mechanica en Wiskunde van de Moskou Lomonosov Rijksuniversiteit. In 1998, na 3 jaar van specialisatie in statistiek onder begeleiding van prof.dr. Y.M.Rosanov en dr. M.V.Kozlov, studeerde hij cum laude af met de scriptie genaamd "Asymptotic behavior of some functionals of random walk and Brownian motion".

Na het afstuderen werkte hij een half jaar in de afdeling van informatietechnologie en communicaties van het ministerie van Transport van de Russische Federatie.

In de jaren 1999 - 2003 was hij assistent in opleiding bij de Technische Universiteit Delft waar hij onder supervisie van prof.dr. P.Groeneboom and dr. H.P.Lopuhaä onderzoek verrichtte naar verschillende aspecten en toepassing van de maximum likelihood methodes.