

Confidence bands for structural relationship models

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Janis Valeinis
aus
Riga, Lettland

Göttingen, 2007

D7

Referent: Prof.Dr.Axel Munk

Koreferent: Prof.Dr.Manfred Denker

Tag der mündlichen Prüfung: 18.01.2007

Contents

1	Introduction	4
2	Structural relationship models	11
3	Probability plots, P-P and Q-Q plots	15
3.1	Introduction	15
3.2	Approximations of P-P and Q-Q processes	18
3.3	Confidence bands for P-P plot via empirical processes	19
4	Confidence bands for general shift function Δ_s	23
5	Empirical likelihood	26
5.1	Nonparametric maximum likelihood	26
5.2	EL in one sample case	28
5.3	EL in two sample case	31
6	Plug-in empirical likelihood in two sample case	35
6.1	Assumptions and Notation	35
6.2	Main results	39
6.2.1	Maximization problem	39
6.2.2	Limiting distribution	42
6.3	Confidence bands for structural relationship models	45
7	Smoothed plug-in EL for Structural relationships	48
7.1	Main results	50
7.1.1	Maximization problem	51
7.1.2	Limiting distribution	52
7.2	Bootstrap confidence bands	55
8	Implementation	57
8.1	Simulation study	57
8.2	Data examples	58
9	Discussion	65
10	Technical tools: Empirical processes	67
10.1	Definitions and notation	67
10.2	Smoothed empirical process	68
10.3	Smoothed empirical distribution function	70

1 Introduction

It is of great interest in statistics to make inference about two sample problems. Consider the two sample case, where X_1, \dots, X_n and Y_1, \dots, Y_m are independent samples, with distribution functions F_1 and F_2 respectively. It may be of interest to check whether the two distribution functions are similar, that can be done visually by graphical inspection. However, statisticians use the goodness-of-fit tests to deal with this problem in a more formal way. For such purpose, the most famous tests are Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling tests.

The topic of this thesis does not concern the problem of goodness-of-fit itself. For example, our goal is not to detect whether F_1 or F_2 can be approximated by some parametric class of distributions, say Normal distribution, which in literature is called composite hypothesis. The simple hypothesis would deal with comparison of F_1 or F_2 with some concrete distribution with known parameters, e.g. Normal distribution with mean zero and variance one.

However, sometimes it is more interesting to check whether the two distribution functions differ from each another by a shift, i.e. we are interested in whether

$$F_1(t) = F_2(t - \mu) \tag{1.1}$$

holds for all $t \in \mathbb{R}$, where μ is some positive constant, usually called a shift or location parameter.

This question arises naturally in medicine, when the impact of a new drug is studied. The positive effect (it means that (1.1) holds for some $\mu > 0$) is welcomed. Usually, there are two groups of patients, X 's are control responses and Y 's are treatment responses. The parameter μ can be regarded as the amount the treatment adds to the potential control response. This situation also concerns the Sono-Histo data analyzed in Section 8.2, where different measurements of the size of pigmented skin tumors are compared. The control response, here, is a histological measurement (also denoted as a gold standard) and the treatments are the 20-MHz-Sonometric and the 50-MHz-Sonometric measurement methods.

Another example considers the relationship between hospitalization, as a measure of morbidity, and mortality (cf. Section 8.2). In this case, the prior interest is whether the distribution of age at hospitalization may be described by the distribution of age at death, or a simple transformation of that distribution.

Although relationship (1.1) is a very simple one, it is mostly of interest in application. Further we may be interested to test whether two samples differ by location and scale parameters, i.e., whether the relationship

$$F_1(t) = F_2\left(\frac{t - \mu}{\sigma}\right), \quad t \in \mathbb{R} \tag{1.2}$$

holds, where μ, σ are some positive constants.

The last model that we will mention is the Lehmann's alternative model, where, for the distribution functions F_1 and F_2 , the structural relationship

$$F_1(t) = 1 - (1 - F_2(t))^{(1/h)}, \quad t \in \mathbb{R} \tag{1.3}$$

holds with some parameter $h > 0$. If F_1 and F_2 are from a family of Lehmann's alternative model (1.3) then the respective hazard functions are proportional. This means that $\lambda_2(t) = h\lambda_1(t)$, where for $i = 1, 2$, $\lambda_i(t) = f_i(t)/(1 - F_i(t))$ and $f_i(t) = F_i'(t)$ is a density function. Under the assumption of Lehmann's alternative for censored data, several equivalence tests have been

derived in Com-Nougue *et al.* (1993) and Wellek (1993) (see also Munk, 1996 for the case of uncensored data).

The importance of structural models such as (1.1)-(1.3) has been emphasized by Freitag *et al.* (2003). They explain that the ranking methods, which are well used in practice, essentially require semiparametric assumptions (such as shift or location model in (1.1)) in order to be consistent and powerful (which is highlighted in Proposition 2.1 in Brunner and Puri (1996)). Therefore, a proper 'model check', e.g. whether the distribution functions F_1 and F_2 comes from location model (1.1), is advisable when using ranking methods. Statistics based on ranks are attractive, because they possess several robustness and invariance properties. For the two-sample case different rank tests for the location model are discussed in Chapter 3, Hettmansperger (1991). Among other literature on nonparametric statistics based on ranks and some structural models see Compagnone and Denker (1996), Hettmansperger (1991) and Manoukian (1986).

To unify the three above mentioned examples, (1.1), (1.2) and (1.3), Freitag (2000), Freitag and Munk (2005) have introduced the notion of a *structural relationship models*, which has the following general form

$$F_1(t) = \phi_2^-(F_2(\phi_1^-(t, h)), h), \quad t \in \mathbb{R}, \quad (1.4)$$

where ϕ_1, ϕ_2 are some functions defined later, ϕ_i^- denotes the inverse function with respect to the first argument, $h \in \mathcal{H} \subseteq \mathbb{R}^l$ is some parameter, l is an integer. Freitag (2000) has derived several goodness-of-fit tests for models of form (1.4) for the precise neighborhood hypothesis. It means that tests are made on the two distribution functions F_1 and F_2 in the following way:

$$H : d(F_1, F_2) > \Delta_0 \quad \text{vs} \quad K : d(F_1, F_2) \leq \Delta_0, \quad (1.5)$$

where $\Delta_0 > 0$ and d is some appropriate metric, e.g. Mallows distance (cf. Mallows, 1972). For constructing the precise neighborhood hypothesis (1.5), the Mallow's distance was first used in Czado and Munk (1998), and further analyzed in Czado and Munk (2000).

Let F_{1n} and F_{2m} denote the cumulative empirical distribution functions of F_1 and F_2 respectively. In (1.4) the unknown parameter h is estimated semiparametrically through Mallow's distance, i.e.

$$\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}_0} \left\{ \frac{1}{b-a} \int_a^b (F_{1n}^{-1}(u) - \phi_1(F_{2m}^{-1}(\phi_2(u, h)), h))^2 du \right\}, \quad (1.6)$$

where h, ϕ_1, ϕ_2 are as defined in (1.4), $0 \leq a < b \leq 1$ are the trimming bounds, and F^{-1} is a left inverse of function F . The test is performed with the plugged-in semiparametric estimator \hat{h} in the test statistic.

We have introduced general structural relationship models and described also the corresponding testing procedure. However, the first step for comparing two distribution functions is a graphical inspection of the data. This is done by means of the P-P (probability-probability) and Q-Q (quantile-quantile) plot, which can be identified with the graphs of the functions $\{F_1(F_2^{-1}(y)) : y \in (0, 1)\}$ and $\{F_1^{-1}(F_2(x)) : x \in \mathbb{R}\}$ respectively. From data we plot such graphs using the cumulative empirical distribution functions F_{1n} and F_{2m} instead of F_1 and F_2 .

Figure 1 shows the empirical P-P and Q-Q plot for simulated $X_i \sim N(0, 1)$, $i = 1, \dots, n$ and $Y_j \sim N(1, 1)$, $j = 1, \dots, m$ with $n = m = 100$, where \sim denotes 'distributed as' and $N(0, 1)$ is the Normal distribution with mean 0 and variance 1.

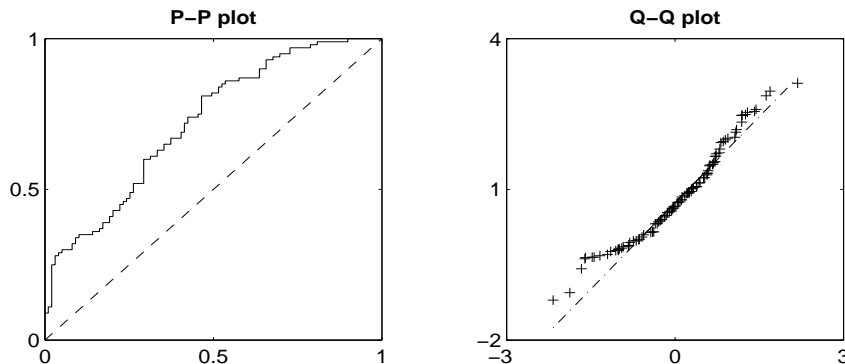


Figure 1: P-P and Q-Q plots for $N(0,1)$ and $N(1,1)$ with $n = m = 100$.

The Q-Q plot indicates that both distributions most likely come from one family of distributions, because the graph can be quite well approximated by a straight line. On the other hand, the P-P plot suggests that there is a positive shift for location model (1.1), because the curve is above the diagonal. See more description in Section 3. For the location model (1.1) P-P plots are clearly more useful than Q-Q plots, because Q-Q plots which are almost identical for both negative and positive shifts. P-P plots are also to be preferred over Q-Q plots when outliers might be present. Finally, P-P plots allow for scale invariant treatment comparisons (cf. Holmgren, 1995).

Both P-P and Q-Q plots only indicate for one or another property. To perform a two sample test we need to add simultaneous confidence bands. Most literature deals with pointwise confidence bands, but this allows only to perform a test in some fixed point of functions defined via P-P or Q-Q plots. To construct simultaneous bands one possibility is to use asymptotic theory of P-P and Q-Q plot processes developed by Beirlant and Deheuvels (1990) among others (further on denoted as *empirical bands*). Alternatively, one can use the well-known Bonferroni's method. From the considerations above we choose P-P plots to construct the bands. For an illustration, see Figure 2 for X 's and Y 's distributed again as $N(0,1)$ and $N(1,1)$ respectively.

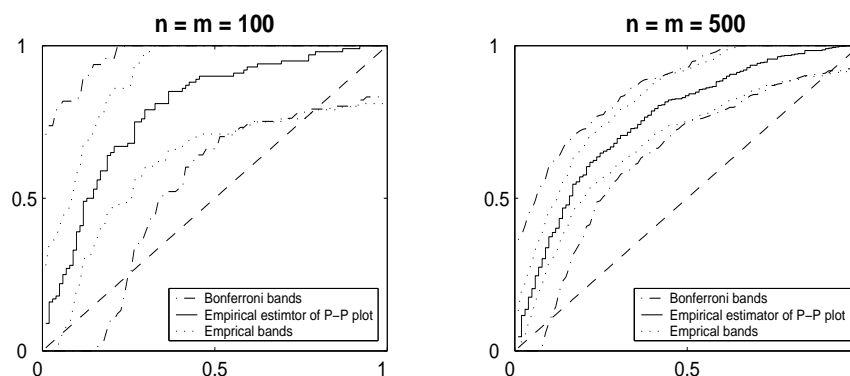


Figure 2: P-P empirical plots of simulated $N(0,1)$ against $N(1,1)$ with simultaneous confidence bands added for several sample sizes $n = m = 100, 500$.

Empirical bands obtained from asymptotic theory are narrower, thus they are better than the Bonferroni's bands. Now, if the diagonal connecting the points (0,0) and (1,1) would fit into the bands, then we would not reject the hypothesis that the two distributions F_1 and F_2 are the same with confidence $1 - \alpha$, where e.g. $\alpha = 0.05$.

The goal of this thesis is to construct such simultaneous confidence bands for a general P-P plot for the structural relationship model (1.4). The idea is to estimate the unknown parameter h semiparametrically through the Mallow's distance as in (1.6). Then to transform the second sample in such a way that we would expect the diagonal to fit in the bands if the distributions F_1, F_2 were the same. For example, the confidence bands for the location model (1.1) shall be constructed for the P-P plot of $F_1(t)$ and $F_2(t - \hat{h})$.

Now we list the most well known methods in statistics literature for the construction of simultaneous confidence bands for P-P or Q-Q plots. Apart from bands obtained from asymptotic theory and Bonferroni's method, there are two main approaches. First, it is possible to construct bands for the parameter itself, if we consider the *general shift* function introduced by Doksum (1974, 1977), Doksum and Sievers (1976),

$$\Delta_s(t) = F_2^{-1}(F_1(t)) - t, \quad t \in \mathbb{R}. \quad (1.7)$$

We can recognize from (1.7) that Δ_s is the Q-Q plot $F_2^{-1}(F_1(t))$ minus the argument t itself. Kolmogorov-Smirnov two sample statistic and its weighted version are proper tools for derivation of confidence bands. For the location model a horizontal line should fit in the bands. This method allows us to construct the bands for general shift function with $X + \Delta_s(X) \sim Y$. Thus, $\Delta_s(\cdot)$ can be regarded as the amount of 'shift' needed to bring the X 's up to the Y 's in distribution. For structural relationship models this method can not be extended in a general way. However, the method is often used when the location model (1.1) is under consideration (see e.g. Aaberge *et al.*, 1985). See Figure 3 for the illustration.

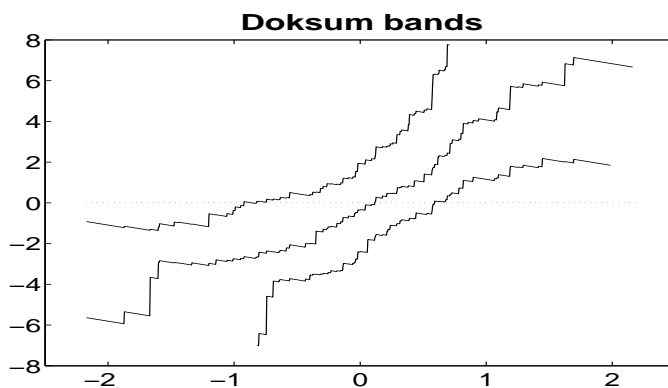


Figure 3: Simultaneous confidence bands for general shift function Δ_s constructed for simulated $N(0,1)$ and $N(1,4)$ samples, $\alpha = 0.05$, $n = m = 100$.

If no diagonal fitted in the bands, with confidence of $1 - \alpha$, where $\alpha = 0.05$, we could reject the hypothesis that the location-scale model holds. If no horizontal line fits in the bands we can reject the hypothesis of the location model (for explanation, see Section 4).

Now we turn again to the idea of first estimating the parameter h and then constructing the confidence bands. The nonparametric maximum likelihood method, also called empirical likelihood (cf. Owen, 1988; 1990), has become surprisingly popular for constructing confidence regions. It appears to have advantages almost above all other existing estimation methods, including the bootstrap. Empirical likelihood methods are very appealing, because they require no pre-specified assumptions about the shape of confidence regions, which is determined automatically by the sample, as pointed out by Owen (1990). Empirical likelihood regions are Bartlett correctable in most cases (cf. Hall and La Scala, 1990; DiCiccio *et al.*, 1991). This means that a simple correction for the mean of the empirical log-likelihood ratio reduces the coverage error from order n^{-1} to order n^{-2} , where n denotes the sample size.

The empirical likelihood method is an interesting alternative to the Normal distribution based symmetric confidence regions, because it avoids the need to estimate the variance term (cf. Claeskens *et al.*, 2003). Confidence intervals based on likelihood ratios also have the additional advantage of respecting the range of the parameter (see Owen, 2001 for a summary of literature on this topic).

Using the empirical likelihood method Einmahl and McKeague (1999) constructed simultaneous confidence bands for multiple quantile-quantile plots from censored samples. For independent samples using empirical likelihood Claeskens *et al.* (2003) obtained a nonparametric estimator of the P-P plot process and constructed the simultaneous bands using the method introduced by Hall and Owen (1993). Hall and Owen's (1993) method allows empirical likelihood to set the shape of the confidence bands and uses bootstrap to set the level. Thus, the pointwise confidence bands are constructed using a level obtained from bootstrapping method. Moreover, Claeskens *et al.* (2003) use the smoothed empirical likelihood method, which amazingly provides some additional advantages in case of constructing the confidence intervals for sample quantiles (cf. Chen and Hall, 1993). Note that P-P processes are closely related to Receiver Operating Characteristic (ROC) curves, which are of great importance in medicine. Nowadays, a lot of projects are connected with developing simultaneous bands for ROC curves (see, for example, Macskassy and Provost, 2004 or Macskassy *et al.*, 2005).

In this thesis we generalize the setup of Qin and Zhao (2000) and Hjort *et al.* (2004) introducing plug-in empirical likelihood for the two sample case. A plug-in version of empirical likelihood allows us to derive pointwise confidence bands for the P-P plot of general structural relationship models. Claesken's *et al.* (2003) results for constructing the confidence bands for the P-P plot follow directly from our results. Although Claeskens *et al.* (2003) state that the P-P and Q-Q plots require different techniques we show that for the independent samples they can be treated in the same way. However, in our context we found P-P advantageous above Q-Q plots. Thus, for the data examples we construct simultaneous bands for P-P plots only.

We complete our work with establishing a smoothed version of plug-in empirical likelihood for the two sample case, thus extending the results of Chen and Hall (1993). Finally, the simultaneous confidence bands have been obtained using the bootstrap method of Hall and Owen (1993) as described before. Now consider the first sample X_1, \dots, X_n . Our condition $nb_1^{3r} \rightarrow 0$, where r is the order of the kernel, as $n \rightarrow \infty$ on the speed of the smoothing parameter b_1 differs from Claesken's *et al.* (2003) conditions. In Section 9 we comment on our conditions.

It is remarkable that for structural relationship models the usual chi-squared distribution with one degree of freedom is the limiting distribution for the test statistic. Thus, we have no

perturbation due to plug-in for structural relationships, that is different in general (cf. Theorems 21, 29).

For an illustration, in Figure 4 we have plotted the bootstrapped simultaneous bands obtained from smoothed plug-in empirical likelihood method for the location model. The data have been taken from previously discussed example with histological and sonometrical measurements of pigmented skin tumors (cf. Section 8.2).

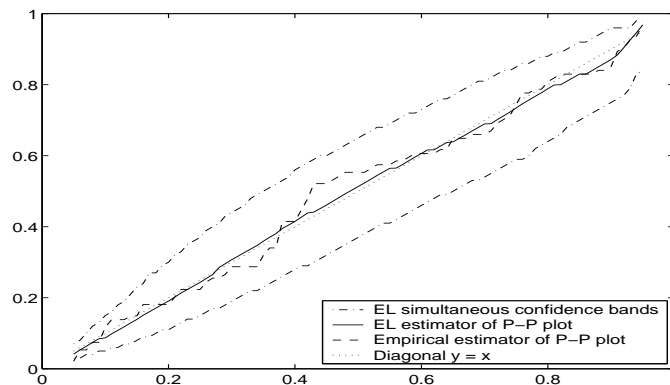


Figure 4: EL simultaneous bands for P-P plot comparing Histological data with 50-MHz-Sonometrical data, $\alpha = 0.05$.

Using the Hall and Owen's (1993) method the simultaneous confidence bands are obtained from constructing the pointwise confidence intervals with the bootstrapped critical value $c^* = 6.57$. Note that the smoothed empirical likelihood estimator of P-P plot is much "smoother" than that of the empirical P-P plot, which is also to be expected. The range of the estimated P-P plot and the confidence bands is always between 0 and 1, which is easy to interpret and compare.

The bands in Figure 4 are only indirectly comparable with the bands for the general shift function Δ_s , which are shown in Figure 5.

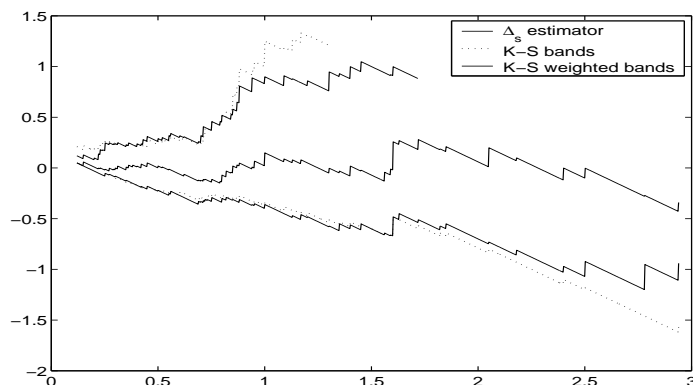


Figure 5: Doksum bands for Δ_s for comparing Histological data with 50-MHz-Sonometrical data, $\alpha = 0.05$.

A horizontal line would fit into the bands in Figure 5. On the other hand the diagonal clearly fits into the simultaneous empirical likelihood bands in Figure 4. We can make the same

conclusion in both cases: for the particular data example we can not reject the hypothesis that the location model holds.

This thesis is organized as follows. In Section 2 we introduce structural relationship models. Sections 3 and 4 are devoted to the introduction of the known results in literature concerning simultaneous confidence bands and structural relationships. As our main task is to construct confidence bands around P-P or Q-Q plots, we introduce them in Section 3 as well as state some theorems concerning asymptotic theory regarding confidence bands for P-P plots. In Section 4 we introduce simultaneous confidence bands for a general shift function. Section 5 defines the empirical likelihood method for one and two sample cases. We establish plug-in empirical likelihood for the two sample case in Section 6 and state our main assumptions and results of the thesis. Confidence bands for structural relationship models follow as a consequence in Section 6.3. Next, we continue with establishing smoothed plug-in empirical likelihood in the two sample case and define the bootstrap confidence bands for structural relationship models. This is completed in Section 7. Finally, in Section 8 we show how the bootstrap method works for constructing simultaneous confidence bands by simulation study and two data examples. We discuss our results in Section 9. The Appendix summarizes some results from empirical process theory and introduces the smoothed empirical processes, which is needed for proofs.

Acknowledgments

I wish to thank my thesis advisor Axel Munk for proposing the problem and supervision. My thanks goes also to Manfred Denker for taking the Koreferat and interesting discussions during my stay in Göttingen. During my time as a Ph.D. student I was a member of the Ph.D. program "Applied Statistics and Empirical Methods". Not only from there, but also from the Graduiertenkolleg "Identifikation in mathematischen Modellen: Synergie stochastischer und numerischer Methoden" I received a financial support and I am grateful for that. I have to thank Krzysztof for always being ready to help with computers and bicycles, which I broke usually in accidents. For a good working spirit and atmosphere I would like to thank especially Achim, Marina, Ta-chao, Mihaela, Mikhail, Leif, Natalia, Rada, Alexei, Hajo and Gudrun. My greatest thanks goes to my family - my wife Daina, my parents and sister in Latvia. Their support and love always gave me belief and strength to finish my Ph.D.

2 Structural relationship models

In nonparametric statistics often statistical tests or other methods of analysis are being performed assuming that some semiparametric model holds. For example, comparing continuous distribution functions often the assumption of location-scale alternative models has been made (cf. e.g. Hettmansperger, 1991). In survival analysis the proportional Hazards model is well-known and analysed (cf. Lehmann, 1953). In this case we have a situation of the Lehmann alternative model. To assess the validity of such models some goodness-of-fit tests based on Mallows-distance have been derived in Freitag (2000), Freitag and Munk (2005) for general "structural semiparametric models". For the two-sample case let us give first some important examples.

Example 1. The classical location-scale model with two distribution functions F_1 and F_2 for independent samples is defined as follows

$$F_1(t) = F_2\left(\frac{t - \mu}{\sigma}\right) := F_2(t, h), \quad t \in \mathbb{R} \quad (2.1)$$

for some parameter $h = (\mu, \sigma)$ with $\sigma > 0$. This relationship can be also expressed through quantile functions

$$F_1^{-1}(u) = F_2^{-1}(u)\sigma + \mu, \quad u \in [0, 1], \quad (2.2)$$

where $F_1^{-1}(u) = \inf\{x : F_1(x) \geq u\}$. If $\sigma \equiv 1$ in (2.1) we have the shift or location model. We have the scale model if $\mu \equiv 0$ in (2.1). In survival analysis the scale model is also denoted as an acceleration model.

Example 2. We say that F_1 and F_2 belong to the family of Lehmann-alternatives if

$$F_1(t) = 1 - (1 - F_2(t))^{(1/h)} =: F_2(t, h), \quad t \in \mathbb{R} \quad (2.3)$$

for some $h > 0$. Using again quantile functions this structural model can be rewritten as

$$F_1^{-1}(u) = F_2^{-1}(1 - (1 - u)^h), \quad u \in [0, 1]. \quad (2.4)$$

If F_1 and F_2 belong to (2.3), then we have the situation of proportional hazard rates in the two-sample case, where the relation

$$\lambda_2(t) = h\lambda_1(t), \quad t \in \mathbb{R},$$

where for $i = 1, 2$, $\lambda_i(t) = f_i(t)/(1 - F_i(t))$ and $f_i(t) = F_i'(t)$ is a density function with respect to distribution function F_i .

Let i.i.d. r.v.'s X_1, \dots, X_n and Y_1, \dots, Y_m have unknown distribution functions F_1 and F_2 respectively which are elements of the set

$$\mathcal{F}^2 := \{F : F \text{ is a c.d.f. and } \int t^2 dF(t) < \infty\}. \quad (2.5)$$

Two-sample models, such as (2.2) or (2.4), which can be represented with quantile functions can be formally written in a general form.

Definition 1 (Freitag and Munk (2005)). Let $\mathcal{H} \subseteq \mathbb{R}^l$, compact and $\phi_1 : \mathbb{R} \times \mathcal{H} \rightarrow \mathbb{R}$, $\phi_2 : [0, 1] \times \mathcal{H} \rightarrow [0, 1]$. We say that F_1 and F_2 in \mathcal{F}^2 are related by a *structural relationship* induced by ϕ_1 and ϕ_2 , if $(F_1, F_2) \in \mathcal{U}_{\phi_1, \phi_2} =: \mathcal{U}$, where the model class \mathcal{U} is given by

$$\mathcal{U} := \{(F_1, F_2) \in \mathcal{F}^2 \mid \exists h \in \mathcal{H} : F_1(\phi_1(F_2^{-1}(\phi_2(u, h))), h), \quad u \in [0, 1]\}. \quad (2.6)$$

Assumption 1. For $F \in \mathcal{F}$ and $h \in \mathcal{H}$ it holds that $\phi_1(F^{-1}(\phi_2(\cdot, h)), h) \in \mathcal{F}$. For the trimming bounds $0 \leq a < b \leq 1$, let $l(h) := \phi_2(a, h)$ and $u(h) := \phi_2(b, h)$ for $h \in \mathcal{H}$. It is assumed that $\phi_2(\cdot, h)$ is strictly isotonic for all $h \in \mathcal{H}$, i.e. there is a map $\phi_2^- : [l(h), u(h)] \times \mathcal{H} \rightarrow [a, b]$ with

$$\phi_2(\phi_2^-(v, h), h) = v, \quad v \in [l(h), u(h)]. \quad (2.7)$$

Likewise, there is an inverse $\phi_1^- : \mathbb{R} \times \mathcal{H} \rightarrow \mathbb{R}$, i.e. for all $h \in \mathcal{H}$,

$$\phi_1(\phi_1^-(t, h), h) = t, \quad t \in \mathbb{R}. \quad (2.8)$$

The functions ϕ_1 and ϕ_2 are twice continuously differentiable with respect to both arguments on $\mathbb{R} \times \mathcal{H}$.

We can also express the structural relationships using distribution functions instead of quantile functions in the definition of \mathcal{U} ,

$$F_1(t) = \phi_2^-(F_2(\phi_1^-(t, h)), h) =: F_2(t, h). \quad (2.9)$$

Remark 1. After a general definition of structural relationships in (2.6) we immediately get the previously described location-scale model (2.1) as a special case, if $\phi_1(t, (h_1, h_2)^T) = h_1 + h_2 t$ and $\phi_2(u, h) \equiv u$. For $\phi_1(t, h) \equiv t$ and $\phi_2(u, h) = 1 - (1 - u)^h$ we obtain the model of Lehmann alternatives introduced in (2.3).

In this work the trimmed Mallows distance between two distribution functions has been used as an appropriate discrepancy measure to estimate the difference between two samples. The properties and asymptotic of Mallows distance have been nicely described in Freitag (2000), Section 5; Czado and Munk (1998, 2001) and Munk and Czado (1998).

For two distribution functions $F_1, F_2 \in \mathcal{F}^2$ the *trimmed Mallows distance* is defined as

$$d_{[a,b]}(F_1, F_2) := \left\{ \frac{1}{b-a} \int_a^b (F_1^{-1}(u) - F_2^{-1}(u))^2 du \right\}^{\frac{1}{2}}, \quad (2.10)$$

where $[a, b] \subseteq [0, 1]$. If $[a, b] = [0, 1]$, we have exactly the L^2 distance introduced by Mallows (1972). Let us mention that the Mallows distance has not been much used in context of goodness-of-fit testing. First in Czado and Munk (1998) Mallows distance has been used constructing tests for the precise neighborhood hypotheses ($H_\Delta : d(F, F_0) > \Delta_0$ vs $K_\Delta : d(F, F_0) \leq \Delta_0$, where $\Delta_0 > 0$ and d is some appropriate metric, e.g. Mallows distance). Recently, Freitag (2000) and Freitag and Munk (2005) analyze asymptotic behavior of those tests in context of survival analysis and censored data in detail.

To get an estimator for the parameter $h \in \mathcal{H}$ in (2.6) we define first general classes of distance functionals, i.e.

$$\begin{aligned} T(F_1, F_2) := d^2(F_1, F_2, \mathcal{U}) &:= \inf_{h \in \mathcal{H}} d^2(F_1, F_2(\cdot, h)) \\ &= \inf_{h \in \mathcal{H}} \left\{ \frac{1}{b-a} \int_a^b (F_1^{-1}(u) - \phi_1(F_2^{-1}(\phi_2(u, h)), h))^2 du \right\} \end{aligned}$$

with $F_2(\cdot, h)$ from (2.9). Now we need the following assumption for the existence of minimum of the functional $d^2(F_1, F_2(\cdot, h))$.

Assumption 2. A minimum h_0 of $d^2(F_1, F_2(\cdot, h))$ exists inside \mathcal{H} . For every minimum h_0 there exists neighborhood $\mathcal{H}_0 \subset \mathcal{H}$ for h_0 , so that h_0 is a unique minimum for $d^2(F_1, F_2(\cdot, h))$ in \mathcal{H}_0 , i.e.

$$h_0 = \operatorname{argmin}_{h \in \mathcal{H}_0} \left\{ \frac{1}{b-a} \int_a^b (F_1^{-1}(u) - \phi_1(F_2^{-1}(\phi_2(u, h)), h))^2 du \right\}. \quad (2.11)$$

If $(F_1, F_2) \in \mathcal{U}$, then there exists a unique parameter h_0 , so that the structural relationship model is satisfied; this h_0 is also the unique minimum of $d^2(F_1, F_2(\cdot, h))$.

For further purpose we need to analyze the behavior of the estimator \hat{h} . Let F_{1n} and F_{2m} be the cumulative empirical distribution functions of F_1 and F_2 respectively.

Remark 2. From Freitag (2000) and Freitag and Munk (2005) follows that the semiparametric estimator \hat{h} obtained as

$$\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}_0} \left\{ \frac{1}{b-a} \int_a^b (F_{1n}^{-1}(u) - \phi_1(F_{2m}^{-1}(\phi_2(u, h)), h))^2 du \right\} \quad (2.12)$$

is a consistent estimator, thus $\hat{h} \rightarrow_p h_0$, where \rightarrow_p means the convergence in probability.

Actually a stronger result holds for the estimator \hat{h} than that stated in Remark 2. For this reason we need a version of strong approximation theorems for the quantile processes, which states that under some conditions

$$\sup_{0 < y < 1} |f(F^{-1}(y))(F_{1n}^{-1}(y) - F_1^{-1}(y))| = O(n^{-1/2}(\log \log n)^{1/2}) \quad a.s. \quad (2.13)$$

For the reference see, for example, Theorem 5.3.1. in Csörgő and Révész (1981).

Remark 3. The estimator \hat{h} can be obtained by letting the derivative of the expression (2.12) equal to zero and finding the root for h . Thus, \hat{h} is in fact a functional of the functions F_{1n} and F_{2m} . From the continuous mapping theorem in (2.12) and (2.13) it follows that $\hat{h} \rightarrow h_0$ almost surely.

Let us shortly comment on almost sure rates of the estimator \hat{h} , needed for Theorems 29 and 37.

Freitag (2000) has shown that \hat{h} as a functional from the empirical distribution functions F_{1n} and F_{2m} is compactly or Hadamard differentiable (see, for example, van der Vaart, 1998, p. 296). Unfortunately from the rates of empirical quantile processes as in (2.13) we can not deduce the rates for Hadamard differentiable functionals in general (for discussion see Dudley, 1994, pages 1-5). However, in the case of the location model we can deduce the almost sure rates, which is highlighted in the next remark.

Remark 4. In the case of the location model an exact expression can be obtained (see also Freitag et al., 2003) as follows. With trimming bounds $0 \leq a < 1 - a \leq 1$ the minimizing expression is

$$\frac{1}{1-2a} \int_a^{1-a} (F_{1n}^{-1}(u) - F_{2m}^{-1}(u) - h)^2 du \quad (2.14)$$

and we have

$$\hat{h} = \frac{1}{1-2a} \int_a^{1-a} (F_{1n}^{-1}(u) - F_{2m}^{-1}(u)) du. \quad (2.15)$$

Now from (2.12) and continuous mapping theorem it follows that $\hat{h} \rightarrow h_0$ almost surely with the rate $O(n^{-1/2}(\log \log n)^{1/2})$.

3 Probability plots, P-P and Q-Q plots

Our aim is to construct simultaneous confidence bands for the P-P plot of structural relationship models. Other approach is to construct the bands for the general shift function Δ_s , which contains also the Q-Q plot process. Therefore in this Section we give a short introduction for P-P and Q-Q plots, and define confidence bands from the asymptotic theory for P-P and Q-Q plot processes.

3.1 Introduction

Graphical methods have always been an important tool for both formal and informal analysis of the data. Due to highly developed computer science technologies, new statistical packages are used in almost all scientific fields that deal with large quantities of raw, empirical data. The simplest diagnostic plots as histograms, empirical distribution functions or box-plots are used to explore location, scale, skewness, kurtosis or other differences in two-sample problems, symmetry or goodness of fit-problems for one sample; analysis of covariance, k -sample or other multivariate procedures. An extensive review and bibliography of graphical methods in non-parametric statistics can be found in, e.g., Doksum (1977), Gnanadesikan (1977), Fisher (1983), Sawitzki (1994).

An important role appertains to such graphical methods that are based on comparisons of the sample distribution or quantile functions, especially in two-sample problems. Powerful statistical techniques based on the sample distribution function have been developed analyzing underlying models, see e.g. Gnanadesikan and Wilk (1968) or Doksum (1977) to elucidate the nature of association between variables (Taguri *et al.*, 1976), and to highlight the structure of multivariate distributions (Tukey and Tukey, 1981). The classical methods include probability-probability (P-P) plot, quantile-quantile (Q-Q) plot, pair charts, receiver operating characteristic (ROC) curves, proportional hazard plots, etc.. Now we will introduce P-P and Q-Q plots, state their main properties and derive simultaneous confidence bands for these plots.

The two random samples will be denoted by X_1, \dots, X_n and Y_1, \dots, Y_m , drawn from underlying populations with univariate, continuous distribution functions F_1 and F_2

The classical P-P plot ($\{F_1(x), F_2(x) : x \in \mathbb{R}\}$) and Q-Q plot ($\{(F_1^{-1}(y), F_2^{-1}(y)) : y \in (0, 1)\}$) are based on univariate distribution functions F_1 and F_2 and are generally used to compare them. The P-P plot can be identified with the graph of the function

$$\{F_1(F_2^{-1}(y)) : y \in (0, 1)\}, \quad (3.1)$$

and the Q-Q plot with the graph of the function

$$\{F_1^{-1}(F_2(x)) : x \in \mathbb{R}\}. \quad (3.2)$$

To show the difference and specific of each of P-P and Q-Q plots, in Figure 6 we have displayed several empirical plots based on Normal distribution with $n = 100$ using the empirical distribution functions F_{1n} , F_{2m} instead of F_1 and F_2 . The P-P plots have been drawn from (3.1), but for the presentation of the Q-Q plots we preferred to draw the quantiles of one sample against the quantiles of another sample as usually done in literature.

The P-P plot has been drawn according to the definition 3.1. However, we prefer to present the Q-Q plot as it has been usually done in literature as a point plot, where the quantiles from one sample have been drawn against the quantiles from another sample.

Although when F_1 coincide with F_2 both plots will be straight lines with slope one, thus P-P and Q-Q plots seemingly have similar behavior, their properties are quite different which is nicely attributed by Gnanadesikan and Wilk (1968), Holmgren (1995). An elementary property of Q-Q plots is that if Y is a linear function of X then the corresponding Q-Q plot will still be linear but with possibly changed location and slope. The P-P plot is not useful to reveal the shift or scale differences of the same distributions, it does not have a linear invariance property. On the other hand it can be easily recognized whether we have a tendency of positive shift (tendency toward increased values on treatment relative to control) or negative only by P-P plots, Q-Q plots are almost identical in these situations (cf. Figure 6).

Another property of Q-Q plots is their sensitivity to differences between F_1 and F_2 in the tails, but P-P plots highlight more difference in the middle. So, Q-Q plot is advantageous, if one is interested to recognize immediately the outliers, but can direct attention away from the vast majority of data points. In this case P-P reveals more general tendency not taking into account the outliers.

The last property to mention is the range of the P-P and Q-Q plots. Q-Q plots can vary substantially in case of various control distributions, making comparisons of treatment effects difficult (cf. Holmgren, 1995). The range of P-P plot is always maintained - it begins at $(0, 0)$ and ends at $(1, 1)$ always, in this matter making them more attractive than Q-Q plots.

The Q-Q plots were first used by Lorenz (1905) for comparing two independent samples. Since then Q-Q plots have been extensively examined, their asymptotic properties profoundly investigated (cf. Gnanadesikan, 1977; Fisher, 1983; Aly, 1986b). The P-P plots became of great interest mainly, because of their relation to Receiver Operating Characteristic (ROC) curves defined as

$$\{1 - F_1(F_2^{-1}(1 - y)) : y \in (0, 1)\}. \quad (3.3)$$

ROC curves are of great importance and have been used in signal theory, psychology, radiology, medicine, etc. (see, e.g. Li *at al.*, 1996; Hsieh and Turnbull, 1996). Nonparametric statistics based on P-P plots and modifications have been studied in Gnanadesikan (1977), Parzen (1993), Beirlant and Deheuvels (1990), Sawitzki (1994), Deheuvels and Einmahl (1992), Nair (1981, 1982), Aly (1986a), Girling (2000), etc.

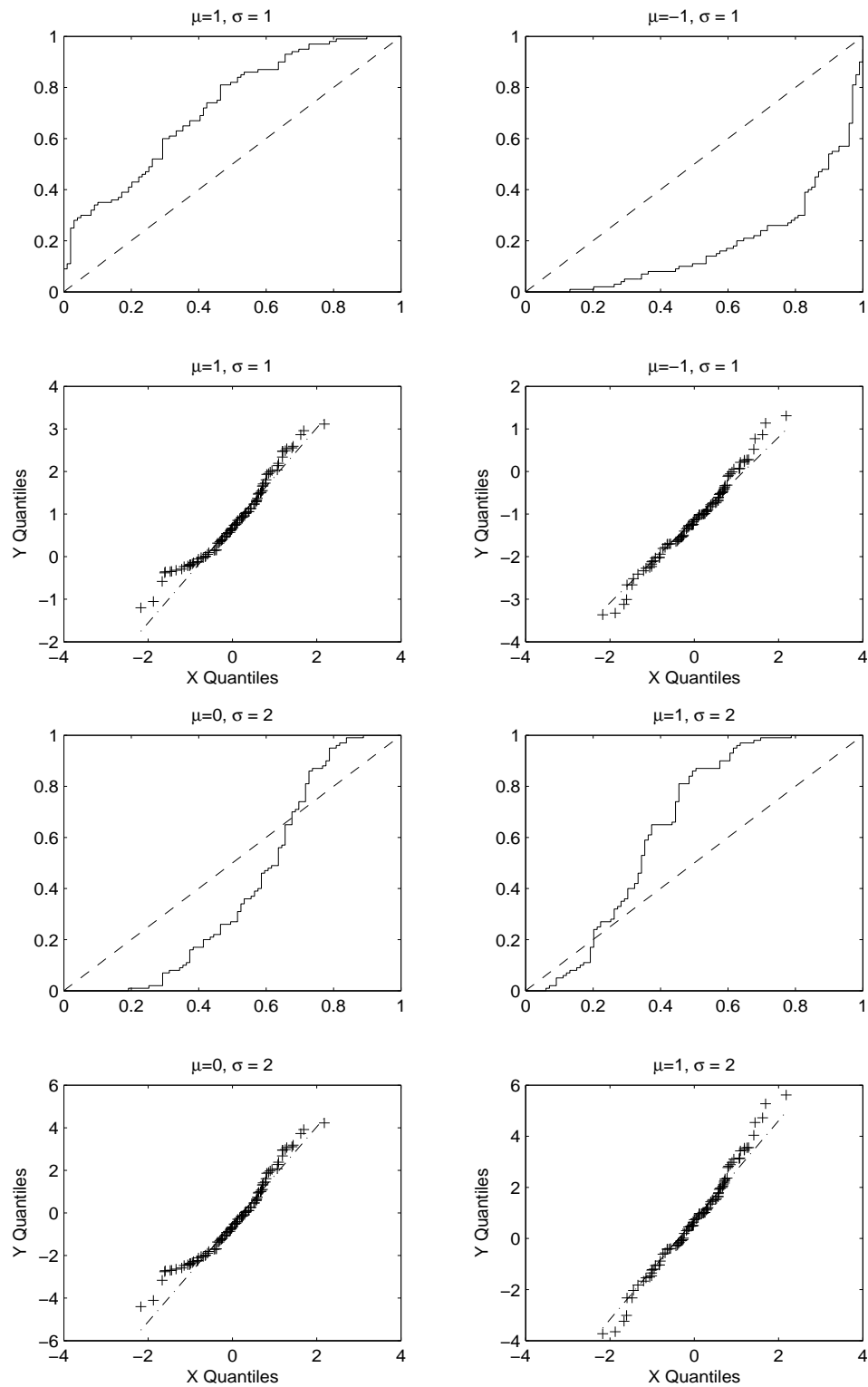


Figure 6: P-P and Q-Q empirical plots of simulated $N(\mu, \sigma)$ against $N(0, 1)$ distribution with sample size $n = m = 100$.

3.2 Approximations of P-P and Q-Q processes

We maintain all the notation from introduction. For asymptotic results we assume that F_1, F_2 have continuous densities f_1, f_2 . First already from the famous Glivenko-Cantelli theorem we obtain that P-P and Q-Q plots converge pointwise to their theoretical counterparts with probability one.

Lemma 5. *As $n, m \rightarrow \infty$, for each $y \in (0, 1)$ we have*

$$|F_{1n}(F_{2m}^{-1}(y)) - F_1(F_2^{-1}(y))| \rightarrow 0 \quad a.s., \quad (3.4)$$

where *a.s.* denotes the almost sure convergence and F_{1n}, F_{2m} are defined in Remark 2. For each $x \in (a, b)$ with

$$a := \inf\{x : F_2(x) > 0\} \quad \text{and} \quad b := \sup\{x : F_2(x) < 1\},$$

we have that

$$|F_{1n}^{-1}(F_{2m}(x)) - F_1^{-1}(F_2(x))| \rightarrow 0 \quad a.s. \quad (3.5)$$

Proof. By the Glivenko-Cantelli theorem when F_j and F_j^{-1} , for $j = 1, 2$ are continuous we obtain for any $\epsilon > 0$ and m large enough

$$F_2(x) - \epsilon \leq F_{2m}(x) < F_2(x) + \epsilon \quad \text{for all } x \in \mathbb{R} \quad a.s.$$

Since F_{2m}^{-1} is the inverse of F_{2m} we obtain that

$$F_2^{-1}(y - \epsilon) \leq F_{2m}^{-1}(y) < F_2^{-1}(y + \epsilon) \quad \text{for all } y \in (0, 1) \quad a.s.,$$

then for each $y \in (0, 1)$ such $\delta > 0$ exists, that

$$F_2^{-1}(y) - \delta \leq F_2^{-1}(y - \epsilon) \leq F_{2m}^{-1}(y) < F_2^{-1}(y + \epsilon) \leq F_2^{-1}(y) + \delta \quad a.s.$$

Hence for each fixed $y \in (0, 1)$,

$$\begin{aligned} |F_{1n}(F_{2m}^{-1}(y)) - F_1(F_2^{-1}(y))| &\leq |F_{1n}(F_{2m}(y)) - F_1(F_{2m}^{-1}(y))| \\ &+ |F_1(F_2^{-1}(y)) - F_1(F_2^{-1}(y))| \rightarrow 0, \quad a.s. \quad n, m \rightarrow \infty. \end{aligned}$$

The second statement (3.5) follows similarly.

Beirlant and Deheuvels (1990) define P-P and Q-Q plot processes as follows

$$\Delta_{n,m}(y) = \sqrt{\frac{nm}{n+m}} (F_{1n}(F_{2m}^{-1}(y)) - F_1(F_2^{-1}(y))), \quad y \in (0, 1). \quad (3.6)$$

and

$$\Gamma_{nm}(x) = \sqrt{\frac{nm}{n+m}} f_1(F_1^{-1}(F_2(x))) (F_{1n}^{-1}(F_{2m}(x)) - F_1^{-1}(F_2(x))), \quad x \in (a_m, b_m). \quad (3.7)$$

At certain conditions, under the null hypothesis $H_0 : F_1 \equiv F_2$, the P-P plot process Δ_{nm} is distribution free (i.e. independent of the common value of $F_1 = F_2$), while the empirical Q-Q plot process is not. However, after the change of scale $x = F_2^{-1}(y)$, it follows from Corrolary 3.1 in Beirlant and Deheuvels (1990) that also the Q-Q plot process is distribution free. They showed that both processes converge in distribution to Brownian bridges with the approximation rate $n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}$, where $m = m(n) \rightarrow \infty$ such that $n, m \rightarrow \infty$, which is optimal in the setting of Berilant and Deheuvels (1990).

3.3 Confidence bands for P-P plot via empirical processes

In the literature the theoretical focus is mainly on obtaining consistency and asymptotic normality of various estimators of P-P and Q-Q plots, therefore offering the necessary tools to construct pointwise confidence intervals. To get an impression of the whole curves or plots, one should construct simultaneous confidence bands. For P-P plot this means we are looking for random functions $v_1(X, Y, t)$ and $v_2(X, Y, t)$ such that as $n, m \rightarrow \infty$

$$P\{v_1(X, Y, t) \leq F_1(F_2^{-1}(t)) \leq v_2(X, Y, t) \text{ for all } t \in [a, b]\} = 1 - \alpha, \quad (3.8)$$

where $0 \leq \alpha \leq 1$ is a given level, say 0.05 or 0.01.

Here we assume again that F_j and F_j^{-1} for $j = 1, 2$ are continuous. For asymptotic results assume that the sample sizes n, m are such that $m/n \rightarrow \lambda > 0$ as $n \rightarrow \infty$. Let us assume that the slope of the curve $F_1(F_2^{-1}(t))$, that is, $f_1(F_2^{-1}(t))/f_2(F_2^{-1}(t))$ is bounded on any subinterval (a, b) of $(0, 1)$, $0 < a < b < 1$.

To construct the simultaneous confidence bands we will use the following general facts from Hsieh and Turnbull (1996).

Lemma 6 (Hsieh and Turnbull (1996), Theorem 2.1). *Under the above conditions we have*

$$\sup_{0 \leq t \leq 1} |F_{1n}(F_{2m}^{-1}(t)) - F_1(F_2^{-1}(t))| \rightarrow 0 \text{ a.s. as } n \rightarrow \infty. \quad (3.9)$$

Proof. Consider the inequality

$$\begin{aligned} \sup_t |F_{1n}(F_{2m}^{-1}(t)) - F_1(F_2^{-1}(t))| &\leq \sup_t |F_{1n}(F_{2m}^{-1}(t)) - F_1(F_{2m}^{-1}(t))| \\ &\quad + \sup_t |F_1(F_{2m}^{-1}(t)) - F_1(F_2^{-1}(t))|. \end{aligned} \quad (3.10)$$

Applying now the Glivenko-Cantelli theorem for the first term on the RHS and the theorem of Dvoretzky, Kiefer and Wolfowitz (1956) and then the Borel-Cantelli lemma for the second term, the lemma is proved.

Theorem 7 (Hsieh and Turnbull (1996), Theorem 2.2). *Under the above conditions, there exists a probability space on which one can define sequences of two independent versions of Brownian bridges $\{B_1^{(n)}(t), B_2^{(m)}(t), 0 \leq t \leq 1\}$ such that*

$$\begin{aligned} \sqrt{m}(F_{1n}(F_{2m}^{-1}(t)) - F_1(F_2^{-1}(t))) &= \sqrt{\lambda}B_1^{(n)}(F_1(F_2^{-1}(t))) + \frac{f_1(F_2^{-1}(t))}{f_2(F_2^{-1}(t))}B_2^{(m)}(t) \\ &\quad + o(m^{-1/2}(\log m)^2) \text{ a.s.} \end{aligned} \quad (3.11)$$

uniformly on $[a, b]$.

Proof. This theorem follows from Theorem 4.4.1 in Csörgő and Révész (1981) and Theorem 3.2.4. in Csörgő (1983). The details can be found in Hsieh and Turnbull (1992).

From Theorem 7 follows that for all points of continuity of H as $n, m \rightarrow \infty$ we have

$$P \left\{ \sup_{a \leq t \leq b} |\sqrt{m}(F_{1n}(F_{2m}^{-1}(t)) - F_1(F_2^{-1}(t)))| \leq x \right\} = H(x),$$

where

$$H(x) = P \left\{ \sup_{a \leq t \leq b} \left| \sqrt{\lambda} B_1^{(n)}(F_1(F_2^{-1}(t))) + \frac{f_1(F_2^{-1}(t))}{f_2(F_2^{-1}(t))} B_2^{(m)}(t) \right| \leq x \right\}.$$

Corollary 8. *Under the null Hypothesis $F_1 \equiv F_2$ we have*

$$P \left\{ \sup_{a \leq t \leq b} \left| \sqrt{m} (F_{1n}(F_{2m}^{-1}(t)) - F_1(F_2^{-1}(t))) \right| \leq x \right\} = P \left(\sup_{a \leq t \leq b} \left| \sqrt{\lambda} B_1^{(n)}(t) + B_2^{(m)}(t) \right| \leq x \right), \quad (3.12)$$

where $B_1^{(n)}$ and $B_2^{(m)}$ are two independent versions of Brownian bridges on $[0, 1]$.

Another method to construct the simultaneous band is to use Bonferroni's method, which is based on the following inequality

$$P \left(\bigcup_{i=1}^g A_i \right) \geq 1 - \sum_{i=1}^g P[\bar{A}_i],$$

where $g \in \mathbb{N}$, which represents number of statements or comparisons, A_i and its complement \bar{A}_i are any event. In particular, if each A_i is the event that a calculated confidence interval for a particular linear combination, then the left-hand side of the inequality is the probability that all the confidence intervals simultaneously cover their respective values. The right-hand side is one minus the sum of the probabilities of each of the intervals missing their true values. Therefore, if simultaneous multiple interval estimates are desired with an overall confidence coefficient $1 - \alpha$, one can construct each interval with confidence coefficient $(1 - \alpha/g)$, and the Bonferroni inequality insures that the overall confidence coefficient is at least $1 - \alpha$.

Now on the interval $[\epsilon, 1 - \epsilon]$, where $\epsilon \rightarrow 0$ for any $\alpha_i \in (0, 1)$ we define c_i as the solution of the equation

$$P \left(\sup_{0 \leq t \leq 1} |B(t)| \leq c_i \right) = 1 - \alpha_i, \quad i = 1, 2, \quad (3.13)$$

where $\{B(t), 0 \leq t \leq 1\}$ denotes a Brownian bridge. The maximum of $|B(t)|$ turns out to be closely allied with the Kolmogorov-Smirnov goodness-of-fit test. The behavior of $\sup_{0 \leq t \leq 1} |B(t)|$ has been well investigated already by Smirnov (1939, 1948), Feller (1948), Kolmogorov (1941) etc.. It has the distribution function of form

$$G(x) = 1 + 2 \sum_{k=1}^{\infty} (-1)^{k-1} \exp(-k^2 x^2) = (2\pi)^{1/2} x^{-1} \sum_{k=1}^{\infty} \exp(-(2k-1)^2 \pi^2 / 8x^2).$$

The values c_1, c_2 in (3.13) can be found for example in Tables from Smirnov (1948).

Theorem 9. *If F_2 is continuous, $\epsilon_m \rightarrow 0$ and $\sqrt{m}\epsilon_m \rightarrow \infty$ as $m \rightarrow \infty$, then*

$$P \left(F_{1n}(F_{2m}^{-1}(t - c_1 m^{-1/2})) - c_2 n^{-1/2} \leq F_1(F_2^{-1}(t)) \leq F_{1n}(F_{2m}^{-1}(t + c_1 m^{-1/2})) + c_2 n^{-1/2} \right. \\ \left. \text{for all } \epsilon_m \leq t \leq 1 - \epsilon_m \right) \geq (1 - \alpha_1)(1 - \alpha_2), \quad (3.14)$$

where $c_1 = c_1(\alpha_1)$ and $c_2 = c_2(\alpha_2)$ are defined by (3.13).

Proof. First the weak convergence of the empirical process gives

$$\begin{aligned} P\left(F_{1n}(x) - c_2 n^{-1/2} \leq F_1(x) \leq F_{1n}(x) + c_2 n^{-1/2} \text{ for all } x\right) \\ = P\left(\sup_{-\infty < x < \infty} |B(F_1(x))| \leq c_2\right) \geq P\left(\sup_{0 \leq t \leq 1} |B(t)| \leq c_2\right). \end{aligned}$$

Csörgő and Révész (1984) (cf. also Csörgő and Horváth (1985)) showed that

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(F_{2m}^{-1}(t + c_1 m^{-1/2}) \leq F_2^{-1}(t) \leq F_{2m}^{-1}(t - c_1 m^{-1/2}) \right. \\ \left. \text{ for all } \epsilon_m \leq t \leq 1 - \epsilon_m\right) = P\left(\sup_{0 \leq t \leq 1} |B(t)| \leq c_1\right). \end{aligned}$$

We simulated some normally distributed random variables with different sample sizes $n = m = 100$ and $n = m = 500$ (cf. Figure 7). We found by simulations that for $\alpha = 0.05$ the critical value $c_\alpha = 1.8995$ for the confidence bands derived in (3.12), i.e. $P(\sup_{a \leq t \leq b} |\sqrt{\lambda} B_1^{(n)}(t) + B_2^{(m)}(t)| \leq c_\alpha) = 1 - \alpha$. For Bonferroni confidence bands derived in Theorem 9 we have chosen $c_1 = c_2 = 1.48$ from equation (3.13), which correspond to $\alpha_1 = \alpha_2 = 0.025$, thus we have at least $(1 - \alpha_1)(1 - \alpha_2) = 0.95$ coverage.

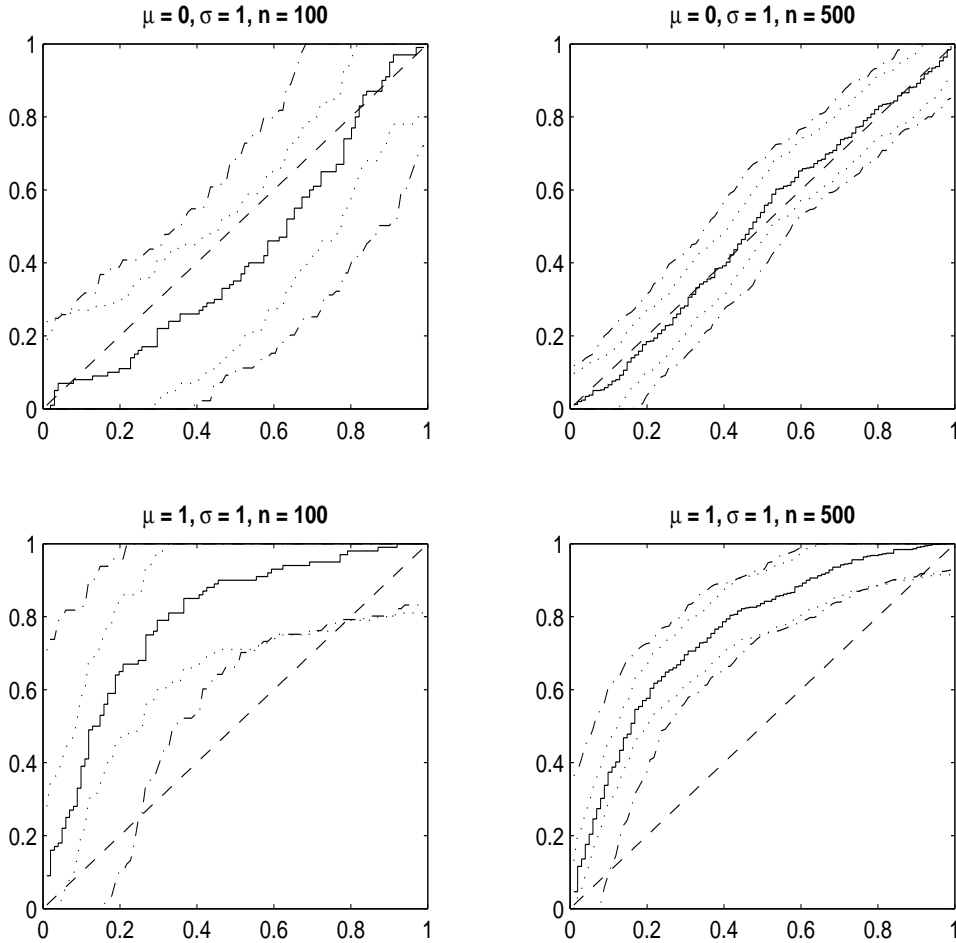


Figure 7: P-P empirical plots of simulated $N(\mu, \sigma)$ against $N(0, 1)$ with simultaneous confidence bands added for several sample sizes $n = m = 100, 500$. The Bonferroni bands (3.14) are plotted with brighter dashed line than the empirical bands (3.12) in all cases.

Bonferroni's method obviously provides wider confidence bands than those obtained in (3.12). It is remarkable that for P-P plots again it is easier to recognize whether there is a positive shift of the underlying distributions in the two sample case. However, we can not decide whether both samples come from the same family of distributions, which could have been recognized when dealing with Q-Q plots. Still if we are interested in such structural relationships as location or scale differences and not in underlying distributions, P-P plots might be more useful.

4 Confidence bands for general shift function Δ_s

Let X_1, \dots, X_n and Y_1, \dots, Y_m be two independent samples with distribution functions F_1 and F_2 . Comparing the distribution of two samples Doksum (1974, 1977) and Doksum and Sievers (1976) have constructed simultaneous confidence bands for general shift function $\Delta_s(x)$ such as $F_1(x) = F_2(x + \Delta_s(x))$ for some function $\Delta_s(x)$. If the X 's are control responses and the Y 's are treatment responses Doksum and Sievers (1976) explains that $\Delta_s(x)$ under certain conditions can be regarded as the amount the treatment adds to a potential control response x (see also Doksum, 1974). Under general conditions ($\Delta_s(x) + x$ should be nondecreasing) it is the only function of x that satisfies $X + \Delta_s(X) \sim Y$, where \sim denotes distributed as. Thus, $\Delta_s(\cdot)$ can be regarded as amount of 'shift' needed to bring the X 's up to the Y 's in distribution.

This structural "general shift" model clearly is a special case of structural relationship models described earlier. We shortly introduce now Doksum's methods here.

Let $F^{-I} = \sup\{x : F(x) \leq u\}$ be the right inverse of F . Then we can write

$$\Delta_s(x) = F_2^{-1}(F_1(x)) - x. \quad (4.1)$$

When $\Delta_s(x) = h$ for some constant $h \in \mathbb{R}$ we have the classical shift or location model (cf. Example 2.1). A natural estimate of $F_2^{-1}(F_1(x))$ is $F_{2m}^{-1}(F_{1n}(x))$, where F_{1n} and F_{2m} denote the empirical distribution functions based on X and Y samples. Doksum and Sievers (1976) state several questions motivating their work which we will cite here: (i) Is the treatment beneficial for all the members of the populations, i.e. is $\Delta_s > 0$ for all x ; (ii) If not, for which part of the population is the treatment beneficial, i.e. what is $\{x : \Delta_s(x) > 0\}$? (iii) Does a shift model hold, i.e. is $\Delta_s(x) = h$, for some h and all x ? (vi) If not, does a shift-scale model hold, i.e. is $\Delta_s(x) = \alpha + \beta x$, for some α and β and for all x ?

Giving now a confidence band $[\Delta_{s*}(x), \Delta_s^*(x)]$ for $\Delta_s(x)$ simultaneously for all x these questions can be answered in the following way: (i) is answered in the affirmative if $\Delta_{s*}(x) > 0$ for all x ; (ii) has solution $\{x : \Delta_{s*}(x) > 0\}$; (iii) is rejected if no horizontal line fits in the confidence band and (vi) has a negative response if no straight line fits in the band.

Remark 10. *This remark is to explain the following: if no straight line fits in the simultaneous bands for the function Δ_s , then we can reject the hypothesis that $\Delta_s(x) = \alpha + \beta x$ for all x with the confidence of $1 - \alpha$, where α is the level of a test and α, β some constants. Under the hypothesis $H_0 : F_1(x) = F_2(\alpha + \beta x)$ we know that the Q-Q plot is a straight line. To see this let for simplicity $X \sim N(\mu, \sigma^2)$. Now*

$$P(X \leq q) = P\left(\frac{X - \mu}{\sigma} \leq \frac{q - \mu}{\sigma}\right) = P\left(Z \leq \frac{q - \mu}{\sigma}\right), \quad (4.2)$$

where $Z \sim N(0, 1)$. This means that the quantile point q of $N(\mu, \sigma^2)$ corresponds to the quantile point $(q - \mu)/\sigma$ of $N(0, 1)$. Clearly under the hypothesis H_0 also the function Δ_s is a straight line. Under H_0 the simultaneous confidence band for the general shift function Δ_s should contain every point on this line with the confidence $1 - \alpha$. Therefore if there is no straight line which would fit into the simultaneous bands, then with $1 - \alpha$ confidence we can reject the hypothesis H_0 .

Let $N = n + m$, $M = mn/N$. Doksum and Sievers (1976) considered bands based on the two sample Kolmogorov-Smirnov statistic

$$T(F_{1n}, F_{2m}) = \sqrt{M} \sup_x |F_{1n}(x) - F_{2m}(x)|. \quad (4.3)$$

Lemma 11 (Doksum and Sievers (1976), Remark 1). *A-level $(1 - \alpha)$ simultaneous distribution-free confidence band for $\Delta_s(x)$ ($-\infty < x < \infty$) is given by*

$$[F_{2m}^{-1}(F_{1n}(x) - K_{S,\alpha}/M^{1/2}) - x, F_{2m}^{-I}(F_{1n}(x) + K_{S,\alpha}/M^{1/2}) - x], \quad (4.4)$$

where $K_{S,\alpha}$ has been chosen from the Kolmogorov-Smirnov tables, so that $P(T(F_{1n}, F_{2m}) \leq K_{S,\alpha}) = 1 - \alpha$.

Proof. Let us give a sketch of the proof. The idea already comes from Switzer (1976), who noted that any two-sample rank test may be inverted in principle to obtain distribution-free confidence sets. Now if $\phi(F_{1n}, F_{2m})$ is a distribution-free level α test function under $H_0 : F_1 = F_2$, then $\{\Delta_s(\cdot) : \phi(F_{1n}, F_{2m}(\Delta_s(y) + y)) = 0\}$ is a distribution-free level $(1 - \alpha)$ confidence region for the response function $\Delta_s(\cdot)$. We get so-called simple bands if we consider distribution-free test statistics $T(F_{1n}, F_{2m})$ with the property that the inequality $T(F_{1n}, F_{2m}) \leq K$ is equivalent to

$$h_*(F_{1n}(x)) \leq F_{2m}(x) \leq h^*(F_{1n}(x))$$

for all x , for some functions h_* and h^* . Clearly for the two-sample Kolmogorov-Smirnov statistic (4.3) we have $h_*(x) = x - K/\sqrt{M}$ and $h^* = x + K/\sqrt{M}$.

Corollary 12. *Let $[t]$ denote the greatest integer less than or equal to t ; let $\langle t \rangle$ be the least integer greater than or equal to t ; let $X(1) < \dots < X(n)$ and $Y(1) < \dots < Y(m)$ denote the order statistics of the X and Y samples, and define $Y(j) = -\infty (j < 0)$ and $Y(j) = \infty (j \geq m + 1)$. Then the band (4.4) can be expressed by*

$$[S_*(x), S^*(x)] = \left[Y \left(\left\langle m \left(\frac{i}{n} - K_{S,\alpha}/\sqrt{M} \right) \right\rangle \right) - x, Y \left(\left[m \left(\frac{i}{n} + K_{S,\alpha}/\sqrt{M} \right) \right] + 1 \right) - x \right] \quad (4.5)$$

for $x \in [X(i), X(i + 1)) (i = 0, 1, \dots, n)$ with $X(0) = -\infty$ and $X(n + 1) = \infty$.

Doksum and Sievers (1976) also considered bands based on a weighted sup norm statistic (further on denoted as weighted bands).

$$W_N = W_N(F_{1n}, F_{2m}) = \sqrt{M} \sup_{\{x:a \leq F_{1n}(x) \leq b\}} \frac{|F_{1n}(x) - F_{2m}(x)|}{\Psi\{H_N(x)\}}, \quad (4.6)$$

where $H_N(x) = \lambda F_{1n}(x) + (1 - \lambda)F_{2m}(x)$, $\lambda = n/N$ and $0 \leq a < b \leq 1$. Choosing $\Psi(t) = \{t(1 - t)\}^{1/2}$ one can get additional advantages, which we do not discuss here (cf. Doksum and Sievers, 1976; Borovkov and Sycheva, 1968).

Lemma 13 (Doksum and Sievers (1976), Remark 2). *Let $P_{F=G}(W_N \leq K) = 1 - \alpha$; then the level $(1 - \alpha)$ simultaneous confidence band for $\Delta_s(x)$ based on W_N with $\Psi(t) = \{t(1 - t)\}^{1/2}$ is*

$$[F_{2m}^{-1}[h^-(F_{1n}(x))] - x, F_{2m}^{-I}[h^+(F_{1n}(x))] - x], \quad x \in \{x : a \leq F_{1n}(x) \leq b\}, \quad (4.7)$$

where

$$\lambda^\pm(u) = \frac{u + 0.5c(1 - \lambda)(1 - 2\lambda u) \pm 0.5\{c^2(1 - \lambda)^2 + 4cu(1 - u)\}^{1/2}}{1 + c(1 - \lambda)^2}$$

and $c = K^2/M$.

When $a = 1 - b = 0$ the values of K can be found in Canner (1975). In Figure 8 both bands derived in (4.5) and in (4.7) are plotted for sample X and sample Y from $N(0, 1)$ and $N(0, 4)$ distributions, respectively. Here $n = m = 100$, $\alpha = 0.05$, and $K = 3.02$ is obtained from Canner (1975). Both bands are quite similar in the middle, but in the tails weighted bands are better, providing narrower confidence bands. In this case it is obvious that a straight line fits in the bands. That means the question (vi) can not be answered negatively (cf. Remark 10) and this is not surprising, because we have a location-scale difference between two Normal distributions.

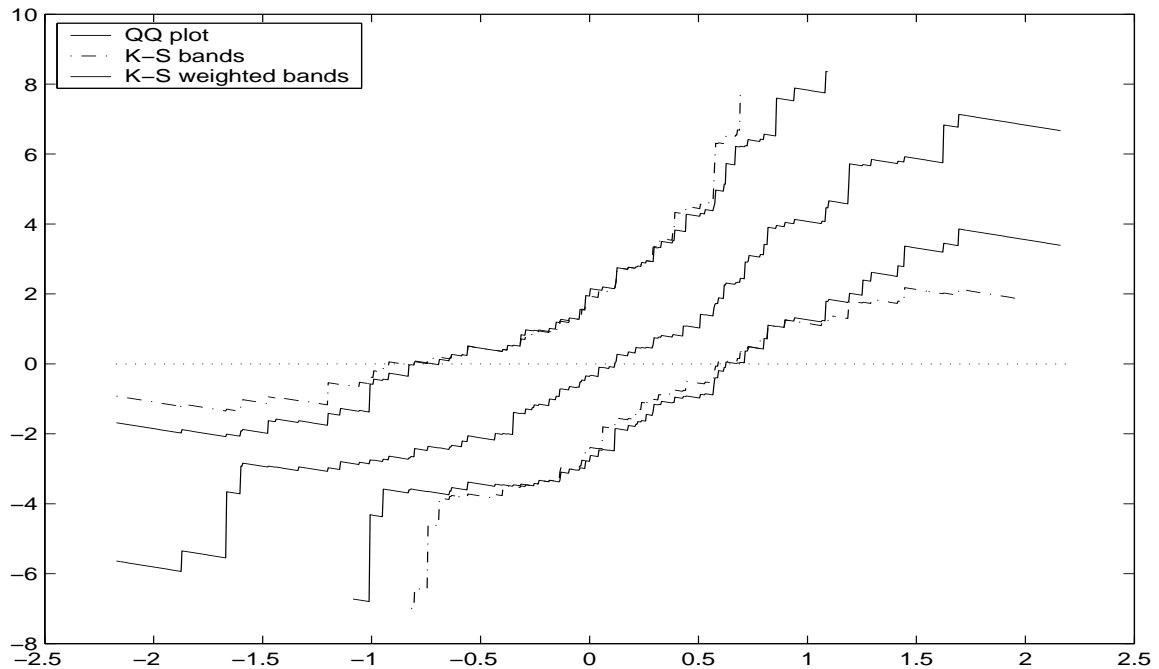


Figure 8: Simultaneous confidence bands for Δ_s constructed for $N(0,1)$ and $N(1,4)$ samples, $\alpha = 0.05$, $n = m = 100$.

5 Empirical likelihood

Maximum likelihood is probably the most important concept for inference in parametric models. For last decades it has been broadly extended to nonparametric and semi-parametric frameworks. Motivated from the earlier work of Thomas and Grukemeier (1975) in the context of estimating survival probabilities, Owen (1988, 1990, 1991) has introduced an "empirical" likelihood ratio statistic for nonparametric problems. The empirical likelihood method amounts to computing the profile likelihood of a general multinomial distribution which has its atoms at data points. Owen's remarkable contribution was to show that the idea has very general applications.

The empirical likelihood method is similar to for the parametric models, although it is computationally more complex. Owen has shown that usually the statistics have limiting chi-square distributions, which is a nonparametric extension of Wilks's (1938) theorem for parametric likelihood ratios. He has also shown how to obtain tests and confidence limits for parameters, expressed as functionals $\theta(F)$ of an unknown distribution function F . Nice accounts of empirical likelihood may be found in Hall and La Scala (1990) and Owen (2001)'s recent book, where one can find also a huge bibliography. Let us shortly cite Hall and La Scala (1990): "The advantages of empirical likelihood over classical methods such as normal approximation are rather obvious; for example, empirical likelihood regions are not shaped in a predetermined way which implies a degree of nonexistent symmetry in the sampling distribution".

Now we list the main advantages of empirical likelihood according to Hall and La Scala (1990):

- (i) Empirical likelihood methods are very appealing especially for construction of confidence regions, since they involve no prespecified assumptions about the shape of the confidence regions which is determined automatically by the sample;
- (ii) Empirical likelihood regions are Bartlett correctable. That is, a simple correction for the mean of the empirical loglikelihood ratio reduces coverage error from order n^{-1} to order n^{-2} , where n denotes the sample size;
- (iii) Empirical likelihood regions are range preserving and transformation respecting.

5.1 Nonparametric maximum likelihood

For this section let X_1, \dots, X_n be a sample of i.i.d. random variables with distribution function F . Our goal is to define the nonparametric maximum likelihood in one-sample case and show that the empirical cumulative distribution function F_n maximizes it. We use $F(x-)$ to denote $P(X < x)$ and so $P(X = x) = F(x) - F(x-)$.

Definition 2. Let $X_1, \dots, X_n \in \mathbb{R}$ be i.i.d. random variables. The nonparametric likelihood of the cumulative distribution function F is

$$L(F) = \prod_{i=1}^n (F(X_i) - F(X_i-)). \quad (5.1)$$

Definition 2 reflects a very literal interpretation of the notion of likelihood. Practically it means that the value $L(F)$ is the probability of getting exactly the observed sample values X_1, \dots, X_n from the distribution function F . One consequence is that $L(F) = 0$ if F is a

continuous distribution. To have a positive nonparametric likelihood, a distribution F must place positive probability on every one of the observed data values.

The next theorem shows that the nonparametric likelihood is maximized by the empirical distribution function F_n . Thus, F_n is a nonparametric maximum likelihood estimate of F .

Theorem 14 (Owen (2001), Theorem 2.1). *Let $X_1, \dots, X_n \in \mathbb{R}$ be i.i.d. r.v.'s with a common distribution function G . Let F_n be their cumulative empirical distribution function and let $G \neq F_n$, then $L(G) < L(F_n)$.*

Proof. Let $z_1 < z_2 < \dots < z_m$ be the distinct values in $\{X_1, \dots, X_n\}$, and let $n_j \geq 1$ be the number of X_i that are equal to z_j . Let $p_j = G(z_j) - G(z_j -)$ and put $\hat{p}_j = n_j/n$. If $p_j = 0$ for any $j = 1, \dots, m$, then $L(G) = 0 < L(F_n)$, so we suppose that all $p_j > 0$, and that for at least one j , $p_j \neq \hat{p}_j$. Now $\log(x) \leq x - 1$ for all $x > 0$ with equality only when $x = 1$. Therefore

$$\log\left(\frac{L(G)}{L(F_n)}\right) = \sum_{j=1}^m n_j \log\left(\frac{p_j}{\hat{p}_j}\right) = n \sum_{j=1}^m \hat{p}_j \log\left(\frac{p_j}{\hat{p}_j}\right) < n \sum_{j=1}^m \hat{p}_j \left(\frac{p_j}{\hat{p}_j} - 1\right) \leq 0, \quad (5.2)$$

and so $L(G) < L(F_n)$.

In parametric inference we may base hypothesis tests and confidence regions on the likelihood ratio. If $L(\eta)$ is much smaller than $L(\hat{\eta})$, then we reject the hypothesis that $\eta_0 = \eta$, and exclude η from our confidence region for η_0 . Under some regularity conditions, Wilks's theorem provides that $-2 \log(L(\eta_0)/L(\hat{\eta}))$ tends to a chisquared distribution as $n \rightarrow \infty$ (cf. Wilks, 1938). This allows us to decide how small $L(\eta)$ must be in order for η to get rejected. The degrees of freedom in the chisquared distribution are usually equal to the dimension of the set of η values. When we want a confidence region for θ we take the image of a confidence region for η . That is

$$\{\theta(\eta) | L(\eta) \geq cL(\hat{\eta})\}, \quad (5.3)$$

where the threshold c is chosen using Wilk's theorem, with degrees of freedom equal to the dimension of the set of θ values. We may also use ratios of the nonparametric likelihood as a basis for hypothesis tests and confidence intervals. For a distribution F , define

$$R(F) = \frac{L(F)}{L(F_n)}, \quad (5.4)$$

through the nonparametric likelihood $L(F)$ of Definition 2. We can proceed analogously as with parametric likelihood. Suppose we are interested in parameter $\theta = T(F)$ for some function T of distribution. This F is a member of a set \mathcal{F} of distributions. In some cases we may take \mathcal{F} to be the set of all distributions on \mathbb{R} . More often a smaller set of distributions is used. Define the profile likelihood ratio function:

$$R(\theta) = \sup\{R(F) | T(F) = \theta, F \in \mathcal{F}\}. \quad (5.5)$$

Now empirical likelihood hypothesis tests reject $H_0 : T(F_0) = \theta_0$, when $R(\theta_0) < r_0$ for some threshold value r_0 . Empirical likelihood confidence regions are of the form

$$\{\theta | R(\theta) \geq r_0\}. \quad (5.6)$$

In many settings, the threshold r_0 may be chosen using an empirical likelihood theorem, a nonparametric analogue of Wilk's theorem.

Empirical likelihood inferences may be recognized as parametric likelihood inferences using a data-determined parametric family. The parametric family involved is the multinomial distribution on the observed values of X_i .

5.2 EL in one sample case

To show the idea of the empirical likelihood method we first outline empirical likelihood as discussed by Owen (1988, 1990) in one sample case. Let x_1, \dots, x_n be i.i.d. observations from a d -variate distribution F having mean μ and nonsingular covariance matrix. The empirical likelihood function is

$$L(F) = \prod_{i=1}^n dF(x_i) = \prod_{i=1}^n p_i, \quad (5.7)$$

where $p_i = dF(x_i) = P(X = x_i)$. Only distributions with an atom of probability on each x_i have nonzero likelihood. From Theorem 14 it follows that empirical likelihood function (5.7) is maximized by the empirical distribution function $F_n(x) = n^{-1} \sum_{i=1}^n I(x_i \leq x)$. The empirical likelihood ratio is then defined as $R(F) = L(F)/L(F_n)$, and clearly this may be written as

$$R(F) = \prod_{i=1}^n np_i.$$

Suppose we want to estimate a parameter $\theta = T(F)$. For simplicity we consider the mean μ of F . To obtain confidence regions for μ , we define the profile empirical likelihood ratio function

$$R_E(\mu) = \sup \left\{ \prod_{i=1}^n np_i \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i x_i = \mu \right\}. \quad (5.8)$$

As noted by Owen (1988, 1990), a unique value for the right-hand side of (5.8) exists, provided that μ is inside the convex hull of the points x_1, \dots, x_n . An explicit expression for $R_E(\mu)$ can be derived by a Lagrange multiplier argument: the maximum of $\prod_{i=1}^n np_i$ subject to the constraints $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$ and $\sum_{i=1}^n p_i x_i = \mu$ is attained when

$$p_i = p_i(\mu) = \frac{1}{n} \frac{1}{1 + \lambda^\tau (x_i - \mu)},$$

where τ stands for the transpose for a vector and $\lambda = \lambda(\mu)$ is a $d \times 1$ vector given as the solution to

$$\sum_{i=1}^n \frac{x_i - \mu}{1 + \lambda^\tau (x_i - \mu)} = 0.$$

Since $\prod_{i=1}^n p_i$ is maximized unconditionally by F_n , it follows that $R_E(\mu)$ is maximized with respect to μ at $\hat{\mu} = \bar{x}$ and that

$$R_E(\mu) = \prod_{i=1}^n \{1 + \lambda^\tau (x_i - \mu)\}^{-1}.$$

The empirical likelihood ratio statistic is $W_E(\mu) = -2 \log R_E(\mu)$, that is,

$$W_E(\mu) = 2 \sum_{i=1}^n \log \{1 + \lambda^\tau (x_i - \mu)\}. \quad (5.9)$$

Owen (1988, 1990) has proved under mild conditions that if $\mu = \mu_0$, then $W_E(\mu_0)$ converges in distribution to $\chi_{(d)}^2$ as $n \rightarrow \infty$. Approximate α -level confidence regions for μ may therefore be obtained as the set of points μ such that $W_E(\mu) \leq c_\alpha$, where c_α satisfies $P(\chi_{(d)}^2 \leq c_\alpha) = \alpha$.

Since then it has been proved that the empirical likelihood approach can be applied for quite general parameters $\theta(F)$, including multidimensional M -estimates, etc. (cf. Owen 2001). A very important extension in the one sample case has been made by Qin and Lawless (1994), which is certainly one of the most remarkable papers for the empirical likelihood method. We shortly describe this general setup.

Consider d -variate i.i.d. random variables x_1, \dots, x_n with unknown distribution function F , and a p -dimensional parameter θ associated with F . We assume that information about θ and F is available in the form of $r \geq p$ functionally independent unbiased estimating functions, that is functions $g_j(x, \theta)$, $j = 1, 2, \dots, r$, such that $\mathbb{E}_F\{g_j(x, \theta)\} = 0$. Applying now empirical likelihood method to this framework by maximizing (5.7) subject to restrictions

$$p_i \geq 0, \quad \sum_i p_i = 1, \quad \sum_i p_i g(x_i, \theta) = 0, \quad (5.10)$$

where $g(x, \theta) = (g_1(x, \theta), \dots, g_r(x, \theta))^T$ and $\mathbb{E}_F\{g(x, \theta)\} = 0$.

Now for a given θ , a unique maximum exists, provided that 0 is inside the convex hull of the points $g(x_1, \theta), \dots, g(x_n, \theta)$. The maximum may be found via Lagrange multipliers. Let

$$H = \sum_{i=1}^n \log p_i + \tau_1 \left(1 - \sum_{i=1}^n p_i\right) - n\lambda^\tau \sum_{i=1}^n p_i g(x_i, \theta), \quad (5.11)$$

where τ_1 and $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_r)^\tau$ are Lagrange multipliers. Taking derivatives with respect to p_i , we have

$$\begin{aligned} \frac{\partial H}{\partial p_i} &= \frac{1}{p_i} - \tau_1 - n\lambda^\tau g(x_i, \theta) = 0, \\ \sum_{i=1}^n p_i \frac{\partial H}{\partial p_i} &= n - \tau_1 \Rightarrow \tau_1 = n \end{aligned}$$

and

$$p_i = \left(\frac{1}{n}\right) \frac{1}{1 + \lambda^\tau g(x_i, \theta)},$$

with the restriction from the third part of (5.10) that

$$0 = \sum_{i=1}^n p_i g(x_i, \theta) = \frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \lambda^\tau g(x_i, \theta)} g(x_i, \theta),$$

from which (see below) λ can be determined in terms of θ .

It is necessary that $0 \leq p_i \leq 1$, which implies that λ and θ must satisfy $1 + \lambda^\tau g(x_i, \theta) \geq 1/n$ for each i . For fixed θ , let $D_\theta = \{\lambda : 1 + \lambda^\tau g(x_i, \theta) \geq 1/n\}$; D_θ is convex and closed, and it is bounded if 0 is inside the convex hull of the $g(x_i, \theta)$'s. Moreover,

$$\frac{\partial}{\partial \lambda} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \lambda^\tau g(x_i, \theta)} g(x_i, \theta) \right\} = -\frac{1}{n} \sum_{i=1}^n \frac{g(x_i, \theta) g^\tau(x_i, \theta)}{\{1 + \lambda^\tau g(x_i, \theta)\}^2} \quad (5.12)$$

is negative definite for λ in D_θ , provided that $\sum_{i=1}^n g(x_i, \theta)g^\tau(x_i, \theta)$ is positive definite. By the inverse theorem, $\lambda = \lambda(\theta)$ is a continuous differentiable function of θ .

Now again solving the maximization problem via Lagrange multipliers we obtain the empirical likelihood ratio

$$l_E(\theta) = \sum_{i=1}^n \log[1 + \lambda^\tau(\theta)g(x_i, \theta)]. \quad (5.13)$$

Clearly choosing $g(x_i, \mu) = x_i - \mu$ we have the case described before with $2l_E(\theta) = W_E(\mu)$, where $W_E(\mu)$ defined in (5.9).

Let $\tilde{\theta} = \arg \min_\theta l(\theta)$. Qin and Lawless (1994) have derived general conditions when the maximization problem described above has a solution and when the ratio statistic $2l_E(\theta_0) - 2l_E(\tilde{\theta})$ has a limiting χ_p^2 distribution.

Lemma 15 (Qin and Lawless, 1994, Lemma 1). *Assume that $\mathbb{E}\{g(x, \theta_0)g^\tau(x, \theta_0)\}$ is positive definite, $\partial g(x, \theta)/\partial\theta$ is continuous in a neighborhood of the true value θ_0 , $\|\partial g(x, \theta)/\partial\theta\|$ and $\|g(x, \theta)\|^3$ are bounded by some integrable function $G(x)$ in the neighborhood, and the rank of $\mathbb{E}\partial g(x, \theta_0)/\partial\theta$ is p . Then, as $n \rightarrow \infty$, with probability 1 $l_E(\theta)$ attains its minimum value at some point $\tilde{\theta}$ in the interior of the ball $\|\theta - \theta_0\| \leq n^{-1/3}$.*

Theorem 16 (Qin and Lawless, 1994, Theorem 2). *In addition to the conditions of Lemma 15, assume that $\partial^2 g(x, \theta)/\partial\theta\partial\theta^\tau$ is continuous in θ in a neighborhood of the true value θ_0 . Then if $\|\partial^2 g(x, \theta)/\partial\theta\partial\theta^\tau\|$ can be bounded by some integrable function $G(x)$ in the neighborhood, then the empirical likelihood ratio statistic for testing $H_0 : \theta = \theta_0$ is*

$$W_E(\theta_0) = 2l_E(\theta_0) - 2l_E(\tilde{\theta}) \rightarrow_d \chi_{(p)}^2$$

as $n \rightarrow \infty$, when H_0 is true, where $l_E(\theta)$ is given by (5.13).

When $r = p$ than the methods developed in Qin and Lawless (1994) are the same as those of Owen (1988, 1990). However, in the case where $r > p$ we can deal with the combination of pieces of information about a distribution. For illustration we give some examples, which are taken from Qin and Lawless (1994).

Example 3. There are situations when the information relating the first and second moments of the variable is given. For example, let y_1, \dots, y_m be i.i.d., univariate observations with mean θ , and suppose that it is known that $\mathbb{E}(y^2) = m(\theta)$, where $m(\cdot)$ is a known function. Our goal is to estimate θ . The information about F can be expressed taking

$$g(y, \theta) = (y - \theta, y^2 - m(\theta))^\tau.$$

Example 4. A nonparametric estimation of a distribution F when information about certain functions of F is available is commonly observed in the literature. For example, Haberman (1984) and Sheehy (1988) consider estimation of $F(x)$ based on i.i.d. sample x_1, \dots, x_n when it is known that $E_F\{T(x)\} = a$, for some specified function $T(\cdot)$. Now taking $g(x) = T(x) = a$; that is, $r = 1$ and the dimension p of θ is 0.

5.3 EL in two sample case

EL has been extended for two sample problems by Qin and Zhao (2000), Qin (1997), Jing (1995), Claeskens *et al.* (2003), Jing and Zhou (2003) and others. We will follow closely the setup of Qin and Zhao (2000), which is general enough for our purposes.

First we describe the general framework. Let i.i.d. r.v.'s X_1, \dots, X_n and Y_1, \dots, Y_m have unknown distribution functions F_1 and F_2 , which are elements of some class \mathcal{F} . We are interested in construction of confidence intervals for some function $t \rightarrow \Delta(t)$ defined on an interval \mathcal{T} (further on we write only Δ). For that reason we will use empirical likelihood, which models F_1 and F_2 by multinomial distributions concentrated of the observations. Let θ_0 be some univariate parameter associated with one of the distributions F_1 or F_2 . We assume that all information about θ_0, Δ, F_1 and F_2 is available in the known form of unbiased estimating functions, i.e.

$$E_{F_1} w_1(X, \theta_0, \Delta, t) = 0, \quad (5.14)$$

$$E_{F_2} w_2(Y, \theta_0, \Delta, t) = 0. \quad (5.15)$$

Remark 17. Let $\Delta = \theta_1 - \theta_0$, where θ_0 and θ_1 are univariate parameters associated with F_1 and F_2 respectively. Then we have exactly the setup of Qin and Zhao (2000). The reason of our formulation of the problem is to have some flexibility in function Δ . This includes such an important case as P-P plots (cf. Example 7) analyzed via empirical likelihood by Claeskens *et al.* (2003) and also Q-Q plots (cf. Example 8). Therefore the results of Claeskens *et al.* (2003) actually follow from the results of Qin and Zhao (2000).

Example 5. Denote $\theta_0 = \int x dF_1(x)$ and $\Delta = \int y dF_2(y) - \int x dF_1(x)$. We can obtain the forms (5.14) and (5.15) by taking

$$w_1(X, \theta_0, \Delta, t) = X - \theta_0, \quad w_2(Y, \theta_0, \Delta, t) = Y - \theta_0 - \Delta.$$

Example 6. For a given t_0 , putting $\theta_0 = F_1(t_0)$ and $\Delta = F_2(t_0) - F_1(t_0)$ again we can obtain the forms (5.14) and (5.15) by taking

$$w_1(X, \theta_0, \Delta, t) = I_{\{X \leq t_0\}} - \theta_0, \quad w_2(Y, \theta_0, \Delta, t) = I_{\{Y \leq t_0\}} - \theta_0 - \Delta.$$

Example 7. Denote $\theta_0 = F_2^{-1}(t)$ and $\Delta = F_1(F_2^{-1}(t))$, which is the P-P plot of functions F_1 and F_2 (cf. Section 3). In this case

$$w_1(X, \theta_0, \Delta, t) = I_{\{X \leq \theta_0\}} - \Delta, \quad w_2(Y, \theta_0, \Delta, t) = I_{\{Y \leq \theta_0\}} - t.$$

This setting also covers the case of ROC curves defined as $\Delta = 1 - F_1(F_2^{-1}(1 - t))$, which is important tool used to summarize the performance of a medical diagnostic test for determining whether a patient has a disease or not.

Example 8. Denote $\theta_0 = F_2(t)$ and $\Delta = F_1^{-1}(F_2(t))$, which is well known as a quantile-quantile Q-Q plot (cf. Section 3). We have

$$w_1(X, \theta_0, \Delta, t) = I_{\{X \leq \Delta\}} - \theta_0, \quad w_2(Y, \theta_0, \Delta, t) = I_{\{Y \leq t\}} - \theta_0.$$

In the two sample case the empirical likelihood function is

$$L(F_1, F_2) = \prod_{i=1}^n dF_1(X_i) \prod_{j=1}^m dF_2(Y_j) = \prod_{i=1}^n p_i \prod_{j=1}^m q_j, \quad (5.16)$$

where $p_i = dF_1(X_i) = \Pr(X = X_i)$ and $q_j = dF_2(Y_j) = \Pr(Y = Y_j)$. It is easy to see that (5.16) is maximized by

$$\prod_{i=1}^n dF_{1n}(X_i) \prod_{j=1}^m dF_{2m}(Y_j) = n^{-n} m^{-m},$$

where F_{1n} and F_{2m} are the respective empirical distribution functions. The empirical likelihood ratio is defined then as

$$R(F_1, F_2) = L(F_1, F_2) / L(F_{1n}, F_{2m}) = \prod_{i=1}^n n p_i \prod_{j=1}^m m q_j.$$

For fixed $t \in \mathcal{T}$ to obtain confidence regions for the function Δ , we define the profile empirical likelihood ratio function

$$R(\Delta, t) = \sup_{\theta, p, q} \prod_{i=1}^n (n p_i) \prod_{j=1}^m (m q_j), \quad (5.17)$$

where $p = (p_1, \dots, p_n)$ and $q = (q_1, \dots, q_m)$ are subject to restrictions

$$p_i \geq 0, \quad \sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i w_1(X_i, \theta, \Delta, t) = 0, \quad (5.18)$$

$$q_j \geq 0, \quad \sum_{j=1}^m q_j = 1, \quad \sum_{j=1}^m q_j w_2(Y_j, \theta, \Delta, t) = 0. \quad (5.19)$$

Since

$$R(\Delta, t) = \sup_{\theta} \sup_{p, q} \prod_{i=1}^n (n p_i) \prod_{j=1}^m (m q_j), \quad (5.20)$$

we first investigate

$$R(\Delta, \theta, t) = \sup_{p, q} \prod_{i=1}^n (n p_i) \prod_{j=1}^m (m q_j), \quad (5.21)$$

where $p = (p_1, \dots, p_n)$ and $q = (q_1, \dots, q_m)$ are subject to restrictions (5.18) and (5.19). For a given θ , a unique maximum exists, provided that 0 is inside the convex hull of the points $w_1(X_i, \theta, \Delta, t)$'s and the convex hull of the $w_2(Y_j, \theta, \Delta, t)$'s. The maximum may be found via Lagrange multipliers. Let

$$\begin{aligned} H = & \sum_{i=1}^n \log p_i + \sum_{j=1}^m \log q_j + \tau_1 \left(1 - \sum_{i=1}^n p_i \right) + \tau_2 \left(1 - \sum_{j=1}^m q_j \right) \\ & - n \lambda_1 \sum_{i=1}^n p_i w_1(X_i, \theta, \Delta, t) - m \lambda_2 \sum_{j=1}^m q_j w_2(Y_j, \theta, \Delta, t), \quad (5.22) \end{aligned}$$

where $\tau_1, \tau_2, \lambda_1$ and λ_2 are Lagrange multipliers. Taking derivatives with respect to p_i and q_i , we have

$$\frac{\partial H}{\partial p_i} = \frac{1}{p_i} - \tau_1 - n \lambda_1 w_1(X_i, \theta, \Delta, t) = 0,$$

$$\sum_{i=1}^n p_i \frac{\partial H}{\partial p_i} = n - \tau_1 \Rightarrow \tau_1 = n$$

thus

$$p_i = \frac{1}{n(1 + \lambda_1(\theta)w_1(X_i, \theta, \Delta, t))}, \quad i = 1, \dots, n. \quad (5.23)$$

In a similar way we get

$$q_j = \frac{1}{m(1 + \lambda_2(\theta)w_2(Y_j, \theta, \Delta, t))}, \quad j = 1, \dots, m. \quad (5.24)$$

From restrictions (5.18) and (5.19) Lagrange multipliers λ_1 and λ_2 can be determined in terms of θ by the equations

$$\frac{1}{n} \sum_{i=1}^n \frac{w_1(X_i, \theta, \Delta, t)}{1 + \lambda_1 w_1(X_i, \theta, \Delta, t)} = 0, \quad (5.25)$$

$$\frac{1}{m} \sum_{j=1}^m \frac{w_2(Y_j, \theta, \Delta, t)}{1 + \lambda_2 w_2(Y_j, \theta, \Delta, t)} = 0. \quad (5.26)$$

As in the one sample case λ_1 and λ_2 can be determined in terms of θ (cf. page 29).

The (profile) empirical likelihood function for Δ is now defined as

$$R_E(\Delta, \theta, t) = \prod_{i=1}^n \left\{ \frac{1}{1 + \lambda_1(\theta)w_1(X_i, \theta, \Delta, t)} \right\} \prod_{j=1}^m \left\{ \frac{1}{1 + \lambda_2(\theta)w_2(Y_j, \theta, \Delta, t)} \right\},$$

By analogy with (5.9) we define the empirical log-likelihood ratio (multiplied by minus two) as

$$-2 \log R_E(\Delta, \theta, t) = 2 \sum_{i=1}^n \log(1 + \lambda_1(\theta)w_1(X_i, \theta, \Delta, t)) + 2 \sum_{j=1}^m \log(1 + \lambda_2(\theta)w_2(Y_j, \theta, \Delta, t)). \quad (5.27)$$

To find $\theta = \theta_E^{(n,m)}$ that maximizes $R(\Delta, t)$ we set $\partial\{-2 \log R_E(\Delta, \theta, t)\}/\partial\theta = 0$. We obtain the empirical likelihood equation as follows:

$$\frac{1}{n} \lambda_1(\theta) \sum_{i=1}^n \frac{\alpha_1(X_i, \theta, \Delta, t)}{1 + \lambda_1(\theta)w_1(X_i, \theta, \Delta, t)} + \frac{1}{m} \lambda_2(\theta) \sum_{j=1}^m \frac{\alpha_2(Y_j, \theta, \Delta, t)}{1 + \lambda_2(\theta)w_2(Y_j, \theta, \Delta, t)} = 0, \quad (5.28)$$

where

$$\alpha_1(X_i, \theta, \Delta, t) = \frac{\partial w_1(X_i, \theta, \Delta, t)}{\partial \theta} \quad \text{and} \quad \alpha_2(Y_j, \theta, \Delta, t) = \frac{\partial w_2(Y_j, \theta, \Delta, t)}{\partial \theta}.$$

Here we assume that $\partial w_1(X_i, \theta, \Delta, t)/\partial \theta$ and $\partial w_2(Y_j, \theta, \Delta, t)/\partial \theta$ exist.

The pointwise EL confidence interval for each fixed $t \in \mathcal{T}$ and for our function of interest Δ has the following form $\Delta : \{R(\Delta, \theta_E^{(n,m)}, t) > c\}$ for the true Δ_0 , where $\theta_E^{(n,m)}$ is a root of (5.28). The constant c can be calibrated using the result of Qin and Zhao (2000).

They make the following assumptions:

(i) $\theta_0 \in \Theta$ and Θ is an open interval;

(ii) $\mathbb{E}_{F_1} w_1^2(x, \theta, \Delta) > 0$ and $\mathbb{E}_{F_2} w_2^2(y, \theta, \Delta) > 0$, $\alpha_1(x, \theta, \Delta)$ and $\alpha_2(y, \theta, \Delta)$ are continuous in a neighborhood of θ_0 , $\alpha_1(x, \theta, \Delta)$ and $w_1^3(x, \theta, \Delta)$ are bounded by some integrable function $G_1(x)$ in this neighborhood, $\alpha_2(y, \theta, \Delta)$ and $w_2^3(y, \theta, \Delta)$ are bounded by some integrable function $G_2(y)$ in this neighborhood, and $E_{F_1} \alpha_1(x, \theta, \Delta)$ and $E_{F_2} \alpha_2(y, \theta, \Delta)$ are nonzero;

(iii) $m/n \rightarrow k$ (as $n, m \rightarrow \infty$) and $0 < k < \infty$.

Theorem 18 (Qin and Zhao (2000), Theorem 1). *If the assumptions (i) through (iii) are satisfied, then there exists a root $\theta_E^{(n,m)}$ of (5.28) such that $\theta_E^{(n,m)}$ is a consistent estimate for θ_0 , $R(\Delta, \theta)$ attains its local maximum value at $\theta_E^{(n,m)}$, and*

$$-2 \log R_E(\Delta, \theta_E^{(n,m)}, t) \rightarrow_d \chi_1^2, \quad (5.29)$$

as $n \rightarrow \infty$, for each fixed $t \in \mathcal{T}$ and \rightarrow_d means the convergence in distribution.

6 Plug-in empirical likelihood in two sample case

Here we establish a generalization of the basic theorem of EL to allow for plug-in estimates of nuisance parameters in the estimating equations in the two sample case. In this context we will closely follow the setup of the paper of Hjort *et al.* (2004), where plug-in empirical likelihood has been derived in the one sample case. The method of profile EL for dealing with nuisance parameters in estimating equations is often not applicable for infinite dimensional nuisance parameters. The necessary smoothness conditions (cf. Owen, 2001, p.54) of estimating functions are not fulfilled for example if we are interested in inference of sample quantiles. And even when it is applicable, implementation can be computationally difficult.

The use of plug-in for nuisance parameters in EL confidence regions is not new. It has been applied recently in various survival analysis contexts by Qin and Jing (2001a, 2001b), Wang and Jing (2001), Li and Wang (2003) and Qin and Tsao (2003). Hjort *et al.* (2004) have provided a wide applicable version of this approach, covering different examples for i.i.d and censored random variables. They consider also slower than \sqrt{n} -rates of convergence and allow the dimension of the parameter (or the number of estimating equations) to increase with the sample size. Calibrating EL confidence regions with plug-in is sometimes intractable due to the complexity of the asymptotics. For such situations Bootstrap approximation has been established by Hjort *et al.* (2004).

Now we are going to establish a similar version of plug-in empirical likelihood for the two sample case. We will not go deep in possible extensions and many examples, instead of keeping in mind our goal - constructing confidence intervals for structural relationship models.

Assume that we have the same setup as in Section 5.3. Here inference about the function Δ is carried out using the estimating functions $w_1(X, \theta_0, \Delta, t, h)$ and $w_2(Y, \theta_0, \Delta, t, h)$ in (5.14) and (5.15), where h is a 'nuisance' parameter (may be infinite) with unknown true value $h_0 = h_0(F_1, F_2) \in \mathcal{H}$. When h_0 is known, it can replace h in the EL ratio function, for each fixed $t \in \mathcal{T}$ leading to confidence interval $\{\Delta : R(\Delta, \theta_E^{(n,m)}, t, h_0) > c\}$ for true Δ_0 . Again the constant c can be calibrated from the fact that

$$-2 \log R_E(\Delta, \theta_E^{(n,m)}, t, h_0) \rightarrow_d \chi_1^2 \quad (6.1)$$

as $n \rightarrow \infty$, for each fixed $t \in \mathcal{T}$ (cf. Theorem 18). Now we establish a plug-in version of EL in which the unknown h_0 is replaced by an estimator \hat{h} , leading to a calibration for $\{\Delta : R(\Delta, \theta_E^{(n,m)}, t, \hat{h}) > c\}$ as a confidence region for Δ_0 .

6.1 Assumptions and Notation

Now we will establish notation and state our assumptions for plug-in empirical likelihood in the two sample case. Assume that $h_0 \in \mathcal{H}$, $\hat{h} \in \bar{H} \subset \mathcal{H}$ and $t \in \mathcal{T}$, where \mathcal{T} is some interval. Remember that for structural relationships defined in (2.6) we have $\mathcal{H} \subseteq \mathbb{R}^l, l > 0$, compact.

For a sequence of random vectors $X_n \in \mathbb{R}^l$ and a sequence a_m we use the stochastic order notation and write $X_n = O_p(a_m)$

$$\lim_{C \rightarrow \infty} \limsup_{m \rightarrow \infty} P(|X_n| > Ca_m) = 0,$$

where \rightarrow_p denotes the convergence in probability. Similarly write $X_n = o_p(a_m)$ if

$$a_m^{-1}|X_n| \rightarrow_p 0.$$

Further on we will use the following notation: $w_1(X, \theta_0, h) := w_1(X, \theta_0, \Delta, t, h)$, $w_2(Y, \theta_0, h) := w_2(Y, \theta_0, \Delta, t, h)$ and $\alpha_1(X, \theta, h) := \alpha_1(X, \theta, \Delta, t, h)$, $\alpha_2(Y, \theta, h) := \alpha_2(Y, \theta, \Delta, t, h)$. Denote

$$M_{1n}(\theta, t, h) := \frac{1}{n} \sum_{i=1}^n w_1(X_i, \theta, h), \quad S_{1n}(\theta, t, h) := \frac{1}{n} \sum_{i=1}^n w_1^2(X_i, \theta, h), \quad (6.2)$$

$$M_{2m}(\theta, t, h) := \frac{1}{m} \sum_{j=1}^m w_2(Y_j, \theta, h), \quad S_{2m}(\theta, t, h) := \frac{1}{m} \sum_{j=1}^m w_2^2(Y_j, \theta, h). \quad (6.3)$$

Assumption 3. (a) $\theta_0 \in \Omega$ and Ω is an open interval; (b) $\frac{m}{n} \rightarrow k$ (as $n, m \rightarrow \infty$) and $0 < k < \infty$.

Assumption 4. (a) For the first sample X_1, \dots, X_n we assume the following:

(A1) $E_{F_1} w_1^2(X, \theta, \hat{h}) > 0$, $\alpha_1(X, \theta, \hat{h})$ is continuous in a neighborhood of θ_0 , $\alpha_1(X, \theta, \hat{h})$ and $w_1^3(X, \theta, \hat{h})$ are bounded by some integrable function $G_1(X)$ in this neighborhood, $E_{F_1} \alpha_1(X, \theta, \hat{h})$ is nonzero.

(A2) For some subset $\bar{\mathcal{H}}$ of \mathcal{H} such that $P\{\hat{h} \in \bar{\mathcal{H}}\} \rightarrow 1$, and for some $\eta \in (1/3; 1/2)$, the class of functions $\mathcal{F} = \{w_1(\cdot, \theta, h) : |\theta - \theta_0| \leq cn^{-\eta}, h \in \bar{\mathcal{H}}\}$ with a positive constant $c < \infty$ has the strong Glivenko-Cantelli property with the almost sure convergence rate

$$\sup_{|\theta - \theta_0| \leq cn^{-\eta}, h \in \bar{\mathcal{H}}} \left| \frac{1}{n} \sum_{i=1}^n \{w_1(X_i, \theta, h) - \mathbb{E}w_1(X, \theta, h)\} \right| = O(\beta_1) \quad a.s., \quad (6.4)$$

where $\beta_1 = o(n^{-\eta})$.

(A3) For the functions $w_1^2(X, \theta, h)$, $w_1^3(X, \theta, h)$, $\alpha_1(X, \theta, h)$ assume that also the strong Glivenko-Cantelli property holds, i.e. (6.4) is true.

(A4) Assume that

$$\mathbb{E}w_1(X, \theta_0, \hat{h}) = O(\beta_2) \quad a.s., \quad (6.5)$$

where $\beta_2 = o(n^{-\eta})$.

(b) Assume for the second sample Y_1, \dots, Y_m that also (A1) - (A4) hold for the functions $w_2(Y, \theta, h)$, $w_2^2(Y, \theta, h)$, $w_2^3(Y, \theta, h)$, $\alpha_2(Y, \theta, h)$.

Assumption 5. (a) For the first sample X_1, \dots, X_n we assume the following:

(B1) $n^{1/2}M_{1n}(\theta_0, t, \hat{h}) \rightarrow_d U_1(t)$, where $U_1(t) \sim N(0, V_1(t))$.

(B2) $\sup_{t \in \mathcal{T}} |S_{1n}(\theta_0, t, \hat{h}) - V_2(t)| \rightarrow_p 0$.

(B3) $\sup_{t \in \mathcal{T}} |n^{-1} \sum_{i=1}^n \alpha_1(X_i, \theta_0, \hat{h}) - V_3(t)| \rightarrow_p 0$.

(B4) $w_1^3(X, \theta_0, \hat{h})$ is bounded by some integrable function $G_{11}(X)$.

(B5) Assume for $\theta \in \{\theta : |\theta - \theta_0| \leq cn^{-\eta}\}$, where c is some positive constant, that

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial S_{1n}(\theta, t, \hat{h})}{\partial \theta} = O_p(1); \quad \frac{1}{n} \sum_{i=1}^n \alpha_1(X_i, \theta, \hat{h}) = O_p(1); \quad \frac{1}{n} \sum_{i=1}^n w_1^3(X_i, \theta, \hat{h}) = O_p(1).$$

(b) Assume for the second sample Y_1, \dots, Y_m that also (B1) - (B5) also hold for the functions $w_2(Y, \theta, h)$, $w_2^2(Y, \theta, h)$, $w_2^3(Y, \theta, h)$, $\alpha_2(Y, \theta, h)$ and $S_{2m}(\theta, t, h)$ with functions $M_1(t)$, $M_2(t)$ and $M_3(t)$ instead of $V_1(t)$, $V_2(t)$ and $V_3(t)$.

Assumption 6 (Technical assumption). For both samples X_1, \dots, X_n and Y_1, \dots, Y_m for $\lambda'_1 \in [0, \lambda_1]$, $\lambda'_2 \in [0, \lambda_2]$ and $\theta \in \{\theta : |\theta - \theta_0| \leq cn^{-\eta}\}$, where c is some positive constant, we assume that

$$\frac{\partial^2 Q_{1n}(\theta, 0, 0)}{\partial \theta^2} = O_p(1); \quad \frac{\partial^2 Q_{2n}(\theta, 0, 0)}{\partial \theta^2} = O_p(1); \quad \frac{\partial^2 Q_{1n}(\theta_0, \lambda'_1, 0)}{\partial \lambda_1^2} = O_p(1)$$

and

$$\frac{\partial^2 Q_{2n}(\theta_0, 0, \lambda'_2)}{\partial \lambda_2^2} = O_p(1); \quad \frac{\partial^2 Q_{3n}(\theta_0, \lambda'_1, 0)}{\partial \lambda_1^2} = O_p(1); \quad \frac{\partial^2 Q_{3n}(\theta_0, 0, \lambda'_2)}{\partial \lambda_2^2} = O_p(1),$$

where the functions $Q_{in}(\theta, \lambda_1, \lambda_2)$ for $i = 1, 2, 3$ are defined later in (6.34), (6.35) and (6.36), respectively.

Let us comment on our set of Assumptions. We analyze only the first sample X_1, \dots, X_n case, because for the second sample Y_1, \dots, Y_m the same considerations hold.

Comment on Assumption 3

Part (a) from the Assumption 3 states that the true parameter θ_0 should be in an open interval. In (b) it has been required that the sample sizes are asymptotically comparable in the sense that $m/n \rightarrow k$, where $0 < k < \infty$ and of course $n, m \rightarrow \infty$. Both assumptions are very natural (see Qin and Lawless, 1994; Qin and Zhao, 2000).

Comment on Assumption 4

The reason we need the Assumption 4, which contains a subset of the assumptions (A1)-(A4), is to show that there exists a solution to the maximization problem, which has been described in Section 5.3. For that reason we need that $\hat{h} \rightarrow h_0$ almost surely with some almost sure rate.

The Assumption (A1) has also been assumed by Qin and Lawless (1994) and Qin and Zhao (2000). However, the assumptions (A2)-(A4) are quite strong and new in case of plug-in empirical likelihood. Conceivably that might be the reason why Hjort *et al.* (2004) assumed the existence of the solution of maximization problem implicitly.

Assumptions (A2) and (A3) mean that the strong law of large numbers holds with the almost sure convergence rate $O(\beta_1)$ for a class of empirical processes which are determined by the parameter $h \in \mathcal{H}$. For a known h_0 the rate is well known and equal to $O(n^{-1/2}(\log n)^{1/2})$ (see, e.g. Serfling, 1980, p.95). It appears that also for quite a general class of empirical processes the almost sure rate for the Glivenko-Cantelli property is the same, $O(n^{-1/2}(\log n)^{1/2})$ (see Alexander (1984) for Vapnik-Červonenkis (VC) classes of sets). Clearly this rate is of order $o(n^{-\eta})$ as required in Assumptions (A2), (A3).

Comment on Assumptions 5 and 6

Assuming an implicit existence of the solution to the maximization problem, we need much weaker assumptions (Assumptions 5 and 6) to derive the limiting distribution of the test statistic. In this case only a consistent estimator \hat{h} of h is required, i.e. we need $\hat{h} \rightarrow_p h_0$.

Assumptions (B1)-(B3) are similar to ones in Hjort *et al.* (2004). They are not too restrictive, but should be checked in each application of interest, because the limiting distribution of $R_E(\Delta, \theta_E^{(n,m)}, t, \hat{h})$ will depend on functions $V_l(t)$ and $M_l(t)$ for $l = 1, 2, 3$ (cf. Theorem 21).

Assumption (B5) is technical, we need it for proving Lemma 26. Actually we could have assumed that the functions $\partial S_{1n}(\theta', t, \hat{h})/\partial \theta$, $\alpha_1(X, \theta, \hat{h})$ and $w_1^3(X, \theta, \hat{h})$ have the weak Glivenko-Cantelli property and are dominated by some integrable functions. Then the Assumption (B5)

would consequently follow. The same concerns the technical Assumption 6.

Let us comment on the Assumption (B4). As already noted by Owen (1988) the existence of $\mathbb{E}|w_1(X, \theta, \hat{h})|^3$ is equivalent to $\sum P(|w_1(X_n, \theta, \hat{h})|^3 > n) < \infty$. It follows by the Borel-Cantelli lemma that $|w_1(X_n, \theta, \hat{h})| < n^{1/3}$ all but finitely often, with probability 1. This implies that

$$\max_{1 \leq i \leq n} |w_1(X_i, \theta, \hat{h})| \leq n^{1/3} \quad (6.6)$$

all but finitely often with probability 1, since $n^{1/3}$ eventually exceeds the largest of the finite collection of $|X_k|$'s that exceed $k^{1/3}$. Therefore we assume that (6.6) holds. We end our discussion for Assumption (B4) with the following lemma:

Lemma 19 (Owen (1990)). *Let $X_i \geq 0$ be i.i.d random variables and define $Z_n = \max_{i \leq n} X_i$. If $\mathbb{E}(X_1^2) < \infty$, then*

$$Z_n = o(n^{1/2}) \quad (6.7)$$

and

$$\frac{1}{n} \sum_{i=1}^n X_i^3 = o(n^{1/2}), \quad (6.8)$$

both with probability 1 as $n \rightarrow \infty$.

Proof. See Owen (1990), p.98. It follows straight forward from Borel-Cantelli lemma and law of large numbers applied to $\sum X_i^2$.

Remark 20. *Hjort et al. (2004) required a weaker assumption, i.e. $\max_{1 \leq i \leq n} |w_1(X_i, \theta_0, \hat{h})| = o_{pr}(n^{1/2})$ instead of (6.6), which naturally follows from Lemma 19. However, they assumed that there exists a solution to the maximization problem described in Section 5.3. Proving Lemma 24 and Lemma 26 we can see that (6.6) is crucial in equations (6.22) and (6.30) for deriving rates of λ_1 and λ_2 .*

Our discussion on assumptions we would like to complete with a note concerning consistent and almost sure estimators \hat{h} of unknown h_0 .

Consistent estimator

Having "only" a consistent estimator \hat{h} we can derive the asymptotic limit distribution of the test statistic, but we can not show that the maximization problem has a solution. Following the main technique of Qin and Lawless (1994), in order to show the maximization problem we need that $\hat{h} \rightarrow h_0$ almost surely with some rate. However, for consistent estimators under the Assumptions 3, 5 and 6 the limiting distribution follows actually from Lemma 27 (see also the proof of Theorem 21 on page 44).

Almost sure estimator

Assume that we have

$$\hat{h} = h_0 + O(\beta) \quad a.s., \quad (6.9)$$

where $\beta = o(n^{-\eta})$ is an almost sure convergence rate. As $\mathbb{E}w_1(X, \theta_0, h_0) = 0$ from the definition of the estimating equations (cf. (5.14) and (5.15)) it follows that $\mathbb{E}w_1(X, \theta_0, \hat{h}) = O(\beta)$ almost surely. Thus, the condition (A4) is verified. Therefore having the almost sure convergence of

an estimator \hat{h} with some rate as in (6.9) under the Assumptions 3 and 4 we can show that the maximization problem has a solution (cf. Lemma 25).

6.2 Main results

Here we state the main results of the work. We divide the results into two parts. In the first part we analyze the maximization problem and show under which conditions it has a solution. The second part is devoted for the limiting distribution of the test statistic. At the beginning we state the main theorem.

Theorem 21. *Assume that Assumptions 3-6 are satisfied, then there exists a root $\theta_E^{(n,m)}$ of (5.28) such that $\theta_E^{(n,m)}$ is a consistent estimate for θ_0 , the function $R_E(\Delta, \theta, t, \hat{h})$ attains its local maximum value at $\theta_E^{(n,m)}$ and*

$$\sqrt{n}(\theta_E^{(n,m)} - \theta_0) \rightarrow_d N\left(0, \frac{M_1 V_1}{V_3^2(t)M_1(t) + kM_3^2(t)V_1(t)}\right), \quad (6.10)$$

$$-2 \log R(\Delta, \theta_E^{(n,m)}, t, \hat{h}) \rightarrow_d \frac{V_3^2(t)M_2(t) + kM_3^2(t)V_2(t)}{M_1(t)V_3^2(t) + kV_1(t)M_3^2(t)} \chi_1^2 \quad (6.11)$$

for $t \in \mathcal{T}$ as $n \rightarrow \infty$.

Proof. For the proof see page 44.

Remark 22. *When $V_1(t) = V_2(t)$ and $M_1(t) = M_2(t)$ we have the standard χ_1^2 limit distribution and there is no perturbation due to plug-in. When it is not the case, the functions can be estimated via consistent estimators. As already noted by Hjort et al. (2004) it is not possible to say anything in general about the estimation of $V_1(t)$, which will depend on the structure of the specific application. They examine bootstrap approach which can be applied when $V_1(t)$ is difficult to estimate by other means and show that $\hat{V}_2(t) = S_{1n}(\hat{\theta}, t, \hat{h})$ consistently estimates $V_2(t)$ under some additional assumptions. We have even more difficult situation, because $V_3(t)$ and $M_3(t)$ should be estimated in general as well.*

6.2.1 Maximization problem

Corollary 23. *From Assumptions (A2)-(A4) it follows that for $\theta \in \{\theta : |\theta - \theta_0| \leq cn^{-\eta}\}$, where $c > 0$ is some constant*

$$S_{1n}(\theta, t, \hat{h}) = O(1); \quad \frac{1}{n} \sum_{i=1}^n w_1^3(X_i, \theta, \hat{h}) = O(1); \quad \frac{1}{n} \sum_{i=1}^n \alpha_1(X_i, \theta, \hat{h}) = O(1) \quad a.s. \quad (6.12)$$

$$M_{1n}(\theta, t, \hat{h}) = O(n^{-\eta}) \quad a.s. \quad (6.13)$$

Proof. The statement (6.12) follow directly from Assumptions (A2) and (A3). From Taylor expansion with $\theta' \in [\theta_0, \theta - \theta_0]$ we have

$$M_{1n}(\theta, t, \hat{h}) = M_{1n}(\theta_0, t, \hat{h}) + \frac{1}{n} \sum_{i=1}^n \alpha_1(X_i, \theta', \hat{h})(\theta - \theta_0), \quad (6.14)$$

On the other hand due to Assumption (A4)

$$M_{1n}(\theta_0, t, \hat{h}) = \mathbb{E}w_1(X, \theta_0, \hat{h}) + O(\beta_1) = o(n^{-\eta}) \quad a.s. \quad (6.15)$$

Thus,

$$M_{1n}(\theta, t, \hat{h}) = o(n^{-\eta}) + O(1)O(n^{-\eta}) = O(n^{-\eta}) \quad a.s. \quad (6.16)$$

Lemma 24. *Suppose that $1/3 < \eta < 1/2$ and the Assumption 4 is satisfied. Then*

$$\lambda_1(\theta) = O(n^{-\eta}), \quad \lambda_2(\theta) = O(m^{-\eta}) \quad a.s. \quad (6.17)$$

uniformly about $\theta \in \{\theta : |\theta - \theta_0| \leq cn^{-\eta}\}$, where c is some positive constant. Consequently, we also have

$$\lambda_1(\theta_0) = O(\beta_1), \quad \lambda_2(\theta_0) = O(\beta_1) \quad a.s. \quad (6.18)$$

Proof. It is similar to the proof of Owen (1990), when $\mathbb{E}|w_1(X, \theta, \hat{h})|^3 < \infty$ and $|\theta - \theta_0| \leq cn^{-\eta}$. The probabilities p_i appearing in the definition of $R_E(\Delta, \theta, t)$ in (5.17) are given by $p_i = n^{-1}(1 + \lambda_1 w_1(X_i, \theta, \hat{h}))^{-1}$. Therefore, $1 + \lambda_1 w_1(X_i, \theta, \hat{h}) \geq 0$, and so

$$|1 + \lambda_1 w_1(X_i, \theta, \hat{h})|^{-1} \geq (1 + |\lambda_1| \max_{1 \leq i \leq n} |w_1(X_i, \theta, \hat{h})|)^{-1}. \quad (6.19)$$

By definition of λ_1 and using (6.19) we have,

$$\begin{aligned} 0 &= n^{-1} \left| \sum_{i=1}^n \{ \lambda_1(\theta) w_1^2(X_i, \theta, \hat{h}) (1 + \lambda_1(\theta) w_1(X_i, \theta, \hat{h}))^{-1} - w_1(X_i, \theta, \hat{h}) \} \right| \\ &\geq |\lambda_1(\theta)| (1 + |\lambda_1(\theta)| \max_{1 \leq i \leq n} |w_1(X_i, \theta, \hat{h})|)^{-1} S_{1n}(\theta, t, \hat{h}) - M_{1n}(\theta, t, \hat{h}). \end{aligned} \quad (6.20)$$

Therefore, $|\lambda_1(\theta)| S_{1n}(\theta, t, \hat{h}) \leq (1 + |\lambda_1(\theta)| \max_{1 \leq i \leq n} |w_1(X_i, \theta, \hat{h})|) M_{1n}(\theta, t, \hat{h})$, or, equivalently,

$$|\lambda_1(\theta)| (S_{1n}(\theta, t, \hat{h}) - \max_{1 \leq i \leq n} |w_1(X_i, \theta, \hat{h})| M_{1n}(\theta, t, \hat{h})) \leq M_{1n}(\theta, t, \hat{h}). \quad (6.21)$$

This is an important equation, from which we can derive different rates of $\lambda_1(\theta)$ and $\lambda_1(\theta_0)$. It follows directly from Corrolary 23 and (6.6) that

$$|\lambda_1|(O(1) - o(n^{1/3})O(n^{-\eta})) \leq O(n^{-\eta}) \quad a.s. \quad (6.22)$$

So, $\lambda_1 = O(n^{-\eta})$ almost surely.

Lemma 25. *Assume that the Assumption 4 is satisfied with η as in Lemma 24. Then with probability tending to 1, there exists a root $\theta_E^{(n,m)}$ of (5.28) such that*

$$|\theta_E^{(n,m)} - \theta_0| = O(n^{-\eta}), \quad (6.23)$$

and $-2 \log R_E(\Delta, \theta, t, \hat{h})$ attains its local maximum value at $\theta_E^{(n,m)}$.

Proof. The idea is similar to Qin and Lawless (1994). Put $\lambda_1 = \lambda_1(\theta)$. From (5.25) we have

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n w_1(X_i, \theta, \hat{h}) \left(1 - \lambda_1(\theta) w_1(X_i, \theta, \hat{h}) + \frac{\lambda_1^2(\theta) w_1^2(X_i, \theta, \hat{h})}{1 + \lambda_1(\theta) w_1(X_i, \theta, \hat{h})} \right) \\ &= M_{1n}(\theta, t, \hat{h}) - \lambda_1(\theta) S_{1n}(\theta, t, \hat{h}) + \frac{1}{n} \sum_{i=1}^n \frac{\lambda_1^2(\theta) w_1^3(X_i, \theta, \hat{h})}{1 + \lambda_1(\theta) w_1(X_i, \theta, \hat{h})}. \end{aligned} \quad (6.24)$$

The modulo from the last term is bounded by

$$\frac{1}{n} \sum_{i=1}^n |w_1^3(X_i, \theta, \hat{h})| |\lambda_1(\theta)|^2 |1 + \lambda_1(\theta)w_1(X_i, \theta, \hat{h})|^{-1} = O(1)O(n^{-2\eta})O(1) = O(n^{-2\eta}) \quad a.s.$$

which follows from Assumption 4 and the fact that $\max_{1 \leq i \leq n} |\lambda_1(\theta)w_1(X_i, \theta, \hat{h})| = O(n^{-\eta})o(n^{1/3}) = o(1)$. So,

$$\lambda_1(\theta) = \frac{M_{1n}(\theta, t, \hat{h})}{S_{1n}(\theta, t, \hat{h})} + O(n^{-2\eta}) = O(n^{-\eta}). \quad (6.25)$$

uniformly about $\theta \in \{\theta : |\theta - \theta_0| \leq cn^{-\eta}\}$. Put

$$H_{1E}(\Delta, \theta, t, \hat{h}) = \sum_{i=1}^n \log(1 + \lambda_1(\theta)w_1(X_i, \theta, \hat{h})), \quad (6.26)$$

$$H_{2E}(\Delta, \theta, t, \hat{h}) = \sum_{j=1}^m \log(1 + \lambda_2(\theta)w_2(Y_j, \theta, \hat{h})). \quad (6.27)$$

Hence, since we can write

$$\log(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3} \frac{x^3}{(1 + \xi)^3},$$

where ξ lies between 0 and x , by (6.25) and Taylor expansion for $\xi_{n,i} \in [0, \lambda_1(\theta)w_1(X_i, \theta, \hat{h})]$, where $\max_{1 \leq i \leq n} |\lambda_1(\theta)w_1(X_i, \theta, \hat{h})| = O(n^{-\eta})o(n^{1/3}) = o(1)$ we have

$$\begin{aligned} & H_{1E}(\Delta, \theta, t, \hat{h}) \\ &= \sum_{i=1}^n \lambda_1(\theta)w_1(X_i, \theta, \hat{h}) - \frac{1}{2} \sum_{i=1}^n \lambda_1^2(\theta)w_1^2(X_i, \theta, \hat{h}) + n\lambda_1^3(\theta) \frac{1}{n} \sum_{i=1}^n \frac{w_1^3(X_i, \theta, \hat{h})}{(1 + \xi_{n,i})^3} \quad a.s. \\ &= \frac{n}{2} \left(\frac{1}{n} \sum_{i=1}^n w_1(X_i, \theta, \hat{h}) \right)^2 \left(\frac{1}{n} \sum_{i=1}^n w_1^2(X_i, \theta, \hat{h}) \right)^{-1} + O(n)O(n^{-3\eta})O(1) \quad a.s. \\ &= \frac{n}{2} \left(\frac{1}{n} \sum_{i=1}^n w_1(X_i, \theta_0, \hat{h}) + \frac{1}{n} \sum_{i=1}^n \alpha_1(X_i, \theta', \hat{h})(\theta - \theta_0)^2 \right)^2 O(1) + o(n^{1-2\eta}) \quad a.s. \\ &= \frac{n}{2} (O(\beta_2) + O(\beta_1) + O(1)O(n^{-\eta}))^2 + o(n^{1-2\eta}) \quad a.s. \\ &= O(n^{1-2\eta}) \quad a.s.. \end{aligned}$$

Similarly,

$$\begin{aligned} H_{1E}(\Delta, \theta_0, t, \hat{h}) &= \frac{n}{2} \left(\frac{1}{n} \sum_{i=1}^n w_1(X_i, \theta_0, \hat{h}) \right)^2 \left(\frac{1}{n} \sum_{i=1}^n w_1^2(X_i, \theta_0, \hat{h}) \right)^{-1} + o(1) \\ &= O(n)(O(\beta_2) + O(\beta_1))^2 \quad a.s. \end{aligned}$$

From Assumptions (A2)-(A4) we have $O(n)(O(\beta_2) + O(\beta_1))^2 = o(n^{1-2\eta})$. Since $H_{1E}(\Delta, \theta, t, \hat{h})$ is a continuous function about θ as θ belongs to the ball $|\theta - \theta_0| \leq cn^{-\eta}$, $H_{1E}(\Delta, \theta, t, \hat{h})$ has minimum value in the interior of this ball. The same holds for $H_{2E}(\Delta, \theta, t, \hat{h})$. Therefore, when $|\theta - \theta_0| = n^{-\eta}$, with probability tending to one

$$H_{1E}(\Delta, \theta, t, \hat{h}) > H_{1E}(\Delta, \theta_0, t, \hat{h}),$$

$$H_{2E}(\Delta, \theta, t, \hat{h}) > H_{2E}(\Delta, \theta_0, t, \hat{h}).$$

Hence when $|\theta - \theta_0| = n^{-\eta}$, with probability tending to one

$$-2 \log R_E(\Delta, \theta, t, \hat{h}) > -2 \log R_E(\Delta, \theta_0, t, \hat{h}).$$

So, there exists a root $\theta_E^{(n,m)}$ of (5.28) such that $|\theta_E^{(n,m)} - \theta_0| = O(n^{-\eta})$ and $-2 \log R_E(\Delta, \theta, t, \hat{h})$ attains its local minimum value at $\theta_E^{(n,m)}$.

6.2.2 Limiting distribution

Lemma 26. *Suppose that $1/3 < \eta < 1/2$ and Assumption 5 holds. Then*

$$\lambda_1(\theta) = O_p(n^{-\eta}), \quad \lambda_2(\theta) = O_p(m^{-\eta}) \quad (6.28)$$

uniformly about $\theta \in \{\theta : |\theta - \theta_0| \leq cn^{-\eta}\}$, where c is some positive constant.

Proof. Similarly as in proof of Lemma 24 (cf. inequality (6.21)) we get

$$|\lambda_1(\theta)|(S_{1n}(\theta, t, \hat{h}) - \max_{1 \leq i \leq n} |w_1(X_i, \theta, \hat{h})| M_{1n}(\theta, t, \hat{h})) \leq M_{1n}(\theta, t, \hat{h}). \quad (6.29)$$

From Taylor expansion we get for $\theta' \in [\theta, \theta - \theta_0]$

$$S_{1n}(\theta, t, \hat{h}) = S_{1n}(\theta_0, t, \hat{h}) + \frac{1}{n} \sum_{i=1}^n \frac{\partial S_{1n}(\theta', t, \hat{h})}{\partial \theta} (\theta - \theta_0)^2 = O_p(1) + O_p(1)O_p(n^{-\eta}) = O_p(1)$$

and

$$\begin{aligned} M_{1n}(\theta, t, \hat{h}) &= M_{1n}(\theta_0, t, \hat{h}) + \frac{1}{n} \sum_{i=1}^n \alpha_1(X_i, \theta', \hat{h}) (\theta - \theta_0)^2 \\ &= O_p(n^{-1/2}) + O_p(1)O_p(n^{-\eta}) = O_p(n^{-\eta}). \end{aligned}$$

Now from Assumptions 5 follows

$$|\lambda_1|(O_p(1) - o_p(n^{1/3})O_p(n^{-\eta})) \leq O_p(n^{-\eta}). \quad (6.30)$$

So, $\lambda_1 = O_p(n^{-\eta})$.

Lemma 27. *Assume that Assumptions 3, 5 and 6 hold and assume that the maximization problem is solved implicitly, i.e. Lemma 25 is true. Then with $\theta_E^{(n,m)}$ as in Lemma 25 we have*

$$\sqrt{n}(\theta_E^{(n,m)} - \theta_0) \rightarrow_d N \left(0, \frac{M_1 V_1}{M_1(t) V_3^2(t) + k V_1(t) M_3^2(t)} \right), \quad (6.31)$$

$$\lambda_1(\theta_E^{(n,m)}) = -k \frac{M_3(t)}{V_3(t)} \lambda_2(\theta_E^{(n,m)}) + o_p(n^{-1/2}), \quad (6.32)$$

$$\sqrt{n} \lambda_2(\theta_E^{(n,m)}) \rightarrow_d N \left(0, \frac{V_3^2(t)}{k(M_1(t) V_3^2(t) + k V_1(t) M_3^2(t))} \right). \quad (6.33)$$

Proof. Put $\lambda_i = \lambda_i(\theta)$, $\lambda_{iE} = \lambda_i(\theta_E^{(n,m)})$ for $i = 1, 2$, and

$$Q_{1n}(\theta, \lambda_1, \lambda_2) = \frac{1}{n} \sum_{i=1}^n \frac{w_1(X_i, \theta, h)}{1 + \lambda_1 w_1(X_i, \theta, h)}, \quad (6.34)$$

$$Q_{2n}(\theta, \lambda_1, \lambda_2) = \frac{1}{m} \sum_{j=1}^m \frac{w_2(Y_j, \theta, h)}{1 + \lambda_2 w_2(Y_j, \theta, h)}, \quad (6.35)$$

$$Q_{3n}(\theta, \lambda_1, \lambda_2) = \frac{1}{n} \lambda_1 \sum_{i=1}^n \frac{\alpha_1(X_i, \theta, h)}{1 + \lambda_1 w_1(X_i, \theta, h)} + \frac{1}{m} \lambda_2 \sum_{j=1}^m \frac{\alpha_2(Y_j, \theta, h)}{1 + \lambda_2 w_2(Y_j, \theta, h)}. \quad (6.36)$$

From Lemma 25, we have

$$Q_{in}(\theta_E^{(n,m)}, \lambda_{1E}, \lambda_{2E}) = 0 \text{ for } i = 1, 2, 3. \quad (6.37)$$

First note that for $\lambda'_1 \in [0, \lambda_1]$, $\lambda'_2 \in [0, \lambda_2]$ and $\theta' \in [\theta, \theta - \theta_0]$

$$\frac{\partial^2 Q_{3n}(\theta', 0, 0)}{\partial \theta^2} = 0; \quad \frac{\partial^2 Q_{1n}(\theta_0, 0, \lambda'_2)}{\partial \lambda_2^2} = 0; \quad \frac{\partial^2 Q_{2n}(\theta_0, \lambda'_1, 0)}{\partial \lambda_1^2} = 0.$$

From Assumption 6 it follows

$$\frac{\partial^2 Q_{1n}(\theta', 0, 0)}{\partial \theta^2} = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 w_1(X_i, \theta', \hat{h})}{\partial \theta^2} = O_p(1); \quad \frac{\partial^2 Q_{2n}(\theta', 0, 0)}{\partial \theta^2} = \frac{1}{m} \sum_{j=1}^m \frac{\partial^2 w_2(Y_j, \theta', \hat{h})}{\partial \theta^2} = O_p(1);$$

$$\frac{\partial^2 Q_{1n}(\theta_0, \lambda'_1, 0)}{\partial \lambda_1^2} = \frac{2}{n} \sum_{i=1}^n \frac{w_1^3(X_i, \theta_0, \hat{h})}{(1 + \lambda'_1 w_1(X_i, \theta_0, \hat{h}))^3} = O_p(1);$$

$$\frac{\partial^2 Q_{2n}(\theta_0, 0, \lambda'_2)}{\partial \lambda_2^2} = \frac{2}{m} \sum_{j=1}^m \frac{w_2^3(Y_j, \theta_0, \hat{h})}{(1 + \lambda'_2 w_2(Y_j, \theta_0, \hat{h}))^3} = O_p(1);$$

$$\frac{\partial^2 Q_{3n}(\theta_0, \lambda'_1, 0)}{\partial \lambda_1^2} = \frac{2}{n} \lambda'_1 \sum_{i=1}^n \frac{\alpha_1(X_i, \theta_0, \hat{h}) w_1^2(X_i, \theta_0, \hat{h})}{1 + \lambda'_1 w_1(X_i, \theta_0, \hat{h})} - \frac{2}{n} \sum_{i=1}^n \frac{\alpha_1(X_i, \theta_0, \hat{h}) w_1(X_i, \theta_0, \hat{h})}{(1 + \lambda'_1 w_1(X_i, \theta_0, \hat{h}))^2} = O_p(1);$$

$$\frac{\partial^2 Q_{3n}(\theta_0, 0, \lambda'_2)}{\partial \lambda_2^2} = \frac{2}{m} \lambda'_2 \sum_{j=1}^m \frac{\alpha_2(Y_j, \theta_0, \hat{h}) w_2^2(Y_j, \theta_0, \hat{h})}{1 + \lambda'_2 w_2(Y_j, \theta_0, \hat{h})} - \frac{2}{m} \sum_{j=1}^m \frac{\alpha_2(Y_j, \theta_0, \hat{h}) w_2(Y_j, \theta_0, \hat{h})}{(1 + \lambda'_2 w_2(Y_j, \theta_0, \hat{h}))^2} = O_p(1).$$

Therefore by Taylor expansion, Lemma 25 and Lemma 26, we have

$$\begin{aligned} 0 = Q_{in}(\theta_E^{(n,m)}, \lambda_{1E}, \lambda_{2E}) &= Q_{in}(\theta_0, 0, 0) + \frac{\partial Q_{in}(\theta_0, 0, 0)}{\partial \theta} (\theta_E^{(n,m)} - \theta_0) \\ &\quad + \frac{\partial Q_{in}(\theta_0, 0, 0)}{\partial \lambda_1} \lambda_{1E} + \frac{\partial Q_{in}(\theta_0, 0, 0)}{\partial \lambda_2} \lambda_{2E} + O_p(n^{-2\eta}), \quad i = 1, 2, 3. \end{aligned} \quad (6.38)$$

Hence

$$\begin{aligned} Q_{in}(\theta_0, 0, 0) + \frac{\partial Q_{in}(\theta_0, 0, 0)}{\partial \theta} (\theta_E^{(n,m)} - \theta_0) + \frac{\partial Q_{in}(\theta_0, 0, 0)}{\partial \lambda_1} \lambda_{1E} + \frac{\partial Q_{in}(\theta_0, 0, 0)}{\partial \lambda_2} \lambda_{2E} \\ = o_p(n^{-1/2}), \quad i = 1, 2, 3. \end{aligned} \quad (6.39)$$

We have from Assumptions (B1)-(B3)

$$\begin{aligned} \frac{\partial Q_{1n}(\theta_0, 0, 0)}{\partial \theta} &\rightarrow_p V_3(t), \quad \frac{\partial Q_{1n}(\theta_0, 0, 0)}{\partial \lambda_1} \rightarrow_p -V_1(t), \quad \frac{\partial Q_{1n}(\theta_0, 0, 0)}{\partial \lambda_2} = 0, \\ \frac{\partial Q_{2n}(\theta_0, 0, 0)}{\partial \theta} &\rightarrow_p M_3(t), \quad \frac{\partial Q_{2n}(\theta_0, 0, 0)}{\partial \lambda_1} = 0, \quad \frac{\partial Q_{2n}(\theta_0, 0, 0)}{\partial \lambda_2} \rightarrow_p -M_1(t), \\ \frac{\partial Q_{3n}(\theta_0, 0, 0)}{\partial \theta} &= 0, \quad \frac{\partial Q_{3n}(\theta_0, 0, 0)}{\partial \lambda_1} \rightarrow_p V_3(t), \quad \frac{\partial Q_{3n}(\theta_0, 0, 0)}{\partial \lambda_2} \rightarrow_p kM_3(t). \end{aligned}$$

So

$$\begin{pmatrix} \theta_E^{(n,m)} - \theta_0 \\ \lambda_{1E} \\ \lambda_{2E} \end{pmatrix} = -S^{-1} \begin{pmatrix} Q_{1n}(\theta_0, 0, 0) \\ Q_{2n}(\theta_0, 0, 0) \\ 0 \end{pmatrix} + o_p(n^{-1/2}),$$

where

$$S = \begin{pmatrix} V_3(t) & -V_1(t) & 0 \\ M_3(t) & 0 & -M_1(t) \\ 0 & V_3(t) & kM_3(t) \end{pmatrix}.$$

Because

$$S^{-1} = \frac{1}{c_1} \begin{pmatrix} M_1(t)V_3(t) & kV_1(t)M_3(t) & V_1(t)M_1(t) \\ -kM_3^2(t) & kV_3(t)M_3(t) & M_1(t)V_3(t) \\ V_3(t)M_3(t) & -V_3^2(t) & V_1(t)M_3(t) \end{pmatrix}$$

we have

$$\begin{aligned} \theta_E^{(n,m)} - \theta_0 &= -\frac{1}{c_1} (M_1(t)V_3(t)Q_{1n}(\theta_0, 0, 0) + kV_1(t)M_3(t)Q_{2n}(\theta_0, 0, 0)) + o_p(n^{-1/2}), \\ \lambda_1 &= \left(\frac{kM_3(t)}{c_1} \right) (M_3Q_{1n}(\theta_0, 0, 0) - V_3(t)Q_{2n}(\theta_0, 0, 0)) + o_p(n^{-1/2}), \\ \lambda_2 &= -\left(\frac{V_3(t)}{c_1} \right) (M_3Q_{1n}(\theta_0, 0, 0) - V_3(t)Q_{2n}(\theta_0, 0, 0)) + o_p(n^{-1/2}). \end{aligned}$$

The proof of the Lemma 27 follows from the fact that

$$\sqrt{n} \begin{pmatrix} Q_{1n}(\theta_0, 0, 0) \\ Q_{2n}(\theta_0, 0, 0) \end{pmatrix} \rightarrow_d N \begin{pmatrix} V_1(t) & 0 \\ 0 & k^{-1}M_1(t) \end{pmatrix}.$$

Proof of Theorem 21. From Assumption (B4) with $\xi_{n,i} \in [0, \lambda_1(\theta_E^{(n,m)})w_1(X_i, \theta_E^{(n,m)}, \hat{h})]$, where $\max_{1 \leq i \leq n} |\lambda_1(\theta_E^{(n,m)})w_1(X_i, \theta_E^{(n,m)}, \hat{h})| = O_p(n^{-\eta})o_p(n^{1/3}) = o_p(1)$ and $\eta \in (1/3; 1/2)$ we have

$$\frac{n}{3} \lambda_1^3(\theta_E^{(n,m)}) \frac{1}{n} \sum_{i=1}^n \frac{w_1^3(X_i, \theta_E^{(n,m)}, \hat{h})}{(1 + \xi_{n,i})^3} = O(n)O_p(n^{-3\eta})O_p(1) = O_p(n^{-3\eta+1}) = o_p(1). \quad (6.40)$$

From Taylor expansion,

$$\begin{aligned} \log R(\Delta, \theta_E^{(n,m)}, t, \hat{h}) &= -n\lambda_1(\theta_E^{(n,m)})M_{1n}(\theta_E^{(n,m)}, t, \hat{h}) + \frac{n}{2}\lambda_1^2(\theta_E^{(n,m)})S_{1n}(\theta_E^{(n,m)}, t, \hat{h}) \\ &\quad - m\lambda_2(\theta_E^{(n,m)})M_{2m}(\theta_E^{(n,m)}, t, \hat{h}) + \frac{m}{2}\lambda_2^2(\theta_E^{(n,m)})S_{2m}(\theta_E^{(n,m)}, t, \hat{h}) + o_p(1), \end{aligned} \quad (6.41)$$

Now from (5.25) and (5.26) similarly as in (6.24), we have

$$M_{1n}(\theta_E^{(n,m)}, t, \hat{h}) = \lambda_1(\theta_E^{(n,m)})S_{1n}(\theta_E^{(n,m)}, t, \hat{h}) + O_p(n^{-2\eta}).$$

Similarly

$$M_{2m}(\theta_E^{(n,m)}, t, \hat{h}) = \lambda_2(\theta_E^{(n,m)})S_{2m}(\theta_E^{(n,m)}, t, \hat{h}) + O_p(n^{-2\eta}).$$

From Assumption (B2) we have

$$S_{1n}(\theta_E^{(n,m)}, t, \hat{h}) = V_2(t) + o_p(1), \quad S_{2m}(\theta_E^{(n,m)}, t, \hat{h}) = M_2(t) + o_p(1)$$

and from Lemma 27 it follows

$$\begin{aligned} -2 \log R(\Delta, \theta_E^{(n,m)}, t, \hat{h}) &= n\lambda_1^2(\theta_E^{(n,m)})S_{1n}(\theta_E^{(n,m)}, t, \hat{h}) + m\lambda_2^2(\theta_E^{(n,m)})S_{2m}(\theta_E^{(n,m)}, t, \hat{h}) + o_p(1) \\ &= nk^2 \frac{M_3^2(t)}{V_3^2(t)} \lambda_2^2(\theta_E^{(n,m)})V_2(t) + m\lambda_2^2(\theta_E^{(n,m)})M_2(t) + o_p(1) \\ &= (\sqrt{n}\lambda_2(\theta_E^{(n,m)}))^2 \frac{k(V_3^2(t)M_2(t) + kM_3^2(t)V_2(t))}{V_3^2(t)} + o_p(1) \\ &\quad \rightarrow \frac{V_3^2(t)M_2(t) + kM_3^2(t)V_2(t)}{M_1(t)V_3^2(t) + kV_1(t)M_3^2(t)} \chi_1^2. \end{aligned}$$

Remark 28. Qin and Zhao (2000) analyzed only $O_p(\cdot)$ behavior of λ_1 and λ_2 for the maximization problem (instead of almost sure behavior). Proving the results they mostly refer to Qin and Lawless (1994) for the details. However, Qin and Lawless (1994) in their basic and marvelous work in the area of empirical likelihood proved the maximization problem using almost sure behavior, which is essential here (see also Lemma 24 and Lemma 25).

6.3 Confidence bands for structural relationship models

Here we give an important application for plug-in EL established in Section 6 in the two sample case for structural relationship models defined in Section 2, page 12. More specifically, we construct confidence intervals for the function

$$\Delta := \Delta(t) = F_1(\phi_1(F_2^{-1}(\phi_2(t, h)), h)), \quad (6.42)$$

where $t \in \mathcal{T}$, $F_1, F_2 \in \mathcal{F}$, $h \in \mathcal{H} \subseteq \mathbb{R}^l$ with $l > 0$ and functions ϕ_1, ϕ_2 such that in Assumption 1. Actually (6.42) can be seen as a generalization of a P-P plot introduced in Section 3 (see also Example 7, page 31). In this case the estimating functions have the following form:

$$w_1(X, \theta_0, \Delta, t, \hat{h}) = 1_{\{X \leq \theta_0\}} - \Delta, \quad (6.43)$$

$$w_2(Y, \theta_0, \Delta, t, \hat{h}) = 1_{\{Y \leq \phi_1^-(\theta_0, \hat{h})\}} - \phi_2(t, \hat{h}), \quad (6.44)$$

where $\theta_0 = \phi_1(F_2^{-1}(\phi_2(t, \hat{h})), \hat{h})$ and $1_{\{\cdot\}}$ is the indicator function. From the estimating equations (5.14) and (5.15) the system of equations

$$\begin{cases} F_1(\theta_0) &= \Delta, \\ F_2(\phi_1^-(\theta_0, \hat{h})) &= \phi_2(t, \hat{h}) \end{cases} \quad (6.45)$$

follows.

We have the following result.

Theorem 29. *For structural relationship models assume that $\hat{h} = h_0 + O(\beta)$ almost surely, where $\beta = o(n^{-\eta})$. Then Assumptions 4, 5 and 6 hold with the functions*

$$V_1(t) = V_2(t) = \Delta(1 - \Delta), \quad (6.46)$$

$$M_1(t) = M_2(t) = \phi_2(t, h_0)(1 - \phi_2(t, h_0)) \quad (6.47)$$

and there exists a root $\theta_E^{(n,m)}$ of (5.28) such that $\theta_E^{(n,m)}$ is a consistent estimate for θ_0 , the function $R(\Delta, \theta, t, \hat{h})$ attains its local maximum value at $\theta_E^{(n,m)}$ and we have no perturbation due to plug-in, i.e.,

$$-2 \log R(\Delta, \theta_E^{(n,m)}, t, \hat{h}) \rightarrow_d \chi_1^2 \quad (6.48)$$

for $t \in \mathcal{T}$ as $n \rightarrow \infty$.

Proof. First of all note that the class of functions $\{1_{X \leq t}, t \in \mathbb{R}\}$ has a strong Glivenko-Cantelli property (cf. Example 9, page 67 in the Appendix). For both samples X_1, \dots, X_n and Y_1, \dots, Y_m the Assumptions 4, 5 and 6 hold, which follows from almost sure convergence of \hat{h} and Lemma 42, page 68 in the Appendix.

To apply Theorem 21 we need to verify the functions $V_1(t), V_2(t), V_3(t)$ and $M_1(t), M_2(t), M_3(t)$. We start with the first sample X_1, \dots, X_n . It is much easier to check the Assumptions (B1) and (B2), because the function $w_1(X, \theta_0, \Delta, t, \hat{h})$ does not contain the estimator \hat{h} . We easily obtain that

$$V_1(t) = V_2(t) = \Delta(1 - \Delta).$$

For the second sample the situation is a bit more difficult. We have

$$\begin{aligned} & \frac{1}{m} \sum_{j=1}^m w_2^2(Y_j, \theta_0, \Delta, t, \hat{h}) \\ &= \frac{1}{m} \sum_{j=1}^m 1_{\{Y_j \leq \phi_1^-(\theta_0, \hat{h})\}}^2 - 2 \frac{1}{m} \sum_{j=1}^m \phi_2(t, \hat{h}) 1_{\{Y_j \leq \phi_1^-(\theta_0, \hat{h})\}} + \phi_2^2(t, \hat{h}) \\ &= F_{2m}(\phi_1^-(\theta_0, \hat{h})) - 2F_{2m}(\phi_1^-(\theta_0, \hat{h}))\phi_2(t, \hat{h}) + \phi_2^2(t, \hat{h}), \end{aligned}$$

where $F_{2m}(\cdot)$ is the empirical distribution function. Due to the equation system (6.45) it follows that $F_{2m}(\phi_1^-(\theta_0, \hat{h})) \rightarrow F_2(\phi_1^-(\theta_0, h_0)) = \phi_2(t, h_0)$ almost surely (cf. Lemma 42, page 68 in the Appendix). Thus, $M_2(t) = \phi_2(t, h_0)(1 - \phi_2(t, h_0))$.

For the matrix $M_1(t)$ note that

$$\sqrt{m}M_{2m}(\theta_0, t, \hat{h}) = \sqrt{m}M_{2m}(\theta_0, t, h_0) + \sqrt{m}(M_{2m}(\theta_0, t, \hat{h}) - M_{2m}(\theta_0, t, h_0)). \quad (6.49)$$

The first term $\sqrt{m}M_{2m}(\theta_0, t, h_0) \rightarrow_d N(0, F_2(\phi_1^-(\theta_0, h_0))(1 - F_2(\phi_1^-(\theta_0, h_0))))$. To show that the second term in (6.49) converges in probability to zero we can use Skorohod construction (cf. Li and Doss, 1993, p. 788).

Remark 30. *Other way to show that $\sqrt{m}M_{2m}(\theta_0, t, \hat{h}) \rightarrow_d N(0, F_2(\phi_1^-(\theta_0, h_0))(1 - F_2(\phi_1^-(\theta_0, h_0))))$ is to use Lemma 43, page 68 in the Appendix. However, the direct Skorohod construction is plausible here, when considering empirical processes involving the empirical distribution function itself.*

We obtain that

$$M_1(t) = M_2(t) = \phi_2(t, h_0)(1 - \phi_2(t, h_0)).$$

Now we apply Theorem 21. By Remark 22 we know that in the case when $V_1(t) = V_2(t)$ and $M_1(t) = M_3(t)$ we do not have to calculate the functions $V_3(t)$ and $M_3(t)$. It follows that we have no perturbation due to plug-in and

$$-2 \log R(\Delta, \theta_E^{(n,m)}, t, \hat{h}) \rightarrow_d \chi_1^2$$

for $t \in \mathcal{T}$ as $n, m \rightarrow \infty$.

7 Smoothed plug-in EL for Structural relationships

Now we are going to introduce the smoothed plug-in empirical likelihood method which we will use later for simulations in the context of structural relationship models. Amazingly it provides some additional advantages in case of construction of confidence intervals for sample quantiles. Owen (1988) remarked that, when applied to the problem of constructed confidence intervals for a population quantile, the empirical likelihood reproduces precisely the so-called sign-test or binomial-method interval. However, by appropriate smoothing of the empirical likelihood method Chen and Hall (1993) showed that the coverage accuracy may be improved from order $n^{-1/2}$ to n^{-1} . Moreover the smoothed empirical likelihood appears to be Bartlett-correctable. Thus, an empirical correction for scale can reduce the size of coverage error from order n^{-1} to n^{-2} .

Chen and Hall (1993) introduced smoothed empirical likelihood for construction of confidence intervals for quantiles in one population. For constructing confidence bands for P-P plots in the two sample case Claeskens *et al.* (2003) also have used the smoothed version of empirical likelihood.

Now let us introduce the smoothed empirical likelihood for structural relationship model. This can be viewed as an extension of the results of Claeskens *et al.* (2003).

For $j = 1, 2$ let H_j denote a smoothed version of the degenerate distribution function H_0 defined by $H_0(x) = 1$ for $x \geq 0, 0$ otherwise. Define $H_j(t) = \int_{u \leq t} K_j(u) du$, where K_j is a compactly supported r -th order kernel which is commonly used in nonparametric density estimation. That is, for some integer $r \geq 2$ and constant $\kappa \neq 0$, K_j is a function satisfying

$$\int u^k K_j(u) du = \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{if } 1 \leq k \leq r - 1, \\ \kappa, & \text{if } k = r. \end{cases} \quad (7.1)$$

For further use we also define $H_{b_j}(t) = H_j(t/b_j)$, where $b_1 = b_1(n)$ and $b_2 = b_2(m)$ are bandwidth sequences, converging to zero as n, m grows to infinity.

Consider the problem of constructing confidence bands for the function Δ defined in (6.42).

Let $p = (p_1, \dots, p_{n_1})$ and $q = (q_1, \dots, q_{n_2})$ be two vectors consisting of nonnegative numbers adding to one. Define further the estimators

$$\hat{F}_{b_1, p}(\theta_0) = \sum_{i=1}^n p_i H_{b_1}(\theta_0 - X_i) \quad \text{and} \quad \hat{F}_{b_2, q}(\phi_1^-(\theta_0, \hat{h})) = \sum_{j=1}^m q_j H_{b_2}(\phi_1^-(\theta_0, \hat{h}) - Y_j).$$

For this setting we define the profile smoothed empirical likelihood ratio function for Δ as

$$R^{(sm)}(\Delta, \theta_0, t, \hat{h}) = \sup_{p, q} \prod_{i=1}^n n p_i \prod_{j=1}^m m q_j, \quad (7.2)$$

where the latter supremum is subject to the following constraints:

$$\begin{cases} \hat{F}_{b_1, p}(\theta_0) & = \Delta, \\ \hat{F}_{b_2, q}(\phi_1^-(\theta_0, \hat{h})) & = \phi_2(t, \hat{h}). \end{cases} \quad (7.3)$$

The estimating equations have the following form:

$$w_1(X_i, \theta_0, \hat{h}) = H_{b_1}(\theta_0 - X_i) - \Delta, \quad (7.4)$$

$$w_2(Y_j, \theta_0, \hat{h}) = H_{b_2}(\phi_1^-(\theta_0, \hat{h}) - Y_j) - \phi_2(t, \hat{h}). \quad (7.5)$$

Although it is a special case of the main Theorem 21, still we need to review all Lemmas 24, 26, 25, 27 and also main Theorem 21 itself. The reason is that we have the additional smoothing parameters b_1 and b_2 tending to zero as $n, m \rightarrow \infty$ and we need conditions on the rates of the speed. The results of the following lemma will be useful to clarify more on this situation.

Let $f_j = F_j'$ denote the first derivative of F_j , where defined and $f^{(r)}$ be the r -th derivative of the function f .

Assumption 7. For $j = 1, 2$ assume that K_j satisfies (7.1), and is bounded and compactly supported; that f_j and f_j^{r-1} exist in a neighborhood of θ_0 and are continuous at θ_0 , f_j^{r-1} and f_j^r are bounded and $f_j(\theta_0) > 0$.

Assumption 8. For both samples X_1, \dots, X_n and Y_1, \dots, Y_m assume that $nb_1^{3r} \rightarrow 0$ and $mb_2^{3r} \rightarrow 0$ respectively.

Lemma 31. *If Assumption 7 is satisfied we have*

$$\begin{aligned} \mathbb{E}\{w_1(X, \theta, h_0)\} &= F_1(\theta) - F_1(\theta_0) + O(b_1^r), \\ \mathbb{E}\{w_2(Y, \theta, h_0)\} &= F_2(\phi_1^-(\theta, h_0)) - F_2(\phi_1^-(\theta_0, h_0)) + O(b_2^r), \\ \text{var}\{w_1(X, \theta, h_0)\} &= F_1(\theta)(1 - F_1(\theta)) + O(b_1), \\ \text{var}\{w_2(Y, \theta, h_0)\} &= F_2(\phi_1^-(\theta, h_0))(1 - F_2(\phi_1^-(\theta, h_0))) + O(b_2), \\ \mathbb{E}\{\alpha_1(X, \theta, h_0)\} &= f_1(\theta) + O(b_1^r), \\ \mathbb{E}\{\alpha_2(Y, \theta, h_0)\} &= f_2(\phi_1^-(\theta, h_0))\phi_1^-(\theta_0, h_0)' + O(b_2^r). \end{aligned}$$

Proof. Let us do some simple calculations involving Taylor expansions. We will do it only for the function $w_2(Y, \theta_0, h_0)$. For the function $w_1(X, \theta_0, h_0)$ similar considerations hold. Integration by parts and variable transformation gives

$$\begin{aligned} &\mathbb{E}\left\{H_2\left(\frac{\phi_1^-(\theta, h_0) - Y}{b_2}\right)\right\} - F_2(\phi_1^-(\theta, h_0)) \\ &= \int_{-\infty}^{+\infty} H_2\left(\frac{\phi_1^-(\theta, h_0) - y}{b_2}\right) dF_2(y) - F_2(\phi_1^-(\theta, h_0)) \\ &= \int_{-\infty}^{\infty} \{F_2(\phi_1^-(\theta, h_0) - b_2u) - F_2(\phi_1^-(\theta, h_0))\} K_2(u) du \\ &= \int_{-\infty}^{\infty} \left\{f_2(\phi_1^-(\theta, h_0))(-b_2u) + \dots + \frac{1}{r!} f_2^{(r-1)}(\phi_1^-(\theta, h_0))(-b_2u)^r + o(b_2^r)\right\} K_2(u) du \\ &= O(b_2^r). \end{aligned}$$

Again by integration by parts and variable change for $\mu \in [\phi_1^-(\theta, h_0) - b_2u, \phi_1^-(\theta, h_0)]$ we have

$$\begin{aligned} \mathbb{E}\left\{H_2^2\left(\frac{\phi_1^-(\theta, h_0) - Y}{b_2}\right)\right\} &= \int_{-\infty}^{+\infty} H_2^2\left(\frac{\phi_1^-(\theta, h_0) - y}{b_2}\right) dF_2(y) \\ &= 2 \int_{-\infty}^{\infty} \{F_2(\phi_1^-(\theta, h_0) - b_2u)\} H_2(u) dH_2(u) \\ &= F_2(\phi_1^-(\theta, h_0)) \int_{-\infty}^{\infty} dH_2^2(u) - \int_{-\infty}^{\infty} b_2u f_2(\mu) H_2(u) dH_2(u) \\ &= F_2(\phi_1^-(\theta, h_0)) + O(b_2). \end{aligned}$$

At last we have

$$\begin{aligned}
\mathbb{E}\{\alpha_2(Y, \theta, h_0)\} &= \int_{-\infty}^{+\infty} \frac{\partial H_{b_1}(\theta - x)}{\partial \theta} dF(x) \\
&= \frac{1}{b_1} \int_{-\infty}^{+\infty} K_2\left(\frac{\phi_1^-(\theta, h_0) - x}{b_1}\right) \phi_1^-(\theta, h_0)'_{\theta} dF(x) \\
&= \phi_1^-(\theta, h_0)'_{\theta} \int_{-\infty}^{+\infty} f_2(\phi_1^-(\theta, h_0) - b_1 y) K_2(y) dy \\
&= f_2(\phi_1^-(\theta, h_0)) \phi_1^-(\theta, h_0)'_{\theta} + O(b_2^r).
\end{aligned}$$

Remember that in case of empirical likelihood in the two sample case without smoothing we have $\mathbb{E}w_1(X, \theta_0, h_0) = 0$. From Lemma 31 it follows that $\mathbb{E}w_1(X, \theta_0, h_0) = O(b_1^r)$. Therefore we have to review our results in order to find at what speed b_1 and b_2 should converge to zero asymptotically when the sample sizes $n, m \rightarrow \infty$.

We assume again the set of Assumptions 3 - 6 in case of smoothed empirical likelihood for structural relationships, except the Assumption (A4). From Lemma 31 it follows that we need a slight change due to the smoothing parameters b_1 and b_2 in our Assumption (A4), i.e. we need to assume the following:

(A4') For the functions $w_1(X, \theta_0, \hat{h})$ and $w_2(X, \theta_0, \hat{h})$ assume that

$$Ew_1(X, \theta_0, \hat{h}) = O(\beta_2) + O(b_1^r), \quad (7.6)$$

$$Ew_2(Y, \theta_0, \hat{h}) = O(\beta_2) + O(b_2^r), \quad (7.7)$$

where $\beta_2 = o(n^{-\eta})$.

Later we will verify on which conditions Assumptions 3 - 6 with the modified assumption (A4') hold. But now let us review all lemmas and theorems derived in Section 6.2 in order to find out what convergence rates of the smoothing parameters b_1 and b_2 we need to require. We will not rewrite all proofs, but still indicate the difference due to smoothing, when needed.

7.1 Main results

We separate again the maximization problem and the limiting distribution of the test statistic.

7.1.1 Maximization problem

Corollary 32. *From Assumptions (A2), (A4) it follows that for $\theta \in \{\theta : |\theta - \theta_0| \leq cn^{-\eta}\}$, where $c > 0$ is some constant*

$$S_{1n}(\theta, t, \hat{h}) = O(1); \quad \frac{1}{n} \sum_{i=1}^n w_1^3(X_i, \theta, \hat{h}) = O(1); \quad \frac{1}{n} \sum_{i=1}^n \alpha_1(X_i, \theta, \hat{h}) = O(1) \quad a.s. \quad (7.8)$$

$$M_{1n}(\theta, t, \hat{h}) = O(b_1^r) + O(n^{-\eta}) \quad a.s. \quad (7.9)$$

Proof. The statement (7.8) follows directly from Lemma 31 and Assumptions (A3), (A4). For example

$$\frac{1}{n} \sum_{i=1}^n \alpha_1(X_i, \theta, \hat{h}) = \mathbb{E}\alpha_1(X, \theta, \hat{h}) + O(\beta_1) \quad a.s. \quad (7.10)$$

$$= f_1(\theta) + O(b_1^r) + O(\beta_2) + O(\beta_1) = O(1) \quad a.s. \quad (7.11)$$

Lemma 33. *Let $1/3 < \eta < 1/2$. Suppose that Assumption 7 and Assumption 4 with modified condition (A_4') are satisfied. Then*

$$\lambda_1(\theta) = O(b_1^r) + O(n^{-\eta}), \quad \lambda_2(\theta) = O(b_2^r) + O(m^{-\eta}) \quad a.s. \quad (7.12)$$

uniformly about $\theta \in \{\theta : |\theta - \theta_0| \leq cn^{-\eta}\}$, where c is some positive constant. Consequently, we also have

$$\lambda_1(\theta_0) = O(b_1^r) + O(\beta_2), \quad \lambda_2(\theta_0) = O(b_2^r) + O(\beta_2) \quad a.s. \quad (7.13)$$

Proof. Similarly as in (6.21) we have

$$|\lambda_1(\theta)| |(S_{1n}(\theta, t, \hat{h}) - \max_{1 \leq i \leq n} |w_1(X_i, \theta, \hat{h})| M_{1n}(\theta, t, \hat{h}))| \leq M_{1n}(\theta, t, \hat{h}).$$

Let C_1 denote an upper bound to $|H|$. Then $C_2 = C_1 + \Delta$ is an upper bound to $|w_1(X_i, \theta, \hat{h})|$ for all i and b_1 . It follows directly from Corrolary 32 that

$$|\lambda_1|(O(1) - C_2(O(b_1^r) + O(n^{-\eta}))) \leq O(b_1^r) + O(n^{-\eta}) \quad a.s. \quad (7.14)$$

So, $\lambda_1 = O(b_1^r) + O(n^{-\eta})$ almost surely.

Lemma 34. *Assume that Assumption 7 and Assumption 4 with the modified condition (A_4') are satisfied and η is as in Lemma 33. Then with probability tending to 1, there exists a root $\theta_E^{(n,m)}$ of (5.28) such that*

$$|\theta_E^{(n,m)} - \theta_0| = O(n^{-\eta}), \quad (7.15)$$

and $-\log R_E^{(sm)}(\Delta, \theta, t, \hat{h})$ attains its local maximum value at $\theta_E^{(n,m)}$.

Proof. Similarly as (6.25) we get

$$\lambda_1(\theta) = \frac{M_{1n}(\theta, t, \hat{h})}{S_{1n}(\theta, t, \hat{h})} + O(b_1^r + n^{-\eta})^2 = O(b_1^r + n^{-\eta}) \quad (7.16)$$

uniformly about $\theta \in \{\theta : |\theta - \theta_0| \leq cn^{-\eta}\}$. By (7.16) and Taylor expansion with $\theta' \in [\theta_0, \theta - \theta_0]$ we have similarly as in proof of Lemma 25

$$\begin{aligned} & H_{1E}(\Delta, \theta, t, \hat{h}) \\ &= \frac{n}{2} \left(\frac{1}{n} \sum_{i=1}^n w_1(X_i, \theta, \hat{h}) \right)^2 \left(\frac{1}{n} \sum_{i=1}^n w_1^2(X_i, \theta, \hat{h}) \right)^{-1} + O(n)O(b_1^r + n^{-\eta})^3 \quad a.s. \\ &= \frac{n}{2} \left(\frac{1}{n} \sum_{i=1}^n w_1(X_i, \theta_0, \hat{h}) + \frac{1}{n} \sum_{i=1}^n \alpha_1(X_i, \theta', \hat{h})(\theta - \theta_0)^2 \right)^2 O(1) + O(n)O(b_1^r + n^{-\eta})^3 \quad a.s. \\ &= \frac{n}{2} \left(\mathbb{E}(w_1(X, \theta_0, \hat{h})) + O(\beta_1) + O(1)O(n^{-\eta}) \right)^2 + O(n)O(b_1^r + n^{-\eta})^3 \quad a.s. \\ &= \frac{n}{2} (O(b_1^r) + O(\beta_2) + O(\beta_1) + O(n^{-\eta}))^2 + O(n)O(b_1^r + n^{-\eta})^3 \quad a.s. \end{aligned}$$

Similarly,

$$\begin{aligned} & H_{1E}(\Delta, \theta_0, t, \hat{h}) \\ &= \frac{n}{2} \left(\frac{1}{n} \sum_{i=1}^n w_1(X_i, \theta_0, \hat{h}) \right)^2 \left(\frac{1}{n} \sum_{i=1}^n w_1^2(X_i, \theta_0, \hat{h}) \right)^{-1} + O(n)O(b_1^r + O(\beta_2) + O(\beta_1))^3 \quad a.s. \\ &= \frac{n}{2} (O(b_1^r) + O(\beta_2) + O(\beta_1))^2 + O(n)O(b_1^r + O(\beta_2) + O(\beta_1))^3 \quad a.s. \end{aligned}$$

From the proof of Lemma 25 we can conclude for $|\theta - \theta_0| = n^{-\eta}$, with probability tending to one

$$-2 \log R_E^{(sm)}(\Delta, \theta, t, \hat{h}) > -2 \log R_E^{(sm)}(\Delta, \theta_0, t, \hat{h}).$$

So, there exists a root $\theta_E^{(n,m)}$ of (5.28) such that $|\theta_E^{(n,m)} - \theta_0| = O(n^{-\eta})$ and $-2 \log R_E^{(sm)}(\Delta, \theta, t, \hat{h})$ attains its local minimum value at $\theta_E^{(n,m)}$.

7.1.2 Limiting distribution

Lemma 35. *Suppose that $1/3 < \eta < 1/2$ and Assumptions 5 and 7 are satisfied. Then*

$$\lambda_1(\theta) = O_p(b_1^r) + O_p(n^{-\eta}), \quad \lambda_2(\theta) = O_p(b_2^r) + O_p(m^{-\eta}) \quad (7.17)$$

uniformly about $\theta \in \{\theta : |\theta - \theta_0| \leq cn^{-\eta}\}$, where c is some positive constant.

Proof. Similarly as in proof of Lemma 33 (cf. inequality (6.21)) we have

$$|\lambda_1(\theta)|(S_{1n}(\theta, t, \hat{h}) - \max_{1 \leq i \leq n} |w_1(X_i, \theta, \hat{h})| M_{1n}(\theta, t, \hat{h})) \leq M_{1n}(\theta, t, \hat{h}). \quad (7.18)$$

From Taylor expansion we get for $\theta' \in [\theta, \theta - \theta_0]$

$$\begin{aligned} M_{1n}(\theta, t, \hat{h}) &= M_{1n}(\theta_0, t, \hat{h}) + \frac{1}{n} \sum_{i=1}^n \alpha_1(X_i, \theta', \hat{h})(\theta - \theta_0)^2 \\ &= O(b_1^r) + O_p(n^{-1/2}) + O_p(1)O_p(n^{-\eta}) = O_p(b_1^r) + O_p(n^{-\eta}). \end{aligned}$$

Thus, from Corollary 32 it follows

$$|\lambda_1|(O_p(1) - C_2 O_p(b_1^r + n^{-\eta})) \leq O_p(b_2^r + n^{-\eta}),$$

So, $\lambda_1 = O_p(b_1^r + n^{-\eta})$.

Lemma 36. *Assume that Assumptions 3, 5, 6, 7 and 8 hold. Assume additionally that the maximization problem has a solution, i.e. lemma 34 is true. Then with $\theta_E^{(n,m)}$ such as in Lemma 34 we have*

$$\sqrt{n}(\theta_E^{(n,m)} - \theta_0) \rightarrow_d N \left(0, \frac{M_1 V_1}{M_1(t) V_3^2(t) + k V_1(t) M_3^2(t)} \right), \quad (7.19)$$

$$\lambda_1(\theta_E^{(n,m)}) = -k \frac{M_3(t)}{V_3(t)} \lambda_2(\theta_E^{(n,m)}) + o_p(n^{-1/2}), \quad (7.20)$$

$$\sqrt{n} \lambda_2(\theta_E^{(n,m)}) \rightarrow_d N \left(0, \frac{V_3^2(t)}{k(M_1(t) V_3^2(t) + k V_1(t) M_3^2(t))} \right). \quad (7.21)$$

Proof. The proof is the same as the proof of Lemma 27. By Taylor expansion, Lemma 34 and Lemma 35, we have

$$\begin{aligned} 0 &= Q_{in}(\theta_E^{(n,m)}, \lambda_{1E}, \lambda_{2E}) = Q_{in}(\theta_0, 0, 0) + \frac{\partial Q_{in}(\theta_0, 0, 0)}{\partial \theta} (\theta_E^{(n,m)} - \theta_0) \\ &\quad + \frac{\partial Q_{in}(\theta_0, 0, 0)}{\partial \lambda_1} \lambda_{1E} + \frac{\partial Q_{in}(\theta_0, 0, 0)}{\partial \lambda_2} \lambda_{2E} + O_p(b_1^r + n^{-\eta})^2, \quad i = 1, 2, 3. \end{aligned} \quad (7.22)$$

Due to Assumption 8 on bandwidths b_1 and b_2 we have

$$\begin{aligned} Q_{in}(\theta_0, 0, 0) + \frac{\partial Q_{in}(\theta_0, 0, 0)}{\partial \theta} (\theta_E^{(n,m)} - \theta_0) + \frac{\partial Q_{in}(\theta_0, 0, 0)}{\partial \lambda_1} \lambda_{1E} + \frac{\partial Q_{in}(\theta_0, 0, 0)}{\partial \lambda_2} \lambda_{2E} \\ = o_p(n^{-1/2}), \quad i = 1, 2, 3. \end{aligned} \quad (7.23)$$

We have again

$$\begin{aligned} \theta_E^{(n,m)} - \theta_0 &= -\frac{1}{c_1} (M_1(t)V_3(t)Q_{1n}(\theta_0, 0, 0) + kV_1(t)M_3(t)Q_{2n}(\theta_0, 0, 0)) + o_p(n^{-1/2}), \\ \lambda_1 &= \left(\frac{kM_3(t)}{c_1} \right) (M_3Q_{1n}(\theta_0, 0, 0) - V_3(t)Q_{2n}(\theta_0, 0, 0)) + o_p(n^{-1/2}), \\ \lambda_2 &= -\left(\frac{V_3(t)}{c_1} \right) (M_3Q_{1n}(\theta_0, 0, 0) - V_3(t)Q_{2n}(\theta_0, 0, 0)) + o_p(n^{-1/2}). \end{aligned}$$

and the lemma follows from the fact that

$$\sqrt{n} \begin{pmatrix} Q_{1n}(\theta_0, 0, 0) \\ Q_{2n}(\theta_0, 0, 0) \end{pmatrix} \rightarrow_d N \begin{pmatrix} V_1(t) & 0 \\ 0 & k^{-1}M_1(t) \end{pmatrix}.$$

Theorem 37. *Assume that $\hat{h} = h_0 + O(\beta)$ almost surely, where $\beta = o(n^{-1/3})$. For structural relationship models with estimating equations (7.4) and (7.5) let Assumptions 3, 7 and 8 hold. Then Assumption 4 holds with the modified condition (A_4'), Assumptions 5 and 6 hold. Moreover, there exists a root $\theta_E^{(n,m)}$ of (5.28) such that $\theta_E^{(n,m)}$ is a consistent estimate for θ_0 , the function $R_E^{(sm)}(\Delta, \theta, t, \hat{h})$ attains its local maximum value at $\theta_E^{(n,m)}$,*

$$V_1(t) = V_2(t) = \Delta(1 - \Delta); \quad (7.24)$$

$$M_1(t) = M_2(t) = \phi_2(t, h_0)(1 - \phi_2(t, h_0)); \quad (7.25)$$

$$V_3(t) = f_1(\theta_0); \quad M_3(t) = f_2(\phi_1^-(\theta, h_0))\phi_1^-(\theta_0, h_0)'_{\theta} \quad (7.26)$$

and there is no perturbation due to plug-in, that means

$$-2 \log R_E^{(sm)}(\Delta, \theta_E^{(n,m)}, t, \hat{h}) \rightarrow_d \chi_1^2 \quad (7.27)$$

for $t \in \mathcal{T}$ as $n \rightarrow \infty$.

Proof. Is similar to the proof of Theorem 29. From Corollary 45, page 70 in the Appendix we know that the class of functions $\{H_{b_2}(t - y), t \in \mathbb{R}\}$ has the Glivenko-Cantelli property. Clearly Assumptions 5 and 6 hold (see also the proof of Theorem 29).

Now let us verify the functions $V_1(t), V_2(t), V_3(t)$ and $M_1(t), M_2(t), M_3(t)$. It is easy to check that

$$V_1(t) = V_2(t) = \Delta(1 - \Delta),$$

because the first estimating equation (7.4) involves only the function w_1 , which for structural relationship models does not involve nuisance parameters. The situation with functions $M_1(t)$ and $M_2(t)$ is more difficult. Corollary 45, page 70 in the Appendix and Lemma 48, page 71 in the Appendix state that smoothed empirical processes have Glivenko-Cantelli and Donsker

properties. Heuristically this combined with Lemma 42 and Lemma 43 on the page 68 in the Appendix gives us the desired result, but let us make it more precise. Let us check now Assumption (B2) for $S_{2m}(\theta_0, t, \hat{h})$:

$$\begin{aligned} S_{2m}(\theta_0, t, \hat{h}) &= \frac{1}{m} \sum_{j=1}^m w_2^2(Y_j, \theta_0, \Delta, t, \hat{h}) \\ &= \frac{1}{m} \sum_{j=1}^m H_{b_2}^2(\phi_1^-(\theta_0, \hat{h}) - Y_j) - 2\phi_2(t, \hat{h}) \frac{1}{m} \sum_{j=1}^m H_{b_2}(\phi_1^-(\theta_0, \hat{h}) - Y_j) + \phi_2^2(t, \hat{h}) \\ &= \tilde{F}'_{2m}(\phi_1^-(\theta_0, \hat{h})) - 2\tilde{F}_{2m}(\phi_1^-(\theta_0, \hat{h}))\phi_2(t, \hat{h}) + \phi_2^2(t, \hat{h}), \end{aligned}$$

where \tilde{F}' and \tilde{F} are some perturbed empirical distribution functions defined in (10.3), page 69 in the Appendix. From Corollary 45 and Lemma 42 it follows that $\tilde{F}'_{2m}(\phi_1^-(\theta_0, \hat{h})) \rightarrow F_2(\phi_1^-(\theta_0, h_0)) = \phi_2(t, h_0)$ and $\tilde{F}_{2m}(\phi_1^-(\theta_0, \hat{h})) \rightarrow \phi_2(t, h_0)$ almost surely. This leads to the conclusion that $M_2(t) = \phi_2(t, h_0)(1 - \phi_2(t, h_0))$.

Now we need to treat Assumption (B1), i.e. we have to analyze the behavior of $\sqrt{m}M_{2m}(\theta_0, t, \hat{h})$ defined in (6.3). Note that the class $\{H_{b_2}(t - y), t \in \mathbb{R}\}$ is Donsker (cf. Lemma 48). Applying now Lemma 43 we get that

$$\sqrt{m}M_{2m}(\theta_0, t, \hat{h}) \rightarrow_d \sqrt{m}M_{2m}(\theta_0, t, h_0)$$

which is distributed as $N(0, \phi_2(t, h_0)(1 - \phi_2(t, h_0)))$. It is easy to see from Lemma 31 that $V_3 = f_1(\theta_0)$, $M_3 = \phi_1^-(\theta_0, h_0)f_2(\phi_1^-(\theta_0, h_0))$.

Regarding the limiting distribution we have $V_1(t) = V_2(t)$ and $M_1(t) = M_2(t)$ again. Thus, the limiting test statistic has asymptotically the chi-square distribution with one degree of freedom. Still we need to check out which conditions on the bandwidths b_1 and b_2 we need.

First note that from Lemma 35 and Assumption 8 with $\eta \in (1/3; 1/2)$ we have

$$\frac{n}{3} \lambda_1^3(\theta_E^{(n,m)}) \frac{1}{n} \sum_{i=1}^n w_1^3(X_i, \theta_E^{(n,m)}, \hat{h}) = O(n)O_p(b_1^r + n^{-\eta})^3 O_p(1) = o_p(1). \quad (7.28)$$

Using Taylor expansion, we have

$$\begin{aligned} \log R(\Delta, \theta_E^{(n,m)}, t, \hat{h}) &= -n\lambda_1(\theta_E^{(n,m)})M_{1n}(\theta_E^{(n,m)}, t, \hat{h}) + \frac{n}{2}\lambda_1^2(\theta_E^{(n,m)})S_{1n}(\theta_E^{(n,m)}, t, \hat{h}) \\ &\quad - m\lambda_2(\theta_E^{(n,m)})M_{2m}(\theta_E^{(n,m)}, t, \hat{h}) + \frac{m}{2}\lambda_2^2(\theta_E^{(n,m)})S_{2m}(\theta_E^{(n,m)}, t, \hat{h}) + o_p(1). \end{aligned} \quad (7.29)$$

Now from (5.25) and (5.26) similarly as in (6.24), we have

$$M_{1n}(\theta_E^{(n,m)}, t, \hat{h}) = \lambda_1(\theta_E^{(n,m)})S_{1n}(\theta_E^{(n,m)}, t, \hat{h}) + O_p(b_1^r + n^{-\eta})^2$$

and

$$M_{2m}(\theta_E^{(n,m)}, t, \hat{h}) = \lambda_2(\theta_E^{(n,m)})S_{2m}(\theta_E^{(n,m)}, t, \hat{h}) + O_p(b_1^r + n^{-\eta})^2.$$

We have

$$\begin{aligned} S_{1n}(\theta_E^{(n,m)}, t, \hat{h}) &= V_2(t) + o_p(1) = V_1(t) + o_p(1), \\ S_{2m}(\theta_E^{(n,m)}, t, \hat{h}) &= M_2(t) + o_p(1) = M_1(t) + o_p(1). \end{aligned}$$

and using Assumption 8 from Lemma 36 similarly as in the proof of Theorem 21 it follows

$$-2 \log R(\Delta, \theta_E^{(n,m)}, t, \hat{h}) \rightarrow_d \chi_1^2.$$

Remark 38. *We need Assumption 8 for the equation (7.28) to hold. For deriving the test statistic this is crucial.*

7.2 Bootstrap confidence bands

It is well known that pointwise confidence intervals are not optimal in the sense that one can not judge about behavior of the whole curve or function of interest. Nevertheless for the i.i.d case empirical likelihood has been constructed in literature mostly in the context of pointwise intervals. See for example Owen (1988, 1990, 1992), Qin and Lawless (1994), Chen and Hall (1993), etc. If data are censored, a simultaneous inference is done usually in a natural way, because in survival analysis a supremum over the Brownian bridges usually is involved in asymptotic of test statistics (cf. Einmahl and McKeague, 1999; McKeague and Zhao, 2002, etc.).

To construct a simultaneous confidence region for Δ over the interval (a, b) , we will use the bootstrap confidence region without losing advantages of the automated shape-determination by empirical likelihood method. It means that we use empirical likelihood to set the shape of the confidence bands and use the bootstrap to set the level. This method is introduced by Hall and Owen (1993) and has been also used by Claeskens *et al.* (2003). Let us shortly present it here.

First we nonparametrically draw a resample from each of the two samples separately to obtain $\{X_1^*, \dots, X_n^*\}$ and $\{Y_1^*, \dots, Y_m^*\}$. Then we form the bootstrap smoothed empirical log-likelihood ratio

$$\begin{aligned} -2 \log R_E^{*(sm)}(\Delta, \theta_E^{*(n,m)}, t, \hat{h}) &= 2 \sum_{i=1}^n \log(1 + \lambda_1^*(\theta_E^{*(n,m)}) w_1(X_i^*, \theta_E^{*(n,m)}, \hat{h})) \\ &\quad + 2 \sum_{j=1}^m \log(1 + \lambda_2^*(\theta_E^{*(m,m)}) w_2(Y_j^*, \theta_E^{*(n,m)}, \hat{h})), \end{aligned} \quad (7.30)$$

where, for the structural relationships

$$w_1(X_i^*, \theta_E^{*(n,m)}, \hat{h}) = H_{b_1}(\theta_E^{*(n,m)} - X_i^*) - \Delta, \quad (7.31)$$

$$w_2(Y_j^*, \theta_E^{*(n,m)}, \hat{h}) = H_{b_2}(\phi_1^-(\theta_E^{*(n,m)}, \hat{h}) - Y_j^*) - \phi_2(t, \hat{h}). \quad (7.32)$$

The bootstrap values of $\lambda_1^*(\theta_E^{*(n,m)})$, $\lambda_2^*(\theta_E^{*(n,m)})$ and $\theta_E^{*(n,m)}$ are the solutions to the equations (5.25), (5.26) and (5.28).

To proceed we need to give a remark concerning the smoothed empirical likelihood estimator of P-P plot itself.

Remark 39. *The maximum smoothed empirical likelihood estimator is that value of $\hat{\Delta}$ for which the smoothed empirical likelihood function is maximized,*

$$\hat{\Delta} = \arg \max_{\Delta} R_E^{*(sm)}(\Delta, \theta_E^{*(n,m)}, t, \hat{h}). \quad (7.33)$$

Theorem 40. [Claeskens *et al.* (2003), Theorem 1]. *If $\phi_1(\theta, \hat{h}) = \theta$, $\phi_2(t, \hat{h}) = t$ in (6.42) and we have the setup of Claeskens *et al.* (2003), then assuming some regularity conditions (similar to those in our Assumption 7), if $nb_1^{2r} \rightarrow 0$, the smoothed likelihood estimator $\hat{\Delta}(t)$ satisfies*

$$\sqrt{n}(\hat{\Delta}(t) - \Delta(t)) \rightarrow_d N(0, \Delta(t)(1 - \Delta(t)) + (\Delta'(t))^2 t(1 - t)n/m).$$

Proof. Claeskens *et al.* (2003) heuristically shows that the result follows from Lloyd and Yong (1999).

To construct a simultaneous confidence band for the P-P plot process Δ defined in (6.42) from structural relationships over the interval (a, b) , we first construct an appropriate bootstrap critical value c^* such that

$$P(-2 \log R_E^{*(sm)}(\hat{\Delta}, \theta_E^{(n,m)}, t, \hat{h})) \leq c^* \text{ for } a \leq t \leq b | X, Y \} = 1 - \alpha,$$

where $\hat{\Delta} = \hat{\Delta}(t)$ is the smoothed empirical likelihood estimator defined in (7.33). And then apply it for the pointwise confidence bands. Hence the bootstrap confidence band \mathcal{C} consists of those curves $R_E^{(sm)}(\cdot, \theta_E^{(n,m)}, t, \hat{h})$ such that the corresponding log likelihood ratio statistic stays below c^* over the interval (a, b) , that is,

$$\mathcal{C} = \{-2 \log R_E^{(sm)}(\cdot, \theta_E^{(n,m)}, t, \hat{h}) : -2 \log R_E^{(sm)}(\Delta, \theta_E^{(n,m)}, t, \hat{h}) \leq c^* \text{ for } a \leq t \leq b\}.$$

Note that both estimators $\hat{\Delta}$ and \hat{h} are based on the initial samples X_1, \dots, X_n and Y_1, \dots, Y_m . For constructing empirical likelihood confidence bands such bootstrap methods have been introduced first by Hall and Owen (1993) and after applied also by Claeskens *et al.* (2003).

8 Implementation

In this Section we show the implementation of the theory developed using the location model. First applying the smoothed empirical likelihood method we simulate pointwise coverage probabilities. Further we analyze the two data examples, construct and compare the bands for the general shift function Δ_s (further on denoted as Doksum's bands) and the smoothed empirical likelihood bands.

8.1 Simulation study

We provide here simulations for the location model from structural relationships defined in (2.6). We generated 1,000 pseudorandom samples from $F_1 = N(0, 1)$ and $F_2 = N(1, 1)$. Thus, we have a location model with functions $\phi_1(\theta, h) = \theta + h$, $\phi_2(t, h) = t$ and

$$w_1(X_i, \theta, \hat{h}) = H_{b_1}(\theta - X_i) - \Delta, \quad (8.1)$$

$$w_2(Y_j, \theta, \hat{h}) = H_{b_2}(\theta - \hat{h} - Y_j) - t, \quad (8.2)$$

where \hat{h} is a minimizer of Mallows distance functional $T(F_{1n}, F_{2m})$ defined in (2.11) with trimming bounds $a = 0.05$, $b = 1 - a = 0.95$ and $\mathcal{H} \subseteq \mathbb{R}$, i.e.

$$\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}} \left\{ \int_{0.05}^{0.95} (F_{1n}^{-1}(u) - F_{2m}^{-1}(u) - h)^2 du \right\}.$$

For the location model we know that $\hat{h} = h_0 + O(n^{-1/2}(\log \log n)^{1/2})$ almost surely (cf. Remark 3, p. 3). Thus, Theorem 37 is justified.

Throughout we smooth using a biquadratic kernel

$$K_j(t) = \frac{15}{16}(1 - t^2)^2 1_{\{|t| \leq 1\}}, \quad (8.3)$$

and we employ bandwidths $b_1 = n^{-3/20}$ and $b_2 = m^{-3/20}$ which also have been analyzed in Claeskens *et al.* (2003).

For the construction of smoothed empirical likelihood ratio confidence interval, the explicit maximization of the likelihood function can be avoided. To construct the empirical likelihood function as described in Section 5.3, equations (5.25), (5.26) and (5.28) need to be solved, which is however not an easy task, because of complexity of these equations. We act here similarly as Claeskens *et al.* (2003) solving these equations in two stages. In the first stage, we fix θ and solve for $\lambda_1 = \lambda_1(\theta)$ and $\lambda_2 = \lambda_2(\theta)$ from equations (5.25), (5.26). In the second stage, we obtain $\hat{\theta}$ as a solution to equation (5.28), where now $\lambda_j = \lambda_j(\theta)$. This method provides identical solutions to a direct "one-stage" solution to equations (5.25), (5.26) and (5.28).

In order to solve say equation (5.25) for $\lambda_1(\theta)$ we use the following considerations. The value of $\lambda_1(\theta)$ may be found by numerical search. We know that $\lambda_1(\theta)$ solves

$$\frac{1}{n} \sum_{i=1}^n \frac{w_1(X_i, \theta, \hat{h})}{1 + \lambda_1(\theta)w_1(X_i, \theta, \hat{h})} = 0. \quad (8.4)$$

The left side of (8.4) equals $w_1(X_i, \theta, \hat{h})$ at $\lambda_1(\theta) = 0$. It is strictly decreasing in $\lambda_1(\theta)$, as may be found by differentiation (cf. (5.12)). The question is which interval should be taken

to search for $\lambda_1(\theta)$? Monotonicity of (8.4) makes a bisection approach workable, but it is slow. Owen (2001) suggests to use safeguard methods, like Brent's method or some versions of Newton's method, with a superlinear rate of convergence to the solution. But in which interval to look for $\lambda_1(\theta)$? We know that every $p_i > 0$, and of course every $p_i < 1$. Assume now that $w_1(X_1, \theta, \hat{h}) \leq w_1(X_2, \theta, \hat{h}) \leq \dots \leq w_1(X_n, \theta, \hat{h})$. From (5.23) a bracketing interval may be found by alternately setting to 1 the weight on the minimum and maximum observations. Thus, we may start the search knowing that

$$\frac{1 - n^{-1}}{w_1(X_n, \theta, \hat{h})} < \lambda_1(\theta) < \frac{1 - n^{-1}}{w_i(X_1, \theta, \hat{h})}.$$

Using the consideration above for location model confidence intervals are calculated for values of t equal to 0.1, 0.3, 0.5, 0.7 and 0.9 (cf. Tables 1, 2).

Table 1. Coverage accuracy of confidence intervals with nominal level 0.95 (location estimated each time).

	t = 0.1	t=0.3	t=0.5	t=0.7	t=0.9
n=m=50	0.9670	0.9770	0.9780	0.9840	0.9710
n=m=70	0.9780	0.9790	0.9800	0.9790	0.9760
n=m=100	0.9690	0.9760	0.9790	0.9740	0.9770
n=m=200	0.9590	0.9610	0.9510	0.9530	0.9650

Table 2. Coverage accuracy of confidence intervals with nominal level 0.95 (location = 1, known).

	t = 0.1	t=0.3	t=0.5	t=0.7	t=0.9
n=m=50	0.9370	0.9380	0.9360	0.9320	0.9300
n=m=70	0.9390	0.9370	0.9390	0.9500	0.9480
n=m=100	0.9400	0.9259	0.9330	0.9260	0.9400
n=m=200	0.9130	0.8980	0.9110	0.9120	0.9160

8.2 Data examples

Here we analyze the location model with ϕ_1 , ϕ_2 and w_1 , w_2 as in the simulation study (see (8.1) and (8.2)). The same kernels and bandwidths have been used as well. The level is always $\alpha = 0.05$. For construction of the bootstrap simultaneous confidence bands (cf. Section 7.2) we always use 200 bootstrap resamples. The trimming bounds are $a = 0.05$ and $b = 1 - a = 0.95$.

Data example I: Morbidity-Mortality data

For an illustration we use some data from Aaberge *et al.* (1985), where the relationship between hospitalization, as a measure of morbidity, and mortality is examined. The data have been taken from a general medical department in Oslo, Norway. The use of health and social services increases with age, therefore the interest of this two sample problem is to check whether the morbidity and mortality are related.

To compare hospitalization and mortality a comparison has been made between the attained age of the patients at the time of admission and attained age at death in the general population. Let X_i be the attained age of hospitalized person i and let Y_j be attained age at death of person j in the general population. Independent samples X_1, \dots, X_n and Y_1, \dots, Y_m are obtained with cumulative distribution functions F_1 and F_2 . Let us cite now Aaberge *et al.* (1985):

”In the first step of the analysis of hospitalization, we are not willing to be specific about the distribution functions F_1 and F_2 . We were interested primarily in whether the distribution of age at hospitalization may be described by the distribution of age at death or by a simple transformation of that distribution. If that is the case, results from the theory of life tables may be applied in the next step of the analysis of hospitalization; for instance, well-known parametric distribution functions may be fitted to data on hospitalization”.

The patient material in this study consisted of 367 consecutive admissions, 176 males and 191 females for 1980. Data on mortality in Oslo were provided by the Central Bureau of Statistics of Norway and consisted of 6140 deaths, 2989 males and 3151 females. Frequency distributions for data on admissions and deaths are given in Table 3.

Table 3. Number of hospitalized and number of deaths by age and sex.

Age	Females		Males	
	Hospitalized	Dead	Hospitalized	Dead
-29	3	25	5	49
30-34	3	15	5	20
35-39	0	16	4	22
40-44	2	20	8	23
45-49	6	31	6	57
50-54	12	63	5	111
55-59	17	98	14	215
60-64	11	170	28	293
65-69	26	277	35	423
70-74	32	403	27	509
75-79	32	546	21	513
80+	47	1487	18	754
Total	191	3151	176	2989

We did not have an access to the original data. Therefore having only the Table 3 we simulated the data uniformly in each interval. First we construct Doksum’s simultaneous bands (4.5) and (4.7) for function Δ_s (cf. Figure 9).

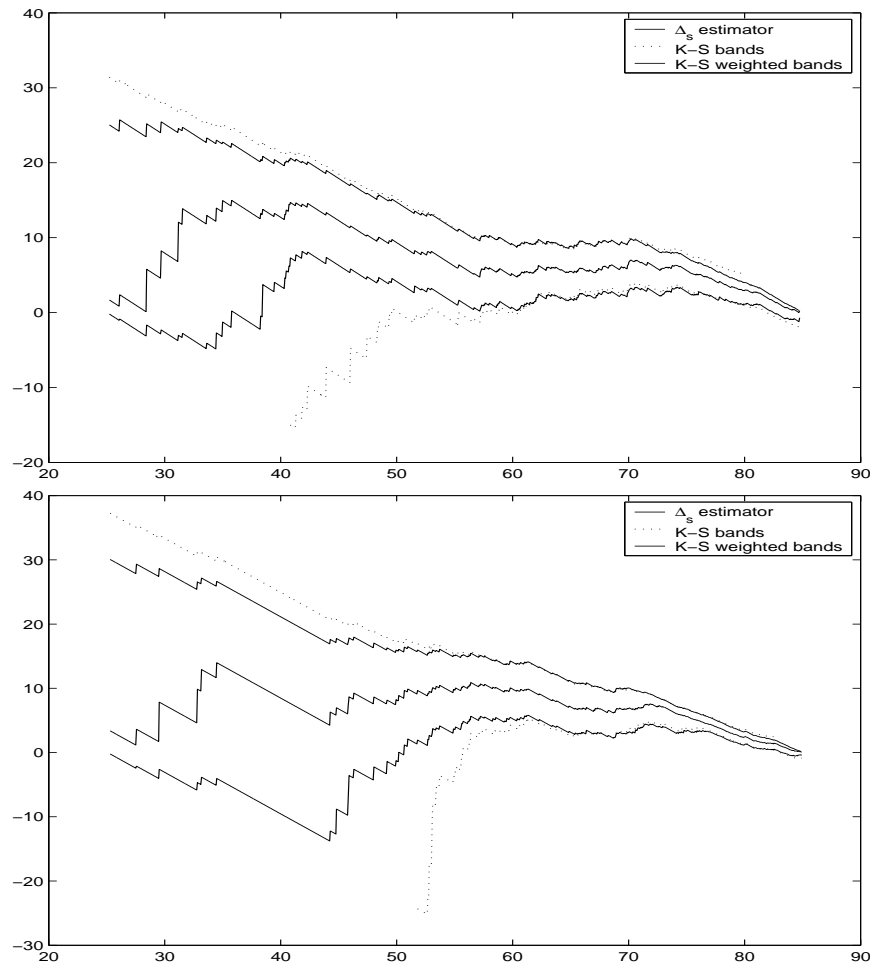


Figure 9: Doksum's bands for Δ_s for mortality data, males and females in upper and lower graph respectively.

We can conclude that for most t the function $\Delta_s(t) = h$, where h is some constant. Thus, the location model holds because we can draw a horizontal line, which "almost" fits into the bands. Only in the right tail the confidence bands become pretty narrow not allowing a horizontal line to fit.

Now we apply the bootstrap method and construct simultaneous confidence bands as described in Section 7.2 to the Mortality data set. We obtained the bootstrapped critical value $c^* = 7.95$ for the male and $c^* = 6.28$ for the female mortality data. The estimated location parameter was $\hat{h} = 6.05$ for males and $\hat{h} = 5.52$ for females. The confidence bands are shown in Figure 10.

The confidence bands for males and females have the same tendency. Mostly the diagonal fits into the bands, thus we can not reject the location model. However, in the tails (mostly in the right tail) the confidence bands are below the diagonal (especially for the female data).

Although both methods provide similar results, we find the empirical likelihood bands preferable. First, the Doksum's bands for males and females seem to be quite similar. The empirical likelihood method reveals that the situation is quite different (see Figure 10). Second, it is easier to find the region, where the location model holds from the bands in Figure 10 than from those in Figure 9.

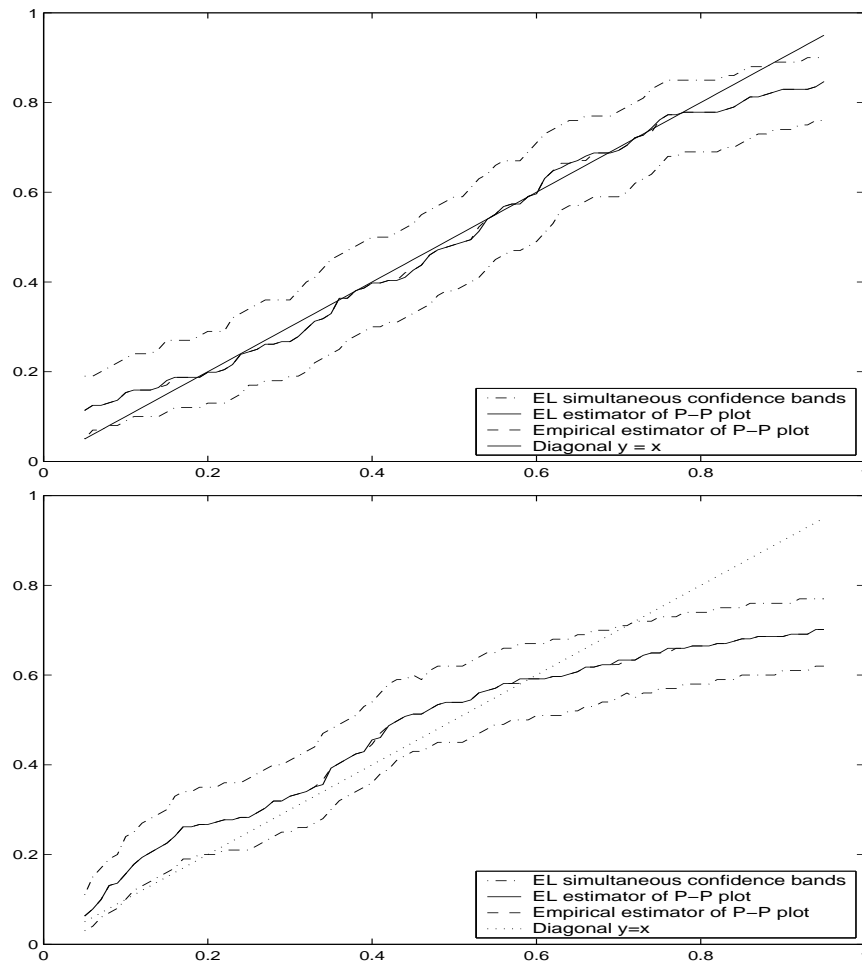


Figure 10: EL simultaneous bands for P-P plot comparing mortality data, males and females in the upper and lower graph respectively.

Data Example II: Sono-Histo data

This statistical analysis concerns a study on 20-MHz-Sonometric and 50-MHz-Sonometric measurements of the size of pigmented skin tumors in 94 patients. For a short illustration of the data see Figure 11.

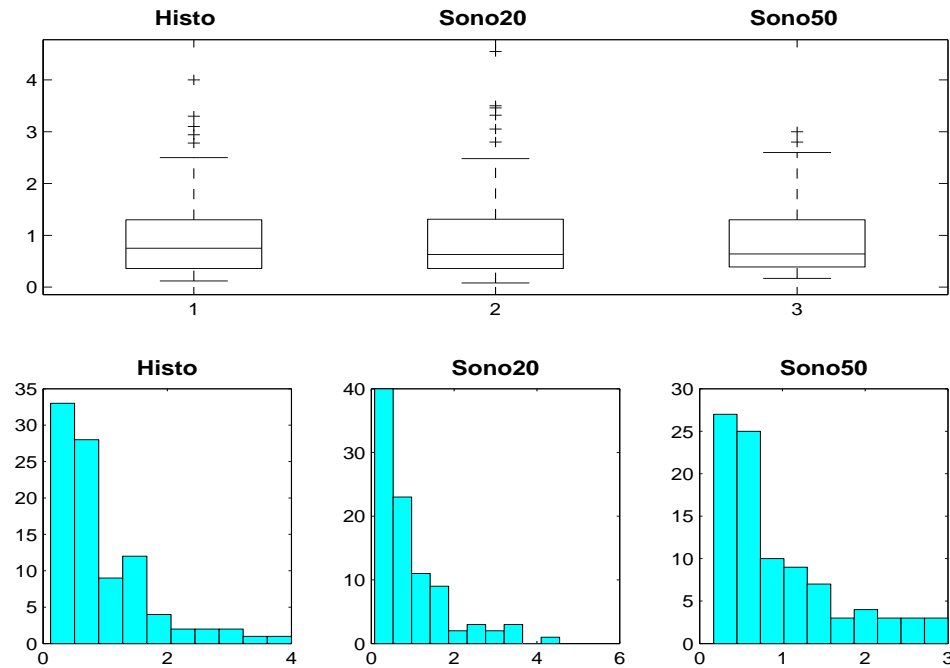


Figure 11: Box plots and histograms of the Sono-Histo data in the upper and lower graph respectively.

For comparing the effectiveness of these two measurement methods with respect to the gold standard (histological measurement) we will draw both simultaneous confidence bands for function Δ_s and the bootstrapped empirical likelihood confidence bands obtained in Section 7.2. We start with Doksum's bands for function Δ_s , which are shown in Figure 12.

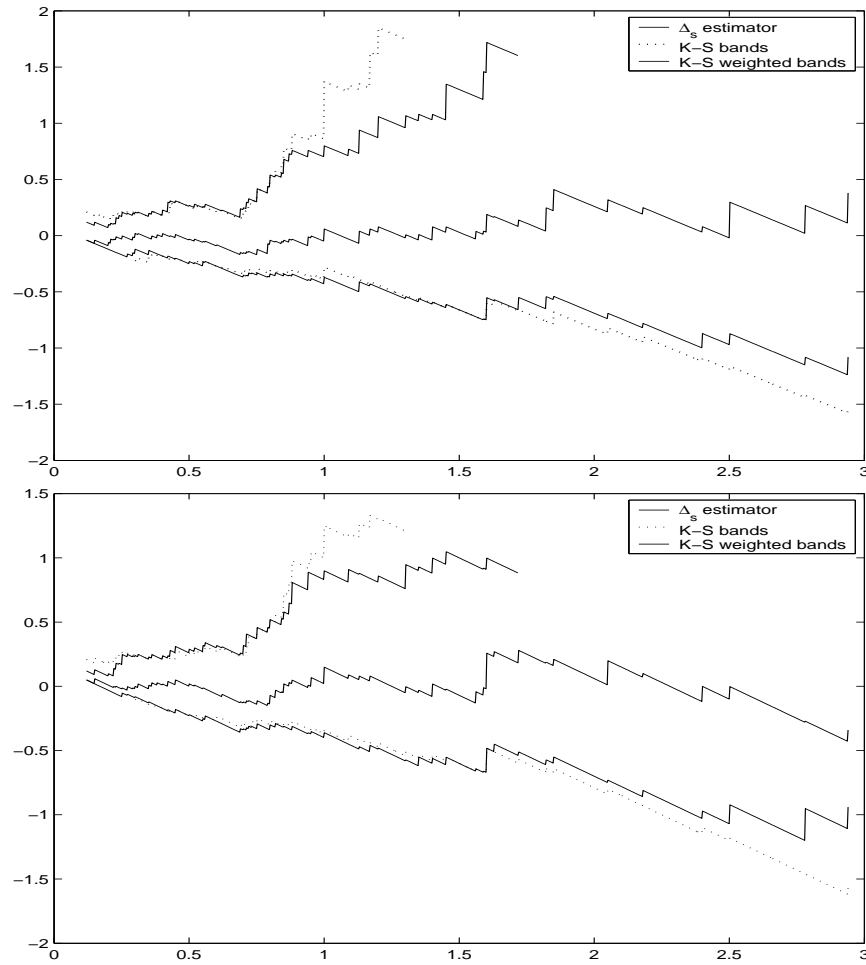


Figure 12: Doksum's bands for Δ_s for comparing Histological data with 20-MHz-Sonometrical and 50-MHz-Sonometrical data in the upper and lower graph respectively.

We see that the weighted Kolmogorov-Smirnov bands are better than the unweighted ones, especially in tails. A horizontal line would fit into the bands, thus with the confidence of $1 - \alpha$ we can not reject the hypothesis that there exists such h with $\Delta_s(t) = h$ for all t .

In Figure 13 the simultaneous confidence bands are constructed for the P-P plot using smoothed two sample plug-in empirical likelihood method. As we could already expect from the Doksum's bands a diagonal fits into the bands. Comparing the histological and 20-MHz-Sonometrical measurements we obtain that the bootstrapped critical value $c^* = 7.38$. For histological and 50-MHz-Sonometrical measurements we got $c^* = 6.57$.

We recognize from the Section 3 that the range of P-P plots and the confidence bands is always between 0 and 1. The function Δ_s with added confidence bands for mortality data ranges approximately from -10 to 30 units and for the Sono-histo data approximately from -2 to 2. Indeed, it is easier to interpret and compare confidence bands in Figure 13 than those in Figure 12. We mentioned already the range preserving property of empirical likelihood. This can be seen if we compare the confidence bands derived from asymptotic theory and Bonferroni's method (cf. Figure 7) with the empirical likelihood bands (cf. Figure 13).

Note that the smoothed empirical likelihood estimator of P-P plot really is much "smoother" than that of the empirical P-P plot, which is also to be expected. We can conclude that both

measurement methods do well in comparison with the histological measurement. They differ only by a small constant ($\hat{h} \approx 0.02$ in case of the Sono20-Histo data and $\hat{h} \approx 0.01$ in case of the Sono50-Histo data).

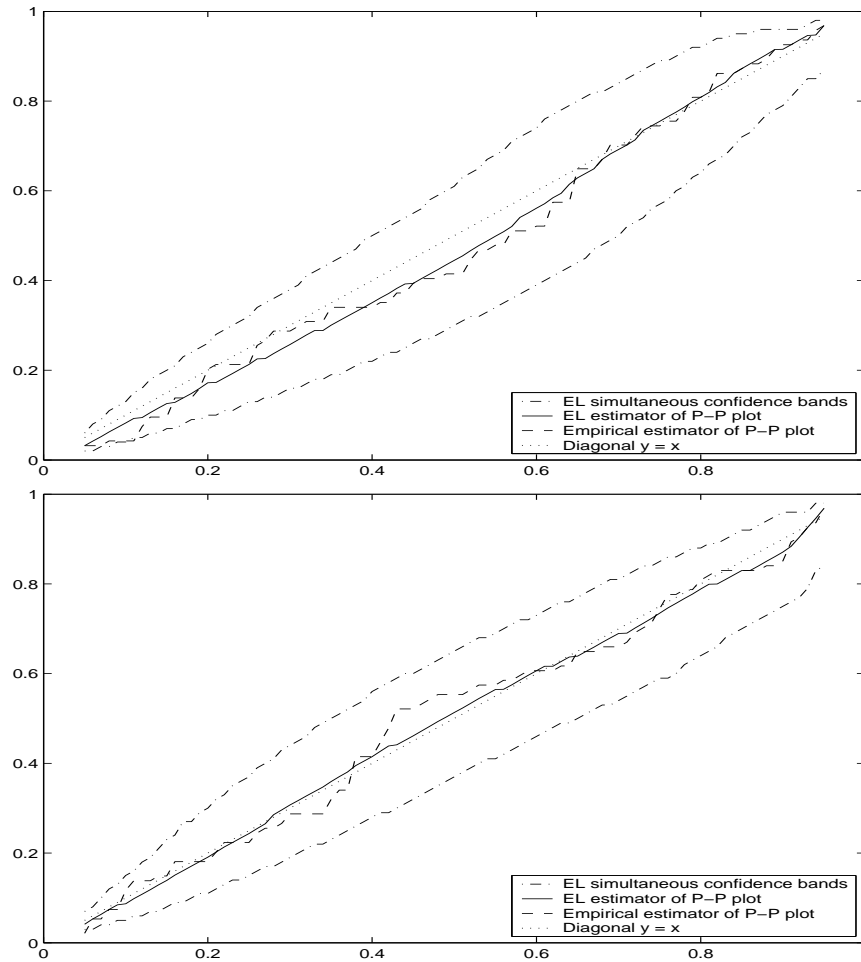


Figure 13: EL simultaneous bands for P-P plot comparing Histological data with 20-MHz-Sonometrical and 50-MHz-Sonometrical data in the upper and lower graph respectively.

9 Discussion

Description of the results

In this thesis we establish smoothed plug-in empirical likelihood for the two-sample case. This is a generalization of several papers: Chen and Hall (1993), Qin and Lawless (1994), Qin and Zhao (2000) and Hjort *et al.* (2004). We have derived the conditions on which the maximization problem has a solution, which is new in the case of two sample plug-in empirical likelihood. Our conditions are slightly better than those in case of known h_0 , i.e. for the strong Glivenko Cantelli property we require the almost sure convergence rate of order $O(\beta_1)$ (cf. Assumption 4), where $\beta_1 = o(n^{-\eta})$ with $\eta \in (1/3, 1/2)$. For a known h_0 the convergence rate is $O(n^{-1/2}(\log n)^{1/2})$ (see e.g. Serfling, 1980, p.95). Thus, we allow slightly slower rates for the Glivenko Cantelli property of empirical processes indexed in $h \in \mathcal{H}$, where h is the unknown parameter for structural relationship models.

Based on our established version of smoothed plug-in empirical likelihood for the two-sample case, we were able to construct the confidence intervals for the structural relationship models introduced by Freitag (2000), Freitag and Munk (2005). More specifically we construct the confidence bands for the general P-P plot for structural relationships. These models generalize the famous location, location-scale models, which are often met in different context in statistical problems. Our regularity conditions are the usual ones, mostly considered in literature. However, our conditions $nb_1^{3r} \rightarrow 0$ and $mb_2^{3r} \rightarrow 0$ on the speed of the smoothing parameters b_1, b_2 , where n, m are the sample sizes and r is the order of the kernel, differ from the results known in literature. This we discuss in detail further.

Comparison of our results with the results in literature

The results of Claeskens *et al.* (2003), where smoothed empirical likelihood is introduced for derivation of the confidence bands for P-P plot, follow directly from the thesis. To see this, let us rewrite Lemma 36 with V_1, M_1 and V_3, M_3 as in Lemma 37.

Lemma 41. *Assume that the conditions of Theorem 37 are satisfied, and $\theta_E^{(n,m)}$ is as that in Lemma 34. Then*

$$\sqrt{n}(\theta_E^{(n,m)} - \theta_0) \rightarrow_d N\left(0, \frac{\Delta(1-\Delta)\phi_2(t, h_0)(1-\phi_2(t, h_0))}{c_1}\right), \quad (9.1)$$

$$\lambda_1(\theta_E^{(n,m)}) = -k \frac{\phi_1^-(\theta_0, h_0)' f_2(\phi_1^-(\theta_0, h_0))}{f_1(\theta_0)} \lambda_2(\theta_E^{(n,m)}) + o_p(n^{-1/2}), \quad (9.2)$$

$$\sqrt{n}\lambda_2(\theta_E^{(n,m)}) \rightarrow_d N\left(0, \frac{f_1^2(\theta_0)}{kc_1}\right), \quad (9.3)$$

where $c_1 = \phi_2(t, h_0)(1 - \phi_2(t, h_0))f_1^2(\theta_0) + k\Delta(1 - \Delta)(\phi_1^-(\theta_0, h_0)' f_2(\phi_1^-(\theta_0, h_0)))^2$.

In case of a known $h = h_0$ and $\phi_1(\theta, h_0) = \theta$, $\phi_2(t, h_0) = t$ the central Lemma 2 in Claeskens *et al.* (2003) follows from the Lemma 41 in the thesis and also the limit distribution for the test statistic is the same. Note that Claeskens *et al.* (2003) use different normalization factor, i.e. instead of \sqrt{n} they use $\sqrt{n+m}$.

There are several points we would like to mention regarding the papers of Claeskens *et al.* (2003) and Jing and Zhou (2003), which have similar problematic, conditions and technique of proving the results. Both of them used smoothed EL constructing the confidence intervals for P-P plots in the two sample case and for the difference of quantiles in the one sample case respectively.

Claeskens *et al.* (2003) require that in our notation (for simplicity, we consider conditions for one sample X_1, \dots, X_n only)

$$nb_1^{4r} \rightarrow 0 \quad \text{and} \quad nb_1^{2r}/\ln n \rightarrow \infty \quad \text{as} \quad n \rightarrow \infty. \quad (9.4)$$

On the other hand Jing and Zhou (2003) require that

$$nb_1^{4r} \rightarrow 0, n^{4s-1}b_1^4 \rightarrow \infty \quad \text{as} \quad n \rightarrow \infty, \quad 1/3 < s < 1/2. \quad (9.5)$$

Chen and Hall (1993) besides some regularity conditions (smoothness of the underlying distribution function) require only that $nb_1^{2r} \rightarrow 0$ for test statistic to have an asymptotic χ_1^2 distribution. This condition on bandwidth b_1 even does not intersect with the conditions in (9.4) or (9.5). Moreover, for the point estimator itself, Theorem 1 in Claeskens *et al.* (2003) is based under the condition $nb_1^{2r} \rightarrow 0$ (cf. Theorem 40). Important is that on the conditions (9.4) the equation (7.28) in this thesis may not hold, which is essential for deriving the limiting distribution.

Mostly in literature the maximization problem has been analyzed with θ such that $\{\theta \in |\theta - \theta_0| \leq cn^{-\eta}\}$ for some $1/3 < \eta < 1/2$ (cf. Qin and Zhao, 2000 or Qin and Lawless, 1994). However, Claeskens *et al.* (2003) analyze the maximization problem for θ such that $|\theta - \theta_0| < \delta$, where $\delta := O(b_1^r + b_2^r)$. Jing and Zhou (2003) consider $\delta := b_1^r + n^{-s}$ for $1/3 < s < 1/2$. This notation can lead to other conditions for the bandwidth b_1 and b_2 , which actually should not be required.

Possible extensions

- The simultaneous confidence bands obtained from the asymptotic theory of P-P and Q-Q plots (cf. Figure 7) could be extended for plug-in estimates. Burke *et al.* (1978) derive the asymptotic for the Kolmogorov-Smirnov test statistic in the one sample when parameters are estimated.
- We constructed the confidence bands for P-P plots. As explained in Section 5.3 it is possible to construct the confidence bands for Q-Q plots in a similar way. This can be further compared with the bands derived by Einmahl and McKeague (1999).

10 Technical tools: Empirical processes

10.1 Definitions and notation

A stochastic process is a collection of random variables $X_t, t \in \mathcal{T}$ on the same probability space, indexed by an arbitrary index set \mathcal{T} . An empirical process is a stochastic process based on a random sample. For example consider a random sample X_1, \dots, X_n of i.i.d. real random variables with distribution F . The empirical distribution function is

$$\mathbb{F}_n(t) = n^{-1} \sum_{i=1}^n 1_{\{X_i \leq t\}}, \quad (10.1)$$

where the index t is allowed to vary over $T = \mathbb{R}$, the real line.

More generally consider a random sample X_1, \dots, X_n from a probability distribution P on a measurable space $(\mathcal{X}, \mathcal{A})$. The empirical distribution is the discrete uniform measure on the observations. We denote it by $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$, where δ_x is the measure which assigns mass 1 at x and zero elsewhere. For a measurable function $f : \mathcal{X} \rightarrow \mathbb{R}$, we write $\mathbb{P}_n f$ for the expectation of f under the empirical measure, and Pf for the expectation under P . Thus,

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^n f(X_i), \quad Pf = \int_{\mathcal{X}} f(x) dP(x).$$

By the law of large numbers, the sequence \mathbb{P}_n converges almost surely to Pf , for every f such that Pf is defined. The abstract Glivenko-Cantelli theorems make this result uniform in f ranging over a class of functions. A class \mathcal{F} of measurable functions $f : \mathcal{X} \rightarrow \mathbb{R}$ is called *P-Glivenko Cantelli* if

$$|\mathbb{P}_n f - Pf|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{P}_n f - Pf| \rightarrow 0 \text{ a.s.},$$

The empirical process evaluated at f is defined as $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf)$. By the multivariate central limit theorem, given any set of measurable functions f_i with $Pf_i^2 < \infty$,

$$(\mathbb{G}_n f_1, \dots, \mathbb{G}_n f_k) \rightarrow_d (\mathbb{G}_P f_1, \dots, \mathbb{G}_P f_k),$$

where the vector on the right possesses a multivariate normal distribution with mean zero and covariances

$$\mathbb{E} \mathbb{G}_P f \mathbb{G}_P g = Pfg - PfPg. \quad (10.2)$$

The abstract Donsker theorems make this result "uniform" in classes of functions. A class \mathcal{F} of measurable functions $f : \mathcal{X} \rightarrow \mathbb{R}$ is called *P-Donsker* if the sequence of processes $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ converges in distribution to a tight limit process in the space $l^\infty(\mathcal{F})$ (the space of bounded functions with $f \in \mathcal{F}$). Then the limit process is a Gaussian process \mathbb{G}_P with zero mean and covariance function as given in the preceding display and is known as a *P-Brownian Bridge*. The Donsker property includes the requirement that the sample paths $f \rightarrow \mathbb{G}_n f$ are uniformly bounded for every n and every realization of X_1, \dots, X_n . This is the case, for instance, if the class \mathcal{F} has a finite and integrable *envelope function* F : a function such that $|f(x)| \leq F(x) < \infty$.

Example 9. *Empirical distribution function.* Setting $\mathcal{X} = \mathbb{R}$, we can now re-express \mathbb{F}_n defined in (10.1) as the empirical process $\{\mathbb{P}_n f, f \in \mathcal{F}\}$, where $\mathcal{F} = \{1_{\{x \leq t\}}, t \in \mathbb{R}\}$. Thus, one can view

the stochastic process \mathbb{F}_n as indexed by either $t \in \mathbb{R}$ or $f \in \mathcal{F}$. Glivenko (1933) and Cantelli (1993) showed that

$$\sup_{t \in \mathbb{R}} |\mathbb{F}_n(t) - F(t)| \rightarrow 0 \text{ a.s.}$$

Another way of saying this is that the sample paths of \mathbb{F}_n get uniformly closer to F as $n \rightarrow \infty$.

Donsker's (1952) theorem tells us that $\mathcal{F} = \{1_{\{x \leq t\}}, t \in \mathbb{R}\}$ is Donsker for all probability measures which are based on some real distribution function F . With $f(x) = 1_{\{x \leq t\}}$ and $g(x) = 1_{\{x \leq s\}}$,

$$\mathbb{E}[f(X)g(X)] - \mathbb{E}f(X)\mathbb{E}g(X) = F(s \wedge t) - F(s)F(t).$$

The language of Glivenko-Cantelli classes and Donsker classes appears to be convenient to state the "regularity conditions" needed in the asymptotic analysis of many statistical procedures. In the following we state some results from van der Vaart (1998), Section 19.

Lemma 42. *If a class \mathcal{F} of functions is a P -Glivenko Cantelli class of measurable functions and a sequence $\hat{f}_n \rightarrow f_0$ almost surely, where $\hat{f}_n \in \mathcal{F}$ and \hat{f}_n is dominated (i.e. there exists an integrable function, say $g(x)$ with $|\hat{f}_n(x)| \leq g(x)$ for each n and almost every x). Then*

$$P_n \hat{f}_n \rightarrow P f_0 \text{ a.s.}$$

Proof. As the class \mathcal{F} is Glivenko Cantelli, then $|\mathbb{P}_n \hat{f}_n - P \hat{f}_n| \rightarrow 0$ almost surely for every sequence \hat{f}_n that are contained in \mathcal{F} . If \hat{f}_n converges almost surely to a function f_0 and the sequence is dominated (or uniformly integrable), so that $P \hat{f}_n \rightarrow P f_0$ almost surely, then it follows that $\mathbb{P}_n \hat{f}_n \rightarrow P f_0$ almost surely.

Lemma 42 covers an important case of controlling random sequences of the form $\sum_{i=1}^n f_{n,\hat{\theta}_n}(X_i)$ for functions $f_{n,\theta}$ that change with n and depend on an estimated parameter. Similar principle applies to Donsker classes of functions.

Lemma 43 (van der Vaart 1998, Lemma 19.24). *Suppose that \mathcal{F} is a P -Donsker class of measurable functions and \hat{f}_n is a sequence of random functions that take values in \mathcal{F} such that $\int (\hat{f}_n(x) - f_0(x))^2 dP(x)$ converges in probability to 0 for some $f_0 \in L_2(P)$. Then $\mathbb{G}_n(\hat{f}_n - f_0) \rightarrow_p 0$ and hence*

$$\mathbb{G}_n \hat{f}_n \rightarrow_d \mathbb{G}_P f_0.$$

Proof. For a Donsker class \mathcal{F} , the empirical process $\mathbb{G}_n f$ converges in distribution to a P -Brownian bridge process $\mathbb{G}_P f$ uniformly in $f \in \mathcal{F}$. One can show that the limiting process has uniformly continuous sample paths with respect to the variance semimetric (cf. van der Vaart 1998, Lemma 18.15, p.262). The uniform convergence combined with the continuity yield the weak convergence $\mathbb{G}_n \hat{f}_n \rightarrow_d \mathbb{G}_P f_0$ for every sequence \hat{f}_n of random functions that are contained in \mathcal{F} and converges in the variance semimetric to a function f_0 .

10.2 Smoothed empirical process

Let X_1, \dots, X_n be a sequence of random variables with common law P . P is assumed to be a probability measure on \mathcal{B} , the usual Borel σ -algebra. The n -th empirical measure is

$$\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i},$$

where δ_x denotes the unit mass at x . It is a discrete measure putting mass $1/n$ at each of the observations. It is also the non-parametric maximum likelihood estimator for the common law P of the observations. If the underlying measure P is smooth it is natural to use a smoothed version of the empirical measure as an estimator of P , rather than the empirical measure itself.

Let \mathcal{F} be a class of real-valued measurable functions on \mathbb{R} . Let us remind that the empirical process indexed by \mathcal{F} is denoted by

$$(\mathbb{P}_n - P)f := \int f(d\mathbb{P}_n - dP), \quad f \in \mathcal{F}.$$

Let now $H_n, n \geq 1$, be a sequence of distribution functions; assume that $H_n \rightarrow_p H_0$ (i.e., $\int f dH_n \rightarrow \int f dH_0$ whenever $f : \mathbb{R} \rightarrow \mathbb{R}$ is continuous and bounded), where H_0 is the distribution for δ_0 . Define now the perturbed empirical distribution function

$$\tilde{F}_n(x) := n^{-1} \sum_{i=1}^n H_n(x - X_i). \quad (10.3)$$

The estimator \tilde{F}_n differs from the usual empirical distribution function $\mathbb{F}_n(x) = n^{-1} \sum_{i=1}^n H_0(x - X_i)$ in that the mass n^{-1} is no longer concentrated at X_i , but is distributed around X_i according to H_n . The assumption $H_n \rightarrow_d H_0$ helps insure that the asymptotic behavior of $\tilde{F} - F$ will be close of that of $\mathbb{F}_n - F$, where F is the distribution function of X .

Winter (1973) and Yamato (1973) have shown that if P is continuous then \tilde{F}_n has the *Glivenko Cantelli* property, that is,

$$\sup_{t \in \mathbb{R}} |\tilde{F}_n(t) - F(t)| \rightarrow 0 \quad a.s., \quad (10.4)$$

which is a clear consequence of Glivenko-Cantelli theorem (see e.g. Tucker (1967), p.127-128).

Now consider a *general* perturbed empirical process

$$(\tilde{\mathbb{P}}_n - P)f := \int f(d\tilde{\mathbb{P}}_n - dP), \quad f \in \mathcal{F},$$

where $\tilde{\mathbb{P}}_n$ denotes the perturbed probability measure associated with \tilde{F}_n and \mathcal{F} a general class of measurable functions not necessarily of the form $\{1_{(-\infty, x]}, x \in \mathbb{R}\}$ as required in (10.4).

It is natural to consider perturbed empirical measure $\tilde{\mathbb{P}}$ as a convolution by $\tilde{\mathbb{P}} := \mu_n * \mathbb{P}_n$, where μ_n is a sequence of probability measures on \mathbb{R} converging weakly to zero ($\mu_n \rightarrow_d \delta_0$). Note that this includes kernel smoothing, by choosing μ_n so that the densities $d\tilde{F}_n/d\lambda$ take the form

$$\frac{d\tilde{F}_n}{d\lambda} = n^{-1} \sum_{i=1}^n b_n^{-1} K\left(\frac{x - X_i}{b_n}\right), \quad (10.5)$$

where K is some fixed kernel and b_n is a sequence of positive reals tending to zero.

Theorem 44 (Yukich (1989)). *Let \mathcal{F} be a Glivenko-Cantelli class of measurable functions that is closed under translation. Then*

$$\left| \int f(d\tilde{\mathbb{P}}_n - dP) \right|_{\mathcal{F}} \rightarrow 0 \quad a.s. \quad \Leftrightarrow \quad |B(n)|_{\mathcal{F}} := \left| \int \int f(x+y) - f(y) dP(y) \mu_n(x) \right|_{\mathcal{F}} \rightarrow 0 \quad a.s.$$

From this general Theorem we now can deduce the Winter-Yamato version of the Glivenko-Cantelli theorem for perturbed empirical measures.

Corollary 45 (Winter (1973); Yamato (1973)). Let $\mathcal{F} := \{1_{(-\infty, t]} : t \in \mathbb{R}\}$. Then

$$\left| \int f(d\tilde{\mathbb{P}} - dP) \right|_{\mathcal{F}} = \sup_{t \in \mathbb{R}} |\mathbb{F}_n(t) - F(t)| \rightarrow 0 \quad a.s.$$

Proof. Clearly \mathcal{F} is closed under translation. By continuity of P , $\forall \epsilon > 0 \exists \delta > 0$ such that $P\{[a, b]\} < \frac{1}{2}\epsilon$ whenever $b - a < \delta$. Find $n_0 := n_0(\delta)$ such that $\forall n \geq n_0, \mu_n\{x : |x| > \delta\} < \frac{1}{2}\epsilon$. Then $\forall n \geq n_0$,

$$|B(n)|_{\mathcal{F}} = \sup_y \int P\{(y, x + y)\} d\mu_n(x) \leq \int_{-\delta}^{\delta} \sup_y P\{(y, x + y)\} d\mu_n(x) + \frac{1}{2}\epsilon \leq \epsilon.$$

This result implies that if F has a density and \hat{f}_n is a "typical" non-parametric density estimator and $\hat{F}_n(x) := \int_{-\infty}^x \hat{f}_n(t) dt$, then \hat{F}_n converges uniformly to F almost surely.

Among others weak invariance principles for perturbed processes has been discussed in Yukich (1989, 1992), van der Vaart (1994) and Radulovic and Wegkamp (2000). We will need a result proved by van der Vaart (1994).

Theorem 46 (van der Vaart (1994)). Let \mathcal{F} be a Donsker class of measurable functions that is closed under translation. Let μ_n be random, signed measures of uniformly bounded variation that converge weakly in probability to the Dirac measure at zero. If

$$\sup_f P \left(\int (f(X + y) - f(X)) d\mu_n(y) \right)^2 \rightarrow_p 0, \quad (10.6)$$

and

$$\sup_f \sqrt{n} \left| P \int (f(X + y) - f(X)) d\mu_n(y) \right| \rightarrow_p 0, \quad (10.7)$$

then $\sqrt{n}(\tilde{\mathbb{P}} - \mathbb{P}_n) = o_p(1)$ and the sequence $\sqrt{n}(\tilde{\mathbb{P}} - P)$ converges weakly in distribution in $l^\infty(\mathcal{F})$ to a tight Brownian bridge process. If the measures μ_n are non-random, then the two conditions are necessary.

Remark 47. The main improvements over Theorem 3.1 by Yukich (1992) are that the class \mathcal{F} is not assumed uniformly bounded, that the smoothing measures are allowed to be random and have negative mass. Elimination of the boundedness condition ensures that the theorem also applies to, for instance, moments of the smoothed empirical. Random μ_n would occur in the case of kernel smoothing with bandwidth b_n chosen dependent on X_1, \dots, X_n , for instance by cross-validation.

10.3 Smoothed empirical distribution function

Let X_1, \dots, X_n be i.i.d random sample having distribution function F . Similarly as in the Section 10.2 let H denote a smoothed version of the degenerate distribution function H_0 defined by $H_0(x) = 1$ for $x \geq 0, 0$ otherwise. Define $H(t) = \int_{u \leq t} K(u) du$, where K is a compactly supported r -th order kernel which is commonly used in nonparametric density estimation. That is, for some integer $r \geq 2$ and constant $\kappa \neq 0$, K is a function satisfying

$$\int u^k K(u) du = \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{if } 1 \leq k \leq r - 1, \\ \kappa, & \text{if } k = r \end{cases} \quad (10.8)$$

For further use we also define $H_n(t) = H(t/b_n)$, where b_n is a bandwidth sequence, converging to zero as n grows to infinity. Now the smoothed empirical distribution function \hat{F}_n is defined by

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n H_n(t - X_i).$$

For small enough bandwidths b_n , the empirical cdf. \mathbb{F}_n defined in (10.1) and the smoothed cdf. \hat{F}_n are almost indistinguishable. Now let $f = F'$ denote the first derivative of F , where defined.

Lemma 48. *Assume that K satisfies (10.8), and is bounded and compactly supported; that $f^{(r-1)}$ is bounded; $b_n \rightarrow 0$ and*

$$\sup_t \sqrt{n} |\mathbb{E} \hat{F}_n(t) - F(t)| \rightarrow_p 0, \quad (10.9)$$

then

$$\sqrt{n} \sup_t |\hat{F}_n(t) - F(t)| \rightarrow_p 0,$$

and in particular, the smoothed empirical process $\{\sqrt{n}(\hat{F}_n - F)(t), t \in \mathbb{R}\}$ converges weakly to a tight Brownian bridge in $D(\mathbb{R})$.

Proof. According to Theorem 46, we only have to check that

$$\sup_t \int \left[\int (1_{\{x+\epsilon \leq t\}} - 1_{\{x \leq t\}}) dH_n(\epsilon) \right]^2 dF(x) \rightarrow 0.$$

After applications of Jensen's inequality and Fubini's theorem, we can bound the term on the left in the preceding display by

$$\begin{aligned} & \sup_t \int (F(t) - F(t - b_n x)) dH(x) \\ &= \sup_t \int \left\{ f(t)(b_n x) + \frac{1}{2} f'(t)(b_n x)^2 + \dots + \frac{1}{r!} f^{(r-1)}(t)(b_n x)^r + o(b_n^r) \right\} K(x) dx \\ &= O(b_n^r), \end{aligned}$$

which tends to zero as $b_n \rightarrow 0$.

Remark 49. *Considering weak convergence of empirical copula processes Fermanian et al. (2004), p.11. proved Lemma 48 on conditions that F is Lipschitz, $b_n \rightarrow \infty$ and (10.9) holds. We do not see reason to assume that F is Lipschitz.*

The assumption on the bias term (10.9) in the statement of the preceding lemma can be handled by means of some smoothness assumptions on F and regularity of K and b_n :

Lemma 50. *Assume that K as in (10.8), $f^{(r-1)}$ is bounded, $\lim_{n \rightarrow \infty} \sqrt{n} b_n^{r-1} = 0$. Then we have*

$$\sup_t \sqrt{n} |\mathbb{E} \hat{F}_n(t) - F(t)| \rightarrow_p 0.$$

Proof. The result follows readily after a Taylor expansion.

List of symbols

\rightarrow_p	Convergence in probability, page 13
\rightarrow_d	Convergence in distribution, page 34
<i>a.s.</i>	Almost sure convergence, page 18
$O_p(\cdot)$	Stochastic order symbol, page 35
$o_p(\cdot)$	Stochastic order symbol, page 36
$[t]$	Greatest integer less than or equal to t , page 24
$\langle t \rangle$	Least integer greater than or equal to t , page 24
\sim	Distributed as, page 5
$a \wedge b$	$\min(a, b)$, page 68
\mathcal{U}	Class of structural relationship models, page 12
\mathcal{F}	Set of distribution functions, page 11
\mathcal{H}	Set of structural parameter, page 12
$d(F_1, F_2)$	Mallows distance metric, page 12
$\phi_1, \phi_2, \phi_1^-, \phi_2^-$	Functions for structural relationships, page 12
$B_1^n(t), B_2^m(t)$	Sequences of Brownian bridges, page 19
$l^\infty(\mathcal{F})$	Space of bounded functions with $f \in \mathcal{F}$, page 67
K_1, K_2	Kernels, page 48
b_1, b_2	Smoothing parameters, page 48
\mathcal{T}	Open interval, page 31
δ_x	Dirac measure, page 67
$f^{(m)}$	m -th derivative of f , page 49
$1_{\{cond\}}$	Indicator for some expression $\{cond\}$, page 45

Bibliography

- Aaberge, R., Laake, K. and Laake, P. (1985). On the problem of measuring the distance between distribution functions: analysis of hospitalization versus mortality. *Biometrics*, **41**, 515-523.
- Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.*, **12**, 1041-1067.
- Aly, E.-E. A. A. (1986a). Quantile-quantile plots under random censorship. *J. Statist. Plann. Inference*, **15**, 123-128.
- Aly, E.-E. A. A. (1986b). Strong approximations of the Q-Q process. *J. Multivariate Anal.*, **20**, 114-128.
- Beirlant, J. and Deheuvels P. (1990). On the approximation of P-P and Q-Q plot processes by Brownian bridges. *Statist. Probab. Lett.*, **9**, 241-251.
- Borovkov, A. A. and Sycheva, N. M. (1968). On asymptotically optimal non-parametric criteria. *Theory Prob. Applic.*, **13**, 359-393.
- Brunner E. and Puri, M. L. (1996). Nonparametric methods in design and analysis of experiments, In: S. Gosh and C.R. Rao (Eds) *Handbook of Statistics*, **13**, 631-703.
- Burke, M. D., Csörgő, M., Csörgő, S. and Révész, P. (1979). Approximations of the empirical process when parameters are estimated. *Ann. Probab.*, **7**, 790-810.
- Canner, P. L. (1975). A simulation study of one- and two-sample Kolmogorov-Smirnov statistics with a particular weight function. *J. Amer. Statist. Assoc.*, **70**, 209-211.
- Cantelli, F. P. (1933). Sulla determinazione empirica delle leggi di probabilità. *Giornale dell'Istituto Italiano degli Attuari*, **4**, 421-424.
- Chen, S. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.*, **21**, 1166-1181.
- Claeskens, G., Jing, B.-Y., Peng, L. and Zhou, W. (2003). Empirical likelihood confidence regions for comparison distributions and ROC curves. *Canad. J. Statist.*, **31**, 173-190
- Com-Nougue, C., Rodary, C. and Patte, C. (1993). How to establish equivalence when data are censored: a randomized trial of treatments for B non-Hodgkin lymphoma. *Stat. Med.*, **12**, 1353-1364.
- Compagnone, D. and Denker, M. (1996). Nonparametric tests for scale and location. *J. Nonpar. Stat.*, **7**, 123-154.
- Csörgő, M. (1983). *Quantile Processes with Statistical Applications*. SIAM, Philadelphia.
- Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York.
- Czado, C. and Munk, A. (1998). Assessing the similarity of distributions - finite sample performance of the empirical Mallows distance. *J. Stat. Comput. Simul.*, **60**, 319-346.
- Czado, C. and Munk, A. (2001). Bootstrap methods for the nonparametric assessment of pop-

- ulation bioequivalence and similarity of distributions. *J. Stat. Comput. Simul.*, **68**, 243-280.
- Deheuvels, P. and Einmahl J. H. J. (1992). Approximations and two-sample tests based on P-P and Q-Q plots of the Kaplan-Meier estimators of lifetime distributions. *J. Multivariate Anal.*, **43**, 200-217.
- DiCiccio, T. J., Hall, P. and Romano, J.P. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.*, **19**, 1053-1061.
- Doksum, K (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.*, **2**, 267-277.
- Doksum, K. (1977). Some graphical methods in statistics. A review and some extensions. *Statist. Neerlandica*, **31**, 53-68.
- Doksum, K. A. and Sievers, G. L. (1976) Plotting with confidence: graphical comparisons of two populations. *Biometrika*, **63**, 421-434.
- Donsker, M. D. (1952). Justification and extension of Doob's heuristic approach to the Komogorov-Smirnov theorems. *Ann. Math. Statistics*, **23**, 277-281.
- Dudley, R. G. (1994). The order of the remainder in derivatives of composition and inverse operators for p -variation norms. *Ann. Statist.*, **22**, 1-20.
- Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution and of the classical multinomial estimator. *Ann. Math. Statist.*, **27**, 642 - 669.
- Einmahl, J. H. J. and McKeague, I. W. (1999). Confidence tubes for multiple quantile plots via empirical likelihood. *Ann. Statist.*, **27**, 1348-1367.
- Feller W. (1948). On the Kolmogorov-Smirnov limit theorems for empirical distributions, *Ann. Math. Statist.*, **19**, 177-189.
- Fermanian, J.-D., Radulovic D. and Wegkamp M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, **10**, 847-860.
- Fisher, N.I. (1983). Graphical methods in nonparametric statistics: a review and annotated bibliography. *Internat. Statist. Rew.*, **51**, 25-58.
- Freitag, G. (2000). *Validierung von Modellen in der Überlebenszeitanalyse*, Ph.D. Dissertation, University of Bochum, Bochum.
- Freitag, G., Munk, A. and Vogt, M. (2003). Assessing structural relationships between distributions - a quantile process approach based on Mallow's distance. In: *Recent Advances and Trends in Nonparametric Statistics*. Ed.: Akritas, M. G., Politis, D. N., Amsterdam: Elsevier B. V., 123-137.
- Freitag, G. and Munk, A. (2005). On Hadamard differentiability in k -sample semiparametric models—with applications to the assessment of structural relationships. *J. Multivariate Anal.*, **94**, 123-158.
- Girling, A. J. (2000). Rank statistics expressible as integrals under P-P plots and receiver operating characteristic curves. *J. R. Stat. Soc. Ser. B* **62**, 367-382.
- Glivenko, V. (1933). Sulla determinazione empirica della leggi di probabilità. *Giornale dell'Istituto*

Italiano Degli Attuari, 4:92-99.

Gnanadesikan, R. (1977). *Methods for statistical data analysis of multivariate observations*. John Wiley & Sons, New York-London-Sydney. A Wiley Publication in Applied Statistics.

Gnanadesikan, R. and Wilk, M.B. (1968). Probability plotting methods for the analysis of data. *Biometrika*, **55**, 1-17.

Hall, P. and La Scala B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.*, **58**, 109-127.

Hall, P. and Owen A. B. (1993). Empirical likelihood confidence bands in density estimation. *J. Comput. Graph. Statist.*, **2**, 273-289.

Hettmansperger, T. P. (1991). *Statistical Inference Based on Ranks*. Krieger Publishing Company, Malabar-Florida.

Hjort N. L., McKeague I. W. and Van Keilegom I. (2004). Extending the Scope of Empirical Likelihood. *Technical Report*. Université Catholique de Louvain, Louvain-la-Neuve.

Holmgren E. C. (1995). The P-P plot as a method of comparing treatments effects. *J. Amer. Statist. Assoc.*, **90**, 360-365.

Hsieh, F. and Turnbull B. W. (1992). Non- and semi-parametric estimation of the receiver operating characteristic curve. Technical Report 1026, School of Operations Research, Cornell Univ.

Hsieh, F. and Turnbull B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann. Statist.*, **24**, 25-40.

Jing, B-Y. (1995). Two-sample empirical likelihood method. *Statist. Probab. Lett.*, **24**, 315-319.

Jing, B-Y. and Wang, Z. (2003). Smoothed empirical likelihood confidence intervals for the difference of quantiles. *Statist. Sinica*, **13**, 83-95.

Jing, B-Y. and Zhou, W. (2003). Smoothed empirical likelihood confidence intervals for the difference of quantiles. *Statist. Sinica*, **13**, 83-95.

Lehmann, E.L. (1953). The power of rank tests. *Ann. Math. Stat.*, **24**, 23-43.

Li, G. and Doss, H. (1993). Generalized Pearson-Fisher chi-square goodness-of-fit tests, with applications to models with life history data. *Ann. Statist.*, **21**, 772-797.

Li, G., Tiwari, R. C. and Wells, M.T. (1996). Quantile comparison functions in two-sample problems, with application to comparisons of diagnostic markers. *J. Amer. Statist. Assoc.*, **91**, 689-698.

Li, G. and Wang, Q.-H. (2003) Empirical likelihood regression analysis for right censored data. *Statist. Sinica*, **13**, 51-68.

Lorenz, M. O. (1905). Methods on measuring the concentration of wealth. *J. Amer. Statist. Assoc.*, **70**, 209-219.

Macskassy, S. A. and Provost, F. (2004). *Confidence Bands for ROC Curves: Methods and an Empirical Study*. In Proceedings of the First Workshop on ROC Analysis in AI (ROCAI-2004)

at ECAI-2004.

Macaskassy, S. A., Provost, F. and Rosset, S. (2005). *ROC Confidence Bands: An Empirical Evaluation*. To appear in Proceedings of the 22nd International Conference on Machine Learning (ICML 2005), Bonn, Germany, 7-11 August, 2005.

Mallows, C. L. (1972). A note on asymptotic joint normality. *Ann. Math. Stat.*, **43**, 508-515.

Manoukian, E. B. (1986). *Mathematical Nonparametric Statistics*. Gordon and Breach Science Publishers, New York.

McKeague, I. W. and Zhao, Y. (2002). Simultaneous confidence bands for ratios of survival functions via empirical likelihood. *Statist. Probab. Lett.*, **60**, 405-415.

Munk, A. (1996). Equivalence and interval testing for Lehmann's Alternative. *J. Amer. Statist. Assoc.*, **91**, 1187-1196.

Munk, A. and Czado, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **60**, 223-241.

Nair, V. N. (1981). Plots and tests for goodness of fit with randomly censored data. *Biometrika*, **68**, 99-103.

Nair, V. N. (1982). Q-Q plots with confidence bands for comparing several populations. *Scand. J. Statist.*, **9**, 193-200.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237-249.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18**, 90-120.

Owen, A. (1991). Empirical likelihood for linear models. *Ann. Statist.*, **19**, 1725-1747.

Owen, A. (2001). *Empirical likelihood*. Chapman & Hall, Boca Raton, FL.

Parzen, E. (1993). Change PP plot and continuous sample quantile function. *Comm. Statist. Theory Methods*, **22**, 3287-3304.

Qin, B.-Y. and Lawless J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.*, **22**, 300-325.

Qin, G. and Jing, B.-Y. (2001a). Empirical likelihood for censored linear regression. *Scand. J. Statist.*, **28**, 661-673.

Qin, G. and Jing, B.-Y. (2001b). Censored partial linear models and empirical likelihood. *J. Multivariate Anal.*, **78**, 37-61.

Qin, G. and Tsao, M. (2003). Empirical likelihood inference for median regression models for censored survival data. *J. Multivariate Anal.*, **85**, 416-430.

Qin, J. (1994). Semi-empirical likelihood ratio confidence intervals for the difference of two sample means. *Ann. Inst. Statist. Math.*, **46**, 117-126.

Qin, Y. S. (1997). Semi-parametric likelihood ratio confidence intervals for various differences of two populations. *Statist. Probab. Lett.*, **33**, 135-143.

Qin, Y. S. and Zhao L. C. (2000). Empirical likelihood ratio confidence intervals for various

differences of two populations. *Systems Sci. Math. Sci.*, **13**, 23-30.

Radulovic D. and Wegkamp M. (2000). Weak convergence of smoothed empirical processes: beyond Donsker classes. *High dimensional probability, II (Seattle, WA, 1999)*, 89-105.

Sawitzki, G. (1994). Diagnostic plots for one-dimensional data. In P. Dirschedl and R. Ostermann (Eds.), *Papers collected on the Occasion of the 25th Conf. on Statistical Computing at Schloss Reisenburg*. Physica, Heidelberg.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.

Smirnov, N. V. (1939). On the deviation of the empirical distribution function (in Russian), *Mat. Sbornik*, **6**, 3-26.

Smirnov, N. V. (1948). Table for estimating the goodness of fit of empirical distributions, *Ann. Math. Statist.*, **19**, 279-281.

Switzer, P. (1976). Confidence procedures for two-sample problems. *Biometrika*, **63**, 13-25.

Taguri, M., Hiramatsu, M., Kittaka, T. and Wakimoto, K. (1976). Graphical representation of correlation analysis of ordered data by linked vector pattern. *J. Jap. Statist. Soc.*, **6**, 17-25.

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *J. Amer. Statist. Assoc.*, **70**, 865-871.

Tucker, H. G. (1967). *A graduate course in probability*. Academic Press, New York.

Tukey, P.A., and Tukey, J.W. (1981). Chapters 10-12 in *Interpreting Multivariate Data*, Ed. V. Barnett, 189-275.

van der Vaart, A. (1994). Weak convergence of smoothed empirical processes. *Scand. J. Statist.*, **4**, 501-504

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press. Wang, Q.-H. and Jing, B.-Y. (2001). Empirical likelihood for a class of functionals of survival distribution with censored data. *Ann. Inst. Statist. Math.*, **53**, 517-527.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, **9**, 60-62.

Winter, B. B. (1973). Strong uniform consistency of integrals of density estimators. *Canad. J. Statist.*, 163-173.

Wellek, S. (1993). A log-rank test for equivalence of two survivor functions. *Biometrics*, **49**, 877-881.

Yamato, H. (1973). Uniform convergence of an estimator of a distribution function. *Bull. Math. Statist.*, **15**, 69-78.

Yukich, J. E. (1989). A note on limit theorems for perturbed empirical processes. *Stoch. Procs. Appl.*, **35**, 163-173.

Yukich, J. E. (1992). Weak convergence of smoothed empirical processes. *Scand. J. Statist.* **19**, 271-279.