Statistica Sinica **17**(2007), 1165-1189

INFERENCE UNDER PEAKEDNESS RESTRICTIONS

Javier Rojo, José Batún-Cutz and Ramón Durazo-Arvizu

Rice University, CIMAT and Northwestern University

Abstract: A distribution function F is more peaked about a known point a than the distribution G is about the known point b if $F((x+a)^-) - F(-x+a) \geq G((x+b)^-) - G(-x+b)$ for every x. The statistical concept of dispersion plays an important role in the theory and practice of statistics. For example, in statistical genetics, the effect of a gene on a phenotype of interest can be ascertained by regressing the squared phenotypical differences on the proportion of identical by descent alleles shared by pairs of siblings (Haseman-Elston (1972)). This paper proposes estimators for the distribution functions F and/or G, when F is more "peaked" than G. The estimators are shown to be strongly uniformly consistent, their asymptotic distribution theory is discussed, and an asymptotic test for equality in peakedness is provided. The case of censored data is also considered. Data from various national and international studies are used to illustrate the new procedures.

Key words and phrases: Kaplan-Meier, peakedness ordering, Stochastic ordering, symmetry, weak convergence.

1. Introduction

Body Mass Index (BMI) is a commonly used measure of obesity, and studies have been conducted to correlate morbidity and mortality to BMI. The Department of Preventive Medicine and Epidemiology at Loyola University in Chicago, instituted an International Collaborative Study on Hypertension in Blacks (ICSHIB). This is an epidemiological multicenter, cross-sectional study of hypertension and associated risk factors in populations of African descent. Figure 1A shows the BMI density estimates for the populations of men and women from Nigeria and from Maywood, IL., and the corresponding empirical distributions functions are shown in Figure 1B. The density estimates suggest that a "shift" and a "spread" change have occurred for both African-American groups. The location shift is likely due to environmental factors while the change in "spread" can be attributed to an increase in genetic variability in the population.

Various concepts of spread, concentration, or dispersion have appeared in the literature. For example, Brown and Tukey (1946), Fraser (1957), Bickel and Lehmann (1979), Lehmann (1988), Doksum (1969) and Shaked (1980), define

F to be more dispersive than G if $F^{-1}(u) - F^{-1}(v) \geq G^{-1}(u) - G^{-1}(v)$ for every $u > v$. Shaked (1982), Bartoszewicz (1985a,b, 1986), Oja (1981) and Rojo and He (1991), among others, have discussed various characterizations and properties of the dispersive order, and Rojo (1995) has considered the problem of estimating F^{-1} or G^{-1}, or both, when F is more dispersed than G.

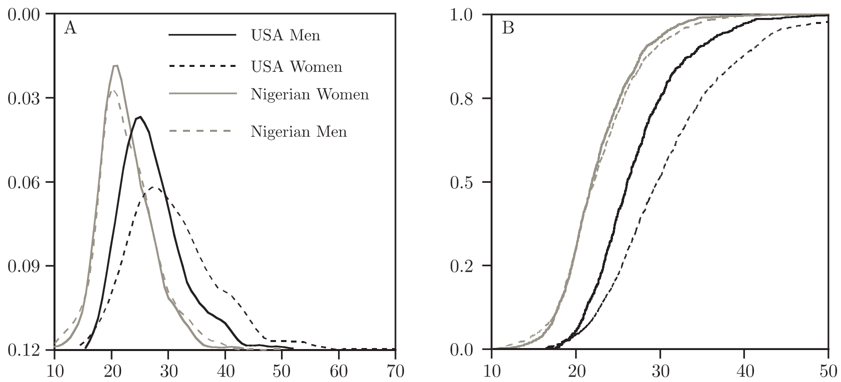

Figure 1. Density and cumulative distribution function estimates of BMI for four groups in the ICSHIB study.

Birnbaum (1948) took a different approach and defined a distribution function F to be more peaked about the point a than the distribution function G is about the point b if

$$F((x+a)^{-}) - F(-x+a) \geq G((x+b)^{-}) - G(-x+b) \tag{1.1}$$

for all $x \geq 0$, where, $h(x^{-}) = \lim_{\varepsilon \downarrow 0} h(x-\varepsilon)$. Proschan (1965), Karlin (1968), Bickel and Lehmann (1979), Shaked (1980, 1982) and Schweder (1982), among others, have considered properties of the ordering defined by (1.1), and have also studied connections with other orderings. When F and G are assumed symmetric, (1.1) can be seen to be equivalent to

$$F^{-1}(u) - F^{-1}\Big(\frac{1}{2}\Big) \geq (\leq)\ G^{-1}(u) - G^{-1}\Big(\frac{1}{2}\Big)$$

depending on whether $u \geq (\leq)1/2$. If a and b in (1.1) represent, respectively, the means of F and G, then (1.1) implies that the variance of F is smaller than the variance of G.

The concept of spread, peakedness, or dispersion permeates the theory and applications of statistics. For example, in the quantitative trait linkage analysis for sib-paired data literature, the Haseman-Elston model (1972), and its modifications (see e.g., Elston, Boxbaum and Olson (2000)), are used to test for linkage between a candidate locus and a specific phenotype. The model expresses the expected value of the squared phenotypic differences as a linear function of the proportion of alleles shared identical-by-descent (IBD) at the locus of interest. The method of Haseman and Elston (1972) is based on the regression model $E(X_i|\pi_i) = \alpha+\beta\pi_i$, where X_i is the squared sib-pair difference for the i^{th} sib-pair conditional on π_i, and π_i is the proportion of alleles shared identical by descent ($\pi_i = 0$, 1/2, or 1). Writing $Y_{1i} = \mu + g_{1i} + \varepsilon_{1i}$ and $Y_{2i} = \mu + g_{2i} + \varepsilon_{2i}$, where Y_{1i} and Y_{2i} represent, respectively, the phenotype values for siblings one and two, where μ is the population mean, and g_{ij} and ε_{ij} are the genetic and the residual effects, respectively, then

$$E(X_j|\pi_j) = \delta_\varepsilon^2 + 2(1 - \pi_j)\delta_g^2.$$

Here $\delta_\varepsilon^2 = E((\varepsilon_{1i} - \varepsilon_{2i})^2)$, and δ_g^2 represents the variance in the trait due to allelic variation at the locus of interest. Thus, when linkage exists, siblings sharing two alleles IBD at the locus of interest will tend to be more similar than siblings sharing one allele IBD, and siblings sharing one allele IBD will, in turn, be more similar than siblings sharing no alleles IBD. It is clear from this model that "similarity" is measured in terms of the spread of the distribution of the differences in the siblings' phenotypical measurements. In this paper, we use the order defined through (1.1), applied to the distribution of the differences of siblings' phenotypes, to compare the similarity of siblings with 0, 1 and 2 alleles IBD.

Since a and b will be assumed known, they are set equal to 0 without loss of generality. In the linkage example to be considered in Section 5, this assumption of known a and b is not as limiting as it may appear to be. Since it is customary to make the assumption that the siblings' phenotypes follow a bivariate normal distribution with marginals having the same mean, known or unknown, the difference of the phenotypes is always symmetric about zero. Even in the absence of the bivariate normal assumption, genetics models in common use, see e.g., Liu (1988), Table 15.7, yield a zero mean for the phenotypic differences.

The goals of this paper are to develop estimators for F and G that satisfy (1.1), and to delineate their asymptotic theory.

El Barmi and Rojo (1997) studied the nonparametric maximum likelihood estimators of F and G under the assumption that F and G are discrete distributions satisfying (1.1), and tests were given to test the hypothesis of homogeneity of F and G against the alternative that F and G satisfy (1.1). They, however, did not examine the asymptotic distribution theory of the estimators, and the case of censored data was not considered. The organization of this paper is as follows: Sections 2 and 3 consider the one-sample and the two-sample problems respectively, where strong uniform consistent estimators are constructed and their asymptotic theory is developed. Section 4 considers the case of right-censored data, and Section 5 illustrates the procedures with data from various national and international biomedical studies. In addition, a test for equality in peakedness is discussed in Sections 3 and 4. Finally, Section 6 discusses the results of computer simulations which compare the mean squared error of the new estimators with that of the empirical distribution function. It is worthwhile mentioning that, for positive stochastically ordered random variables, our estimators defined by (2.1), (3.1), (4.1) and (4.3), reduce to those discussed in Rojo (1995) and Rojo and Ma (1996). Thus, if the positive random variables X and Y with respective distribution functions F and G satisfy the constraint that $F(x) \geq G(x)$, for all $x > 0$, then F is more peaked about 0 than G is. In that case, estimator (2.1) reduces, for example in the one-sample problem case, to $F_n^*(x) = \max\{F_n(x), G(x)\}$, which is the estimator considered in Rojo (1995) and Rojo and Ma (1996). All technical details have been relegated to an appendix.

2. One-Sample Problem

Suppose that F and G satisfy (1.1) with $a = b = 0$. Let $X_1, \ldots, X_n$ be a random sample from the distribution F. To fix ideas, G is first considered a known continuous distribution. The case where G may be discontinuous and unknown will be considered later in the two-sample problem, but with less technical detail, as the technical arguments are similar. It is clear that the empirical distribution function F_n need not satisfy (1.1), and estimators which satisfy (1.1) may be needed.

At least two approaches have been proposed in the literature to address the problem of estimating a distribution function F subject to constraints. One general approach, the Nonparametric Maximum Likelihood Estimator (NPMLE), maximizes the empirical likelihood subject to the given constraints. (See, e.g., Oh (2004)). While this approach yields useful results in many cases, such an approach typically yields algorithmic procedures that are difficult to analyze. More importantly, however, in many cases the NPMLE fails minimal optimality criteria such as consistency. Several examples in the literature illustrate this point. There are additional reasons for preferring our procedures to the NPMLE. One is that, in contrast with the discrete case, the NPMLE is non-unique when dealing with continuous distributions. To see this, note that peakedness is equivalent to stochastic order of the absolute values. But, under the restriction of stochastic ordering, the NPMLE remains undefined to the left of the first order statistic, the restriction being that the NPMLE must be greater than or equal to the benchmark distribution to the left of the first order statistic but without specifying exactly how. A second additional reason is that, in the case of stochastic ordering, the NPMLE can be shown to be pointwise smaller than estimators similar to the one proposed here, and its mean squared error is larger. See, e.g., Rojo and Ma (1996). In particular, if peakedness is defined with respect to zero and the random variables are non-negative, then our estimators reduce to those discussed in Rojo and Ma (1996), and hence are better in terms of bias and mean squared error than the NPMLE. Finally, our Lemma 2.1 provides strong support for our procedures. That is, our estimator is the closest estimator to the empirical distribution function that satisfies the constraints of the problem. Thus, in particular, the NPMLE must have more absolute bias than our estimator (2.1).

A second approach consists of finding functionals of the empirical distribution function F_n that satisfy the constraints of the problem, and which are "closest" to the empirical distribution in some sense. For example, Kiefer and Wolfowitz (1976), Wang (1986, 1987a,b, 1988) and Rojo (1998) provide examples of this approach. More precisely, the idea is to find the majorant and/or minorant of F_n closest to F_n and such that the constraints of the problem are satisfied.

In this paper, we opt for the second approach. In the present context, however, attention can be restricted to estimators $\hat{F}_n$ satisfying (1.1), and are such that $\hat{F}_n \le F_n$ for $x < 0$ and $\hat{F}_n \ge F_n$ for $x \ge 0$. For suppose that $\hat{F}_n$ satisfies (1.1) and $\hat{F}_n > F_n$ for some $x < 0$ or $\hat{F}_n < F_n$ for some $x \ge 0$. Then the estimator $\hat{F}_n^*(x) = \min(\hat{F}_n(x), F_n(x))$ for $x < 0$ and $\hat{F}_n^*(x) = \max(\hat{F}_n(x), F_n(x))$ for $x \ge 0$, satisfies (1.1) and is closer to F_n than $\hat{F}_n$. Now, note that when F satisfies (1.1), then $F(x) \ge G(x) - G(-x) + F(-x)$ for $x > 0$ and $F(x) \le F(-x) - G(-x) + G(x)$ for $x < 0$. This motivates the following estimator,

$$\hat{F}_n(x) = \begin{cases} \min(F_n(x), 1 - G(-x) + G(x)), & x \le 0 \\ \max(F_n(x), \sup\limits_{0 \le y \le x} (G(y) - G(-y) + \hat{F}_n(-y))), & x > 0. \end{cases} \tag{2.1}$$

The *supremum* in (2.1) is needed to guarantee the monotonicity of $\hat{F}_n$. It can be shown that $\hat{F}_n$ is the closest distribution to F_n in the following sense.

Lemma 2.1. *Let f be right-continuous and g be continuous, both nondecreasing, with $0 \le f, g \le 1$. Define*

$$\hat{f}(x) = \begin{cases} \min(f(x), 1 - g(-x) + g(x)), & x \le 0 \\ \max(f(x), \sup\limits_{0 \le y \le x} (g(y) - g(-y) + \hat{f}(-y))), & x > 0. \end{cases} \tag{2.2}$$

Then $\hat{f}(x)$ *is the closest right-continuous nondecreasing function to* f, $0 \le \hat{f} \le 1$, *with*

(1) $\hat{f} \le f$ *for all* $x \le 0$;

(2) $\hat{f} \ge \max\{f(x), g(x) - g(-x) + h(-x))\}$ *for all* $x > 0$, *where* $h(x) = \min(f(x), 1 - g(-x) + g(x))$;

(3) $\hat{f}(x) - \hat{f}(-x) \ge g(x) - g(-x)$ *for all* $x \ge 0$;

(4) $f(x) - f(-x) \ge g(x) - g(-x)$ *for all* $x \ge 0$, *then* $\hat{f} = f$.

As a consequence of the lemma, $\hat{F}_n$ is the closest estimator to F_n satisfying properties (1)−(4). Also note that $\hat{F}_n(x) \to 1$ as $x \to \infty$ and $\hat{F}_n(x) \to 0$ as $x \to -\infty$. For an arbitrary function g, we write $\Delta g(x) = g(x) - g(-x^-), x \in R$.

It is not difficult to write an explicit computational expression for $\hat{F}_n(x)$. Note that if $x < 0$ and $X_{(i)} \le x < X_{(i+1)}$, then $\hat{F}_n(x) = \min\{i/n, 1 + \Delta G(x)\}$. If $x \ge 0$, and $X_{(i)} \le x < X_{(i+1)}$, let j be the unique integer such that $X_{(j)} \le -x < X_{(j+1)}$, and let k be the unique integer such that $X_{(k)} \le 0 < X_{(k+1)}$. Since $\sup_{0 \le y \le x}(\min(F_n(-y) + \Delta G(y), 1)) = \min(\sup_{0 \le y \le x}(F_n(-y) + \Delta G(y)), 1)$, then $\hat{F}_n(x) = \max\{i/n, \min\{1, \max\{j/n + \Delta G(x), \max_{j+1 \le l \le k}\{l/n - \Delta G(X_{(l)})\}\}\}\}$.

Before proceeding to the asymptotic theory of $\hat{F}_n$, note that following ideas similar to those that lead to Lemma 2.1, it is possible to define the estimator,

$$F_n^*(x) = \begin{cases} \min(F_n(x), \inf_{x \le y \le 0}(F_n^*(-y) + \Delta G(y))), & x < 0 \\ \max(F_n(x), \Delta G(x)), & x \ge 0. \end{cases} \tag{2.3}$$

It turns out that F_n^* satisfies (1.1), it is a distribution function, and it is the closest distribution to F_n with the properties that (1) $F_n^* \le F_n$ for $x \le 0$; (2) $F_n^* \ge F_n$ for $x > 0$; and (3) $F_n^*(x) \le \min(F_n(x), \max(F_n(-x) + \Delta G(x), 0))$ for all $x < 0$. The proof of the statement is similar to the proof of Lemma 2.1 and will not be repeated. $\hat{F}_n$ and F_n^* share similar properties. They are strongly uniformly consistent, converge weakly, under suitable conditions, to a Gaussian process, and both render F_n inadmissible with respect to a large class of loss functions. However, there is no clear choice between $\hat{F}_n$ and F_n^* in terms of their mean squared error properties: $\hat{F}_n$ dominates F_n^* on the left tail, while the reverse situation occurs on the right tail. This is not unlike what happens in several other situations. For example, when estimating an IFRA distribution, the estimator of Wang (1987a) behaves better than the isotonic regression estimator on the right tail, while the latter has smaller mean squared error than the former on the left tail of the distribution. We have opted for providing the asymptotic theory for $\hat{F}_n$ only, but the asymptotic theory for F_n^* is similar. Since any convex combination of $\hat{F}_n$ and F_n^* satisfies (1.1), our simulation work will also examine the mean squared error properties of $(\hat{F}_n + F_n^*)/2$. The strong uniform convergence of this average is immediately inherited from that of $\hat{F}_n$ and F_n^*. It will turn out that the average of $\hat{F}_n$ and F_n^* behaves uniformly better than $F_n, \hat{F}_n$ and F_n^* in terms of the mean squared error and, thus, it is a better choice than $\hat{F}_n$ or F_n^*.

Returning our attention to the asymptotic theory of $\hat{F}_n$, we now consider the strong uniform consistency of $\hat{F}_n$. The following lemma validates our estimator in the sense that it shows that it is closer to the true distribution than the empirical is.

Lemma 2.2. *Suppose that F and G satisfy* (1.1) *with G known. Then, for every n, $|\hat{F}_n(x) - F(x)| \le |F_n(x) - F(x)|$ for all $x \le 0$, and $|\hat{F}_n(x) - F(x)| \le \max\{|F_n(x) - F(x)|, \sup_{0\le y\le x} |F_n(-y) - F(-y)|\}$ for every $x > 0$.*

Thus, $\hat{F}_n$ is pointwise closer to F than F_n is for all $x \le 0$ and every n. Strong uniform convergence of $\hat{F}_n$ follows immediately from the previous lemma.

Theorem 2.3. *Suppose that F and G satisfy* (1.1) *with G known. Then, for every n, $\sup_x |\hat{F}_n(x) - F(x)| \le \sup_x |F_n(x) - F(x)|$.*

Since the distribution of $\sup_x |F_n(x) - F(x)|$ is independent of F for F continuous, while the distribution of $\sup_x |\hat{F}_n - F|$ is not, it follows that for loss functions of the form $L(F, G) = V(\sup_x |F - G|)$ where $V(\cdot)$ is nondecreasing on $(0, \infty)$, $\hat{F}_n$ dominates F_n in risk. Consider now the asymptotic distribution of $\hat{F}_n$. Let $H(y) = \Delta G(y) + F(-y)$.

Theorem 2.4. *Suppose that F and G satisfy* (1.1) *with F continuous and G known.*

(i) *If $x \le 0$, or $x > 0$ with $F(x) > \sup_{0\le y\le x} H(y)$,*

$$\sqrt{n}(\hat{F}_n(x) - F(x)) \xrightarrow{D} N(0, F(x)(1 - F(x))).$$

(ii) *If $x > 0$, with $F(x) = H(x)$, and F is strictly increasing on $(x - \eta, x)$ for some $\eta > 0$,*

$$\sqrt{n}(\hat{F}_n(x) - F(x)) \xrightarrow{D} \max(X, Y),$$

where (X, Y) follows a mean zero bivariate normal distribution with Variance$(X) = F(x)\bar{F}(x)$, Variance$(Y) = F(-x)\bar{F}(-x)$ and $\operatorname{Cov}(X, Y) = F(-x)\bar{F}(x)$, where $\bar{F} = 1 - F$.

Pointwise confidence intervals may be computed using Theorem 2.4. To develop confidence bands, the asymptotic behavior of the process $\{\sqrt{n}(\hat{F}_n(t) - F(t)), -\infty < t < \infty\}$ as $n \to \infty$ is needed. Hereafter, the process $\{B(t), -\infty <$

$t < \infty\}$ will denote the Brownian motion with $E(B(t)) = 0$ and $\mathrm{Cov}\,(B(s), B(t)) = \min(F(s), F(t)) - F(s)F(t)$.

Theorem 2.5. *Let* (1.1) *hold with $a = b = 0$, and suppose F is continuous and strictly increasing.*

(i) *If $\Delta F(x) > \Delta G(x)$ for all $x > 0$ then the process $\{\sqrt{n}(\hat{F}_n(x) - F(x)),\ -\infty < x < \infty\}$ converges weakly to the process $\{B(x), -\infty < x < \infty\}$.*

(ii) *If $\Delta F(x_0) = \Delta G(x_0)$ for some $x_0 \neq 0$, while $\Delta F(x) \not\equiv \Delta G(x)$, then the process $\{\sqrt{n}(\hat{F}_n(x) - F(x)), -\infty < x < \infty\}$ does not converge weakly.*

(iii) *If $\Delta F(x) = \Delta G(x)$ for all x and F is strictly increasing, then the process $\{\sqrt{n}(\hat{F}_n(x) - F(x)), -\infty < x < \infty\}$ converges weakly to the process $\{B(x)I_{\{x<0\}} + \max(B(x), B(-x))\ I_{\{x\geq 0\}}, -\infty < x < \infty\}$.*

Asymptotic confidence bands can be constructed for the distribution function F in the cases dealt with in (i) and (iii) of the theorem. In the first case, the distribution of $\sup_x |B(F(x))|$, as discussed in Billingsley (1999), provides the asymptotically exact confidence bands. In the case of (iii), the same approach will yield asymptotically conservative bands for F. This follows since

$$\sup_x |B(F(x))I_{\{x<0\}} + \max(B(F(x)), B(F(-x)))I_{\{x\geq 0\}}| \leq \sup_x |B(F(x))|, \quad (2.4)$$

and hence the level of confidence of the asymptotic coverage of the confidence bands obtained from the distribution of $\sup_x |B(F(x))|$ is a lower bound for the exact asymptotic level obtained from the distribution of the limiting process in (iii). Thus an asymptotic test for equality of F and G in peakedness can be obtained easily from (iii).

Corollary 2.6. *For testing that F and G are equal in peakednes, a conservative test may be obtained by rejecting the null hyphotesis when $\sup_x |\sqrt{n}(\hat{F}_n(x) - F(x))| > k_\alpha$ where k_α is the $1 - \alpha$ quantile of $\sup_x |B(F(x))|$.*

3. The Two Sample Problem

Here $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are independent random samples from F and G respectively, where F and G satisfy (1.1) with $a = b = 0$. As an estimator of G based on $Y_1, \ldots, Y_m$ we use the empirical distribution function $G_m(\cdot)$. To estimate F, define

$$\hat{F}_{n,m}(x) = \begin{cases} \min(F_n(x), 1 + \Delta G_m(x)), & x \leq 0 \\ \max(F_n(x), \sup\limits_{0\leq y\leq x} (\Delta G_m(y) + \hat{F}_{n,m}(-y^-))), & x > 0, \end{cases} \quad (3.1)$$

where F_n is the empirical distribution function based on $X_1, \dots, X_n$. Similar arguments to those used to demonstrate that (2.1) defines a cumulative distribution function show that (3.1) does as well. In addition, $\hat{F}_{n,m}(x^-) - \hat{F}_{n,m}(-x) \geq G_m(x^-) - G_m(-x)$. In case F_n and G_m satisfy (1.1) for all $x \geq 0$, it is desirable that $\hat{F}_{n,m}(x) = F_n(x)$ for all x. To see that in fact $\hat{F}_{n,m}(x) = F_n(x)$ for all x, suppose that $F_n(x^-) - F_n(-x) \geq G_m(x^-) - G_m(-x)$ for all $x > 0$. Then, for $y < 0$, $F_n(y) \leq F_n(-y^-) + \Delta G_m(y) \leq 1 + \Delta G_m(y)$. It follows that for $x < 0$, $\hat{F}_{n,m}(x) = F_n(x)$. On the other hand, when $x > 0$,

$$\begin{aligned}\hat{F}_{n,m}(x) &= \max\{F_n(x), \sup_{0 \leq y \leq x} (\Delta G_m(y) + \hat{F}_{n,m}(-y^-))\} \\ &= \max\{F_n(x), \sup_{0 \leq y \leq x} (\Delta G_m(y) + F_n(-y^-))\}.\end{aligned}$$

$$\begin{aligned}\text{Now, } F_n(y) &= \lim_{\varepsilon \downarrow 0} F_n(y+\varepsilon) \geq \lim_{\varepsilon \downarrow 0} F_n((y+\varepsilon)^-) \\ &\geq \lim_{\varepsilon \downarrow 0} (-\Delta G_m(-(y+\varepsilon)) + F_n(-(y+\varepsilon))) \\ &= \Delta G_m(y) + F_n(-y^-).\end{aligned}$$

Therefore, $F_n(x) = \sup_{0 \leq y \leq x} F_n(y) \geq \sup_{0 \leq y \leq x}(\Delta G_m(y) + F_n(-y^-))$, and hence, $\hat{F}_{n,m}(x) = F_n(x)$. The computational formula for $\hat{F}_{n,m}(x)$ is as follows.

For $x < 0$, suppose that $X_{(i)} \leq x < X_{(i+1)}$ and $Y_{(j)} \leq x < Y_{(j+1)}$, $Y_{(k)} < -x \leq Y_{(k+1)}$, $j \leq k$; then $\hat{F}_{n,m}(x) = \min\{i/n, 1 - [(k-j)/m]\}$. For $x > 0$, suppose that $X_{(i)} \leq x < X_{(i+1)}$. Let j be the unique integer such that $X_{(j)} < -x \leq X_{(j+1)}$ and let k be the unique integer such that $X_{(k)} < 0 \leq X_{(k+1)}$. Then, $\hat{F}_{n,m}(x) = \max\{i/n, \min\{1, \max\{j/n + \Delta G_m(x), \max_{j+1 \leq l \leq k}\{l/n - \Delta G_m(X_{(l)})\}\}\}\}$.
The following Theorem provides the strong uniform convergence of $\hat{F}_{n,m}$.

Theorem 3.1. *Let F and G be continuous distributions that satisfy* (1.1) *and let $\hat{F}_{n,m}$ be defined by* (3.1). *Then, with probability one,* $\sup_x |\hat{F}_{n,m}(x) - F(x)| \to 0$ *as $n, m \to \infty$.*

Theorem 3.2 provides the asymptotic distribution of $\hat{F}_{n,m}$, and is analogous to Theorem 2.4.

Theorem 3.2. *Suppose that F and G satisfy* (1.1), *with F and G continuous.*

(i) *If $x \leq 0$, or $x > 0$ with $F(x) > \sup_{0 \leq y \leq x} H(y)$, then as $n, m \to \infty$,*

$$\sqrt{n}(\hat{F}_{n,m}(x) - F(x)) \xrightarrow{D} N(0, F(x)(1 - F(x))).$$

(ii) *If* $x > 0$, *with* $F(x) = \sup_{0 \le y \le x} H(y)$, *then as* $n, m \to \infty$,

$$\sqrt{n}(\hat{F}_{n,m}(x) - F(x)) \xrightarrow{D} \max(X, Y),$$

where (X, Y) *is distributed as in* (ii) *of Theorem* 2.4.

The weak convergence of $\{\sqrt{n}(\hat{F}_{n,m}(x) - F(x)), -\infty < x < \infty\}$ is the content of the next theorem, and will serve as the basis for an asymptotic test for equality in peakedness.

Theorem 3.3. *Suppose* (1.1) *holds with* F *and* G *continuous, and* F *strictly increasing.*

(i) *If* $\Delta F(x) > \Delta G(x)$ *for all* $x > 0$, *then* $\{\sqrt{n}(\hat{F}_{n,m}(x) - F(x)), -\infty < x < \infty\}$ *converges weakly to the process* $\{B(x), -\infty < x < \infty\}$, *as* $n, m \to \infty$.

(ii) *If* $\Delta F(x_0) = \Delta G(x_0)$ *for some* $x_0 \neq 0$, *while* $\Delta F(x) \not\equiv \Delta G(x)$, *then the process* $\{\sqrt{n}(\hat{F}_{n,m}(x) - F(x)), -\infty < x < \infty\}$ *does not converge weakly.*

(iii) *If* $\Delta F(x) = \Delta G(x)$ *for all* $x \ge 0$, *then the process* $\{\sqrt{n}(\hat{F}_{n,m}(x) - F(x)), -\infty < x < \infty\}$ *converges weakly as* $n, m \to \infty$ *to the process* $\{S(x), -\infty < x < \infty\}$, *where* $S(x) = B(x)I_{\{x<0\}} + \max(B(x), B(-x))I_{\{x \ge 0\}}$.

The previous Theorem, parts (i) and (iii), can thus provide asymptotic confidence bands for $F(x)$ based on the estimator $\hat{F}_{n,m}$, and item (iii), as a consequence of (2.4), allows for a conservative Kolmogorov-Smirnov type test of the hypothesis that $\Delta F(x) = \Delta G(x)$ for all x. This last is stated as a corollary.

Corollary 3.4. *For testing that* $\Delta F(x) = \Delta G(x)$ *for all* $x > 0$, *a conservative test is given by rejecting when* $\sup_x |\sqrt{n}(\hat{F}_{n,m}(x) - F(x))| > k_\alpha$ *where* k_α *is as defined in Corollary* 2.6.

4. The Case of Censored Data

Let $X_1, \ldots, X_n$ be a random sample from the distribution F. As in Csörgő and Horváth (1983), an independent sample $Y_1, \ldots, Y_n$ with left-continuous distribution L censors the distribution F on the right. Thus, the available data consists of the pairs (Z_i, δ_i), $i = 1, 2, \ldots, n$, where Z_i is the minimum of X_i and Y_i, and δ_i is the indicator function of the event $\{X_i \le Y_i\}$. Let $\tilde{F}_n$ denote the Kaplan-Meier product-limit estimator of F.

For F^* a probability distribution, let $T_{F^*} = \inf\{t : F^*(t) = 1\}$, and assume henceforth that F and L do not have jumps in common. The proposed estimators are obtained by replacing the empirical cumulative distribution functions F_n and G_m by their Kaplan-Meier counterparts, and the asymptotic distribution theory for the estimators is based on that asymptotic theory. Let $T^* = \min(T_F, T_L)$. The strong uniform convergence of $\tilde{F}_n$ on $(-\infty, T^*]$ has been demonstrated by

Stute and Wang (1993). Precisely, $\tilde{F}_n$ is strongly uniformly consistent for F on $(-\infty, T^*]$ if and only if either $F\{T^*\} = 0$ or $F\{T^*\} > 0$ but $L(T^{*-}) < 1$. Note that F and L are not required to be continuous. Under stronger conditions on F and L, the weak convergence of $\{\sqrt{n}(\tilde{F}_n(t) - F(t)), -\infty < x < T\}$, for $T < T^*$, to a mean zero Gaussian process Z^* with covariance function $Cov(Z^*(s), Z^*(t)) = C(s)(1 - F(s))(1 - F(t))$, $s \le t$, where $C(s) = \int_{-\infty}^{s} dF(t)/[(1 - F(t))^2(1 - L(t))]$, $s < T_F$, was demonstrated by Breslow and Crowley (1974). Suppose that G is known, and that F and G satisfy (1.1) with $a = b = 0$. Define

$$\tilde{F}_n^*(x) = \min(\tilde{F}_n(x), 1 + \Delta G(x)) I_{x \le 0} + \max(\tilde{F}_n(x), \sup_{0 \le y \le x} (\Delta G(y) + \tilde{F}_n^*(-y))) I_{x>0}. \tag{4.1}$$

The strong uniform convergence of $\tilde{F}_n^*$ follows from Stute and Wang (1993) and the following.

Theorem 4.2. *When* (1.1) *holds,* $\sup_{x \le T^*} |\tilde{F}_n^*(x) - F(x)| \le \sup_{x \le T^*} |\tilde{F}_n(x) - F(x)|$ *for every* n.

If in addition F is continuous, it follows from Theorem 4.2 that $\tilde{F}_n^*$ is better than $\tilde{F}_n$ for loss functions of the form $L(\delta, F) = v(\sup_{x \le T^*} |\delta(x) - F(x)|)$, where v is nondecreasing.

Consider now the weak convergence of $\tilde{F}_n^*$. Both F and G are assumed to be continuous. Define $Z_n(x) = \sqrt{n}(\tilde{F}_n^*(x) - F(x))$ for $-\infty < x \le T$, where $T < T^*$ and $W(T) < 1$, with $1 - W(t) = (1 - F(t))(1 - L(t))$. Suppose that $\Delta F(x) > \Delta G(x)$ for every $x \ge 0$. Select, for arbitrary k, $t_1 < \cdots < t_k = T$, and consider the random vectors $(Z_n(t_1), \ldots, Z_n(t_k))$ and $(S_n^*(t_1), \ldots, S_n^*(t_k))$, where $S_n^*(x) = \sqrt{n}(\tilde{F}_n(x) - F(x))$. It follows from Földes and Rejtő (1981), see also Aly, Csörgő and Horváth (1985), that almost surely as $n \to \infty$,

$$\sup_{-\infty < x \le T} |\tilde{F}_n(x) - F(x)| = O\Big(n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}\Big). \tag{4.2}$$

As a consequence of (4.2), $(Z_n(t_1), \ldots, Z_n(t_k)) = (S_n^*(t_1), \ldots, S_n^*(t_k))$, eventually with probability one. Therefore, the finite-dimensional distributions of Z_n converge to those of the weak limit of the Kaplan-Meier process.

Arguments similar to those that led to tightness of $\{\sqrt{n}(\hat{F}_n(x) - F(x)), -\infty < x < \infty\}$ in Section 2, show that eventually, with probability one, $\sup_{t \le s \le t+\delta} |Z_n(s) - Z_n(t)| = \sup_{t \le s \le t+\delta} |S_n^*(s) - S_n^*(t)|$. Since $\{S_n^*(x), -\infty < x \le T\}$ is tight, weak convergence of Z_n follows.

Theorem 4.3. *Suppose F and L are continuous with $T < T^*$ and $W(T) < 1$. Let $\tilde{F}_n^*$ be defined by* (4.1). *Suppose that F is strictly increasing.*

(i) *If $\Delta F(x) > \Delta G(x)$ for all x, then $\{\sqrt{n}(\tilde{F}_n^*(x) - F(x)), -\infty < x \le T\}$ converges weakly to the weak limit Z^* of the Kaplan-Meier process.*

(ii) *If* $\Delta F(x_0) = \Delta G(x_0)$ *for some* $x_0 < T$ *with* $\Delta F(x) \not\equiv \Delta G(x)$*, then* $\{\sqrt{n}(\tilde{F}_n^*(x) - F(x)), -\infty < x \leq T\}$ *does not converge weakly.*

(iii) *If* $\Delta F(x) = \Delta G(x)$ *for all* $x \geq 0$*, then* $\{\sqrt{n}(\tilde{F}_n^*(x) - F(x)), -\infty < x \leq T\}$ *converges weakly to* $\{S(x), -\infty < x \leq T\}$*, where* $S(x) = Z^* I_{\{x<0\}} + \max\{Z^*(x),\ Z^*(-x)\} I_{\{x\geq 0\}}$.

In the two-sample problem let F, L, T, T^* and C be as in Theorem 4.3. Also, let $X_1^{'}, \ldots, X_m^{'}$ be a random sample from G which is censored on the right by a (left-continuous) distribution function $L^{'}$. Let $T^{**} = \min(T_G, T_{L^{'}})$, and let $T^{'} < T^{**}$. Let $C^{'}$ be defined by $C^{'}(s) = \int_{-\infty}^{s} dG(t)/[(1-G(t))^2(1-L^{'}(t))]$, $s < T^{'}$, and let $\{Z^{'}(t), -\infty < t < T^{'}\}$ be a mean zero Gaussian process with covariance function $Cov(Z^{'}(s), Z^{'}(t)) = C^{'}(s)(1-G(s))(1-G(t))$, $s \leq t$. Let $\tilde{G}_m$ be the Kaplan-Meier estimator of G based on $X_1^{'}, \ldots, X_m^{'}$. Define

$$\tilde{F}_{n,m}^*(x) = \min(\tilde{F}_n(x),\ 1 + \Delta\tilde{G}_m(x)) I_{\{x\leq 0\}} + \max(\tilde{F}_n(x), \sup_{0\leq y\leq x} (\Delta\tilde{G}_m(y) + \tilde{F}_{n,m}(-y^-))) I_{\{x>0\}}.$$

The proofs of Theorems 4.4, 4.5 are omitted as they follow from those of Theorems 3.2, 3.3.

Theorem 4.4. *With probability one,* $\sup_{t<T^*} |\tilde{F}_{n,m}^*(t) - F(t)| \to 0$ *as* $n, m \to \infty$.

Theorem 4.5. *Let the asumptions of Theorem* 4.3 *hold, and* $W_{n,m}(x) = \sqrt{n}(\tilde{F}_{n,m}^*(x) - F(x))$.

(i) *If* $\Delta F(x) > \Delta G(x)$ *for* $x > 0$*,* $\{W_{n,m}(x), -\infty < x \leq T\}$ *converges weakly, as* $m, n \to \infty$ *to the Gaussian process* Z^* *of Theorem* 4.3*, where* $T < T^{'} = \min(T^*, T^{**})$.

(ii) *If* $\Delta F(x_0) = \Delta G(x_0)$ *for some* x_0 *and* $\Delta F(x) \not\equiv \Delta G(x)$*,* $\{W_{n,m}(x), -\infty < x \leq T\}$ *does not converge weakly as* $n, m \to \infty$.

(iii) *If* $\Delta F(x) = \Delta G(x)$ *for* $x \geq 0$ *and* $T < T^{'} = \min(T^*, T^{**})$*,* $\{W_{n,m}(x), -\infty < x \leq T\}$ *converges weakly, as* $m, n \to \infty$*, to* $\{S(x), -\infty < x \leq T\}$*, defined in (iii) of Theorem* 4.3.

5. Genetics Applications

The statistical concept of dispersion has played an important role in statistical genetics as a way to measure genetic variability in a population of interest. Typically, this genetic variability is manifested through the variability in a specific phenotype. This section deals with a data set to illustrate the estimators.

Quantitative Trait Linkage Analysis.- The search for genes that control quantitative traits continues to be an important problem in the area of statistical genetics. Although there are various methods to locate these genes, depending on type and availability of genomic information, a common non-parametric method is the sib-pair method introduced by Haseman and Elston (1972). The essence of this method is that the variation of the phenotype value between sibs is related to the proportion of alleles shared identical by descent (IBD). The method of Haseman and Elston (1972) makes precise these observations, and is based on the regression model discussed in the introduction: $E(X_i|\pi_i) = \alpha + \beta\pi_i$.

When the null hypothesis of no linkage is rejected, this model makes precise the intuitive notion that the more genetically similar the sibs are, as measured by the proportion of alleles IBD, the more similar they will be in their phenotypical value. That is, siblings who share two alleles IBD have a more peaked distribution of phenotypic differences than siblings who share one allele IBD, and these, in turn, are more similar than siblings who share zero alleles IBD. Using this approach Fabsitz, Carmelli and Hewitt (1992) concluded that monozygotic twins are more similar then dizygotic twins in terms of their BMI. Mooser, Scheer, Marcovina, Wang, Guerra, Cohen and Hobbs (1997) examined the relationship between levels of Lipoprotein (a), or Lp(a), in African-American families and a highly polymorphic glycoprotein, apolipoprotein (a). High levels of Lp(a) lead to premature atherosclerosis, and therefore understanding the causes that lead to high levels of Lp(a) has been of utmost importance. Analyzing the plasma Lp(a) levels of 257 sibling pairs from 49 independent African-American families from the Dallas metroplex area, and using the methodology developed by Haseman and Elston (1972), it was concluded that the Lp(a) plasma levels were much more similar in the siblings who inherited two alleles IBD, than in the groups of siblings who inherited zero or one allele IBD.

Following a similar approach, and assuming that the conclusions reached by Mooser et al. (1997) apply as well to Caucasian subjects, we examined the same relationship utilizing a data set collected on 71 Caucasian families in the Dallas metroplex area that consists of 75 sib-pair phenotype differences, of which 38 observations are in the group of pairs of siblings with zero alleles IBD, 34 observations in the group with one allele IBD, and 13 observations in the group with two alleles IBD. Figure 2A shows the empirical distribution functions for each of the three IBD groups for the Caucasian population. It is clear from Figure 2A, that the condition (1.1) is not satisfied by the empiricals of the groups IBD0 and IBD1. Our methodology modifies the empirical distribution function for the group of one allele shared IBD with the restriction that it must be more peaked about 0 than the distribution for the group of zero alleles shared IBD. Figure 2B presents the empirical and the new distribution function for the groups IBD0 and IBD1, calculated using the estimators defined earlier.

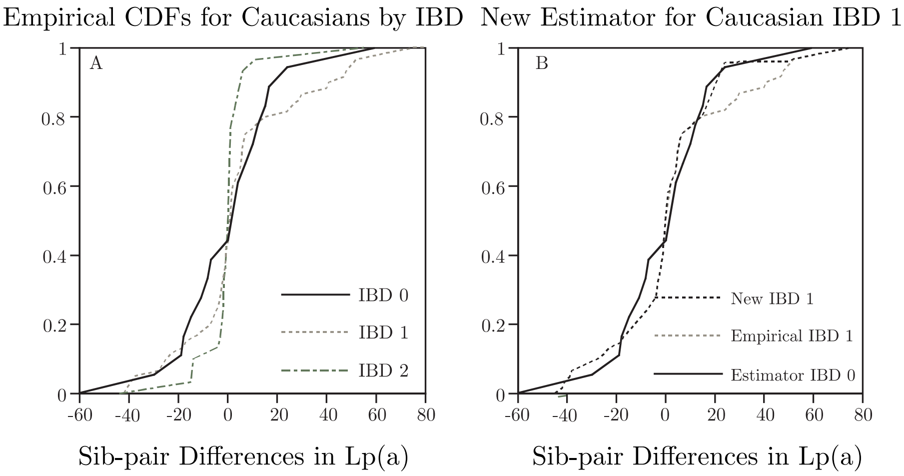

Figure 2. Estimators of the distribution of the Sib-pair differences: A without restrictions; B with peakedness restrictions.

6. Computer simulation

Computer simulations were performed to study the mean squared error (MSE) behavior of the proposed estimators. Experiments were performed for a wide class of underlying distributions. Gauss was used to run 10,000 repetitions of the experiments in double precision. In the one-sample problem, Theorem 2.1 states that the new estimator is closer to the true distribution than the empirical, in terms of the sup norm, and perhaps suggests that the new estimator is also closer, point-wise, to the true distribution than the empirical distribution function. This of course is true for $x < 0$ as demostrated by Theorem 2.1. Similarly, the estimator defined through (2.3) can be shown to be point-wise closer to F than the empirical distribution function for $x \geq 0$. As pointed out in the discussion leading to Lemma 2.2, the average $(\hat{F}_n + F_n^*)/2$ of the estimators defined by (2.1) and (2.3) behaves better in MSE than either $\hat{F}_n$ or F_n^*. Therefore, the simulation work presented in Figure 3 shows the ratios of the mean squared error of the empirical distribution function to the mean squared error of $(\hat{F}_n + F_n^*)/2$. Four cases are considered in Figure 3. In the one-sample problem, Figure 3 illustrates the gains in MSE of the estimator $(\hat{F}_n + F_n^*)/2$ over the empirical distribution function for the case that F is the standard normal distribution and G represents the normal distribution with mean zero and variance 1.21. The distributions used to generate the second graph in the top row are: F is the standard Cauchy distribution, G is Cauchy centered at zero and scale equal to 1.1, and the censoring distribution is the standard exponential. For the two-sample problem, the second row of Figure 3, the first case considers the logistic distribution with

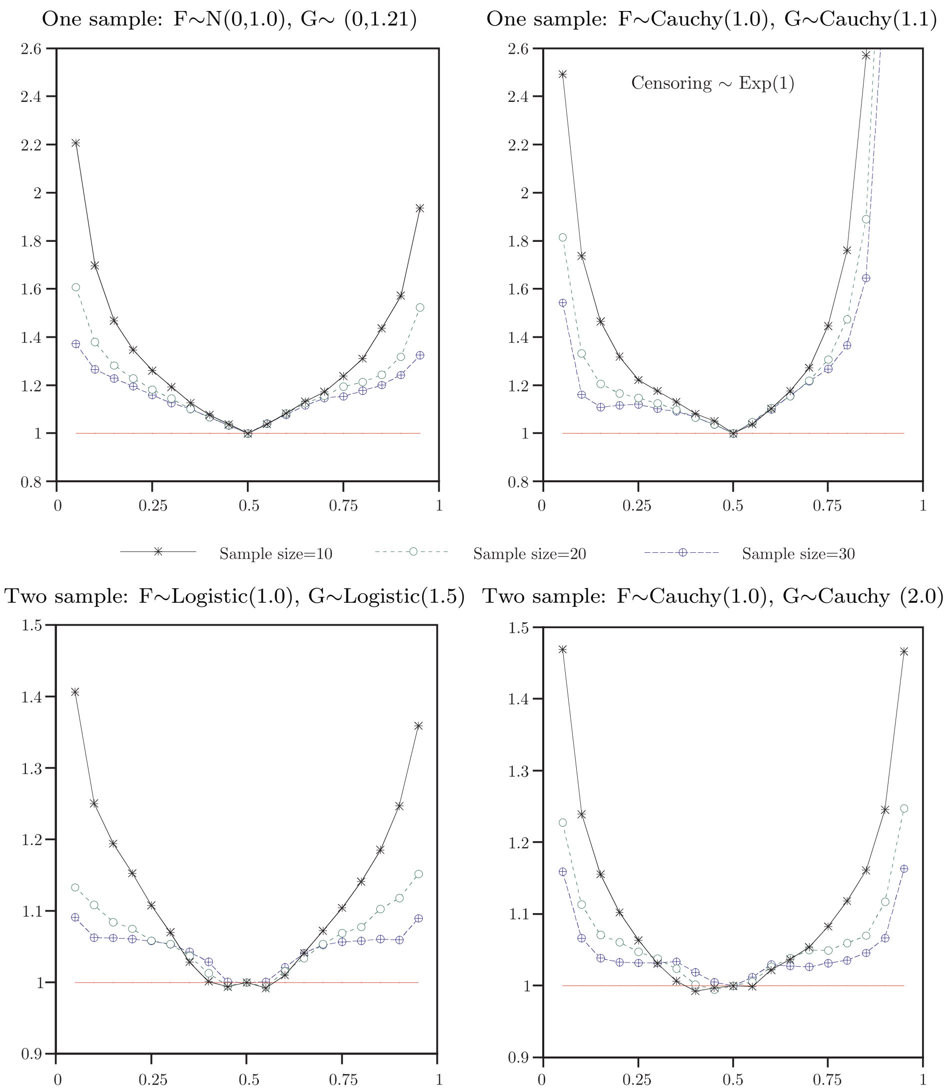

Figure 3. Ratios of Mean Squared Errors of the Empirical Distribution Function or the Kaplan-Meier Estimator to the Constrained Estimators. In the two sample problem, sample sizes are equal.

F the standard logistic, and G the logistic centered at zero and with scale parameter equal to 1.5; the second case takes F as a standard Cauchy distribution while G is a Cauchy centered at zero and scale parameter equal to 2. It is clear from Figure 3, that the new estimator behaves uniformly better than the empirical in the one-sample problem - even under right censoring - while in the two sample problem it behaves better than F_n except in a neighborhood of 0. The results are representative of the simulation results that were obtained and these results can be accessed at http://www.stat.rice.edu/ ~jrojo. The ratios of MSE are evaluated at 19 quantiles (0.05, 0.10, ..., 0.95) of the distribution function F.

7. Appendix

The following notation, and simple observation (5), will be used throughout the Appendix.

(1) $H(x) = \Delta G(x) + F(-x)$, and $H_n(x) = \Delta G(x) + \hat{F}_n(-x)$.
(2) For arbitrary probability distribution function F^*, $\bar{F}^* = 1 - F^*$.
(3) Let $\{A_n\}_{n=1}^{\infty}$ represent a sequence of events. Then, A_n occurs eventually with probability one - denoted as $ewp1$ - if $P(\liminf_{n\to\infty} A_n) = 1$.
(4) $e_n(F(x)) = F_n(x) - F(x)$ and $\hat{e}_n(F(x)) = \hat{F}_n(x) - F(x)$.
(5) Condition (1.1) implies that $\bar{H}(x) \geq \bar{F}(x)$ for all $x \geq 0$.

Proof of Lemma 2.1. By (2.2), $\hat{f}$ is nondecreasing with $\hat{f} \leq f$ for $x \leq 0$ and $\hat{f}(x) \geq \max\{f(x), \Delta g(x) + \hat{f}(-x)\}$ for $x > 0$. The right continuity of $\hat{f}$ for $x \leq 0$ follows from the right continuity of f and the continuity of g. To confirm the right continuity for $x > 0$, consider

$$\begin{aligned}\hat{f}(x+\epsilon) &= \max\{f(x+\epsilon), \sup_{0\leq y\leq x+\epsilon} (\Delta g(y) + \hat{f}(-y))\} \\ &\leq \max\{f(x+\epsilon), \max\{\sup_{0\leq y\leq x} (\Delta g(y) + \hat{f}(-y)), \Delta g(x+\epsilon) + \hat{f}(-x)\}\},\end{aligned}$$

and let $\epsilon \downarrow 0$. To show (3), consider, for $x > 0$ and $\epsilon > 0$,

$$\begin{aligned}\hat{f}(x-\epsilon) - \hat{f}(-x) &\geq \max(f(x-\epsilon), \Delta g(x-\epsilon) + \hat{f}(-x+\epsilon)) - \hat{f}(-x) \\ &\geq \Delta g(x-\epsilon) + \hat{f}(-x+\epsilon)) - \hat{f}(-x).\end{aligned}$$

Letting $\epsilon \downarrow 0$, (3) follows from the right continuity of $\hat{f}$ and the continuity of g.

To prove (4), when f satisfies (1.1), $f(-x) \leq f(x^-) - \Delta g(x) \leq 1 - \Delta g(x)$, and thus for $x < 0$, $\hat{f}(x) = f(x)$. Similarly, for $y > 0$, $f(y^-) \geq \Delta g(y) + f(-y)$ for all $y > 0$, and hence $f(x) \geq \sup_{0\leq y\leq x}(\Delta g(y) + \hat{f}(-y))$, since $\hat{f}(x) \leq f(x)$ for all $x \leq 0$. Then $\hat{f} = f$ if (1.1) holds.

It remains to show that $\hat{f}$ is the closest function to f satisfying (1)−(4). Let f^* be another function satisfying the assumptions and conclusions of the lemma. Using (3), for $x \leq 0$, $f^*(x) \leq f^*(-x) + \Delta g(x) \leq 1 + \Delta g(x)$, and by (1) $f^* \leq f$ for $x \leq 0$. Therefore $f^* \leq \hat{f}$ for $x \leq 0$. For $x > 0$, since $f^* \geq f$ and $f^*(x) = \sup_{0\leq y\leq x} f^*(y) \geq \sup_{0\leq y\leq x}(\Delta g(y) + h(-y))$, where $h(x) = \min(f(-x), 1 + \Delta g(x))$, it follows that $f^*(x) \geq \hat{f}(x)$ for all $x > 0$. Since $\hat{f}(0) = f(0)$, $f^*(x) \geq \hat{f}(x)$ for all x.

Proof of Lemma 2.2. Suppose first that $x \leq 0$, and define $A(x) = \{e_n(F(x)) \geq 0\}$. Then,

$$\begin{aligned}|\hat{e}_n(F(x))| &= |\min(e_n(F(x)), \bar{H}(-x))| \\ &= \min(e_n(F(x)), \bar{H}(-x)) I_{A(x)} + |e_n(F(x))| I_{A^c(x)} \\ &\leq |F_n(x) - F(x)|, \end{aligned} \tag{7.1}$$

where the identity follows since $\bar{H}(-x) \geq \bar{F}(-x) > 0$ for $x \leq 0$. It follows from (7.1) that $\hat{F}_n$ is closer to F than the empirical, for each $x \leq 0$ and each n, and therefore

$$\sup_{x \leq 0} |\hat{e}_n(F(x))| \leq \sup_{x \leq 0} |e_n(F(x))|. \tag{7.2}$$

Now suppose that $x \geq 0$ and observe that under (1.1), $F(y) \geq H(y)$ for all $y \geq 0$, and hence $F(x) \geq \sup_{0 \leq y \leq x} H(y)$ for all $x \geq 0$. Therefore, the result follows as a consequence of (7.2) and the following

$$\begin{aligned}|\hat{e}_n(F(x))| &= |\max(F_n(x), \sup_{0 \leq y \leq x} H_n(y)) - \max(F(x), \sup_{0 \leq y \leq x} H(y))| \\ &\leq \max(|e_n(F(x))|, \sup_{0 \leq y \leq x} |\hat{e}_n(F(-y)|) \\ &\leq \max(|e_n(F(x))|, \sup_{0 \leq y \leq x} |e_n(F(-y))|).\end{aligned}$$

Proof of Theorem 2.4. Suppose first that $x \leq 0$. Then $\hat{e}_n(F(x)) = \min(e_n(F(x)), \bar{H}(-x))$.
Since $\bar{H}(-x) \geq \bar{F}(-x) > 0$, *ewp*1, $\hat{e}_n(F(x)) = e_n(F(x))$. It follows that for $x \leq 0$, $\sqrt{n}\hat{e}_n(F(x)) \xrightarrow{D} N(0, F(x)\bar{F}(x))$.
If $x > 0$, and $F(x) - \sup_{0 \leq y \leq x} H(y) = a > 0$, then

$$\begin{aligned}\hat{e}_n(F(x)) &= \max(e_n(F(x)), \sup_{0 \leq y \leq x} (H_n(y) - F(x))) \\ &= \max(e_n(F(x)), \sup_{0 \leq y \leq x} (H(y) + \hat{e}_n(F(-y)) - F(x)).\end{aligned}$$

But, *ewp*1, $\sup_{0 \leq y \leq x}(H(y) + \hat{e}_n(F(-y)) - F(x) \leq -a + \sup_{0 \leq y \leq x}(e_n(F(-y))$. Therefore, *ewp*1, $\sqrt{n}(\hat{e}_n(F(x)) = \sqrt{n}(e_n(F(x))$, and hence $\sqrt{n}(\hat{e}_n(F(x)) \xrightarrow{D} N(0, F(x)\bar{F}(x))$. On the other hand, if $x > 0$ and $F(x) = H(x)$, so that $F(x) = \sup_{0 \leq y \leq x} H(y)$, then $\hat{e}_n(F(x)) = \max(e_n(F(x)), \sup_{0 \leq y \leq x}(H_n(y)) - F(x))$, and

$$\sup_{0 \leq y \leq x} (H_n(y)) - F(x) = \sup_{0 \leq y \leq x} \{H(y) + \min\{e_n(F(-y)), \bar{H}(y)\}\} - F(x). \tag{7.3}$$

Since $\bar{H}(y) \geq \bar{F}(y) > 0$, it follows that *ewp*1, $\min(e_n(F(-y)), \bar{H}(y)) = e_n(F(-y))$. Therefore, the right side of (7.3) equals, *ewp*1, $\sup_{0 \leq y \leq x}(H(y) - F(x) + e_n(F(-y))$.

If in addition F is strictly increasing on $(x-\eta, x)$ for some $\eta > 0$, then for $0 < \delta < \eta$,

$$\sup_{0\le y\le x}(H(y) - F(x) + e_n(F(-y)) = \max\{\sup_{0\le y\le x-\delta}(H(y) - F(x) + e_n(F(-y)), \sup_{x-\delta<y\le x}(H(y) - F(x) + e_n(F(-y))\}.$$

Since $F(x) > F(x-\delta) \ge \sup_{0\le y\le x-\delta} H(y)$ for every $\delta < \eta$, and since $\sup_y |e_n(F(y))| \to 0$ with probability one, it follows that letting $\delta \downarrow 0$, $ewp1$, $\sup_{0\le y\le x}(H(y) - F(x) + e_n(F(-y)) = e_n(F(-x))$. Therefore, for $x > 0$, with $F(x) = \sup_{0\le y\le x} H(y)$, $ewp1$,

$$\sqrt{n}(\hat{e}_n(F(x)) = \max(\sqrt{n}(e_n(F(x)), \sqrt{n}(e_n(F(-x))) \xrightarrow{D} \max(X, Y),$$

where (X, Y) follows a mean zero bivariate normal distribution with Variance(X) $= F(x)\bar{F}(x)$, Variance$(Y) = F(-x)\bar{F}(-x)$, and Cov$(X, Y) = F(-x)\bar{F}(x)$.

Proof of Theorem 2.5. Using arguments similar to those used to prove (i) in Theorem 2.4, it follows immediately that the finite-dimensional distributions of the process $\{\sqrt{n}(\hat{e}_n(F(t)), -\infty < t < \infty\}$ converge to the finite-dimensional distributions of the process $\{B(t), -\infty < t < \infty\}$. To prove tightness, and hence weak convergence of the process $\{\sqrt{n}(\hat{e}_n(F(x)), -\infty < x < \infty\}$, we follow Billingsley (1999), Theorem 15.5. Define $Z_n(x) = \sqrt{n}\hat{e}_n(F(x))$. Note that $P(|Z_n(0)| > a) = P(|ne_n(F(0))| > \sqrt{n}a) \le Var(nF_n(0))/na^2 = F(0)\bar{F}(0)/a^2$. Therefore, for every positive η there exists $a > 0$ such that $P(|Z_n(0)| > a) \le \eta$ for every $n \ge 1$. It remains to show that, for every positive ε and η, there exists δ, $0 < \delta < 1$, and an integer n_0 such that for every t and all $n \ge n_0$,

$$P\Big(\sup_{t\le s\le t+\delta}|Z_n(s) - Z_n(t)| \ge \varepsilon\Big) \le \eta\delta. \tag{7.4}$$

For that purpose, suppose first that $t < 0$ and select δ small enough that $t + \delta < 0$. Then $\sup_{t\le s\le t+\delta}|Z_n(s) - Z_n(t)| = \sup_{t\le s\le t+\delta}\sqrt{n}|\min(e_n(F(s)), \bar{H}(-s)) - \min(e_n(F(t)), \bar{H}(-t))|$. Since $\bar{H}(-s) > \bar{F}(-s) \ge \bar{F}(-t) > 0$, and $\bar{H}(-t) > \bar{F}(-t) > 0$, then, $ewp1$, $\sup_{t\le s\le t+\delta}|Z_n(s) - Z_n(t)| = \sup_{t\le s\le t+\delta}\sqrt{n}|e_n(F(s)) - e_n(F(t))|$. Since the process $\{\sqrt{n}(e_n(F(s)), -\infty < s < \infty\}$ is tight, (7.4) is satisfied for $t < 0$. That is, for every positive ε and η, there exits $\delta > 0$ and an integer n_0 such that for $t < 0$, $P(\sup_{t\le s\le t+\delta}|Z_n(s) - Z_n(t)| \ge \varepsilon) \le \eta\delta$. If $t > 0$, then

$$\sup_{t\le s\le t+\delta}|Z_n(s) - Z_n(t)| = \sup_{t\le s\le t+\delta}\sqrt{n}|\max(e_n(F(s)), \sup_{0\le y\le s}(H_n(y)) - F(s)) - \max(e_n(F(t)), \sup_{0\le y\le t}(H_n(y) - F(t))|. \tag{7.5}$$

Since $\sup_{0\le y\le s}(H_n(y))-F(s)=\sup_{0\le y\le s}(H(y)-F(s)+\min(e_n(F(-y)),\bar{H}(y)))$, and $\bar{H}(y)>\bar{F}(y)\ge\bar{F}(s)>0$ for all $y\le s$, and since $F(s)>\sup_{0\le y\le s}H(y)$, it follows from the strong uniform convergence of F_n to F, that, *ewp*1, the first max on the right side of (7.5) equals $e_n(F(s))$. Similar arguments yield the result that, *ewp*1, $\max(e_n(F(t)),\sup_{0\le y\le t}(H_n(y)-F(t)))=e_n(F(t))$. It follows that, *ewp*1,

$$\sup_{t\le s\le t+\delta}|Z_n(s)-Z_n(t)|=\sup_{t\le s\le t+\delta}\sqrt{n}|e_n(F(s))-e_n(F(t))|. \tag{7.6}$$

The tightness of the empirical process then implies (7.4) for $t>0$. The case $t=0$ follows immediately from the argument used for the case $t>0$, after recalling that $\hat{F}_n(0)=F_n(0)$. Hence $\{Z_n(t),-\infty<t<\infty\}$ is tight and therefore converges weakly.

(ii) Without loss of generality suppose that $x_0>0$ with $\Delta F(x_0)=\Delta G(x_0)$. Let $1\le k\le n$ be such that $X_{(k)}\le x_0<X_{(k+1)}$, where $X_{(i)}$ denotes the $i^{\underline{th}}$ order statistic and $X_{(n+1)}\equiv\infty$. Without loss of generality, suppose that there is $\gamma>0$ such that $\Delta F(x)>\Delta G(x)$ for every $x\in(x_0,x_0+\gamma)$. Note that $\sup_{x_0\le s\le x_0+\delta}|\sqrt{n}\hat{e}_n(F(s))-\sqrt{n}\hat{e}_n(F(x_0))|=\sup_{x_0\le s\le x_0+\delta}\sqrt{n}|(\max(F_n(s),\sup_{0\le y\le s}(H_n(y)))-F(s))-(\max(F_n(x_0),\sup_{0\le y\le x_0}(H_n(y))-F(x_0)))|$. For $s_n=x_0+\min(\delta,\gamma,X_{(k+1)}-x_0)/2$, *ewp*1, $\max(F_n(s_n),\sup_{0\le y\le s_n}(H_n(y)))-F(s_n)=e_n(F(s_n))$. Also, *ewp*1, $\max(e_n(F(x_0)),\sup_{0\le y\le x_0}(H_n(y)-F(x_0)))=\max(e_n(F(x_0)),e_n(F(-x_0)))$. It follows that, *ewp*1,

$$\begin{aligned}
&\sup_{x_0\le s\le x_0+\delta}|\sqrt{n}\hat{e}_n(F(s))-\sqrt{n}\hat{e}_n(F(x_0))|\\
&\quad\ge\sqrt{n}|e_n(F(s_n))-\max(e_n(F(x_0)),e_n(F(-x_0)))|\\
&\quad=\sqrt{n}|\max(e_n(F(x_0))-e_n(F(s_n)),e_n(F(-x_0))-e_n(F(s_n)))|\\
&\quad=\sqrt{n}|\max(F(s_n)-F(x_0),e_n(F(-x_0))-e_n(F(s_n)))|\\
&\quad\ge\sqrt{n}\max(0,e_n(F(-x_0))-e_n(F(s_n))).
\end{aligned}$$

Thus, $P(\sup_{x_0\le s\le x_0+\delta}|\sqrt{n}(\hat{e}_n(F(s))-\hat{e}_n(F(x_0))|\ge\varepsilon)\ge P(\sqrt{n}\max(0,e_n(F(-x_0))-e_n(F(s_n)))\ge\varepsilon)\to 1-\Phi(\varepsilon/\sigma)$, where σ^2 denotes the limiting variance of $\sqrt{n}(e_n(F(-x_0))+e_n(F(s_n)))$, which equals $F(-x_0)\bar{F}(-x_0)+F(x_0)\bar{F}(x_0)+2F(-x_0)\bar{F}(x_0)$. It follows that the process $\{Z_n(t),-\infty<t<\infty\}$ is not tight and hence cannot converge weakly.

(iii) If $\Delta F(x)=\Delta G(x)$ for all $x\ge 0$, then $\hat{F}_n$ and Z_n become, respectively,

$$\hat{F}_n(x)=\begin{cases}\min(F_n(x),1+\Delta F(x)), & x\le 0\\ \max(F_n(x),\sup\limits_{0\le y\le x}(\Delta F(y)+\hat{F}_n(-y))), & x>0,\ \text{and}\end{cases}$$

$$Z_n(x) = \begin{cases} \sqrt{n}\min(e_n(F(x)), \bar{F}(-x))), & x \le 0 \\ \sqrt{n}\max(e_n(F(x)), \sup\limits_{0\le y\le x}(\Delta F(y) + \hat{F}_n(-y) - F(x))), & x > 0. \end{cases}$$

Since $\bar{F}(-x) > 0$, *ewp*1, $Z_n(x) = \sqrt{n}e_n(F(x))$ for all $x \le 0$. Also, for any $\delta > 0$,

$$\sup_{0\le y\le x}(\Delta F(y) + \hat{F}_n(-y) - F(x)) = \max\Big\{\sup_{0\le y\le x-\delta}(\Delta F(y) + \hat{F}_n(-y) - F(x)),$$
$$\sup_{x-\delta\le y\le x}(\Delta F(y) + \hat{F}_n(-y) - F(x))\Big\}.$$

Since F_n is strongly uniformly consistent and F is strictly increasing, arguments similar to the one used to prove (ii) of Theorem 2.4 yield that, *ewp*1, $\sup_{0\le y\le x}(\Delta F(y) + \hat{F}_n(-y) - F(x)) = e_n(F(-x))$. It follows that the finite-dimensional distributions of $\{Z_n(x), -\infty < x < \infty\}$ converge to the finite-dimensional distributions of the process $\{B(x)I_{\{x<0\}} + \max(B(x), B(-x))I_{\{x\ge 0\}}, -\infty < x < \infty\}$. It remains to prove (7.4). For $t < 0$, we can proceed exactly as indicated just after (7.4). For $t \ge 0$, *ewp*1,

$$\sup_{t\le s\le t+\delta}|Z_n(s) - Z_n(t)| = \sup_{t\le s\le t+\delta}\sqrt{n}\,\Big|\max(e_n(F(s)), e_n(F(-s)))$$
$$- \max(e_n(F(t)), e_n(F(-t)))\Big|.$$

Also, $\{\sqrt{n}e_n(F(x))I_{\{x<0\}} + \sqrt{n}\max(e_n(F(x)), \ e_n(F(-x))I_{\{x\ge 0\}})\}$ converges weakly by the Continuous Mapping Theorem. Thus, for $\varepsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ and n_0 such that

$$P\{\sup_{t\le s\le t+\delta}\sqrt{n}|\max(e_n(F(s)), e_n(F(-s))) - \max(e_n(F(t)), e_n(F(-t)))| \ge \varepsilon\} \le \eta\delta$$

for all $n \ge n_0$. Thus, the process $\{Z_n(t), -\infty < t < \infty\}$ is tight and hence converges weakly to the process $\{B(x)I_{\{x<0\}} + \max(B(x), B(-x))I_{\{x\ge 0\}}, -\infty < x < \infty\}$.

Proof of Theorem 3.1. Consider first the case $x \le 0$. Then $|\hat{F}_{n,m}(x) - F(x)| = |\min(e_n(F(x)), \bar{F}(x) + \Delta G_m(x))|$. Now $\Delta G_m(x) + \bar{F}(x) = \Delta G_m(x)) + \Delta G(x) - \Delta G(x) + \bar{F}(x)$. Since $\Delta G(x) + \bar{F}(x) \ge \bar{F}(-x) > 0$, and $e_m(G(-x^-))$ and $e_m(G(x))$ go to zero with probability one, it follows that, as $n, m \to \infty$, *ewp*1,

$$\begin{aligned} &|\min(e_n(F(x)), \Delta G_m(x) + \bar{F}(x))| \\ &\quad = |e_n(F(x)|I_{\{e_n(F(x))<0\}} + \min(e_n(F(x)), \Delta G_m(x) + \bar{F}(x))I_{\{e_n(F(x))\ge 0\}} \\ &\quad \le |e_n(F(x))|. \end{aligned}$$

For $x>0$, write $|\hat{F}_{n,m}(x)-F(x)|=|\max(e_n(F(x)),\sup_{0\le y\le x}(\Delta G_m(y)+\hat{F}_{n,m}(-y^-)-F(x)))|$. When $F(x)=\sup_{0\le y\le x}H(y)$, using Lemma 1 in Rojo (1998) and a triangle inequality

$$\begin{aligned}|\hat{F}_{n,m}(x)-F(x)| &\le \max\{|e_n(F(x))|,|\sup_{0\le y\le x}(\Delta G_m(y)+\hat{F}_{n,m}(-y^-))-F(x)|\}\\ &=\max\{|e_n(F(x))|,|\sup_{0\le y\le x}(\Delta G_m(y)+\hat{F}_{n,m}(-y^-))-\sup_{0\le y\le x}H(y)|\}\\ &\le \max\{|e_n(F(x))|,\sup_{0\le y\le x}|e_m(G(y))|+\sup_{0\le y\le x}|e_m(G(-y^-)|\\ &\quad+\sup_{0\le y\le x}|\hat{F}_{n,m}(-y^-)-F(-y)|\}\text{ and,}\end{aligned}\tag{7.7}$$

$$\begin{aligned}&\sup_{0\le y\le x}|\hat{F}_{n,m}(-y^-)-F(-y)|\\ &\quad=\sup_{0\le y\le x}|\min(F_n(-y^-),\bar{F}(-y)-\Delta G_m(y))|\\ &\quad=\sup_{0\le y\le x}|\min\{e_n(F(-y^-)),\bar{H}(y)-e_m(G(y))+e_m(G(-y^-))\}|.\end{aligned}$$

Since $\bar{H}(y)\ge\bar{F}(y)\ge\bar{F}(x)>0$, and using the strong uniform convergence of G_m to G and of F_n to F, it then follows that, *ewp*1, as $n,m\to\infty$,

$$\min(e_n(F(-y^-)),\bar{H}(y)-e_m(G(y))+e_m(G(-y^-))=e_n(F(-y^-)).$$

Therefore, as $n,m\to\infty$, *ewp*1, $\sup_{0\le y\le x}|\hat{F}_{n,m}(-y^-)-F(-y)|=\sup_{0\le y\le x}|e_n(F(-y^-)|$. This, together with (7.7), imply that if $x>0$ and $F(x)=\sup_{0\le y\le x}H(y)$, then $\hat{F}_{n,m}(x)\to F(x)$ as $n,m\to\infty$. On the other hand, if $x>0$ and $F(x)>\sup_{0\le y\le x}H(y)$, consider $|\hat{F}_{n,m}(x)-F(x)|=|\max(e_n(F(x)),\sup_{0\le y\le x}(\Delta G_m(y)+\hat{F}_{n,m}(-y^-))-F(x))|$. It can be seen as before that, *ewp*1, $\hat{F}_{n,m}(-y^-)-F(-y)=e_n(F(-y^-))$ and hence, *ewp*1,

$$\max(e_n(F(x)),\sup_{0\le y\le x}(e_m(G(-y^-))+\hat{F}_{n,m}(-y)-F(x)))=e_n(F(x)).\tag{7.8}$$

It follows that $\hat{F}_{n,m}(x)$ converges to $F(x)$ with probability one for every x. Since F is continuous and $\hat{F}_{n,m}(x)$ is right-continuous, the strong uniform convergence then follows from a lemma in Chung (1974), page 133.

Proof of Theorem 3.3. (i) If $x\le 0$, then $\hat{F}_{n,m}(x)-F(x)=\min(e_n(F(x)),\bar{F}(x)+\Delta G_m(x))$. Now, $\bar{F}(x)+\Delta G_m(x)=\bar{H}(-x)+\Delta G_m(x)-\Delta G(x)$. Since $\bar{H}(-x)\ge\bar{F}(-x)>0$, and $e_m(G(-x^-)$ and $e_m(G(x))$ converge to zero with probability one, it follows that as $n,m\to\infty$, *ewp*1, $\hat{F}_{n,m}(x)-F(x)=e_n(F(x))$. Similarly, when $x>0$ and $F(x)>H(x)$ for every x, so that $F(x)>\sup_{0\le y\le x}H(y)$, then

$$\hat{F}_{n,m}(x)-F(x)=\max(e_n(F(x)),\sup_{0\le y\le x}(\Delta G_m(y)+\hat{F}_{n,m}(-y^-)-F(x))).\tag{7.9}$$

It follows from (7.8), that (7.9) equals $e_n(F(x))$ *ewp*1. These results immediately imply that under the assumption in (i), the finite-dimensional distributions of $\{\sqrt{n}(\hat{F}_{n,m}(x)-F(x)), -\infty < x < \infty\}$ converge to the finite-dimensional distributions of $\{B(x), -\infty < x < \infty\}$. It remains to prove tightness of the sequence $\{\sqrt{n}(\hat{F}_{m,n}(x) - F(x)), -\infty < x < \infty\}$. Define $Z_{n,m}(t) = \sqrt{n}(\hat{F}_{n,m}(t) - F(t))$. Suppose first that $t < 0$ and select $\delta > 0$, small enough so that $t + \delta < 0$. Then

$$\begin{aligned}
&\sup_{t\le s\le t+\delta} |Z_{n,m}(s) - Z_{n,m}(t)| \\
&\quad = \sup_{t\le s\le t+\delta} \sqrt{n}\Big| \min(e_n(F(s)), \bar{F}(s) - G_m(-s^-) + G_m(s)) \\
&\qquad - \min(e_n(F(t)), \bar{F}(t) - G_m(-t^-) + G_m(t))|.
\end{aligned}$$

Since $\bar{H}(-s) > \bar{F}(-s) \ge \bar{F}(-t) > 0$ and $\bar{H}(-t) > \bar{F}(-t) > 0$, by the strong uniform convergence of G_m, it follows that, *ewp*1,

$$\sup_{t\le s\le t+\delta} |Z_{n,m}(s) - Z_{n,m}(t)| = \sup_{t\le s\le t+\delta} |\sqrt{n}(e_n(F(s)) - \sqrt{n}(e_n(F(t))|.$$

Since $\{\sqrt{n}e_n(F(t)), -\infty < t < \infty\}$ is tight, it follows that, for every $\varepsilon > 0$, $\eta > 0$, there is $\delta > 0$ and an n_0 such that, for $t < 0$ and $n \ge n_0$,

$$P(\sup_{t\le s\le t+\delta} |Z_{n,m}(s) - Z_{m,n}(t)| \ge \varepsilon) \le \eta\delta. \tag{7.10}$$

If $t > 0$, similar arguments show that $\sup_{t\le s\le t+\delta} |Z_{n,m}(s) - Z_{n,m}(t)| = \sqrt{n}\sup_{t\le s\le t+\delta} |e_n(F(s)) - e_n(F(t)|$, *ewp*1. Thus, the weak convergence of $\{\sqrt{n}$ $(\hat{F}_{n,m}(t) - F(t)), -\infty < t < \infty\}$ follows.

(ii) Without loss of generality suppose that $x_0 > 0$ with $\Delta F(x_0) = \Delta G(x_0)$. Let $1 \le k \le n$ be such that $X_{(k)} \le x_0 < X_{(k+1)}$ where $X_{(i)}$ denotes the $i^{\underline{th}}$ order statistic. Without loss of generality, suppose there is $\gamma > 0$ such that $\Delta F(x) > \Delta G(x)$ for every $x \in (x_0, x_0 + \gamma)$. Define $s_n = x_0 + \min(\delta, \gamma, X_{(k+1)} - x_0)/2$. Then

$$\begin{aligned}
&P(\sup_{x_0\le s\le x_0+\delta} |\sqrt{n}(\hat{F}_{n,m}(s) - F(s)) - \sqrt{n}(\hat{F}_{n,m}(x_0) - F(x_0))| \ge \varepsilon) \\
&\quad \ge P(|\sqrt{n}(\hat{F}_{n,m}(s_n) - F(s_n)) - \sqrt{n}(\hat{F}_{n,m}(x_0) - F(x_0))| \ge \varepsilon).
\end{aligned}$$

Since $\hat{F}_{n,m}(x) \to \hat{F}_n(x)$ as $m \to \infty$, the result follows as in the proof of (ii) of Theorem 2.5.

(iii) The proof is analogous to that of (iii) in Theorem 2.5 once $m \to \infty$, since then $\hat{F}_{n,m}(x) \to \hat{F}_n(t)$ uniformly with probability one, reducing the problem to the one-sample case.

Proof of Theorem 4.3. The proof of (i) was given before stating the theorem, and (ii) and (iii) follow from arguments similar to those in proofs of (ii) and (iii) of Theorem 2.5.

Acknowledgements

This research was supported in part by NIH through the BioStatistical Laboratory, by NSF through the MRCE program Grant no. RII-8802973 and Grant no. DMS-0532346, by NASA grant NAG10-287, and Research Project No 39017-E of Conacyt-México. This work was conceived while the first author visited the Statistics Department at UC Berkeley, the Department of Preventive Medicine at the Loyola Medical Center in Chicago, and held a visiting position in the Statistics Department of SMU. Dr. Victor Pérez Abreu kindly facilitated the collaboration between J. Rojo and J. L. Batún. Drs. Rudy Guerra and Jonathan Cohen provided the data on paired siblings through the support of NIH grant R01 HL-53917.

References

Aly, E. E., Csörgő, M. and Horváth, L. (1985). Strong approximations of the quantile process of the product-limit estimator. *J. Multivariate Anal.* **16**, 185-210.

Bartoszewicz, J. (1985a). Moment inequalities for order statistics from ordered families of distributions. *Metrika* **32**, 383-389.

Bartoszewicz, J. (1985b). Dispersive ordering and monotone failure rate distributions. *Adv. Appl. Probab.* **17**, 472-47.

Bartoszewicz, J. (1986). Dispersive ordering and the total time on test transformation. *Statist. Probab. Lett.* **4**, 285-288.

Bickel, P. J. and Lehmann, E. L. (1979). Dispersive statistics for nonparametric models IV. Spread. In *Contrib. Statist.*, Jaroslaw Hajek Memorial Volume (Edited by J. Jureckova), 33-40. Reidel, Dordrecht.

Billingsley, P. (1999). *Convergence of Probability Measures.* 2nd edition. Wiley, New York.

Birnbaum, Z. W. (1948). On random variables with comparable peakedness. *Ann. Math. Statist.* **19**, 6-81.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimators under random censorship. *Ann. Statist.* **2**, 437-453.

Brown, G. and Tukey, J. W. (1946). Some distributions of sample means. *Ann. Math. Statist.* **7**, 1-12.

Chung, K. L. (1974). *A Course in Probability Theory.* Academic Press, New York.

Csörgő, S. and Horváth, L. (1983). The rate of strong uniform consistency for the product-limit estimator. *Z. Wahrsch. Verw. Gebiete* **62**, 411-426.

Doksum, K. (1969). Starshaped transformations and the power of rank tests. *Ann. Math. Statist.* **40**, 1167-1176.

El Barmi, H. and Rojo, J. (1997). Likelihood Ratio Test for Peakedness in Multinomial Populations. *J. Nonparametr. Stat.* **7**, 221-237.

Elston, R. C., Boxbaum, S. and Olson, M. (2000). Haseman and Elston Revisited. *Genetic Epidem.* **19**, 1-17.

Fabsitz, R. R., Carmelli, D. and Hewitt, J. K. (1992). Evidence for independent genetic influence on obesity in middle age. *Int. J. Obesity* **16**, 657-666.

Földes, A. and Rejtő, L. (1981). A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete* **56**, 75-86.

Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics.* Wiley, New York.

Haseman, J. K. and Elston, R. C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2**, 2-19.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 475-487.

Karlin, S. (1968). *Total Positivity.* Stanford University Press, CA.

Kiefer, J. and Wolfowitz, J. (1976). Asymptotic minimax estimation of concave and convex distributions funtions. *Z. Wahrsch. Verw. Gebiete* **34**, 73-85.

Lehmann, E. L. (1988). Comparing location experiments. *Ann. Statist.* **16**, 521-533.

Liu, B. H. (1988). *Statistical Genomics, Linkage, Mappings and QTL Analyisis.* CRC Press, New York.

Mooser, V., Scheer, D., Marcovina, S. M., Wang, J., Guerra, R., Cohen, J. and Hobbs, H. H. (1997). The Apo(a) Gene is the Major Determinant in Plasma Lp(a) Levels in African Americans. *Am. J. Hum. Genet.* **61**, 402-417.

Oh, M. (2004). Inference for peakedness ordering between two distributions. *J. Korean Statist. Soc.* **33**, 303-312.

Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scand. J. Statist.* **8**, 154-168.

Proschan, F. (1965). Peakedness of distributions of convex combinations. *Ann. Math. Statist.* **36**, 1703-1706.

Rojo, J. and He, G. Z. (1991). New properties and characterizations of the dispersive ordering. *Statist. Probab. Lett.* **11**, 365-372.

Rojo, J. (1995). Nonparametric Quantile Estimation Under Order Constraints. *J. Nonparametr. Stat.* **5**, 185-200.

Rojo, J. and Ma, Z. (1996). On the Estimation of Stochastically Ordered Survival Functions. *J. Statist. Comput. Simulation* **55**, 1-21.

Rojo, J. (1998). Estimation of the Quantile Function of an IFRA Distribution. *Scand. J. Statist.* **25**, 293-310.

Schweder, T. (1982). On the dispersion of mixtures. *Scand. J. Statist.* **9**, 165-169.

Shaked, M. (1980). On mixtures from exponential families. *J. Roy. Statist. Soc. Ser. B* **42**, 192-198.

Shaked, M. (1982). Dispersive orderings of distributions. *J. Appl. Probab.* **19**, 310-320.

Stute, W. and Wang, J. L. (1993). The strong law under random censorship. *Ann. Statist.* **21**, 1591-1607.

Wang, J. L. (1986). Asymptotically minimax estimators for distributions with increasing failure rate. *Ann. Statist.* **14**, 1113-1131.

Wang, J. L. (1987a). Estimators of a distribution function with increasing failure rate average. *J. Statist. Plann. Inference* **16**, 415-427.

Wang, J. L. (1987b). Estimating IFRA and NBU survival curves based on censored data. *J. Scand. J. Statist.* **14**, 199-210.

Wang, J. L. (1988). Optimal estimation of star-shaped distributions functions. *Statist. Decisions* **6**, 21-32.

Department of Statistics, Rice University, Houston, TX 77251-1892.

E-mail: jrojo@stat.rice.edu

Department of Statistics and Probability, CIMAT, Guanajuato, Gto., 36240 México.

E-mail: batun@cimat.mx

Northwestern University, Medical School, Center for Healthcare Studies, Div. General Internal Medicine, Chicago, IL 60611, U.S.A.

E-mail: rdurazo@lumc.edu

(Received December 2004; accepted January 2006)