Glivenko-Cantelli Properties of some Generalized Empirical DF's and Strong Convergence of Generalized L-Statistics *

R. Helmers[1], P. Janssen[2] and R. Serfling[3]

[1] Centre for Mathematics and Computer Science, Kruislaan 413, SJ-1098 Amsterdam, The Netherlands
[2] Limburgs Universitair Centrum, B-3610 Diepenbeek, Belgium
[3] Department of Mathematical Sciences, Johns Hopkins University, Baltimore, MD 21218, USA

Summary. We study a nonclassical form of empirical df H_n which is of U-statistic structure and extend to H_n the classical exponential probability inequalities and Glivenko-Cantelli convergence properties known for the usual empirical df. An important class of statistics is given by $T(H_n)$, where $T(\cdot)$ is a generalized form of L-functional. For such statistics we prove almost sure convergence using an approach which separates the functional-analytic and stochastic components of the problem and handles the latter component by application of Glivenko-Cantelli type properties. Classical results for U-statistics and L-statistics are obtained as special cases without addition of unnecessary restrictions. Many important new types of statistics of current interest are covered as well by our result.

1. Introduction

Let $X_1, \ldots, X_n$ be independent r.v.'s having common df F and let h be a measurable function from $\mathbb{R}^m$ to $\mathbb{R}$. Define the associated df $H_F(y) = P_F\{h(X_1, \ldots, X_m) \leq y\}$, $y \in \mathbb{R}$ and empirical df

$$H_n(y) = n_{(m)}^{-1} \sum 1\{h(X_{i_1}, \ldots, X_{i_m}) \leq y\}, \quad y \in \mathbb{R},$$

where the sum is taken over all $n_{(m)} = n(n-1)\ldots(n-m+1)$ m-tuples $(i_1, \ldots, i_m)$ of distinct elements from $\{1, \ldots, n\}$. For each fixed y, $H_n(y)$ is a U-statistic with mean $H_F(y)$; in the case $m = 1$ and $h(x) = x$, H_n reduces to the usual empirical df F_n.

Statistics of the form $T(H_n)$ were investigated by Serfling (1984) for $T(\cdot)$ an L-functional and by Janssen et al. (1984) for $T(\cdot)$ a more general type of L-functional. Certain Glivenko-Cantelli properties for H_n and asymptotic normality results for $T(H_n)$ were established.

The present paper provides further Glivenko-Cantelli results for the generalized empirical df H_n (Sect. 2) and strong convergence results for the "generalized L-statistics" discussed above (Sects. 3 and 4). Results for the multi-sample case are indicated also.

Our key tool in establishing Glivenko-Cantelli results is an extension to H_n of the exponential probability inequality of Dvoretzky et al. (1956) for the Kolmogorov distance between F_n and F. Our strong convergence results generalize those of Hoeffding (1961) for U-statistics and of van Zwet (1980) for L-statistics, without adding unnecessary restrictions.

In our treatment of the strong convergence problem for generalized L-statistics, we separate the functional-analytic and stochastic components of the problem. As a result, one can obtain results for statistics $T(H_n)$ by substituting H_n for G_n into basic convergence theorems proved for the functional $T(\cdot)$ defined on deterministic sequences $\{G_n\}$. This permits flexibility and easier generalization in the choice of (nonclassical) empirical df to be used and illuminates the fundamental issues involved in the convergence problem.

It should be noted that results on the behavior of H_n typically apply also to the empirical df H_{F_n} given by putting F_n for F in the definition of H_F. For each fixed y, $H_{F_n}(y)$ is a von Mises statistic.

The closeness of H_n and H_{F_n} may be evaluated through the relations

(i) $$n^m(H_n - H_{F_n}) = (n^m - n_{(m)})(H_n - \tilde{H}_n),$$

where $\tilde{H}_n(y)$ is the average of terms $1\{h(X_{i_1}, \ldots, X_{i_m}) \leq y\}$ with at least one equality $i_a = i_b$, $a \neq b$, and

(ii) $$n^m - n_{(m)} = O(n^{m-1}).$$

We shall leave results for H_{F_n} implicit from those stated for H_n.

2. Glivenko-Cantelli Results

Here we establish results on the almost sure convergence to 0 of $\|H_n - H_F\|_\infty$, where $\|f\|_\infty$ denotes $\sup |f(x)|$. Our first result is an analogue of the exponential probability inequality of Dvoretzky, Kiefer and Wolfowitz (1956) for $\|F_n - F\|_\infty$. Their inequality, which serves as a lemma in our development, is the following.

Lemma 2.1. *There exists a finite constant C_0, not depending on F, such that*

(2.1) $$P\{\|F_n - F\|_\infty > d\} \leq C_0 \exp\{-2nd^2\}, \quad d > 0, \; n \geq 1.$$

We shall utilize Lemma 2.1 to establish a related inequality for the moment-generating function of $\|F_n - F\|_\infty$, which is evidently a novel result for the classical empirical df.

Lemma 2.2. *There exists a finite constant C_1, not depending on F, such that*

(2.2) $$E\{\exp s \|F_n - F\|_\infty\} \leq (1 + C_1 s n^{-1/2}) \exp\{s^2/8n\}, \quad s > 0, \; n \geq 1.$$

Proof. By a well-known identity and an application of Lemma 2.1, we have

$$(2.3) \qquad E\{\exp s\,\|F_n-F\|_\infty\} = \int_0^\infty P\{\exp(s\,\|F_n-F\|_\infty)>t\}\,dt$$

$$= 1 + \int_1^{e^s} P\{\|F_n-F\|_\infty>(\log t)/s\}\,dt$$

$$\leqq 1 + C_0 \int_1^{e^s} \exp\{-2n(\log t)^2/s^2\}\,dt,$$

where we have used the fact that $0 \leqq \|F_n-F\|_\infty \leqq 1$. The integral in (2.3) is easily evaluated as (here $N(\mu,\sigma^2)$ denotes a normal distribution with mean μ and variance σ^2)

$$\left(\frac{\pi}{2}\right)^{1/2} sn^{-1/2}\, P\{0\leqq N(s^2/4n,\,s^2/4n)\leqq s\}\,\exp\{s^2/8n\},$$

so that (2.2) follows with $C_1 = C_0\left(\dfrac{\pi}{2}\right)^{1/2}$. $\square$

We now extend Lemma 2.2 to the empirical *df* H_n.

Lemma 2.3. *For the constant* C_1 *in* (2.2), *we have*

$$(2.4)$$
$$E\{\exp s\,\|H_n-H_F\|_\infty\} \leqq (1+C_1\,s[n/m]^{-1/2})\exp\{s^2/8[n/m]\}, \qquad s>0,\ n\geqq m.$$

(Here $[\cdot]$ denotes greatest integer part.)

Proof. Utilizing a representation for *U*-statistics given by Hoeffding (1963) (or see Serfling (1980), p. 180), we have the representation

$$(2.5) \qquad\qquad H_n-H_F = \frac{1}{n!}\sum_{i=1}^{n!}(H_n^{(i)}-H_F),$$

where each $H_n^{(i)}$ is a classical empirical *df* based on $[n/m]$ terms $h(X_{i_1},\ldots,X_{i_m})$, $h(X_{i_{m+1}},\ldots,X_{i_{2m}}),\ldots$ corresponding to a particular permutation $(i_1,i_2,\ldots,i_n)$ of $(1,\ldots,n)$. By convexity of the exponential function and Lemma 2.2, we obtain (2.4). $\square$

Applying Lemma 2.3 in connection with the elementary relation

$$(2.6) \qquad P\{\|H_n-H_F\|_\infty>d\} \leqq e^{-sd}\,E\{e^{s\|H_n-H_F\|_\infty}\}, \qquad s>0,$$

and with $s=4[n/m]\,d$, we obtain

Theorem 2.1. *There exists a finite constant C_1, not depending on F, or h, such that*

(2.7)
$$P\{\|H_n - H_F\|_\infty > d\} \leq (1 + 4C_1 [n/m]^{1/2} d) \exp\{-2[n/m] d^2\}, \quad d > 0, \; n \geq m.$$

It should be noted that Theorem 2.1, specialized to the case $m = 1$, is not as sharp as Lemma 2.1. This is due to the indirect method of proof using moment-generating functions and perhaps also due to a lack of sharpness in the bound obtained in Lemma 2.2. Nevertheless, for practical purposes in typical applications, Theorem 2.1 in its present form is sufficiently powerful. For example, it immediately yields, via the Borel-Cantelli lemma, the following rate for the almost sure convergence of $\|H_n - H_F\|_\infty$ to 0.

Corollary 2.1. *There exists a finite constant C_m, not depending on F or otherwise on h, such that with probability 1*

(2.8)
$$\limsup_{n \to \infty} \left(\frac{n}{\log n}\right)^{1/2} \|H_n - H_F\|_\infty < C_m.$$

(For example, $C_m > (m/2)^{1/2}$ suffices.) In the case $m = 1$, it follows by the LIL for $\|F_n - F\|_\infty$ (due to Chung (1949); see also Serfling (1980), p. 62, for discussion) that the limsup in (2.8) remains finite with $(\log n)$ replaced by $(\log\log n)$. Such refinements of (2.8) involve tools more delicate than Lemma 2.1 and Theorem 2.1 and will not be pursued here.

Our next result treats the almost sure behavior of *weighted* discrepancies between H_n and H_F, extending the strengthened Glivenko-Cantelli theorem for F_n given by Wellner (1977).

Theorem 2.2. *Let q be a nonnegative continuous function in $[0, 1]$ which is nondecreasing in $[0, \delta]$ and nonincreasing in $[1 - \delta, 1]$, for some $\delta > 0$, and satisfies*
$$\int_0^1 [q(t)]^{-1} dt < \infty. \text{ Then with probability 1}$$

(2.9)
$$\|(H_n - H_F)/q \circ H_F\|_\infty \to 0, \quad n \to \infty.$$

Proof. We follow the technique of Wellner (1977). Let $\varepsilon > 0$ be given and choose $M < \delta$ such that $\int_0^M [q(t)]^{-1} dt < \varepsilon$ and $\int_{1-M}^1 [q(t)]^{-1} dt < \varepsilon$ and such that M and $1 - M$ are continuity points of H_F^{-1}. Put $\theta_0 = H_F^{-1}(M)$ and $\theta_1 = H_F^{-1}(1 - M)$. Now write

$$\|(H_n - H_F)/q \circ H_F\|_\infty \leq \sup_{y < \theta_0} \frac{H_n(y)}{q \circ H_F(y)} + \sup_{y < \theta_0} \frac{H_F(y)}{q \circ H_F(y)}$$

$$+ \sup_{\theta_0 \leq y \leq \theta_1} \left|\frac{H_n(y) - H_F(y)}{q \circ H_F(y)}\right| + \sup_{y > \theta_1} \frac{1 - H_n(y)}{q \circ H_F(y)} + \sup_{y > \theta_1} \frac{1 - H_F(y)}{q \circ H_F(y)}$$

$$=: A_1 + A_2 + A_3 + A_4 + A_5.$$

Now

$$A_3 \leqq (\sup_{\theta_0 \leqq y \leqq \theta_t} [q \circ H_F(y)]^{-1}) \|H_n - H_F\|_\infty \to 0 \, wp \, 1$$

by Theorem 2.1. And, since $y < \theta_0 \Rightarrow H_F(y) < M$, we have

$$A_2 \leqq \sup_{0 < t < M} \frac{t}{q(t)} \leqq \sup_{0 < t < M} \int_0^t [q(u)]^{-1} du = \int_0^M [q(u)]^{-1} du < \varepsilon.$$

Similarly, we have $A_5 < \varepsilon$. And $wp \, 1$

$$A_1 \leqq n_{(m)}^{-1} \sum \sup_{y < \theta_0} \frac{1\{h(X_{i_1}, \ldots, X_{i_m}) \leqq y\}}{q \circ H_F(y)}$$

$$\leqq n_{(m)}^{-1} \sum \frac{1\{h(X_{i_1}, \ldots, X_{i_m}) \leqq \theta_0\}}{q \circ H_F(h(X_{i_1}, \ldots, X_{i_m}))} \to \int_{-\infty}^{\theta_0} [q \circ H_F(y)]^{-1} dH_F(y) < \varepsilon,$$

using the SLLN for U-statistics (Hoeffding (1961); or see Serfling (1980)) and the fact that

$$h(X_{i_1}, \ldots, X_{i_m}) \leqq y < \theta_0 \Rightarrow h(X_{i_1}, \ldots, X_{i_m}) < H_F^{-1}(M)$$

$$\Rightarrow H_F(h(X_{i_1}, \ldots, X_{i_m})) < M < \delta.$$

Similarly, A_4 has an a.s. limit $< \varepsilon$.

Therefore, for every $\varepsilon > 0$, $\limsup_{n \to \infty} \|(H_n - H_F)/q \circ H_F\|_\infty < 5\varepsilon$. $\square$

As an application of Theorem 2.2, we have

Corollary 2.2. Let $p \geqq 1$. If $\int (H_F(1 - H_F))^{p(1-\varepsilon)} < \infty$ for some $0 < \varepsilon < 1$, then with probability 1

$$(2.10) \qquad\qquad \|H_n - H_F\|_p \to 0, \quad n \to \infty.$$

(Here $\|g\|_p$ denotes $(\int |g(x)|^p \, dx)^{1/p}$.)

Proof. We write

$$\|H_n - H_F\|_p \leqq \|(H_n - H_F)/q \circ H_F\|_\infty \cdot \|q \circ H_F\|_p$$

and apply Theorem 2.2 with $q(t) = (t(1-t))^{1-\varepsilon}$. $\square$

It is easily checked that the assumption of Corollary 2.2 is satisfied if $E|h|^{p^{-1}+\delta} < \infty$ for some $\delta > 0$ (cf. Lemma 2.2.1 of Helmers (1982)).

Let μ_n and μ_F denote the probability measures on the class **B** of Borel sets in $\mathbb{R}$, associated with the df's H_n and H_F. Following Gaenssler (1983), for any subclass **C** of **B**, we define the *empirical* **C**-*discrepancy*

$$(2.11) \qquad D_n(\mathbf{C}, \mu_F) := \sup_{C \in \mathbf{C}} |\mu_n(C) - \mu_F(C)|,$$

and we assume that this quantity is Borel-measurable. Let $\mathbf{X}_{(n)} = (X_{n,1}, \ldots, X_{n,n})$ denote the vector of order statistics of $X_1, \ldots, X_n$.

Theorem 2.3. $\{D_n(\mathbf{C}, \mu_F)\}_{n \geq m}$ *is a reverse submartingale with respect to the sequence of σ-fields* $\mathbf{A}_n = \sigma\{\mathbf{X}_{(n)}, X_{n+1}, X_{n+2}, \ldots\}$, *i.e., for* $m \leq k \leq n$, *with probability* 1

$$(2.12) \qquad E\{D_k(\mathbf{C}, \mu_F) \mid \mathbf{A}_n\} \geq D_n(\mathbf{C}, \mu_F).$$

Proof. Use the reverse martingale property (Serfling (1980), p. 180) of the U-statistic $\mu_n(C) - \mu_F(C)$, for each fixed $C \in \mathbf{C}$, to get

$$E\{D_k(\mathbf{C}, \mu_F) \mid \mathbf{A}_n\} \geq \sup_{C \in \mathbf{C}} |E\{(\mu_k(C) - \mu_F(C)) \mid \mathbf{A}_n\}| = D_n(\mathbf{C}, \mu_F). \qquad \square$$

In the case $m = 1$ this reduces to Lemma 5 of Gaenssler (1983).

We now examine extensions to the multi-sample case. Consider c independent collections of independent observations $\{X_1^{(1)}, \ldots, X_{n_1}^{(1)}\}, \ldots, \{X_1^{(c)}, \ldots, X_{n_c}^{(c)}\}$ taken from df's $F^{(1)}, \ldots, F^{(c)}$, respectively.

Let a kernel $h(x_1^{(1)}, \ldots, x_{m_1}^{(1)}; \ldots; x_1^{(c)}, \ldots, x_{m_c}^{(c)})$ mapping $\mathbb{R}^{m_1 + \cdots + m_c}$ to $\mathbb{R}$ be given, put $\underline{F} = (F^{(1)}, \ldots, F^{(c)})$, denote by H_F the df of $h(X_1^{(1)}, \ldots, X_{m_1}^{(1)}; \ldots; X_1^{(c)}, \ldots, X_{m_c}^{(c)})$, and define the relevant empirical df by

$$(2.13) \qquad H_{\underline{n}}(y) = \left[\prod_{j=1}^{c} (n_j)_{(m_j)}\right]^{-1} \sum \mathbf{1}\{h(X_{i_{11}}^{(1)}, \ldots, X_{i_{1m_1}}^{(1)}; \ldots; X_{i_{c1}}^{(c)}, \ldots, X_{i_{cm_c}}^{(c)}) \leq y\},$$

$y \in \mathbb{R}$, where $\underline{n} = (n_1, \ldots, n_c)$ and the sum is taken over all $(n_j)_{(m_j)} = n_j(n_j - 1) \ldots (n_j - m_j + 1)$ m_j-tuples $(i_{j1}, \ldots, i_{jm_j})$ of distinct elements from $\{1, \ldots, n_j\}$, $1 \leq j \leq c$. Finally, put $k_{\underline{n}} = \min\{[n_1/m_1], \ldots, [n_c/m_c]\}$.

It turns out that Theorem 2.1 has the following extension.

Theorem 2.1*. *There exists a finite constant* C_1^*, *not depending on F or h, such that*

$$(2.14) \qquad P\{\|H_{\underline{n}} - H_F\|_\infty > d\} \leq (1 + 4C_1^* k_{\underline{n}}^{1/2} d) \exp\{-2k_{\underline{n}} d^2\}, \qquad d > 0, \quad \underline{n} \geq \underline{m}.$$

Proof. Using an extension to generalized U-statistics by Serfling (1985a) of a representation given by Hoeffding (1963) for U-statistics, we have as an extension of (2.5) the representation

$$(2.15) \qquad H_{\underline{n}} - H_F = \left(1 \Big/ \left[\prod_{j=1}^{c} (n_j)!\right]\right) \sum_{i=1}^{\prod_{j=1}^{c} (n_j)!} (H_{\underline{n}}^{(i)} - H_F),$$

where each $H_{\underline{n}}^{(i)}$ is a classical empirical df based on $k_{\underline{n}}$ i.i.d. terms having df H_F. The remainder of the proof is analogous to that of Theorem 2.1. $\square$

However, getting an extension of Corollary 2.1 is somewhat more complicated when c may be > 1. For this purpose, following Serfling (1985a) we introduce the following constraint on the array $\{(n_1, \ldots, n_c)\}$.

Definition 2.1. An array $\{(n_1, \ldots, n_c)\} \in \mathbb{N}^c$ satisfies *Condition A* if

$$(2.16) \qquad \frac{\log \max(n_1, \ldots, n_c)}{\min(n_1, \ldots, n_c)} \to 0 \qquad \text{as } \min(n_1, \ldots, n_c) \to \infty.$$

(This is trivially satisfied in the case $c = 1$ and in general is not very restrictive.)

The notation "$\min(n_1, \ldots, n_c) \xrightarrow{(A)} \infty$" shall denote restriction under Condition A.

Corollary 2.1*. *There exists a finite constant $C_{\underline{m}}$, not depending on F or otherwise on h, such that with probability 1*

$$(2.17) \qquad \limsup_{\min(n_1, \ldots, n_c) \xrightarrow{(A)} \infty} k_{\underline{n}}^{1/2} (\log n_1 \ldots n_c)^{-1/2} \| H_{\underline{n}} - H_F \|_\infty < C_{\underline{m}}.$$

(Here $\underline{m} = (m_1, \ldots, m_c)$.)

The proof is a straightforward application of the Borel-Cantelli lemma.

For weighted discrepancies, the proof of Theorem 2.2 carries over to the present situation and yields.

Theorem 2.2*. *Under the assumptions of Theorem 1, we have with probability 1*

$$(2.18) \qquad \|(H_{\underline{n}} - H_F)/q \circ H_F\|_\infty \to 0, \qquad \min(n_1, \ldots, n_c) \xrightarrow{(A)} \infty.$$

To extend Theorem 2.3, we must restrict attention to *ordered* elements from the partially ordered array $\{\underline{n}\}$. We define $\mu_{\underline{n}}$, μ_F and $D_{\underline{n}}(C, \mu_F)$ in similar fashion as before, and we define $A_{\underline{n}}$ to be the σ-algebra generated by $A_{n_j}^{(j)}$, $i \leq j \leq c$, where, with $X_{(n_j)}^{(j)} = (X_{n_j, 1}^{(j)}, \ldots, X_{n_j, n_j}^{(j)})$ the vector of order statistics of $X_1^{(j)}, \ldots, X_{n_j}^{(j)}$,

$$A_{n_j}^{(j)} = \{X_{(n_j)}^{(j)}, X_{n_j+1}^{(j)}, X_{n_j+2}^{(j)}, \ldots\}.$$

Then it is known (see Sen (1977), p. 288) that for fixed $B \in \mathbf{B}$, the generalized U-statistic array $\{\mu_{\underline{n}}(B) - \mu_F(B)\}$ is a reverse martingale w.r.t. the σ-field $A_{\underline{n}}$, i.e., with probability 1, $E\{\mu_{\underline{k}}(B) - \mu_F(B) \mid A_{\underline{n}}\} = \mu_{\underline{n}}(B) - \mu_F(B)$, for $\underline{n} \geq \underline{k} \geq \underline{m}$. Hence we have

Theorem 2.3*. $\{D_{\underline{n}}(C, \mu_F)\}$ *is a reverse submartingale w.r.t. the σ-field $A_{\underline{n}}$, i.e., for $\underline{m} \leq \underline{k} \leq \underline{n}$, with probability 1*

$$(2.19) \qquad E\{D_{\underline{k}}(C, \mu_F) \mid A_{\underline{n}}\} \geq D_{\underline{n}}(C, \mu_F).$$

Remark 2.1. In this section we have presented certain Glivenko-Cantelli results having general interest and application. Some further a.s. results of a more specialized nature are developed in the subsequent sections of this paper; see Lemmas 3.2 and 3.3.

3. A Strong Law for Generalized L-Statistics

Let H_F and H_n be defined as previously in terms of a kernel $h: \mathbb{R}^m \to \mathbb{R}$ and $X_1, \ldots, X_n$ i.i.d. F, and let $W_{n,1} \leqq \ldots \leqq W_{n,n_{(m)}}$ denote the ordered values of $h(X_{i_1}, \ldots, X_{i_m})$ taken over the $n_{(m)}$ m-tuples $(i_1, \ldots, i_m)$ of distinct elements from $\{1, \ldots, n\}$. A wide class of parameters of F can be represented usefully as $T(H_F)$, where $T(\cdot)$ is a functional of the general form

$$(3.1) \qquad T(G) = \int_0^1 g \circ T_t(G) \, dK(t), \qquad G \,\partial\, df,$$

where $g: \mathbb{R} \to \mathbb{R}$ is Borel-measurable, K is a signed measure on $(0, 1)$, and for each t in the support of K, $T_t(\cdot)$ is a classical L-functional: $T_t(G)$ $= \int_0^1 G^{-1}(s) \, dM_t(s)$, where $G^{-1}(s) = \inf\{x: G(x) \geqq s\}$ and M_t is a signed measure on $(0, 1)$. This functional was introduced by Janssen et al. (1984), who established asymptotic normality results for the corresponding statistics $T(H_n)$ for estimation of $T(H_F)$.

Our purpose here is to establish strong convergence, and in fact we shall let $K(\cdot)$ in (3.1) depend on n, thus considering the functional

$$(3.2) \qquad T_n(G) = \int_0^1 g \circ T_t(G) \, dK_n(t).$$

We shall take the measures $dK(t)$, $dK_n(t)$ and $dM_t(s)$ to be the differentials of

$$(3.3) \qquad K(t) = \int_0^t J(u) \, du + \sum_{j=1}^d a_j \mathbf{1}\{t \leqq p_j\}, \qquad 0 < p_j < 1, \; j = 1, \ldots, d,$$

$$(3.4) \qquad K_n(t) = \int_0^t J_n(u) \, du + \sum_{j=1}^d a_j \mathbf{1}\{t \leqq p_j\}, \qquad 0 < p_j < 1, \; j = 1, \ldots, d,$$

and

$$(3.5) \qquad M_t(s) = \int_0^s J_t(u) \, du + \sum_{j=1}^{d_t} a_{tj} \mathbf{1}\{s \leqq p_{tj}\}, \qquad 0 < p_{tj} < 1, \; j = 1, \ldots, d_t,$$

thus permitting both smooth and discrete weighting of quantiles $G^{-1}(s)$ in $T_t(G)$ and of functionals $g \circ T_t(G)$ in $T(\cdot)$. This covers essentially all cases of interest. Note that in the special case of (3.2) corresponding to $T_t(G) = G^{-1}(t)$, the statistic $T_n(H_n)$ may be expressed in the form

$$(3.6) \qquad \sum_{i=1}^{n_{(m)}} c_{ni} \, g(W_{n,i}).$$

Conversely, any statistic of form (3.6) may be represented as $T_n(H_n)$ with $T_i(G)$ $= G^{-1}(t)$ and some choice of $dK_n(t)$ (not necessarily unique) satisfying c_{ni}

$$= \int_{(i-1)/n_{(m)}}^{i/n_{(m)}} dK_n(t); \text{ e.g., take } (3.4) \text{ with } d=0 \text{ and } J_n(t)=n_{(m)}c_{ni} \text{ for } (i-1)/n_{(m)} < t$$

$\leq i/n_{(m)}$, $0 < t < 1$. The class of statistics given by (3.6) may be termed "linear combinations of functions of generalized order statistics." For the case $g(x) = x$ it gives the "generalized L-statistics" (which includes both U-statistics and classical L-statistics) introduced by Serfling (1984), who established asymptotic normality results. For the special case $m=1$, $h(x)=x$, (3.6) gives the class of "linear combinations of functions of order statistics", for which van Zwet (1980) proved strong convergence results under very tight conditions, confining attention to the case that $dK_n(t) = J_n(t) dt$. The present treatment extends van Zwet's by allowing limits $dK(t)$ to have a discrete component (considering $K(\cdot)$ as a limit of $K_n(\cdot)$), and considering arbitrary kernels h and functionals $T_i(G)$ other than $G^{-1}(t)$, and retains van Zwet's general scope of allowing the transformation $g(\cdot)$ to be an arbitrary Borel-measurable function. We establish a very general and far-reaching *deterministic* result (Theorem 3.1), which we apply in Corollaries 3.1 and 3.2 to two broad classes of statistics: Examples 3.1, treating the general class given by (3.6), and Examples 3.2, treating some spread estimators of Bickel and Lehmann (1979).

To obtain (under appropriate assumptions) that $wp1$

$$(3.7) \qquad\qquad T_n(H_n) - T_n(H_F) \to 0, \qquad n \to \infty,$$

and also that $T_n(H_F) \to T(H_F)$, we build upon the very insightful treatment of van Zwet (1980), but also separate the roles played by functional analysis and Glivenko-Cantelli theory. First we show that (3.7) holds *deterministically* with H_F replaced by an arbitrary df G and $\{H_n\}$ replaced by an arbitrary sequence $\{G_n\}$ converging weakly to G and satisfying other technical restrictions. This is the "functional analysis" part. Then we show that $\{H_n\}$ satisfies $wp1$ the conditions imposed on $\{G_n\}$. This is the "Glivenko-Cantelli" part. The idea of separating the functional-analytic and probabilistic components of the strong convergence problem for L-statistics was introduced by Wellner (1977), who obtained a general strong law for L-statistics as a corollary of extended Glivenko-Cantelli theorems which he developed for the empirical df F_n, such as we have extended in Sect. 2 to the empirical df's H_n and $H_{\underline{n}}$. However, the assumptions of his theorem are not quite sharp, imposing for example $(1+\varepsilon)$-moment conditions for the sample mean. The development by van Zwet (1980) leads to an appropriately sharp strong law but entails unseparated functional-analytic and Glivenko-Cantelli components. However the separation of these components permits in a straightforward fashion important extensions to nonclassical versions of empirical df's such as our H_n. Therefore, we take special care to achieve this separation, without sacrificing sharpness. We shall first develop the functional-analytic theory for the functional $T_n(G)$ given by (3.2), then treat the matter of substitution of an empirical df H_n or $H_{\underline{n}}$.

We now state for the case of *continuous* g our key functional-analytic theorem for the functional $T_n(\cdot)$ defined by (3.2), (3.4) and (3.5), and the functional $T(\cdot)$

given by (3.1), (3.3) and (3.5). (Extension to the general case of arbitrary Borel-measurable g requires some additional details and development, which we discuss briefly at the conclusion of Sect. 4.). The relevant functions $\{J_n(t), 0<t<1\}$, $\{J_t(s), 0<s<1\}$ for fixed t, and $\{T_t(G), 0<t<1\}$ will be viewed as functions $f:(0,1)\to\mathbb{R}$ and we shall denote the $L_p((0,1), \mathbf{B}_{(0,1)}, \lambda)$ norms of f by $\|f\|_p$

$$=\left(\int_0^1 |f(t)|^p\,dt\right)^{1/p} \text{ in the case } 1\leq p<\infty \text{ and } =\text{ess sup}\,|f| \text{ in the case } p=\infty, \text{ where}$$

λ denotes Lebesgue measure on $((0,1), \mathbf{B}_{(0,1)})$. For the functionals $T_t(G), 0<t<1$, it will be convenient to use the notation $m_G(t)=T_t(G)$, $0<t<1$. Due to the complexity of the functional $T_n(G)$ as well as of the restrictions on the sequence $\{G_n\}$ approaching G, it will be convenient to state the corresponding sets of assumptions prior to the formulation of the theorem.

The conditions to be met by the functional $T_n(G)$ are

Assumptions (3.8). *Consider the sequence of functionals* $\{T_n(\cdot), n\geq 1\}$ *given by*

(3.2), (3.3) *and* (3.4), *put* $A_0=\bigcup_{n=1}^{\infty} A_n$, *with* $A_n=\{t: 0<t<1, J_n(t)\neq 0\}$ *and* A_t

$=\{s: 0<s<1, J_t(s)\neq 0\}, 0<t<1,$ *and let* G *be a given df. We assume, with* $1\leq p$ $\leq\infty, p^{-1}+q^{-1}=1, 1\leq\tilde{p}\leq\infty, \tilde{p}^{-1}+\tilde{q}^{-1}=1,$ *that*

(3.8.a) $\|J_t\|_{\tilde{p}}<\infty,$ *each* $t\in A_0\cup\{p_1,\ldots,p_d\}$;

(3.8.b) $\sup_n \|J_n\|_p<\infty$ *if* $1<p\leq\infty$;

or $\{J_n(\cdot), n\geq 1\}$ *is uniformly integrable if* $p=1$;

(3.8.c) G^{-1} *is continuous at* $s=p_{tj}, 1\leq j\leq d_t,$ *for all* $t\in A_0$

        except for t *in a* λ*-null set, and for* $t=p_1,\ldots,p_d$. $\quad\square$

For a sequence of df's $\{G_n\}$ for which $T_n(G_n)-T_n(G)\to 0$ is to be established, we will require

Assumptions (3.9). *Let* $\{G_n\}$ *and* G *be* df*'s and* $q, \tilde{q}, A_0$ *and* A_t *as in Assumptions* (3.8). *We assume* $G_n\Rightarrow G$ *and*

(3.9.a) $\{(G_n^{-1})^{\tilde{q}}, n\geq 1\}$ *is uniformly integrable on* A_t,

        each $t\in A_0\cup\{p_1,\ldots,p_d\},$ *if* $1\leq\tilde{q}<\infty$; *or*

        $\sup_n \|G_n^{-1}\mathbf{1}_{A_t}\|_\infty<\infty,$ *each* $t\in A_0\cup\{p_1,\ldots,p_d\},$ *if* $\tilde{q}=\infty$;

(3.9.b) $\{(g\circ m_{G_n})^q, n\geq 1\}$ *is uniformly integrable on* A_0 *if* $1\leq q<\infty$; *or*

        $\sup_n \|(g\circ m_{G_n})\mathbf{1}_{A_0}\|_\infty$ *if* $q=\infty$. $\quad\square$

Theorem 3.1. (i) *Assume* g *continuous. Suppose that the sequence of functionals* $\{T_n(\cdot), n\geq 1\}$ *given by* (3.2), (3.4) *and* (3.5) *satisfies Assumptions* (3.8) *for a df* G, *and that the sequence of* df*'s* $\{G_n\}$ *satisfies Assumptions* (3.9) *for* G. *Then*

(3.10) $T_n(G_n)-T_n(G)\to 0, \quad n\to\infty.$

(ii) *Suppose that the conditions of part* (i) *are satisfied and that there exists a function*

(3.11) $$J \in L_p((0, 1), \mathbf{B}_{(0, 1)}, \lambda)$$

such that

(3.12) $$\int_0^\tau J_n(s)\,ds \to \int_0^\tau J(s)\,ds, \quad all \ \tau \in (0, 1).$$

Then, for $T(\cdot)$ *given by* (3.1), (3.3) *and* (3.5) *we have, assuming* $\|(g \circ m_G)\,1_{A_0}\|_q < \infty$,

(3.13) $$T_n(G) \to T(G), \quad n \to \infty.$$

The proof of Theorem 3.1 is deferred to Sect. 4.

Turning now to the *stochastic* component of our development, we put $\{H_n\}$ and H_F for $\{G_n\}$ and G. The main issue, then, with Assumptions (3.8) assumed with respect to H_F, is to verify that with probability 1 $\{H_n\}$ is a sequence of df's satisfying Assumptions (3.9) and thus to conclude that $T(H_n) \to T(H_F)$ $wp1$. Besides the probability space $((0, 1), \mathbf{B}_{(0, 1)}, \lambda)$ which is central to the functional-analytic treatment, we now also have a probability space $(\Omega, \mathbf{A}, \mathbb{P})$ on which our basic r.v.'s $\{X_i\}$ are defined. Thus we shall be speaking of almost sure convergence in two possible senses, a.s. [$\mathbb{P}$] or a.s. [λ]. Our goal is to establish that $T_n(H_n) \to T(H_F)$ a.s. [$\mathbb{P}$].

We shall use the property of empirical df's that the support of H_n must belong to that of H_F. The support of a df G is defined to be $S(G) = \{x : G(x + \varepsilon) > G(x - \varepsilon), \text{ all } \varepsilon > 0\}$. The following is easily checked.

Lemma 3.1. *If* G_0 *and* G_1 *are (right-continuous) df's with* $S(G_1) \subset S(G_0)$, *then*

(3.14) $$G_1^{-1}(t) = G_0^{-1}(G_0(G_1^{-1}(t))), \quad t \in (0, 1).$$

In verifying conditions (3.9.a) or (3.9.b), it is convenient (and unrestrictive from a practical standpoint) to assume that the sets A_0 and A_t, $0 < t < 1$, are intervals in (0, 1). The results we give below will be specific to the cases that the interval in question is either the open interval (0, 1) or (contained in) a closed interval $[a, b]$ in (0, 1).

Lemma 3.2. *Let* $1 \leq \tilde{q} \leq \infty$. *If, for all* $t \in A_0 \cup \{p_1, \ldots, p_d\}$, *either all* $A_t = (0, 1)$ *and* $\|H_F^{-1}\|_{\tilde{q}} < \infty$ *or all* $A_t \subset [a, b]$, *then a.s.* [$\mathbb{P}$] *the sequence* $\{H_n\}$ *satisfies* (3.9.a); *i.e., for* n *sufficiently large* $\{(H_k^{-1})^{\tilde{q}}, k \geq n\}$ *is uniformly integrable on* A_t, *each* $t \in A_0 \cup \{p_1, \ldots, p_d\}$ *if* $1 \leq \tilde{q} < \infty$, *and* $\sup_{k \geq n} \|H_k^{-1} 1_{A_t}\|_\infty < \infty$, *each* $t \in A_0 \cup \{p_1, \ldots, p_d\}$ *if* $\tilde{q} = \infty$.

Proof. First consider the case $A_t = (0, 1)$, $t \in A_0 \cup \{p_1, \ldots, p_d\}$. For $\tilde{q} = \infty$, we have by Lemma 3.1 that

$$\|H_n^{-1}\|_\infty = \sup_t |H_n^{-1}(t)| = \sup_t |H_F^{-1} \circ H_F \circ H_n^{-1}(t)| \leq \|H_F^{-1}\|_\infty.$$

For $\bar{q} < \infty$, we have

$$\|H_n^{-1}\|_{\bar{q}}^{\bar{q}} = \int_0^1 |H_n^{-1}(t)|^{\bar{q}} dt = \int_{\mathbb{R}} |y|^{\bar{q}} dH_n(y) = n_{(m)}^{-1} \sum |h(X_{i_1}, \ldots, X_{i_m})|^{\bar{q}},$$

so that $\|H_n^{-1}\|_{\bar{q}}^{\bar{q}}$ is a U-statistic with kernel $|h|^{\bar{q}}$ and a.s. $[\mathbb{P}]$ converges to $\|H_F^{-1}\|_{\bar{q}}^{\bar{q}} < \infty$ by the SLLN for U-statistics. Hence, by a standard result, a.s. $[\mathbb{P}]$ the functions $\{(H_k^{-1})^{\bar{q}}, k \geq n\}$ are uniformly integrable on $(0, 1)$, for n sufficiently large.

If $A_t \subset [a, b]$, for all $t \in A_0 \cup \{p_1, \ldots, p_d\}$, take $0 < a - \varepsilon$, $b + \varepsilon < 1$ continuity points of H_F^{-1}. Then a.s. $[\mathbb{P}]$ $\sup_t |H_n^{-1}(t) \mathbf{1}_{A_t}| \leq |H_F^{-1}(a - \varepsilon)| + |H_F^{-1}(b + \varepsilon)|$ for all n sufficiently large. Therefore in this case (3.9.a) is immediate. $\square$

Verification of (3.9.b) in general form is rather troublesome, but in practice it can be checked ad hoc for typical cases of the functionals $\{T_t(G), t \in A_0\}$, under effective restrictions on g and A_0 which do not preclude the motivating applications for this investigation. For such purposes the following result is useful.

Lemma 3.3. *Assume g continuous. Suppose that* (3.8.a), (3.8.c) *and the conditions of Lemma 3.2 are satisfied. Assume that a.s.* $[\mathbb{P}]$

$$(3.15) \qquad\qquad \limsup_{n \to \infty} \sup_{t \in A_0} |T_t(H_n)| < \infty.$$

Let $1 \leq q \leq \infty$ be given. Then a.s. $[\mathbb{P}]$ the sequence $\{H_n\}$ satisfies (3.9.b); i.e., for n sufficiently large $\{(g \circ m_{H_k})^q, k \geq n\}$ is uniformly integrable on A_0 if $1 \leq q < \infty$; or $\sup_{k \geq n} \|(g \circ m_{H_k}) \mathbf{1}_{A_0}\|_\infty < \infty$ if $q = \infty$.

Proof. We have a.s. $[\mathbb{P}]$

$$(3.16) \qquad \|(g \circ m_{H_n}) \mathbf{1}_{A_0}\|_\infty \leq \sup_{t \in A_0} |g \circ T_t(H_n)| \leq \sup_{x \in B(H_n)} |g(x)|$$

where $B(H_n) = \{T_t(H_n), t \in A_0\}$. By (3.15), for all but a finite number of n's, $B(H_n)$ is contained in a compact set, whence by continuity of g the term on the right in (3.16) is bounded for n sufficiently large. Thus for $q = \infty$ the conclusion of the lemma follows from (3.16). If $1 \leq q < \infty$ (3.16) implies for $\varepsilon > 0$ a.s. $[\mathbb{P}]$ boundedness of $\|(g \circ H_n^{-1}) \mathbf{1}_{A_0}\|_{q(1+\varepsilon)}$ for n sufficiently large. This is sufficient for the desired uniform integrability. $\square$

With these tools, we now can give two very general convergence results. We first deal in Corollary 3.1 with strong convergence of $T_n(H_n)$, with $T_n(\cdot)$ given by (3.2), for the important special case $T_t(G) = G^{-1}(t)$. Secondly, Corollary 3.2 deals with the strong convergence of $T_n(H_n)$, with $T_t(G)$ an L-functional. Both results are specializations of Theorem 3.1 to the stochastic sequence $\{H_n\}$.

Note also, that, with H_F and $\{H_n, n \geq 1\}$ for G and $\{G_n, n \geq 1\}$, we have by our Glivenko-Cantelli result (Corollary 2.1) that $H_n \Rightarrow H_F$ a.s. $[\mathbb{P}]$.

Corollary 3.1. *Assume g continuous. Let $T(\cdot)$ and $\{T_n(\cdot), n \geq 1\}$ be given by (3.1) and (3.2) with $T_t(G) = G^{-1}(t)$ and $dK(t)$, $dK_n(t)$ as in (3.3) and (3.4). Suppose $\{J_n(\cdot), n \geq 1\}$ satisfies (3.8.b), H_F satisfies (3.8.c) and that either $A_0 = (0, 1)$ and $\|g \circ H_F^{-1}\|_q < \infty$ or $A_0 \subset [a_0, b_0] \subset (0, 1)$, where $1 \leq p \leq \infty$, $p^{-1} + q^{-1} = 1$. Also let $J(\cdot)$ be determined by (3.11) and (3.12). Then a.s. [$\mathbb{P}$]*

$$(3.17) \qquad T_n(H_n) \to T(H_F), \qquad n \to \infty.$$

Proof. To apply Theorem 3.1, we verify that Assumptions (3.8) and (3.9) hold a.s. [$\mathbb{P}$] for $\{H_n\}$. Since $T_t(G) = G^{-1}(t)$, we have $J_t(s) = 0$, each $t \in A_0 \cup \{p_1, \ldots, p_d\}$, so $A_t = \phi$ and (3.8.a) and (3.9.a) are vacuous. Since, $m_G = G^{-1}$, (3.9.b) simply means that a.s. [$\mathbb{P}$] $\{(g \circ H_k^{-1})^q, k \geq n\}$ is uniformly integrable on A_0 for n sufficiently large if $1 \leq q < \infty$; or $\sup_{k \geq n} \|(g \circ H_k^{-1}) 1_{A_0}\|_\infty < \infty$ if $q = \infty$. For the case $A_0 = (0, 1)$, (3.9.b) follows by repeating the first part of the proof of Lemma 3.2 with H_n^{-1} replaced by $g \circ H_n^{-1}$ and H_F^{-1} replaced by $g \circ H_F^{-1}$. If $A_0 \subset [a_0, b_0]$ we argue as in the proof of Lemma 3.3. □

Remark 3.1. Note that our corollaries remain valid if we take $A_0 = \bigcup_{n=N}^{\infty} A_n$ for some $N \in \mathbb{N}$.

Examples 3.1

Linear combinations of functions of generalized order statistics. Corollary 3.1 essentially covers the class of statistics (3.6), which includes some classical collections of statistics as well as new varieties of recent interest in the literature. We present several examples.

(i) *Linear combinations of functions of order statistics.* Here $m = 1$, $h(x) = x$. For this specific choice and taking $A_0 = (0, 1)$ Corollary 3.1 reduces to Corollary 2.1 of van Zwet (1980), specialized to the case g continuous. (Following the proof of Theorem 3.1, in Sect. 4, we discuss extension to the case of arbitrary Borel-measurable functions $g(\cdot)$.) Confining attention to the case $g(x) = x$, we mention the *sample mean* (take $J_n(t) \equiv 1$, $d = 0$, and assume $\|F^{-1}\|_1 < \infty$, i.e., F has finite mean), the *sample median* (take $J_n(t) \equiv 0$, $d = 1 = a_1$, $p_1 = \frac{1}{2}$, and assume that F^{-1} is continuous at $\frac{1}{2}$, i.e., that $\frac{1}{2}$ is the unique solution of $F(x-) \leq \frac{1}{2} \leq F(x)$), *trimmed means* $(J_n(t) = (n/(n - 2[\alpha n])) 1\{[\alpha n]/n \leq t \leq (n - [\alpha n])/n\}$, $d = 0)$, *Gini's mean difference* $(J_n(t) = (n/(n-1))(4t - 2)$, $d = 0$ and assume $\|F^{-1}\|_1 < \infty)$, etc.

(ii) *U-statistics.* Apply Corollary 3.1 with $g(x) = x$, $J_n(t) \equiv 1$, $d = 0$, assuming $\|H_F^{-1}\|_1 < \infty$, i.e., $E|h| < \infty$. Hence the corollary reduces to the strong law of large numbers for U-statistics, obtained by Hoeffding (1961).

(iii) *Trimmed U-statistics.* Apply Corollary 3.1 with $g(x) = x$,

$$J_n(t) = (n_{(m)}/(n_{(m)} - 2[\alpha n_{(m)}])) 1\{[\alpha n_{(m)}] \leq n_{(m)} t \leq n_{(m)} - [\alpha n_{(m)}]\}, \qquad d = 0.$$

In particular, a *"trimmed variance"* is given by taking $m = 2$, $h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$.

(iv) *Another spread measure.* Apply Corollary 3.1 with $g(x) = x$, $J_n(t) \equiv 0$, $d = 1$ $= a_1$, $p_1 = \frac{1}{2}$, and $m = 2$, $h(x_1, x_2) = |x_1 - x_2|$. This yields a spread measure considered by Bickel and Lehmann (1979), the relevant parameter being the median of the distribution of $|X_1 - X_2|$.

(v) *Generalized Hodges-Lehmann location estimators.* Apply Corollary 3.1 with $g(x) = x$, $J_n(t) \equiv 0$, $d = 1 = a_1$ and $h(x_1, \ldots, x_m) = m^{-1}(x_1 + \ldots + x_m)$. (For $m = 1$ this gives the sample median and for $m = 2$ the classical Hodges-Lehmann estimator.). $\square$

Corollary 3.2. *Assume g continuous. Let $T(\cdot)$ and $\{T_n(\cdot), n \geq 1\}$ be given by (3.1) and (3.2) with $dK(t)$, $dK_n(t)$ and $dM_t(s)$ as in (3.3)–(3.5). Suppose that the weight functions satisfy (3.8.a) and (3.8.b) and that H_F satisfies (3.8.c), and $\|(g \circ m_{H_F}) \mathbf{1}_{A_0}\|_q < \infty$. Further suppose for all $t \in A_0 \cup \{p_1, \ldots, p_d\}$ either $A_t = (0, 1)$ and $\|H_F^{-1}\|_{\tilde{q}} < \infty$ or $A_t \subset [a, b]$; where $1 \leq p \leq \infty$, $p^{-1} + q^{-1} = 1$, $1 \leq \tilde{p} \leq \infty$, $\tilde{p}^{-1} + \tilde{q}^{-1} = 1$. If a.s. [$\mathbb{P}$] $T_t(H_n)$, $t \in A_0$, satisfies (3.15), then a.s. [$\mathbb{P}$]*

$$(3.18) \qquad\qquad T_n(H_n) \to T(H_F), \qquad n \to \infty.$$

Proof. We only have to verify that Assumptions (3.9) hold a.s. [$\mathbb{P}$] for $\{H_n\}$. For (3.9.a) we use Lemma 3.2 and (3.9.b) is immediate from Lemma 3.3. $\square$

Examples 3.2

Some spread estimators of Bickel and Lehmann. We consider here some functionals $T(\cdot)$ for which $T(H_n)$ gives statistics of the type proposed by Bickel and Lehmann (1979) as measures of spread for a *df F* not necessarily symmetric about any known point, and for which asymptotic normality has been proved by Janssen, Serfling and Veraverbeke (1984).

(i) Consider the functional

$$T(F) = \int_0^1 |F^{-1}(t) - F^{-1}(1 - t)|^\gamma \, dK(t).$$

Bickel and Lehmann (1979) suggest the case $\gamma = 2$ and $K(\cdot)$ uniform on $[\beta, 1 - \beta]$, where $0 < \beta < \frac{1}{2}$, giving

$$(3.19) \qquad T_\beta(F) = (1 - 2\beta)^{-1} \int_\beta^{1-\beta} [F^{-1}(t) - F^{-1}(1 - t)]^2 \, dt.$$

The corresponding statistic is given by

$$(3.20) \qquad T_{\beta n}(F_n) = (n - 2[\beta n])^{-1} \sum_{k = [\beta n] + 1}^{n - [\beta n]} (X_{n, k} - X_{n, n-k})^2.$$

Then (3.20) is of form (3.2) with the specializations $h(x) = x$ (making $H_F = F$ and $H_n = F_n$); $g(x) = x^2$; $J_n(t) = (n/(n - 2[\beta n])) \mathbf{1}\{[\beta n]/n \leq t \leq (n - [\beta n])/n\}$;

$A_0 \subset [\beta - \eta, \ 1 - \beta + \eta] \subset (0, 1)$ for some $\eta > 0$; $d = 0$; and for each $t \in A_0$, $J_t(s) = 0$, $A_t = \phi$, $d_t = 2$, $a_{t1} = 1 = -a_{t2}$, $p_{t1} = t = 1 - p_{t2}$, i.e., $T_t(G) = G^{-1}(t) - G^{-1}(1 - t)$. To apply Corollary 3.2 we take $p = \infty$, $q = 1$ and need only verify the assumptions $\|m_F \mathbf{1}_{[\beta - \eta, \ 1 - \beta + \eta]}\|_1 < \infty$ and (3.15).

Now in the present case $m_F(t) = F^{-1}(t) - F^{-1}(1 - t)$, and we trivially have

$$\|m_F \mathbf{1}_{[\beta - \eta, \ 1 - \beta + \eta]}\|_1 \leqq 2(|F^{-1}(\beta - \eta)| + |F^{-1}(1 - \beta + \eta)|) < \infty.$$

Condition (3.15) reduces to a.s. [$\mathbb{P}$] that

$$\limsup_{n \to \infty} \ \sup_{t \in [\beta - \eta, \ 1 - \beta + \eta]} \ |F_n^{-1}(t) - F_n^{-1}(1 - t)| < \infty,$$

which follows easily by arguments used above. Finally note that the difference between

$$T_{\beta n} = (n - 2[\beta n])^{-1} \sum_{k = [\beta n] + 1}^{n - [\beta n]} (X_{n, k} - X_{n, n - k + 1})^2,$$

the spread estimator for (3.19) proposed by Bickel and Lehmann (1979), and $T_{\beta n}(F_n)$ tends a.s. to zero. Therefore, we obtain, *without any assumption on* F, that a.s. [$\mathbb{P}$] $T_{\beta n} \to T(F)$, $n \to \infty$.

(ii) Consider the functional

$$T_{\alpha\beta}(G) = \left(\frac{1}{1 - \alpha - \beta}\right) \int_\alpha^{1 - \beta} \left[G^{-1}\left(\frac{t + 1}{2}\right)\right]^2 dt,$$

where $0 < \alpha < \frac{1}{2} < 1 - \beta < 1$. Bickel and Lehmann (1979) introduced another spread statistic which is given by $T_{\alpha\beta}(H_n)$, for the kernel $h(x_1, x_2) = x_1 - x_2$. In this case H_F is the symmetric (about 0) df of $X - X'$, for X and X' independent r.v.'s with df F. The functional $T_{\alpha\beta}(\cdot)$ is of form (3.1) with $g(x) = x^2$, $J(t) = \mathbf{1}_{A_0}(t)/(1 - \alpha - \beta)$, $A_0 = [\alpha, 1 - \beta]$, $d = 0$; and for each $t \in A_0$, $J_t(s) \equiv 0$, $A_t = \phi$, $d_{t1} = 1 = a_{t1}$, $p_{t1} = \frac{1}{2}(t + 1)$, i.e., $T_t(G) = G^{-1}\left(\frac{t + 1}{2}\right)$. To apply Corollary 3.2, we take $p = \infty$, $q = 1$ and we have

$$\|m_{H_F} \mathbf{1}_{[\alpha, \ 1 - \beta]}\|_1 = \left\|H_F^{-1}\left(\frac{t + 1}{2}\right) \mathbf{1}_{[\alpha, \ 1 - \beta]}\right\|_1 \leqq \left|H_F^{-1}\left(\frac{\alpha + 1}{2}\right)\right| + \left|H_F^{-1}\left(1 - \frac{\beta}{2}\right)\right| < \infty.$$

Condition (3.15) reduces to a.s. [$\mathbb{P}$] $\limsup_{n \to \infty} \ \sup_{t \in [\alpha, \ 1 - \beta]} \left|H_n^{-1}\left(\frac{t + 1}{2}\right)\right| < \infty$, which again follows by previously used arguments. Thus we conclude, *without any assumptions on* F *or* H_F, that a.s. [$\mathbb{P}$] $T_{\alpha\beta}(H_n) \to T_{\alpha\beta}(H_F)$, $n \to \infty$. $\quad\square$

Extensions to the multi-sample case are straightforward, now that the functional-analytic and "Glivenko-Cantelli" parts of the problem have been separated. The relevant basic Glivenko-Cantelli theorem giving $\|H_n - H_F\| \to 0$, and

thus $H_n \Rightarrow H_F$ a.s. [$\mathbb{P}$], is available in Corollary 2.1* of Sect. 2. The SLLN for U-statistics used in Lemma 3.2 becomes replaced by the one for "generalized" U-statistics (Sen (1977) and McConnell (1987)). Introducing other appropriate notions regarding convergence with multi-dimensional indices, and imposing appropriate restrictions, one can obtain for the functional given by (3.1) that

a.s. [$\mathbb{P}$] $T(H_n) \to T(H_F)$ as $\min(n_1, \ldots, n_c) \xrightarrow{(A)} \infty$. The detailed treatment of the multi-sample case will be pursued elsewhere. $\square$

4. Proof and Extensions

In the proof of Theorem 3.1 the following basic convergence lemma will be instrumental.

Lemma 4.1. *Let* $1 \leq r \leq \infty$ *and* $r^{-1} + s^{-1} = 1$. *Let the r.v.'s* ξ, $\{\xi_n\}$ *and* $\{\eta_n\}$ *on a probability space* $(\Omega, A, \mathbb{P})$ *satisfy* $\xi_n \xrightarrow{p} \xi$ *and either*

(i) $1 \leq r < \infty$, $\{|\xi_n|^r\}$ *is uniformly integrable and* $\sup_n \|\eta_n\|_s < \infty$,

or

(ii) $r = \infty$, $\sup_n \|\xi_n\|_\infty < \infty$ *and* $\{\eta_n\}$ *is uniformly integrable.*

Then $\|(\xi_n - \xi)\eta_n\|_1 \to 0, n \to \infty$. $\square$

Proof of Theorem 3.1. We first consider convergence of $m_{G_n}(t) = T_t(G_n)$, $n \to \infty$, for a fixed $t \in A_0 \cup \{p_1, \ldots, p_d\}$. Note that

$$m_{G_n}(t) = \int_0^1 G_n^{-1}(s) J_t(s) \, ds + \sum_{j=1}^{d_t} a_{tj} G_n^{-1}(p_{tj}).$$

Note that the convergence $G_n \Rightarrow G$ implies (see Lemma 1.5.6 of Serfling (1980) and its proof) that

(4.1.a) $G_n^{-1}(t) \to G^{-1}(t)$,   all continuity points t of G^{-1},

and hence

(4.1.b) $\lambda\{t: G_n^{-1}(t) \to G^{-1}(t), n \to \infty\} = 0$,

i.e., G_n^{-1} converges to G^{-1} a.s. [λ] and hence in λ-measure. We apply Lemma 4.1 with $(\Omega, A, \mathbb{P}) = ((0, 1), \mathbf{B}_{(0,1)}, \lambda)$ and $\xi = G^{-1} 1_{A_t}$, $\xi_n = G_n^{-1} 1_{A_t}$, $\eta_n = J_t$, $r = \tilde{q}$, $s = \tilde{p}$. The hypotheses of the lemma are verified using (3.8.a), (3.9.a) and (4.1). This yields

$$\int_0^1 G_n^{-1}(s) J_t(s) \, ds \to \int_0^1 G^{-1}(s) J_t(s) \, ds$$

for all $t \in A_0 \cup \{p_1, \ldots, p_d\}$. The same is true, by (3.8.c) and (4.1.a) for the other part of $T_t(G_n)$ and hence also for $T_t(G_n)$ itself for all $t \in A_0 \cup \{p_1, \ldots, p_d\}$, except for t in a λ-null set not including $\{p_1, \ldots, p_d\}$. Thus we have proved that for $t = p_1, \ldots, p_d$ and almost all t in A_0,

$$(4.2) \qquad\qquad m_{G_n}(t) \to m_G(t).$$

Whence, by the continuity of g,

$$(4.3) \qquad\qquad g \circ m_{G_n} \to g \circ m_G \quad \text{in } \lambda\text{-measure.}$$

Now we apply Lemma 4.1 again, with the same probability space as above but $\xi = (g \circ m_G) 1_{A_0}$, $\xi_n = (g \circ m_{G_n}) 1_{A_0}$, $\eta_n = J_n$, $r = q$, $s = p$. The hypotheses of the lemma are verified using (3.8.b), (3.9.b) and (4.3), and we thus have proved

$$(4.4) \qquad\qquad \int_0^1 [g \circ T_t(G_n) - g \circ T_t(G)] \, J_n(t) \, dt \to 0, \qquad n \to \infty.$$

Applying continuity of g in connection with (4.2) for $t = p_1, \ldots, p_d$, and combining with (4.4), we obtain (3.10), completing the proof of part (i) of the theorem.

Regarding part (ii) we use the assumptions (3.8.b), (3.11), (3.12) and the fact that $g \circ T_t(G)$ is assumed to be in L_q in combination with Sects. IV.8.11, IV,13.23, IV.13.25 and IV.13.27 of Dunford and Schwartz (1958) to obtain

$$\int_0^1 g \circ T_t(G) \, J_n(t) \, dt \to \int_0^1 g \circ T_t(G) \, J(t) \, dt, \qquad n \to \infty,$$

and hence (3.13). $\quad\square$

Remark 4.1. (i) Theorem 3.1 is a deterministic version and generalization, for g continuous, of Theorem 2.1 and Corollary 2.1 of van Zwet (1980).

(ii) As mentioned already in Example 3.1(i), our Corollary 3.1 with $m = 1$, $h(x) = x$ and $A_0 = (0, 1)$ reduces to Corollary 2.1 of van Zwet (1980), except that for simplicity of presentation we have restricted g to be continuous (note that we do not require H_F^{-1} to be continuous). Following van Zwet's approach in spirit this condition can be removed by an application of Luzin's theorem, giving the following extension.

Corollary 4.1. *Corollary* 3.1 *remains valid for arbitrary Borel-measurable g satisfying*

$$(4.5) \qquad\qquad g \text{ is continuous at } H_F^{-1}(p_j), \qquad j = 1, \ldots, d$$

and, for the case $A_0 \subset [a_0, b_0] \subset (0, 1)$,

$$(4.6) \qquad g \text{ is bounded on } [H_F^{-1}(a_0) - \varepsilon, H_F^{-1}(b_0) + \varepsilon], \qquad \text{for some } \varepsilon > 0.$$

Proof. To prove Corollary 3.1 the continuity of g is used to show, in the case $T_t(G) = G^{-1}(t)$, the validity of (4.3) and the boundedness of the r.h.s. in (3.17).

The latter is still valid by (4.6). To show that (4.3) is still valid, it suffices to prove that a.s. [$\mathbb{P}$]

$$\lim_{n \to \infty} \lambda\{t \in A_0 \colon |g \circ H_n^{-1}(t) - g \circ H_F^{-1}(t)| > \delta\} = 0, \quad \text{every } \delta > 0.$$

By Luzin's theorem we have for any $\varepsilon > 0$ the existence of a Borel set B and a continuous function $\tilde{g}$ with $H_F(B) < \varepsilon$ and such that $g = \tilde{g}$ on $\mathbb{R} \setminus B$. Therefore, with $B_n = \{t \colon H_n^{-1}(t) \in B\}$,

(4.6) $\displaystyle\limsup_{n \to \infty} \lambda\{t \in A_0 \colon |g \circ H_n^{-1}(t) - g \circ H_F^{-1}(t)| > \delta\} \leq H_F(B) + \limsup_{n \to \infty} H_n(B)$

$\displaystyle + \limsup_{n \to \infty} \lambda\{t \in A_0 \colon |\tilde{g} \circ H_n^{-1}(t) - \tilde{g} \circ H_F^{-1}(t)| > \delta\}.$

Since $H_n^{-1}(t) \to H_F^{-1}(t)$, $n \to \infty$, a.s. [$\mathbb{P}$] for all continuity points t of $H_F^{-1}(\cdot)$, the last term in the r.h.s. of (4.6) equals zero. The sum of the first two terms is less than 2ε. Hence the proof is complete. $\square$

(iii) We finally note that our general *deterministic* result, Theorem 3.1, can also be stated without the continuity assumption on g. This development involves formulation of additional convergence concepts for sequences $\{G_n, n \geq 1\}$ and is omitted for brevity here. Complete details can be found in Helmers et al. (1985) and Serfling (1985 b).

References

Bickel, P.J., Lehmann, E.L.: Descriptive statistics for non-parametric models. IV. Spread. In: Jurečková, J. (ed.) Contributions to statistics. Hájek Memorial Volume, pp. 33–40. Prague: Academia 1979

Chung, K.L.: An estimate concerning the Kolmogorov limit distribution. Trans. Am. Math. Soc. 67, 36–50 (1949)

Dunford, N., Schwartz, J.T.: Linear operators, vol. I. New York: Wiley 1958

Dvoretzky, A., Kiefer, J., Wolfowitz, J.: Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann. Math. Statist. 27, 642–669 (1956)

Gaenssler, P.: Empirical processes. Hayward, Calif.: Institute of Mathematical Statistics, Lecture Notes-Monograph Series, vol. 3 (1983)

Helmers, R.: Edgeworth expansions for linear combinations of order statistics. Mathematical Centre Tracts, vol. 105. Amsterdam: Mathematisch Centrum 1982

Helmers, R., Janssen, P., Serfling, R.: Glivenko-Cantelli properties of some generalized empirical df's and strong convergence of generalized L-statistics. Technical Report No. 460, Dept. of Math. Sciences, Johns Hopkins University, Baltimore (1985)

Hoeffding, W.: The strong law of large numbers for U-statistics. Univ. of North Carolina Institute of Statistics Mimeo Series, No. 302 (1961)

Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. 58, 13–30 (1963)

Janssen, P., Serfling, R., Veraverbeke, N.: Asymptotic normality for a general class of statistical functions and application to measure of spread. Ann. Stat. 12, 1369–1379 (1984)

McConnell, T.R.: Two-parameter strong laws and maximal inequalities for U-statistics. Proc. of the Royal Soc. of Edinburgh 107 A, 133–151 (1987)

Sen, P.K.: Almost sure convergence of generalized U-statistics. Ann. Probab. 5, 287–290 (1977)

Serfling, R.J.: Approximation theorems of mathematical statistics. New York: Wiley 1980

Serfling, R.J.: Generalized L-, M- and R-statistics. Ann. Stat. **12**, 76–86 (1984)

Serfling, R.J.: A Bahadur representation for quantiles of empirical df's of generalized U-statistic structure. Technical Report No. 453, Dept. of Math. Sciences, Johns Hopkins University, Baltimore (1985a)

Serfling, R.J.: A note on convergence of functions of random elements. Technical Report No. 459, Dept. of Math. Sciences, Johns Hopkins University, Baltimore (1985b)

van Zwet, W.R.: A strong law for linear functions of order statistics. Ann. Probab. **8**, 986–990 (1980)

Wellner, J.A.: A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. Ann. Stat. **5**, 473–480 (1977). Correction, ibid. **6**, 1394 (1977)