Stanley, R. M. (1972). Boundary crossing probabilities for the Kolmogorov–Smirnov statistics. *Ann. Math. Statist.* **43**, 664–668.

Steck, G. P. (1969). The Smirnov two-sample tests as rank tests. *Ann. Math. Statist.* **40**, 1449–1466.

Steck, G. P. (1971). Rectangular probabilities for uniform order statistics and the probability that the empirical distribution function lies between two distribution functions. *Ann. Math. Statist.* **42**, 1–11.

Steck, G. P. (1974). A new formula for $P(R_i \leq b_i,\ 1 \leq i \leq m/m,\ n,\ F = G^k)$. *Ann. Probability* **2**, 147–154.

Stute, W. (1982). The oscillation behavior of empirical processes. *Ann. Probability* **10**, 86–107.

Suzuki, G. (1967). On exact probabilities of some generalized Kolmogorov's D-statistics. *Ann. Inst. Statist. Math.* **19**, 373–388.

Takács, L. (1964). The use of a ballot theorem in order statistics. *J. Appl. Probability* **1**, 389–392.

Takács, L. (1965a). Applications of a ballot theorem in physics and in order statistics. *J. Roy. Statist. Soc. Ser. B* **27**, 130–137.

Takács, L. (1965b). The distributions of some statistics depending on the deviations between empirical and theoretical distribution functions. *Sankhyā Ser. A* **27**, 93–100.

Takács, L. (1967). *Combinatorial methods in the theory of stochastic processes.* Wiley, New York.

Takács, L. (1970). Combinatorial methods in the theory of order statistics. In: M. L. Puri, ed., *Nonparametric Techniques in Statistical Inference.* Cambridge Univ. Press, London, pp. 359–384.

Takács, L. (1971). On the comparison of a theoretical and an empirical distribution function. *J. Appl. Probability* **8**, 321–330.

Vincze, I. (1961). On two-sample tests based on order statistics. In: *Proc. Fourth Berkeley Sympos. Math. Statist. and Probability, Vol. I.* Univ. of California Press, Berkeley, CA, pp. 695–705.

Vincze, I. (1967). On some questions connected with two-sample tests of Smirnov type. In: *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability, Vol. I.* Univ. of California Press, Berkeley, CA, pp. 654–666.

Vincze, I. (1968). On the power of the Kolmogorov–Smirnov two-sample test and related non-parametric tests. *Studies in Math. Statist.: Theory and Appl.* (Sympos. Budapest, 1964). Akad. Kiadó, Budapest, pp. 201–210.

Vincze, I. (1970). On Kolmogorov–Smirnov type distribution theorems. In: M. L. Puri, ed., *Nonparametric Techniques in Statistical Inference.* Cambridge Univ. Press, London, pp. 385–401.

Vincze, I. (1972). On some results and problems in connection with statistics of the Kolmogorov–Smirnov type. In: *Proc. Sixth Berkeley Sympos. Math. Statist. and Probability, Vol. I.* Univ. of California Press, Berkeley, CA, pp. 459–470.

Wald, A. and Wolfowitz, J. (1939). Confidence limits for continuous distribution functions. *Ann. Math. Statist.* **10**, 105–118.

Walsh, J. E. (1963). Bounded probability properties of Kolmogorov–Smirnov and similar statistics for discrete data. *Ann. Inst. Statist. Math.* **15**, 153–158.

Wellner, J. A. (1977a). A martingale inequality for the empirical process. *Ann. Probability* **5**, 303–308.

Wellner, J. A. (1977b). A Glivenko–Cantelli theorem and strong laws of large numbers for functions of order statistics. *Ann. Statist.* **5**, 473–480. Correction: **6** (1978), 1391.

Wellner, J. A. (1977c). A law of the iterated logarithm for functions of order statistics. *Ann. Statist.* **5**, 481–494.

Wellner, J. A. (1977d). Distributions related to linear bounds for the empirical distribution function. *Ann. Statist.* **5**, 1003–1016.

Wellner, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. Verw. Gebiete* **45**, 73–88.

Whittle, P. (1961). Some exact results for one-sided distribution tests of the Kolmogorov–Smirnov type. *Ann. Math. Statist.* **32**, 499–505.

Yu, G. C. S. (1971). Pitman efficiencies of Kolmogorov–Smirnov tests. *Ann. Math. Statist.* **42**, 1595–1605.

Yu, G. C. S. (1975). Bounds on the power for some Rényi-type statistics. *J. Amer. Statist. Assoc.* **70**, 233–237.

Yu, G. C. S. (1977). Power bounds on some non-parametric test procedures for censored data. *Sankhyā Ser. B*, **39**, 279–283.

P. R. Krishnaiah and P. K. Sen, eds., *Handbook of Statistics, Vol. 4*
© Elsevier Science Publishers (1984) 431–462

20

Invariance Principles for Empirical Processes

Miklós Csörgő

1. Introduction: Basic notions and definitions

Let $X_1, X_2, \ldots$ be independent identically distributed random variables (i.i.d. rv) on a probability space $(\Omega, \mathcal{A}, P)$ with values in a sample space $(R, \mathcal{B})$. Denote by μ the probability distribution of X_1 on $\mathcal{B}$, i.e.,

$$\mu(B) = P\{\omega \in \Omega : X_1(\omega) \in B\} \quad \text{for all } B \in \mathcal{B}. \tag{1.1}$$

For each $\omega \in \Omega$ and $B \in \mathcal{B}$ let $n\mu_n(B)$ be the number of those $X_1(\omega), \ldots, X_n(\omega)$ which fall into the set B. The number $\mu_n(B)$ is called the *empirical measure* of B for the random sample $X_1, \ldots, X_n$, conveniently written as

$$\mu_n(B) = n^{-1} \sum_{j=1}^{n} 1_B(X_j) \tag{1.2}$$

where

$$1_B(x) = \begin{cases} 1 & \text{if } x \in B, \\ 0 & \text{if } x \notin B. \end{cases} \tag{1.3}$$

The corresponding *empirical measure process* β_n is defined by

$$\beta_n(B) = n^{1/2}(\mu_n(B) - \mu(B)), \quad B \in \mathcal{B}. \tag{1.4}$$

Usually $X_1, X_2, \ldots$ will be random vectors in the Euclidean space $\mathbb{R}^d$ $(d \geq 1)$, i.e., $R = \mathbb{R}^d$, and $\mathcal{B}$ is the Borel subsets of $\mathbb{R}^d$. In this case let F be the right continuous distribution function of X_1, i.e., $F(x) = \mu((-\infty, x]) = P\{\omega \in \Omega : X_1(\omega) \in (-\infty, x)\}$, where $(-\infty, x]$ is a d-dimensional interval with $x \in \mathbb{R}^d$. The corresponding *empirical distribution function* $\mu_n((-\infty, x])$ of the random sample $X_1, \ldots, X_n$ will be denoted by $F_n(x)$, i.e., for each $\omega \in \Omega$ $nF_n(x)$ is the number of those $X_j(\omega) = (X_{j1}(\omega), \ldots, X_{jd}(\omega))$ $(j = 1, \ldots, n)$

Research supported by a NSERC Canada Grant at Carleton University, Ottawa.

whose components are less than or equal to the corresponding components of $x=(x_1,\dots,x_d)\in\mathbb{R}^d$, conveniently written as

$$F_n(x)=n^{-1}\sum_{j=1}^{n}\prod_{i=1}^{d}1_{(-\infty,x_i]}(X_{ji})\,. \tag{1.5}$$

Whence on denoting $\beta_n((-\infty,x])$ by $\beta_n(x)$, the empirical measure process of (1.4) in terms of these distribution functions F, F_n is

$$\beta_n(x)=n^{1/2}(F_n(x)-F(x)),\quad x\in\mathbb{R}^d\ (d\geq 1)\,, \tag{1.6}$$

and it will be simply called the *empirical process*. If X_1 is uniformly distributed over the unit cube $I^d=[0,1]^d$ $(d\geq 1)$ then for F_n, F, β_n we use the symbols E_n, λ, α_n and the corresponding *uniform empirical process* then is

$$\alpha_n(y)=n^{1/2}(E_n(y)-\lambda(y)),\quad y\in I^d\ (d\geq 1)\,, \tag{1.7}$$

where $\lambda(y)=\Pi_{i=1}^d y_i$ with $y=(y_1,\dots,y_d)\in I^d$.

In the context of *continuous* distribution functions F on $\mathbb{R}^d$ the uniform empirical process occurs the following way. Let $\mathcal{F}$ be *the class of continuous distribution functions* on $\mathbb{R}^d$, and let $\mathcal{F}_0$ be *the subclass consisting of every member of* $\mathcal{F}$ *which is a product of its associated one-dimensional marginal distribution functions.* Let $y_i=F_{(i)}(x_i)$ $(i=1,\dots,d)$ be *the i-th marginal* distribution of $F\in\mathcal{F}$ and let $F_{(i)}^{-1}(y_i)=\inf\{x_i\in\mathbb{R}^1\colon F_{(i)}(x_i)\geq y_i\}$ be its *inverse.* Define the mapping $L^{-1}: I^d\to\mathbb{R}^d$ by

$$L^{-1}(y)=L^{-1}(y_1,\dots,y_d)=(F_{(1)}^{-1}(y_1),\dots,F_{(d)}^{-1}(y_d)),\quad y=(y_1,\dots,y_d)\in I^d\,. \tag{1.8}$$

Then (cf. (1.6) and (1.7)), *whenever* $F\in\mathcal{F}_0$,

$$\alpha_n(y)=\beta_n(L^{-1}(y)),\quad y=(y_1,\dots,y_d)\in I^d\ (d\geq 1)\,, \tag{1.9}$$

i.e., if $F\in\mathcal{F}_0$, then the empirical process $\beta_n(L^{-1}(y))=\alpha_n(y)$ is distribution free (does not depend on the distribution function F). *In statistical terminology* we say that when we are testing the *independence null hypothesis*

$$H_0\colon\ F\in\mathcal{F}_0\quad\text{versus the alternative}\quad H_1\colon\ F\in\mathcal{F}-\mathcal{F}_0\quad (d\geq 2)\,, \tag{1.10}$$

then the null distribution of $\beta_n(L^{-1}(y))$ is that of $\alpha_n(y)$, i.e., the same for all $F\in\mathcal{F}_0$ and for $d=1$ with F simply continuous. Otherwise, i.e., if H_1 obtains, the empirical process β_n is a function of F and so will be also its distribution.

Hoeffding (1948), and Blum, Kiefer and Rosenblatt (1961) suggested an *alternate empirical process* for handling H_0 of (1.10). Let F_{ni} be the *marginal*

empirical distribution function of the i-th component of X_j $(j=1,\dots,n)$, i.e.,

$$F_{ni}(x_i)=n^{-1}\sum_{j=1}^{n}1_{(-\infty,x_i]}(X_{ji}),\quad i=1,\dots,d\,, \tag{1.11}$$

and define

$$T_n(x)=T_n(x_1,\dots,x_d)=n^{1/2}\Big(F_n(x)-\prod_{i=1}^{d}F_{ni}(x_i)\Big),\quad d\geq 2\,, \tag{1.12}$$

with F_n as in (1.5). In terms of the mapping L^{-1} of (1.8) we define t_n, the uniform version of T_n, by

$$\begin{aligned} t_n(y)=T_n(L^{-1}(y))&=n^{1/2}\Big(F_n(F_{(1)}^{-1}(y_1),\dots,F_{(d)}^{-1}(y_d))-\prod_{i=1}^{d}F_{ni}(F_{(i)}^{-1}(y_i))\Big)\\ &=n^{1/2}\Big(E_n(y)-\prod_{i=1}^{d}E_{ni}(y_i)\Big)\,, \end{aligned} \tag{1.13}$$

where $E_{ni}(y_i)$ $(i=1,\dots,d)$ is the i-th uniform empirical distribution function of the i-th component of $L(X_j)=(F_{(1)}(X_{j1}),\dots,F_d(X_{jd}))$ $(j=1,\dots,n)$, i.e., L is the inverse of L^{-1} of (1.8). Consequently H_0 of (1.10) is equivalent to $H_0\colon F(L^{-1}(y))=\Pi_{i=1}^d y_i=\lambda(y)$, i.e., given H_0, $T_n(L^{-1}(y))=t_n(y)$ is distribution free. Hence, in order to study the distribution of T_n under H_0, we may take F to be the uniform distribution on I^d $(d\geq 2)$ and study the distribution of t_n instead.

If F is *a continuous distribution function* on $\mathbb{R}^1$ then its inverse function F^{-1} *will be called the quantile function* Q of F, i.e.,

$$\begin{aligned} Q(y)=F^{-1}(y)&=\inf\{x\in\mathbb{R}^1\colon F(x)\geq y\}\\ &=\inf\{x\in\mathbb{R}^1\colon F(x)=y\},\quad 0\leq y\leq 1\,. \end{aligned} \tag{1.14}$$

Thus $F(Q(y))=y\in[0,1]$, and if we put $U_1=F(X_1)$, then U_1 is a uniformly distributed rv on $[0,1]$. Also $U_1=F(X_1)$, $U_2=F(X_2),\dots$ are independent uniform-$[0,1]$ rv. Let $X_{1:n}\leq X_{2:n}\leq\cdots\leq X_{n:n}$ be the order statistics of a random sample on F which, in turn, induce the uniform-$[0,1]$ order statistics $U_{1:n}=F(X_{1:n})\leq U_{2:n}=F(X_{2:n})\leq\cdots\leq U_{n:n}=F(X_{n:n})$ of $U_1,U_2,\dots,U_n$. The empirical distribution function F_n of (1.5) can be written as

$$F_n(x)=\begin{cases}0 & \text{if } X_{1:n}>x\,,\\ k/n & \text{if } X_{k:n}\leq x<X_{k+1:n}\,,\\ 1 & \text{if } X_{n:n}\leq x\,,\end{cases} \tag{1.15}$$

$x\in\mathbb{R}^1$, and the uniform empirical distribution function E_n of $U_1,\dots,U_n$ as

$$E_n(y)=F_n(Q(y))=\begin{cases}0 & \text{if } U_{1:n}\,,\\ k/n & \text{if } U_{k:n}\leq y<U_{k+1:n}\,,\\ 1 & \text{if } U_{n:n}\leq y\,,\end{cases} \tag{1.16}$$

$0 \leq y \leq 1$. Then (1.7) takes up the simple form

$$\alpha_n(y) = \beta_n(Q(y)) = n^{1/2}(E_n(y) - y), \quad 0 \leq y \leq 1. \tag{1.17}$$

In terms of the latter empirical distribution functions F_n and E_n we define now the *empirical quantile function* Q_n by

$$\begin{aligned} Q_n(y) = F_n^{-1}(y) &= \inf\{x \in \mathbb{R}^1 : F_n(x) \geq y\} \\ &= X_{k:n} \quad \text{if } (k-1)/n < y \leq k/n \quad (k = 1, \ldots, n) \end{aligned} \tag{1.18}$$

and the uniform empirical quantile function U_n by

$$\begin{aligned} U_n(y) = E_n^{-1}(y) &= \inf\{u \in [0,1] : F_n(Q(u)) \geq y\} \\ &= U_{k:n} = F(X_{k:n}) \quad \text{if } (k-1)/n < y \leq k/n \quad (k = 1, \ldots, n) \\ &= F(Q_n(y)). \end{aligned} \tag{1.19}$$

If F is an *absolutely continuous distribution function* on $\mathbb{R}^1$, then let $f = F'$ be its density function (with respect to Lebesgue measure). We define the *quantile process* ρ_n by

$$\rho_n(y) = n^{1/2} f(Q(y))(Q_n(y) - Q(y)), \quad y \in (0,1), \tag{1.20}$$

and the corresponding *uniform quantile process* u_n by

$$u_n(y) = n^{1/2}(U_n(y) - y). \tag{1.21}$$

A simple relationship like that of (1.17) for α_n and β_n does not exist for u_n and ρ_n. However by (1.19) we have

$$\rho_n(y) = n^{1/2} f(Q(y))(Q(F(Q_n(y))) - Q(y)) = u_n(y)(f(Q(y))/f(Q(\theta_{y,n}))), \tag{1.22}$$

where $U_n(y) \wedge y < \theta_{y,n} < U_n(y) \vee y$. Since $u_n(k/n) = -\alpha_n(U_{k:n})$ $(k = 1, \ldots, n)$, it is reasonable to expect that the asymptotic distribution theory of α_n and u_n should be the same. This, in turn, implies that via (1.22) ρ_n should also have the same kind of asymptotic theory if f is 'nice'. We are going to see in Section 4 that this actually is true under appropriate conditions of f.

Let $C(t) = \int_{\mathbb{R}^d} \exp(i\langle t, x\rangle)\,\mathrm{d}F(x)$ be the characteristic function of $F(x)$ on $\mathbb{R}^d$, where $\langle t, x\rangle = \sum_{k=1}^d t_k x_k$, the usual inner product of $t = (t_1, \ldots, t_k)$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. With F_n as in (1.5) the *empirical characteristic function* C_n of the sample $X_1, \ldots, X_n$ is defined by

$$C_n(t) = n^{-1} \sum_{k=1}^n \exp(i\langle t, X_k\rangle) = \int_{\mathbb{R}^d} \exp(i\langle t, x\rangle)\,\mathrm{d}F_n(x), \quad t \in \mathbb{R}^d. \tag{1.23}$$

Next we define a number of Gaussian processes which play a basic role in the asymptotic distribution of some of the empirical processes introduced so far.

Wiener process: A real valued separable Gaussian process

$$\{W(x); x \in \mathbb{R}_+^d\} = \{W(x_1, \ldots, x_d); 0 \leq x_i < \infty, i = 1, \ldots, d\}$$

with continuous sample paths is called a Wiener process if $EW(x) = 0$ and

$$EW(x)W(y) = \lambda(x \wedge y) \quad \text{with } \lambda(x \wedge y) = \prod_{i=1}^d (x_i \wedge y_i),$$

where $x = (x_1, \ldots, x_d)$, $y = (y_1, \ldots, y_d) \in \mathbb{R}_+^d$ $(d \geq 1)$.

Brownian bridge: $\{B(x); x \in I^d\} = \{W(x) - \lambda(x)W(1, \ldots, 1); x \in I^d\}$. Whence $EB(x) = 0$ and $EB(x)B(y) = \lambda(x \wedge y) - \lambda(x)\lambda(y)$, $x = (x_1, \ldots, x_d)$, $y = (y_1, \ldots, y_d) \in I^d$ $(d \geq 1)$, where $\lambda(x) = \lambda(x \wedge x)$.

Kiefer process: $\{K(x,t); (x,t) \in I^d \times \mathbb{R}_+^1\} = \{W(x,t) - \lambda(x)W(1, \ldots, 1, t); x \in I^d, t \geq 0\}$. Whence $EK(x,t) = 0$ and

$$EK(x, t_1)K(y, t_2) = (t_1 \wedge t_2)(\lambda(x \wedge y) - \lambda(x)\lambda(y)),$$

where $x = (x_1, \ldots, x_d)$, $y = (y_1, \ldots, y_d) \in I^d$, $t_1, t_2 \geq 0$.

For a proof of existence of the multitime parameter Wiener process $\{W(x); x \in \mathbb{R}_+^d\}$, which is sometimes called the Yeh–Wiener process or Brownian (Wiener) sheet, we refer to Yeh (1960), Čencov (1956), or Csörgő and Révész (1981a, Section 1.11). A Kiefer process at integer valued arguments $t = n$ can be also viewed as the partial sum process of a sequence of *independent* Brownian bridges $\{B_i(x); x \in I^d\}_{i=1}^\infty$:

$$\{K(x,n); x \in I^d, n = 1, 2, \ldots\} = \left\{\sum_{i=1}^n B_i(x); x \in I^d, n = 1, 2, \ldots\right\},$$

and

$$\{B_n(x); x \in I^d\} = \{K(x,n) - K(x, n-1); x \in I^d\} \quad (n = 1, 2, \ldots) \tag{1.24}$$

is a sequence of independent Brownian bridges.

For further properties of a Kiefer process we refer to M. Csörgő and P. Révész (1981a, Section 1.15 and Theorem S.1.15.1).

Strong approximations of empirical processes are going to be described in Section 2. For distribution theory of tests based on the sample distribution function on $\mathbb{R}^1$ we refer to Durbin (1973a). A direct treatment of the empirical process on $\mathbb{R}^1$ and many of its statistical applications can be seen in this volume by Csáki (1982b), and Doksum and Yandell (1982) (cf. Also Csáki, 1977a, 1977b, 1982a). For the parameters estimated empirical process on $\mathbb{R}^d$ $(d \geq 1)$ we refer to Durbin (1973b, 1976), Neuhaus (1976) and M. D. Burke, M. Csörgő, S. Csörgő and P. Révész (1979). For a theory of, and further references on strong and weak convergence of the empirical characteristic function we refer to S. Csörgő (1980, 1981a,b).

Recent work on the limiting distribution of and critical values for the multivariate Cramér–von Mises and Hoeffding–Blum–Kiefer–Rosenblatt independence criteria is reviewed in Section 3.

An up-to-date review of strong and weak approximations of the quantile process, including that of weak convergence in $\|\cdot/q\|$-metrics, is given in Section 4. For further readings, references on this subject and its statistical applications, we refer to Chapters 4, 5 in Csörgő and Révész (1981a), Csörgő and Révész (1981), M. Csörgő, (1983), and Csörgő, Csörgő, Horváth and Révész (1982). For the parameters estimated quantile process we refer to Carleton Mathematical Lecture Note No. 31 by Aly and Csörgő (1981) and references therein. For recent results on the product-limit empirical and quantile processes we refer to Carleton Mathematical Lecture Note No. 38 by Aly, Burke, Csörgő, Csörgő and Horváth (1982) and references therein and to Chapter VIII in M. Csörgő (1983).

2. Strong and weak approximations of empirical processes by Gaussian processes

There is an excellent recent survey of results concerning empirical processes and measures of i.i.d. rv by Gaenssler and Stute (1979). Chapter 4 of Csörgő and Révész (1981a) is also devoted to the same subject on $\mathbb{R}^1$. For further references, in addition to the ones mentioned in this exposition, we refer to these sources. Even with the latter references added, this review is not and, indeed, does not intend to be complete. In this section we are going to list, or mention, essentially only those results for α_n and β_n which are best possible or appear to be the best available ones so far in their respective categories.

First, on strong approximations of the *uniform empirical process* α_n of (1.7) or, equivalently, that of $\alpha_n = \beta_n(L^{-1})$ of (1.9) in terms of a *sequence* of Brownian bridges $\{B_n(x); x \in I^d\}_{n=1}^{\infty}$ we have

THEOREM 2.1. *Let $X_1, \dots, X_n$ $(n = 1, 2, \dots)$ be independent random d-vectors, uniformly distributed on I^d, or with distribution function $F \in \mathscr{F}_0$ on $\mathbb{R}^d$. Let α_n be as in* (1.7) *or in* (1.9). *Then one can construct a probability space for $X_1, X_2, \dots$ with a sequence of Brownian bridges $\{B_n(y); y \in I^d\ (d \geq 1)\}_{n=1}^{\infty}$ on it so that:*

(i) *for all n and x we have (cf. Komlós, Major and Tusnády,* 1975a)

$$P\{\sup_{y\in I^1} |\alpha_n(y) - B_n(y)| > n^{-1/2}(C \log n + x)\} \leq Le^{-\lambda x}, \tag{2.1}$$

where C, L, λ are positive absolute constants,

(ii) *for all n and x we have (cf. Tusnády,* 1977a)

$$P\{\sup_{y\in I^2} |\alpha_n(y) - B_n(y)| > n^{-1/2}(C \log n + x) \log n\} \leq Le^{-\lambda x}, \tag{2.2}$$

where C, L, λ are positive absolute constants,

(iii) *for any $\lambda > 0$ there exists a constant $C > 0$ such that for each n (cf. Csörgő and Révész,* 1975a)

$$P\{\sup_{y\in I^d} |\alpha_n(y) - B_n(y)| > C(\log n)^{3/2} n^{-1/2(d+1)}\} \leq n^{-\lambda}, \quad d \geq 1. \tag{2.3}$$

The constants of (2.1) *for example can be chosen as* $C = 100$, $L = 10$, $\lambda = 1/50$ *(cf. Tusnády,* 1977b,c).

COROLLARY 2.1. (2.1), (2.2), (2.3) *in turn imply*

$$\sup_{y\in I^1} |\alpha_n(y) - B_n(y)| \overset{\text{a.s.}}{=} O(n^{-1/2} \log n), \tag{2.4}$$

$$\sup_{y\in I^2} |\alpha_n(y) - B_n(y)| \overset{\text{a.s.}}{=} O(n^{-1/2} \log^2 n), \tag{2.5}$$

$$\sup_{y\in I^d} |\alpha_n(y) - B_n(y)| \overset{\text{a.s.}}{=} O(n^{-1/2(d+1)}(\log n)^{3/2}), \quad d \geq 1. \tag{2.6}$$

Due to Bártfai (1966) and/or to the Erdős–Rényi (1970) theorem, the $O(\cdot)$ rate of convergence of (2.4) is best possible (cf. Komlós, Major, Tusnády, 1975a, or Theorem 4.4.2 in Csörgő and Révész, 1981a).

Next, on strong approximations of the *uniform empirical process* α_n of (1.7) or, equivalently, that of $\beta_n(L^{-1})$ of (1.9) in terms of a *single* Gaussian process, the Kiefer process $\{K(x, t); (x, t) \in I^d \times \mathbb{R}^1_+\}$, we have

THEOREM 2.2. *Let $X_1, \dots, X_n$ $(n = 1, 2, \dots)$ be independent random d-vectors, uniformly distributed on I^d, or with distribution function $F \in \mathscr{F}_0$ on $\mathbb{R}^d$. Let α_n be as in* (1.7) *or in* (1.9). *Then one can construct a probability space for $X_1, X_2, \dots$ with a Kiefer process $\{K(y, t); (y, t) \in I^d \times \mathbb{R}^1_+\}$ on it so that:*

(i) *for all n and x we have (cf. Komlós, Major and Tusnády,* 1975a)

$$P\{\sup_{1\leq k\leq n} \sup_{y\in I^1} |k^{1/2}\alpha_k(y) - K(y, k)| > (C \log n + x) \log n\} < Le^{-x\lambda}, \tag{2.7}$$

where C, L, λ are positive absolute constants,

(ii) *for any $\lambda > 0$ there exists a constant $C > 0$ such that for each n (cf. Csörgő and Révész,* 1975a)

$$P\{\sup_{1\leq k\leq n} \sup_{y\in I^d} |k^{1/2}\alpha_k(y) - K(y, k)| > Cn^{(d+1)/2(d+2)} \log^2 n\} \leq n^{-\lambda}, \quad d \geq 1. \tag{2.8}$$

COROLLARY 2.2. (2.7), (2.8) *in turn imply*

$$n^{-1/2} \sup_{1\leq k\leq n} \sup_{y\in I^1} |k^{1/2}\alpha^k(y) - K(y, k)| \overset{\text{a.s.}}{=} O(\log^2 n/n^{1/2}), \tag{2.9}$$

$$n^{-1/2} \sup_{1\le k\le n} \sup_{y\in I^d} |k^{1/2}\alpha_k(y) - K(y,k)| \overset{\text{a.s.}}{=} O(n^{-1/(2d+4)} \log^2 n), \quad d \geq 1 . \tag{2.10}$$

The first result of the type of (2.4) is due to Brillinger (1969) with the a.s. rate of convergence $O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4})$. Kiefer (1969) was first to call attention to the desirability of viewing the one dimensional (in y) empirical process $\alpha_n(y)$ as a two-time parameter stochastic process in y and n and that it should be a.s. approximated in terms of an appropriate two-time parameter Gaussian process. Müller (1970) introduced $\{K(y,t); (y,t) \in I^1 \times \mathbb{R}^1_+\}$ and proved a corresponding two dimensional weak convergence of $\{\alpha_n(y); y \in I^1, n = 1, 2, \ldots\}$ to the latter stochastic process. Kiefer (1972) gave the first strong approximation solution of the type (2.9) with the a.s. rate of convergence $O(n^{-1/6}(\log n)^{2/3})$.

Both Corollaries 2.1 and 2.2 imply the weak convergence of α_n to a Brownian bridge B on the Skorohod space $D[0,1]^d$. Writing $k = [ns]$ ($s \in [0,1]$), $[ns]^{1/2}\alpha_{[ns]}(y)/n^{1/2}$ is a random element in $D[0,1]^{d+1}$ for each integer n, and (2.10) implies also

Corollary 2.3. $[n\cdot]^{1/2}\alpha_{[n\cdot]}(\cdot)/n^{1/2} \xrightarrow{D} K(\cdot,\cdot)$ *on* $D[0,1]^{d+1}$.

For $d = 1$ the latter result is essentially the above mentioned result of Müller (1970) (cf. also discussion of the latter on page 217 in Gaenssler and Stute, 1979).

Corollary 2.2 also provides strong invariance principles, i.e. laws like the Glivenko–Cantelli theorem, LIL, etc. are inherited by $\alpha_n(y)$ from $K(y,n)$ or vice versa (cf., e.g., Section 5.1, Theorem S.5.1.1 in Csörgő and Révész, 1981a). The main difference between the said Corollaries is that such strong laws like the ones mentioned do not follow from Corollary 2.1. In the latter we have no information about the finite dimensional distributions in n of the sequence of Brownian bridges $\{B_n(y)\}_{n=1}^{\infty}$. On the other hand, the inequalities (2.1), (2.2) and (2.3) can be used to estimate the rates of convergence for the distributions of some functionals of α_n (to those of a Brownian bridge B), and those of the appropriate Prohorov distances of measures generated by the sequences of stochastic processes $\{\alpha_n(y); y \in I^d\}_{n=1}^{\infty}$ and $\{B_n(y); y \in I^d\}_{n=1}$ (cf., e.g., Komlós, Major and Tusnády, 1975b; Theorem 1.16 in M. Csörgő, 1981a; Theorem 2.3.1 in Gaenssler and Stute, 1979 and references therein; Borovkov, 1978; and review of the latter paper by M. Csörgő, 1981, MR 81j; 60044; the latter two to be interpreted in terms of $\{\alpha_n(y); y \in I^d\}_{n=1}^{\infty}$ and $\{B_n(y); y \in I^d\}_{n=1}^{\infty}$ instead of the there considered partial sum and Wiener processes).

When the distribution function F of (1.6) is not a product of its marginals for all $x \in \mathbb{R}^d$ ($d \geq 2$), then strong and weak approximations of β_n can be described in terms of the following *Gaussian processes associated with the distribution function* $F(x)$ ($x \in \mathbb{R}^d$, $d \geq 2$).

Brownian bridge B_F *associated with* F *on* $\mathbb{R}^d$ ($d \geq 2$): A separable d-parameter real valued Gaussian process with the following properties

$$EB_F(x) = 0, \qquad EB_F(x)B_F(y) = F(x \wedge y) - F(x)F(y),$$
$$\lim_{x_i\to-\infty} B_F(x_1, \ldots, x_d) = 0 \quad (i = 1, \ldots, d),$$
$$\lim_{(x_1,\ldots,x_d)\to(\infty,\ldots,\infty)} B_F(x_1, \ldots, x_d) = 0,$$

where for $x, y \in \mathbb{R}^d$ we write $x \wedge y = (x_1 \wedge y_1, \ldots, x_d \wedge y_d)$.

Kiefer process $K_F(\cdot,\cdot)$ *on* $\mathbb{R}^d \times [0,\infty)$ *associated with the distribution function* F *on* $\mathbb{R}^d$ ($d \geq 2$): A separable $(d+1)$-parameter real valued ($x \in \mathbb{R}^d$, $0 \leq t < \infty$) Gaussian process with the following properties:

$$K_F(x, 0) = 0,$$
$$\lim_{x_i\to-\infty} K_F(x_1, \ldots, x_d, t) = 0 \quad (i = 1, \ldots, d),$$
$$\lim_{(x_1,\ldots,x_d)\to(\infty,\ldots,\infty)} K_F(x_1, \ldots, x_d, t) = 0,$$
$$EK_F(x,t) = 0 \quad \text{and} \quad EK_F(x,t_1)K_F(y,t_2) = (t_1 \wedge t_2)(F(x \wedge y) - F(x)F(y))$$

for all $x, y \in \mathbb{R}^d$ and $t_1, t_2 \geq 0$.

A more tractable description of B_F and K_F can be given in terms of the mapping $L : \mathbb{R}^d \to I^d$ defined by

$$L(x) = L(x_1, \ldots, x_d) = (F_{(1)}(x_1), \ldots, F_{(d)}(x_d)) = (y_1, \ldots, y_d) \in I^d, \quad (x_1, \ldots, x_d) \in \mathbb{R}^d, \tag{2.11}$$

the inverse map of L^{-1} of (1.8), where, just as in the latter map, $y_i = F_{(i)}(x_i)$ ($i = 1, \ldots, d$) are the i-th marginals of F. It is well known that (cf. p. 293 in Wichura, 1973; or Lemma 1 in Philipp and Pinzur, 1980; or Lemma 3.2 in Moore and Spruill, 1975) there is a d-variate distribution function G on I^d with uniform marginals on $[0,1]$ such that

$$F(x) = G(L(x)), \qquad G(y) = F(L^{-1}(y)), \tag{2.12}$$

i.e., G on I^d has uniform marginals $y_i = F_{(i)}(x_i)$ ($i = 1, \ldots, d$) on $[0,1]$. (We note for example that if $F \in \mathcal{F}_0$ then $G(y) = \lambda(y)$ with $\lambda(\cdot)$ as in (1.7), i.e., *in the latter case* $G(y)$ *is the uniform distribution function on* I^d.)

Now consider the d-parameter Wiener process W_G associated with the distribution function G on I^d, defined as follows.

Wiener process W_G *on* I^d *associated with the distribution function* G *on* I^d ($d \geq 2$): A real valued d-parameter Gaussian process $\{W_G(y); y \in I^d\}$ with $EW_G(y) = 0$, $EW_G(x)W_G(y) = G(x \wedge y)$, and $W_G(y_1, \ldots, y_d) = 0$ whenever $y_i = 0$ ($i = 1, \ldots, d$).

Then the d-parameter Gaussian process

$$\{B_G(y);\ y\in I^d\}=\{W_G(y_1,\ldots,y_d)-G(y_1,\ldots,y_d)W_G(1,\ldots,1);\ y=(y_1,\ldots,y_d)\in I^d\} \tag{2.13}$$

is a *Brownian bridge process on I^d associated with G on I^d*, and *the Brownian bridge process B_F associated with F on $\mathbb{R}^d$ can be represented* via (2.12) and (2.13) *as*

$$\{B_F(x);\ x\in\mathbb{R}^d\}=\{W_G(L(x))-G(L(x))W_G(1,\ldots,1);\ x\in\mathbb{R}^d\}$$

and

$$\{B_F(L^{-1}(y));\ y\in I^d\}=\{B_G(y);\ y\in I^d\}. \tag{2.14}$$

Consider also the $(d+1)$-parameter Wiener process $W_G(\cdot,\cdot)$ on $I^d\times[0,\infty)$ associated with the distribution function G on I^d, defined as follows.

Wiener process $W_G(\cdot,\cdot)$ on $I^d\times[0,\infty]$ associated with the distribution function G on I^d $(d\geq 2)$: A real valued $(d+1)$-parameter Gaussian process $\{W_G(y,t);\ y\in I^d,\ t\geq 0\}$ with $W_G(y_1,\ldots,y_d,t)=0$ whenever any of $y_1,\ldots,y_d$ or t is zero, $EW_G(y,t)=0$, and $EW_G(y,t_1)W_G(x,t_2)=(t_1\wedge t_2)G(x\wedge y)$.

Then the $(d+1)$-parameter Gaussian process

$$\{K_G(y,t);\ y\in I^d,t\geq 0\}=\{W_G(y_1,\ldots,y_d,t)-G(y_1,\ldots,y_d)W_G(1,\ldots,1,t);\ y=(y_1,\ldots,y_d)\in I^d,t\geq 0\} \tag{2.15}$$

is a *Kiefer process on $I^d\times[0,\infty]$ associated with G on I^d*, and *the Kiefer process $K_F(\cdot,\cdot)$ associated with F on $\mathbb{R}^d$ can be represented* via (2.12) and (2.15) *as*

$$\{K_F(x,t);\ x\in\mathbb{R}^d,t\geq 0\}=\{W_G(L(x),t)-G(L(x))W_G(1,\ldots,1,t);\ x\in\mathbb{R}^d,\ r\geq 0\}$$

and

$$\{K_F(L^{-1}(y),t);\ y\in I^d,\ t\geq 0\}=\{K_G(y,t);\ y\in I^d,\ t\geq 0\}. \tag{2.16}$$

We note that if $F\in\mathcal{F}_0$ or $d=1$, then the latter Kiefer processes $K_G(\cdot,\cdot)$ and $K_F(\cdot,\cdot)$ coincide with our originally defined Kiefer process $K(\cdot,\cdot)$ on $I^d\times\mathbb{R}^1_+$. The same is true concerning our originally defined Brownian bridge B, and the Brownian bridges B_G and B_F in the context of $F\in\mathcal{F}_0$. Note also that in general β_n of (1.16) can be written as

$$\beta_n(L^{-1}(y))=n^{1/2}(F_n(L^{-1}(y))-F(L^{-1}(y))),\quad =n^{1/2}(F_n(L^{-1}(y))-G(y)),\quad y\in I^d, \tag{2.17}$$

which, in turn, reduces to the equality of (1.9) whenever $F\in\mathcal{F}_0$ or $d=1$.

As far as we know, the best available strong approximation of the empirical process β_n of (1.6) or, equivalently, that of $\beta_n(L^{-1})$ of (2.17) is (for more recent information we refer to Borisov, 1982).

THEOREM 2.3 (Philipp and Pinzur 1980). *Let $X_1,\ldots,X_n$ $(n=1,2,\ldots)$ be independent d-vectors on $\mathbb{R}^d$ with distribution function F. Let β_n be as in* (1.6) *or, equivalently, as in* (2.17). *Then one can construct a probability space for $X_1,X_2,\ldots$ with a Kiefer process $\{K_F(x,t);\ x\in\mathbb{R}^d,\ t\geq 0\}$ associated with F on it such that*

$$P\{\sup_{1\leq k\leq n}\sup_{x\in\mathbb{R}^d}|k^{1/2}\beta_k(x)-K_F(x,k)|>C_1n^{(1/2)-\lambda}\} = P\{\sup_{1\leq k\leq n}\sup_{y\in I^d}|k^{1/2}\beta_k(L^{-1}(y))-K_G(y,k)|>C_1n^{(1/2)-\lambda}\}\leq C_2n^{-(1+1/36)} \tag{2.18}$$

for $\lambda=1/(5000d^2)$, where C_1, C_2 are positive constants depending only on F and d.

COROLLARY 2.4. (2.18) *in turn implies*

$$n^{-1/2}\sup_{1\leq k\leq n}\sup_{x\in\mathbb{R}^d}|k^{1/2}\beta_k(x)-K_F(x,k)| = n^{-1/2}\sup_{1\leq k\leq n}\sup_{y\in I^d}|k^{1/2}\beta_k(L^{-1}(y))-K_G(y,k)|\overset{\text{a.s.}}{=}O(n^{-\lambda}) \tag{2.19}$$

with λ as in Theorem 2.3.

REMARK 2.1. We note that Philipp and Pinzur (1980) state only (2.19) in their Theorem 1. S. Csörgő (1981b) noted that going through their proof one can see that they had in fact proved the somewhat stronger (2.18).

REMARK 2.2. Theorem 2.3 and Corollary 2.4 are best available in the sense that in them there are no assumptions made on F. The a.s. rates of convergence of Corollary 2.2 are of course better than that (2.19), but in the former F is assumed to be uniformly distributed on I^d. While in case of $d=1$ the latter assumption is not a restriction (for $d=1$ and F continuous we have (1.9), and if F is arbitrary in the latter case, then (2.1) and (2.7) remain true (cf. Remark 1 in S. Csörgő, 1981a)), for $d\geq 2$ it is. Csörgő and Révész (1975b) actually proved (2.18) with the better rate $n^{(d+1)/(2d+4)}\log^2 n$ (cf. (2.8)) replacing its present rate of $n^{(1/2)-\lambda}$ $(\lambda=1/(5000d^2))$, but only for a class of d-variate distribution functions satisfying a rather strict regularity condition.

Clearly, for all fixed $t>0$,

$$\{t^{-1/2}K_F(L^{-1}(y),t);\ y\in I^d\}=\{t^{-1/2}K_G(y,t);\ y\in I^d\} \overset{\mathcal{D}}{=}\{B_F(L^{-1}(y));\ y\in I^d\}=\{B_G(y);\ y\in I^d\}. \tag{2.20}$$

Therefore Corollary 2.4 implies

COROLLARY 2.5. $\beta_n(L^{-1}(\cdot)) \xrightarrow{D} B_G(\cdot)$ *on* $D[0,1]^d$. (2.21)

Also on writing $k = [ns]$ $(s \in [0,1])$, $[ns]^{1/2}\beta_{[ns]}(L^{-1}(y))/n^{1/2}$ is a random element in $D[0,1]^{d+1}$ for each integer n, and (2.19) implies

COROLLARY 2.6. $[n\cdot]^{1/2}\beta_{[n\cdot]}(L^{-1}(\cdot))/n^{1/2} \xrightarrow{D} K_G(\cdot,\cdot)$ *on* $D[0,1]^{d+1}$. (2.22)

The result of (2.21) was first proved by Dudley (1966) (cf. also Neuhaus, 1971; Straf, 1971; Bickel and Wichura, 1971; Theorem 2.1.3 in Gaenssler and Stute, 1979, and discussion of the latter theorem therein). The result of (2.22) was first proved by Bickel and Wichura (1971) (cf. Theorems 2.1.4 and 2.1.5 in Gaenssler and Stute, 1979; see also Neuhaus and Sen, 1977).

A common property of the quoted results so far is that they are uniform approximations of measures over intervals only (of I^d or those of $\mathbb{R}^d$ mapped onto I^d). Concerning now the more general problem of approximating the empirical measure process β_n of (1.4) by appropriate Gaussian measure processes, the question is over how rich a class of subsets $\mathscr{C} \subset \mathscr{B}$ of $(R, \mathscr{B})$ (cf. paragraph one of Section 1) could we possibly have theorems like, for example, Theorems 2.1, 2.2 and 2.3. Let $(R, \mathscr{B}, \mu) = (I^d, \mathscr{B}, \lambda)$, λ the uniform Lebesgue measure on I^d, i.e., $X_1, \ldots, X_n$ $(n = 1, 2, \ldots)$ are distributed as in Theorems 2.1 and 2.2. In this case write $\alpha_n(B) = n^{1/2}(\lambda_n(B) - \lambda(B))$, $B \in \mathscr{B}$, instead of β_n of (1.4). Then, while it is true (cf. Philipp, 1973) that

$$\limsup_{n\to\infty} \sup_{B\in\mathscr{C}} (2n \log\log n)^{1/2} \alpha_n(B) \stackrel{\text{a.s.}}{=} \tfrac{1}{2}, \qquad (2.23)$$

where $\mathscr{C} = \mathscr{C}(2)$ is the class of convex sets of I^2, it is also known that the law of the iterated logarithm (2.23) fails for $\mathscr{C} = \mathscr{C}(d)$, the class of convex sets of I^d $(d \geq 3)$. The latter negative result for $d = 3$ was only recently proved by Dudley (1982) (for a discussion of previous results see e.g. Gaenssler and Stute 1979). In spite of (2.23) dimension two $(d = 2)$ is also critical, for Dudley (1982) showed also that if $\mathscr{C}$ is the collection of lower layers in I^2 (a lower layer in I^2 is a set B such that if $(x, y) \in B$, $u \leq x$ and $v \leq y$, then $(u, v) \in B$), then the law of the iterated logarithm (2.23) fails again (for previous mostly negative results and references along these lines for higher dimensions we refer to Stute, 1977; and Gaenssler and Stute, 1979). The same kind of negative results hold true concerning the problem of central limit theorem for $\alpha_n(B)$ (cf. Dudley, 1979), i.e., concerning empirical measure processes on I^d or $\mathbb{R}^d$, for the central limit theorem as well as law of the iterated logarithm the critical dimension is 2 for the lower layers and 3 for the convex sets. Hence any extension of results like those of Theorems 2.1, 2.2 and Corollaries 2.1, 2.2 in terms of uniform distances over a class of sets $\mathscr{C}$ of I^d, other than the intervals already considered, can only be true for somewhat restricted classes $\mathscr{C} \subset \mathscr{B}$. Révész (1976a,b) extended (2.6) and (2.10) over sets in I^d, defined by differentiability conditions. For example, instead of (2.6) we have (Révész, 1976a)

$$\sup_{B\in\mathscr{C}} |\alpha_n(B) - B_n(B)| \stackrel{\text{a.s.}}{=} O(n^{-1/19}) \qquad (2.24)$$

and, instead of (2.10) we can have (Révész, 1976a)

$$n^{-1/2} \sup_{B\in\mathscr{C}} |n^{1/2}\alpha_n(B) - K(B, n)| \stackrel{\text{a.s.}}{=} O(n^{-1/50}), \qquad (2.25)$$

where $\mathscr{C}$ is the class of those Borel sets of I^2 which have twice differentiable boundaries, and $\{B_n(B); B \in \mathscr{C}\}_{n=1}^{\infty}$ respectively $\{K(B, n); B \in \mathscr{C}, n \geq 1\}$ are Gaussian measure processes with mean zero and covariance function $EB_n(B)B_n(D) = \lambda(B \cap D) - \lambda(B)\lambda(D)$ $(n = 1, 2, \ldots)$ respectively $EK(B, n)K(D, m) = (n \wedge m)(\lambda(B \cap D) - \lambda(B)\lambda(D))$ for all $B, D \in \mathscr{C}$. Similar extensions hold true over sets in I^d $(d \geq 2)$ with differentiable boundaries (cf. Révész, 1976b; Ibero, 1979a,b).

A common feature of the results of Theorem 2.2, Corollary 2.2 and that of the first one of these types by Kiefer (1972) is not only that they improve (they imply for instance functional laws of iterated logarithm (cf., e.g., Section 5.1 in Csörgő and Révész, 1981a)) and are conceptually simpler than the original weak convergence result of Donsker (1952) on empirical distribution functions, but they also avoid the problem of measurability and topology caused by the fact that $D[0,1]^d$ endowed with the supremum norm is not separable (cf., e.g., Billingsley, 1968, p. 153). This idea of proving a.s. or in probability invariance principles à la Kiefer (1972) also works for distribution functions on $\mathbb{R}^d$ (cf. Theorem 2.3 and its predecessors by Révész (1976a, Theorem 3), and Csörgő and Révész (1975b, Theorem 1)) and, as we have just seen in (2.24) and (2.25), for uniform distances over sets of I^d, defined by differentiability conditions (see also Révész, 1976b; Ibero, 1979a,b). Recently Dudley and Philipp (1981) used the same idea to reformulate and strengthen the results of Dudley (1978, 1981a,b), Kuelbs (1976) on empirical measure processes while removing their previously assumed measurability conditions. They do this by proving invariance principles for sums of not necessarily measurable random elements with values in a not necessarily separable Banach space and by showing that empirical measure processes fit easily into the latter setup. We refer for example to Theorems 1.5 and 7.1 in Dudley and Philipp (1981) which can be viewed as far reaching generalizations (with slower but adequate rates of convergence) of Theorems 2.2, 2.3 and their Corollaries 2.2, 2.4, and also that of (2.25), in terms of Kiefer Measures $\{K_\mu(B, n); B \in \mathscr{C}, n \geq 1\}$ associated with probability measures μ on $(R, \mathscr{B})$ over a subclass $\mathscr{C}$ (of some generality) of $\mathscr{B}$.

The strong and weak approximations of the empirical characteristic function C_n of (1.23) can be accomplished, on $\mathbb{R}^1$ in terms of Gaussian processes built on Kiefer and Brownian bridge processes (cf. S. Csörgő, 1981a) and on $\mathbb{R}^d$ $(d \geq 2)$ in terms of Gaussian processes built on Kiefer and Brownian bridge processes associated with F on $\mathbb{R}^d$ (cf. S. Csörgő, 1981b). For further references we refer to the just mentioned two papers of S. Csörgő.

3. On the limiting distribution of and critical values for the multivariate Cramér–von Mises and Hoeffding–Blum–Kiefer–Rosenblatt independence criteria

A study of empirical and quantile processes on $\mathbb{R}^1$ with the help of strong approximation methodologies is given in Chapters 4 and 5 of Csörgő and Révész (1981a). We are also going to touch upon some of these problems in the light of some recent developments in Section 4. of this exposition. An excellent *direct* theoretical and statistical study of the empirical process on $\mathbb{R}^1$ can be seen in this volume by Csáki (1982b). The latter is recommended as parallel reading to the material covered in this paper. In this section we add details to Theorems 2.1, 2.2 and Corollaries 2.1, 2.2 while studying Cramér–von Mises functionals of the empirical processes α_n (cf. (1.7) and (1.9)) and $t_n = T_n(L^{-1})$ (cf. (1.13)).

When testing the null hypothesis H_0 of (1.10), or that of F being a completely specified continuous distribution function on $\mathbb{R}^1$, one of the frequently used statistics is the Cramér–von Mises statistic $\omega_{n,d}^2$, defined by

$$\omega_{n,d}^2 = \int_{R^d} \beta_n^2(x) \prod_{i=1}^{d} \mathrm{d}F_{(i)}(x_i) = \int_{I^d} \alpha_n^2(y) \prod_{i=1}^{d} \mathrm{d}y_i$$
$$= n^{-1} \sum_{k=1,j=1}^{n} \left\{ \prod_{i=1}^{d} (1-(y_{ki} \vee y_{ji})) - \prod_{i=1}^{d} 2^{-1}(1-y_{ki}^2) - \prod_{i=1}^{d} 2^{-1}(1-y_{ji}^2) + 3^{-d} \right\}, \tag{3.1}$$

$d \geq 1$, where $(y_{j1}, \ldots, y_{jd})_{j=1}^n$ with $y_{ji} = F_{(i)}(X_{ji})$ $(i = 1, \ldots, d)$ are the observed values of the random sample $X_j = (X_{j1}, \ldots, X_{jd})$, $j = 1, \ldots, n$. One rejects H_0 of (1.10), or that of F being a given continuous distribution function on $\mathbb{R}^1$, if for a given random sample $X_1, \ldots, X_n$ on F the computed value of $\omega_{n,d}^2$ is too large for a given level of significance (fixed size Type I error). Naturally, in order to be able to compute the value of $\omega_{n,d}^2$ for a sample, H_0 of (1.10), i.e. the marginals of F, or F itself on $\mathbb{R}^1$, will have to be completely specified (simple statistical hypothesis). While it is true that the distribution of $\omega_{n,d}^2$ will not depend on the specific form of these marginals (cf. (1.9)), the problem of finding and tabulating this distribution is not an easy task at all.

Let $V_{n,d}$ be the distribution function of the rv $\omega_{n,d}^2$, i.e.,

$$V_{n,d}(x) = P\{\omega_{n,d}^2 \leq x\}, \quad 0 < x < \infty. \tag{3.2}$$

Csörgő and Stachó (1979) gave a recursion formula for the exact distribution function $V_{n,1}$ of the rv $\omega_{n,1}^2$. The latter in principle is applicable to tabulating $V_{n,1}$ exactly for any given n. Naturally, much work has already been done to compile tables for $V_{n,1}$. A survey and comparison of these can be found in Knott (1974), whose results prove to be the most accurate so far. All these results and tables are based on some kind of an approximation of $V_{n,1}$. As to higher

dimensions $d \geq 2$, no analytic results appear to be known about the exact distribution function $V_{n,d}$. It follows by (2.6) that we have

$$\lim_{n\to\infty} V_{n,d}(x) = P\{\omega_d^2 \leq x\} := V_d(x), \quad 0 < x < \infty, \ d \geq 1, \tag{3.3}$$

where $\omega_d^2 = \int_{I^d} B^2(y)\,\mathrm{d}y$, $\{B(y); y \in I^d\}$ a Brownian bridge, and $\mathrm{d}y = \prod_{i=1}^d \mathrm{d}y_i$ from now on.

For the sake of describing the speed of convergence of the distribution functions $\{V_{n,d}\}_{n=1}^\infty$ to the distribution function V_d of ω_d^2 (cf. (3.3)) we define

$$\Delta_{n,d} = \sup_{0<x<\infty} |V_{n,d}(x) - V_d(x)|. \tag{3.4}$$

S. Csörgő (1976) showed that $\Delta_{n,1} = \mathrm{O}(n^{-1/2} \log n)$ and, on the basis of his complete asymptotic expansion for the Laplace transform of the rv $\omega_{n,1}^2$ (cf. (3.1)), he conjectured that $\Delta_{n,1}$ is of order $1/n$. Indeed, the latter turned out to be correct (cf. Corollary 1: $\Delta_{n,1} = \mathrm{O}(n^{-1})$, in Cotterill and M. Csörgő, 1982), and it can be deduced from the ground breaking work of Götze (1979). Actually the latter work when combined with Dugue (1969), and Bhattacharya and Ghosh (1978) implies (cf. Section 2 in Cotterill and M. Csörgő, 1982) an asymptotic expansion of arbitrary order for the distribution function $V_{n,d}$ of (3.2) and also that

$$\Delta_{n,d} = \mathrm{O}(n^{-1}), \quad d \geq 1, \tag{3.5}$$

(cf. Corollary 3 in Cotterill and M. Csörgő, 1982).

An extensive tabulation of the distribution function V_1 (cf. (3.3)) can be found in the monograph of Martynov (1978), where the theory and applications of a wide range of univariate Cramér–von Mises types statistics are also surveyed.

There appear to be no tables available for the distribution function $V_{n,d}$ $(d \geq 2)$ (cf. (3.2)). Hence, and in the light of the just quoted result of (3.5), tables for the distribution function V_d $(d \geq 2)$ of (3.3) are of special interest. Durbin (1970) tabulated V_d for $d = 2$, and Krivyakova, Martynov and Tyurin (1977) for $d = 3$. Using the characteristic function of the distribution function V_d (cf. Dugue, 1969; Durbin, 1970), Cotterill and M. Csörgő (1982) obtained a recursive equation for the cumulants of the rv ω_d^2 and, using the first six of these cumulants in the Cornish–Fisher asymptotic expansion, tabulated its critical values for $d = 2, 3, \ldots, 50$ at various levels of significance. These critical values are within 3% of Durbin's values for $d = 2$ and those of Krivyakova, Martynov and Tyrin for $d = 3$. We note also that errors in the said tables for higher dimensions should be further reduced due to the fact that cumulants of ω_d^2 are $\mathrm{O}(e^{-d})$ (cf. Corollary 7 and Remark 3.2 in D. S. Cotterill and M. Csörgő, 1982). As far as we know, for the present there exist no further tables of V_d for $d \geq 4$.

As mentioned already, for the sake of computing the value of $\omega^2_{n,d}$ for a sample, H_0 of (1.10) will have to be completely specified. An alternate route to testing for H_0 of (1.10) can be based on the empirical process $t_n = T_n(L^{-1})$ of (1.13) which will not require the specification of the marginals of F under H_0, i.e., it will work also when H_0 of (1.10) is a *composite* statistical hypothesis.

For the sake of describing the latter approach due to Hoeffding (1948), and Blum, Kiefer and Rosenblatt (1961), we define the sequence of Gaussian processes $\{T^{(n)}(y); y \in I^d\}_{n=1}$ by

$$\{T^{(n)}(y); y \in I^d\} = \Big\{B_n(y) - \sum_{i=1}^{d} B_n(1, \ldots, 1, y_i, 1, \ldots, 1) \prod_{j\neq i} y_j;\ y = (y_1, \ldots, y_d) \in I^d (d \geq 2)\Big\} \tag{3.6}$$

where $\{B_n(y); y \in I^d(d\geq 2)\}_{n=1}^{\infty}$ is a sequence of Brownian bridges.

Define also the Gaussian process $\{T(y, t); y \in I^d, t \geq 0\}$ by

$$\{T(y, t); y \in I^d, t \geq 0\} = \Big\{K(y, t) - \sum_{i=1}^{d} K(1, \ldots, 1, y_i, 1, \ldots, 1, t) \times \prod_{j\neq i} y_j;\ y \in I^d(d\geq 2), t \geq 0\Big\}, \tag{3.7}$$

where $\{K(y, t); y \in I^d(d\geq 2), t \geq 0\}$ is a Kiefer process.

Obviously $ET^{(n)}(y) = ET(y, t) = 0$, and simple but somewhat tedious calculations yield the covariance functions

$$ET^{(n)}(x)T^{(n)}(y) = \prod_{i=1}^{d} (x_i \wedge y_i) + (d-1) \prod_{i=1}^{d} x_i y_i - \sum_{i=1}^{d} (x_i \wedge y_i) \prod_{j\neq i} x_j y_j := \rho(x, y) \quad \text{for all } n, \tag{3.8}$$

and

$$ET(x, s)T(y, t) = (s \wedge t)\rho(x, y), \tag{3.9}$$

where $x = (x_1, \ldots, x_d)$, $y = (y_1, \ldots, y_d) \in I^d$ $(d \geq 2)$ and $s, t \geq 0$.

Strong approximations of $t_n = T_n(L^{-1})$ in terms of the latter Gaussian processes follow quite directly by Theorems 2.1 and 2.2. The following results are known (cf. Theorems 3 and 4 in M. Csörgő, 1979).

THEOREM 3.1. (Csörgő, 1979). *Let $X_1, \ldots, X_n$ $(n = 1, 2, \ldots)$ be independent random d-vectors on $\mathbb{R}^d$ with distribution function $F \in \mathscr{F}_0$ and let t_n be as in (1.13). Then one can construct a probability space for $X_1, X_2, \ldots$ with a sequence of Gaussian processes $\{T^n(y);\ y \in I^d\ (d\geq 2)\}_{n=1}^{\infty}$, defined as in (3.6), and a Gaussian process $\{T(y, t); y \in I^d\ (d \geq 2), t \geq 0\}$, defined as in (3.7), on it so that*

(i) *for all n and x we have*

$$P\{\sup_{y\in I^2} |t_n(y) - T^{(n)}(y)| > n^{-1/2}(C \log n + x) \log n\} \leq L\, e^{-\lambda x} \tag{3.10}$$

where C, L and λ are positive absolute constants,

(ii) *for any $\lambda > 0$ there exists a constant $C > 0$ such that*

$$P\{\sup_{y\in I^d} |t_n(y) - T^{(n)}(y)| > C(\log n)^{3/2} n^{-1/2(d+1)}\} \leq n^{-\lambda}, \quad d \geq 2, \tag{3.11}$$

and

$$P\{\sup_{1\leq k<n}\ \sup_{y\in I^d} |k^{1/2}t_k(y) - T(y, k)| > Cn^{(d+1)/2(d+2)} \log^2 n\} \leq n^{-\lambda}, \quad d \geq 2. \tag{3.12}$$

COROLLARY 3.1. (3.10), (3.11), (3.12) *in turn imply*

$$\sup_{y\in I^2} |t_n(y) - T^{(n)}(y)| \overset{\text{a.s.}}{=} O(n^{-1/2} \log^2 n), \tag{3.13}$$

$$\sup_{y\in I^d} |t_n(y) - T^{(n)}(y)| \overset{\text{a.s.}}{=} O(n^{-1/2(d+1)}(\log n)^{3/2}), \quad d \geq 2, \tag{3.14}$$

$$n^{-1/2} \sup_{1\leq k\leq n}\ \sup_{y\in I^d} |k^{1/2}t_k(y) - T(y, k)| \overset{\text{a.s.}}{=} O(n^{-1/(2d+4)} \log^2 n), \quad d \geq 2. \tag{3.15}$$

It follows from (3.6) and (3.7), or by (3.8) and (3.9), that for each n

$$\{T^{(n)}(y); y \in I^d (d \geq 2)\} \overset{\mathcal{D}}{=} \{n^{-1/2}T(y, n); y \in I^d\ (d \geq 2)\} \overset{\mathcal{D}}{=} \{T(y, 1); Y \in I^d\ (d \geq 2)\}. \tag{3.16}$$

Define the Gaussian process $\{T(y); y \in I^d\ (d \geq)\}$ by

$$\{T(y); y \in I^d\ (d \geq 2)\} = \{T(y, t); y \in I^d\ (d \geq 2), t = 1\} \overset{\mathcal{D}}{=} \Big\{B(y) - \sum_{i=1}^{d} B(1, \ldots, 1, y_i, 1, \ldots, 1) \prod_{j\neq i} y_j;\ y = (y_1, \ldots, y_d) \in I^d\ (d \geq 2)\Big\}, \tag{3.17}$$

where $\{B(y); y \in I^d\ (d \geq 2)\}$ is a Brownian bridge. Thus $T(\cdot)$ has mean zero and covariance function $\rho(\cdot, \cdot)$ of (3.8), and weak convergence of t_n to the Gaussian process T of (3.17) on the Skorohod space $D[0, 1]^d$ follows by (3.14) say. Also, a Corollary 2.3 type weak convergence of $[n\cdot]^{1/2}t_{[n\cdot]}(\cdot)/n^{1/2}$ to $T(\cdot, \cdot)$ of (3.7) on $D[0, 1]^{d+1}$ follows by (3.15).

Blum, Kiefer and Rosenblatt (1961) proposed the following Cramér–von Mises type test statistic for H_0 of (1.10):

$$C_{n,d} = \int_{\mathbb{R}^d} T_n^2(x) \prod_{i=1}^{d} dF_{(i)}(x_i) = \int_{I^d} t_n^2(y)\, dy, \quad d \geq 2. \tag{3.18}$$

One rejects H_0 of (1.10) if for a given random sample $X_1, \ldots, X_n$ on F the computed value of $C_{n,d}$ is too large for a given level of significance.

Let $\Gamma_{n,d}$ be the distribution function of the rv $C_{n,d}$, i.e.,

$$\Gamma_{n,d}(x) = P\{C_{n,d} \leq x\}, \quad 0 < x < \infty, \quad d \geq 2. \tag{3.19}$$

Then, by (3.14) say, we have

$$\lim_{n\to\infty} \Gamma_{n,d}(x) = P\{C_d \leq x\} := \Gamma_d(x), \quad 0 < x < \infty, \quad d \geq 2, \tag{3.20}$$

where $C_d = \int_{I^d} T^2(y)\,dy$ with $\{T(y); y \in I^d \ (d \geq 2)\}$ as in (3.17).

There does not seem to be anything known about the exact distribution function $\Gamma_{n,d}$ of the rv $C_{n,d}$. As to the speed of convergence in (3.20) via (3.10) and (3.11) we get (cf. Theorem 1 in Cotterill and Csörgő, 1980)

$$\nabla_{n,d} := \sup_{0<x<\infty} |\Gamma_{n,d}(x) - \Gamma_d(x)| = \begin{cases} O(n^{-1/2}\log^2 n) & \text{if } d = 2, \\ O(n^{-1/(2d+2)}(\log n)^{3/2}) & \text{if } d \geq 3. \end{cases} \tag{3.21}$$

As far as we know the rates of convergence in (3.21) are the only ones available so far.

Concerning tables for the distribution function Γ_d, for $d = 2$, Blum, Kiefer and Rosenblatt (1961) obtained the characteristic function of the distribution function Γ_d of the rv C_d and tabulated its distribution via numerical inversion of the said characteristic function. The statistic $C_{n,d}$ of (3.18) itself cannot be computed unless $F \in \mathscr{F}_0$ of H_0 of (1.10) is completely specified. Hoeffding (1948), and Blum, Kiefer and Rosenblatt (1961) suggested, as critical region for H_0 of (1.10) when it is viewed as a *composite* statistical hypothesis, large values of

$$\hat{C}_{n,d} = \int_{\mathbb{R}^d} T_n^2(x)\,dF_n(x), \quad d \geq 2, \tag{3.22}$$

or those of

$$\tilde{C}_{n,d} = \int_{\mathbb{R}^d} T_n^2(x) \prod_{i=1}^{d} dF_{ni}(x_i), \quad d \geq 2. \tag{3.23}$$

These two statistics are equivalent to $C_{n,d}$ in that both converge in distribution to the rv C_d. This was already noted by Blum, Kiefer and Rosenblatt (1961), and for a detailed proof of this statement we refer to Section 4 in Cotterill and Csörgő (1980). Recently DeWet (1980) studied a version of (3.23) in the case of $d = 2$ with some nonnegative weight functions multiplying the integrand T_n^2 of $\tilde{C}_{n,d}$. Koziol and Nemec (1979) studied $\hat{C}_{n,d}$ of (3.22) and its

performance (power properties) in testing for independence with bivariate normal observations.

As to tables for the distribution function Γ_d for $d \geq 2$, Cotterill and Csörgő (1980, Section 4) find an expression for the characteristic function of the rv C_d, $d \geq 2$, via utilizing the representation of the stochastic process $\{T(y); y \in I^d \ (d \geq 2)\}$ of (3.17) in terms of Brownian bridges. This in turn enables them to find the first five cumulants of the rv C_d, and using these in the Cornish–Fisher asymptotic expansion, they tabulated its critical values for $d = 2, \ldots, 20$ at the 'usual' levels of significance. These tables and details as to how to calculate approximate critical values of the rv C_d for all $d \geq 2$ are given in Sections 5 and 6 of the said paper. Compared with the figures of Blum, Kiefer and Rosenblatt (1961) for $d = 2$, the Cornish–Fisher approximation seems to work quite well. For $d > 2$ we do not know of any other tables for the rv C_d.

Another approach to this problem was suggested by Deheuvels (1981), who showed that the Gaussian process $\{T(y); y \in I^d \ (d \geq 2)\}$ of (3.17) which approximates the empirical process t_n of (1.13) (cf. Theorem 3.1) can be decomposed into $2^d - d - 1$ independent Gaussian processes whose covariance functions are of the same structure for all $d \geq 2$ as that of $T(y)$ for $d = 2$. If tables for the Cramér–von Mises functionals of these $2^d - d - 1$ independent rv were available, then one could test asymptotically independently whether there are dependence relationships within each subset of the coordinates of $X \in \mathbb{R}^d$, $d \geq 2$.

4. On strong and weak approximations of the quantile process

In this section we are going to give an up-to-date summary of strong and weak invariance principles for the quantile process ρ_n of (1.20). For further readings, references on this subject and its applications to statistics we refer to Doksum (1974), Doksum and Sievers (1976), Doksum, Fenstad and Aaberge (1977), Parzen (1979, 1980), Chapters 4, 5 in Csörgő and Révész (1981a), Csörgő and Révész (1981b), M. Csörgő (1981b, 1983), and Csörgő, Csörgő, Horváth and Révész (1982). Random variables are $\mathbb{R}^1$-valued throughout this section. We start with comparing the general quantile process ρ_n of (1.20) to its corresponding uniform version, the uniform quantile process u_n of (1.21) (cf. also (1.22), and (1.14)–(1.19) for definitions used in this section).

THEOREM 4.1 (Csörgő and Révész, 1978). *Let $X_1, X_2, \ldots$ be i.i.d. rv with a continuous distribution function F and assume*

(i) *is twice differentiable on (a, b), where $a = \sup\{x: F(x) = 0\}$, $b = \inf\{x: F(x) = 1\}$, $-\infty \leq a < b \leq +\infty$,*

(ii) *$F'(x) = f(x) > 0$ on (a, b),*

(iii) *for some $\gamma > 0$ we have*

$$\sup_{0<y<1} y(1-y)\frac{|f'(Q(y))|}{f^2(Q(y))} \leq \gamma.$$

Then, with $\delta_n = 25n^{-1}\log\log n$,

$$\sup_{\delta_n \le y \le 1-\delta_n} |\rho_n(y) - u_n(y)| \overset{\text{a.s.}}{=} O(n^{-1/2}\log\log n). \tag{4.1}$$

If, in addition to (i), (ii) *and* (iii), *we also assume*

(iv) $A = \lim_{x\downarrow a} f(x) < \infty$, $B = \lim_{x\uparrow b} f(x) < \infty$,

(v) *one of* (v, α) $A \wedge B > 0$, (v, β) *if* $A = 0$ (*resp.* $B = 0$) *then f is nondecreasing* (*resp. nonincreasing*) *on an interval to the right of a* (*resp. to the left of b*), *then, if* (v, α) *obtains,*

$$\sup_{0\le y\le 1} |\rho_n(y) - u_n(y)| \overset{\text{a.s.}}{=} O(n^{-1/2}\log\log n), \tag{4.2}$$

and if (v, β) *obtains,*

$$\sup_{0\le y\le 1} |\rho_n(y) - u_n(y)| \overset{\text{a.s.}}{=} \begin{cases} O(n^{-1/2}\log\log n) & \text{if } \gamma < 1, \\ O(n^{-1/2}(\log\log n)^2) & \text{if } \gamma = 1, \\ O(n^{-1/2}(\log\log n)^{\gamma}(\log n)^{(1+\varepsilon)(\gamma-1)}) & \text{if } \gamma > 1, \end{cases} \tag{4.3}$$

where $\varepsilon > 0$ *is arbitrary, and* γ *is as in* (iii).

The above theorem also implies approximations of ρ_n in terms of appropriate sequences of Brownian bridges $\{B_n(y), 0 \le y \le 1\}$ (cf. Csörgő and Révész, 1978; Section 3.1 in M. Csörgő, 1983) due to

Theorem 4.2 (Csörgő and Révész, 1975c, 1978). *For an i.i.d. sequence of rv* $X_1, X_2, \ldots$ *there exists a probability space with a sequence of Brownian bridges* $\{B_n\}$ *on it such that*

$$\sup_{0\le y\le 1} |u_n(y) - B_n(y)| \overset{\text{a.s.}}{=} O(n^{-1/2}\log n). \tag{4.4}$$

Naturally, from the above two theorems it follows that

$$\rho_n(\cdot) \overset{D}{\to} B(\cdot) \tag{4.5}$$

in Skorohod's space $D[0, 1]$.

Let $q(y) \ge 0$ be a continuous function on $[0, 1]$ which is strictly positive on $(0, 1)$, nondecreasing on $[0, \frac{1}{2}]$, and symmetric about $y = \frac{1}{2}$, and let

$$h(y) = \left(y(1-y)\log\log\frac{1}{y(1-y)}\right)^{1/2}, \quad 0 \le y \le 1. \tag{4.6}$$

Define also $g(y) = q(y)/h(y)$ so that

$$g(y) = q(y)/h(y) \to \infty \quad \text{as } y \to 0. \tag{4.7}$$

Then $g(y) = g(1-y)$ by definition.

Shorack (1979) showed that with q and g as above, the condition (4.7) is sufficient for O'Reilly's (1974) sufficient condition on q for

$$\sup_{1/(n+1)\le y\le n/(n+1)} |(u_n(y) - B_n(y))/q(y)| \overset{P}{\to} 0 \tag{4.8}$$

to be true with the Brownian bridges B_n of (4.4). For an up-to-date discussion of O'Reilly's (1974) theorems in the light of strong approximations we refer to Chapter V in M. Csörgő (1983). Similar work to that of O'Reilly's was done earlier by Chibisov (1964), and Pyke and Shorack (1968). S. Csörgő (1982) showed that on assuming conditions (i), (ii), (iii), (iv) and (v) of Theorem 4.1 and the condition (4.7), we have

$$\sup_{1/(n+1)\le y\le n/(n+1)} |(\rho_n(y) - B_n(y))/q(y)| \overset{P}{\to} 0 \tag{4.9}$$

with B_n as in (4.4), provided that γ of condition (iii) is less than 1. Otherwise, i.e., if in (iii) $\gamma \ge 1$, growth conditions had to be introduced for the function $G(t) = \inf\{g(y): 0 \le y \le t\} \uparrow \infty$ as $t \downarrow 0$. M. Csörgő (1983, Theorem 5.1.1) verified (4.9) only under the conditions (i), (ii), (iii) of Theorem 4.1 and (4.7). Stute (1982) proved (4.9) with $q \equiv 1$ under (4.7) and only the conditions (iv), (v) of Theorem 4.1 as follows: we have (4.9) with $q \equiv 1$ if (v, α) obtains, or if (v, β) obtains, provided that in the latter case both

$$g(y)f(Q(y/\lambda))/f(Q(y)) \to \infty \quad \text{as } y \to 0 \tag{4.10}$$

and (symmetrically)

$$g(y)f(Q(1-y)/\lambda)/f(Q(1-y)) \to \infty \quad \text{as } y \to 0 \tag{4.11}$$

for each $\lambda \ge 1$ (note that in Stute's (1982) case $g = 1/h$ on account of $q \equiv 1$). Shorack (1982) announced the latter result with q and g as in (4.7) (for a proof we may, for example, refer to M. Csörgő, 1983, Corollary 5.3.2). All the afore-mentioned results concerning (4.9) are contained in

Theorem 4.3 (M. Csörgő, 1983, Theorem 5.3.1). *Let a, b be as in Theorem* 4.1 *and assume that F has a continuous density function* $F' = f$ *that is positive on* (a, b), *the support of F. Let q be any given O'Reilly weight function with* (4.7). *Then, as* $n \to \infty$, *with the sequence of Brownian bridges* $\{B_n\}$ *of* (4.4) *we have* (4.9) *under* (4.7), *provided that with g of the latter the following assumption also holds true*:

For any given $0 < \eta < 1$ *and* $\varepsilon > 0$ *there exist* $0 < c < 1$ *and* n_0 *such that*

$$P\Big\{\sup_{1/(n+1)\le y\le c}\ \sup_{\theta_{y,n}} \frac{f(Q(y))}{f(Q(\theta_{y,n}))}\frac{1}{g(y)} > \varepsilon\Big\} \le \eta$$

and similarly

$$P\Big\{\sup_{1-c\le y\le n/(n+1)}\ \sup_{\theta_{y,n}} \frac{f(Q(y))}{f(Q(\theta_{y,n}))}\frac{1}{g(y)} > \varepsilon\Big\} \le \eta \tag{4.12}$$

for all $n \ge n_0$, *where* $U_n(y) \wedge y < \theta_{y,n} < U_n(y) \vee y$.

All the afore quoted results concerning (4.9) can be put in terms of weak convergence on $D[0, 1]$, *provided* we redefine u_n (and hence also ρ_n) to be equal to zero on $[0, 1/(n+1))$ and $(n/(n+1), 1]$.

The sufficient conditions of Theorem 4.3 (cf. (4.12)) for the weak approximation of ρ_n on $[1/(n+1), n/(n+1)]$ are nearly necessary as well. For convenience, a weight function q will be called an O'Reilly weight function from now on if $g = q/h$ satisfies (4.7). We have

Theorem 4.4 (Csörgő, Csörgő, Horváth and Révész, 1982). *Let a, b be as in Theorem 4.1 and assume that F has a continuous density function $F' = f$ that is positive on (a, b), the support of F. If for any given g of an O'Reilly weight function q we have that the rv*

$$\Big(f\Big(Q\Big(\frac{1}{n}\Big)\Big)\Big/f(Q(\theta_{1/n,n}))\Big)\Big/g\Big(\frac{1}{n}\Big)(\log\log n)^{1/2},$$

or (4.13)

$$\Big(f\Big(Q\Big(\frac{n-1}{n}\Big)\Big)\Big/f(Q(\theta_{(n-1)/n,n}))\Big)\Big/g\Big(\frac{1}{n}\Big)(\log\log n)^{1/2}$$

is, contrary to (4.12), bounded away from zero in probability as $n\to\infty$, then the rv of (4.9) is also bounded away from zero in probability for any sequence of Brownian bridges $\{B_n\}$.

As to the problem of weak convergence of ρ_n/q in sup-norm metric over $[0, 1]$, we first observe that for a Brownian bridge B

$$\sup_{0<y\le 1/n} |B(y)/q(y)| \xrightarrow{P} 0 \quad\text{and}\quad \sup_{1-1/n\le y<1} |B(y)/q(y)| \xrightarrow{P} 0$$

as $n\to\infty$ for any O'Reilly weight function q. Hence the only way for ρ_n/q to converge weakly to B/q over $[0, 1]$ in sup-norm metric is that we have the latter two statements holding true also with ρ_n replacing B in them. In this context we have

Theorem 4.5 (Csörgő, Csörgő, Horváth and Révész, 1982). *Let a, b be as in Theorem 4.1 and assume that F has a continuous density function $F' = f$ that is positive on (a, b), the support of F. Let q be any given O'Reilly weight function so that as $n\to\infty$*

$$\sup_{0\le y\le 1/n}\ \sup_{\theta_{y,n}} \frac{n^{-1/2}}{q(y)}\frac{f(Q(y))}{f(Q(\theta_{y,n}))} \xrightarrow{P} 0,$$
$$\sup_{(n-1)/n\le y\le 1}\ \sup_{\theta_{1,n}} \frac{n^{-1/2}}{q(y)}\frac{f(Q(y))}{f(Q(\theta_{y,n}))} \xrightarrow{P} 0, \tag{4.14}$$

where $\theta_{y,n}$ is as in (4.12). Then, as $n\to\infty$,

$$\sup_{0\le y\le 1/n} |\rho_n(y)/q(y)| \xrightarrow{P} 0, \qquad \sup_{(n-1)/n\le y\le 1} |\rho_n(y)/q(y)| \xrightarrow{P} 0. \tag{4.15}$$

If, on the other hand, for the given q and g there exists a sequence of rv $\tau_n \le 1/n$ so that the rv

$$n^{-1/2}(f(Q(\tau_n))/f(Q(\theta_{\tau_n,n})))/q(\tau_n), \quad or$$
$$n^{-1/2}(f(Q(1-\tau_n))/f(Q(\theta_{1-\tau_n,n})))/q(\tau_n) \tag{4.16}$$

with $\tau_n \wedge U_{1:n} \le \theta_{\tau_n,n} \le \tau_n \vee U_{1:n}$ and $(1-\tau_n)\wedge U_{n:n} \le \theta_{1-\tau_n,n} \le (1-\tau_n)\vee U_{n:n}$ is bounded away from zero in probability as $n\to\infty$, then so will be also the rv of (4.15).

We now mention some *implications* (examples) of interest which follow from Theorems 4.3, 4.4 and 4.5:

(1) We have already noted that Theorem 4.3 implies the Stute (1982)–Shorack (1982) result: (4.9) holds true under the conditions (iv) and (v) of Theorem 4.1 if (v, α) obtains, or if (v, β) obtains and (4.10), (4.11) are also assumed. If (v, α) obtains then F has finite support and hence (4.9) holds immediately. If (v, β) obtains and, say, A of (iv) is zero, then for small enough c we have

$$\sup_{1/(n+1)\le y\le c}\ \sup_{\theta_{y,n}} \frac{f(Q(y))}{f(Q(\theta_{y,n}))}\frac{1}{g(y)} \le \sup_{1/(n+1)\le y\le c} \frac{f(Q(y))}{f(Q(y/\lambda))}\frac{1}{g(y)}$$

for some $\lambda \ge 1$ and all $n \ge 1$ with probability arbitrarily near to one by Remark 1 in Wellner (1978). Hence (4.17) implies the first condition of (4.12) and the Stute (1982)–Shorack (1982) theorem follows from Theorem 4.3.

(2) (Taken from Csörgő, Csörgő, Horváth and Révész, 1982). For any O'Reilly weight function q such that $q(y)\to 0$ we have $\sup_{0<y\le 1/n} 1/q(y) = \infty$. Hence in case of the uniform quantile process u_n we can choose $\{\tau_n\}$ of (4.16) such that for any given constant $K > 0$ we have $(1/n^{1/2}q(\tau_n)) > K$. Consequently, for any q with $q(0) = 0$ and for any sequence of Brownian bridges $\{B_n\}$ we have

$$P\{\sup_{0\le y\le 1} |(u_n(y) - B_n(y))/q(y)| = \infty\} = 1.$$

(3) (Taken from Csörgő, Csörgő, Horváth and Révész, 1982). As to the

problem of having

$$\sup_{0\leq y\leq 1} |(\rho_n(y)-B_n(y))/q(y)| \xrightarrow{P} 0 \tag{4.17}$$

with B_n as in (4.4), we note that choosing the g function of q appropriately, for some specific distributions (4.17) might turn out to be true. For example in case of $F(x)=1-e^{-x}$, $x\geq 0$, we may choose $g(y)=Q(y)a(y)/(\log\log(1/1-y))^{1/2}$, where now $Q(y)=\log(1/(1-y))$, $a(y)\to\infty$ as $y\to 1$, and then (4.17) will hold true. Naturally a similar statement holds true symmetrically for $F(x)=1-e^{x}$, $x\leq 0$. We note also that with the same g function as above, (4.17) will hold true also for the Weibull distribution.

(4) As mentioned already, M. Csörgő (1983, Theorem 5.1.1) verified (4.9) only under the conditions (i), (ii), (iii) of Theorem 4.1, and noted also that Theorem 4.3 also contained the latter result. In order to see this, consider

$$\sup_{1/(n+1)\leq y\leq n/(n+1)} \sup_{\theta_{y,n}} \frac{f(Q(y))}{f(Q(\theta_{y,n}))} \leq \sup_{1/(n+1)\leq y\leq n/(n+1)} \left(\frac{U_n(y)\vee y}{U_n(y)\wedge y}\,\frac{1-(U_n(y)\wedge y)}{1-(U_n(y)\vee y)}\right)^{\gamma}, \tag{4.18}$$

where γ is as in (iii) of Theorem 4.1, and the inequality is by Lemma 1 in Csörgő and Révész (1978). Using now Lemma 2 in Wellner (1978) on the right hand side rv of the inequality of (4.18) it follows (cf. e.g., (2.8) in M. Csörgő, 1983) that

$$\lim_{\varepsilon\to\infty} \overline{\lim_{n\to\infty}}\, P\left\{ \sup_{1/(n+1)\leq y\leq n/(n+1)} \sup_{\theta_{y,n}} \frac{f(Q(y))}{f(Q(\theta_{y,n}))} > \varepsilon \right\} = 0\,. \tag{4.19}$$

Hence condition (4.12) of Theorem 4.3 is satisfied and consequently (4.9) holds true with *any* O'Reilly weight function q under conditions (i), (ii), (iii) of Theorem 4.1.

(5) (Taken from Csörgő, Csörgő, Horváth and Révész, 1982). Given the conditions (i), (ii) of Theorem 4.1 and replacing its condition (iii) by requiring the existence of the limits (cf. (vi) of Theorem 4.7)

$$\lim_{y\downarrow 0} y\frac{f'(Q(y))}{f^2(Q(y))}=\gamma_1, \qquad \lim_{y\uparrow 1}(1-y)\frac{f'(Q(y))}{f^2(Q(y))}=\gamma_2,$$

where γ_1 and γ_2 are real numbers, then it can be shown (cf. Theorem 3.A in Parzen (1980), or page 7 of Seneta (1976), or Mason (1982)) that we have

$$\begin{aligned} &f(Q(y))=y^{\gamma_1}L_1(y) \quad \text{as } y\downarrow 0,\\ &f(Q(y))=(1-y)^{\gamma_2}L_2(y) \quad \text{as } y\uparrow 1, \end{aligned} \tag{4.20}$$

where L_1 and L_2 are slowly varying functions at 0 resp. at 1. If we simply

assume the forms of (4.20) *for* $f(Q)$ *on the tails*, then these are weaker assumptions on f than that of (iii), for then we do not require the existence of f' on (a, b), the support of F. So let us *assume that* $f(Q)$ *is as in* (4.20) *on the tails*, and consider its first statement (regarding the second one, similar conclusions will hold true). It follows from Corollary on page 274 in Feller (1966) that for a slowly varying function L_1 we have: for any $\varepsilon>0$ there exist some positive constants K_1, K_2 and $0<y_0<1$ such that we have

$$K_1y^{\varepsilon}<L_1(y)<K_2y^{-\varepsilon} \quad \text{for all } 0<y\leq y_0\,. \tag{4.21}$$

Consider now (cf. first statement of (4.20) with $\gamma_1\equiv\gamma$)

$$\begin{aligned} \sup_{1/(n+1)\leq y\leq c} \sup_{\theta_{y,n}} \frac{f(Q(y))}{f(Q(\theta_{y,n}))}\frac{1}{q(y)} &= \sup_{1/(n+1)\leq y\leq c} \sup_{\theta_{y,n}} \left(\frac{y}{\theta_{y,n}}\right)^{\gamma} \frac{L_1(y)}{L_1(\theta_{y,n})}\frac{1}{g(y)}\\ &\leq \sup_{1/(n+1)\leq y\leq c} \sup_{\theta_{y,n}} \left(\frac{y}{\theta_{y,n}}\right)^{\gamma} \frac{K_2y^{-\varepsilon}}{K_1\theta^{\varepsilon}_{y,n}}\frac{1}{g(y)}\\ &= \sup_{1/(n+1)\leq y\leq c} \sup_{\theta_{y,n}} \left(\frac{y}{\theta_{y,n}}\right)^{\gamma+\varepsilon} \frac{K_2}{K_1}\frac{1}{y^{2\varepsilon}}\frac{1}{g(y)}. \end{aligned} \tag{4.22}$$

It follows from (2.8) in M. Csörgő (1983) that for both $\gamma+\varepsilon>0$ and $\gamma+\varepsilon<0$ we have

$$\lim_{a\to\infty} \overline{\lim_{n\to\infty}}\, P\left\{ \sup_{1/(n+1)\leq y\leq c} \sup_{\theta_{y,n}} \left(\frac{y}{\theta_{y,n}}\right)^{\gamma+\varepsilon} > a \right\} = 0\,. \tag{4.23}$$

Hence choosing $g(y)=y^{-\delta}$, $\delta>0$ and then $\varepsilon>0$ of (4.21) so that $\varepsilon<\delta/2$, then the first condition of (4.12) holds true by (4.22) and (4.23) combined. This means that having assumed (4.20), which is weaker than (i), (ii), (iii) of Theorem 4.1 combined, the statement of (4.9) holds only with $g(y)=y^{-\delta}$ for $\delta>0$, arbitrary otherwise (cf. the last sentence of our example (4)).

(6) We note that the conditions discussed in (4) and (5) for the validity of (4.9) cannot, in general, insure also the validity of (4.17). Namely we have the following (Csörgő, Csörgő, Horváth and Révész, 1982)

Observation. *If* $\overline{\lim}_{y\to 0} f(Q(y))/q(y)=\infty$ *or* $\overline{\lim}_{y\to 1} f(Q(y))/q(y)=\infty$, *then*

$$P\{\sup_{0\leq y\leq 1} |\rho_n(y)/q(y)|=\infty\}=1\,. \tag{4.24}$$

Next we mention some *new strong approximations* of ρ_n. First we recall that conditions (i), (ii), (iii) of Theorem 4.1 imply (4.1), and given also the tail monitonicity assumptions of (iv), (v) we have also (4.2) and (4.3). We have just seen in Examples (4) and (5) that conditions (i), (ii), (iii) of Theorem 4.1 alone, or the somewhat weaker assumptions of (4.20) for the tail behaviour of $f(Q)$, imply (4.9) (in (4) without any further restrictions on q, while in (5) with $g(y)$ of the form $y^{-\delta}$ $(\delta > 0)$ only). As to the possibility of extending the statement of (4.1) over a wider range than $[\delta_n, 1-\delta_n]$, $\delta_n = 25n^{-1}\log\log n$, but using only the assumptions (i), (ii), (iii) of Theorem 4.1 and not those of its conditions (iv), (v) which, when combined with (i), (ii), (iii), made (4.2) and (4.3) possible, we have

THEOREM 4.6 (Csörgő, Csörgő, Horváth and Révész, 1982). *Assume the conditions* (i), (ii), (iii) *of Theorem* 4.1. *Then*

$$\sup_{1/(n+1)\leq y\leq n/(n+1)} |\rho_n(y) - u_n(y)| \overset{\text{a.s.}}{=} \begin{cases} O(n^{-1/2}(\log\log n)^{1+\gamma}) & \text{if } \gamma \leq 1\,, \\ O(n^{-1/2}(\log\log n)(\log n)^{(1+\varepsilon)(\gamma-1)}) & \text{if } \gamma > 1\,, \end{cases} \tag{4.25}$$

where $\varepsilon > 0$ is arbitrary, and γ is as in condition (iii).

It is clear from Theorem 4.6 that, when proving (4.2) and (4.3) the conditions (iv) and (v) of Theorem 4.1 come into play only because of the tail regions $[0, 1/(n+1))$, $(n/(n+1), 1]$. Having replaced δ_n of (1.8) by $1/(n+1)$ in (4.25), we have only paid the price of slightly weakened rates of convergence. While they render (4.2) and (4.3) true, the extra conditions (iv) and (v) of Theorem 4.1 are somewhat disjoint from that of (iii). Next we modify the latter somewhat for the sake of seeking further insight into the effect of the tail behaviour of the density-quantile function $f(Q)$ on a statment like (4.25). We are going to formulate this over the interval $[0, \frac{1}{2}]$ only and note that similar statements can be made over $[\frac{1}{2}, 1]$.

THEOREM 4.7 (Csörgő, Csörgő, Horváth and Révész, 1982). *Assume the conditions* (i), (ii), (iii) *of Theorem* 4.1 *and, instead of its conditions* (iv), (v), *we assume now that*

(vi) $$\lim_{y\downarrow 0} y\frac{f'(Q(y))}{f^2(Q(y))} = \gamma_1\,.$$

Then, if $\gamma_1 > 0$, we have (4.3), *and if $\gamma_1 > 0$ we have*

$$\sup_{n^{-\alpha}\leq y\leq 1/2} |\rho_n(y) - u_n(y)| \xrightarrow{\text{a.s.}} 0 \quad \text{as } n\to\infty\,, \tag{4.26}$$

provided that $\alpha < 1 + 1/(2|\gamma_1|)$. On the other hand, when $\gamma_1 < 0$, for $\alpha > 1 + 1/(2|\gamma_1|)$ there exists positive constants $K = K(\alpha)$ and $\lambda = \lambda(\alpha)$ such that

$$\overline{\lim_{n\to\infty}}\, n^{-\lambda} \sup_{n^{-\alpha}\leq y\leq 1/2} |\rho_n(y)| \geq K \quad \text{a.s.} \tag{4.27}$$

One of the interesting consequences of Theorem 4.1 is the following law of iterated logarithm (LIL) for ρ_n:

$$\overline{\lim_{n\to\infty}} \left(\frac{2}{\log\log n}\right)^{1/2} \sup_{0\leq y\leq 1} |\rho_n(y)| \overset{\text{a.s.}}{=} 1\,. \tag{4.28}$$

An interesting consequence of (4.26) of Theorem 4.7 is that it throws new light on the latter LIL. We have

COROLLARY 4.1 (Csörgő, Csörgő, Horváth and Révész, 1982). *Assume the conditions* (i), (ii), (iii) *of Theorem* 4.1 *and condition* (vi) *of Theorem* 4.7. *Then if $\gamma_1 > 0$ we have* (4.28), *and if $\gamma_1 < 0$ we have*

$$\overline{\lim_{n\to\infty}} \left(\frac{2}{\log\log n}\right)^{1/2} \sup_{n^{-\alpha}\leq y\leq 1/2} |\rho_n(y)| \overset{\text{a.s.}}{=} \begin{cases} 1 & \text{if } \alpha < 1 + 1/(2|\gamma_1|)\,, \\ \infty & \text{if } \alpha > 1 + 1/(2|\gamma_1|)\,. \end{cases} \tag{4.29}$$

PROOF. We have $\overline{\lim}_{n\to\infty} (2/\log\log n)^{1/2} \sup_{n^{-\alpha}\leq y\leq 1/2} |u_n(y)| \overset{\text{a.s.}}{=} 1$.

Now the first statement of (4.29) follows from the latter combined with (4.26). The second statement of (4.29) is by (4.17).

We note that Corollary 4.1 implies the non existence of LIL for ρ_n over $[0, 1]$ under (i), (ii), (iii) and (vi) if $\gamma_1 < 0$. On the other hand it follows from Theorem 3 in Mason (1982) that if we replace the weight function $f(Q)$ in $n^{-1/2}\rho_n$ by y^{ε}, $\varepsilon > 0$, then

$$\lim_{n\to\infty} \sup_{0\leq y\leq 1/2} y^{\varepsilon}|Q_n(y) - Q(y)| \overset{\text{a.s.}}{=} 0$$

for every $\varepsilon > 0$, given the conditions (i), (ii), (iii) and (vi), i.e., under the latter conditions we always have a Mason type Glivenko–Cantelli theorem for $(Q_n(y) - Q(y))$.

Summarizing the main features of the problem of strong approximation of ρ_n by u_n over $[0, 1]$ in general, we have seen so far that under the conditions (i), (ii), (iii), (iv), (v) of Theorem 4.1 we have (4.2) and (4.3), on dropping the conditions (iv) and (v) we have (4.25), and when we replace the conditions (iv), (v) by that of (vi) we have (4.26) and (4.27). Now we give an example which will amount to saying that for results like (4.2) and (4.3) neither the conditions (iv), (v), nor the condition (vi) are necessary. This example is due to Parzen (1979, page 116). Continuing the numbering of examples of this section, we now have

(7) Parzen's example (1979) (Result of (4.32) is quoted from Csörgő, Csörgő, Horváth and Révész, 1982): Let

$$1 - F(x) = \exp(-x - C\sin x), \quad x \geq 0,\ 0.5 < C < 1.$$

Letting $x = Q(y)$ $(0 \leq y \leq 1)$ we get

$$-\log(1-y)=Q(y)+C\sin Q(y)$$

and

$$f(Q(y))=(1-y)(1+C\cos Q(y)).$$

Hence

$$Q(y)\le|\log(1-y)|+C \tag{4.30}$$

and

$$f(Q(y))\le(1-y)(1+C). \tag{4.31}$$

Also

$$f'(Q(y))=-(1-y)((1+C\cos Q(y))^2+C\sin Q(y)).$$

Clearly then

$$\sup_{0<x<\infty}F(x)(1-F(x))\frac{|f'(x)|}{f^2(x)}=\sup_{0<y<1}y(1-y)\frac{|f'(Q(y))|}{f^2(Q(y))}\le\frac{1}{1-C},$$

i.e., *conditions* (i), (ii), (iii) *of Theorem* 4.1 *are satisfied. Hence by Theorem* 4.6 *we have* (4.25) *with* $\gamma=1/(1-C)$. *On the other hand, as* $y\to1$, $Q(y)\to\infty$ *and* $f(Q(y))$ *oscillates. Hence conditions* (iv), (v) *of Theorem* 4.1 *are not satisfied. Also, as* $y\to1$,

$$(1-y)\frac{f'(Q(y))}{f^2(Q(y))}$$

oscillates, i.e., the right tail version limit requirement of condition (vi) *is also not satisfied. Nevertheless in case of this example we have*

$$\sup_{0\le y\le1}|\rho_n(y)-u_n(y)|\overset{\text{a.s.}}{=}O(n^{-1/2}(\log\log n)^{1/(1-C)}(\log n)^{(1+\varepsilon)C/(1-C)}) \tag{4.32}$$

where $C\in(0.5,1)$ *as above.*

For a discussion of Bahadur's (1966) representation of sample quantiles and extension of Kiefer's (1970) theory of deviations between the sample quantile and empirical processes in the light of Theorem 4.1, we refer to Section 5.2 in Csörgő and Révész (1981a) and Chapter VI in M. Csörgő (1983).

References

Aly, E.-E., Burke, M. D., Csörgő, M., Csörgő, S. and Horváth, L. (1982). On the product-limit empirical and quantile processes: A collection of four papers. Carleton Mathematical Lecture Note No. 38, Carleton University, Ottawa.

Aly, E.-E. and Csörgő, M. (1981). Three papers on quantiles and the parameters estimated quantile process. Carleton Mathematical Lecture Note No. 31, Carleton University, Ottawa.

Bahadur, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 577–580.

Bártfai, P. (1966). Die Bestimmung der zu einem wiederkehrenden Process gehörenden Verteilungsfunktion aus den mit Fehlern behafteten Daten einer Einzigen Relation. *Studia Sci., Math. Hung.* **1**, 161–168.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6**, 434–451.

Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42**, 1656–1670.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **32**, 485–498.

Borisov, I. S. (1982). An approximation of empirical fields. *Coll. Math. Soc. J. Bolyai* **32**, *Nonparametric Statistical Inference*, Budapest 1980.

Borovkov, A. A. (1978). Rate of convergence and large deviations in invariance principle. In: *Proceedings of the International Congress of Mathematicians* (Helsinki, 1978), 725–731. Acad. Sci. Fennica, Helsinki (1980).

Brillinger, D. L. (1969). An asymptotic representation of the sample distribution function. *Bull. Amer. Math. Soc.* **75**, 545–547.

Burke, M. D., Csörgő, M., Csörgő, S. and Révész, P. (1979). Approximations of the empirical process when parameters are estimated. *Ann. Probability* **7**, 790–810.

Čencov, N. N. (1956). Wiener random fields depending on several parameters. *Dokl. Akad. Nank. SSSR* **106**, 607–609.

Chibisov, D. (1964). Some theorems on the limiting behaviour of empirical distribution function. *Selected Translations Math. Statist. Probability* **6**, 147–156.

Cotterill, D. S. and Csörgő, M. (1980). On the limiting distribution of and critical values for the Hoeffding–Blum–Kiefer–Rosenblatt independence criterion. Carleton Mathematical Lecture Note No. 24, Carleton University, Ottawa.

Cotterill, D. S. and Csörgő, M. (1982). On the limiting distribution of and critical values for the multivariate Cramér–von Mises statistic. *Ann. Statist.* **10**, 233–244.

Csáki, E. (1977a). Investigations concerning the empirical distribution function. *Magyar Tud. Akad. Mat. Fiz. Oszt. Kozl.* **23**, 239–327. English transl. in *Selected Transl. Math. Statist. and Probability* **15**, 229–317. Amer. Math. Soc., Providence, R.I. 1981.

Csáki, E. (1977b). The law of the iterated logarithm for normalized empirical distribution function. *Z. Wahrsch. Verw. Gebiete* **38**, 147–167.

Csáki, E. (1982a). On the standardized empirical distribution function. *Coll. Math. Soc. J. Bolyai* **32**, *Nonparametric Statistical Inference*, Budapest 1980.

Csáki, E. (1982b). Empirical distribution function. *This Volume.*

Csörgő, M. (1979). Strong approximations of the Hoeffding, Blum, Kiefer, Rosenblatt empirical process. *J. Multivariate Anal.* **9**, 84–100.

Csörgő, M. (1981a). Gaussian processes, strong approximations: an interplay. Colloques Internationaux Du Centre National De La Recherche Scientifique No. 307, *Aspects Statistiques Et Aspects Physiques Des Processus Gaussiens*, Saint-Flour 22–29 juin 1980, 131–229, Éditions du CNRS, Paris 1981.

Csörgő, M. (1981b). On a test for goodness-of-fit based on the empirical probability measure of Foutz and testing for exponentiality. In: D. Dugue et al., Eds., *Analytical Methods in Probability Theory*, Lecture Notes in Mathematics 861, Springer, Berlin, pp. 25–34.

Csörgő, M. (1981). MR81j:60044.

Csörgő, M. (1983). *Quantile Processes with Statistical Applications.* Regional Conference Series in Applied Mathematics, Philadelphia, SIAM (To appear).

Csörgő, M., Csörgő, S., Horváth, L. and Révész, P. (1982). On weak and strong approximations of the quantile process. To appear in *Proceedings of the Seventh Conference on Probability Theory* (Brasov, Aug. 29 – Sept. 4, 1982).

Csörgő, M. and Révész, P. (1975a). A new method to prove Strassen type laws of invariance principle, II. *Z. Wahrsch. Verw. Gebiete* **31**, 261–269.

Csörgő, M. and Révész, P. (1975b). A strong approximation of the multivariate empirical process. *Studia Sci. Math. Hungar.* **10**, 427–434.

Csörgő, M. and Révész, P. (1975c). Some notes on the empirical distribution function the quantile process. *Coll. Math. Soc. J. Bolyai* **11**, *Limit Theorems of Probability Theory* (P. Révész, Ed.). North-Holland, Amsterdam–London, pp. 59–71.

Csörgő, M. and Révész, P. (1978). Strong approximations of the quantile process. *Ann. Statist.* **6**, 882–894.
Csörgő, M. and Révész, P. (1981a). *Strong Approximations in Probability and Statistics.* Academic Press, New York - Akadémiai Kiadó, Budapest.
Csörgő, M. and Révész, P. (1981b). Quantile processes and sums of weighted spacings for composite goodness-of-fit. In: M. Csörgő et al., eds., *Statistics and Related Topics.* North-Holland, Amsterdam, pp. 69–87.
Csörgő, S. (1976). On an asymptotic expansion for the von Mises ω^2 statistic. *Acta Sci. Math. (Szeged)* **38**, 45–67.
Csörgő, S. (1980). On the quantogram of Kendall and Kent. *J. Appl. Probab.* **17**, 440–447.
Csörgő, S. (1981a). Limit behaviour of the empirical characteristic function. *Ann. Probab.* **9**, 130–144.
Csörgő, S. (1981b). Multivariate empirical characteristic function. *Z. Wahrsch. Verw. Gebiete* **55**, 203–229.
Csörgő, S. (1982). On general quantile processes in weighted sup-norm metrics. *Stoch. Process. Appl.* **12**, 215–220.
Csörgő, S. and Stacho, L. (1979). A step toward an asymptotic expansion for the Cramér–von Mises statistic. *Coll. Math. Soc. J. Boyai* **21**, Analytic Function Methods in Probability Theory (B. Gyires, Ed.), 53–65, North-Holland, Amsterdam.
Deheuvels, P. (1981). An asymptotic decomposition for multivariate distribution-free tests of independence. *J. Multivariate Anal.* **11**, 102–113.
DeWet, T. (1980). Cramér–von Mises tests for independence. *J. Multivariate Anal.* **10**, 38–50.
Doksum, Kjell (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.* **2**, 267–277.
Doksum, K. A., Fenstad, G. and Aaberg, R. (1977). Plots and tests for symmetry. *Biometrika* **64**, 473–487.
Doksum, K. A. and Sievers, G. L. (1976). Plotting with confidence: Graphical comparisons of two populations. *Biometrika* **63**, 421–434.
Doksum, K. A. and Yandell, B. S. (1982). Tests for exponentiality. *This volume.*
Donsker, M. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **23**, 277–283.
Dudley, R. M. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Ill. J. Math.* **10**, 109–126.
Dudley, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6**, 899–929; Correction (1979), ibid. **7**, 909–911.
Dudley, R. M. (1979). Lower layers in $\mathbb{R}^2$ and convex sets in $\mathbb{R}^3$ are not GB classes. *Probability in Banach Spaces II,* Lecture Notes in Mathematics 709. Springer, Berlin, pp. 97–102.
Dudley, R. M. (1981a). Donker classes of functions. In: M. Csörgő et al., eds., *Statistics and Related Topics.* North-Holland, Amsterdam, pp. 341–352.
Dudley, R. M. (1981b). Vapnik–Červonenkis Donsker classes of functions. Colloques Internationaux Du Centre National De La Recherche Scientifique No. 307, *Aspects Statistiques Et Aspects Physiques Des Processus Gaussiens,* Saint-Flour 22–29 juin 1980, 251–269, Editions du CNRS, Paris.
Dudley, R. M. (1982). Empirical and Poisson processes on classes of sets or functions too large for central limit theorems. Preprint.
Dudley, R. M. and Philipp, W. (1981). Invariance principles for sums of Banach space valued random elements and empirical processes. Preprint.
Dugue, D. (1969). Characteristic functions of random variables connected with Brownian motion and of the von Mises multidimensional ω_n^2. In: P. R. Krishnaiah, ed., *Multivariate Analysis 2.* Academic Press, New York, 289–301.
Durbin, J. (1970). Asymptotic distributions of some statistics based on the bivariate sample distribution function. In: M. L. Puri, ed., *Nonparametric Techniques in Statistical Inference.* Cambridge University Press, London, pp. 435–451.
Durbin, J. (1973a). *Distribution Theory for Tests Based on the Sample Distribution Function.* Regressional Conference Series in Applied Mathematics 9, Philadelphia, SIAM.

Durbin, J. (1973b). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1**, 279–290.
Durbin, J. (1976). Kolmogorov–Smirnov tests when parameters are estimated. In: P. Gaenssler and P. Révész, eds., *Empirical Distributions and Processes,* Lecture Notes in Mathematics 566. Springer, Berlin, pp. 33–44.
Erdős, P. and Rényi, A. (1970). On a new law of large numbers. *J. Analyse Math.* **23**, 103–111.
Feller, W. (1966). *An Introduction to Probability Theory and Its Applications, Vol. II.* Wiley, New York.
Gaenssler, P. and Stute, W. (1979). Empirical processes: a survey of results for independent and identically distributed random variables. *Ann. Probab.* **7**, 193–243.
Götze, F. (1979). Asymptotic expansions for bivariate von Mises functionals. *Z. Wahrsch. Verw. Gebiete* **50**, 333–355.
Hoeffding, W. (1948). A nonparametric test of independence. *Ann. Math. Statist.* **19**, 546–557.
Ibero, M. (1979a). Approximation forte du processus empirique fonctionnel multidimensionnel. *Bull. Sci. Math.* **103**, 409–422.
Ibero, M. (1979b). Intégrale stochastique par rapport au processus empirique multidimensionell. *C.R. Acad. Sci. Paris Sér. A* **288**, 165–168.
Kiefer, J. (1969). Old and new methods for studying order statistics and sample quantiles. In: *Proc. First International Conf. Nonparametric Inference.* Cambridge University Press 1970, London, pp. 349–357.
Kiefer, J. (1970). Deviations between the sample quantile process and the sample D.F. In: M. L. Puri, ed., *Nonparametric Techniques in Statistical Inference.* Cambridge University Press, London, pp. 299–319.
Kiefer, J. (1972). Skorohod embedding of multivariate RV's and the sample D.F. *Z. Wahrsch. Verw. Gebiete* **24**, 1–35.
Knott, M. (1974). The distribution of the Cramér–von Mises statistic for small sample sizes. *J. Roy. Statist. Soc. Ser. B.* **36**, 430–438.
Komlós, J., Major, P. and Tusnády, G. (1975a). An approximation of partial sums of independent R.V.'s and the sample D.F. I. *Z. Wahrsch. Verw. Gebiete* **32**, 111–131.
Komlós, J., Major, P. and Tusnády, G. (1975b). Weak convergence and embedding. *Coll. Math. Soc. J. Bolyai* **11**, Limit Theorems of Probability Theory (P. Révész, Ed.), North-Holland, Amsterdam–London, pp. 149–165.
Koziol, J. A. and Nemec, A. F. (1979). On a Cramér–von Mises type statistic for testing bivariate independence. *Canad. J. Statist.* **7**, 43–52.
Krivyakova, E. N., Martynov, G. V. and Tyurin Yu. N. (1977). On the distribution of the ω^2 statistic in the multidimensional case. *Th. Probab. Appl.* **22**, 406–410.
Kuelbs, J. (1976). A strong convergence theorem for Banach space valued random variables. *Ann. Probab.* **4**, 744–771.
Martynov, G. V. (1978). *Omega-Square Criteria* (in Russian). Nauka, Moscow.
Mason, D. M. (1982). Some characterizations of almost sure bounds for weighted multidimensional empirical distributions and a Glivenko–Cantelli theorem for sample quantiles. *Z. Wahrsch. Verw. Gebiete* **59**, 505–513.
Moore, D. S. and Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests. *Ann. Statist.* **4**, 599–616.
Müller, D. W. (1970). On Glivenko–Cantelli convergence. *Z. Wahrsch. Verw. Gebiete* **16**, 195–210.
Neuhaus, G. (1976). Weak convergence under contiguous alternatives of the empirical process when parameters are estimated: The D_k approach. In: P. Gaenssler and P. Révész, eds., *Empirical Distributions and Processes,* Lecture Notes in Mathematics 566, Springer, Berlin, pp. 68–82.
Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42**, 1285–1295.
Neuhaus, G. and Sen, P. K. (1977). Weak convergence of tail-sequence processes for sample distributions and averages. *Mitt. Math. Sem. Giessen* **123**, 25–35.
O'Reilly, N. (1974). On the weak convergence of empirical processes in sup-norm metrics. *Ann. Probability* **2**, 624–651.

Parzen, E. (1979). Nonparametric statistical data modelling. *J. Amer. Statist. Assoc.* **58**, 13–30.
Parzen, E. (1980). Quantile functions, convergence in quantile, and extreme value distribution theory. Texas A & M University Technical Report No. B-3.
Philipp, W. (1973). Empirical distribution functions and uniform distribution mod 1. In: C. F. Osgood, ed., *Diophantine Approximation and its Applications*. Academic Press, New York.
Philipp, W. and Pinzur, L. (1980). Almost sure approximation theorems for the multivariate empirical process. *Z. Wahrsch. Theorie Verw. Gebiete* **54**, 1–13.
Pyke, R. and Shorack, G. R. (1968). Weak convergence of the two-sample empirical process and a new approach to Chenoff–Savage theorems. *Ann. Math. Statist.* **39**, 755–771.
Révész, P. (1976a). On strong approximation of the multidimensional empirical process. *Ann. Probab.* **4**, 729–743.
Révész, P. (1976b). Three theorems of multivariate empirical process. In: P. Gaenssler and P. Révész, eds., *Empirical Distributions and Processes*, Lecture Notes in Mathematics 566. Springer, Berlin, pp. 106–126.
Seneta, E. (1976). Regularly Varying Functions. In: Lecture Notes in Mathematics 508, Springer, Berlin.
Shorack, G. R. (1979). Weak convergence of empirical and quantile processes in sup-norm metrics via KMT-construction. *Stoch. Process. Appl.* **9**, 95–98.
Shorack, G. R. (1982). Weak convergence of the general quantile process in $\|\cdot/q\|$-metrics (preliminary report). *IMS Bull.* **11**, Abstract 82t-2.
Straf, M. L. (1971). Weak convergence of stochastic processes with several parameters. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2**, University of California Press, pp. 187–221.
Stute, W. (1977). Convergence rates for the isotrope discrepancy. *Ann. Probab.* **5**, 707–723.
Stute, W. (1982). The oscillation behaviour of empirical processes. *Ann. Probab.* **10**, 86–107.
Tusnády, G. (1977a). A remark on the approximation of the sample DF in the multidimensional case. *Periodica Math. Hungar.* **8**, 53–55.
Tusnády, G. (1977b). *A Study of Statistical Hypotheses* (in Hungarian). Candidatus Dissertation, Hungarian Academy of Sciences.
Tusnády, G. (1977c). Strong invariance principles. In: J. R. Barra et al., eds., *Recent Developments in Statistics*. North-Holland, Amsterdam, pp. 289–300.
Wellner, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. Verw. Gebiete* **45**, 73–88.
Wichura, M. (1973). Some Strassen-type laws of the iterated logarithm for multiparameter stochastic processes with independent increments. *Ann. Probab.* **1**, 272–296.
Yeh, I. (1960). Wiener measure in space of functions of two variables. *Trans. Amer. Math. Soc.* **95**, 433–450.

PART IV

ESTIMATION PROCEDURES