The Annals of Statistics
1989, Vol. 17, No. 4, 1447–1471

BOOTSTRAPPING EMPIRICAL FUNCTIONS

BY SÁNDOR CSÖRGŐ[1] AND DAVID M. MASON[2]

University of Szeged and University of Delaware

We develop the complete bootstrapped parallel to the asymptotic theory of weighted empirical and quantile processes. Utilizing this parallel theory, we present a general body of techniques to establish the asymptotic validity of the bootstrap method of constructing confidence bands for statistical functions. These techniques are demonstrated to be applicable to the construction of asymptotic bootstrap confidence bands for a variety of concrete functions.

1. Introduction and discussion. Efron (1979) introduced the bootstrap method of constructing confidence intervals for a real valued population parameter $\theta(F)$. Given independent observations $X_1, \dots, X_n$ from F, this method consists of approximating the sampling distribution of an appropriate parameter estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ of $\theta(F)$ by means of the sampling distribution of the quantities $\hat{\theta}_{m,n} = \hat{\theta}_m(\tilde{X}_1, \dots, \tilde{X}_m)$, where the $m \geq 1$ observations $\tilde{X}_1, \dots, \tilde{X}_m$ are sampled independently from the finite population $X_1, \dots, X_n$ with distribution function $F_n(x) = n^{-1}\#\{k\colon\ 1 \leq k \leq n,\ X_k \leq x\}$, $-\infty < x < \infty$, i.e., $\tilde{X}_1, \dots, \tilde{X}_m$ are conditionally independent random variables (rv's) with common distribution function F_n, given $X_1, \dots, X_n$. In less than a decade the literature on the practical and theoretical aspects of the bootstrap approximation has become enormous.

Consider also the quantile function belonging to F,

$$\begin{aligned} Q(s) &= \inf\{x\colon F(x) \geq s\}, \qquad 0 < s < 1, \\ Q(0) &= Q(0+), \qquad Q(1) = Q(1-) \end{aligned} \tag{1.1}$$

and, with $X_{1,n} \leq \cdots \leq X_{n,n}$ denoting the order statistics of $X_1, \dots, X_n$, its empirical counterpart

$$Q_n(s) = \begin{cases} X_{k,n}, & \text{if } (k-1)/n < s \leq k/n;\ k = 1, \dots, n, \\ X_{1,n}, & \text{if } s = 0. \end{cases} \tag{1.2}$$

The bootstrapped empirical and quantile processes are, respectively,

$$m^{1/2}\{\tilde{F}_{m,n}(x) - F_n(x)\}, \qquad -\infty < x < \infty, \tag{1.3}$$

and

$$m^{1/2}\{\tilde{Q}_{m,n}(s) - Q_n(s)\}, \qquad 0 \leq s \leq 1, \tag{1.4}$$

Received May 1988; revised January 1989.

[1]Partially supported by the Hungarian National Foundation for Scientific Research, Grants 1808/86 and 457/88.

[2]Partially supported by the Alexander von Humboldt Foundation while the author was visiting the University of Munich and by NSF Grant DMS-88-03209.

AMS 1980 *subject classifications.* Primary 62G15, 62E20; secondary 62E25, 60F17, 60F05.

Key words and phrases. Bootstrap, confidence band, weighted empirical and quantile processes.

where $\tilde{F}_{m,n}(x) = m^{-1}\#\{k: 1 \le k \le m, \tilde{X}_k \le x\}$ and $\tilde{Q}_{m,n}(s) = \tilde{X}_{k,m}$ if $(k-1)/m < s \le k/m$, $k = 1, \dots, m$, and $\tilde{Q}_{m,n}(0) = \tilde{X}_{1,m}$, with $\tilde{X}_{1,m} \le \cdots \le \tilde{X}_{m,m}$ standing for the order statistics of the bootstrap sample $\tilde{X}_1, \dots, \tilde{X}_m$ with resampling size m.

Among many other things, Bickel and Freedman (1981) established the weak convergence of the processes in (1.3) and (1.4), the latter on a proper subinterval of $[0, 1]$ not containing the endpoints, and from these results they were able to deduce the asymptotic validity of the bootstrap method of forming confidence *bands* for F and Q. Shorack (1982) gave a simple proof for the first process in (1.3). [See also Shorack and Wellner (1986), Section 23.1.] The Bickel and Freedman result for $m^{1/2}(\tilde{F}_{m,n}(\cdot) - F_n(\cdot))$ has been subsequently generalized for empirical processes based on observations in $\mathbb{R}^d$, $d > 1$, and in very general sample spaces and indexed by various classes of sets and functions. [See, for example, Beran (1984), Beran and Millar (1986), Beran, Le Cam and Millar (1987), Gaenssler (1987), Lohse (1987) and Sheehy and Wellner (1986).] A certain final result is due to Giné and Zinn (1988). The Bayesian analogue of the Bickel and Freedman theory has been established by Lo (1987).

In the present paper we pursue further the above line of the Bickel and Freedman (1981) study and consider the validity of the bootstrap for general empirical functions on the real line containing as special cases the empirical distribution function and the empirical quantile function. This means that our ultimate aim is to show the asymptotic validity of bootstrap confidence-band estimation of functions on the real line generally different from F and Q. Our prime examples here will be various reliability, econometric and moment-type functions. In order to enable us to make our program more clear we now introduce some notations. Any convergence and order relations will be understood throughout as $n \to \infty$ if not specified otherwise.

Let $R_F(\cdot)$ be a statistical function of interest defined in an interval $I \subseteq \mathbb{R}$ and let $R_n(\cdot) = R_n(\cdot; X_1, \dots, X_n)$ be an appropriate estimator of $R_F(\cdot)$ on I. We can allow I to be the union of a finite number of disjoint (finite or infinite) intervals. Typically, for the process

$$r_n(\cdot) = n^{1/2}(R_n(\cdot) - R_F(\cdot)) \tag{1.5}$$

one can find a sequence of copies $\mathscr{G}_F^{(n)}(\cdot)$ of a separable Gaussian process $\mathscr{G}_F(\cdot)$ on I, i.e., $\mathscr{G}_F^{(n)}(\cdot) =_{\mathscr{D}} \mathscr{G}_F(\cdot)$ for each $n \ge 1$, such that on an appropriate probability space $(\Omega, \mathscr{A}, P)$,

$$P\Big\{\sup_{t \in I} |\mathscr{G}_F(t)| < \infty\Big\} = 1 \quad \text{and} \quad \sup_{t \in I} \big|r_n(t) - \mathscr{G}_F^{(n)}(t)\big| \to_P 0. \tag{1.6}$$

Consequently, given $0 < \alpha < 1$, we have (on any probability space where the X's are defined)

$$\Pr\big\{R_n(t) - cn^{-1/2} \le R_F(t) \le R_n(t) + cn^{-1/2},\ t \in I\big\} \to 1 - \alpha,$$

provided that $G_F(c) = 1 - \alpha$ and $c = c(\alpha, F)$ is a continuity point of the distribution function

$$G_F(x) = P\Big\{\sup_{t \in I} |\mathscr{G}_F(t)| \le x\Big\}, \qquad x \ge 0. \tag{1.7}$$

This means that $\{R_n(t) \pm cn^{-1/2},\ t \in I\}$, is an asymptotically correct $(1-\alpha)$ 100% confidence band for the statistical function R_F.

It is rare that this method of forming asymptotically correct confidence bands is feasible, since there are only a few cases when $c = c(\alpha, F)$ is independent of F and its analytical form is known. The most well-known case when this is true is the choice $R_F = F$, $R_n = F_n$ and F is continuous.

Consider the bootstrapped version of the empirical function $R_n(\cdot)$ given by $\tilde{R}_{m,n}(\cdot) = R_m(\cdot;\ \tilde{X}_1, \dots, \tilde{X}_m)$. Suppose we were able to show that on an appropriate extension $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P})$ of the above probability space $(\Omega, \mathscr{A}, P)$ there exist a sequence of versions $\tilde{\mathscr{G}}_F^{(m)}$ of the process $\mathscr{G}_F$, i.e., $\{\tilde{\mathscr{G}}_F^{(m)}(t):\ t \in I\} =_{\mathscr{D}} \{\mathscr{G}_F(t):\ t \in I\}$ for each n and a sequence of versions $\hat{R}_{m,n}$ of the process $\tilde{R}_{m,n}$ such that

(1.8a) $\quad$ the sequences $\{X_n\}_{n=1}^{\infty}$ and $\left\{\tilde{\mathscr{G}}_F^{(m)}\right\}_{n=1}^{\infty}$ are independent,

(1.8b)
$$\left\{\left(R_n(s), \tilde{R}_{m,n}(t)\right):\ s, t \in I\right\} =_{\mathscr{D}} \left\{\left(R_n(s), \hat{R}_{m,n}(t)\right):\ s, t \in I\right\} \quad \text{for each } n,$$

(1.8c)
$$\sup_{t \in I}\left| m^{1/2}\left(\hat{R}_{m,n}(t) - R_n(t)\right) - \tilde{\mathscr{G}}_F^{(m)}(t)\right| \to_{\tilde{P}} 0,$$
where $m = m(n) \to \infty$ at an appropriate rate.

(Here and in what follows $=_{\mathscr{D}}$ stands for the equality of all finite dimensional distributions of the stochastic processes on the two sides.) From (1.6) and (1.8) we can conclude that whenever x is a continuity point of G_F in (1.7), then, by Proposition 1 in the Appendix,

$$\Pr\left\{\sup_{t \in I} m^{1/2}\left|\tilde{R}_{m,n}(t) - R_n(t)\right| \le x | X_1, \dots, X_n\right\} \to_{\Pr} G_F(x) \tag{1.9}$$

(on any probability space) for the same $m = m(n)$ sequence. Now fix $0 < \alpha < 1$ and suppose we can show that

(1.10)
G_F is continuous at both $c = c(\alpha, F) = \inf\{x:\ G_F(x) \ge 1 - \alpha\}$ and $d = d(\alpha, F) = \sup\{x:\ G_F(x) \le 1 - \alpha\}$.

Given then the observations $X_1, \dots, X_n$, for each m let $c_m = c_m(X_1, \dots, X_n)$ be defined as

$$c_m = \inf\left\{x:\ \Pr\left\{\sup_{t \in I} m^{1/2}\left|\tilde{R}_{m,n}(t) - R_n(t)\right| \le x | X_1, \dots, X_n\right\} \ge 1 - \alpha\right\}.$$

By Proposition 2 in the Appendix we then conclude from (1.6) and (1.9) that

$$\Pr\left\{\sup_{t \in I} n^{1/2}\left|R_n(t) - R_F(t)\right| \le c_m\right\} \to 1 - \alpha, \tag{1.11}$$

provided $m = m(n) \to \infty$ at the rate required by (1.8). Hence from (1.11) we see that an asymptotically $(1-\alpha)\,100\%$ confidence band for R_F is given by

$$\{R_n(t) \pm c_m n^{-1/2},\ t \in I\}. \tag{1.12}$$

We note, however, that since, given $X_1, \dots, X_n$, $\sup\{m^{1/2}|\tilde{R}_{m,n}(t) - R_n(t)|: t \in I\}$ can take on as many as n^m possible values, which is typically an astronomically large number, c_m must in most practical situations be estimated by Monte Carlo simulation.

Bickel and Freedman (1981) established the validity of the above procedure [spelled out very similarly but nevertheless somewhat differently also by Beran (1984)] in two cases. One is when $R_F = F$ is an arbitrary distribution function, $I = (-\infty, \infty)$ and $R_n = F_n$. The other is when $R_F = Q$ as given in (1.1), $I = [a, b] \subset (0, 1)$, $a < b$, and $R_n = Q_n$. Concerning these results, see also Section 3.

The requirement in (1.10) is satisfied in any conceivable statistical situation. Indeed, it follows from Theorem 1 of Tsirel'son (1975) that if $\mathscr{G}_F$ is any separable, mean-zero, almost surely bounded Gaussian process such that $\operatorname{Var} \mathscr{G}_F(t) > 0$ for some point t of I, then G_F is continuous on the whole half-line $(0, \infty)$. Without further mention, this will be the case for any concrete example in the present paper.

On the other hand, to show that (1.6) holds for the processes r_n in (1.5) always requires a specific approach depending on the particular nature of the statistical function R_F and the structure of the empirical function R_n. This is also the case when one must in addition establish the bootstrap counterpart in (1.8). Subsequently, to state and prove a general theorem concerning when the above bootstrap program works is a hopeless venture. Rather than make such an attempt, the purpose of this paper is to provide a body of techniques that should prove useful in establishing (1.8) for a variety of empirical functions of statistical interest. We will demonstrate how to apply our techniques by showing the validity of (1.8) in Sections 3 and 4 for a number of concrete examples.

The philosophy of the bootstrap principle includes the appealing heuristic idea that bootstrapped versions $\tilde{r}_{m,n} = m^{1/2}(\tilde{R}_{m,n} - R_n)$ of processes r_n behave asymptotically the same way as the original processes r_n *under the same regularity conditions on the underlying distribution*, and, therefore, under the (preferably optimal) regularity conditions of (1.6) we also have the final bootstrap confidence-band statement in (1.11). [Of course, the idea is not true literally in general. See the counterexamples of Bickel and Freedman (1981).] One of the most powerful techniques of establishing a result (1.6) on the real line (usually under optimal regularity conditions) is the weighted approximation method of Csörgő, Csörgő, Horváth and Mason (1986a) [see also Mason and van Zwet (1987)], which has been used in various contexts in Csörgő, Csörgő, Horváth and Mason (1986a, b, c), and a weaker version of the idea in Csörgő, Csörgő, and Horváth (1986). The origin of this method goes back to the works of Chibisov, Pyke and Shorack, Shorack and O'Reilly. [See the references in the above papers and in Shorack and Wellner (1986).] Now, almost to the contrary of the negative claims in the preceding paragraph, the principal message or "metatheorem" of the present paper intends to be something like the following:

Whatever result (1.6) *can be proved by the weighted approximation method under some regularity conditions on F the bootstrap is automatically valid under the same conditions, that is, we also have* (1.1).

We have just defined our understanding of the asymptotic validity of the bootstrap in the present confidence-band context. With the notable exception of Efron (1979) himself, most authors justify the bootstrap by proving that the statement corresponding to (1.9) holds in the stronger sense of almost sure convergence rather than convergence in probability. Our approach cannot be adapted to produce this. The in probability version of the justification of the bootstrap is in fact more universally applicable than the almost sure version. There are cases when the bootstrap construction of confidence intervals and bands cannot be justified by almost sure convergence while it can be in the in probability sense. After presenting the main results in Sections 2 and 3 and the examples in Section 4, we show in Section 5 that our approach contains as a by-product the asymptotic validity of bootstrapping the mean in the critical case when F has an infinite variance but still belongs to the domain of attraction of a normal law, i.e., (2.16) below holds. This result was proved earlier by Athreya (1985), who showed by completely different direct methods that asymptotically equivalent versions of (5.1) and (5.2) hold, also in probability only. We point out there that (5.1) and (5.2) do not always hold almost surely only under (2.16). However, we view (1.9) as an intermediate step only and the difference between almost sure and stochastic convergence here disappears when passing to (1.11). Note that our point of view is closely related to what Hall (1986) writes in the second half of the third paragraph of his introduction.

This last issue seems to have some bearing on the difference between the original, very powerful method and our method of validating the bootstrap. The original method, starting with Bickel and Freedman (1981) and followed and extended by Beran (1984), Sheehy and Wellner (1986) and others, requires some uniformity in F in the convergence of the original limit theorem (1.6) [and thereby yields the almost sure version of (1.9)]. Leaving aside the problem that none of the processes corresponding to Examples 3–6 in Section 4 can be written as a single functional of $n^{1/2}(F_n - F)$ or as $n^{1/2}(F_n - F)$ indexed by a single class of functions, the uniformity requirement generally stipulates more regularity on F than the original convergence theorem needs. But then there is no chance to have the bootstrap result under the same regularity conditions. This is the reason, for example, why Theorem 3.2 for the bootstrapped empirical process indexed by functions on the line does not follow from any general result of the same kind: It holds under the same conditions as the original limit theorem. These conditions (which are optimal in certain "stationary" cases) are individual for the given F. This point is most transparent again in the mean example of Section 5: The indexing function is the inverse of F and condition (2.16) as well as the normalizing sequence in the result (5.1) depend on the whole tail behavior of the individual F. The counterexamples of Bickel and Freedman (1981) and Beran (1984) do show that uniformity is necessary in some situations. We do not think it is necessary in all situations. In fact, the fine result of Giné and Zinn

(1988) shows that it is not necessary even for almost sure convergence. Similarly, our method is not universal either to handle all the problems on the real line, to where it is admittedly restricted.

We emphasize that the bootstrap does solve previously inaccessible problems. With the exception of Example 1, with continuous F, in Section 3 and Examples 3 and 7 in Section 4, confidence bands have not been previously available in the other nine examples, due to the intractability of the limiting process.

In order to guard against situations where large variances at certain t points (usually close to the extreme endpoints of I) dominate and make the bands $R_n(\cdot) \pm c_m n^{-1/2}$ uselessly wide, a referee has suggested the bands

$$\left\{R_n(t) \pm d_m n^{-1/2}\sigma_n(t),\ t \in J\right\}, \tag{1.13}$$

where

$$d_m = \inf\left\{x\colon \Pr\left\{\sup_{t\in J} m^{1/2}\frac{|\tilde{R}_{m,n}(t) - R_n(t)|}{\sigma_n(t)} \le x | X_1, \dots, X_n\right\} \ge 1 - \alpha\right\},$$

$\sigma_n(\cdot)$ is a uniformly consistent estimator of the standard deviation function $\sigma(\cdot)$ of the limiting process $\mathscr{G}_F(\cdot)$ and $J \subseteq I$ is the union of a finite number of intervals where $\sigma(\cdot)$ is bounded away from zero and infinity, under which conditions the theory works, as competitors to the above constant-width bands, together with the possibility of other weighted or transformed bands.

The problem of deciding which one of the two types of bands in (1.12) and (1.13), or perhaps yet another type, is better will depend on the type of the statistical function R_F to be estimated as well as on the underlying F and the intervals I or J of interest. It would be hard to get a general recipe that would work in all situations. Each case should be examined separately. While there is nothing against the natural choice $m = n$ for the bootstrap sample size, which is almost forced by condition (2.11) below, the same can be said about the speed of convergence of the true coverage probabilities to the nominal $1 - \alpha$. Of course, the smaller α is, the larger the number of Monte Carlo simulations for the computation of c_m or d_m must be to obtain reasonable bands. The investigation of the complexity of these interesting statistical problems of practical importance would require very extensive simulation studies and is out of the scope of the present paper.

In Example 7 we refer to some concrete experience of one of us suggesting that 3000 bootstrap replications should suffice for $1 - \alpha = 0.9$ and $m = n \approx 100$ for quantile type functions (which is usually the bad case) on very long intervals. We illustrate Example 12 by constructing bands for the probability generating function of the famous horsekick data of von Bortkiewitz (1898).

2. Weighted approximations to the bootstrapped uniform empirical and quantile processes. Let $(\Omega, \mathscr{A}, P)$ be the probability space constructed in Csörgő, Csörgő, Horváth, and Mason (1986a). This space carries a sequence of independent uniform (0,1) rv's $U_1, U_2, \dots$ and a sequence of Brownian bridges

$B_1, B_2, \ldots$ such that for the uniform empirical and quantile processes $\alpha_n(s) = n^{1/2}(G_n(s) - s)$ and $\beta_n(s) = n^{1/2}(s - U_n(s))$, $0 \le s \le 1$, where $G_n(s) = n^{-1}\#\{k: 1 \le k \le n, U_k \le s\}$ and, with $U_{1,n} \le \cdots \le U_{n,n}$ denoting the order statistics of $U_1, \ldots, U_n$, $U_n(s) = U_{k,n}$ if $(k-1)/n < s \le k/n$, $k = 1, \ldots, n$, and $U_n(0) = U_{1,n}$, we have

$$(2.1)\qquad \begin{aligned} \sup_{0<s<1} \frac{|\alpha_n(s) - B_n^*(s)|}{(s(1-s))^{1/2-\nu_1}} &= O_P(n^{-\nu_1}), \\ \sup_{\lambda/n \le s \le 1-\lambda/n} \frac{|\beta_n(s) - B_n(s)|}{(s(1-s))^{1/2-\nu_2}} &= O_P(n^{-\nu_2}) \end{aligned}$$

for all $0 < \lambda < \infty$ and $0 \le \nu_1 < \frac{1}{4}$, $0 \le \nu_2 < \frac{1}{2}$, where $B_n^*(s) = B_n(s)$ if $1/n \le s \le 1 - 1/n$ and zero otherwise. For the same construction we also have

$$(2.1')\qquad \begin{aligned} \sup_{0 \le s \le 1} |\alpha_n(s) - B_n(s)| &= O\big(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}\big) \quad \text{a.s.}, \\ \sup_{0 \le s \le 1} |\beta_n(s) - B_n(s)| &= O(n^{-1/2}\log n) \quad \text{a.s.} \end{aligned}$$

[Alternatively, we could choose $(\Omega, \mathscr{A}, P)$ to be the space constructed by Mason and van Zwet (1987) with ν_1 and ν_2 transposed in (2.1) and (2.2) below and the rate sequences transposed in (2.1′).]

Now extend $(\Omega, \mathscr{A}, P)$ to obtain a probability space $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P})$ which, besides $\{U_i\}$ and $\{B_i\}$, carries another sequence of independent uniform $(0,1)$ rv's $\xi_1, \xi_2, \ldots$ and another sequence of Brownian bridges $\tilde{B}_1, \tilde{B}_2, \ldots$ such that the two sets of random elements $\{U_i\}_{i=1}^{\infty} \cup \{B_i(s): 0 \le s \le 1\}_{i=1}^{\infty}$ and $\{\xi_i\}_{i=1}^{\infty} \cup \{\tilde{B}_i(s): 0 \le s \le 1\}_{i=1}^{\infty}$ are independent and that, besides (2.1), for the processes $e_m(s) = m^{1/2}(E_m(s) - s)$ and $k_m(s) = m^{1/2}(s - \xi_m(s))$, $0 \le s \le 1$, where $E_m(s) = m^{-1}\#\{k: 1 \le k \le m, \xi_k \le s\}$ and, with $\xi_{1,m} \le \cdots \le \xi_{m,m}$ denoting the order statistics of $\xi_1, \ldots, \xi_m$, $\xi_m(s) = \xi_{k,m}$ if $(k-1)/m < s \le k/m$, $k = 1, \ldots, m$, and $\xi_m(0) = \xi_{1,m}$, we also have, as $m \to \infty$, the parallel to (2.1′) and

$$(2.2)\qquad \begin{aligned} \sup_{0<s<1} \frac{|e_m(s) - \tilde{B}_m^*(s)|}{(s(1-s))^{1/2-\nu_1}} &= O_{\tilde{P}}(m^{-\nu_1}), \\ \sup_{\lambda/m \le s \le 1-\lambda/m} \frac{|k_m(s) - \tilde{B}_m(s)|}{(s(1-s))^{1/2-\nu_2}} &= O_{\tilde{P}}(m^{-\nu_2}) \end{aligned}$$

for all $0 < \lambda < \infty$ and $0 \le \nu_1 < \frac{1}{4}$, $0 \le \nu_2 < \frac{1}{2}$, where $\tilde{B}_m^*(s) = \tilde{B}_m(s)$ if $1/m \le s \le 1 - 1/m$ and zero otherwise. Of course $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P})$ can be obtained by taking the product of $(\Omega, \mathscr{A}, P)$ with itself. Now P can be replaced by $\tilde{P}$ in (2.1).

REMARK 1. Using the fact that for each $0 < \delta \le \frac{1}{2}$ both

$$a^{-\delta} \sup_{0 \le s \le a} |B(s)| s^{-1/2+\delta} \quad \text{and}$$

$$a^{-\delta} \sup_{0 \le s \le a} |B(1-s)| s^{-1/2+\delta} \to_{\mathscr{D}} \sup_{0 \le s \le 1} |W(s)| s^{-1/2+\delta}$$

as $a \downarrow 0$ where W is a standard Wiener process on $[0,1]$ [cf. Csörgő and Mason (1985)], it is easy to show that the first weighted approximations in (2.1) and (2.2) remain true when $0 < \nu_1 < \frac{1}{4}$ and B_n^* and $\tilde{B}_m^*$ are replaced by B_n and $\tilde{B}_m$, respectively.

For convenient reference later on we list a number of facts on the linearity of the uniform empirical distribution and quantile functions:

$$\sup_{0<s<1} G_n(s)/s + \sup_{0<s<1} (1 - G_n(s))/(1 - s) = O_{\tilde{P}}(1), \tag{2.3}$$

$$\sup_{U_{1,n} \le s < 1} s/G_n(s) + \sup_{0<s<U_{n,n}} (1 - s)/(1 - G_n(s)) = O_{\tilde{P}}(1), \tag{2.4}$$

$$\sup_{0<s<1} s/U_n(s) + \sup_{0<s<1} (1 - s)/(1 - U_n(s)) = O_{\tilde{P}}(1) \tag{2.5}$$

and for any $0 < \rho < \infty$,

$$\sup_{\rho/n \le s < 1} U_n(s)/s + \sup_{0<s\le 1-\rho/n} (1 - U_n(s))/(1 - s) = O_{\tilde{P}}(1). \tag{2.6}$$

All these follow from Remark 1 of Wellner (1978). Of course, all these statements hold also for the empirical distribution and quantile functions $E_m(\cdot)$ and $\xi_m(\cdot)$ of the $\{\xi_i\}$ sequence.

Introduce

$$G_{m,n}(s) = E_m(G_n(s)) \quad \text{and} \quad U_{m,n}(s) = U_n(\xi_m(s)), \qquad 0 \le s \le 1. \tag{2.7}$$

It is easy to show that given independent rv's $X_1, \dots, X_n$ (on any probability space) with common distribution function F, for the bootstrapped empirical distribution function $\tilde{F}_{m,n}$ and quantile function $\tilde{Q}_{m,n}$ given in (1.3) and (1.4) we have for all $n, m \ge 1$,

$$\begin{aligned} &\left\{\left(\tilde{F}_{m,n}(x), \tilde{Q}_{m,n}(s), F_n(y), Q_n(t)\right): -\infty < x,\ y < \infty, 0 \le s,\ t \le 1\right\} \\ &\quad =_{\mathscr{D}} \left\{\left(G_{m,n}(F(x)), Q(U_{m,n}(s)), G_n(F(y)), Q(U_n(t))\right): \right. \\ &\qquad\qquad\qquad \left. -\infty < x,\ y < \infty, 0 \le s,\ t \le 1\right\}. \end{aligned} \tag{2.8}$$

[This distributional equivalence was used implicitly by Bickel and Freedman (1981) and by Shorack (1982).] For this reason, by some abuse of language, we shall sometimes refer to the processes

$$\begin{aligned} \alpha_{m,n}(s) &= m^{1/2}(G_{m,n}(s) - G_n(s)) \quad \text{and} \\ \beta_{m,n}(s) &= m^{1/2}(U_n(s) - U_{m,n}(s)), \qquad 0 \le s \le 1, \end{aligned} \tag{2.9}$$

as the bootstrapped uniform empirical and quantile processes, respectively. For these processes we need the "bootstrapped" versions of (2.1) and (2.2). Set $l(n) = n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}$ for the rate sequence figuring in (2.1′).

THEOREM 2.1. *For any sequence $m = m(n) \to \infty$ of positive integers and for each $0 \le \nu < \frac{1}{4}$,*

$$\sup_{U_{1,n} \le s < U_{n,n}} \left|\alpha_{m,n}(s) - \tilde{B}_m^*(s)\right| / (s(1 - s))^{1/2-\nu} = O_{\tilde{P}}\left((m \wedge n)^{-\nu}\right) \tag{2.10}$$

and

$$\sup_{0 \le s \le 1} \left|\alpha_{m,n}(s) - \tilde{B}_m(s)\right| = O(l(m) \vee l(n)) \quad a.s. \tag{2.10'}$$

and whenever $m = m(n)$ *satisfies the condition that for two constants* $0 < C_1 < C_2$,

$$C_1 m \le n \le C_2 m, \qquad n = 1, 2, \dots, \tag{2.11}$$

for any $0 < \lambda < \infty$ *and* $0 \le \nu < \frac{1}{4}$,

$$\sup_{\lambda/m \le s \le 1 - \lambda/m} \left|\beta_{m,n}(s) - \tilde{B}_m(s)\right| / (s(1-s))^{1/2-\nu} = O_{\tilde{P}}(m^{-\nu}) \tag{2.12}$$

and

$$\sup_{0 \le s \le 1} \left|\beta_{m,n}(s) - \tilde{B}_m(s)\right| = O(l(m)) \quad a.s. \tag{2.12'}$$

Proof. First we consider (2.10). Choose any $0 \le \nu < \frac{1}{4}$. Since (2.3) implies that

$$\sup_{0 < s < 1} G_n(s)(1 - G_n(s))/(s(1-s)) = O_{\tilde{P}}(1),$$

it is sufficient to prove that

$$S_{m,n}^{(\nu)} := \sup_{U_{1,n} \le s < U_{n,n}} \Delta_{m,n}^{(\nu)}(s) = O_{\tilde{P}}\left((n \wedge m)^{-\nu}\right), \tag{2.13}$$

where

$$\Delta_{m,n}^{(\nu)}(s) = \left|\alpha_{m,n}(s) - \tilde{B}_m^*(s)\right| / (G_n(s)(1 - G_n(s)))^{1/2-\nu}.$$

Observe that

$$\begin{aligned} S_{m,n}^{(\nu)} \le & \sup_{U_{1,n} \vee 1/m \le s < U_{n,n} \wedge (1-1/m)} \Delta_{m,n}^{(\nu)}(s) + \sup_{U_{1,n} \le s < U_{1,n} \vee 1/m} \Delta_{m,n}^{(\nu)}(s) \\ & + \sup_{U_{n,n} \wedge (1-1/m) \le s < U_{n,n}} \Delta_{m,n}^{(\nu)}(s) =: S_{m,n,1}^{(\nu)} + S_{m,n,2}^{(\nu)} + S_{m,n,3}^{(\nu)}, \end{aligned}$$

where $S_{m,n,2}^{(\nu)}$ is defined to be zero if $U_{1,n} \vee (1/m) = U_{1,n}$ and $S_{m,n,3}^{(\nu)}$ is defined to be zero if $U_{n,n} \wedge (1 - 1/m) = U_{n,n}$.

Notice that

$$\begin{aligned} S_{m,n,1}^{(\nu)} \le & \sup_{U_{1,n} \vee 1/m \le s < U_{n,n} \wedge (1-1/m)} \frac{\left|\alpha_{m,n}(s) - \tilde{B}_m^*(G_n(s))\right|}{(G_n(s)(1 - G_n(s)))^{1/2-\nu}} \\ & + \sup_{U_{1,n} \vee 1/m \le s < U_{n,n} \wedge (1-1/m)} \frac{\left|\tilde{B}_m(G_n(s)) - \tilde{B}_m(s)\right|}{(G_n(s)(1 - G_n(s)))^{1/2-\nu}} \\ \le & \sup_{0 < s < 1} \left|e_m(s) - \tilde{B}_m^*(s)\right| / (s(1-s))^{1/2-\nu} \\ & + \sup_{U_{1,n} \le s < U_{n,n}} \left|\tilde{B}_m(G_n(s)) - \tilde{B}_m(s)\right| / (G_n(s)(1 - G_n(s)))^{1/2-\nu}. \end{aligned}$$

By (2.2) the first term on the right side is $O_{\tilde{P}}(m^{-\nu})$, while the second term is

$$\leq 2 \max_{1 \leq i \leq n-1} \sup_{U_{i,n} \leq s < U_{i+1,n}} |\tilde{B}_m(i/n) - \tilde{B}_m(s)|/(i/n)^{1/2-\nu}$$

$$+ 2 \max_{1 \leq i \leq n-1} \sup_{U_{i,n} \leq s < U_{i+1,n}} |\tilde{B}_m(i/n) - \tilde{B}_m(s)|/(1 - i/n)^{1/2-\nu}.$$

In the proof of Theorem 2.2 of Csörgő, Csörgő, Horváth and Mason (1986a) it was shown that these last two terms are $O_{\tilde{P}}(n^{-\nu})$. Hence we have

$$S_{m,n,1}^{(\nu)} = O_{\tilde{P}}\big((m \wedge n)^{-\nu}\big). \tag{2.14}$$

Next we consider $S_{m,n,2}^{(\nu)}$. Note that since $\tilde{B}_m^*(s) = 0$ for $0 \leq s < 1/m$,

$$S_{m,n,2}^{(\nu)} = \sup_{U_{1,n} \leq s < U_{1,n} \vee 1/m} |\alpha_{m,n}(s)|/\big(G_n(s)(1 - G_n(s))\big)^{1/2-\nu}$$

$$\leq \sup_{0 < s < G_n(1/m)} |e_m(s)|/\big(s(1-s)\big)^{1/2-\nu}.$$

Choose any $\rho > 1$. Notice that whenever $G_n(1/m) \leq \rho/m$, this last expression is

$$\leq \sup_{0 < s < \rho/m} |e_m(s)|/\big(s(1-s)\big)^{1/2-\nu}$$

$$\leq m^{-\nu} \sup_{0 < s < \rho/m} (ms)^{1/2+\nu}/(1-s)^{1/2-\nu}$$

$$+ m^{-\nu}\{mE_m(\rho/m)\}/\big\{(m\xi_{1,m})^{1/2-\nu}(1 - \rho/m)^{1/2-\nu}\big\}$$

$$\leq 2m^{-\nu} + O_{\tilde{P}}(m^{-\nu})mE_m(\rho/m)$$

for large enough m and by (2.5). Since $E(mE_m(\rho/m)) = \rho$, the last two terms are $O_{\tilde{P}}(m^{-\nu})$. On the other hand, by Markov's inequality $\tilde{P}\{G_n(1/m) \leq \rho/m\} \geq 1 - 1/\rho$ for all $\rho > 1$ and $m \geq 1$, and therefore an elementary argument now establishes that $S_{m,n,2}^{(\nu)} = O_{\tilde{P}}(m^{-\nu})$. An analogous proof shows that we also have $S_{m,n,3}^{(\nu)} = O_{\tilde{P}}(m^{-\nu})$. The last two relations and (2.14) imply (2.13) and hence the first statement of the theorem.

To prove (2.12), first we observe that

$$|\beta_{m,n}(s) - \tilde{B}_m(s)| \leq |k_m(s) - \tilde{B}_m(s)| + \left(\frac{m}{n}\right)^{1/2} |\beta_n(\xi_m(s)) - B_n(\xi_m(s))|$$

$$+ \left(\frac{m}{n}\right)^{1/2} |B_n(\xi_m(s)) - B_n(s)| + \left(\frac{m}{n}\right)^{1/2} |B_n(s) - \beta_n(s)|$$

$$:= \nabla_m^{(1)}(s) + \cdots + \nabla_m^{(4)}(s).$$

Choose any $0 \leq \nu < \frac{1}{4}$. By (2.2) we have

$$\sup_{\lambda/m \leq s \leq 1-\lambda/m} \nabla_m^{(1)}(s)/\big(s(1-s)\big)^{1/2-\nu} = O_{\tilde{P}}(m^{-\nu})$$

and by (2.1) along with assumption (2.11) we obtain

$$\sup_{\lambda/m \leq s \leq 1-\lambda/m} \nabla_m^{(4)}(s)/\big(s(1-s)\big)^{1/2-\nu} = O_{\tilde{P}}(m^{-\nu}).$$

Choose any $1 < \rho < \infty$ and set

$A_m^{(\lambda)}(\rho) = \{s/\rho \le \xi_m(s) \text{ and } \xi_m(s) \le 1 - (1-s)/\rho \text{ for } \lambda/m \le s \le 1 - \lambda/m\}$.

Notice that on the event $A_m^{(\lambda)}(\rho)$,

$$\sup_{\lambda/m \le s \le 1-\lambda/m} \nabla_m^{(2)}(s)/(s(1-s))^{1/2-\nu}$$

$$\le \rho^{1-2\nu}(m/n)^{1/2} \sup_{\lambda/m \le s \le 1-\lambda/m} \frac{|\beta_n(\xi_m(s)) - B_n(\xi_m(s))|}{(\xi_m(s)(1-\xi_m(s)))^{1/2-\nu}}$$

$$\le \rho^{1-2\nu}(m/n)^{1/2} \sup_{\lambda/(\rho m) \le t \le 1-\lambda/(\rho m)} |\beta_n(t) - B_n(t)|/(t(1-t))^{1/2-\nu},$$

which by (2.1) and (2.11) is $O_{\tilde{P}}(n^{-\nu}) = O_{\tilde{P}}(m^{-\nu})$. But (2.5) as applied to $\xi_m(\cdot)$ implies that

$$\lim_{\rho\to\infty} \liminf_{m\to\infty} \tilde{P}\{A_m^{(\lambda)}(\rho)\} = 1.$$

Hence

$$\sup_{\lambda/m \le s \le 1-\lambda/m} \nabla_m^{(2)}(s)/(s(1-s))^{1/2-\nu} = O_{\tilde{P}}(m^{-\nu}).$$

Finally, to establish that

$$\sup_{\lambda/m \le s \le 1-\lambda/m} \nabla_m^{(3)}(s)/(s(1-s))^{1/2-\nu} = O_{\tilde{P}}(m^{-\nu}),$$

one requires a routine, though very lengthy, modification of the corresponding part of the proof of Theorem 2.2 of Csörgő, Csörgő, Horváth and Mason (1986a). For the sake of brevity these details are omitted.

The primed versions (2.10′) and (2.12′) follow from (2.1′) and the corresponding primed parallel to (2.2) by usual strong approximation methods. Most of the details can be read out from Lo (1987). □

REMARK 2. Combining the fact cited in Remark 1 with (2.6), it is straightforward to show that (2.10) remains true for all $0 < \nu < \frac{1}{4}$ when the supremum is taken over $[0,1]$ and $\tilde{B}_m^*$ is replaced by $\tilde{B}_m$. Similarly, since for any choice of $0 < \lambda_1 < \lambda_2 < 1$ one has trivially that both

$$\sup_{\lambda_1 a \le s \le \lambda_2 a} |B(s)|s^{-1/2} \text{ and } \sup_{\lambda_1 a \le s \le \lambda_2 a} |B(1-s)|s^{-1/2} \to_{\mathscr{D}} \sup_{\lambda_1 \le s \le \lambda_2} |W(s)|s^{-1/2}$$

as $a \downarrow 0$, it is routine to prove, using (2.6) and assuming (2.11), that when $\nu = 0$, (2.10) still holds when the supremum is taken over $[0,1]$.

Next we state the Chibisov–O'Reilly theorem for the bootstrapped uniform empirical and quantile processes in the same generality as given in Csörgő, Csörgő, Horváth and Mason (1986a) for the ordinary processes. These results can be deduced from Theorem 2.1 in exactly the same way as the corresponding results for the ordinary processes were derived from (2.1) and (2.2) [cf. the first proof of Theorem 4.2.1 and Corollary 4.3.1 in Csörgő, Csörgő, Horváth and Mason (1986a)].

Let $\mathscr{Q}$ denote the class of positive functions q defined on $(0,1)$ such that for any $q \in \mathscr{Q}$ there exists a $0 < \delta < \frac{1}{2}$ and an $\varepsilon > 0$ such that $q(s) \geq \varepsilon$ for all $\delta < s < 1 - \delta$ and both $q(s)$ and $q(1-s)$ are nondecreasing on $(0, \delta]$. A function $q \in \mathscr{Q}$ is called a Chibisov–O'Reilly function if and only if

$$\int_0^{1/2} s^{-1} e^{-cq^2(s)/s}\, ds < \infty \quad \text{and} \quad \int_0^{1/2} s^{-1} e^{-cq^2(1-s)/s}\, ds < \infty$$

for all $c > 0$.

THEOREM 2.2. *Let $m = m(n)$ be any sequence of positive integers converging to infinity. For any $q \in \mathscr{Q}$,*

$$\sup_{0<s<1} \left|\alpha_{m,n}(s) - \tilde{B}_m(s)\right|/q(s) \to_{\tilde{P}} 0$$

if and only if q is a Chibisov–O'Reilly function. Also, if $q \in \mathscr{Q}$ is a Chibisov–O'Reilly function and if condition (2.11) *is satisfied, then for any $0 < \lambda < \infty$,*

$$\sup_{\lambda/m \leq s \leq 1-\lambda/m} \left|\beta_{m,n}(s) - \tilde{B}_m(s)\right|/q(s) \to_{\tilde{P}} 0.$$

We note that the second statement is in fact true under the weaker assumption that $q \in \mathscr{Q}$ is such that both $q(s)/s^{1/2}$ and $q(1-s)/s^{1/2}$ converge to infinity as $s \downarrow 0$.

The linearity statements (2.3)–(2.6) were useful in proving Theorem 2.1. They play very important roles in various contexts, in particular when proving concrete examples of (1.6). Used in conjunction with Theorem 2.1, their analogues for $G_{m,n}$ and $U_{m,n}$ in (2.7) contained in the next theorem are just as important when proving concrete examples of (1.8).

THEOREM 2.3. *Let $m = m(n)$ denote any sequence of positive integers converging to infinity. Then*

$$\sup_{0<s<1} G_{m,n}(s)/s + \sup_{0<s<1} (1 - G_{m,n}(s))/(1-s) = O_{\tilde{P}}(1),$$

$$\sup_{U_{m,n}(0) \leq s < 1} s/G_{m,n}(s) + \sup_{0<s<U_{m,n}(1)} (1-s)/(1 - G_{m,n}(s)) = O_{\tilde{P}}(1),$$

$$\sup_{0<s<1} s/U_{m,n}(s) + \sup_{0<s<1} (1-s)/(1 - U_{m,n}(s)) = O_{\tilde{P}}(1),$$

and whenever there exists a constant $0 < C < \infty$ such that $m/n \leq C$ for all $n \geq 1$, then for any $0 < \lambda < \infty$,

$$\sup_{\lambda/m \leq s < 1} U_{m,n}(s)/s + \sup_{0<s\leq 1-\lambda/m} (1 - U_{m,n}(s))/(1-s) = O_{\tilde{P}}(1).$$

A complete proof would require showing that all eight terms that appear are stochastically bounded. This can be seen by simple separate arguments and observations, using the original statements (2.3)–(2.6). We omit the details.

Now we formulate a convergence theorem for the bootstrapped uniform empirical process indexed by functions. This is the bootstrapped analogue of Corollary 3.2 in Csörgő, Csörgő, Horváth and Mason (1986a). This corollary is a special case of Theorem 3.2 there and follows from the first relation in (2.1). Following an appropriately simplified version of the proof given there, with obvious small modifications requiring (2.11), relation (2.10) of Theorem 2.1 implies the result.

THEOREM 2.4. *If $\mathscr{L}$ is a class of functions l defined on* $(0,1)$ *such that each l can be written as $l = l_1 - l_2$, where l_1 and l_2 are nondecreasing left-continuous functions on* $(0,1)$ *and*

$$\sup_{l\in\mathscr{L}}\sup_{0<s\le\delta}\big(|l_1(s)|+|l_2(s)|+|l_1(1-s)|+|l_2(1-s)|\big)\frac{s^{1/2}}{L(s)}\to 0$$

as $\delta\downarrow 0$, where L is a fixed positive nonincreasing function defined on $(0,\frac{1}{2}]$ *and slowly varying at zero, then*

$$\sup_{l\in\mathscr{L}}\left|\int_0^1 l(s)\,d\alpha_{m,n}(s)-\int_{1/m}^{1-1/m} l(s)\,d\tilde{B}_m(s)\right|\Big/L\left(\frac{1}{m}\right)\to_{\tilde{P}} 0,$$

provided condition (2.11) *holds.*

Recall (1.1) and let X be a rv with a nondegenerate distribution function F. In this case, the function

$$L(s)=\left(\int_s^{1-s}\int_s^{1-s}(u\wedge v-uv)\,dQ(u)\,dQ(v)\right)^{1/2} \tag{2.15}$$

is positive and finite for $0<s\le\gamma$ with some $0<\gamma\le\frac{1}{2}$, and for $\gamma\le s\le\frac{1}{2}$ we put $L(s)=L(\gamma)$. Let $I(\cdot)$ be the indicator function. The classical normal convergence criterion

$$\lim_{x\to\infty} x^2P\{|X|\ge x\}/E\big(|X|^2I(|X|<x)\big)=0 \tag{2.16}$$

can be completely described in terms of Q [Csörgő, Csörgő, Horváth and Mason (1986b)] and it follows in particular from (2.16) that L is slowly varying at zero. The two statements of the last result of the present section are the bootstrapped counterparts of Theorem 2.1 and Corollary 2.1 in Csörgő, Csörgő, Horváth and Mason (1986b), the second one following from the first and the first one following from Theorem 2.4 with almost verbatim proofs as given for the original results.

THEOREM 2.5. *Let F be nondegenerate and assume* (2.11). *If* (2.16) *holds, then with L in* (2.15),

$$\sup_{0\le t\le 1}\left|\int_0^t \alpha_{m,n}(s)\,dQ(s)-\int_0^t \tilde{B}_m^*(s)\,dQ(s)\right|\Big/L\left(\frac{1}{m}\right)\to_{\tilde{P}} 0.$$

If $EX^2<\infty$, then

$$\sup_{0\le t\le 1}\left|\int_0^t \alpha_{m,n}(s)\,dQ(s)-\int_0^t \tilde{B}_m(s)\,dQ(s)\right|\to_{\tilde{P}} 0.$$

3. The general bootstrapped empirical process on the real line. The two examples below show the asymptotic validity of bootstrap confidence-band estimation of distribution and quantile functions. Example 1, in the special case $q \equiv 1$, and Example 2 are originally due to Bickel and Freedman (1981).

EXAMPLE 1. Let F be an arbitrary distribution function and $m = m(n)$ be a sequence of positive integers such that $m(n) \to \infty$. It follows from Theorem 2.2 and (2.8) that on the probability space $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P})$ of Section 2, for each $n \geq 1$,

$$\begin{aligned}\{(F_n(x), \tilde{F}_{m,n}(y)) \colon -\infty < x, y < \infty\}& \\ =_{\mathscr{D}} \{(F_n(x), \hat{F}_{m,n}(y)) \colon -\infty < x, y < \infty\},&\end{aligned} \tag{3.1}$$

where F_n on the right-hand side is defined in terms of $X_1 = Q(U_1), \dots, X_n = Q(U_n)$ and hence $F_n(x) = G_n(F(x))$, $-\infty < x < \infty$, and where $\hat{F}_{m,n}(y) = G_{m,n}(F(y))$, $-\infty < y < \infty$, and

$$\sup_{-\infty < t < \infty} \left| m^{1/2}\{\hat{F}_{m,n}(t) - F_n(t)\} - \tilde{B}_m(F(t)) \right| / q(F(t)) \to_{\tilde{P}} 0,$$

where q is any Chibisov–O'Reilly function. This approximation with the weight $q(F(\cdot))$ in the bottom often provides a quick and easy means to establish the validity of the bootstrap in many situations.

EXAMPLE 2. Assume that F has a continuous density f such that $f(Q(t)) > 0$ for any $a \leq t \leq b$, where $0 < a < b < 1$. Let $m = m(n)$ be a sequence of positive integers such that condition (2.11) is satisfied. Elementary arguments based on Theorem 2.2 and (2.8) show that on the probability space $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P})$ of Section 2, for each $n \geq 1$,

$$\{(Q_n(s), \tilde{Q}_{m,n}(t)) \colon a \leq s, t \leq b\} =_{\mathscr{D}} \{(Q_n(s), \hat{Q}_{m,n}(t)) \colon a \leq s, t \leq b\} \tag{3.2}$$

and

$$\sup_{a \leq t \leq b} \left| m^{1/2}\{\hat{Q}_{m,n}(t) - Q_n(t)\} - \tilde{B}_m(t)/f(Q(t)) \right| \to_{\tilde{P}} 0,$$

where $\hat{Q}_{m,n}(t) = Q(U_{m,n}(t))$, $a \leq t \leq b$.

Next we formulate the generalized anålogue of Theorem 2.4 for an arbitrary distribution function F and a sequence of integers $m = m(n) \to \infty$. For simplicity we assume $L \equiv 1$. This is the bootstrapped analogue of Theorem 1.1 in Csörgő, Csörgő, Horváth and Mason (1986c). If we change m to n in the proof of the latter result, we see that it remains valid word for word and hence shows that Theorem 2.4 and (2.8) imply this analogue.

THEOREM 3.1. *Let $\mathscr{H}_F$ be a class of Borel measurable functions h defined on $(Q(0), Q(1))$, the support of F, such that the class $\mathscr{L} = \{l(s) = h(Q(s)), 0 < s < 1\colon h \in \mathscr{H}_F\}$ satisfies the conditions of Theorem 2.4 with $L \equiv 1$. Then on the space*

$(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P})$ *of Section* 2 *we have* (3.1) *and*

$$\sup_{h \in \mathscr{H}_F} \left| \int_{-\infty}^{\infty} h(x)\, d\hat{f}_{m,n}(x) - \int_{Q(1/m)}^{Q(1-1/m)} h(x)\, d\tilde{B}_m(F(x)) \right| \to_{\tilde{P}} 0$$

as $n \to \infty$, *where*

$$\hat{f}_{m,n}(x) = m^{1/2}\{\hat{F}_{m,n}(x) - F_n(x)\}. \tag{3.3}$$

This result is mathematically quite nice in the sense that it provides a meaningful approximation even in the case when the supremum over $\mathscr{H}_F$ of the approximating Gaussian sequence goes to infinity. In this case, which shows the strength of the approximation, the two sequences blow up together. The final step allows us to extend the integration to $(-\infty, \infty)$ in the approximating sequence and ensures at the same time that it remains bounded. With a view toward applications, just as in the primary limit theory, this we do for classes $\mathscr{H}_F$ admitting a Euclidean parametrization. Having Theorem 3.1 above, the proof of Theorem 1.2 in Csörgő, Csörgő, Horváth and Mason (1986c) is again valid word for word to give its bootstrapped analogue:

THEOREM 3.2. *Let* $\mathscr{H}_F = \{h_{\mathbf{t}}(\cdot): \mathbf{t} \in [a, b]^d\}$ *be a function class satisfying the conditions of Theorem* 3.1, *where* $[a, b]$ *is a finite interval and* $d \geq 1$ *is an integer. Assume that the function*

$$d^2_{\mathscr{H}}(\mathbf{s}, \mathbf{t}) = d^2_{\mathscr{H}_F}(\mathbf{s}, \mathbf{t}) = \int_{-\infty}^{\infty} \{(h_{\mathbf{s}}(x) - h_{\mathbf{t}}(x))^2\, dF(x)\}$$

is continuous on $[a, b]^d \times [a, b]^d$ *and let* $N_{d_{\mathscr{H}}}(\varepsilon)$ *be the minimum number of* $d_{\mathscr{H}}$*-balls with centers in* $[a, b]^d$ *and radii at most* $\varepsilon > 0$ *that cover* $[a, b]^d$, *where a* $d_{\mathscr{H}}$*-ball with center* $\mathbf{t}$ *and radius* $\delta > 0$ *is the set* $\{\mathbf{s}: d_{\mathscr{H}}(\mathbf{s}, \mathbf{t}) < \delta\}$. *If, in addition, the metric entropy condition*

$$\int_0^{\bar{d}_{\mathscr{H}}} (\log N_{d_{\mathscr{H}}}(\varepsilon))^{1/2}\, d\varepsilon < \infty$$

is also satisfied, where $\bar{d}_{\mathscr{H}} = \sup\{d_{\mathscr{H}}(\mathbf{s}, \mathbf{t}): \mathbf{s}, \mathbf{t} \in [a, b]^d\}$ *is the* $d_{\mathscr{H}}$*-diameter of* $[a, b]^d$, *then, on the probability space* $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P})$ *of Section* 2 *we have* (3.1) *and*

$$\sup_{\mathbf{t} \in [a, b]^d} \left| \int_{-\infty}^{\infty} h_{\mathbf{t}}(x)\, d\hat{f}_{m,n}(x) - \int_{-\infty}^{\infty} h_{\mathbf{t}}(x)\, d\tilde{B}_m(F(x)) \right| \to_{\tilde{P}} 0$$

as $n \to \infty$, *where* $m = m(n)$ *satisfies condition* (2.11) *and* $\hat{f}_{m,n}$ *is given in* (3.3).

As mentioned in Csörgő, Csörgő, Horváth and Mason (1986c), the entropy condition is the presently available nicest sufficient condition for the sample continuity of the Gaussian process $\{\int_{-\infty}^{\infty} h_{\mathbf{t}}(x)\, dW(F(x)),\ t \in [a, b]^d\}$, where W

is a standard Wiener process and is implied by the simpler condition that

$$\int_0^{\delta} \frac{\phi_{\mathscr{H}}(\varepsilon)}{\sqrt{\varepsilon \log \varepsilon^{-1}}}\, d\varepsilon < \infty \quad \text{for some } \delta > 0, \tag{3.4}$$

where, with $\|\cdot\|$ standing for the maximum norm in $\mathbb{R}^d$, $\phi_{\mathscr{H}}(\varepsilon) = \sup\{d_{\mathscr{H}}(\mathbf{s},\mathbf{t})$: $\mathbf{s},\mathbf{t} \in [a, b]^d$, $\|\mathbf{s} - \mathbf{t}\| \le \varepsilon\}$.

4. Examples: Bootstrapping reliability, concentration and moment-type functions. Throughout this section we assume (2.11).

EXAMPLES 3–6. As discussed in Csörgő, Csörgő and Horváth (1986), one of the most important problems of statistical reliability theory and survival analysis is the estimation of the mean residual life function and the total time on test function, while the estimation of the Lorenz curve and its inverse, the Goldie concentration curve, of a distribution is of utmost importance in economic concentration theory. The primary limit theory for the corresponding empirical "reliability and concentration" processes is worked out in Csörgő, Csörgő and Horváth (1986). In particular, (1.6) is proved for these four processes in Theorems 4.1, 6.2, 11.2 and 13.5 there, respectively, under optimal conditions. That the bootstrap method of constructing confidence bands for these four functions is asymptotically valid was also proved in Section 17 of Csörgő, Csörgő and Horváth (1986). However, this was done [with almost sure convergence in (1.9)] under a moment condition stronger than the necessary second-order moment conditions of the just listed four primary limit theorems. Now these bootstrap results follow from Theorems 2.1, 2.2 and 2.3 and the second statement of Theorem 2.5 under the original optimal conditions of the primary limit theorems. This can be seen by more or less repeating the proofs of the primary theorems. The details for the mean residual life and the total time on test functions can be found in the unpublished report Csörgő, Csörgő and Mason (1984), which is a preliminary version of the present paper.

It should be pointed out that for one of the four functions, the mean residual life, it is possible to construct confidence bands without the bootstrap [see Corollary 23.5.1 in Shorack and Wellner (1986)].

EXAMPLE 7. Another reliability function of interest is, for $0 < p < 1$ fixed, the $(1-p)$-percentile residual life function $R^{(p)}(t) = Q(1 - p(1 - F(t))) - t$, $t \ge 0$, where $F(0) = 0$. The case $p = \frac{1}{2}$ gives what is called the median residual lifetime, a notion competitive to that of the mean residual lifetime in Example 3 above. Note that $R^{(1/2)}(0) = Q(\frac{1}{2})$. The primary estimation theory based on the empirical counterpart $R_n^{(p)}(t) = Q_n(1 - p(1 - F_n(t))) - t$, $t \ge 0$, is given in Csörgő and Csörgő (1987) together with numerous references to applications. Assume that F has a density f which is positive and continuous on the interval $(Q(1-p) - \varepsilon, Q(1 - p(1 - F(T))) + \varepsilon)$, where $T < Q(1)$ and $\varepsilon > 0$ is arbitrarily small. Complementing the proof of Theorem 1 in Csörgő and Csörgő (1987) with trivial details, we see that (2.1′) and (2.10′) and (2.12′) of Theorem 2.1, with the rates ignored, in conjunction with (2.8) imply that (3.1) and (3.2) are jointly true and

$$\sup_{0 \le t \le T} \left| n^{1/2}\left(R_n^{(p)}(t) - R^{(p)}(t)\right) - \frac{pB_n(F(t)) - B_n(1 - p(1 - F(t)))}{f\left(R^{(p)}(t) + t\right)} \right| \to_{\tilde{P}} 0 \tag{4.1}$$

and in obvious notation, provided condition (2.11) holds,

$$\sup_{0 \le t \le T} \left| m^{1/2}\left(R_{m,n}^{(p)}(t) - R_n^{(p)}(t)\right) - \frac{p\tilde{B}_m(F(t)) - \tilde{B}_m(1 - p(1 - F(t)))}{f\left(R^{(p)}(t) + t\right)} \right| \to_{\tilde{P}} 0. \tag{4.2}$$

Here the variance function of the approximating Gaussian processes is $\sigma^2(t) = p(1-p)(1 - F(t))/f^2(R^{(p)}(t) + t)$, so the choice of $\sigma_n(t)$ for (1.13) is obvious with F_n, f_n and $R_n^{(p)}$ replacing F, f and $R^{(p)}$, where f_n is an appropriate density estimator.

Similarly as for the mean residual life mentioned above, it is shown by Csörgő and Csörgő (1987) under more stringent conditions on f that it is possible to construct asymptotically distribution-free bands for $R^{(p)}(\cdot)$, which also avoid density estimation together with the bootstrap, if the unknown distribution of $\sup\{|pB(s) - B(1 - p(1 - s))|: 0 \le s \le 1\}$ is simulated once and for all, where $B(\cdot)$ is a Brownian bridge. [This simulation study is reported in Csörgő and Viharos (1988).] However, this is achieved by letting $T = T_n$ converge to $Q(1)$, at some rate depending on Q near 1, which in certain examples may cause very wide bands even for large n and in spite of the fact that they can be constructed on very short intervals only. This and another type of band, also not requiring the bootstrap but involving density estimation, together with the two types of bootstrap bands of (1.12) and (1.13) as applied to estimation of $R^{(1/2)}(\cdot)$ are investigated by Csörgő and Viharos (1988) using a data set consisting of $n = 840$ British strike duration times. The conclusion is that the bootstrap bands, based on 3000 bootstrap simulations with $m = n$, are much better, this time the constant-width band in (1.12) being nicer than the "equal precision" or standardized band of (1.13).

In a somewhat different but very closely related situation, where the bootstrap is definitely needed, one of us has constructed constant-width bootstrap bands for $F(\cdot)$ and $R^{(1/2)}(\cdot)$ for the Channing House data extensively investigated by Efron (1981). We chose $m = n = 97$ and with 3000 bootstrap simulations obtained very nice bands with nominal $1 - \alpha = 0.9$ in both cases [see Csörgő (1988) for details].

EXAMPLE 8. Consider estimating the moment generating function

$$M(t) = Ee^{tx} = \int_{-\infty}^{\infty} e^{tx}\, dF(x)$$

on some t interval $[a, b]$, the choice of which depends on the interval I^* where $M(\cdot)$ is finite. (Roughly, one has to halve I^* if it is finite [see Csörgő (1982) or Csörgő, Csörgő, Horváth and Mason (1986c)].) The sample version of it is

$$M_n(t) = \frac{1}{n} \sum_{j=1}^{n} e^{tX_j} = \int_{-\infty}^{\infty} e^{tx}\, dF_n(x)$$

and since in Section 3 of Csörgő, Csörgő, Horváth and Mason (1986c) it is shown that the conditions of Theorem 3.2 are satisfied, we have the conclusion of that theorem with $h_t(x) = \exp(tx)$ and $d = 1$.

We note that many results and problems in Teugels' (1985) monograph suggest that the estimation of $M(\cdot)$ is of fundamental importance in insurance mathematics.

EXAMPLE 9. As pointed out in Section 4 of Csörgő, Csörgő, Horváth and Mason (1986c), the estimation of the Hall translated moment functions $H(t) = E|X + t|^p$ or $H^*(t) = E(|X + t|^p \operatorname{sgn}(X + t))$, $-\infty < t < \infty$, where $p > 1$ is a fixed number, may be of some statistical interest. Let $a < b$ be arbitrary finite numbers and assume that $E|X|^{2p} < \infty$. Then by Section 4 in Csörgő, Csörgő, Horváth and Mason (1986c), the conditions of Theorem 3.2 are again satisfied and we have the conclusion with $h_t(x) = |x + t|^p$ or $h_t(x) = |x + t|^p \operatorname{sgn}(x + t)$ and $d = 1$.

EXAMPLES 10 AND 11. Following Section 5 in Csörgő, Csörgő, Horváth and Mason (1986c), suppose $P\{X > 0\} = 1$ and consider estimating the moment function $K(t) = EX^t$ and the generalized mean function $L(t) = (K(t))^{1/t}$ on some t interval $[a, b]$, where $a < 0 < b$, under the assumption that $EX^{2a} + EX^{2b} < \infty$. Then, for the case of the moment function, by Example 8 we have the conclusion of Theorem 3.2 with $h_t(x) = x^t$. Substituting this conclusion and Theorem 2.2 for the corresponding primary results, the proof of Theorem 5.1 in Csörgő, Csörgő, Horváth and Mason (1986c) gives that

$$\sup_{a \le t \le b} \left| m^{1/2} \left\{ \left(\int_0^{\infty} x^t\, d\hat{F}_{m,n}(x) \right)^{1/t} - \left(\int_0^{\infty} x^t\, dF_n(x) \right)^{1/t} \right\} - \frac{L(t)}{K(t)} \int_0^{\infty} \frac{x^t - 1}{t}\, d\tilde{B}_m(F(x)) \right| \to_{\hat{P}} 0.$$

Since in all the above examples the corresponding primary limit theorems, i.e., the corresponding forms of (1.6), hold with approximating Gaussian sequences which are distributionally the same as those in the above conclusions [more specifically, they are the same stochastic integrals with respect to the Brownian bridges B_n from (2.1) in all the Examples 3–6, 8–11, while in Example 7 we have (4.1) and (4.2)], our final conclusion is that the bootstrap confidence bands (1.12) and (1.13) are asymptotically valid under the stated minimal conditions in each case. [Of course, for (1.13) one has to produce $\sigma_n(t)$, but this is usually routine as in Example 7.] The same is true in the last example that follows.

EXAMPLE 12. Kocherlakota and Kocherlakota (1986) have recently proposed that for discrete random variables statistical inference can reasonably be based on empirical probability generating functions. For simplicity, let X have nonnegative integer values and consider

$$P(t) = Et^X = \sum_{k=0}^{\infty} t^k P\{X = k\} = \int_0^{\infty} t^x \, dF(x), \qquad -1 \le t \le 1.$$

Its sample version is

$$P_n(t) = \frac{1}{n} \sum_{j=1}^{n} t^{X_j} = \int_0^{\infty} t^x \, dF_n(x), \qquad -1 \le t \le 1.$$

We are not aware that the primary limit theorem, i.e., the present special form of (1.6), was proved earlier. However, assuming $EX^2 < \infty$, it is easy to see that for $\mathscr{H} = \mathscr{H}_F = \{t^x, 0 \le x < \infty: \ -1 \le t \le 1\}$ we have $d_{\mathscr{H}}(s, t) \le (EX^2)^{1/2}|s - t|$, $-1 \le s,\ t \le 1$, which implies (3.4) and that the other conditions of Theorem 3.2 are also satisfied. Hence by Theorem 1.2 in Csörgő, Csörgő, Horváth and Mason (1986c) we get

$$\sup_{-1 \le t \le 1} \left| \int_0^{\infty} t^x \, d\alpha_n(F(x)) - \int_0^{\infty} t^x \, dB_n(F(x)) \right| \to_P 0$$

with the Brownian bridges B_n from (2.1). The approximating sequence here consists of copies of a mean-zero Gaussian process with covariance function $P(st) - P(s)P(t)$, $-1 \le s,\ t \le 1$. Also, we have the conclusion of Theorem 3.2 with $a = -1$, $b = 1$, $d = 1$ and $h_t(x) = t^x$.

For illustration we have chosen the (in)famous numbers 3, 5, 7, 9, 10, 18, 6, 14, 11, 9, 5, 11, 15, 6, 11, 17, 12, 15, 8, 4 of yearly deaths by horsekicks in the Prussian army recorded by von Bortkiewitz (1898) over the 20 years 1875–1894. Figure 1 depicts the constant-width band $\{P_{20}(t) \pm 0.155/\sqrt{20},\ t \in I\}$ of (1.12) by dashed lines and the standardized band $\{P_{20}(t) \pm 0.482(P_{20}(t^2) - P_{20}^2(t))^{1/2}/\sqrt{20},\ t \in I\}$ of (1.13) by dashed–dotted lines with nominal asymptotic coverage probabilities $1 - \alpha = 0.9$, where $I = [-0.9, -0.1] \cup [0.1, 0.9]$ and where $c_m = 0.155$ and $d_m = 0.482$ with $m = n = 20$ were obtained by 1000 Monte Carlo trials requiring 2.5 min of computer time on the (slow) R-55 machine of the University of Szeged. The middle solid curve is $P_{20}(t)$. Since $P_n(-1) = n^{-1}(\#\{\text{even observations}\} - \#\{\text{odd observations}\})$, the variance is largest at $t = -1$ and nearby [at $t = -1$ its estimator is $1 - P_n^2(-1)$], while it is zero at $t = 1$ and must be nearly zero at $t = 0$ if there is no zero observation. These features are reflected by the present bands.

We stopped the simulation at 100 and 500 replications, where the corresponding c_m values were 0.145 and 0.152 and those of d_m were 0.475 and 0.485. At

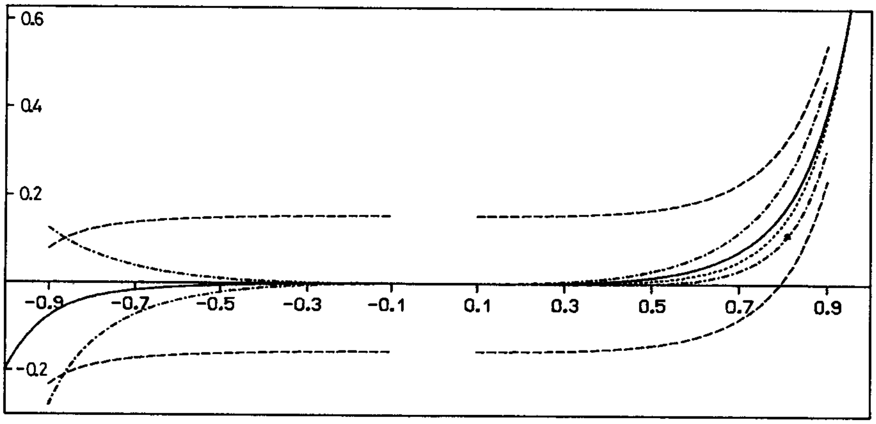

FIG. 1. *Deadly horsekicks in the Prussian army.*

$1 - \alpha = 0.75, 0.80, 0.85$ and 0.95 the discrepancies were similarly small. As a rule of thumb, 500 repetitions seem to suffice here.

Since this data set introduces the Poisson distribution to many students [Andrews and Herzberg (1985), page 17], we also plotted with dotted lines the generating function $P(t) = \exp(\lambda(t-1))$ of the Poisson (λ) distribution with $\lambda = \hat{\lambda} = 9.8$, the sample mean. This is 0.007 at $t = 0.5$ and practically zero below $t = 0.5$. Although this is not a bootstrap test (which would be easy to do and would lead to a similar picture), the picture speaks in favour of a Poisson distribution. [See, however, the interesting discussion in Section 9.5 of Bishop, Fienberg and Holland (1975).] It would be of interest to find out by simulation studies if $(m =)\ n = 20$ is sufficiently close to infinitely for distributions near to Poisson (10).

Similar results can be obtained for constructing bootstrap bands for the derivatives of $P(\cdot)$.

5. Bootstrapping the mean. Bickel and Freedman (1981) and Singh (1981) proved that the bootstrap of the mean works if the underlying distribution has a finite variance. Bretagnolle (1983) and Athreya (1985, 1987) proved that the naive bootstrap $(m \equiv n)$ of the mean fails when F is in the domain of attraction of a nonnormal stable law and hence it has infinite variance, and the bootstrap of the mean can work in this case only if $m/n \to 0$. The borderline situation between the two cases is when F has an infinite variance but still belongs to the domain of attraction of a normal law, the latter happening if and only if (2.16) holds. In this situation, using an ingenious direct proof based on an extension of the classical characteristic function method, Athreya (1985) has shown that bootstrapping the mean still works with $m \equiv n$. As another indication of the mathematical strength of the present general approach, we now point out that his result is contained in Theorem 2.5 for any $m = m(n)$ satisfying (2.11).

Indeed, introducing the sample mean $\bar{X}_n = n^{-1}\Sigma_{k=1}^n X_k$, by (2.8) we have for any $n \ge 1$ that

$$\left(\frac{1}{n^{1/2}}\sum_{j=1}^{n}(X_j - EX), \frac{1}{m^{1/2}}\sum_{j=1}^{m}(\tilde{X}_j - \bar{X}_n)\right)$$

$$=_{\mathscr{D}}\left(-\int_0^1 \alpha_n(s)\,dQ(s), -\int_0^1 \alpha_{m,n}(s)\,dQ(s)\right)$$

and for the independent standard normal rv's,

$$Z_n = -\int_{1/n}^{1-1/n} B_n(s)\,dQ(s)/L\left(\frac{1}{n}\right) \quad \text{and}$$

$$\tilde{Z}_m = -\int_{1/m}^{1-1/m} \tilde{B}_m(s)\,dQ(s)/L\left(\frac{1}{m}\right),$$

where $L(\cdot)$ is given in (2.15), we have by the proof of Corollary 2.2 in Csörgő, Csörgő, Horváth and Mason (1986b) and by Theorem 2.5 that

$$\left|-\int_0^1 \alpha_n(s)\,dQ(s)/L\left(\frac{1}{n}\right) - Z_n\right| \to_{\tilde{P}} 0$$

and

$$\left|-\int_0^1 \alpha_{m,n}(s)\,dQ(s)/L\left(\frac{1}{m}\right) - \tilde{Z}_m\right| \to_{\tilde{P}} 0.$$

Since L is slowly varying, by condition (2.11) we can replace $L(1/m)$ by $L(1/n)$ here. Furthermore, we have

$$\left\{n^{-1}\sum_{j=1}^{n} Q^2(U_j) - \left(n^{-1}\sum_{j=1}^{n} Q(U_j)\right)^2\right\}\Big/ L^2\left(\frac{1}{n}\right) \to_{\tilde{P}} 1.$$

[In the critical case when F has an infinite variance, this can be inferred, for example, either from Athreya (1985) or Csörgő and Mason (1987).] Introducing now the sample variance

$$S_n^2 = \frac{1}{n}\sum_{j=1}^{n} X_j^2 - \left(\frac{1}{n}\sum_{j=1}^{n} X_j\right)^2,$$

as the present forms of (1.9), under (2.16) and (2.11) we obtain

$$\text{(5.1)} \qquad \Pr\left\{\sum_{j=1}^{m}(\tilde{X}_j - \bar{X}_n)\Big/\left(m^{1/2}L\left(\frac{1}{n}\right)\right) \le x | X_1, \dots, X_n\right\} \to_{\Pr} \Phi(x)$$

and

$$\text{(5.2)} \qquad \Pr\left\{\sum_{j=1}^{m}(\tilde{X}_j - \bar{X}_n)/(m^{1/2}S_n) \le x | X_1, \dots, X_n\right\} \to_{\Pr} \Phi(x)$$

for any $x \in \mathbb{R}$, where Φ is the standard normal distribution function.

Assume for the rest of this section that $m \equiv n$. Bickel and Freedman (1981) and Singh (1981) have shown that (5.1) and (5.2) hold almost surely if $\mathrm{Var}(X) < \infty$. Recently, Giné and Zinn (1989) have proved that convergence in (5.1) takes place almost surely only if $\mathrm{Var}(X) < \infty$. Motivated by an idea of theirs, we now show that the same statement is true regarding (5.2). Moreover, we prove that (5.2) is true with $m = n$ if and only if (2.16) holds.

First suppose that for all $x \in \mathbb{R}$,

$$\Pr\left\{ \sum_{j=1}^{n} \left(\tilde{X}_j - \overline{X}_n\right)/\left(n^{1/2}S_n\right) \le x | X_1, \dots, X_n \right\} \to \Phi(x) \tag{5.3}$$

with probability 1. Choose any realization of $X_1, X_2, \dots$ such that (5.3) holds for all $x \in \mathbb{R}$ and set

$$\tilde{Y}_n = \left(\tilde{X}_n - \overline{X}_n\right)/\left(n^{1/2}S_n\right),$$

where $\tilde{X}_n$ has distribution function F_n. Obviously,

$$E_{F_n}\tilde{Y}_n = 0 \quad \text{and} \quad E_{F_n}\left(\tilde{Y}_n\right)^2 = \frac{1}{n},$$

so that by Chebyshev's inequality,

$$P_{F_n}\left\{|\tilde{Y}_n| > \varepsilon\right\} \to 0$$

for all $\varepsilon > 0$. Hence n independent copies of $\tilde{Y}_n$ form an infinitesimal array. Therefore, on account of (5.3) we must have [cf. Gnedenko and Kolmogorov (1954), Theorem 1, page 126]

$$nP_{F_n}\left\{|\tilde{Y}_n| > \varepsilon\right\} \to 0 \tag{5.4}$$

for all $\varepsilon > 0$. Since $nP_{F_n}(A)$ is necessarily a nonnegative integer for any event A, (5.4) implies that

$$\max_{1 \le j \le n} |X_j - \overline{X}_n|/\left(n^{1/2}S_n\right) \to 0. \tag{5.5}$$

We easily see that (5.5) in turn implies that

$$\max_{1 \le j \le n} X_j^2 \Big/ \sum_{k=1}^{n} X_k^2 \to 0. \tag{5.6}$$

Thus we have shown that (5.3) holding almost surely forces (5.6) to be true with probability 1. By Lemma 4.1 of Maller and Resnick (1984) this can happen only if $0 < EX^2 < \infty$.

Next suppose that (5.3) holds in probability. Then for every subsequence $\{n_1\} \subset \{n\}$ there exists a further subsequence $\{n_2\} \subset \{n_1\}$ such that (5.3) holds almost surely along $\{n_2\}$. Repeating the above argument we get that (5.6) is true along $\{n_2\}$ almost surely. This says that

$$\max_{1 \le j \le n} X_j^2 \Big/ \sum_{k=1}^{n} X_k^2 \to_P 0. \tag{5.7}$$

From Theorem 2 of Breiman (1965) it can be inferred that (5.7) is equivalent to (2.16).

APPENDIX

Passage to (1.9) and (1.11). In Proposition 1 we work on the space $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P})$ of (1.8). Then (1.9) results from (1.8b).

PROPOSITION 1. *If* (1.6) *and* (1.8) *hold, then at each continuity point x of G_F,*

$$Y_n(x) = \tilde{P}\Big\{\sup_{t\in I}|\hat{r}_{m,n}(t)| \le x | X_1, \dots, X_n\Big\} \to_{\tilde{P}} G_F(x). \tag{A.1}$$

PROOF. For any $\varepsilon > 0$ consider the event

$$A_{m,n}(\varepsilon) = \Big\{\sup_{t\in I}|\hat{r}_{m,n}(t) - \tilde{\mathscr{G}}_F^{(m)}(t)| > \varepsilon\Big\}.$$

First we claim that

$$\tilde{P}\{A_{m,n}(\varepsilon)|X_1, \dots, X_n\} \to_{\tilde{P}} 0. \tag{A.2}$$

Setting $B_{m,n}(\varepsilon, \eta) = \{\tilde{P}\{A_{m,n}(\varepsilon)|X_1, \dots, X_n\} > \eta\}$, we see by restricting the domain of integration from $\tilde{\Omega}$ to $B_{m,n}(\varepsilon, \eta)$ that

$$\tilde{P}\{A_{m,n}(\varepsilon)\} = E\big(\tilde{P}\{A_{m,n}(\varepsilon)|X_1, \dots, X_n\}\big) \ge \eta\tilde{P}\{B_{m,n}(\varepsilon, \eta)\}.$$

Hence (1.8c) implies that $\tilde{P}\{B_{m,n}(\varepsilon, \eta)\} \to 0$ for any $\eta > 0$ and this is (A.2). Also, obvious probability inequalities and (1.8a) give that

$$G_F(x - \varepsilon) - \tilde{P}\{A_{m,n}(\varepsilon)|X_1, \dots, X_n\} \le Y_n(x) \le G_F(x + \varepsilon) + \tilde{P}\{A_{m,n}(\varepsilon)|X_1, \dots, X_n\}$$

almost surely. These inequalities and (A.2) clearly imply (A.1). □

PROPOSITION 2. *If* (1.6), (1.9) *and* (1.10) *hold, then we have* (1.11) *and*

$$G_F(c_m) \to_{\Pr} 1 - \alpha \tag{A.3}$$

and if $c = d$, then

$$c_m \to_{\Pr} c. \tag{A.4}$$

PROOF. Choose any $\varepsilon > 0$ such that $c - \varepsilon$ and $d + \varepsilon$ are also continuity points of G_F. Putting $\tilde{r}_{m,n}(t) = m^{1/2}(\tilde{R}_{m,n}(t) - R_n(t))$, observe that

$$\Pr\{c_m < c - \varepsilon\} \le \Pr\Big\{\Pr\Big\{\sup_{t\in I}|\tilde{r}_{m,n}(t)| \le c - \varepsilon | X_1, \dots, X_n\Big\} \ge 1 - \alpha\Big\},$$

but since by (1.9),

$$\Pr\Big\{\sup_{t\in I}|\tilde{r}_{m,n}(t)| \le c - \varepsilon | X_1, \dots, X_n\Big\} \to_{\Pr} G_F(c - \varepsilon) < 1 - \alpha,$$

we see that $\Pr\{c_m < c - \varepsilon\} \to 0$. Similarly,

$$\Pr\{c_m > d + \varepsilon\} \leq \Pr\left\{\Pr\left\{\sup_{t \in I} |\tilde{r}_{m,n}(t)| \leq d + \varepsilon | X_1, \dots, X_n\right\} < 1 - \alpha\right\},$$

$$\Pr\left\{\sup_{t \in I} |\tilde{r}_{m,n}(t)| \leq d + \varepsilon | X_1, \dots, X_n\right\} \to_{\Pr} G_F(d + \varepsilon) > 1 - \alpha,$$

which implies that $\Pr\{c_m > d + \varepsilon\} \to 0$. Thus we have

$$\Pr\{c - \varepsilon \leq c_m \leq d + \varepsilon\} \to 1,$$

which yields $\Pr\{G_F(c - \varepsilon) \leq G_F(c_m) \leq G_F(d + \varepsilon)\} \to 1$. Clearly, all three statements follow. □

Acknowledgments. We thank Miklós Csörgő for his contribution to the preliminary version [Csörgő, Csörgő and Mason (1984)] of this paper. He has decided to withdraw from the present continuation of the project. We also thank Peter Gaenssler for pointing out two gaps and an associated inaccuracy in that preliminary version, which are filled in the Appendix. We are grateful to Mihály Tóth Abonyi for programming the horsekick example. The criticism of three referees and an Associate Editor was helpful in shaping the final version.

REFERENCES

ANDREWS, D. F. and HERZBERG A. M. (1985). *Data*. Springer, New York.

ATHREYA, K. B. (1985). Bootstrap of the mean in the infinite variance case, II. Technical Report 86-21, Dept. Statistics, Iowa State Univ.

ATHREYA, K. B. (1987). Bootstrap of the mean in the infinite variance case. *Ann. Statist.* **15** 724–731.

BERAN, R. (1984). Bootstrap methods in statistics. *Jahresber. Deutsch. Math.-Verein.* **86** 14–30.

BERAN, R. and MILLAR, P. W. (1986). Confidence sets for a multivariate distribution. *Ann. Statist.* **14** 431–443.

BERAN, R., LE CAM, L. and MILLAR, P. W. (1987). Convergence of stochastic empirical measures. *J. Multivariate Anal.* **23** 159–168.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BISHOP, Y. M. M., FIENBERG, S. E. and HOLLAND, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, Mass.

BREIMAN, L. (1965). On some limit theorems similar to the arc-sin law. *Theory Probab. Appl.* **10** 323–331.

BRETAGNOLLE, J. (1983). Lois limites du bootstrap de certaines fonctionelles. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **19** 281–296.

CSÖRGŐ, M. and CSÖRGŐ, S. (1987). Estimation of percentile residual life. *Oper. Res.* **35** 598–606.

CSÖRGŐ, M. and MASON, D. M. (1985). On the asymptotic distribution of weighted uniform empirical and quantile processes in the middle and on the tails. *Stochastic Process. Appl.* **21** 119–132.

CSÖRGŐ, M., CSÖRGŐ, S. and HORVÁTH, L. (1986). *An Asymptotic Theory for Empirical Reliability and Concentration Processes. Lecture Notes in Statist.* **33**. Springer, Berlin.

CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. M. (1986a). Weighted empirical and quantile processes. *Ann. Probab.* **14** 31–85.

CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. M. (1986b). Normal and stable convergence of integral functions of the empirical distribution function. *Ann. Probab.* **14** 86–118.

CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. M. (1986c). Supnorm convergence of the empirical process indexed by functions and applications. *Probab. Math. Statist.* **7** 13–26.

CSÖRGŐ, M., CSÖRGŐ, S. and MASON, D. M. (1984). Bootstrapping empirical functions. Technical Report No. 42, Lab. Res. Statist. Probab., Carleton Univ., Ottawa.

CSÖRGŐ, S. (1982). The empirical moment generating function. In *Nonparametric Statistical Inference* (B. V. Gnedenko, M. L. Puri and I. Vincze, eds.) 139–150. North-Holland, Amsterdam.

CSÖRGŐ, S. (1988). Estimation in the proportional hazards model of random censorship. *Statistics* **19** 437–463.

CSÖRGŐ, S. and MASON, D. M. (1987). Approximations of weighted empirical processes with applications to extreme, trimmed and self-normalized sums. In *Proc. First World Congress of Bernoulli Soc.* (Yu. V. Prohorov and V. V. Sazonov, eds.) **2** 811–819. VNU Sci. Press, Utrecht.

CSÖRGŐ, S. and VIHAROS, L. (1988). Confidence bands for the percentile residual lifetimes. *J. Statist. Plann. Inference*. To appear.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

EFRON, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.* **76** 312–319.

GAENSSLER, P. (1987). Bootstrapping empirical measures indexed by Vapnik–Chervonenkis classes of sets. In *Probability Theory and Mathematical Statistics* 467–481. VNU Sci. Press, Utrecht.

GINÉ, E. and ZINN, J. (1988). Bootstrapping general empirical measures. *Ann. Probab.* To appear.

GINÉ, E. and ZINN, J. (1989). Necessary conditions for the bootstrap of the mean. *Ann. Statist.* **17** 684–691.

GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Cambridge, Mass.

HALL, P. (1986). On the bootstrap and confidence intervals. *Ann. Statist.* **14** 1431–1452.

KOCHERLAKOTA, S. and KOCHERLAKOTA, K. (1986). Goodness of fit tests for discrete distributions. *Comm. Statist. A—Theory Methods* **15** 815–829.

LO, A. Y. (1987). A large sample study of the Bayesian bootstrap. *Ann. Statist.* **15** 360–375.

LOHSE, K. (1987). Consistency of the bootstrap. *Statist. Decisions* **5** 353–366.

MALLER, R. A. and RESNICK, S. I. (1984). Limiting behaviour of sums and the term of maximum modulus. *Proc. London Math. Soc.* (3) **49** 385–422.

MASON, D. M. and VAN ZWET, W. R. (1987). A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.* **15** 871–884.

SHEEHY, A. and WELLNER, J. A. (1986). Uniformity in P of some limit theorems for empirical measures and processes. Unpublished.

SHORACK, G. R. (1982). Bootstrapping robust regression. *Comm. Statist. A—Theory Methods* **11** 961–972.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

TEUGELS, J. L. (1985). *Selected Topics in Insurance Mathematics*. Katholieke Univ. te Leuven.

TSIREL'SON, J. L. (1975). The density of the distribution of the maximum of a Gaussian process. *Theory Probab. Appl.* **20** 847–856.

VON BORTKIEWITZ, L. (1898). *Das Gesetz der kleinen Zahlen*. Teubner, Leipzig.

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.

BOLYAI INSTITUTE
UNIVERSITY OF SZEGED
ARADI VÉRTANÚK TERE 1
H-6720 SZEGED
HUNGARY

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF DELAWARE
NEWARK, DELAWARE 19716