The Annals of Statistics
1977, Vol. 5, No. 5, 1039–1046

ON THE UNIFORMITY OF SEQUENTIAL PROCEDURES[1]

BY RAYMOND J. CARROLL

University of North Carolina at Chapel Hill

An extension of a central limit theorem for a mean of a random number of observations is given. A natural application occurs in the area of fixed-width confidence intervals. We provide an example which shows that the standard procedure does not preserve the intended coverage probability uniformly over nontrivial sets of distribution functions. The major weak convergence result is used to provide conditions for and a simple proof of such uniformity. The results are also shown to hold for M-estimates of location.

1. Introduction. This paper has evolved from a study of the uniformity of central limit theorem approximations for sequential fixed-width confidence intervals and related ranking and selection procedures. The main weak convergence result concerns a process

$$W_n(s, t) = n^{-\frac{1}{2}} \sum_{i=1}^{[ns]} \{G_t(X_i) - EG_t(X_1)\},$$

where $\{G_t\}$ is a family of functions. We prove W_n converges to a process in $C_2[0, \infty)$ and obtain the same result for a random sample size version $W_n(sN(t)/n, t)$ taking special care to assure the process is an element of $D_2[0, \infty)$. The topologies and conditions for convergence will be taken from Billingsley (1968) and Bickel and Wichura (1971), hereafter denoted B–W. The methods will extend to include dependence among $X_1, X_2, \cdots$, but we make no special effort in this direction.

In Section 3, we first find an example of a compact space of distributions on which a sequential central limit theorem is not uniform. This failure is due to the lack of tightness in a process closely related to W_n. The weak convergence results are then used to provide conditions under which the approximation is uniform. In Section 4 the results are extended beyond sample averages, the example being M-estimates, a class of robust estimates.

Throughout, "$\Rightarrow$" denotes weak convergence.

2. Weak convergence. Suppose $X_1, X_2, \cdots$ are i.i.d. uniform (0, 1) random variables and $\mathscr{G} = \{G_t : \int_0^1 G_t(x)\,dx = 0, 0 \leqq t < \infty\}$ a family of functions. We consider the following assumptions, with integrals taken over (0, 1).

(A1) $G_t(x)$ is continuous in t for each x and on each interval $[0, c]$ there exists

Received August 1974; revised September 1976.

[1] This research was supported in part by the Office of Naval Research Contract N00014-67-A-0001 and in part by the Air Force Office of Scientific Research under Grant No. AFOSR-75-2796.

AMS 1970 *subject classifications.* Primary 62L99; Secondary 62F07.

Key words and phrases. Sequential confidence intervals, weak convergence, two-dimensional processes.

$M, \delta > 0$ for which $\int |G_t(x)|^{2+\delta}\,dx$ is bounded and

$$\int |G_{t_2}(x) - G_{t_1}(x)|^4\,dx \leqq M|t_2 - t_1|^{1+\delta}\,, \qquad 0 \leqq t_1, t_2 \leqq c\,.$$

(A2) There exists an integer-valued stochastic process (in $D[0, \infty)$) $N = N(t) = N_n(t)$ and a stochastic process $y = y(t)$ in $C[0, \infty)$ with $N/n \to y$ in probability on $D[0, c]$.

(A3) There exists a second integer-valued process M in $D[0, \infty)$ with $M(t) \leqq N(t)$ and $M/n \to y$ in probability on $D[0, c]$.

We define three stochastic processes by

$$\begin{aligned} W_n(s, t) &= n^{-\frac{1}{2}} \textstyle\sum_{i=1}^{[ns]} G_t(X_i) \\ W_n{}^*(s, t) &= W_n(sN(t)/n, t) \\ H_n &= \sup\{|W_n{}^*(s, t) - W_n{}^*(1, t)| : 0 \leqq t \leqq c,\ M(t)/N(t) \leqq s \leqq 1\}\,. \end{aligned}$$

THEOREM 1. *Under assumption* (A1) *there is a process* W *in* $C_2[0, \infty)$ *with* $W_n \Rightarrow W$ *in* $D_2[0, \infty)$.

PROOF. It suffices to show weak convergence on $D_2[0, c]$ for an arbitrary c. Define $V_n(s, t) = W_n(s, t) + V_n{}^{(1)}(s, t)$, where

$$V_n{}^{(1)}(s, t) = (ns - [ns])G_t(X_{[ns]+1})/n^{\frac{1}{2}}\,.$$

V_n is an element of $C_2[0, \infty)$, so that we need only show that (i) W_n is tight and (ii) $V_n{}^{(1)} \Rightarrow 0$, for in this circumstance V_n would then be tight. Let $B = [s_1, s_2) \times [t_1, t_2)$ be any block (B–W, page 1658), where ns_1 and ns_2 are integers. Define $W_n(B) = W_n(s_1, t_1) - W_n(s_1, t_2) - W_n(s_2, t_1) + W_n(s_2, t_2)$. By (A1),

$$\begin{aligned} E|W_n(B)|^4 &\leqq M|t_2 - t_1|^{1+\delta}\{|s_2 - s_1|/n + (s_2 - s_1)^2\} \\ &\leqq M_1|t_2 - t_1|^{1+\delta}|s_2 - s_1|^{1+\delta} \end{aligned}$$

since $|s_2 - s_1| \geqq 1/n$. Hence for any pair of blocks B, C,

$$E|W_n(B)W_n(C)|^2 \leqq M_2|t_2 - t_1|^{1+\delta}|s_2 - s_1|^{1+\delta}\,,$$

proving by the extension to Theorem 3 of B–W (their page 1665) that W_n is tight. To prove that $V_n{}^{(1)} \Rightarrow 0$, it suffices to prove that a process $\nu_n \Rightarrow 0$, where $\nu_n(t) = \sup\{|G_t(X_i)| : 1 \leqq i \leqq n\}/n^{\frac{1}{2}}$. Since $\int |G_t(x)|^{2+\delta}\,dx$ is bounded, the finite dimensional distributions of ν_n converge in probability to zero, so it suffices to prove the tightness of ν_n. Letting $n^{\frac{1}{2}}c_n = E|G_{t_2}(X_1) - G_{t_1}(X_1)|$,

$$\begin{aligned} |\nu_n(t_2) - \nu_n(t_1)| &\leqq \sup\nolimits_{1 \leqq i \leqq n}\{|G_{t_2}(X_i) - G_{t_1}(X_i)| - n^{\frac{1}{2}}c_n\}/n^{\frac{1}{2}} + c_n \\ &\leqq n^{-\frac{1}{2}} \textstyle\sum_{i=1}^{n}\{|G_{t_2}(X_i) - G_{t_1}(X_i)| - n^{\frac{1}{2}}c_n\} + c_n \qquad (2.1) \\ &= \beta_n(t_2, t_1) \quad \text{(say)}. \end{aligned}$$

Thus for any $\varepsilon > 0$, for n sufficiently large and δ small,

$$\begin{aligned} &\Pr\{\min\{|\nu_n(t_2) - \nu_n(t_1)|, |\nu_n(t_3) - \nu_n(t_2)|\} > \varepsilon, \text{ some } t_3 - t_1 < \delta, t_1 \leqq t_2 \leqq t_3\} \\ &\quad \leqq \Pr\{\min\{\beta_n(t_2, t_1), \beta_n(t_3, t_2)\} > \varepsilon/2, \text{ some } t_3 - t_1 < \delta, t_1 \leqq t_2 \leqq t_3\}\,. \end{aligned}$$

Now, $E|\beta_n(t_2, t_1)|^4 \leqq M_3|t_2 - t_1|^{1+\delta}$ by (A1) so that ν_n is tight (Theorem 3 of B–W) and the proof is complete. □

$W_n{}^*$ is not always an element of $D_2[0, \infty)$. Happily, this is not a real problem as the following corollary shows that the process $W_n{}^*$ is "between" two processes both in $D_2[0, \infty)$ and both with the same weak limit. To this end, define

$$N^*(t) = \sup\{N(v): j/n \leqq v, t < (j+1)/n \text{ for some } j\}$$
$$N_*(t) = \inf\{N(v): j/n \leqq v, t < (j+1)/n \text{ for some } j\}.$$

LEMMA 1. *Assume* (A1) *and* (A2). *Then both* N^*/n *and* N_*/n *converge to* y *in probability on* $D[0, c]$ *for all* c. *Further, if we replace* N *in the definition of* $W_n{}^*$ *by either* N^* *or* N_*, *the resultant* $W_n{}^* \in D_2[0, \infty)$ *and* $W_n{}^* \Rightarrow W^*$. *Finally,* $H_n \to 0$ *in probability, where in the definition of* H_n, N *is replaced by* N^* *and* M *is replaced by* N_*.

PROOF. Both N^* and N_* are elements of $D[0, \infty)$ and the fact that they are constant on intervals guarantees the new processes $W_n{}^*$ are elements of $D_2[0, \infty)$. By Billingsley (Theorem 5.5), $N^*/n \Rightarrow y$ and $(W_n, N^*/n) \Rightarrow (W, y)$; similarly for N_*. Consider N^* only; we first must show that $W_n{}^* \Rightarrow W^*$ on $D_2[0, c]$ for arbitrary c. Now, the function h_n on $D_2[0, \infty) \times D[0, \infty)$ defined by

$$(x, z) \to x(sz^*(t), t)$$

is continuous at points $(x, z) \in C_2[0, \infty) \times C[0, \infty)$. This follows because if $x_m \to x$, $z_m \to z$, the convergence is uniform. Further, $h_m(x_m, z_m) \to h(x, z) = x(sz(t), t)$. Thus by Billingsley (Theorem 5.5), $W_n{}^* \Rightarrow W^*$. For the second part, note that for any $\eta > 0$, with probability approaching one as $n \to \infty$,

$$\begin{aligned} H_n \leqq \ & \sup_t |W_n{}^*(1, t) - W_n{}^*(1-\eta, t)| \\ & + \sup_{1-\eta \leqq s \leqq 1} \min\{\sup_t |W_n{}^*(s, t) - W_n{}^*(1, t)|, \\ & \qquad \sup_t |W_n{}^*(1-\eta, t) - W_n{}^*(s, t)|\}, \end{aligned} \tag{2.2}$$

the suprema on t over $[0, c]$. The first term on the right side of (2.2) converges in probability to zero as $n \to \infty$, $\eta \to 0$ by Chebychev's inequality and (A1), while the second term is bounded by the $D_2[0, c]$ modulus of continuity of B–W and hence also converges in probability to zero. □

A little thought shows that (A2) and (A3) could be weakened to $N/n \Rightarrow y$, $M/n \Rightarrow y$, $(W_n, N/n) \Rightarrow (W, y)$ and $(W_n, M/n) \Rightarrow (W, y)$.

3. An application. The sequential fixed-width confidence interval rules introduced by Chow and Robbins (1965) may be described as follows. The space of possible distributions is $\mathscr{F} = \{F_\theta : \theta \in I\}$, where θ is an indexing parameter, which we will take to be a scalar. We observe i.i.d. observations $Y_1, Y_2, \cdots$ from F_θ. Assume $n^{\frac{1}{2}}(\bar{Y}_n - \mu(\theta))/h(\theta) \Rightarrow N(0, 1)$, where h is continuous in θ. We wish to construct a confidence interval for $\mu(\theta)$ of length $2d$ which has coverage probability at least $1 - 2\alpha$ throughout $\mathscr{F}$. Let $\Phi(b) = 1 - \alpha$ and for a small constant $\alpha^* > 0$, define

$$\sigma(\theta) = \max\{a^*, h(\theta)\} \qquad s_{n\theta}^2 = \max\{a^*, n^{-1}\sum_1^n (Y_i - \bar{Y}_n)^2\}$$
$$n_d(\theta) = (b\sigma(\theta)/d)^2 \qquad N_d(\theta) = \inf\{n: n \geqq (bs_{n\theta}/d)^2\}.$$

The stopping rule $N = N_d(\theta)$ differs slightly from that of Chow and Robbins (1965), who show

$$\text{(3.1)} \qquad \inf_{F\in\mathscr{F}} \lim_{d\to 0} \Pr\{|\bar{X}_N - \mu(\theta)| \leqq d\} = 1 - 2\alpha .$$

This is rather unsatisfactory as it gives no guide to the user in deciding how small d must be for any particular θ. One of the appealing features of the t-test is that it holds its level *uniformly* over wide classes of distributions; one naturally asks whether these confidence interval procedures have the same property, i.e., one seeks conditions on $\mathscr{F}$ for which

$$\text{(3.2)} \qquad \lim_{d\to 0} \inf_{F\in\mathscr{F}} \Pr\{|\bar{X}_N - \mu(\theta)| \leqq d\} = 1 - 2\alpha .$$

The difference in the two is that if one believes the true distribution is a member of a particular parametric family $\mathscr{F} = \{F_\theta, \theta \in I\}$ and if I is compact (because of constraints on the process), there will be a real interest in assuring that the accuracy of probabilistic approximations does not depend on the "true" value of θ in this particular experiment, for a slight change in conditions may change this value of θ. The following example shows that care must be taken, that (3.2) may fail while (3.1) holds.

Let $U_1, U_2, \cdots$ be i.i.d. uniform (0, 1) random variables. Define for $0 \leqq \theta \leqq 1$ $a(\theta) = (\frac{1}{2})^{(1-\theta)/\theta}$, I_A the indicator of the event A, and

$$X_i(\theta) = (\theta/1-\theta)^{\frac{1}{2}}(1 - a(\theta))^{-1} I_{(U_i > a(\theta))} .$$

Thus for $\theta < 1$,

$$\mu(\theta) = EX_1(\theta) = (\theta/1-\theta)^{\frac{1}{2}}, \qquad \sigma^2(\theta) = \operatorname{Var} X_1(\theta) = (\theta/1-\theta)(2^{(1-\theta)/\theta} - 1)^{-1},$$

while for $\theta = 1$, $\mu(\theta) = \sigma^2(\theta) = 0$. Then

$$N_d(\theta)^{-1} \sum_1^{N_d(\theta)} (X_i(\theta) - \mu(\theta))$$

is a stochastic process in $D[0, 1]$, but choosing $\theta(d) = d^{-4}/(1 + d^{-4})$ we obtain

$$\begin{aligned} \lim_{d\to 0} \sup_{0\leqq\theta\leqq 1} &\Pr\{|N_d(\theta)^{-1} \textstyle\sum_1^{N_d(\theta)}(X_i(\theta) - \mu(\theta))| \geqq d\} \\ &\geqq \lim_{d\to 0} \Pr\{X_k(\theta(d)) = 0, k = 1, \cdots, (ba^*/d)^2\} \\ &= \lim_{d\to 0} (2^{-d^4})^{(ba^*/d)^2} = 1 , \end{aligned}$$

the inequality following because $X_k(\theta(d)) = 0$ for $k = 1, \cdots, (ba^*/d)^2$ means $N_d(\theta(d)) = (ba^*/d)^2$ and $\mu(\theta(d)) = d^{-2} \geqq d$. However, from Chow and Robbins (1965), it is clear that

$$\sup_{0\leqq\theta\leqq 1} \lim_{d\to 0} \Pr\{|N_d(\theta)^{-1} \textstyle\sum_1^{N_d(\theta)} (X_i(\theta) - \mu(\theta))| \geqq d\} \leqq 2(1 - \Phi(b)) .$$

The example fails because the distribution functions are not sufficiently close, and it appears that a weak convergence approach will most naturally describe the necessary tightness. We will suppose the distribution functions F_θ have inverses G_θ and take $\mu(\theta) \equiv 0$ for convenience of notation. The probability integral transforms gives

$$n^{-1} \sum_1^n Y_i =_D n^{-1} \sum_1^n G_\theta(X_i) \,,$$

where $X_1, X_2, \cdots$ are i.i.d. uniform (0, 1) random variables and the equality is in distribution ($s^2_{n\theta}$ and $N_d(\theta)$ may also be described in terms of G_θ). The latter sum is a stochastic process in θ and the weak convergence framework of Section 2 now applies. Lemma 1 becomes relevant because $N_d(\cdot)$ is not necessarily a stochastic process in $D[0, \infty)$, nor is a process corresponding to $W_n^*(1, \cdot)$.

THEOREM 2. *Let θ range over a compact indexing set Υ. Suppose* (A1) *and the following hold*:

(A4) $\lim_{n_0\to\infty} \Pr\{|s^2_{n\theta}/\sigma^2(\theta) - 1| > \varepsilon$ *for some* $0 \leqq \theta \leqq c, n \geqq n_0\} = 0$, *where* c *is arbitrary but finite.*

Then (3.2) *holds.*

We need the following proposition.

PROPOSITION 1. *Suppose $V_n \Rightarrow V$ in $D[0, \infty)$, where $V(t)$ is normally distributed with mean zero and variance at most M_*, a finite constant. Suppose* $\Pr\{V \in C[0, \infty)\} = 1$. *Then for any compact set* $\mathscr{T}$

$$\lim_{n\to\infty} \sup_{t\in\mathscr{T}} |\Pr\{|V_n(t)| \leqq b\} - \Pr\{|V(t)| \leqq b\}| = 0 \,.$$

PROOF. Let $A_t = \{x \in D[0, \infty) : |x(t)| \leqq b\}$ and $\mathscr{A} = \{A_t : t \in \mathscr{T}\}$. By Theorem 3 of Topsøe (1967), we have to show that if $\delta_n \to 0$, $t_n \in \mathscr{T}$,

$$P_V(\textstyle\bigcap_{n=1}^\infty (\partial A_{t_n})^{\delta_n}) = 0 \,.$$

Assume $\mathscr{A}$ is a V-continuity class. Then since $\{t_n\}$ has a limit point, this is shown to hold by following the method of proof of Topsøe's Theorem 8. To verify that $\mathscr{A}$ is a V-continuity class, one must show $P_V(\partial A_t) = 0$ for each t. But

$$\begin{aligned} P_V(\partial A_t) &= P_V(\partial(A_t \cap C)) \\ &= P_V\{x \in C : |x(t)| = b\} = 0 \,. \end{aligned}$$ □

PROOF OF THEOREM 2. Fix $c > 0$ and define on intervals $[0, a_{1n})$, $[a_{1n}, a_{2n})$, $\cdots$, $[a_{kn}, c]$ each of length at most $\exp(-n^2)$

$$\begin{aligned} s^2_{n2}(\theta) &= \sup\{s^2_{n\theta} : a_{jn} \leqq \theta \leqq a_{j+1,n}\} \\ s^2_{n1}(\theta) &= \inf\{s^2_{n\theta} : a_{jn} \leqq \theta \leqq a_{j+1,n}\} \end{aligned}$$

if $a_{jn} \leqq \theta < a_{j+1,n}$. Define

$$\begin{aligned} N_d^{(1)}(\theta) &= \inf\{n : n \geqq (bs_{n1}(\theta)/d)^2\} \\ N_d^{(2)}(\theta) &= \sup\{n : n \geqq (bs_{n2}(\theta)/d)^2\} \,. \end{aligned}$$

Clearly, $N_d^{(1)}(\theta) \leqq N_d(\theta) \leqq N_d^{(2)}(\theta)$ and (A2) and (A3) both hold with the replacements $n = n_d(c)$, $N = N_d^{(2)}(\theta)$, $M = N_d^{(1)}(\theta)$, and $y \equiv 1$. Thus the conclusion to Lemma 1 holds with the process W_n^* replaced by

$$W_d^{**}(s, \theta) = b(dN_d^{(1)}(\theta))^{-1} \sum_{i=1}^{[sN_d^{(2)}(\theta)]} G_\theta(X_i) .$$

Thus, for any $\varepsilon > 0$ and d sufficiently small, since

$$\sup \left\{ |W_d^{**}(s, \theta) - W_d^{**}(1, \theta)| : 0 \leqq \theta \leqq c, \frac{N_d^{(1)}(\theta)}{N_d^{(2)}(\theta)} \leqq s \leqq 1 \right\} \to_p 0 ,$$

we obtain

$$\begin{aligned} (3.3) \qquad \Pr\{|N_d^{-1}(\theta) \textstyle\sum_1^{N_d(\theta)} G_\theta(X_i)| \leqq d\} &= \Pr\{|b(dN_d(\theta))^{-1} \textstyle\sum_1^{N_d(\theta)} G_\theta(X_i)| \leqq b\} \\ &\geqq \Pr\{|W_n^{**}(1, \theta)| \leqq b - \varepsilon\} - \varepsilon . \end{aligned}$$

Now, the process $W_d^{**}(1, \cdot)$ converges weakly on $D[0, \infty)$ to a process (easily shown to be Gaussian with variances 1) W^{**} on $C[0, \infty)$, so that Proposition 1 (with b replaced by $b - \varepsilon$) says that as $d \to 0$, the last term in (3.3) is bounded uniformly in $\theta \in I$ by $2(\Phi(b - \varepsilon)) - (1 + 2\varepsilon)$. Letting $\varepsilon \to 0$ completes the proof. □

As a special case, consider the location-scale family $F((x - \mu)/\sigma)$. Then, $\Pr\{|\bar{X}_N - \mu| \leqq d \mid \mu, \sigma\} = \Pr\{|\bar{X}_N| \leqq d \mid \mu = 0, \sigma\}$. Thus, with $\theta = \sigma$, the conditions of Theorem 2 hold if $\int x^4 \, dF(x) < \infty$ since $G_{\mu,\sigma}(x) = \mu + \sigma G(x)$.

The same uniformity problem treated here arises in the context of ranking and selection. Letting CS denote a correct selection (such as correctly selecting the stochastically largest of k populations), most papers only show

$$\inf_{\mathscr{F}} \lim_{d \to 0} \Pr\{CS\} \geqq P^* .$$

Finally, results similar to that of Theorem 2 can be obtained by the simple expedient of requiring all the steps in Chow and Robbins (1965) to hold uniformly in $\mathscr{F}$, such as the convergence of N/n, the central limit theorem for sample sums, and uniform continuity in probability. However, such an approach will not obtain Theorem 1 and will not suffice when $(W_n, N/n)$ converges jointly to (W, y) and y is nonrandom. This problem has recently arisen in recent unpublished work of Swanepoel and Carroll, the former obtaining a class of sequential selection rules, the latter showing that the stopping times of these rules converge weakly to a random variable.

4. Extensions. The results of the previous sections continue to hold if the sample mean and variance are replaced by robust location estimators $T_{n\theta}$ and their associated variance estimates $g_{n\theta}^2$. The key idea here is that many robust estimators T_n look very much like sample averages of *bounded* random variables (Carroll (1975)). In the next example, we show the results hold for a whole class of smooth M-estimates.

DEFINITION. $\{T_{n\theta}\}$ converges to zero almost surely uniformly (denoted $T_{n\theta} \to 0$ (a.s.u.)) if for all c,

$$\sup\{|T_{n\theta}| : 0 \leqq \theta \leqq c\} \to 0 \quad \text{(a.s.)}.$$

EXAMPLE. Here we define $X_1, X_2, \cdots$ as i.i.d. uniform (0, 1) random variables and $T_{n\theta}$ by $\sum_1^n \psi(G_\theta(X_i) - T_{n\theta}) = 0$. ψ is bounded and nondecreasing with two bounded continuous derivatives. Further, $\psi'(x) > 0$ in a neighborhood of zero and $\psi'(x) = 0$ outside an interval $[-k, k]$. These ψ functions include smoothed versions of the Huber M-estimate. $g_{n\theta}^2$ is defined by

$$g_{n\theta}^2 = n^{-1} \sum_1^n \psi^2(G_\theta(X_i) - T_{n\theta})\{n^{-1} \sum_1^n \psi'(G_\theta(X_i) - T_{n\theta})\}^{-2} .$$

We again have the $\{T_{n\theta}\}$ with "mean" zero; more precisely, $\int \psi(x)\, dF_\theta(x) = 0$.

We assume that for every $c > 0$, there exists C_*, C_{**} such that

(B1) $\sup\{|G_{\theta_1}(x) - G_{\theta_2}(x)|/|\theta_2 - \theta_1| : 0 \leqq |x|, \theta_1, \theta_2 \leqq c\} \leqq C_*$

(B2) $\inf\{|\int \psi(x + \varepsilon)\, dF_\theta(x)| : 0 \leqq \theta \leqq c\} > 0$ for all $\varepsilon \neq 0$

(B3) $\inf\{F_\theta(C_{**}) - F_\theta(-C_{**}) : 0 \leqq \theta \leqq c\} > 0$,

LEMMA 2. *Under* (B1)—(B3), *the conclusion to Theorem* 2 *holds.*

PROOF. By detailed Taylor expansions one shows that

$$n^{\frac{1}{2}}|T_{n\theta} - n^{-1} \sum_1^n \psi(G_\theta(X_i))/\int \psi'(x)\, dF_\theta(x)| \to 0 \quad \text{(a.s.u.)}.$$

Now define $\rho_\theta(x) = a_1(\theta)\psi^2(x) + a_2(\theta)\psi'(x) + a_3(\theta)\psi(x)$, where

$$\begin{aligned} a_1(\theta) &= (E\psi'(x))^{-2} \\ a_2(\theta) &= -2E\psi^2(X)(E\psi'(X))^{-3} \\ a_3(\theta) &= -a_2 E\psi''(X) - a_1 E\psi(X)\psi'(X) , \end{aligned} \tag{4.1}$$

and the expectations are under F_θ. Using (4.1), and Taylor's theorem, one shows

$$g_{n\theta}^2 - h(\theta) - n^{-1} \sum_1^n \rho_\theta(G_\theta(X_i)) + \int \rho_\theta(x)\, dF_\theta(x)| \to 0 \quad \text{(a.s.u.)}, \tag{4.2}$$

where $h(\theta) = \int \psi^2(x)\, dF_\theta(x)\{\int \psi'(x)\, dF_\theta(x)\}^{-2}$ is clearly continuous in θ.

Now reconsider the proof of Theorem 1. One can, for example, redefine W_n by

$$\begin{aligned} W_n(s, t) &= n^{-\frac{1}{2}} T_{[ns]}(G_t(X_1), \cdots, G_t(X_n)) \quad && (s \geqq \tfrac{1}{2}) \\ &= W_n(\tfrac{1}{2}, t) && \text{(otherwise)}. \end{aligned} \tag{4.3}$$

(B1) shows that (A1) holds for the sample means generated by $\psi(G_\theta(X))(a_1(\theta))^{\frac{1}{2}}$ while (4.2) shows that (A4) holds for $s_{n\theta}^2 = \max\{a^*, g_{n\theta}^2\}$. Because of (4.1), the weak convergence arguments in Theorems 1 and 2 apply to processes such as (4.3), and the proof is complete. □

Acknowledgment. The author wishes to thank Professors S. Gupta and H. Rootzén for their encouragement and advice.

REFERENCES

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **12** 1656–1670.

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

CARROLL, R. J. (1975). An asymptotic representation for M-estimators and linear functions of order statistics. Institute of Statistics Mimeo Series # 1006, Univ. of North Carolina at Chapel Hill.

CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **38** 30.

TOPSØE, F. (1967). On the connection between *P*-continuity and *P*-uniformity in weak convergence. *Theor. Probability Appl.* **12** 241–250.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514