On the Uniformity of Sequential Ranking Procedures*

by

Raymond J. Carroll

Department of Statistics
University of North Carolina at Chapel Hill

May, 1976

Institute of Statistics Mimeo Series No. 1067

* This research was supported in part by the Office of Naval Research Contract N00014-67-A-00014 and in part by the Air Force Office of Scientific Research under Grant No. AFOSR-75-2796.

SUMMARY

On the Uniformity of Sequential Ranking Procedures*

by

Raymond J. Carroll

In the class of ranking procedures based on sequential confidence intervals of fixed-length, the probability statements are not uniform in the scale parameter. Further the available generalizations to rankings based on stochastic orderings are not uniform over the parameter space Ω. Two methods are proposed for solving such problems; the first is based on the theory of weak convergence while the second is more direct but only solves the problem when Ω is compact.

American Mathematical Society 1970 subject classifications: Primary 62L99; Secondary 62F07.

Key words and phrases: Ranking and Selection, Sequential Confidence Intervals, Uniformity in Weak Convergence.

---

* This research was supported in part by the Office of Naval Research Contract N00014-67-A-00014 and in part by the Air Force Office of Scientific Research under Grant No. AFOSR-75-2796.

1. Introduction

In a certain class of sequential ranking and selection procedures based on fixed-length confidence intervals (Geertsema (1972), Chow and Robbins (1965)), the standard work assumes an unknown d.f. $F_{\mu,\sigma}(x) = F(\sigma^{-1}(x-\mu))$ and provides, for each fixed F and fixed k-vector $\underset{\sim}{\sigma}$, a statement $\lim_{d\to 0} \inf_{\underset{\sim}{\mu}} \Pr\{CS|\underset{\sim}{\mu},\underset{\sim}{\sigma}\} = P^*$, where d is the length of the preference zone $\Omega(d)$, the infimum is taken over $\Omega(d)$ and CS denotes a correct selection. However, the main historical concern in ranking theory has been lower bounds for $\Pr\{CS|\underset{\sim}{\mu},\underset{\sim}{\sigma}\}$ uniform in both $\Omega(d)$ and certain sets of nuisance parameters, so that in fact one requires

$$\lim_{d\to 0} \inf_{\underset{\sim}{\mu},\underset{\sim}{\sigma}} \Pr\{CS|\mu,\sigma\} = P^* . \tag{1}$$

In the general problem, the distributions take the form $F_{\mu,\sigma}(x) = F_\sigma(x-\mu)$, where σ ranges over an indexing set F (which is taken here as a subset of $[0,\infty)$); the ranking might be with respect to EX. Taking $\mu \equiv 0$, this example includes the case of stochastic orderings.

Although important applications of our results are in ranking theory, it turns out that if one applies the sequential rule defined below (with b chosen so that $\int \Phi^{k-1}(x+b)d\Phi(x) = P^*$) to each of the k independent populations and chooses the population with the largest observed value of the ranking statistic $T_{n\theta}$, then to show (1) it is sufficient to show (4) below, i.e., to show the uniformity of the corresponding sequential fixed-width confidence interval. The problem of uniform bounds for the vast number of such confidence intervals is itself an important and neglected area, as the usual results show that under $F_\sigma(x-\mu)$,

$$\text{(2)} \qquad \inf_{\sigma} \lim_{d\to 0} \Pr\{\text{correct coverage}\} \geq P^* \ .$$

The object of this work then is to find sequential selection and confidence interval procedures and conditions under which (1) and (4) hold for all compact indexing sets F. That compactness alone is not sufficient to guarantee (1) and (4) is shown in the counter-example of Section 2; the problem is that a certain stochastic process in $D[0,1]$ (Billingsley (1968)) fails to be tight. This leads to the presentation in Section 3 (see Theorem 1) of a weak convergence theory which guarantees (1) and (4) in the sample means case. In Section 4, two examples are given which illustrate the use of Theorem 1. Finally, in Section 5, a generalization of Theorem 1 is presented.

Tong (1971) considered a problem similar to (1) in the fixed sample size case, making use of results on the uniformity of convergence (in θ) in the Central Limit Theorem (Parzen (1954)). These results are not applicable in our case since the number of observations is random; hence, this paper might also be considered a slight extension of Parzen's work to the sequential case.

To set the notation for our solution to the confidence interval and ranking problems, we assume there are i.i.d. observations $X_1, X_2, \ldots$ from a d.f. $F_\theta(x)$. One defines for each n a location-scale equivariant statistic $T_{n\theta}$ for which, under F_θ , $\sqrt{n}\big(T_{n\theta}-\mu(\theta)\big)/h(\theta) \xrightarrow{L} N(0,1)$ and the ranking is with respect to $\mu(\theta)$. By translation invariance, one may assume $\mu(\theta) \equiv 0$. Hence, θ may be considered a scalar and $h(\theta)$ will satisfy a Lipschitz condition. For the rest of this paper, assume for convenience that $h(\theta) \equiv \theta$. One also defines a sequence $g^2_{n\theta}$ of location invariant, scale equivariant statistics for which $g^2_{n\theta} \to \theta$ almost surely. Finally, define for $a^* > 0$,

$$\sigma(\theta) = \max\{a^*,\theta\}, \qquad s^2_{n\theta} = \max\{a^*, g^2_{n\theta}\}$$

$$n_d(\theta) = [b\sigma(\theta)/d]^2$$
$$N_d(\theta) = \inf\{n\geq 5 : n \geq (bs_{n\theta}/d)^2\}.$$

Although $N_\theta(d)$ is slightly different from Chow and Robbins' (1965) rule, choosing a^* very small results in no great change. The problem is to show for all compact F in $[0,\infty)$ that

(4) $$\lim_{d\to 0} \inf_{\theta\in F} \Pr\{|T_{N_d(\theta),\theta}|\leq d\} = 2\Phi(b) - 1,$$

where θ is the standard normal d.f. Note that θ need not be a scale parameter, and F_θ need not be known.

2. An Example

The following example (which can easily be extended to the ranking problem) shows that even if $T_{n\theta}$ is the sample mean and $g_{n\theta}^2$ the sample variance, it is possible for $F = [0,1]$ that (4) fails but (2) does not. Let $U_1, U_2, \ldots$ be i.i.d. uniform $(0,1)$ random variables. Define for $0\leq\theta\leq 1$ $a(\theta) = (\frac{1}{2})^{(1-\theta)/\theta}$, I_A the indicator of the event A, and

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

Thus for $\theta < 1$,

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

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

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

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

$$\lim_{d\to 0} \sup_{0\leq\theta\leq 1} \Pr\{|N_d(\theta)\sum_1^{N_d(\theta)} \left(X_i(\theta)-\mu(\theta)\right)|\geq d\} \geq \lim_{d\to 0} \Pr\{X_k\left(\theta(d)\right)=0,\ k=1,\ldots,(ba^*/d)^2\}$$
$$= \lim_{d\to 0} (2^{-d^4})^{(ba^*/d)^2} = 1,$$

the inequality following because $X_k(\theta(d)) = 0$ for $k = 1,\dots,(ba^*/d)^2$ means $N_d(\theta(d)) = (ba^*/d)^2$ and $\mu(\theta(d)) = d^{-2} \geq d$. However, from Anscombe (1952), it is clear that

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

3. Weak Convergence

Suppose the distributions F_θ have inverses G_θ which are continuous in θ. The crucial idea is that if $X_1, X_2, \dots$ are i.i.d. uniform (0,1) random variables, the statistics $T_{n\theta} = T_n(G_\theta(X_1),\dots,G_\theta(X_n))$ form a stochastic process in θ. To exploit this idea, the first example considered is that of the sample mean $T_{n\theta} = n^{-1} \sum_1^n G_\theta(X_i)$. $g_{n\theta}^2$ will be the sample variance. Consider the following assumptions:

(A1) On each finite interval, $\int G_\theta^4(x)dx$ is bounded and there exist positive constants M_0, δ_0 such that

$$\int (G_{\theta_1}(x)-G_{\theta_2}(x))^4 dx \leq M_0|\theta_2-\theta_1|^{1+\delta_0}.$$

(A2) For arbitrary positive constants ε, M,

$$\lim_{n_0\to\infty} \Pr\{|s_{n\theta}^2/\sigma^2(\theta) - 1| > \varepsilon \text{ for some } 0\leq\theta\leq M,\ n\geq n_0\} = 0.$$

Note that (A1) and (A2) hold in the important special case of scale orderings $(F_{\mu,\theta}(x) = F(\frac{x-\mu}{\theta}))$ as long as $\int x^4 dF(x) < \infty$. The key is the following lemma.

<u>Lemma 1</u>. 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 F

$$\lim_{n\to\infty} \sup_{t\in F} \left|\Pr\{|V_n(t)|\le b\} - \Pr\{|V(t)|\le b\}\right| = 0.$$

<u>Proof</u>: Let $A_t = \{x \in D[0,\infty) : |x(t)|\le b\}$ and $\mathcal{A} = \{A_t : t\in F\}$. By Theorem 3 of Topsøe (1967), we have to show that if $\delta_n \to 0$, $t_n \in F$,

$$P_V\left(\bigcap_{n=1}^{\infty} (\partial A_{t_n})^{\delta_n}\right) = 0. \tag{5}$$

Assume $\mathcal{A}$ is a V-continuity class. Then since $\{t_n\}$ has a limit point (5) is shown to hold by following the method of proof of Topsøe's Theorem 8. To verify that $\mathcal{A}$ is a V-continuity class, one must show $P_V(\partial A_t) = 0$ for each t. But

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

<u>Theorem 1</u>. Under (A1) and (A2), if F is compact, (4) and hence (1) hold.

<u>Proof</u>: We freely use the notation and results of Bickel and Wichura (1971), said paper hereafter denoted by B-W. Fix $M^* > 0$ and define on intervals $[0,a_{1n}),[a_{1n},a_{2n}),\ldots,[a_{kn},M^*]$ each of length at most $\exp(-n^2)$

$$g_{n2}^2(\theta) = \sup\{g_{n\theta}^2 : a_{jn} \le \theta \le a_{j+1,n}\}$$

$$g_{n1}^2(\theta) = \inf\{g_{n\theta}^2 : a_{jn} \le \theta \le a_{j+1,n}$$

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

$$N_d^{(1)}(\theta) = \inf\{n\ge 5 : n \ge (bs_{n1}(\theta)/d)^2\}$$

$$N_d^{(2)}(\theta) = \inf\{n\ge 5 : n \ge (bs_{n2}(\theta)/d)^2\}.$$

Clearly $N_d^{(1)}(\theta) \le N_d(\theta) \le N_d^{(2)}(\theta)$ and the above rules are members of $D[0,\infty)$. Define $n_d^{(1)}(\theta)$ and $n_d^{(2)}(\theta)$ in a manner similar to the above and for $0 \le t \le 2$, $0 \le \theta \le M^*$, let

$$V_d(t,\theta) = \left(\sigma^2(M^*)n_d^{(2)}(M^*)\right)^{-\frac{1}{2}} \sum_{i=1}^{[tn_d^{(2)}(M^*)]} G_\theta(X_i),$$

where $[\circ]$ is the greatest integer function. V_d is an element of $D_2(M^*) = D([0,2] \times [0,M^*])$ and for fixed t, $V_d(t,\circ)$ is an element of $C[0,M^*]$. Let $B = [s,t) \times [\theta_1,\theta_2)$ be any block (B-W, page 1658), where ns and nt are integers. Then, by (A1),

$$E|V_d(B)|^4 \le M_1\ |\theta_2-\theta_1|^{1+\delta_0}\ \{|t-s|/n + |t-s|^2\}$$

$$\le M_2\ |\theta_2-\theta_1|^{1+\delta_0}\ |t-s|^{1+\delta_0}$$

since $|t-s| \ge 1/n$. Hence for any pair of blocks B, C, the Schwarz inequality shows

$$E|V_d(B)V_d(C)|^2 \le M_2 \, |\theta_2-\theta_1|^{1+\delta_0} \, |t-s|^{1+\delta_0} ,$$

proving by the Corollary to Theorem 3 of B-W that $\{V_d\}$ is tight. By Theorems 2 and 4 of B-W, V_d converges weakly to a Gaussian process V in $D_2(M^*)$. Now define

$$V_d^{(1)}(t,\theta) = V_d\left(tn_d^{(2)}(\theta)/n_d^{(2)}(M^*),\ \theta\right).$$

$\{V_d^{(1)}\}$ are also in $D_2(M^*)$ and the sequence is tight since if (t_1,θ_1), (t_2,θ_2) are in a rectangle R of diameter $\delta > 0$, $\left(t_i \frac{n_d^{(2)}(\theta_i)}{n_d^{(2)}(M^*)},\ \theta_i\right)$ $(i=1,2)$ are in a rectangle with the same center as R but of diameter $c\delta$, for a fixed constant c and d sufficiently small. Here, we have used the fact that $n_d^{(2)}(\theta)/n_d^{(2)}(M^*) \ge (a^*/2M^*)^2$. Note that $n_d^{(2)}(\theta)$ was defined precisely to keep $V_d^{(1)}$ in D_2. Next define $V_d^{(2)}(t,\theta) = \left(\frac{n_d^{(2)}(M^*)\sigma^2(M^*)}{n_d^{(2)}(\theta)\sigma^2(\theta)}\right)^{\frac{1}{2}} V_d^{(1)}(t,\theta)$. Then $V_d^{(2)}$ is tight in $D_2(M^*)$ and its finite dimensional distributions converge to those of a Gaussian process V with covariance function when $t = 1$ and $\theta_1 < \theta_2$ given by

$$\int \frac{G_{\theta_1}(x)G_{\theta_2}(x)}{\sigma^2(\theta_2)}\,dx.$$

Hence $V_d^{(2)}(1,\theta_2) - V_d^{(2)}(1,\theta_1)$ converges to a Gaussian random variable with mean zero and variance

$$\int \left\{\frac{G^2_{\theta_2}(x)}{\sigma^2(\theta_2)} + \frac{G^2_{\theta_1}(x)}{\sigma^2(\theta_1)} - \frac{2G_{\theta_1}(x)G_{\theta_2}(x)}{\sigma^2(\theta_2)}\right\} dx \le M_3|\theta_2-\theta_1|^{\varepsilon} ,$$

for some $\varepsilon > 0$. The inequality here follows from (A1) and the fact that $|\sigma^2(\theta_2) - \sigma^2(\theta_1)| \le M_4|\theta_2-\theta_1|$ Hence, by Billingsley (1968, page

97) for fixed t, $V_d^{(2)}(t,\circ)$ converges weakly to a Gaussian process on $C[0,M]$. Now define $V_d^{(3)}(t,\theta) = V_d^{(2)}\left(t\,\frac{N_d^{(2)}(\theta)}{n_d^{(2)}(\theta)},\ \theta\right)$. Note that condition (A2) shows that for $\varepsilon > 0$

(6) $$\lim_{d \downarrow 0} \Pr\left\{\sup\left|\frac{N_d^{(2)}(\theta)}{n_d^{(2)}(\theta)} - 1\right| > \varepsilon : 0 \le \theta \le M\right\} = 0.$$

By (6) and the Kolmogorov inequality (see Anscombe (1952)), one shows that for fixed t and θ,

$$V_d^{(2)}(t,\theta) - V_d^{(3)}(t,\theta) \to 0 \quad \text{(in probability)}.$$

$V_d^{(3)}$ (and hence $V_d^{(2)} - V_d^{(3)}$) is shown to be tight by (6) and the same kind of argument used to prove the tightness of $V_d^{(1)}$; hence for fixed t, $V_d^{(3)}(t,\circ)$ converges weakly to a Gaussian element of $C[0,M]$. We need the following result. Define

$$W_d = \sup\left\{|V_d^{(3)}(t,\theta) - V_d^{(3)}(1,\theta)| : 0 \le \theta \le M,\ \frac{N_d^{(1)}(\theta)}{N_d^{(2)}(\theta)} \le t \le 1\right\}.$$

Proposition 1. $W_d \to 0$ in probability.

Proof of Proposition 1: First note that (A2) implies

(7) $$\sup\{|N_d^{(1)}(\theta)/N_d^{(2)}(\theta) - 1| : 0 \le \theta \le M\} \xrightarrow{p} 0.$$

Let $\eta > 0$ be arbitrary. Then

$$|V_d^{(3)}(t,\theta) - V_d^{(3)}(1,\theta)| \le |V_d^{(3)}(1,\theta) - V_d^{(3)}(1-\eta,\theta)|$$
$$+ \min\{|V_d^{(3)}(t,\theta) - V_d^{(3)}(1,\theta)|,\ |V_d^{(3)}(1-\eta,\theta) - V_d^{(3)}(t,\theta)|\},$$

so that by (7), with probability approaching one,

$$(8)\quad W_d \le \sup_\theta |V_d^{(3)}(1,\theta)-V_d^{(3)}(1-\eta,\theta)| + \sup_{1-\eta\le t\le 1} \min\{\sup_\theta|V_d^{(3)}(t,\theta)-V_d^{(3)}(1,\theta)|,\ \sup_\theta|V_d^{(3)}(1-\eta,\theta)-V_d^{(3)}(t,\theta)|\}.$$

The first term on the right hand side of (8) converges in probability to zero as $\eta, d \to 0$ by Chebychev's inequality, while the second is bounded by the modulus of continuity w_η'' of B-W and hence converges to zero in probability as $\eta \to 0$.

Returning to the proof of Theorem 1, we see that the conclusion of Proposition 1 also holds for the process

$$V_d^{(4)}(t,\theta) = b\left(dN_d^{(1)}(\theta)\right)^{-1} \sum_{i=1}^{[tN_d^{(2)}(\theta)]} G_\theta(X_i)$$

since $b^2\sigma^2(\theta)\left(d^2N_d^{(1)}(\theta)\right)^{-1} \xrightarrow{P} 1$ uniformly on $0 \le \theta \le M$. Thus, for all $\varepsilon > 0$ and d sufficiently small

$$(9)\quad \Pr\{|N_d^{-1}(\theta)\ \sum_1^{N_d(\theta)} G_\theta(X_i)|\ge d\} = \Pr\left\{|\frac{b}{dN_d(\theta)}\sum_1^{N_d(\theta)} G_\theta(X_i)|\ge b\right\}$$
$$\le \Pr\{|V_d^{(4)}(1,\theta)| \ge b-\varepsilon\} + \varepsilon.$$

Since $V_d^{(4)}(1,\ )$ converges weakly on $D[0,M^*]$ to a Gaussian process (call it V) on $C[0,M^*]$ and M^* is arbitrary, the convergence is on $D[0,\infty)$ to a Gaussian process V which is in $C[0,\infty)$ with probability one. Thus, Lemma 1 (with b replaced by $b-\varepsilon$) says that as $d \to 0$, the last term in (9) is bounded uniformly in $\theta \in F$ by $2\left(1-\Phi(b-\varepsilon)\right) + 2\varepsilon$. Letting $\varepsilon \to 0$ completes the proof.

In the sample means case, (A1) and (A2) hold if $\int G_\theta^4(x)dx$ is bounded on each finite interval and there exists a function H with finite fourth moment such that $|G_{\theta_2}(x)-G_{\theta_1}(x)| \le |\theta_2-\theta_1|H(x)$.

4. Applications

Theorem 1 holds for many types of estimators, two of which (the sample median and a smooth M-estimate) we illustrate here. The key idea is that Theorem 1 will hold when $T_{n\theta}$ can be expanded as a sample mean plus a uniformly small order term.

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

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

Example 4.1. Here we define $X_1, X_2, \ldots$ as i.i.d. uniform $(0,1)$ random variables and $T_{n\theta}$ by $\sum_1^n \psi(G_\theta(x_i) - T_{n\theta}) = 0$. Here 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 (Huber (1964)). $g_{n\theta}^2$ 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 make the following assumptions for every $C > 0$:

(B1) There exists C_* such that

$$\sup\{|G_{\theta 1}(x)-G_{\theta 2}(x)|/|\theta_2-\theta_1| \ : \ 0\le|x|,\theta_1,\theta_2\le C\} \le C_*$$

(B2) For every ε, there exists $C_*^! > 0$ such that

$$\inf\{|\int \psi(x+\varepsilon)dF_\theta(x)| \ : \ 0\le\theta\le C\} \ge C_*^!$$

(B3) There exist positive constants c_1, c_2 such that

$$\inf\{F_\theta(c_1) - F_\theta(-c_1) \ : \ 0\le\theta\le C\} \ge c_2 \ .$$

Condition (B1) is weaker than the sufficient condition for the sample means given at the end of Section 3, while (B2) and (B3) are quite reasonable.

Lemma 2. Under (B1) - (B3), the conclusion to Theorem 1 holds.

Proof: We first show $|T_{n\theta}| \to 0$ (a.s.u.). By a Taylor expansion and the fact that ψ' vanishes outside $[-k,k]$ for any ε, there exists $\eta_1 > 0$ such that

$$(10) \quad |n^{-1} \textstyle\sum_1^n \psi(G_\theta(X_i)-\varepsilon) - n^{-1} \sum_1^n \psi(G_{\theta_0}(X_i)-\varepsilon)|$$

$$\le \sup_x \psi'(x)\, n^{-1} \textstyle\sum_1^n |G_\theta(X_i)-G_{\theta_0}(X_i)|\, I_{\{\eta_1\le X_i\le 1-\eta_1\}}$$

$$\le M_1\, |\theta-\theta_0|,$$

where M_1 depends only on C, and $0\le\theta\le C$. Now choose a set of such parameters θ_0 by $\{\theta_{in} = i/n^{\frac{1}{2}},\ 0\le i\le cn^{\frac{1}{2}}\}$. By Theorem 1 of Hoeffding (1963) and the Borel-Cantelli Lemma,

(11) $\sup\left\{\left|n^{-1}\sum_1^n \psi\left(G_\theta(X_i)-\varepsilon\right) - \int \psi(x-\varepsilon)dF_\theta(x)\right| : \theta\epsilon\{\theta_{in}\}\right\} \to 0$ (a.s.).

Since $\int \psi'(x-\varepsilon)dF_\theta(x) > \eta_0$, (10) and (11) give $|T_{n\theta}| \to 0$ (a.s.u.). Again by a Taylor expansion,

$$T_{n\theta}\, n^{-1}\sum_1^n \psi'\left(\xi_{\theta n}(X_i)\right) = n^{-1}\sum_1^n \psi\left(G_\theta(X_i)\right),$$

where $\xi_{\theta n}(X_i)$ is between $G_\theta(X_i)$ and $G_\theta(X_i) - T_{n\theta}$. Since $\psi'(x) > 0$ in a neighborhood of zero and (B3) holds, there is a positive constant η_2 such that

$$\inf\{n^{-1}\sum_1^n \psi'\left(\xi_{\theta n}(X_i)\right) : 0\le\theta\le C\} \ge \eta_2$$

almost surely as $n \to \infty$. One now shows

(12) $$n^{\frac{1}{4}}\, |T_{n\theta}| \to 0 \quad \text{(a.s.u.)}$$

by showing $n^{-3/4}\sum_1^n \psi\left(G_\theta(X_i)\right) \to 0$ (a.s.u.); this is accomplished by following the steps in (10) and (11). Now using (12) and the fact that ψ'' is bounded, one more Taylor expansion shows

(13) $n^{\frac{1}{2}}\, \left|T_{n\theta} - n^{-1}\sum_1^n \psi\left(G_\theta(X_i)\right) \Big/ \int \psi'(x)dF_\theta(x)\right| \to 0$ (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

$$a_1(\theta) = \left(E\psi'(x)\right)^{-2}$$

$$a_2(\theta) = -2E\psi^2(X)\left(E\psi'(X)\right)^{-3}$$

$$a_3(\theta) = -a_2 E\psi''(X) - a_1 E\psi(X)\psi'(X),$$

and the expectations are under F_θ . Using (13), Taylor's theorem, and the same type of arguments as in (10) and (11), one shows

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

where $h(\theta) = \int \psi^2(x)dF_\theta(x)\{\int \psi'(x)dF_\theta(x)\}^{-2}$ is clearly Lipschitz in θ on each subinterval $0 \le \theta \le M$.

Now reconsider the proof of Theorem 1. One can redefine all the processes there in terms of $T_{n\theta}$. For example,

(15) $$V_d(t,\theta) = \left(\sigma^2(M^*)n_d^{(2)}(M^*)\right)^{-\frac{1}{2}} T_{[tn_d^{(2)}(M^*)],\theta} \quad \text{if } t \ge \tfrac{1}{2}$$
$$= V_d(\tfrac{1}{2},\theta) \quad \text{if } t < \tfrac{1}{2}.$$

Because of (B1), (A1) holds for the sample means generated by $\psi(G_\theta(X))/a_1(\theta)$, while (14) shows that (A2) holds for $g_{n\theta}^2$. Because of (13), all the weak convergence arguments in Theorem 1 apply to processes such as (15), and the proof is complete.

It is clear that this type of result can be extended to a more general class of smooth ψ functions and to the Huber ψ, but we will stay with Lemma 2 for the sake of brevity. Further, the scale equivariant version given by

$$0 = n^{-1}\sum_1^n \psi\left(q_{n\theta}^{-1}(G_\theta(X_i) - T_{n\theta})\right)$$

could also be considered. Here $q_{n\theta} = q_n(G_\theta(X_1), \ldots, G_\theta(X_n))$ is location invariant and scale equivariant; an example would be the interquartile range. The proof of Lemma 2 will still be applicable if we can write $q_{n\theta}$ as a sum of i.i.d. variates plus remainder term as in (13), say $q_{n\theta} = n^{-1}\sum_1^n H(G_\theta(X_i)) + o(n^{-\frac{1}{2}})$ (a.s.u.). The next example includes implicit conditions under which the interquartile range admits such an expansion.

Example 4.2. Let $X_{1n} < X_{2n} < \ldots < X_{nn}$ be the order statistics from the uniform sample $X_1, X_2, \ldots, X_n$. Let $0 < p < 1$ and define

$$F_{n\theta}(x) = n^{-1} \sum_1^n I_{\{G_\theta(X_i)<x\}}, \qquad F_n(x) = n^{-1} \sum_1^n I_{\{X_i<x\}}.$$

Bahadur (1966) has shown that if $k_n = np + (n^{\frac{1}{2}} \log n)$,

$$X_{k_n,n} - p = \{k_n/n - F_n(p)\} + R_n,$$

where $n^{\frac{1}{2}} R_n \to 0$ (a.s.). Following Kiefer (1967, page 1324), it is possible to prove the following.

Proposition 2. Define $\xi_\theta = G_\theta(p)$, and suppose that uniformly on $0 \le \theta \le C$, for some $\varepsilon > 0$,

$$\inf\{F_\theta'(x) : \xi_\theta - \varepsilon < x < \xi_\theta + \varepsilon\} > 0$$

$$\sup F''(x) : \xi_\theta - \varepsilon < x < \xi_\theta + \varepsilon\} < \infty.$$

Then

$$n^{\frac{1}{2}} \left| G_\theta(X_{k_n,n}) - \xi_\theta - (k_n/n - F_{n\theta}(\xi_\theta))/F'(\xi_\theta) \right| \to 0 \quad \text{(a.s.u.)}.$$

This proposition gives an (a.s.u.) expansion for the interquartile range, the median, and the variance estimate for the median suggested by Geertsema (1972). As in the previous example, it suffices to verify Theorem 1 when the random variables are given by

$$G_\theta(X_i) = (\tfrac{1}{2} - I_{\{X_i<\frac{1}{2}\}})/F_\theta'(\xi_\theta)$$

and $g_{n\theta}^2$ is given by the expansion for Geertsema's variance estimate. Under the conditions of Proposition 2, (A1) and (A2) reduce to requiring that $F_\theta'(\xi_\theta)$ is Lipschitz in θ of order $\alpha > \frac{1}{4}$.

5. A General Result

The Theorem we actually gave in Section 3 is stronger than stated. We place it here in its fullest generality as it should be of independent interest in the theory of weak convergence of partial sum processes with random indices.

Theorem 2. Suppose (A1) holds and there exists elements $N_d^{(1)}(\theta)$, $N_d^{(2)}(\theta)$ and $n_d(\theta)$ in $D[0,\infty)$ with $\{n_d(\theta)\}$ non-random and $N_d^{(1)}(\theta) \le N_d(\theta) \le N_d^{(2)}(\theta)$. Suppose further that

$$N_d^{(2)}(\theta)/n_d(\theta) \Rightarrow 1 \qquad \text{on } D[0,\infty)$$

$$N_d^{(1)}(\theta)/n_d(\theta) \Rightarrow 1 \qquad \text{on } D[0,\infty)$$

$$d^2 n_d(\theta)/\sigma^2(\theta) \Rightarrow b^2 \qquad \text{on } D[0,\infty),$$

where "$\Rightarrow$" is weak convergence. Then (4) and (1) hold. Further, if we define

$$V_d(t) = \left(\sigma^2(\theta) n_d(\theta)\right)^{-\frac{1}{2}} \sum_1^{N_d^{(2)}(\theta)} G_\theta(X_i),$$

then V_d converges weakly to a Gaussian process V on $D[0,\infty)$ whose sample paths are in $C[0,\infty)$ with probability one.

ACKNOWLEDGEMENT

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

REFERENCES

ANSCOMBE, F.J. (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc. 48*, 600-617.

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

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

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. New York: John Wiley & Sons, Inc.

CHOW, Y.S. and ROBBINS, H. (1968). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist. 36*, 463-467.

GEERTSEMA, J.C. (1972). Nonparametric sequential procedures for selecting the best of k populations. *J. Am. Statist. Assoc. 67*, 614-616.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Am. Statist. Assoc. 58*, 13-30.

HUBER, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist. 35*, 73-101.

KIEFER, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist. 38*, 1323-1342.

LINDVALL, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0,\infty)$. *J. Appl. Prob. 10*, 109-121.

PARZEN, E. (1954). On uniform convergence of families of sequences of random variables. *Univ. of Calif. Publ. in Statist. 2*, 23-54.

TONG, Y.L. (1971). On the consistency of single-stage ranking procedures. *Ann. Institute Statist. Math. 22*, 271-284.

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