JOURNAL OF MULTIVARIATE ANALYSIS 14, 83–93 (1984)

On a Class of Stopping Times for M-Estimators

WINFRIED STUTE

University of Siegen, Siegen, West Germany

Communicated by J. Pfanzagl

For a given score function $\psi = \psi(x, \theta)$, let θ_n be Huber's M-estimator for an unknown population parameter θ. Under some mild smoothness assumptions it is known that $n^{1/2}(\theta_n - \theta)$ is asymptotically normal. In this paper the stopping times $\tau_c(m) = \inf\{n \geqslant m : n^{1/2} |\theta_n - \theta| > c\}$ associated with the sequence of confidence intervals for θ are investigated. A useful representation of M-estimators is derived, which is also appropriate for proving laws of the iterated logarithm and Donsker-type invariance principles for $(\theta_n)_n$.

0. INTRODUCTION AND MAIN RESULT

Let $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}$, be some parametric family of probability distributions on the real line. Suppose that $\xi_1, \ldots, \xi_n, \ldots$, is a sequence of independent random variables with common distribution $P = P_{\theta_0}$ for some unknown $\theta_0 \in \Theta$. Lai [4] then investigated the possibility of constructing a sequence of confidence regions $C_n(\xi_1, \ldots, \xi_n)$ such that

$$\mathbb{P}_\theta(\theta \in C_n(\xi_1, \ldots, \xi_n) \text{ for all } n \geqslant m) \geqslant 1 - \alpha, \tag{1}$$

where $m \in \mathbb{N}$ and $1 - \alpha$ is a prescribed coverage probability. Here $\mathbb{P}_\theta$, $\theta \in \Theta$, denotes a probability measure on the sample space such that $\xi_1, \ldots, \xi_n$ have distribution P_θ under $\mathbb{P}_\theta$. Suppose that $\theta_n = T_n(\xi_1, \ldots, \xi_n)$ is a sequence of estimates for θ_0 such that under $\mathbb{P}_{\theta_0}$

$$n^{1/2}(\theta_n - \theta_0) \to N(0, \sigma^2_{\theta_0}), \tag{2}$$

in distribution as $n \to \infty$, where $N(0, \sigma^2_{\theta_0})$ is a centered normal random variable with variance $\sigma^2_{\theta_0} > 0$. Assume for simplicity that $\sigma^2 = \sigma^2_\theta$ is independent of θ, and let c_α denote the $1 - \alpha/2$ quantile of $N(0, \sigma^2)$. Then,

Received May 14, 1982.
AMS 1980 subject classifications: Primary 60G40, 62F25; Secondary 60F05, 60F15.
Key words and phrases: M-estimator, stopping time, representation, law of the iterated logarithm, invariance principle.

0047-259X/84 $3.00

asymptotically, the intervals $C_n := [\theta_n - c_\alpha n^{-1/2}, \theta_n + c_\alpha n^{-1/2}]$ form a sequence of confidence regions with coverage probability $1-\alpha$,

$$\mathbb{P}_\theta(\theta \in C_n) \to 1-\alpha \qquad \text{as} \quad n \to \infty. \tag{3}$$

Condition (3) has to be clearly distinguished from (1). In fact, in most cases the sequence $(\theta_n)_n$ fulfills a law of the iterated logarithm so that for the above C_n,

$$\mathbb{P}_\theta(\theta \in C_n \text{ for all } n \geqslant m) = 0, \qquad m \in \mathbb{N}. \tag{4}$$

Similarly, a sequential test based on $[\theta_0' - c_\alpha n^{-1/2}, \theta_0' + c_\alpha n^{-1/2}]$ leads one with probability to a rejection of the hypothesis $\theta_0 = \theta_0'$, though it is true.

Now, fix $\theta \in \Theta$ and define $A_m := \{\theta \in C_n \text{ for all } n \geqslant m\}$, $m \in \mathbb{N}$. Then $\omega \in A_m$ if and only if $\tau_{c_\alpha}(\theta, m, \omega) = \infty$, where $\tau_c(\theta, m)$ denotes the stopping time,

$$\tau_c = \tau_c(\theta, m) := \inf\{n \geqslant m : n^{1/2} |\theta_n - \theta| > c\}, \qquad c > 0.$$

Under (4) τ_c is almost surely finite. So it makes sense to investigate $\mathbb{E}(\tau_c) = \mathbb{E}_\theta(\tau_c)$, the expected time for θ_n to leave the interval(s) $[\theta - cn^{-1/2}, \theta + cn^{-1/2}]$ for the first time after m, if θ is the true parameter. In a hypothesis-testing framework τ_{c_α} is the first time after m, such that the hypothesis $\theta_0 = \theta$ is rejected on a level α against a two-sided alternative. Information about $\mathbb{E}(\tau_c)$ provides one further tool for judging procedures based on $(\theta_n)_n$.

If θ is the true parameter, then clearly estimates are preferred which in the mean yield a large τ_c, $c > 0$, while it shoud be small if θ is the wrong parameter. In this paper we shall deal with the class of M-estimators as introduced by Huber [3]. To be specific, let $\psi(x, \theta)$, $x \in \mathbb{R}$, $\theta \in \Theta$, be a (smooth) score function which is Fisher-consistent, i.e.,

$$\int \psi(x, \theta)\, P_\theta(dx) = 0 \qquad \text{for all } \theta \in \Theta.$$

Denote with F_n the empirical distribution function (d.f.) of the sample $\xi_1, \ldots, \xi_n$ and assume that the equation,

$$\int \psi(x, \theta)\, F_n(dx) = 0 \qquad (\text{within } 1/n), \tag{5}$$

has a root θ_n. We shall always deal with functions ψ which are nonincreasing in θ for each fixed x. Hence θ_n is essentially unique. Under a smoothness assumption θ_n is a minimum contrast estimator as considered by Pfanzagl [5].

It is known (cf. Huber [3]) that (2) holds with

$$\sigma_{\theta_0}^2 = \frac{\int \psi^2(x, \theta_0)\, P_{\theta_0}(dx)}{[\partial/\partial\theta \int \psi(x, \theta)\, P_{\theta_0}(dx)]_{\theta=\theta_0}^2}.$$

It is the purpose of this paper to prove the following:

THEOREM. *Let ψ be a smooth-bounded score function, nonincreasing in θ for each $x \in \mathbb{R}$. Then for each $\theta \in \Theta$*

(i) $\mathbb{E}_\theta(\tau_c(m, \theta)) < \infty$ *for each* $c < \sigma_\theta$ *and every* $m \in \mathbb{N}$.

(ii) $\mathbb{E}_\theta(\tau_c(m, \theta)) = \infty$ *for each* $c > \sigma_\theta$ *and* $m \geqslant m_0(c, \theta)$, *where* $m_0(c, \theta)$ *is some integer increasing with* c *as* $c \downarrow \sigma_\theta$.

Remark. The function $x \to \psi(x, \theta_0)/[\ldots]$ is the influence function of the statistical functional defining the θ_n's. Hence the theorem may be viewed as giving a sort of stochastic interpretation of the role played by the influence function in parameter estimation. When θ_n is the (nonrobust) sample mean, part (ii) of the theorem is due to Chow, Robbins, and Teicher [1], with $m_0(c, \theta) = 1$ and $c \geqslant \sigma_\theta$. Their method of proof is strongly based on a second order version of Wald's identity. In our case such an argument is not applicable, since under $\mathbb{P}_\theta \theta_n$ is of the form $n(\theta_n - \theta) = S_n + R_n$, where S_n is a partial sum of independent identically distributed (i.i.d.) zero-mean random variables and R_n is a nonvanishing error term. There is little known about the dependence structure of the sequence $(R_n)_n$. The only information will be contained in an estimate showing that $n^{-1/2} R_n$ is small with high probability, at least for large n. For small n we shall have no control on the value of R_n. For the sample mean part (i) is due to Gundy and Siegmund [2]. For testing the hypothesis $\theta_0 = \theta_0'$, the constant $c = c_\alpha$ is usually larger than $\sigma_{\theta_0'}$; hence, $\mathbb{E}_{\theta_0'}(\tau_c(m, \theta_0')) = \infty$.

EXAMPLES. (1) The location case: $F_\theta(x) = P_\theta(-\infty, x] = F(x - \theta)$ with some known or unknown F. In most cases F is to be considered symmetric. For invariance reasons take ψ of the form $\psi(x, \theta) = \psi_0(x - \theta)$, where ψ_0 is a smooth-bounded, nondecreasing function of one variable. To obtain Fisher-consistency (with F symmetric), ψ_0 has to be chosen skew-symmetric. Also, σ_θ is independent of θ. The fundamental paper of Huber [3] treats the question how to find minimax solutions for ψ_0 if F is only approximately known.

(2) For

$$\begin{aligned} \psi_0(x) &= -1, & x &< 0, \\ &= 0, & x &= 0, \\ &= p/(1-p), & x &> 0, \end{aligned}$$

the resulting estimate is the p-quantile of the distribution $(0 < p < 1)$. Though ψ_0 is not smooth the assertion of the theorem is also valid for this particular ψ_0. This is because for the proof of the theorem one needs, for a given d.f. F, smoothness of the function $H(\theta) = \int \psi_0(x-\theta) F(dx)$ together with some appropriate fluctuation bounds for the associated empirical process. But $H(\theta) = (p - F(\theta - 0))/(1-p)$ is smooth whenever F is smooth, and the fluctuation may be controlled using results of Stute [7]. For smooth empirical processes related results are contained in Stute [8].

1. Lemmas and Proofs

Fix $\theta \in \Theta$ and replace $\mathbb{P}_\theta$ by $\mathbb{P}$. As mentioned in the Introduction $n(\theta_n - \theta) \equiv T_n$ will be of the form $T_n = S_n + R_n$, S_n being a sum of n i.i.d. random variables. The error term R_n will be asymptotically smaller than S_n. We have thus to study stopping times,

$$\tau_c = \tau_c(m) = \inf\{n \geqslant m : |T_n| > cn^{1/2}\}, \qquad m \in \mathbb{N}, \quad c > 0, \tag{6}$$

under a control condition for $(R_n)_n$. For our purposes it will be enough to consider

$$\sum_1^\infty n^2 \mathbb{P}(n^{-1/2} |R_n| > \varepsilon) < \infty \qquad \text{for each } \varepsilon > 0. \tag{7}$$

In fact, it will be shown that for θ_n we have $\mathbb{P}(n^{-1/2} |R_n| > \varepsilon) \to 0$ exponentially fast.

The following two lemmas will be needed for part (i) of the theorem. For this, let $\eta_1, \eta_2, \ldots,$ be a sequence of i.i.d. random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Assume $\mathbb{E}(\eta_i) = 0$ and $\mathbb{E}(\eta_i^2) = 1$, with $|\eta_i| \leqslant M$ for some finite $M \geqslant 4$. Define $S_n = \sum_1^n \eta_i$ and let $(R_n)_n$ be some sequence of random variables satisfying (7). Assume that $(\mathcal{F}_n)_n$ is an increasing family of sub-σ-fields of $\mathcal{A}$ such that both $(S_n)_n$ and $(R_n)_n$ are adapted to $(\mathcal{F}_n)_n$, and such that $\eta_{n+1}, \eta_{n+2}, \ldots,$ are independent of $\mathcal{F}_n$ for each $n \geqslant 1$. Put $T_n = S_n + R_n$ and let $\tau_c(m)$ be defined by (6). Then $\tau_c(m)$ is a stopping time w.r.t. $(\mathcal{F}_n)_n$. Similarly, put $r_c(m) := \inf\{n \geqslant m : |S_n| > cn^{1/2}\}$. The following estimate extends, in a sense, the lemma in Gundy and Siegmund [2].

Lemma 1. *For $c < 1$ and $m \in \mathbb{N}$ we have*

$$\mathbb{E}(r_c(m)) \leqslant M^2 m/(1-c)^2.$$

Proof. Fix $m \in \mathbb{N}$ and let, for $n > m$, $r = r_c(m) \wedge n$. By the second order version of Wald's equation,

$$\begin{aligned}\mathbb{E}(r) = \mathbb{E}(S_r^2) &= \mathbb{E}(S_{r-1}^2) + 2\mathbb{E}(S_{r-1}\eta_r) + \mathbb{E}(\eta_r^2)\\ &\leqslant \mathbb{E}(S_{r-1}^2) + 2M\sqrt{\mathbb{E}(S_{r-1}^2)} + M^2.\end{aligned}$$

Since

$$S_{r-1}^2 = S_{r-1}^2 1_{\{r>m\}} + S_{r-1}^2 1_{\{r=m\}},$$

we obtain

$$\begin{aligned}\mathbb{E}(r) &\leqslant c^2\mathbb{E}(r) + m - 1 + 2Mc\sqrt{\mathbb{E}(r)} + 2M\sqrt{m-1} + M^2\\ &\leqslant c^2\mathbb{E}(r) + 2Mc\sqrt{m\mathbb{E}(r)} + mM^2,\end{aligned}$$

where the first inequality follows from $\sqrt{a+b} \leqslant \sqrt{a} + \sqrt{b}$; $a, b \geqslant 0$. It follows that with $x^2 = mM^2/\mathbb{E}(r)$, $1 \leqslant c^2 + 2cx + x^2 = (c+x)^2$, and therefore $\mathbb{E}(r) \leqslant M^2 m/(1-c)^2$. The assertion now follows by letting n tend to infinity. ∎

Lemma 1 will be needed to bound the mean of $\tau_c = \tau_c(m)$. This is the content of

LEMMA 2. *Let* (7) *be satisfied. Then*

$$\mathbb{E}(\tau_c) < \infty \qquad \textit{for all } 0 < c < 1.$$

Proof. We shall only consider the case $m = 1$. The proof for a general m will be similar. Now, for each $\varepsilon > 0$, let $t_\varepsilon := \sup\{n \geqslant 1 : n^{-1/2}|R_n| > \varepsilon\}$. By (7) and the Borel–Cantelli lemma $n^{-1/2}R_n \to 0$, almost surely, i.e., t_ε is well defined. Observe that t_ε is not a stopping time. Furthermore, put

$$s_c(\varepsilon) = s_c := \inf\{n > t_\varepsilon : |S_n| > cn^{1/2}\},$$

which is finite (almost surely) according to the finiteness of t_ε and the law of the iterated logarithm for $(S_n)_n$. Since $t_\varepsilon = k$ implies $k^{-1/2}|R_k| > \varepsilon$ we get from (7),

$$\sum_1^\infty k\mathbb{P}(t_\varepsilon = k) < \infty,$$

i.e., t_ε has finite expectation. To prove the lemma it therefore remains to show

$$\int_{\{\tau_c > t_\varepsilon\}} \tau_c \, d\mathbb{P} < \infty,$$

for some appropriate small $\varepsilon > 0$. Now, choose $\varepsilon > 0$ so small that $c^* := c + \varepsilon < 1$. Then, on $\{\tau_c > t_\varepsilon\}$,

$$|S_n| \leqslant cn^{1/2} + \varepsilon n^{1/2} = c^* n^{1/2},$$

for $t_\varepsilon < n < \tau_c$, i.e., $s_{c^*} \geqslant \tau_c$ and hence,

$$\int_{\{\tau_c > t_\varepsilon\}} \tau_c \, d\mathbb{P} \leqslant \int_{\{\tau_c > t_\varepsilon\}} s_{c^*} \, d\mathbb{P} \leqslant \int s_{c^*} \, d\mathbb{P}.$$

Let us show that the last integral is finite. But

$$\int s_{c^*} \, d\mathbb{P} = \sum_1^\infty \int_{\{t_\varepsilon = k\}} s_{c^*} \, d\mathbb{P}.$$

Fix k and write, for $n > k$,

$$S_n = S_k + \eta_{k+1} + \cdots + \eta_n = S_k + S_k^n.$$

Given that $\eta_i = x_i$ for $1 \leqslant i \leqslant k$, the condition,

$$|x_1 + \cdots + x_k + \eta_{k+1} + \cdots + \eta_n| \leqslant c^* n^{1/2},$$

implies

$$\begin{aligned} |S_k^n| = |\eta_{k+1} + \cdots + \eta_n| &\leqslant Mk + c^* n^{1/2} \\ &= \sqrt{n-k} \left[\frac{Mk + c^* n^{1/2}}{\sqrt{n-k}} \right]. \end{aligned}$$

Now, there exists some large constant C and some $c^{**} < 1$ such that the term in brackets is less than or equal to c^{**}, provided that $n \geqslant Ck^2$. Thus, defining

$$s_{c^{**}}(k) := \inf\{n \geqslant Ck^2 : |S_k^n| > c^{**} \sqrt{n-k}\},$$

we get, on $\{t_\varepsilon = k\}$, $s_{c^*} \leqslant s_{c^{**}}(k)$. Hence,

$$\int_{\{t_\varepsilon = k\}} s_{c^*} \, d\mathbb{P} \leqslant \int_{\{t_\varepsilon = k\}} s_{c^{**}}(k) \, d\mathbb{P} \leqslant \int_{\{k^{-1/2}|R_k| > \varepsilon\}} s_{c^{**}}(k) \, d\mathbb{P}.$$

Since R_k is $\mathscr{F}_k$-measurable, we have by independence of $\eta_{k+1}, \eta_{k+2}, \ldots$, that the last integral is equal to

$$\mathbb{P}(k^{-1/2} |R_k| > \varepsilon) \int s_{c^{**}}(k) \, d\mathbb{P}.$$

Hence it remains to show

$$\sum_{1}^{\infty} \mathbb{P}(k^{-1/2}|R_k| > \varepsilon) \int s_{c^{**}}(k)\, d\mathbb{P} < \infty.$$

By stationarity the last integral is equal to $\mathbb{E}(r_{c^{**}}(Ck^2 - k))$. The result now follows from Lemma 1 and (7). ∎

If $\eta_1, \ldots,$ have finite variance σ^2, then $\mathbb{E}(\tau_c) < \infty$ for each $0 < c < \sigma$. The following lemma will be needed for the second part of the theorem.

LEMMA 3. *Under the assumptions of Lemma 2, for each $c > 1$ there exists some $m_0(c)$, which increases as $c \downarrow 1$, such that*

$$\mathbb{E}(\tau_c(m)) = \infty \qquad \textit{for all } m \geqslant m_0(c).$$

Proof. Choose $\varepsilon > 0$ such that $c^* = c - \varepsilon > 1$, and let t_ε be defined as in the proof of Lemma 2. Fix $m \in \mathbb{N}$ and suppose that $\mathbb{E}(\tau_c) < \infty$, $\tau_c = \tau_c(m)$. Remember $T_n = S_n + R_n$. By the second order version of Wald's identity,

$$\mathbb{E}(\tau_c) = \mathbb{E}(S_{\tau_c}^2) \geqslant \int_{\{\tau_c > t_\varepsilon\}} (T_{\tau_c} - R_{\tau_c})^2\, d\mathbb{P}.$$

Clearly, $|T_{\tau_c}| > c\tau_c^{1/2}$ and $|R_{\tau_c}| \leqslant \varepsilon\tau_c^{1/2}$, and hence,

$$(T_{\tau_c} - R_{\tau_c})^2 \geqslant c^*\tau_c \qquad \text{on} \quad \{\tau_c > t_\varepsilon\}.$$

Thus

$$\mathbb{E}(\tau_c) \geqslant c^* \int_{\{\tau_c > t_\varepsilon\}} \tau_c\, d\mathbb{P}.$$

On the other hand,

$$\begin{aligned} \int_{\{\tau_c \leqslant t_\varepsilon\}} \tau_c\, d\mathbb{P} &\leqslant \int_{\{\tau_c \leqslant t_\varepsilon\}} t_\varepsilon\, d\mathbb{P} \leqslant \int_{\{m \leqslant t_\varepsilon\}} t_\varepsilon\, d\mathbb{P} = \sum_{k=m}^{\infty} k\mathbb{P}(t_\varepsilon = k) \\ &\leqslant \sum_{k=m}^{\infty} k\mathbb{P}(k^{-1/2}|R_k| > \varepsilon) \leqslant \frac{1}{m} \sum_{1}^{\infty} k^2\mathbb{P}(k^{-1/2}|R_k| > \varepsilon) \\ &\equiv A/m. \end{aligned}$$

In summary, we get

$$\mathbb{E}(\tau_c) \geqslant c^*\mathbb{E}(\tau_c) - c^* \int_{\{\tau_c \leqslant t_\varepsilon\}} \tau_c\, d\mathbb{P} \geqslant c^*\mathbb{E}(\tau_c) - c^*A/m,$$

and therefore,

$$\mathbb{E}(\tau_c) \leqslant \frac{c^*A}{m(c^*-1)} \equiv B.$$

But, since $\tau_c \geqslant m$, $\mathbb{E}(\tau_c) \geqslant m$, so that $m \leqslant B$. This is impossible for $m^2 > c^*A/(c^*-1)$. Thus the assertion holds with $m_0(c) = \langle \sqrt{c^*A/(c^*-1)} \rangle$, an integer. ▮

We now show, that under $\mathbb{P}_\theta$,

$$n(\theta_n - \theta) \equiv T_n = S_n + R_n,$$

where $(R_n)_n$ satisfies (7). As before, let $\psi = \psi(x, \theta)$ be any smooth and bounded nonincreasing score-function, and let F be a fixed d.f. Suppose that the equation $\int \psi(x, \theta) F(dx) = 0$ has a root θ_0. Define

$$H(\theta) = \int \psi(x, \theta) F(dx),$$

and let

$$\alpha_n(\theta) := n^{1/2} \int \psi(x, \theta)[F_n(dx) - F(dx)],$$

denote the empirical process pertaining to the family $\{\psi(\cdot, \theta): \theta \in \Theta\}$. Write

$$H(\theta_n) = (\theta_n - \theta_0) H'(\theta_0) + (\theta_n - \theta_0)^2 H''(\theta^*)/2,$$

for some θ^* between θ_0 and θ_n. Hence, up to an error term $\mathcal{O}(1)$,

$$\begin{aligned} T_n H'(\theta_0) \equiv n(\theta_n - \theta_0) H'(\theta_0) &= -n^{1/2}\alpha_n(\theta_0) - n^{1/2}[\alpha_n(\theta_n) - \alpha_n(\theta_0)] \\ &- n(\theta_n - \theta_0)^2 H''(\theta^*)/2 = \sum_1^n \eta_i + R_n, \end{aligned}$$

where

$$R_n = -n^{1/2}[\alpha_n(\theta_n) - \alpha_n(\theta_0)] - n(\theta_n - \theta_0)^2 H''(\theta^*)/2$$

and $\eta_i = -\psi(\xi_i, \theta_0)$. Of course, $\eta_1, \eta_2, \dots$, are i.i.d. random variables with zero means and variance $\int \psi^2(x, \theta_0) F(dx)$. The boundedness of ψ yields boundedness of the η's. For $\mathcal{F}_n$ we may take $\sigma(\xi_1, \dots, \xi_n)$. Thus according to Lemmas 2 and 3 it remains to show that for every $\varepsilon > 0$ both

$$\mathbb{P}(|\alpha_n(\theta_n) - \alpha_n(\theta_0)| > \varepsilon) \to 0 \tag{8}$$

and

$$\mathbb{P}(n^{1/2}(\theta_n - \theta_0)^2 > \varepsilon) \to 0, \tag{9}$$

exponentially fast as $n \to \infty$. The second assertion may be of interest in itself. For (8) and (9) the following exponential bound (see, e.g., Strassen and Dudley [6]) will be useful.

LEMMA 4. *Let $\eta_1, \ldots, \eta_n$ be independent random variables with expectation zero and bounded by some constant $M > 0$. Then for every $\varepsilon > 0$,*

$$\mathbb{P}\left(\left|\sum_1^n \eta_i\right| > \varepsilon\right) \leqslant 2\exp[-\varepsilon^2/(9nM^2)].$$

Now, for $0 < \delta < 1$,

$$\{\theta_n - \theta_0 > \delta\} = \{\theta_n > \theta_0 + \delta\} \subset \left\{\int \psi(x, \theta_0 + \delta)\, F_n(dx) > 0\right\}.$$

But

$$n^{1/2} \int \psi(x, \theta_0 + \delta)\, F_n(dx) = \alpha_n(\theta_0 + \delta) - \alpha_n(\theta_0) + n^{1/2} H(\theta_0 + \delta) + \alpha_n(\theta_0).$$

By boundedness and smoothness, $|(\partial/\partial\theta)\, \psi(x, \theta)| \leqslant C < \infty$ for all $x \in \mathbb{R}$ and θ (in a neighbourhood of θ_0). In this situation Lemma 6 from Stute [8] yields a stochastic upper bound for the fluctuation of α_n in a (small) neighbourhood Θ_0 of θ_0.

LEMMA 5. *For each (small enough) $0 < \delta < 1$ and every $s > 0$,*

$$\mathbb{P}\Big(\sup_{\substack{|\theta - \theta'| \leqslant \delta \\ \theta, \theta' \in \Theta_0}} |\alpha_n(\theta) - \alpha_n(\theta')| \geqslant 64 s\delta(\ln \delta^{-1})^{1/2}\Big) \leqslant C_0 \exp(-s^2/9),$$

where C_0 is some constant depending on Θ_0 but not on δ, s, and n.

This implies, in particular, that for $\delta = \delta_n = \varepsilon^{1/2}/n^{1/4}$,

$$\mathbb{P}(|\alpha_n(\theta_0 + \varepsilon/n^{1/2}) - \alpha_n(\theta_0)| \geqslant n^{-1/4}(\ln n)^2) \to 0,$$

exponentially fast (i.e., $\mathbb{P}(\cdot) = O(\exp(-n^\lambda))$ for some $\lambda > 0$). Furthermore,

$$n^{1/2} H(\theta_0 + \delta_n) = n^{1/2}\delta_n \frac{H(\theta_0 + \delta_n) - H(\theta_0)}{\delta_n} \to -\infty,$$

as $-n^{1/4}$. Thus it remains to show that $\mathbb{P}(|\alpha_n(\theta_0)| \geqslant c_n) \to 0$, exponentially fast, whenvever $c_n \to +\infty$ as $n^{1/4}$. However, this is an easy consequence of Lemma 4. The case $\{\theta_n - \theta_0 < -\delta\}$ is treated similarly, whence (9). Condition (8) is an easy consequence of (9) and Lemma 5. ∎

2. Some Limit Results

We conclude this paper by providing one further application of our representation of $n(\theta_n - \theta) = S_n + R_n$, where S_n is a partial sum of independent identically distributed bounded random variables and R_n is an error term satisfying (7).

First it follows from the Borel–Cantelli lemma that along with $(S_n)_n$, the sequence $(\theta_n)_n$ fulfills a law of the iterated logarithm,

$$\limsup_{n\to\infty} \frac{n^{1/2}|\theta_n - \theta|}{(2\ln\ln n)^{1/2}\sigma_\theta} = 1 \qquad \text{with probability one.}$$

Second, defining the process G_n by

$$G_n(t) = (n\sigma_\theta^{-1})^{1/2}[\theta_{\langle nt\rangle} - \theta], \qquad 0 \leqslant t \leqslant 1,$$

then $G_n \to B$ in distribution as $n \to \infty$, where B is a standard Brownian motion. This follows at once from Donsker's invariance principle applied to S_n, since $n^{-1/2}\sup_{1\leqslant m\leqslant n}|R_m|$ is asymptotically negligible. In fact, for $\varepsilon > 0$ and $n_0 \geqslant 1$ we have

$$\mathbb{P}(n^{-1/2}\sup_{1\leqslant m\leqslant n}|R_m| > \varepsilon) \leqslant \mathbb{P}(n^{-1/2}\sup_{1\leqslant m\leqslant n_0}|R_m| > \varepsilon) + \sum_{n_0\leqslant m}\mathbb{P}(m^{-1/2}|R_m| > \varepsilon),$$

which, by (7), can be made arbitrarily small by choosing n_0 large and then letting n tend to infinity.

References

[1] Chow, Y. S., Robbins, H., and Teicher, H. (1965). Moments of randomly stopped sums. *Ann. Math. Statist.* **36** 789–799.

[2] Gundy, R., and Siegmund, D. (1967). On a stopping rule and the central limit theorem. *Ann. Math. Statist.* **38** 1915–1917.

[3] Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

[4] Lai, T. L. (1976). On confidence sequences. *Ann. Statist.* **4** 265–280.

5] PFANZAGL, J. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika* **14** 249–272.

6] STRASSEN, V., AND DUDLEY, R. M. (1969). *The Central Limit Theorem and ε-Entropy.* Lecture Notes in Math., No. 89, pp. 224–231. Springer-Verlag Berlin/New York.

7] STUTE, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86–107.

8] STUTE, W. (1981). Parameter estimation in smooth empirical processes. In *Math. Forschungsber.* 64/81. University of Siegen, West Germany.