The Annals of Probability
1989, Vol. 17, No. 2, 737–759

CHARACTERIZATION OF THE CLUSTER SET OF THE LIL SEQUENCE IN BANACH SPACE[1]

BY KENNETH S. ALEXANDER

University of Southern California

Let $S_n = X_1 + \cdots + X_n$, where $X_1, X_2, \ldots$ are iid Banach-space-valued random variables with weak mean 0 and weak second moments. Let K be the unit ball of the reproducing kernel Hilbert space associated to the covariance of X. We show that the cluster set of $\{S_n/(2n \log\log n)^{1/2}\}$ either is empty or has the form αK, where $0 \le \alpha \le 1$. A series condition is given which determines the value of α. In a companion paper, examples are given to show that all $\alpha \in [0, 1]$ do occur.

1. Introduction. Let $X, X_1, X_2, \ldots$ be iid random variables taking values in a separable Banach space $(B, \|\cdot\|)$, $S_n := X_1 + \cdots + X_n$, and $a_n := (2n \log\log n)^{1/2}$. X is said to satisfy the bounded law of the iterated logarithm ($X \in$ BLIL) if S_n/a_n is bounded almost surely, and the compact law of the iterated logarithm ($X \in$ CLIL) if there exists a compact $D \subset B$ such that

$$C(\{S_n/a_n\}) = D \quad \text{a.s.} \tag{1.1}$$

and

$$d(S_n/a_n, D) \to 0 \quad \text{a.s.,} \tag{1.2}$$

where $C(\{y_n\})$ is the cluster set of the sequence $\{y_n\}$, and $d(y, D) := \inf\{\|y - z\|: z \in D\}$ is the distance from y to D.

When S_n/a_n stays bounded, the one-dimensional LIL tells us that $X \in WM_0^2$, that is, $Ef(X) = 0$ and $Ef(X)^2 < \infty$ for all f in the dual B^*.

Kuelbs (1976) showed that there is only one possible compact cluster set D when $X \in$ CLIL: The unit ball K of the reproducing kernel Hilbert space $H_P \subset B$ associated to the covariance of X. That is,

$$K = \left\{ \int x f(x)\, dP(x) \colon f \in \overline{B^*}, \|f\|_2 \le 1 \right\}, \tag{1.3}$$

where $\|\cdot\|_2$ is the $L^2(P)$ norm, $\overline{B^*}$ is the closure of B^* in $L^2(P)$ and P is the law of X. K exists as a subset of B whenever $X \in WM_0^2$. It is easy to show that an equivalent definition of K is

$$K = \{ y \in B \colon f(y) \le \|f\|_2 \text{ for all } f \in B^* \}, \tag{1.4}$$

and that

$$\sup\{\|y\| \colon y \in K\} = \sigma := \sup\{\|f\|_2 \colon f \in B_1^*\} < \infty, \tag{1.5}$$

Received August 1987; revised June 1988.

[1]Research supported by National Science Foundation Grant DMS-87-02906.

AMS 1980 *subject classifications*. Primary 60B12; secondary 60F15.

Key words and phrases. Law of the iterated logarithm, cluster set, Banach-space-valued random variables.

where B_1^* is the unit ball of B^*. For details and more about H_P, see Goodman, Kuelbs and Zinn (1981).

In contrast to the finite-dimensional situation in infinite-dimensional Banach space (1.1) can hold, but (1.2) fail, for a nonempty $D \subset B$, or we can have $C(\{S_n/a_n\}) \neq K$ a.s. when $X = \text{BLIL}$. (Examples from the literature will be described below.) This motivates the question we examine here: What are the possible values of the cluster set

$$A := C(\{S_n/a_n\}),$$

and when does each occur?

Some remarkable recent results of Ledoux and Talagrand (1988) put our question in clearer perspective. They showed that $X \in \text{BLIL}$ if and only if

$$E\|X\|^2/\log\log\|X\| < \infty, \tag{1.6}$$

$$X \in WM_0^2 \tag{1.7}$$

and

$$\{S_n/a_n\} \text{ is bounded in probability,} \tag{1.8}$$

while $X \in \text{CLIL}$ if and only if (1.6) holds,

$$\left\{f(X)^2\colon f \in B_1^*\right\} \text{ is uniformly integrable} \tag{1.7'}$$

and

$$S_n/a_n \to 0 \quad \text{in probability.} \tag{1.8'}$$

It is known [see Ledoux and Talagrand (1988)] that (1.7′) is equivalent to the compactness of K. It is also equivalent to the total boundedness (or equivalently, compactness, since B_1^* is closed) of B_1^* in $L^2(P)$.

de Acosta and Kuelbs (1983) showed that (1.7) and (1.8′) imply that $A = K$ a.s., while in Hilbert space, (1.6) and (1.7) imply (1.1) and (1.2) with $D = K$, even if K is not compact. In contrast, however, examples of the following types have been constructed:

(i) $X \in \text{BLIL}$ and $A = K$ a.s., but K is not compact [Goodman, Kuelbs, and Zinn (1981)].

(ii) X bounded, $X \in \text{BLIL}$, K compact and $A = K$ a.s., but $X \notin \text{CLIL}$. Thus necessarily (1.8′) fails, and $\{S_n/a_n\}$ clusters at points of K but makes bounded excursions well outside of K [Kuelbs (1981)].

(iii) X bounded, $X \in \text{BLIL}$ and K compact, but $A = \varnothing$ a.s. and $0 < \liminf\|S_n/a_n\| \le \limsup\|S_n/a_n\| < \infty$. Thus again (1.8′) fails, and $\{S_n/a_n\}$ moves around in a spherical shell but never clusters anywhere [Kuelbs (1981)].

In all of these examples, the cluster set is either $\varnothing$ or K a.s. In a companion paper [Alexander (1987)] it is shown that for each $\alpha \in [0,1]$, there exists a bounded, c_0-valued random variable X for which the cluster set is αK a.s. The following, our main result, shows that $\varnothing$ and these multiples of K are the only possible cluster sets and characterizes those random variables X which have a given cluster set.

Some notation needed is

$$(1.9) \qquad n_k := n_k(\gamma) := [\gamma^k], \qquad I_k := I_k(\gamma) := [n_k, n_{k+1}).$$

When there is no ambiguity or when a result does not depend upon γ, we will suppress the γ in this notation. Throughout most of this article we will think of γ as fixed but not yet specified.

THEOREM 1.1. *Suppose $X \in WM_0^2$. Define $\alpha \in [0,1]$ by*

$$(1.10) \qquad \alpha^2 := \sup\left\{\beta \geq 0: \sum_{k=1}^{\infty} k^{-\beta}\mathbb{P}[\|S_n/a_n\| < \delta \text{ for some } n \in I_k] = \infty \text{ for all } \delta > 0\right\}$$

whenever this set is not empty. Then

$$A = \begin{cases} \alpha K & a.s. \text{ if the set in (1.10) is not empty,} \\ \varnothing & a.s. \text{ if the set in (1.10) is empty.} \end{cases}$$

Consider the case of real-valued X with $EX = 0$, $EX^2 = \sigma^2$. A normal approximation would say

$$\mathbb{P}\left[|S_{n_k}/a_{n_k}| > \beta^{1/2}\sigma\right] \approx \exp(-\beta \log\log n_k) \approx k^{-\beta}.$$

Some proofs of the one-dimensional LIL are of course based on this, together with the Borel–Cantelli lemma and the fact that $\Sigma_k k^{-\beta}$ diverges exactly when $\beta \leq 1$; it follows that the cluster set is $[-\sigma, \sigma]$. In Theorem 1.1 the same series $\Sigma_k k^{-\beta}$ appears, but with each term multiplied by a probability. Typically this probability will be very close to 0 or 1, according as δ is greater than or less than the median of $\|S_{n_k}/a_{n_k}\|$. Thus heuristically Theorem 1.1 says that in Banach space we still determine the cluster set from convergence or divergence of $\Sigma k^{-\beta}$, but include only those k for which $\|S_{n_k}/a_{n_k}\|$ is small in probability.

All proofs are postponed until Section 2.

Theorem 1.1 remains true, with mild additional assumptions, if B is not separable, but the technicalities involved obscure the main ideas, so we will do the proof for the separable case and then mention the modifications necessary for the general case.

Under the stronger moment assumption, there is an alternative, more natural characterization of those random variables for which the cluster set is empty: If $EX = 0$ and $E\|X\|^2 < \infty$, then $A = \varnothing$ a.s. if and only if $\liminf E\|S_n/a_n\| > 0$ [Alexander (1987)].

The following are fairly direct consequences of Theorem 1.1. Corollary 1.2 was proved by Goodman, Kuelbs and Zinn (1981) under an additional assumption on X.

COROLLARY 1.2. *Suppose $X \in WM_0^2$ and suppose that for each $\varepsilon > 0$ there exist $\delta > 0$ and a subset $J(\varepsilon)$ of the positive integers such that*

$$\mathbb{P}[\|S_n/a_n\| \leq \varepsilon] \geq \delta \quad \text{for all } n \in J(\varepsilon)$$

and

$$\sum_{n \in J(\varepsilon)} 1/n(\log n)^{\beta} = \infty \quad \text{for all } \beta < 1.$$

Then $A = K$ *a.s.*

COROLLARY 1.3 [de Acosta and Kuelbs (1983)]. *Suppose* $X \in WM_0^2$ *and* $S_n/a_n \to 0$ *in probability. Then* $A = K$ *a.s.*

2. Proof of the characterization. Before we begin, it is worth examining the heuristics for why K should generally be the cluster set of $\{S_n/a_n\}$. Suppose B is partitioned into finitely many sets E_j, most of which have small diameter, with the jth having (small) probability p_j and let y_j be a point of the jth set. Further, let

$$P_n := n^{-1} \sum_{i=1}^{n} \delta_{X_i} \quad \text{and} \quad \nu_n := n^{1/2}(P_n - P)$$

be the nth empirical measure and empirical process,

$$f_j := \nu_n(E_j)/p_j b_n, \qquad b_n := (2 \log\log n)^{1/2},$$

and let f be the function given by

$$f := f_j \quad \text{on } E_j.$$

Then, approximating each X_i in E_j by y_j, we expect that roughly

$$S_n/a_n \approx \sum_j y_j(nP_n(E_j))/a_n \approx \sum_j y_j \nu_n(E_j)/b_n \approx \sum_j y_j f_j p_j \approx \int y f(y)\, dP(y).$$

Here the second approximation uses the fact that $\sum_j y_j p_j \approx EX = 0$. Setting

$$u_f := \int y f(y)\, dP(y),$$

we observe that

$$u_f = u_{\pi(f)}, \tag{2.1}$$

where $\pi(f)$ is the projection of f onto $\overline{B^*}$ in $L^2(P)$. Thus we obtain the form in (1.3); we may think of $f(x)$ as the signed normalized "density" of excess or deficient points X_i near x, with the excess or deficiency being relative to expectation, and we may think of the partition as producing a simple-function approximation to this "density."

Further, by Bernstein's inequality [(2.14) below; see Bennett (1962)], if $f \in B^*$, then

$$\mathbb{P}\left[\nu_n(E_j)/p_j b_n \approx f_j\right] \approx \exp\left(-f_j^2 p_j \log\log n\right).$$

Since the $\nu_n(E_j)$ are approximately independent, we expect that

$$(2.2)\quad \mathbb{P}\big[S_n/a_n \approx u_f\big] \approx \exp\Big(-\sum_j f_j^2 p_j \log\log n\Big) = \exp\big(-\|f\|_2^2 \log\log n\big).$$

Summing over a geometrically increasing subsequence of values of n, we get convergence if and only if $\|f\|_2 > 1$, which leads to the cluster set K.

Our proof makes this argument rigorous, by showing that the errors made in the approximations above are sufficiently small sufficiently often. As (2.2) hints, the success of this approach depends in a sense only on $\|f\|_2$, which leads to the possible forms αK for the cluster set, since $\alpha K = \{u_f\colon \|f\|_2 \le \alpha\}$.

Our first lemma is due to Kuelbs (1976); we present here an alternative proof which works even if B is not separable. Let

$$K^\delta := \{y \in B\colon d(y, K) < \delta\}.$$

LEMMA 2.1. *If $X \in WM_0^2$, then $A \subset K$ a.s.*

PROOF. Observe that

$$\begin{aligned}&[A \not\subset K]\\ &\quad\subset \bigcup_{k,\, m \ge 1} \big[d(S_k/a_k, K) \ge 3\sigma/m;\ d(S_n/a_n, S_k/a_k) < \sigma/m \text{ i.o. (in } n)\big]\\ &\quad:= \bigcup_{k,\, m\ge 1} C_{km}.\end{aligned}$$

If $\omega' := (X_1, \dots, X_k)$ satisfies

$$(2.3)\qquad d(S_k/a_k, K) > 3\sigma/m,$$

then there exists an $f = f_{\omega'} \in B^*$ such that $f \le 1$ on $K^{\sigma/m}$ but $f \ge 1$ on the ball $B(S_k/a_k, 2\sigma/m)$. Since $K^{\sigma/m} \supset (1 + m^{-1})K$, it follows that $f \le m/(m+1)$ on K, so $\|f\|_2 \le m/(m+1)$. Therefore, letting $\omega'' := (X_{k+1}, X_{k+2}, \dots)$,

$$\mathbb{P}\big[\omega''\colon f((S_n - S_k)/a_{n-k}) \ge 1 \text{ i.o. in } n\big] = 0.$$

But if $(\omega', \omega'') \in C_{km}$, then $d(S_k/a_k, (S_n - S_k)/a_{n-k}) \le 2\sigma/m$ i.o. Hence from the definition of f, $\mathbb{P}[\omega''\colon (\omega', \omega'') \in C_{km}] = 0$ for each ω' satisfying (2.3), and the lemma follows. □

The following is an immediate corollary, as in Kuelbs (1981), Lemma 1.

LEMMA 2.2. *If K is separable, then there exists a nonrandom closed set $D \subset K$ such that $A = D$ a.s.*

To establish Theorem 1.1, it is therefore sufficient to establish the following result, which shows that whether or not $u_h \in A$ a.s. depends only on $\|h\|_2$.

THEOREM 2.3. *Suppose $X \in WM_0^2$ and $h \in \overline{B^*}$. The following are equivalent*:

(i) $u_h \in A$ *a.s.*

(ii) *For each* $\beta < \|h\|_2^2$ (*for* $\beta = 0$ *if* $h = 0$) *and each* $\delta > 0$, *there exists* $\gamma > 1$ *such that*

$$\sum_{k=1}^{\infty} k^{-\beta} \mathbb{P}\left[\|S_n/a_n\| < \delta \text{ for some } n \in I_k(\gamma)\right] = \infty. \tag{2.4}$$

(iii) *For each* $\beta < \|h\|_2^2$ (*for* $\beta = 0$ *if* $h = 0$) *and each* $\delta > 0$ *and* $\gamma > 1$, (2.4) *holds*.

The proof of this theorem depends on careful choice of a partition of B, as in the heuristic preceding Lemma 2.1. For this we will need some notation and definitions. We say a partition $\Pi = \{E_0, \dots, E_J\}$ of B is *bounded* if all but one of its blocks are bounded sets. E_0 will always denote the unique unbounded block. Given Π, we let $\mathscr{S}$ denote the (finite) σ-algebra it generates, $p_j := P(E_j)$, and, when $p_0 > 0$,

$$y_0 := -p_0^{-1} P(E_0^c) E(X | E_0^c).$$

When $X \in WM_0^2$ and $p_0 > 0$, y_0 is the weak mean of X given $X \in E_0$; that is,

$$f(y_0) = E\big(f(X) | X \in E_0\big) \quad \text{for all } f \in B^*. \tag{2.5}$$

Note that $E(X|\mathscr{S})$ is well defined a.s. for $X \in E_0^c$; in a convenient abuse of notation we define

$$E(X|\mathscr{S}) := y_0 \quad \text{when } X \in E_0.$$

Though we have not yet specified it, let us consider the partition Π as fixed, with $p_0 < 1$, and set

$$X' := E(X|\mathscr{S}), \qquad X'' := X - E(X|\mathscr{S}),$$

$$S_n' := \sum_{i=1}^{n} X_i', \qquad S_n'' := \sum_{i=1}^{n} X_i''.$$

By combining blocks if necessary, we may always assume $p_j > 0$ for all $1 \le j \le J$.

Heuristically, S_n'/a_n is the signal, and S_n''/a_n the noise, in S_n/a_n. The following special construction will help us to elucidate the dependencies between the signal and the noise, and between the bounded and unbounded parts of the signal, which we will show are quite weak. The idea is that, thinking of $E_0, \dots, E_J$ as "bins," one can construct $\{S_n, n \ge 1\}$ by first deciding which r.v.'s X_i go in which bin; the numbers of the r.v.'s $X_1, \dots, X_n$ in each bin have a multinomial distribution for fixed n. Given this information, the r.v.'s in bin E_j form an iid sequence with law $P(\cdot | E_j)$, so we may then choose their locations in E_j independently with that distribution.

Further, it is useful (see the proof of Lemma 2.15) to know that in allocating the r.v.'s among the bins, one may first decide which of $X_1, \dots, X_n$ go in E_0, then do an independent allocation of the remaining ones among $E_1, \dots, E_J$.

Thus let $\tilde{X}, \tilde{X}_1, \tilde{X}_2, \ldots,$ be iid with distribution $P(\cdot|E_0^c)$, and let $\tilde{P}_n$ be the corresponding empirical measure. Let $\{(\tilde{T}_{in}, \ldots, \tilde{T}_{jn}): n \geq 1\}$ be random variables with the same joint distribution as $\{(n\tilde{P}_n(E_1), \ldots, n\tilde{P}_n(E_J)): n \geq 1\}$, that is, multinomial with parameters n, $p_1/(1-p_0), \ldots, p_J/(1-p_0)$ for each n. Let $\{T_{0n}: n \geq 1\}$ be independent of the $\tilde{T}_{jn}$ with the same joint distribution as $\{nP_n(E_0): n \geq 1\}$ and let

$$T_{jn} := \tilde{T}_{j, n - T_{n0}}, \qquad 1 \leq j \leq J, \, n \geq 1.$$

Then

$$\{(T_{0n}, \ldots, T_{Jn}): n \geq 1\} \stackrel{d}{=} \{(nP_n(E_0), \ldots, nP_n(E_J)): n \geq 1\}.$$

For each $0 \leq j \leq J$ let $\{\xi_{jl}: l \geq 1\}$ be iid with distribution $P(\cdot|E_j)$, independent of the $\tilde{T}_{jn}$ and T_{0n} and independent for distinct j. We may then assume that

$$S_n = \sum_{j=0}^{J} \sum_{l=1}^{T_{jn}} \xi_{jl}, \qquad S_n' = \sum_{j=0}^{J} \sum_{l=1}^{T_{jn}} E\xi_{jl}, \qquad S_n'' = \sum_{j=0}^{J} \sum_{l=1}^{T_{jn}} (\xi_{jl} - E\xi_{jl}),$$

where, continuing our abuse of notation, $E\xi_{0l}$ should be interpreted as the weak mean y_0.

The following fact is isolated here for easy reference, as it will be used several times. The proof is completely straightforward, so we omit it.

LEMMA 2.4. *Let $\{F_n, n \geq 1\}$ be any sequence of events and $\beta \geq 0$. Then convergence or divergence of*

$$\sum_{k=1}^{\infty} k^{-\beta} \mathbb{P}[F_n \text{ occurs for some } n \in I_k(\gamma)]$$

does not depend on $\gamma > 1$.

The following result underlies many of the operations performed in our proofs.

LEMMA 2.5 [Kuelbs (1981)]. *Let $y \in B$. Then $y \in A$ a.s. if and only if*

$$\sum_{k=1}^{\infty} \mathbb{P}[\|S_n/a_n - y\| < \delta \text{ for some } n \in I_k] = \infty \quad \text{for all } \delta > 0.$$

We can now take one of the main steps in proving Theorem 1.1, by showing that if S_n/a_n goes near u_h infinitely often, it is because the noise S_n''/a_n becomes small infinitely often when the signal S_n'/a_n is near u_h. The underlying idea is then that for any g with $\|g\|_2 < \|h\|_2$, since the signal is near u_g at least as often as it is near u_h and since the signal and noise are nearly independent, the noise will also become small infinitely often when the signal is near u_g, which ensures $u_g \in A$ a.s.

PROPOSITION 2.6. *Let $X \in \overline{WM_0^2}$ and $\theta > 0$, and let Γ be a bounded partition of B. Suppose $h \in \overline{B^*}$ and $u_h \in A$ a.s. Then the partition Π can be chosen* *so that*

$$\Pi \text{ refines } \Gamma \tag{2.6}$$

and

$$\sum_{k=1}^{\infty} \mathbb{P}[\|S_n'/a_n - u_h\| < \theta \text{ and } \|S_n''/a_n\| < \theta \text{ for some } n \in I_k] = \infty. \tag{2.7}$$

PROOF. We will construct a sequence $\Pi_0, \Pi_1, \dots$ of bounded partitions of B, each a refinement of the previous one, in such a way that eventually (2.7) holds.

Let $\Pi_0 := \{B\}$ and $\Pi_1 := \Gamma$. Given an increasing sequence $\Pi_0, \Pi_1, \dots,$ let

$$Y_i^{(l)} := E(X_i|\mathscr{S}_l) - E(X_i|\mathscr{S}_{l-1}), \qquad i, l \geq 1,$$

where $\mathscr{S}_l$ is the finite σ-algebra generated by Π_l, and let

$$U_n^{(l)} := \sum_{i=1}^{n} Y_i^{(l)}.$$

Let $U_n^{(0)} = Y_i^{(0)} = 0$.

Since $Y^{(1)} := E(X|\mathscr{S}_1)$ is finite-dimensional with mean 0 and weak (hence strong) second moments, $\{a_n^{-1}U_n^{(1)}\}$ is, with probability 1, a bounded sequence in a finite-dimensional space, so there exists $v_1 \in B$ a.s. such that (v_1, u_h) is a cluster point of $\{a_n^{-1}(U_n^{(1)}, S_n)\}$. Applying Lemma 2.2 to the random variable $(Y^{(1)}, X)$, we see that v_1 may be taken nonrandom.

Suppose now that for some $k \geq 1$, $\Pi_0, \dots, \Pi_k$ have been constructed so that

(i) Π_j refines Π_{j-1} for each $j \leq k$ and

(ii) some $(v_1, \dots, v_k, u_h) \in B^{k+1}$, with $\|v_l\| \geq 2\theta/3$ for all $2 \leq l \leq k$, is, with probability 1, a cluster point of $\{a_n^{-1}(U_n^{(1)}, \dots, U_n^{(k)}, S_n)\}$.

These are true for $k = 1$. If

$$\left\| u_h - \sum_{l=1}^{k} v_l \right\| < \theta, \tag{2.8}$$

then, since by (ii) $(\sum_{l=1}^k v_l, u_h - \sum_{l=1}^k v_l)$ is a cluster point of $a_n^{-1}(S_n', S_n'')$ a.s., (2.6) with $\Pi = \Pi_k$ follows from (i) and (2.7) from Lemma 2.5 applied to (X', X''). Thus we must show (2.8) holds for some $k \geq 1$.

One may think of

$$u_h = v_1 + \cdots + v_k + \left(u_h - \sum_{l=1}^{k} v_l \right)$$

as a decomposition of u_h corresponding to the decomposition

$$S_n = U_n^{(1)} + \cdots + U_n^{(k)} + \left(S_n - \sum_{l=1}^{k} U_n^{(l)} \right)$$

of S_n into increments arising from the successive partitioning of B. The idea is to choose the refinements so that if (2.8) fails, a quantity, which roughly speaking approximates the "total variation" $\sum_{l=1}^k \|v_l\|$, grows at least linearly in k, while the corresponding quantity for the decomposition of $a_n^{-1}S_n$ can only grow like $k^{1/2}$; this limits the size of k for which the two quantities can be equal, forcing (2.8) to hold eventually.

Fix k and suppose (i) and (ii) hold but (2.8) fails. Let $f_2,\dots,f_{k+1}\in B_1^*$ be such that

$$f_l(v_l)\ge 2\theta/3 \quad \text{for all } 2\le l\le k, \tag{2.9}$$

$$f_{k+1}\left(u_h-\sum_{l=1}^{k}v_l\right)=\left\|u_h-\sum_{l=1}^{k}v_l\right\|. \tag{2.10}$$

Let Π_{k+1} be a bounded partition which refines Π_k and satisfies

$$E\big(f_{k+1}(X)-E(f_{k+1}(X)|\mathscr{S}_{k+1})\big)^2<\theta^2/9. \tag{2.11}$$

Letting

$$S_n^{(l)}:=\sum_{j=1}^{l}U_n^{(j)},$$

we have by (2.11)

$$\limsup_{n\to\infty} f_{k+1}\big(a_n^{-1}(S_n-S_n^{(k+1)})\big)<\theta/3 \quad \text{a.s.,}$$

while by (ii), (2.10) and failure of (2.8),

$$\limsup_{r\to\infty} f_{k+1}\big(a_{m_r}^{-1}(S_{m_r}-S_{m_r}^{(k)})\big)\ge f_{k+1}\left(u_h-\sum_{l=1}^{k}v_l\right)\ge\theta \quad \text{a.s.}$$

for some random sequence (m_r) along which the limit behavior in (ii) occurs. Therefore

$$\limsup_{r\to\infty} f_{k+1}\big(a_{m_r}^{-1}U_{m_r}^{(k+1)}\big)\ge 2\theta/3 \quad \text{a.s.} \tag{2.12}$$

Now $\{a_n^{-1}U_n^{(k+1)}\}$ is a.s. a bounded sequence in a finite-dimensional subspace of B, so from (ii) and (2.12) we obtain $v_{k+1}\in B$ such that $\|v_{k+1}\|\ge 2\theta/3$ and $(v_1,\dots,v_{k+1},u_h)$ is, with probability 1, a cluster point of $a_n^{-1}(U_n^{(1)},\dots,U_n^{(k+1)},S_n)$ in B^{k+2}. As with v_1, v_{k+1} can be taken nonrandom, establishing (i) and (ii) for $k+1$.

Thus one can make (i) and (ii) hold for arbitrary large k, so long as (2.8) continues to fail. Consider the random variables

$$Z_i:=\sum_{l=2}^{k}f_l(Y_i^{(l)}),\qquad R_n:=\sum_{i=1}^{n}Z_i.$$

The summands in Z_i are orthogonal [see (2.5)] so

$$EZ_1^2=\sum_{l=2}^{k}Ef_l^2(Y_1^{(l)})\le\sum_{l=2}^{k}Ef_l^2(X)\le(k-1)\sigma^2$$

and hence

$$\limsup_{n\to\infty} a_n^{-1} R_n \le (k-1)^{1/2}\sigma \quad \text{a.s.}$$

But by (ii), $t := \sum_{l=2}^{k} f_l(v_l)$ is a cluster point of $a_n^{-1}R_n$ a.s., while by (2.9), $t \ge 2(k-1)\theta/3$. Therefore $2(k-1)\theta/3 \le (k-1)^{1/2}\sigma$ so we have shown

$$\text{(i) and (ii) imply } k \le (3\sigma/2\theta)^2 + 1.$$

This means that (2.8) must hold for some finite k, which, as we have mentioned, proves the proposition. □

The next lemma is also standard; we include it for ready reference. Define

$$L_k := L_k(\gamma) := I_{k-1}(\gamma) \cup I_k(\gamma) \cup I_{k+1}(\gamma).$$

LEMMA 2.7. *Let $Y, Y_1, Y_2, \dots$ be iid B-valued random variables with weak mean 0, and $s^2 := \sup_{f \in B_1^*} Ef^2(Y)$. Let N be an I_k-valued stopping time and let $M > 0$. Then*

$$(2.13)\quad \mathbb{P}\left[\left\|\sum_{i=n_k}^{N} Y_i\right\| > Mn_k^{1/2}\right] \le 2\mathbb{P}\left[\left\|\sum_{i=n_k}^{n_{k+1}} Y_i\right\| > \left(M - 2(\gamma-1)^{1/2}s\right)n_k^{1/2}\right].$$

The same result holds if I_k, n_k, n_{k+1} and γ are replaced by L_k, n_{k-1}, n_{k+2} and γ^3.

PROOF. Let $F = \{f_1, f_2, \dots\}$ be a countable norm-determining subset of B_1^*. When the first event in (2.13) occurs, define T to be the least $l \ge 1$ such that

$$f_l\left(\sum_{i=n_k}^{N} Y_i\right) > Mn_k^{1/2}.$$

By Chebyshev's inequality,

$$\mathbb{P}\left[f_l\left(\sum_{i=n+1}^{n_{k+1}} Y_i\right) \ge 2(\gamma-1)^{1/2} s n_k^{1/2}\right] \le 1/2$$

for all $l \ge 1$ and $n \in I_k$, so the lemma follows by conditioning on N and T. □

We will need the following variant of Bernstein's inequality [see Bennett (1962)]: For $Y_1, \dots, Y_n$ iid mean zero real random variables bounded in magnitude by $b > 0$ with $\operatorname{var}(Y_1) \le s^2$, $\delta > 0$ and $0 < M < 3n^{1/2}s^2\delta/b$,

$$(2.14)\qquad \mathbb{P}\left[\left|n^{-1/2}\sum_{i=1}^{n} Y_i\right| > M\right] \le 2\exp\left(-(1-\delta)M^2/2s^2\right).$$

LEMMA 2.8. *For every $\delta > 0$,*

$$\sum_{k=1}^{\infty} \mathbb{P}\left[\|(S_n' - S_{n_{k-1}}')/a_{n_{k-1}}\| > \delta \text{ for some } n \in L_k(\gamma)\right] < \infty,$$

provided $\gamma - 1$ is sufficiently small. The same result holds if n_{k-1} and L_k are replaced by n_k and I_k.

PROOF. Let H be the (finite-dimensional) span of the range of X'. There exist $f_1, \dots, f_N \in B_1^*$ such that for $y \in H$, $\max_l f_l(y) \geq \|y\|/2$. Hence by Lemma 2.7 and Bernstein's inequality (2.14), for large k,

$$\begin{aligned}
&\mathbb{P}\left[\|(S_n' - S_{n_{k-1}}')/a_{n_{k-1}}\| > \delta \text{ for some } n \in L_k\right] \\
&\leq \sum_{k=1}^{N} 2\mathbb{P}\left[f_l(S_{n_{k+2}}' - S_{n_{k-1}}')/a_{n_{k-1}} > \delta/2\right] \\
&\leq 4N\exp\left(-\delta^2(\log\log n_{k-1})/8(\gamma^3 - 1)\sigma^2\right).
\end{aligned}$$

One now need only take γ to satisfy $\delta^2/8(\gamma^3 - 1)\sigma^2 > 1$. □

LEMMA 2.9. *Suppose $h_n \in B^*$ and $h_n \to h$ a.s. Then h has a version $\tilde{h}$ such that*:

(i) *$\tilde{h}$ is linear.*
(ii) *$\tilde{h}(y) = h(y)$ for all y such that $h_n(y)$ converges to finite limit.*

PROOF. Let $F := \{y \in B\colon h_n(y)$ converges to a finite limit$\}$ and let W_F be a Hamel basis for the (not necessarily closed) subspace F. Define h_0 on F by $h_0(y) := \lim_n h_n(y)$. Then $P(F) = 1$ and $h = h_0$ a.s. on F. Let W_B be an extension of W_F to a Hamel basis for B. Let $\tilde{h}$ be the linear function on B determine by the values

$$\tilde{h}(v) := \begin{cases} h_0(v) & \text{if } v \in W_F, \\ 0 & \text{if } v \in W_B \setminus W_F. \end{cases}$$

Since h_0 is linear on F, we have $h_0 = \tilde{h}$ on F, and the lemma follows. □

Henceforth, therefore, we will assume that all elements of $\overline{B^*}$ mentioned are linear. The linear version of h need not, of course, be continuous.

LEMMA 2.10. *Suppose $h \in \overline{B^*}$. For each $\eta > 0$ there exists $\tau > 0$, not depending on the partition Π, such that*

$$\mathbb{P}\left[\|S_n'/a_n - u_h\| < \tau\right] \leq \exp\left(-(1 - \eta)\|h\|_2^2 \log\log n\right)$$

for all sufficiently large n.

PROOF. Since $h \in \overline{B^*}$, we cannot have $u_h = u_\psi$ for any function ψ with $\|\psi\|_2 < \|h\|_2$. Thus $\|h\|_2 = \inf\{\alpha\colon u_h \in \alpha K\}$. Therefore given $\eta > 0$ there exists $\varphi \in B^*$ such that $\varphi < 1$ on $(1 - \eta/3)\|h\|_2 K$ but $\varphi(u_h) = 1 + \delta$ for some $\delta > 0$.

Letting $\tau := \delta/\|\varphi\|_{B^*}$ we have by Bernstein's inequality (2.14)

$$\begin{aligned}\mathbb{P}[\|S_n'/a_n - u_h\| < \tau] &\le \mathbb{P}[|\varphi(S_n'/a_n) - (1+\delta)| < \tau\|\varphi\|_{B^*}] \\ &\le \mathbb{P}[|\varphi(S_n'/a_n)| > 1] \\ &\le \exp(-(1-\eta/3)(\log\log n)/\|\varphi\|_2^2).\end{aligned}$$

But

$$\|\varphi\|_2 = \varphi(u_{\varphi/\|\varphi\|_2}) \le \sup_{y\in K}\varphi(y) \le ((1-\eta/3)\|h\|_2)^{-1}$$

and the lemma follows. □

The next lemma is an immediate consequence of Lemma 3.2 of de Acosta and Kuelbs (1983), since $\|u_h - EX'h(X')\| \le \sup_{g\in B_1^*} Eg(X)h(X'') \le \sigma(Eh^2(X''))^{1/2}$.

LEMMA 2.11. *Suppose $h \in \overline{B^*}$, $\delta > \sigma(Eh^2(X''))^{1/2}$ and $\eta > 0$. Then*

$$\mathbb{P}[\|S_n'/a_n - u_h\| < \delta] \ge \exp(-(1+\eta)\|h\|_2^2 \log\log n)$$

for all sufficiently large n.

The next lemma is a technical variation on Proposition 2.6.

LEMMA 2.12. *Suppose $X \in WM_0^2$, $\theta > 0$, $0 \le \mu < \beta \le 1$ (or $\mu = \beta = 0$), Λ is a bounded partition of B and for every $\delta > 0$,*

$$\sum_{k=1}^{\infty} k^{-\beta}\mathbb{P}[\|S_n/a_n\| < \delta \textit{ for some } n \in I_k] = \infty. \tag{2.15}$$

Then the bounded partition Π can be chosen so that Π refines Λ and

$$\sum_{k=1}^{\infty} k^{-\mu}\mathbb{P}[\|S_n''/a_n\| < \theta \textit{ for some } n \in I_k] = \infty.$$

PROOF. For $\mu = \beta = 0$, this is an immediate consequence of Lemma 2.5 and Proposition 2.6, so we may assume $\beta > 0$.

Let $Y, Y_1, Y_2, \dots$ be iid with $\mathbb{P}[Y = 1] = \mathbb{P}[Y = -1] = 1/2$, and let $V_n := \sum_{i=1}^n Y_i$. Let $\lambda \in (\mu^{1/2}, \beta^{1/2})$. It suffices to consider θ small enough so that

$$(\lambda - 3\theta)^2 \ge \mu. \tag{2.16}$$

By (2.15) and Lemma 2.11,

$$\sum_{k=1}^{\infty} \mathbb{P}[|V_{n_k}/a_{n_k} - \lambda| < \delta]\mathbb{P}[\|S_n/a_n\| < \delta \text{ for some } n \in I_k] = \infty$$

for all $\delta > 0$, so by independence and Lemmas 2.8 and 2.4,

$$\sum_{k=1}^{\infty} \mathbb{P}[|V_n/a_n - \lambda| + \|S_n/a_n\| < 3\delta \text{ for some } n \in I_k] = \infty$$

for all $\delta > 0$.

Thus by Lemma 2.5, $(\lambda, 0)$ is a.s. in the cluster set of $\{a_n^{-1}(V_n, S_n)\}$. By Proposition 2.6, applied with $\Gamma = \{\{-1\},\{1\}\} \times \Lambda$, there is a partition Π of B such that for the decomposition $(Y, X) = (Y', X') + (Y'', X'')$ corresponding to the partition $\{\{-1\},\{1\}\} \times \Pi$, we have $Y'' = 0$ and

$$\sum_{k=1}^{\infty} \mathbb{P}[|V_n/a_n - \lambda| + \|S_n'/a_n\| < \theta \text{ and } \|S_n''/a_n\| < \theta \text{ for some } n \in I_k] = \infty.$$

By Lemma 2.8, independence, Bernstein's inequality (2.14) and Lemma 2.4, then,

$$\begin{aligned}\infty &= \sum_{k=1}^{\infty} \mathbb{P}\left[|V_{n_k}/a_{n_k} - \lambda| < 2\theta \text{ and } \|S_n''/a_n\| < \theta \text{ for some } n \in I_k\right] \\ &\leq \sum_{k=1}^{\infty} k^{-(\lambda-3\theta)^2}\mathbb{P}[\|S_n''/a_n\| < \theta \text{ for some } n \in I_k]\end{aligned}$$

and the lemma follows from (2.16). □

LEMMA 2.13. *Let $\xi_1, \xi_2, \dots$ be iid with $n^{-1}\sum_{i=1}^n \xi_i \to 0$ a.s., let $\delta > 0$ and let $\{m_k\}$ be an increasing sequence of positive integers such that for some $\varepsilon > 0$, $m_k \geq \varepsilon \sum_{j=1}^{k-1} m_j$ for all $k \geq 1$. Then*

$$\sum_{k=1}^{\infty} \mathbb{P}\left[\left\|\sum_{i=1}^{m_k} \xi_i\right\| > \delta m_k\right] < \infty.$$

PROOF. We may assume $\varepsilon < 1$. Let $r_k := \sum_{j=1}^k m_j$. If

$$\left\|\sum_{i=r_{k-1}+1}^{r_k} \xi_i\right\| > \delta m_k,$$

then either

$$\|S_{r_k}/m_k\| > \delta/2 \quad \text{or} \quad \|S_{r_{k-1}}/m_k\| > \delta/2.$$

Since $m_k \geq \varepsilon r_k/2$, the former implies $\|S_{r_k}/r_k\| > \varepsilon\delta/4$, while the latter implies $\|S_{r_{k-1}}/r_{k-1}\| > \varepsilon\delta/2$. The assumed SLLN says each of these two events can happen only finitely often, so the lemma follows from the Borel–Cantelli lemma. □

Of course for separable B the assumption that $n^{-1}\sum_{i=1}^n \xi_i \to 0$ a.s. is equivalent to the assumption that $E\|\xi\| < \infty$ and $E\xi = 0$, but this is not true for nonseparable B.

Next we will give some facts about our special construction of the random variables X_i. Define

$$
\begin{aligned}
t_{jn} &:= [np_j],\\
m_{jn} &:= \left[8p_j^{1/2}a_n\right],\\
W_n &:= \{r \in \mathbb{Z}_+ : |r - t_{0n}| \le m_{0n}\},\\
R_n &:= \left\{(r_1,\dots,r_J) \in \mathbb{Z}_+^J : |r_j - t_{jn}| \le m_{jn} \text{ for all } 1 \le j \le J\right\},\\
Q_n &:= W_n \times R_n,\\
U_n &:= (T_{1n},\dots,T_{Jn}),\\
V_n &:= (T_{0n},\dots,T_{Jn})
\end{aligned}
$$

and define the event

$$C_k := [V_n \in Q_n \text{ for all } n \in L_k].$$

LEMMA 2.14. $\sum_{k=1}^{\infty}\mathbb{P}(C_k^c) < \infty$.

PROOF. We may assume $\gamma^3 < 2$. Then for large k, by Lemma 2.7 and Bernstein's inequality (2.14),

$$
\begin{aligned}
&\mathbb{P}\left[|T_{jn} - t_{jn}| > m_{jn} \text{ for some } n \in L_k\right]\\
&\quad\le \mathbb{P}\left[|T_{jn} - ET_{jn}| > 4p_j^{1/2}a_n \text{ for some } n \in L_k\right]\\
&\quad\le 2\mathbb{P}\left[n_{k+2}^{-1/2}|T_{jn_{k+2}} - ET_{jn_{k+2}}| > 2p_j^{1/2}(2\log\log n_{k-1})^{1/2}\right]\\
&\quad\le 4\exp(-3\log\log n_{k-1})\\
&\quad\le 4k^{-2}
\end{aligned}
$$

and the lemma follows. □

Now define events

$$
\begin{aligned}
A_k(\delta) &:= [\|S_n''/a_n\| < \delta \text{ for some } n \in I_k],\\
A_k^*(\delta) &:= [\|S_n''/a_n\| < \delta \text{ for some } n \in L_k],\\
B_k(\delta, h) &:= \left[\|S_{n_k}'/a_{n_k} - u_h\| < \delta\right].
\end{aligned}
$$

Let $Y, Y_1, Y_2, \dots$ be an independent copy of the sequence $X, X_1, X_2, \dots$ and recall that $\tilde{X}, \tilde{X}_1, \tilde{X}_2, \dots$ are iid with distribution $P(\cdot|E_0^c)$. Let

$$Y = Y' + Y'' \quad \text{and} \quad \tilde{X} = \tilde{X}' + \tilde{X}''$$

be the decompositions induced by the partition Π.

The next two lemmas provide the required degree of independence between the signal S_n' and the noise S_n''. The underlying idea is the following: $B_k(\delta, h)$ affects $A_k(\delta)$ only through the values of the T_{jn_k}, which give the numbers of random variables $\xi_{jl} - E\xi_{jl}$ to be summed in forming S_{n_k}''. By Lemma 2.14, the variation in the T_{jn_k} is $O(a_{n_k})$. But the SLLN says that changing the number of summands by $O(a_{n_k})$ should change the sum S''_{n_k} by only $o(a_{n_k})$; the effect of this on $A_k(\delta)$ is negligible.

Of course the SLLN does not apply to the unbounded variable $\xi_{01} - E\xi_{01}$ in general, which introduces some complications.

Throughout our remaining proofs, we will make statements which, though we do not explicitly say so each time, are valid only for sufficiently large k.

LEMMA 2.15. *Suppose $h \in \overline{B^*}$ satisfies $\|h\|_2 \le 1$, and $\delta > \sigma E(h^2(X''))^{1/2}$, where σ is as given in* (1.5). *Suppose also that $p_0 < 1/100$. Then*

$$\sum_{k=2}^{\infty} \left[\mathbb{P}(A_k(\delta)) - \mathbb{P}(A_k^*(4\delta)|B_{k-1}(\delta, h))\right]^+ < \infty, \tag{2.17}$$

provided $\gamma - 1$ is small enough.

PROOF. The idea is that the conditional probability in (2.17) can be expressed [see (2.24)] as a convex combination of conditional probabilities $\mathbb{P}(A_k^*(4\delta)|V_{n_{k-1}} = q)$. Terms with $q \notin Q_{n_{k-1}}$ contribute negligibly [see (2.24) and (2.25)], so it is enough to show that, up to summable error terms, these latter conditional probabilities exceed $\mathbb{P}(A_k(\delta))$ uniformly over $q \in Q_{n_{k-1}}$. First, this uniformity is shown to hold (modulo changing δ to 3δ) when the conditioning is only on the first coordinate q_0, not on all of q [see (2.20)]. For this a coupling of the conditional and unconditional distributions of portions of the sequence S''_n is used. Then, using another coupling, it is shown that (modulo changing 3δ to 4δ) the conditional probability is the same, up to summable error terms, for all q with a given first coordinate q_0 [see (2.21)]. This gives uniformity over all $q \in Q_{n_{k-1}}$.

Suppose first that $p_0 > 0$. Fix $k \ge 2$ and $q = (q_0, \dots, q_J) \in Q_{n_{k-1}}$. Let N be the least n such that $T_{0n} = q_0$. N is finite since $T_{0n}/n \to p_0$ a.s. Then in distribution, jointly for all $n \in L_k$,

$$S''_{N+n-n_{k+1}} \overset{d}{=} G_n := \sum_{l=1}^{q_0} (\xi_{0l} - E\xi_{0l}) + \sum_{i=1}^{N-q_0} \tilde{X}''_i + \sum_{i=1}^{n-n_{k-1}} Y''_i.$$

This just says that the X''_i's which occur after N may be replaced by $Y''_1, Y''_2, \dots,$ which is true because N is a stopping time. In contrast, if we condition on $T_{0n_{k-1}} = q_0$, then jointly for all $n \in L_k$,

$$S''_n \overset{d}{=} H_n := \sum_{l=1}^{q_0} (\xi_{0l} - E\xi_{0l}) + \sum_{i=1}^{n_{k-1}-q_0} \tilde{X}''_i + \sum_{i=1}^{n-n_{k-1}} Y''_i.$$

The r.v.'s G_n and H_n give us our coupling between the conditional and unconditional distributions of certain groups of the r.v.'s S''_n. The key property is that the unbounded parts of G_n and H_n are the same, while the SLLN heuristic preceding this lemma works for the bounded parts.

Specifically, N is independent of $\{\tilde{X}_i'': i \geq 1\}$, so

$$\max_{n \in I_k} \|G_n - H_n\| \stackrel{d}{=} \left\| \sum_{i=1}^{|N-n_{k-1}|} \tilde{X}_i'' \right\|.$$

Further, if $n \in I_k$ and $m = n - N + n_{k-1}$, then either $m \in L_k$ or $|n_{k-1} - N| > n_k - n_{k-1} > (\gamma - 1)n_{k-1}/2$; the latter implies the event C_{k-1}^c [see (2.19)]. Therefore, provided $\gamma - 1$ is small,

$$\begin{aligned} \mathbb{P}(A_k(\delta)) &\leq \mathbb{P}\left[\|S''_{N+m-n_{k-1}}/a_{n_{k+1}}\| < \delta \text{ for some } m \in L_k\right] \\ &\quad + \mathbb{P}\left[|n_{k-1} - N| > (\gamma - 1)n_{k-1}/2\right] \\ &\leq \mathbb{P}\left[\|G_m/a_m\| < 2\delta \text{ for some } m \in L_k\right] + \mathbb{P}(C_{k-1}^c) \\ &\leq \mathbb{P}\left[\|H_n/a_n\| < 3\delta \text{ for some } n \in L_k\right] \\ &\quad + \mathbb{P}\left[\left\| \sum_{i=1}^{|N-n_{k-1}|} \tilde{X}_i'' \right\| > \delta a_{n_{k-1}}\right] + \mathbb{P}(C_{k-1}^c). \end{aligned} \tag{2.18}$$

It is easily verified that

$$|N - n_{k-1}| > 17p_0^{-1/2}a_{n_{k-1}} \text{ implies } C_{k-1}^c. \tag{2.19}$$

In fact, if $N < r := n_{k-1} - [17p_0^{-1/2}a_{n_{k-1}}]$, then since $q_0 \in W_0$,

$$T_{0r} \geq q_0 \geq t_{0n_{k-1}} - m_{0n_{k-1}} \geq n_{k-1}p_0 - 1 - 8p_0^{1/2}a_{n_{k-1}},$$

while

$$t_{0r} \leq rp_0 \leq n_{k-1}p_0 - 17p_0^{1/2}a_{n_{k-1}} + 1,$$

so

$$T_{0r} - t_{0r} \geq 9p_0^{1/2}a_{n_{k-1}} - 2 > m_{0r},$$

and C_{k-1} fails. Similarly, $N > s := n_{k-1} + [17p_0^{-1/2}a_{n_{k-1}}]$ implies $T_{0s} - t_{0s} < m_{0s}$.

Thus, letting $m_k := [17p_0^{-1/2}a_{n_{k-1}}]$, we have by (2.18), (2.19) and an analog of Lemma 2.7,

$$\begin{aligned} \mathbb{P}(A_k(\delta)) &\leq \mathbb{P}\left(A_k^*(3\delta) | T_{0n_{k-1}} = q_0\right) \\ &\quad + 2\mathbb{P}\left[\left\| \sum_{i=1}^{m_k} k_1\tilde{X}_i'' \right\| > \delta a_{n_{k-1}}/2\right] + 2\mathbb{P}(C_{k-1}^c). \end{aligned} \tag{2.20}$$

Letting

$$\lambda_k := 2\mathbb{P}\left[\left\| \sum_{i=1}^{m_k} \tilde{X}_i'' \right\| > \delta a_{n_{k-1}}/2\right] + 2\mathbb{P}(C_{k-1}^c),$$

we see from Lemmas 2.13 and 2.14 that $\sum_{k=1}^{\infty}\lambda_k < \infty$.

We wish to condition on all of $V_{n_{k-1}}$, not just on $T_{0n_{k-1}}$. To this end, let $q' \in Q_{n_{k-1}}$ with $q_0' = q_0$. Because of the latter, we can make another coupling in which only the bounded parts differ, as follows. Given $V_{n_{k-1}} = q$, jointly for all $n \in L_k$ we have

$$S_n'' \stackrel{d}{=} G_n^* := \sum_{j=0}^{J} \sum_{l=1}^{q_j} (\xi_{jl} - E\xi_{jl}) + \sum_{i=1}^{n-n_{k-1}} Y_i'',$$

while given $V_{n_{k-1}} = q'$,

$$S_n'' \stackrel{d}{=} H_n^* := \sum_{j=0}^{J} \sum_{l=1}^{q_j'} (\xi_{jl} - E\xi_{jl}) + \sum_{i=1}^{n-n_{k-1}} Y_i''.$$

Therefore as in (2.18) and (2.20),

$$\begin{aligned} &\mathbb{P}\big(A_k^*(3\delta) | V_{n_{k-1}} = q'\big) \\ &\quad \le \mathbb{P}\big(A_k^*(4\delta) | V_{n_{k-1}} = q\big) + \mathbb{P}\left[\sum_{j=1}^{J} \left\| \sum_{l=1}^{|q_j - q_j'|} (\xi_{jl} - E\xi_{jl}) \right\| > \delta a_{n_{k-1}} \right] \\ (2.21) \quad &\quad \le \mathbb{P}\big(A_k^*(4\delta) | V_{n_{k-1}} = q\big) + \sum_{j=1}^{J} 2\mathbb{P}\left[\left\| \sum_{l=1}^{2m_{jn_{k-1}}} (\xi_{jl} - E\xi_{jl}) \right\| > \delta a_{n_{k-1}}/2J \right] \\ &\quad := \mathbb{P}\big(A_k^*(4\delta) | V_{n_{k-1}} = q\big) + \lambda_k', \end{aligned}$$

where $\sum_{k=1}^{\infty} \lambda_k' < \infty$ by Lemma 2.13. It follows that

$$\begin{aligned} &\mathbb{P}\big(A_k^*(3\delta) | T_{0n_{k-1}} = q_0\big) \\ &\quad \le \sum_{r \in R_{n_{k-1}}} \mathbb{P}\big(A_k^*(3\delta) | V_{n_{k-1}} = (q_0, r)\big) \mathbb{P}\big(U_{n_{k-1}} = r | T_{0n_{k-1}} = q_0\big) \\ (2.22) \quad &\qquad + \mathbb{P}\big(U_{n_{k-1}} \notin R_{n_{k-1}} | T_{0n_{k-1}} = q_0\big) \\ &\quad \le \mathbb{P}\big(A_k^*(4\delta) | V_{n_{k-1}} = q\big) + \lambda_k' + \mathbb{P}\big(U_{n_{k-1}} \notin R_{n_{k-1}} | T_{0n_{k-1}} = q_0\big). \end{aligned}$$

We must bound the last term on the right-hand side of (2.22). Now for $n = n_{k-1}$,

$$(2.23) \qquad \mathbb{P}(U_n \notin R_n | T_{0n} = q_0) \le \sum_{j=1}^{J} \mathbb{P}\big[|\tilde{T}_{j,n-q_0} - t_{jn}| > m_{jn}\big].$$

Since $q_0 = [np_0] + 8\mu p_0^{1/2} a_n$ for some μ with $|\mu| < 1$ when $q_0 \in W_n$, and since $p_0 < 1/100$,

$$\begin{aligned} |t_{jn} - E\tilde{T}_{j,n-q_0}| &\le |np_j - (n - q_0)p_j/(1 - p_0)| + 1 \\ &\le 8|\mu| p_0^{1/2} p_j a_n/(1 - p_0) + 2 \\ &\le p_j^{1/2} a_n. \end{aligned}$$

Hence the right-hand side of (2.23) is bounded above by

$$\sum_{j=1}^{J} \mathbb{P}\left[|\tilde{T}_{j,\,n-q_0} - E\tilde{T}_{j,\,n-q_0}| > 6p_j^{1/2}a_n\right] \le 2J\exp(-3\log\log n).$$

Thus for $n = n_{k-1}$,

$$\mathbb{P}\left(U_{n_{k-1}} \notin R_{n_{k-1}} | T_{0n_{k-1}} = q_0\right) \le k^{-2}. \tag{2.24}$$

Using Lemma 2.11 and the proof of Lemma 2.14 gives

$$\begin{aligned}
&\mathbb{P}\left(V_{n_{k-1}} \notin Q_{n_{k-1}} | B_{k-1}(\delta, h)\right) \\
&\quad \le \mathbb{P}\left(V_{n_{k-1}} \notin Q_{n_{k-1}}\right)/\mathbb{P}(B_{k-1}(\delta, h)) \\
&\quad \le 4\exp(-5\log\log n_{k-1})/\exp\left(-2\|h\|_2^2 \log\log n_{k-1}\right) \\
&\quad \le k^{-2}.
\end{aligned} \tag{2.25}$$

Since $q \in Q_{n_{k-1}}$ is arbitrary, combining this fact with (2.20), (2.22) and (2.24) gives

$$\begin{aligned}
\mathbb{P}(A_k(\delta)) &\le \sum_{q \in Q_{n_{k-1}}} \mathbb{P}(A_k(\delta))\mathbb{P}\left(V_{n_{k-1}} = q | B_{k-1}(\delta, h)\right) \\
&\quad + \mathbb{P}\left(V_{n_{k-1}} \notin Q_{n_{k-1}} | B_{k-1}(\delta, h)\right) \\
&\le \sum_{q \in Q_{n_{k-1}}} \mathbb{P}\left(A_k^*(4\delta) | V_{n_{k-1}} = q\right)\mathbb{P}\left(V_{n_{k-1}} = q | B_{k-1}(\delta, h)\right) \\
&\quad + \lambda_k + \lambda'_k + 2k^{-2} \\
&\le \mathbb{P}(A_k^*(4\delta) | B_{k-1}(\delta, h)) + \lambda_k + \lambda'_k + 2k^{-2}
\end{aligned} \tag{2.26}$$

and the lemma follows.

If $p_0 = 0$, a simpler version of this same proof, with no need to first condition on $T_{0n_{k-1}}$, works. □

The next lemma provides a converse to Lemma 2.15.

LEMMA 2.16. *Under the hypotheses of Lemma* 2.15,

$$\sum_{k=1}^{\infty} \left[\mathbb{P}(A_k(\delta) | B_{k-1}(\delta, h)) - \mathbb{P}(A_k^*(4\delta))\right]^+ < \infty,$$

provided $\gamma - 1$ *is sufficiently small.*

PROOF. This proof is similar to that of Lemma 2.15, so we will continue with the notation of that proof and omit some of the details. In particular, let us assume $p_0 > 0$.

Fix $q \in Q_{n_{k-1}}$. Analogously to (2.21), we obtain for all $r \in R_{n_{k-1}}$,

$$\mathbb{P}\left(A_k(\delta) | V_{n_{k-1}} = q\right) \le \mathbb{P}\left(A_k^*(2\delta) | V_{n_{k-1}} = (q_0, r)\right) + \lambda'_k$$

so that

$$\begin{aligned}
&\mathbb{P}\big(A_k(\delta)|V_{n_{k-1}} = q\big)\\
&\quad\le \sum_{r\in R_{n_{k-1}}} \mathbb{P}\big(A_k^*(2\delta)|V_{n_{k-1}} = (q_0, r)\big)\\
&\qquad\qquad \times\mathbb{P}\big(V_{n_{k-1}} = (q_0, r)|T_{0n_{k-1}} = q_0, U_{n_{k-1}} \in R_{n_{k-1}}\big) + \lambda'_k\\
&\quad= \mathbb{P}\big(A_k^*(2\delta)|T_{0n_{k-1}} = q_0, U_{n_{k-1}} \in R_{n_{k-1}}\big) + \lambda'_k.
\end{aligned}$$

Therefore, since $q \in Q_{n_{k-1}}$ is arbitrary,

$$\begin{aligned}
&\mathbb{P}(A_k(\delta)|B_{k-1}(\delta, h))\\
&\quad\le \sum_{q\in Q_{n_{k-1}}} \mathbb{P}\big(A_k(\delta)|V_{n_{k-1}} = q\big)\mathbb{P}\big(V_{n_{k-1}} = q|B_{k-1}(\delta, h)\big)\\
&\qquad+\mathbb{P}\big(V_{n_{k-1}} \notin Q_{n_{k-1}}|B_{k-1}(\delta, h)\big)\\
&\quad\le \sum_{q_0\in W_{n_{k-1}}} \mathbb{P}\big(A_k^*(2\delta)|T_{0n_{k-1}} = q_0, U_{n_{k-1}} \in R_{n_{k-1}}\big)\\
&\qquad\qquad\times\mathbb{P}\big(T_{0n_{k-1}} = q_0, U_{n_{k-1}} \in R_{n_{k-1}}|B_{k-1}(\delta, h)\big)\\
&\qquad+\lambda'_k + k^{-2},
\end{aligned}\tag{2.27}$$

where (2.25) has been used. By (2.24),

$$\begin{aligned}
&\mathbb{P}\big(A_k^*(2\delta)|T_{0n_{k-1}} = q_0, U_{n_{k-1}} \in R_{n_{k-1}}\big)\\
&\quad\le \mathbb{P}\big(A_k^*(2\delta), T_{0n_{k-1}} = q_0\big)/(1 - k^{-2})\mathbb{P}\big(T_{0n_{k-1}} = q_0\big)\\
&\quad\le \mathbb{P}\big(A_k^*(2\delta)|T_{0n_{k-1}} = q_0\big) + 2k^{-2},
\end{aligned}$$

which with (2.27) shows that

$$\begin{aligned}
&\mathbb{P}(A_k(\delta)|B_{k-1}(\delta, h))\\
&\quad\le \sum_{q_0\in W_{n_{k-1}}} \mathbb{P}\big(A_k^*(2\delta)|T_{0n_{k-1}} = q_0\big)\\
&\qquad\qquad\times\mathbb{P}\big(T_{0n_{k-1}} = q_0|B_{k-1}(\delta, h)\big) + 3k^{-2} + \lambda'_k.
\end{aligned}\tag{2.28}$$

An argument using the variables G_n^* and H_n^*, similar to that which produced (2.20), gives

$$\mathbb{P}\big(A_k^*(2\delta)|T_{0n_{k-1}} = q_0\big) \le \mathbb{P}\big(A_k^*(4\delta)\big) + \lambda_k.$$

With (2.28) this leads to

$$\mathbb{P}\big(A_k(\delta)|B_{k-1}(\delta, h)\big) \le \mathbb{P}\big(A_k^*(4\delta)\big) + \lambda_k + 3k^{-2} + \lambda'_k$$

and the lemma is proved. □

PROOF OF THEOREM 2.3. The equivalence of (ii) and (iii) follows from Lemma 2.4, so we must show (i) implies (ii), and (iii) implies (i).

Suppose first $u_h \in A$ a.s. If $h = 0$, then (ii) follows from Lemma 2.5, so suppose $h \neq 0$. Let $0 \leq \beta < \|h\|_2^2$ and let $\eta > 0$ satisfy $(1-\eta)\|h\|_2^2 > \beta$. Let τ be as in Lemma 2.10, and $\theta < \tau/2$.

Now $h \in \overline{B^*}$, so h is a limit in $L^2(P)$, and a.s., of a sequence h_n of bounded linear functions on B. We may assume h satisfies (i) and (ii) of Lemma 2.9. Then regardless of what bounded partition is used, we have

$$\lim_{n\to\infty} h_n(E(X|\mathscr{P})) = \lim_{n\to\infty} E(h_n(X)|\mathscr{P}) = E(h(X)|\mathscr{P}),$$

which, by (ii) of Lemma 2.9, means

$$h(E(X|\mathscr{P})) = E(h(X)|\mathscr{P}) \quad \text{a.s.}$$

From this and the linearity of h it follows readily that there exists a bounded partition Γ, with unbounded block of probability less than $1/100$, such that $(Eh^2(X - E(X|\mathscr{P})))^{1/2} < 2\theta/\sigma$ whenever Π is a refinement of Γ. Here σ is as in (1.5).

By Proposition 2.6, there exists a refinement Π of Γ such that (2.7) holds. Let $\gamma - 1$ be small enough so that Lemma 2.8 applies with $\delta = \theta$, Lemma 2.14 applies with $\delta = 12\theta$ and Lemma 2.16 applies with $\delta = 3\theta$, and small enough so that for large k and all $y \in B$, $\|y/a_n - u_h\| < \theta$ for some $n \in L_k$ implies $\|y/a_{n_{k-1}} - u_h\| < 2\theta$. Then by Lemma 2.8 and (2.7),

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k(\theta) \cap B_{k-1}(3\theta, h)) = \infty$$

so that by Lemma 2.10,

$$\sum_{k=1}^{\infty} k^{-\beta}\mathbb{P}(A_k(3\theta)|B_{k-1}(3\theta, h)) = \infty.$$

Applying Lemma 2.16, we obtain

$$\sum_{k=1}^{\infty} k^{-\beta}\mathbb{P}(A_k^*(12\theta)) = \infty,$$

from which it follows easily that

$$\sum_{k=1}^{\infty} k^{-\beta}\mathbb{P}(A_k(12\theta)) = \infty. \tag{2.29}$$

Since X' is finite-dimensional with a second moment, $S_n'/a_n \to 0$ in probability. Therefore from (2.29)

$$\sum_{k=1}^{\infty} k^{-\beta}\mathbb{P}(A_k(12\theta))\mathbb{P}(B_{k-1}(12\theta, 0)) = \infty.$$

Hence by Lemma 2.15,

$$\sum_{k=1}^{\infty} k^{-\beta}\mathbb{P}\big(A_k^*(48\theta) \cap B_{k-1}(12\theta, 0)\big) = \infty.$$

Then from Lemma 2.8,

$$\begin{aligned}\infty &= \sum_{k=1}^{\infty} k^{-\beta}\mathbb{P}\big(A_k^*(48\theta);\ \|S_n'/a_n\| < 13\theta \text{ for all } n \in L_k\big)\\ &\le \sum_{k=1}^{\infty} k^{-\beta}\mathbb{P}\big(\|S_n/a_n\| < 61\theta \text{ for some } n \in L_k\big).\end{aligned} \tag{2.30}$$

The latter is then also infinite for L_k replaced by I_k; since θ is arbitrary, (ii) follows.

Conversely, suppose (iii) holds. Let $\varepsilon > 0$ and $\varphi := (1-\varepsilon)h$. Let $\beta, \mu, \eta > 0$ satisfy $\|h\|_2^2 > \beta > \mu > (1+\eta)\|\varphi\|_2^2$ ($\beta = \varphi = \eta = 0$ if $h = 0$). Let $\theta > 0$ and let Γ be as in the above proof that (i) implies (ii). Let $\gamma - 1$ be small enough so that (a) Lemma 2.15 applies with $\delta = \theta$ and h replaced by φ, (b) Lemma 2.8 applies with $\delta = \theta$ and (c) for large k and all $y \in B$, $\|y/a_{n_{k-1}} - u_\varphi\| < 2\theta$ implies $\|y/a_n - u_\varphi\| < 3\theta$ for all $n \in L_k$. By Lemma 2.12, there exists a bounded partition Π which refines Γ and satisfies

$$\sum_{k=1}^{\infty} k^{-\mu}\mathbb{P}\big(A_k(\theta)\big) = \infty.$$

By Lemma 2.11 we have $k^{-\mu} \le 2\mathbb{P}(B_{k-1}(\theta, \varphi))$ so that by Lemma 2.15,

$$\sum_{k=1}^{\infty} \mathbb{P}\big(A_k^*(4\theta) \cap B_{k-1}(\theta, \varphi)\big) = \infty.$$

By Lemma 2.8 and property (c) of $\gamma - 1$, this implies that

$$\sum_{k=1}^{\infty} \mathbb{P}\big[\|S_n''/a_n\| < 4\theta,\ \|S_n'/a_n - u_\varphi\| < 3\theta \text{ for some } n \in I_k\big] = \infty.$$

Since $\theta > 0$ is arbitrary, this and Lemma 2.5 show that $u_\varphi \in A$ a.s. Since $\varepsilon > 0$ is arbitrary, this shows $u_h \in A$ a.s. □

Since Theorem 1.1 is immediate from Theorem 2.3 and Corollary 1.3 is immediate from Theorem 1.1, all that remains is to prove Corollary 1.2.

PROOF OF COROLLARY 1.2. The assumptions imply that for each $\beta < 1$ and $\varepsilon > 0$,

$$\infty = \sum_{k=1}^{\infty} k^{-\beta} \sum_{n \in J(\varepsilon) \cap I_k} 1/n$$

so that

$$\begin{aligned}\infty &= \sum_{k=1}^{\infty} k^{-\beta}|I_k \cap J(\varepsilon)|/|I_k| \\ &\le \sum_{k=1}^{\infty} k^{-\beta} 1_{[I_k \cap J(\varepsilon) \neq \varnothing]} \\ &\le \delta^{-1} \sum_{k=1}^{\infty} k^{-\beta} \mathbb{P}[\|S_n/a_n\| \le \varepsilon \text{ for some } n \in I_k].\end{aligned}$$

This makes α [of (1.10)] equal to 1. □

3. The nonseparable case. Here the problem is best reformulated as follows. Let $\mathscr{F}$ be a collection of functions on a separable space T with Borel field $\mathscr{B}_T$, and Z a T-valued random variable with law P. Suppose $\mathscr{F}$ is bounded both pointwise and in $L^2(P)$. Let $B = l^\infty(\mathscr{F})$, endowed with the sup norm $\|\cdot\|_{\mathscr{F}}$, and $X = \delta_Z$, so that X is B-valued. Any random variable ξ taking values in a separable Banach space E can be placed in this context by taking $T = E$, $Z = \xi$ and $\mathscr{F} = E_1^*$, where E_1^* is the unit ball of the dual of E. Suppose $\|\sum_{i=1}^n c_i \delta_{Z_i} - u\|_{\mathscr{F}}$ is measurable for all n, all $L^2(P)$-continuous $u \in l^\infty(\mathscr{F})$ and all constants c_i. This makes all events used in our proofs measurable when they need to be. Let

$$K = \{u_\varphi \colon E\varphi(Z) = 0, \|\varphi\|_2 \le 1\},$$

where $u_\varphi \in l^\infty(\mathscr{F})$ is given by

$$u_\varphi(f) = \int f(y)\varphi(y)\, dP(y).$$

Most proofs remain the same, with span($\mathscr{F}$) and its closure in $L^2(P)$ generally used where B^* and $\overline{B^*}$ were previously used. Some definitions change, as follows. $\{E_0, \dots, E_J\}$ is now a partition of T, and

$$\begin{aligned}X'(f) &:= E(f(Z)|\mathscr{S}), \\ X''(f) &:= X(f) - X'(f) = f(Z) - E(f(Z)|\mathscr{S}).\end{aligned}$$

The ξ_{jl} now take values in $E_j \subset T$ with distribution $P_{(j)} := P(\cdot|E_j)$; we assume they are coordinates on the product space $(T^\infty, \mathscr{B}_T^\infty, P_{(j)}^\infty)$, to avoid measurability difficulties.

All of this avoids the use of the previous definitions of X' and the description (1.3) of K, which would involve integrating over the possibly nonseparable Banach space $l^\infty(\mathscr{F})$ using a measure not in general defined for all Borel sets. The reformulation is natural in the context of empirical processes, where nonseparable space of the form $l^\infty(\mathscr{F})$ arises naturally but T is generally separable. With the new definition of X' and X'', Lemma 2.9 is not needed.

When Lemma 2.13 is applied, in a nonseparable space boundedness of ξ_{j1} does not guarantee that ξ_{j1} satisfies the strong law of large numbers. But it can be shown that if X satisfies the SLLN, then so does $\delta_{\xi_{j1}}$ for each $1 \le j \le J$. We may take E_0 to have the form $\{y \colon \|y\|_{\mathscr{F}} > M\}$ for some $M > 0$. Therefore it is enough to add the assumption in Theorem 1.1, that either X, or for all $M > 0$, $X1_{[\|X\| \le M]}$, satisfies the SLLN.

It should be pointed out that in this nonseparable context, (1.4) need not be valid. Here compactness of K is equivalent to total boundedness of $\mathscr{F}$ and not necessarily to (1.7′).

That Lemma 2.7 remains valid in the nonseparable context follows from Lemma 3.2 of Alexander (1984).

In the nonseparable case the analog of (2.10) can in general only be made to hold to within an arbitrarily small ε but this is enough for the proof to work.

Acknowledgments. The author would like to thank Jim Kuelbs for helpful conversations and Ulrich Krengel for assistance with Lemma 2.9.

REFERENCES

ALEXANDER, K. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041–1067.

ALEXANDER, K. (1987). Unusual cluster sets for the LIL sequence in Banach space. *Ann. Probab.* To appear.

BENNETT, G. (1962). Probability inequalities for the sum of bounded random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

DE ACOSTA, A. and KUELBS, J. (1983). Some results on the cluster sets $C(\{S_n/a_n\})$ and the LIL. *Ann. Probab.* **11** 102–122.

GOODMAN, V., KUELBS, J. and ZINN, J. (1981). Some results on the LIL in Banach space with applications to weighted empirical processes. *Ann. Probab.* **9** 713–752.

KUELBS, J. (1976). A strong convergence theorem for Banach space valued random variables. *Ann. Probab.* **4** 744–771.

KUELBS, J. (1981). When is the cluster set of S_n/a_n empty? *Ann. Probab.* **9** 377–394.

LEDOUX, M. and TALAGRAND, M. (1988). Characterization of the law of the iterated logarithm in Banach spaces. *Ann. Probab.* **16** 1242–1264.

DEPARTMENT OF MATHEMATICS DRB 306
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90089-1113