The Annals of Probability
1994, Vol. 22, No. 1, 17–27

LARGE DEVIATIONS, MODERATE DEVIATIONS AND LIL FOR EMPIRICAL PROCESSES[1]

BY LIMING WU

Wuhan University

Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. r.v.'s with values in a measurable space $(E, \mathscr{E})$ of law μ, and consider the empirical process $L_n(f) = (1/n)\Sigma_{k=1}^n f(X_k)$ with f varying in a class of bounded functions $\mathfrak{F}$. Using a recent isoperimetric inequality of Talagrand, we obtain the necessary and sufficient conditions for the large deviation estimations, the moderate deviation estimations and the LIL of $L_n(\cdot)$ in the Banach space of bounded functionals $\ell_\infty(\mathfrak{F})$. The extension to the unbounded functionals is also discussed.

1. Introduction and main results. Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. r.v.'s defined on a probability space $(\Omega, \mathscr{B}, \mathbb{P})$ with values in a measurable space $(E, \mathscr{E})$, of law μ. Consider the empirical measures

$$L_n := \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k}, \qquad n \geq 1,$$

which are random elements of $M_1(E)$, the space of probability measures on $(E,\mathscr{E})$. The well-known Sanov theorem tells us that $\mathbb{P}(L_n \in \cdot)$ satisfies, as $n \to \infty$, the large deviation principle (LDP) on $M_1(E)$ equipped with the τ-topology $\sigma(M_1(E), b\mathscr{E})$ [where $b\mathscr{E}$ denotes the space of bounded and measurable real functions on $(E, \mathscr{E})$], with speed $1/n$ and with the rate function given by

$$h(\nu;\mu) = \begin{cases} \int_E \frac{d\nu}{d\mu}\log\frac{d\nu}{d\mu}\,d\mu, & \text{if } \nu \ll \mu, \\ +\infty, & \text{otherwise}, \end{cases} \tag{1.1}$$

which is the relative entropy of ν w.r.t. μ.

Now let $(\lambda(n))_{n\geq 1}$ be an increasing sequence of positive numbers so that

$$\lambda(n) \to \infty \quad \text{and} \quad \frac{\lambda(n)}{\sqrt{n}} \to 0. \tag{1.2}$$

It is a simple consequence of the Cramér method, as we showed in [8], Section 4.1, that

$$\mathbb{P}\left(\frac{\sqrt{n}}{\lambda(n)}(L_n - \mu) \in \cdot\right)$$

Received September 1992.
[1]Partially supported by the NSF of China
AMS 1991 *subject classifications*. 60F10, 60B12, 60G50.
Key words and phrases. Large deviations, moderate deviations, law of iterated logarithm (LIL), isoperimetric inequality, Smirnov–Kolmogorov theorem.

satisfies the LDP on $M_b(E)$, the space of signed measures of finite variation on $(E, \mathscr{E})$ equipped with the τ-topology, with speed $\lambda^{-2}(n)$ and with the rate function given by

$$(1.3)\quad \mathbf{I}(\nu) = \begin{cases} \dfrac{1}{2}\displaystyle\int\left(\frac{d\nu}{d\mu}\right)^2 d\mu, & \text{if } \nu \ll \mu \text{ with } \nu(E) = 0, \\ +\infty & \text{otherwise,} \end{cases}$$

where $\nu \in M_b(E)$. This is the so-called *moderate deviation* (MD).

By the contraction principle in LD theory, for every $\mathbf{f} = (f_1, \dots, f_d)\colon E \to \mathbb{R}^d$, with $f_i \in b\mathscr{E}$,

$$(L_n - \mu)(\mathbf{f}) = \int \mathbf{f}\, d(L_n - \mu) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{f}(X_k)$$

also satisfies the LD and MD estimations. However, in nonparametric statistics, we need the uniform estimations of $(L_n - \mu)(f)$ over a class of functions $\mathfrak{F}$. This is the objective of this article.

More precisely, given a class of functions $\mathfrak{F} \subseteq b\mathscr{E}$, let $\ell_\infty(\mathfrak{F})$ be the space of all bounded real functions on $\mathfrak{F}$ with sup norm $\|F\|_{\mathfrak{F}} = \sup_{f\in\mathfrak{F}} |F(f)|$. This is a nonseparable Banach space if $\mathfrak{F}$ is infinite. Every $\nu \in M_b(E)$ corresponds to an element $\nu^{\mathfrak{F}}$ in $\ell_\infty(\mathfrak{F})$ given by $\nu^{\mathfrak{F}}(f) = \nu(f) := \int f\, d\nu$ for all $f \in \mathfrak{F}$. Our aim is then to establish the LD and MD estimations of $(L_n - \mu)^{\mathfrak{F}}$ in $\ell_\infty(\mathfrak{F})$.

If $\mathfrak{F}$ is finite, these estimations follow from the two results recalled above. However, if $\mathfrak{F}$ is infinite, neither the two results above nor the Cramér theorem about LD and the recent result of Ledoux [4] about MD can be applied directly to the present setting because of the nonseparability of $\ell_\infty(\mathfrak{F})$.

Throughout this paper we assume that $\mathfrak{F}$ is *countable* or, instead that *the processes* $\{(L_n - \mu)(f);\ f \in \mathfrak{F}\}$ *are separable*, to avoid various measurability problems (see Dudley [2] for the treatments of such questions in the general case). We work actually in the following setting:

$$\text{(H1)}\qquad 0 \le f \le 1 \quad \text{for all } f \text{ in } \mathfrak{F}.$$

Let $(\varepsilon_n)_{n\ge1}$ be a sequence of i.i.d. r.v.'s defined still on $(\Omega, \mathfrak{F}, \mathbb{P})$ with $\mathbb{P}(\varepsilon_n = \pm 1) = 1/2$, independent of $(X_n)_{n\ge1}$. Introduce the following quantities:

$$(1.4)\qquad H(n) = \mathbb{E}\left\| \sum_{k=1}^{n} \varepsilon_k f(X_k) \right\|_{\mathfrak{F}};$$

$$(1.5)\qquad \overline{H}(n) = \mathbb{E}\left\| \sum_{k=1}^{n} f(X_k) - n\mu(f) \right\|_{\mathfrak{F}};$$

$$(1.6)\qquad H(n, \eta) = \mathbb{E}\left\| \sum_{k=1}^{n} \varepsilon_k f(X_k) \right\|_{\mathfrak{F}_\eta}$$

where

$$\mathfrak{F}_\eta = \left\{ f - g;\ f, g \in \mathfrak{F} \text{ and } d_2(f,g) := \left(\int (f-g)^2 \, d\mu \right)^{1/2} \le \eta \right\}.$$

Using the finite approximation and a recent isoperimetric inequality of Talagrand, we shall establish the following theorem.

THEOREM 1. *Under* (H1), *the following properties are equivalent*:

(i) $\mathbb{P}(L_n^{\mathfrak{F}} \in \cdot)$ *as* $n \to +\infty$ *satisfies the LDP on* $\ell_\infty(\mathfrak{F})$ *with speed* $1/n$ *and with the rate function given by*

$$h_{\mathfrak{F}}(F) = \inf\{h(\nu;\mu) | \nu \in M_1(E) \text{ and } \nu^{\mathfrak{F}} = F \text{ on } \mathfrak{F}\}; \tag{1.7}$$

(ii) $(\mathfrak{F}, d_2)$ *is totally bounded and*

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{H(n,\eta)}{n} = 0; \tag{1.8}$$

(iii) $(\mathfrak{F}, d_2)$ *is totally bounded and* $(L_n - \mu)^{\mathfrak{F}} \to 0$ *in probability in* $\ell_\infty(\mathfrak{F})$.

THEOREM 2. *Assume* (H1). *Let* $(\lambda(n);\ n \ge 1)$ *be as in* (1.2). *The following properties are equivalent*:

(i) $\mathbb{P}([\sqrt{n}/\lambda(n)](L_n - \mu)^{\mathfrak{F}} \in \cdot)$ *satisfies LDP on* $\ell_\infty(\mathfrak{F})$ *with speed* $\lambda^{-2}(n)$ *and with the rate function given by*

$$\mathbf{I}_{\mathfrak{F}}(F) = \inf\{\mathbf{I}(\nu) | \nu \in M_b(E) \text{ and } \nu^{\mathfrak{F}} = F \text{ on } \mathfrak{F}\}; \tag{1.9}$$

(ii) $(\mathfrak{F}, d_2)$ *is totally bounded and*

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{H(n,\eta)}{\sqrt{n}\,\lambda(n)} = 0; \tag{1.10}$$

(iii) $(\mathfrak{F}, d_2)$ *is totally bounded and* $[\sqrt{n}/\lambda(n)](L_n - \mu)^{\mathfrak{F}} \to 0$ *in probability in* $\ell_\infty(\mathfrak{F})$.

From Theorem 2, we can derive the following:

COROLLARY 3. *Assume* (H1). *The compact LIL of Strassen holds for* $(L_n - \mu)(\cdot)$ *in* $\ell_\infty(\mathfrak{F})$ (*i.e.*, $\mathbf{K} := [\mathbf{I}_{\mathfrak{F}} \le 1/2]$ *is compact in* $\ell_\infty(\mathfrak{F})$ *and with probability* 1, *the limit set of the sequence* $[\sqrt{n}/\lambda(n)](L_n - \mu)^{\mathfrak{F}}$ *is exactly* $\mathbf{K}$), *if and only if one of the equivalent properties in Theorem* 2 *holds for* $\lambda(n) = \sqrt{2 \log\log n}$.

REMARKS.

(i) When $E = \mathbb{R}$ and $\mathfrak{F} = \{(-\infty, t];\ t \in \mathbb{R}\}$, Theorem 1 was shown in [8] as a direct consequence of the Sanov theorem for the τ-topology.

(ii) If $\mathfrak{F}$ is a Donsker class, then $(\mathfrak{F}, d_2)$ is totally bounded and

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{H(n,\eta)}{\sqrt{n}} = 0,$$

by Ledoux and Talagrand ([6], Theorem 14.6, page 405). Then the three results above hold.

(iii) Corollary 3 (even improved versions) was obtained by Ledoux and Talagrand [5] and Yurich [10].

(iv) The extension of Theorems 1 and 2 to the unbounded case will be given in Section 3.

2. Proofs of the main results. Our main tool will be the following isoperimetric bound established recently by Talagrand [7], Theorem 3.5.

LEMMA 1 (Talagrand). *Assume* (H1). *Set* $\sigma := \sup_{f \in \mathfrak{F}} \int (f - \mu(f))^2 \, d\mu)^{1/2}$ *and* $S := n\sigma^2 + H(n)$, *where* $H(n)$ *is given in* (1.4). *Then for some universal constant* $L > 0$, *we have*

$$\forall\, t \geq LH(n): \mathbb{P}\left(\left\| \sum_{k=1}^{n} f(X_k) - n\mu(f) \right\|_{\mathfrak{F}} \geq t \right) \leq \exp(-\phi_{L,S}(t)), \tag{2.1}$$

where the function $\phi_{L,S}(t)$ *for* $t \geq 0$ *is given by*

$$\phi_{L,S}(t) = \frac{t^2}{L^2 S}, \qquad if\, t \leq LS, \tag{2.2}$$

$$\phi_{L,S}(t) = \frac{t}{L}\left(\log \frac{et}{LS}\right)^{1/2}, \qquad if\, t \geq LS. \tag{2.2$'$}$$

We also require the following lemma.

LEMMA 2. *Let* $(\xi_n)_{n \geq 1}$ *be a sequence of i.i.d. r.v.'s with values in a Banach space* $(B, \|\cdot\|)$ *(not necessarily separable) defined on a complete probability space* $(\Omega, \mathfrak{F}, \mathbb{P})$, *and let* (a_n) *be a sequence of positive numbers increasing to infinity. Assume that there is a countable subset D of the unit ball of* B' *such that* $\|x\| = \sup_{f \in D} f(x)$ *for all x in B.*

If $S_n / a_n \to 0$ *in probability and* $a_n^{-p} \mathbb{E} \max_{k \leq n} \|\xi_k\|^p \to 0$, *then* $a_n^{-p} \mathbb{E} \|S_n\|^p$ *tends to* 0, *where* $p \in (0, +\infty)$.

PROOF (Following the proof of Lemma 7.2 in [6]). By Proposition 6.8 of [6], page 156,

$$\mathbb{E} \max_{k \leq n} \|S_k\|^p \leq 2 \cdot 4^p \mathbb{E} \max_{k \leq n} \|\xi_k\|^p + 2 \cdot (4t_0)^p,$$

where $t_0 = t_0(n) = \inf\{t > 0;\ \mathbb{P}(\max_{k \leq n} \|S_k\| > t) \leq (2 \cdot 4^p)^{-1}\}$. Now by the Ottaviani inequality we have

$$\mathbb{P}\left(\max_{k \leq n} \|S_k\| > s + t \right) \leq \frac{\mathbb{P}(\|S_n\| > t)}{1 - \max_{k \leq n} \mathbb{P}(\|S_k\| > s)}.$$

Taking $s = t = \varepsilon a_n$ in this inequality, were ε is arbitrary, we see clearly that $t_0(n) \le \varepsilon a_n$ for all n large enough. Hence the desired convergence follows. □

We turn now to the following:

PROOF OF THEOREM 1. We begin with the implication (i) ⇒ (iii). It is well known that the total boundedness of $(\mathfrak{F}, d_2)$ follows from the relative inf-compactness of $\mathbf{I}_{\mathfrak{F}}$ on $\ell_\infty(\mathfrak{F})$. We show now that the relative inf-compactness of $\mathbf{I}_{\mathfrak{F}}$ on $\ell_\infty(\mathfrak{F})$ follows that of $h_{\mathfrak{F}}$ (assumed by the definition of LDP). In fact, for every $a > \varepsilon > 0$, let

$$A_{a,\varepsilon} = \{\nu \in M_b(E); \mathbf{I}(\nu) \le a \text{ and } \nu^+(E) \wedge \nu^-(E) \ge \varepsilon\},$$

and set $\tilde{\nu}^{\pm} = (\nu^{\pm}(E))^{-1} \cdot \nu^{\pm}$; we have $|\nu|(E) \le 2\sqrt{a}$ and

$$h_\mu(\tilde{\nu}^+) + h_\mu(\tilde{\nu}^-) \le 2\mathbf{I}(\nu) \le 2a.$$

Then $A^{\mathfrak{F}}_{a,\varepsilon} := \{\nu^{\mathfrak{F}}; \nu \in A_{a,\varepsilon}\}$ is relatively compact in $\ell_\infty(\mathfrak{F})$. Now observe that as $\varepsilon \to 0$ the distance between $A^{\mathfrak{F}}_{a,\varepsilon}$ and $[\mathbf{I} \le a]^{\mathfrak{F}}$ tends to zero. Consequently this last set is also relatively compact, showing the claim above and then the total boundedness of $(\mathfrak{F}, d_2)$.

From the assumed LDP and the Borel–Cantelli lemma, we have

$$(L_n - \mu)^{\mathfrak{F}} \to 0 \quad \text{a.s. in } \ell_\infty(\mathfrak{F}),$$

which yields (iii).

(iii) ⇒ (ii). By Lemma 2, (iii) implies

$$\frac{\overline{H}(n)}{n} = \mathbb{E}\| L_n(f) - \mu(f) \|_{\mathfrak{F}} \to 0.$$

By the symmetrization inequality (see Lemma 6.3 in [5], page 152),

$$\frac{H(n)}{2} \le \tilde{H}(n) \le 2H(n). \tag{2.3}$$

Now (ii) follows from the obvious fact that $H(n, \eta) \le 2H(n)$.

It remains to prove the key part, (ii) ⇒ (i). If $\mathfrak{F}$ is finite, this is a direct consequence of Sanov's theorem as noted at the beginning of this paper. In general we have only to establish that

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\| L_n(\cdot) - \mu(\cdot) \|_{\mathfrak{F}_\eta} \ge \delta\right) = -\infty, \forall\, \delta > 0. \tag{2.4}$$

In fact, assume (2.4). For each $\eta > 0$, we can take a *finite* η-net $\mathfrak{F}^\eta$ of $\mathfrak{F}$ [i.e., $\mathfrak{F}^\eta \subseteq \mathfrak{F}$ and $\forall\, f \in \mathfrak{F}$, $\exists\, g \in \mathfrak{F}^\eta$ so that $d_2(f, g) \le \eta$], and we can apply the comparison technique to show the LDP. For example let us show the upper bound in the LDP of this theorem under (2.4).

For this purpose note first that $[h_{\mathfrak{F}} \le a]$, $a \ge 0$, are compact in $\ell_\infty(\mathfrak{F})$ by the total boundedness of $(\mathfrak{F}, d_2)$ and the Arzelà–Ascoli theorem. Next let $\mathbf{C}$ be an arbitrary closed set in $\ell_\infty(\mathfrak{F})$ and set $b := \inf_{F \in \mathbf{C}} h_{\mathfrak{F}}(F)$. We need to show

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\big(L_n^{\mathfrak{F}} \in \mathbf{C}\big) \le -b.$$

If $b = 0$, there is nothing to be shown. Assume then $b > 0$.

For any $0 < a < b$, let $\mathbf{K}$ denote the level set $[h_{\mathfrak{F}} \le a]$, let h^η denote the rate function corresponding to $\mathfrak{F}^\eta$ defined by (1.7) and let $p_\eta(F)$ denote the restriction of F to $\mathfrak{F}^\eta$. Set $\mathbf{K}^\eta = [h^\eta \le a]$. We obviously have $\mathbf{K}^\eta = p_\eta(\mathbf{K})$. Thus for every $\delta > 0$, there is $\eta > 0$ such that for all $F \in \ell_\infty(\mathfrak{F})$, if

$$p_\eta(F) \in \mathbf{K}^\eta(\delta)$$

and

$$\sup\{|F(f) - F(g)|;\ f, g \in \mathfrak{F} \text{ and } d_2(f,g) < \eta\} < \delta \quad \text{then} \quad F \in \mathbf{K}(3\delta),$$

where $\mathbf{K}(\varepsilon)$ [resp. $\mathbf{K}^\eta(\varepsilon)$] is the ε-neighborhood of $\mathbf{K}$ in $\ell_\infty(\mathfrak{F})$ [resp., of $\mathbf{K}^\eta$ in $\ell_\infty(\mathfrak{F}^\eta)$]. Now take $\delta > 0$ so that $\mathbf{C} \cap \mathbf{K}(3\delta) = \varnothing$ (possible since $\mathbf{K}$ is compact). Then for all η small enough, we have, by the claim above and the finite case,

$$\begin{aligned}\mathbb{P}\big(L_n^{\mathfrak{F}} \in \mathbf{C}\big) &\le \mathbb{P}\big(L_n^{\mathfrak{F}} \notin \mathbf{K}(3\delta)\big) \le \mathbb{P}\big(L_n^{\mathfrak{F}^\eta} \notin \mathbf{K}^\eta(\delta)\big) + \mathbb{P}\big(\|L_n(f)\|_{\mathfrak{F}_\eta} \ge \delta\big) \\ &\le \exp[-na + o(n)] + \exp[-L(\eta) \cdot n],\end{aligned}$$

where $L(\eta) \to +\infty$ as $\eta \to 0$ by (2.4). As a $(< b)$ is arbitrary, we get the desired upper bound. The lower bound based on (2.4) can be established in a similar way (easier!).

We turn now to prove the key estimation (2.4) by applying Lemma 1 of Talagrand. For this purpose set $\mathfrak{F}'_\eta := \{(f+1)/2;\ f \in \mathfrak{F}_\eta\}$, which satisfies the assumption of Lemma 1. Setting $H'(n,\eta) = \mathbb{E}\|\sum_{k=1}^n \varepsilon_k f(X_k)\|_{\mathfrak{F}'_\eta}$, we have

$$|2H'(n,\eta) - H(n,\eta)| \le \mathbb{E}\left|\sum_{k=1}^n \varepsilon_k\right| \le \sqrt{n}. \tag{2.5}$$

Consequently $n\delta/2 \ge L \cdot H'(n,\eta)$ if η is sufficiently small and n large enough, by our assumption (1.8) in (ii), where L is the universal constant in Lemma 1.

Applying Lemma 1 to $\mathfrak{F}'_\eta$, we get

$$\begin{aligned}\mathbb{P}\big(\|L_n(\cdot) - \mu(\cdot)\|_{\mathfrak{F}_\eta} \ge \delta\big) &= \mathbb{P}\big(\|L_n(\cdot) - \mu(\cdot)\|_{\mathfrak{F}'_\eta} \ge \delta/2\big) \\ &\le \exp(-\phi_{L,S}(n\delta/2)),\end{aligned}$$

where $S = n\eta^2/2 + H'(n,\eta)$. We see clearly that $\phi_{L,S}(n\delta/2)$ is given by expression (2.2)$'$ if η is small enough; hence

$$\lim_{\eta \to 0} \lim_{n \to \infty} n^{-1} \phi_{L,S}(n\delta/2) \to +\infty.$$

Consequently (2.4) follows. □

PROOF OF THEOREM 2. (i) $\Rightarrow$ (iii). The total boundedness follows from the inf-compactness of $\mathbf{I}_{\mathfrak{F}}$ on $\ell_\infty(\mathfrak{F})$. The LDP implies convergence a.s. in (iii).

(iii) $\Rightarrow$ (ii). A direct consequence of Lemma 2 as in the proof of Theorem 1.

(ii) $\Rightarrow$ (i). As in the proof of Theorem 1, we have only to establish

$$(2.6)\qquad \lim_{\eta\to 0}\limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log\mathbb{P}\left(\frac{1}{\sqrt{n}\,\lambda(n)}\left\|\sum_{k=1}^n f(X_k)-n\mu(f)\right\|_{\mathfrak{F}_\eta}\ge\delta\right)=-\infty,$$

for all $\delta>0$. As in the proof of Theorem 1, applying Lemma 1 to $\mathfrak{F}'_\eta$, we get

$$\mathbb{P}\left(\frac{1}{\sqrt{n}\,\lambda(n)}\left\|\sum_{k=1}^n f(X_k)-n\mu(f)\right\|_{\mathfrak{F}_\eta}\ge\delta\right)$$

$$\le\mathbb{P}\left(\frac{1}{\sqrt{n}\,\lambda(n)}\left\|\sum_{k=1}^n f'(X_k)-n\mu(f')\right\|_{\mathfrak{F}'_\eta}\ge\delta/2\right)$$

$$\le\exp\left(-\phi_{L,S}\left(\sqrt{n}\,\lambda(n)\delta/2\right)\right),$$

where $S=n\eta^2+H'(n,\eta)$. It can be seen that $\sqrt{n}\,\lambda(n)\delta/2\ge L\cdot H'(n,\eta)$ for η small enough and n large enough, by our condition (1.10) in (ii) and the fact (2.5). Since $\phi_{L,S}(\sqrt{n}\,\lambda(n)\delta/2)$ is given now by expression (2.2), we have

$$\lim_{\eta\to 0}\lim_{n\to\infty}\lambda^{-2}(n)\phi_{L,S}(n\delta/2)\to+\infty.$$

Consequently (2.6) follows. □

Since Corollary 3 is known, we are content simply to point out that the MD estimation obtained in Theorem 2 gives us the crucial estimation required by the classical proof of LIL (i.e., the Borel–Cantelli lemma *plus* blocking technique; see the proof of Theorem 8.2 in [6] pages 198–203).

3. Extension to the unbounded case. Let $\mathfrak{F}$ be a class of real measurable functions on E so that $\|f(x)\|<+\infty$ for every $x\in E$. Since the isoperimetric inequality of Talagrand in Lemma 1 depends strictly on assumption (H1) (but it is very sharp, perhaps *too* sharp for our purpose), in the unbounded case we should find other tools.

3.1. *Large deviation estimations.*

THEOREM 4. *Suppose that* $\mathfrak{F}$ *is a class of functions in* $L^2(E,\mu)$ *such that*

$$(3.1)\qquad h(\lambda):=\int_E\exp\lambda\|f(x)\|_{\mathfrak{F}}\,d\mu(x)<+\infty\quad\textit{for all }\lambda>0.$$

Then the equivalence of (i), (ii) *and* (iii) *in Theorem* 1 *is valid.*

PROOF. The implications (i) ⇒ (iii) ⇒ (ii) can be done in the same way as in the proof of Theorem 1 in the bounded cases. Notice also that (iii) follows easily from (ii). Thus it remains to prove (iii) ⇒ (i).

For this, we shall apply the comparison technique. For every $N > 0$, consider the class $\mathfrak{F}_N := \{f_N := (f \vee (-N)) \wedge N;\ f \in \mathfrak{F}\}$. As $d_2(f_N, g_N) \le d_2(f, g)$, this new class is again totally bounded in $L^2(E, \mu)$. We claim now that $\mathfrak{F}_N$ satisfies also

$$(3.2)\qquad \frac{1}{n}\overline{H}_N(n) := \frac{1}{n}\mathbb{E}\left\| \sum_{k=1}^{n} g(X_k) - n\mu(g) \right\|_{\mathfrak{F}_N} \to 0.$$

In fact, by Lemma 2 and the symmetrization inequality (2.3), the convergence in probability in (iii) is equivalent to

$$(3.3)\qquad \frac{H(n)}{n} \to 0.$$

Next, by the comparison theorem ([6], Theorem 4.12, page 112),

$$H_N(n) := \mathbb{E}\left\| \sum_{k=1}^{n} \varepsilon_k \cdot f_N(X_k) \right\|_{\mathfrak{F}} \le 2H(n),$$

where the desired convergence (3.2) follows by (3.3) and the inequality (2.3).

Consequently Theorem 1 can be applied to $\mathfrak{F}_N$ for each $N > 0$. By the comparison technique, as in the proof of Theorem 1, to prove the LDP in (i), it will suffice to establish

$$(3.4)\qquad \lim_{N\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(\| L_n(f - f_N) - \mu(f - f_N) \|_{\mathfrak{F}} \ge \delta \big) = -\infty, \qquad \forall\, \delta > 0.$$

To this end we have, by the Markov inequality,

$$\begin{aligned} &\mathbb{P}\big(\| L_n(f - f_N) - \mu(f - f_N) \|_{\mathfrak{F}} \ge \delta \big) \\ &\quad \le \exp(-n\,\delta\lambda) \cdot \mathbb{E}\exp\big(n\lambda \| L_n(f - f_N) - \mu(f - f_N) \|_{\mathfrak{F}} \big) \\ &\quad \le \exp(-n\,\delta\lambda) \cdot \left(\int_E \exp \lambda \| f - f_N - \mu(f - f_N) \|_{\mathfrak{F}}\, d\mu \right)^n, \end{aligned}$$

which implies immediately (3.4) by the dominated convergence theorem and our integrability condition (3.1). □

REMARKS. If $\mathfrak{F}$ is finite, the integrability condition (3.1) can be relaxed to

$$(3.1)'\qquad \int_E \exp \lambda \| f(x) \|_{\mathfrak{F}}\, d\mu(x) < +\infty \quad \text{for some } \lambda > 0,$$

(Condition (3.1)′ is also necessary to the LDP in (i). See [8], Section 4.1, for detailed comments). However, in the infinite case, (3.1) cannot be weakened in general, in the sense that one can find examples which do satisfy (3.1)′ but not the LDP. This was indicated by de Acosta [1] and Gao [3] (see [8], Section 4.1, for the presentation of his result). Notice finally that in the setting of separable Banach spaces, the integrability condition (3.1) is the same as that of the well-known Cramér theorem due to Donsker–Varadhan.

3.2. *MD estimations*. For LD estimation above we use the truncation method where the truncation error is controlled easily by (3.1). For MD estimation, Ledoux developed the truncation method in a finer way, based on an isoperimetric inequality (i.e., Theorem 6.18 in [6]). We present his crucial estimation (see (8) in [4]) in the nonseparable setting.

LEMMA 3 (Ledoux [4]). *Let $(\xi_n)_{n\geq 1}$ be a sequence of i.i.d. r.v.'s defined on a complete probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, with values in a Banach space $(B, \|\cdot\|)$ (not necessarily separable), satisfying the following: There is a countable subset D of the unit ball of B′ such that $\|x\| = \sup_{f\in D}|f(x)|$ for all x in B. Assume that $\{f(\xi_1)^2;\ f\in D\}$ is uniformly integrable. Let $(\lambda(n))_{n\geq 1}$ be an increasing sequence of positive numbers satisfying* (1.2) *and the auxiliary condition*

$$\lambda(nk) \leq Ak^{-\delta-1/2}\lambda(n) \quad \text{for some } \delta > 0. \tag{3.5}$$

[*in other words, $(\lambda(n))$ cannot be too near $\sqrt{n}$*.]

If the following two conditions are fulfilled:

(i) $(\lambda(n)\sqrt{n})^{-1}S_n \to 0$ *in probability*;
(ii) *there is $M > 0$ such that, for all $u > 0$*,

$$\limsup_{n\to\infty} \lambda^{-2}(n)\log(n\mathbb{P}(\|\xi\| > u)) \leq -u^2/M; \tag{3.6}$$

then there exists some constant C_0 depending only on M and A, δ in (3.5) (*and not on ξ*) *such that*

$$\limsup_{n\to\infty} \lambda^{-2}(n)\log\mathbb{P}\big(\|S_n\| > \lambda(n)\sqrt{n}\big) \leq -\frac{1}{C_0\sigma^2}, \tag{3.7}$$

where $\sigma^2 = \sigma^2(\xi) = \sup_{f\in D}\mathbb{E}f(\xi_1)^2$.

The estimation (3.7) is essential (and the difficult part of [4]) for the characterization of the moderate deviations in a separable Banach space in [4]. The proof of (3.7) given by Ledoux works in the nonseparable setting (the separability in [4] serves only for the existence of a finite-dimensional approximation).

Having this crucial estimation, we can easily prove the following theorem.

THEOREM 5. *Suppose that $\mathfrak{F}$ is a class of functions in $L^2(E,\mu)$, and $\lambda(n)$ is as in Lemma 3. Then $\mathbb{P}([\sqrt{n}/\lambda(n)](L_n-\mu)^{\mathfrak{F}} \in \cdot)$ satisfies LDP on $\ell_\infty(\mathfrak{F})$ with speed $\lambda^{-2}(n)$ and with the rate function $\mathbf{I}_{\mathfrak{F}}$ given by* (1.9), *if and only if the* *following three conditions are fulfilled*:

(i) $(\mathfrak{F}, d_2)$ *is totally bounded*;
(ii) $[\sqrt{n}/\lambda(n)](L_n - \mu)^{\mathfrak{F}} \to 0$ *in probability in* $\ell_\infty(\mathfrak{F})$;
(iii) *there exists* $M > 0$ *such that, for all* $u > 0$,

$$\limsup_{n\to\infty} \lambda^{-2}(n)\log\big(n\mu(\|f\|_{\mathfrak{F}} > u)\big) \le -u^2/M.$$

PROOF (Following closely Ledoux [4]). The necessity of (i) and (ii) can be shown in the same way as in the bounded case, and that of (iii) can be established in the same way as in [4].

For the sufficiency part, notice first that if $\mathfrak{F}$ is finite, this is done in [4]. For $\mathfrak{F}$ infinite, we have only to establish (2.6) by the discussion in Section 2, and (2.6) is a direct consequence of (3.7) if one identifies $B = \ell_\infty(\mathfrak{F}_\eta)$ and notes that σ^2 in (3.7) is smaller than η. □

REMARKS. For some usual sequences of $\lambda(n)$ such as $\lambda(n) = n^{1/p}$, with $p > 2$, or $\lambda(n) = \sqrt{2\log\log n}$, Ledoux [4] translated condition (3.7) into the integrability condition of $\|f\|_{\mathfrak{F}}$.

CONCLUDING REMARKS.

(i) For the MD estimation, Lemma 3 of Ledoux is of course applicable to the bounded case, but with the auxiliary condition (3.5) about $\lambda(n)$. It seems that the isoperimetric inequality used by Ledoux is difficult to adapt to the LD estimation. Anyway this paper speaks against the power of the isoperimetric ideas.

(ii) In this paper we reduce the LD or MD estimations to the corresponding convergence in probability. In the theory of empirical processes one can establish this last property by means of metric entropy of $(\mathfrak{F}, d_2)$ [or of $\mathscr{F}$ equipped with another metric] or by means of majorizing measure (see [2] and [6]). For example, one can obtain easily interesting sufficient conditions of the types of Theorems 19 and 20 of [5] from Theorem 18 of [5].

(iii) Of course Theorems 4 and 5 can be applied to the partial sums of a sequence of i.i.d. r.v.'s $(\xi_n)_{n\ge 1}$ with values in a Banach space B (not necessarily separable) satisfying $\|x\| = \sup_{f\in D}|f(x)|$ for some fixed countable subset D of the unit ball of B'. In fact, it suffices to put $E = B$ and $\mathfrak{F} = D$, and notice that $\ell_\infty(\mathfrak{F})$ is isometric to $(B, \|\cdot\|)$.

Acknowledgments. I am grateful to Professor Michel Ledoux for useful discussions and especially for his communication of the recent works of Talagrand [7]. I am very obliged to a referee, who pointed out the possibility of the extension to the unbounded situation and whose comments improved the first version in many aspects.

REFERENCES

[1] DE ACOSTA, A. (1985). Upper bounds for large deviations of dependent random vectors. *Z. Wahrsch. Verw. Gebiete* **69** 551–566.

[2] DUDLEY, R. (1984). A course on empirical processes. *Ecole d'Eté de Probabilités de Saint-Flour. Lecture Notes in Math.* **1097** 2–142. Springer, Berlin.

[3] GAO, F. Q. (1991). Private communication.

[4] LEDOUX, M. (1992). Sur les déviations modérées des sommes de variables aléatoires vectoriels indepenandes de la même loi. *Ann. Inst. H. Poincaré* **28** 267–280.

[5] LEDOUX, M. and TALAGRAND, M. (1989). Comparison theorems, random geometry and some limit theorems for empirical processes. *Ann. Probab.* **17** 596–631.

[6] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces—Isoperimetry and Processes*. Springer, New York.

[7] TALAGRAND, M. (1992). Sharper bounds for empirical processes. Unpublished manuscript.

[8] WU, L. M. (1991). Some general methods of large deviations and applications. Unpublished manuscript.

[9] WU, L. M. (1992). Moderate deviations of dependent random variables related to CLT and LIL. Prépublication 118 du Laboratoire de Probabilités, Univ. Paris VI.

[10] YURICH, J. (1990). The law of the iterated logarithm for empirical processes. In *Probability in Banach Spaces* **6** 265–282. Birkhäuser, Boston.

DEPARTMENT OF MATHEMATICS
WUHAN UNIVERSITY
430072, HUBEI
CHINA