PROBABILITIES OF MODERATE DEVIATIONS
IN A BANACH SPACE

CHANDRAKANT M. DEO AND GUTTI JOGESH BABU

ABSTRACT. We obtain probabilities of moderate deviations for i.i.d. sequences taking values in a separable Banach space under precise necessary and sufficient conditions. The results obtained are new even on the real line.

1. Introduction. Let \( \{X_n\} \) be an i.i.d. sequence of real-valued random variables, \( EX_1 = 0, EX_1^2 = 1 \). Let \( S_n = \sum_1^n X_i \) and let \( t_n \) be a sequence of positive numbers with \( t_n \to \infty \). A theorem on large deviations is the statement that

\[
P\left( \frac{S_n}{\sqrt{n}} > t_n \right) \sim 1 - \Phi(t_n) = \Phi(-t_n) \sim P\left( \frac{S_n}{\sqrt{n}} < -t_n \right)
\]

where \( a_n \sim b_n \) denotes \( \lim_{n \to \infty} (a_n / b_n) = 1 \) and \( \Phi \) is the c.d.f. of the standard normal distribution. Cramér (1938) proved that (1) holds if \( X_1 \) has a moment generating function in a neighborhood of 0 and \( \lim_{n \to \infty} n^{-1/6} t_n = 0 \). He proved further if \( t_n = n^\alpha, \frac{1}{6} < \alpha < \frac{1}{2} \), then (1) holds provided one assumes in addition that a sufficient number of cumulants of \( X_1 \) agree with those of the standard normal. Rubin and Sethuraman (1965) considered the case \( t_n = \sqrt{c \log n}, c > 0 \). In this case, called "moderate deviations" by them, they proved that (1) holds without existence of a moment generating function; it suffices that \( E|X_1|^{2+c+\delta} < \infty \) for some \( \delta > 0 \). Also in this case they proved that \( E|X_1|^p < \infty, \forall p < 2 + c \), is a necessary condition for (1). Recently Slastnikov (1978) has shown that, again with \( t_n = \sqrt{c \log n} \), (1) holds if and only if \( X_1 \) satisfies the tail-condition

\[
\lim_{t \to \infty} t^{2+c}(\log t)^{-(1+c)/2} P(|X_1| > t) = 0.
\]

In this paper we consider probabilities of moderate deviations for random variables taking values in a Banach space and give precise necessary and sufficient conditions which are new even in the one-dimensional case.

2. Main results. For \( B \)-valued r.v.'s (1) is too fine to be generalized. One can only hope to generalize the "logarithmic form" of (1). We begin with a one-dimensional result. Notice, however, that Theorem 2 as specialized to the case \( B = R \) implies a result stronger than Theorem 1, viz. that (3) implies strengthened forms of (4a), (4b) where lim sup is replaced by limit and the inequality \( \langle -1/2\sigma^2 \rangle \) is replaced by equality. Also to put the condition (3) in perspective it is easily verified that

\[
E|X_1|^{2+c} < \infty \Rightarrow (2) \Rightarrow (3) \Rightarrow E|X_1|^p < \infty, \quad \forall p < 2 + c;
\]

with each implication being strict.
THEOREM 1. Let \( \{X_n\} \) be an i.i.d. sequence of real-valued r.v.'s with \( \text{EX}_1 = 0 \), \( \text{EX}_1^2 = \sigma^2 > 0 \). Let \( c > 0 \). If

\[
\lim \sup_{t \to \infty} \frac{\log P[|X_1| > t]}{\log t} < -c - 2
\]

holds then for any sequence \( a_n > 0 \) with \( a_n / \sqrt{n} \to \infty \) and \( a_n \leq \sqrt{cn \log n} \) we have

\[
\lim \sup_{n \to \infty} \frac{n}{a_n^2} \log P[S_n > a_n] \leq -\frac{1}{2\sigma^2},
\]

and

\[
\lim \sup_{n \to \infty} \frac{n}{a_n^2} \log P[S_n < -a_n] \leq -\frac{1}{2\sigma^2}.
\]

Conversely if (4a), (4b) hold with \( a_n = \sigma \sqrt{c \log n} \) then (3) holds.

PROOF. We will use both the big \( O \) and the Vinogradov symbol \( \ll \) according to convenience. Take \( 0 < t_n \ll \sqrt{c \log n} \) and let, without loss of generality, \( \sigma = 1 \). Write \( Y_{i,n} = X_i I[|X_i| < r \sqrt{n} t_n] \) where \( I[\cdot] \) is the indicator function and \( r \) is a positive number to be selected later. Let \( Z_n = \sum_{i=1}^n Y_{i,n} \). Clearly

\[
P[S_n > t_n \sqrt{n}] \leq P[Z_n > t_n \sqrt{n}] + nP[|X_1| > rt_n \sqrt{n}].
\]

For any \( s > 0 \),

\[
P[Z_n > t_n \sqrt{n}] \leq \exp(-st_n \sqrt{n}) \cdot \left\{ \mathbb{E}[\exp(sY_{1,n})]\right\}^n
\]

and thus, taking \( s = t_n / \sqrt{n} \), we have

\[
P[Z_n > t_n \sqrt{n}] \leq \exp(-t_n^2) \cdot \left\{ \mathbb{E}[\exp(t_n^{-1/2}Y_{1,n})]\right\}^n.
\]

Now for all real \( x \), as \( |x| \to \infty \), \( \exp(x) = 1 + x + x^2/2 + O(|x|^3 \exp|x|) \). Hence

\[
\mathbb{E}[\exp(t_n^{-1/2}Y_{1,n})] \leq 1 + t_n^{-1/2} |Y_{1,n}| + (2n)^{-1/2} EY_{1,n}^2 + O(n^{-3/2} t_n^{-3} \exp(t_n^2 r)).
\]

Now suppose (3) holds. Then \( E|X_1|^{2+c'} < \infty \) if \( 0 < c' < c \). Hence, using the fact that \( \text{EX}_1 = 0 \), we have

\[
|\text{E}Y_{1,n}| = |\text{E}\left\{X_1 I[|X_1| > rt_n \sqrt{n}]\right\}| \leq (rt_n \sqrt{n})^{-c'-1} E|X_1|^{2+c'} \ll n^{-1/2-\delta} \quad \text{for some} \ \delta > 0.
\]

Also \( EY_{1,n}^2 \ll EX_1^2 = 1 \). Now to estimate \( E|Y_{1,n}|^3 \) we have \( E|Y_{1,n}|^3 < \infty \) if \( c > 1 \). If \( 0 < c < 1 \), then

\[
E|Y_{1,n}|^3 = E\left[|Y_{1,n}|^{2+c'/2}|Y_{1,n}|^{1-c'/2}\right] \leq (rt_n \sqrt{n})^{1-c'/2} E|Y_1|^{2+c'/2} \ll (t_n \sqrt{n})^{1-c'/2}.
\]

Using this estimate of \( E|Y_{1,n}|^3 \) and, taking \( r = 1/8 \) in case \( c < 1 \) and \( r = (4c)^{-1} \) if \( c > 1 \), one easily gets

\[
t_n^3 n^{-3/2} E|Y_{1,n}|^3 \exp(t_n^2 r) \ll n^{-1-\delta} \quad \text{for some} \ \delta > 0.
\]
Thus from (7),
\[ E \left\{ \exp(t_n n^{-1/2} Y_{1,n}) \right\} \leq 1 + t_n^2 / 2n + O(n^{-1-\delta}) \]
\[ \leq \exp\left(\frac{t_n^2}{2n} + O(n^{-\delta})\right), \]
and hence, \( [E \{ \exp(t_n n^{-1/2} Y_{1,n}) \}]^p \leq \exp\left(\frac{t_n^2}{2} + O(n^{-\delta})\right). \) Thus, by (6),
\[ P\left[ Z_n > t_n \sqrt{n} \right] \leq \exp\left( -\frac{1}{2} t_n^2 + O(n^{-\delta}) \right) \leq \exp\left( -\frac{1}{2} t_n^2 \right). \]

Now let \( \varepsilon > 0. \) By condition (3), \( \exists y_0 \) such that, for all \( y > y_0, \) \( \log P[|X_1| > y] < (\varepsilon - c - 2) \log y. \) Hence, eventually,
\[ t_n^{-2} \log(n P[|X_1| > r t_n \sqrt{n}]) \]
\[ \leq t_n^{-2} \left\{ \log n + (\varepsilon - c - 2) \log r + (\varepsilon - c - 2) \log(t_n \sqrt{n}) \right\} \]
\[ \leq \frac{1}{2} \left( \frac{\varepsilon}{c} - 1 \right) \frac{c \log n}{t_n^2} + \frac{(\varepsilon - c - 2) \log r}{t_n^2} + O\left( \frac{\log t_n}{t_n^2} \right). \]

Therefore, since \( \varepsilon \) is arbitrary, we have
\[ \lim_{n \to \infty} \sup_n t_n^{-2} \log(n P[|X_1| > r t_n \sqrt{n}]) \leq -\frac{1}{2}. \]

Taking \( a_n = t_n \sqrt{n} \) it is now a simple verification that (5), (8) and (9) yield (4a). Proof of (4b) is similar.

It remains to show that (4a), (4b) imply (3). By inequality (5.10) on p. 149 of Feller (1971) we have
\[ n P\left[ |X_1| > \sqrt{c n \log n} \right] \leq P\left[ |S_n| > \sqrt{c n \log n} \right]. \]

Also (4a), (4b) give
\[ \lim_{n \to \infty} \sup_n (c \log n)^{-1} \log P\left[ |S_n| > \sqrt{c n \log n} \right] \leq -\frac{1}{2}. \]

From (10), (11) we get
\[ \lim_{n \to \infty} \sup_n (\log n)^{-1} \log P\left[ |X_1| > \sqrt{c n \log n} \right] \leq -\frac{1}{2}(c + 2). \]

It is easy to see that (12) implies (3). This concludes the proof of Theorem 1. \( \square \)

Using the elegant Theorem 2.1 of Borovkov and Mogul'skii (1978) we now consider the case of random variables taking values in a Banach space. Let \( B \) be a separable Banach space and \( \{ X_n \} \) an i.i.d. sequence of \( B \)-valued random variables. We assume that the central limit theorem holds for \( \{ X_n \} \), i.e. \( \exists \) a zero-mean, \( B \)-valued, Gaussian random variable \( \eta \) such that \( S_n / \sqrt{n} \) converges to \( \eta \) in distribution. It is well known that a moment condition alone (on \( X_1 \)) is not sufficient to ensure the central limit theorem. Assume for simplicity that the support of \( \eta \) is all of \( B \) as equivalently \( \sigma^2 \eta = E[\theta(X_1)^2] > 0 \) for all nonzero \( \theta \in B^* \). Let \( \Lambda \) be the "Cramér transform" of \( \eta \), i.e. for \( x \in B \),
\[ \Lambda(x) = \inf \{ -\theta(x) + \log E[\exp(\theta(\eta))]: \theta \in B^* \} \]
\[ = \inf \{ -\theta(x) + \frac{1}{2} \sigma^2_\theta: \theta \in B^* \}. \]
Furthermore, slightly misusing the notation, we write \( \Lambda(U) = \sup_{x \in U} \Lambda(x) \), \( U \subset B \). For \( c > 0 \), let \( \mathcal{S}_c = \{ U : U \text{ open, convex and } \Lambda(U) \geq -c/2 \} \). It is well known (cf. p. 614 of Bahadur and Zabell (1979)) that
\[
\Lambda(x) = -\frac{1}{2} \| x \|^2_H, \quad x \in H, \\
= -\infty, \quad \text{otherwise},
\]
where \( H \subset B \) is the RKHS associated with \( \eta \) and \( \| \cdot \|_H \) is the RKHS norm. Since closed balls in \( H \) are compact subsets of \( B \) it is easy to see that \( \mathcal{S}_c \) can also be described as the collection of all those open, convex subsets of \( B \) whose closures have nonempty intersection with the closed ball in \( H \) centered at the origin and of radius \( \sqrt{c} \). We now introduce the condition
\[
(13) \quad \lim_{t \to \infty} \sup_{t} \frac{\log P[|\theta(X_i)| > t]}{\log t} < -c - 2, \quad \forall \theta \in B^*.
\]
This condition is implied by e.g. \( E \| X_1 \|^{2+c} < \infty \). The following theorem is the main result of this note.

**Theorem 2.** Suppose (13) holds. Then for each sequence \( \{a_n\} \) of positive numbers with \( a_n/\sqrt{n} \to \infty \) and \( a_n < \sqrt{n \log n} \) and, for each \( U \in \mathcal{S}_c \), we have
\[
(14) \quad (n/a_n^2) \log P[S_n/a_n \in U] \to \Lambda(U).
\]
Conversely if (14) holds with \( a_n = \sqrt{n \log n} \) and for all \( U \in \mathcal{S}_c \) then (13) is true.

**Proof.** For sufficiency we closely follow the proof of Theorem 2.1 in Borovkov and Mogul'skii (1978). Write \( l_n(U) = (n/a_n^2) \log P[S_n/a_n \in U] \). Then, by Lemma 2.2 of [2], we have for each open ball \( U \) in \( B \)
\[
(15) \quad \liminf_{n \to \infty} l_n(U) \geq \Lambda(U).
\]
Since open balls are a basis for the topology of \( B \) it is clear that (15) holds for all open sets in \( B \) because \( \Lambda(U) \) is simply the supremum of \( \Lambda(x) \) over \( x \in U \).

It remains to show that for \( U \in \mathcal{S}_c \)
\[
(16) \quad \limsup_{n \to \infty} l_n(U) < \Lambda(U).
\]
First take \( U \in \mathcal{S}_c \) to be a half-space, i.e. for some \( \theta \in B^* \) and \( d \) real, \( U = \{ x : \theta(x) > d \} \). Suppose \( d > 0 \). By the relation (2.7) in [2] we have \( \Lambda(U) = -(2\sigma_\theta)^{-1}d^2 \) and hence \( U \in \mathcal{S}_c \) implies \( d < \sigma_\theta \sqrt{c} \). Now, in this case,
\[
l_n(U) = \frac{n}{a_n^2} \log P \left\{ \frac{1}{a_n} \sum_{i=1}^{n} \theta(x_i) > d \right\}.
\]
Hence (13) allows us to apply (the one-dimensional) Theorem 1 to conclude
\[
\limsup_{n} l_n(U) \leq -(2\sigma_\theta)^{-1}d^2 = \Lambda(U). \quad \text{if } d < 0,
\]
(16) holds trivially. To go from half-spaces to other sets in \( \mathcal{S}_c \) we use again the arguments in [2]. Take \( U \in \mathcal{S}_c \) and \( \varepsilon > 0 \). Let \( B = \{ x : \Lambda(x) > \Lambda(U) + \varepsilon \} \). Then \( B \) is a closed, convex set disjoint from \( U \). Hence \( \exists \) a half-space \( U' = \{ x : \theta(x) > d \} \) such that \( U \subset U' \) and \( U' \cap B = \emptyset \). So, \( \Lambda(U) < \Lambda(U') < \Lambda(U) + \varepsilon \). Thus \( U' \in \mathcal{S}_c \) and by the foregoing
\[
\limsup_{n} l_n(U) \leq \limsup_{n} l_n(U') < \Lambda(U') < \Lambda(U) + \varepsilon.
\]
Letting \( \varepsilon \to 0 \) we get (16). This concludes the proof of Theorem 2. \( \square \)
It is straightforward to extend Theorem 2 to general locally convex vector spaces under conditions imposed by either Bahadur and Zabell (1979) or Borovkov and Mogul’skii (1978). Perhaps one should make some comments concerning [2]. The argument in the first paragraph on p. 700 is clearly false. Also the existence of the sequence \( t_n \) asserted in the proof of Lemma 2.2 appears to be in doubt. However, given \( \varepsilon > 0 \), it is easy to see that one can find a sequence \( t_n \) (depending upon \( \varepsilon \)) such that the equation (2.3) there holds with the right-hand side decreased by \( \varepsilon \). One now completes the proof with this new right-hand side and finally dispenses with \( \varepsilon \) since it is arbitrary.

3. **Applications.** Probabilities of moderate deviations are used in proving a number of limit theorems. As applications of Theorem 2 we give two theorems here which extend to \( B \)-valued sequences some results of T. L. Lai (1974) and (1976). Theorem 3 deals with the so-called “delayed averages” and Theorem 4 with the “\( r \)-quick” limit points. Given \( r > 0 \), a sequence \( \{\xi_n\} \) of random variables taking values in a metric space \((M, d)\) is said to be \( r \)-quickly relatively compact if \( \forall \varepsilon > 0, \exists \text{ a finite union } U \text{ of } \varepsilon \text{-balls such that} \)

\[
E\left(\sup\{n : \xi_n \notin U\}\right)^r < \infty.
\]

Also \( x \in M \) is called an \( r \)-quick limit point of \( \{\xi_n\} \) if, for each neighborhood \( U \) of \( x \), \( E\{\sup\{n : \xi_n \in U\}\}^r = \infty \).

Let now \( B \) be a separable Banach space, \( \{X_n\} \) an i.i.d. \( B \)-valued sequence of random variables, \( S_n = \sum_1^n X_i \). Assume again that \( S_n/\sqrt{n} \) converges in distribution to a zero-mean, Gaussian random variable \( \eta \). Let \( K \) be the unit ball in the RKHS of \( \eta \).

**Theorem 3.** Let \( 0 < \alpha < 1 \) and let \( E\|X_1\|^{2/\alpha} < \infty \). Then, with probability one, the sequence \( \{(2n^{\alpha}(1-\alpha)\log n)^{-1/2}(S_n + \alpha\log n) - S_n\} \) is relatively compact in \( B \) and has \( K \) as the set of its limit points.

**Theorem 4.** Let \( r > 0 \) and \( E\|X_1\|^{2(r+1)} < \infty \). Then the sequence \( \{(2n \log n)^{-1/2}S_n\} \) is \( r \)-quickly relatively compact in \( B \) and the set of its \( r \)-quick limit points coincides with \( K \).

Given the probability estimates in Theorem 2, the proofs of these two theorems are routine, i.e. they need only standard techniques such as subsequences and maximal inequality. We will omit these proofs.

Notice that empirical processes corresponding to i.i.d. sequences of bounded real-valued random variables can be regarded as normalized partial sums of random variables taking values in a suitable function space such as \( L_2 \). It is well known that the central limit theorem applies in this case. Consequently Theorems 3 and 4 can be applied to draw a variety of conclusions about such empirical processes.

**Acknowledgement.** Thanks are due to the referee for pointing out a number of errors in the presentation.
REFERENCES


DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OTTAWA, OTTAWA, ONTARIO K1N 9B4, CANADA

MATHEMATICS-STATISTICS DIVISION, INDIAN STATISTICAL INSTITUTE, CALCUTTA, INDIA