The Annals of Probability
1984, Vol. 12, No. 2 427–435

ON THE PROBABILITY OF LARGE DEVIATIONS IN BANACH SPACES

BY E. BOLTHAUSEN

Technische Universität Berlin

Probabilities of large deviations for sums of i.i.d. Banach space valued random variables are investigated when the laws of the random variables converge weakly and a uniform exponential integrability condition is satisfied. Furthermore, a discussion of possible improvements of the estimates is given, when the probability is estimated that the sum lies in a convex set.

1. Introduction. Let B be a real separable Banach space, equipped with the Borel-σ-field $\mathscr{B}$ and let $\mathbb{P}$ be the set of probability measures on $(B, \mathscr{B})$. B^* denotes the (topological) dual of B.

If $\mu \in \mathbb{P}$, $\varphi \in B^*$, let $M(\varphi \mid \mu) = \int \exp(\varphi(x))\mu(dx)$ and if $a \in B$, let $h(a \mid \mu) = \sup\{\varphi(a) - \log M(\varphi \mid \mu): \varphi \in B^*\}$. The following result is due to Donsker and Varadhan [6] and Bahadur and Zabell [3]:

THEOREM 1. *If* $\int \exp(t\|x\|)\,\mu(dx) < \infty$ *for all* $t > 0$, *then*

(1.1) *if* $A \subset B$ *is closed,* $\limsup_{n\to\infty}(1/n)\log \mu^{*n}(nA) \le -h(A \mid \mu)$.

(1.2) *If* $A \subset B$ *is open,* $\liminf_{n\to\infty}(1/n)\log \mu^{*n}(nA) \ge -h(A \mid \mu)$,

where μ^{*n} *is the n-fold convolution of* μ *and* $h(A \mid \mu) = \inf\{h(a \mid \mu): a \in A\}$.

We shall prove here the following extension:

THEOREM 2. *Let* $\mu_n, \mu \in \mathbb{P}$, $n \in \mathbb{N}$, *such that* $\{\mu_n\}$ *converges weakly to* μ *and*

(1.3) $$\sup_n \int \exp(t\|x\|)\mu_n(dx) < \infty \quad \text{holds for all} \quad t > 0.$$

Then

(1.4) *if* $A \subset B$ *is closed,* $\limsup_{n\to\infty}(1/n)\log \mu_n^{*n}(nA) \le -h(A \mid \mu)$,

(1.5) *if* $A \subset B$ *is open,* $\liminf_{n\to\infty}(1/n)\log \mu_n^{*n}(nA) \ge -h(A \mid \mu)$.

The special case, where the μ_n are Gaussian, has been treated by Ellis and Rosen [7] and S. Chevet [4]. In this case (1.3) is automatically satisfied. In fact, inspection of Fernique's proof of the existence of exponential moments for Gaussian measures shows that if μ_n, $n \in \mathbb{N}$, are Gaussian and the μ_n converge

Received September 1982; revised September 1983.
AMS 1979 *subject classifications.* 60B12, 60F10.
Key words and phrases. Banach space valued random variables, large deviations.

weakly, then there are constants a, b, $c > 0$, not depending on n, such that

$$\mu_n(\{x\colon \|x\| \geq u\}) \leq a \exp(-bu^2) \quad \text{as} \quad u \geq c.$$

From this (1.3) follows (see [8]).

The proof given here is a rather elementary modification of the Donsker-Varadhan proof. In contrast, the proofs of Ellis/Rosen and Chevet rely on non-trivial properties of Gaussian measures in Banach spaces.

If $B = \mathbb{R}$ and A is an interval, the results which have been obtained are much better than Theorem 1 or 2 (see e.g. Bahadur and Rao [2] or Höglund [9]). Partly, this possibility of improvements depends only on the convexity of A. Although I have only very incomplete results in this direction, it seemed worth pointing out how the convexity of A leads to improvements of (1.1) and (1.2). This is done in Section 4. This has also been investigated by P. Ney [11] in the case $B = \mathbb{R}^n$.

2. The upper estimate. If $\nu, \mu \in \mathbb{P}$ let $k(\nu \mid \mu)$ be the Kullback/Leibler information, i.e. $k(\nu \mid \mu) = \nu(\log(d\nu/d\mu))$ if $\nu << \mu$ and $\nu(|\log(d\nu/d\mu)|) < \infty$ and $k(\nu \mid \mu) = \infty$ else. We write $\mu(f)$ for the expectation of f with respect to μ. Then

$$h(a \mid \mu) = \inf\{k(\nu \mid \mu)\colon \nu(\text{id}) \text{ exists and equals } a\}. \tag{2.1}$$

Here id is the identity mapping $B \to B$ (see [6], Theorem 5.2. (iv)). Although there is in general no $\varphi \in B^*$ with $h(a \mid \mu) = \varphi(a) - \log M(\varphi \mid \mu)$, there is always a $\nu \in \mathbb{P}$ satisfying $\nu(\text{id}) = a$ and $h(a \mid \mu) = k(\nu \mid \mu)$, at least if $h(a \mid \mu) < \infty$. Furthermore, ν is then unique (see Csiszar [5]).

LEMMA 1. *Let μ_n, μ satisfy the condition of the theorem and $a_n \in B$ converge weakly to $a \in B$. Then* $\liminf_{n\to\infty} h(a_n \mid \mu_n) \geq h(a \mid \mu)$.

PROOF. From (1.3) it follows that for any $\varphi \in B^*$

$$\lim_{n\to\infty} M(\varphi \mid \mu_n) = M(\varphi \mid \mu). \tag{2.2}$$

Given $\varepsilon > 0$, there is a $\varphi \in B^*$ with $\varphi(a) - \log M(\varphi \mid \mu) \geq h(a \mid \mu) - \varepsilon$. Therefore, if n is large enough, we have $h(a_n \mid \mu_n) \geq \varphi(a_n) - \log M(\varphi \mid \mu_n) \geq h(a \mid \mu) - 2\varepsilon$. This proves the lemma.

LEMMA 2. *Let $A \subset B$ be closed, then*

$$h(A \mid \mu) \leq \liminf_{n\to\infty} h(A \mid \mu_n).$$

PROOF. We may assume that $\liminf_{n\to\infty} h(A \mid \mu_n) < \infty$. We select a subsequence $\{n_k\}$ with $\lim_{k\to\infty} h(A \mid \mu_{n_k}) = \liminf_{n\to\infty} h(A \mid \mu_n)$. Let $a_k \in A$ satisfy $h(a_k \mid \mu_{n_k}) \leq h(A \mid \mu_{n_k}) + 1/k$ and $\nu_k \in \mathbb{P}$ satisfy $k(\nu_k \mid \mu_{n_k}) = h(a_k \mid \mu_{n_k})$, $\nu_k(\text{id}) = a_k$. From Lemma 5.1 of [6] it follows that the sequence $[\nu_k\}$ is tight and furthermore

$$\limsup_{\rho\uparrow\infty} \sup_k \int_{\|x\|\geq\rho} \|x\| \, \nu_k(dx) = 0.$$

Therefore $\{a_k\}$ is relatively compact. Let $a \in A$ be a limit point of this sequence.

Then by Lemma 1

$$h(A \mid \mu) \le h(a \mid \mu) \le \lim_{k\to\infty} h(a_k \mid \mu_{n_k}) = \lim_{k\to\infty} h(A \mid \mu_{n_k}) = \liminf_{n\to\infty} h(A \mid \mu_n).$$

LEMMA 3. *If $A \subset B$ is open and convex, then*

$$\mu^{*n}(nA) \le \exp(-nh(A \mid \mu)).$$

PROOF. If A is open and convex, then $-h(A \mid \mu) = \lim_{n\to\infty}(1/n)\log \mu^{*n}(nA)$ (see [1], Theorem I 4.8). If A is convex, one has the following subadditivity: $\mu^{*n}(nA)\mu^{*m}(mA) \ge \mu^{*(n+m)}((n+m)A)$. From this, $h(A \mid \mu) = \inf_n(-(1/n)\log \mu^{*n}(nA))$. The lemma follows.

PROOF OF (1.4) IN THE CASE WHERE A IS COMPACT. Take $\varepsilon > 0$ and $A \subset \cup_{j=1}^m U_j$, where U_j are open balls with radius ε and center in A. Then

$$\begin{aligned}
\limsup_{n\to\infty} (1/n) \log \mu_n^{*n}(nA) &\le \limsup_{n\to\infty} (1/n)\log \mu_n^{*n}(\cup_{j=1}^m nU_j) \\
&\le \limsup_{n\to\infty} (1/n)\log \left(\textstyle\sum_{j=1}^m \mu_n^{*n}(nU_j)\right). \\
&\le \max_{1\le j\le m} \limsup_{n\to\infty} (1/n)\log \mu_n^{*n}(nU_j) \\
&\le \max_{1\le j\le m} \limsup_{n\to\infty} (-h(U_j \mid \mu_n)) \quad \text{by Lemma 3} \\
&\le \max_{1\le j\le m} (-\liminf_{n\to\infty} h(\bar U_j \mid \mu_n)) \\
&\le -\min_{1\le j\le m} h(\bar U_j \mid \mu) \quad \text{by Lemma 2} \\
&= -h(\cup_{j=1}^m \bar U_j \mid \mu) \\
&\le -h(A^\varepsilon \mid \mu) \quad \text{where } A^\varepsilon \text{ is the closed } \varepsilon\text{-neighbourhood of } A.
\end{aligned}$$

If $\varepsilon \downarrow 0$, then $h(A^\varepsilon \mid \mu)$ increases to $h(A \mid \mu)$, as follows easily from the compactness of A and the fact that $h(a \mid \mu)$ is lower semicontinuous.

The general noncompact case can now be reduced to the compact case as is done in [6], by just showing that all arguments there work uniformly in n if (1.3) is satisfied.

Let μ_n^n be the n-fold product measure on B^n and θ_n: $B^n \to \mathbb{P}$ be defined by $\theta_n(x_1, \cdots, x_n) = (1/n)\sum_{j=1}^n \delta_{x_j}$ where δ_x is the one point measure in x.

LEMMA 4. *Given any $a > 0$, there is a compact set $C(a) \in \mathbb{P}$ (in the weak topology) with $\mu_n^n(\theta_n \notin C(a)) \le e^{-na}$ for all $n \in \mathbb{N}$.*

PROOF. This follows by a straightforward transcription of the corresponding result where the μ_n do not depend on n (see e.g. [1], Lemma I 7.4).

We construct now a sequence $0 = t_0 < t_1 < \cdots$, such that for $k \in \mathbb{N}$

$$\sup_n \int_{\|x\|\ge t_k} \exp(k\,\|x\|)\mu_n(dx) \le 2^{-k}.$$

Let f: $[0, \infty) \to [0, \infty)$ be such that $f(t)/t$ is continuous and increasing with

$\lim_{t\to\infty} f(t)/t = \infty$ and $f(t_k)/t_k \le k - 1$, $k \in \mathbb{N}$. Then it is easy to see that $\int \exp(f\|x\|)\mu_n(dx) \le 2$ for all n. If $a > 0$, let $G(a) = \{\nu \in \mathbb{P} : \int f(\|x\|)\nu(dx) \le a\}$. Then

$$\mu_n^n(\theta_n \notin G(a)) = \mu_n^n(\{x \in B^n: \textstyle\sum_{j=1}^n f(\|x_j\|) > na\} \le e^{-na}(\mu_n(e^f))^n \le e^{-na+n}. \tag{2.3}$$

Let $\Lambda(a) = \{\nu(\mathrm{id}): \nu \in C(a) \cap G(a)\}$. $C(a)$ is compact and $G(a)$ is closed in $\mathbb{P}$. Furthermore $\nu \to \nu(\mathrm{id})$ restricted to $G(a)$ is continuous. It follows that $\Lambda(a)$ is compact in B. Furthermore

$$\mu_n^{*n}(n\Lambda^c(a)) = \mu_n^n(\theta_n \notin \Lambda(a)) \le \exp(-na) + \exp(-na + n) \le 2\exp(-n(a - 1)).$$

If A is closed in B with $h(A \mid \mu) < \infty$, then

$$\begin{aligned} &\limsup_{n\to\infty} (1/n)\log \mu_n^{*n}(nA) \\ &\quad\le \limsup_{n\to\infty} (1/n)\log(\mu_n^{*n}(n(A \cap \Lambda(a))) + 2\exp(-n(a-1))) \\ &\quad= \limsup_{n\to\infty} (1/n)\log \mu_n^{*n}(n(A \cap \Lambda(a))) \quad \text{if} \quad a > h(A) + 1 \\ &\quad\le -h(A \cap \Lambda(a) \mid \mu) \le -h(A \mid \mu). \end{aligned}$$

So (1.4) is proved.

3. The lower estimate.

LEMMA 5. *Let $A \subset B$ be open, $\varepsilon > 0$, $\mu \in \mathbb{P}$ with $\int \exp(t\|x\|)\mu(dx) < \infty$ for all t. Then there is a $\nu \in \mathbb{P}$ with a bounded continuous everywhere positive density g* w.r.t. *μ, such that $k(\nu \mid \mu) \le h(A \mid \mu) + \varepsilon$ and $\nu(\mathrm{id}) \in A$.*

PROOF. We may assume that $h(A \mid \mu) < \infty$. Then there is a $\nu' \in \mathbb{P}$ with $k(\nu' \mid \mu) \le h(A \mid \mu) + \varepsilon$ and $\nu'(\mathrm{id}) \in A$. Let $g' = d\nu'/d\mu$. If we put $g_n = (n \wedge g') \vee (1/n)$, then $\int g_n \log g_n d\mu \to k(\nu' \mid \mu)$, $\int g_n d\mu \to 1$ and $\int x\, g_n(x)\mu(dx) \to \nu'(\mathrm{id})$. By taking the densities $g_n/\int g_n\, d\mu$, we see that there is a bounded density g'', which is bounded away from 0, such that if $d\nu'' = g'' d\mu$, we have $k(\nu'' \mid \mu) \le h(A \mid \mu) + \varepsilon$, $\nu''(\mathrm{id}) \in A$. Approximating this density pointwise by bounded continuous densities which remain bounded away from 0, we arrive at the desired conclusion.

Let now $\mu_n \to \mu$ as in the statement of the theorem and let g be as in Lemma 5. We put $d\nu_n = g d\mu_n/\int g d\mu_n$. Then $\int g d\mu_n \to \int g d\mu = 1$ and therefore $k(\nu_n \mid \mu_n) \to k(\nu \mid \mu)$ and $\nu_n(\mathrm{id}) \to \nu(\mathrm{id})$.

Given $\varepsilon > 0$ there exists $N \in \mathbb{N}$, such that if $n \ge N$ $k(\nu_n \mid \mu_n) \le h(A \mid \mu) + \varepsilon$ and $\nu_n(\mathrm{id}) \in A$. If f is the function constructed after Lemma 4, we have $\sup_n \int \exp(f(\|x\|))\nu_n(dx) < \infty$ and therefore $\sup_n \int f(\|x\|)\nu_n(dx) < \infty$.

PROOF OF (1.5). Let $b: \mathbb{P} \to B$ be defined by $b(\nu) = \nu(\mathrm{id})$, whenever it exists.

Then

$$\mu_n^{*n}(nA) = \mu_n^n(\theta_n \in b^{-1}(A)) \geq \mu_n^n(\theta_n \in b^{-1}(A) \cap G(a))$$

for any $a > 0$. As b is continuous on $G(a)$ and $\nu \in G(a)$ for sufficiently large a, we have a weak neighbourhood U of ν in $\mathbb{P}$, such that $b^{-1}(A) \cap G(a) \supset U \cap G(a)$.

$$\mu_n^n(\theta_n \in b^{-1}(A) \cap G(a)) \geq \mu_n^n(\theta_n \in U) - \mu_n^n(\theta_n \notin G(a))$$
$$\geq \mu_n^n(\theta_n \in U) - \exp(-n(a-1)).$$

Let U be of the form

$$U = \{\pi \in \mathbb{P} : |\pi(f_1) - \nu(f_1)| < \varepsilon, \cdots, |\pi(f_k) - \nu(f_k)| < \varepsilon\}$$

where $f_1, \cdots, f_k$ are bounded continuous functions on B. As μ_n is equivalent to ν_n, we have

$$\mu_n^n(\theta_n \in U) = \int_{A_n} \exp\left(- \textstyle\sum_{j=1}^n \log \left(\frac{d\nu_n}{d\mu_n}(x_j)\right)\right)\nu_n^n(d\underline{x}),$$

where $A_n = \{\underline{x} \in B^n : |(1/n) \sum_{j=1}^n f_i(x_j) - \nu(f_i)| < \varepsilon, 1 \leq i \leq k\}$.

$$\mu_n^n(\theta_n \in U) = \exp(-nk(\nu \mid \mu)) \int_{A_n} \exp\left(- \textstyle\sum_{j=1}^n \left(\log \frac{d\nu_n}{d\mu_n}(x_j) - k(\nu \mid \mu)\right)\right)\nu_n^n(d\underline{x})$$

$$\geq \int_{A_n \cap B_n(\delta)} \exp(-n\delta)\exp(-nk(\nu \mid \mu))\nu_n^n(d\underline{x})$$

$$= \exp(-n(k(\nu \mid \mu) + \delta))\, \nu_n^n(A_n \cap B_n(\delta))$$

where

$$B_n(\delta) = \left\{\underline{x} \in B^n : \left| \frac{1}{n} \textstyle\sum_{j=1}^n \log \frac{d\nu_n}{d\mu_n}(x_j) - k(\nu \mid \mu)\right| < \delta\right\}$$

and $\delta > 0$ is arbitrary. $\mathrm{Log}(d\nu_n/d\mu_n)$ is bounded and has mean $k(\nu_n \mid \mu_n)$ when integrated over $d\nu_n$. As $k(\nu_n \mid \mu_n) \to k(\nu \mid \mu)$ and $\nu_n(f_j) \to \nu(f_j)$, $1 \leq j \leq k$. We obtain by an application of the Tchebychev inequality that $\nu_n^n(A_n \cap B_n(\delta)) \to 1$ as $n \to \infty$. Therefore, we have

$$\liminf_{n\to\infty} (1/n) \log \mu_n^{*n}(nA) \geq -\min(k(\nu \mid \mu) + \delta, a - 1).$$

As a is arbitrary large, δ arbitrary small and $k(\nu \mid \mu)$ arbitrary close to $h(A \mid \mu)$, (1.5) follows.

4. The use of dominating points in convex sets. We shall discuss here how (1.1), (1.2), (1.4), (1.5) can be improved if A is convex. This can be achieved by the use of so-called dominating points (see P. Ney [11]). Let $A \subset B$ be closed and convex. By using (2.1) and the well-known strong convexity of $k(\nu \mid \mu)$ in ν

one sees that $h(a \mid \mu)$ is strongly convex. Therefore, if $h(A \mid \mu) < \infty$, there exists a unique $a_0 \in A$ with $h(a_0 \mid \mu) = h(A \mid \mu)$ and then a unique $\nu_0 \in \mathbb{P}$ with $k(\nu_0 \mid \mu) = h(A \mid \mu)$ and $\nu_0(\mathrm{id}) = a_0$. From Csiszar [5] (2.8) and Theorem 2.2 it follows that if $\nu \in \mathbb{P}$ is such that $k(\nu \mid \mu) < \infty$ and $\nu(\mathrm{id}) \in A$, then

$$\nu(\log(d\nu_0/d\mu)) \geq k(\nu_0 \mid \mu). \tag{3.1}$$

$$\nu \ll \nu_0. \tag{3.2}$$

Let $A_n = \{\underset{\sim}{x} = (x_1, \cdots, x_n) \in B^n \colon (1/n) \sum_{j=1}^n x_j \in A\}$ and μ^n, ν_0^n be the n-fold product probabilities on B^n of μ resp. ν_0.

PROPOSITION 1.
a) $\mu^{*n}(nA) = \int_{A_n} \exp(-\sum_{j=1}^n \log(d\nu_0/d\mu)(x_j))\, \nu_0^n(d\underset{\sim}{x})$,
b) $\sum_{j=1}^n \log(d\nu_0/d\mu)(x_j) \geq nk(\nu_0 \mid \mu)$ μ^n – a.s. *on* A_n.

PROOF. a) $\mu^{*n}(nA) = \mu^n(A_n)$. It therefore suffices to show that for $1 \leq j \leq n$: $\mu^n(A_n \cap \{\underset{\sim}{x}\colon (d\nu_0/d\mu)(x_j) = 0\}) = 0$. It suffices to take $j = 1$. Let $\Gamma = A_n \cap \{\underset{\sim}{x} \in B^n\colon (d\nu_0/d\mu)(x_1) = 0\}$. If $\mu^n(\Gamma) > 0$, we define a probability measure ρ on $(B, \mathscr{B})$ by $\rho(C) = \int_\Gamma (1/n) \sum_{j=1}^n 1_C(x_j)\mu^n(d\underset{\sim}{x})/\mu^n(\Gamma)$. Then $d\rho/d\mu \leq (\mu^n(\Gamma))^{-1}$ and $\rho(\mathrm{id}) \in A$. Therefore, using (3.2), one has $\rho < < \nu_0$. Let $N = \{x \in B\colon (d\nu_0/d\mu)(x) = 0\}$. Then

$$\rho(N) = \int_\Gamma \frac{1}{n} \sum_{j=1}^n 1_N(x_j)\mu^n(d\underset{\sim}{x})/\mu^n(\Gamma)$$
$$\geq \frac{1}{n} \int_\Gamma 1_N(x_1)\mu^n(d\underset{\sim}{x})/\mu^n(\Gamma) = \frac{1}{n}$$

which is a contradiction.

PROOF OF b). Let $\Gamma = \{\underset{\sim}{x} \in B^n\colon (1/n) \sum_{j=1}^n \log(d\nu_0/d\mu)(x_j) < k(\nu_0 \mid \mu)\} \cap A_n$. *If* $\mu^n(\Gamma) > 0$, we define ρ as in a). Then again $d\rho/d\mu \leq (\mu^n(\Gamma))^{-1}$ and $\rho(\mathrm{id}) \in A$. From (3.1) it follows that $\rho(\log(d\nu_0/d\mu)) \geq k(\nu_0 \mid \mu)$. This contradicts

$$\rho(\log d\nu_0/d\mu) = \int_\Gamma \frac{1}{n} \sum_{j=1}^n \log \frac{d\nu_0}{d\mu} (x_j)\mu^n(d\underset{\sim}{x})/\mu^n(\Gamma) < k(\nu_0 \mid \mu).$$

The propositions may be applied to get upper bounds in the following way: If A is closed and convex, then

$$\mu^{*n}(nA) \leq e^{-nh(A \mid \mu)} \int_\Gamma \exp\left(- \sum_{j=1}^n \left(\log \frac{d\nu_0}{d\mu} (x_j) - h(A \mid \mu)\right)\right)\nu_0^n(d\underset{\sim}{x}) \tag{3.3}$$

where $\Gamma = \{\underset{\sim}{x} \in B^n\colon \sum_{j=1}^n \log(d\nu_0/d\mu)(x_j) \geq h(A \mid \mu)\}$ and this is

$$\leq e^{-nh(A \mid \mu)} \int \exp\left(- \left| \sum_{j=1}^n \log \frac{d\nu_0}{d\mu} (x_j) - h(A \mid \mu) \right| \right)\nu_0^n(d\underset{\sim}{x}).$$

As an application, we prove the following result. Let

$$J = \{a \in B\colon h(a \mid \mu) < \infty\}. \quad x_0 = \mu(\mathrm{id}) \in J \quad \text{and}$$

$$J' = \{\lambda x_0 + (1 - \lambda)a\colon a \in J,\ \lambda \in (0, 1]\} \subset J.$$

THEOREM 3. *If A is closed and convex, $x_0 \notin A$ and $h(A \cap J' \mid \mu) < \infty$, then*

$$\mu^{*n}(nA)\exp(nh(A \mid \mu))$$

$$= O\left(\int \frac{1}{n} \left| \sum_{j=1}^n \left(\log \frac{d\nu_0}{d\mu}(x_j) - h(A \mid \mu)\right) \right| \nu_0^n(d\underline{x})\right) = o(1).$$

REMARK. It seems likely that the condition $h(A \cap J' \mid \mu) < \infty$ is satisfied in all reasonable cases where $h(A \mid \mu) < \infty$ although a proof eludes me. It is certainly satisfied if $h(\mathrm{int}\ A) < \infty$ or if $J - x_0$ is a linear subspace of B, as is true for Gaussian measures.

PROOF OF THEOREM 3. $\mathrm{Log}(d\nu_0/d\mu)$ has expectation $h(A \mid \mu)$ under ν_0. If $h(A \cap J' \mid \mu) < \infty$ is satisfied, there exists a $\nu' \in \mathbb{P}$ with $k(\nu' \mid \mu) < \infty$, $\nu' \sim \mu$ and ν' (id) $\in A$. Therefore, it follows from (3.2) that $\nu_0 \gg \nu' \sim \mu$ and therefore $\nu_0 \sim \mu$. If $\log(d\nu_0/d\mu) = h(A)\ \nu_0 -$ a.s. it follows that $d\nu_0/d\mu = 1$ contradicting $x_0 \notin A$. Therefore we have $\nu_0(\log(d\nu_0/d\mu) = h(A)) < 1$. The theorem then follows from (3.3) and the following.

LEMMA 6. *Let $Y_1, Y_2, \cdots$ be an* i.i.d. *sequence of real valued random variables with $E \mid Y_i \mid < \infty$, $EY_i = 0$, $P(Y_i = 0) < 1$. Then there is a constant $c > 0$, such that*

$$E(\exp(- \mid \textstyle\sum_{j=1}^n Y_j \mid)) \le c\ E \mid (1/n) \sum_{j=1}^n Y_j \mid \quad \textit{for all} \quad n.$$

PROOF. Let $f(x) = \mathrm{sign}(x)(1 - e^{-|x|})(\mathrm{sign}(0) = 1)$. Then $f'(x) = e^{-|x|}$. Let $S_n = \sum_{j=1}^n Y_j$. Then

$$E(S_n f(S_n)) = n\ E(Y_n f(S_n)) = n\ E(Y_n(f(S_n) - f(S_{n-1})))$$

$$\ge n\ E(Y_n^2 f'(S_{n-1} + \theta Y_n);\ \mid Y_n \mid \le \beta)$$

for all $\beta > 0$, where θ is a random variable with $0 \le \theta \le 1$. Now $\exp(- \mid x + t \mid) \ge \exp(- \mid x \mid)e^{-\beta}$ if $\mid t \mid \le \beta$. Therefore

$$E \mid S_n \mid \ge E(S_n f(S_n)) \ge n e^{-\beta} E(\exp(- \mid S_{n-1} \mid)) E(Y_n^2;\ \mid Y_n \mid \le \beta)$$

$E(Y_n^2;\ \mid Y_n \mid \le \beta)$ is a constant, which is > 0 if β is large enough. So the lemma follows.

REMARKS. In any case

$$\int \left| \frac{1}{n} \sum_{j=1}^n \left(\log \frac{d\nu_0}{d\mu}(x_j) - h(A \mid \mu)\right) \right| \nu_0^n(d\underline{x}) = o(1).$$

If one further knows that $\nu_0((\log(d\nu_0/d\mu))^2) < \infty$, it is $O(1/\sqrt{n})$. This is satisfied if A is the closure of an open convex set which is flat at the point at which satisfies $h(a \mid \mu) = h(A \mid \mu)$. We recall that a point $x \in \partial A$ is called a flat point if there is a unique closed hyperplane through x which has A on one side. If $y \in \text{int } A$ then $x \in \partial A$ is a flatpoint if and only if the function $q_y(z) = \inf\{\rho \geq o \colon z - y \in \rho(A - y)\}$ is Gâteaux-differentiable at x. This implies that if x is flat there is a $\varphi \in B^*$ such that for all z with $\varphi(x) = \varphi(z)$ and all $y \in \text{int } A$

$$\text{(3.4)} \quad \inf\{\lambda > 0 \colon \lambda y + (1 - \lambda)(tz + (1 - t)x) \in A\} = o(t) \quad \text{as} \quad t \to 0.$$

If A is the closure of an open convex set B then $B = \text{int } A$ and if $x_0 = \mu(\text{id}) \notin A$ then it easily follows that $h(a \mid \mu) = h(B \mid \mu)$ and if this is smaller than infinity then the unique point a with $h(a \mid \mu) = h(A \mid \mu)$ belongs to ∂A.

THEOREM 4. *If in the above described situation a is a flat point then* $\mu^{*n}(nA)\exp(nh(A \mid \mu)) = O(1/\sqrt{n})$.

PROOF. As is mentioned above, $h(\text{int } A \mid \mu) < \infty$, so the conditions in Theorem 3 are satisfied. If $y \in B$ is any point with $h(y \mid \mu) < \infty$, and z satisfies $\varphi(z) = \varphi(a)$ (φ the above Gâteaux-derivative at a) then

$$\begin{aligned} h(a) &\leq \lambda(t)h(y) + t(1 - \lambda(t))h(z) + (1 - \lambda(t))(1 - t)h(a) \\ &\leq \lambda(t)h(y) + t(1 - \lambda(t))(h(z) - h(a)) + h(a) \end{aligned}$$

where $\lambda(t)$ is the infimum in (3.4). Using (3.4) and $h(y) < \infty$ we see that $h(z) \geq h(a)$. Therefore

$$h(a \mid \mu) = \inf\left\{k(\nu \mid \mu) \colon \int \varphi(x)\nu(dx) = \varphi(a)\right\}$$

and this infimum is attained at ν_0.

From Theorem 3.1 in [5] it follows that there is a $t \in \mathbb{R}$ with $d\nu_0/d\mu = \exp(t\varphi)/M(t\varphi)$. Therefore $\log(d\nu_0/d\mu)$ has moments of any order under ν_0 and so Theorem 4 follows from Theorem 3.

REMARK. In some Banach spaces the condition that all boundary points are flat is quite strong. E.g. balls have this property in L_p-space but not in $C[0, 1]$. On the other hand, even in $C[0, 1]$, balls have many flat boundary points, e.g. in the unit ball every f for which there is a unique $t \in [0, 1]$ with $|f(t)| = \|f\|_\infty = 1$ is a flat point (for this and the other facts on flat points used here, see Köthe [10], Section 26). So the estimate in Theorem 4 might be useful even in such spaces.

REFERENCES

[1] AZENCOTT R. (1980). Grandes déviations et applications. *Lecture Notes in Math.* **774**. Springer, Berlin.

[2] BAHADUR, R. R. and RAO, R. R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

[3] BAHADUR, R. R. and ZABELL, S. L. (1979). Large deviations of the sample mean in general vector spaces. *Ann. Probab.* **7** 587–621.

[4] CHEVET, S. (1982). Gaussian measures and large deviations. Unpublished manuscript
[5] CSISZAR, I. (1975). I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* **3** 146–158.
[6] DONSKER, M. D. and VARADHAN, S. R. S. (1976). Asymptotic evaluation of certain Markov process expectations for large time III. *Comm. Pure Appl. Math.* **29** 389–461.
[7] ELLIS, R. S. and ROSEN, J. S. (1982). Laplace's method for Gaussian integrals with an application to statistical mechanics. *Ann. Probab.* **16** 47–66.
[8] FERNIQUE, M. X. (1970). Intégrabilité des vecteurs Gaussiens. *Acad. Sci., Paris, Compt. Rendus* **270** Ser. A 1698–1699.
[9] HÖGLUND, T. (1979). A unified formulation of the central limit theorem for small and large deviations from the mean. *Z. Wahrsch. verw. Gebiete* **49** 105–117.
[10] KÖTHE, G. (1960). *Topologische Lineare Räume I.* Springer, Berlin.
[11] NEY, P. (1983). Dominating points and asymptotics of large deviations in $\mathbb{R}^d$. *Ann. Probab.* **11** 158–167.

TECHNISCHE UNIVERSITÄT BERLIN
FACHBEREICH MATHEMATIK
STRASSE DES 17. JUNI 135
1000 BERLIN 12 (WEST)