PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 103, Number 4, August 1988

A COMPARISON PRINCIPLE FOR LARGE DEVIATIONS

JOHN R. BAXTER AND NARESH C. JAIN

(Communicated by William D. Sudderth)

ABSTRACT. If $\{\mu_n\}$ and $\{\nu_n\}$ are two sequences of probability measures on a separable metric space, we give conditions under which $\{\mu_n\}$ satisfies a large deviation principle if and only if $\{\nu_n\}$ does. A known and a new theorem follow immediately from the application of this comparison principle to standard results in large deviation theory.

1. Introduction. Let (S, d) be a metric space, where d denotes the metric. $\mathscr{S}$ will denote the σ-algebra of Borel subsets of S. A sequence of probability measures on $(S, \mathscr{S})$ is said to satisfy a large deviation principle (LDP) with rate function I, if

(1.1) for every closed set F

$$\limsup_{n\to\infty}(1/n)\log\mu_n(F) \leq -\inf_{x\in F} I(x);$$

and

(1.2) for every open set G

$$\liminf_{n\to\infty}(1/n)\log\mu_n(G) \geq -\inf_{x\in G} I(x).$$

The rate function I is a function from S to $[0,\infty]$ which satisfies

(1.3a) $I(x)$ is not identically ∞, I is lower semicontinuous,

(1.3b) $\{I \leq a\}$ is compact for all $a < \infty$.

We will observe in §2 (Propositions 1 and 2) that if $\{\mu_n\}$ and $\{\nu_n\}$ are two sequences of probability measures on S which become sufficiently close as $n \to \infty$, then $\{\mu_n\}$ satisfies the LDP with rate function I if and only if $\{\nu_n\}$ does. We then apply these simple observations to obtain a comparison principle for the large deviations of the sample mean when the summands are independent and identically distributed (i.i.d.) and take values in a locally convex topological vector space (Theorem 1). Propositions 1, 2, and Theorem 1 are the main results of this note.

The motivation for Theorem 1 becomes clear in §3, where two applications are given. We first state two standard results on large deviations as Theorems 2 and 3 in §3 for ready reference. These theorems are extensions of the classical Cramér and Sanov theorems by Donsker and Varadhan [**3**]. Theorems 1 and 2 then yield Theorem 4 (due to Bolthausen [**1**]) as an immediate corollary. Likewise, Theorems 1 and 3 immediately give the corresponding perturbation of Theorem 3 without any additional work. Bolthausen's theorem [**1**] is a generalization of a theorem

Received by the editors July 16, 1987.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 60F10, 60G15, 60B12.
Key words and phrases. Large deviations, Banach-space-valued Random Variable.
This work was partly supported by NSF Grant DMS86-03437.

of Chevet [**2**]. For applications of such perturbations of standard large deviation results one may consult Ellis and Rosen [**5**].

For future use, we state the definition of the Prohorov distance between two probability measures μ, ν on (S,d).

$$\rho(\mu,\nu)=\inf\{\delta>0\colon \mu(F)\le\nu(F^\delta)+\delta \text{ for all closed sets } F\}, \tag{1.4}$$

where for any set $A\subset S$, $\delta>0$,

$$A^\delta=\{y\colon d(x,y)<\delta \text{ for some } x\in A\}. \tag{1.5}$$

If different metrics are used, we will write $A^\delta(d)$ to indicate that the metric d is under consideration.

For $\varepsilon>0$, we also define

$$\rho_\varepsilon(\mu,\nu)=\inf\{\delta>0\colon \mu(F)\le\nu(F^\varepsilon)+\delta \text{ for all closed sets } F\}. \tag{1.6}$$

It is easy to check that $\rho^\varepsilon(\mu,\nu)=\rho^\varepsilon(\nu,\mu)$.

Let $C_b(S)$ denote the continuous bounded functions on S. As usual, we say that a sequence of probabilities μ_n on S converges weakly to a limit μ if $\int_S f\,d\mu_n\to\int_S f\,d\mu$, for every $f\in C_b(S)$. We will denote the set of all probability measures on S by $\mathscr{M}_1(S)$, and consider $\mathscr{M}_1(S)$ to be a topological space with the topology corresponding to weak convergence. For a sequence of probability measures, convergence in the Prohorov distance implies weak convergence, and when the underlying space S is separable, the converse holds. Thus for separable S, $\mathscr{M}_1(S)$ is a separable metric space with the Prohorov distance as its metric.

In what follows we will use δ_x to denote the Dirac measure with unit mass at x.

2. The main results.

PROPOSITION 1. *Let $\{\mu_n\}$, $\{\nu_n\}$ be two sequences of probability measures on a metric space (S,d) such that for every closed set F and every $\varepsilon>0$*

$$\limsup_{n\to\infty}(1/n)\log\mu_n(F)\le\limsup_{n\to\infty}(1/n)\log\nu_n(F^\varepsilon), \tag{2.1a}$$

$$\liminf_{n\to\infty}(1/n)\log\nu_n(F)\le\liminf_{n\to\infty}(1/n)\log\mu_n(F^\varepsilon). \tag{2.1b}$$

Then the LDP holds for $\{\mu_n\}$ with rate function I if it holds for $\{\nu_n\}$ with the same rate function.

For the proof of Proposition 1 we need Lemma 1 below. As a convenient notation, for any set A we will define

$$I(A)=\inf_{x\in A}I(x). \tag{2.2}$$

LEMMA 1. *If F is closed, and I is a rate function (satisfying* (1.3)*), then*

$$I(F)=\lim_{\delta\to 0}I(F^\delta).$$

This lemma is well known, and the proof is easy. Note that (1.3b) is used in the proof of this lemma.

PROOF OF PROPOSITION 1. Assume the LDP holds for $\{\nu_n\}$ with rate function I. Let F be a closed set. Since the LDP holds for $\{\nu_n\}$ with rate I, we have by (2.1a) that

$$\limsup_{n\to\infty}(1/n)\log\mu_n(F)\le I(F^\varepsilon)$$

for all $\varepsilon > 0$, and by Lemma 1 we get (1.1) by letting $\varepsilon \to 0$.

If G is an open set, to prove (1.2) we may assume without any loss of generality that $I(G) < \infty$. Let $I(G) = \beta$. Fix $\alpha > \beta$. Then there is a point $x \in G$ with $I(x) < \alpha$. Let B be an open ball around x, $\varepsilon > 0$ such that $D^\varepsilon \subset G$, where D is the closure of B. We have $\mu_n(G) \geq \mu_n(D^\varepsilon)$. Also, by (2.1b)

$$\liminf_{n\to\infty}(1/n)\log\nu_n(D) \leq \liminf_{n\to\infty}(1/n)\log\mu_n(D^\varepsilon).$$

Therefore

$$\liminf_{n\to\infty}(1/n)\log\mu_n(G) \geq \liminf_{n\to\infty}(1/n)\log\nu_n(B) \geq -I(B) \geq -\alpha.$$

Thus (1.2) holds and Proposition 1 is proved.

PROPOSITION 2. *Let $\{\mu_n\}$, $\{\nu_n\}$ be two sequences of probability measures on a metric space (S,d) such that for every $\varepsilon > 0$*

$$\lim_{n\to\infty}(1/n)\log\rho_\varepsilon(\mu_n,\nu_n) = -\infty. \tag{2.4}$$

Then the LDP holds for $\{\mu_n\}$ with rate function I if and only if it holds for $\{\nu_n\}$ with the same rate function.

PROOF. For any $a, b > 0$,

$$\log(a+b) \leq \log 2 + \max(\log a, \log b).$$

For any closed set F, and any $\varepsilon > 0$, by definition (recall that ρ_ε is symmetric) $\mu_n(F) \leq \nu_n(F^\varepsilon) + \rho_\varepsilon(\mu_n,\nu_n)$ and $\nu_n(F) \leq \mu_n(F^\varepsilon) + \rho_\varepsilon(\mu_n,\nu_n)$. (2.1a) and (2.1b) follow at once, and since we may exchange μ_n and ν_n, Proposition 2 is proved.

Before stating Theorem 1, we recall the notion of a locally convex topological vector space. E is a locally convex vector space if there is a family of seminorms $\{||\cdot||_j\}_{j\in J}$ separating points on E, which defines the topology on E. We will assume that E is metrizable, or equivalently that the index set J is countable, say $J = \{1, 2, \ldots\}$. We also assume that E is *separable*. Let $\mathscr{E}$ denote the Borel sets of E. We will denote by d_j the metric on E defined by

$$d_j(x,y) = ||x-y||_j/(1+||x-y||_j)$$

and let

$$\bar{d} = \sum_{j=1}^{\infty} 2^{-j} d_j.$$

The metric $\bar{d}$ is a convenient metric for the topology on E.

If $\{\mu_n\}$ is a sequence of probability measures on $(E,\mathscr{E})$, for each n we will denote by $\bar{\mu}_n$ the probability measure on $(E,\mathscr{E})$ defined by

$$\bar{\mu}_n(A) = \mu_n^{*n}(nA), \qquad A \in \mathscr{E}, \tag{2.5}$$

where μ^{*k} denotes the k-fold convolution of μ. Note that if $X_{n1}, X_{n2}, \ldots, X_{nn}$ are i.i.d. with distribution μ_n, then $\bar{\mu}_n$ denotes the distribution of $(X_{n1}+\cdots+X_{nn})/n$.

THEOREM 1. *Let E be a locally convex topological vector space as above. Let $\{\mu_n\}$, $\{\nu_n\}$ be two sequences of probability measures on $(E, \mathscr{E})$ such that $\rho(\mu_n, \nu_n) \to 0$ as $n \to \infty$, and such that for every j and for each $t \geq 0$, the sequences $\int e^{t\|x\|_j}\, d\mu_n$ and $\int e^{t\|x\|_j}\, d\nu_n$ are bounded in n. Then $\{\bar{\mu}_n\}$ obeys the large deviation principle with rate function I if and only if $\{\bar{\nu}_n\}$ does.*

PROOF. It is enough to show that (2.4) holds, with $S = E$, $d = \bar{d}$, $\mu_n = \bar{\mu}_n$ and $\nu_n = \bar{\nu}_n$. Let F be a closed set, $\varepsilon > 0$. Let j be sufficiently large that if $d' = d_1 + \cdots + d_j$ then $\bar{d} < d' + \varepsilon/2$. Then $F^\varepsilon(\bar{d}) \supset F^{\varepsilon/2}(d')$. Thus it is enough to prove (2.4) when $d = d'$. Let $\|\cdot\| = \|\cdot\|_1 + \cdots + \|\cdot\|_j$, and let $d''(x,y) = \|x - y\|$. Clearly $d' \leq d''$. Thus $F^\varepsilon(d'') \subset F^\varepsilon(d')$. Thus it is enough to prove (2.4) when $d = d''$.

We now define a *coupling* of μ_n, ν_n. Since $\rho(\mu_n, \nu_n) \to 0$, by an extension of a theorem of Strassen, due to Dudley (cf. [**4**, Corollary 18.3]) we know that for each n we can find E-valued random variables X_n, Y_n (on some probability space), such that X_n has distribution μ_n, Y_n has distribution ν_n, and $\bar{d}(X_n, Y_n) \to 0$ in probability as $n \to \infty$. Hence $\|X_n - Y_n\| \to 0$ in probability as $n \to \infty$. For each n, let $(X_n(j), Y_n(j))$ be a sequence of independent $E \times E$-valued random variables such that for every j, $(X_n(j), Y_n(j))$ has the same distribution as (X_n, Y_n). Let $S_n = X_n(1) + \cdots + X_n(n)$, $T_n = Y_n(1) + \cdots + Y_n(n)$, $Z_n(j) = \|X_n(j) - Y_n(j)\|$, and $V_n = Z_n(1) + \cdots + Z_n(n)$. Let $c_n(t)$ denote $E(e^{tZ_n(j)})$. Then $c_n(t) \leq \int e^{2t\|x\|}\, d\mu_n + \int e^{2t\|x\|}\, d\nu_n$. It follows easily that $\{e^{tZ_n(j)} : n \geq 1\}$ is uniformly integrable. Since $Z_n(j) \to 0$ in probability as $n \to \infty$, we have $c_n(t) \to 1$ as $n \to \infty$, for all $t \geq 0$.

For any set $A \in \mathscr{E}$, for any $\varepsilon > 0$,

$$\bar{\mu}_n(A) = P(S_n \in nA) \leq P(T_n \in nA^\varepsilon) + P(V_n \geq n\varepsilon).$$

Since $Ee^{tV_n} = (c_n(t))^n$, $P(V_n \geq n\varepsilon) \leq e^{-n\varepsilon t}(c_n(t))^n$, we have

$$\bar{\mu}_n(A) \leq \bar{\nu}_n(A^\varepsilon) + e^{-n\varepsilon t}(c_n(t))^n.$$

Thus $(1/n)\log \rho_\varepsilon(\bar{\mu}_n, \bar{\nu}_n) \leq -\varepsilon t + \log c_n(t)$. Hence, since t is arbitrary, we see easily that (2.4) holds, so Theorem 1 is proved.

REMARKS. If the sequence $\{\nu_n\}$ in Theorem 1 is constant, i.e. $\nu_n = \nu$ for every n, then the condition $\rho(\mu_n, \nu_n) \to 0$ simply says that $\mu_n \to \nu$ weakly. Let $C_b^j(E)$ denote the space of bounded functions on E which are continuous with respect to the norm $\|\cdot\|_1 + \cdots + \|\cdot\|_j$. It is easy to see that the union of the spaces $C_b^j(E)$ is strongly separating and hence forms a convergence determining class. Thus when $\nu_n = \nu$ for all n, to check that $\rho(\mu_n, \nu_n) \to 0$ it is enough to check that for every j

$$\int_E f\, d\mu_n \to \int_E f\, d\nu, \qquad f \in C_b^j(E).$$

3. Applications.

We first state two standard results from [**3**].

THEOREM 2. *Let E be a separable Banach space with dual E^*, and let μ be a probability measure on E. If $\int \exp(t\|x\|)\, d\mu(x) < \infty$ for every $t > 0$, then $\bar{\mu}_n$ (defined by (2.5) with $\mu_n \equiv \mu$) satisfies the LDP with rate function*

$$I(x) = \sup_{\theta \in E^*} \left\{ \langle \theta, x\rangle - \log \int \exp(\langle \theta, y\rangle)\, d\mu(y) \right\}. \tag{3.1}$$

THEOREM 3. *Let (S,d) be a complete separable metric space, and let $X_1, X_2, \ldots$ be independent random variables with values in S and having common distribution α. Let $\bar{\mu}_n$ denote the distribution of $(\delta_{X_1} + \cdots + \delta_{X_n})/n$ in $\mathscr{M}_1(S)$. Then $\{\bar{\mu}_n\}$ satisfies the LDP with rate function $I_\alpha(\beta)$, where*

$$I_\alpha(\beta) = \int \log(d\beta/d\alpha)\, d\beta, \tag{3.2}$$

if $\beta \ll \alpha$ and $\int |\log(d\beta/d\alpha)|\, d\beta < \infty$; $I_\alpha(\beta) = \infty$, otherwise.

The next theorem is due to Bolthausen [**1**] and we will show how it follows directly from Theorems 1 and 2.

THEOREM 4. *Let E be a separable Banach space. Let $\{\mu_n\}$ be a sequence of probability measures on E such that $\mu_n \to \mu$ weakly. If for every $t > 0$, $\int \exp(t||x||)\, d\mu(x)$ is bounded in n, then $\bar{\mu}_n$ defined in* (2.5) *satisfies the LDP with rate function I defined by* (3.1).

PROOF. In Theorem 1 we take $||\cdot||_j = ||\cdot||$, the Banach norm, for every j. Let $\nu_n = \mu$. Theorem 1 then says that $\{\bar{\mu}_n\}$ satisfies the LDP if $\hat{\mu}_n$ does, and with the same rate function, where

$$\bar{\mu}_n(A) = \mu_n^{*n}(nA), \quad \text{and} \quad \hat{\mu}_n(A) = \mu^{*n}(nA).$$

The rest then follows from Theorem 2, and the proof is complete.

The next theorem is a perturbation of Theorem 3, in the same sense that Theorem 4 is a perturbation of Theorem 2.

THEOREM 5. *Let (S,d) be a complete separable metric space. Let $\{\alpha_n\}$ be a sequence of probability measures on (S,d) such that $\alpha_n \to \alpha$ weakly. For each n, let $X_1(n), X_2(n), \ldots, X_n(n)$ be S-valued independent random variables with common distribution α_n. Let $\bar{\mu}_n$ denote the distribution of $(\delta_{X_1(n)} + \cdots + \delta_{X_n(n)})/n$ on $\mathscr{M}_1(S)$. Then $\{\bar{\mu}_n\}$ satisfies the LDP with rate function I_α given by* (3.2).

PROOF. Let $\{f_j\}_{j\geq 1}$ be a convergence-determining class in $\mathscr{C}_b(S)$. We may choose each f_j to be uniformly continuous with $0 \leq f_j \leq 1$. The topology of weak convergence on $\mathscr{M}_1(S)$ is given by the metric

$$\hat{\rho}(\nu_1, \nu_2) = \sum_{j=1}^{\infty} 2^{-j} \left| \int f_j\, d\nu_1 - \int f_j\, d\nu_2 \right|.$$

Let E denote the linear space of linear functionals on the span of the f_j. Define seminorms $||\cdot||_j$ on E by $||\varphi||_j = |\varphi(f_j)|$, $j \geq 1$. Then E satisfies the conditions of Theorem 1. We may regard $\mathscr{M}_1(S)$ as a subset of E, since a measure ν gives a linear functional $\varphi_\nu(f) = \int f\, d\nu$, and a measure ν in $\mathscr{M}_1(S)$ is uniquely determined by the values $\varphi_\nu(f_j)$, $j = 1, 2, \ldots$. It is easy to see that the topology induced on $\mathscr{M}_1(S)$ by E is the weak topology.

Let $\psi\colon S \to \mathscr{M}_1(S)$ be the map $\psi(x) = \delta_x$. ψ is continuous. Let $\mu_n \equiv \alpha_n \circ \psi^{-1}$, $\mu \equiv \alpha \circ \psi^{-1}$. Then $\mu_n \to \mu$ weakly as $n \to \infty$. We can extend the measures $\bar{\mu}_n$, μ_n, μ from $\mathscr{M}_1(S)$ to E in the usual way, by setting each measure equal to 0 on $E - \mathscr{M}_1(S)$. Clearly we still have $\mu_n \to \mu$ weakly as $n \to \infty$. Hence $\rho'(\mu_n, \mu) \to 0$, where ρ' is the Prohorov distance with respect to the metric $\bar{d} = \sum_{j=\infty}^{\infty} 2^{-j} d_j$ defined in §2. Since $\delta_{X_j(n)}$ has distribution μ_n, $\bar{\mu}_n$ and μ_n satisfy (2.5). We extend

I_α to E by setting $I_\alpha(\varphi) = \infty$ if φ is not in $\mathscr{M}_1(S)$. Let $\bar{\nu}_n(A) = \mu^{*n}(nA)$, $A \in \mathscr{E}$. Then the sequence $\{\bar{\nu}_n\}$ satisfies the LDP on $\mathscr{M}_1(S)$ with rate function I_α, by Theorem 3. It is trivial to see that then $\{\bar{\nu}_n\}$ satisfies the LDP *on* E with rate function I_α. We note that $\int \exp(t||x||_j)\,d\mu_n(x) = \int \exp(tf_j)\,d\alpha_n \leq e^t$. Thus the conditions of Theorem 1 hold, and hence $\{\bar{\mu}_n\}$ satisfies the LDP with rate function I_α on E, and in particular on $\mathscr{M}_1(S)$, so Theorem 5 is proved.

REMARK. We assumed that (S, d) was complete separable in Theorem 5, because Theorem 3 requires this assumption. If we wished to state Theorem 5 as a comparison principle for two general sequences α_n and β_n, the completeness condition could be dropped, since only Theorem 1 is needed in this case. The proof of this comparison principle is very little different from the argument given above, except that one needs to note that the map ψ is *uniformly* continuous between the two metrics used, and thus if $\rho(\alpha_n, \beta_n) \to 0$, where ρ denotes the Prohorov distance using d, then $\rho'(\alpha_n\psi^{-1}, \beta_n\psi^{-1}) \to 0$, where ρ' denotes the Prohorov distance using $\bar{d}$.

ACKNOWLEDGEMENT. We are thankful to S. R. S. Varadhan for the formulation of condition (2.4) in Proposition 2.

REFERENCES

1. E. Bolthausen, *On the probability of large deviations in Banach spaces*, Ann. Probab. **12** (1984), 427–435.
2. S. Chevet, *Gaussian measures and large deviations*, Probability in Banach Spaces. IV, Oberwolfach 1982, Lecture Notes in Math., vol. 990, Springer-Verlag, Berlin and New York, 1983, pp. 30–46.
3. M. D. Donsker and S. R. S. Varadhan, *Asymptotic evaluation of certain Markov process expectations for large time.* III, Comm. Pure Appl. Math. **29** (1976), 389–461.
4. R. M. Dudley, *Convergence of laws on metric spaces, with a view to statistical testing*, Lecture Notes No. 45, Matematisk Institut, Aarhus Universitet, Denmark, 1976.
5. R. S. Ellis and J. S. Rosen, *Laplace's method for Gaussian integrals with an application to statistical mechanics*, Ann. Probab. **10** (1982), 47–66.
6. S. R. S. Varadhan, *Large deviations and applications*, CBMS-NSF Regional Conf. Ser. in Appl. Math., SIAM, Philadelphia, Pa., 1982.

SCHOOL OF MATHEMATICS, UNIVERSITY OF MINNESOTA, MINNEAPOLIS, MINNESOTA 55455