HH

DIAS-STP-93-33
SCU 9404

THERMODYNAMIC PROBABILITY THEORY: SOME ASPECTS OF LARGE DEVIATIONS

J.T. Lewis[a], C.-E. Pfister[b]

[a]Dublin Institute for Advanced Studies
10 Burlington Road
Dublin 4, Ireland
[b]Ecole Polytechnique Fédérale de Lausanne
Département de Mathématiques
CH-1015 Lausanne Switzerland

CERN LIBRARIES, GENEVA
P00020208

1 Introduction

Boltzmann's remarkable formula

$$S = k \ln W, \tag{1.1}$$

relating the thermodynamic entropy S of a macroscopic equilibrium state to the number W of microscopic states which correspond to the macroscopic state, caused problems for theoretical physicists: the entropy is a function of a small number of variables such as u, the internal energy per unit volume, while W must be interpreted as a measure on phase-space. To have any hope of giving the formula a precise meaning, we should pass to the limit in which the volume of the system becomes infinite; in this limit, we might expect a formula

$$s(u) = \lim_{V\to\infty} \frac{k}{V} \ln W \left[\frac{H_V}{V} = u \right], \tag{1.2}$$

where $s(u)$ is the entropy per unit volume and $W[\frac{H_V}{V} = u]$ is the measure of the subset of phase-space on which the Hamiltonian H_V takes the value uV; however, the existence of the limit in (1.2) poses obvious difficulties. These were resolved by Ruelle [R1] in 1965.

Ruelle's idea can be illustrated simply: under mild restrictions on the Hamiltonian H, one can prove that the limit

$$s[J] := \lim_{V\to\infty} \frac{k}{V} \ln W \left[\frac{H_V}{V} \in J \right] \tag{1.3}$$

exists for each open interval J containing the point u; taking the infimum over all such J, we obtain the entropy function $s(u)$:

$$s(u) := \inf_{J \ni u} s[J]. \tag{1.4}$$

The subadditivity argument used to establish the existence of the limit proves also that $u \to s(u)$ is concave. This simple idea, extended to vector-valued Hamiltonians, was developed by Ruelle and Lanford to provide a rigorous treatment of statistical thermodynamics, described in detail in Lanford's 1971 Battelle lectures [L]. Ruelle's idea turned out to have a surprising ramification in probability theory: Lanford used it to give a completely new proof of Cramèr's Theorem; this was the first step in an important development in the theory of large deviations.

The theory of large devations began as an attempt by Khinchin [Kh] in 1928 to refine the central limit theorem for sums of Bernoulli random variables; however, the theory we shall describe starts with Cramèr's refinement [C] of the weak law of large numbers. In a digression, Lanford [L] gives a proof of the following version of Cramèr's Theorem:

Theorem 1.1 *Consider a sequence $X_1, X_2, \ldots$ of bounded, independent, identically distributed random variables taking values in a finite-dimensional vector space E; we denote their common distribution by $\mathbb{K}$ and assume that the support Γ of $\mathbb{K}$ is non-empty. There exists a continuous concave function s on Γ such that, for every convex open set J with $J \cap \Gamma \neq \emptyset$,*

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\,[\,\frac{1}{n}\{X_1 + \ldots + X_n\} \in J\,] = \sup_{x \in J \cap \Gamma} s(x). \tag{1.5}$$

Lanford first proves that ψ, defined by

$$\psi(n; J) := \frac{1}{n} \ln \mathbb{P}\,[\,\frac{1}{n}\{X_1 + \ldots + X_n\} \in J\,] \tag{1.6}$$

is subadditive:

$$\psi(n+m; J) \geq \psi(n; J) + \psi(m; J); \tag{1.7}$$

from this follows the existence of the limit:

$$s[J] := \lim_{n\to n} \frac{\psi(n; J)}{n} = \sup_n \frac{\psi(n; J)}{n}. \tag{1.8}$$

Defining $s(x)$ by

$$s(x) := \inf_{J \ni x} s[J], \tag{1.9}$$

Lanford then proves that

$$s[J] = \sup_{x \in J} s(x). \tag{1.10}$$

This idea was taken up by Bahadur and Zabell [BZ]; they developed it to prove a powerful generalization of Cramèr's Theorem in which the X_j are possibly unbounded and the space E is possibly infinite-dimensional. Azencott [A] and, later, Deuschel and Stroock [DS], systematized these developments.

Under the hypotheses of Theorem 1.1, the function s is strictly concave and attains its supremum at m, the mean of the distribution $\mathbb{K}$; it is a simple consequence of Theorem 1.1 that $\mathbb{K}_n$, the distribution of

$$Y_n := \frac{1}{n}\{\, X_1 + \ldots + X_n \,\}, \tag{1.11}$$

converges to the Dirac measure δ_m in the sense that

$$\int_\Gamma f(x)\,\mathbb{K}_n[dx] \;\to\; f(m) \tag{1.12}$$

as $n \to \infty$ for every continuous function $f : \Gamma \to \mathbb{R}$. In addition, Theorem 1.1 gives the exponential rate at which this limit is approached. It is in this sense that Cramèr's Theorem refines the Weak Law of Large Numbers.

By now, large deviation techniques are used widely for the rigorous treatment of problems in mathematical physics. This development came from a different direction: in the early 'sixties, Donsker proposed using such a refinement of the Weak Law of Large Numbers to give a generalization of Laplacian asymptotics to function-spaces; this was exploited in the thesis of his student Schilder [S]. Here is the idea: let $\{\mathbb{K}_n\}_{n\geq 1}$ be a sequence of probability measures on a space E; if this sequence behaves asymptotically for large n as

$$\mathbb{K}_n[dx] \sim \exp ns(x)\text{“}dx\text{”}, \tag{1.13}$$

where "dx" is some reference measure, then we might expect that

$$\int_E \exp(ng(x))\mathbb{K}_n[dx] \sim \exp(n \sup_{x\in E}\,(g(x)+s(x)). \tag{1.14}$$

This idea was given precise form in the definitive work of Varadhan [V]: the sequence $\{\mathbb{K}_n\}_{n\geq 1}$ obeys a Large Deviation Principle if there exists a function $s : E \to \overline{\mathbb{R}}$ such that

- (LDP 1) the function s is upper semicontinuous;
- (LDP 2) the level sets of s are compact;
- (LDP 3) for each closed set F,

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{K}_n[F] \leq \sup_{x\in F} s(x); \tag{1.15}$$

- (LDP 4) for each open set G,

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mathbb{K}_n[G] \geq \sup_{x\in G} s(x); \tag{1.16}$$

the function s is called the rate-function. (Many authors call the negative of s the rate-function.) Varadhan's Theorem (see Corollary 6.1) gives conditions on a function $g : E \to \mathbb{R}$ so that, if the sequence $\{\mathbb{K}_n\}_{n\geq 1}$ obeys a Large Deviation Principle, we have

$$\lim_{n\to\infty} \frac{1}{n} \ln \int_E \exp(ng(x))\mathbb{K}_n[dx] = \sup_{x\in E}\,(g(x)+s(x)). \tag{1.17}$$

As Ellis has pointed out [E], hypothesis (LDP 2) is not required for the proof of (1.17); however, we shall see that its presence ensures that a Large Devation Principle is a covariant notion: if $\phi : E \to E'$ is continuous and the sequence $\{\mathbb{K}_n\}_{n\geq 1}$ obeys an LDP, then the sequence $\{\mathbb{K}_n \circ \phi^{-1}\}_{n\geq 1}$ of image measures also obeys an LDP. The theorem which relates the two rate-functions is known as the Contraction Principle. To a physicist, hypotheses (LDP 3) and (LDP 4) may seem

strange at first sight; of course, for special sets, together they imply a statement like (1.5). As O'Brien and Vervaat [OV] pointed out, even as they stand, they constitute a statement about the convergence of a sequence of set-functions: the set-function

$$B \to k_n[B] := \frac{1}{n} \ln \mathbb{K}_n[B] \tag{1.18}$$

is an increasing set-function as is the set-function

$$B \to {}^{\vee}s[B] := \sup_{x \in B} s(x); \tag{1.19}$$

together, conditions (LDP 3) and (LDP 4) can be re-stated as

- the sequence $\{k_n\}_{n \geq 1}$ converges in the narrow topology (Bourbaki: *topologie étroite*) on the space of increasing set-functions on E to ${}^{\vee}s$.

The condition (LDP 3) is sometimes weakened to

- (LDP 3′) for each compact set K,

$$\limsup_{n \to \infty} \frac{1}{n} \ln \mathbb{K}_n[K] \leq \sup_{x \in K} s(x); \tag{1.20}$$

together, conditions (LDP 3′) and (LDP 4) can be restated as:

- the sequence $\{k_n\}_{n \geq 1}$ converges in the vague topology (Bourbaki: *topologie vague*) on the space of increasing set-functions on E to ${}^{\vee}s$.

In the hands of Donsker and Varadhan, these theorems of Varadhan proved to be a powerful tool for solving problems which could be formulated in terms of the asymptotic behaviour of a Feynman-Kac integral; perhaps their most spectacular success was their solution of the long-standing Polaron Problem [DV]. Independently of this, large deviation techniques in function spaces were developed in the Soviet Union by Borovkov, Freidlin and Wentzell; see Freidlin and Wentzell [FW], for example. It is not our intention to review these developments; our aim is much more modest: to explore the rôle of thermodynamic functions in the theory of large deviations. Little of what we present is original except, perhaps, the organization of the material and Theorem 7.2.

Entropy is not the only thermodynamic function to arise in the theory; the grand canonical pressure also has a part to play. We saw that entropy enters in Lanford's proof of Cramèr's Theorem; another proof, due to Chernoff [Ch], makes use of the pressure. To see the relevance of the pressure in a more general context, consider Varadhan's formula (1.17) in the case in which E is a vector space and g is a linear functional $x \to \langle t, x \rangle$; then the left-hand side of (1.17) can be identified with the pressure p and the right-hand side with the conjugate of $-s$, so that

$$p(t) = \sup_{x \in E} (\langle t, x \rangle + s(x)); \tag{1.21}$$

if s is concave, then (1.21) can be inverted to give

$$s(x) = -\sup_{t \in E^*} (\langle t, x \rangle - p(t)). \tag{1.22}$$

Now we turn to a description of the contents of this paper. The probability measures which are studied in the theory of large deviations are the distributions of

random variables taking values in a topological space X, so that they are measures on a topological space; moreover, the topological structure of the space X enters the theory as well its Borel structure. It has been our policy to introduce hypotheses about the topological structure only when we need them.

We begin in Section 2.1 with a description of the general setting. We have seen that the traffic between statistical mechanics and the theory of large deviations has been two-way; it is not surprising then that there is a primitive principle which occurs in them both. We identify this in Section 2.2 as the Principle of the Largest Term; this principle is used repeatedly in what follows and is what gives the theory its special character. Our starting point in this work was to generalize slightly the Ruelle-Lanford definition of entropy, introducing upper and lower functions; it turns out that others have had the same idea: Takahashi [T] and Orey [O] independently defined these functions in the context of ergodic theory. The upper and lower deviation functions are defined and studied in Section 2.3; we have adopted Orey's terminology. In Section 2.4, we introduce the concept of the concentration of a sequence of measures. It turns out that the upper deviation function determines a set on which a sequence of probability measures is eventually concentrated. We have used these results in our work on the equivalence of ensembles [LPS].

A Vague Large Deviation Principle is said to hold when (LDP 1), (LDP 3′) and (LDP 4) are satisfied; VLDPs are studied in Section 3.1. When the upper and lower deviation functions coincide, we call the function which takes their common values the *Ruelle-Lanford function*. This definition, introduced in Section 3.2, is motivated by the Ruelle-Lanford definition of entropy; we reserve the name 'entropy' for those RL-functions which are concave. It turns out that a VLDP is an automatic consequence of the existence of the RL-function; however, if the space X is not locally compact, the rate-function may not be unique; nevertheless, there is a sense in which the RL-function is the optimal rate-function. Some illustrations of the concepts introduced in Sections 2 and 3 are given in the examples in Section 4. We remark in passing that in some applications (in statistical mechanics of lattice spin systems, for example), the space X is compact; in which case, a full theory of large deviations is provided by Sections 2, 3 and part of Section 7.

A Narrow Large Deviation Principle is said to hold when (LDP1), (LDP3) and (LDP 4) are satisfied; NLDPs are studied in Section 5.1. A substitute for compactness of the space X in the theory of large deviations is a condition on the sequence of measures called *exponential tightness*; this is discussed in Section 5.2. In Section 5.3 we return to the subject of concentration of measures, this time under the hypothesis of exponential tightness.

A Large Deviation Principle is said to hold when all four conditions (LDP 1), (LDP 2), (LDP 3) and (LDP 4) are satisfied; LDPs are studied in Section 6.1; in Section 6.2 we give proofs of Varadhan's Theorem and the Contraction Principle.

Up to this point in the paper only the topological and the Borel structures of the space X have been used; when, in addition, the space X has a convex structure, the theory of large deviations becomes substantially more powerful: pressure enters the arena, joining entropy. This is studied in Section 7.

Apart from Lanford's seminal work, referred to above, we have been influenced by Azencott's exposition [A] and by the paper of O'Brien and Vervaat [OV]; as a general reference, we found the book by Deuschel and Stroock [DS] invaluable.

2 The Principle of the Largest Term

2.1 The General Setting

Let X be a Hausdorff topological space. (We require X to be Hausdorff so that the compact sets are closed.) Let $\mathcal{P}(X)$ denote the collection of all subsets of X, $\mathcal{G}(X)$ the open, $\mathcal{F}(X)$ the closed and $\mathcal{K}(X)$ the compact. Let $\mathcal{B}(X)$ denote the collection of Borel subsets of X, the smallest σ-field containing $\mathcal{G}(X)$. We find it useful to adopt the following convention: A always denotes an arbitrary subset, G an open, F a closed, K a compact and B a Borel subset. We write $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ and use the following notation: if $f : X \to \overline{\mathbb{R}}$ is an arbitrary function, we write

$$\sup_{x \in A} f(x) := \sup\{f(x) : x \in A\}; \tag{2.1}$$

by convention, we put

$$\sup_{x \in \emptyset} f(x) := -\infty. \tag{2.2}$$

For a,b in $\overline{\mathbb{R}}$, we define

$$a \vee b := \max\{a, b\}, \tag{2.3}$$

$$a \wedge b := \min\{a, b\}. \tag{2.4}$$

We shall require some elementary facts about some special set-functions called sup-integrals and their relations with upper semicontinuous functions. We state these results without proof; for further details, see Vervaat [Ve].

Let $f : X \to \overline{\mathbb{R}}$ be an arbitrary function; the *sup-integral* of f is the set-function ${}^{\vee}f : \mathcal{P}(X) \to \overline{\mathbb{R}}$ defined by

$${}^{\vee}f[A] := \sup_{x \in A} f(x). \tag{2.5}$$

The sup-integral ${}^{\vee}f$ is an increasing set-function and ${}^{\vee}f[\emptyset] = -\infty$. There is an inverse construction: let c be an increasing set-function defined on a collection of subsets containing $\mathcal{G}(X)$ and satisfying $c[\emptyset] = -\infty$; the *inf-derivative* of c is the function ${}^{\wedge}c : X \to \overline{\mathbb{R}}$ defined by

$${}^{\wedge}c(x) := \inf_{G \ni x} c[G]. \tag{2.6}$$

We will need the following results:

Lemma 2.1 *The sup-integral ${}^{\vee}f : \mathcal{P}(X) \to \overline{\mathbb{R}}$ of a function $f : X \to \overline{\mathbb{R}}$ satisfies*

$${}^{\vee}f[A_1 \cup A_2] = {}^{\vee}f[A_1] \vee {}^{\vee}f[A_2]; \tag{2.7}$$

if, in addition, f is usc, then

$$\inf_{G \supset A} {}^{\vee}f[G] = {}^{\vee}f[A]; \tag{2.8}$$

in general, we have $f \leq {}^{\wedge\vee}f$ with $f = {}^{\wedge\vee}f$ if and only if f is upper semicontinuous (usc).

Lemma 2.2 *Let $c : \mathcal{G}(X) \to \overline{\mathbb{R}}$ be an increasing set-function; then the function ${}^{\wedge}c : X \to \overline{\mathbb{R}}$ is usc; in general, we have $c[G] \geq {}^{\vee\wedge}c[G]$ with equality if and only if $c = {}^{\vee}f$ for some usc function $f : X \to \overline{\mathbb{R}}$.*

Let $f : X \to \overline{\mathbb{R}}$ be an arbitrary function; the *usc-regularization* of f is the usc function $f^\diamond$ defined by

$$f^\diamond := {}^\wedge{}^\vee f. \tag{2.9}$$

A function $g : X \to \overline{\mathbb{R}}$ is lower semicontinuous (lsc) if and only if $-g$ is usc. The *lsc-regularization* of f is the lsc function $f_\diamond$ defined by

$$f_\diamond := -(-f)^\diamond. \tag{2.10}$$

In general, we have

$$f_\diamond \leq f \leq f^\diamond. \tag{2.11}$$

2.2 The Principle of the Largest Term

Let $\{\mathbb{M}_n\}_{n\geq 1}$ be a sequence of positive measures on $\mathcal{B}(X)$ which are *locally finite* (that is, for each x in X, there exists $G_x \in \mathcal{G}(X)$ such that $G_x \ni x$ and $\mathbb{M}[G_x] < \infty$); let $\{V_n\}_{n\geq 1}$ be a *scale*, an increasing sequence of positive real numbers diverging to $+\infty$ as $n \to \infty$. We denote $\{\mathbb{M}_n\}_{n\geq 1}$ by $\mathbb{M}_\circ$ and $\{V_n\}_{n\geq 1}$ by $V_\circ$. We are interested in the asymptotics of $\mathbb{M}_\circ$ on the scale $V_\circ$; we make this precise, beginning with some definitions:

Definition 2.1 *For B in $\mathcal{B}(X)$, put*

$$m_n[B] := \frac{1}{V_n} \ln \mathbb{M}_n[B], \tag{2.12}$$

$$\overline{m}[B] := \limsup_{n\to\infty} m_n[B], \tag{2.13}$$

$$\underline{m}[B] := \liminf_{n\to\infty} m_n[B]. \tag{2.14}$$

The following properties of the set-functions $\underline{m}$, $\overline{m}$ are easily proved. Property (c) is the key to the development; we refer to it as the *Principle of the Largest Term* since it is an abstract version of that principle in statistical mechanics.

Lemma 2.3 *On $\mathcal{B}(X)$, we have*

(a) $-\infty \leq \underline{m}[B] \leq \overline{m}[B] \leq +\infty$;

(b) if $B_1 \subset B_2$, then $\underline{m}[B_1] \leq \underline{m}[B_2]$ and $\overline{m}[B_1] \leq \overline{m}[B_2]$;

(c) $\overline{m}[B_1 \cup B_2] = \overline{m}[B_1] \vee \overline{m}[B_2]$.

Proof: Because of its importance, we give a proof of (c). For $j = 1, 2$, we have

$$\mathbb{M}_n[B_j] \leq \mathbb{M}_n[B_1 \cup B_2] \leq \mathbb{M}_n[B_1] + \mathbb{M}_n[B_2] \tag{2.15}$$

so that

$$\mathbb{M}_n[B_1] \vee \mathbb{M}_n[B_2] \leq \mathbb{M}_n[B_1 \cup B_2] \leq 2\mathbb{M}_n[B_1] \vee \mathbb{M}_n[B_2]; \tag{2.16}$$

it follows that

$$\overline{m}[B_1 \cup B_2] = \limsup_{n\to\infty} (m_n[B_1] \vee m_n[B_2]). \tag{2.17}$$

But for each pair $\{a_n\}_{n\geq 1}, \{b_n\}_{n\geq 1}$ of sequences in $\overline{R}$, we have

$$\limsup_{n\to\infty} (a_n \vee b_n) = (\limsup_{n\to\infty} a_n) \vee (\limsup_{n\to\infty} b_n), \tag{2.18}$$

and (c) follows from (2.17) and (2.18). □

2.3 Upper and Lower Deviation Functions

We introduce two functions on X, the upper and lower deviation functions; we establish the two basic inequalities of the theory, (c) of Lemma 2.4 and (b) of Lemma 2.5. Our terminology agrees with that of Orey [O].

Definition 2.2 *The* upper deviation function $\overline{\mu}$ *and the* lower deviation function $\underline{\mu}$, *are defined on* X *by*

$$\overline{\mu}(x) :=^{\wedge} \overline{m}(x) := \inf_{G \ni x} \overline{m}[G], \tag{2.19}$$

$$\underline{\mu}(x) :=^{\wedge} \underline{m}(x) := \inf_{G \ni x} \underline{m}[G]. \tag{2.20}$$

Remark: The definitions of $\overline{\mu}(x)$ and $\underline{\mu}(x)$ use the family of open neighbourhoods of x; because, by (*b*) of Lemma 2.3, the set-functions $\overline{m}$ and $\underline{m}$ are increasing, this family may be replaced by any base of Borel neighbourhoods of x.

Lemma 2.4 *The upper and lower deviation functions* $\overline{\mu}$ *and* $\underline{\mu}$ *of the pair* $(\mathbb{M}_\circ, V_\circ)$ *satisfy:*

(a) $\underline{\mu}(x) \leq \overline{\mu}(x)$;

(b) $\underline{\mu}$ *and* $\overline{\mu}$ *are usc functions;*

(c) $\underline{m}[G] \geq \sup_{x \in G} \underline{\mu}(x)$;

(d) $\overline{m}[G] \geq \sup_{x \in G} \overline{\mu}(x)$.

Proof: Property (*a*) follows from (*a*) of Lemma 2.4, the corresponding property of $\underline{m}$ and $\overline{m}$. Properties (b), (c) and (d) are consequences of properties of inf-derivatives of set-functions, see Lemma 2.2. □

The lower bound (d) for $\overline{m}$ on open sets is rarely used; of greater importance is the upper bound for $\overline{m}$ on compact sets, (*b*) of Lemma 2.5, which is a consequence of the Principle of the Largest Term:

Lemma 2.5 *Let* K *be a compact subset of* X; *then*

(a) if K *is nonempty and if there is a real number* a *such that* $a > {}^{\vee}\overline{\mu}[K]$, *then there exists* G *open such that* $G \supset K$ *and* $\overline{m}[G] < a$.

(b)

$$\overline{m}[K] \leq {}^{\vee}\overline{\mu}[K] := \sup_{x \in K} \overline{\mu}(x). \tag{2.21}$$

Proof: Fix $a > {}^{\vee}\overline{\mu}[K]$ then $a > \overline{\mu}(x)$ for all x in K. Since

$$\overline{\mu}(x) := \inf_{G \ni x} \overline{m}[G], \tag{2.22}$$

for each $x \in K$ there exists an open set G_x such that

$$\overline{m}[G_x] < a. \tag{2.23}$$

Since K is compact, there exists a finite subcover $\{G_{x_1}, ..., G_{x_d}\}$ of the cover $\{G_x : x \in K\}$ of K. Put $G = G_{x_1} \cup ... \cup G_{x_d}$; then

$$\overline{m}[G] = \overline{m}[G_{x_1}] \vee ... \vee \overline{m}[G_{x_d}] < a. \tag{2.24}$$

This establishes (a). To prove (b), note that if ${}^{\vee}\overline{\mu}[K] = +\infty$ or $K = \emptyset$, then (b) is trivially true. Assume ${}^{\vee}\overline{\mu}[K] < \infty$ and $K \neq \emptyset$ and choose $a > {}^{\vee}\overline{\mu}[K]$; using (a) and the fact that $\overline{m}$ is an increasing set-function, we have $\overline{m}[K] < a$ and (b) follows on taking the infimum over $\{a : a > {}^{\vee}\overline{\mu}[K]\}$. □

Let $g : X \to \mathbb{R}$ be a locally bounded measurable function; we define a new sequence $\mathbb{M}_\circ^g$ of positive locally finite measures on $\mathcal{B}(X)$ by

$$\mathbb{M}_n^g[B] := \int_B e^{V_n g(x)}\, \mathbb{M}_n[dx]. \tag{2.25}$$

Define for the pair $(\mathbb{M}_\circ^g, V_\circ)$ set-functions m_n^g, $\overline{m}^g$, $\underline{m}^g$ by

$$m_n^g[B] := \frac{1}{V_n} \ln \mathbb{M}_n^g[B], \tag{2.26}$$

$$\overline{m}^g[B] := \limsup_{n\to\infty} m_n^g[B], \tag{2.27}$$

$$\underline{m}^g[B] := \liminf_{n\to\infty} m_n^g[B]. \tag{2.28}$$

Define also the upper and lower deviation functions $\overline{\mu}^g$ and $\underline{\mu}^g$ on X by

$$\overline{\mu}^g(x) :={}^{\wedge}\overline{m}^g(x) := \inf_{G\ni x} \overline{m}^g[G], \tag{2.29}$$

$$\underline{\mu}^g(x) :={}^{\wedge}\underline{m}^g(x) := \inf_{G\ni x} \underline{m}^g[G]. \tag{2.30}$$

Lemma 2.6 *Let $g : X \to \mathbb{R}$ be a locally bounded measurable function; the upper and lower deviation functions for the pair $(\mathbb{M}_\circ^g, V_\circ)$ are related to those for $(\mathbb{M}_\circ, V_\circ)$ by*

$$\overline{\mu}(x) + g_\diamond(x) \leq \overline{\mu}^g(x) \leq \overline{\mu}(x) + g^\diamond(x) \tag{2.31}$$

$$\underline{\mu}(x) + g_\diamond(x) \leq \underline{\mu}^g(x) \leq \underline{\mu}(x) + g^\diamond(x). \tag{2.32}$$

Proof: We give the proof of the first inequality in (2.31): fix x and an open neighbourhood G of x; then

$$m_n^g[G] \geq m_n[G] + \inf_{y\in G} g(y) \geq m_n[G] + \inf_{y\in G'} g(y) \tag{2.33}$$

for all $G' \supset G \ni x$, so that

$$\overline{m}^g[G] \geq \overline{m}[G] + \inf_{y\in G'} g(y). \tag{2.34}$$

Taking the infimum over $G \ni x$ and then the supremum over $G' \ni x$, we get

$$\overline{\mu}^g(x) \geq \overline{\mu}(x) + \sup_{G'\in x} \inf_{y\in G'} g(y) = \overline{\mu}(x) + g_\diamond(x). \tag{2.35}$$

□

Corollary 2.1 *Let $g : X \to \mathbb{R}$ be continuous; then the upper and lower deviation functions for the pair $(\mathbb{M}_\circ^g, V_\circ)$ are related to those for the pair $(\mathbb{M}_\circ, V_\circ)$ by*

$$\overline{\mu}^g(x) = \overline{\mu}(x) + g(x), \tag{2.36}$$

$$\underline{\mu}^g(x) = \underline{\mu}(x) + g(x). \tag{2.37}$$

2.4 Concentration of Measures on Compact Spaces

Our first application of the Principle of the Largest Term is to the investigation of the behaviour of the measures $\mathbb{M}_n$ as $n \to \infty$. The upper deviation function determines a set on which the measures are eventually concentrated (Theorem 2.2); how useful this is depends on how well we have chosen the scale V_o. The requirement that X be compact will be removed when we return to the topic in Section 5.3.

Definition 2.3 *Let $\mathbb{M}_o$ be a sequence of probability measures on $\mathcal{B}(X)$; we say that $\mathbb{M}_o$ is* eventually concentrated *on a set A if, for each open neighbourhood G of A, we have*

$$\lim_{n\to\infty} \mathbb{M}_n[G] = 1. \tag{2.38}$$

Remark: A sequence $\mathbb{M}_o$ of probability measures on $\mathcal{B}(X)$ *converges narrowly* to a measure $\mathbb{M}$ if and only if $\mathbb{M}_n[f] \to \mathbb{M}[f]$ for every bounded continuous function f on X. In the probability literature, 'narrow convergence' is called 'weak convergence'; use of the latter term can cause confusion in the context of functional analysis, so we avoid it. Our usage here is consistent with that in section 5 where we define narrow convergence for sequences of increasing set functions.

The next theorem implies the following statement: If $A = \{x_o\}$ and $\mathbb{M}_o$ is eventually concentrated on A, then $\mathbb{M}_o$ converges narrowly to the Dirac measure δ_{x_0}; that is, $\mathbb{M}_o$ obeys a weak law of large numbers.

Theorem 2.1 *Let $\mathbb{M}_o$ be a sequence of probability measures on $\mathcal{B}(X)$ and let B be a Borel subset of X such that*

$$\lim_{n\to\infty} \mathbb{M}_n[B] = 1. \tag{2.39}$$

Suppose that $\mathbb{M}_o$ is eventually concentrated on a subset N of X. If $f : X \to \overline{\mathbb{R}}$ is lower semicontinuous and bounded below on B, then

$$\inf_{x\in N} f(x) \le \liminf_{n\to\infty} \int_B f(x)\mathbb{M}_n[dx]; \tag{2.40}$$

if $g : X \to \overline{\mathbb{R}}$ is upper semicontinuous and bounded above on B, then

$$\limsup_{n\to\infty} \int_B g(x)\mathbb{M}_n[dx] \le \sup_{x\in N} g(x). \tag{2.41}$$

Proof: We give a proof of the upper bound; the lower bound can be deduced by applying the upper bound with $-f$ in place of g. For each $n \ge 1$ and any G, we have

$$\int_B g(x)\mathbb{M}_n[dx] \le (\sup_{x\in G} g(x))\mathbb{M}_n[B \cap G] + (\sup_{x\in B} g(x))\mathbb{M}_n[B\backslash G]. \tag{2.42}$$

For any open neighbourhood G of N, we have

$$\lim_{n\to\infty} \mathbb{M}_n[G] = 1; \tag{2.43}$$

since

$$\lim_{n\to\infty} \mathbb{M}_n[B] = 1, \tag{2.44}$$

it follows that, given $\epsilon > 0$, we have

$$\mathbb{M}_n[B \cap G] \geq 1 - \epsilon, \tag{2.45}$$

and

$$\mathbb{M}_n[B \backslash G] \leq \epsilon, \tag{2.46}$$

for all n sufficiently large. Thus we have

$$\limsup_{n \to \infty} \int_B g(x) \mathbb{M}_n[dx] \leq \sup_{x \in G} g(x) \tag{2.47}$$

for every open neighbourhood G of N. Using the upper semicontinuity of g and (2.8) of Lemma 2.1, we have

$$\inf_{G \supset N} \sup_{x \in G} g(x) = \sup_{x \in N} g(x). \tag{2.48}$$

□

Theorem 2.2 *Let X be compact, let $\mathbb{M}_o$ be a sequence of probability measures on $\mathcal{B}(X)$. Let V_o be a scale and $\overline{\mu}$ the upper deviation function for the pair $(\mathbb{M}_o, V_o)$; then*

(a) the set

$$N_{\overline{\mu}} := \{ x \in X : \overline{\mu}(x) = 0 \} \tag{2.49}$$

is a non-empty compact subset;

(b) the sequence $\mathbb{M}_o$ is eventually concentrated on the set $N_{\overline{\mu}}$.

Proof: Since $\mathbb{M}_n[X] = 1$ for all $n \geq 1$, we have

$$\overline{m}[X] = 0. \tag{2.50}$$

Since X is open, we have

$$\overline{\mu}(x) \leq \overline{m}[X] = 0 \tag{2.51}$$

so that

$$\overline{\mu}[X] \leq 0. \tag{2.52}$$

Since X is compact, we have

$$\overline{\mu}[X] \geq \overline{m}[X] = 0; \tag{2.53}$$

hence

$$\overline{\mu}[X] = 0. \tag{2.54}$$

Since $\overline{\mu}$ is usc, by (*b*) of Lemma 2.4, and X is compact by hypothesis, the set $N_{\overline{\mu}}$ is non-empty and closed, hence compact. This proves (a).

Let G be an open neighbourhood of $N_{\overline{\mu}}$; then $X \backslash G$ is compact, since X is, so that $\overline{\mu}[X \backslash G]$ is attained on $X \backslash G$ since $\overline{\mu}$ is usc. But $\overline{\mu}(x) \leq 0$, so that $\overline{\mu}[X \backslash G] < 0$, since $N_{\overline{\mu}} \cap X \backslash G = \emptyset$. Using (b) of Lemma 2.5, we have

$$\limsup_{n \to \infty} \frac{1}{V_n} \ln \; M_n[X \backslash G] \leq \delta < 0 \tag{2.55}$$

so that

$$\lim_{n \to \infty} \; M_n[X \backslash G] = 0 \tag{2.56}$$

and hence

$$\lim_{n \to \infty} \mathbb{M}_n[G] = 1. \tag{2.57}$$

□

3 Vague Large Deviation Principles and Ruelle-Lanford Functions

3.1 Vague Large Deviation Principles

Definition 3.1 *A pair* $(\mathbb{M}_o, V_o)$ *obeys a* vague large deviation principle *(VLDP) if there exists an usc function* $s : X \to \overline{\mathbb{R}}$ *such that*

$$\overline{m}[K] \leq \sup_{x \in K} s(x) \ , \ K \in \mathcal{K}(X), \tag{3.1}$$

$$\underline{m}[G] \geq \sup_{x \in G} s(x) \ , \ G \in \mathcal{G}(X). \tag{3.2}$$

The function s *is called the* rate-function.

Remark: We follow O'Brien and Vervaat [OV] in using this terminology. Our usage differs from that of Deuschel and Stroock [DS] in two respects: the measures they consider are always probability measures; what we call a 'vague' large deviation principle, they call a 'weak' large deviation principle.

Definition 3.2 *Let* $\mathcal{C}[X]$ *denote the space of increasing set-functions* $c : \mathcal{B}(X) \to \overline{\mathbb{R}}$ *which satisfy* $c[\emptyset] = -\infty$*; the* vague topology *on* $\mathcal{C}[X]$ *is the topology generated by the sub-base consisting of the following families:*

$$\{c : c[G] > x\} \ , \ x \in \mathbb{R} \ , \ G \in \mathcal{G}(X), \tag{3.3}$$

$$\{c : c[K] < x\} \ , \ x \in \mathbb{R} \ , \ K \in \mathcal{K}(X). \tag{3.4}$$

A sequence $\{c_n\}_{n \geq 1}$ converges to c in $\mathcal{C}[X]$ in the vague topology *if and only if*

$$\limsup_{n \to \infty} c_n[K] \leq c[K] \ , \ K \in \mathcal{K}(X), \tag{3.5}$$

$$\liminf_{n \to \infty} c_n[G] \geq c[G] \ , \ G \in \mathcal{G}(X). \tag{3.6}$$

Remark: The definition of a VLDP can be re-stated as follows: the pair $(\mathbb{M}_o, V_o)$ obeys a VLDP with rate-function s if and only if there exists an usc function $s : X \to \overline{\mathbb{R}}$ such that the sequence $\{m_n\}_{n \geq 1}$ of increasing set-functions, defined at (2.12), converges in the vague topology to the sup-integral ${}^{\vee}s$.

There is a sense in which the upper and lower deviation functions yield optimal bounds: Lemma 3.1 shows that $\underline{\mu}$ is the greatest function with property (3.2); Lemma 3.2 shows that, provided the space X is locally compact, $\overline{\mu}$ is the least usc function with property (3.1).

Lemma 3.1 *Let* X *be a Hausdorff topological space . Suppose that* $s : X \to \overline{\mathbb{R}}$ *is such that*

$$\underline{m}[G] \geq \sup_{x \in G} s(x) \tag{3.7}$$

for all $G \in \mathcal{G}(X)$ *; then*

$$\underline{m}[G] \geq \sup_{x \in G} s^{\diamond}(x) \tag{3.8}$$

for all $G \in \mathcal{G}(X)$ *and*

$$s^{\diamond}(x) \leq \underline{\mu}(x) \tag{3.9}$$

for all $x \in X$.

Proof: By (2.11), we have

$$s(x) \leq s^{\diamond}(x) \tag{3.10}$$

for all $x \in X$, but

$$s^{\diamond}(x) := \inf_{G \ni x} \sup_{y \in G} s(y) \leq \inf_{G \ni x} \underline{m}[G] =: \underline{\mu}(x) \tag{3.11}$$

so that

$$\sup_{x \in G} s^{\diamond}(x) \leq \sup_{x \in G} \underline{\mu}(x) \leq \underline{m}[G], \tag{3.12}$$

by (c) of Lemma 2.4 . □

Lemma 3.2 *Suppose that X is a locally compact Hausdorff topological space and that $r : X \to \overline{\mathbb{R}}$ is such that*

$$\overline{m}[K] \leq \sup_{x \in K} r(x) \tag{3.13}$$

for all $K \in \mathcal{K}(X)$; then

$$\overline{m}[K] \leq \sup_{x \in K} r^{\diamond}(x) \tag{3.14}$$

for all $K \in \mathcal{K}(X)$ and

$$r^{\diamond}(x) \geq \overline{\mu}(x) \tag{3.15}$$

for all $x \in K$.

Proof: Since X is a locally compact Hausdorff topological space, the family of compact neighbourhoods of a point x is a base of neighbourhoods at x; by the remark preceding Lemma 2.4, we can use such a base to compute $\overline{\mu}(x)$. Hence we have

$$\overline{\mu}(x) = \inf_{K \ni x} \overline{m}[K] \leq \inf_{K \ni x} \sup_{y \in K} r(y) = r^{\diamond}(x). \tag{3.16}$$

By (2.11), we have

$$\overline{m}[K] \leq \sup_{x \in K} r(x) \leq \sup_{x \in K} r^{\diamond}(x). \tag{3.17}$$

□

Corollary 3.1 *Let X be a locally compact Hausdorff topological space; suppose that the pair (M_o, V_o) obeys a VLDP with rate-function s, then the rate-function is unique and is equal to*

$$s(x) = \mu(x) := \underline{\mu}(x) = \overline{\mu}(x) \tag{3.18}$$

for all $x \in X$.

Proof: By definition, the function s is usc so that

$$s(x) = s^{\diamond}(x); \tag{3.19}$$

using Lemma 3.1 and Lemma 3.2, we have

$$s(x) \leq \underline{\mu}(x) \leq \overline{\mu}(x) \leq s(x). \tag{3.20}$$

□

Remark: When the topological space X is not locally compact, the vague topology in the space of increasing set-functions is not Hausdorff: there may be more than one rate-function for a VLDP. We mention without proof the following result.

Define the set-function $\overline{m}_{\mathfrak{h}}$ as follows: on $\mathcal{G}(X)$, $\overline{m}_{\mathfrak{h}}[G] := \sup_{K\subset G} \overline{m}[K]$; then on $\mathcal{K}(X)$, $\overline{m}_{\mathfrak{h}}[K] := \inf_{G\supset K} \overline{m}_{\mathfrak{h}}[G]$; finally, extend the function to every A by setting $\overline{m}_{\mathfrak{h}}[A] := \sup_{K\subset A} \overline{m}_{\mathfrak{h}}[K]$. Let X be a Hausdorff topological space; there exists a VLDP for the pair $(M_\circ, V_\circ)$ if and only if

$$\overline{m}[K] \;\leq\; \underline{m}[G] \tag{3.21}$$

for all K and for all G with $K \subset G$, in which case every usc function s satisfying

$$\hat{}\overline{m}_{\mathfrak{h}}(x) \;\leq\; s(x) \;\leq\; \hat{}\underline{m}(x) \;=\; \underline{\mu}(x) \tag{3.22}$$

is a rate-function; see example (h) in Section 4.

3.2 Ruelle-Lanford Functions

In probability theory as well as in statistical mechanics, there are important cases where it can be verified directly (by a sub-additivity argument, for example) that the upper and lower deviation functions coincide. This motivates the following definition:

Definition 3.3 *A pair* $(\mathbb{M}_\circ, V_\circ)$ *has a* Ruelle-Lanford function *(R-L function)* μ *if*

$$\overline{\mu}(x) = \underline{\mu}(x) \tag{3.23}$$

for all x *in* X*; in which case we put*

$$\mu(x) := \overline{\mu}(x) = \underline{\mu}(x). \tag{3.24}$$

The next theorems summarize the results of the previous section when the R-L function exists.

Theorem 3.1 *Let* X *be a Hausdorff topological space,* $\mathbb{M}_\circ$ *a sequence of locally finite positive measures on* $\mathcal{B}(X)$ *and* $V_\circ$ *a scale. If the pair* $(\mathbb{M}_\circ, V_\circ)$ *has a RL-function* μ*, then it satisfies a VLDP with rate-function* μ*:*

$$\overline{m}[K] \leq \sup_{x\in K} \mu(x)\ ,\ K \in \mathcal{K}(X), \tag{3.25}$$

$$\underline{m}[G] \geq \sup_{x\in G} \mu(x)\ ,\ G \in \mathcal{G}(X). \tag{3.26}$$

Theorem 3.2 *Let* X *be a Hausdorff topological space,* $g : X \to \mathbb{R}$ *a continuous function,* $\mathbb{M}_\circ$ *a sequence of locally finite positive measures on* $\mathcal{B}(X)$ *and* $V_\circ$ *a scale. If the pair* $(\mathbb{M}_\circ, V_\circ)$ *has a RL-function* μ*, then the pair* $(\mathbb{M}_\circ^g, V_\circ)$ *has a RL-function* μ^g *and* μ^g *is given by*

$$\mu^g(x) = \mu(x) + g(x)\ ,\ x \in X. \tag{3.27}$$

The following result is a special case of Corollary 6.1.

Corollary 3.2 *Let* X *be compact and* $g : X \to \mathbb{R}$ *continuous; if the pair* $(\mathbb{M}_\circ, V_\circ)$ *has a RL-function* μ*, then*

$$\lim_{n\to\infty} \frac{1}{V_n} \ln \int_X e^{V_n g(x)} \mathbb{M}_n[dx] = \sup_{x\in X}(\mu(x) + g(x)). \tag{3.28}$$

Proof: By Theorem 3.2, the pair $(\mathbb{M}_\circ^g, V_\circ)$ has a RL-function μ^g. Since X is compact, we have

$$\overline{m}^g[X] \leq \sup_{x\in X} \mu^g(x); \tag{3.29}$$

since X is open, we have

$$\underline{m}^g[X] \geq \sup_{x\in X} \mu^g(x); \tag{3.30}$$

hence we have

$$\overline{m}^g[X] = \underline{m}^g[X] = \sup_{x\in X} \mu^g(x). \tag{3.31}$$

It follows that

$$\lim_{n\to\infty} \frac{1}{V_n} \ln \int_X e^{V_n g(x)} \mathbb{M}_n[dx] = \sup_{x\in X} \mu^g(x) = \sup_{x\in X}(\mu(x) + g(x)). \tag{3.32}$$

□

Remark: Versions of Varadhan's Theorem, such as Corollary 3.2 , are used in two directions:

- the investigation of the asymptotic behaviour of the integral

$$\int_X e^{V_n g(x)} \mathbb{M}_n[dx] \tag{3.33}$$

  is converted into a variational problem;

- suppose E is a vector space and the limit on the left-hand side of (3.28) can be evaluated explicitly in the case where g is a linear functional on E; the right-hand side is then the conjugate of $-\mu$ and an expression for μ can be obtained by taking the conjugate of both sides, provided we can show that μ is concave.

The final result of this section examines when a VLDP exists for the image measures $\mathbb{M}_n \circ \varphi^{-1}$. Theorem 3.3 below is an example of a *contraction principle.*

Theorem 3.3 *Let X and X' be locally compact Hausdorff topological spaces and let $\varphi : X \to X'$ be continuous and such that the inverse image of a compact subset of X' is a compact subset of X. Suppose that the pair $(\mathbb{M}_\circ, V_\circ)$ obeys a VLDP with rate-function $s : X \to \overline{\mathbb{R}}$; then the pair $(\mathbb{M}'_\circ, V_\circ)$, where $\mathbb{M}'_n := \mathbb{M}_n \circ \varphi^{-1}$, obeys a VLDP with rate-function $\sigma : X' \to \overline{\mathbb{R}}$, where*

$$\sigma(x') := \sup_{\{y:\varphi(y)=x'\}} s(y); \tag{3.34}$$

σ is the RL-function for $(\mathbb{M}'_\circ, V_\circ)$.

Proof: We have

$$\overline{m}'[K] = \overline{m}[\varphi^{-1}K] \leq \sup_{x\in\varphi^{-1}K} s(x) = \sup_{x'\in K} \sup_{\{y:\varphi(y)=x'\}} s(y) = \sup_{x'\in K} \sigma(x') \tag{3.35}$$

for all $K \in \mathcal{K}(X')$. A similar calculation yields

$$\underline{m}'[G] \geq \sup_{x'\in G} \sigma(x') \tag{3.36}$$

for all $G \in \mathcal{G}(X')$. It remains to show that σ is usc. Below, K' always denotes a compact neighbourhood of $x' \in X'$; since the space X' is locally compact, a function f is usc if and only if

$$f(x') = \inf_{K' \ni x'} \sup_{y' \in K'} f(y'). \tag{3.37}$$

Fix $x' \in X'$ and put

$$\mathcal{K}_0 = \{K' \in \mathcal{K}(X') : K' \ni x'\}. \tag{3.38}$$

We have

$$\sigma^\diamond(x') = \inf_{K' \in \mathcal{K}_o} \sup_{y' \in K'} \sigma(y') = \inf_{K' \in \mathcal{K}_0} \sup_{y \in \varphi^{-1} K'} s(y). \tag{3.39}$$

Put

$$F := \{x \in X : s(x) \geq \sigma^\diamond(x')\}; \tag{3.40}$$

If $\sigma^\diamond(x') = -\infty$, then $\sigma^\diamond(x') = \sigma(x')$ by (2.11); let $\sigma^\diamond(x') > -\infty$. F is closed because s is usc and non-empty because $\varphi^{-1}K'$ is compact. Since X' is Hausdorff, we have

$$\bigcap_{K' \in \mathcal{K}_o} K' = \{x'\} \tag{3.41}$$

so that

$$\bigcap_{K' \in \mathcal{K}_o} \varphi^{-1} K' = \varphi^{-1}\{x'\}. \tag{3.42}$$

Since $\mathcal{K}_o$ is closed under finite intersections, so is $\varphi^{-1}\mathcal{K}_o$; let $\mathcal{K}_F := \varphi^{-1}\mathcal{K}_o \cap F$. If $K' \in \mathcal{K}_o$, then

$$\sup_{y \in \varphi^{-1} K'} s(y) \geq \sigma^\diamond(x') \tag{3.43}$$

so that if $K \in \mathcal{K}_F$, then K is non-empty: $\mathcal{K}_F$ is a family of compact subsets having the finite intersection property; it follows that

$$\bigcap_{K \in \mathcal{K}_F} K = \varphi^{-1}\{x'\} \cap F \neq \emptyset. \tag{3.44}$$

and hence $\sigma(x') \geq \sigma^\diamond(x')$ so that σ is usc. □

4 Examples

In this section, we consider several concrete situations. Let $X = \mathbb{R}$ with the usual topology; choose $V_n = n$ for the scale. We compute the RL-function in some elementary cases:

(a) $\mathbb{M}_n$ is Lebesgue measure for all n; then $\mu(x) = 0$.

(b) $\mathbb{M}_n$ is absolutely continuous with respect to Lebesgue measure with density $\rho_n(x) = \exp(-nx^2/2)$; then $\mu(x) = -x^2/2$.

(c) $\mathbb{M}_n$ the probability measure got by normalizing the measure in (b); now, $\rho_n(x) = (n/2\pi)^{\frac{1}{2}} \exp(-nx^2/2)$. The RL-function is again given by $\mu(x) = -x^2/2$. However, knowing that the $\mathbb{M}_n$ are probability measures gives us useful information. Indeed, μ attains its maximum at $x = 0$, hence $\mathbb{M}_n$ converges to δ_0.

(d) Let δ_x be the Dirac measure at x and let $\mathbb{M}_n = e^{an}\delta_1 + e^{bn}\delta_n$; then

$$\mu(x) = \begin{cases} a, & \text{if } x = 1 \\ -\infty, & \text{otherwise.} \end{cases} \tag{4.1}$$

(e) $\mathbb{M}_n$ is Lebesgue measure for all $n \geq 1$. Let $X' = [-2, 2]$ and define $\phi : X \to X'$ by $\phi(x) := \tanh x$; then ϕ is an injective local homeomorphism. On X', define $\mathbb{M}'_n := \mathbb{M}_n \circ \phi^{-1}$; then the RL-function μ' for the pair $(\mathbb{M}'_o, V_o)$ is given by

$$\mu'(x) = \begin{cases} -\infty, & \text{if } -2 \leq x < -1 \text{ or } 1 < x \leq 2, \\ \infty, & \text{if } x = -1 \text{ or } x = 1, \\ 0, & \text{if } -1 < x < 1. \end{cases} \tag{4.2}$$

Theorem 3.3 does not hold here. Notice that the function

$$f(x') \;:=\; \sup\ \{\mu(y) : \phi(y) = x'\} \tag{4.3}$$

is not usc and $f(-1) = f(+1) = -\infty$.

(f) Let $X = (0,1) \cup (2,3)$ with the usual topology and let $X' = (0,1)$ with the usual topology; define $\phi : X \to X'$ by

$$\phi(x) := \begin{cases} x, & \text{if } x \in (0,1), \\ x/4, & \text{if } x \in (2,3). \end{cases} \tag{4.4}$$

For $n \geq 2$, put $V_n := n$ and

$$\mathbb{M}_n := e^{bn}\delta_{2+\frac{1}{n}}; \tag{4.5}$$

then $\mu(x) = -\infty$ for all x. Define $\mathbb{M}'_n$ by

$$\mathbb{M}'_n := \mathbb{M}_n \circ \phi^{-1}; \tag{4.6}$$

then

$$\mu'(x') = \begin{cases} b & \text{if } x' = \frac{1}{2}, \\ -\infty & \text{otherwise.} \end{cases} \tag{4.7}$$

Theorem 3.3 fails to hold. Here the function ϕ is a surjective local homeomorphism but, in both examples (e) and (f), there are compact subsets of X' whose inverse images in X are not compact.

(g) Let $X = \mathbb{R}$ equipped with the Sorgenfrey topology [K] which we denote by τ (this topology is also known as the half-open topology). A base for this topology is formed by all intervals of the form [a,b) with $a < b$. The τ-topology is strictly finer than the usual topology; nevertheless, both topologies generate the same Borel structure. In the τ-topology, the sets $(-\infty, a)$, [a,b) and [b,∞) are both open and closed; a subset K is *τ-compact* if and only if it is compact in the usual topology and contains no strictly increasing infinite sequence. The set $\{0\} \cup \{\frac{1}{n} : n \geq 1\}$ is τ-compact, but the set $\{1 - \frac{1}{n} : n \geq 1\} \cup \{1\}$ is not. The τ-topology is not metrizable and (X, τ) is not locally compact. Let $V_n := n$ and let

$$\mathbb{M}_n := \delta_{\frac{1}{n}} + \delta_{1-\frac{1}{n}}; \tag{4.8}$$

then the RL-function is given by

$$\mu_\tau(x) = \begin{cases} 0 & \text{if } x = 0\ , \\ -\infty & \text{otherwise.} \end{cases} \tag{4.9}$$

With the usual topology, the RL-function μ is given by

$$\mu(x) = \begin{cases} 0 & \text{if } x = 0 \text{ or } x = 1 \ , \\ -\infty & \text{otherwise.} \end{cases} \tag{4.10}$$

(h) Same example as (g), but with $\mathbb{M}_n$ the Lebesgue measure for all n. Then

$$\hat{\underline{m}}(x) = \mu(x) = 0, \tag{4.11}$$

and (see remark at the end of section 3.1)

$$\hat{\overline{m}}_{\natural}(x) = -\infty. \tag{4.12}$$

Thus any usc non-positive function s is a rate-function.

5 Narrow Large Deviation Principles and Exponential Tightness

5.1 Narrow Large Deviation Principles

Definition 5.1 *A pair* $(\mathbb{M}_o, V_o)$ *obeys a* narrow large deviation principle *(NLDP) if there exists an usc function* $s : X \to \overline{\mathbb{R}}$ *such that*

$$\overline{m}[F] \ \leq \ \sup_{x \in F} s(x) \ , \ F \in \mathcal{F}(X), \tag{5.1}$$

$$\underline{m}[G] \ \geq \ \sup_{x \in G} s(x) \ , \ G \in \mathcal{G}(X). \tag{5.2}$$

The function s is called the rate-function.

Definition 5.2 *Let $\mathcal{C}[X]$ denote the space of increasing set-functions $c : \mathcal{B}(X) \to \overline{\mathbb{R}}$ which satisfy $c[\phi] = -\infty$; the* narrow topology *on $\mathcal{C}[X]$ is the topology generated by the sub-base consisting of the following families:*

$$\{c : c[G] > x\} \ , \ x \in \mathbb{R} \quad G \in \mathcal{G}(X), \tag{5.3}$$

$$\{c : c[F] < x\} \ , \ x \in \mathbb{R} \ \ F \in \mathcal{F}(X). \tag{5.4}$$

A sequence $\{c_n\}_{n \geq 1}$ converges to c in $\mathcal{C}[X]$ in the narrow topology *if and only if*

$$\limsup_{n \to \infty} c_n[F] \leq c[F] \ , \ F \in \mathcal{F}(X), \tag{5.5}$$

$$\liminf_{n \to \infty} c_n[G] \geq c[G] \ , \ G \in \mathcal{G}(X). \tag{5.6}$$

Remark: The definition of an NLDP can be re-stated as follows: the pair $(\mathbb{M}_o, V_o)$ obeys an NLDP with a rate-function s if and only if there exists an usc function s such that the sequence $\{m_n\}_{n \geq 1}$ of increasing set-functions, defined at (2.12), converges in the narrow topology to the sup-integral $^{\vee}s$.

Lemma 5.1 is the analogue of Lemma 3.2 for closed sets.

Lemma 5.1 *Let X be a regular topological space. Suppose that $r : X \to \overline{\mathbb{R}}$ is such that*

$$\overline{m}[F] \leq \sup_{x \in F} r(x) \tag{5.7}$$

for all $F \in \mathcal{F}(X)$; then

$$\overline{m}[F] \leq \sup_{x \in F} r^{\diamond}(x) \tag{5.8}$$

for all $F \in \mathcal{F}(X)$ and

$$r^{\diamond}(x) \geq \overline{\mu}(x) \tag{5.9}$$

for all $x \in X$.

Proof: Since X is regular, the family of closed neighbourhoods of a point x is a base of neighbourhoods at x; by the remark preceding Lemma 2.4, we can use such a base to compute $\overline{\mu}(x)$. Hence we have

$$\overline{\mu}(x) = \inf_{F \ni x} \overline{m}[F] \leq \inf_{F \ni x} \sup_{y \in F} r(y) = r^{\diamond}(x). \tag{5.10}$$

By (2.11), we have

$$\overline{m}[F] \leq \sup_{x \in F} r(x) \leq \sup_{x \in F} r^{\diamond}(x). \tag{5.11}$$

□

Corollary 5.1 *Let X be a regular topological space; suppose that the pair $(\mathbb{M}_{\circ}, V_{\circ})$ obeys an NLDP with rate-function s, then the pair has a RL-function μ and*

$$s(x) = \mu(x) \tag{5.12}$$

for all $x \in X$.

Proof: Same proof as Corollary 3.1. □

Theorem 5.1 *Let X be a Hausdorff topological space and let the pair $(\mathbb{M}_{\circ}, V_{\circ})$ obey an NLDP with RL-function μ as rate-function. If $g : X \to \mathbb{R}$ is continuous and both conditions*

$$\limsup_{a \to \infty} \overline{m}^{g}[\{x : g(x) > a\}] = -\infty \tag{5.13}$$

$$\limsup_{a \to -\infty} \overline{m}^{g}[\{x : g(x) < a\}] = -\infty \tag{5.14}$$

hold, then the pair $(\mathbb{M}_{\circ}^{g}, V_{\circ})$ obeys an NLDP with RL-function $\mu^{g} = \mu + g$ as rate-function. In particular, we have

$$\overline{m}^{g}[X] = \underline{m}^{g}[X] = \sup_{x \in X}\{\mu(x) + g(x)\}. \tag{5.15}$$

Proof: We know from Theorem 3.2 that the pair $(\mathbb{M}_{\circ}^{g}, V_{\circ})$ obeys a VLDP with RL-function $\mu^{g} = \mu + g$. We must prove the upper bound for closed sets. Let F be an arbitrary closed set; for any $a, b \in \overline{\mathbb{R}}$, $a < b$, we put

$$L(a, b) := \{x \in X : a \leq g(x) \leq b\}. \tag{5.16}$$

Then

$$\overline{m}^{g}[F] = \overline{m}^{g}[F \cap L(-\infty, a)] \vee \overline{m}^{g}[F \cap L(a, b)] \vee \overline{m}^{g}[F \cap L(b, \infty)]; \tag{5.17}$$

if $\overline{m}^g[F] = -\infty$, then there is nothing to prove; if $\overline{m}^g[F] > -\infty$, then by (5.13) and (5.14) we can choose $a, b \in \mathbb{R}$ such that

$$\overline{m}^g[F] = \overline{m}^g[F \cap L(a,b)]. \tag{5.18}$$

Given ϵ, we divide the interval $[a,b]$ into a finite number of closed intervals $[a_i, b_i]$ of length at most ϵ. On $L(a_i, b_i)$, $\mu(x) \leq \mu(x) + g(x) - a_i$ and $g(x) \leq b_i$; therefore

$$\begin{aligned} \overline{m}^g[F \cap L(a_i,b_i)] &\leq \overline{m}[F \cap L(a_i,b_i)] + b_i \\ &\leq {}^{\vee}\mu[F \cap L(a_i,b_i)] + b_i \\ &\leq {}^{\vee}\mu^g[F \cap L(a_i,b_i)] + \epsilon. \end{aligned} \tag{5.19}$$

From (c) of Lemma 2.3 we get

$$\overline{m}^g[F] \leq {}^{\vee}\mu^g[F \cap L(a,b)] + \epsilon \leq {}^{\vee}\mu^g[F] + \epsilon; \tag{5.20}$$

since ϵ is arbitrary, the theorem follows . □.

We end this subsection by stating a contraction principle and giving in the remark after Theorem 5.2 a necessary and sufficient condition for an NLDP.

Theorem 5.2 *Let X and X' be regular topological spaces and let $\varphi : X \to X'$ be continuous. Suppose the pair $(\mathbb{M}_\circ, V_\circ)$ obeys an NLDP with rate-function s; then the pair $(\mathbb{M}'_\circ, V_\circ)$, where*

$$\mathbb{M}'_n := \mathbb{M}_n \circ \varphi^{-1}, n \geq 1, \tag{5.21}$$

obeys an NLDP with rate-function $\sigma^\diamond$, where $\sigma^\diamond$ is the usc regularization of the function

$$x' \to \sigma(x') := \sup_{\{y:\varphi(y)=x'\}} s(y); \tag{5.22}$$

$\sigma^\diamond$ is the R-L function of the pair $(\mathbb{M}'_\circ, V_\circ)$.

Proof: For all $F \in \mathcal{F}(X')$, we have

$$\begin{aligned} \overline{m}'[F] &= \overline{m}[\varphi^{-1}F] \\ &\leq \sup_{x \in \varphi^{-1}F} s(x) \\ &= \sup_{x' \in F} \sup_{\{y:\varphi(y)=x'\}} s(x). \end{aligned} \tag{5.23}$$

A similar calculation yields

$$\overline{m}'[G] \geq \sup_{x' \in G} \sup_{\{y:\varphi(y)=x'\}} s(x) \tag{5.24}$$

for all $G \in \mathcal{G}(X')$. The stated result now follows from Lemmas 3.1, 5.1 and Corollary 5.1. □

Remark: We mention without proof the following result. Let X be a regular topological space. The narrow topology in the space of increasing set-functions is separated. Let $\overline{m}$ be given; define the set-function $\overline{m}^\flat$ as follows: on $\mathcal{F}(X)$, $\overline{m}^\flat[F] := \sup_{G \supset F} \overline{m}[F]$; then extend the function to every A by setting $\overline{m}^\flat[A] := \sup_{F \subset A} \overline{m}^\flat[F]$. Then exists an NLDP with usc rate-function $s =^\wedge \overline{m} =^\wedge \underline{m}$ for the pair $(M_\circ, V_\circ)$ if and only if

$$\overline{m}[F] \leq \underline{m}[G] , \tag{5.25}$$

for all F and G with $F \subset G$, and

$$\sup_{K \subset F} \overline{m}^\flat[K] = \overline{m}^\flat[F] , \tag{5.26}$$

for all F.

5.2 Exponential Tightness

The concept of exponential tightness plays an important rôle, since it is often the case that the only way to establish an NLDP is to start with a VLDP and prove that exponential tightness holds. There is a bonus which comes from following this route: the level-sets of the rate-function are automatically compact (Lemma 5.3). We return to this useful property in section 6.

Definition 5.3 *A pair* $(\mathbb{M}_o, V_o)$ *is* exponentially tight *if there exists a sequence* $\{K_n\}_{n\geq 1}$ *of compact subsets of* X *such that*

$$\limsup_{n\to\infty} \overline{m}[X\backslash K_n] = -\infty. \tag{5.27}$$

Lemma 5.2 *Suppose that the pair* $(\mathbb{M}_o, V_o)$ *is exponentially tight and that the function* $s : X \to \overline{\mathbb{R}}$ *is such that*

$$\overline{m}[K] \leq \sup_{x\in K} s(x) \tag{5.28}$$

for all $K \in \mathcal{K}(X)$*; then*

$$\overline{m}[F] \leq \sup_{x\in F} s(x) \tag{5.29}$$

for all $F \in \mathcal{F}(X)$.

Proof: By hypothesis, there exists a sequence of compact subsets of X such that

$$\limsup_{n\to\infty} \overline{m}[X\backslash K_n] = -\infty. \tag{5.30}$$

For each K_n and each $F \in \mathcal{F}(X)$, we have

$$\begin{aligned} \overline{m}[F] &= \overline{m}[F\cap K_n] \vee \overline{m}[F\backslash K_n] \\ &\leq \overline{m}[F\cap K_n] \vee \overline{m}[X\backslash K_n]. \end{aligned} \tag{5.31}$$

But $F \cap K_n$ is compact, so that $\overline{m}[F\cap K_n] \leq {}^{\vee}s[F]$. Taking the superior limit over n in (5.31), we have

$$\overline{m}[F] \leq {}^{\vee}s[F]. \tag{5.32}$$

□

Lemma 5.3 *Suppose that the pair* $(\mathbb{M}_o, V_o)$ *is exponentially tight; then the level sets*

$$\{x : \overline{\mu}(x) \geq a\}, \{x : \underline{\mu}(x) \geq a\}, a \in \mathbb{R} \tag{5.33}$$

of the upper deviation function $\overline{\mu}$ *and the lower deviation function* $\underline{\mu}$ *are compact.*

Proof: Since $\overline{\mu}$ and $\underline{\mu}$ are usc their level sets are closed; since

$$\{x : \underline{\mu}(x) \geq a\} \subset \{x : \overline{\mu}(x) \geq a\} \tag{5.34}$$

it is enough to prove that there exists a compact set K_a containing $\{x : \overline{\mu}(x) \geq a\}$ for each $a \in \mathbb{R}$. Since the pair $(\mathbb{M}_o, V_o)$ is exponentially tight, there exists $K_n \in \mathcal{K}(X)$ such that

$$\overline{m}[X\backslash K_n] < a. \tag{5.35}$$

But $X\backslash K_n$ is open, so for each $x \in X\backslash K_n$, we have

$$\overline{\mu}(x) \leq \overline{m}[X\backslash K_n] < a. \tag{5.36}$$

Hence we have

$$X \backslash K_n \subset \{x : \overline{\mu}(x) < a\} \tag{5.37}$$

so that

$$\{x : \overline{\mu}(x) \geq a\} \subset K_n. \tag{5.38}$$

□

Lemma 5.4 *Let $g : X \to \mathbb{R}$ be a locally bounded measurable function. Suppose that the pair $(\mathbb{M}_o, V_o)$ is exponentially tight and that*

$$\limsup_{a \to \infty} \overline{m}^g[\{x \in X : g(x) \geq a\}] = -\infty; \tag{5.39}$$

then the pair $(\mathbb{M}_o^g, V_o)$ is also exponentially tight.

Proof: Put $G_a := \{x \in X : g(x) > a\}$; since the pair $(\mathbb{M}_o, V_o)$ is exponentially tight, there exists a sequence $\{K_n\}_{n \geq 1}$ of compact subsets of X such that

$$\limsup_{n \to \infty} \overline{m}[X \backslash K_n] = -\infty. \tag{5.40}$$

For each $n \geq 1$, we have

$$\overline{m}^g[(X \backslash G_a) \cap (X \backslash K_n)] \leq a + \overline{m}[X \backslash K_n] \tag{5.41}$$

so that

$$\begin{aligned} \overline{m}^g[X \backslash K_n] &= \overline{m}^g[(X \backslash G_a) \cap (X \backslash K_n)] \vee \overline{m}^g[G_a \cap (X \backslash K_n)] \\ &\leq (a + \overline{m}[X \backslash K_n]) \vee \overline{m}^g[G_a]. \end{aligned} \tag{5.42}$$

Given $b \in \mathbb{R}$, choose a such that

$$\overline{m}^g[G_a] < b \tag{5.43}$$

and then choose n_0 such that, for all $n \geq n_0$, we have

$$a + \overline{m}[X \backslash K_n]] < b; \tag{5.44}$$

then

$$\overline{m}^g[X \backslash K_n] < b \tag{5.45}$$

for all $n \geq n_0$. □

The next theorem is one of the main results of the theory. It summarizes the above results.

Theorem 5.3 *(a) Let X be a Hausdorff topological space. Suppose that the pair $(\mathbb{M}_o, V_o)$ obeys a VLDP with rate-function s and is exponentially tight; then the pair satisfies an NLDP with rate-function s.*

(b) If furthermore the topological space X is regular, then the rate-function is unique; it is the R-L function μ of the pair $(\mathbb{M}_o, V_o)$ and hence has compact level-sets. Moreover if $g : X \to \mathbb{R}$ is continuous and

$$\limsup_{a \to \infty} \overline{m}^g[\{x \in X : g(x) \geq a\}] = -\infty; \tag{5.46}$$

then the pair $(\mathbb{M}_o^g, V_o)$ satisfies an NLDP with rate-function $\mu + g$.

Proof: the result follows from Lemma 5.2, Corollary 5.1, Lemma 5.3, Corollary 2.1 and Lemma 5.4. □

5.3 Concentration of Exponentially Tight Measures

We return now to the important subject of the concentration of measures; it turns out that exponential tightness of the pair $(\mathbb{M}_\circ, V_\circ)$ is a substitute for the compactness of the space X so that we are able to improve on the results of Section 2.3.

Theorem 5.4 *Let $\mathbb{M}_\circ$ be a sequence of probability measures on $\mathcal{B}(X)$. Suppose that the pair $(\mathbb{M}_\circ, V_\circ)$ is exponentially tight; then*

(a) the set

$$N_{\overline{\mu}} := \{x \in X, \overline{\mu}(x) = 0\} \tag{5.47}$$

is non-empty and compact;

(b) the sequence $\mathbb{M}_\circ$ is eventually concentrated on the set $N_{\overline{\mu}}$.

Proof: Fix $a \in (-\infty, 0)$; then there exists a compact set K_a such that

$$\overline{m}[X \backslash K_a] < a < 0, \tag{5.48}$$

since the pair is exponentially tight. Since $\mathbb{M}_n[X] = 1$ for all $n \geq 1$, we have $\overline{m}[X] = 0$. Thus we have, by Lemma 2.4 (d),

$$\begin{aligned} 0 = \overline{m}[X] &= \overline{m}[K_a] \vee \overline{m}[X \backslash K_a] \\ &= \overline{m}[K_a] \\ &\leq \sup_{x \in K_a} \overline{\mu}(x) \\ &\leq \sup_{x \in X} \overline{\mu}(x) \\ &\leq \overline{m}[X] = 0. \end{aligned} \tag{5.49}$$

Since $\overline{\mu}$ is usc and K_a is compact, there exists $x_a \in K_a$ with

$$\overline{\mu}(x_a) = \sup_{x \in K_a} \overline{\mu}(x) = 0. \tag{5.50}$$

By Lemma 5.3, the level-sets of $\overline{\mu}$ are compact; since $N_{\overline{\mu}} = \{x \in X : \overline{\mu}(x) \geq 0\}$, (a) is proved.

Fix $a \in (-\infty, 0)$; the level-set $L_a := \{x \in X : \overline{\mu}(x) \geq a\}$ is compact and non-empty. Let G be an open neighbourhood of $N_{\overline{\mu}}$; there are two cases to be considered: (a) $(X \backslash G) \cap L_a = \emptyset$; since $\overline{\mu}(x) < a$ on $X \backslash G$,

$$^{\vee}\overline{\mu}[(X \backslash G)] \leq a < 0. \tag{5.51}$$

(b) $(X \backslash G) \cap L_a \neq \emptyset$; then the supremum of $\overline{\mu}(x)$ on $X \backslash G$ is taken on the compact set $(X \backslash G) \cap L_a$; since $\overline{\mu}$ is usc, there exists $x_a \in (X \backslash G) \cap L_a$ such that

$$^{\vee}\overline{\mu}[(X \backslash G)] = \overline{\mu}(x_a) < 0 \tag{5.52}$$

because $N_{\overline{\mu}}$ is disjoint from $X \backslash G$. In either case, we have

$$^{\vee}\overline{\mu}[X \backslash G] < 0. \tag{5.53}$$

Since $X \backslash G$ is closed, we have by Lemma 5.2

$$\overline{m}[X \backslash G] \leq {}^{\vee}\overline{\mu}[X \backslash G] < 0 \tag{5.54}$$

so that

$$\lim_{n \to \infty} \mathbb{M}_n[X \backslash G] = 0 \tag{5.55}$$

and hence

$$\lim_{n \to \infty} \mathbb{M}_n[G] = 1. \tag{5.56}$$

□

6 Large Deviation Principles and Varadhan's Theorems

6.1 Large Deviation Principles

We have already seen that compactness of its level-sets is a useful property for a rate-function to have; moreover, if the topological space X is regular and if exponential tightness holds, then the rate-function, when it exists, automatically has this property (Theorem 5.3 which is recalled below as Theorem 6.1). These circumstances justify making the next definition, Varadhan's original one [V].

Definition 6.1 *A pair* $(\mathbb{M}_\circ, V_\circ)$ *obeys a* large deviation principle *(LDP) if there exists a function* $s : X \to \overline{\mathbb{R}}$ *such that*

(LDP 1) *the function* s *is usc;*
(LDP 2) *the level sets of* s *are compact;*
(LDP 3) $\overline{m}[F] \leq \sup_{x \in F} s(x)$, $F \in \mathcal{F}(X)$;
(LDP 4) $\underline{m}[G] \geq \sup_{x \in G} s(x)$, $G \in \mathcal{G}(X)$.

The function s *is called the* rate-function.

Theorem 6.1 *Let* X *be a Hausdorff topological space, and let the pair* $(\mathbb{M}_\circ, V_\circ)$ *be exponentially tight and obey a VLDP with R-L function* μ *as rate-function. Suppose that* $g : X \to \mathbb{R}$ *is continuous and obeys the condition*

$$\limsup_{n \to \infty} \overline{m}^g[\{x \in X : g(x) \geq a\}] = -\infty; \tag{6.1}$$

then the pair $(\mathbb{M}_\circ^g, V_\circ)$ *satisfies an LDP with R-L function* $\mu + g$. *If* X *is regular, then the rate-function of an LDP is unique.*

The next theorem is one whose proof makes use of methods which are outside the scope of this paper, but is a result of which the reader should be aware [LS].

Theorem 6.2 *Let* X *be a complete metric space; if the pair* $(\mathbb{M}_\circ, V_\circ)$ *obeys an LDP, then it is exponentially tight.*

6.2 Varadhan's Theorems

In this section we give the main theorems which hold when the pair $(\mathbb{M}_\circ, V_\circ)$ obeys an LDP. There is no assumption of exponential tightness.

Theorem 6.3 *Let* X *be a Hausdorff topological space, and let the pair* $(\mathbb{M}_\circ, V_\circ)$ *obey an LDP with RL-function* μ *as rate-function; let* $g : X \to \mathbb{R}$ *be continuous; if*

$$\limsup_{a \to \infty} \overline{m}^g[\{x \in X : g(x) \geq a\}] = -\infty, \tag{6.2}$$

then the pair $(\mathbb{M}_\circ^g, V_\circ)$ *obeys an LDP with rate-function* $\mu^g = \mu + g$.

Proof: We verify the hypothesis of Theorem 5.1, and then prove that the level-sets are compact. We must show that

$$\limsup_{a \to -\infty} \overline{m}^g[\{x \in X : g(x) \leq a\}] = -\infty. \tag{6.3}$$

If $\mu(x) \equiv -\infty$, then this is obvious; otherwise there exists $b \in \mathbb{R}$ such that $\{x \in X : \mu(x) \geq b\}$ is a non-empty compact set, and therefore g attains its minimum, a^*, on this set. Choose $\epsilon > 0$; the open set $G := \{x \in X : g(x) > a^* - \epsilon\}$ contains $\{x \in X : \mu(x) \geq b\}$, so that, for each $a < a^* - \epsilon$, we have

$$\begin{aligned} \overline{m}^g[\{x \in X : g(x) \leq a\}] &\leq a + \overline{m}[X \backslash G] \\ &\leq a + b, \end{aligned} \tag{6.4}$$

from which (6.3) follows. It remains to prove that, for each $b \in \mathbb{R}$, the set

$$F_b := \{x \in X : \mu^g(x) \geq b\} \tag{6.5}$$

is compact. Put $c(a) := \overline{m}^g[\{g(x) > a\}]$; since $\{g(x) > a\}$ is open,

$$c(a) \geq {}^{\vee}\mu^g[\{g(x) > a\}] \geq \mu^g(x) \tag{6.6}$$

for all $x \in \{g(x) > a\}$. Given b, we can choose a such that $c(a) < b$ (see (6.2); therefore

$$F_b \cap \{g(x) > a\} = \emptyset, \tag{6.7}$$

and

$$F_b = F_b \cap (X \backslash \{g(x) > a\}) \subset \{x \in X : \mu(x) \geq b - a\} \tag{6.8}$$

which is compact by hypothesis. But F_b is closed because μ^g is usc, so that F_b is compact. □

Corollary 6.1 *Let X be a Hausdorff topological space, and let the pair $(\mathbb{M}_\circ, V_\circ)$ obey an LDP with RL-function μ as rate-function; let $g : X \to \mathbb{R}$ be continuous; if*

$$\limsup_{a \to \infty} \overline{m}^g[\{x \in X : g(x) \geq a\}] = -\infty, \tag{6.9}$$

then

$$\lim_{n \to \infty} \frac{1}{V_n} \ln \int_X e^{V_n g(x)} \mathbb{M}_n[dx] = \sup_{x \in X} (\mu(x) + g(x)). \tag{6.10}$$

Proof: As for Corollary 3.2. □

Finally we prove the contraction principle.

Theorem 6.4 *Let X and X' be two regular topological spaces. Let $\varphi : X \to X'$ be continuous and let the pair $(\mathbb{M}_\circ, V_\circ)$ obey an LDP on X with RL-function μ; then the pair $(\mathbb{M}'_\circ, V_\circ)$, where $\mathbb{M}'_n := \mathbb{M}_n \circ \varphi^{-1}$, satisfies an LDP on X' with RL-function σ given by*

$$\sigma(x') = \sup_{\{y : \varphi(y) = x'\}} \mu(y). \tag{6.11}$$

Proof: From Theorem 5.2 we know that the pair $(\mathbb{M}'_\circ, V_\circ)$ satisfies an NLDP with a rate-function which is the usc regularization of σ; it remains to show that σ is already usc and, in addition, has compact level sets. By (2.11) $\sigma^\diamond(x') = \sigma(x')$ if $\sigma^\diamond(x') = -\infty$. Below, F' always denotes a closed neighbourhood of $x' \in X'$; since the space X' is regular, a function f is usc if and only if

$$f(x') = \inf_{F' \ni x'} \sup_{y' \in F'} f(y'). \tag{6.12}$$

Let

$$\sigma^{\diamond}(x') = \inf_{F' \ni x'} \sup_{y' \in F'} \sigma(y') > -\infty. \tag{6.13}$$

If (6.13) holds, the closed set $\varphi^{-1}F'$ is nonempty, and has a nonempty intersection with the compact set $\{\mu(x) \geq b\}$, provided $b < \sigma^{\diamond}(x')$. Since μ is usc, there exists $y(F') \in (\varphi^{-1}F') \cap \{\mu(x) \geq b\}$ with

$$\mu(y(F')) = \sup_{y \in \varphi^{-1}F'} \mu(y) = \sup_{y' \in F'} \sigma(y'). \tag{6.14}$$

Since $(\varphi^{-1}F') \cap \{\mu(x) \geq b\}$ is compact, the net $\{y(F') : F' \ni x'\}$ has a cluster point y^*, say; let $\{y(F') : F' \ni x'\}$ now denote a subnet converging to y^*. Since the net $\{\varphi(y(F')) : F' \ni x'\}$ converges to x' and φ is continuous, we have

$$\varphi(y^*) = x'. \tag{6.15}$$

By (6.13), and because μ is usc,

$$\begin{aligned} \sigma(x') \leq \sigma^{\diamond}(x') &= \inf_{F' \ni x'} \mu(y(F')) \\ &\leq \lim_{F'} \mu(y(F')) \\ &\leq \mu(y^*) \leq \sup_{\{x:\varphi(x)=x'\}} \mu(x) = \sigma(x'); \end{aligned} \tag{6.16}$$

this shows that $\sigma = \sigma^{\diamond}$. Since σ is usc, its level sets are closed and the inverse image under φ of $\{\sigma(x') \geq a\}$, $a > -\infty$, has a non-empty intersection K with the compact level set $\{\mu(x) \geq b\}$, $b < a$. But the level set $\{\sigma(x') \geq a\}$ is the image under φ of the compact set K and hence is compact. □

7 Convexity

7.1 The Scaled Generating Function

Up to this point in our development, only the topological structure and the Borel structure of the space X have been used; when, in addition, the space X has a convex structure, the theory of large deviations becomes substantially more powerful. Eventually, we will assume X to be a closed convex subset of (E, τ), a locally convex topological vector space E over $\mathbb{R}$ equipped with a Hausdorff topology τ; initially, we will study large deviations on (E, τ).

Let E^* be the topological dual of E; choose a topology τ^* on E^* so that the pair (E, E^*) is in duality: the topological dual of (E^*, τ^*) is E. There are several possible candidates for τ and τ^* so that the pair (E, E^*) is in duality, the coarest being the weak topologies $\sigma(E, E^*)$ and $\sigma(E^*, E)$. It should be noted that the topology τ^* on E^* plays a minor rôle in the development. As before, $\mathbb{M}_o$ is a sequence of locally finite positive measures on the Borel subsets $\mathcal{B}(E)$ of (E, τ). The function

$$t \to m_n^t[E] \tag{7.1}$$

is defined on E^* as an $\overline{\mathbb{R}}$-valued function. It is convex and $m_n^t[E] > -\infty$; this follows from Hölder's inequality applied to

$$\exp\{V_n m_n^t[E]\} = \int_E e^{V_n \langle t, x \rangle} \mathbb{M}_n[dx] \tag{7.2}$$

and the fact that $\mathbb{M}_n^t[E] > 0$.

Lemma 7.1 *Let s, t be elements of E^*; suppose that for some $\alpha > 0$ there exists $c \in \mathbb{R}$ such that*

$$m_n^{t+\alpha s}[E] \leq c \tag{7.3}$$

for all $n \in \mathbb{N}$; then

$$\limsup_{a \to \infty} \overline{m}^t[\{x : \langle s, x\rangle > a\}] = -\infty. \tag{7.4}$$

Proof: Let $s \in E^*$ and $a > 0$; for $\alpha > 0$, we have

$$\mathbb{M}_n^t[\{x : \langle s, x\rangle > a\}] \leq e^{-V_n \alpha a} \int_E e^{V_n \langle \alpha s, x\rangle} \mathbb{M}_n^t[dx] \tag{7.5}$$

so that

$$\begin{aligned} \overline{m}^t[\{x : \langle s, x\rangle > a\}] &\leq -\alpha a + \overline{m}^{t+\alpha s}[E] \\ &\leq -\alpha a + c. \end{aligned} \tag{7.6}$$

□

Corollary 7.1 *Let E be $\mathbb{R}^d$ with the usual topology and E^* identified with $\mathbb{R}^d$. Suppose that there is an open neighbourhood V of 0 in E and a constant c such that, for all $t \in V$ and all $n \in \mathbb{N}$, we have*

$$m_n^t[E] \leq c; \tag{7.7}$$

then the pair $(\mathbb{M}_o, V_o)$ is exponentially tight.

Proof: Let $\{e_j : j = 1, ..., d\}$ be a standard basis for $E = \mathbb{R}^d$; the sets

$$K_n := \{x \in E : |\langle e_j, x\rangle| \leq n, j = 1, \ldots, d\} \tag{7.8}$$

are compact for all $n \in \mathbb{N}$. It follows from Lemma 7.1 and (c) of Lemma 2.4 that

$$\limsup_{a \to \infty} \overline{m}[E \backslash K_n] = -\infty. \tag{7.9}$$

□

Definition 7.1 *If*

$$\overline{m}^t[E] = \underline{m}^t[E] \tag{7.10}$$

for all $t \in E^*$, then we say that the scaled generating function *$p : E^* \to \overline{\mathbb{R}}$ exists for the pair $(\mathbb{M}_o, V_o)$ and is given by*

$$p(t) := \overline{m}^t[E] = \underline{m}^t[E]. \tag{7.11}$$

Remark: The scaled generating function is the grand canonical pressure of statistical mechanics; we shall sometimes refer to it as the *pressure*.

We shall require some results about conjugate functions which we state without proof; proofs may be found in Brønsted [B]. An $\overline{\mathbb{R}}$-valued function is *proper* if it never takes the value $-\infty$ and it is not identically $+\infty$. Let $f : E \to \overline{\mathbb{R}}$ be any function on E; the conjugate function $f^* : E^* \to \overline{\mathbb{R}}$ is defined by

$$f^*(t) := \sup_{x \in E} \{\langle t, x\rangle - f(x)\} \tag{7.12}$$

Similarly, let g be defined on E^*; then its conjugate function $g^* : E \to \overline{\mathbb{R}}$ is defined by

$$g^*(x) := \sup_{t \in E^*} \{\langle t, x \rangle - g(t)\}. \tag{7.13}$$

Since both f^* and g^* are upper envelopes of families of affine functions, they are automatically convex and l.s.c. in the weak topologies; they may, however, fail to be proper. It is always the case that $f^{**} \leq f$, but it is possible that $f^{**}(x) < f(x)$ for some x. We have the following important result:

Lemma 7.2 *If f is a proper l.s.c. convex function, then f^* is a proper l.s.c. convex function and $f^{**} = f$.*

Henceforth we shall assume that the following hypothesis holds:

Hypothesis (P): *the scaled generating function $p : E^* \to \overline{\mathbb{R}}$ exists for the pair $(\mathbb{M}_o, V_o)$ and is a proper convex function.*

Remarks: (1) Notice that we do not assume that p is l.s.c.
(2) The convexity of p is, in fact, a consequence of the convexity of the function $t \to \overline{m}^t[E]$.

Lemma 7.3 *Assume that (P) holds; for all t in E^* and all x in E, we have*

$$\underline{\mu}^t(x) = \underline{\mu}(x) + \langle t, x \rangle \leq p(t), \tag{7.14}$$

$$\overline{\mu}^t(x) = \overline{\mu}(x) + \langle t, x \rangle \leq p(t), \tag{7.15}$$

$$\overline{\mu}(x) \leq -(-\overline{\mu})^{**}(x) \leq -p^*(x). \tag{7.16}$$

Proof: Since E is open, we have

$$\underline{\mu}^t(x) \leq \underline{m}^t[E] = p(t) \tag{7.17}$$

and

$$\overline{\mu}^t(x) \leq \overline{m}^t[E] = p(t) \tag{7.18}$$

from the definitions of $\underline{\mu}^t$ and $\overline{\mu}^t$ and (P); the remaining parts of (7.14) and (7.15) follow from Corollary 2.1. Taking the supremum over x in (7.15), we get

$$(-\overline{\mu})^*(t) \leq p(t); \tag{7.19}$$

taking conjugates, we get

$$p^*(x) \leq (-\overline{\mu})^{**}(x) \leq -\overline{\mu}(x). \tag{7.20}$$

□

Definition 7.2 *Let $f : E^* \to \overline{\mathbb{R}}$ be convex; then dom f, the* essential domain *of f, is the set*

$$dom\, f := \{t \in E^* : f(t) < \infty\}. \tag{7.21}$$

A point t of dom f is an algebraic interior point *of dom f if, for all s in E^*, there exists $\epsilon > o$ such that the segment $\{u : u = t + \alpha s, |\alpha| < \epsilon\}$ is contained in dom f.*

Theorem 7.1 *Let the pair* $(\mathbb{M}_o, V_o)$ *obey an LDP and let* μ *be its RL-function. Assume that* (P) *holds; then*

(a) for each t *in the algebraic interior of* $dom\, p$*, the pair* $(\mathbb{M}_o^t, V_o)$ *obeys an LDP, its RL-function is* μ^t *and*

$$p(t) = (-\mu)^*(t); \tag{7.22}$$

if, in addition, the RL-function μ *is concave, then*

(b) μ *is u.s.c. in the* $\sigma(E, E^*)$*-topology and if, furthermore,* $dom\, p = E^*$*, then*

$$\mu(x) = -p^*(x) \tag{7.23}$$

for all x *in* E*, and* p *is l.s.c. in the* $\sigma(E^*, E)$*-topology.*

Proof: Let t be in the algebraic interior of $dom\, p$ and choose $\epsilon > 0$ such that the segment $\{(1+\alpha)t : |\alpha| < \epsilon\}$ is in $dom\, p$. On the segment $\{u : u = (1+\alpha)\, t, |\alpha| \leq \epsilon/2\}$, the function p is proper and bounded above:

$$p((1+\alpha)t) \leq C < \infty \ , \ |\alpha| < \frac{\epsilon}{2}.$$

Thus Lemma 7.1 implies that the hypothesis (6.2) of Theorem 6.3 is satisfied; this proves (a), since Corollary 6.1 applies. If we use the weak topologies $\sigma(E, E^*)$ and $\sigma(E^*, E)$, then the pair (E, E^*) is in duality; thus the closed convex sets in the topologies τ and $\sigma(E, E^*)$ are the same, ([RR] chap.2), and μ is u.s.c. in the $\sigma(E, E^*)$-topology. Finally, if dom $p = E^*$, then a) holds for all t showing that p is l.s.c. in the $\sigma(E^*, E)$; the duality result now implies that

$$\mu(x) = -p^*(x) \tag{7.24}$$

for all x in E. □

Remarks: (1) In order to conclude that (a) and (b) hold, it is enough to assume that the pair $(\mathbb{M}_o, V_o)$ satisfies an NLDP and that for t in $dom\, p$ there is an $\epsilon > 0$ for which the segment $\{u : u = (1+\alpha)t, |\alpha| < \epsilon\}$ is in $dom\, p$.
(2) The existence and finiteness of p imply that μ can be computed as $-p^*$ whenever μ is concave.
(3) Assume that (P) holds with $dom\, p = E^*$ and introduce the normalized version $\mathbb{K}_n^t$ of the tilted measure $\mathbb{M}_n^t$:

$$\mathbb{K}_n^t := \exp\{-V_n m_n^t[E]\}\mathbb{M}_n^t. \tag{7.25}$$

Let κ^t denote the R-L function of the pair $(\mathbb{K}_o^t, V_o)$; then κ^t satisfies

$$\kappa^t(x) = \mu(x) + \langle t, x\rangle - p(t) \leq 0, \tag{7.26}$$

where μ is the RL-function of the pair $(\mathbb{M}_o, V_o)$. For y in E, we have

$$\mu(y) = -p^*(y) \tag{7.27}$$

if there exists s in E^* such that

$$\kappa^s(y) = 0, \tag{7.28}$$

and this is the case if and only if

$$\langle s, y\rangle - p(s) = -\mu(y). \tag{7.29}$$

In statistical mechanics, this is a statement of the equivalence of ensembles (see also the example below in Section 7.2). Notice that the sequence $\mathbb{K}_o^s$ is eventually concentrated on the set

$$\{x \in E : \kappa^s(x) = 0\}. \tag{7.30}$$

7.2 Weak Law of Large Numbers and the Differentiability of the Pressure

In Theorem 7.1, we assumed the existence of μ and showed that, provided an LDP holds, we have

$$p(t) = (-\mu)^*(t) \tag{7.31}$$

at all algebraic interior points of $dom\, p$; in particular, if $dom\, p = E^*$, then

$$p = (-\mu)^*. \tag{7.32}$$

If, in addition, μ is concave, we have

$$p = (-\mu)^* \text{ and } \mu = -p^* \tag{7.33}$$

by duality. This suggests that it might be possible to prove the existence of μ by showing by a different route that (7.33) holds. To do that, we prove a weak law of large numbers by a another method than that of sections 2.3 and 5.3, based on p rather than on μ; from this weak law of large numbers, the result follows easily. However, we need stronger measure-theoretic assumptions; first, we explain these.

Definition 7.3 *Let* $\mathbb{K}$ *be a probability measure on* $\mathcal{B}(X)$ *and let*

$$F_0 := \{\cap F \,:\, F \in \mathcal{F}(X) : \mathbb{K}[F] = 1\}; \tag{7.34}$$

if

$$\mathbb{K}[F_0] = 1, \tag{7.35}$$

we say that $\mathbb{K}$ *has the* support property.

Radon measures have the support property; therefore we assume below that our space X is a *Radon space* (that is, a Hausdorff topological space on which every Borel measure is a Radon measure [Sc]). A compact metrizable space is Radon, for example. A typical situation is the following one: X is a closed convex subset of E, a locally convex Hausdorff topological space over the reals, and X is a Polish space with respect to the topology which it inherits as a subset of E; then X is Radon.

Example: Consider a set-up which arises in statistical mechanics: let $\mathbb{Z}^d$ be the d-dimensional integer lattice; over each site j in $\mathbb{Z}^d$ we have a copy Ω_j of some fixed finite set S and Ω is the product space. On each Ω_j we have the normalized counting measure ρ_j and we denote the product measure on Ω by ρ. We equip each Ω_j with the discrete topology and Ω with the product topology so that Ω is a Polish space. We choose for E the space $\mathcal{M}(\Omega)$ of all bounded signed measures on Ω; we choose for τ the narrow topology, the $\sigma(E, C_b(\Omega))$- topology (here $C_b(\Omega)$ is the space of bounded continuous functions on Ω). The topological dual E^* of (E, τ) is precisely $C_b(\Omega)$. Let $\mathcal{M}_1^+(\Omega)$ be the space of probability measures on Ω; equipped with the narrow topology τ, the space $(\mathcal{M}_1^+(\Omega), \tau)$ is Polish. Let ω be an element of Ω; there is a natural action of the group $\mathbb{Z}^d$ on Ω: for each j in $\mathbb{Z}^d$, let θ_j denote the map

$$(\theta_j\omega)(i) = \omega(i - j). \tag{7.36}$$

This action lifts to an action on the space of measures. Let X be the space of translation – invariant probability measures. Let $\{\Lambda_n\}_{n\geq 1}$ be an increasing sequence of cubes in $\mathbb{Z}^d$ with $|\Lambda_n| \to \infty$ when $n \to \infty$; for every ω, we define ω_n^{per} to

be the configuration which coincides with ω on Λ_n and which is extended to $\mathbb{Z}^d$ periodically; let $\{Z_n\}_{n\geq 1}$ be a sequence of random variables on Ω defined by

$$Z_n(\omega) := \frac{1}{|\Lambda_n|} \sum_{j\in\Lambda_n} \delta_{\theta_j(\omega_n^{per})} \tag{7.37}$$

and let $\mathbb{K}_n := \rho \circ Z_n^{-1}$ be the distribution of Z_n under ρ. We have, of course, $\mathbb{K}_n[X] = 1$ for each $n \geq 1$. In this example, the statements

$$\mu(x) = -p^*(x), \tag{7.38}$$
$$p(t) = (-\mu)^*(t), \tag{7.39}$$

are simply the two expressions of the variational principle of statistical mechanics (see [R2]). Indeed, if we write the above statements explicitly, we have

$$\mu(x) = \inf_{t\in C_b(\Omega)} (p(t) - \int_\Omega t(\omega)x[d\omega]), \tag{7.40}$$
$$p(t) = \sup_{x\in X} (\mu(x) + \int_\Omega t(\omega)x[d\omega]). \tag{7.41}$$

In this case, we recognise $\mu(x)$ as the specific Kolmogorov entropy of the measure x.

From now on, we suppose that $\mathbb{K}_o$ is a sequence of probability measures and that (P) holds with dom $p = E^*$. It follows that we may normalize the tilted measures $\mathbb{K}_n^t$, at least for n sufficiently large, by dividing by $\exp(V_n m_n^t[E])$ and we do so. Let X be a closed convex subset of E and suppose that $\mathbb{K}_n^t[X] = 1$, where $\mathbb{K}_n^t$ now denotes the normalized tilted measure. The notion of convergence which we adopt for sequences of probability measures on X is narrow convergence. Finally, we require the notion of weak differentiability of a convex function:

Definition 7.4 *Let $f : E^* \to \overline{\mathbb{R}}$ be convex; we say that f is* weakly differentiable at t *if t is an algebraic interior point of dom f and if there exists an element of E, denoted by grad $f(t)$, such that, for all s in E^*, we have*

$$\frac{d}{d\alpha} f(t + \alpha s) \mid_{\alpha=0} = \langle s, grad\, f(t)\rangle. \tag{7.42}$$

Theorem 7.2 *Let X be a closed convex subset of E, a locally convex Hausdorff topological space over the reals; let X be a Radon space with respect to the topology which it inherits as a subset of E. Let $\mathbb{K}_o$ be a sequence of probability measures on X and let hypothesis (P) be satisfied with dom $p = E^*$; then we have:*

(a) If p is weakly differentiable at t and if $\mathbb{K}_o^t$ is a relatively compact set, then $\mathbb{K}_o^t$ obeys a weak law of large numbers: $\mathbb{K}_o^t$ converges to the Dirac mass at grad $p(t)$.

(b) If $\mathbb{K}_o^t$ converges to the Dirac mass at y, then the R-L function $\kappa(y)$ at y exists for the pair $(\mathbb{K}_o, V_o)$ and

$$p^*(y) = (-\kappa)(y). \tag{7.43}$$

Before starting on the proof of Theorem 7.2, we establish a simple lemma on convex functions. Let g be a proper convex function on the real line. The quotient

$$\frac{g(\alpha + h\beta) - g(\alpha)}{h}$$

is monotone decreasing as $h \downarrow 0$; we define $\delta g(\alpha; \beta)$ by

$$\delta g(\alpha; \beta) := \lim_{h \downarrow 0} \frac{g(\alpha + h\beta) - g(\alpha)}{h}. \tag{7.44}$$

The function $\beta \to \delta g(\alpha; \beta)$ is positively homogeneous and

$$-\delta g(\alpha; -\beta) \leq \delta g(\alpha; \beta)\ ,\ \beta \in \mathbb{R}. \tag{7.45}$$

The function g is differentiable at α if and only if we have equality in (7.45) for all $\beta \in \mathbb{R}$.

Lemma 7.4 *Let $f_n, n \geq 1$, and f be proper convex functions on $\mathbb{R}$ with dom f_n = dom f_n = dom $f = \mathbb{R}$. Suppose that $\{f_n\}_{n \geq 1}$ converges pointwise to f and that f is differentiable at α; then*

$$\lim_{n \to \infty} \frac{f_n(\alpha + h_n\beta) - f_n(\alpha)}{h_n} = \delta f(\alpha; \beta) \tag{7.46}$$

for any decreasing sequence $\{h_n\}_{n \geq 1}, h_n \downarrow 0$.

Proof: For $\mu \in [0, 1]$, the convexity of f_n implies that

$$\begin{aligned} f_n(\alpha) - f_n(\alpha - \gamma) &\leq \frac{f_n(\alpha + \mu\gamma) - f_n(\alpha)}{\mu} \\ &\leq f_n(\alpha + \gamma) - f_n(\alpha). \end{aligned} \tag{7.47}$$

Choosing $\mu = h_n/\nu$, with $\nu > 0$ and $\gamma = \nu\beta$, we have

$$\begin{aligned} \frac{f_n(\alpha) - f_n(\alpha - \nu\beta)}{\nu} &\leq \frac{f_n(\alpha + h_n\beta) - f_n(\alpha)}{h_n} \\ &\leq \frac{f_n(\alpha + \nu\beta) - f_n(\alpha)}{\nu}. \end{aligned} \tag{7.48}$$

Taking lim sup and lim inf over n and then letting ν decrease to zero, we get

$$\begin{aligned} -\delta f(\alpha; -\beta) &\leq \liminf_{n \to \infty} \frac{f_n(\alpha + h_n\beta) - f_n(\alpha)}{h_n} \\ &\leq \limsup_{n \to \infty} \frac{f_n(\alpha + h_n\beta) - f_n(\alpha)}{h_n} \leq \delta f(\alpha; \beta). \end{aligned} \tag{7.49}$$

□

Proof of Theorem 7.2: Let $s \in E^*$ be given; let $p_n(t)$ denote the function $t \to k_n^t[X]$. The measures $\mathbb{K}_n^t$ are normalized, so, by definition, we have

$$\exp\{V_n(p_n(t + s/V_n) - p_n(t))\} = \int_X e^{\langle s, x \rangle} \mathbb{K}_n^t[dx]. \tag{7.50}$$

Let $\pi_s : X \to \mathbb{R}$ be the mapping defined by $x \to \pi_s(x) := \langle s, x \rangle$ and let ρ_n be the image measure defined by

$$\rho_n := \mathbb{K}_n^t \circ \pi_s^{-1}; \tag{7.51}$$

for every $\beta \in \mathbb{R}$, we have

$$\int_X e^{\langle \beta s, x\rangle} \mathbb{K}_n^t[dx] = \int_{\mathbb{R}} e^{\beta\alpha} \rho_n[d\alpha] \tag{7.52}$$

and Lemma 7.4 implies that

$$\lim_{n\to\infty} \int_{\mathbb{R}} e^{\beta\alpha} \rho_n[d\alpha] = \exp \beta\langle s, grad\, p(t)\rangle. \tag{7.53}$$

Using the theory of the Laplace transform [F], we deduce that $\{\rho_n\}_{n\geq 1}$ converges weakly to a Dirac measure concentrated at $\langle s, grad\, p(t)\rangle$. Define the set $C(s,\epsilon)$ by

$$C(s,\epsilon) := \{x \in X :| \langle s, x\rangle - \langle s, grad\, p(t)\rangle \,| \leq \epsilon\}; \tag{7.54}$$

this is a closed subset of X and if $\mathbb{K}^t$ is a cluster point of the sequence $\mathbb{K}_o^t$, then

$$\begin{aligned} 1 &= \limsup_{n\to\infty} \rho_n[\{\alpha :| \alpha - \langle s, grad\, p(t)\rangle \,| \leq \epsilon\}] \\ &= \limsup_{n\to\infty} \mathbb{K}_n^t[C(s,\epsilon)] \\ &\leq \mathbb{K}^t[C(s,\epsilon)]. \end{aligned} \tag{7.55}$$

The intersection of all $C(s,\epsilon)$ over all s and all $\epsilon > 0$ reduces to the set consisting of the single point $grad\, p(t)$ hence $\mathbb{K}^t$ is a Dirac measure concentrated at $grad\, p(t)$. This proves (a). The proof of (b) is straightforward: let G be any open neighbourhood of y; then

$$\lim_{n\to\infty} \mathbb{K}_n^s[G] = 1, \tag{7.56}$$

which implies that $\kappa^s(y) = 0$; since the measures are normalised, statement (7.14) of Lemma 7.3 now reads:

$$\underline{\kappa}^t(x) = \underline{\kappa}(x) - p(t) + \langle t, x\rangle \leq 0. \tag{7.57}$$

Replacing x by y and t by s, we get

$$\begin{aligned} 0 &= \underline{\kappa}(y) - p(s) + \langle s, y\rangle \\ &= \underline{\kappa}(y) + p^*(y); \end{aligned} \tag{7.58}$$

similarly, we have

$$0 = \overline{\kappa}(y) + p^*(y). \tag{7.59}$$

□

Finally, we consider the particular case when $E = \mathbb{R}^n = E^*$, equipped with the usual topology. The set-up is as described in the paragraph preceding the statement of Theorem 7.2, with $\mathbb{K}_n$ and $\mathbb{K}_n^t$ normalized measures. We assume that (P) holds and that p is l.s.c. For $t \in dom\, p$, the subdifferential $\partial p(t)$ of p at t is defined to be the set-valued mapping given by

$$\partial p(t) := \{x \in X : p(t+s) \geq p(t) + \langle s, x\rangle, \forall s \in E^*\}. \tag{7.60}$$

An element of $\partial p(t)$ is called a *subgradient*; the set $\partial p(t)$ may be empty, but if p is continuous and finite at t then $\partial p(t)$ is non-empty. The set $\partial p(t)$ is a closed convex subset of X and $x \in \partial p(t)$ if and only if

$$p(t) + p^*(x) = \langle t, x\rangle. \tag{7.61}$$

By definition, we have

$$dom\,\partial p := \{t \in E^* : \partial p(t) \neq \emptyset\}; \tag{7.62}$$

the duality relation above implies

$$x \in \partial p(t) \text{ if and only if } t \in \partial p^*(t), \tag{7.63}$$

and therefore

$$dom\,p^* \supset range\,\partial p \supset dom\,\partial p^*. \tag{7.64}$$

We shall use below the following fact about $range\,\partial p$:

$$dom\,p^* \supset range\,\partial p \supset ri(dom\,p^*). \tag{7.65}$$

Recall that $riC \neq \emptyset$ for every non-empty convex set $C \subset \mathbb{R}^n$; if $x \in riC$ and $y \in \overline{C}$, then

$$(1-\mu)x + \mu y \in riC, 0 \leq \mu \leq 1. \tag{7.66}$$

Recall also that a proper convex function f is said to be *essentially smooth* if it satisfies the following three conditions for $C = int(dom\,f)$:
(a) *C is non-empty;*
(b) *f is differentiable throughout C;*
(c) $\lim_{j\to\infty} \| grad\,f(t_j) \| = +\infty$ *whenever $\{t_j\}_{j\geq 1}$ is a sequence in C converging to a boundary point of C.*

Theorem 7.3 *Let X be a closed convex subset of $E = \mathbb{R}^d$ equipped with the usual topology; suppose that hypothesis (P) holds, and that zero belongs to $int(dom\,p)$. If p is essentially smooth, then the pair $(\mathbb{K}_o, V_o)$ obeys an LDP with RL-function κ given by*

$$\kappa(x) = -p^*(x). \tag{7.67}$$

Proof: Since zero is an interior point of $dom\,p$, it follows from Corollary 7.1 that the pair $(\mathbb{K}_o, V_o)$ is exponentially tight, and therefore, given any $\epsilon > 0$ there exists a compact subset $\widehat{K_\epsilon}$ and N_ϵ so that

$$\mathbb{K}_n[\widehat{K_\epsilon}] \geq 1 - \epsilon \ , \forall n \geq N_\epsilon; \tag{7.68}$$

since every measure $\mathbb{K}_n$ is tight, we can find another compact subset K_ϵ such that (7.68) holds for any n with K_ϵ instead of $\widehat{K_\epsilon}$. This shows that $\mathbb{K}_o$ is relatively compact. We show that $\kappa(x) = -p^*(x)$. Let $x \in ridom\,p^*$; then there exists $t \in dom\,p$ such that $x \in \partial p(t)$. The point t is in the interior of $dom\,p$ since p is essentially smooth and hence $x = grad\,p(t)$. Theorem 7.2 implies that $\kappa(x) = -p^*(x)$. If $x \notin dom\,p^*$, then

$$\overline{\kappa}^t(x) = \overline{\kappa}(x) - p(t) + \langle t, x\rangle \leq 0 \tag{7.69}$$

implies that $\overline{\kappa}(x) = -\infty$; similarly, we have $\underline{\kappa}(x) = -\infty$. Thus $\kappa(x) = -p^*(x)$ holds for $x \in ridom\,p^*$ and for $x \notin dom\,p^*$. Let $y \in dom\,p^* \backslash ridom\,p^*$ and $x \in ridom\,p^*$; the convexity of p^* and (7.66) imply

$$\limsup_{\mu\to 1} p^*((1-\mu)x + \mu y) \leq p^*(y); \tag{7.70}$$

on the other hand, since p^* is lsc,

$$\liminf_{\mu \to 1} p^*((1-\mu)x + \mu y) \geq p^*(y); \tag{7.71}$$

therefore

$$\lim_{\mu \to 1} \kappa((1-\mu)x + \mu y) = \lim_{\mu \to 1} \underline{\kappa}((1-\mu)x + \mu y) = -p^*(y) \leq \underline{\kappa}(y), \tag{7.72}$$

because $\underline{\kappa}$ is usc. But (7.69) implies $\overline{\kappa}(y) \leq -p^*(y)$. □

Remark: Theorem 7.3 is known as the Gärtner-Ellis Theorem (see [G], [E]).

Acknowledgements: We thank W.Vervaat for sending us pre-prints of his papers, G.Ben Arous for drawing our attention to the work of Takahashi, J.V.Pulé for discussions in the early stages of this work and T.C.Dorlas for sending us a copy of an unpublished manuscript on Radon measures; above all, we thank W.G.Sullivan for reading critically several earlier drafts of this paper and for his suggestions which improved some of the proofs.

References

[A] Azencott,R.: Grandes déviations et applications in Lecture Notes in Mathematics **774**, 2-176 Springer, Berlin (1980)

[B] Brønsted,A.: Conjugate convex functions in topological vector spaces *Mat.Fys.Medd.Dan.Vid.Selk.* **34**, no.2, 3-27 (1964)

[BZ] Bahadur,R.R., Zabel,S.L.: Large deviations of the sample mean in general vector spaces *Ann. Prob.* **7**, 587-621 (1979)

[C] Cramèr,H.: Sur un théorème-limite de la théorie des probabilités, Actualités scientifiques et industrielles **736**, 5-23 (1938)

[Ch] Chernoff,H.: A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations *Ann. Math. Stat.* **23** 493-507 (1952)

[DS] Deuschel,J.-D., Stroock D.W.: Large Deviations. Academic Press, New-York (1989)

[DV] Donsker,M.D.,Varadhan,S.R.S.: Asymptotics for the Polaron *Comm. Pure Appl. Math.* **36**,505-528 (1983)

[E] Ellis,R.S.: Entropy, Large Deviations and Statistical Mechanics. Spinger, Berlin (1985)

[F] Feller,W.: An Introduction to Probability Theory and its Applications, vol.2. Wiley, New-York (1971)

[FW] Freidlin,M.L., Wentzell,A.D.: Random Perturbations of Dynamical Systems. Springer, Berlin (1984)

[G] Gärtner,J.: On large deviations from the invariant measure *Th. Prob. Appl.* **22**, 24-39 (1977)

[K] Kelley,J.L.: General Topology. Van Nostrand, Princeton (1955)

[Kh] Khinchin,A.Ya.: Über einen neuen Grenzwertsatz der Wahrscheinlichkeitsrechnung *Math. Annalen* **101**, 745-752 (1929)

[L] Lanford,O.E.: Entropy and equilibrium states in classical statistical mechanics, in Lecture Notes in Physics **20**, 1-113 Springer, Berlin (1973)

[LS] Lynch,J., Sethuraman,J.: Large deviations for processes with independent increments *Ann. Prob.* **15**, 610-627 (1987)

[LPS] Lewis,J.T., Pfister,C.-E., Sullivan,W.G.: Large deviations and the thermodynamic formalism: a new proof of the eqivalence of ensembles, preprint, DIAS-STP-93-24, to appear in Leuven Conference Proceedings: Micro, Meso and Macroscopic Approaches to Physics (eds. M.Fannes, C.Maes and A.Verbeure) Plenum, New York (1994)

[O] Orey,S.: Large deviations in ergodic theory, in Seminar on Stochastic Processes, 195-249. Birkhäuser, Boston (1986)

[OV] O'Brien,G.L., Vervaat,W.: Capacities, large deviations and loglog laws, in Stable Processes (eds. S.Cambanis, G.Samorodnitsky and M.S.Taqqu) 43-83 (1991)

[R1] Ruelle,D.: Correlation functionals *J. Math. Phys.* **6**, 201-220 (1965)

[R2] Ruelle,D.: Thermodynamic Formalism: the Mathematical Structures of Classical Equilibrium Statistical Mechanics. Addison-Wesley, Reading, Mass. (1978)

[RR] Robertson,A.P., Robertson,W.J.: Topological Vector Spaces. Cambridge University Press, Cambridge (1964)

[S] Schilder,M.: Some asymptotics formulae for Wiener integrals *Trans. Amer. Math. Soc.* **125**, 63-85 (1966)

[Sc] Schwartz,L.: Radon Measures. Oxford University Press, Oxford (1973)

[T] Takahashi,Y.: Entropy functional (free energy) for dynamical systems and their random perturbations, in Proceedings Taniguchi Symposium on Stochastic Analysis (ed. K. Itô), 437-467. Kinokuniya and North Holland, Tokyo (1982)

[V] Varadhan,S.R.S.: Asymptotic probabilities and differential equations *Commun. pure and appl. math.* **19**, 261-286 (1966)

[Ve] Vervaat,W.: Random upper semicontinuous functions and extremal processes, CWI Report MS-R8801, Centrum voor Wiskunde en Informatica, Amsterdam (1989)