typo 8.80

DIAS-STP 94-43

ENTROPY, CONCENTRATION OF PROBABILITY AND CONDITIONAL LIMIT THEOREMS

J.T. Lewis[a], C.-E. Pfister[b], W.G. Sullivan[ac]

[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

[c]University College
Department of Mathematics
Belfield, Dublin 4, Ireland

Abstract: We provide a framework in which a class of conditional limit theorems can be proved in an unified way. We introduce three concepts: a concentration set for a sequence of probability measures, generalizing the Weak Law of Large Numbers; conditioning with respect to a sequence of sets which satisfies a regularity condition; the asymptotic behaviour of the information gain of one sequence of probability measures with respect to another. These concepts are required for the statement of our main abstract result, Theorem 5.1, which describes the asymptotic behaviour of the information gain of a sequence of conditioned measures with respect to a sequence of tilted measures. Provided certain natural convexity assumptions are satisfied, it follows that conditional limit theorems are valid in great generality; this is the content of Theorem 6.1. We give several applications of the formalism, both for independent and weakly dependent random variables, extending in all cases previously known results. For the empirical measure, we provide a conditional limit theorem and give an alternative proof of the Large Deviation Principle. We discuss also the problem of equivalence of ensembles for lattice models in Statistical Mechanics.

Mathematics subject classification: 60B10, 60B12, 60F05, 60F10, 60K35, 82B05, 82B20.

1 Introduction

The purpose of this paper is to develop a framework in which conditional limit theorems can be proved. We have in mind a class of limit theorems of which the following, due to van Campenhout and Cover, is an early example:

Theorem 1.1 ([CC]) *Let $Y_1, Y_2, \ldots$ be i.i.d. random variables having uniform probability mass on the range $\{1, 2, \ldots, m\}$. Then, for $1 \leq \alpha \leq m$ and for all $x \in \{1, 2, \ldots m\}$, we have*

$$\lim_{\substack{n\to\infty \\ n\alpha \text{ integer}}} \operatorname{Prob}\{Y_1 = x \mid \frac{1}{n}\sum_{i=1}^{n} Y_i = \alpha\} = \beta^{\lambda}(x), \tag{1.1}$$

where

$$\beta^{\lambda}(x) = \frac{e^{\lambda x}}{\sum_{k=1}^{m} e^{\lambda k}} \tag{1.2}$$

and the constant λ is chosen to satisfy the constraint $\sum_k k\beta^{\lambda}(k) = \alpha$.

A landmark in the development of such theorems is the paper by Csiszár [C], in which several important concepts are introduced. When one attempts to generalize Csiszár's results, one finds the need to make distinctions which do not arise in the i.i.d. setting. For example, information gain arises in [C] in two ways: it serves as the rate–function of the empirical distribution and as the tool used to compare probability measures through the Kemperman–Pinsker inequality; only the second of these functions survives in the general setting. We introduce three concepts in our analysis of the structure of conditional limit theorems:

- a concentration set for a sequence of probability measures, generalizing the Weak Law of Large Numbers;
- conditioning with respect to a sequence of sets which satisfies a regularity condition;
- the asymptotic behaviour of the information gain of one sequence of probability measures with respect to another.

These concepts are related to ones introduced by Csiszár [C]: the first is related to the generalized I–projection; the second, to the Sanov property; the third, to the concept of asymptotically quasi–independence. Concentration of measures and regular conditioning sequences of sets are defined and studied in part I. Fundamental to all this is the notion of the Ruelle–Lanford function (RL–function) through which we express the large deviation aspects of the problem [LP]. In part II, we study some properties of the information gain of a sequence of conditioned measures with respect to a sequence of tilted measures. Here substantial use is made of convexity theory. In part III, on the basis of the results of parts I and II, we prove conditional limit theorems and study the question of equivalence of ensembles in Statistical Mechanics. We give also an alternative proof of the Large Deviation Principle for empirical measures. For the reader's convenience, we summarize below our main results. First we set the notation and recall some basic facts. We follow essentially the setting of [LP]; however, in the course of this work, we have found it useful to

take a slightly more general point of view and this has led to some modifications of the framework established in [LP].
Throughout the paper, $(X, \mathcal{B})$ is a measurable space and $\mathcal{B}$ is the collection of measurable subsets of X. It is essential for our purposes that X have some topological structure; we assume the minimum required for our purposes:

- X is a Hausdorff topological space;
- each point x of X has a local base of measurable subsets (that is, each open set containing x contains a measurable neighbourhood of x).

Often we choose $\mathcal{B}$ to be the Borel σ–algebra of X, but this is not always the case; there are some interesting examples in which $\mathcal{B}$ is not the Borel σ–algebra. This approach obviates the discussion of non–measurable sets, required in [C].

We denote the closure of a subset A of X by $\mathrm{cl}\, A$ and its interior by $\mathrm{int}\, A$. We adopt the following convention: G always denotes a measurable neighbourhood and B a measurable subset.

We use $\overline{\mathbb{R}}$ to denote the extended real line: $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$; for a,b in $\overline{\mathbb{R}}$, we define

$$a \vee b := \max\{a, b\}, \tag{1.3}$$

$$a \wedge b := \min\{a, b\}. \tag{1.4}$$

If $f : X \to \overline{\mathbb{R}}$ is an arbitrary function, we put

$$\sup_{x \in \emptyset} f(x) := -\infty. \tag{1.5}$$

Let $\{\mathbb{M}_n\}_{n \geq 1}$ be a sequence of positive measures on $\mathcal{B}$ which are locally finite (that is, for each x in X, there exists G_x such that $G_x \ni x$ and $\mathbb{M}[G_x] < \infty$); let $\{V_n\}_{n \geq 1}$ be a **scale**, that is, an increasing sequence of positive real numbers diverging to $+\infty$ as $n \to \infty$. We are interested in the asymptotics of $\{\mathbb{M}_n\}$ on the scale $\{V_n\}$ as n diverges. Define the set–function

$$m_n[B] := \frac{1}{V_n} \ln \mathbb{M}_n[B]; \tag{1.6}$$

let

$$\overline{m}[B] \; := \; \limsup_{n \to \infty} m_n[B], \tag{1.7}$$

$$\underline{m}[B] \; := \; \liminf_{n \to \infty} m_n[B]. \tag{1.8}$$

The following properties of the set–functions $\underline{m}$, $\overline{m}$ are easily proved. Property (1.13) below is the key to the development; we refer to it as the **Principle of the Largest Term**. It is a consequence of

$$\limsup_{n \to \infty} (a_n \vee b_n) = (\limsup_{n \to \infty} a_n) \vee (\limsup_{n \to \infty} b_n), \tag{1.9}$$

valid for each pair $\{a_n\}_{n \geq 1}, \{b_n\}_{n \geq 1}$ of sequences in $\overline{\mathbb{R}}$.

Lemma 1.1 *On $\mathcal{B}$, we have*

$$-\infty \leq \underline{m}[B] \leq \overline{m}[B] \leq +\infty; \tag{1.10}$$

if $B_1 \subset B_2$, *then*

$$\underline{m}[B_1] \ \leq \ \underline{m}[B_2] \ , \tag{1.11}$$

and

$$\overline{m}[B_1] \ \leq \ \overline{m}[B_2] \ ; \tag{1.12}$$

furthermore, for all B_1 *and* B_2 *in* $\mathcal{B}$, *we have*

$$\overline{m}[B_1 \cup B_2] = \overline{m}[B_1] \vee \overline{m}[B_2]. \tag{1.13}$$

Following Orey [O], we exploit the topology of X to derive from $\overline{m}$ and $\underline{m}$ two auxiliary functions on X, **the lower and upper deviation functions** :

$$\underline{\mu}(x) := \inf_{G_x} \underline{m}[G_x] \ , \ \overline{\mu}(x) := \inf_{G_x} \overline{m}[G_x] \, . \tag{1.14}$$

Because the set-functions $\overline{m}$ and $\underline{m}$ are increasing, the definition (1.14) of $\underline{\mu}(x)$ and $\overline{\mu}(x)$ is independent of the choice of the local base $\{G_x\}$ of measurable neighbourhoods of x.

Definition 1.1 *A pair* $(\{\mathbb{M}_n\}, \{V_n\})$ *has a* **Ruelle-Lanford function** *(RL-function)* μ *if*

$$\overline{\mu}(x) = \underline{\mu}(x) \tag{1.15}$$

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

$$\mu(x) := \overline{\mu}(x) = \underline{\mu}(x). \tag{1.16}$$

Lemma 1.2 is elementary; nevertheless, it contains the two fundamental inequalities of Large Deviation Theory.

Lemma 1.2 *The RL-function* μ *is upper semicontinuous (u.s.c.) and*

$$\underline{m}[B] \geq \sup_{x \in \mathrm{int}\, B} \mu(x) \quad , \quad \text{any } B \in \mathcal{B}; \tag{1.17}$$

$$\overline{m}[B] \leq \sup_{x \in \mathrm{cl} B} \mu(x) \quad , \quad B \text{ relatively compact.} \tag{1.18}$$

Lemma 1.2 can be regarded as an abstract version of Ruelle's treatment of entropy in Statistical Mechanics [Ru1]. Ruelle gave a precise mathematical interpretation of Boltzmann's remarkable formula

$$S = k \ln W \ , \tag{1.19}$$

relating the entropy S of a macroscopic equilibrium state to a measure W of the size of the set of microscopic states corresponding to the macroscopic state. Lanford [L] made explicit the connection with Large Deviations.

Often one needs a stronger version of (1.18), valid for all B in $\mathcal{B}$. Together with a compactness property for μ, the strengthened bounds constitute a Large Deviation Principle (LDP) as defined by Varadhan in [Va1] (see also [Va2], [A],[DS]):

A pair $(\{\mathbb{M}_n\},\{V_n\})$ obeys a **Large Deviation Principle** with **rate–function** s if there exists an u.s.c. function $s : X \to \overline{\mathbb{R}}$ whose level sets $\{x : \; s(x) \geq a\}$, $a \in \mathbb{R}$, are compact, and such that

$$\underline{m}[B] \geq \sup_{x \in \operatorname{int} B} s(x) \quad , \quad \text{any } B \in \mathcal{B}; \tag{1.20}$$

$$\overline{m}[B] \leq \sup_{x \in \operatorname{cl} B} s(x) \quad , \quad \text{any } B \in \mathcal{B};. \tag{1.21}$$

Note: For the remainder of this Introduction, we specialize our results to the case in which the space X is compact; this yields simpler statements. In the main part of the paper, the theorems are stated and proved without this restriction.

The thermodynamic entropy is a concave function; this is not necessarily the case with μ in our general context. However, as with the thermodynamic entropy, there is a "maximum principle" associated with μ: the set on which μ attains its maximum is a concentration set for the sequence $\{\mathbb{M}_n\}$. We say that a sequence $\{\mathbb{M}_n\}$ of probability measures is **eventually concentrated on the set** A if, for any measurable neighbourhood G of A, we have

$$\lim_n \mathbb{M}_n[G] = 1. \tag{1.22}$$

For a sequence $\{\mathbb{M}_n\}$ of probability measures, an RL–function is necessarily non-positive. If an RL–function μ exists, the sequence $\{\mathbb{M}_n\}$ is eventually concentrated on the set

$$A = \{x \in X : \; \mu(x) = 0\} \tag{1.23}$$

on which μ takes its maximum value. If A in (1.23) is a singleton, then (1.22) means that the sequence $\{\mathbb{M}_n\}$ satisfies a Weak Law of Large Numbers. The word "entropy" in the title of this paper refers to the RL–function.

In the rest of this Introduction, we shall assume that an RL–function μ exists for the pair $(\{\mathbb{M}_n\},\{V_n\})$, and that the $\mathbb{M}_n$ are probability measures. Let $C_0 \in \mathcal{B}$ be a non-empty set, and let $C := \operatorname{cl} C_0$ be its closure. We say that $C_0 \in \mathcal{B}$ is **LD–regular** if

1. for n sufficiently large, $0 < \mathbb{M}_n[C_0]$;
2. the limit $\lim_n m_n[C_0]$ exists, is finite and $\lim_n m_n[C_0] = \sup_{x \in C} \mu(x)$.

The notion of an LD–regular set C_0 is closely related to the Sanov property of Csiszár [C]; it coincides with it when C_0 is convex and μ concave on $C = \mathrm{clC_0}$. More generally, we say that a sequence of sets $\{C_n\}$ is **LD–regular** if

1. $\{C_n\}$ is a decreasing sequence of measurable sets and, for n sufficiently large, $0 < \mathbb{M}_n[C_n]$;
2. the closed set $C := \cap_n \operatorname{cl} C_n$ is non–empty;
3. the limit $\lim_n m_n[C_n]$ exists, is finite and $\lim_n m_n[C_n] = \sup_{x \in C} \mu(x)$.

The notion of LD–regularity of a sequence is not a notion of convergence since we do not require that C_n be eventually contained in any neighbourhood of C. Given an LD–regular sequence $\{C_n\}$, we study in Section 3 the corresponding sequence of **conditioned measures** $\{\mathbb{M}_n^C\}$,

$$\mathbb{M}_n^C[B] := \mathbb{M}_n[B|C_n] \ , \ B \in \mathcal{B}. \tag{1.24}$$

In general, we cannot determine the RL–function of this sequence of probability measures; it is possible that it does not exist. However, we prove (see Theorem 3.1) the following useful result:

Theorem 1.2 *Let X be compact and $\{C_n\}$ be LD–regular. Then the sequence of conditioned measures $\{\mathbb{M}_n^C\}$ is eventually concentrated on the non–empty compact set*

$$N_C := \{x \in C : \mu(x) = \sup_{y \in C} \mu(y)\}. \tag{1.25}$$

Part II is devoted to the study of the tilted measures and the comparison of these measures with the conditioned ones. To introduce the tilted measures, we need a convex structure for the space X. It is natural to work with a dual pair (E^*, E) of locally convex topological vector spaces with pairing $(x', x) \in E^* \times E \mapsto \langle x', x\rangle \in \mathbb{R}$. We require that the measurable space $(X, \mathcal{B})$ be a closed convex subset of E with the induced topology. Furthermore, we require that the maps $x \mapsto \langle x', x\rangle$ be $\mathcal{B}$–measurable for every $x' \in E^*$. For convenience, we extend μ to all E by setting $\mu(x) := -\infty$ for $x \in E\backslash X$. A typical example is the following: $(\Omega, \mathcal{F})$ is a measurable space and E is the space $\mathcal{M}(\Omega)$ of all finite signed measures on $(\Omega, \mathcal{F})$; E^* is the space $\mathcal{C}_b(\Omega)$ of all bounded $\mathcal{F}$–measurable functions on Ω; the pairing is given by the bilinear form

$$\langle x', x\rangle = \int_\Omega x'(\omega)x[d\omega] \ , \ x' \in \mathcal{C}_b(\Omega) \ , \ x \in \mathcal{M}(\Omega) \ ; \tag{1.26}$$

the topology on $\mathcal{M}(\Omega)$ is the $\sigma(E, E^*)$–topology: a sequence $\{x_n\}$ of measures converges to a measure x if and only if

$$\lim_n \int_\Omega f(\omega)x_n[d\omega] = \int_\Omega f(\omega)x[d\omega] \ \text{ all } f \in \mathcal{C}_b \ ; \tag{1.27}$$

X is the closed convex subset $\mathcal{M}_1^+(\Omega)$ of all probability measures on $(\Omega, \mathcal{F})$ equipped with the induced topology, and $\mathcal{B}$ is the σ–algebra generated by the maps $x \mapsto \langle x', x\rangle$, $x' \in E'$.

We define on E^* a function p , the **scaled generating function**,

$$p(x') := \lim_n \frac{1}{V_n} \ln \int_X e^{V_n\langle x', x\rangle} \mathbb{M}_n[dx] \ . \tag{1.28}$$

We assume throughout this paper that p is well-defined, possibly non–finite; it is automatically convex. The **essential domain** of p is the subset of E^* defined by

$$\operatorname{dom} p := \{x' \in E^* : p(x') \text{ is finite}\}. \tag{1.29}$$

For X compact, Varadhan's Theorem implies that $\operatorname{dom} p = E^*$ and that p is the conjugate of the function $-\mu$:

$$p(x') = (-\mu)^*(x') := \sup_{x \in E}\{\langle x', x\rangle + \mu(x)\}. \tag{1.30}$$

For any $x' \in \operatorname{dom} p$, we define the **tilted measure** $\mathbb{M}_n^{x'}$ by the formula

$$\mathbb{M}_n^{x'}[B] := \frac{\int_B e^{V_n\langle x', x\rangle}\mathbb{M}_n[dx]}{\int_X e^{V_n\langle x', x\rangle}\mathbb{M}_n[dx]}, \quad B \in \mathcal{B}. \tag{1.31}$$

An RL–function $\mu^{x'}$, given by

$$\mu^{x'}(x) = \mu(x) + \langle x', x\rangle - p(x'), \tag{1.32}$$

exists for the pair $(\{\mathbb{M}_n^{x'}\}, \{V_n\})$, and the sequence $\{\mathbb{M}_n^{x'}\}$ is eventually concentrated on the non–empty compact set

$$N^{x'} := \{x \in X : \mu^{x'}(x) = 0\}. \tag{1.33}$$

The central concept of part II is the notion of asymptotically I–nullness. In order to compare the asymptotic properties of two sequences of probability measures, we make use of the information gain. Recall that the **information gain** $\mathcal{H}(\lambda_1|\lambda_2)$ of two probability measures λ_1 and λ_2 defined on the same space $(\Omega, \mathcal{F})$ is

$$\mathcal{H}(\lambda_1|\lambda_2) := \begin{cases} \int_\Omega \ln h(\omega)\lambda_1[d\omega], & \text{if } \lambda_1[d\omega] = h(\omega)\lambda_2[d\omega], \\ +\infty, & \text{otherwise.} \end{cases} \tag{1.34}$$

Let $\{\mathbb{K}_n\}$ and $\{\mathbb{Q}_n\}$ be two sequences of probability measures. We say that the sequence $\{\mathbb{K}_n\}$ is **asymptotically I–null to the sequence $\{\mathbb{Q}_n\}$ on the scale $\{V_n\}$** if

$$\lim_n \frac{1}{V_n}\mathcal{H}(\mathbb{K}_n|\mathbb{Q}_n) = 0. \tag{1.35}$$

The notion of asymptotically I–nullness is a generalization of the notion of asymptotically quasi–independence, introduced in [C]; it is not a symmetric relation. To get some feeling for its significance, consider a sequence $\{\mathbb{Q}_n\}$ which is eventually concentrated on a set A at an **exponential rate** on the scale $\{V_n\}$: instead of (1.22), the stronger statement

$$\limsup_n \frac{1}{V_n}\ln \mathbb{Q}_n[X\backslash G] < 0 \tag{1.36}$$

holds for any measurable neighbourhood G of A . If, in addition, (1.35) holds, then the sequence $\{\mathbb{K}_n\}$ is eventually concentrated on the set A, not necessarily at an exponential rate on the scale V_n; this is the content of Theorem 2.3

One of the main result of part II is Lemma 5.1 which gives the following bounds: if $\{C_n\}$ is LD–regular and X compact, then

$$\begin{aligned} 0 \ &\leq \ \inf_{x \in N_C}\{-\mu(x) - p^*(x)\} \leq \liminf_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) \\ &\leq \ \limsup_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) \leq \sup_{x \in N_C}\{p(x') - \langle x', x\rangle - \mu(x)\}\ ; \end{aligned} \tag{1.37}$$

here p^* is the conjugate of p . The next theorem, a special case of Theorem 5.1, follows immediately from these inequalities.

Theorem 1.3 *Let X be compact and $\{C_n\}$ LD-regular. If $N_C \subset N^{x'}$, then $\{\mathbb{M}_n^C\}$ is asymptotically I-null to $\{\mathbb{M}_n^{x'}\}$.*

The condition $N_C \subset N^{x'}$ has an immediate geometric interpretation: when μ is concave and (1.30) holds, each x in $N^{x'}$ satisfies

$$\mu(x) + \langle x', x\rangle = p(x') \geq \mu(y) + \langle x', y\rangle \ , \ \ \forall y \in E \ ; \tag{1.38}$$

writing $y = x + z$, we have

$$-\mu(x+z) \geq -\mu(x) + \langle x', z\rangle \ , \ \ \forall z \in E \ ; \tag{1.39}$$

in the language of convex analysis, x' is a subgradient of $-\mu$ at x. When C is convex, we verify the condition $N_C \subset N^{x'}$ by showing the existence of a subgradient x'; see Section 6. This result, Theorem 6.1, is a theorem of convex analysis; it is a consequence of the Hahn–Banach Theorem.

Theorem 1.4 *Let X be compact and $\{C_n\}$ LD-regular. Let $C = \cap_n \mathrm{cl}\, C_n$ be convex and let μ coincide on C with its concave envelope. If $\mathrm{int}\, C$ is non-empty or μ is continuous at some point of C, then there exists $x' \in E^*$ such that $N_C \subset N^{x'}$ and x' is a subgradient of $-\mu$ at x for all $x \in N_C$.*

In part III, we apply the general formalism to prove conditional limit theorems. The spaces X, E and E^* are as above. Let $(S, \mathcal{S})$ be a standard Borel space; for each $i \in \mathbb{Z}^d$, let $(\Omega_i, \mathcal{F}_i)$ be a copy of $(S, \mathcal{S})$ and define $(\Omega, \mathcal{F})$ as the product space. There is a natural action on Ω of $\mathbb{Z}^d$ as group of translations. This action lifts to the space of random variables on Ω, and to the space of probability measures $\mathcal{M}_1^+$ on Ω; the action of translation by $j \in \mathbb{Z}^d$ is denoted in all cases by θ_j. Let $\mathcal{F}_{\mathrm{qloc}}$ be the space of quasilocal functions on $(\Omega, \mathcal{F})$; we equip the space $\mathcal{M}_1^+$ with the $\sigma(\mathcal{M}_1^+, \mathcal{F}_{\mathrm{qloc}})$-topology: a sequence $\{\nu_n\}$ converges to ν if and only if for every $f \in \mathcal{F}_{\mathrm{qloc}}$

$$\lim_n \int_\Omega f(\omega)\nu_n[d\omega] = \int_\Omega f(\omega)\nu[d\omega]. \tag{1.40}$$

We choose an increasing sequence $\{\Lambda_n\}$ of cubes in $\mathbb{Z}^d$, each cube Λ_n being centered at the origin and having cardinality $V_n = (2n+1)^d$. On the space of random variables and on the space $\mathcal{M}_1^+$, we define the averaging operation

$$\mathcal{A}_n := \frac{1}{V_n} \sum_{j \in \Lambda_n} \theta_j. \tag{1.41}$$

Let $\varphi : \Omega \to X$ be a random variable whose distribution is given by the probability measure $\beta \in \mathcal{M}_1^+$. Define for each n the random variable $T_n : \Omega \to X$ by

$$T_n(\omega) := \mathcal{A}_n \varphi(\omega) \tag{1.42}$$

and put $\mathbb{M}_n := \beta \circ T_n^{-1}$. . In Section 8, we consider the case where β is a weakly dependent translation invariant probability measure. (Weak-dependence is defined in Section 8.1; examples of weakly dependent measures are Gibbs measures defined by a local specification with an absolutely summable potential.) On the space $\mathcal{M}_1^{+,\theta}$

of translation invariant probability measures, the specific information gain $h(\lambda|\beta)$ is well-defined; on the space $\mathcal{F}_{\mathrm{qloc}}$, the scaled generating function

$$p(f|\beta) := \lim_n \frac{1}{V_n} \ln \int_\Omega \exp\{\sum_{j\in\Lambda_n} f(\theta_j\omega)\}\beta[d\omega] \tag{1.43}$$

is well-defined. In fact, these two functions are conjugate to each other; this statement is the content of the variational principle in Statistical Mechanics. Let f belong to $\mathcal{F}_{\mathrm{qloc}}$; for any $\lambda \in \mathcal{M}_1^{+,\theta}$, we have

$$\int_\Omega f(\omega)\lambda[d\omega] \le p(f|\beta) + h(\lambda|\beta). \tag{1.44}$$

We say that λ is an (f,β)–**equilibrium state** if

$$\int_\Omega f(\omega)\lambda[d\omega] = p(f|\beta) + h(\lambda|\beta). \tag{1.45}$$

The set of such states is non-empty and convex; it is not necessarily a singleton. Suppose that, for $x' \in E^*$, the function $f_\varphi^{x'} : \Omega \to \mathbb{R}$ defined by

$$f_\varphi^{x'}(\omega) := \langle x', \varphi(\omega)\rangle \tag{1.46}$$

is quasilocal; we define

$$p_n(x') := \frac{1}{V_n} \ln \int_\Omega \exp\{\sum_{j\in\Lambda_n} f_\varphi^{x'}(\theta_j\omega)\}\beta[d\omega] \tag{1.47}$$

and set

$$\beta_n^{x'}[d\omega] := \exp\{\sum_{j\in\Lambda_n} f_\varphi^{x'}(\theta_j\omega) - V_n p_n(x')\}\beta[d\omega], \tag{1.48}$$

and

$$\beta_n^C[d\omega] := \beta[d\omega|T_n \in C_n], \tag{1.49}$$

where $C_n \in \mathcal{B}$ is a sequence of sets with $\beta[T_n \in C_n] > 0$. Recall that for every $\rho \in \mathcal{M}_1^{+,\theta}$

$$h(\rho|\beta) = \lim_n \frac{1}{V_n}\mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\rho|\beta) \tag{1.50}$$

exists, and is non-negative.

Theorem 1.5 *In the above setting, assume that there exists $x' \in E^*$ such that the function $f_\varphi^{x'}$ is quasilocal and*

$$\lim_n \frac{1}{V_n}\mathcal{H}(\beta_n^C|\beta_n^{x'}) = 0. \tag{1.51}$$

Then the set of limit points of the sequence

$$\{\mathcal{A}_n\beta[\ \cdot\ |T_n \in C_n]\} \tag{1.52}$$

is non-empty, and any limit point β^C satisfies the identity

$$h(\beta^C|\beta) = -\lim_n \frac{1}{V_n} \ln \beta[T_n \in C_n] = \int_\Omega f_\varphi^{x'}(\omega)\beta^C[d\omega] - p(x'). \tag{1.53}$$

In particular β^C is an $(f_\varphi^{x'},\beta)$–equilibrium state.

We give two applications of this theorem, one when φ is a quasilocal $\mathbb{R}^k$–valued function (Theorem 8.4), and another one when $\varphi(\omega) = \delta_\omega$, where δ_ω is the Dirac mass at ω, so that T_n is the empirical measure (Theorem 8.5). Our formalism yields an alternative proof of the Large Deviation Principle for empirical measures: there is a natural embedding of the space of probability measures in the unit ball of the dual of the Banach space of quasilocal functions on Ω equipped with the weak*–topology; since the unit ball is compact in that topology, we have an LDP when the random variable T_n is regarded as taking values in the unit ball; by a simple argument, we show that $\mu(x) > -\infty$ implies that x is a translation invariant probability measure. In the special case where β is a product measure and φ depends only on the value of ω at 0, $\varphi(\omega) = \varphi(\omega_0)$, we can use an argument due to Csiszár [C] to prove directly a theorem extending the van Campenhout– Cover Theorem (Theorem 7.2).

In Section 9, we deal with the question of equivalence of ensembles in Statistical Mechanics [Gi]. Theorem 8.4 is reformulated in the standard framework of lattice systems, and the connection with the theory of Gibbs states is made. We give proofs of the results announced in [LPS1] and [LPS2]. Since Gibbs' time, many proofs have been offered of the equivalence of ensembles. We refer to [LPS1] and [LPS2] for some (incomplete) remarks on the history of the question. The recent works [DSZ], [RZ] and [G1] all approach the problem of equivalence of ensembles through the Large Deviation Principle for empirical measures; our large deviation analysis, based on Theorem 8.4, is less technical and yet more natural; it has the merit of yielding more precise results. The main advantage of the large deviation analysis, common to both approaches, is that it permits the treatment of systems with phase transitions.
The essential features of our approach are these: we concentrate attention on a sequence $\{T_n\}$ of generalized energy functions taking values in $\mathbb{R}^k$; we apply our formalism to the sequence $\mathbb{M}_n$ of probability distributions on $\mathbb{R}^k$, where $\mathbb{M}_n$ is the distribution of T_n. In this case, the Ruelle-Lanford function μ is concave and is precisely the thermodynamic entropy, the scaled generating function is the grand canonical pressure and x' is the generalized chemical potential which now lies in $\mathbb{R}^k$. We prove that, provided the sequence $\{C_n\}$ of sets we use for conditioning is LD-regular, the set of limit-points of the sequence

$$\{\mathcal{A}_n\beta[\,\cdot\,|T_n \in C_n]\} \tag{1.54}$$

of averaged conditioned measures is non-empty and each limit-point β^C is an equilibrium state characterized by the generalized chemical potential x'. Moreover, x' is characterized as a subgradient of $-\mu$ at any point of the non-empty compact set N_C. In typical situations in statistical mechanics, the thermodynamic entropy is C^1 on the interior of its essential domain, and then x' is given by $x' = -\mathrm{grad}\mu(x)$, $x \in N_C$. We obtain very satisfactory results concerning a subclass of translation invariant microcanonical states; to extend these results to non–translation invariant states is an open problem. The theory of Large Deviations works well, even in the presence of phase transitions, because of its thermodynamic character: it exploits the properties of thermodynamic potentials, the RL–function and the scaled generating function. On the other hand, it seems that its thermodynamic character restricts it to those equilibrium states which are translation invariant. It is an interesting and difficult problem to consider limits of the sequence

$$\{\beta[\,\cdot\,|T_n \in C]\} \tag{1.55}$$

of conditioned measures without averaging; the techniques of large deviations do not apply when the limits are not translation-invariant because the rate-function of the distribution of the empirical measure is non-trivial on the translation-invariant measures alone.

Acknowledgement: This work was partially supported by the European Union under the Human Capital and Mobility Scheme.(EU contract CHRX-CT93-0411) and by Fonds National Suisse de la Recherche Scientifique.

Part I: Concentration of Probability and Conditioning

2 Concentration of Probability

Throughout this section, we assume that $\{\mathbb{M}_n\}$ is a sequence of probability measures. Given a scale $\{V_n\}$, the upper deviation function determines a set on which the measures are eventually concentrated (Theorem 2.2 and the comment following it); the usefulness of this information about $\{\mathbb{M}_n\}$ depends on how well we have chosen the scale $\{V_n\}$.

Definition 2.1 *Let $\{\mathbb{M}_n\}$ be a sequence of probability measures on $\mathcal{B}$; we say that $\{\mathbb{M}_n\}$ is* **eventually concentrated** *on a set A if, for each measurable neighbourhood G of A, we have*

$$\lim_{n\to\infty} \mathbb{M}_n[G] = 1 . \tag{2.1}$$

This definition is a hypothesis of the following theorem which provides, via Lemma 5.1, the essential bounds for our main results, Theorem 5.1 and Theorem 6.1.

Theorem 2.1 *Let $\{\mathbb{M}_n\}$ be a sequence of probability measures on $\mathcal{B}$, let $f : X \to \overline{\mathbb{R}}$ be a measurable function and let $\{B_n\}$ be a sequence of measurable subsets of X such that*

$$\lim_{n\to\infty} \mathbb{M}_n[B_n] = 1 . \tag{2.2}$$

Suppose that $\{\mathbb{M}_n\}$ is eventually concentrated on a subset N of X, and that each open set containing N contains a measurable neighbourhood of N.

a) If f is lower semicontinuous and uniformly bounded below on B_n for n sufficiently large, then

$$\inf_{x\in N} f(x) \le \liminf_{n\to\infty} \int_{B_n} f(x)\, \mathbb{M}_n[dx] . \tag{2.3}$$

b) If f is upper semicontinuous and uniformly bounded above on B_n for n sufficiently large, then

$$\limsup_{n\to\infty} \int_{B_n} f(x)\, \mathbb{M}_n[dx] \le \sup_{x\in N} f(x) . \tag{2.4}$$

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 f. Let G be a measurable neighbourhood of N; for each sufficiently large n we have

$$\int_{B_n} f(x)\, \mathbb{M}_n[dx] \le [\sup_{x\in G} f(x)]\, \mathbb{M}_n[B_n \cap G] + [\sup_{x\in B_n} f(x)]\, \mathbb{M}_n[B_n \backslash G] \tag{2.5}$$

and

$$\lim_{n\to\infty} \mathbb{M}_n[G] = 1 ; \tag{2.6}$$

since

$$\lim_{n\to\infty} \mathbb{M}_n[B_n] = 1, \tag{2.7}$$

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

$$\mathbb{M}_n[B_n \cap G] \geq 1 - \varepsilon , \tag{2.8}$$

and

$$\mathbb{M}_n[B_n \backslash G] \leq \varepsilon , \tag{2.9}$$

for all n sufficiently large. Thus we have

$$\limsup_{n\to\infty} \int_{B_n} f(x)\, \mathbb{M}_n[dx] \leq \sup_{x\in G} f(x) \tag{2.10}$$

for every measurable neighbourhood G of N. Since every open set containing N contains a measurable neighbourhood G of N, the upper semicontinuity of f implies

$$\inf_{G\supset N} \sup_{x\in G} f(x) = \sup_{x\in N} f(x) , \tag{2.11}$$

and hence the lemma follows. □

Later, we shall make use of the fact that any compact set N has the property that any open set containing N contains a measurable neighbourhood of N. Of course, when $\mathcal{B}$ is the Borel σ–algebra of X, this property holds for an arbitrary subset N. We mention also the following particular case of Theorem 2.1:
Recall ([S]) that a sequence $\{\mathbb{M}_n\}$ of Radon measures *converges narrowly* to a Radon measure $\mathbb{M}$ if and only if, for every bounded u.s.c. function f on X, we have

$$\limsup_n \mathbb{M}_n[f] \leq \mathbb{M}[f] . \tag{2.12}$$

The Weak Law of Large Numbers: *Let $\mathcal{B}$ be the Borel σ–algebra of the Hausdorff space X; let $\{\mathbb{M}_n\}$be a sequence of Radon measures on $\mathcal{B}$ which is eventually concentrated on the set $N = \{x_*\}$; then $\{M_n\}$ converges narrowly to the Dirac measure δ_{x_*}.*

To proceed further, we need to be able to identify a set on which a sequence of probability measures is eventually concentrated; we shall prove that the null-set of an RL-function is such a set. However, we are not always able to compute the RL-function on a given scale, even when we can prove it exists; for that reason, the following result is important.

Theorem 2.2 *Let $\{\mathbb{M}_n\}$ be a sequence of probability measures on $\mathcal{B}$ and let $\{V_n\}$ be a scale. Suppose there exists a function s which is u.s.c., has compact level–sets and the upper bound*

$$\overline{m}[B] \leq \sup_{x\in \mathrm{cl}\, B} s(x) \tag{2.13}$$

holds. Then

a) the set

$$N_s := \{x \in X, s(x) \geq 0\} \tag{2.14}$$

is non-empty and compact;

b) the sequence $\{\mathbb{M}_n\}$ is eventually concentrated on the set N and, for any measurable neighbourhood G of N, we have

$$\limsup_n \frac{1}{V_n} \ln \mathbb{M}_n[X\backslash G] \; < \; 0 \; . \tag{2.15}$$

Proof: Applying the upper bound to the set $B = X$, we have

$$\overline{m}[X] \leq \sup_{x \in X} s(x). \tag{2.16}$$

Since $\{\mathbb{M}_n\}$ is a sequence of probability measures, we have $\mathbb{M}_n[X] = 1$; hence

$$\sup_{x \in X} s(x) \geq 0. \tag{2.17}$$

Since s is u.s.c. and has compact level–sets, the supremum of s is attained on any closed set, in particular, on X; thus the set

$$N_s = \{x \in X, s(x) \geq 0\}$$

is a non–empty compact subset of X and (a) is proved.

Fix a in $(-\infty, 0)$; the level-set $L_a := \{x \in X : s(x) \geq a\}$ is compact and non-empty. Let G be a measurable neighbourhood of N_s; there are two cases to be considered: (a) $\mathrm{cl}\,(X\backslash G) \cap L_a$ is empty; since $s(x) < a$ on $\mathrm{cl}\,(X\backslash G)$, we have

$$\sup_{x \in \mathrm{cl}\,(X\backslash G)} s(x) \leq a < 0 \; ; \tag{2.18}$$

(b) $\mathrm{cl}\,(X\backslash G) \cap L_a$ is non–empty; then $\mathrm{cl}\,(X\backslash G) \cap L_a$ is a non–empty compact set and, since s is u.s.c., there exists $x_a \in \mathrm{cl}\,(X\backslash G) \cap L_a$ such that

$$\sup_{x \in \mathrm{cl}\,(X\backslash G)} s(x) = s(x_a) < 0, \tag{2.19}$$

because N_s is disjoint from $\mathrm{cl}\,(X\backslash G)$. In either case, we have

$$\sup_{x \in \mathrm{cl}\,(X\backslash G)} s(x) < 0 \tag{2.20}$$

so that

$$\overline{m}[X\backslash G] \leq \sup_{x \in \mathrm{cl}\,(X\backslash G)} s(x) < 0 \; , \tag{2.21}$$

establishing (2.15). It follows that

$$\lim_{n \to \infty} \mathbb{M}_n[X\backslash G] = 0 \tag{2.22}$$

and hence

$$\lim_{n \to \infty} \mathbb{M}_n[G] = 1. \tag{2.23}$$

□

Comment: In order to have the best result, one must find the smallest possible function s with the properties mentioned in Theorem 2.2. If the upper deviation function $\overline{\mu}$ has compact level–sets and if it has the upper bound property (2.13), then $\overline{\mu}$ is the best possible choice for Theorem 2.2 whenever the space X is regular (Lemma 5.1 in [LP]). Moreover, if the pair $(\{\mathbb{M}_n\}, \{V_n\})$ is **exponentially tight**, that is, if there exists a sequence $\{K_n\}_{n\geq 1}$ of measurable relatively compact subsets of X such that

$$\limsup_{n\to\infty} \overline{m}[X\backslash K_n] = -\infty, \tag{2.24}$$

then the upper deviation function $\overline{\mu}$ has the upper bound property (2.13) and has compact level–sets (see Lemmas 5.2 and 5.3 in [LP]; there the proofs are given in the case of $\mathcal{B}$ the Borel σ–algebra of X, but they hold with easy modifications in the general case). Thus the hypotheses of Theorem 2.2 are satisfied with $s = \overline{\mu}$ in the following cases:

- X is compact;
- $(\{\mathbb{M}_n\}, \{V_n\})$ is exponentially tight;
- the pair $(\{\mathbb{M}_n\}, \{V_n\})$ obeys an LDP with RL–function μ as rate-function.

Notice that, in Theorem 2.2, we proved a little more than that the sequence $\{\mathbb{M}_n\}$ is eventually concentrated on the set N_s : in proving (2.15), we established a bound on the rate at which the measure of a set in the complement of N_s goes to zero. The behaviour described by (2.15) is worth naming.

Definition 2.2 *Let $\{\mathbb{M}_n\}$ be a sequence of probability measures; if, for every measurable neighbourhood G of N, we have*

$$\limsup_{n} \frac{1}{V_n} \ln \mathbb{M}_n[X\backslash G] \; < \; 0 \; , \tag{2.25}$$

we say that the sequence $\{\mathbb{M}_n\}$ is **eventually concentrated on N at an exponential rate on the scale $\{V_n\}$**.

Definition 2.3 *Let $\{\mathbb{S}_n\}$ and $\{\mathbb{Q}_n\}$ be two sequences of probability measures on the same space. The sequence $\{\mathbb{S}_n\}$ is* **asymptotically I–null to the sequence $\{\mathbb{Q}_n\}$ on the scale $\{V_n\}$** *if*

$$\lim_{n} \frac{1}{V_n} \mathcal{H}(\mathbb{S}_n|\mathbb{Q}_n) = 0, \tag{2.26}$$

where $\mathcal{H}(\mathbb{S}_n|\mathbb{Q}_n)$ is the specific information gain of $\mathbb{S}_n$ with respect to $\mathbb{Q}_n$.

Theorem 2.3 *Let $\{\mathbb{Q}_n\}$ be a sequence of probability measures which is eventually concentrated on N at an exponential rate on the scale $\{V_n\}$. If $\{\mathbb{S}_n\}$ is a sequence of probability measures which is asymptotically I–null to $\{\mathbb{Q}_n\}$ on the scale $\{V_n\}$, then $\{\mathbb{S}_n\}$ is eventually concentrated on N.*

Proof: We note that

$$\begin{aligned}\mathcal{H}(\mathbb{S}_n|\mathbb{Q}_n) &\geq \mathbb{S}_n[G]\ln\frac{\mathbb{S}_n[G]}{\mathbb{Q}_n[G]} + \mathbb{S}_n[X\backslash G]\ln\frac{\mathbb{S}_n[X\backslash G]}{\mathbb{Q}_n[X\backslash G]} \\ &\geq -\ln 2 - \mathbb{S}_n[X\backslash G]\ln\mathbb{Q}_n[X\backslash G].\end{aligned} \tag{2.27}$$

Since

$$\limsup_n \frac{1}{V_n}\ln\mathbb{Q}_n[X\backslash G] < 0, \tag{2.28}$$

there exists $\delta > 0$ such that, for all n sufficiently large, we have

$$\frac{1}{V_n}\ln\mathbb{Q}_n[X\backslash G] < -\delta. \tag{2.29}$$

Thus we have

$$\frac{1}{V_n}\mathcal{H}(\mathbb{S}_n|\mathbb{Q}_n) \geq \frac{-\ln 2}{V_n} + \delta\cdot\mathbb{S}_n[X\backslash G] \geq \frac{-\ln 2}{V_n}; \tag{2.30}$$

but

$$\lim_n \frac{1}{V_n}\mathcal{H}(\mathbb{S}_n|\mathbb{Q}_n) = 0 \tag{2.31}$$

by hypothesis, so that

$$\lim_n \mathbb{S}_n[X\backslash G] = 0 \tag{2.32}$$

and

$$\lim_n \mathbb{S}_n[G] = 1. \tag{2.33}$$

□

3 Conditioning

Throughout this section, we use $\mathbb{M}_n$ to denote a positive measure (not necessarily normalised). We make the standing assumption that an RL–function μ exists for the pair $(\{\mathbb{M}_n\}, \{V_n\})$.
Given a measurable set C for which $\mathbb{M}_n[C]$ is strictly positive for all sufficiently large n, we can construct a sequence of probability measures $\mathbb{M}_n[\,\cdot\mid C\,]$ by conditioning on the set C. We are interested in finding a set on which the conditioned measures are eventually concentrated; if we could compute the RL–function μ_C for the conditioned measures using the RL–function μ, we could use the fact that a sequence of probability measures is eventually concentrated on the null–set of its RL–function. In some cases this computation can be carried out, yielding the result that the sequence of conditioned measures is eventually concentrated on the set

$$N_C := \{x \in \operatorname{cl} C : \ \mu(x) = \sup_{y\in\operatorname{cl} C}\mu(y)\}\,. \tag{3.1}$$

It turns out that we can prove this concentration property in a much wider setting than that in which we can compute μ_C; this motivates the following definition:

Definition 3.1 *Let μ be the RL-function of the pair $(\{\mathbb{M}_n\}, \{V_n\})$. A sequence $\{C_n\}$ of sets is* **LD–regular** *if*

a) $\{C_n\}$ is a decreasing sequence of measurable sets, and $0 < \mathbb{M}_n[C_n] < \infty$ for all sufficiently large n;

b) the closed set

$$C := \bigcap_n \mathrm{cl}\, C_n$$

is non-empty;

c) the limit $\lim_n m_n[C_n]$ exists, is finite and

$$\lim_n m_n[C_n] = \sup_{x \in C} \mu(x).$$

Lemma 3.1 *Let $\{C_n\}$ and $\{D_n\}$ be LD–regular sequences. Then $\{C_n \cup D_n\}$ is an LD–regular sequence.*

Proof: Since, for any sets A and B, we have

$$\mathrm{cl}\,(A \cup B) = \mathrm{cl}\, A \cup \mathrm{cl}\, B\ , \tag{3.2}$$

it follows that

$$\bigcap_n \mathrm{cl}\,(C_n \cup D_n) = \bigcap_n (\mathrm{cl}\, C_n \cup \mathrm{cl}\, D_n) = (\bigcap_n \mathrm{cl}\, C_n) \cup (\bigcap_n \mathrm{cl}\, D_n)\ . \tag{3.3}$$

We have

$$\liminf_n\ m_n[C_n \cup D_n]\ \geq\ \liminf_n\ m_n[C_n]\ \vee\ \liminf_n\ m_n[D_n]\ . \tag{3.4}$$

On the other hand, by the principle of the largest term (see (1.9)), we have

$$\limsup_n\ m_n[C_n \cup D_n]\ =\ \limsup_n\ m_n[C_n]\ \vee\ \limsup_n\ m_n[D_n]\ . \tag{3.5}$$

Hence we have

$$\lim_n\ m_n[C_n \cup D_n]\ =\ \sup_{x \in \mathrm{cl}\,(C \cup D)} \mu(x)\ . \tag{3.6}$$

□

For any sequence $\{C_n\}$ such that $0 < \mathbb{M}_n[C_n] < \infty$, we define the **conditioned measures** $\mathbb{M}_n^C$ by

$$\mathbb{M}_n^C[B] := \mathbb{M}_n[B|C_n]\ \ ,\ \ B \in \mathcal{B}. \tag{3.7}$$

We set

$$\overline{m}_C[B] := \limsup_{n \to \infty} \frac{1}{V_n} \ln \mathbb{M}_n^C[B] \tag{3.8}$$

and define the upper deviation function as before by

$$\overline{\mu}_C(x) := \inf_{G \ni x} \overline{m}_C[G]. \tag{3.9}$$

The main result on the concentration of probability of the sequence $\{\mathbb{M}_n^C\}$ is contained in the next theorem.

Theorem 3.1 *Let the pair $(\{\mathbb{M}_n\}, \{V_n\})$ obey an LDP with RL–function μ. Let $\{C_n\}$ be an LD–regular sequence with $C := \cap_n \mathrm{cl}\, C_n$. Then*

a) the upper deviation function $\overline{\mu}_C$ has compact level–sets and, for each measurable set B, we have

$$\overline{m}_C[B] \le \sup_{x\in \mathrm{cl}\, B} \overline{\mu}_C(x)\,; \tag{3.10}$$

b) the sequence $\{\mathbb{M}_n^C\}$ of conditioned measures is eventually concentrated on the non–empty compact subset

$$N_C := \{x \in C :\ \mu(x) = \sup_{y\in C} \mu(y)\}; \tag{3.11}$$

c) if the sets C_n are relatively compact, then it is sufficient to suppose the existence of an RL–function for the pair $(\{\mathbb{M}_n\}, \{V_n\})$ in order that a) and b) hold.

Remark: Essentially the same theorem holds when the sequence $\{C_n\}$ satisfies a) and b) of Definition 3.1 and c) is replaced by the weaker statement that

$$-\infty < \liminf_n m_n[C_n] = \limsup_n m_n[C_n] < \infty. \tag{3.12}$$

In that case, one must replace the set N_C by

$$\{x \in C :\ \mu(x) \ge \lim_n m_n[C_n]\}. \tag{3.13}$$

Proof: Let $\alpha := \lim_n m_n[C_n]$; by hypothesis, α is finite. We assume that $\overline{m}_C[B] > -\infty$, otherwise there is nothing to prove. Since $\{C_n\}$ is decreasing, for any index k, we have

$$\begin{aligned} \overline{m}_C[B] &= \limsup_n m_n[B\cap C_n] - \alpha \\ &\le \sup_{x\in \mathrm{cl}\,(B\cap C_k)} \mu(x) - \alpha. \end{aligned} \tag{3.14}$$

Since $\mathrm{cl}\,(B\cap C_k) \subset \mathrm{cl}\, B \cap \mathrm{cl}\, C_k$, we have

$$-\infty < \overline{m}_C[B] + \alpha \le \inf_k \sup_{x\in \mathrm{cl}\, B\cap \mathrm{cl}\, C_k} \mu(x); \tag{3.15}$$

hence the compact level–set

$$K := \{x : \mu(x) \ge \overline{m}_C[B] + \alpha\} \tag{3.16}$$

is non–empty and has a non–empty intersection with $\mathrm{cl}\, B$. Let G be any open neighbourhood of K; by definition of K, and because μ attains its maximum on every closed set, we have

$$\sup_{x\in X\backslash G} \mu(x) < \overline{m}_C[B] + \alpha. \tag{3.17}$$

The upper bound property for closed sets implies that

$$\overline{m}_C[X\backslash G] \le \overline{m}[X\backslash G] - \alpha \le \sup_{x\in X\backslash G} \mu(x) - \alpha < \overline{m}_C[B], \tag{3.18}$$

and

$$\begin{aligned}\overline{m}_C[B] &= \overline{m}_C[B\backslash G] \vee \overline{m}_C[B \cap G] \\ &\leq \overline{m}_C[X\backslash G] \vee \overline{m}_C[B \cap G] \\ &\leq \overline{m}_C[\mathrm{cl}\, B \cap G].\end{aligned} \tag{3.19}$$

Given $\varepsilon > 0$, there exists for each x an open set $G_x \ni x$ such that

$$\overline{m}_C[G_x] \leq \overline{\mu}_C(x) + \varepsilon\ ; \tag{3.20}$$

since $K \cap \mathrm{cl}\, B$ is compact, we can cover $K \cap \mathrm{cl}\, B$ by a finite number of these open sets, say $G_{x_1}, \ldots, G_{x_k}$. Let U be an open neighbourhood of K; then

$$G := [U\backslash\mathrm{cl}\, B] \cup [U \cap (G_{x_1} \cup \cdots \cup G_{x_k})] \tag{3.21}$$

is also an open neighbourhood of K and, by the principle of the largest term,

$$\overline{m}_C[B] \leq \overline{m}_C[\mathrm{cl}\, B \cap G] \leq \overline{m}_C[\cup_i G_{x_i}] \leq \sup_i \overline{\mu}_C(x_i) + \varepsilon \leq \sup_{x\in\mathrm{cl}\, B} \overline{\mu}_C(x) + \varepsilon. \tag{3.22}$$

For any k, if $x \notin \mathrm{cl}\, C_k$, then there exists a measurable neighbourhood $G \ni x$ with

$$\overline{m}_C[G] = -\infty. \tag{3.23}$$

On the other hand, for any measurable neighbourhood $G \ni x$ we have

$$\overline{m}_C[G] \leq \overline{m}[G] - \alpha. \tag{3.24}$$

Consequently, for any k

$$\overline{\mu}_C(x) \leq \begin{cases} \mu(x) - \alpha & \text{if } x \in \mathrm{cl}\, C_k, \\ -\infty & \text{otherwise;} \end{cases} \tag{3.25}$$

hence

$$\overline{\mu}_C(x) \leq s(x) := \begin{cases} \mu(x) - \alpha & \text{if } x \in C, \\ -\infty & \text{otherwise.} \end{cases} \tag{3.26}$$

Since s has compact level–sets, the same is true for $\overline{\mu}_C$. Using Theorem 2.2 with the function s, we conclude that the conditioned measures are eventually concentrated on the non–empty compact set

$$N_C := \{x \in C : \mu(x) = \sup_{y\in C} \mu(y)\}. \tag{3.27}$$

To prove the last statement c), we notice that we used the upper bound property only in (3.14) and (3.18). Lemma 1.2 covers (3.14). Since C_1 contains all C_n, we have

$$\overline{m}_C[X\backslash G] = \overline{m}_C[C_1\backslash G]. \tag{3.28}$$

Therefore, we only need the upper bound for the relatively compact set $C_1\backslash G$, and

$$\overline{m}_C[X\backslash G] \leq \overline{m}[C_1\backslash G] - \alpha \leq \sup_{x\in\mathrm{cl}\, C_1\backslash G} \mu(x) - \alpha < \overline{m}_C[B], \tag{3.29}$$

since G is a neighbourhood of

$$\{x : \mu(x) \geq \overline{m}_C[B] + \alpha\}. \tag{3.30}$$

□

Example: a) We consider several variants of the following example. Let $X := [-1,+1]$, $V_n := n$, and

$$\mathbb{M}_n := \mathrm{e}^{-an}\delta_0 + \mathrm{e}^{-an}\delta_{1-\frac{1}{n}} + \mathrm{e}^{-an^2}\delta_{-1+\frac{1}{n}} + (1 - 2\mathrm{e}^{-an} - \mathrm{e}^{-an^2})\delta_{-\frac{1}{2}}, \tag{3.31}$$

with a some given positive real number. The RL–function of the pair $(\{\mathbb{M}_n\},\{V_n\})$ is

$$\mu(x) := \begin{cases} 0 & \text{if } x = -\frac{1}{2}\ , \\ -a & \text{if } x = 0 \text{ or } x = 1\ , \\ -\infty & \text{otherwise.} \end{cases} \tag{3.32}$$

Let $\{C_n\}$ be the sequence of sets defined by

$$C_n := \{0\} \cup (1 - \frac{2}{n}, 1). \tag{3.33}$$

We have $C = \cap_n \mathrm{cl}\, C_n = \{0\} \cup \{1\}$ and $\lim_n m_n[C_n] = -a$; hence the sequence $\{C_n\}$ is LD–regular. Since X is compact, we can apply Theorem 3.1; the sequence $\{\mathbb{M}_n^C\}$ of conditioned measures is eventually concentrated on C. Moreover, this sequence converges (in the narrow topology) to $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. We have similar conclusions if we replace the sequence $\{C_n\}$ by the sequence $\{C'_n\}$, with

$$C'_n := (1 - \frac{2}{n}, 1). \tag{3.34}$$

Here $C' = \{1\}$, and the sequence $\{\mathbb{M}_n^{C'}\}$ converges to δ_1. In this case, we have $\mathbb{M}_n[\{1\}|C'_n] = 0$ and $\cap_n C'_n = \emptyset$.

b) We consider the same example in the space $X' := [-1,+1)$. Now the RL–function is

$$\mu'(x) := \begin{cases} 0 & \text{if } x = -\frac{1}{2}\ , \\ -a & \text{if } x = 0\ , \\ -\infty & \text{otherwise.} \end{cases} \tag{3.35}$$

The sequence $\{C_n\}$ is still LD–regular, but now $C = \{0\}$; the conditioned measures $\mathbb{M}_n^C$ are the same as before. We cannot apply Theorem 3.1 because we do not have an LDP,

$$-a = \overline{m}[[\frac{1}{2}, 1)] \not\leq \sup_{x \in [\frac{1}{2},1)} \mu'(x). \tag{3.36}$$

If we consider the sequence $\{C'_n\}$, we have $\lim_n m_n[C'_n] = -a$ and

$$C' = \bigcap_n \mathrm{cl} C'_n = \bigcap_n [1 - \frac{2}{n}, 1) = \emptyset. \tag{3.37}$$

Hence the sequence $\{C'_n\}$ is not LD–regular.

Part II: Tilted Measures and Convexity

We introduce the tilted measures and we compare them with the conditioned measures, using the notion of an asymptotically I-null sequence. To develop the theory we require that the space X be a closed convex subset of a locally convex topological real vector space.

4 Tilting

4.1 Convex Structure

We fix the setting for the next three sections. Let (E, τ) be a locally convex Hausdorff topological vector space E over $\mathbb{R}$. The topological dual of (E, τ) is denoted by E^*. We choose a topology τ^* on E^* so that the pair (E, E^*) is in duality: the topological dual of (E^*, τ^*) is E. Elements of E are denoted by x and those of E^* by x'; the pairing between E and E^* is denoted by $(x', x) \mapsto \langle x', x \rangle$. We require that the space X be a closed convex subset of E equipped with the induced topology. As before, $(X, \mathcal{B})$ is also a measurable space, and each point x in X has a local base of measurable neighbourhoods; moreover, we require that the maps $x \mapsto \langle x', x \rangle$ be $\mathcal{B}$-measurable for every $x' \in E^*$.

Some important examples of the above setting are of the following kind: E and E' are real topological spaces and $\langle \cdot, \cdot \rangle$ is a bilinear map $E' \times E \to \mathbb{R}$ so that

a) for each $x \neq 0$ of E there exists $x' \in E'$ with $\langle x', x \rangle \neq 0$;

b) for each $x' \neq 0$ of E' there exists $x \in E$ with $\langle x', x \rangle \neq 0$.

When conditions *a*) and *b*) are satisfied, we say that (E, E') is a *dual pair*. For the topology τ, we choose the $\sigma(E, E')$-topology which is generated by the base of (closed) neighbourhoods

$$\{x : \sup_{1 \leq i \leq n} |\langle x'_i, x \rangle| \leq 1\} \quad (x'_i \in E'); \tag{4.1}$$

the topology $\sigma(E, E')$ is locally convex and Hausdorff since condition *a*) holds. The topological dual E^* of (E, τ) is the set of all continuous linear forms on E. By definition of the topology $\sigma(E, E')$, E^* contains the set E'. Since *a*) and *b*) hold, the topological dual of (E, τ) is $E^* \equiv E'$; if we choose for the topology τ^* on E^* the $\sigma(E^*, E)$-topology, then the topological dual of (E^*, τ^*) is E. (See, for example, [RR] or [B].) We require that the maps $x \mapsto \langle x', x \rangle$ be $\mathcal{B}$-measurable for every x' in E^*, and that X be a closed convex subset of E. This implies that each point x in X has a local base of open (respectively closed) convex measurable neighbourhoods. Typical examples are:

(I) $E = \mathbb{R}^d$, with $\mathbb{R}^d$ equipped with the Euclidean topology; in this case, we have $E^* = \mathbb{R}^d$, the bilinear form is the Euclidean scalar product, the set X is a closed convex subset of $\mathbb{R}^d$ and $\mathcal{B}$ is the Borel σ-algebra of X.

(II) Let $(\Omega, \mathcal{F})$ be a Polish space and $E = \mathcal{M}(\Omega, \mathcal{F})$, the space of finite signed measures on $(\Omega, \mathcal{F})$. Let $E' = C_b(\Omega)$ be the space of bounded continuous functions on Ω. The bilinear form is

$$\langle x', x \rangle = \int_\Omega x'(\omega) x[d\omega] \quad , \quad x' \in C_b(\Omega)\,, x \in \mathcal{M}(\Omega). \tag{4.2}$$

The $\sigma(E, C_b(\Omega))$–topology coincides with the topology of narrow convergence, the set $X = \mathcal{M}_1^+(\Omega, \mathcal{F})$ is the space of probability measures on the Polish space $(\Omega, \mathcal{F})$ and $\mathcal{B}$ is the σ–algebra generated by the maps $x \mapsto \langle x', x \rangle$, $x' \in C_b(\Omega)$. In this case, the set X is a Polish space and $\mathcal{B}$ is the Borel σ–algebra of X.

(III) Let $(\Omega, \mathcal{F})$ be a measurable space and $E = \mathcal{M}(\Omega, \mathcal{F})$, the space of finite signed measures on $(\Omega, \mathcal{F})$. Let E' be the space of all bounded $\mathcal{F}$–measurable functions on Ω. The bilinear form is

$$\langle x', x \rangle = \int_\Omega x'(\omega) x[d\omega] \quad , \quad x' \in C_b(\Omega)\,, x \in \mathcal{M}(\Omega). \tag{4.3}$$

We choose the $\sigma(E, E')$–topology: a sequence $\{x_n\}$ of measures converges to a measure x in this topology if and only if

$$\lim_n \int_B x_n[d\omega] = \int_B x[d\omega] \text{ for all } B \in \mathcal{B}(\Omega). \tag{4.4}$$

The set $X = \mathcal{M}_1^+(\Omega, \mathcal{F})$ is the space of probability measures on $(\Omega, \mathcal{F})$ and $\mathcal{B}$ is the σ–algebra generated by the maps $x \mapsto \langle x', x \rangle$, $x' \in E'$.

Finally, we recall two definitions of convex analysis. Let g be any function $g : E \to \overline{\mathbb{R}}$; the **conjugate function** g^* **of** g is defined on E^* by

$$g^*(x') := \sup_{x \in E} \{\langle x', x \rangle - g(x')\}. \tag{4.5}$$

Similarly, the conjugate function $f^* : E \to \overline{\mathbb{R}}$ of $f : E^* \to \overline{\mathbb{R}}$ is defined by

$$f^*(x) := \sup_{x' \in E^*} \{\langle x', x \rangle - f(x')\}. \tag{4.6}$$

The functions g^* and f^* are always lower semicontinuous and convex. Let f be convex on E^*; the **subdifferential** ∂f **of** f **at** x' is the subset of E given by

$$\partial f(x') := \{x \in E : f(x' + y') \geq p(x') + \langle y', x \rangle, \forall y' \in E^*\}. \tag{4.7}$$

4.2 Tilted Probability Measures

Let $\{\mathbb{M}_n\}$ be a sequence of measures on $(X, \mathcal{B})$ and $\{V_n\}$ a scale. We suppose always that

$$p(x') := \lim_n \frac{1}{V_n} \ln \int_X e^{V_n \langle x', x \rangle} \mathbb{M}_n[dx] \tag{4.8}$$

exists (but is not necessarily finite). The function $p : E^* \to \overline{\mathbb{R}}$, called the **scaled generating function**, is necessarily convex; the **essential domain** of p, $\operatorname{dom} p$, is defined by

$$\operatorname{dom} p := \{x' \in E^* : p(x') \in \mathbb{R}\}. \tag{4.9}$$

For all $x' \in \operatorname{dom} p$, we have

$$0 < \int_X e^{V_n\langle x',x\rangle} \mathbb{M}_n[dx] < \infty \tag{4.10}$$

for n sufficiently large; for those x', we define the **tilted (probability) measure** $\mathbb{M}_n^{x'}$ by

$$\mathbb{M}_n^{x'}[B] := \frac{\int_B e^{V_n\langle x',x\rangle} \mathbb{M}_n[dx]}{\int_X e^{V_n\langle x',x\rangle} \mathbb{M}_n[dx]}. \tag{4.11}$$

Theorem 4.1 *Let E, E^* and X be as above. Let x' be an interior point of* $\operatorname{dom} p$, *and suppose that the pair $(\{\mathbb{M}_n\}, \{V_n\})$ obeys an LDP with RL-function μ which is not identically $-\infty$ and bounded above on X. Then*

a) An RL-function $\mu^{x'}$, given by

$$\mu^{x'}(x) = \mu(x) + \langle x', x\rangle - p(x') \leq 0; \tag{4.12}$$

exists for the pair $(\{\mathbb{M}_n^{x'}\}, \{V_n\})$, and the sequence $\{\mathbb{M}_n^{x'}\}$ is eventually concentrated on the non-empty compact set

$$N^{x'} := \{x \in X : \mu^{x'}(x) = 0\}. \tag{4.13}$$

b) If $\operatorname{dom} p = E^*$ *and $\mu(x) = -p^*(x)$, then the concentration set $N^{x'}$ coincides with the subdifferential ∂p of p at x'.*

Proof: For t small and positive and a non-negative, we have

$$\frac{1}{V_n} \ln \int_{\{\langle x',x\rangle \geq a\}} e^{V_n\langle x',x\rangle} \mathbb{M}_n[dx] \leq -ta + \frac{1}{V_n} \ln \int_X e^{V_n(1+t)\langle x',x\rangle} \mathbb{M}_n[dx]\ ; \tag{4.14}$$

thus

$$\lim_{a\to\infty} (\limsup_n \frac{1}{V_n} \ln \int_{\{\langle x',x\rangle \geq a\}} e^{V_n\langle x',x\rangle} \mathbb{M}_n[dx]) = -\infty. \tag{4.15}$$

Using Theorem 6.3 of [LP] we conclude that the pair $(\{\mathbb{M}_n^{x'}\}, \{V_n\})$ obeys an LDP with RL-function

$$\mu^{x'}(x) = \mu(x) + \langle x', x\rangle - p(x') \leq 0. \tag{4.16}$$

Theorem 2.2 implies that $\{\mathbb{M}_n^{x'}\}$ is eventually concentrated on the non-empty compact set

$$N^{x'} = \{x \in X : p(x') = \mu(x) + \langle x', x\rangle\}. \tag{4.17}$$

This proves a). The second part of the theorem is an elementary exercise in convex analysis. For any y' and any $x \in N^{x'}$, we have

$$\begin{aligned} p(y') &\geq \mu(x) + \langle y', x\rangle \\ &= p(x') - \langle x', x\rangle + \langle y', x\rangle \\ &= p(x') + \langle y' - x', x\rangle; \end{aligned} \tag{4.18}$$

hence $x \in \partial p(x')$. Let $-\mu(x) = p^*(x)$ and $x \in \partial p(x')$; we have

$$p(y') \geq p(x') + \langle y' - x', x\rangle = p(x') - \langle x', x\rangle + \langle y', x\rangle; \tag{4.19}$$

thus

$$\begin{aligned} \langle x', x\rangle - p(x') &\geq \sup_{y'}(\langle y', x\rangle - p(y')) \\ &= p^*(x) \\ &= -\mu(x). \end{aligned} \tag{4.20}$$

Since $\mu^{x'}$ is non-positive, this implies that $x \in N^{x'}$. □

5 Asymptotically I–Null Sequences

We compare the sequence of conditioned probability measures $\{\mathbb{M}_n^C\}$ with the sequence of tilted measures $\{\mathbb{M}_n^{x'}\}$. Lemma 5.1 gives upper and lower bounds on the specific information gain

$$\lim_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}). \tag{5.1}$$

Lemma 5.1 *Let E, E^* and X be as above, and let μ be the RL–function of the pair $(\{\mathbb{M}_n\},\{V_n\})$. Let C_n be a sequence of measurable sets such that $\alpha := \lim_n m_n[C_n]$ is finite. Assume that the sequence $\{\mathbb{M}_n^C\}$ is eventually concentrated on the non-empty subset N and that any open neighbourhood of N contains a measurable neighbourhood of N. For x' in* $\operatorname{dom} p$, *we have:*

a) If the continuous linear functional $x \mapsto \langle x', x\rangle$ is uniformly bounded below on C_n for n sufficiently large, then

$$\limsup_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) \le \sup_{x\in N}\{p(x') - \langle x', x\rangle - \alpha\}. \tag{5.2}$$

b) Let $\widehat{\mu}$ be the concave envelope of the RL–function μ. If the continuous linear functional $x \mapsto \langle x', x\rangle$ is uniformly bounded above on C_n for n sufficiently large and $N \subset \{x : \mu(x) \ge \alpha\}$, then

$$\begin{aligned}\liminf_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) &\ge [p(x') - (-\mu)^*(x')] \\ &+ \inf_{x\in N}\{\widehat{\mu}(x) - \mu(x)\} \ge 0,\end{aligned} \tag{5.3}$$

both terms on the right-hand side being nonnegative. If, in addition, $\operatorname{dom} p = E^*$, *then*

$$\liminf_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) \ge \inf_{x\in N}\{-\mu(x) - p^*(x)\} \ge 0. \tag{5.4}$$

Proof: a) Let us introduce the notation

$$p_n(x') := \frac{1}{V_n}\ln\int_X e^{V_n\langle x',x\rangle}\mathbb{M}_n[dx]. \tag{5.5}$$

We have
$$\frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) = -\int_{C_n}\langle x', x\rangle\mathbb{M}_n^C[dx] + p_n(x') - m_n[C_n] \tag{5.6}$$

so that

$$\begin{aligned}\limsup_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) \le &- \liminf_n \int_{C_n}\langle x', x\rangle\mathbb{M}_n^C[dx] \\ &+ p(x') - \alpha.\end{aligned} \tag{5.7}$$

Using Theorem 2.1, we have

$$\liminf_n \int_{C_n}\langle x', x\rangle\mathbb{M}_n^C[dx] \ge \inf_{x\in N}\langle x', x\rangle; \tag{5.8}$$

thus

$$\limsup_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) \leq -\inf_{x\in N}\langle x',x\rangle + p(x') - \alpha$$
$$\leq \sup_{x\in N}\{p(x') - \langle x',x\rangle - \alpha\}. \tag{5.9}$$

b) We have

$$\frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) = -\int_{C_n}\langle x',x\rangle \mathbb{M}_n^C[dx] + p_n(x') - m_n[C_n] \tag{5.10}$$

so that

$$\liminf_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) \geq -\limsup_n \int_{C_n}\langle x',x\rangle \mathbb{M}_n^C[dx] + p(x') - \alpha. \tag{5.11}$$

Using Theorem 2.1, we have

$$\limsup_n \int_{C_n}\langle x',x\rangle \mathbb{M}_n^C[dx] \leq \sup_{x\in N}\langle x',x\rangle. \tag{5.12}$$

Since

$$0 \geq \mu^{x'}(x) = \langle x',x\rangle + \mu(x) - p(x') \;,\; \forall x \in E\,, \tag{5.13}$$

we get

$$(-\mu)^*(x') \leq p(x') \tag{5.14}$$

by taking the supremum over x. Thus we have

$$\liminf_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) \geq -\sup_{x\in N}\langle x',x\rangle + p(x') - \alpha \tag{5.15}$$
$$= [p(x') - (-\mu)^*(x')]$$
$$+ [-\sup_{x\in N}\langle x',x\rangle + (-\mu)^*(x') - \alpha].$$

But, since $\mu(x) \geq \alpha$ on N, we have

$$-\sup_{x\in N}\langle x',x\rangle + (-\mu)^*(x') - \alpha \geq \inf_{x'}\{-\sup_{x\in N}\langle x',x\rangle + (-\mu)^*(x')\} - \alpha \tag{5.16}$$
$$= \inf_{x\in N}\inf_{x'}\{-\langle x',x\rangle + (-\mu)^*(x')\} - \alpha$$
$$= \inf_{x\in N}\{-(-\mu)^{**}(x)\} - \alpha$$
$$\geq \inf_{x\in N}\{\widehat{\mu}(x) - \mu(x)\}.$$

From (5.13), we have

$$p^*(x) \leq -\mu(x) \tag{5.17}$$

by taking the supremum over x'. From (5.15), using the hypothesis that $\mu(x) \geq \alpha$ on N , we have

$$\liminf_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) \geq -\sup_{x\in N}\langle x',x\rangle + p(x') - \alpha \tag{5.18}$$
$$\geq -\sup_{x\in N}\sup_{x'}\{\langle x',x\rangle - p(x') + \alpha\}$$
$$\geq \inf_{x\in N}\{-p^*(x) - \mu(x)\} \geq 0.$$

□

Lemma 5.1 gives estimates which enable us to compare a sequence of conditioned measures with a sequence of tilted measures.

Theorem 5.1 *Let the pair* $(\{\mathbb{M}_n\}, \{V_n\})$ *obey an LDP with RL-function* μ *as rate-function. Let* $\{C_n\}$ *be an LD-regular sequence with* $C := \cap_n \mathrm{cl}\, C_n$, *and* $\mathrm{dom}\, p = E^*$. *Then*

(i) If the map $x \mapsto \langle x', x\rangle$ *is bounded below on* C_n *for* n *sufficiently large and*

$$N_C = \{x \in C : \sup_{y\in C} \mu(y) = \mu(x)\} \subset N^{x'} = \{x \in X : p(x') = \mu(x) + \langle x', x\rangle\},$$

then the sequence $\{\mathbb{M}_n^C\}$ *of conditioned measures is asymptotically I-null to the sequence* $\{\mathbb{M}_n^{x'}\}$ *of tilted measures on the scale* $\{V_n\}$:

$$\lim_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C | \mathbb{M}_n^{x'}) = 0.$$

(ii) If the map $x \mapsto \langle x', x\rangle$ *is bounded above on* C_n *for* n *sufficiently large and the sequence* $\{\mathbb{M}_n^C\}$ *of conditioned measures is asymptotically I-null to the sequence* $\{\mathbb{M}_n^{x'}\}$ *of tilted measures on the scale* $\{V_n\}$, *then* p *is conjugate to* $-\mu$ *at* x':

$$p(x') = \sup_{x\in E}\{\langle x', x\rangle + \mu(x)\}.$$

If the sets C_n *are relatively compact, then it is not necessary to have an LDP; it is sufficient to assume the existence of an RL-function* μ.

Proof: By Theorem 3.1, we can apply Lemma 5.1 with $N = N_C$. (Since N_C is compact, every open neighbourhood of N_C contains a measurable neighbourhood of N_C.) Thus

$$\begin{aligned} 0 \le \limsup_n \frac{1}{V_n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) &\le \sup_{x\in N_C}\{p(x') - \langle x', x\rangle - \sup_{y\in C}\mu(y)\} \\ &\le \sup_{x\in N_C}\{p(x') - \langle x', x\rangle - \mu(x)\} \\ &= 0, \end{aligned} \tag{5.19}$$

since $N_C \subset \{x \in X : p(x') = \mu(x) + \langle x', x\rangle\}$. If the C_n are relatively compact, then Theorem 3.1 is still valid if we assume only the existence of the RL-function. The second statement of the theorem follows from Lemma 5.1 b),
since $N_C \subset \{x : \mu(x) \ge \lim_n m_n[C_n]\}$. □

6 Convexity

We investigate the consequences of the assumption that C is a convex set and μ a concave function. The main result is that, under this assumption, we can fulfil the hypotheses of Theorem 5.1. References on convex functions are [B], [R] and [ET].

6.1 Convexity and LD–Regularity

Let S be a subset of a real vector space E; we say that a point $y \in E$ is **linearly accessible in S from x** if there exists $x \in S$, $x \neq y$, such that

$$]x,y[\, := \{\alpha x + (1-\alpha)y : 0 < \alpha < 1\} \subset S. \tag{6.1}$$

Lemma 6.1 *Let the pair $(\{\mathbb{M}_n\}, \{V_n\})$ obey an LDP with RL–function μ.*

a) Let B be a subset of E with non-empty interior. If μ is concave on $\operatorname{cl} B$ and each point y in $\operatorname{cl} B$ is linearly accessible in $\operatorname{int} B$ from some x with $\mu(x)$ finite, then B is LD–regular.

b) Let B be a convex subset of E with non-empty interior. If μ is concave on $\operatorname{cl} B$ and finite at some point of $\operatorname{int} B$, then B is LD–regular.

If the set B is relatively compact, then in order that statements a) and b) hold it is not necessary to have an LDP; it is sufficient to assume the existence of an RL–function μ.

Proof: Since we have an LDP, we have

$$\sup_{x \in \operatorname{int B}} \mu(x) \le \underline{m}[B] \le \overline{m}[\operatorname{cl} B] \le \sup_{x \in \operatorname{cl} B} \mu(x). \tag{6.2}$$

We prove that

$$\sup_{x \in \operatorname{int B}} \mu(x) = \sup_{x \in \operatorname{cl} B} \mu(x). \tag{6.3}$$

Let $y \in \operatorname{cl} B$. By assumption, there exists $x \in \operatorname{int} B$ such that $]x,y[\subset \operatorname{int} B$ and $\mu(x)$ is finite. Let $x_\alpha := \alpha x + (1-\alpha)y$. Since μ is concave on $\operatorname{cl} B$, for all $\alpha > 0$, we have

$$\mu(x_\alpha) \ge \alpha\mu(x) + (1-\alpha)\mu(y), \tag{6.4}$$

which implies

$$\liminf_{\alpha \to 0} \mu(x_\alpha) \ge \mu(y) . \tag{6.5}$$

Since μ is u.s.c., we have

$$\limsup_{\alpha \to 0} \mu(x_\alpha) \le \mu(y), \tag{6.6}$$

which implies

$$\lim_{\alpha \to 0} \mu(x_\alpha) = \mu(y). \tag{6.7}$$

Since, for all α different from zero, the point x_α is in $\operatorname{int} B$, equation (6.3) holds. It remains to show that $\sup_{x \in \operatorname{cl} B} \mu(x)$ is finite. This supremum is not $-\infty$, since μ is finite at some point of B. Since μ is u.s.c., concave and finite at some point of B, a standard result of convexity theory implies that $\mu(x) < \infty$ for all x (see, for example, [ET]). Since μ has compact level–sets, the supremum of μ on $\operatorname{cl} B$ is attained and is therefore finite. Statement b) is a consequence of convexity theory: if B is convex and has a non-empty interior, then all points of $\operatorname{cl} B$ are linearly accessible in $\operatorname{int} B$ from any given point $x \in \operatorname{int} B$. Hence the result follows from a). If B is relatively compact, then (6.2) still holds; therefore a) is still true. □

6.2 Convexity and $N_C \subset N^{x'}$

We denote by $\widehat{\mu}$ the concave envelope of the RL–function μ; it is convenient to extend $\widehat{\mu}$ to all of E by putting $\mu(x) = -\infty$ for $x \in E \backslash X$. Let B be a convex subset of E and set $C := \mathrm{cl}\, B$; let μ coincide with its concave envelope on C. Maximizing μ on C is equivalent to minimizing $(-\widehat{\mu}) + h_C$ on E, where h_C is the indicator function of C:

$$h_C(x) := \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{if } x \notin C. \end{cases} \tag{6.8}$$

Lemma 6.2 *Let B be a convex subset, $C := \mathrm{cl}\, B$, and assume that*

$$N_C = \{x \in C : \sup_{y \in C} \mu(y) = \mu(x)\} \tag{6.9}$$

is non-empty and $\sup_{y \in C} \mu(y)$ is finite. Assume further that the RL–function μ coincides with its concave envelope $\widehat{\mu}$ on C. If either C has an interior point or μ is continuous at some point of C, then

a) there exists x' in E^* which is bounded below on C;

b) $x' \in \partial(-\mu)(x)$ for all $x \in N_C$;

c) $-x' \in \partial(h_C)(x)$ for all $x \in N_C$;

d) N_C is a subset of the subdifferential

$$\partial(-\mu)^*(x') = \{z \in E : \langle x', z \rangle = -\mu(z) + (-\mu)^*(x')\}. \tag{6.10}$$

Remark: The subgradients of $h_C(x)$ have a simple geometrical interpretation: if $0 \neq x' \in \partial h_C(x)$, then (see (6.24))

$$0 \geq \langle x', z - x \rangle \ , \ \ \forall z \in C; \tag{6.11}$$

thus x' is the exterior normal to the closed half–plane $\{z \in E : \langle x', z \rangle \leq \langle x', x \rangle\}$ containing C. Conversely, if $x \in C$ and the closed half–plane $\{z \in E : \langle x', z \rangle \leq \langle x', x \rangle\}$ contains C, then $x' \in \partial h_C(x)$.

Proof : We set

$$f_1(x) := \begin{cases} -\widehat{\mu}(x) & x \in X, \\ \infty & x \in E \backslash X, \end{cases} \tag{6.12}$$

and

$$f_2(x) := h_C(x). \tag{6.13}$$

The convex function f_1 is a closed function (that is, it is lower semicontinuous), it never takes the value $-\infty$ and $\mathrm{dom} f_1 = \{x \in E : f_1(x) < \infty\}$ is non-empty. Since C is closed, f_2 is a closed convex function, as is $f := f_1 + f_2$. Let x belong to N_C; by hypothesis, x is a minimum of f on E so that the definition of subgradient implies that $0 \in \partial f(x)$. The heart of the proof is to show that

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x); \tag{6.14}$$

if this holds, then there exists $x' \in E^*$ such that $x' \in \partial f_1(x)$ and $-x' \in \partial f_2(x)$. We follow [ET] for the proof of (6.14), a consequence of the Hahn–Banach Theorem.

Since either C has an interior point or μ is continuous at some point of C, it follows that there exists $z \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$ where f_1 or f_2 is continuous, say f_1. Let y' be any subgradient in $\partial(f_1 + f_2)(x)$; this means that $f_1(x) < \infty$, $f_2(x) < \infty$ and, for all $y \in E$, we have

$$f_1(y) + f_2(y) \geq f_1(x) + f_2(x) + \langle y', y - x \rangle. \tag{6.15}$$

Let

$$y \mapsto g(y) := f_1(y) - f_1(x) - \langle y', y - x \rangle; \tag{6.16}$$

the function g is a closed convex function on E which is continuous at z. Let us consider the convex sets $\mathcal{C}_1$ and $\mathcal{C}_2$ in $E \times \mathbb{R}$:

$$\mathcal{C}_1 := \{(y, a) : \; g(y) \leq a\}, \tag{6.17}$$

and

$$\mathcal{C}_2 := \{(y, a) : \; a \leq f_2(x) - f_2(y)\}. \tag{6.18}$$

Relation (6.15) implies that $\mathcal{C}_1$ and $\mathcal{C}_2$ have only boundary–points in common; since $\mathcal{C}_1$ is the epigraph of g and g is continuous at z, the set $\mathcal{C}_1$ has a non–empty interior $\operatorname{int} \mathcal{C}_1$. We can separate $\mathcal{C}_2$ and $\operatorname{int} \mathcal{C}_1$ by a closed hyperplane. The hyperplane cannot be vertical; indeed, if the hyperplane were vertical, then we could separate $\operatorname{dom} f_1$ and $\operatorname{dom} f_2$; this is impossible since there exists $z \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$ which is a continuity point of f_1. Consequently, the separating hyperplane is of the form

$$y \mapsto \langle y'', y \rangle + \alpha, \quad y'' \in E^* \,, \; \alpha \in \mathbb{R}, \tag{6.19}$$

and for all y we have

$$f_2(x) - f_2(y) \leq \langle y'', y \rangle + \alpha \leq f_1(y) - f_1(x) - \langle y', y - x \rangle. \tag{6.20}$$

Putting $y = x$, we get $\alpha = -\langle y'', x \rangle$, and hence

$$f_2(y) \geq f_2(x) + \langle -y'', y - x \rangle \;\; , \forall y, \tag{6.21}$$

and

$$f_1(y) \geq f_1(x) + \langle y' + y'', y - x \rangle \;\; , \forall y. \tag{6.22}$$

Therefore we can decompose y' into $y' = (y' + y'') + (-y'')$ with $y' + y'' \in \partial f_1$ and $-y'' \in \partial f_2$. In our case, we have $y' = 0$ and we set $x' := y''$. Hence there exists $x' \in E^*$ such that $-x' \in \partial f_2(x)$; that is,

$$h_C(u) \geq \langle \, -x' \, , u - x \rangle \;\; , \;\; \forall u \in E, \tag{6.23}$$

which is equivalent to

$$\langle x', u \rangle \geq \langle x', x \rangle \;\; , \;\; \forall u \in C. \tag{6.24}$$

The functional x' is therefore bounded below on C and a) is proved.

The rest of the proof is elementary. We show that $x' \in \partial f_1(y)$ for any $y \in N_C$, which implies that

$$f_1^*(x') + f_1(y) = \langle x', y \rangle \;\; , \;\; \forall y \in N_C. \tag{6.25}$$

Indeed, let x and y be distinct elements of N_C; since $f_1(x) = f_1(y)$ and $\langle x', y-x\rangle \geq 0$ by (6.24), for all z, we have

$$\begin{aligned} f_1(z) &\geq f_1(x) + \langle x', z - x\rangle \\ &= f_1(y) + \langle x', z-y\rangle + \langle x', y-x\rangle \\ &\geq f_1(y) + \langle x', z-y\rangle. \end{aligned} \tag{6.26}$$

Therefore (6.25) holds, and N_C is a subset of

$$\{z \in C : \langle x', z\rangle = (-\widehat{\mu})(z) + (-\widehat{\mu})^*(x')\}. \tag{6.27}$$

On N_C, we have $\widehat{\mu}(z) = \mu(z)$; we show that $(-\widehat{\mu})^*(x') = (-\mu)^*(x')$. For any $z \in E$, we have $\widehat{\mu}(z) \geq \mu(z)$; using (6.25), we have

$$\begin{aligned} (-\mu)^*(x') &\leq (-\widehat{\mu})^*(x') \\ &= \langle x', x\rangle + \mu(x) \\ &\leq (-\mu)^*(x'). \end{aligned} \tag{6.28}$$

Hence d) holds. We have

$$(-\mu)^*(x') + \mu(y) = \langle x', y\rangle \quad , \quad \forall y \in N_C, \tag{6.29}$$

so that $x' \in \partial(-\mu)(y)$ for all $y \in N_C$: hence b) and c) hold. □

Finally, we recall the following useful result which relates $-\mu$ to the conjugate function of p (see [DS] or [LP]).

Lemma 6.3 *Let the pair* $(\{\mathbb{M}_n\}, \{V_n\})$ *obey an LDP with RL-function* μ. *If* $\operatorname{dom} p = E^*$, *then*

$$p(x') = (-\mu)^*(x'). \tag{6.30}$$

If, in addition, μ *is concave, then* $-\mu$ *and* p *are conjugate functions:*

$$p(x') = (-\mu)^*(x') \text{ and } -\mu(x) = p^*(x). \tag{6.31}$$

6.3 Convexity and Asymptotically I–null Sequences

We summarize the results obtained so far. For convenience, we recall the setting. $(X, \mathcal{B})$ is a measurable space. There exists a locally convex Hausdorff topological vector space (E, τ) over $\mathbb{R}$ with topological dual (E^*, τ^*), so that the pair (E, E^*) is in duality; X is a closed convex subset of E, equipped with the induced topology. The measurable and topological structures on X are compatible in the following sense: each point $x \in X$ has a local base of measurable neighbourhoods; the maps $x \mapsto \langle x', x\rangle$ are $\mathcal{B}$-measurable for every $x' \in E^*$, where $\langle x', x\rangle$ denotes the pairing between E and E^*.

Theorem 6.1 *The setting is as above. Assume that the pair* $(\{\mathbb{M}_n\}, \{V_n\})$ *obeys an LDP with RL-function* μ. *Let* $\{C_n\}$ *be an LD-regular sequence with* $C = \cap \operatorname{cl} C_n$. *Assume that* C *is convex and that the concave envelope of* μ *coincides with* μ *on* C; *assume further that either* $\operatorname{int} C$ *is non-empty or* μ *is continuous at some point of* C. *Then*

a) The sequence of conditioned measures $\{\mathbb{M}_n^C\}$ *is concentrated on the non-empty compact set*

$$N_C = \{x \in C : \mu(x) = \sup_{y \in C} \mu(y)\}; \tag{6.32}$$

there exists $x' \in \cap_{x \in N_C} \partial(-\mu)(x)$ *bounded below on* C *such that*

$$N_C \subset \partial(-\mu)^*(x'). \tag{6.33}$$

b) If, in addition, x' *is bounded below on* C_n *for* n *sufficiently large and* $p(x') = (-\mu)^*(x')$, *then* N_C *is a subset of*

$$N^{x'} = \{x \in E : \mu(x) + \langle x', x \rangle - p(x') = 0\}$$

and the sequence of conditioned measures $\{\mathbb{M}_n^C\}$ *is asymptotically I-null to the sequence of normalized tilted measures* $\{\mathbb{M}_n^{x'}\}$ *on the scale* $\{V_n\}$.

If the sets C_n *are relatively compact, then it is not necessary to have an LDP; it is sufficient to assume the existence of an RL-function* μ.

Proof: By Theorem 3.1, the conditioned measures are eventually concentrated on the non-empty compact subset N_C. The second part of a) follows from Lemma 6.2. In particular, (6.33) reads

$$N_C \subset \{x \in E : \langle x', x \rangle = -\mu(x) + (-\mu)^*(x')\}. \tag{6.34}$$

Therefore $p(x') = (-\mu)^*(x')$ implies $N_C \subset N^{x'}$. If $x \in N_C$, then $(-\mu)^*(x')$ is finite; hence $p(x')$ is finite. Therefore b) follows from Theorem 5.1. □

Part III: Conditional Limit Theorems

7 Independent Random Variables

We show how the formalism developed in parts I and II can be applied to prove conditional limit theorems. In 7.1, we give a general result, a direct consequence of Theorem 5.1. In 7.2, we give a concrete application for discrete random variables. In the final subsection, 7.3, we study a counter–example where the hypotheses of Theorem 6.1 are not verified.

7.1 General Case

Let $(S, \mathcal{S})$ be a standard Borel space and β a probability measure on $(S, \mathcal{S})$. For each $i \in \mathbb{N}$, let $(\Omega_i, \mathcal{F}_i)$ be a copy of the space $(S, \mathcal{S})$ and define $(\Omega, \mathcal{F})$ to be the product space. Let Y_i, $i = 1, \ldots$, be independent random variables,

$$Y_i : \Omega \to S \quad , \quad Y_i(\omega) := \omega_i, \tag{7.1}$$

with common law β. We consider another random variable $\varphi : S \to X$ with values in a measurable space $(X, \mathcal{B})$ which satisfies the hypotheses of part II. For each $n \in \mathbb{N}$, we take $V_n = n$ and define $T_n : \Omega \to X$ by

$$T_n(\omega) := \frac{1}{n} \sum_{j=1}^{n} \varphi(\omega_j). \tag{7.2}$$

The distribution of T_n on X is $\mathbb{M}_n$. A typical example is

$$\varphi : S \to \mathcal{M}_1^+(S) \quad , \quad \varphi(s) := \delta_s, \tag{7.3}$$

where δ_s is the Dirac mass at $s \in S$. Here T_n is the empirical distribution; this case has been extensively studied, see for example [BZ], [GOR], [C] and [A]. The scaled generating function $p(x')$ exists for all $x' \in E^*$ but is not necessarily finite; it is given by

$$p(x') = \lim_n \frac{1}{n} \ln \int_X \mathrm{e}^{n\langle x', x\rangle} \mathbb{M}_n[dx] = \ln \int_S \mathrm{e}^{\langle x', \varphi(s)\rangle} \beta[ds]. \tag{7.4}$$

The function p is automatically convex, a consequence of Hölder's inequality. If $\{x'_k\}$ is a sequence converging to x', then Fatou's Lemma implies that

$$\liminf_k \ln \int_S \mathrm{e}^{\langle x'_k, f(s)\rangle} \beta[ds] \geq \ln \int_S \mathrm{e}^{\langle x', f(s)\rangle} \beta[ds]. \tag{7.5}$$

Hence p is a closed convex function on E^*. Let G be a convex neighbourhood of $\frac{1}{2}(x+y) \in X$; there exist convex neighbourhoods $G_1 \ni x$ and $G_2 \ni y$ such that $\frac{1}{2}G_1 + \frac{1}{2}G_2 \subset G$. Since the random variables Y_j are independent, we have

$$\mathbb{M}_n[G_1] \cdot \mathbb{M}_n[G_2] \leq \mathbb{M}_{2n}[G]. \tag{7.6}$$

From this inequality, the existence and concavity of the RL–function μ follow immediately using the standard subadditivity argument [L].

Theorem 7.1 *Let $Y_1, Y_2 \dots$ be a sequence of i.i.d. random variables taking their values in a standard Borel space $(S, \mathcal{S})$ with common law β. Let $\varphi : S \to X$ be a random variable with values in a the measurable space X which satisfies the hypotheses of part II. Let $\{C_n\}$ be an LD–regular sequence in X with $C := \cap_n \mathrm{cl}\, C_n$. If there exists $x' \in \mathrm{dom}\, p$ such that the sequence of conditioned measures $\{\mathbb{M}_n^C\}$ is asymptotically I–null to the sequence of tilted measures $\{\mathbb{M}_n^{x'}\}$, then the law λ_n^C of Y_1 conditioned with respect to the event $\{T_n \in C_n\}$ converges in information as $n \to \infty$ to the probability measure*

$$\beta^{x'}[ds] := \frac{e^{\langle x', \varphi(s)\rangle}\beta[ds]}{\int_S e^{\langle x', \varphi(t)\rangle}\beta[dt]}\,; \tag{7.7}$$

that is,

$$\lim_{n\to\infty} \mathcal{H}(\lambda_n^C | \beta^{x'}) = 0. \tag{7.8}$$

Proof: The idea of the proof comes from [C]. Let ν be the infinite product measure on Ω with all factors equal to β and let $\nu_n^{x'} = \prod_{j\geq 1}\rho_j$, where the first n factors are equal to the measure $\beta^{x'}$ and the remaining ones to the measure β. We observe that the conditioned measure $\mathbb{M}_n^C$ is the image under T_n of the measure

$$\nu[\;\cdot\;|T_n \in C_n], \tag{7.9}$$

and that the tilted measure $\mathbb{M}_n^{x'}$ is the image under T_n of the product measure $\nu_n^{x'}$. By a change of variable, we have

$$\mathcal{H}(\mathbb{M}_n^C | \mathbb{M}_n^{x'}) = \mathcal{H}(\nu[\;\cdot\;|T_n \in C_n] | \nu_n^{x'}). \tag{7.10}$$

The law λ_n^C of Y_1 conditioned with respect to the event $\{T_n \in C_n\}$ is equal to the marginal of $\nu[\;\cdot\;|T_n \in C_n]$; therefore the theorem follows from Theorem 5.1 and Lemma 10.2. □

7.2 Lattice Case

Let $Y_1, Y_2, \dots$ be a sequence of i.i.d. integer-valued random variables with common law β. We assume that the variables have maximal span one. For φ, we choose the identity function, so that

$$T_n(\omega) := \frac{1}{n}\sum_{j=1}^{n} Y_j(\omega). \tag{7.11}$$

Here $X = E = \mathbb{R}$ with its Borel structure, and $E^* = \mathbb{R}$; the scaled generating function p is given by

$$p(x') = \ln \int_{\mathbb{R}} e^{x' \cdot t}\beta[dt]\,, \tag{7.12}$$

and we define as above the tilted measure

$$\beta^{x'}[ds] := \frac{e^{x' \cdot s}\beta[ds]}{\int_{\mathbb{R}} e^{x' \cdot t}\beta[dt]}\,. \tag{7.13}$$

Let μ be the RL–function of the pair $(\{\mathbb{M}_n\}, \{V_n\})$, let $\mathbb{M}_n = \beta \circ T_n^{-1}$ and $V_n = n$. By definition of μ, for any point $x \in \mathbb{R}$, we have

$$\liminf_n m_n[\{x\}] \leq \limsup_n m_n[\{x\}] \leq \mu(x). \tag{7.14}$$

Proving that the singleton $C := \{x\}$ is an LD–regular set is equivalent to proving that

$$\liminf_n m_n[\{x\}] = \lim_n m_n[\{x\}] = \mu(x). \tag{7.15}$$

There is a simple case where this can be done, namely, when we can show that

$$\liminf_n m_n[\{x\}] = 0. \tag{7.16}$$

Indeed, if this happens, we must have $\lim_n m_n[\{x\}] = \mu(x)$ since μ is nonpositive. Another elementary remark is that whenever we have

$$\lim_n m_n[\{x\}] = \mu(x), \tag{7.17}$$

we have also an analogous result for the tilted case for any x' such that $p(x')$ is finite:

$$\lim_n m_n^{x'}[\{x\}] = \mu^{x'}(x). \tag{7.18}$$

We make use of these two remarks to prove

Lemma 7.1 *Let x be a rational number with the properties:*

a) there exists $k \in \mathbb{N}$ so that $\mathrm{Prob}[\{T_k = x\}] > 0$*;*

b) there exists x' in the interior of $\mathrm{dom}\, p$ *such that* $\mathrm{grad}\, p(x') = x$.

Then we have

$$\lim_n \frac{1}{kn} \ln \mathbb{M}_{kn}[\{x\}] = \mu(x). \tag{7.19}$$

Proof: Let Z_i, $i = 1, 2, \ldots$, be i.i.d. random variables with common law $\beta^{x'}$; by choice of x', we have $\mathbb{E}[Z_i] = x$. The distribution of $\frac{1}{n}\sum_{j=1}^n Z_j$ is the tilted measure $\mathbb{M}_n^{x'}$. Since x' is in the interior of the essential domain of p, all moments of Z_i are finite and, by the Central Limit Theorem for lattice distributions (see [F]), we have

$$\mathrm{Prob}[\sum_{j=1}^{kn} Z_j = knx] = \mathbb{M}_{kn}^{x'}[\{x\}] = O(\frac{1}{\sqrt{kn}}). \tag{7.20}$$

It follows that

$$0 = \lim_n \frac{1}{kn} \ln \mathbb{M}_{kn}^{x'}[\{x\}] = \mu^{x'}(x). \tag{7.21}$$

But, clearly, we have

$$\begin{aligned} \mu^{x'}(x) &= x' \cdot x - p(x') + \mu(x) \qquad\qquad (7.22)\\ &= \lim_n \frac{1}{kn} \ln \mathbb{M}_{kn}^{x'}[\{x\}] = x' \cdot x - p(x') + \lim_n \frac{1}{kn} \ln \mathbb{M}_{kn}[\{x\}], \end{aligned}$$

which means precisely that $\{x\}$ is LD–regular for the pair $(\{\mathbb{M}_{kn}\}, \{V_{kn}\})$. □

Theorem 7.2 *Let $Y_1, Y_2 \ldots$ be a sequence of i.i.d. integer-valued random variables with maximal span one and common law β. Let* $\mathrm{dom}\, p$ *be the essential domain of the closed convex function $p : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$,*

$$p(x') = \ln \int_{\mathbb{R}} e^{x' \cdot s} \beta[ds]. \tag{7.23}$$

Assume that $\mathrm{dom}\, p$ *contains a neighbourhood of the origin. Let x be a rational number with the properties:*

a) there exists $k \in \mathbb{N}$ so that $\mathrm{Prob}[\{T_k = x\}] > 0$*;*

b) there exists x' in the interior of $\mathrm{dom}\, p$ *such that* $\mathrm{grad}\, p(x') = x$.

Then the law λ^x_{kn} of Y_1 conditioned with respect to the event $\{T_{kn} = x\}$ converges in information for $n \to \infty$ to the probability measure $\beta^{x'}$. Consequently, we have

$$\lim_{n\to\infty} \sum_{r \in \mathbb{Z}} |\lambda^x_{kn}[\{r\}] - \beta^{x'}[\{r\}]| = 0. \tag{7.24}$$

Proof: Since $\mathrm{dom}\, p$ contains a neighbourhood of the origin, we have an LDP for the pair $(\{\mathbb{M}_{kn}\}, \{V_{kn}\})$. The function p is differentiable on the interior of its essential domain, which implies that μ is strictly concave and continuous on

$$\Delta := \{y \in \mathbb{R} : y = \mathrm{grad}\, p(x'),\ x' \in \mathrm{int}\,(\mathrm{dom}\, p)\}. \tag{7.25}$$

The set $C = \{x\}$ is LD–regular (Lemma 7.1). We have $N_C = \{x\}$; by hypothesis, x is in Δ and this implies that x' is the unique subgradient to $(-\mu)$ at x and thus satisfies the hypothesis of part (b) of Theorem 6.1; hence Theorem 7.1 applies. The final statement says that we have convergence in the total variation metric as a consequence of the Kemperman–Pinsker inequality (see Proposition 10.3) □

7.3 A Counter–Example

The following example, inspired by a model from Statistical Mechanics, the Curie–Weiss model, shows that if the RL–function is not concave, then the conclusions of Theorem 6.1 need not obtain; this example shows that there are measures obtained by conditioning on convex sets, which are not equal to tilted measures. We also illustrate the fact that, for $|N_C| > 1$, it is necessary to study large deviations on a scale smaller than the scale $\{V_n\}$ in order to determine more precisely the concentration set of the conditioned measures.

Let $S := \{-1, +1\}$ and, for each $b \in [-1, +1]$, let β^b be the probability measure on S defined by

$$\beta^b[\{+1\}] := \frac{1+b}{2}\ ,\ \beta^b[\{-1\}] := \frac{1-b}{2}. \tag{7.26}$$

For each $j \in \mathbb{N}$, let (Ω_j, λ^b_j) be a copy of (S, β^b); let Ω be the infinite product space $\prod_{j\geq 1} \Omega_j$ and let λ^b be the infinite product measure $\prod_{j\geq 1} \lambda^b_j$. Let Y_j, $j \in \mathbb{N}$, be the random variable defined on Ω by

$$Y_j : \Omega \to \{-1, +1\}\ ,\ Y_j(\omega) := \omega_j, \tag{7.27}$$

and for each $n \in \mathbb{N}$ let

$$T_n(\omega) := \frac{1}{n}\sum_{j=1}^{n} Y_j(\omega) = \frac{1}{n}\sum_{j=1}^{n} \omega_j. \tag{7.28}$$

The random variable T_n is distributed according to the probability measure ν_n,

$$\nu_n[d\omega] := \frac{e^{n\frac{\alpha}{2}T_n^2(\omega)}\lambda^0[d\omega]}{\int_\Omega e^{n\frac{\alpha}{2}T_n^2(\omega')}\lambda^0[d\omega']}. \tag{7.29}$$

In this example $V_n = n$, $E = E^* = \mathbb{R}$ and $X = [-1, +1] \subset \mathbb{R}$. The distribution of T_n on X is $\mathbb{M}_n = \nu_n \circ T_n^{-1}$. Let s be defined on $[0, 1]$ by

$$s(y) := \begin{cases} 0 & \text{if } y = 0, \\ -y\ln y - (1-y)\ln(1-y) & \text{if } 0 < y < 1, \\ 0 & \text{if } y = 1. \end{cases} \tag{7.30}$$

The RL–function μ of the pair $(\{\mathbb{M}_n\}, \{V_n\})$ is

$$\mu(x) = s(\frac{1+x}{2}) + \frac{\alpha x^2}{2} - p_\alpha, \tag{7.31}$$

where p_α is the constant

$$p_\alpha = \sup_{x\in[0,1]} \{s(\frac{1+x}{2}) + \frac{\alpha x^2}{2}\}. \tag{7.32}$$

For each $x' \in \mathbb{R}$, the tilted measure $\mathbb{M}_n^{x'}$ is the image by the map T_n of the measure

$$\nu_n^{x'}[d\omega] := \frac{e^{n(\frac{\alpha}{2}T_n^2(\omega)+x'T_n(\omega))}\lambda^0[d\omega]}{\int_\Omega e^{n(\frac{\alpha}{2}T_n^2(\omega')+x'T_n(\omega'))}\lambda^0[d\omega']}. \tag{7.33}$$

The scaled generating function $p_\alpha(x')$ can be computed by Varadhan's Theorem,

$$p_\alpha(x') = \sup_{x\in[0,1]} \{s(\frac{1+x}{2}) + \frac{\alpha x^2}{2} + x' \cdot x\}. \tag{7.34}$$

For $\alpha > \frac{1}{2}$, the RL–function μ is not concave: the RL–function attains its maximum value at the points $\pm m^*$, where $m^* = m^*(\alpha)$ is the positive root of the equation

$$\tanh 2\alpha x = x. \tag{7.35}$$

The concave envelope of μ,

$$\widehat{\mu}(x) = \begin{cases} \mu(x) & \text{if } x \in [-1, -m^*], \\ 0 & \text{if } x \in [-m^*, m^*], \\ \mu(x) & \text{if } x \in [m^*, 1], \end{cases} \tag{7.36}$$

is strictly larger than μ on the open interval $(-m^*, m^*)$.

We choose the parameter $\alpha > \frac{1}{2}$. Let $C := [-a, +a]$, with $0 < a < m^*$. The set C is LD–regular; since the RL–function is symmetric, the concentration set is given by

$$N_C = \{x \in C : \mu(x) = \sup_{y\in C} \mu(y)\} = \{-a, +a\}. \tag{7.37}$$

For any $x' \in \mathbb{R}$, we have

$$\lim_n \frac{1}{n}\mathcal{H}(\mathbb{M}_n^C|\mathbb{M}_n^{x'}) = \lim_n \frac{1}{n}\mathcal{H}(\nu_n[\ \cdot\ |\{T_n \in C\}]|\nu_n^{x'}) \geq -\mu(a) > 0. \tag{7.38}$$

A simple analysis shows that the tilted measures are concentrated on the subset $N^{x'}$, where

$$N^{x'} = \begin{cases} \{b_*(x')\},\ b_*(x') \in (-1, -m^*) & \text{if } x' < 0, \\ \{-m^*, +m^*\} & \text{if } x' = 0, \\ \{b_*(x')\},\ b_*(x') \in (m^*, 1) & \text{if } x' > 0. \end{cases} \tag{7.39}$$

One can prove that the sequence $\{\nu_n^{x'}\}$ converges to a probability measure, denoted by $\nu_\infty^{x'}$. Since $\nu_\infty^{x'}$ is invariant under any finite permutation of the ω_j, it follows from de Finetti's Theorem (see, for example, [F]) that

$$\nu_\infty^{x'} = \begin{cases} \lambda^{b_*(x')} & \text{if } x' \neq 0, \\ \frac{1}{2}\lambda^{-m^*} + \frac{1}{2}\lambda^{m^*} & \text{if } x' = 0. \end{cases} \tag{7.40}$$

The conditioned measure $\nu_n[\ \cdot\ |\{T_n \in C\}]$ has a limit as n tends to infinity, which is also invariant under any finite permutation of the ω_j, so that again we can use de Finetti's Theorem. We show that we find different limiting measures by choosing different LD-regular sequences $\{C_n\}$ converging to C. For example, we choose $\varepsilon_n > 0$, such that $\varepsilon_n \downarrow 0$ and $n\varepsilon_n \to \infty$ faster than $\ln n$, as $n \to \infty$ and $C_n := [-a, a + \varepsilon_n]$; the sequence $\{C_n\}$ converges to C and is LD-regular, thus the concentration set for the sequence of conditioned measures $\{\mathbb{M}_n^C\}$ is again $N_C = \{-a, +a\}$. However, a finer analysis based on Lemma 7.2 shows that the measures $\mathbb{M}_n^C$ converge to a Dirac measure δ_a concentrated at a, so that $\lim_{n\to\infty} \nu_n[\ \cdot\ |\{T_n \in C\}] = \lambda^a$. Notice that λ^a is not of the form

$$\int_{\mathbb{R}} \lambda^{b_*(x')}\rho[dx'], \tag{7.41}$$

with ρ a probability measure on $\mathbb{R}$.

Lemma 7.2 *Let J be an open subinterval of $[-1, 1]$ and let $D_\infty := [a-\delta, a] \subset D_n := [a - \delta, a + \varepsilon_n] \subset J$, where $\delta > 0$ and $\varepsilon_n \downarrow 0$ as $n \to \infty$. With $\mu'(x)$ denoting the derivative of the RL-function, if $\inf_{x\in J} \mu'(x)$ is strictly positive and $n\varepsilon_n/\ln n \to \infty$, then*

$$\lim_{n\to\infty} \frac{\mathbb{M}_n[D_\infty]}{\mathbb{M}_n[D_n]} = 0. \tag{7.42}$$

Proof: We prove the lemma for the case where the function μ is given by

$$\mu(x) := g(x) + s(\frac{1+x}{2}), \tag{7.43}$$

with $g : [-1, 1] \to \mathbb{R}$ a continuously differentiable function. The measure $\mathbb{M}_n$ is in this case (we can omit normalization because the limit involves a ratio)

$$\mathbb{M}_n[B] := \sum_{k=0}^{n} \rho_n(k)\delta_{\frac{2k}{n}-1}[B], \tag{7.44}$$

where

$$\rho_n(k) := \binom{n}{k} e^{ng(\frac{2k}{n}-1)}\ . \tag{7.45}$$

From Stirling's formula, we have the following estimate: given $0 < \alpha < \alpha' < 1$, there exist $K_0 > 0$ and $n_0 \in \mathbb{N}$ such that, for $n > n_0$ and k such that $\alpha < \frac{k}{n} < \alpha'$, we have

$$|\ln \binom{n}{k} - ns(\frac{k}{n})| < K_0 \ln n. \tag{7.46}$$

From this it follows that there exist $K_1 > 0$ and $n_1 \in \mathbb{N}$ such that, for $n > n_1$ and k such that $\frac{2k}{n} - 1 \in J$, we have

$$|\ln \rho_n(k) - n\mu(\frac{2k}{n} - 1)| < K_1 \ln n. \tag{7.47}$$

Define the measures $\widehat{\mathbb{M}}_n$, $n \geq 1$, on the subsets of $[-1,1]$ by

$$\widehat{\mathbb{M}}_n[B] := \sum_{k=0}^{n} \mathrm{e}^{n\mu(\frac{2k}{n}-1)} \delta_{\frac{2k}{n}-1}[B]; \tag{7.48}$$

for $B \subset J$ and $n > n_1$, we have

$$\mathrm{e}^{-K_1 \ln n} \widehat{\mathbb{M}}_n[B] \leq \mathbb{M}_n[B] \leq \mathrm{e}^{K_1 \ln n} \widehat{\mathbb{M}}_n[B]. \tag{7.49}$$

Now let B be an interval of length $|B|$; by the principle of the largest term, we have

$$\max_{k \in \phi^{-1}B} \mu(\frac{2k}{n} - 1) \leq \frac{1}{n} \ln \widehat{\mathbb{M}}_n[B] \leq \max_{k \in \phi^{-1}B} \mu(\frac{2k}{n} - 1) + \frac{\ln(n|B|)}{n}, \tag{7.50}$$

where $\phi^{-1}B := \{k : \frac{2k}{n} - 1 \in B\}$. Thus there exist $K_2 > 0$ and $n_2 \in \mathbb{N}$ such that, for $n > n_2$, we have

$$0 \leq \frac{\mathbb{M}_n[D_\infty]}{\mathbb{M}_n[D_n]} \leq \mathrm{e}^{(K_2 \ln n + \ln \delta)} \cdot \mathrm{e}^{\left(n\{\max_{k \in \phi^{-1}D_\infty} \mu(\frac{2k}{n}-1) - \max_{k \in \phi^{-1}D_n} \mu(\frac{2k}{n}-1)\}\right)}. \tag{7.51}$$

Since μ is continuously differentiable, we have

$$\{\max_{k \in \phi^{-1}D_\infty} \mu(\frac{2k}{n} - 1) - \max_{k \in \phi^{-1}D_n} \mu(\frac{2k}{n} - 1)\} \leq -\frac{\varepsilon_n}{2} \cdot \inf_{x \in J} \mu'(x); \tag{7.52}$$

taking $n \to \infty$, the result follows. □

Let δ be strictly positive, so that

$$[-a, a + \varepsilon_n] = [-a, -a + \delta] \cup (-a + \delta, a - \delta) \cup [a - \delta, a + \varepsilon_n]; \tag{7.53}$$

we write the measure $\mathbb{M}_n^C$ as

$$\begin{aligned} \mathbb{M}_n^C[\;\cdot\;] \;&=\; \mathbb{M}_n[\;\cdot\;|[-a, -a + \delta]] \cdot \frac{\mathbb{M}_n[[-a, -a + \delta]]}{\mathbb{M}_n[C_n]} \\ &+\; \mathbb{M}_n[\;\cdot\;|(-a + \delta, a - \delta)] \cdot \frac{\mathbb{M}_n[(-a + \delta, a - \delta)]}{\mathbb{M}_n[C_n]} \\ &+\; \mathbb{M}_n[\;\cdot\;|[a - \delta, a + \varepsilon_n]] \cdot \frac{\mathbb{M}_n[[a - \delta, a + \varepsilon_n]]}{\mathbb{M}_n[C_n]}. \end{aligned} \tag{7.54}$$

Using $\mathbb{M}_n[[-a, -a + \delta]] = \mathbb{M}_n[[a - \delta, a]]$, it follows immediately that

$$\mathbb{M}_n^C \to \delta_a \;,\; n \to \infty. \tag{7.55}$$

8 Conditional Limit Theorems: Stationary Case

We now consider dependent random variables. The special case of Markov conditioning is addressed in [CCC]; our results cover a much wider range of applications. The main result, proved in 8.2, is a conditional limit theorem for stationary sequences of random variables $\{Y_i\}$ indexed by the points of the lattice $\mathbb{Z}^d$. Two concrete applications of this theorem are given in 8.3 and 8.4. In particular, we give in 8.4 a new proof of the LDP for the empirical measure. In the first subsection, 8.1, we define the notion of equilibrium state. For that purpose, we introduce the concept of a weakly dependent measure and recall the basic properties of the specific information gain.

8.1 Equilibrium States

Let $(S, \mathcal{S})$ be a standard Borel space; for each $i \in \mathbb{Z}^d$, let $(\Omega_i, \mathcal{F}_i)$ be a copy of $(S, \mathcal{S})$ and define $(\Omega, \mathcal{F})$ as the product space. We set $|i| := \max_k |i_k|$ if $i = (i_1, \ldots i_d) \in \mathbb{Z}^d$ and use $|\Lambda|$ to denote the cardinality of a subset $\Lambda \subset \mathbb{Z}^d$. $(\Omega_\Lambda, \mathcal{F}_\Lambda)$ denotes the product space $(\prod_{i\in\Lambda} \Omega_i, \prod_{i\in\Lambda} \mathcal{F}_i)$ and $(\Omega, \mathcal{F})$ stands for $(\Omega_{\mathbb{Z}^d}, \mathcal{F}_{\mathbb{Z}^d})$. We write $f \in \mathcal{F}_\Lambda$ to mean that the function f is $\mathcal{F}_\Lambda$-measurable. For $i \in \mathbb{Z}^d$, we have the translation operator θ_i acting on $\mathbb{Z}^d$ by $j \mapsto j + i$. This lifts to

$$\theta_i : \Omega_\Lambda \longrightarrow \Omega_{\theta_i\Lambda} \quad \text{with} \quad (\theta_i\omega)_j = \omega_{j-i}\,. \tag{8.1}$$

For $f \in \mathcal{F}_\Lambda$, we define $\theta_i f \in \mathcal{F}_{\theta_i\Lambda}$ by $\theta_i f(\omega) = f(\theta_{-i}\omega)$. For the measure ν, we define $\theta_i\nu$ so that $\int f\, d(\theta_i\nu) = \int(\theta_{-i}f)\, d\nu$ obtains. For any bounded function $\varphi : \Omega \to \mathbb{R}^k$, $\varphi = (f_1, \ldots, f_k)$, we set

$$|\varphi(\omega)| := \max_{j=1,\ldots,k} |f_j(\omega)|, \quad \|\varphi\| := \sup_{\omega\in\Omega} |\varphi(\omega)|. \tag{8.2}$$

Definition 8.1 *A real valued function f on Ω is called* **local** *if f is $\mathcal{F}_\Lambda$-measurable for some finite Λ. The symbol $\mathcal{F}_{\text{loc}}$ denotes the space of all* bounded *local functions; the closure of $\mathcal{F}_{\text{loc}}$ with respect to the norm $\|\cdot\|$ is denoted $\mathcal{F}_{\text{qloc}}$ and $f \in \mathcal{F}_{\text{qloc}}$ is called* **quasilocal**. *The vector space of all finite signed measures on $(\Omega, \mathcal{F})$ is denoted by $\mathcal{M}$; the probability measures, by $\mathcal{M}_1^+$; translation invariant probability measures, by $\mathcal{M}_1^{+,\theta}$.*

For $f \in \mathcal{F}_{\text{qloc}}$ and $\nu \in \mathcal{M}$, there is the natural pairing

$$\langle f, \nu\rangle = \int_\Omega f(\omega)\nu[d\omega]. \tag{8.3}$$

Equipped with the $\sigma(\mathcal{F}_{\text{qloc}}, \mathcal{M})$-topology, $\mathcal{M}$ and $\mathcal{F}_{\text{qloc}}$ are mutually dual locally convex Hausdorff topological vector spaces. The set $\mathcal{M}_1^+$ is a convex subset of $\mathcal{M}$ and $\mathcal{M}_1^{+,\theta}$ is a convex subset of $\mathcal{M}_1^+$. Though $\mathcal{F}_{\text{qloc}}$ and $\mathcal{F}_{\text{loc}}$ induce different topologies on $\mathcal{M}$, the topology on $\mathcal{M}_1^+$ induced by $\mathcal{F}_{\text{loc}}$ coincides with the topology $\sigma(\mathcal{M}_1^+, \mathcal{F}_{\text{qloc}})$. A sequence $\{\nu_n\}$ of probability measures on Ω converges to the probability measure ν in this topology if and only if

$$\lim_{n\to\infty} \int f\, d\nu_n = \int f\, d\nu \tag{8.4}$$

for each $f \in \mathcal{F}_{\mathrm{loc}}$. Sometimes this convergence is called τ**−convergence**. *Unless otherwise specified, the discussion of limits of probability measures below employs this topology.*

We fix once and for all a sequence of finite subsets $\Lambda_n \subset \mathbb{Z}^d$, $n \in \mathbb{N}$,

$$\Lambda_n := [-n, n]^d \subset \mathbb{Z}^d \tag{8.5}$$

such that eventually any finite subset of $\mathbb{Z}^d$ is contained in Λ_n; the corresponding scale is $\{V_n := (2n+1)^d\}$. The complement of Λ_n in $\mathbb{Z}^d$ is written $\overline{\Lambda}_n := \mathbb{Z}^d \backslash \Lambda_n$. The restriction of a probability measure λ to a σ−algebra $\mathcal{B}$ is denoted $\lambda_{|\mathcal{B}}$. We set

$$c_n(\lambda) := \sup\{| \ln \left(\frac{\mathbb{E}_\lambda(f)}{\mathbb{E}_{\lambda_{|\mathcal{F}_{\Lambda_n}} \otimes \lambda_{|\mathcal{F}_{\overline{\Lambda}_n}}}(f)} \right) | : \ f \geq 0 \text{ in } \mathcal{F}_{\mathrm{loc}}\}. \tag{8.6}$$

(If the numerator and denominator are both zero, then the quotient in the definition of $c_n(\nu)$ is defined to be 1.)

Definition 8.2 *A translation invariant probability measure λ on $(\Omega, \mathcal{F})$ is* **weakly dependent** *if*

$$\lim_{n \to \infty} \frac{1}{V_n} c_n(\lambda) = 0. \tag{8.7}$$

Remark: In the case of a local specification defined by an absolutely summable potential, Gibbs measures are weakly dependent (see Section 9). In [Su], a similar definition is introduced ((3.8) of [Su]); however, the present formulation using (8.6) is more convenient.

In this setting, we define two important functions: the specific information gain $h(\ \cdot\ |\beta)$ on $\mathcal{M}_1^{+,\theta}$ and the scaled generating function $p(\ \cdot\ |\beta)$ for the empirical measure, defined on $\mathcal{F}_{\mathrm{qloc}}$. In the following, β is a fixed translation invariant weakly dependent probability measure. Let λ be an element of $\mathcal{M}_1^+$; we set

$$\mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\lambda|\beta) := \mathcal{H}(\lambda_{|\mathcal{F}_{\Lambda_n}} | \beta_{|\mathcal{F}_{\Lambda_n}}). \tag{8.8}$$

Definition 8.3 *A probability measure λ has* **specific information gain** $h(\lambda|\beta)$ *relative to the probability measure β, if*

$$h(\lambda|\beta) := \lim_{n \to \infty} \frac{1}{V_n} \mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\lambda|\beta) \tag{8.9}$$

exists.

Theorem 8.1 *Let $\beta \in \mathcal{M}_1^{+,\theta}$ be a weakly dependent probability measure.*

a) For any $\lambda \in \mathcal{M}_1^{+,\theta}$, the specific information gain

$$h(\lambda|\beta) := \lim_{n\to\infty} \frac{1}{V_n} \mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\lambda|\beta) \tag{8.10}$$

exists as a nonnegative number or $+\infty$ and $h(\,\cdot\,|\beta)$ is a lower semicontinuous affine function on $\mathcal{M}_1^{+,\theta}$.

b) The level-sets of $h(\,\cdot\,|\beta)$ in $\mathcal{M}_1^{+,\theta}$ are compact.

The proof is given in Section 10.2.

Definition 8.4 *Let f be an element of $\mathcal{F}_{\mathrm{qloc}}$ and let $\beta \in \mathcal{M}_1^{+,\theta}$ be weakly dependent. We define*

$$p(f|\beta) := \limsup_n \frac{1}{V_n} \ln \int_\Omega \exp\{\sum_{j\in\Lambda_n} f(\theta_j\omega)\}\beta[d\omega]. \tag{8.11}$$

The empirical measure is the probability measure given by

$$\frac{1}{V_n} \sum_{j\in\Lambda_n} \delta_{\theta_j\omega}. \tag{8.12}$$

Since

$$\sum_{j\in\Lambda_n} f(\theta_j\omega) = V_n\langle f, \frac{1}{V_n} \sum_{j\in\Lambda_n} \delta_{\theta_j\omega}\rangle, \tag{8.13}$$

the function $p(\,\cdot\,|\beta)$ is the scaled generating function for the empirical measure defined on the probability space $(\Omega, \mathcal{F}, \beta)$.

Theorem 8.2 *Let $\beta \in \mathcal{M}_1^{+,\theta}$ be a weakly dependent probability measure.*

a) For any $f \in \mathcal{F}_{\mathrm{qloc}}$, $p(f|\beta)$ given by (8.11) exists as a limit, not just as a limit superior, and $p(\,\cdot\,|\beta)$ is a lower semicontinuous convex function on $\mathcal{F}_{\mathrm{qloc}}$.

b) $p(\,\cdot\,|\beta)$ and $h(\,\cdot\,|\beta)$ are conjugate functions:

$$\begin{aligned} p(f|\beta) &= \sup\{\langle f,\lambda\rangle - h(\lambda|\beta) : \lambda \in \mathcal{M}_1^{+,\theta}\}, \\ h(\lambda|\beta) &= \sup\{\langle f,\lambda\rangle - p(f|\beta) : f \in \mathcal{F}_{\mathrm{qloc}}\}. \end{aligned} \tag{8.14}$$

c) For any $\alpha \in \mathcal{M}_1^{+,\theta}$ and $f \in \mathcal{F}_{\mathrm{qloc}}$, we have

$$\langle f,\alpha\rangle \le p(f|\beta) + h(\alpha|\beta). \tag{8.15}$$

The proofs of a) and b) are consequences of the results of Section 8.4. One can give a direct and straightforward demonstration of a) along the lines of the proof of Lemma 8.2, while c) follows from b).

Definition 8.5 *Let $f \in \mathcal{F}_{\mathrm{qloc}}$ and $\beta \in \mathcal{M}_1^{+,\theta}$ be weakly dependent. A translation invariant probability measure α is an* **(f,β)–equilibrium state** *if*

$$\langle f,\alpha\rangle = p(f|\beta) + h(\alpha|\beta). \tag{8.16}$$

In this section we shall make use of the following two operations. Let λ be any probability measure on Ω. For any $k \in \mathbb{N}$, we define a periodic probability measure $\mathcal{P}_k\lambda$ by

$$\mathcal{P}_k\lambda := \prod_{j \in \mathbb{Z}^d} \theta_{(2k+1)j}(\lambda_{|\mathcal{F}_{\Lambda_k}}). \tag{8.17}$$

We call $\mathcal{P}_k$ the **blocking operation**. We define also the **averaging operation** $\mathcal{A}_k$ by

$$\mathcal{A}_k := \frac{1}{V_k} \sum_{j \in \Lambda_k} \theta_j, \tag{8.18}$$

acting on the space of measures or on the space of quasilocal functions; by definition of the action of $\mathbb{Z}^d$, we have

$$\langle \mathcal{A}_n f, \lambda \rangle = \langle f, \mathcal{A}_n \lambda \rangle. \tag{8.19}$$

It is immediate that ν is a limit point of the sequence $\{\lambda_k\}$ if, and only if, ν is a limit point of $\{\mathcal{P}_k\lambda_k\}$. An elementary estimate shows that ν is a limit point of $\{\mathcal{A}_k\lambda_k\}$ if, and only if, ν is a limit point of $\{\mathcal{A}_k(\mathcal{P}_k\lambda_k)\}$. Notice that $\mathcal{A}_k\mathcal{P}_k\lambda$ is an *ergodic* probability measure. To see this, it suffices to note that, for any $g \in \mathcal{F}_{\mathrm{loc}}$, we have

$$\int_\Omega g(\omega) \cdot \theta_i g(\omega) (\mathcal{A}_k\mathcal{P}_k\lambda)[d\omega] = \left(\int_\Omega g(\omega)(\mathcal{A}_k\mathcal{P}_k\lambda)[d\omega] \right)^2, \tag{8.20}$$

provided $|i|$ is large enough.

Lemma 8.1 *Let $\beta \in \mathcal{M}_1^{+,\theta}$ be weakly dependent.*

a) For any probability measure λ, we have

$$\left(\mathcal{H}_{\mathcal{F}_{\Lambda_m}}(\lambda|\beta) - c_m(\beta) - \ln V_m\right)/V_m \leq h(\mathcal{A}_m\mathcal{P}_m\lambda|\beta) \leq \left(\mathcal{H}_{\mathcal{F}_{\Lambda_m}}(\lambda|\beta) + c_m(\beta)\right)/V_m. \tag{8.21}$$

b) For any sequence $\{\lambda_n\}$ of probability measures, we have

$$\limsup_n \frac{1}{V_n} \mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\lambda_n|\beta) = \limsup_n h(\mathcal{A}_n\mathcal{P}_n\lambda_n|\beta), \tag{8.22}$$

and the corresponding equality with lim inf *obtains.*

The proof of a) is given in Section 10.2, while b) is a direct consequence of a).

8.2 Conditional Limit Theorem: General Case

Let β be a translation invariant weakly dependent probability measure. We consider a random variable $\varphi : \Omega \to X$, defined on the probability space $(\Omega, \mathcal{F}, \beta)$ with values in a measurable space $(X, \mathcal{B})$, which satisfies the hypothesis of part II (see Section 4.1). For each n, we define the random variable $T_n : \Omega \to X$ by

$$T_n(\omega) := \mathcal{A}_n\varphi(\omega). \tag{8.23}$$

If for $x' \in E^*$ the function $f_\varphi^{x'} : \Omega \to \mathbb{R}$ defined by

$$f_\varphi^{x'}(\omega) := \langle x', \varphi(\omega) \rangle \tag{8.24}$$

is quasilocal, then

$$\lim_n p_n(x') := \lim_n \frac{1}{V_n} \ln \int_\Omega \exp\{\sum_{j\in\Lambda_n} f_\varphi^{x'}(\theta_j\omega)\}\beta[d\omega] = p(x') \tag{8.25}$$

exists. We set

$$\beta_n^{x'}[d\omega] := \exp\{\sum_{j\in\Lambda_n} f_\varphi^{x'}(\theta_j\omega) - V_n p_n(x')\}\beta[d\omega], \tag{8.26}$$

and

$$\beta_n^C[d\omega] := \beta[d\omega | T_n \in C_n], \tag{8.27}$$

where $C_n \in \mathcal{B}$ is a sequence of sets with $\beta[T_n \in C_n] > 0$. Recall that for every $\rho \in \mathcal{M}_1^{+,\theta}$

$$h(\rho|\beta) = \lim_n \frac{1}{V_n}\mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\rho|\beta) \tag{8.28}$$

exists, and is non–negative.

Theorem 8.3 *In the above setting, assume that there exists $x' \in E^*$ such that the function $f_\varphi^{x'}$ is quasilocal, and*

$$\lim_n \frac{1}{V_n}\mathcal{H}(\beta_n^C|\beta_n^{x'}) = 0. \tag{8.29}$$

Then the set of limit points of the sequence

$$\{\mathcal{A}_n\beta[\ \cdot\ |T_n \in C_n]\} \tag{8.30}$$

is non–empty, and any limit point β^C satisfies the identity

$$h(\beta^C|\beta) = -\lim_n \frac{1}{V_n}\ln\beta[T_n \in C_n] = \int_\Omega f_\varphi^{x'}(\omega)\beta^C[d\omega] - p(x'). \tag{8.31}$$

In particular β^C is an $(f_\varphi^{x'},\beta)$–equilibrium state.

Proof: We have

$$\frac{1}{V_n}\mathcal{H}(\beta_n^C|\beta_n^{x'}) = \frac{1}{V_n}\mathcal{H}(\beta_n^C|\beta) - \int_\Omega f_\varphi^{x'}(\omega)\mathcal{A}_n\beta_n^C[d\omega] + p_n(x') \tag{8.32}$$

and

$$\frac{1}{V_n}\mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\beta_n^C|\beta) \le \frac{1}{V_n}\mathcal{H}(\beta_n^C|\beta) = -\frac{1}{V_n}\ln\beta[T_n \in C_n]. \tag{8.33}$$

Since $f_\varphi^{x'}$ is quasilocal, $\sup_\omega |f_\varphi^{x'}(\omega)| = \|f_\varphi^{x'}\| < \infty$ and $\lim_n p_n(x') = p(x')$ exists and is finite. Using Lemma 8.1, (8.33) and (8.32) we get

$$\begin{aligned}
0 &\le \liminf_n \frac{1}{V_n}\mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\beta_n^C|\beta) = \liminf_n h(\mathcal{A}_n\mathcal{P}_n\beta_n^C|\beta) \\
&\le \liminf_n -\frac{1}{V_n}\ln\beta[T_n \in C_n] \le \limsup_n -\frac{1}{V_n}\ln\beta[T_n\in C_n] \\
&= \limsup_n \Big(\frac{1}{V_n}\mathcal{H}(\beta_n^C|\beta_n^{x'}) + \int_\Omega f_\varphi^{x'}(\omega)\mathcal{A}_n\beta_n^C[d\omega] - p_n(x')\Big) \\
&\le \limsup_n \int_\Omega f_\varphi^{x'}(\omega)\mathcal{A}_n\beta_n^C[d\omega] - p(x') \le \|f_\varphi^{x'}\| - p(x') < \infty.
\end{aligned} \tag{8.34}$$

Since the level–sets of the specific information gain are compact, the set of limit points of the sequence $\{\mathcal{A}_n \mathcal{P}_n \beta_n^C\}$ is non–empty. This set coincides with the set of limit points of the sequence $\{\mathcal{A}_n \beta_n^C\}$. Let β^C be such a limit point. Since $f_\varphi^{x'}$ is quasilocal,

$$\lim_n \int f_\varphi^{x'} d\mathcal{A}_n \beta_n^C = \int f_\varphi^{x'} d\beta^C; \tag{8.35}$$

the lower semicontinuity of the specific information gain on $\mathcal{M}_1^{+,\theta}$ implies that

$$\begin{aligned} h(\beta^C|\beta) &\leq \liminf_n -\frac{1}{V_n} \ln \beta[T_n \in C_n] \leq \limsup_n -\frac{1}{V_n} \ln \beta[T_n \in C_n] \quad (8.36) \\ &\leq \int_\Omega f_\varphi^{x'}(\omega) \beta^C[d\omega] - p(x'). \end{aligned}$$

Since β^C is translation invariant, Theorem 8.2 implies the reversed inequality

$$h(\beta^C|\beta) \geq \int_\Omega f_\varphi^{x'}(\omega) \beta^C[d\omega] - p(x'). \tag{8.37}$$

□

The above result may be applied to sequences which approximate $\{\mathcal{A}_n \varphi(\omega)\}$. We have the following

Corollary 8.1 *Assume the hypotheses of Theorem 8.3 are satisfied with $\{T_n\}$ which does not equal $\{\mathcal{A}_n \varphi(\omega)\}$, but such that*

$$\lim_{n\to\infty} \|\langle x', T_n(\omega)\rangle - \langle x', \mathcal{A}_n \varphi(\omega)\rangle\| = 0. \tag{8.38}$$

Then the conclusions of the Theorem still obtain.

Remark: A crucial step of the proof is the use of the lower semicontinuity of the specific information gain on $\mathcal{M}_1^{+,\theta}$. The following example shows that this property fails to hold on $\mathcal{M}_1^+$, even if the limiting measure is translation invariant. We construct a sequence of probability measures $\{\lambda_n\}$ such that $\{\lambda_n\}$ converges (the topology is that induced by $\mathcal{F}_{\text{qloc}}$) to a translation invariant probability measure and

$$\lim_n h(\lambda_n|\beta) < h(\lim_n \lambda_n|\beta). \tag{8.39}$$

Consider $\mathbb{Z}^1$ with $\Omega_i = \{0, 1\}$. We have $\Lambda_n = \{-n, \ldots, n\}$ and $V_n = 2n+1$. Let β be the product probability measure with $\beta[\{1\}] = a$, for all $i \in \mathbb{Z}$, where $0 < a < 1$, and let ν be the product probability measure with $\nu[\{\omega_i = 1\}] = a'$ for all $i \in \mathbb{Z}$, where $0 < a' < 1$ and $a' \neq a$. Define

$$\nu_k[\{\omega_i = 1\}] = \begin{cases} a' & \text{if } i^2 < k; \\ a & \text{otherwise.} \end{cases} \tag{8.40}$$

Then $\{\nu_n\}$ converges to ν. Set $\lambda_n := \mathcal{P}_n \nu_n$; $\{\lambda_n\}$ converges to ν. Note that $\mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\nu_n|\beta)$ is of order $2\sqrt{n}$, so that Lemma 8.1 implies the equality

$$\lim_{n\to\infty} h(\lambda_n|\beta) = \lim_{n\to\infty} \frac{1}{V_n} \mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\nu_n|\beta) = 0. \tag{8.41}$$

But $h(\nu|\beta)$ is non–zero.

8.3 Conditional Limit Theorem: Euclidean Case

We give an application of Theorem 8.3 when the measurable space $(X, \mathcal{B})$ is a compact subset of $\mathbb{R}^k$ with its Borel structure. We fix a translation invariant weakly dependent probability measure β on $(\Omega, \mathcal{F})$. Let $\varphi : \Omega \to \mathbb{R}^k$ be a quasilocal function. Recall the norms defined in (8.2). Let $E = E^* = \mathbb{R}^k$ with the pairing given by the Euclidean scalar product $\langle x', x\rangle$. Let $X \subset E$ be a closed ball at the origin with max norm radius larger than $\|\varphi\|$. The distribution on X of $T_n = \mathcal{A}_n\varphi$ is $\mathbb{M}_n = \beta \circ T_n^{-1}$.

Lemma 8.2 *The RL-function μ of the pair $(\{\mathbb{M}_n\}, \{V_n\})$ exists; it is a concave function on $\mathbb{R}^k$. The scaled generating function*

$$p(x') := \lim_n \frac{1}{V_n} \ln \int_X e^{\{V_n\langle x', x\rangle\}} \mathbb{M}_n[dx], \tag{8.42}$$

exists; it is a closed convex function with $\operatorname{dom} p = \mathbb{R}^k$.

Proof: Let $B_\varepsilon(a)$ be an open ball of radius ε and center $a \in \mathbb{R}^k$. Let a_0, a_1, a_2 be elements of R^k which satisfy $a_0 + a_1 = 2a_2$ and let $0 < \varepsilon'' < \varepsilon' < \varepsilon$; we shall prove

$$\underline{m}[B_\varepsilon(a_2)] \geq \frac{\overline{m}[B_{\varepsilon''}(a_0)] + \underline{m}[B_{\varepsilon''}(a_1)]}{2} \,. \tag{8.43}$$

If we put $a_0 = a_1 = a_2$ in (8.43), we have

$$\underline{m}[B_\varepsilon(a_0)] \geq \overline{m}[B_{\varepsilon''}(a_0)] \,, \tag{8.44}$$

which implies $\underline{\mu} \geq \overline{\mu}$. To see that the resulting function, denoted by μ, is concave, we note that (8.43) implies the inequality

$$\mu(a_2) \geq \frac{\mu(a_0) + \mu(a_1)}{2} \,. \tag{8.45}$$

Since μ is upper semicontinuous, this implies the concavity of μ. The existence of p is a consequence of Varadhan's Theorem; since X is compact, we have an LDP. Therefore

$$\begin{aligned} p(x') &= \lim_n \frac{1}{V_n} \ln \int_X e^{\{V_n\langle x', x\rangle\}} \mathbb{M}_n[dx] \\ &= \sup_{x\in X}\{\langle x', x\rangle + \mu(x)\}. \end{aligned} \tag{8.46}$$

We return to the proof of (8.43). If φ were a function of a single coordinate and β a product measure, then this would follow from (7.6); we have to show that, under the existing assumptions, the inequality got from (7.6) by taking logarithms and dividing by V_n continues to hold up to a small correction (which vanishes as n goes to infinity) provided we work on a sufficiently coarse scale. Given the box Λ_n, we define for $n > m$ the sublattice $\Lambda_{n|m}$, whose points are the centers of all translates of Λ_m by multiples of $2m + 1$ which stay inside Λ_n:

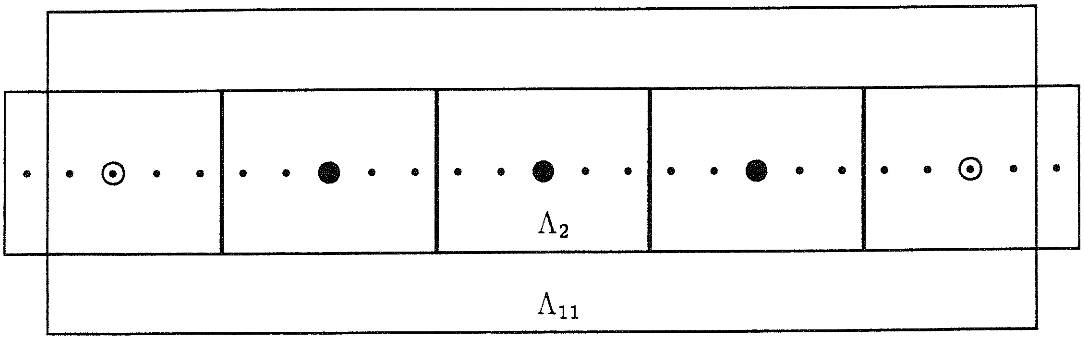
$$\Lambda_{n|m} := \{(2m + 1)j : j \in \Lambda_{(n-m)\operatorname{div}(2m+1)}\}, \tag{8.47}$$

$$q(n|m) := |\Lambda_{n|m}| = \Big(2\{(n - m)\operatorname{div}(2m + 1)\} + 1\Big)^d \,, \tag{8.48}$$

where div denotes the integer part of the quotient. Then

$$\bigcup_{j\in\Lambda_{n|m}} \theta_j\Lambda_m \subset \Lambda_n \quad \text{and} \quad |\bigcup_{j\in\Lambda_{n|m}} \theta_j\Lambda_m| = q(n|m)V_m\,, \tag{8.49}$$

since the translates $\theta_j\Lambda_m$ are disjoint.



$d=1$: boxes Λ_2 and Λ_{11}; large • are the points of $\Lambda_{11|2}$; $q(11|2)=3$.

Enumerate the points of $\Lambda_{n|m}$ by i_j, $j=1,\ldots,q(n|m)$, so that 1 corresponds to the origin in $\Lambda_{n|m}$. Because $q(n|m)$ is odd, we omit the origin from sums involving $\Lambda_{n|m}$. Making use of the convex relationship $a_2=(a_0+a_1)/2$, we have for $n>m$

$$\|(T_n(\omega)-a_2)V_n - \sum_{j=2}^{q(n|m)} (\theta_{i_j}T_m(\omega)-a_{j\,\mathrm{mod}\,2})V_m\| \le \tag{8.50}$$

$$\big(V_n-(q(n|m)-1)V_m\big)(\|a_2\|+\|\varphi\|)\,.$$

For any $m\in\mathbb{N}$, we can find $G_m\in\mathcal{F}_{\Lambda_m}$ so that

$$\lim_m \frac{1}{V_m}\sup_\omega |G_m(\omega)-\sum_{j\in\Lambda_m}\varphi(\theta_j\omega)|=0. \tag{8.51}$$

Therefore there exists M so that $m\ge M$ implies

$$\frac{1}{V_m}\sup_\omega |G_m(\omega)-\sum_{j\in\Lambda_m}\varphi(\theta_j\omega)|\le \min\{\varepsilon'-\varepsilon'',\varepsilon-\varepsilon'\}. \tag{8.52}$$

For $m\ge M$, it follows from the two estimates above that

$$\limsup_n \|\{T_n(\omega)-a_2\} - \frac{V_m}{V_n}\sum_{j=2}^{q(n|m)}\{\theta_{i_j}\frac{1}{V_m}G_m(\omega)-a_{j\,\mathrm{mod}\,2}\}\|\le\varepsilon-\varepsilon', \tag{8.53}$$

and

$$T_m^{-1}[B_{\varepsilon''}(a_{j\,\mathrm{mod}\,2})]\subset\{\omega:\frac{1}{V_m}G_m(\omega)\in B_{\varepsilon'}(a_{j\,\mathrm{mod}\,2})\}. \tag{8.54}$$

Then there exists N_m such that $n\ge N_m$ implies

$$\{\omega:\theta_{i_j}\frac{1}{V_m}G_m(\omega)\in B_{\varepsilon'}(a_{j\,\mathrm{mod}\,2}),\ j=2,\ldots,q(n|m)\}\subset T_n^{-1}[B_\varepsilon(a_2)]\,. \tag{8.55}$$

Using the fact that β is weakly dependent, we can write

$$\frac{\ln \mathbb{M}_n[B_\epsilon(a_2)]}{V_n} \geq \frac{\ln \mathbb{M}_m[B_{\epsilon''}(a_0)] + \ln \mathbb{M}_m[B_{\epsilon''}(a_1)]}{2V_m} \frac{V_m\,(q(n|m)-1)}{V_n} - \frac{q(n|m)}{V_n} c_m(\beta). \tag{8.56}$$

To deduce (8.43) from (8.56), select a sequence $\{m_k\}$ so that $\ln \mathbb{M}_{m_k}[B_{\epsilon''}(a_0)]/V_{m_k}$ converges to $\overline{m}[B_{\epsilon''}(a_0)]$. Using this sequence and the fact that $(q(n|m_k)-1)V_{m_k}/V_n$ goes to 1 as $n \to \infty$, (8.43) follows. □

Theorem 8.4 *Let $\beta \in \mathcal{M}_1^{+,\theta}$ be weakly dependent. Let φ be a quasilocal $\mathbb{R}^k$-valued random variable on the probability space $(\Omega, \mathcal{F}, \beta)$. The distribution of $T_n = \mathcal{A}_n\varphi$ on $\mathbb{R}^k$ is $\mathbb{M}_n$, and μ is the RL-function of the pair $(\{\mathbb{M}_n\}, \{V_n\})$. Let $\{C_n\}$ be an LD-regular sequence in $\mathbb{R}^k$ with respect to μ. Assume there exists $x' \in \mathbb{R}^k$ such that $N_C \subset N^{x'}$. Then the following hold:*

a) The set of limit points of the sequence

$$\{\mathcal{A}_n\beta[\ \cdot\ |T_n \in C_n]\} \tag{8.57}$$

is non-empty.

b) Any limit point of the sequence (8.57) is an $(f_\varphi^{x'}, \beta)$-equilibrium state, with $f_\varphi^{x'} = \langle x', \varphi \rangle$.

c) x' is a subgradient of $-\mu$ for any point of the non-empty compact set $N_C = \{x \in \mathbb{R}^k : \mu(x) = \sup_{y \in C} \mu(y)\}$.

In particular, if B is a convex set containing an interior point where μ is finite, then the sequence $\{C_n \equiv B\}$ is LD-regular and there exists x' so that $N_C \subset N^{x'}$ with $C := \mathrm{cl}B$.

Proof: The pair $(\{\mathbb{M}_n\}, \{V_n\})$ has an RL-function μ and a scaled generating function p, with $\mathrm{dom}\, p = \mathbb{R}^k$, and

$$p(x') = (-\mu)^*(x'). \tag{8.58}$$

Since X is compact, we have an LDP; we apply Theorem 6.1. The sequence of measures $\{\mathbb{M}_n^C\}$ is eventually concentrated on the non-empty compact set N_C; there exists $x' \in \cap_{x \in N_C} \partial(-\mu)(x) \subset \mathbb{R}^k$, such that

$$\lim_n \frac{1}{V_n} \mathcal{H}(\mathbb{M}_n^C | \mathbb{M}_n^{x'}) = 0. \tag{8.59}$$

The theorem is now a consequence of Theorem 8.2; Lemma 6.1 covers the particular case. □

8.4 Conditional Limit Theorem: Empirical Measure

We consider here a case where the random variable $\varphi : \Omega \to X$ takes its values in the space of probability measures $\mathcal{M}_1^+$. Putting

$$\varphi(\omega) := \delta_\omega \ , \ \delta_\omega \text{ Dirac mass at } \omega\ , \tag{8.60}$$

the random variable T_n,

$$T_n(\omega) := \mathcal{A}_n\varphi(\omega), \tag{8.61}$$

is called the **empirical measure**. The main result of this section is a simple proof of the LDP for empirical measures.

There is a natural embedding of $\mathcal{M}_1^+$ in the unit ball of the dual $\mathcal{F}_{\mathrm{qloc}}^*$ of the Banach space of quasilocal functions $\mathcal{F}_{\mathrm{qloc}}$ equipped with the norm $\|f\| = \sup_\omega |f(\omega)|$. We take advantage of the fact that the norm unit ball of $\mathcal{F}_{\mathrm{qloc}}^*$,

$$\{x \in \mathcal{F}_{\mathrm{qloc}}^* : |\langle f, x \rangle| \leq \|f\| \ \text{ for all } f \in \mathcal{F}_{\mathrm{qloc}}\}, \tag{8.62}$$

is compact in the $\sigma(\mathcal{F}^*_{\mathrm{qloc}}, \mathcal{F}_{\mathrm{qloc}})$–topology. We always choose this topology for $\mathcal{F}^*_{\mathrm{qloc}}$. In general, this space contains $\mathcal{M}$ as a proper subspace and the subspace topology on $\mathcal{M}$ is the topology introduced in Section 8.1, so that we have an embedding of $\mathcal{M}$ into $\mathcal{F}^*_{\mathrm{qloc}}$. On $\mathcal{F}_{\mathrm{qloc}}$, we consider the $\sigma(\mathcal{F}_{\mathrm{qloc}}, \mathcal{F}^*_{\mathrm{qloc}})$–topology which differs in general from the $\sigma(\mathcal{F}_{\mathrm{qloc}}, \mathcal{M})$–topology. With this topology, the spaces $\mathcal{F}_{\mathrm{qloc}}$ and $\mathcal{F}^*_{\mathrm{qloc}}$ are in duality. With the notations of Section 4, we have $E^* = \mathcal{F}_{\mathrm{qloc}}$, $E = \mathcal{F}^*_{\mathrm{qloc}}$ and we choose X as the norm unit ball (8.62). The σ–algebra $\mathcal{B}$ on X is generated by all maps

$$x \mapsto \langle f, x\rangle \quad , \quad f \in \mathcal{F}_{\mathrm{qloc}}, \tag{8.63}$$

where $\langle f, x\rangle$ is the pairing between $\mathcal{F}_{\mathrm{qloc}}$ and $\mathcal{F}^*_{\mathrm{qloc}}$. We fix a translation invariant weakly dependent probability measure β on $(\Omega, \mathcal{F})$ and consider the empirical measure T_n on the probability space $(\Omega, \mathcal{F}, \beta)$ with values in the compact space X. The distribution of T_n on X is $\mathbb{M}_n = \beta \circ T_n^{-1}$.

The proof of existence and concavity of the RL–function for $(\{\mathbb{M}_n\}, \{V_n\})$ is essentially the same as the proof of Lemma 8.2 since a base of neighbourhoods of $x^* \in X$ is of the form

$$B(x^*, \mathbf{f}, \varepsilon) := \{x \in X : |\langle f_j, x\rangle - \langle f_j, x^*\rangle| < \varepsilon_j, j = 1, ..., k\}, \tag{8.64}$$

where $\mathbf{f} = (f_1, \ldots, f_k)$ with each $f_j \in \mathcal{F}_{\mathrm{loc}}$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_k)$ with each $\varepsilon_j > 0$ and k an arbitrary positive integer. (It is sufficient to consider only $f_j \in \mathcal{F}_{\mathrm{loc}}$ because X is a norm bounded subset of $\mathcal{F}^*_{\mathrm{qloc}}$.) The RL–function μ for $(\{\mathbb{M}_n\}, \{V_n\})$ can at each $x \in X$ be approximated arbitrarily closely by an RL–function $\mu_{\mathbf{f}}$ for some finite vector $\mathbf{f} = (f_1, \ldots, f_k)$ with each $f_j \in \mathcal{F}_{\mathrm{loc}}$.

Lemma 8.3 *The RL–function μ associated to the empirical measure, regarded as a random variable with values in X, exists and is concave.*

Since X is compact we have an LDP; we compute the scaled generating function by Varadhan's Theorem:

$$\begin{aligned} p(x'|\beta) &= \lim_n \frac{1}{V_n} \ln \int_\Omega \exp\{V_n\langle x', T_n(\omega)\rangle\}\beta[d\omega] \\ &= \lim_n \frac{1}{V_n} \ln \int_X e^{V_n\langle x', x\rangle}\, \mathbb{M}_n[dx] \\ &= \sup_{x\in X}\{\langle x', x\rangle + \mu(x)\}, \end{aligned} \tag{8.65}$$

for any $x' \in \mathcal{F}_{\mathrm{qloc}}$. This identity shows that $p(\cdot|\beta)$ is the conjugate function of $-\mu$; since $-\mu$ is closed convex and proper, the converse is also true: $-\mu$ is the conjugate of $p(\,\cdot\,|\beta)$.

Remark: It is possible to treat on the same footing the following variants of the empirical measure. Define

$$T_n^c(\omega) = \mathcal{A}_n \mathcal{P}_n \varphi(\omega); \tag{8.66}$$

T_n^c is called the **periodic empirical measure**. More generally, define

$$T_n^\bullet(\omega) := \mathcal{A}_n \mathbb{Q}_n(\omega), \tag{8.67}$$

where $\mathbb{Q}_n(\omega)$ is any probability measure on Ω which is a measurable function of ω and whose marginal distribution on $\mathcal{F}_{\Lambda_n}$ coincides with δ_ω. The empirical measure (8.61) and the periodic empirical measure (8.66) are special cases of (8.67). In all cases, the RL–functions and scaled generating functions are the same, so that we have the same large deviations results on the scale $\{V_n\}$. Indeed, for any $f \in \mathcal{F}_{\text{loc}}$ and any $i \in \mathbb{Z}^d$, we have

$$\langle \theta_i f, \delta_\omega \rangle = \langle \theta_i f, \mathbb{Q}_n(\omega) \rangle, \tag{8.68}$$

if n is large enough; hence

$$\lim_n \| \langle f, T_n^\bullet(\ \cdot \ ) \rangle - \langle f, T_n(\ \cdot \ ) \rangle \| = 0. \tag{8.69}$$

Lemma 8.4 *Let $x \in X$ satisfy $\mu(x) > -\infty$. Then x is translation invariant and $\langle f, x \rangle \geq 0$ for all nonnegative $f \in \mathcal{F}_{\text{loc}}$; moreover, we have $\langle x, 1 \rangle = 1$.*

Proof: We show translation invariance; the other parts are similar. Assume x is not translation invariant. Then there exist $i \in \mathbb{Z}^d$, $\varepsilon > 0$ and $f \in \mathcal{F}_{\text{loc}}$ so that with $g := f - \theta_i f$, we have $\langle x, g \rangle > \varepsilon$. Notice that

$$\lim_n \sup_\omega |\mathcal{A}_n g| = 0. \tag{8.70}$$

Therefore, for all ω and all n sufficiently large, we have

$$|\langle g, T_n(\omega) \rangle| = |\mathcal{A}_n g(\omega)| < \varepsilon; \tag{8.71}$$

this implies that the neighbourhood $\{u \in X : \langle u, g \rangle > \varepsilon\}$ of x has $\overline{m}_\beta = -\infty$, hence $\mu(x) = -\infty$. □

Lemma 8.5 *For $f_1, f_2 \in \mathcal{F}_{\text{qloc}}$, we have $|p(f_1|\beta) - p(f_2|\beta)| \leq \|f_1 - f_2\|$. Also, for any $i \in \mathbb{Z}^d$, we have $p(f_1 + f_2 - \theta_i f_2|\beta) = p(f_1|\beta)$.*

Proof: The inequality follows from

$$p(f_j|\beta) = \sup_{x \in X} \langle f_j, x \rangle + \mu(x) \tag{8.72}$$

and $|\langle x, f_1 - f_2 \rangle| \leq \|f_1 - f_2\|$ for $x \in X$. The second part follows because $\mu(x) = -\infty$ unless x is translation invariant. □

Lemma 8.6 *If $\nu \in X$ is a translation invariant probability measure, then the conjugate function of $p(\ \cdot \ |\beta)$ at ν, $p_\beta^*(\nu)$, satisfies*

$$-\mu(\nu) = p_\beta^*(\nu) \leq h(\nu|\beta). \tag{8.73}$$

Proof: By definition, we have

$$p_\beta^*(\nu) = \sup_{f \in \mathcal{F}_{\text{qloc}}} \left(\langle f, \nu \rangle - \lim_{n \to \infty} \frac{1}{V_n} \ln \int \exp(\sum_{i \in \Lambda_n} \theta_i f(\omega)) \beta[d\omega] \right). \tag{8.74}$$

We have shown already that $p(\,\cdot\,|\beta)$ and $-\mu$ are conjugate functions. From Lemma 8.5 and the continuity of $\langle \nu, f\rangle$, it follows that the supremum over $\mathcal{F}_{\rm loc}$ yields the same value. Take any $f \in \mathcal{F}_{\Lambda_m}$. Then, by Proposition 10.2, we have

$$\langle f,\nu\rangle - \frac{1}{V_n}\ln\int \exp(\sum_{i\in\Lambda_n}\theta_i f(\omega))\beta[d\omega] = \tag{8.75}$$

$$\frac{1}{V_n}\left(\langle\sum_{i\in\Lambda_n}\theta_i f,\nu\rangle - \ln\int\exp(\sum_{i\in\Lambda_n}\theta_i f(\omega))\beta[d\omega]\right) \le \frac{1}{V_n}\mathcal{H}_{\mathcal{F}_{\Lambda_{n+m}}}(\nu|\beta)\,,$$

because $\sum_{i\in\Lambda_n}\theta_i f \in \mathcal{F}_{\Lambda_{n+m}}$. Now

$$\liminf_{n\to\infty}\frac{1}{V_n}\mathcal{H}_{\mathcal{F}_{\Lambda_{n+m}}}(\nu|\beta) = h(\nu|\beta), \tag{8.76}$$

so $f \in \mathcal{F}_{\rm loc}$ implies

$$\langle f,\nu\rangle - \lim_{n\to\infty}\frac{1}{V_n}\ln\int\exp(\sum_{i\in\Lambda_n}\theta_i f(\omega))\beta[d\omega] \le h(\nu|\beta). \tag{8.77}$$

Then (8.73) follows. □

For any $x^* \in X$ with $\mu(x^*) > -\infty$, each measurable neighbourhood C of x^* and each positive integer n with $\beta[T_n \in C] > 0$, we consider the probability measure

$$\beta_{n,C}[\,\cdot\,] := \beta[\,\cdot\,|T_n\in C], \quad \beta^{\mathcal{A}}_{n,C} := \mathcal{A}_n\mathcal{P}_n\beta_{n,C}\,. \tag{8.78}$$

Lemma 8.7 *Let $x^* \in X$ with $\mu(x^*) > -\infty$, where β is a translation invariant weakly dependent probability measure. Let C be a measurable neighbourhood of x^* in X. Then, for n sufficiently large, we have*

$$h(\beta^{\mathcal{A}}_{n,C}|\beta) \le \frac{c_n(\beta)}{V_n} - \ln\beta[T_n\in C]/V_n. \tag{8.79}$$

If, in addition, C is closed and convex, then any limit point y of the sequence $\{\beta^{\mathcal{A}}_{n,C}\}$ satisfies $y \in C$.

Proof: From Lemma 8.1, we have

$$\begin{aligned} V_n\, h(\beta^{\mathcal{A}}_{n,C}|\beta) - c_n(\beta) &\le \mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\beta[\,\cdot\,|T_n\in C]|\beta) \\ &\le \mathcal{H}(\beta[\,\cdot\,|T_n\in C]|\beta) \\ &= -\ln\beta[\omega : T_n \in C]. \end{aligned} \tag{8.80}$$

Assume, in addition, that C is closed and convex. Let y be an element of $X\setminus C$; then there exists $f\in\mathcal{F}_{\rm qloc}$ so that

$$\langle f,x\rangle \le 0 \text{ for all } x\in C, \quad \langle f,y\rangle = 1. \tag{8.81}$$

Given $f\in\mathcal{F}_{\rm qloc}$ satisfying the above, one can find $f^*\in\mathcal{F}_{\rm loc}$ satisfying the same condition, so we may assume without loss of generality that we have $f\in\mathcal{F}_{\rm loc}$ satisfying (8.81), say $f\in\mathcal{F}_{\Lambda_{m^*}}$. An elementary calculation shows that

$$\|\mathcal{A}_n f - \mathcal{A}_{n-m^*}f\| \le 2\|f\|\frac{V_n - V_{n-m^*}}{V_n}. \tag{8.82}$$

We have defined the action of $\mathcal{A}_n$ so that

$$\langle f, T_n(\omega)\rangle = \langle \mathcal{A}_n f, \delta_\omega\rangle, \quad \langle f, \beta^{\mathcal{A}}_{n,C}\rangle = \langle \mathcal{A}_n f, \mathcal{P}_n \beta_{n,C}\rangle. \tag{8.83}$$

Because $\langle \mathcal{A}_{n-m^*} f, \mathcal{P}_n\beta_{n,C}\rangle = \langle \mathcal{A}_{n-m^*} f, \beta_{n,C}\rangle$, we have

$$|\langle f, \beta^{\mathcal{A}}_{n,C}\rangle - \langle \mathcal{A}_{n-m^*} f, \beta_{n,C}\rangle| \leq 2\|f\|\frac{V_n - V_{n-m^*}}{V_n}. \tag{8.84}$$

Now $\beta_{n,C}$ is supported by those $\omega \in \Omega$ which satisfy $\langle \mathcal{A}_n f, \delta_\omega\rangle \leq 0$; hence we have

$$\langle \mathcal{A}_n f, \beta_{n,C}\rangle \leq 0. \tag{8.85}$$

From the four above displayed formulas we deduce that

$$\langle f, \beta^{\mathcal{A}}_{n,C}\rangle \leq 4\|f\|\frac{V_n - V_{n-m^*}}{V_n}. \tag{8.86}$$

Since $(V_n - V_{n-m^*})/V_n \to 0$ as $n \to \infty$, if x is a limit point of $\{\beta^{\mathcal{A}}_{n,C}\}$, then $\langle f, x\rangle \leq 0$; that is, $y \in X \setminus C$ implies y is not a limit point of $\{\beta^{\mathcal{A}}_{n,C}\}$. □

For the probability measure $\nu \in X$ from (8.73), we always have $-\mu(\nu) \leq h(\nu|\beta)$. A periodic probability measure ν which is not translation invariant has $\mu(\nu) = -\infty$, but one may have $h(\nu|\beta) < \infty$. In the case of translation invariance, h coincides with $-\mu$.

Lemma 8.8 *If $\mu(x^*) > -\infty$, then x^* is a translation invariant probability measure. If $\nu \in X$ is a translation invariant probability measure, then*

$$\mu(\nu) = -h(\nu|\beta). \tag{8.87}$$

Proof: Let x^* be an element of X such that $\mu(x^*) > -\infty$. For any closed convex neighbourhood $C \ni x^*$, we have

$$\mu(x^*) \leq \underline{m}[C]; \tag{8.88}$$

for $\varepsilon > 0$ and all n sufficiently large, we have

$$-\infty < \mu(x^*) \leq \frac{1}{V_n}\ln\beta[T_n \in C] + \varepsilon. \tag{8.89}$$

For those n, Lemma 8.7 implies that, for $\beta^{\mathcal{A}}_{n,C}$ defined in (8.78), we have

$$h(\beta^{\mathcal{A}}_{n,C}|\beta) \leq \frac{c_n(\beta)}{V_n} - \mu(x^*) + \varepsilon < \infty; \tag{8.90}$$

since the level–sets of $h(\;\cdot\;|\beta)$ are compact on the space $\mathcal{M}_1^{+,\theta}$, for each closed convex neighbourhood C of x^* the sequence $\{\beta^{\mathcal{A}}_{n,C}\}$ has at least one limit point which is a probability measure which we denote by β_C; Lemma 8.7 shows that β_C belongs to C. The lower semicontinuity of the specific information gain on $\mathcal{M}_1^{+,\theta}$ implies that

$$h(\beta_C|\beta) \leq -\mu(x^*) + \varepsilon < \infty. \tag{8.91}$$

The net of probability measures $\{\beta_C\}$ parameterized by the closed convex neighbourhoods of x^* ordered by inclusion has limits points which are probability measures because of (8.91). By definition, this net converges to x^*. Since the topology is Hausdorff, the limit is unique and is thus a probability measure. If $\nu \in \mathcal{M}_1^{+,\theta}$ with $\mu(\nu) = -\infty$, then (8.73) implies (8.87); otherwise, the reverse inequality follows from (8.79) and the lower semicontinuity of the specific information gain. □

We consider again the setting of Section 8.1 with the empirical measure T_n regarded as a random variable defined on the probability space $(\Omega, \mathcal{F}, \beta)$, where $\beta \in \mathcal{M}_1^{+,\theta}$ is weakly dependent. We have the following result:

Theorem 8.5 *Let $\beta \in \mathcal{M}_1^{+,\theta}$ be a weakly dependent probability measure. Then the empirical measure T_n defined on the probability space $(\Omega, \mathcal{F}, \beta)$ with values in the space of probability measures $\mathcal{M}_1^+$ satisfies an LDP with RL–function μ given by*

$$\mu(\nu) = \begin{cases} -h(\nu|\beta) & \nu \in \mathcal{M}_1^{+,\theta}, \\ -\infty & \nu \in \mathcal{M}_1^+ \backslash \mathcal{M}_1^{+,\theta}. \end{cases} \tag{8.92}$$

Proof: We know that we have an LDP on the space X. Since the empirical measure T_n takes its values in $\mathcal{M}_1^+$, the RL–function at $\nu \in \mathcal{M}_1^+$ is equal to the above RL–function; it remains to show the upper bound for closed sets. Let B be a measurable set in $\mathcal{M}_1^+$; denote by $\overline{B}$ its closure in $\mathcal{M}_1^+$ and by $\overline{B}_X$ its closure in X. We have $\overline{B} = \overline{B}_X \cap \mathcal{M}_1^+$; thus from Lemma 8.8 and the LDP on X we have

$$\overline{m}[B] \leq \sup_{x \in \overline{B}_X} \mu(x) = \sup_{x \in \overline{B}} \mu(x). \tag{8.93}$$

□

The proof of Theorem 8.4 and Corollary 8.1 yield the following result.

Theorem 8.6 *Let $\beta \in \mathcal{M}_1^{+,\theta}$ be weakly dependent. Let $\{T_n\}$ be the empirical measure (8.61) or one of the the variants of the empirical measure given by (8.66) and (8.67). Let $\mathbb{M}_n = \beta \circ T_n^{-1}$ and let μ be the RL–function of the pair $(\{\mathbb{M}_n\}, \{V_n\})$. Let $\{C_n\}$ be an LD–regular sequence in X such that there exists $x' \in \mathcal{F}_{\mathrm{qloc}}$ so that $N_C \subset N^{x'}$, where $C := \cap \mathrm{cl}\, C_n$. Then the following hold:*

a) The set of limit points of the sequence

$$\{\mathcal{A}_n \beta[\ \cdot\ | T_n \in C_n]\} \tag{8.94}$$

is non–empty.

b) x' is a subgradient of $-\mu$ for any point of the non–empty compact set $N_C = \{x \in \mathcal{M}_1^+ : \mu(x) = \sup_{y \in C} \mu(y)\}$.

In particular, if B is a convex set containing an interior point where μ is finite, then the sequence $\{C_n \equiv B\}$ is LD–regular and there exists x' so that $N_C \subset N^{x'}$ with $C := \mathrm{cl} B$.

9 Equivalence of Ensembles for Lattice Systems

As an application of Theorem 8.4, we discuss the question of equivalence of ensembles in the case of lattice systems, giving proofs of the results announced in [LPS1] and [LPS2].

The setting is that of Section 8.1. An **interaction** $\Phi = \{\Phi_\Lambda\}$ is a family of $\mathcal{F}_\Lambda$–measurable functions Φ_Λ, indexed by the finite nonempty subsets Λ of $\mathbb{Z}^d$, such that

$$\theta_i \Phi_\Lambda = \Phi_{\theta_i \Lambda}\ ,\ \forall \Lambda,\ \forall i \in \mathbb{Z}^d. \tag{9.1}$$

A potential is **absolutely summable** if

$$\|\Phi\| := \sum_{\Lambda \ni 0} \|\Phi_\Lambda\| < \infty\,, \tag{9.2}$$

where $\|\Phi_\Lambda\|$ is the supremum of $|\Phi_\Lambda(\omega)|$ over Ω. Another norm used with lattice models is given by

$$\|\Phi\|_{\#} := \sum_{\Lambda \ni 0} \|\Phi_\Lambda\|/|\Lambda|\,. \tag{9.3}$$

For each interaction Φ, we define a quasilocal function f_Φ by

$$f_\Phi := \sum_{\Lambda \ni 0} \frac{1}{|\Lambda|}\Phi_\Lambda. \tag{9.4}$$

It is convenient to allow the interaction to be $\mathbb{R}^k$–valued. For a given $\mathbb{R}^k$–valued interaction Φ, we define

$$U_\Lambda(\omega) := \sum_{A:A\subset\Lambda} \Phi_A(\omega)\,, \qquad W_\Lambda(\omega) := \sum_{A:\, \substack{A\cap\Lambda\neq\emptyset \\ A\setminus\Lambda\neq\emptyset}} \Phi_A(\omega)\,, \tag{9.5}$$

and the families of mappings $\{S_n\}$, $\{S_n^\eta\}$, and $\{S_n^c\}$ of Ω into $\mathbb{R}^k$ by

$$V_n S_n(\omega) = (U_{1,n}(\omega), \ldots, U_{k,n}(\omega))\,; \tag{9.6}$$

$$V_n S_n^\eta(\omega) = (U_{1,n}(\omega) + W_{1,n}(\omega_{\Lambda_n}\eta_{\overline{\Lambda}_n}), \ldots, U_{k,n}(\omega) + W_{k,n}(\omega_{\Lambda_n}\eta_{\overline{\Lambda}_n}))\,. \tag{9.7}$$

The first index of $U_{i,n}$ refers to the coordinate of $\mathbb{R}^k$ and the notation $\omega_{\Lambda_n}\eta_{\overline{\Lambda}_n}$ means the point of Ω whose Λ_n coordinates are taken from ω while the $\overline{\Lambda}_n \equiv \mathbb{Z}^d \setminus \Lambda_n$ coordinates are taken from η; $\{S_n^c\}$ is $\{S_n^\eta\}$ with η arising from $\omega \in \Lambda_n$ by periodic continuation to $\mathbb{Z}^d$. The sequence $\{S_n\}$ corresponds to the energy with *free* boundary conditions; $\{S_n^\eta\}$, to *fixed* boundary condition η; $\{S_n^c\}$, to *cyclic* boundary conditions.

The next estimates allow us to make the connection with the results of Section 8.3. We define the A–boundary of the set Λ_n as the subset of $\mathbb{Z}^d$

$$\partial_A \Lambda_n := \{j \in \Lambda_n : A + j \not\subset \Lambda_n\}. \tag{9.8}$$

The following properties are obvious from the definition:

$$\frac{|\partial_A \Lambda_n|}{V_n} \leq 1, A \subset \mathbb{Z}^d, \tag{9.9}$$

$$\lim_n \frac{|\partial_A \Lambda_n|}{V_n} = 0, A \subset \mathbb{Z}^d. \tag{9.10}$$

We have

$$\begin{aligned} \|\sum_{j\in\Lambda_n} \theta_j f_\Phi - \sum_{A\subset\Lambda_n} \Phi_A\| &= \|\sum_{j\in\Lambda_n}\sum_{A\ni j} \frac{1}{|A|}\Phi_A - \sum_{j\in\Lambda_n}\sum_{\substack{A\ni j\\ A\subset\Lambda_n}} \frac{1}{|A|}\Phi_A\| \\ &\le \sum_{j\in\Lambda_n}\sum_{\substack{A\ni j\\ A\not\subset\Lambda_n}} \frac{1}{|A|}\|\Phi_A\| \\ &\le \sum_{A\ni 0} |\partial_A\Lambda_n| \frac{1}{|A|}\|\Phi_A\|. \end{aligned} \tag{9.11}$$

Since

$$\frac{|\partial_A\Lambda_n|}{V_n}\frac{1}{|A|}\|\Phi_A\| \le \frac{1}{|A|}\|\Phi_A\|, \tag{9.12}$$

and

$$\sum_{A\ni 0} \frac{1}{|A|}\|\Phi_A\| = \|\Phi\|_{\#} < \infty, \tag{9.13}$$

it follows from the Bounded Convergence Theorem that

$$\begin{aligned} \lim_n \|\sum_{j\in\Lambda_n} \frac{1}{V_n}\theta_j f_\Phi - \sum_{A\subset\Lambda_n} \frac{1}{V_n}\Phi_A\| &\le \lim_n \sum_{A\ni 0} \frac{|\partial_A\Lambda_n|}{V_n}\frac{1}{|A|}\|\Phi_A\| \\ &= \sum_{A\ni 0} \lim_n \frac{|\partial_A\Lambda_n|}{V_n}\frac{1}{|A|}\|\Phi_A\| = 0. \end{aligned} \tag{9.14}$$

A similar estimate can be derived for $\{S_n^\eta\}$ or $\{S_n^c\}$ instead of $\{S_n\}$ when the interaction is absolutely summable because

$$\begin{aligned} \sum_{A:\substack{A\cap\Lambda_n\neq\emptyset\\ A\setminus\Lambda_n\neq\emptyset}} \|\Phi_A\| &\le \sum_{j\in\Lambda_n}\sum_{\substack{A\ni j\\ A\not\subset\Lambda_n}} \|\Phi_A\| \\ &\le \sum_{A\ni 0} |\partial_A\Lambda_n|\, \|\Phi_A\|. \end{aligned} \tag{9.15}$$

As above, we have

$$\lim_n \sum_{A\ni 0} |\partial_A\Lambda_n|\, \|\Phi_A\|/V_n = 0. \tag{9.16}$$

The next Lemma follows directly from these estimates.

Lemma 9.1 *Let $\beta \in \mathcal{M}_1^{+,\theta}$ be weakly dependent and Φ be an $\mathbb{R}^k$-valued interaction.*

a) If $\|\Phi\|_{\#} < \infty$, then the RL-function of $(\{\beta\circ S_n^{-1}\}, \{V_n\})$ exists and is equal to the RL-function μ of $(\{\beta\circ T_n^{-1}\}, \{V_n\})$ with $T_n = \mathcal{A}_n f_\Phi$.

b) If $\|\Phi\| < \infty$, then the RL-functions of $(\{\beta\circ S_n^{\eta\,-1}\}, \{V_n\})$ and $(\{\beta\circ S_n^{c\,-1}\}, \{V_n\})$ exist and are equal to the RL-function μ of $(\{\beta\circ T_n^{-1}\}, \{V_n\})$ with $T_n = \mathcal{A}_n f_\Phi$.

The above shows that the RL-function determined by $\mathcal{F}_{\text{qloc}}$ yields the RL-functions for potentials. For the $\mathbb{R}^k$ valued $f \in \mathcal{F}_{\Lambda_{m^*}}$, one defines the absolutely convergent potential Φ^f by

$$\Phi_\Lambda^f = \begin{cases} \theta_i f & \text{if } \theta_i\Lambda = \Lambda_{m^*}, \\ 0 & \text{otherwise,} \end{cases} \tag{9.17}$$

Using (9.4) to define f_{Φ^f}, we compute

$$f_{\Phi^f} = \mathcal{A}_{m^*} f; \tag{9.18}$$

thus the RL–fuction determined by f_{Φ^f} coincides with that determined by f and $p(f_{\Phi^f}|\beta) = p(f|\beta)$ for all weakly dependent $\beta \in \mathcal{M}_1^{+,\theta}$. More generally, for $f \in \mathcal{F}_{\text{qloc}}$ there exists a sequence $\{f_j\}$ in $\mathcal{F}_{\text{loc}}$ so that $f = \sum f_j$ and $\sum \|f_j\| < \infty$. Define $\Phi^f = \sum \Phi_{f_j}$, where Φ_{f_j} is given by (9.17) with m^* the smallest positive integer for which $f_j \in \mathcal{F}_{\Lambda_{m^*}}$. Then $\|\Phi^f\|_\# < \infty$ and f_{Φ^f} determines the same RL–function as that determined by f.

Henceforth in this Section, $\mathbb{P}$ denotes a given translation invariant probability measure which is the product of its marginals $\mathbb{P}_{\{j\}}$, $j \in \mathbb{Z}^d$. Let Ψ be a real-valued absolutely summable interaction. For finite $\Lambda \subset \mathbb{Z}^d$, define

$$r_\Lambda(\omega) := e^{U_\Lambda(\omega)+W_\Lambda(\omega)} \tag{9.19}$$

with U_Λ and W_Λ given by (9.5) with Ψ instead of Φ. For each finite subset Λ of $\mathbb{Z}^d$, we introduce a probability kernel $\gamma_\Lambda : \mathcal{F} \times \Omega \to [0,\infty]$ given by

$$\gamma_\Lambda(F,\omega) := \frac{\int_\Omega 1_F(\eta_\Lambda \omega_{\overline{\Lambda}}) r_\Lambda(\eta_\Lambda \omega_{\overline{\Lambda}}) \mathbb{P}_{\mathcal{F}_\Lambda}(d\eta)}{\int_\Omega r_\Lambda(\eta_\Lambda \omega_{\overline{\Lambda}}) \mathbb{P}_{\mathcal{F}_\Lambda}(d\eta)}, \tag{9.20}$$

where 1_F is the indicator function of $F \in \mathcal{F}$.

Definition 9.1 *Let Ψ be a real-valued absolutely summable interaction. Let $\mathbb{P} \in \mathcal{M}_1^{+,\theta}$ be the product of its marginals $\mathbb{P}_{\{j\}}$, $j \in \mathbb{Z}^d$. A probability measure λ on $(\Omega, \mathcal{F})$ is a $(\Psi,\mathbb{P})$-***Gibbs state** *if and only if*

$$\mathbb{E}_\lambda(1_F|\mathcal{F}_{\mathbb{Z}^d\setminus\Lambda})(\omega) = \gamma_\Lambda(F,\omega)\ \lambda - a.s. \tag{9.21}$$

for each finite subset Λ of $\mathbb{Z}^d$ and each $F \in \mathcal{F}$.

It is known that a translation invariant probability measure λ is a $(f_\Psi,\mathbb{P})$–equilibrium state if and only if λ is a $(\Psi,\mathbb{P})$–Gibbs state. For a proof, see [Ru2] or [G2]. Notice that a Gibbs state is not necessarily translation invariant.

Lemma 9.2 *Let Ψ be a real-valued, absolutely summable interaction. Any $(\Psi,\mathbb{P})$-Gibbs state is a weakly dependent probability measure with*

$$c_n(\lambda) \le 4 \sum_{A \ni 0} |\partial_A \Lambda_n|\, \|\Psi_A\|. \tag{9.22}$$

Proof: Define

$$w(n) := \sum_{A \ni 0} |\partial_A \Lambda_n|\, \|\Psi_A\|. \tag{9.23}$$

We have

$$e^{-2w(n)} r_{\Lambda_n}(\omega_{\Lambda_n}\eta_{\overline{\Lambda}_n}) \le r_{\Lambda_n}(\omega_{\Lambda_n}\omega_{\overline{\Lambda}_n}) \le e^{2w(n)} r_{\Lambda_n}(\omega_{\Lambda_n}\eta_{\overline{\Lambda}_n}). \tag{9.24}$$

If $f \in \mathcal{F}_{\text{loc}}$ is positive, then

$$\begin{aligned} \int f(\omega)\lambda[d\omega] &= \int \gamma_{\Lambda_n}(f,\omega)\lambda_{|\mathcal{F}_{\overline{\Lambda}_n}}[d\omega] \qquad (9.25) \\ &\le e^{4w(n)} \int \left(\frac{\int f(\omega'_{\Lambda_n}\omega_{\overline{\Lambda}_n}) r_{\Lambda_n}(\omega'_{\Lambda_n}\eta_{\overline{\Lambda}_n}) \mathbb{P}_{|\mathcal{F}_{\Lambda_n}}[d\omega']}{\int r_{\Lambda_n}(\omega'_{\Lambda_n}\eta_{\overline{\Lambda}_n}) \mathbb{P}_{|\mathcal{F}_{\Lambda_n}}[d\omega']} \right) \lambda_{|\mathcal{F}_{\overline{\Lambda}_n}}[d\omega]. \end{aligned}$$

Integrating this inequality with respect to $\lambda[d\eta]$, we get

$$\int f(\omega)\lambda[d\omega] \le e^{4w(n)} \int f(\omega'_{\Lambda_n}\omega_{\overline{\Lambda}_n}) \lambda_{|\mathcal{F}_{\Lambda_n}}[d\omega'_{\Lambda_n}] \otimes \lambda_{|\mathcal{F}_{\overline{\Lambda}_n}}[d\omega_{\overline{\Lambda}_n}]. \tag{9.26}$$

A similar lower bound can be proved; this shows that $c_n(\lambda) \le 4w(n)$. □

Theorem 9.1 is a more elaborate version of Theorem 5.2 of [LPS1]; see also [LPS2].

Theorem 9.1 *Let Ψ be an absolutely summable real-valued interaction and β a translation invariant $(\Psi, \mathbb{P})$-Gibbs state on $(\Omega, \mathcal{F})$. Let Φ be an absolutely summable $\mathbb{R}^k$-valued interaction, and $C \subset \mathbb{R}^k$ a closed convex subset such that μ is finite at an interior point of C, where μ is the common RL-function of Lemma 9.1. Then*

a) Each of the sequences

$$\{\mathcal{A}_n\beta[\ \cdot\ |S_n \in C]\}\ ,\ \{\mathcal{A}_n\beta[\ \cdot\ |S_n^\eta \in C]\} \text{ or } \{\mathcal{A}_n\beta[\ \cdot\ |S_n^c \in C]\}, \tag{9.27}$$

has at least one limit point.

b) There exists $x' \in \mathbb{R}^k$ such that any limit point of the sequences (9.27) is an $(\langle x', \Phi\rangle + \Psi, \mathbb{P})$-Gibbs state.

c) The generalized chemical potential x' is a subgradient of $-\mu$ for any point of the non-empty compact set $N_C = \{x \in \mathbb{R}^k : \mu(x) = \sup_{y\in C} \mu(y)\}$.

d) If $\|\Phi\|_\# < \infty$ instead of $\|\Phi\| < \infty$, then any limit point of the sequence $\{\mathcal{A}_n\beta[\ \cdot\ |S_n \in C]\}$ is an $(f_\Phi^{x'}, \beta)$-equilibrium state, with $f_\Phi^{x'} := \langle x', f_\Phi\rangle$.

Proof: The theorem is essentially a corollary of Theorem 8.4. Any limit point λ of the sequences (9.27) is an $(f_\Phi^{x'}, \beta)$-equilibrium state:

$$\int_\Omega f_\Phi^{x'}(\omega)\lambda[d\omega] = p(f_\Phi^{x'}|\beta) + h(\lambda|\beta). \tag{9.28}$$

It is not difficult to show that

$$p(f_\Phi^{x'}|\beta) = p(f_\Phi^{x'} + f_\Psi|\mathbb{P}) - p(f_\Psi|\mathbb{P}), \tag{9.29}$$

and, from Lemma 10.1 (see also [G2]), that

$$h(\lambda|\beta) = h(\lambda|\mathbb{P}) - \int_\Omega f_\Psi(\omega)\mathbb{P}[d\omega] + p(f_\Psi|\mathbb{P}). \tag{9.30}$$

From these identities, we conclude that

$$\int_\Omega \{f_\Phi^{x'}(\omega) + f_\Psi(\omega)\}\lambda[d\omega] = p(f_\Phi^{x'} + f_\Psi|\mathbb{P}) + h(\lambda|\mathbb{P}); \tag{9.31}$$

that is, λ is an $(f_\Phi^{x'} + f_\Psi, \mathbb{P})$-equilibrium state, and therefore also an $(\langle x', \Phi\rangle + \Psi, \mathbb{P})$-Gibbs state. □

Remark: Convexity is used to prove LD-regularity and the existence of x'; equivalence of ensembles can be proved in special cases without assuming convexity of the conditioning set. Equivalence of ensembles for the empirical measure can be proved in the case of absolutely convergent potentials in essentially the same way as was done with $\mathcal{F}_{\mathrm{qloc}}$. Instead of $\mathcal{F}_{\mathrm{qloc}}^*$, one embeds $\mathcal{M}_1^+$ in the norm unit ball of the Banach space dual of the space of absolutely convergent interactions.

10 Information Gain and Specific Information Gain

10.1 Information Gain

Let $(\Omega, \mathcal{F})$ be a measurable space. Let λ and β be two probability measures on $(\Omega, \mathcal{F})$. The information gain $\mathcal{H}(\lambda|\beta)$ of λ with respect to β is defined by

$$\mathcal{H}(\lambda|\beta) := \begin{cases} \int_\Omega f \ln f(\omega)\beta[d\omega], & \text{if } \lambda[d\omega] = f(\omega)\beta[d\omega], \\ +\infty, & \text{otherwise,} \end{cases} \tag{10.1}$$

with $0 \ln 0 := 0$. If $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{F}$, then

$$\mathcal{H}_\mathcal{B}(\lambda|\beta) := \mathcal{H}(\lambda_{|\mathcal{B}}|\beta_{|\mathcal{B}}), \tag{10.2}$$

where $\lambda_{|\mathcal{B}}$ and $\beta_{|\mathcal{B}}$ denote the restrictions of the measures to $\mathcal{B}$.

Proposition 10.1 *For probability measures λ and β on the measurable space $(\Omega, \mathcal{F})$, we have*

$$\mathcal{H}(\lambda|\beta) \geq 0. \tag{10.3}$$

If $\mathcal{B}_1, \mathcal{B}_2$ are sub-σ-algebras of $\mathcal{F}$ and $\mathcal{B}_1 \subset \mathcal{B}_2$, then

$$\mathcal{H}_{\mathcal{B}_1}(\lambda|\beta) \leq \mathcal{H}_{\mathcal{B}_2}(\lambda|\beta). \tag{10.4}$$

There exists a sequence $\{\mathcal{F}_{\Lambda_n}\}$ of finite *sub-σ-algebras of $\mathcal{F}$ such that*

$$\lim_{n\to\infty} \mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\lambda|\beta) = \mathcal{H}(\lambda|\beta). \tag{10.5}$$

Proof: With $\psi(f) = f \ln f - f + 1$ we have $\psi(f) \geq 0$ and $\int f \ln f \, d\beta = \int \psi(f)\, d\beta$ when $f \geq 0$ and $\int f \, d\beta = 1$, which proves (10.3). Since ψ is convex, (10.4) follows from Jensen's inequality. For (10.5), if λ is not absolutely continuous with respect to β, take $\mathcal{F}_n$ to be the σ-algebra generated by any set $A \in \mathcal{F}$ with $\lambda[A] > 0$ and $\beta[A] = 0$. Otherwise, let f denote a Radon–Nikodym derivative of λ with respect to β. Define $\mathcal{B}_n$ to be the σ-algebra of subsets of the real line generated by sets of the form

$$\{x \in \mathbb{R} : k/2^n \leq x < (k+1)/2^n\} \text{ for } k = 0, \ldots, 4^n. \tag{10.6}$$

Let $\mathcal{F}_n = f^{-1}\mathcal{B}_n$. Let f_n be the β-conditional expectation of f with respect to $\mathcal{F}_n$. Then $\lim_{n\to\infty} f_n = f$, β-almost surely. If $\mathcal{H}(\lambda|\beta) < \infty$, the Legesgue Dominated Convergence Theorem implies the convergence of $\{\mathcal{H}_{\mathcal{F}_n}(\lambda|\beta)\}$ to $\mathcal{H}(\lambda|\beta)$. Conversely, if $\{\mathcal{H}_{\mathcal{F}_n}(\lambda|\beta)\}$ is bounded above, $\mathcal{H}(\lambda|\beta)$ is finite. □

Elementary considerations yield the following.

Lemma 10.1 *If $\lambda = f\beta$, g is λ integrable and $\int e^g \, d\beta = 1$, then*

$$\mathcal{H}(\lambda|e^g\beta) + \int g \, d\lambda = \mathcal{H}(\lambda|\beta). \tag{10.7}$$

Proposition 10.2 *For probability measures λ and β on the measurable space $(\Omega, \mathcal{F})$, we have*

$$\mathcal{H}(\lambda|\beta) = \sup_{g\in\mathcal{F}_b} \int g \, d\lambda - \ln \int e^g \, d\beta, \tag{10.8}$$

where $\mathcal{F}_b$ denotes the set of bounded, $\mathcal{F}$ measurable, real valued functions.

Proof: If $\mathcal{H}(\lambda|\beta) < \infty$, we have

$$\sup_{g\in\mathcal{F}_b} \int g\,d\lambda - \ln\int e^g\,d\beta = \sup_{\substack{g\in\mathcal{F}_b \\ \int e^g\,d\beta=1}} \int g\,d\lambda = \sup_{\substack{g\in\mathcal{F}_b \\ \int e^g\,d\beta=1}} \mathcal{H}(\lambda|\beta) - \mathcal{H}(\lambda|e^g\beta) \quad (10.9)$$

from (10.7). This shows that $\mathcal{H}(\lambda|\beta)$ is an upper bound for the right hand side of (10.8). To show that it is the supremum, we first remark that if λ is not absolutely continuous with respect to β, we can select A with $\lambda[A] > 0$ and $\beta[A] = 0$. Define

$$g_c(\omega) = \begin{cases} c & \text{if } c \in A, \\ 0 & \text{otherwise.} \end{cases} \quad (10.10)$$

The supremum over $\{g_c\}$ gives an infinite value to the right hand side of (10.8). If $\lambda = f\beta$, define

$$g_n(\omega) = \begin{cases} \ln f & \text{if } |\ln f| \le n, \\ n\,\mathrm{sign}(f) & \text{otherwise.} \end{cases} \quad (10.11)$$

Then $\lim_{n\to\infty}\int e^{g_n}\,d\beta = 1$ and $\lim_{n\to\infty}\int g_n\,d\lambda = \mathcal{H}(\lambda|\beta)$, which shows that the right hand side of (10.8) can be no less than $\mathcal{H}(\lambda|\beta)$. □

Lemma 10.2 *Let $(\Omega_1,\mathcal{F}_1)$ and $(\Omega_2,\mathcal{F}_2)$ be measurable spaces. Let $\Omega = \Omega_1\times\Omega_2$ with $\mathcal{F}$ the corresponding product σ-algebra. Let λ and β be probability measures on $(\Omega,\mathcal{F})$ with λ_1,λ_2 and β_1,β_2 denoting the restrictions to $\mathcal{F}_1,\mathcal{F}_2$ considered as sub-σ-algebras of $\mathcal{F}$. Assume $\beta = \beta_1\otimes\beta_2$. Then we have*

$$\mathcal{H}(\lambda|\beta) = \mathcal{H}(\lambda|\lambda_1\otimes\lambda_2) + \mathcal{H}(\lambda_1|\beta_1) + \mathcal{H}(\lambda_2|\beta_2)\,. \quad (10.12)$$

Proof: If $\mathcal{H}(\lambda|\beta) < \infty$, we have $\mathcal{H}(\lambda_1|\beta_1) < \infty$ and $\mathcal{H}(\lambda_2|\beta_2) < \infty$ from(10.4). If we write $\lambda = f(\omega_1,\omega_2)\beta$ and $\lambda_1 = f_1(\omega_1)\beta_1$, $\lambda_2 = f_2(\omega_2)\beta_2$, the relation (10.12) is a formal equality involving integrals. Since three of the four terms are finite, the fourth must be finite and equality obtains. Similarly, if the terms on the right hand side of (10.12) are finite, one can deduce the remaining Radon–Nikodym derivative and equality obtains. □

Lemma 10.3 *For $j = 1,\ldots,m$, let $\alpha_j > 0$, $\sum_1^m \alpha_i = 1$, and let λ_j be a probability measure with $\mathcal{H}(\lambda_j|\beta) < \infty$. Then*

$$\sum_1^m \alpha_j\mathcal{H}(\lambda_j|\beta) + \sum_1^m \alpha_j\ln\alpha_j \le \mathcal{H}(\sum_1^m \alpha_j\lambda_j|\beta) \le \sum_1^m \alpha_j\mathcal{H}(\lambda_j|\beta). \quad (10.13)$$

Proof: Let $\overline{\lambda} = \sum\alpha_j\lambda_j$. On $\{1,\ldots,m\}\times\Omega$, define the probability measures λ^* and $\overline{\lambda^*}$ so that $\lambda^*[\{j\}\times B] = \alpha_j\lambda[B]$ and $\overline{\lambda^*}[\{j\}\times B] = \alpha_j\overline{\lambda}[B]$. A straightforward calculation similar to that of the previous lemma shows that

$$\sum_1^m \alpha_j\mathcal{H}(\lambda_j|\beta) - \mathcal{H}(\sum_1^m \alpha_j\lambda_j|\beta) = \mathcal{H}(\lambda^*|\overline{\lambda^*}). \quad (10.14)$$

This proves the second inequality of (10.13); the first follows by noting that, for fixed j, the Radon–Nikodym derivative of λ^* with respect to $\overline{\lambda^*}$ is equal to or greater than $1/\alpha_j$ so that

$$\mathcal{H}(\lambda^*|\overline{\lambda^*}) = \sum_1^m \alpha_j\ln\frac{d\lambda_j}{d\overline{\lambda}} \le \sum_1^m \alpha_j\ln 1/\alpha_j\,. \quad (10.15)$$

□

Proposition 10.3 *For probability measures λ and β on the measurable space $(\Omega, \mathcal{F})$, we have*

$$\|\lambda - \beta\|^2_{\mathrm{TV}} \le 2\mathcal{H}(\lambda|\beta), \tag{10.16}$$

where the norm is the total variation norm.

Proof: If $\mathcal{H}(\lambda|\beta) = \infty$, there is nothing to prove. Otherwise, let $\lambda = f\beta$. Let $A = \{\omega : f(\omega) \ge 1\}$. Let $\mathcal{F}_A$ denote the (finite) σ-algebra generated by A. With $x = \lambda[A]$ and $y = \beta[A]$, we have

$$\|\lambda - \beta\|_{\mathrm{TV}} = 2(x - y). \tag{10.17}$$

If $x = 0$ or $y = 0$ or $x = 1$ or $y = 1$, then $\lambda = \beta$. Otherwise, we have

$$\mathcal{H}_{\mathcal{F}_A}(\lambda|\beta) = x \ln \frac{x}{y} + (1 - x) \ln \frac{1 - x}{1 - y}. \tag{10.18}$$

Since $\mathcal{H}(\lambda|\beta) \ge \mathcal{H}_{\mathcal{F}_A}(\lambda|\beta)$, it suffices to show that

$$2(x - y)^2 \le x \ln \frac{x}{y} + (1 - x) \ln \frac{1 - x}{1 - y}. \tag{10.19}$$

But this is a consequence of the fact that the convex function

$$u(x) := x \ln x + (1 - x) \ln(1 - x) - 2x^2 \tag{10.20}$$

satisfies $u(x) - u(y) \ge (x - y)u'(y)$ when $0 < x < 1$ and $0 < y < 1$. □

10.2 Specific Information Gain

We prove Theorem 8.1 and Lemma 8.1. The setting and the notations are those of Section 8.1.

Lemma 10.4 *Let $\beta \in \mathcal{M}_1^{+,\theta}$ with $c_m(\beta) < \infty$. Then, for $n > m$, the restriction of $\beta_{|\mathcal{F}_{\Lambda_n}}$ to the σ-algebra $\mathcal{F}_{\Lambda_n}$ and the probability measure*

$$\bigotimes_{i \in \Lambda_{n|m}} \theta_i \beta_{|\mathcal{F}_{\Lambda_m}} \otimes \beta_{|\mathcal{F}_{\Lambda_n \setminus \cup_j \theta_j \Lambda_m}} \tag{10.21}$$

are mutually absolutely continuous. The absolute value of the logarithm of the Radon–Nikodym derivative is bounded almost surely by $q(n|m)c_m(\beta)$.

Proof: Consider the analog of (8.6) with $\beta_{|\mathcal{F}_{\Lambda_n}}$ and $\beta^* := \beta_{|\mathcal{F}_{\Lambda_n \setminus \Lambda_m}} \otimes \beta_{|\mathcal{F}_{\Lambda_m}}$. Let $A \in \mathcal{F}_{\Lambda_n}$; since $c_m(\beta)$ is finite, $\beta_{|\mathcal{F}_{\Lambda_n}}(A) = 0$ if and only if $\beta^*(A) = 0$. (8.6) implies the a.s. bound of $c_m(\beta)$ for the absolute value of the logarithm of the Radon–Nikodym derivative. One repeats the argument for subsets of $\Lambda_n \setminus \Lambda_m$. □

Proof of Theorem 8.1

Proof of a): For $n > m$, we have from Lemmas 10.1, 10.2, 10.4 and translation invariance the inequality

$$\mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\lambda|\beta) \ge q(n|m)\big(\mathcal{H}_{\mathcal{F}_{\Lambda_m}}(\lambda|\beta) - c_m(\beta)\big). \tag{10.22}$$

We have $c_m(\beta)/V_m \to 0$ as $m \to \infty$, and $q(n|m)V_m/V_n \to 1$ as $n \to \infty$. Therefore we have

$$\liminf_n \frac{\mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\lambda|\beta)}{V_n} \geq \frac{\mathcal{H}_{\mathcal{F}_{\Lambda_m}}(\lambda|\beta) - c_m(\beta)}{V_m}. \tag{10.23}$$

Now take lim sup over m. The lower semicontinuity of $\mathcal{H}_{\mathcal{F}_{\Lambda_m}}(\cdot|\beta)$ and (10.23) imply that $h(\cdot|\beta)$ is lower semicontinuous. Indeed, let $\{\lambda_k\}$ be a net in $\mathcal{M}_1^{+,\theta}$ converging to λ; we have

$$\liminf_k h(\lambda_k|\beta) \geq \liminf_k \frac{\mathcal{H}_{\mathcal{F}_{\Lambda_m}}(\lambda_k|\beta) - c_m(\beta)}{V_m} \geq \frac{\mathcal{H}_{\mathcal{F}_{\Lambda_m}}(\lambda|\beta) - c_m(\beta)}{V_m}; \tag{10.24}$$

hence we have $\liminf_k h(\lambda_k|\beta) \geq h(\lambda|\beta)$. The affine character of $h(\lambda|\beta)$ is a consequence of Lemma 10.3. For the proof of b), see for example [G2]. □

Proof of Lemma 8.1

Fix the integer $m > 0$. For $n \geq m$ we have

$$r := 2 + (n-m)\operatorname{div}(2m+1) \quad \text{implies} \quad \Lambda_n \subset \sum_{j\in\Lambda_r} \theta_{(2m+1)j}(\theta_i\Lambda_m) \tag{10.25}$$

for any $i \in \Lambda_m$. Then, with $i \in V_m$, (10.12), (10.7) and Lemma 10.4 yield

$$\mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\theta_i\mathcal{P}_m\lambda|\beta) \leq V_r\Big(\mathcal{H}_{\theta_i\mathcal{F}_m}(\lambda|\beta) + c_m(\beta)\Big), \tag{10.26}$$

because $\mathcal{H}_{\theta_i\mathcal{F}_m}(\lambda|\beta) = \mathcal{H}_{\theta_i\mathcal{F}_m}(\theta_i\mathcal{P}_m\lambda|\beta)$. From translation invariance, we deduce that $\mathcal{H}_{\theta_i\mathcal{F}_m}(\lambda|\beta) = \mathcal{H}_{\mathcal{F}_{\Lambda_m}}(\lambda|\beta)$. From (10.26) and Lemma 10.3, we then have

$$\mathcal{H}_{\mathcal{F}_{\Lambda_n}}(\mathcal{A}_m\mathcal{P}_m\lambda|\beta) \leq V_r\Big(\mathcal{H}_{\mathcal{F}_{\Lambda_m}}(\lambda|\beta) + c_m(\beta)\Big). \tag{10.27}$$

We divide by V_n and use $\lim_n V_n/V_r = V_m$. The other bound follows from (10.23) applied to $\mathcal{A}_m\mathcal{P}_m\lambda$ and Lemma 10.3. □

References

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

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

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

[C] I.Csiszár, Sanov property, generalized I-projection and a conditional limit theorem, *Ann.Prob.* **12**, 768-793 (1984)

[CC] J.M.van Campenhout and T.M.Cover, Maximum Entropy and Conditional Probability, *IEEE Trans. Inform. Theory* **4**, 483-489 (1981)

[CCC] I.Csiszár, T.M.Cover and B.-S. Choi, Conditional Limit Theorems under Markov Conditioning, *IEEE Trans. Inform. Theory* **33**, 788-801 (1987)

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

[DSZ] J.-D.Deuschel, D.W.Stroock and H.Zessin, Microcanonical distribution for lattice gases, *Commun. Math. Phys.* **139**, 83 (1991)

[ET] I.Ekeland and R.Temam, Analyse convexe et problèmes variationnels, Dunod, Gauthier–Villars, Paris–Bruxelles–Montréal (1974)

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

[G1] H.-O.Georgii, Large deviations and maximum entropy principle for interacting random fields on $\mathbb{Z}^d$, *Ann. Prob.* **21**, 1845-1875 (1993)

[G2] H.-O.Georgii, *Gibbs Measures and Phase Transitions.* Walter de Gruyter (de Gruyter Studies in Mathematics, Vol. 9), Berlin–New York, (1988)

[Gi] J.W.Gibbs, Elementary Principles of Statistical Mechanics, Yale University Press, New Haven (1902)

[GOR] P.Groeneboom, J.Oosterhoff and F.H.Ruymgaart, Large Deviation Theorems for Empirical Probability Measures *Ann.Prob.* **7**, 553-586 (1979)

[L] O.E.Lanford, Entropy and equilibrium states in classical mechanics, in Statistical Mechanics and Mathematical Problems, A. Lenard, ed., Lecture Notes in Physics 20, Springer,Berlin (1973)

[LP] J.T.Lewis, C.-E.Pfister, Thermodynamic Probability Theory: some aspects of large deviations, preprint DIAS-93-33 (1993)

[LPS1] J.T.Lewis, C.-E.Pfister and W.G.Sullivan, Large Deviations and the Thermodynamic Formalism: a new proof of the equivalence of ensembles, in On Three Levels, M. Fannes, C. Maes, A. Verbeure eds., Plenum Press 183-193, (1994)

[LPS2] J.T.Lewis, C.-E.Pfister and W.G.Sullivan, The Equivalence of Ensembles for Lattice Systems: Some examples and a Counterexample *J.Stat.Phys.* **77**, 397-419 (1994)

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

[R] R.T.Rockafellar, Convex Analysis, Princeton University Press, Princeton (1970)

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

[RZ] S.Roelly, H.Zessin, The Equivalence of Equilibrium Principles in Statistical Mechanics and some Applications to Large Particle Systems, *Expositiones Mathematicae* **11**, 385-405 (1993)

[Ru1] D.Ruelle, Correlation functionals *J. Math. Phys.* **6**, 201-220 (1965)

[Ru2] D.Ruelle, Thermodynamic Formalism: the Mathematical Structures of Classical Equilibrium Statistical Mechanics. Addison-Wesley, Reading, Mass. (1978)

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

[Su] W.G.Sullivan, Specific Information Gain for Interacting Markov Processes *Z.Wahr.verw.Geb.* **37**, 77-90 (1976)

[Va1] S.R.S.Varadhan, Asymptotic probabilities and differential equations, *Comm. Pure App. Math.* **19**, 261 (1966)

[Va2] S.R.S.Varadhan, Large Deviations, in Lecture Notes in Mathematics **1362** ed. P.L. Hennequin, Spinger, Berlin (1988)