AN EXTENSION OF SANOV'S THEOREM. APPLICATION TO THE GIBBS CONDITIONING PRINCIPLE

CHRISTIAN LÉONARD[1,2], JAMAL NAJIM[1]

Abstract. A large deviation principle is proved for the empirical measures of independent identically distributed random variables with a topology based on functions having only some exponential moments. The rate function differs from the usual relative entropy: It involves linear forms which are no longer measures. Following D.W. Stroock and O. Zeitouni, the Gibbs Conditioning Principle (GCP) is then derived with the help of the previous result. Besides a rather direct proof, the main improvements with respect to already published GCP's are the following: Convergence holds in situations where the underlying log-Laplace transform (the pressure) may not be steep and the constraints are built on energy functions admitting only some finite exponential moments. Basic techniques from Orlicz spaces theory appear to be a powerful tool.

(1) *Modal-X, Université Paris 10, Bâtiment G*
200, Av. de la République, 92001 Nanterre cedex, France
(2) *CMAP, Ecole Polytechnique, 91128 Palaiseau cedex, France*

E-mail: *Christian.Leonard@nospam.u-paris10.fr, najim@nospam.tsi.enst.fr*

Math. Subj. Class.: 60F10, 60G57 (primary); 46E30 (secondary)
Key words: Large deviations, empirical measures, Sanov's theorem, Gibbs conditioning principle, Orlicz spaces.

1. Introduction

Let $(Y_i)_{i\geq 1}$ be a sequence of independent identically distributed random variables with common law μ on a measurable space $(\Sigma, \mathcal{A})$. The empirical measures

$$L_n^Y = \frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i}$$

(δ_a is the Dirac measure at a) are random elements in the set $\mathcal{P}$ of the probability measures on $(\Sigma, \mathcal{A})$.

Date: May 22, 2003, Revised version.

1.1. **Sanov's theorem.** Sanov's theorem describes the limiting behaviour of $\frac{1}{n}\log \mathbb{P}(L_n^Y \in \cdot)$ as n tends to infinity, by means of a Large Deviation Principle (LDP) whose good rate function is given for any $\nu \in \mathcal{P}$ by

$$H(\nu \mid \mu) = \int_\Sigma \log\left(\frac{d\nu}{d\mu}\right) d\nu \quad \text{if} \quad \nu \ll \mu,$$

and ∞ otherwise: The relative entropy of ν with respect to μ. For this LDP, the topology on $\mathcal{P}$ is $\sigma(\mathcal{P}, B)$: The coarsest topology which makes the evaluations $\nu \in \mathcal{P} \mapsto \int_\Sigma f\, d\nu \in \mathbb{R}$ continuous for all f in the space B of the bounded measurable functions on Σ.

In the present paper, this LDP is extended by considering a stronger topology where B is replaced by the space $\mathcal{L}_\tau$ of all functions f with some finite exponential moments with respect to μ :

$$\int_\Sigma e^{a|f|}\, d\mu < \infty, \quad \text{for some } a > 0. \tag{1.1}$$

We identify this space as the Orlicz space associated with the following norm:

$$\|f\|_\tau = \inf\left\{a > 0, \quad \int_\Sigma \tau\Big(\frac{f}{a}\Big)\, d\mu \le 1\right\} \quad \text{where} \quad \tau(x) = e^{|x|} - |x| - 1.$$

A precise description of the space of continuous linear forms of this non-reflexive Banach space allows us to define the state space for the extended Sanov theorem (see Section 2 for details). This space is not $\mathcal{P}$ anymore, but a different set $\mathcal{Q}$ of all non-negative continuous linear forms on $\mathcal{L}_\tau$ with a unit mass. In particular, the effective domain of the relative entropy H is a strict subset of $\mathcal{Q}$. The topology on $\mathcal{Q}$ is $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ and the rate function I has the following form (for any $\ell \in \mathcal{Q}$ such that $I(\ell) < \infty$)

$$I(\ell) = H(\ell^a \mid \mu) + \sup\left\{\langle \ell^s, f\rangle;\ f, \int_\Sigma e^f\, d\mu < \infty\right\},$$

where $\ell = \ell^a + \ell^s$ is uniquely decomposed into the sum of a probability measure ℓ^a which is absolutely continuous with respect to μ, and a non-negative continuous linear form ℓ^s on $\mathcal{L}_\tau$ which is not σ-additive (if non null).

The space of singular forms is the annihilator of the space M_τ of all functions admitting *all* exponential moments (see Section 2.1). In particular, the "mass" of ℓ^s is $\langle \ell^s, \mathbf{1}\rangle = 0$, although $\ell^s \ge 0$ may not be zero. For more details on this subject see [12] and [14]. In the context of Csiszár's example (Section 3.4), the singular parts are approximated in a certain sense by probability measures (Proposition 3.10 and Remark 3.6). For a precise statement of the extended Sanov theorem, see Theorem 3.2 below.

1.2. **Gibbs conditioning principle.** The Gibbs conditioning principle describes the limiting behaviour as n tends to infinity of the law of k tagged particles $Y_1, \dots, Y_k$ under the constraint that L_n^Y belongs to some subset A_0 of $\mathcal{P}$ with $\mathbb{P}(L_n^Y \in A_0)$ positive for all $n \geq 1$. Typically, the expected result is

$$\lim_{n\to\infty} \mathbb{P}\Big((Y_1, \dots, Y_k) \in \cdot \mid L_n^Y \in A_0\Big) = \nu_*^k(\cdot), \tag{1.2}$$

where ν_* minimizes $\nu \mapsto H(\nu \mid \mu)$ subject to $\nu \in A_0$. This question is of interest in statistical mechanics since this conditional law is a canonical distribution. A typical conditioning set is

$$A_0 = \left\{ \nu \in \mathcal{P}; \int_\Sigma \varphi \, d\nu = E \right\}, \tag{1.3}$$

where φ is an energy function on the one-particle phase space Σ.
The close relationship between Sanov's theorem and the Gibbs conditioning principle is well known. It has been exploited by I. Csiszár in [6] and by D. W. Stroock and O. Zeitouni in [21]. As in [6] and [21], we shall not be able to handle the difficult and important case where $\mathbb{P}(L_n^Y \in A_0) = 0$ for $n \geq 1$. We follow [21] by introducing blow-ups A_δ of $A_0 = \bigcap_{\delta>0} A_\delta$ such that $\mathbb{P}(L_n^Y \in A_\delta) > 0$ for all $n \geq 1$ and $\delta > 0$.
With the extended Sanov theorem in hand, rather than its usual version, the proof of the Gibbs conditioning principle is more direct and its assumptions can be significantly relaxed. On one hand, conditioning sets A_0 as in (1.3) are naturally built on energy functions φ in $\mathcal{L}_\tau$, that is satisfying (1.1). On the other hand, the following restriction, assumed in [21]: For all $\delta > 0$, $\lim_{n\to\infty} \nu_*^n(\{L_n^Y \in A_\delta\}) = 1$, is removed. As a consequence, it is proved that the Gibbs conditioning principle still holds in situations where a lack of steepness of the pressure $\beta \mapsto \log \int_\Sigma e^{\beta\varphi} \, d\mu$ occurs (see Csiszár's example in Sections 3.4 and 4.4). The Gibbs conditioning principle we have obtained is stated in Theorem 4.2.

Let us emphasize that the introduction of Orlicz spaces and specifically of basic duality results for Orlicz spaces (recalled in Section 2) is of prime necessity to get direct proofs of our main results.

1.3. **About the literature.** I. N. Sanov [19] proved the LDP for L_n^Y with $\Sigma = \mathbb{R}$ and the weak topology on $\mathcal{P}$. This LDP is extended to the situation where Σ is a Polish space by M. D. Donsker and S. R. S. Varadhan [9] and R. R. Bahadur and S. Zabell [3] for the weak topology. P. Groeneboom, J. Oosterhoff and F. H. Ruymgaart [13] dropped the Polish requirement and considered Hausdorff spaces. They obtained Sanov's theorem for the so-called τ-topology: $\sigma(\mathcal{P}, B)$. A. de Acosta improved this result and simplified the proof in [7]. In [6], I. Csiszár proved Sanov's theorem in a general setting by means of an alternative approach based on projection in information.

In [10], P. Eichelsbacher and U. Schmock consider the U-empirical measures $L_n^{Y,k} = \frac{1}{n\cdots(n-k+1)}\sum_{(i_1,\dots,i_k)}\delta_{(X_{i_1},\dots,X_{i_k})}$ where the sum is taken over all k-tuples in $\{1,\dots,n\}^k$ with pairwise distinct components. The special case $k=1$ is an extension of Sanov's theorem: The rate function is the usual relative entropy. The lower bound is obtained for a topology which is slightly weaker than $\sigma(\mathcal{P}_\tau,\mathcal{L}_\tau)$ and the upper bound holds for a topology on $\mathcal{P}$ which is slightly weaker than $\sigma(\mathcal{P}_\tau,\mathcal{M}_\tau)$, where $\mathcal{M}_\tau$ stands for the space of all functions f which admit finite exponential moments of all orders with respect to μ :

$$\int_\Sigma e^{a|f|}\,d\mu < \infty, \quad \text{for all } a>0, \tag{1.4}$$

and $\mathcal{P}_\tau$ stands for the set of all probability measures which integrate all functions in $\mathcal{L}_\tau$ or $\mathcal{M}_\tau$. Our result for $k=1$ is stronger, but on the other hand, the LDP for $L_n^{Y,k}$ in [10] for $k\geq 2$ is far from being trivial and is not a consequence of our results.

As already mentioned, the Gibbs conditioning principle (GCP) has been studied by I. Csiszár in [6] and by D. W. Stroock and O. Zeitouni in [21]. An updated presentation of the second approach is available in the textbook of A. Dembo and O. Zeitouni ([8], Section 7.3). In [4], E. Bolthausen and U. Schmock proved a GCP for the occupation measures of uniformly ergodic Markov chains. Based on [6], A. Aboulalaâ [1] obtained a GCP for the empirical measures L_n^Y of Markov jump processes. With the LDP for $L_n^{Y,k}$ in hand, P. Eichelsbacher and U. Schmock [10] derived a GCP, following the approach of [21]. They obtained it for k tagged particles with an energy function φ in $\mathcal{M}_\tau$ (see (1.3) and (1.4)) and for a topology on $\mathcal{P}_\tau$ which is slightly weaker than $\sigma(\mathcal{P}_\tau,\mathcal{M}_\tau)$.

In [6], I. Csiszár obtained alternative results with an alternative powerful approach. In particular, he proved the convergence in information of the conditioned laws which implies their convergence in variation and the notion of generalized I-projection is introduced so that the GCP holds even with energy functions satisfying (1.1).

1.4. **Outline of the paper.** We recall in Section 2 a few definitions and results about Orlicz spaces: It is natural and worthy to express exponential moment conditions ((1.1) and (1.4)) in terms of Orlicz spaces (see Section 2.3).
We prove the LDP in Section 3. The main result is Theorem 3.2 which states the LDP and describes the associated rate function.
We study the Gibbs Conditioning Principle in Section 4 whose main statement is Theorem 4.2.

2. Definitions and results about Orlicz spaces

In this section, elementary facts about Orlicz spaces and their dual spaces are recalled for future use, without proof.

2.1. **Basic definitions and results.** A *Young function* θ is an even, convex, $[0,\infty]$-valued function satisfying $\lim_{s\to\infty}\theta(s)=\infty$ and $\theta(s_0)<\infty$ for some $s_0>0$. Let μ be a probability measure on the measurable space $(\Sigma,\mathcal{A})$. Consider the following vector spaces of measurable functions:

$$\begin{aligned}\mathcal{L}_\theta &= \left\{f:\Sigma\to\mathbb{R},\ \exists a>0, \int_\Sigma \theta\Big(\frac{f}{a}\Big)\,d\mu<\infty\right\},\\ \mathcal{M}_\theta &= \left\{f:\Sigma\to\mathbb{R},\ \forall a>0, \int_\Sigma \theta\Big(\frac{f}{a}\Big)\,d\mu<\infty\right\}.\end{aligned}$$

The spaces L_θ and M_θ correspond to $\mathcal{L}_\theta$ and $\mathcal{M}_\theta$ when μ-almost everywhere equal functions are identified. Consider the following Luxemburg norm on L_θ:

$$\|f\|_\theta = \inf\left\{a>0,\quad \int_\Sigma \theta\Big(\frac{f}{a}\Big)\,d\mu\le 1\right\}. \tag{2.1}$$

Then, $(L_\theta,\|.\|_\theta)$ is a Banach space called the Orlicz space associated with θ. M_θ is a subspace of L_θ. If θ is a finite function, M_θ is the closure of the space of step functions $\sum_{i=1}^n a_i\mathbf{1}_{A_i}$ under $\|\cdot\|_\theta$. For references, see [2], [17]. Let θ^* be the convex conjugate of the Young function θ:

$$\theta^*(t)=\sup_{s\in\mathbb{R}}\{st-\theta(s)\}.$$

As θ^* is a Young function, one can consider the Orlicz space L_{θ^*}.
Hölder's inequality holds between L_θ and L_{θ^*}: For all $f\in L_\theta$ and $g\in L_{\theta^*}$,

$$fg\in L^1(\mu)\quad\text{and}\quad \int_\Sigma |fg|\,d\mu\le 2\|f\|_\theta\|g\|_{\theta^*}. \tag{2.2}$$

A Young function θ satisfies the Δ_2-condition if there exists $K>0, s_0\ge 0$ such that for all $s\ge s_0$, $\theta(2s)\le K\theta(s)$. If θ satisfies the Δ_2-condition, then $M_\theta=L_\theta$, see ([17], Corollary 5, p. 77).

2.2. **Duality in Orlicz spaces.** By (2.2), any g in L_{θ^*} defines a continuous linear form on L_θ for the duality bracket $\langle f,g\rangle=\int fg\,d\mu$. In the general case, the topological dual space of $(L_\theta,\|.\|_\theta)$ may be larger than L_{θ^*}. Nevertheless, we always have the following result:

Theorem 2.1. *Let θ be a finite Young function and θ^* its convex conjugate. The topological dual space of M_θ can be identified, by means of the previous duality bracket, with L_{θ^*}: $M'_\theta\simeq L_{\theta^*}$.*

For a proof of this result, see [17] or ([15], Section 4).

Remark 2.1. If θ satisfies the Δ_2-condition, then $L'_\theta \simeq L_{\theta^*}$

As L_θ is a Riesz space (see [5]), one can define the absolute value $|\ell|$ of any $\ell \in L'_\theta$.

Definition 2.2. *Let $\ell \in L'_\theta$, ℓ is said to be μ-singular if there exists a sequence $A_1 \supset A_2 \supset A_3 \supset \cdots$ of measurable sets such that*

$$\lim_k \mu(A_k) = 0 \quad \textit{and} \quad \langle |\ell|, \mathbf{1}_{\Sigma \setminus A_k} \rangle = 0, \quad \forall k \geq 1.$$

Let us denote by L^s_θ the subspace of all μ-singular elements of L'_θ.

Theorem 2.3. *Let θ be a finite Young function, the topological dual space L'_θ of $(L_\theta, \|\cdot\|_\theta)$ is the direct sum:*

$$L'_\theta \simeq (L_{\theta^*} \cdot \mu) \oplus L^s_\theta.$$

Therefore any continuous linear form ℓ on L_θ is uniquely decomposed as $\ell = \ell^a + \ell^s$, where ℓ^a and ℓ^s are continuous, $\frac{d\ell^a}{d\mu} \in L_{\theta^*}$ and ℓ^s is μ-singular.

For a proof of this result, see ([14], Theorem 2.2) or ([15], Theorem 4.3).
ℓ^a is called the absolutely continuous part of ℓ and ℓ^s its singular part.

Proposition 2.4. *Let θ be a finite Young function. Then for any $f \in M_\theta$ and $\ell^s \in L^s_\theta$, we have $\langle \ell^s, f \rangle = 0$*

For a proof of this result, see [14] or ([15], Proposition 4.2).

2.3. **Orlicz spaces and exponential moment conditions.** Consider

$$\gamma(s) = e^s - s - 1 \quad \text{and} \quad \tau(s) = \gamma(|s|). \tag{2.3}$$

Then, τ is a Young function and the two following equivalences are straightforward:

$$\left(\exists a > 0, \int e^{a|f|}\, d\mu < \infty\right) \quad \Leftrightarrow \quad f \in \mathcal{L}_\tau,$$

$$\left(\forall a > 0, \int e^{a|f|}\, d\mu < \infty\right) \quad \Leftrightarrow \quad f \in \mathcal{M}_\tau.$$

In case $f \in \mathcal{L}_\tau$, we shall say that f admits *some* exponential moments; in case $f \in \mathcal{M}_\tau$, we will say that f admits *all* its exponential moments.

3. An extension of Sanov's Theorem

The main result of this section is Theorem 3.2 which states the LDP and describes the associated rate function. The LDP is partially proved in Section 3.2 via a projective limit technique which yields a convex conjugate rate function Θ^*. In Section 3.3, we identify the rate function by comparing it to an auxiliary function J and by using a result of R. T. Rockafellar on the representation of convex functionals. In Section 3.4, we study an example due to

Csiszár in order to show the existence of singular parts.

3.1. **Statement of the extended Sanov theorem.** Let us consider $\mathcal{L}_\tau$ and its algebraic dual space $\mathcal{L}_\tau^*$. Note that almost everywhere equal functions are not identified when dealing with $\mathcal{L}_\tau$. Consider the collection of linear forms on $\mathcal{L}_\tau^*$ denoted by $G_f : \ell \mapsto \langle \ell, f\rangle$, $f \in \mathcal{L}_\tau$. Denote by $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$ the coarsest topology on $\mathcal{L}_\tau^*$ which makes all the G_f's continuous and by $\mathcal{E}$ the smallest σ-field on $\mathcal{L}_\tau^*$ which makes them measurable. We are interested in the large deviation behaviour of

$$L_n^Y = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i} \in \mathcal{L}_\tau^*,$$

where $\{Y_i\}_{i\geq 1}$ is a sequence of Σ-valued, independent and identically μ-distributed variables.

An identity between function spaces. The following identifications prevail in the remainder of the paper

$$L_{\tau^*} \subset L'_\tau \subset L_\tau^* \subset \mathcal{L}_\tau^*, \tag{3.1}$$

where

- $\cdot$ L_{τ^*} is the Orlicz space associated with the Young function τ^*
- $\cdot$ L'_τ (resp. L_τ^*) is the topological (resp. algebraic) dual of L_τ and
- $\cdot$ $\mathcal{L}_\tau^*$ is the algebraic dual of $\mathcal{L}_\tau$.

For the first identification, take $f \in L_{\tau^*}$, then $f\mu \in L'_\tau$ by Hölder's inequality (2.2): We write $L_{\tau^*} = L_{\tau^*}.\mu$ for short. The second identification is straighforward. For the third identification, let $\ell \in L_\tau^*$ and consider $\tilde{\ell}$ defined on $\mathcal{L}_\tau^*$ by $\langle \tilde{\ell}, f\rangle = \langle \ell, \dot{f}\rangle$ where $f \in \mathcal{L}_\tau$ and $\dot{f} \in L_\tau$ is the equivalence class of f with respect to μ-almost everywhere equality. The form $\tilde{\ell}$ is well-defined and the third identification holds.

The state space. The state space of the extended Sanov theorem is

$$\mathcal{Q} \stackrel{\triangle}{=} \{\ \ell \in \mathcal{L}_\tau^*;\ \ell \geq 0,\ \langle \ell, \mathbf{1}\rangle = 1\ \}.$$

It is endowed with $\mathcal{E}_\mathcal{Q}$: The σ-field induced by $\mathcal{E}$ on $\mathcal{Q}$. Note that $\mathcal{L}_\tau$, $\mathcal{Q}$, $\mathcal{E}$ and $\mathcal{E}_\mathcal{Q}$ depend on μ.

The rate function. The rate function of the extended Sanov theorem is

$$I(\ell) = \begin{cases} \int_\Sigma \log\left(\frac{d\ell^a}{d\mu}\right) d\ell^a + \sup_{f\in D_\mu} \langle \ell^s, f\rangle & \text{if } \ell \in \mathcal{Q}\cap L'_\tau \\ \infty & \text{otherwise.} \end{cases},$$

where $\ell = \ell^a + \ell^s$ is the decomposition stated in Theorem 2.3, $D_\mu = \{f \in \mathcal{L}_\tau;\ \mathbb{E}\, e^{f(Y)} < \infty\}$ and $\mathbb{E}$ stands for the expectation with respect to μ.

Remark 3.1. Due to (3.1), the set $\mathcal{Q}\cap L'_\tau$ is well-defined.

Definition 3.1. *The above rate function $I(\ell)$ is the* extended relative entropy *of ℓ with respect to μ.*

We shall denote $I(\ell) = I_a(\ell) + I_s(\ell) = I_a(\ell^a) + I_s(\ell^s)$ where:

$$\begin{aligned} I_a(\ell) &= \int_\Sigma \log\big(\frac{d\ell^a}{d\mu}\big)\, d\ell^a, \\ I_s(\ell) &= \sup\{\langle \ell^s, f\rangle;\ f, \mathbb{E}\, e^{f(Y)} < \infty\}. \end{aligned}$$

where ℓ^a and ℓ^s are the absolutely continuous and singular parts of ℓ (see Theorem 2.3). The following theorem is the main result of the section.

Theorem 3.2. (Extended Sanov theorem). *The empirical measures $\{L_n^Y\}_{n\geq 1}$ satisfy the LDP in $\mathcal{Q}$ endowed with the σ-field $\mathcal{E}_\mathcal{Q}$ and the topology $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ with the rate function I. This means that*

(1) *for all measurable closed subset F of $\mathcal{Q}$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n^Y \in F) \leq -\inf_{\ell\in F} I(\ell),$$

(2) *for all measurable open subset G of $\mathcal{Q}$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n^Y \in G) \geq -\inf_{\ell\in G} I(\ell).$$

Moreover, I is convex and is a good rate function: Its sublevel sets $\{I \leq \alpha\}$, $\alpha \geq 0$, are compact.

Proof. The LDP in $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$ is proved in Lemma 3.4 with the rate function Θ^* which is convex as a convex conjugate. By Proposition 3.8, $\Theta^* = I$. But the domain of I is included in $\mathcal{Q}$, thus the LDP holds on $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ (for this argument see [8], Lemma 4.1.5(b)). This completes the proof of the theorem. $\square$

Remark 3.2. Let $I(\ell) < \infty$, then $\ell \in L'_\tau$, $\ell^a \geq 0$ and by Proposition 2.4, $\langle \ell^a, \mathbf{1}\rangle = 1$. Hence $\frac{d\ell^a}{d\mu}$ is a probability density and I_a is close to the usual relative entropy $H(\cdot \mid \mu)$. The difference lies in the fact that I_a is defined over $\mathcal{L}_\tau^*$ whereas H is defined over $\mathcal{P}$.

Remark 3.3. The trace of the decomposition of L'_τ into absolutely continuous and singular components (Theorem 2.3) on $\mathcal{Q}$ is $\mathcal{Q}\cap L'_\tau = \Big(L_{\tau^*}\cap\mathcal{P}\Big)\oplus(L_\tau^s\cap\{\ell \geq 0\})$.

Remark 3.4. We cannot expect the LDP with the good rate function $H(\cdot \mid \mu)$. Indeed, A. Schied proved in [20] that if the topology is too wide (the so-called τ_ϕ-topology where ϕ admit only some exponential moments), then $\{H \leq \alpha\}$ is no longer compact. The same argument holds in our context: $\{\ell \in \mathcal{Q}; H(\ell^a) \leq \alpha\} = \{f\mu; f \in L_{\tau^*}, f \geq 0, \int_\Sigma f\, d\mu = 1, \int_\Sigma f \ln f\, d\mu \leq \alpha\} + (L_\tau^s \cap \{\ell \geq 0\})$ which is not compact.

Let $\mathcal{P}_\tau$ denote the set of all probability measures which integrate all functions in $\mathcal{M}_\tau$: $\mathcal{P}_\tau = \{\nu \in \mathcal{P}; \int_\Sigma |f|\, d\nu < \infty, \forall f \in \mathcal{M}_\tau\}$. Let us endow it with the σ-field $\sigma\big(\nu \mapsto \int_\Sigma f\, d\nu; f \in \mathcal{M}_\tau\big)$ and with the topology $\sigma(\mathcal{P}_\tau, \mathcal{M}_\tau)$.

Corollary 3.3. *The empirical measures* $\{L_n^Y\}_{n\geq 1}$ *satisfy the LDP in* $(\mathcal{P}_\tau, \sigma(\mathcal{P}_\tau, \mathcal{M}_\tau))$ *with the good rate function* $H(\cdot \mid \mu)$.

This is in accordance with the result obtained by P. Eichelsbacher and U. Schmock ([10], Theorem 1.8).

Proof. This is a direct consequence of the contraction principle applied to the transformation $\ell \in \mathcal{Q} \to \ell_{|\mathcal{M}_\tau} \in \mathcal{M}_\tau^*$. Indeed, by Proposition 2.4 we have $\ell_{|\mathcal{M}_\tau} = \ell^a_{|\mathcal{M}_\tau} = \ell^a$ (where the last equality is an identification). Hence, $\inf\{I(\ell); \ell_{|\mathcal{M}_\tau} = \nu\} = \inf\{I_a(\nu) + I_s(\ell^s); \ell^a = \nu\} = I_a(\nu) = H(\nu \mid \mu)$. The result follows from the obvious continuity of the considered transformation. $\square$

3.2. **Proof of the LDP.** Lemma 3.4 below states the LDP with the rate function Θ^* expressed as the convex conjugate of

$$\Theta(f) = \log \mathbb{E}\, e^{f(Y)} = \log \int_\Sigma e^f\, d\mu \in (-\infty, \infty],\ \ f \in \mathcal{L}_\tau.$$

Lemma 3.4. *The empirical measures* $\{L_n^Y\}_{n\geq 1}$ *satisfy the LDP (in the sense of Theorem 3.2) in* $\mathcal{L}_\tau^*$ *endowed with the* σ*-field* $\mathcal{E}$ *and the topology* $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$ *with the good rate function* $\Theta^*(\ell) = \sup_{f\in\mathcal{L}_\tau}\{\langle \ell, f\rangle - \Theta(f)\}$.

Proof. It is based on Dawson-Gärtner's projective limit approach. By Theorem 4.6.9 in [8], it is sufficient to check that for all $d \geq 1$ and $f_1, \dots, f_d \in \mathcal{L}_\tau$, $(\langle L_n^Y, f_1\rangle, \dots, \langle L_n^Y, f_d\rangle)$ satisfies a LDP.
But $(\langle L_n^Y, f_1\rangle, \dots, \langle L_n^Y, f_d\rangle) = \frac{1}{n}\sum_{i=1}^n (f_1(Y_i), \dots, f_d(Y_i)) = \frac{1}{n}\sum_{i=1}^n \vec{f}(Y_i)$ where $\vec{f}(x) = \big(f_1(x), \dots, f_d(x)\big)$ is a $\mathbb{R}^d$-valued function. Then, $\{\vec{f}(Y_i)\}$ is a sequence of i.i.d., $\mathbb{R}^d$-valued random variables. Since $f_1, \dots, f_d \in \mathcal{L}_\tau$, $\vec{f}(Y_i)$ admits exponential moments. By Cramér's theorem in Polish spaces (Theorem 6.1.3 and Corollary 6.1.6 in [8]), $\frac{1}{n}\sum_{i=1}^n \vec{f}(Y_i)$ satisfies the LDP in $\mathbb{R}^d$ with the good rate function

$$I_d(x) = \sup_{\lambda\in\mathbb{R}^d}\{\lambda\cdot x - \log \mathbb{E}\, e^{\lambda\cdot\vec{f}(Y)}\}, \qquad x \in \mathbb{R}^d.$$

By Dawson-Gärtner's theorem, $L_n^Y = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ satisfies the LDP with the good rate function, given for any $\ell \in \mathcal{L}_\tau^*$, by:
$\sup\{\sum_{i=1}^d \lambda_i\langle \ell, f_i\rangle - \log \mathbb{E}\, e^{\sum \lambda_i f_i(Y)};\ d \geq 1,\ \lambda \in \mathbb{R}^d,\ f_1, \dots, f_d \in \mathcal{L}_\tau\} = \sup_{f\in\mathcal{L}_\tau}\{\langle \ell, f\rangle - \log \mathbb{E}\, e^{f(Y)}\} = \Theta^*(\ell)$, which is the desired result. $\square$

3.3. **Identification of the rate function.** In this section, we prove the identity $I = \Theta^*$ in Proposition 3.8. Lemmas 3.5, 3.6 and 3.7 are preliminary results for the proof of this identity.
Let us consider $J(\ell) = \sup_{f \in \mathcal{L}_\tau} \left\{ \langle \ell - \mu, f \rangle - \int_\Sigma \gamma(f)\, d\mu \right\}$, where γ is given by (2.3).

Lemma 3.5. *Let $\ell \in \mathcal{L}_\tau^*$, then*
(1) $\Theta^*(\ell) < \infty \Rightarrow \ell \in L_\tau^*$,
(2) $J(\ell) \leq \Theta^*(\ell)$,
(3) $\Theta^*(\ell) < \infty \Rightarrow \ell \in \mathcal{Q} \cap L'_\tau$.

Proof. 1. Let $f \in \mathcal{L}_\tau$ be such that $f = 0$ μ-a.e. Then, for any real λ, $\Theta(\lambda f) = 0$ and $\lambda \langle \ell, f \rangle \leq \Theta^*(\ell) + \Theta(\lambda f) = \Theta^*(\ell)$. Hence, $\Theta^*(\ell) < \infty$ implies $\langle \ell, f \rangle = 0$. Thus ℓ is constant over the equivalence classes, that is $\ell \in L_\tau^*$.
2. As for all $t \geq 0$, $\log t \leq t - 1$, we have: $-\mathbb{E}\, e^{f(Y)} + 1 \leq -\log \mathbb{E}\, e^{f(Y)}$ and

$$J(\ell) = \sup_{f \in \mathcal{L}_\tau} \left\{ \langle \ell, f \rangle - \int_\Sigma (e^f - 1)\, d\mu \right\} \leq \Theta^*(\ell).$$

3. As $\Theta^*(\ell) < \infty$ implies $\ell \in L_\tau^*$, let us consider $\ell \in L_\tau^*$.
For all $f \in L_\tau$, $\langle \ell - \mu, f \rangle \leq J(\ell) + \int_\Sigma \gamma(f)\, d\mu$. As $\gamma(s) \leq \tau(s) = \gamma(|s|)$, we have $\langle \ell - \mu, f \rangle \leq J(\ell) + \int_\Sigma \tau(f)\, d\mu$. Choosing $\eta = \pm 1/\|f\|_\tau$ when $f \neq 0$, the definition of the Luxemburg norm (2.1) yields $\int_\Sigma \tau(\eta f)\, d\mu = 1$ which implies $|\langle \ell - \mu, f \rangle| \leq (J(\ell) + 1)\|f\|_\tau$. This inequality still holds with $\|f\|_\tau = 0$. Therefore $\ell - \mu \in L_\tau^*$ is $\|\cdot\|_\tau$-continuous: $\ell \in L'_\tau$, since $J(\ell) \leq \Theta^*(\ell) < \infty$.
Suppose that $\langle \ell, \mathbf{1} \rangle = a \neq 1$. Then $\Theta^*(\ell) \geq \langle \ell, \lambda \mathbf{1} \rangle - \log \mathbb{E}\, e^{\lambda \mathbf{1}} = \lambda(a - 1)$ which tends to ∞ as λ tends to infinity with the sign of $a - 1$. Therefore, $\langle \ell, \mathbf{1} \rangle = 1$ if $\Theta^*(\ell) < \infty$.
Suppose now that there exists $f \geq 0$ with $\langle \ell, f \rangle < 0$. Let $\lambda \geq 0$, then $\Theta^*(\ell) \geq \langle \ell, -\lambda f \rangle - \log \mathbb{E}\, e^{-\lambda f} \geq \langle \ell, -\lambda f \rangle$ tends to ∞ as λ tends to ∞. Thus $\Theta^*(\ell) < \infty$ implies $\ell \geq 0$ and Lemma 3.5 is proved. □

Lemma 3.6. *Let $\ell \in L'_\tau$. Then, for all $f \in L_\tau$*
(1) $\lim_{n \to \infty} \langle \ell, f_n \rangle = \langle \ell^a, f \rangle$ *where (f_n) is any sequence of bounded measurable functions which converges pointwise to f and such that $|f_n| \leq |f|$, for all $n \geq 1$,*
(2) $\lim_{n \to \infty} \langle \ell, \mathbf{1}_{\{|f| > n\}} f \rangle = \langle \ell^s, f \rangle$.

Proof. 1. Since f_n is bounded, $\langle \ell^s, f_n \rangle = 0$ (see Proposition 2.4). Therefore $\langle \ell, f_n \rangle = \langle \ell^a, f_n \rangle = \int f_n \frac{d\ell^a}{d\mu}\, d\mu$ with $\frac{d\ell^a}{d\mu} \in L_{\tau^*}$. The limit follows from the dominated convergence theorem.
2. We have $\langle \ell, \mathbf{1}_{\{|f| > n\}} f \rangle = \langle \ell^s, \mathbf{1}_{\{|f| > n\}} f \rangle + \langle \ell^a, \mathbf{1}_{\{|f| > n\}} f \rangle$. The dominated convergence theorem implies that $\lim_{n \to \infty} \langle \ell^a, \mathbf{1}_{\{|f| > n\}} f \rangle = 0$ and since $\mathbf{1}_{\{|f| \leq n\}} f$ is bounded, $\langle \ell^s, f \rangle = \langle \ell^s, \mathbf{1}_{\{|f| > n\}} f \rangle$ (see Proposition 2.4). □

Lemma 3.7. *For all ℓ in L'_τ, $\Theta^*(\ell) = \Theta^*(\ell^a) + \sup\{\langle \ell^s, f\rangle;\ f \in \mathrm{dom}\,\Theta\}$ where* $\mathrm{dom}\,\Theta = \{f \in \mathcal{L}_\tau, \Theta(f) < \infty\}$ *is the effective domain of* Θ.

Remark 3.5. Clearly, $\mathrm{dom}\,\Theta = D_\mu$.

Proof. We first introduce some notations which are customary in convex analysis. Let A be a convex subset of L_τ and let ℓ be in L'_τ. The convex indicator function of A is $\delta(f|A) = \begin{cases} 0 & \text{if } f \in A \\ +\infty & \text{otherwise} \end{cases}$ and its convex conjugate $\delta^*(\ell|A) = \sup_{f\in L_\tau}\{\langle \ell, f\rangle - \delta(f|A)\} = \sup_{f\in A}\langle \ell, f\rangle$ is called the support functional of A. For any $\ell \in L'_\tau$, we have

$$\begin{aligned}\Theta^*(\ell) &= \sup_{f\in L_\tau}\{\langle \ell^a, f\rangle - \Theta(f) + \langle \ell^s, f\rangle - \delta(f|\mathrm{dom}\,\Theta)\} \\ &\leq \Theta^*(\ell^a) + \delta^*(\ell^s|\mathrm{dom}\,\Theta).\end{aligned}$$

To prove the converse, let $f, g \in L_\tau$. For $n \geq 1$, define $u_n = f_n + g\mathbf{1}_{\{|g|>n\}}$ with $f_n = (-n \vee f \wedge n)\mathbf{1}_{\{|f|\leq n\}}$. Then

$$\Theta^*(\ell) \geq \langle \ell, u_n\rangle - \Theta(u_n) = \langle \ell, f_n\rangle - \Theta(u_n) + \langle \ell, g\mathbf{1}_{\{|g|>n\}}\rangle.$$

Since $e^{u_n} \leq 1 + e^f + e^g$, it follows from the dominated convergence theorem that $\Theta(u_n) \to \Theta(f)$. Hence, $\Theta^*(\ell) \geq \langle \ell^a, f\rangle - \Theta(f) + \langle \ell^s, f\rangle$ by Lemma 3.6. This completes the proof of the proposition. □

Proposition 3.8. *The identity $\Theta^* = I$ holds on $\mathcal{L}^*_\tau$.*

Proof. By Lemma 3.5, the effective domain of Θ^* is included in $\mathcal{Q} \cap L'_\tau$ and by Lemma 3.7, for all $\ell \in L'_\tau$, $\Theta^*(\ell) = \Theta^*(\ell^a) + I_s(\ell^s)$. Taking Remark 3.3 into account, it remains to prove that for all $\ell \in \mathcal{P} \cap L_{\tau^*}$, $\Theta^*(\ell) = H(\ell \mid \mu)$. Let $\ell = h\mu$ belong to $\mathcal{P} \cap L_{\tau^*}$, that is $h \in L_{\tau^*}$, $h \geq 0$, $\int_\Sigma h\,d\mu = 1$. By a direct computation, we have, for all $f \in \mathrm{dom}\ \Theta$,

$$\Theta(f) = \inf_{\lambda\in\mathbb{R}}\left\{-\lambda - 1 + e^\lambda \int_\Sigma e^f\,d\mu\right\}. \tag{3.2}$$

Therefore,

$$\begin{aligned}
\Theta^*(h\mu) &= \sup_{f\in\mathrm{dom}\,\Theta}\{\langle h\mu, f\rangle - \Theta(f)\}\\
&\overset{(a)}{=} \sup_{\lambda\in\mathbb{R}, f\in\mathrm{dom}\,\Theta}\left\{\langle h\mu, f\rangle + \lambda + 1 - e^{\lambda}\int_\Sigma e^f\,d\mu\right\}\\
&\overset{(b)}{=} \sup_{\lambda\in\mathbb{R}, f\in\mathrm{dom}\,\Theta}\left\{\langle h\mu, \lambda + f\rangle - \int_\Sigma (e^{\lambda+f}-1)\,d\mu\right\}\\
&= \sup_{g\in\mathrm{dom}\,\Theta}\left\{\int_\Sigma hg\,d\mu - \int_\Sigma (e^{g}-1)\,d\mu\right\}\\
&\overset{(c)}{=} \int_\Sigma (h\log h - h + 1)\,d\mu\\
&\overset{(d)}{=} \int_\Sigma h\log h\,d\mu = H(h\mu \mid \mu),
\end{aligned}$$

where (a) comes from (3.2), (b) and (d) follow from the fact that μ and $h\mu$ are probability measures and (c) follows from a general result of R. T. Rockafellar ([18], Theorem 2), noting that the convex conjugate of $e^s - 1$ is $t\log t - t + 1$. This completes the proof of the proposition. □

3.4. **Csiszár's example.** In this section, we encounter a minimizer of the extended relative entropy under a linear constraint with a non null singular part. We deal here with a probability distribution μ which already appears in ([6], Example 3.2) and in ([8], Exercise 7.3.11). Let μ be the probability measure on $\Sigma = [0,\infty)$ defined by $\mu(dy) = c\frac{e^{-y}}{1+y^3}\,dy$. Let $\{Y_i\}$ be a sequence of i.i.d. $[0,\infty)$-valued random variables, with distribution μ. We consider:

$$L_n^Y = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i} \quad \text{and} \quad \hat{S}_n = \frac{1}{n}\sum_{i=1}^n Y_i.$$

By the usual Sanov theorem, L_n^Y satisfies the LDP with the good rate function $H(\cdot \mid \mu)$ in $(\mathcal{P}, \sigma(\mathcal{P}, B))$. By Cramér's theorem, $\hat{S}_n$ also satisfies the LDP with the good rate function $\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}\{x\lambda - \Lambda(\lambda)\}$, where $\Lambda(\lambda) = \log\int_{[0,\infty)} c\frac{e^{(\lambda-1)y}}{1+y^3}\,dy$. One can ask if the Contraction Principle (CP) holds between L_n^Y and $\hat{S}_n$. Let us denote $u : [0,\infty) \to [0,\infty)$, $u(y) = y$, so that $\langle L_n^Y, u\rangle = \hat{S}_n$. As u is not bounded, one cannot apply the CP to the usual Sanov theorem to obtain:

$$\inf\{H(\nu \mid \mu),\, \nu\in\mathcal{P},\, \langle \nu, u\rangle = x\} = \Lambda^*(x), \quad x \geq 0. \tag{3.3}$$

It turns out that this equality holds (see Proposition 3.10 below) but that the infimum is not attained in $\mathcal{P}$ when x is large. With the notations of Section 2.3, u belongs to $\mathcal{L}_\tau(\mu)$. Therefore, $G_u : \mathcal{L}_\tau^*(\mu) \to \mathbb{R}$, $G_u(\ell) = \langle \ell, u\rangle$ is a

$\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$-continuous linear form. One can notice here the advantage of using the $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$-topology which is wider than the $\sigma(\mathcal{P}, B)$-topology. By Theorem 3.2 and the CP, $G_u(L_n^Y) = \frac{1}{n}\sum_{i=1}^n Y_i = \hat{S}_n$ satisfies the LDP with the good rate function

$$I'(x) = \inf\{I(\ell); \ell \in \mathcal{Q}; \langle \ell, u\rangle = x\}.$$

As I is a good rate function, there exists at least one minimizing argument $\ell_x \in \mathcal{Q}$ satisfying $I(\ell_x) = I'(x)$ and $\langle \ell_x, u\rangle = x$. By the uniqueness of the rate function ([8], Lemma 4.1.4), $I' = \Lambda^*$. Therefore, the following identification holds:

$$\Lambda^*(x) = \inf\{I(\ell); \ell \in \mathcal{Q}; \langle \ell, u\rangle = x\} = I(\ell_x), \tag{3.4}$$

for some $\ell_x \in \mathcal{Q}$ satisfying $\langle \ell_x, u\rangle = x$.

Proposition 3.9. *Let us denote $x_* = \Lambda'(1^-)$.*

(1) *For any $0 < x < x_*$, there exists a unique minimizer ℓ_x in (3.4). It is given by $\ell_x = \nu_x$ where*

$$\nu_x(dy) = \exp(\lambda_x y - \Lambda(\lambda_x))\mu(dy)$$

and $\lambda = \lambda_x$ is the unique solution of $\Lambda'(\lambda) = x$.

(2) *For $x = x_*$, the statement (1) still holds with $\lambda_{x_*} = 1$, $\Lambda'(1^-) = x_*$ and $\nu_{x_*} = \nu_*$ given by*

$$\nu_*(dy) = e^{y-\Lambda(1)}\,\mu(dy) = \frac{c'}{1+y^3}\,dy.$$

(3) *For all $x \geq x_*$ and all minimizing arguments ℓ_x of (3.4), we have $\ell_x^a = \nu_*$.*
Moreover, $\langle \nu_*, u\rangle = x_*$, $\langle \ell_x^s, u\rangle = x - x_*$ and

$$I(\ell_x) = H(\nu_*|\mu) + I_s(\ell_x^s), \tag{3.5}$$

with $H(\nu_*|\mu) = \Lambda^*(x_*)$ and $I_s(\ell_x^s) = x - x_*$.

This proposition means that when $x > x_*$, the minimizers of (3.4) cannot be probability measures. The contribution of the absolutely continuous part is stopped at x_*: $\langle \nu_*, u\rangle = x_*$. It is the singular parts (not unique in general, see Prop. 4.4) which fill the gap between x_* and x: $\langle \ell_x^s, u\rangle = x - x_*$. Moreover the contribution of these singular parts appear in the rate function (see (3.5)). Finally, $I_s(\ell^s) = \sup_{f\in D_\mu}\langle \ell^s, f\rangle$ implies that these singular parts are non-null, whenever $x > x_*$.

Proof of Proposition 3.9. Proof of 1. and 2. Clearly, for all $0 < x \leq x_*$, we have $\langle \nu_x, u\rangle = x$. Let ℓ be such that $\langle u, \ell\rangle = x$ and (without loss of generality)

$I(\ell) < \infty$. Then, $\ell = \ell^a + \ell^s$ with $\ell^a \in \mathcal{P}$ and $\ell^s \geq 0$. Let us denote: $\langle \ell^a, u\rangle = x'$. We have $\langle \ell^s, u\rangle = x - x' \geq 0$ and

$$\begin{aligned} I(\ell) - I(\nu_x) &= H(\ell^a \mid \mu) - H(\nu_x \mid \mu) + I(\ell^s) \\ &= H(\ell^a \mid \nu_x) + \int \log\left(\frac{d\nu_x}{d\mu}\right) d(\ell^a - \nu_x) + I(\ell^s) \\ &= H(\ell^a \mid \nu_x) + I(\ell^s) - \lambda_x(x - x') \\ &\overset{(a)}{\geq} H(\ell^a \mid \nu_x) + \langle \ell^s, u\rangle - \lambda_x(x - x') \\ &\overset{(b)}{\geq} H(\ell^a \mid \nu_x) \\ &\geq 0, \end{aligned}$$

where (a) follows from $I(\ell^s) = \sup\{\langle \ell^s, v\rangle; v \in D_\mu\} \geq \langle \ell^s, u\rangle$ and (b) follows from $\langle \ell^s, u\rangle - \lambda_x(x - x') = (1 - \lambda_x)(x - x') \geq 0$.

For the equality to hold, it is necessary that $\ell^a = \nu_x$. Hence, $I(\ell^s) = 0$ which in turn implies that $\ell^s = 0$. Finally, ν_x is the unique minimizer of (3.4).

Proof of 3.: For all $\lambda \leq 1$, both $\Lambda(\lambda)$ and $\Lambda'(\lambda)$ are finite, while $\Lambda(\lambda) = \infty$, when $\lambda > 1$. As $\Lambda'(1^-) = x_*$ is finite, Λ is not steep. Standard convexity arguments lead to $\Lambda^*(x_*) = x_* - \Lambda(1)$ and an easy computation yields

$$\Lambda^*(x) = \Lambda^*(x_*) + x - x_*,\ x \geq x_*. \tag{3.6}$$

The rest of the proof is divided into three steps.

Step 1: Let ℓ_x (resp. ℓ_y) be any minimizing argument of $I'(x)$ (resp. $I'(y)$): $I(\ell_x) = I'(x) = \inf\{I(\ell), \langle \ell, u\rangle = x\}$. Let us prove that for all $0 \leq \alpha, \beta \leq 1$, $\alpha + \beta = 1$, the following identity holds:

$$\forall x, y \geq x_*, \quad I(\alpha\ell_x + \beta\ell_y) = \alpha I(\ell_x) + \beta I(\ell_y). \tag{3.7}$$

By definition of ℓ_x and ℓ_y, we get $\langle \ell_x, u\rangle = x$ and $\langle \ell_y, u\rangle = y$ and by (3.6), $I(\ell_x) = \Lambda^*(x) = (x - x_*) + \Lambda^*(x_*)$. Similarily, $I(\ell_y) = (y - x_*) + \Lambda^*(x_*)$. The convexity of I implies that

$$\begin{aligned} I(\alpha\ell_x + \beta\ell_y) &\leq \alpha I(\ell_x) + \beta I(\ell_y) = \Lambda^*(x_*) + \alpha x + \beta y - x_* \\ &= \Lambda^*(\alpha x + \beta y) = I'(\alpha x + \beta y). \end{aligned}$$

But $I'(\alpha x + \beta y) = \inf\{I(\ell); \ell \in L'_\tau; \langle \ell, u\rangle = \alpha x + \beta y\}$ and $\alpha\ell_x + \beta\ell_y$ satisfies the constraint $\langle \alpha\ell_x + \beta\ell_y, u\rangle = \alpha x + \beta y$. Thus $I'(\alpha x + \beta y) \leq I(\alpha\ell_x + \beta\ell_y)$ and (3.7) holds.

Step 2: Let us show that for any $x, y \geq x_*$ and ℓ_x (resp. ℓ_y) minimizing argument of $I'(x)$ (resp. $I'(y)$), we have $\ell_x^a = \ell_y^a \overset{\triangle}{=} \nu$ where $\ell_x = \ell_x^a + \ell_x^s$ (resp. $\ell_y = \ell_y^a + \ell_y^s$).

By the definition of I, we have:

$$\begin{array}{rcl} I(\alpha\ell_x+\beta\ell_y) & = & I_a(\alpha\ell_x^a+\beta\ell_y^a)+I_s(\alpha\ell_x^s+\beta\ell_y^s),\\ \alpha I(\ell_x)+\beta I(\ell_y) & = & \alpha I_a(\ell_x^a)+\beta I_a(\ell_y^a)+\alpha I_s(\ell_x^s)+\beta I_s(\ell_y^s). \end{array}$$

The convexity of I_a and I_s implies

$$\left\{ \begin{array}{ll} I_a(\alpha\ell_x^a+\beta\ell_y^a) & \leq \alpha I_a(\ell_x^a)+\beta I_a(\ell_y^a),\\ I_s(\alpha\ell_x^s+\beta\ell_y^s) & \leq \alpha I_s(\ell_x^s)+\beta I_s(\ell_y^s). \end{array}\right.$$

By (3.7), $I(\alpha\ell_x+\beta\ell_y)=\alpha I(\ell_x)+\beta I(\ell_y)$. Therefore, equality must hold in the two previous inequalities. But due to the strict convexity of I_a, $I_a(\alpha\ell_x^a+\beta\ell_y^a)=\alpha I_a(\ell_x^a)+\beta I_a(\ell_y^a)$ implies that $\ell_x^a=\ell_y^a=\nu$.

STEP 3: Let us show that for all $x\geq x_*$, $I_a(\ell_x^a)=\Lambda^*(x_*)$ and $I_s(\ell_x^s)=x-x_*$. Considering $\nu_*(dy)=e^{y-\Lambda(1)}\mu(dy)$, one shows that $\langle\nu_*,u\rangle=x_*$, $I(\nu_*)=I_a(\nu_*)=\Lambda^*(x_*)$. Hence, ν_* satisfies (3.4) at x_*. Thus, $\nu_*=\nu$ and for all $x\geq x_*$, $\ell_x^a=\nu_*$. It follows that $I_a(\ell_x^a)=\Lambda^*(x_*)$ and $I_s(\ell_x^s)=x-x_*$. □

Proposition 3.10. *The equality (3.3) holds for all $x\geq 0$.*

Remark 3.6. In the proof below, we show that $\nu_n=(1-\frac{1}{n})\nu_*+\frac{1}{n}\frac{\mathbf{1}_{I_n}}{\mu(I_n)}\mu$ is a sequence satisfying

$$H(\nu_n|\mu)>\Lambda^*(x),\quad \langle\nu_n,\mu\rangle=x\quad\text{and}\quad\lim_{n\to\infty}H(\nu_n|\mu)=\Lambda^*(x).$$

In Proposition 3.9 (3), it is shown that the minimizers ℓ_x have the form $\ell_x=\nu_*+\ell_x^s$. Therefore, $\frac{1}{n}\frac{\mathbf{1}_{I_n}}{\mu(I_n)}\mu$ contributes asymptotically to ℓ_x^s in the sense that

$$\lim_{n\to\infty}\langle\frac{1}{n}\frac{\mathbf{1}_{I_n}}{\mu(I_n)}\mu,u\rangle=\langle\ell_x^s,u\rangle=x-x_*.$$

Proof of Proposition 3.10. For $x=0$, $\Lambda^*(0)=\infty$ and there is no $\nu\in\mathcal{P}$ such that $\nu\ll\mu$ and $\langle\nu,u\rangle=0$. For $0<x\leq x_*$, the desired equality is a consequence of Proposition 3.9, (1) and (2). Let us now consider the case $x>x_*$. First note that

$$\begin{array}{l} \inf\{H(\nu\mid\mu),\,\nu\in\mathcal{P},\,\langle\nu,u\rangle=x,\}\\ \quad=\;\inf\{I(\nu),\,\nu\in\mathcal{P},\,\langle\nu,u\rangle=x\}\\ \quad\geq\;\inf\{I(\ell),\;\ell\in\mathcal{Q},\,\langle\ell,u\rangle=x\}=\Lambda^*(x). \end{array}$$

In particular, $H(\nu|\mu)\geq\Lambda^*(x)$ if $\langle\nu,u\rangle=x$ (in fact, $>$ by Proposition 3.9, (3)). Therefore, it is sufficient to exhibit a minimizing sequence (ν_n) satisfying $\nu_n\in\mathcal{P}$, $\langle\nu_n,u\rangle=x$ and $\lim_{n\to\infty}H(\nu_n\mid\mu)=\Lambda^*(x)$. We take it of the form

$$\nu_n=(1-\frac{1}{n})\nu_*+\frac{1}{n}\frac{\mathbf{1}_{I_n}}{\mu(I_n)}\mu,$$

where the interval I_n must be chosen such that $\langle \nu_n, u\rangle = x$. As $\langle \nu_*, u\rangle = x_*$, I_n must satisfy

$$\frac{\int_{I_n} u\,d\mu}{\mu(I_n)} = x_* + n(x - x_*).$$

Consider $I_n(t) = [x_* + n(x - x_*) - t, x_* + n(x - x_*) + 1]$ and

$$\phi(t) = \frac{\int_{I_n(t)} u\,d\mu}{\mu(I_n(t))} = \frac{\int_{I_n(t)} yf(y)\,dy}{\mu(I_n(t))},$$

where f is μ's density. Simple computations yield

$$\phi(0) > x_* + n(x - x_*) \quad \text{and} \quad \phi(1) < x_* + n(x - x_*).$$

As ϕ is continuous, there exists $\alpha_n \in [0,1]$ such that $\phi(\alpha_n) = x_* + n(x - x_*)$. We denote by $I_n \triangleq I_n(\alpha_n)$. We now estimate $H(\nu_n|\mu)$.

$$\begin{aligned} H(\nu_n \mid \mu) &= \int_{[0,\infty)\setminus I_n} \log\left(\frac{d\nu_n}{d\mu}\right) d\nu_n + \int_{I_n} \log\left(\frac{d\nu_n}{d\mu}\right) d\nu_n \\ &\le H(\nu_* \mid \mu) + \int_{I_n} \log\left(\frac{d\nu_n}{d\mu}\right) d\nu_n = \Lambda^*(x_*) + \int_{I_n} \log\left(\frac{d\nu_n}{d\mu}\right) d\nu_n, \end{aligned}$$

and

$$\frac{d\nu_n}{d\mu}(y) = e^{y-\Lambda(1)}(1 + \frac{1}{n}) + \frac{1_{I_n}(y)}{n\mu(I_n)}. \tag{3.8}$$

We shall use the following inequality.

$$(a+b)\log(a+b) \le (a\log a)\left(1 + \frac{b}{a}\right)^2, \quad \forall a \ge e, b > 0. \tag{3.9}$$

Expressions (3.8) and (3.9) yield

$$\begin{aligned} &\int_{I_n} \log\left(\frac{d\nu_n}{d\mu}\right) d\nu_n \\ &\quad \le \int_{I_n} \frac{1}{n\mu(I_n)} \log\left(\frac{1}{n\mu(I_n)}\right) \left(1 + e^{y-\Lambda(1)}(1+\frac{1}{n})n\mu(I_n)\right)^2 d\mu(y). \end{aligned}$$

But if $y \in I_n$ then $e^{y-\Lambda(1)}(1+\frac{1}{n})n\mu(I_n) \xrightarrow[n\to\infty]{} 0$. On the other hand, $\lim_{n\to\infty} \frac{1}{n}\log\frac{1}{n\mu(I_n)} = x - x_*$. Consequently,

$$\lim_{n\to\infty} \int_{I_n} \frac{1}{n\mu(I_n)} \log\left(\frac{1}{n\mu(I_n)}\right) \left(1 + e^{y-\Lambda(1)}(1+\frac{1}{n})n\mu(I_n)\right)^2 d\mu(y) = x - x_*,$$

and the proof is completed. $\square$

4. The Gibbs Conditioning Principle

In this section, we apply Theorem 3.2 to derive the following Gibbs conditioning principle:

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{P}\Big((Y_1,\dots,Y_k)\in\cdot\mid L_n^Y\in A_\delta\Big)=\nu_*^k(\cdot).$$

This is stated in Theorem 4.2, the main result of the section. This result holds true without any underlying law of large numbers. Namely, the following equation might not hold:

$$\lim_{n\to\infty}\nu_*^n\{L_n^Y\in A_\delta\}=1\quad\text{for all }\delta>0,\tag{4.1}$$

as shown in Remark 4.5. Moreover, ν_* might not belong to the minimizers of the set $\{I(\ell);\ \ell\in A_0\}$ where A_0 and $(A_\delta)_{\delta>0}$ are subset of $\mathcal{Q}$ specified in Assumptions (A-0) and (A-1) below.

These features are illustrated via Csiszár's example in Section 4.4. In our presentation, we shall closely follow the framework of Section 7.3.5 in [8].

4.1. **Notations and statement of the assumptions.**

As before, let us consider $L_n^Y=\frac{1}{n}\sum_{i=1}^n\delta_{Y_i}\in\mathcal{L}_\tau^*$ where $\{Y_i\}_{i\geq 1}$ is a sequence of Σ-valued independent and identically μ-distributed variables. Let μ^n be the product measure induced by μ on Σ^n and Q_n be the probability measure induced by μ^n on $(\mathcal{Q},\mathcal{E}_\mathcal{Q})$, where $\mathcal{Q}$ is equipped with the topology $\sigma(\mathcal{Q},\mathcal{L}_\tau)$ and its σ-field $\mathcal{E}_\mathcal{Q}$:

$$Q_n(A)=\mu^n\{L_n^y\in A\},\quad A\in\mathcal{E}_\mathcal{Q}.$$

We are interested in the limiting behaviour of the distribution of $(Y_1,\cdots,Y_k)$ under the conditioning constraint $\{L_n^Y\in A_\delta\}$, for $n\to\infty$ followed by $\delta\to 0$. We denote this distribution:

$$\mu^n_{Y^k|A_\delta}(\cdot)=\mu^n\Big((y_1,\dots,y_k)\in\cdot\mid L_n^y\in A_\delta\Big).\tag{4.2}$$

In case $k=1$, we write $\mu^n_{Y^k|A_\delta}=\mu^n_{Y|A_\delta}$. We follow D. W. Strook and O. Zeitouni in [21] by considering the constraint set $\{L_n^Y\in A_\delta\}$ rather than $\{L_n^Y\in A_0\}$ where A_δ is a blow-up of A_0. By Assumption (A-1) below, A_δ must satisfy $Q_n(A_\delta)>0$ whereas A_0 may be a Q_n-negligible set. The following conventions prevail in this section: $\Gamma_n=\{L_n^Y\in\Gamma\}$, A° is the $\sigma(\mathcal{Q},\mathcal{L}_\tau)$-interior of A and $I(A)=\inf\{I(\ell);\ell\in A\}$ for all $A\subset\mathcal{Q}$.

Assumption A-0. *The set A_0 can be written $A_0=\bigcap_{\delta>0}A_\delta$, where $(A_\delta)_{\delta>0}$ is a family of nested measurable $\sigma(\mathcal{Q},\mathcal{L}_\tau)$-closed sets satisfying*

$$I(A_\delta^\circ)\leq I(A_0),\quad\textit{for all }\delta>0.\tag{4.3}$$

Two important cases where (4.3) is satisfied may be considered:

1. $A_0\subset A_\delta^\circ$, for all $\delta>0$.
2. $A_\delta=A_0$ for all $\delta>0$, and $I(A_0^\circ)=I(A_0)$.

Remark 4.1. The topology $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ and the σ-field $\mathcal{E}_\mathcal{Q}$ which appear in the statement of the assumption (A-0) are both wider than the usual τ-topology $\sigma(\mathcal{P}, B)$, and the σ-field $\mathcal{B}^{cy}$ (see ([8], Section 6.2) for its definition). Hence more open sets and more measurable sets are available. As an example, consider the family defined by

$$A_\delta = \{\ell \in \mathcal{Q};\ |\langle \ell, u\rangle - 1| \leq \delta\}, \qquad A_0 = \{\ell \in \mathcal{Q};\ \langle \ell, u\rangle = 1\},$$

where u satisfies condition (1.1). This family satisfies Assumption (A-0).

The following assumption is the counterpart of Assumption (A-1) in ([8], Section 7.3).

Assumption A-1. $I(A_0) < \infty$ *and for all* $\delta > 0$, $n \geq 1$, $Q_n(A_\delta) > 0$.

Remark 4.2. Equation (4.1) which is part of the assumptions in [8] (see also [10], Condition 1.16) enforces a law of large numbers under the minimizing law ν_* which appears in (1.2). This is not required in the present approach: By Assumption (A-0), one can apply Sanov's lower bound (see the proof of Lemma 4.3 below) so that no underlying law of large numbers is necessary. Moreover, there exist cases where (4.1) fails while (A-1) is still satisfied (see Remark 4.5 below).

4.2. **Convex constraints.** The set of minimizers is denoted by

$$\mathcal{M} \stackrel{\triangle}{=} \{\ell \in A_0;\ I(\ell) = I(A_0)\}.$$

The following result states that $\mathcal{M}$ has a special form when the constraint A_0 is convex.

Lemma 4.1. *Suppose that* A_0 *is convex, then*

$$\mathcal{M} = \nu_* + \mathcal{S},$$

where ν_* *is a probability measure and* $\mathcal{S}$ *is a set of singular parts. In other words, if* $\ell \in \mathcal{M}$*, then* $\ell = \nu_* + \ell^s$ *where* ν_* *is* ℓ*'s absolutely continuous part and* $\ell^s \in \mathcal{S}$ *is* ℓ*'s singular part.*

Remark 4.3. In this case, ν_* is the I-generalized projection of μ over the set of constraint A_0 in the sense of Csiszár (see [6]).

Proof of Lemma 4.1. Let ℓ and $\tilde{\ell}$ stand in $\mathcal{M}$. By the convexity of A_0 and I, we have

$$I(A_0) \leq I(\alpha\ell + \beta\tilde{\ell}) \leq \alpha I(\ell) + \beta I(\tilde{\ell}) = I(A_0),$$

for all $\alpha, \beta \geq 0$ such that $\alpha + \beta = 1$. Similarly, as I_a and I_s are convex, we obtain:

$$\begin{cases} I_a(\alpha\ell^a + \beta\tilde{\ell}^a) & \leq & \alpha I_a(\ell^a) + \beta I_a(\tilde{\ell}^a), \\ I_s(\alpha\ell^s + \beta\tilde{\ell}^s) & \leq & \alpha I_s(\ell^s) + \beta I_s(\tilde{\ell}^s). \end{cases}$$

Suppose that at least one of these inequalities is strict. We obtain by summing: $I(A_0) < \alpha I(\ell) + \beta I(\tilde{\ell}) = I(A_0)$ which is false. Hence, $I_a(\alpha\ell^a + \beta\tilde{\ell}^a) = \alpha I_a(\ell^a) + \beta I_a(\tilde{\ell}^a)$ and $I_s(\alpha\ell^s + \beta\tilde{\ell}^s) = \alpha I_s(\ell^s) + \beta I_s(\tilde{\ell}^s)$. As I_a is strictly convex, we obtain $\ell^a = \tilde{\ell}^a \stackrel{\triangle}{=} \nu_*$. As I_s is not strictly convex, ℓ^s and $\tilde{\ell}^s$ may differ. $\square$

We shall see in Section 4.4 that within the scope of Csiszár's example, $\mathcal{M} = \nu_* + \mathcal{S}$ where $\mathcal{S}$ is not reduced to a single point.

4.3. **Convergence of $\mu^n_{Y^k|A_\delta}$ to a probability distribution.** In this section, it is assumed that Σ is a separable metric space and that its σ-field is its Borel σ-field. Denote by $C_b(\Sigma^k)$ the set of continuous and bounded functions over Σ^k. Theorem 4.2 below is the counterpart of Corollary 7.3.5 in [8]. It is a corollary of Lemma 4.3.

Theorem 4.2. *Let us assume that (A-0), (A-1) hold, Σ is a separable metric space and the constraint set A_0 is convex. Then, for all f in $C_b(\Sigma^k)$, we have:*

$$\langle \mu^n_{Y^k|A_\delta}, f\rangle \to \langle \nu^k_*, f\rangle$$

for $n \to \infty$ followed by $\delta \to 0$ where ν_* is the common absolutely continuous part of the elements of $\mathcal{M}$ (see Lemma 4.1).

Theorem 4.2 improves Corollary 7.3.5 in [8] in two directions:

(1) The constraint sets A_δ can be based on functions with possibly infinite exponential moments, for instance
$$A_\delta = \{\ell \in \mathcal{Q}; |\langle \ell, u\rangle - 1| \le \delta\} \quad \text{with} \quad u \in \mathcal{L}_\tau.$$
Such functions u may grow quite fast.

(2) It has been previously remarked that the present Assumptions (A-0) and (A-1) do not require an underlying law of large numbers. For an illustration of the benefit, see Section 4.4 below and in particular Remark 4.5.

Remark 4.4. In the case $k = 1$, the convergence even holds for all $f \in M_\tau$, that is
$$\langle \mu^n_{Y|A_\delta}, f\rangle \to \langle \nu_*, f\rangle \quad \text{for} \quad f \in M_\tau.$$
However, this result relies on a finer estimate than Lemma 4.3 below. The estimate and the convergence theorem can be found in ([16], Lemma 2.10 and Theorem 2.11).

The following lemma is the counterpart of Theorem 7.3.3 in [8].

Lemma 4.3. *Assume (A-0) and (A-1). Then, $\mathcal{M}$ is a nonempty $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$-compact subset of $\mathcal{Q}$ and for any open measurable subset $\Gamma \in \mathcal{E}_{\mathcal{Q}}$ with $\mathcal{M} \subset \Gamma$, we have:*
$$\limsup_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mu^n\big(L^Y_n \notin \Gamma | L^Y_n \in A_\delta\big) < 0.$$

Proof. Standard arguments yield $\mathcal{M} \neq \emptyset$ and $\mathcal{M} = A_0 \cap \{I \leq I(A_0)\}$. As I is a good rate function, $\{I \leq I(A_0)\}$ is a compact set. A_0 being closed (see (A-0)), it follows that $\mathcal{M}$ is compact. As $(A_\delta)_{\delta>0}$ is a nested family of measurable sets, we obtain

$$\begin{aligned}\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mu^n\big(L_n^Y \notin \Gamma | L_n^Y \in A_\delta\big)\\ \leq \lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log Q_n\big(\Gamma^c \cap A_\delta\big) - \lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log Q_n\big(A_\delta\big).\end{aligned} \tag{4.4}$$

With the help of the upper and lower bounds of Theorem 3.2, the same argument as in [8] (Sanov's upper bound) yields:

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log Q_n\big(\Gamma^c \cap A_\delta\big) < -I(A_0). \tag{4.5}$$

On the other hand, by Sanov's lower bound, we obtain for all $\delta > 0$

$$\liminf_{n\to\infty}\frac{1}{n}\log Q_n\big(A_\delta\big) \geq -\inf\{I(\ell), \ell \in A_\delta^\circ\}. \tag{4.6}$$

Combining these arguments with (4.3), we obtain:

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log Q_n\big(A_\delta\big) \geq -I(A_0)$$

One completes the proof of the lemma using this inequality together with (4.5) in (4.4). $\square$

Proof of Theorem 4.2. As A_0 is assumed to be convex, by Lemma 4.1, $\mathcal{M}$ is decomposed as $\nu_* + \mathcal{S}$.
Let the function $f(x_1,\dots,x_k) = \prod_{i=1}^k f_i(x_i)$ be fixed where each $f_i \in C_b(\Sigma)$. By the definition of $\mu^n_{Y^k|A_\delta}$ (see (4.2)),

$$\frac{d\mu^n_{Y^k|A_\delta}}{d\mu^k}(y_1,\dots,y_k) = \int_{\Sigma^{n-k}}\frac{\mathbf{1}_{A_{\delta,n}}(y_1,\dots,y_n)}{Q_n(A_\delta)}\,\mu(dy_{k+1})\cdots\mu(dy_n). \tag{4.7}$$

where $A_{\delta,n} = \{L_n^Y \in A_\delta\}$. Consider

$$\Gamma(\eta) = \bigcap_{i=1}^k\{\ell \in \mathcal{Q};\ |\langle \ell, f_i\rangle - \langle \nu_*, f_i\rangle| < \eta\},$$

and let $\Gamma_n(\eta) = \{L_n^Y \in \Gamma(\eta)\}$. Then $\Gamma(\eta)$ satisfies the assumptions of Lemma 4.3, since it is open measurable and $\mathcal{M} \subset \Gamma(\eta)$. Let us prove this inclusion. If $\ell \in \mathcal{M}$, then $\ell = \nu_* + \ell^s$ by the assumption on $\mathcal{M}$. As $f_i \in M_\tau$ for $1 \leq i \leq k$, $\langle \ell^s, f_i\rangle = 0$ by Proposition 2.4. Hence $\langle \ell, f_i\rangle = \langle \nu_*, f_i\rangle$ and $\mathcal{M} \subset \Gamma(\eta)$. Therefore,

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mu^n\big(L_n^Y \notin \Gamma | L_n^Y \in A_\delta\big) = 0$$

by Lemma 4.3. The rest of the proof follows step by step the proof of Corollary 7.3.5 in [8]. Hence Theorem 4.2 is proved. □

4.4. **Back to Csiszár's example.** Within the scope of Section 3.4, we are interested in the limiting behaviour of $\mu^n_{Y|A_\delta(x)}$ where

$$\begin{aligned} A_\delta(x) &= \{\ell \in \mathcal{Q};\quad |\langle \ell, u\rangle - x| \le \delta\} \quad \text{with} \quad u(y) = y, \\ A_0(x) &= \{\ell \in \mathcal{Q};\quad \langle \ell, u\rangle = x\}. \end{aligned}$$

The sets of constraint are $\{L^Y_n \in A_\delta(x)\} = \{(y_1, \cdots, y_n);\ |\frac{1}{n}\sum_1^n y_i - x| \le \delta\}$ and $\mu^n_{Y|A_\delta(x)}$ represents the law of Y_1 under the constraint that the mean $\frac{1}{n}\sum_1^n Y_i$ is close to x. Let us denote by $\mathcal{M}_x$ the corresponding set of minimizers of (3.4).

Proposition 4.4. *For any $x \ge x_*$, ℓ belongs to $\mathcal{M}_x$ if and only if*
(1) *$\ell^a = \nu_*$ with $\nu_*(dy) = e^{y-\Lambda(1)}\,\mu(dy)$,*
(2) *$\langle \ell^s, u\rangle = x - x_*$ where $u(y) = y$, $y \ge 0$,*
(3) *$\sup\{\langle \ell^s, f\rangle; f, \int_{[0,\infty)} e^f\,d\mu < \infty\} = \langle \ell^s, u\rangle$.*

In particular, for any $x > x_*$, there are infinitely many elements in $\mathcal{M}_x$.

Proof. A careful look will convince the reader that the equivalence is already proved in Proposition 3.9.

Let us show that there are infinitely many minimizers when $x > x_*$. Because of the item 3) of the proposition, it is sufficient to prove that the gauge function $p(g) = \inf\{\lambda > 0;\ g/\lambda \in D_\mu\}$ of $D_\mu = \{f; \int_{[0,\infty)} e^f\,d\mu < \infty\}$ is not Gâteaux-differentiable at u (for this argument see for instance [11], p. 123). This means that there exists $f \in \mathcal{L}_\tau$ such that

$$\lim_{t\to 0,\ t>0} \frac{p(u+tf) - p(u)}{t} \ne \lim_{t\to 0,\ t<0} \frac{p(u+tf) - p(u)}{t}. \tag{4.8}$$

Consider

$$f(y) = \begin{cases} ay & \text{if} \quad y \in \cup_{n\ge 0}[2n, 2n+1) \\ -by & \text{if} \quad y \in \cup_{n\ge 0}[2n+1, 2n+2) \end{cases} \quad \text{where} \quad a \ne b,\ a > 0,\ b > 0.$$

A straightforward computation yields

$$\lim_{t\to 0,\ t>0} \frac{p(u+tf) - p(u)}{t} = a \quad \text{and} \quad \lim_{t\to 0,\ t<0} \frac{p(u+tf) - p(u)}{t} = -b.$$

Hence (4.8) holds and the proposition is proved. □

For an alternative proof and more details, see ([15], Proposition 10.5).

Applying Theorem 4.2 to $\mu^n_{Y|A_\delta(x)}$, we see that for all $x > x_* : \mathcal{M}_x = \nu_* + \mathcal{S}_x$ and for any $f \in C_b([0,\infty))$, $\langle \mu^n_{Y|A_\delta(x)}, f\rangle$ tends to $\langle \nu_*, f\rangle$, as $n \to \infty$, followed by $\delta \to 0$.

Moreover, the convergence of $\langle \mu^n_{Y|A_\delta(x)}, f\rangle$ to $\langle \nu_*, f\rangle$ even holds for $f \in M_\tau([0,\infty))$ (see Remark 4.4).

Remark 4.5. In this example, one can easily check that:

$$\begin{aligned} \lim_{n\to\infty} \nu^n_*\{L^Y_n \in A_\delta(x_*)\} &= 1 \quad \text{for } \delta > 0, \\ \lim_{n\to\infty} \nu^n_*\{L^Y_n \in A_\delta(x)\} &= 0 \quad \text{for } x > x_* \text{ and } \delta \in (0, x - x_*). \end{aligned}$$

Hence, Equation (4.1) which enforces a LLN under ν_*, is not satisfied when $x > x_*$ whereas the convergence of $\mu^n_{Y|A_\delta(x)}$ toward ν_* still occurs. Note that the approach developped by I. Csiszár in [6], which is based on the convergence in information also does not rely on such a restrictive LLN-assumption.

References

[1] A. Aboulalaâ. The Gibbs principle for Markov jump processes. *Stochastic Process. Appl.*, 64:257–271, 1996.

[2] R. A. Adams. *Sobolev Spaces.* Academic Press, New York, 1975.

[3] R.R. Bahadur and S. Zabell. Large deviations of the sample means in general vector spaces. *Ann. Probab.*, 7:587–621, 1979.

[4] E. Bolthausen and U. Schmock. On the maximum entropy principle for uniformly ergodic Markov chains. *Stochastic Process. Appl.*, 33:1–27, 1989.

[5] N. Bourbaki. *Eléments de Mathématique - Intégration 1-4.* Hermann, Paris, second edition, 1961.

[6] I. Csiszár. Sanov property, generalized I-projection and a conditionnal limit theorem. *Ann. Probab.*, 12(3):768–793, 1984.

[7] A. de Acosta. On large deviations of empirical measures in the τ-topology. *J. Appl. Probab.*, 31A (special volume):41–47, 1994.

[8] A. Dembo and O. Zeitouni. *Large Deviations Techniques And Applications.* Springer Verlag, New York, second edition, 1998.

[9] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov processes expectations for large time, III. *Comm. Pure Appl. Math.*, 29:389–461, 1976.

[10] P. Eichelsbacher and U. Schmock. Large deviations of U-empirical measures in strong topologies and applications. To be published in *Annales de l'Institut Henri Poincaré*, probabilités et statistiques in 2002.

[11] J.R. Giles. Convex analysis with application in the differentiation of convex functions. In *Research Notes in Math. 58.* Pitman, London, 1982.

[12] E. Giner. Décomposabilité et dualité des espaces d'orlicz, (in french). *Travaux du Séminaire d'Analyse Convexe, Montpellier*, 6, 1976.

[13] P. Groeneboom, J. Oosterhoff, and F.H. Ruymgaart. Large deviation theorems for empirical probability measures. *Ann. Probab.*, 7:553–586, 1979.

[14] Andrzej Kozek. Convex integral functionals on Orlicz spaces. *Comment. Math. Prace Mat.*, 21(1):109–135, 1980.

[15] C. Léonard. Minimizers of energy functional under not very integrable constraints. To appear in Journal of Convex Analysis, 2002.

[16] J. Najim. *Grandes déviations pour certaines mesures empiriques.* PhD thesis, Université de Paris 10- Nanterre, 2001.

[17] M. Rao and Z. D. Ren. *Theory of Orlicz spaces.* Marcel Dekker Inc., New-York, 1991.

[18] R. T. Rockafellar. Integrals which are convex functionals. *Pacific J. Math.*, 24(3):525–539, 1968.

[19] I. N. Sanov. On the probability of large deviations of random variables. *Selected transl. Math. Statist. and Prob.*, 1:214–244, 1961.

[20] A. Schied. Cramér's condition and Sanov's theorem. *Statist. Probab. Lett.*, 39:55–60, 1998.

[21] D. W. Stroock and O. Zeitouni. Microcanonical distributions, Gibbs states, and the equivalence of ensembles. In R. Durrett and H. Kesten, editors, *Festchrift in honour of F. Spitzer*, pages 399–424, Basel, Switzerland, 1991. Birkhäuser.