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 situation where the underlying log-Laplace transform (the pressure) may not be steep; for a topology based on test functions admitting only some finite exponential moments and with constraints built on energy functions admitting only some finite exponential moments.

(1) *Modal-X, Université Paris 10, Nanterre*
(2) *CMAP, Ecole Polytechnique, Palaiseau*

E-mail: leonard@modalx.u-paris10.fr, najim@modalx.u-paris10.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 a 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: March 2000.

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

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

and $+\infty$ 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 the functions f with some finite exponential moment with respect to μ :

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

The state space for the extended Sanov's theorem is not $\mathcal{P}$ anymore, but the larger set $\mathcal{Q}$ of all the nonnegative linear forms on $\mathcal{L}_\tau$ with a unit mass. 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 nonnegative linear form ℓ^s on $\mathcal{L}_\tau$ which is not σ-additive (if it is non null). In particular, its mass is $\langle \ell^s, \mathbf{1}\rangle = 0$, although $\ell^s \geq 0$ may not be zero. For a precise statement of the extended Sanov's 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 microcanonical 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 [7] and by D. W. Stroock and O. Zeitouni in [22]. As in [7] and [22], 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 [22] by introducing blow-ups A_δ of $A_0 = \cap_{\delta>0} A_\delta$ such that $\mathbb{P}(L_n^Y \in A_\delta) > 0$ for all $n \geq 1$ and $\delta > 0$.
With the extended Sanov's theorem in hand, rather than the usual one, 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 [22]: for all $\delta > 0$, $\lim_{n\to\infty} \nu_*^n(\{L_n^Y \in A_\delta\}) = 1$; is removed here. 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 4.6). Our main results about the Gibbs conditioning principle are Theorem 4.7 and Corollary 4.5.

1.3. **About the literature.** I. N. Sanov [20] 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 [10] and R. R. Bahadur and S. Zabell [3] for the weak topology. P. Groeneboom, J. Oosterhoff and F. H. Ruymgaart [16] dropped the Polish requirement and 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 [8]. In [7], I. Csiszár proves Sanov's theorem in a general setting by means of an alternative approach based on projection in information. In [11], 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 the k-tuples in $\{1,\dots,n\}$ 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 the functions f which admit finite exponential moments with respect to μ, of all orders

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

and $\mathcal{P}_\tau$ stands for the set of all probability measures which integrate all the functions in $\mathcal{L}_\tau$ or $\mathcal{M}_\tau$.

As already mentioned, the Gibbs conditioning principle (GCP) has been studied by I. Csiszár in [7] and by D. W. Stroock and O. Zeitouni in [22]. An updated presentation of the second approach is available in the textbook of A. Dembo and O. Zeitouni ([9], Section 7.3). In [4], E. Bolthausen and U. Schmock have proved a GCP for the occupation measures of uniformly ergodic Markov chains. Based on [7], A. Aboulalaâ [1] has obtained a GCP for the empirical measures L_n^Y of Makov jump processes. With the LDP for $L_n^{Y,k}$ in hand, P. Eichelsbacher and U. Schmock [11] derive a GCP, following the approach of [22]. They obtain 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 $\sigma(\mathcal{P}_\tau, \mathcal{M}_\tau)$.

Let us sum up the comparison between the results of this article and the already existing literature (to the best of our knowledge). The stronger version of Sanov's theorem in the literature is obtained by Eichelsbacher and Schmock in [11]: the rate function is the relative entropy and the topology for the full LDP is close to $\sigma(\mathcal{P}_\tau, \mathcal{M}_\tau)$. We improve this result strengthening the topology which becomes $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ where $\mathcal{Q}$ contains $\mathcal{P}$. Our rate function I is not the relative entropy anymore.
On the other hand, the LDP for $L_n^{Y,k}$ in [11] for $k \geq 2$ is far from being trivial and is not a consequence of our results.
For the GCP, the situation is more complex. In [11], the GCP is obtained for constraints and topologies based on functions satisfying (1.4): if the energy function φ admits only some finite exponential moments, an additional assumption ([11], Condition 1.16) might fail, see Remark 4.4 below. We extend this result in Theorem 4.7, in the case of only one ($k = 1$) tagged particle, since we obtain the GCP for constraints and topologies based on functions only satisfying (1.1). Note that the limiting objects are not probability measures anymore, in the general case. But our proof only gives a partial result (see Corollary 4.5) when $k \geq 2$: the topology is weaker than [11]'s one, although the energy function might only admit some finite exponential moments.
I. Csiszár has obtained alternative results with an alternative powerful approach. In particular, the convergence in information of the conditioned laws implies their convergence in variation. The constraints in [7] have to be convex with additional assumptions. This is the reason why a large amount of preliminary work has to be performed in [1] to be in a position to apply Csiszár's GCP; on the other hand, [1]'s main result is a simple consequence of our Theorem 4.7. In [7], the notion of generalized I-projection is introduced so that the GCP holds even with energy functions satisfying (1.1). For a discussion on more subtle differences between [7] and the present paper, see Section 4.8 below.

1.4. **Outline of the paper.** We remind in Section 2 a few definitions and results about Orlicz spaces. It is natural 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 and compare our results to Csiszár's ones in Section 4.8.

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:

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

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:

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

$(L_\theta,\|.\|_\theta)$ is a Banach space called the Orlicz space associated with θ. M_θ is a subspace of L_θ. It 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 ([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 nonincreasing sequence (A_k) 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 the 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

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

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

$$\left(\exists a > 0, \int_\Sigma e^{a|f(y)|}\, \mu(dy) < +\infty\right) \quad \Leftrightarrow \quad f \in \mathcal{L}_\tau$$

$$\left(\forall a > 0, \int_\Sigma e^{a|f(y)|}\, \mu(dy) < +\infty\right) \quad \Leftrightarrow \quad f \in \mathcal{M}_\tau$$

In case $f \in \mathcal{L}_\tau$, it is said that f admit *some* exponential moment; in case $f \in \mathcal{M}_\tau$, it is said that f admit *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's 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 $\mathcal{L}_\tau^* : G_f : \ell \mapsto \langle \ell, f \rangle$, $f \in \mathcal{L}_\tau$. 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 of the extended Sanov's 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 of the extended Sanov's 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 $I_a : \mathcal{L}_\tau^* \to [0, \infty]$ and $I_s : \mathcal{L}_\tau^* \to [0, \infty]$ are given by:

$$
\begin{aligned}
I_a(\ell) &= \begin{cases} \int_\Sigma \log\left(\frac{d\ell^a}{d\mu}\right) d\ell^a & \text{if } \ell \in \mathcal{Q} \cap L'_\tau \\ +\infty & \text{otherwise.} \end{cases} \\
I_s(\ell) &= \begin{cases} \sup\{\langle \ell^s, f\rangle;\ f, \mathbb{E}\, e^{f(Y)} < \infty\} & \text{if } \ell \in L'_\tau \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}
$$

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's 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 level sets $\{I \leq \alpha\}$, $\alpha \geq 0$, are compact.

Proof. The LDP in $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$ is proved in Lemma 3.3 with the rate function Θ^* which is convex as a convex conjugate. By Proposition 3.7, $\Theta^* = I$. But the domain of I is included in $\mathcal{Q}$, so that the LDP holds on $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$, for this argument see ([9], Lemma 4.1.5(b)). This completes the proof of the theorem. □

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 $\mathcal{Q} \cap L'_\tau = \Big(L_{\tau^*} \cap \mathcal{P}\Big) \oplus L_\tau^s$.

Remark 3.4. Note that while considering the LDP on $(\mathcal{M}_\tau^*, \sigma(\mathcal{M}_\tau^*, \mathcal{M}_\tau))$, no singular linear form appears since $M'_\tau \simeq L_{\tau^*}$. Moreover, the effective domain of I is included in $\mathcal{P}$ and the LDP holds in $(\mathcal{P}, \sigma(\mathcal{P}, \mathcal{M}_\tau))$. Hence, when the topology is based on functions with all exponential moments finite, the LDP holds in $\mathcal{P}$ with the usual rate function $H(.|\mu)$. This is in accordance with the result obtained by P. Eichelsbacher and U. Schmock ([11], Theorem 1.8).

We cannot expect the LDP with the good rate function $H(.|\mu)$. Indeed, A. Schied proved in [21] that if the topology is too wide (the so-called τ_ϕ-topology where ϕ admit only some exponential moments), then $\{H \le \alpha\}$ is no longer compact. The same argument holds in our context: $\{\ell \in \mathcal{Q}; I_a(\ell) \le \alpha\} = \{f\mu; f \in L_{\tau^*}, f \ge 0, \int_\Sigma f\,d\mu = 1, \int_\Sigma f \ln f\,d\mu \le \alpha\} + L^s_\tau$ which is not compact.

3.2. **Proof of the LDP.** Lemma 3.3 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 3.3. *The empirical measures $\{L^Y_n\}_{n\ge1}$ 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 a projective limit approach. By the ([9], Corollary 4.6.9) of Dawson-Gärtner's theorem, it is sufficient to check that for all $d \ge 1$ and $f_1, \cdots, f_d \in \mathcal{L}_\tau$, $(\langle L^Y_n, f_1\rangle, \dots, \langle L^Y_n, f_d\rangle)$ satisfies some LDP.
But $(\langle L^Y_n, f_1\rangle, \dots, \langle L^Y_n, f_d\rangle) = \frac{1}{n}\sum_{i=1}^n (f_1(Y_i), \dots, f_d(Y_i)) = \frac{1}{n}\sum_{i=1}^n \vec{f}(Y_i)$ 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 $f_1, \cdots, f_d \in \mathcal{L}_\tau$, $\vec{f}(Y_i)$ admits exponential moments. By Cramér's theorem in Polish spaces (Theorem 6.1.3 and Corollary 6.1.6 in [9]), $\frac{1}{n}\sum_{i=1}^n \vec{f}(Y_i)$ 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, $L^Y_n = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ satisfies the LDP with the good rate function, given for any $\ell \in \mathcal{L}^*_\tau$, by:
$\sup\{\sum_{i=1}^d \lambda_i\langle \ell, f_i\rangle - \log \mathbb{E}\, e^{\sum \lambda_i f_i(Y)};\; d \ge 1,\; \lambda \in \mathbb{R}^d,\; f_1, \cdots, f_d \in \mathcal{L}_\tau\} = \sup_{f\in\mathcal{L}_\tau}\{\langle \ell, f\rangle - \log \mathbb{E}\, e^{f(Y)}\} = \Theta^*(\ell)$, which is the desired result. $\square$

3.3. **Identification of the rate function.** In this section, we prove the identity $I = \Theta^*$ at Proposition 3.7. Lemmas 3.4, 3.5 and 3.6 are preminilary results for the proof of this identity.
Let us consider $J(\ell) = \sup_{f\in L_\tau}\left\{\langle \ell - \mu, f\rangle - \int_\Sigma \gamma(f)\,d\mu\right\}$, where γ is given by (2.3).

Lemma 3.4. *Let $\ell \in \mathcal{L}^*_\tau$, then*

1. $\Theta^*(\ell) < \infty \Rightarrow \ell \in L^*_\tau$
2. $J(\ell) \le \Theta^*(\ell)$
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 $\lambda \langle \ell, f \rangle \le \Theta^*(\ell) + \Theta(\lambda f) = \Theta^*(\ell)$. Hence, $\Theta^*(\ell) < \infty$ implies $\langle \ell, f \rangle = 0$. Thus ℓ is constant over the equivalence classes, that is $f \in L_\tau^*$.
2. As for all $t \ge 0$, $\quad \log t \le t - 1$, we have $-\mathbb{E}\, e^{f(Y)} + 1 \le -\log \mathbb{E}\, e^{f(Y)}$ and

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

3. As $\Theta^*(\ell) < \infty$ implies $J(\ell) < \infty$ and $\ell \in L_\tau^*$, let us consider $\ell \in L_\tau^*$.
For all $f \in L_\tau$, $\langle \ell - \mu, f \rangle \le J(\ell) + \int_\Sigma \gamma(f)\, d\mu$. As $\gamma(s) \le \tau(s) = \gamma(|s|)$, we have $\langle \ell - \mu, f \rangle \le J(\ell) + \int_\Sigma \tau(f)\, d\mu$. Choosing $\eta = \pm 1/\|f\|_\tau$ when $f \ne 0$, by the definition of the Luxemburg norm (2.1), one gets $\int_\Sigma \tau(\eta f)\, d\mu = 1$ which implies $|\langle \ell - \mu, f \rangle| \le (J(\ell) + 1)\|f\|_\tau$. This inequality still holds with $\|f\|_\tau = 0$. Therefore $\ell - \mu \in L_\tau^*$ is $\|\cdot\|_\tau$-continuous: $\ell \in L_\tau'$.
Suppose that $\langle \ell, \mathbf{1} \rangle = a \ne 1$. Then $\Theta^*(\ell) \ge \langle \ell, \lambda \mathbf{1} \rangle - \log \mathbb{E}\, e^{\lambda \mathbf{1}} = \lambda(a-1)$ which tends to $+\infty$ 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 $f > 0$ with $\langle \ell, f \rangle < 0$. Let $\lambda \ge 0$, then $\Theta^*(\ell) \ge \langle \ell, -\lambda f \rangle - \log \mathbb{E}\, e^{-\lambda f} \ge \langle \ell, -\lambda f \rangle$ tends to $+\infty$ as λ tends to $+\infty$. Thus $\Theta^*(\ell) < \infty$ implies $f \ge 0$ and Lemma 3.4 is proved. □

Lemma 3.5. *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 functions which converges pointwise to f and such that* $\quad |f_n| \le |f|$, *for all $n \ge 1$*
2. $\lim_{n\to\infty} \langle \ell, \mathbf{1}_{\{|f|>n\}} f \rangle = \langle \ell^s, f \rangle$

Proof. 1. Since f 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 $\mathbf{1}_{\{|f|\le n\}} f$ is bounded, $\langle \ell^s, f \rangle = \langle \ell^s, \mathbf{1}_{\{|f|>n\}} f \rangle$ (see Proposition 2.4). □

Lemma 3.6. *For all ℓ in L_τ', $\Theta^*(\ell) = \Theta^*(\ell^a) + \sup\{\langle \ell^s, f \rangle;\ f \in \operatorname{dom}\Theta\}$ where $\operatorname{dom}\Theta = \{f \in \mathcal{L}_\tau, \Theta(f) < \infty\}$ is the effective domain of Θ.*

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

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 the support functional of

A. For any $\ell \in L'_\tau$, we have

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

To prove the converse inequality, let $u, v \in L_\tau$, $n \ge 1$ and consider $f_{u,v,n} = \mathbf{1}_{\{|v|\le n\}}(-n \vee u \wedge n) + \mathbf{1}_{\{|v|>n\}}v$. Then,

$$\begin{aligned}\Theta^*(\ell) &\ge \langle \ell, f_{u,v,n}\rangle - \Theta(f_{u,v,n}) \\ &\ge \langle \ell, \mathbf{1}_{\{|v|\le n\}}(-n \vee u \wedge n)\rangle - \Theta(f_{u,v,n}) + \langle \ell, \mathbf{1}_{\{|v|>n\}}v\rangle\end{aligned}$$

As $\mathbf{1}_{\{|v|\le n\}}(-n \vee u \wedge n)$ is bounded, Proposition 2.4 yields $\langle \ell, \mathbf{1}_{\{|v|\le n\}}(-n \vee u \wedge n)\rangle = \langle \ell^a, \mathbf{1}_{\{|v|\le n\}}(-n \vee u \wedge n)\rangle$. Hence, $\Theta^*(\ell) \ge \langle \ell^a, \mathbf{1}_{\{|v|\le n\}}(-n \vee u \wedge n)\rangle - \Theta(f_{u,v,n}) + \langle \ell, \mathbf{1}_{\{|v|>n\}}v\rangle \ge \langle \ell^a, u\rangle - \Theta(u) + \langle \ell^s, v\rangle$ where the last inequality follows from the dominated convergence theorem and from Lemma 3.5. This completes the proof of the proposition. □

Proposition 3.7. *The identity $\Theta^* = I$ holds on $\mathcal{L}^*_\tau$.*

Proof. By Lemma 3.4, the effective domain of Θ^* is included in $\mathcal{Q} \cap L'_\tau$ and by Lemma 3.6, 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 $h \in L_{\tau^*}$, $h \ge 0$, $\int_\Sigma h\, d\mu = 1$. 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

([19], Theorem 2), noting that the convex conjugate of $e^s - 1$ is $t\log t - t + 1$. This completes the proof of the proposition. $\square$

3.4. **Csiszàr'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 ([7], Example 3.2) and in ([9], 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's theorem, L_n^Y satisfies the LDP with the good rate function $H(\cdot|\mu)$ in $(\mathcal{P}, \sigma(\mathcal{P}, B))$. By Cramér's theorem, $\hat{S}_n$ also satisfies the LDP with the good rate function Λ^*, where Λ^* is the convex conjugate of $\Lambda(\lambda) = \log\int_{[0,\infty)} c\frac{e^{(\lambda-1)y}}{1+y^3}\,dy$. 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's theorem and the following identity might fail to happen:

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

By 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 3.3 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 one can identify the rate functions I' and Λ^* to get, for any $x \in [0, +\infty)$:

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

for some $\ell_x \in \mathcal{Q}$ such that $\langle \ell_x, u\rangle = x$.
One is not sure that (3.3) holds, whereas (3.4) always holds. Roughly speaking, as $I(\ell) = H(\ell^a|\mu) + I_s(\ell^s)$, we have $\inf\{H(P \mid \mu); P \in \mathcal{P}, \langle P, u\rangle = x\} \leq \inf_{\langle \ell, u\rangle = x} I(\ell)$ and it is the contribution of the singular part in I which allows (3.4) to hold. The following proposition makes this statement precise and yields, as a by-product, the existence of non-null singular parts.

Proposition 3.8. *Let $x_* = \Lambda'(1)$. There exists a unique probability measure*

$$\nu_*(dy) = e^{y-\Lambda(1)}\,\mu(dy)$$

such that for all $x \geq x_*$ and all minimizing argument 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

$$I(\ell_x) = \int_{[0,\infty)} \log\left(\frac{d\nu_*}{d\mu}\right) d\nu_* + I_s(\ell_x^s) \tag{3.5}$$

with $\int_{[0,\infty)} \log\left(\frac{d\nu_*}{d\mu}\right) d\nu_* = \Lambda^*(x_*)$ 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 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.

One can prove that for $0 < x \leq x_*$, ℓ_x is a probability measure ν_x such that $\langle \nu_x, u\rangle = x$ and $I(\ell_x) = H(\nu_x \mid \mu)$. For more details, see [7], ([15], Proposition 10.5) and Section 4.6 below.

Proof of Proposition 3.8. For all $\lambda \leq 1$, both $\Lambda(\lambda)$ and $\Lambda'(\lambda)$ are finite, while $\Lambda(\lambda) = \infty$, for all $\lambda > 1$. Yet Λ is not steep and $x_* = \Lambda'(1)$ is the highest reached slope of Λ. 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, $I(\ell_y) = (y - x_*) + \Lambda^*(y^*)$. The convexity of I implies that

$$I(\alpha\ell_x + \beta\ell_y) \leq \alpha I(\ell_x) + \beta I(\ell_y) = \Lambda^*(x_*) + \alpha x + \beta y - x_* = I'(\alpha x + \beta y).$$

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 $\ell_x^a = \ell_y^a \stackrel{\triangle}{=} \nu_*$ 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:

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

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 holds in the two previous inequalities. But due to the strict convexity of I_a, $I_a(\alpha\ell_x^a + \beta\ell_y^a) = \alpha I_a(\ell_x^a) + \beta I_a(\ell_y^a)$, which 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 $I_s(\ell_x^s) = x - x_*$. □

4. The Gibbs Conditioning Principle

Let μ be a probability measure on $(\Sigma, \mathcal{A})$. 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 $\mathcal{E}_{\mathcal{Q}}$:

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

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:

$$\mu^n_{Y^k|A_\delta}(\cdot) = \mu^n\Big((Y_1, \ldots, Y_k) \in \cdot \mid L_n^Y \in A_\delta\Big)$$

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 [22] 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 definition of $\mu^n_{Y|A_\delta}$ yields:

$$\frac{d\mu^n_{Y|A_\delta}}{d\mu}(y_1) = \int_{\Sigma^{n-1}} \frac{1_{A_{\delta,n}}(y_1, \ldots, y_n)}{Q_n(A_\delta)} \mu(dy_2)\cdots\mu(dy_n)$$

As the density $\frac{d\mu^n_{Y|A_\delta}}{d\mu}$ is bounded, it belongs to L_{τ^*}. Due to identification (3.1), we shall consider $\mu^n_{Y|A_\delta}$ as an element of L'_τ in the sequel. The following conventions prevail in this section: $\Gamma_n = \{L_n^Y \in \Gamma\}$, $A_{\delta,n} = \{L_n^Y \in A_\delta\}$ and A° is the $\sigma(\mathcal{Q}, L_\tau)$-interior of A.

4.1. **Outline of the section.** Assumptions are stated in Section 4.2. Preliminary estimates are established in Section 4.3. Lemma 4.1 is based on the extended Sanov's theorem (Theorem 3.2). Lemma 4.2 allows the control of large values of $\langle f, L_n^Y \rangle$ outside an arbitrary set containing the minimizers of the extended entropy.
In Section 4.4, we describe the set of minimizers $\mathcal{M} = \{\ell \in A_0;\ I(\ell) = I(A_0)\}$ when A_0 is convex. In general this set is not reduced to a unique minimizer.
Section 4.5 is devoted to the convergence of $\mathbb{P}((Y_1, \dots, Y_k) \in \cdot \mid L_n^Y \in A_0)$ towards a probability. Theorem 4.4 states the convergence when the test function are not necessarily bounded but satisfy (1.4) and Corollary 4.5 establishes the weak convergence.
In Section 4.6, we revisit Csiszár's example. We exhibit a GCP in a situation where the log-Laplace transform is not steep; as a consequence, the set of minimizers is not reduced to a single point.
Section 4.7 is devoted to the study of $\mathbb{P}(Y_1 \in \cdot \mid L_n^Y \in A_0)$ when test functions satisfy (1.1). Theorem 4.7 is the main result of the section. It is proved there that the cluster points of $\mathbb{P}(Y_1 \in \cdot \mid L_n^Y \in A_0)$ belong to the closure of the convex hull of $\mathcal{M}$.
We compare our results with Csiszár's results in Section 4.8.

4.2. **Statement of the assumptions.** Let us denote $I(A) = \inf\{I(\ell); \ell \in A\}$ for all $A \subset \mathcal{Q}$.

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

$$I(A_\delta^\circ) \le I(A_0), \quad \textit{for all } \delta > 0. \tag{4.1}$$

Two important cases where (4.1) 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 $\mathcal{E}$ 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 ([9], Section 6.2) for its definition). Hence more open sets and more measurable sets are available. As an example, consider the family defined by $A_\delta = \{\ell \in \mathcal{Q}; |\langle \ell, u \rangle - 1| \le \delta\}$ where $u \in \mathcal{L}_\tau$. Then $(A_\delta)_{\delta>0}$ satisfies (A-0), with $A_0 = \{\ell \in \mathcal{Q}; \langle \ell, u \rangle = 1\}$.

The following assumption is the counterpart of Assumption (A-1) in ([9], Section 7.3).

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

Remark 4.2. The equation ([9],(7.3.2)) which is part of the assumptions in [9] (see also [11], Condition 1.16) enforce 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 4.1 below) so that no underlying law of large numbers is necessary. Moreover, there exist cases where ([9],(7.3.2)) fails to hold while (A-1) is still satisfied (see Remark 4.4 below).

4.3. **Preliminary estimates.** In this section, we prove estimates at Lemma 4.1 and Lemma 4.2 which will be used later in order to obtain convergence results for $\{\mu^n_{Y^k|A_\delta}\}$. The set of minimizers is denoted by

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

The following lemma is the counterpart of Theorem 7.3.3 in [9].

Lemma 4.1. *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 $\Gamma \in \mathcal{E}$ with $\mathcal{M} \subset \overset{\circ}{\Gamma}$, we have:*

$$\limsup_{\delta\to 0}\limsup_{n\to\infty} \frac{1}{n}\log \mu^n\big(L_n^Y \notin \Gamma | L_n^Y \in A_\delta\big) < 0.$$

Proof. Standard arguments yield $\mathcal{M} \neq \emptyset$ and $\mathcal{M} = A_0 \cap \{I \le I(A_0)\}$. As I is a good rate function, $\{I \le I(A_0)\}$ 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

$$\begin{aligned}\limsup_{\delta\to 0}\limsup_{n\to\infty} &\frac{1}{n}\log \mu^n\big(L_n^Y \notin \Gamma | L_n^Y \in A_\delta\big)\\ &\le \lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log Q_n\big(\Gamma^c \cap A_\delta\big) - \lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log Q_n\big(A_\delta\big)\end{aligned} \tag{4.2}$$

The same argument as in [9] (Sanov's upper bound) yields:

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log Q_n\big(\Gamma^c \cap A_\delta\big) < -I(A_0) \tag{4.3}$$

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

$$\liminf_{n\to\infty}\frac{1}{n}\log Q_n\big(A_\delta\big) \ge -\inf\{I(\ell), \ell \in A_\delta^\circ\}. \tag{4.4}$$

Combining these arguments with (4.1), we obtain:
$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log Q_n\big(A_\delta\big) \ge -I(A_0)$
One completes the proof of the theorem, using this inequality together with (4.3) in (4.2). □

Lemma 4.2. *Assume (A-0) and (A-1) and let $\mathcal{M}$ and Γ be as in Lemma 4.1. For all $f \in \mathcal{L}_\tau$, we have:* $\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{E}[|\langle f, L_n^Y\rangle|\mathbf{1}_{\Gamma^c}(L_n^Y)\big| L_n^Y \in A_\delta] = 0$.

Proof. For all $n \geq 1$, $\delta > 0$,

$$\begin{aligned}
& \mathbb{E}[|\langle f, L_n^Y \rangle| \mathbf{1}_{\Gamma^c}(L_n^Y) \big| A_{\delta,n}] \\
\leq \; & \mathbb{E}[|\langle f, \delta_{Y_1} \rangle| \mathbf{1}_{\Gamma^c}(L_n^Y) \big| A_{\delta,n}] \\
= \; & \int_{\Sigma^n} |f(y_1)| \mathbf{1}_{\Gamma_n^c}(y_1, \cdots, y_n) \mu(dy_1 \cdots dy_n | A_{\delta,n}) \\
= \; & \int_{\Sigma^n} |f(y_1)| \mathbf{1}_{\Gamma_n^c}(y_1, \cdots, y_n) \mathbf{1}_{A_{\delta,n}}(y_1, \cdots, y_n) \frac{\mu(\,dy_1) \cdots \mu(\,dy_n)}{Q_n(A_\delta)} \\
= \; & \int_{\Sigma} |f(y_1)| \Big\{ \int_{\Sigma^{n-1}} g_{n,\delta}(y_1, \cdots, y_n) \mu(\,dy_2) \cdots \mu(\,dy_n) \Big\} \mu(\,dy_1)
\end{aligned}$$

with $g_{n,\delta}(y_1, \cdots, y_n) = \frac{\mathbf{1}_{\Gamma_n^c \cap A_{\delta,n}}(y_1, \cdots, y_n)}{Q_n(A_\delta)}$. Let us denote

$$u_{n,\delta}(y_1) = \int_{\Sigma^{n-1}} g_{n,\delta}(y_1, \cdots, y_n) \mu(dy_2) \cdots \mu(dy_n).$$

We have shown that: $\mathbb{E}[|\langle f, L_n^Y \rangle| \mathbf{1}_{\Gamma^c}(L_n^Y) \big| A_{\delta,n}] \leq \int_\Sigma |f(y_1)| u_{n,\delta}(y_1) \mu(\,dy_1) \leq 2\|f\|_\tau \|u_{n,\delta}\|_{\tau^*}$, where the last inequality is Hölder's inequality in Orlicz spaces (see (2.2))

It remains to prove that $\|u_{n,\delta}\|_{\tau^*} \to 0$ for $n \to \infty$ followed by $\delta \to 0$. By the definition of the Luxemburg norm, $\int_\Sigma \tau^*\big(\frac{u}{\epsilon}\big)\, d\mu \leq 1$ implies that $\|u\|_{L_{\tau^*}} \leq \epsilon$. Therefore, it is sufficient to prove that for any $\epsilon > 0$

$$\limsup_\delta \limsup_n \int_\Sigma \tau^* \Big(\frac{u_{n,\delta}(y_1)}{\epsilon} \Big) \mu(dy_1) \leq 1 \tag{4.5}$$

As τ^* is convex, Jensen's inequality yields:

$$\tau^* \Big(\frac{u_{n,\delta}(y_1)}{\epsilon} \Big) = \tau^* \Big(\frac{\int_{\Sigma^{n-1}} g_{n,\delta}(y_1, .)\, d\mu^{n-1}}{\epsilon} \Big) \leq \int_{\Sigma^{n-1}} \tau^* \Big(\frac{g_{n,\delta}(y_1, .)}{\epsilon} \Big)\, d\mu^{n-1}$$

Integrating on both sides, we obtain:

$$\int_\Sigma \tau^* \Big(\frac{u_{n,\delta}(y_1)}{\epsilon} \Big) \mu(dy_1) \leq \int_{\Sigma^n} \tau^* \Big(\frac{g_{n,\delta}(y_1, \cdots, y_n)}{\epsilon} \Big) \mu(\,dy_1) \cdots \mu(\,dy_n).$$

Recall that $\tau^*(t) = (|t| + 1)\log(|t| + 1) - |t|$. As $\tau^*(0) = 0$,

$$\tau^* \Big(\frac{g_{n,\delta}(y_1, \cdots, y_n)}{\epsilon} \Big) = \mathbf{1}_{\Gamma_n^c \cap A_{\delta,n}}(y_1, \cdots, y_n)\ \tau^* \Big(\frac{1}{\epsilon Q_n(A_\delta)} \Big)$$

and

$$\begin{aligned}
\int_{\Sigma^n} \tau^* \Big(\frac{g_{n,\delta}}{\epsilon} \Big)\, d\mu^n &= Q_n(\Gamma^c \cap A_\delta) \tau^* \Big(\frac{1}{\epsilon Q_n(A_\delta)} \Big) \\
&= \frac{1}{\epsilon} \mu^n(\Gamma_n^c | A_{\delta,n})\ \epsilon\ Q_n(A_\delta) \tau^* \Big(\frac{1}{\epsilon Q_n(A_\delta)} \Big)
\end{aligned}$$

As $t\tau^*(1/t) = (1+t)\log(1+1/t) - 1 \le 2\log(2/t)$ for $0 < t \le 1$, we obtain:

$$\begin{aligned}
&\int_{\Sigma^n} \tau^*\left(\frac{g_{n,\delta}}{\epsilon}\right) d\mu^n \\
\le\ & \mu^n(\Gamma_n^c|A_{\delta,n})\frac{2}{\epsilon}\log\left(\frac{2}{\epsilon Q_n(A_\delta)}\right) \\
=\ & \mu^n(\Gamma_n^c|A_{\delta,n})\frac{2}{\epsilon}\log\left(\frac{2}{\epsilon}\right) + \mu^n(\Gamma_n^c|A_{\delta,n})\frac{2}{\epsilon}|\log Q_n(A_\delta)| \qquad (4.6)
\end{aligned}$$

where $0 < \epsilon \le 1$. The first term of the right hand side of (4.6) tends to 0 by Lemma 4.1. The second term can be rewritten

$$\frac{2}{\epsilon}\mu^n(\Gamma_n^c|A_{\delta,n})|\log Q_n(A_\delta)| = \frac{2}{\epsilon}\ n\mu^n(\Gamma_n^c|A_{\delta,n})\left|\frac{1}{n}\log Q_n(A_\delta)\right|$$

By (4.4) and (4.1), we have: $\limsup_\delta \limsup_n |\frac{1}{n}\log Q_n(A_\delta)| \le I(A_0)$, and by Lemma 4.1 we obtain: $\limsup_\delta \limsup_n n\mu^n(\Gamma_n^c|A_{\delta,n}) = 0$.
Finally, for every fixed $\epsilon > 0$, the second term of the right hand side of (4.6) tends to 0. Hence, (4.5) holds and the proof of the lemma is completed. □

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

Lemma 4.3. *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.

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

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

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

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

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. □

We shall see in Section 4.6 that within the scope of Csiszár's example (see Section 3.4), $\mathcal{M} = \nu_* + \mathcal{S}$ where $\mathcal{S}$ is not reduced to a single point.

4.5. **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. Let us define the following set of test functions:

$$\mathcal{F} = \left\{ f : \Sigma^k \to \mathbb{R};\ f(x_1, \cdots, x_k) = f_1(x_1) g(x_2, \cdots, x_k),\ f_1 \in M_\tau(\Sigma),\ g \in C_b(\Sigma^{k-1}) \right\}$$

The main restriction is that only one coordinate function f_i belongs to M_τ, the other ones being bounded. We choose the first one for the sake of simplicity, since it is obvious by exchangeability that our results still hold with $\mathcal{F}$ replaced by the space of its symetrized functions. Note that $\mathcal{F}$ is a subset of $M_\tau(\Sigma^k)$. As the Radon-Nikodym's derivative $\frac{d\mu^n_{Y^k|A_\delta}}{d\mu^n}$ is bounded under Assumption (A-1), the dual bracket $\langle \mu^n_{Y^k|A_\delta}, f \rangle$, $n \geq 1, \delta > 0$, is well-defined for any $f \in \mathcal{F}$. Theorem 4.4 and Corollary 4.5 below are counterparts of Corollary 7.3.5 in [9]. They are corollaries of Lemma 4.1 and Lemma 4.2.

Theorem 4.4. *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 $\mathcal{F}$, 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 4.3).

Remark 4.3. If $k = 1$, then $\mathcal{F} = M_\tau(\Sigma)$ and $\mu^n_{Y|A_\delta} \to \nu_*$ with respect to $\sigma(M^*_\tau, M_\tau)$. Note that M_τ does not separate L'_τ. But, identifying the nonseparated elements, we obtain $\sigma(L'_\tau, M_\tau) = \sigma(M'_\tau, M_\tau) = \sigma(L_{\tau^*}, M_\tau)$ where the last equality is Theorem 2.1.

Theorem 4.4 improves Corollary 7.3.5 in [9] in three directions:

1. The constraint sets A_δ can be based on functions with possibly infinite exponential moments, for instance
$$A_\delta = \{\ell \in \mathcal{Q}; |\langle \ell, u \rangle - 1| \leq \delta\} \quad \text{with} \quad u \in \mathcal{L}_\tau.$$
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 4.6 below and in particular Remark 4.4.
3. It handles a convergence slightly more general than the weak convergence (as it induces Corollary 4.5)

Corollary 4.5. *Under the assumptions of Theorem 4.4, $\mu^n_{Y^k|A_\delta} \to \nu^k_*$ weakly for $n \to \infty$ followed by $\delta \to 0$.*

Here, weakly means with respect to $\sigma(\mathcal{P}(\Sigma^k), C_b(\Sigma^k))$.

Proof of Corollary 4.5. Assume that Theorem 4.4 is proved and consider the function $f(x) = \prod_1^k f_i(x_i)$ where $f_i \in C_b(\Sigma)$, the set of continuous and bounded functions over Σ. Since $f \in \mathcal{F}$, we have $\langle \mu^n_{Y^k|A_\delta}, f\rangle \to \langle \nu^k_*, f\rangle$. One completes the proof noting that, as Σ^k is a separable metric space, $C_b(\Sigma)^{\otimes k}$ is convergence determining for $\mathcal{P}(\Sigma^k)$. □

Proof of Theorem 4.4. As A_0 is assumed to be convex, by Lemma 4.3, $\mathcal{M}$ is decomposed as $\nu_* + \mathcal{S}$.
Let the function $f(x) = \prod_{i=1}^k f_i(x_i)$ be fixed where $f_1 \in M_\tau(\Sigma)$ and $f_i \in B$, $i \geq 2$. We first control the growth of f_1 with the help of Lemma 4.2 outside some set Γ. Then, we use the same combinatorial technique as in the proof of Corollary 7.3.5 in [9] to work within Γ. By the definition of $\mu^n_{Y^k|A_\delta}$,

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

Consider

$$\Gamma(\eta) = \cap_{i=1}^k \{\ell \in \mathcal{L}^*_\tau;\ |\langle \ell, f_i\rangle - \langle \nu_*, f_i\rangle| \leq \eta\} \cap \{\ell;\ |\langle \ell, |f_1|\rangle - \langle \nu_*, |f_1|\rangle| \leq \eta\}$$

and let $\Gamma_n(\eta) = \{L^Y_n \in \Gamma(\eta)\}$. Then $\Gamma(\eta)$ satisfies the assumptions of Lemma 4.1: $\Gamma^\circ(\eta) \subset \mathcal{M}$. In fact, 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 $\mathcal{M} \subset \Gamma^\circ(\eta)$.
In the following computation, we shall frequently use Γ and Γ_n instead of $\Gamma(\eta)$ and $\Gamma_n(\eta)$. Since f can be considered as a function from Σ^n to $\mathbb{R}$, (4.7) and a straightforward application of Fubini's theorem yield:

$$\langle f, \mu^n_{Y^k|A_\delta}\rangle = \langle f, \mu^n_{Y^n|A_\delta}\rangle = \langle f\mathbf{1}_{\Gamma_n}, \mu^n_{Y^n|A_\delta}\rangle + \langle f\mathbf{1}_{\Gamma_n^c}, \mu^n_{Y^n|A_\delta}\rangle$$

Our goal is to show that $\lim_{n\to\infty}\langle f, \mu^n_{Y^k|A_\delta}\rangle = \langle \nu^k_*, f\rangle$. The same computation as before yields:

$$\begin{aligned}\langle f, \mu^n_{Y^k|A_\delta}\rangle - \langle f, \nu^k_*\rangle = {}& \langle f\mathbf{1}_{\Gamma_n^c}, \mu^n_{Y^n|A_\delta}\rangle - \langle f, \nu^k_*\rangle \mu^n\big(L^Y_n \notin \Gamma | L^Y_n \in A_\delta\big)\\ &+ \langle f\mathbf{1}_{\Gamma_n}, \mu^n_{Y^n|A_\delta}\rangle - \langle f, \nu^k_*\rangle \mu^n\big(L^Y_n \in \Gamma | L^Y_n \in A_\delta\big)\end{aligned} \tag{4.8}$$

We first deal with the first term of the right hand side of (4.8). By Lemma 4.1, $\lim_{\delta\to 0}\lim_{n\to\infty}\mu^n\big(L^Y_n \notin \Gamma | L^Y_n \in A_\delta\big) = 0$ and

$$\begin{aligned}&|\langle f\mathbf{1}_{\Gamma_n^c}, \mu^n_{Y^n|A_\delta}\rangle|\\ = {}& \Big|\int_{\Sigma^n}\prod_{i=1}^k f_i(y_i)\frac{\mathbf{1}_{A_{\delta,n}\cap\Gamma_n^c}(y_1,\dots,y_n)}{Q_n(A_\delta)}\,\mu(dy_1)\cdots\mu(dy_n)\Big|\\ \leq {}& \Big(\prod_{i=2}^k \|f_i\|_\infty\Big)\int_{\Sigma^n}|f_1(y_1)|\frac{\mathbf{1}_{A_{\delta,n}\cap\Gamma_n^c}(y_1,\dots,y_n)}{Q_n(A_\delta)}\,\mu(dy_1)\cdots\mu(dy_n)\end{aligned}$$

where the last integral converges to zero by Lemma 4.2. Hence, for any $\eta > 0$

$$\lim_{\delta\to 0}\lim_{n\to\infty}\Big\{\langle f\mathbf{1}_{\Gamma_n^c(\eta)}, \mu^n_{Y^n|A_\delta}\rangle - \langle f, \nu_*^k\rangle\mu^n\big(L_n^Y \notin \Gamma(\eta)|L_n^Y \in A_\delta\big)\Big\} = 0 \qquad (4.9)$$

Let us consider now the second term of the right hand side of (4.8). It is worthy to compare $\langle f\mathbf{1}_{\Gamma_n}, \mu^n_{Y^n|A_\delta}\rangle$ with $\mathbb{E}[\prod_1^k\langle f_i\ , L_n^Y\rangle\ \mathbf{1}_{\Gamma_n}|A_{\delta,n}]$. In fact, the definition of Γ_n gives information on quantities like $\langle f_i\ , L_n^Y\rangle$. Using the same combinatorial technique as in the proof of Corollary 7.3.5 in [9], one obtains

$$\begin{aligned}&|\langle f\mathbf{1}_{\Gamma_n}, \mu^n_{Y^n|A_\delta}\rangle - \mathbb{E}[\prod_{i=1}^k\langle f_i, L_n^Y\rangle\mathbf{1}_{\Gamma_n}|A_{\delta,n}]|\\ &\quad\le 2\Big(1-\tfrac{n!}{n^k(n-k)!}\Big)\Big(\textstyle\prod_{i=2}^k\|f_i\|_\infty\Big)\int_{\Sigma^n}|f_1(y_1)|\mathbf{1}_{\Gamma_n}\,\mu^n_{Y^n|A_\delta}(dy_1\cdots dy_n)\end{aligned} \qquad (4.10)$$

It remains to control the last integral:
$\int_{\Sigma^n}|f_1(y_1)|\,\mu^n_{Y^n|A_\delta}(dy_1\cdots dy_n) = \int_{\Sigma^n}\langle|f_1|, L_n^Y\rangle\,\mu^n_{Y^n|A_\delta}(dy_1\cdots dy_n) \le \langle|f_1|,\nu_*\rangle+\eta$ where the last inequality comes from the definition of Γ (this is why a control over $|f_1|$ appears in the definition of Γ). Thus, for any $\delta > 0$

$$\big|\langle f\mathbf{1}_{\Gamma_n}, \mu^n_{Y^n|A_\delta}\rangle - \mathbb{E}[\prod_{i=1}^k\langle f_i, L_n^Y\rangle\mathbf{1}_{\Gamma_n}|A_{\delta,n}]\ \big| \le C\Big(1-\frac{n!}{n^k(n-k)!}\Big)\xrightarrow[n\to\infty]{} 0$$

and for any $\eta > 0$

$$\lim_{\delta\to 0}\lim_{n\to\infty}\Big\{\langle\mu^n_{Y^k|A_\delta}, f\mathbf{1}_{\Gamma_n(\eta)}\rangle - \mathbb{E}[\prod_{i=1}^k\langle f_i, L_n^Y\rangle\mathbf{1}_{\Gamma_n(\eta)}|A_{\delta,n}]\Big\} = 0 \qquad (4.11)$$

Finally, the definition of Γ yields

$$\begin{aligned}
&\big|\mathbb{E}[\prod_{i=1}^{k}\langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n(\eta)}|A\delta, n] - \langle f, \nu_*^k\rangle \mu^n\big(L_n^Y \in \Gamma(\eta)|L_n^Y \in A_\delta\big)\big| \\
&= \ \big|\mathbb{E}[\prod_{i=1}^{k}\langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n} - \prod_{1}^{k}\langle f_i, \nu_*\rangle \ \mathbf{1}_{\Gamma_n}|A_{\delta,n}] \ \big| \\
&= \ \big|\mathbb{E}[\langle f_1, L_n^Y\rangle \prod_{i=2}^{k}\langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n} - \langle f_1, \nu_*\rangle \prod_{i=2}^{k}\langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n} \\
&\quad +\langle f_1, \nu_*\rangle\langle f_2, L_n^Y\rangle \prod_{i=3}^{k}\langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n} - \langle f_1, \nu_*\rangle\langle f_2, \nu_*\rangle \prod_{i=3}^{k}\langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n} + \cdots] \ \big| \\
&\leq \ \eta \prod_{i=2}^{k} \|f_i\|_\infty + \eta|\langle f_1, \nu_*\rangle| \prod_{i=3}^{k} \|f_i\|_\infty + \cdots \\
&\leq \ C\eta
\end{aligned} \tag{4.12}$$

Hence, letting η tend to zero, (4.9), (4.11) and (4.12) yield

$$\lim_{\delta\to 0}\lim_{n\to\infty}\left\{\langle \mu^n_{Y^k|A_\delta}, f\rangle - \langle f, \nu_*^k\rangle\right\} = 0$$

The convergence is then proved for every $f = \prod_1^k f_i$ where $f_1 \in M_\tau$ and $f_i \in B$, $i \geq 2$. In particular, let $f_i \in C_b(\Sigma)$, $i \geq 2$, then $\langle \mu^n_{Y^k|A_\delta}, f_1 \prod_2^k f_i\rangle \to \langle (\nu_*)^k, f_1 \prod_2^k f_i\rangle$ which can be rewritten

$$\langle m_{f_1}, g\rangle \to \langle \nu_*, f_1\rangle\langle (\nu_*)^{k-1}, \prod_2^k f_i\rangle \tag{4.13}$$

where $m_{f_1}(dy_2\cdots dy_k) = \int_{z\in\Sigma} f_1(z)\mu^n_{Y^k|A_\delta}(dzdy_2\cdots dy_k)$ is a signed measure and $g = \prod_2^k f_i$. As Σ is a separable metric space, $C_b(\Sigma)^{k-1}$ is convergence determining for $\mathcal{P}(\Sigma^k)$ and (4.13) holds for every $g \in C_b(\Sigma^{k-1})$. This completes the proof of Theorem 4.4. $\square$

As an attempt to extend this result to a class of test functions wider than $\mathcal{F}$, for instance $M_\tau(\Sigma^k)$, one may try to substitute in the above proof $\Gamma(\eta)$ by a set like $\Gamma_* = \{\ell \in \mathcal{L}_\tau^*; |\langle \ell^{a\otimes k} - (\nu_*)^{\otimes k}, f\rangle| \leq \eta\}$ with f in $\mathcal{M}_\tau(\Sigma^k)$. But, $\ell \mapsto \ell^{a\otimes k}$ is not continuous with respect to the topologies $\sigma(\mathcal{M}_\tau^*, \mathcal{M}_\tau)$ and $\sigma(\mathcal{M}_\tau(\Sigma^k)^*, \mathcal{M}_\tau(\Sigma^k))$, for such an argument see ([9], Exercice 7.3.18). As a consequence, it is not clear that $\mathcal{M} \subset \Gamma_*^\circ$ and Lemma 4.1 may not be available.

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

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

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 4.6. *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 3.8.

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 [12], p. 123). This means that there exists $f \in \mathcal{L}_\tau$ such that

$$\lim_{t\to 0,\ t>0} \frac{p(u+tf) - p(u)}{t} \ne \lim_{t\to 0,\ t<0} \frac{p(u+tf) - p(u)}{t} \tag{4.14}$$

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 (4.14) holds and the proposition is proved. $\square$

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

Applying Theorem 4.4 to $\mu^n_{Y|A_\delta(x)}$, we see that for all $x > x_* : \mathcal{M}_x = \nu_* + \mathcal{S}_x$ and for any $f \in \mathcal{F}$, $\langle \mu^n_{Y|A_\delta(x)}, f\rangle$ tnds to $\langle \nu_*, f\rangle$, as $n \to \infty$, followed by $\delta \to 0$.

Remark 4.4. In this example, one can easily check that:

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

Hence, Assumption (7.3.2) in [9] and the corresponding assumption in [11], which enforce a LLN under ν_*, are not satisfied when $x > x_*$ whereas the convergence of $\mu^n_{Y|A_\delta(x)}$ toward ν_* still happens: Assumption (A-1) is a weaker assumption.
Note that the approach developped by I. Csiszár in [7], which is based on the convergence in information (see Section 4.8), also does not rely on such a restrictive LLN-assumption.

Remark 4.5. Assume that $x > x_*$, then no singular part is detected by the convergence with test functions in $\mathcal{F}$. This is not surprising since by Remark 4.3, $\mathcal{F} = M_\tau$ if $k = 1$ and by Proposition 2.4, $\langle \ell^s, f\rangle = 0$ if $f \in M_\tau$.
As the singular part of a minimizer ℓ gives information on the gap between x and x_* (recall that $I_s(\ell^x_s) = x - x_*$), one can ask if it is possible to have a convergence result more precise than Theorem 4.4 which would take singular parts into account. This is the goal of the next section.

4.7. **More about the limiting behaviour of $\mu^n_{Y|A_\delta}$.** We consider here the case where $k = 1$ and where the topology is $\sigma(L'_\tau, L_\tau)$. It has been seen in Sections 4.5 and 4.6 that

- no singular parts appear when dealing with the topology $\sigma(L'_\tau, M_\tau)$ (see Remark 4.3): $\langle \mu^n_{Y|A_\delta}, f\rangle \to \langle \ell^a, f\rangle$, for all $f \in M_\tau$
- the singular parts are of interest (see Remark 4.5).
- there might be several minimizers (see Proposition 4.6).

Hence, several questions arise:

- When considering the topology $\sigma(L'_\tau, L_\tau)$, does $\mu^n_{Y|A_\delta}$ admit a unique limit ℓ, where ℓ is a minimizer?
- Does $\mu^n_{Y|A_\delta}$ visit several minimizers? Does it visit all of them?

Our answer to these questions is very partial. We just show that the set of cluster points of $\{\mu^n_{Y|A_\delta}\}_{n,\delta}$ is nonempty and is a subset of the closed convex hull of $\mathcal{M}$. As the convergence is of a special type (first $n \to \infty$ then $\delta \to 0$), we shall make precise what a cluster point is.

We say that $\nu \in L'_\tau$ is a (n,δ)-cluster point of $\{\mu^n_{Y|A_\delta}\}_{n,\delta}$ if there exists a family $\{\nu_\delta\}_{\delta>0} \subset L'_\tau$ such that

- for any $\delta > 0$, there exists a subnet of the sequence $\{\mu^n_{Y|A_\delta}\}_{n\geq 1}$ converging to ν_δ for the $\sigma(L'_\tau, L_\tau)$-topology.
- there exists a nonincreasing sequence $\delta_n \to 0$ such that there exists a subnet of $\{\nu_{\delta_n}\}_{n\geq 1}$ converging to ν for the $\sigma(L'_\tau, L_\tau)$-topology.

Remark 4.6. Since L_τ is not separable (τ grows too fast: it does not satisfy the Δ_2-condition), the unit strong ball of L'_τ does not admit a countable base of neighbourhoods for the topology $\sigma(L'_\tau, L_\tau)$. For any sequence in the unit ball of

L'_τ, the existence of a converging subsequence is not guaranteed and one must deal with subnets (for references, see the Appendix; see also [13] and [18]).

The following theorem is a counterpart of Corollary 7.3.5 in [9].

Theorem 4.7. *Assume (A-0) and (A-1). Let $\mathcal{C}$ be the set of (n,δ)-cluster points of $\{\mu^n_{Y|A_\delta}\}_{n,\delta}$, then $\mathcal{C}$ is a nonempty subset of $\overline{\mathrm{co}}(\mathcal{M})$, the closed convex hull of $\mathcal{M}$.*

The proof of Theorem 4.7 is based on the following lemma.

Lemma 4.8. *For any $\delta > 0$, $\{\mu^n_{Y|A_\delta}\}_{n\geq 1}$ is relatively $\sigma(L'_\tau, L_\tau)$-compact.*

In this lemma, δ is fixed and we are looking for a result valid uniformly in f. Thus, Lemma 4.1 and Lemma 4.2 are not directly useful. Nevertheless, the following results can be proved easily by mimicking their proofs. Let $\Gamma_{a,f} \stackrel{\triangle}{=} \{\ell \in L'_\tau, |\langle \ell, f\rangle| \leq a\}$, where $f \in L_\tau$ and $a \geq 0$. As $\mathcal{M}$ is compact, for all f there exists a such that $\mathcal{M} \subset \Gamma_{a,f}$. Moreover, one can prove that for any $\delta > 0$, $f \in L_\tau$

$$\lim_{a\to\infty} \limsup_n \frac{1}{n} \log \mu^n \left(L^Y_n \notin \Gamma_{a,f} | L^Y_n \in A_\delta\right) = -\infty \tag{4.15}$$

and there exists $a > 0$ such that

$$\mathbb{E}[|\langle f, L^Y_n\rangle| \mathbf{1}_{\Gamma^c_{a,f}}(L^Y_n) \big| L^Y_n \in A_\delta] \to_{n\to\infty} 0 \tag{4.16}$$

These statements are analogous to Lemma 4.1 and Lemma 4.2 with a fixed δ.

Proof of Lemma 4.8. Let f be in L_τ and choose a in accordance with (4.16), then

$$\begin{aligned} & |\langle f, \mu^n_{Y|A_\delta}\rangle| \\ = \ & |\mathbb{E}[\langle f, L^Y_n\rangle \big| L^Y_n \in A_\delta]| \\ \leq \ & \mathbb{E}[|\langle f, L^Y_n\rangle| \mathbf{1}_{\Gamma_{a,f}}(L^Y_n) \big| L^Y_n \in A_\delta] + \mathbb{E}[|\langle f, L^Y_n\rangle| \mathbf{1}_{\Gamma^c_{a,f}}(L^Y_n) \big| L^Y_n \in A_\delta] \\ \leq \ & a + \mathbb{E}[|\langle f, \delta_{Y_1}\rangle| \mathbf{1}_{\Gamma^c_{a,f}}(L^Y_n) \big| L^Y_n \in A_\delta] \end{aligned}$$

By (4.16), $\lim_n \mathbb{E}[|\langle f, \delta_{Y_1}\rangle| \mathbf{1}_{\Gamma^c_{a,f}}(L^Y_n) \big| L^Y_n \in A_\delta] = 0$. It follows that:

$$\sup_n |\langle f, \mu^n_{Y|A_\delta}\rangle| < \infty, \text{ for all } f \in L_\tau \text{ and } \delta > 0$$

By the uniform boudedness theorem and Banach-Alaoglu's theorem, $\{\mu^n_{Y|A_\delta}\}_{n\geq 1}$ is relatively compact and there exists a subnet of $\{\mu^n_{Y|A_\delta}\}_{n\geq 1}$ converging to $\nu_\delta \in L'_\tau$. □

Proof of Theorem 4.7. We first show that $\mathcal{C}$ is nonempty. Let us take f in L_τ. As $\ell \mapsto |\langle \ell, f\rangle|$ is continuous and $\mathcal{M}$ is a compact set, there exists $a > 0$ such that $\mathcal{M} \subset \{\ell \in L'_\tau, |\langle \ell, f\rangle| \ < \ a \ \}$. Let $\Gamma = \{\ell \in L'_\tau, |\langle \ell, f\rangle| \leq a\}$ then $\mathcal{M} \subset \overset{\circ}{\Gamma}$, and

$$\begin{aligned} |\langle f, \mu^n_{Y|A_\delta}\rangle| &= |\mathbb{E}[\langle f, L^Y_n\rangle \big| A_{\delta,n}]| \\ &\leq \mathbb{E}[|\langle f, L^Y_n\rangle|\mathbf{1}_{\Gamma_n}\big|A_{\delta,n}] + \mathbb{E}[|\langle f, L^Y_n\rangle|\mathbf{1}_{\Gamma^c_n}\big|A_{\delta,n}] \\ &\leq a + \mathbb{E}[|\langle f, \delta_{Y_1}\rangle|\mathbf{1}_{\Gamma^c_n}\big|A_{\delta,n}] \end{aligned}$$

By Lemma 4.2, $\lim_\delta \lim_n \mathbb{E}[|\langle f, \delta_{Y_1}\rangle|\mathbf{1}_{\Gamma^c}(L^Y_n)\big|L^Y_n \in A_\delta] = 0$. Thus, for any $f \in L_\tau$

$$\limsup_\delta \limsup_n |\langle f, \mu^n_{Y|A_\delta}\rangle| < \infty \tag{4.17}$$

Let us take $\delta > 0$ and let ν_δ be a cluster point of $\{\mu^n_{Y|A_\delta}\}_{n\geq 1}$. By Lemma 4.8 and Proposition A.1, there exists a subsequence $\{\mu^{\phi_\delta(n)}_{Y|A_\delta}\}_{n\geq 1}$ of $\{\mu^n_{Y|A_\delta}\}_{n\geq 1}$ such that $\langle \mu^{\phi_\delta(n)}_{Y|A_\delta}, f\rangle \to_n \langle \nu_\delta, f\rangle$. It follows that $|\langle f, \nu_\delta\rangle| \leq \limsup_n |\langle f, \mu^n_{Y|A_\delta}\rangle|$ and $\limsup_\delta |\langle f, \nu_\delta\rangle| \leq \limsup_\delta \limsup_n |\langle f, \mu^n_{Y|A_\delta}\rangle|$.
With (4.17), it comes out that for any nonincreasing sequence $\{\delta_n\}_{n\geq 1}$ converging to zero, $\limsup_n |\langle f, \nu_{\delta_n}\rangle| \leq \limsup_\delta |\langle f, \nu_\delta\rangle| < \infty$.
Thus, $\sup_n |\langle f, \nu_{\delta_n}\rangle| < \infty$ for any $f \in L_\tau$. By the uniform boudedness theorem and Banach-Alaoglu's theorem, $\{\nu_{\delta_n}\}_{n\geq 1}$ is relatively $\sigma(L'_\tau, L_\tau)$-compact and there exists a subnet of $\{\nu_{\delta_n}\}_{n\geq 1}$ converging to $\nu \in L'_\tau$. Therefore, $\mathcal{C} \subset L'_\tau$ is nonempty.

Let us show now that $\mathcal{C} \subset \overline{\text{co}}(\mathcal{M})$. Suppose that $\xi \notin \overline{\text{co}}(\mathcal{M})$. By Hahn-Banach theorem, there exists $f \in L_\tau$ such that $\langle \xi, f\rangle > a$ and $\langle \ell, f\rangle < a$ for all $\ell \in \overline{\text{co}}(\mathcal{M})$. Let $\Gamma \overset{\triangle}{=} \{\ell \in L'_\tau, \langle \ell, f\rangle \leq a\}$, then $\mathcal{M} \subset \Gamma^\circ$ and

$$\begin{aligned} \langle \mu^n_{Y|A_\delta} f\rangle - \langle \xi, f\rangle &= \mathbb{E}[\langle f, L^Y_n\rangle - \langle \xi, f\rangle | A_{\delta,n}] \\ &= \mathbb{E}[\langle L^Y_n, f\rangle \mathbf{1}_{\Gamma_n}|A_{\delta,n}] - \langle \xi, f\rangle \mu^n\big(L^Y_n \in \Gamma|A^n_\delta\big) + \epsilon(n,\delta) \end{aligned}$$

where $\epsilon(n,\delta) = \mathbb{E}[\big\{\langle L^Y_n, f\rangle - \langle \xi, f\rangle\big\}\mathbf{1}_{\Gamma^c_n}|A_{\delta,n}]$. Due to Lemma 4.1 and Lemma 4.2, $\lim_\delta \lim_n \epsilon(n,\delta) = 0$. The mere definition of Γ yields $\mathbb{E}[\langle L^Y_n, f\rangle \mathbf{1}_{\Gamma_n}|A_{\delta,n}] \leq a$. As $\overline{\text{co}}(\mathcal{M}) \subset \Gamma$, we have:
$\lim_\delta \lim_n \mu^n\big(L^Y_n \in \Gamma|A_{\delta,n}\big) = 1$ so that $\lim_\delta \lim_n \langle \xi, f\rangle \mu^n\big(L^Y_n \in \Gamma|A^n_\delta\big) > a$.
Finally, we obtain: $\liminf_\delta \liminf_n |\langle \mu^n_{Y|A_\delta} f\rangle - \langle \xi, f\rangle| > 0$, that is $\xi \notin \mathcal{C}$. And Theorem 4.7 is proved. □

4.8. **A comparison with I. Csiszár's results.** In [7], I. Csiszár has studied the convergence of $\mu^n_{Y|A}$ in an information-theoretic sense. Namely, if ν_* is the I-generalised projection of μ over the set of constraint A, then

$$\mu^n_{Y|A} \xrightarrow[n\to\infty]{\text{information}} \nu_*.$$

This means that $\frac{1}{n}H(\mu^n_{Y^n|A}|\nu^n_*) \xrightarrow[n\to\infty]{} 0$. In ([6], Lemma 3.1), it is proved that this result induces the convergence in variation of $\mu^n_{Y|A}$ to ν_* and

$$\langle \mu^n_{Y|A}, f\rangle \xrightarrow[n\to\infty]{} \langle \nu_*, f\rangle \quad \text{for all} \quad f \in L_\tau(\nu_*)$$

where $L_\tau(\nu_*) = \{f;\ \exists a > 0, \int e^{a|f|}\, d\nu_* < \infty\}$. Note that the space $L_\tau(\nu_*)$ is based on ν_* and not on μ.

In the framework of the present article, let us assume that the set of minimizers is of the following form: $\mathcal{M} = \nu_* + \mathcal{S}$. Then, ν_* is the I-generalized projection of μ on A_0 in Csiszár's sense (see [7]). If we want to compare our results with the results in [7], we have to compare the set of test functions in [7]: $L_\tau(\nu_*)$, with: $M_\tau(\mu)$ or $L_\tau(\mu)$. No general results are available.

Indeed, let $A_0 = \{\mu\}$, then $\nu_* = \mu$ and $L_\tau(\nu_*) = L_\tau(\mu)$.

On the other hand, in the scope of Csiszár's example, Theorem 3.1 in [7] yields: $\langle \mu^n_{Y|A}, f\rangle \xrightarrow[n\to\infty]{} \langle \nu_*, f\rangle$, for all $f \in L_\tau(\nu_*)$. Nevertheless, no comparison can be made between $L_\tau(\nu_*)$ and $M_\tau(\mu)$. Indeed, consider $f = \sum_1^\infty n\mathbf{1}_{[a_n,b_n]}$ where $a_n = n - e^{-n}$ and $b_n = n + e^{-n}$, then one can check that $f \in L_\tau(\nu_*)$ but $f \notin M_\tau(\mu)$. Conversely, let $f(y) = y^{3/4}$. Then, $f \in M_\tau(\mu)$ but $f \notin L_\tau(\nu_*)$.

The nature of the results in [6]-[7] and in the present article are different. Note that the convergence in an information-theoretic sense can be as strong as in Theorem 4.7 when no singular part appears. In any case, this convergence does not "see" singular parts.

References

[1] A. Aboulalaâ. The Gibbs principle for Markov jump processes. *Stoch. Proc. 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. *Stoch. Proc. Appl.*, 33:1–27, 1989.

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

[6] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.

[7] I. Csiszár. Sanov property, generalized I-projection and a conditionnal limit theorem. *The Annals of Probability*, 12(3):768–793, 1984.

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

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

[10] 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.

[11] P. Eichelsbacher and U. Schmock. Large deviations of products of empirical measures and U-empirical measures in strong topologies. Preprint, 1996.
[12] J.R. Giles. Convex analysis with application in the differentiation of convex functions. In *Research Notes in Math. 58*. Pitman, London, 1982.
[13] J. L. Kelley. *General Topology.* Springer Verlag, New York, 1955.
[14] A. Kozek. Convex integral functionals on Orlicz spaces. *Ann. Soc. Math. Polonae, Series I*, pages 109–135, 1979.
[15] C. Léonard. Minimizers of energy functionals under not very integrable constraints. Preprint, 2000.
[16] J. Ooesterhoff P. Groeneboom and F.H. Ruymgaart. Large deviation theorems for empirical probability measures. *Ann. Probab.*, 7:553–586, 1979.
[17] M. Rao and Z. D. Ren. *Theory of Orlicz spaces.* Marcel Dekker Inc., New-York, 1991.
[18] M. Reed and B. Simon. *Methods of Modern Mathematical Physics I.* Academic Press, New York, 1980.
[19] R. T. Rockafellar. Integrals which are convex functionals. *Pacific J. Math.*, 24(3):525–539, 1968.
[20] I. N. Sanov. On the probability of large deviations of random variables. *Selected transl. Math. Statist. and Prob.*, 1:214–244, 1961.
[21] A. Schied. Cramer's condition and Sanov's theorem. *Statistics and Probability Letters*, 39:55–60, 1998.
[22] 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.

Appendix A. About subnets

Recall that a subnet $(y_\alpha)_{\alpha\in A}$ of a sequence $(u_n)_{n\geq 1}$ is a family $(y_\alpha)_{\alpha\in A}$ where A is a right-filtering set such that there exists $p : A \to \mathbb{N}$ satisfying:

- $\alpha \leq \beta \quad \Rightarrow \quad p(\alpha) \leq p(\beta)$.
- for all $n \in \mathbb{N}$, there exists α_n such that $p(\alpha_n) \geq n$.
- $y_\alpha = u_{p(\alpha)}$

The subnet (y_α) is said to be convergent to y_∞ if for every neighborhood V of y_∞, there exists α_0 such that $\alpha \geq \alpha_0$ implies $y_\alpha \in V$. The differences between a subnet and a subsequence are twofold: A can be huge compared to $\mathbb{N}$ and the order $\leq$ over A is partial.

Proposition A.1. *Let $\{\mu^{p(\beta)}_{Y|A_\delta}\}_{\beta\in B}$ be a subnet converging to ν_δ. Then, for all $f \in L_\tau$, there exists a subsequence $\{\mu^{\phi_\delta(n)}_{Y|A_\delta}\}$ satisfying*

$$\lim_{n\to\infty} \langle \mu^{\phi_\delta(n)}_{Y|A_\delta}, f\rangle = \langle \nu_\delta, f\rangle \tag{A.1}$$

Proof. Consider the following neighbourhood of ν_δ, $V_n = \{\xi \in L'_\tau, |\langle \xi - \nu_\delta, f\rangle| < \frac{1}{n}\}$. The mere definition of a subnet yields: $\exists \beta_n \in B,\ p(\beta_n) \geq n,\ \mu^{p(\beta_n)}_{Y|A_\delta} \in V_n$. One can choose a nondecreasing sequence of β_n hence $\langle \mu^{p(\beta_n)}_{Y|A_\delta}, f\rangle \to \langle \nu_\delta, f\rangle$ and the proposition is proved. $\square$

Remark A.1. In the preceding proposition, it is necessary to deal with nets. By the Bolzano-Weirstass Theorem, we can extract a converging subsequence from $\langle \mu^n_{Y|A_\delta}, f\rangle$ yet we can not be sure that this subsequence converges to $\langle \nu_\delta, f\rangle$.This information is of interest since we need the following kind of control:

$$|\langle f, \nu_\delta\rangle| \leq \limsup_n |\langle f, \mu^n_{Y|A_\delta}\rangle|$$