Bernoulli **8**(6), 2002, 721–743

An extension of Sanov's theorem: application to the Gibbs conditioning principle

CHRISTIAN LÉONARD[1*2] and JAMAL NAJIM[1**]

[1]*Modal-X, Université Paris 10, Bâtiment G 200, Av. de la République, 91002 Nanterre cedex, France. E-mail:* **Christian.Leonard@u-paris10.fr,* ***najim@tsi.enst.fr*
[2]*CMAP, Ecole Polytechnique, 91128 Palaiseau cedex, France*

A large-deviation principle is proved for the empirical measures of independent and 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 Stroock and Zeitouni, the Gibbs conditioning principle (GCP) is then derived with the help of the previous result. Apart from a more direct proof than has previously been available, the main improvements with respect to GCPs already published 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 the theory of Orlicz spaces appear to be a powerful tool.

Keywords: empirical measures; Gibbs conditioning principle; large deviations; Orlicz spaces; Sanov's theorem

1. Introduction

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

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

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

1.1. Sanov's theorem

Sanov's theorem describes the limiting behaviour of $n^{-1}\log \mathbb{P}(L_n^Y \in \cdot)$ 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

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

and ∞ otherwise: this is 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 $\nu \in \mathcal{P} \mapsto \int_\Sigma f\,\mathrm{d}\nu \in \mathbb{R}$ 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:

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

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

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

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 no longer $\mathcal{P}$, but a different set $\mathcal{Q}$ of all non-negative continuous linear forms on $\mathcal{L}_\tau$ with 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 form (for any $\ell \in \mathcal{Q}$ such that $I(\ell) < \infty$)

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

where $\ell = \ell^{\mathrm{a}} + \ell^{\mathrm{s}}$ 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 $\langle \ell^{\mathrm{s}}, \mathbf{1}\rangle = 0$, although $\ell^{\mathrm{s}} \geqslant 0$ may not be zero. For more details on this subject, see Giner (1976) and Kozek (1980). 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 Savov theorem, see Theorem 3.2 below.

1.2. Gibbs conditioning principle

The Gibbs conditioning principle (GCP) describes the limiting behaviour as n tends to infinity of the law of k tagged particles $Y_1, \ldots, Y_k$ 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 $n \geqslant 1$. Typically, the expected result is

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

where ν_* minimizes $\nu \mapsto H(\nu|\mu)$ 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

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

where φ is an energy function on the one-particle phase space Σ.

The close relationship between Sanov's theorem and the GCP is well known. It has been exploited by Csiszár (1984) and by Stroock and Zeitouni (1991). As in those works, we shall not be able to handle the difficult and important case where $\mathbb{P}(L_n^Y \in A_0) = 0$ for $n \geqslant 1$. We follow Stroock and Zeitouni (1991) in 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 $n \geqslant 1$ and $\delta > 0$.

With the extended Sanov theorem in hand, rather than its usual version, the proof of the GCP is more direct and its assumptions can be significantly relaxed. On the 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 restriction, assumed by Stroock and Zeitouni (1991), that 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 GCP still holds in situations where a lack of steepness of the pressure $\beta \mapsto \log \int_\Sigma \mathrm{e}^{\beta\varphi} \, \mathrm{d}\mu$ occurs (see Csiszár's example in Sections 3.4 and 4.4). The GCP 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 in obtaining direct proofs of our main results.

1.3. Review of the literature

Sanov (1961) 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 Donsker and Varadhan (1976) and Bahadur and Zabell (1979) for the weak topology. Groeneboom *et al*. (1979) dropped the Polish requirement and considered Hausdorff spaces. They obtained Sanov's theorem for the so-called τ-topology, $\sigma(\mathcal{P}, B)$. De Acosta (1994) improved this result and simplified the proof. Csiszár (1984) proved Sanov's theorem in a general setting by means of an alternative approach based on projection in information. Eichelsbacher and Schmock (2002) consider the U-empirical measures

$$L_n^{Y,k} = \frac{1}{n \cdots (n-k+1)} \sum_{(i_1,\ldots,i_k)} \delta_{(X_{i_1},\ldots,X_{i_k})},$$

where the sum is taken over all k-tuples in $\{1, \ldots, n\}^k$ 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,

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

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 Eichelsbacher and Schmock (2002) for $k \geqslant 2$ is far from trivial and is not a consequence of our results.

As already mentioned, the GCP has been studied by Csiszár (1984) and by Stroock and Zeitouni (1991). An updated presentation of the latter study is available in Dembo and Zeitouni (1998, Section 7.3). Bolthausen and Schmock (1989) proved a GCP for the occupation measures of uniformly ergodic Markov chains. Based on Csiszár (1984), Aboulalaâ (1996) obtained a GCP for the empirical measures L_n^Y of Markov jump processes. With the LDP for $L_n^{Y,k}$ in hand, Eichelsbacher and Schmock (2002) derive a GCP, following the approach of Stroock and Zeitouni (1991). 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 $\sigma(\mathcal{P}_\tau, \mathcal{M}_\tau)$.

Csiszár (1984) obtained alternative results with a different, powerful approach. In particular, he proved the convergence in information of the conditioned laws which implies their convergence in variation, and introduced the notion of generalized *I*-projection 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 worthwhile to express exponential moment conditions ((1.1) and (1.4)) in terms of such 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 GCP in Section 4, the main statement of which is Theorem 4.2.

2. Orlicz spaces

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

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:

$$\mathcal{L}_\theta = \left\{ f : \Sigma \to \mathbb{R},\ \exists a > 0,\ \int_\Sigma \theta\left(\frac{f}{a}\right) \mathrm{d}\mu < \infty \right\},$$

$$\mathcal{M}_\theta = \left\{ f : \Sigma \to \mathbb{R},\ \forall a > 0,\ \int_\Sigma \theta\left(\frac{f}{a}\right) \mathrm{d}\mu < \infty \right\}.$$

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 also the following Luxemburg norm on L_θ:

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

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 $\|.\|_\theta$. For more details and further reading, see Adams (1975) and Rao and Ren (1991). 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 $g \in L_{\theta^*}$,

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

A Young function θ satisfies the Δ_2 condition if there exist $K > 0$ and $s_0 \geqslant 0$ such that, for all $s \geqslant s_0$, $\theta(2s) \leqslant K\theta(s)$. If θ satisfies the Δ_2 condition, then $M_\theta = L_\theta$ (see Rao and Ren 1991, 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 $\langle f, g\rangle = \int fg\, \mathrm{d}\mu$. 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, see Rao and Ren (1991).

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

As L_θ is a Riesz space (see Bourbaki 1961), 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* $A_1 \supset A_2 \supset A_3 \supset \ldots$ *of measurable sets such that*

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

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*

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

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

For a proof of this result, see Kozek (1980, Theorem 2.2). 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* $\ell^{\mathrm{s}} \in L^{\mathrm{s}}_\theta$, *we have* $\langle \ell^{\mathrm{s}}, f \rangle = 0$.

For a proof of this result, see Kozek (1980).

2.3. Orlicz spaces and exponential moment conditions

Consider

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

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

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

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

If $f \in \mathcal{L}_\tau$, we shall say that f admits *some* exponential moments; if $f \in \mathcal{M}_\tau$, 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 due to 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. 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 every G_f 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 $\{Y_i\}_{i\geqslant 1}$ is a sequence of Σ-valued independent and identically μ-distributed variables.

Henceforth, the identifications

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

prevail, where L_{τ^*} is the Orlicz space associated with the Young function τ^*, L_τ' (L_τ^*) is the topological (algebraic) dual of L_τ, and $\mathcal{L}_\tau^*$ 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 $L_{\tau^*} = L_{\tau^*} \cdot \mu$ for short. The second identification is straightforward. 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 theorem is

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

This 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 theorem is

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

where $\ell = \ell^{\mathrm{a}} + \ell^{\mathrm{s}}$ is the decomposition stated in Theorem 2.3, $D_\mu = \{f \in \mathcal{L}_\tau; \mathbb{E}\mathrm{e}^{f(Y)} < \infty\}$ 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 write $I(\ell) = I_{\mathrm{a}}(\ell) + I_{\mathrm{s}}(\ell) = I_{\mathrm{a}}(\ell^{\mathrm{a}}) + I_{\mathrm{s}}(\ell^{\mathrm{s}})$, where

$$I_{\mathrm{a}}(\ell) = \int_\Sigma \log\left(\frac{\mathrm{d}\ell^{\mathrm{a}}}{\mathrm{d}\mu}\right) \mathrm{d}\ell^{\mathrm{a}},$$

$$I_{\mathrm{s}}(\ell) = \sup\{\langle \ell^{\mathrm{s}}, f\rangle;\, f,\, \mathbb{E}\mathrm{e}^{f(Y)} < \infty\}.$$

The following theorem is the main result of the section.

Theorem 3.2 (Extended Sanov theorem). *The empirical measures $\{L_n^Y\}_{n\geqslant 1}$ 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:*

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

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

(ii) *for all measurable open subsets 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 $\{I \leqslant \alpha\}$, $\alpha \geqslant 0$, are compact.

Proof. The LDP in $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$ is proved in Lemma 3.4 with the convex rate function Θ^*. By Proposition 3.8, $\Theta^* = I$. But the domain of I is included in $\mathcal{Q}$, thus the LDP holds on $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ (see Dembo and Zeitouni 1998, Lemma 4.1.5(b)). This completes the proof of the theorem. □

Remark 3.2. Let $I(\ell) < \infty$. Then $\ell \in L_\tau'$, $\ell^{\mathrm{a}} \geqslant 0$ and, by Proposition 2.4, $\langle \ell^{\mathrm{a}}, \mathbf{1}\rangle = 1$. Hence $\mathrm{d}\ell^{\mathrm{a}}/\mathrm{d}\mu$ is a probability density and I_{a} is close to the usual relative entropy $H(\cdot|\mu)$. 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

$$\mathcal{Q} \cap L_\tau' = (L_{\tau^*} \cap \mathcal{P}) \oplus (L_\tau^{\mathrm{s}} \cap \{\ell \geqslant 0\}).$$

Remark 3.4. We cannot expect the LDP to hold under the strong topology $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$ with the entropy $H(\cdot\,|\mu)$ as a good rate function. Indeed, Schied (1998) proved that if the topology is too wide (the so-called τ_ϕ-topology, where ϕ admits only some exponential moments), then $\{H \leqslant \alpha\}$ is no longer compact. The same argument holds in our context:

$$\{\ell \in \mathcal{Q};\, H(\ell^{\mathrm{a}}) \leqslant \alpha\} = \{f\mu;\, f \in L_{\tau^*},\, f \geqslant 0,\, \int_\Sigma f\,\mathrm{d}\mu = 1,\, \int_\Sigma f \ln f\,\mathrm{d}\mu \leqslant \alpha\} + (L_\tau^{\mathrm{s}} \cap \{\ell \geqslant 0\}),$$

which is not compact.

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

This is in accordance with the result obtained by Eichelsbacher and Schmock (2002, 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 $\ell_{|\mathcal{M}_\tau} = \ell^{\mathrm{a}}_{|\mathcal{M}_\tau} = \ell^{\mathrm{a}}$ (where the last equality is an identification). Hence,

$$\inf\{I(\ell);\ \ell_{|\mathcal{M}_\tau} = \nu\} = \inf\{I_{\mathrm{a}}(\nu) + I_{\mathrm{s}}(\ell^{\mathrm{s}});\ \ell^{\mathrm{a}} = \nu\} = I_{\mathrm{a}}(\nu) = H(\nu|\mu).$$

The result follows from the obvious continuity of the transformation considered. □

3.2. Proof of the large-deviation principle

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

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

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. The proof is based on the Dawson–Gärtner projective limit approach. By Theorem 4.6.9 in Dembo and Zeitouni (1998), it is sufficient to check that, for all $d \geq 1$ and $f_1, \ldots, f_d \in \mathcal{L}_\tau$, $(\langle L_n^Y, f_1\rangle, \ldots, \langle L_n^Y, f_d\rangle)$ satisfies an LDP. But

$$(\langle L_n^Y, f_1\rangle, \ldots, \langle L_n^Y, f_d\rangle) = \frac{1}{n}\sum_{i=1}^n (f_1(Y_i), \ldots, f_d(Y_i)) = \frac{1}{n}\sum_{i=1}^n \mathbf{f}(Y_i),$$

where $\mathbf{f}(x) = f_1(x), \ldots, f_d(x))$ is a $\mathbb{R}^d$-valued function. Then, $\{\mathbf{f}(Y_i)\}$ is a sequence of independent and identically distributed, $\mathbb{R}^d$-valued random variables. Since $f_1, \ldots, f_d \in \mathcal{L}_\tau$, $\mathbf{f}(Y_i)$ admits exponential moments. By Cramér's theorem on Polish spaces (Dembo and Zeitouni 1998, Theorem 6.1.3 and Corollary 6.1.6), $n^{-1}\sum_{i=1}^n \mathbf{f}(Y_i)$ satisfies the LDP in $\mathbb{R}^d$ with the good rate function

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

By the Dawson–Gärtner theorem, $L_n^Y = n^{-1}\sum_{i=1}^n \delta_{Y_i}$ satisfies the LDP with the good rate function given, for any $\ell \in \mathcal{L}_\tau^*$, by

$$\sup\left\{\sum_{i=1}^{\mathrm{d}} \lambda_i \langle \ell, f_i \rangle - \log \mathbb{E}e^{\Sigma \lambda_i f_i(Y)};\ d \geqslant 1,\ \lambda \in \mathbb{R}^d,\ f_1, \ldots, f_d \in \mathcal{L}_\tau\right\}$$
$$= \sup_{f \in \mathcal{L}_\tau} \{\langle \ell, f \rangle - \log \mathbb{E}e^{f(Y)}\} = \Theta^*(\ell),$$

which is the desired result. □

3.3. Identification of the rate function

We prove the identity $I = \Theta^*$ in Proposition 3.8. Lemmas 3.5–3.7 are required preliminary results.

Let us consider $J(\ell) = \sup_{f \in \mathcal{L}_\tau}\{\langle \ell - \mu, f \rangle - \int_\Sigma \gamma(f)\,\mathrm{d}\mu\}$, where γ is given by (2.3).

Lemma 3.5. *Let $\ell \in \mathcal{L}_\tau^*$. Then:*

(i) $\Theta^*(\ell) < \infty \Rightarrow \ell \in L_\tau^*$;
(ii) $J(\ell) \leqslant \Theta^*(\ell)$;
(iii) $\Theta^*(\ell) < \infty \Rightarrow \ell \in \mathcal{Q} \cap L_\tau'$.

Proof. (i) Let $f \in \mathcal{L}_\tau$ be such that $f = 0$ μ-almost everywhere. Then, for any real λ, $\Theta(\lambda f) = 0$ and $\lambda\langle \ell, f \rangle \leqslant \Theta^*(\ell) + \Theta(\lambda f) = \Theta^*(\ell)$. Hence, $\Theta^*(\ell) < \infty$ implies $\langle \ell, f \rangle = 0$. Thus ℓ is constant over the equivalence classes, that is, $\ell \in L_\tau^*$

(ii) Since, for all $t \geqslant 0$, $\log t \leqslant t - 1$, we have $-\mathbb{E}e^{f(Y)} + 1 \leqslant -\log \mathbb{E}e^{f(Y)}$ and

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

(iii) As $\Theta^*(\ell) < \infty$ implies $\ell \in L_\tau^*$, let us consider $\ell \in L_\tau^*$.

For all $f \in L_\tau$, $\langle \ell - \mu, f \rangle \leqslant J(\ell) + \int_\Sigma \gamma(f)\mathrm{d}\mu$. As $\gamma(s) \leqslant \tau(s) = \gamma(|s|)$, we have $\langle \ell - \mu, f \rangle \leqslant J(\ell) + \int_\Sigma \tau(f)\mathrm{d}\mu$. Choosing $\eta = \pm 1/\|f\|_\tau$ when $f \neq 0$, the definition of the Luxemburg norm (2.1) yields $\int_\Sigma \tau(\eta f)\mathrm{d}\mu = 1$, which implies $|\langle \ell - \mu, f \rangle| \leqslant (J(\ell) + 1)\|f\|_\tau$. This inequality still holds with $\|f\|_\tau = 0$. Therefore $\ell - \mu \in L_\tau^*$ is $\|\cdot\|_\tau$-continuous: $\ell \in L_\tau'$, since $J(\ell) \leqslant \Theta^*(\ell) < \infty$.

Suppose that $\langle \ell, \mathbf{1} \rangle = a \neq 1$. Then $\Theta^*(\ell) \geqslant \langle \ell, \lambda \mathbf{1} \rangle - \log \mathbb{E}e^{\lambda \mathbf{1}} = \lambda(a - 1)$ 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 $f \geqslant 0$ with $\langle \ell, f \rangle < 0$. Let $\lambda \geqslant 0$. Then $\Theta^*(\ell) \geqslant \langle \ell, -\lambda f \rangle - \log \mathbb{E}e^{-\lambda f} \geqslant \langle \ell, -\lambda f \rangle$ tends to ∞ as λ tends to ∞. Thus $\Theta^*(\ell) < \infty$ implies $\ell \geqslant 0$ and the lemma is proved. □

Lemma 3.6. *Let $\ell \in L_\tau'$. Then, for all $f \in L_\tau$:*

(i) $\lim_{n\to\infty}\langle \ell, f_n\rangle = \langle \ell^{\mathrm{a}}, f\rangle$, *where* (f_n) *is any sequence of bounded measurable functions which converges pointwise to* f *and such that* $|f_n| \leqslant |f|$, *for all* $n \geqslant 1$.
(ii) $\lim_{n\to\infty}\langle \ell, \mathbf{1}_{\{|f|>n\}} f\rangle = \langle \ell^{\mathrm{s}}, f\rangle$.

Proof. (i) Since f_n is bounded, $\langle \ell^{\mathrm{s}}, f_n\rangle = 0$ (see Proposition 2.4). Therefore $\langle \ell, f_n\rangle = \langle \ell^{\mathrm{a}}, f_n\rangle = \int f_n \cdot (\mathrm{d}\ell^{\mathrm{a}}/\mathrm{d}\mu)\cdot \mathrm{d}\mu$, with $\mathrm{d}\ell^{\mathrm{a}}/\mathrm{d}\mu \in L_{\tau^*}$. The limit follows from the dominated convergence theorem.

(ii) We have $\langle \ell, \mathbf{1}_{\{|f|>n\}} f\rangle = \langle \ell^{\mathrm{s}}, \mathbf{1}_{\{|f|>n\}} f\rangle + \langle \ell^{\mathrm{a}}, \mathbf{1}_{\{|f|>n\}} f\rangle$. The dominated convergence theorem implies that $\lim_{n\to\infty}\langle \ell^{\mathrm{a}}, \mathbf{1}_{\{|f|>n\}} f\rangle = 0$. Since $\mathbf{1}_{\{|f|\leqslant n\}} f$ is bounded, $\langle \ell^{\mathrm{s}}, f\rangle = \langle \ell^{\mathrm{s}}, \mathbf{1}_{\{|f|>n\}} f\rangle$ (see Proposition 2.4). □

Lemma 3.7. *For all* ℓ *in* L'_τ, $\Theta^*(\ell) = \Theta^*(\ell^{\mathrm{a}}) + \sup\{\langle \ell^{\mathrm{s}}, f\rangle;\ f \in \operatorname{dom}\Theta\}$, *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 notation which is 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

$$\delta(f|A) = \begin{cases} 0 & \text{if } f \in A,\\ +\infty & \text{otherwise;}\end{cases}$$

its convex conjugate,

$$\delta^*(\ell|A) = \sup_{f\in L_\tau}\{\langle \ell, f\rangle - \delta(f|A)\} = \sup\nolimits_{f\in A}\langle \ell, f\rangle,$$

is called the support functional of A.

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

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

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

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

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

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 $\ell \in L'_\tau$, $\Theta^*(\ell) = \Theta^*(\ell^{\mathrm{a}}) + I_{\mathrm{s}}(\ell^{\mathrm{s}})$. Taking Remark 3.3 into account, it remains to prove that, for all $\ell \in \mathcal{P} \cap L_{\tau^*}$, $\Theta^*(\ell) = H(\ell|\mu)$. Let $\ell = h\mu$ belong to $\mathcal{P} \cap L_{\tau^*}$, that is, $h \in L_{\tau^*}$, $h \geqslant 0$, $\int_\Sigma h\,\mathrm{d}\mu = 1$. By a direct computation, we have, for all $f \in \operatorname{dom}\Theta$,

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

Therefore,

$$\begin{aligned}
\Theta^*(h\mu) &= \sup_{f\in\operatorname{dom}\Theta}\{\langle h\mu, f\rangle - \Theta(f)\} \\
&= \sup_{\lambda\in\mathbb{R}, f\in\operatorname{dom}\Theta}\left\{\langle h\mu, f\rangle + \lambda + 1 - \mathrm{e}^\lambda\int_\Sigma \mathrm{e}^f\,\mathrm{d}\mu\right\} && (3.3a)\\
&= \sup_{\lambda\in\mathbb{R}, f\in\operatorname{dom}\Theta}\left\{\langle h\mu, \lambda + f\rangle - \int_\Sigma (\mathrm{e}^{\lambda+f} - 1)\,\mathrm{d}\mu\right\} && (3.3b)\\
&= \sup_{g\in\operatorname{dom}\Theta}\left\{\int_\Sigma hg\,\mathrm{d}\mu - \int_\Sigma (\mathrm{e}^g - 1)\,\mathrm{d}\mu\right\} \\
&= \int_\Sigma (h\log h - h + 1)\,\mathrm{d}\mu && (3.3c)\\
&= \int_\Sigma h\log h\,\mathrm{d}\mu = H(h\mu|\mu), && (3.3d)
\end{aligned}$$

where (3.3a) comes from (3.2), (3.3b) and (3.3d) follow from the fact that μ and $h\mu$ are probability measures and (3.3c) follows from a general result due to Rockafellar (1968, Theorem 2), noting that the convex conjugate of $\mathrm{e}^s - 1$ is $t\log t - t + 1$. This completes the proof. □

3.4. Csiszár's example

In this subection, 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 Csiszár (1984, Example 3.2) and in Dembo and Zeitouni (1998, Exercise 7.3.11). Let μ be the probability measure on $\Sigma = [0, \infty)$ defined by $\mu(\mathrm{d}y) = C\mathrm{e}^{-y}(1 + y^3)^{-1}\,\mathrm{d}y$. Let $\{Y_i\}$ be a sequence of independent and identically distributed, $[0, \infty)$-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 $H(\cdot\,|\mu)$ 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 $\Lambda(\lambda) = \log\int_{[0,\infty)} c\mathrm{e}^{(\lambda-1)y}(1 + y^3)^{-1}\,\mathrm{d}y$. One can ask if the contraction principle holds between L_n^Y and $\hat{S}_n$. Let us write $u\colon [0, \infty) \to [0, \infty)$,

$u(y) = y$, so that $\langle L_n^Y, u\rangle = \hat{S}_n$. As u is not bounded, one cannot apply the contraction principle to the usual Sanov theorem to obtain

$$\inf\{H(\nu|\mu),\, \nu \in \mathcal{P},\, \langle \nu, u\rangle = x\} = \Lambda^*(x), \qquad x \geqslant 0. \tag{3.4}$$

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. In the notation 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 observes 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 contraction principle, $G_u(L_n^Y) = n^{-1}\sum_{i=1}^n Y_i = \hat{S}_n$ 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 (Dembo and Zeitouni 1998, Lemma 4.1.4), $I' = \Lambda^*$. Therefore, the identification

$$\Lambda^*(x) = \inf\{I(\ell);\, \ell \in \mathcal{Q};\, \langle \ell, u\rangle = x\} = I(\ell_x) \tag{3.5}$$

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

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

(i) *For any* $0 < x < x_*$, *there exists a unique minimizer* ℓ_x *in* (3.5). *It is given by* $\ell_x = \nu_x$, *where*

$$\nu_x(\mathrm{d}y) = \exp(\lambda_x y - \Lambda(\lambda_x))\mu(\mathrm{d}y)$$

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

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

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

(iii) *For all* $x \geqslant x_*$ *and all minimizing arguments* ℓ_x *of* (3.5)*, we have* $\ell_x^{\mathrm{a}} = \nu_*$. *Moreover,* $\langle \nu_*, u\rangle = x_*$, $\langle \ell_x^{\mathrm{s}}, u\rangle = x - x_*$ *and*

$$I(\ell_x) = H(\nu_*|\mu) + I_{\mathrm{s}}(\ell_x^{\mathrm{s}}), \tag{3.6}$$

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

This proposition means that when $x > x_*$, the minimizers of (3.5) cannot be probability measures. The contribution of the absolutely continuous part vanishes at x_*: $\langle \nu_*, u\rangle = x_*$. It is the singular parts (not unique in general, see Proposition 4.4) which fill the gap between x_* and x: $\langle \ell_x^{\mathrm{s}}, u\rangle = x - x_*$. Moreover, the contribution of these singular parts appears in the rate function (see (3.6)). Finally, $I_{\mathrm{s}}(\ell^{\mathrm{s}}) = \sup_{f \in D_\mu}\langle \ell^{\mathrm{s}}, f\rangle$ implies that these singular parts are non-null whenever $x > x_*$.

Proof. We prove (i) and (ii) together. Clearly, for all $0 < x \leqslant x_*$, 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, $\ell = \ell^{\mathrm{a}} + \ell^{\mathrm{s}}$, with $\ell^{\mathrm{a}} \in \mathcal{P}$ and $\ell^{\mathrm{s}} \geqslant 0$. Let $\langle \ell^{\mathrm{a}}, u \rangle = x'$. We have $\langle \ell^{\mathrm{s}}, u \rangle = x - x' \geqslant 0$ and

$$\begin{aligned} I(\ell) - I(\nu_x) &= H(\ell^{\mathrm{a}}|\mu) - H(\nu_x|\mu) + I(\ell^{\mathrm{s}}) \\ &= H(\ell^{\mathrm{a}}|\nu_x) + \int \log\left(\frac{\mathrm{d}\nu_x}{\mathrm{d}\mu}\right) \mathrm{d}(\ell^{\mathrm{a}} - \nu_x) + I(\ell^{\mathrm{s}}) \\ &= H(\ell^{\mathrm{a}}|\nu_x) + I(\ell^{\mathrm{s}}) - \lambda_x(x - x') \\ &\geqslant H(\ell^{\mathrm{a}}|\nu_x) + \langle \ell^{\mathrm{s}}, u \rangle - \lambda_x(x - x') &\quad (3.7\mathrm{a}) \\ &\geqslant H(\ell^{\mathrm{a}}|\nu_x) &\quad (3.7\mathrm{b}) \\ &\geqslant 0, \end{aligned}$$

where (3.7a) follows from $I(\ell^{\mathrm{s}}) = \sup\{\langle \ell^{\mathrm{s}}, v \rangle;\ v \in D_\mu\} \geqslant \langle \ell^{\mathrm{s}}, u \rangle$ and (3.7b) follows from $\langle \ell^{\mathrm{s}}, u \rangle - \lambda_x(x - x') = (1 - \lambda_x)(x - x') \geqslant 0$.

For the equality to hold, it is necessary that $\ell^{\mathrm{a}} = \nu_x$. Hence $I(\ell^{\mathrm{s}}) = 0$, which in turn implies that $\ell^{\mathrm{s}} = 0$. Finally, ν_x is the unique minimizer of (3.5).

We now turn to the proof of (iii). For all $\lambda \leqslant 1$, both $\Lambda(\lambda)$ and $\Lambda'(\lambda)$ are finite; when $\lambda > 1$, $\Lambda(\lambda) = \infty$. 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

$$\Lambda^*(x) = \Lambda^*(x_*) + x - x_*,\ x \geqslant x_*. \tag{3.8}$$

The rest of the proof is divided into three steps.

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

$$\forall x, y \geqslant x_*, \qquad I(\alpha\ell_x + \beta\ell_y) = \alpha I(\ell_x) + \beta I(\ell_y). \tag{3.9}$$

By definition of ℓ_x and ℓ_y, we obtain $\langle \ell_x, u \rangle = x$ and $\langle \ell_y, u \rangle = y$ and, by (3.8), $I(\ell_x) = \Lambda^*(x) = (x - x_*) + \Lambda^*(x_*)$. Similarly, $I(\ell_y) = (y - x_*) + \Lambda^*(x_*)$. The convexity of I implies that

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

But $I'(\alpha x + \beta y) = \inf\{I(\ell);\ \ell \in L'_\tau;\ \langle \ell, u \rangle = \alpha x + \beta y\}$ and $\alpha\ell_x + \beta\lambda_y$ satisfies the constraint $\langle \alpha\ell_x + \beta\ell_y, u \rangle = \alpha x + \beta y$. Thus $I'(\alpha x + \beta y) \leqslant I(\alpha\ell_x + \beta\ell_y)$ and (3.9) holds.

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

By the definition of I, we have:

$$I(\alpha\ell_x + \beta\ell_y) = I_\mathrm{a}(\alpha\ell_x^\mathrm{a} + \beta\ell_y^\mathrm{a}) + I_\mathrm{s}(\alpha\ell_x^\mathrm{s} + \beta\ell_y^\mathrm{s}),$$

$$\alpha I(\ell_x) + \beta I(\ell_y) = \alpha I_\mathrm{a}(\ell_x^\mathrm{a}) + \beta I_\mathrm{a}(\ell_y^\mathrm{a}) + \alpha \mathrm{I}_\mathrm{s}(\ell_x^\mathrm{s}) + \beta I_\mathrm{s}(\ell_y^\mathrm{s}).$$

The convexity of I_a and I_s implies

$$\begin{aligned} I_\mathrm{a}(\alpha\ell_x^\mathrm{a} + \beta\ell_y^\mathrm{a}) &\leqslant \alpha I_\mathrm{a}(\ell_x^\mathrm{a}) + \beta I_\mathrm{a}(\ell_y^\mathrm{a}), \\ I_\mathrm{s}(\alpha\ell_x^\mathrm{s} + \beta\ell_y^\mathrm{s}) &\leqslant \alpha I_\mathrm{s}(\ell_x^\mathrm{s}) + \beta I_\mathrm{s}(\ell_y^\mathrm{s}). \end{aligned} \tag{3.10}$$

By (3.9), $I(\alpha\ell_x + \beta\ell_y) = \alpha I(\ell_x) + \beta I(\ell_y)$. Therefore, equality must hold in (3.10). But due to the strict convexity of I_a, $I_\mathrm{a}(\alpha\ell_x^\mathrm{a} + \beta\ell_y^\mathrm{a}) = \alpha I_\mathrm{a}(\ell_x^\mathrm{a}) + \beta I_\mathrm{a}(\ell_y^\mathrm{a})$ implies that $\ell_x^\mathrm{a} = \ell_y^\mathrm{a} = \nu$.

Step 3. Let us show that, for all $x \geqslant x_*$, $I_\mathrm{a}(\ell_x^\mathrm{a}) = \Lambda^*(x_*)$ and $I_\mathrm{s}(\ell_x^\mathrm{s}) = x - x_*$. Considering $\nu_*(\mathrm{d}y) = \mathrm{e}^{y-\Lambda(1)}\mu(\mathrm{d}y)$, one shows that $\langle \nu_*, u\rangle = x_*$, $I(\nu_*) = I_\mathrm{a}(\nu_*) = \Lambda^*(x_*)$. Hence, ν_* satisfies (3.5) at x_*. Thus, $\nu_* = \nu$ and for all $x \geqslant x_*$, $\ell_x^\mathrm{a} = \nu_*$. It follows that $I_\mathrm{a}(\ell_x^\mathrm{a}) = \Lambda^*(x_*)$ and $I_\mathrm{s}(\ell_x^\mathrm{s}) = x - x_*$. □

Proposition 3.10. *Equality* (3.4) *holds for all* $x \geqslant 0$.

Remark 3.6. In the proof below, we show that

$$\nu_n = \left(1 - \frac{1}{n}\right)\nu_* + \frac{1}{n}\frac{\mathbf{1}_{I_n}}{\mu(I_n)}\mu \tag{3.11}$$

is a sequence satisfying

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

In Proposition 3.9(iii), it is shown that the minimizers ℓ_x have the form $\ell_x = \nu_* + \ell_x^\mathrm{s}$. Therefore, the second term on the right-hand side of (3.11) contributes asymptotically to ℓ_x^s in the sense that

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

Proof. 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 $0 < x \leqslant x_*$, the desired equality is a consequence of (i) and (ii) in Proposition 3.9. Let us now consider the case $x > x_*$. First note that

$$\begin{aligned} \inf\{H(\nu|\mu), \nu \in \mathcal{P}, \langle \nu, u\rangle = x,\} &= \inf\{I(\nu), \nu \in \mathcal{P}, \langle \nu, u\rangle = x\} \\ &\geqslant \inf\{I(\ell), \ell \in \mathcal{Q}, \langle \ell, u\rangle = x\} = \Lambda^*(x). \end{aligned}$$

In particular, $H(\nu|\mu) \geqslant \Lambda^*(x)$ if $\langle \nu, u\rangle = x$ (in fact, this is a strict inequality by Proposition 3.9(iii)). Therefore, it is sufficient to exhibit a minimizing sequence (ν_n) satisfying $\nu_n \in \mathcal{P}$, $\langle \nu_n, u\rangle = x$ and $\lim_{n\to\infty} H(\nu_n|\mu) = \Lambda^*(x)$. We take it in the form (3.11), 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

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

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

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

where f is the density of μ. Simple computations yield

$$\phi(0) > x_* + n(x - x_*), \qquad \phi(1) < x_* + n(x - x_*).$$

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

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

and

$$\frac{\mathrm{d}\nu_n}{\mathrm{d}\mu}(y) = \mathrm{e}^{y-\Lambda(1)}\left(1 + \frac{1}{n}\right) + \frac{1_{I_n}(y)}{n\mu(I_n)}. \tag{3.12}$$

We shall use the inequality

$$(a + b)\log(a + b) \leqslant (a \log a)\left(1 + \frac{b}{a}\right)^2, \qquad \forall a \geqslant \mathrm{e},\ b > 0. \tag{3.13}$$

Expressions (3.12) and (3.13) yield

$$\int_{I_n} \log\left(\frac{\mathrm{d}\nu_n}{\mathrm{d}\mu}\right)\mathrm{d}\nu_n \leqslant \int_{I_n} \frac{1}{n\mu(I_n)} \log\left(\frac{1}{n\mu(I_n)}\right)\left(1 + \mathrm{e}^{y-\Lambda(1)}\left(1 + \frac{1}{n}\right)n\mu(I_n)\right)^2 \mathrm{d}\mu(y).$$

But if $y \in I_n$ then

$$\mathrm{e}^{y-\Lambda(1)}\left(1 + \frac{1}{n}\right)n\mu(I_n) \underset{n\to\infty}{\longrightarrow} 0.$$

On the other hand,

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{n\mu(I_n)} = x - x_*.$$

Consequently,

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

and the proof is complete. □

4. The Gibbs conditioning principle

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

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

This is stated in Theorem 4.2, the main result of the section. This result holds true without any underlying law of large numbers (LLN) – that is to say, the equation

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

might not hold, 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 subsets of $\mathcal{Q}$ specified in Assumptions A and B below.

These features are illustrated via Csiszár's example in Section 4.4. In our presentation, we shall closely follow the framework of Dembo and Zeitouni (1998, Section 7.3.5).

4.1. Notation and assumptions

As before, let us consider $L_n^Y=n^{-1}\sum_{i=1}^n\delta_{Y_i}\in\mathcal{L}_\tau^*$, where $\{Y_i\}_{i\geqslant 1}$ 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 $\mathcal{E}_\mathcal{Q}$:

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

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

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

For $k=1$, we write $\mu^n_{Y^k|A_\delta}=\mu^n_{Y|A_\delta}$. We follow Stroock and Zeitouni (1991) in 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 B 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. *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*

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

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

(i) $A_0 \subset A_\delta^\circ$, for all $\delta > 0$;
(ii) $A_\delta = A_0$ 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 Assumption A are both wider than the usual τ-topology $\sigma(\mathcal{P}, \mathcal{B})$, and the σ-field $\mathcal{B}^{cy}$ (see Dembo and Zeitouni 1998, Section 6.2). Hence more open sets and more measurable sets are available. As an example, consider the family defined by

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

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

The following assumption is the counterpart of assumption (A-1) in Dembo and Zeitouni (1998, Section 7.3).

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

Remark 4.2. Equation (4.1), which is part of Dembo and Zeitouni's (1998) assumptions (see also Eichelsbacher and Schmock 2002, Condition 1.16), enforces an LLN under the minimizing law ν_* which appears in (1.2). This is not required in the present approach: by Assumption A, one can apply Sanov's lower bound (see the proof of Lemma 4.3 below) so that no underlying LLN is necessary. Moreover, there exist cases where (4.1) fails while Assumption B 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 $\ell = \nu_* + \ell^{\mathrm{s}}$, where ν_* is the absolutely continuous part of ℓ and $\ell^{\mathrm{s}} \in \mathcal{S}$ is the singular part of ℓ.

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 (1984).

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

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

for all $\alpha, \beta \geqslant 0$ such that $\alpha + \beta = 1$. Similarly, as I_{a} and I_{s} are convex, we obtain:

$$I_{\mathrm{a}}(\alpha\ell^{\mathrm{a}} + \beta\tilde{\ell}^{\mathrm{a}}) \leqslant \alpha I_{\mathrm{a}}(\ell^{\mathrm{a}}) + \beta I_{\mathrm{a}}(\tilde{\ell}^{\mathrm{a}}),$$

$$I_{\mathrm{a}}(\alpha\ell^{\mathrm{a}} + \beta\tilde{\ell}^{\mathrm{s}}) \leqslant \alpha I_{\mathrm{s}}(\ell^{\mathrm{s}}) + \beta I_{\mathrm{s}}(\tilde{\ell}^{\mathrm{s}}).$$

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

We shall see in Section 4.4 that wihtin 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 subsection, it is assumed that Σ is a separable metric space and that its σ-field is its Borel σ-field. Denote by $C_{\mathrm{b}}(\Sigma^k)$ the set of continuous and bounded functions over Σ^k. The following theorem is the counterpart of Corollary 7.3.5 in Dembo and Zeitouni (1998). It is a corollary of Lemma 4.3.

Theorem 4.2. *Suppose that Assumptions A and B hold, Σ is a separable metric space and the constraint set A_0 is convex. Then, for all f in $C_{\mathrm{b}}(\Sigma^k)$, 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 Dembo and Zeitouni's Corollary 7.3.5 in two directions:

(1) The constraint sets A_δ can be based on functions with possibly infinite exponential moments, for instance

$$A_\delta = \{\ell \in \mathcal{Q}; |\langle \ell, u\rangle - 1| \leqslant \delta\} \qquad \text{with } u \in \mathcal{L}_\tau.$$

Such functions u may grow quite fast.

(2) It has been previously remarked that Assumptions A and B do not require an underlying LLN. For an illustration of why this is useful, 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,

$$\langle \mu^n_{Y|A_\delta}, f\rangle \to \langle \nu_*, f\rangle \qquad \text{for } f \in M_\tau.$$

However, this result relies on a finer estimate than Lemma 4.3 below. The estimate and the convergence theorem can be found in Najim (2001, Lemma 2.10 and Theorem 2.11).

The following lemma is the counterpart of Dembo and Zeitouni's (1998) Theorem 7.3.3.

Lemma 4.3. *Suppose that Assumptions A and B hold. Then $\mathcal{M}$ is a non-empty $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$-compact subset of $\mathcal{Q}$ and, for any open measurable subset $\Gamma \in \mathcal{E}_\mathcal{Q}$ with $\mathcal{M} \subset \Gamma$, we have*

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

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

$$\limsup_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mu^n(L^y_n \notin \Gamma | L^Y_n \in A_\delta)$$
$$\leqslant \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log Q_n(\Gamma^c \cap A_\delta) - \lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log Q_n(A_\delta). \quad (4.4)$$

With the help of the upper and lower bounds of Theorem 3.2, the same argument as in Dembo and Zeitouni's (1998) – Sanov's upper bound – yields

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

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

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(A_\delta) \geqslant -\inf\{I(\ell), \ell \in A^\circ_\delta\}. \quad (4.6)$$

Combining these arguments with (4.3), we obtain

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

The proof of the lemma is completed by using this inequality together with (4.5) in (4.4)

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 $f(x_1, \ldots, x_k) = \prod_{i=1}^k f_i(x_i)$ be fixed, where each $f_i \in C_b(\Sigma)$. By the definition of $\mu^n_{Y^k|A_\delta}$ (see (4.2)),

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

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 $\Gamma(\eta)$ satisifies 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 $\ell = \nu_* + \ell^{\mathrm{s}}$ by the assumption on $\mathcal{M}$. As $f_i \in M_\tau$ for $1 \leqslant i \leqslant k$, $\langle \ell^{\mathrm{s}}, f_i \rangle = 0$ by Proposition 2.4. Hence $\langle \ell, f_i \rangle = \langle \nu_*, f_i \rangle$ and $\mathcal{M} \subset (\eta)$. Therefore,

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

by Lemma 4.3. The rest of the proof follows step by step the proof of Dembo and Zeitouni's (1998) Corollary 7.3.5. □

4.4. Csiszár's example revisited

Within the scope of Section 3.4, we are interested in the limiting behaviour of $\mu^n_{Y|A_\delta(x)}$ where

$$A_\delta(x) = \{\ell \in \mathcal{Q};\ |\langle \ell, u \rangle - x| \leqslant \delta\} \qquad \text{with } u(y) = y,$$

$$A_0(x) = \{\ell \in \mathcal{Q};\ \langle \ell, u \rangle = x\}.$$

The sets of constraints are

$$\{L_n^Y \in A_\delta(x)\} = \left\{ (y_1, \ldots, y_n);\ \left| \frac{1}{n} \sum_1^n y_i - x \right| \leqslant \delta \right\}$$

and $\mu^n_{Y|A_\delta(x)}$ represents the law of Y_1 under the constraint that the mean $n^{-1}\Sigma_1^n Y_i$ is close to x. Let us denote by $\mathcal{M}_x$ the corresponding set of minimizers of (3.5).

Propostion 4.4. *For any $x \geqslant x_*$, ℓ belongs to $\mathcal{M}_x$ if and only if*

(i) $\ell^{\mathrm{a}} = \nu_*$ *with* $\nu_*(\mathrm{d}y) = \mathrm{e}^{y - \Lambda(1)} \mu(\mathrm{d}y)$,
(ii) $\langle \ell^{\mathrm{s}}, u \rangle = x - x_*$, *where* $u(y) = y$, $y \geqslant 0$,
(iii) $\sup\{\langle \ell^{\mathrm{s}}, f \rangle;\ f,\ \int_{[0,\infty)} \mathrm{e}^f \, \mathrm{d}\mu < \infty\} = \langle \ell^{\mathrm{s}}, u \rangle$.

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 has already been proved in Proposition 3.9.

Let us show that there are infinitely many minimizers when $x > x_*$. Because of item (iii) of the proposition, it is sufficient to prove that the gauge function $p(g) = \inf\{\lambda > 0;\ g/\lambda \in D_\mu\}$ of $D_\mu = \{f;\ \int_{[0,\infty)} \mathrm{e}^f \mathrm{d}\mu < \infty\}$ is not Gâteaux-differentiable at u (for this argument see, for instance, Giles 1982, p. 123). This means that there exists $f \in \mathcal{L}_\tau$ such that

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

Consider

$$f(y) = \begin{cases} ay & \text{if } y \in \cup_{n\geqslant 0}[2n, 2n+1), \\ -by & \text{if } y \in \cup_{n\geqslant 0}[2n+1, 2n+2), \end{cases}$$

where $a \neq b$, $a > 0$, $b > 0$. A straightforward computation yields

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

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

For an alternative proof and further details, see Léonard (2002, Section 6).

Applying Theorem 4.2 to $\mu^n_{Y|A_\delta(x)}$, we see that for all $x > x_*$ such that $\mathcal{M}_x = \nu_* + \mathcal{S}_x$, and for any $f \in C_{\mathrm{b}}([0, \infty))$, $\langle \mu^n_{Y|A_\delta(x)}, f\rangle$ 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

$$\lim_{n\to\infty} \nu^n_*\{L^Y_n \in A_\delta(x_*)\} = 1 \qquad \text{for } \delta > 0,$$

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

Hence (4.1), which enforces an LLN under ν_*, is not satisfied when $x > x_*$, whereas the convergence of $\mu^n_{Y|A_\delta(x)}$ towards ν_* still occurs. Note that the approach developed by Csiszár (1984), which is based on convergence in information, also does not rely on such a restrictive LLN assumption.

References

Aboulalaâ, A. (1996) The Gibbs principle for Markov jump processes. *Stochastic Process. Appl.*, **64**, 257–271.

Adams, R.A. (1975) *Sobolev Spaces*. New York: Academic Press.

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

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

Bourbaki, N. (1961) *Eléments de Mathématique – Intégration 1–4*, 2nd edn. Paris: Hermann.

Csiszár, I. (1984) Sanov property, generalized *I*-projection and a conditional limit theorem. *Ann. Probab.*, **12**, 768–793.

de Acosta, A. (1994) On large deviations of empirical measures in the τ-topology. In J. Galambos and J. Gani (eds), *Studies in Applied Probability: Essays in Honour of Lajos Takács*, Journal of Applied Probability special volume 31A, pp. 41–47. Sheffield: Applied probability Trust.

Dembo, A. and Zeitouni, O. (1998) *Large Deviations Techniques and Applications*, 2nd edn. New York: Springer-Verlag.
Donsker, M.D. and Varadan, S.R.S. (1976) Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math*., **29**, 389–461.
Eichelsbacher, P. and Schmock, U. (2002) Large deviations of *U*-empirical measures in strong topologies and applications. *Ann. Inst. H. Poincaré Probab. Statist*., **38**, 779–797.
Giles, J.R. (1982) *Convex Analysis with Application in the Differentiation of Convex Functions*, Research Notes in Math. 58. Boston: Pitman.
Giner, E. (1976) Décomposabilité et dualité des espaces d'Orlicz. In *Travaux du Séminaire d'Analyse Convexe*, Vol. 6. Montpellier: Université des Sciences et Techniques du Languedoc, UER des Mathématiques.
Groeneboom, P., Oosterhoff, J. and Ruymgaart, F.H. (1979) Large deviation theorems for empirical probability measures. *Ann. Probab.* **7**, 553–586.
Kozek, A. (1980) Convex integral functionals on Orlicz spaces. *Comment. Math. Prace Mat*., **21**, 109–135.
Léonard, C. (2002) Minimizers of energy functionals under not very integrable constraints. *J. Convex Anal*., **9**(2).
Najim, J. (2001) Grandes déviations pour certaines mesures empiriques. Doctoral thesis, Université de Paris 10, Nanterre.
Rao, M. and Ren, Z.D. (1991) *Theory of Orlicz Spaces*. New York: Marcel Dekker.
Rockafellar, R.T. (1968) Integrals which are convex functionals. *Pacific J. Math*., **24**, 525–539.
Sanov, I.N. (1961) On the probability of large deviations of random variables. *Select. Transl. Math. Statist. Probab*., **1**, 213–244.
Schied, A. (1998) Cramér's condition and Sanov's theorem. *Statist. Probab. Lett*., **39**, 55–60.
Stroock, D.W. and Zeitouni, O. (1991) Microcanonical distributions, Gibbs states, and the equivalence of ensembles. In R. Durrett and H. Kesten (eds), *Festchrift in Honour of F. Spitzer*, pp. 399–424. Basel: Birkhäuser.

Received April 2001 and revised February 2002