ESAIM: PS
October 2005, Vol. 9, p. 283–306
DOI: 10.1051/ps:2005016

ESAIM: Probability and Statistics

CONDITIONAL PRINCIPLES FOR RANDOM WEIGHTED MEASURES

Nathael Gozlan[1]

Abstract. In this paper, we prove a conditional principle of Gibbs type for random weighted measures of the form $L_n = \frac{1}{n}\sum_{i=1}^n Z_i \delta_{x_i^n}$, $(Z_i)_i$ being a sequence of i.i.d. real random variables. Our work extends the preceding results of Gamboa and Gassiat (1997), in allowing to consider thin constraints. Transportation-like ideas are used in the proof.

Mathematics Subject Classification. 60E15, 60F10.

Received October 25, 2004. Revised April 7, 2005.

1. Introduction

1.1. Convex methods for solving ill posed inverse problems

Consider the so called Moment-Problem: *Find a finite measure Q on $\mathcal{X}$ satisfying*

$$\int_{\mathcal{X}} \Phi(x)\,\mathrm{d}Q(x) \in C,$$

where $\mathcal{X}$ is a Polish space, $\Phi = (\varphi_1, \ldots, \varphi_k)$ a vector valued function and C a convex subset of $\mathbb{R}^k$. Such problems appear in many physical contexts such as tomography, spectroscopy, astronomy, etc.

In order to select an element of

$$\mathcal{S}(\Phi, C) := \left\{ Q \in \mathcal{M}(\mathcal{X}) : \int_{\mathcal{X}} \Phi(x)\,\mathrm{d}Q(x) \in C \right\},$$

where $\mathcal{M}(\mathcal{X})$ is the set of finite Borel measures on $\mathcal{X}$, a classical method consists to choose as a solution the measure Q^* that minimises a certain convex cost function $\mathrm{I}(\,.\,)$ over $\mathcal{S}(\Phi, C)$.

When dealing with probability measures, one of the most popular methods is the Minimization of Entropy method (ME), *i.e.* $\mathrm{I}(\,.\,)$ is defined as the Kullback-Leibler distance with respect to some reference probability R on $\mathcal{X}$:

$$\mathrm{I}(P) = \mathrm{H}\,(\,P\,|\,R\,) := \begin{cases} \int_{\mathcal{X}} \frac{\mathrm{d}P}{\mathrm{d}R} \log\left(\frac{\mathrm{d}P}{\mathrm{d}R}\right) \mathrm{d}R & \text{if } P \ll R, P \in \mathcal{P}(\mathcal{X}) \\ +\infty & \text{otherwise,} \end{cases}$$

Keywords and phrases. Large deviations, transportation cost inequalities, conditional laws of large numbers, minimum entropy methods.

[1] Université Paris 10-Nanterre, équipe MODAL'X, UFR SEGMI, 200 avenue de la République, 92001 Nanterre, Cedex, France; `nathael.gozlan@u-paris10.fr`

where $\mathcal{P}(\mathcal{X})$ denotes the set of probability measures on $\mathcal{X}$.

In the renowned articles [5, 6], I. Csiszar derived precise results on the algebraic form of the minimizer (the so called *I-projection*) and in [7], he gave an axiomatic justification for the ME method.

More recently in [2, 3], Borwein and Lewis have studied the minimization of γ-divergences under linear constraints, that is the minimization of functionals $\mathrm{I}(\,.\,)$ having the following form:

$$\mathrm{I}(Q) = \int_{\mathcal{X}} \gamma\left(\frac{\mathrm{d}Q_a}{\mathrm{d}R}\right) \mathrm{d}R + b_\psi Q_s^+(\mathcal{X}) - a_\psi Q_s^-(\mathcal{X}),$$

where R is a probability measure on $\mathcal{X}$, $\gamma : \mathbb{R} \to [0, +\infty]$ is a convex function, Q_a is the absolutely continuous part of Q with respect to R, and $Q_s = Q_s^+ - Q_s^-$ is the Jordan decomposition of the singular part of Q (see Sect. 2 for the definition of a_ψ and b_ψ). For these functionals, they obtained precise results on the algebraic expression of the minimizers (see [2, 3] and [8] Ths. 2.2 and 2.4). The interest of γ-divergences lies in the fact that a good choice of γ makes it possible to impose additional non-linear constraints to the density of the solution (see [8] for more information on this subject).

1.2. **A probabilistic interpretation of these methods**

Large deviations theory furnishes a nice interpretation of relative entropy and I-projections: Sanov Theorem and Gibbs Conditioning Principle.

Let us briefly recall these well known results:

- (**Sanov theorem**) If $(X_i)_i$ is a sequence of independent and identically distributed random variables with law R taking values in some polish space $\mathcal{X}$, then the empirical distribution

$$N_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \tag{1.1}$$

  satisfies a large deviations principle with good rate function $\mathrm{H}(\,.\,|\,R\,)$ in $\mathcal{P}(\mathcal{X})$ equipped with the τ-topology. (see [10], Chap. 6).

- (**I-projections**) For every subset A of $\mathcal{P}(\mathcal{X})$, define

$$\mathrm{H}(\,A\,|\,R\,) := \inf\{\mathrm{H}(\,P\,|\,R\,) : P \in A\}.$$

  If A is a convex subset of $\mathcal{P}(\mathcal{X})$ such that $\mathrm{H}(\,A\,|\,R\,) < +\infty$, a probability measure $R^* \in A$ such that

$$\mathrm{H}(\,R^*\,|\,R\,) = \mathrm{H}(\,A\,|\,R\,)$$

  is called the *I-projection of R on A*. Thanks to the strict convexity of $\mathrm{H}(\,.\,|\,R\,)$, if such R^* exists, it is unique. If A is a convex subset that is closed for the τ-topology, then R admits an I-projection R^* on A (this is an easy consequence of the lower semi-continuity of $\mathrm{H}(\,.\,|\,R\,)$ and the compactness of the sublevel sets $\{\mathrm{H}(\,.\,|\,R\,) \leq t\}, t \geq 0$).

- (**Gibbs conditioning principle**) Let A be a measurable subset of $\mathcal{P}(\mathcal{X})$ that is closed and convex and suppose that $\mathrm{H}(\,A\,|\,R\,) = \mathrm{H}(\overset{\circ}{A}|R) < +\infty$, then $R^{\otimes n}(N_n \in A) > 0$ for all n sufficiently large, and

$$\mathcal{L}(X_1|N_n \in A) = \frac{\mathbb{E}_{R^{\otimes n}}[N_n \mathbb{1}_A(N_n)]}{R^{\otimes n}(N_n \in A)} \xrightarrow[n\to+\infty]{} R^*,$$

  where R^* is the I-projection of R on A (see [10] Sect. 7.3).

In other words, Gibbs Conditioning Principle expresses: *When forcing the empirical measure of* $(X_1, X_2, \ldots, X_n)$ *to belong to* A*, the law of* X_1 *is modified in such a way that it converges to the I-projection of* R *on* A*.*

In [15], Gamboa and Gassiat have established that a large classe of γ-divergences enjoys the same kind of properties: they govern the large deviations of random measures and in this framework some type of Gibbs conditional principle holds.

Before stating their results, let us introduce some notations:

For every probability measure ν on $\mathbb{R}^d$, let Z_ν, Λ_ν and Λ_ν^* denote respectively the Laplace transform, the logarithmic moment generating function and the Cramer transform of ν, defined by:

$$\begin{aligned}
&\forall s \in \mathbb{R}^d, \quad Z_\nu(s) = \int \exp\langle s, x\rangle \,\mathrm{d}\nu(x) \in \mathbb{R}^+ \cup \{+\infty\}\\
&\forall s \in \mathbb{R}^d, \quad \Lambda_\nu(s) = \log Z_\nu(s) \in \mathbb{R} \cup \{+\infty\}\\
&\forall t \in \mathbb{R}^d, \quad \Lambda_\nu^*(t) = \sup_{s\in\mathbb{R}^n} \{\langle s, t\rangle - \Lambda_\nu(s)\} \in \mathbb{R}^+ \cup \{+\infty\}.
\end{aligned}$$

Recall that the domain of a convex function $f : V \to \mathbb{R} \cup \{+\infty\}$, denoted by $\operatorname{dom} f$ is the set defined by:

$$\operatorname{dom} f = \{x \in V : f(x) < +\infty\}.$$

Theorem 1.1 (Gamboa and Gassiat, [15] Th. 3.4)**.** *Let* $\mathcal{X}$ *be a compact metric space,* R *a probability measure on* $\mathcal{X}$ *and* $(x_i^n)_{i=1\ldots n,\, n\in\mathbb{N}^*} \subset \mathcal{X}$ *such that,*

$$\frac{1}{n}\sum_{i=1}^n \delta_{x_i^n} \xrightarrow[n\to+\infty]{} R,$$

in the weak topology.

Let μ *be a probability measure on* $\mathbb{R}$ *such that* $\operatorname{dom} Z_\mu =]-\alpha, \beta[$*, with* $\alpha, \beta > 0$*.*

If $(Z_i)_i$ *is a sequence of independent identically distributed random variables with* $\mathcal{L}(Z_i) = \mu$*, then the random weighted measures*

$$L_n = \frac{1}{n}\sum_{i=1}^n Z_i \delta_{x_i^n} \tag{1.2}$$

satisfy a large deviations principle on $\mathcal{M}(\mathcal{X})$ *equipped with the topology of weak convergence, with good rate function*

$$\mathrm{I}_\mu\left(Q\,|R\right) = \int_{\mathcal{X}} \Lambda_\mu^*\left(\frac{\mathrm{d}Q_a}{\mathrm{d}R}\right)\mathrm{d}R + \alpha Q_s^-(\mathcal{X}) + \beta Q_s^+(\mathcal{X}). \tag{1.3}$$

(See also [10] Th. 7.2.3, [12, 22] for a more general result.)

Furthermore, assuming that $\mu^{\otimes n}(L_n \in \mathcal{S}(\Phi, C)) > 0$ for all n large enough, and letting

$$R_n = \mathbb{E}_{\mu^{\otimes n}}[L_n | L_n \in \mathcal{S}(\Phi, C)] := \frac{\mathbb{E}[L_n \mathbb{1}_{\mathcal{S}(\Phi,C)}(L_n)]}{\mu^{\otimes n}(L_n \in \mathcal{S}(\Phi, C))},$$

they showed, under appropriate assumptions, that R_n converges to R^* which is the unique minimizer of $\mathrm{I}_\mu\left(\,.\,|R\right)$ over $\mathcal{S}(\Phi, C)$. (See [15] and Sect. 3 Th. 3.1 for precise statements, and for more general results, see the recent article [20] by Léonard.)

But, in this framework, it is easily seen that

$$\big(\mu^{\otimes n}(L_n \in \mathcal{S}(\Phi, C)) > 0, \quad \text{for all } n \text{ large enough } \big) \quad \Leftrightarrow \quad \overset{\circ}{C} \neq \emptyset.$$

The aim of this paper is to study the case $\overset{\circ}{C} = \emptyset$ (*thin constraints*).

1.3. **The problem of thin constraints**

When studying conditional objects of the form

$$R_n = \mathbb{E}[Z_n | Z_n \in A],$$

where $(Z_n)_n$ is a sequence of random measures and A a given subset of $\mathcal{M}(\mathcal{X})$, the main difficulty is, of course, to give a meaning to R_n when $\mathbb{P}(Z_n \in A) = 0$ and when one cannot use an explicit desintegration of the measure.

In the case of Gibbs Conditioning Principle, the classical mean to avoid this problem is to state the convergence in a double limit formulation (see [24] or [10] Sect. 7.3):

$$\lim_{\varepsilon \to 0} \lim_{n \to +\infty} \mathbb{E}[N_n | N_n \in A_\varepsilon] = R^*$$

where N_n is defined by (1.1), A_ε denotes an enlargement of A and R^* the I-projection of R on A.

In [4], Cattiaux and the author investigated the convergence in a stronger simple limit formulation

$$\lim_{n \to +\infty} \mathbb{E}[N_n | N_n \in A_{\varepsilon_n}] = R^* \tag{1.4}$$

where $(\varepsilon_n)_n$ is a sequence converging slowly to 0 (see Ths. 2.19 and 2.24 of [4]).

We shall here follow the same route. Our main result (Th. 3.2) is the analog of (1.4) in the setting of random weighted measures. We will prove that,

$$\lim_{n \to +\infty} \mathbb{E}_{\mu^{\otimes n}}[L_n | L_n \in \mathcal{S}(\Phi, C^{\varepsilon_n})] = R^*,$$

where L_n is defined by (1.2), C^ε denotes a closed blowup of C and R^* is the unique minimizer of $\mathrm{I}_\mu\,(\,.\,|R)$ over $\mathcal{S}(\Phi, C)$. As in [4], we allow small enlargement of size $\varepsilon_n \gg \frac{1}{\sqrt{n}}\cdot$

Though the statement of Theorem 3.2 is the same as (1.4), the proof is completely different. In [4], the authors took advantage of a remarkable inequality by I. Csiszar, namely

$$\forall n \in \mathbb{N}^*, \quad \mathrm{H}\,(\,\mathcal{L}(X_1 | N_n \in A_\varepsilon)\,|\,R^*_\varepsilon\,) \leq -\frac{1}{n}\log\left(\mathbb{P}(N_n \in A_\varepsilon)\mathrm{e}^{n\,\mathrm{H}(\,A_\varepsilon\,|R\,)}\right), \tag{1.5}$$

where R^*_ε is the I-projection of R on the closed convex set A_ε (see [6] (2.17) Th. 1). In Proposition 5.1, we will obtain, in the setting of random weighted measures, an inequality similar to (1.5). The proof of Proposition 5.1 relies on new ingredients. One of the main ingredient (Prop. 3.1) is inspired by some mass-transportation ideas, and gives some uniform control for the fluctuation of the mean around μ. These results can be extended to a more general study of mass-transportation inequalities for the W_1 Wasserstein distance (see [16], chaps. 6 and 7). The other tools are an exact deviation lower bound (Lem. 5.3) and a Bernstein-like inequality (Lem. 5.2).

This paper is organized as follows:

Section 2: this section is devoted to γ-divergences minimization: we recall Borwein and Lewis results on this subject (Th. 2.1) and present the Minimization of Entropy on the Mean (M.E.M.) (Th. 2.2) approach of Gamboa and Gassiat;
Section 3: main results;
Section 4: transportation-like inequalities and explicit examples;
Section 5: we apply the preceding inequalities to prove our main result;
Appendix: proof of Theorem 2.2 on M.E.M.

Several applications of these results can be considered: superresolution, fast simulation of rare events, calibration in finance. Some will be treated in [16].

2. Minimization of γ-divergences and the MEM procedure

In this section, the following assumptions hold:

Assumption 1.

(1) *$\mathcal{X}$ is a compact metric space; the set $\mathcal{M}(\mathcal{X})$ of finite Borel measures on $\mathcal{X}$ is endowed with the topology of weak convergence;*
(2) *R is a probability measure on $\mathcal{X}$ having full support;*
(3) *$\Phi = (\varphi_1, \ldots, \varphi_k) : \mathcal{X} \to \mathbb{R}^k$ is a continuous function on $\mathcal{X}$ with linearly independent components;*
(4) *C is a convex compact subset of $\mathbb{R}^k$.*

Recall that

$$\mathcal{S}(\Phi, C) = \left\{ Q \in \mathcal{M}(\mathcal{X}) : \int_{\mathcal{X}} \Phi(x)\,\mathrm{d}Q(x) \in C \right\}.$$

Theorem 2.1 (Borwein-Lewis, [3]). *Let $\gamma : \mathbb{R} \to [0, +\infty]$ be a closed convex function and denote by $a_\gamma < b_\gamma$ the endpoints of $\operatorname{dom}\gamma$. Suppose γ is differentiable and strictly convex on the interior of its domain and such that the minimum of γ is 0, attained at some point y_0 of the interior of $\operatorname{dom}\gamma$.*

Let ψ denote the convex conjugate of γ, i.e.

$$\psi(s) = \gamma^*(s) = \sup_{t \in \mathbb{R}} \{st - \gamma(t)\},$$

and denote by $a_\psi < 0 < b_\psi$ the endpoints of $\operatorname{dom}\psi$.

Suppose there is $Q_0 \in \mathcal{S}(\Phi, C)$ such that $Q_0 \ll R$ and $\frac{\mathrm{d}Q_0}{\mathrm{d}R} \in]a_\gamma, b_\gamma[$ R a.s.

Then the functional $\mathrm{I}_\gamma\,(\,.\, | R)$, defined on $\mathcal{M}(\mathcal{X})$ by

$$\mathrm{I}_\gamma\,(Q\,|R) = \int_{\mathcal{X}} \gamma\left(\frac{\mathrm{d}Q_a}{\mathrm{d}R}\right) \mathrm{d}R + b_\psi Q_s^+(\mathcal{X}) - a_\psi Q_s^-(\mathcal{X}),$$

attains its minimum on $\mathcal{S}(\Phi, C)$.

Further each minimizer R^* of $\mathrm{I}_\gamma\,(\,.\, | R)$ on $\mathcal{S}(\Phi, C)$ is of the form:

$$R^* = g^* R + \sigma,$$

where

- *$g^*(x) = \psi' \langle v^*, \Phi(x) \rangle$;*
- *v^* is the unique minimizer of $H(v) = \int_{\mathcal{X}} \psi\langle v, \Phi(x)\rangle\,\mathrm{d}R(x) - \inf_{y \in C} \langle v, y\rangle$;*
- *σ is singular with respect to R.*

Moreover, if v^* is an interior point of $\left\{v : \int_{\mathcal{X}} \psi\langle v, \Phi(x)\rangle \,\mathrm{d}R(x) < +\infty\right\}$*, then the unique minimizer of* $\mathrm{I}_\gamma\left(. \mid R\right)$ *on* $\mathcal{S}(\Phi, C)$ *is* $R^* = g^*R$*. This is in particular the case when* $\operatorname{dom}\psi = \mathbb{R}$.

(For a proof, see [3] or the appendix A of [8] and for generalizations, see [18,19].)

The following theorem presents the Minimization of Entropy on the Mean (M.E.M) procedure developed in [9, 13–15] by Dacunha-Castelle, Gamboa an Gassiat, which gives another point of view on γ-divergences minimization. We need the following

Assumption 2.

(1) *μ is a probability measure on $\mathbb{R}$ such that* $\operatorname{dom}\Lambda_\mu =]-\alpha, \beta[$ *with* $\alpha, \beta \in \mathbb{R}_+^* \cup \{+\infty\}$*;*

(2) $(x_i^n)_{i=1\dots n, n\in\mathbb{N}^*} \subset \mathcal{X}$ *is such that* $\frac{1}{n}\sum_{i=1}^n \delta_{x_i^n} \xrightarrow[n\to+\infty]{} R$*;*

(3) *there is* $g_0 : \mathcal{X} \to]a_\mu, b_\mu[$ *continuous, such that* $g_0 R \in \mathcal{S}(\Phi, C)$*, where* $a_\mu < b_\mu$ *are the endpoints of the closed convex hull of the support of μ;*

(4) *the function H defined on $\mathbb{R}^k$ by:*

$$H(v) = \int_{\mathcal{X}} \Lambda_\mu\langle v, \Phi(x)\rangle \,\mathrm{d}R(x) - \inf_{y\in C}\langle v, y\rangle,$$

has a unique minimizer v^* belonging to the interior of its domain.

We put together here different results proved in [14, 15] (Th. 2.1) with a slight refinement at points 4 and 5:

Theorem 2.2. *For all $n \in \mathbb{N}^*$, let $L_n : \mathbb{R}^n \to \mathcal{M}(\mathcal{X})$ be defined by* $L_n(z) = \frac{1}{n}\sum_{i=1}^n z_i \delta_{x_i^n}$.

For all $\varepsilon \geq 0$, let $C^\varepsilon = \{x \in \mathbb{R}^k : \exists y \in C, d_\infty(x, y) \leq \varepsilon\}$ *where* $d_\infty(x, y) = \max(|x_i - y_i|, i = 1\dots k)$.

For all $n \geq 1$ and $\varepsilon \geq 0$, let

$$\Pi_n(C^\varepsilon) = \left\{\nu \in \mathcal{P}(\mathbb{R}^n) : \mathbb{E}_\nu\left[\int_{\mathcal{X}} \Phi \,\mathrm{d}L_n\right] \in C^\varepsilon\right\}.$$

Then, under Assumptions 1 and 2, it holds:

(1) *There is $n_0 \geq 1$ such that for all $\varepsilon \geq 0$, $\mu^{\otimes n}$ has an I-projection $\mu^*_{n,\varepsilon}$ on $\Pi_n(C^\varepsilon)$, i.e. $\mu^*_{n,\varepsilon}$ is the unique probability measure belonging to $\Pi_n(C^\varepsilon)$ satisfying*

$$\mathrm{H}\left(\mu^*_{n,\varepsilon} \,\middle|\, \mu^{\otimes n}\right) = \inf\left\{\mathrm{H}\left(\nu \,\middle|\, \mu^{\otimes n}\right), \nu \in \Pi_n(C^\varepsilon)\right\}.$$

(2) *For $n \geq n_0$, $\mu^*_{n,\varepsilon}$ has the following expression:*

$$\mu^*_{n,\varepsilon} = \frac{\exp\langle w^*_{n,\varepsilon}, \,.\,\rangle}{Z_{\mu^{\otimes n}}(w^*_{n,\varepsilon})}\mu^{\otimes n} \qquad \textit{with} \qquad w^*_{n,\varepsilon} = \begin{bmatrix} \langle v^*_{n,\varepsilon}, \Phi(x_1^n)\rangle \\ \vdots \\ \langle v^*_{n,\varepsilon}, \Phi(x_n^n)\rangle \end{bmatrix}$$

and $v^*_{n,\varepsilon}$ is a minimizer of the function $H_{n,\varepsilon}$ defined on $\mathbb{R}^k$ by

$$H_{n,\varepsilon}(v) = \frac{1}{n}\sum_{i=1}^n \Lambda_\mu\langle v, \Phi(x_i^n)\rangle - \inf_{y\in C^\varepsilon}\langle v, y\rangle. \tag{2.1}$$

(3) *For all $n \geq n_0$, one has:*

$$R^*_{n,\varepsilon} := \mathbb{E}_{\mu^*_{n,\varepsilon}}[L_n] = \frac{1}{n}\sum_{i=1}^n \Lambda'_\mu\langle v^*_{n,\varepsilon}, \Phi(x_i^n)\rangle \delta_{x_i^n}.$$

(4) *For every sequence* $\varepsilon_n \in \mathbb{R}^+$ *converging to* 0, v^*_{n,ε_n} *converges to* v^* *(the unique minimiser of* H*).*
(5) *For every sequence* $\varepsilon_n \in \mathbb{R}^+$ *converging to* 0*, the sequence* R^*_{n,ε_n} *weakly converges to* R^* *the unique minimizer of* $\mathrm{I}_\mu(\,.\,|R)$ *on* $\mathcal{S}(\Phi, C)$*, which satisfies:*

$$R^* = \Lambda'_\mu\langle v^*, \Phi(\,.\,)\rangle R.$$

(A proof of this result will be given in the appendix.)

Remark 1.
- We will simply write μ^*_n, R^*_n, v^*_n etc., instead of $\mu^*_{n,0}$, $R^*_{n,0}$, $v^*_{n,0}$ etc.
- The measures $R^*_{n,\varepsilon}$ will be called the *M.E.M. estimators*.
- When $\operatorname{dom}\Lambda_\mu = \mathbb{R}$, Assumption 2 (4) is automatically fulfilled.
- When Assumption 2 (4) does not hold, the M.E.M estimators do not converge in general, (see [15] Th. 2.1 for precise results on the accumulation points).
- Assume that $\operatorname{dom}\Lambda_\mu = \mathbb{R}$ and let $\overline{R}_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}$. One can show that the measure $R^*_{n,\varepsilon}$ is the unique minimizer of the functional:

$$\mathrm{I}\left(Q\,|\overline{R}_n\right) = \int_{\mathcal{X}} \Lambda^*_\mu\left(\frac{\mathrm{d}Q}{\mathrm{d}\overline{R}_n}\right)\mathrm{d}R_n$$

  under the constraint $Q \in \mathcal{S}(\Phi, C^\varepsilon)$ (see Prop. V.10 of [16]).

3. Main results

The result we want to extend is the following:

Theorem 3.1 (Gamboa and Gassiat, [15] Th. 2.3)**.** *Under Assumptions 1 and 2, if* C *has a nonempty interior, then the* Bayesian estimator

$$R_n := \frac{\mathbb{E}_{\mu^{\otimes n}}\left[L_n \mathbb{1}_{\mathcal{S}(\Phi,C)}(L_n)\right]}{\mu^{\otimes n}(L_n \in \mathcal{S}(\Phi, C))}$$

is well defined for all n *sufficiently large and weakly converges to* R^**, the unique minimizer of* $\mathrm{I}_\mu(\,.\,|R)$ *on* $\mathcal{S}(\Phi, C)$*.*

Our main result is the following

Theorem 3.2. *Suppose Assumptions 1 and 2 are fulfilled, and let* $(\varepsilon_n)_n$ *be a sequence of positive real numbers converging to* 0 *and such that* $\lim\limits_{n\to+\infty} n\varepsilon_n^2 = +\infty$*.*

Then, the Bayesian estimator

$$R_{n,\varepsilon_n} := \frac{\mathbb{E}_{\mu^{\otimes n}}\left[L_n \mathbb{1}_{\mathcal{S}(\Phi,C^{\varepsilon_n})}(L_n)\right]}{\mu^{\otimes n}(L_n \in \mathcal{S}(\Phi, C^{\varepsilon_n}))}$$

is well defined for all n *sufficiently large and weakly converges to* R^**, the unique minimizer of* $\mathrm{I}_\mu(\,.\,|R)$ *on* $\mathcal{S}(\Phi, C)$*.*

Let us introduce some additional notations:
- For every $u \in \operatorname{dom} Z_\mu$, μ_u is the probability measure defined by:

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu}(x) = \frac{\exp(ux)}{Z_\mu(u)},$$

  and for all $n \geq 2$ and all $u \in (\operatorname{dom} Z_\mu)^n$,

$$\mu_u^{\otimes n} = \mu_{u_1} \otimes \cdots \otimes \mu_{u_n}.$$

- Θ is the set of nonnegative, nondecreasing, continuous, concave, unbounded functions defined on $\mathbb{R}^+$ and vanishing at 0.

The proof of Theorem 3.2 makes use of the following proposition, whose proof is very close to the one of Bobkov-Götze theorem on $\mathcal{T}_1$-transportation inequality (*cf.* [1] Th. 3.1):

Proposition 3.1. *For every compact interval $K \subset]-\alpha,\beta[$, there is $\theta_K \in \Theta$ such that, for all $u \in K$ and $\nu \in \mathcal{P}(\mathbb{R})$:*

$$\left|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u(x)\right| \leq \theta_K(\mathrm{H}\,(\,\nu\,|\,\mu_u\,)).$$

Remark 2. If μ is such that $\Lambda''_\mu(t) \leq M$ for all $t \in \mathbb{R}$, (for example if μ has a compact support or μ is a gaussian measure), one can take $\theta_K(x) = \sqrt{2Mx}$. In this case, the preceding inequality can be seen as a particular case of the $\mathcal{T}_1$-transportation inequality (*cf.* [1] Th. 3.1). Other explicit bounds can be found in Section 4.2.

Using well known methods of information theory, we will deduce from this result an upper bound for the total variation distance between R_{n,ε_n} and R^*_{n,ε_n} of the following form:

$$\|R_{n,\varepsilon_n} - R^*_{n,\varepsilon_n}\|_{TV} \leq \theta\left(\frac{-1}{n}\log\left[\mu^{\otimes n}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n \in C^{\varepsilon_n}\right)\mathrm{e}^{\mathrm{H}\left(\mu^*_{n,\varepsilon_n}|\mu^{\otimes n}\right)}\right]\right),$$

where $\theta \in \Theta$ does not depend on n (Prop. 5.1).

Finally, we will majorize the right hand side thanks to an exact deviation lower bound (Lem. 5.3) and a Berstein-like inequality (Lem. 5.2). The convergence of the Bayesian estimators R_{n,ε_n} will then follow from the convergence of the M.E.M estimators R^*_{n,ε_n}.

4. Transportation-like inequalities

4.1. General results

Recall that Θ denotes the set of nonnegative, nondecreasing, continuous, concave, unbounded functions defined on $\mathbb{R}^+$ and vanishing at 0. We will need the following lemma:

Lemma 4.1. *Let $k : [0,r[\to \mathbb{R}_+$, $r \in \mathbb{R}^*_+ \cup \{+\infty\}$ be such that $\lim_{s\to 0} k(s) = 0$ and $\lim_{s\to r} k(s) = +\infty$.*

Then the function θ defined for all $a \in \mathbb{R}_+$ by $\theta(a) = \inf_{s\in]0,r[}\left\{\frac{a}{s} + k(s)\right\}$ belongs to Θ.

Proof.

- For all $a \geq 0$, $s \mapsto \frac{a}{s} + k(s)$ is nonnegative, so $\theta(a) = \inf_{0<s<r}\left\{\frac{a}{s} + k(s)\right\} \in \mathbb{R}_+$, and θ is well defined on $\mathbb{R}_+$. Moreover $\theta(0) = \inf_{0<s<r}\{k(s)\}$ and $\lim_{s\to 0} k(s) = 0$, thus $\theta(0) = 0$.
- The function θ being an infimum of affine functions, it is concave. As θ is finite over $\mathbb{R}^+$, θ is continuous over $]0,+\infty[$.
- If $0 \leq a \leq a' < r$, then for all $0 < s < r$, $\frac{a}{s} + k(s) \leq \frac{a'}{s} + k(s)$, thus, taking the inf at both sides, one obtains $\theta(a) \leq \theta(a')$ and θ is therefore nondecreasing.
- Let $(a_n)_n$ be such that $a_n \xrightarrow[n\to+\infty]{} 0$. One has for all $0 < s < r$: $\theta(a_n) \leq \frac{a_n}{s} + k(s)$, so $\limsup_{n\to+\infty}\theta(a_n) \leq k(s)$. As $\inf_{0<s<r} k(s) = 0$, we get $\limsup_{n\to+\infty}\theta(a_n) = 0$ and θ is continuous at 0.
- Finally, let $(a_n)_n$ be such that $a_n \xrightarrow[n\to+\infty]{} +\infty$ and let us prove that $\theta(a_n) \xrightarrow[n\to+\infty]{} +\infty$. The function θ being nondecreasing, it suffices to show that $\theta(a_n)$ is unbounded. For all n, $s \mapsto \frac{a_n}{s} + k(s)$ tends to $+\infty$ when s tends to 0 or r, thus there is a number s_n such that $\theta(a_n) = \frac{a_n}{s_n} + k(s_n)$. Consequently, we get

$$\limsup_{n\to+\infty}\theta(a_n) \geq \limsup_{n\to+\infty}\frac{a_n}{s_n} \vee \limsup_{n\to+\infty} k(s_n).$$

If s_n is bounded, $\limsup_{n\to+\infty} \frac{a_n}{s_n} = +\infty$ and if s_n is not ($r = +\infty$), $\limsup_{n\to+\infty} k(s_n) = +\infty$. In both cases, $\theta(a_n)$ is unbounded. $\square$

Proof of Proposition 3.1.
(1) For all $u \in]-\alpha, \beta[$,

$$Z_{\mu_u}(s) = \frac{\int \exp(sx)\exp(ux)\,\mathrm{d}\mu(x)}{Z_\mu(u)} = \frac{Z_\mu(u+s)}{Z_\mu(u)}$$

so dom $Z_{\mu_u} =]-\alpha-u, \beta-u[$.

Let $s \in]-\alpha-u, \beta-u[$, and write

$$s\left(\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u(x)\right) = \int h_s(x)\,\mathrm{d}\nu(x) + \log\int \mathrm{e}^{s(x-\int y\,\mathrm{d}\mu_u(y))}\,\mathrm{d}\mu_u(x),$$

denoting

$$h_s(x) = s\left(x - \int y\,\mathrm{d}\mu_u(y)\right) - \log\int \mathrm{e}^{s(x-\int y\,\mathrm{d}\mu_u(y))}\,\mathrm{d}\mu_u(x).$$

Clearly,

$$\int \exp h_s\,\mathrm{d}\mu_u = 1.$$

Thanks to the following variational formula for relative entropy (see *e.g.* [21], Chap. 1, Prop. 4),

$$\mathrm{H}(\nu\,|\,\mu_u) = \sup\left\{\int h\,\mathrm{d}\nu : \int \exp h\,\mathrm{d}\mu_u \le 1\right\},$$

one gets

$$\int h_s\,\mathrm{d}\nu \le \mathrm{H}(\nu\,|\,\mu_u).$$

Moreover, noticing that $\Lambda'_\mu(u) = \int y\,\mathrm{d}\mu_u(y)$, one gets easily

$$\log\int \mathrm{e}^{s(x-\int y\,\mathrm{d}\mu_u(y))}\,\mathrm{d}\mu_u(x) = \Lambda_\mu(s+u) - \Lambda_\mu(u) - s\Lambda'_\mu(u) := q(s,u),$$

and $q(s,u)$ is non-negative due to the convexity of Λ_μ.

Thus, for all $s \in]0, \beta-u[$, one has

$$\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u(x) \le \frac{\mathrm{H}(\nu\,|\,\mu_u)}{s} + \frac{q(s,u)}{s}$$

and for $s \in]0, \alpha+u[$

$$\int x\,\mathrm{d}\mu_u(x) - \int x\,\mathrm{d}\nu(x) \le \frac{\mathrm{H}(\nu\,|\,\mu_u)}{s} + \frac{q(-s,u)}{s}.$$

Let $K = [a,b] \subset]\alpha, \beta[$ and $r = \min(\alpha+a, \beta-b) \in \mathbb{R}_+^* \cup \{+\infty\}$, then for all $0 < s < r$ one has

$$\left|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u(x)\right| \le \frac{\mathrm{H}(\nu\,|\,\mu_u)}{s} + \frac{q(s,u)+q(-s,u)+s^2}{s}.$$

Let

$$k(s) = \frac{\max_{u\in K}(q(s,u)+q(-s,u))+s^2}{s},$$

then for all $u \in K$

$$\left|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u(x)\right| \leq \frac{\mathrm{H}\,(\,\nu\,|\,\mu_u\,)}{s} + k(s).$$

Taking the inf over $0 < s < r$, one obtains

$$\left|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u(x)\right| \leq \theta_K(\mathrm{H}\,(\,\nu\,|\,\mu_u\,)),$$

with θ_K defined by

$$\theta_K(a) = \inf_{0<s<r}\left\{\frac{a}{s} + k(s)\right\}.$$

(2) Let us check that k satisfies the assumptions of Lemma 4.1.

If $r = +\infty$, then $k(s) \geq s$ and so $\lim\limits_{s\to+\infty} k(s) = +\infty$.

If $r = \alpha + a < +\infty$, then

$$k(s) \geq \frac{q(-s,a)}{s} = \frac{\Lambda_\mu(a-s) - \Lambda_\mu(a)}{s} + \Lambda'_\mu(a).$$

As $\lim\limits_{s\to\alpha+a} \Lambda_\mu(a-s) = +\infty$, $\lim\limits_{s\to\alpha+a} k(s) = +\infty$.

If $r = \beta - b < +\infty$, one gets similarly $\lim\limits_{s\to\beta-b} k(s) = +\infty$. In all cases, $\lim\limits_{s\to r} k(s) = +\infty$.

Let us verify that $\lim\limits_{s\to 0} k(s) = 0$. Let $0 < s_n < r$ be such that $s_n \xrightarrow[n\to+\infty]{} 0$. For all n, there is $u_n \in K$ such that

$$k(s_n) = \frac{q(s_n,u_n) + q(-s_n,u_n)}{s_n} + s_n.$$

Let us assume that for all n, $k(s_n) \geq \varepsilon > 0$. As $(u_n)_n$ is a bounded sequence, there exists ϕ such that $u_{\phi(n)} \to u_0 \in K$. But Λ''_μ being nonnegative, Taylor formula yields

$$q(s_{\phi(n)},u_{\phi(n)}) + q(-s_{\phi(n)},u_{\phi(n)}) \leq s^2_{\phi(n)} \sup\{\Lambda''_\mu(u) : u \in [u_{\phi(n)} - s_{\phi(n)}, u_{\phi(n)} + s_{\phi(n)}]\},$$

which implies that $k(s_{\phi(n)}) \xrightarrow[n\to+\infty]{} 0$. Contradiction, so $\lim\limits_{s\to 0} k(s) = 0$ and $\theta_K \in \Theta$. □

Corollary 4.1. *For every compact interval $K \subset]-\alpha,\beta[$, one has*

$$\forall u \in K^n, \quad \forall \nu \in \mathcal{P}(\mathbb{R}^n), \quad \frac{1}{n}\left\|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u^{\otimes n}(x)\right\|_1 \leq \theta_K\left(\frac{\mathrm{H}\,(\,\nu\,|\,\mu_u^{\otimes n}\,)}{n}\right),$$

denoting $\mu_u^{\otimes n} = \mu_{u_1} \otimes \cdots \otimes \mu_{u_n}$ and $\|x\|_1 = \sum_{i=1}^n |x_i|$.

Proof. We will denote by $\nu_1, \nu_2, \ldots, \nu_n$, the one dimensional marginales of ν.

One has:

$$\begin{aligned}\frac{1}{n}\left\|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u^{\otimes n}(x)\right\|_1 &= \frac{1}{n}\sum_{i=1}^n\left|\int x_i\,\mathrm{d}\nu(x) - \int x_i\,\mathrm{d}\mu_u^{\otimes n}(x)\right| \\ &= \frac{1}{n}\sum_{i=1}^n\left|\int x\,\mathrm{d}\nu_i(x) - \int x\,\mathrm{d}\mu_{u_i}(x)\right|.\end{aligned}$$

As for all $i \in \{1,\dots,n\}$, $u_i \in K$, Proposition 3.1 yields:

$$\left|\int x\,\mathrm{d}\nu_i(x) - \int x\,\mathrm{d}\mu_{u_i}(x)\right| \le \theta_K(\mathrm{H}\,(\,\nu_i\,|\,\mu_{u_i}\,)).$$

So

$$\frac{1}{n}\left\|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u^{\otimes n}(x)\right\|_1 \le \frac{1}{n}\sum_{i=1}^n \theta_K(\mathrm{H}\,(\,\nu_i\,|\,\mu_{u_i}\,)).$$

The function θ_K being concave, one gets thanks to Jensen inequality:

$$\frac{1}{n}\left\|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u^{\otimes n}(x)\right\|_1 \le \theta_K\left(\frac{\sum_{i=1}^n \mathrm{H}\,(\,\nu_i\,|\,\mu_{u_i}\,)}{n}\right).$$

But according to the following identity (see [6] (2.11))

$$\mathrm{H}\left(\,\nu\,\middle|\,\mu_u^{\otimes n}\,\right) = \mathrm{H}\,(\,\nu\,|\,\nu_1\otimes\cdots\otimes\nu_n\,) + \sum_{i=1}^n \mathrm{H}\,(\,\nu_i\,|\,\mu_{u_i}\,),$$

one has

$$\sum_{i=1}^n \mathrm{H}\,(\,\nu_i\,|\,\mu_{u_i}\,) \le \mathrm{H}\left(\,\nu\,\middle|\,\mu_u^{\otimes n}\,\right).$$

As θ_K is nondecreasing, one obtains:

$$\frac{1}{n}\left\|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u^{\otimes n}(x)\right\|_1 \le \theta_K\left(\frac{\mathrm{H}\,(\,\nu\,|\,\mu_u^{\otimes n}\,)}{n}\right). \qquad \square$$

4.2. Some explicit bounds

Proposition 4.1. *Let μ be such that $\Lambda''_\mu(u) \le M$ for all $u \in \mathbb{R}$, then for all $\nu \in \mathcal{P}(\mathbb{R})$ and all $u \in \mathbb{R}$*

$$\left|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u(x)\right| \le \sqrt{2M\,\mathrm{H}\,(\,\nu\,|\,\mu_u\,)}.$$

Proof. Thanks to Taylor-Lagrange formula, for all $u, s \in \mathbb{R}$, there exists a such that:

$$q(s,u) = \Lambda_\mu(u+s) - \Lambda_\mu(u) - s\Lambda'_\mu(u) = \frac{s^2}{2}\Lambda''_\mu(a)$$

thus $q(s,u) \le \frac{s^2 M}{2}$, and one can take $k(s) = \frac{sM}{2}$. A simple calculus yields then $\theta(x) = \sqrt{2Mx}$. $\square$

Examples.

– μ has its support included in $[a,b]$: the support of μ_u is also in $[a,b]$ and $\Lambda''_\mu(u) = \mathrm{Var}(\mu_u) \le (b-a)^2$. In this case, one can take

$$\theta(x) = (b-a)\sqrt{2x}.$$

– $\mu = Z^{-1}\mathrm{e}^{-U}\mathrm{d}x$, with $U'' \ge c > 0$: the probability measure μ satisfies then the following Poincaré inequality:

$$\mathrm{Var}_\mu(f) \le \frac{1}{c}\int (f')^2(x)\,\mathrm{d}\mu(x).$$

But $\mu_u = \frac{e^{-U+ux}}{ZZ_\mu(u)}dx$ and $V = U(x) + ux$ also satisfies $V'' \geq c > 0$, so that μ_u satisfies the same Poincaré inequality as μ. In particular, choosing $f(x) = x$, one obtains:

$$\Lambda''_\mu(u) = \mathrm{Var}(\mu_u) = \mathrm{Var}_{\mu_u}(x) \leq \frac{1}{c}\cdot$$

In this case, one can thus take

$$\theta(x) = \sqrt{\frac{2x}{c}}\cdot$$

The following lemma will enable us, in certain cases, to majorize the function θ by another function enjoying the same properties as θ except concavity.

Lemma 4.2. *Let $k : [0, +\infty[\to \mathbb{R}_+$ be a $\mathcal{C}^2$ function such that $k(0) = k'(0) = 0$ and $k'' \geq c > 0$.*
Define $\Psi(t) = \int_0^t uk''(u)\,\mathrm{d}u = tk'(t) - k(t)$, then

(1) *For all $a \in \mathbb{R}^+$,*

$$\theta(a) = \inf_{s\in\mathbb{R}_+}\left\{\frac{a}{s} + \frac{k(s)}{s}\right\} = k'(\Psi^{-1}(a)).$$

(2) *Moreover, for all $a \in \mathbb{R}^+$, $\theta(a) \leq k'\left(\sqrt{\frac{2a}{c}}\right)\cdot$*

Proof.

1) For all $a > 0$, $g_a : s \mapsto \frac{a}{s} + \frac{k(s)}{s}$ goes to $+\infty$ when s goes to 0 or $+\infty$, thus g_a attains its minimum at a point s_a such that $g'_a(s_a) = 0$, that is to say $\Psi(s_a) = a$. The function Ψ being increasing, $s_a = \Psi^{-1}(a)$, and this remains true for $a = 0$.

Moreover

$$\theta(a) = \frac{a}{s_a} + \frac{k(s_a)}{s_a} = \frac{k'(s_a)s_a - k(s_a)}{s_a} + \frac{k(s_a)}{s_a} = k'(s_a) = k'(\Psi^{-1}(a)).$$

2) $a = \int_0^{s_a} uk''(u)\,\mathrm{d}u \geq \int_0^{s_a} cu\,\mathrm{d}u = c\frac{s_a^2}{2}\cdot$

Thus $s_a \leq \sqrt{\frac{2a}{c}a}$ and k' being increasing, one has

$$\theta(a) = k'(s_a) \leq k'\left(\sqrt{\frac{2a}{c}}\right)\cdot \qquad \square$$

Examples.

– μ is the Poisson distribution with parameter $\lambda > 0$: $\Lambda_\mu(u) = \lambda(e^u - 1)$ and $\Lambda_\mu(u+s) + \Lambda_\mu(u-s) - 2\Lambda_\mu(u) = 2\lambda e^u[\cosh(s) - 1]$.

Let $M > 0$ and define $k(s) = 2\lambda e^M[\cosh(s) - 1]$. It follows from the proof of Proposition 3.1 that for all $u \in [-M, M]$ and all $\nu \in \mathcal{P}(\mathbb{R})$,

$$\left|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u(x)\right| \leq \theta_M(\mathrm{H}\,(\,\nu\,|\,\mu_u\,)),$$

with $\theta_M(a) = \inf\left\{\frac{a}{s} + \frac{k(s)}{s}\right\}$.

Moreover $k'(s) = 2\lambda e^M \sinh(s)$ and $k''(s) = 2\lambda e^M \cosh(s) \geq 2\lambda e^M$, thus, from the preceding lemma, it holds:

$$\theta_M(a) \leq 2\lambda e^M \sinh\sqrt{\frac{e^{-M}a}{\lambda}}\cdot$$

Consequently, for all $u \in [-M, M]$ and $\nu \in \mathcal{P}(\mathbb{R})$:

$$\left|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u(x)\right| \le 2\lambda \mathrm{e}^M \sinh\sqrt{\frac{\mathrm{e}^{-M}\,\mathrm{H}\,(\,\nu\,|\,\mu_u\,)}{\lambda}}.$$

– μ is the exponential distribution with parameter λ: Adapting slightly the proof of the preceding lemma, one gets:

For all $u \le b < \lambda$ and $\nu \in \mathcal{P}(\mathbb{R})$ such that $\mathrm{H}\,(\,\nu\,|\,\mu_u\,) < 1$,

$$\left|\int x\,\mathrm{d}\nu(x) - \int x\,\mathrm{d}\mu_u(x)\right| \le \frac{2}{\lambda - b}\frac{\sqrt{\mathrm{H}\,(\,\nu\,|\,\mu_u\,)}}{1-\mathrm{H}\,(\,\nu\,|\,\mu_u\,)}.$$

5. Conditional Principle

5.1. Majorization of the total variation distance between the M.E.M. estimator and the Bayesian estimator

According to Theorem 2.2, there is n_0 such that for all $n \ge n_0$, $\mu^*_{n,\,\varepsilon}$ is well defined for all $\varepsilon \ge 0$ and $\mu^*_{n,\,\varepsilon} = \mu^{\otimes n}_{w^*_{n,\,\varepsilon}}$.

Lemma 5.1. *For every sequence ε_n of nonnegative numbers converging to* 0*, there is $m \ge n_0$ and a compact interval $K \subset]-\alpha, \beta[$ such that*

$$\forall n \ge m, \quad w^*_{n,\,\varepsilon_n} \in K^n \qquad \textit{and} \qquad \forall x \in \mathcal{X}, \quad \langle v^*, \Phi(x)\rangle \in K.$$

Proof. According to Theorem 2.2 (2):

$$w^*_{n,\,\varepsilon_n} = \begin{bmatrix} \langle \Phi(x_1^n), v^*_{n,\,\varepsilon_n}\rangle \\ \vdots \\ \langle \Phi(x_n^n), v^*_{n,\,\varepsilon_n}\rangle \end{bmatrix}.$$

The function Φ being continuous on the compact set $\mathcal{X}$, there is $N > 0$ such that $\|\Phi(x)\| \le N$ for all $x \in \mathcal{X}$.

For all $i \in \{1, \dots, n\}$, it follows from Cauchy-Schwarz inequality that

$$\left|\langle \Phi(x_i^n), v^*_{n,\,\varepsilon_n}\rangle - \langle \Phi(x_i^n), v^*\rangle\right| \le N\|v^*_{n,\,\varepsilon_n} - v^*\|,$$

and so

$$\inf_{x\in\mathcal{X}}\langle v^*, \Phi(x)\rangle - N\|v^*_{n,\,\varepsilon_n} - v^*\| \le (w^*_{n,\,\varepsilon_n})_i \le \sup_{x\in\mathcal{X}}\langle v^*, \Phi(x)\rangle + N\|v^*_{n,\,\varepsilon_n} - v^*\|.$$

According to Assumption 2 (4), $v^* \in \overset{\circ}{\operatorname{dom}}\, H$. Now, it is easily seen that

$$\overset{\circ}{\operatorname{dom}}\, H = \left\{v \in \mathbb{R}^k : \forall x \in \mathcal{X}, \quad \langle v, \Phi(x)\rangle \in]-\alpha, \beta[\right\}.$$

Thanks to the compactness of $\mathcal{X}$, one has

$$-\alpha < \inf_{x\in\mathcal{X}}\langle v^*, \Phi(x)\rangle \le \sup_{x\in\mathcal{X}}\langle v^*, \Phi(x)\rangle < \beta.$$

According to Theorem 2.2 (5), $v^*_{n,\,\varepsilon_n}$ converges in $\mathbb{R}^k$ to v^*, and the result follows easily. □

Lemma 5.2. *There are $M > 0$ and $n_1 \geq n_0$ such that for all $\varepsilon > 0$ and $n \geq n_1$,*

$$\mu_n^* \left(\int_{\mathcal{X}} \Phi \, \mathrm{d}L_n \in C^\varepsilon \right) \geq 1 - 2k \exp\left[-\frac{n\varepsilon^2}{2M(2M+\varepsilon)} \right].$$

(Recall that Φ is $\mathbb{R}^k$-valued.)

Proof.
First step. Let us show that for all compact interval $K \subset]-\alpha, \beta[$, there is $M > 0$ such that for all $u \in K$ and $j \geq 2$:

$$\int \left| z - \int x \, \mathrm{d}\mu_u(x) \right|^j \mathrm{d}\mu_u(z) \leq j! M^j.$$

Denoting $\tau(x) = \mathrm{e}^{|x|} - 1 - |x|$, and $J(u, M) = \int \tau\left(\frac{z - \int x \, \mathrm{d}\mu_u(x)}{M} \right) \mathrm{d}\mu_u(z)$, one gets easily that

$$\sup_{u \in K} J(u, M) \xrightarrow[M \to +\infty]{} 0.$$

Therefore there is $M > 0$ such that $\sup_{u \in K} J(u, M) \leq 1$.

Now,

$$J(u, M) = \sum_{j=2}^{+\infty} \frac{\int \left| z - \int x \, \mathrm{d}\mu_u(x) \right|^j \mathrm{d}\mu_u(z)}{M^j j!},$$

hence for all $u \in K$ and $k \geq 2$, one has

$$\frac{\int \left| z - \int x \, \mathrm{d}\mu_u(x) \right|^j \mathrm{d}\mu_u(z)}{M^j j!} \leq J(u, M) \leq 1.$$

Second step. Let us show that for all compact interval $K \subset]-\alpha, \beta[$, and all $N > 0$, there is $M > 0$ such that for all sequence $Z_1, \ldots, Z_n$ of independent random variables such that $\mathcal{L}(Z_i) = \mu_{u_i}$ with $u_i \in K$ and all sequence $\alpha_1, \ldots, \alpha_n$ of real numbers with $|\alpha_i| \leq N$, one has

$$\forall \varepsilon > 0, \quad \mathbb{P}\left(\left| \bar{Z} - m \right| > \varepsilon \right) \leq 2 \exp\left[-\frac{n\varepsilon^2}{2M(2M+\varepsilon)} \right],$$

where $\bar{Z} = \frac{1}{n} \sum_{i=1}^n \alpha_i Z_i$ and $m = \mathbb{E}\left[\bar{Z} \right]$.

According to the first step, there is $M_0 > 0$ depending only on K such that for all i,

$$\forall j \geq 2, \quad \mathbb{E}\left[|Z_i - \mathbb{E}[Z_i]|^j \right] \leq j! M_0^j.$$

From this, it follows that for all i

$$\forall j \geq 2, \quad \mathbb{E}\left[|\alpha_i (Z_i - \mathbb{E}[Z_i])|^j \right] \leq j! (M_0 N)^j.$$

Letting $M = M_0 N$, the results follows from Bernstein inequality (see *e.g.* [25], 2.2.11 p. 103).

Third step. Now let us prove the lemma.

Let $c_n = (c_{n,1}, \ldots, c_{n,k}) := \mathbb{E}_{\mu_n^*}\left[\int_{\mathcal{X}} \Phi \, \mathrm{d}L_n\right] \in C$.

Then

$$\begin{aligned}\mu_n^*\left(\int_{\mathcal{X}} \Phi \, \mathrm{d}L_n \in C^{\varepsilon}\right) &\geq \mu_n^*\left(\left\|\int_{\mathcal{X}} \Phi \, \mathrm{d}L_n - c_n\right\|_{\infty} \leq \varepsilon\right)\\ &= 1 - \mu_n^*\left(\left\|\int_{\mathcal{X}} \Phi \, \mathrm{d}L_n - c_n\right\|_{\infty} > \varepsilon\right)\\ &\geq 1 - \sum_{j=1}^{k} \mu_n^*\left(z : \left|\frac{1}{n}\sum_{p=1}^{n} z_i \phi_p(x_i^n) - c_{n,p}\right| > \varepsilon\right).\end{aligned}$$

The functions ϕ_p being continuous on the compact $\mathcal{X}$, there is $N > 0$ such that $|\phi_p(x)| \leq N$ for all p and x. Moreover, according to Lemma 5.1 applied to the sequence $\varepsilon_n = 0$, there is $n_1 \geq n_0$ and a compact interval $K \subset\,]-\alpha, \beta[$ such that for all $n \geq n_1$, $w_n^* \in K^n$. Thus, according to the second step, one can conclude that there is $M > 0$ such that for all $\varepsilon > 0$ and all integer $n \geq n_1$, one has:

$$\mu_n^*\left(\int_{\mathcal{X}} \Phi \, \mathrm{d}L_n \in C^{\varepsilon}\right) \geq 1 - 2k \exp\left[-\frac{n\varepsilon^2}{2M(2M+\varepsilon)}\right]. \qquad \square$$

Now we are ready to prove the

Proposition 5.1. *Let ε_n be a sequence of positive real numbers converging to 0 such that $n\varepsilon_n^2 \to +\infty$.*

Then the following holds

1. *There is $n_2 \geq n_0$ such that for all $n \geq n_2$, R_{n,ε_n} and R^*_{n,ε_n} are well defined.*
2. *There is $\theta \in \Theta$ such that for all $n \geq n_2$:*

$$\left\|R_{n,\varepsilon_n} - R^*_{n,\varepsilon_n}\right\|_{TV} \leq \theta\left(\frac{-1}{n}\log\left[\mu^{\otimes n}\left(\int_{\mathcal{X}} \Phi \, \mathrm{d}L_n \in C^{\varepsilon_n}\right) \mathrm{e}^{\mathrm{H}\left(\mu^*_{n,\varepsilon_n} \,|\, \mu^{\otimes n}\right)}\right]\right).$$

Proof.

(1) For $n \geq n_0$, μ_n^* and μ^*_{n,ε_n} are well defined.

Moreover, according to Lemma 5.2, there is $n_1 \geq n_0$ and $M > 0$ such that for all $n \geq n_1$,

$$\mu_n^*\left(\int_{\mathcal{X}} \Phi \, \mathrm{d}L_n \in C^{\varepsilon_n}\right) \geq 1 - 2k \exp\left[-\frac{n\varepsilon_n^2}{2M(2M+\varepsilon_n)}\right].$$

As $n\varepsilon_n^2 \xrightarrow[n\to+\infty]{} +\infty$, it is clear that $\mu_n^*\left(\int_{\mathcal{X}} \Phi \, \mathrm{d}L_n \in C^{\varepsilon_n}\right) \xrightarrow[n\to+\infty]{} 1$. In particular, there is $m_1 \geq n_1$ such that for all $n \geq m_1$, $\mu_n^*\left(\int_{\mathcal{X}} \Phi \, \mathrm{d}L_n \in C^{\varepsilon_n}\right) > 0$. As $\mu^{\otimes n}$ is equivalent to μ_n^*, it follows that for all $n \geq m_1$, $\mu^{\otimes n}\left(\int_{\mathcal{X}} \Phi \, \mathrm{d}L_n \in C^{\varepsilon_n}\right) > 0$ and thus R_{n,ε_n} is well defined.

(2) According to Lemma 5.1, there are a compact interval $K \subset]-\alpha, \beta[$ and $m_2 \geq n_0$ such that for all $n \geq m_2$, $w^*_{n,\varepsilon_n} \in K^n$. Let $\nu_{n,\varepsilon_n} \in \mathcal{P}(\mathbb{R}^n)$ be defined by $\nu_{n,\varepsilon_n} = \frac{\mathbb{1}_{\mathcal{S}(\Phi, C^{\varepsilon_n})}(L_n)}{\mu^{\otimes n}(L_n \in \mathcal{S}(\Phi, C^{\varepsilon_n}))} . \mu^{\otimes n}$. According to Corollary 4.1, one has for all $n \geq n_2 = \max(m_1, m_2)$, letting $\theta = \theta_K$:

$$\frac{1}{n}\left\|\int x \, \mathrm{d}\nu_{n,\varepsilon_n}(x) - \int x \, \mathrm{d}\mu^*_{n,\varepsilon_n}(x)\right\|_1 \leq \theta\left(\frac{\mathrm{H}\left(\nu_{n,\varepsilon_n} \,\middle|\, \mu^*_{n,\varepsilon_n}\right)}{n}\right). \tag{5.1}$$

But

$$\begin{aligned}\|R_{n,\,\varepsilon_n}-R^*_{n,\,\varepsilon_n}\|_{TV} &= \left\|\frac{1}{n}\sum_{i=1}^n\left(\int z_i\,\mathrm{d}\nu_{n,\,\varepsilon_n}(z)-\int z_i\,\mathrm{d}\mu^*_{n,\,\varepsilon_n}(z)\right).\delta_{x_i^n}\right\|_{TV}\\ &\le \frac{1}{n}\sum_{i=1}^n\left|\int z_i\,\mathrm{d}\nu_{n,\,\varepsilon_n}(z)-\int z_i\,\mathrm{d}\mu^*_{n,\,\varepsilon_n}(z)\right|\\ &= \frac{1}{n}\left\|\int x\,\mathrm{d}\nu_{n,\varepsilon_n}(x)-\int x\,\mathrm{d}\mu^*_{n,\,\varepsilon_n}(x)\right\|_1.\end{aligned}$$

Thus, according to (5.1), for all $n\ge n_2$,

$$\|R_{n,\,\varepsilon_n}-R^*_{n,\,\varepsilon_n}\|_{TV}\le\theta\left(\frac{\mathrm{H}\left(\nu_{n,\,\varepsilon_n}\,\middle|\,\mu^*_{n,\,\varepsilon_n}\right)}{n}\right).$$

But one easily sees that

$$\nu_{n,\,\varepsilon_n}\in\Pi_n(C^{\varepsilon_n}).$$

Applying Csiszar inequality (*cf.* [5] Th. 2.2), it holds

$$\mathrm{H}\left(\nu_{n,\,\varepsilon_n}\,\middle|\,\mu^{\otimes n}\right)\ge\mathrm{H}\left(\nu_{n,\,\varepsilon_n}\,\middle|\,\mu^*_{n,\,\varepsilon_n}\right)+\mathrm{H}\left(\mu^*_{n,\,\varepsilon_n}\,\middle|\,\mu^{\otimes n}\right).$$

Moreover, a simple calculus yields

$$\mathrm{H}\left(\nu_{n,\varepsilon_n}\,\middle|\,\mu^{\otimes n}\right)=-\log\mu^{\otimes n}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon_n}\right),$$

hence

$$\mathrm{H}\left(\nu_{n,\,\varepsilon_n}\,\middle|\,\mu^*_{n,\,\varepsilon_n}\right)\le-\log\left[\mu^{\otimes n}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon_n}\right)\mathrm{e}^{\mathrm{H}\left(\mu^*_{n,\,\varepsilon_n}\,|\,\mu^{\otimes n}\right)}\right]$$

and θ being nondecreasing, one obtains for all $n\ge n_2$:

$$\|R_{n,\,\varepsilon_n}-R^*_{n,\,\varepsilon_n}\|_{TV}\le\theta\left(\frac{-1}{n}\log\left[\mu^{\otimes n}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon_n}\right)\mathrm{e}^{\mathrm{H}\left(\mu^*_{n,\,\varepsilon_n}\,|\,\mu^{\otimes n}\right)}\right]\right).\qquad\square$$

5.2. **Convergence of the Bayesian estimators**

We need the following lemma

Lemma 5.3. *As soon as* $\mu^*_n\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^\varepsilon\right)>0$*, it holds:*

$$\begin{aligned}\frac{1}{n}\log\left[\mu^{\otimes n}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^\varepsilon\right)\mathrm{e}^{\mathrm{H}\left(\mu^*_n\,|\,\mu^{\otimes n}\right)}\right]\ge\frac{\mathrm{H}\left(\mu^*_n\,|\,\mu^{\otimes n}\right)}{n}\left(1-\frac{1}{\mu^*_n\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^\varepsilon\right)}\right)\\ +\frac{1}{n}\log\mu^*_n\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^\varepsilon\right)-\frac{1}{ne}\frac{1}{\mu^*_n\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^\varepsilon\right)}.\end{aligned}$$

Proof. (see also [11], Ex. 3.3.23 p. 76) The probability measure $\mu^{\otimes n}$ being equivalent to μ^*_n,

$$\mu^*_n\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^\varepsilon\right)>0\Rightarrow\mu^{\otimes n}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^\varepsilon\right)>0.$$

It holds:

$$\begin{aligned}\frac{1}{n}\log\mu^{\otimes n}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)&=\frac{1}{n}\log\int\mathbb{1}_{C^{\varepsilon}}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\right)\,\mathrm{d}\mu^{\otimes n}=\frac{1}{n}\log\int\mathbb{1}_{C^{\varepsilon}}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\right)\frac{\mathrm{d}\mu^{\otimes n}}{\mathrm{d}\mu_n^*}\,\mathrm{d}\mu_n^*\\&=\frac{1}{n}\log\int\frac{\mathrm{d}\mu^{\otimes n}}{\mathrm{d}\mu_n^*}\frac{\mathbb{1}_{C^{\varepsilon}}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\right)}{\mu_n^*(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon})}\,\mathrm{d}\mu_n^*+\frac{1}{n}\log\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right).\end{aligned}$$

As $\frac{\mathbb{1}_{C^{\varepsilon}}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\right)}{\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}\mu_n^*$ is a probability measure, Jensen inequality yields

$$\frac{1}{n}\log\int\frac{\mathrm{d}\mu^{\otimes n}}{\mathrm{d}\mu_n^*}\frac{\mathbb{1}_{C^{\varepsilon}}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\right)}{\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}\,\mathrm{d}\mu_n^*\geq\frac{1}{n}\int\log\frac{\mathrm{d}\mu^{\otimes n}}{\mathrm{d}\mu_n^*}\frac{\mathbb{1}_{C^{\varepsilon}}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\right)}{\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}\,\mathrm{d}\mu_n^*.$$

Moreover, letting $I_n=\frac{1}{n}\int\log\frac{\mathrm{d}\mu^{\otimes n}}{\mathrm{d}\mu_n^*}\frac{\mathbb{1}_{C^{\varepsilon}}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\right)}{\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}\,\mathrm{d}\mu_n^*$, one has

$$\begin{aligned}I_n&=\frac{1}{n\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}\int\log\frac{\mathrm{d}\mu^{\otimes n}}{\mathrm{d}\mu_n^*}\,\mathrm{d}\mu_n^*-\frac{1}{n}\int\log\frac{\mathrm{d}\mu^{\otimes n}}{\mathrm{d}\mu_n^*}\frac{\mathbb{1}_{(C^{\varepsilon})^c}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\right)}{\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}\,\mathrm{d}\mu_n^*\\&=-\frac{1}{n\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}\,\mathrm{H}\left(\mu_n^*\,|\,\mu^{\otimes n}\right)+\frac{1}{n}\int\frac{\mathrm{d}\mu_n^*}{\mathrm{d}\mu^{\otimes n}}\log\frac{\mathrm{d}\mu_n^*}{\mathrm{d}\mu^{\otimes n}}\frac{\mathbb{1}_{(C^{\varepsilon})^c}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\right)}{\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}\,\mathrm{d}\mu^{\otimes n}.\end{aligned}$$

But $x\mapsto x\log(x)$ is always greater than $-\dfrac{1}{e}$, so

$$\begin{aligned}\frac{1}{n}\int\frac{\mathrm{d}\mu_n^*}{\mathrm{d}\mu^{\otimes n}}\log\frac{\mathrm{d}\mu_n^*}{\mathrm{d}\mu^{\otimes n}}\frac{\mathbb{1}_{(C^{\varepsilon})^c}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\right)}{\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}\,\mathrm{d}\mu^{\otimes n}&\geq-\frac{\mu^{\otimes n}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\notin C^{\varepsilon}\right)}{ne\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}\\&\geq-\frac{1}{ne\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}.\end{aligned}$$

Hence

$$\begin{aligned}\frac{1}{n}\log\mu^{\otimes n}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)\geq&-\frac{\mathrm{H}\left(\mu_n^*\,|\,\mu^{\otimes n}\right)}{n\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}+\frac{1}{n}\log\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)\\&-\frac{1}{ne}\frac{1}{\mu_n^*\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n\in C^{\varepsilon}\right)}\end{aligned}$$

and adding $\dfrac{\mathrm{H}\left(\mu_n^*\,|\,\mu^{\otimes n}\right)}{n}$ at both sides, the result follows. □

Proof of Theorem 3.2. It suffices to show that

$$\|R_{n,\,\varepsilon_n}-R^*_{n,\,\varepsilon_n}\|_{TV}\xrightarrow[n\to+\infty]{}0.$$

According to Proposition 5.1 (2), there are $\theta\in\Theta$ and n_2 such that for all $n\geq n_2$

$$\|R_{n,\,\varepsilon_n} - R^*_{n,\,\varepsilon_n}\|_{TV} \le \theta\left(\frac{-1}{n}\log\left[\mu^{\otimes n}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n \in C^{\varepsilon_n}\right)\mathrm{e}^{\mathrm{H}\left(\mu^*_{n,\,\varepsilon_n}\,|\mu^{\otimes n}\right)}\right]\right).$$

The function θ being continuous nondecreasing vanishing at 0 it suffices to majorize

$$B_n := \frac{-1}{n}\log\left[\mu^{\otimes n}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n \in C^{\varepsilon_n}\right)\mathrm{e}^{\mathrm{H}\left(\mu^*_{n,\,\varepsilon_n}\,|\mu^{\otimes n}\right)}\right]$$

by a quantity converging to 0.

Let us write

$$B_n = B_n^1 + B_n^2,$$

with

$$B_n^1 = \frac{-1}{n}\log\left[\mu^{\otimes n}\left(\int_{\mathcal{X}}\Phi\,\mathrm{d}L_n \in C^{\varepsilon_n}\right)\mathrm{e}^{\mathrm{H}\left(\mu^*_n\,|\mu^{\otimes n}\right)}\right],$$

and

$$B_n^2 = \frac{1}{n}\left[\mathrm{H}\left(\mu^*_n\,|\,\mu^{\otimes n}\right) - \mathrm{H}\left(\mu^*_{n,\,\varepsilon_n}\,|\,\mu^{\otimes n}\right)\right].$$

A simple calculus yields

$$\frac{\mathrm{H}\left(\mu^*_n\,|\,\mu^{\otimes n}\right)}{n} = \frac{1}{n}\sum_{i=1}^n\left[\langle\Phi(x_i^n), v_n^*\rangle\Lambda'_\mu\langle\Phi(x_i^n), v_n^*\rangle - \Lambda_\mu\langle\Phi(x_i^n), v_n^*\rangle\right]$$

and

$$\frac{\mathrm{H}\left(\mu^*_{n,\,\varepsilon_n}\,|\,\mu^{\otimes n}\right)}{n} = \frac{1}{n}\sum_{i=1}^n\left[\langle\Phi(x_i^n), v^*_{n,\,\varepsilon_n}\rangle\Lambda'_\mu\langle\Phi(x_i^n), v^*_{n,\,\varepsilon_n}\rangle - \Lambda_\mu\langle\Phi(x_i^n), v^*_{n,\,\varepsilon_n}\rangle\right].$$

Using Assumption 2 (2), Theorem 2.2 (4), and Lemma 5.1 one easily concludes that $\frac{\mathrm{H}\left(\mu^*_n\,|\mu^{\otimes n}\right)}{n}$ and $\frac{\mathrm{H}\left(\mu^*_{n,\,\varepsilon_n}\,|\mu^{\otimes n}\right)}{n}$ converge to the same limit ℓ, as n tends to $+\infty$[1]:

$$\ell = \int\langle\Phi(x), v^*\rangle\Lambda'_\mu\langle\Phi(x), v^*\rangle\,\mathrm{d}R(x) - \int\Lambda_\mu\langle\Phi(x), v^*\rangle\,\mathrm{d}R(x).$$

In particular,

$$B_n^2 \xrightarrow[n\to+\infty]{} 0.$$

Finally, thanks to Lemmas 5.2 and 5.3, one can easily see that B_n^1 is majorized by a quantity converging to 0. $\square$

Appendix A. Proof of Theorem 2.2

This proof of Theorem 2.2 is contained in several parts of the paper by Gamboa and Gassiat [9, 13–15]. For the sake of completeness and reader's convenience, we give below a complete proof, slightly improving their results for points (4) and (5).

We need the following results to prove Theorem 2.2:

Proposition A.1. *$C \subset \mathbb{R}^k$ is convex and compact, A is a $k \times n$ matrix and ν a probability measure on $\mathbb{R}^n$. Suppose* $\operatorname{dom} Z_\mu$ *is open and let $\mathcal{S}_\nu$ denote the support of ν.*

If

$$A^{-1}(C) \cap \overset{\circ}{\mathrm{co}}\,\mathcal{S}_\nu \neq \emptyset,$$

[1] Remark: $\ell = \mathrm{I}_\mu\left(R^*\,|R\right)$.

($\overset{\circ}{\mathrm{co}}\,\mathcal{S}_\nu$ denoting the interior of the convex hull of $\mathcal{S}_\nu$), then ν has an I-projection ν^* on

$$\Pi(C) = \{\alpha \in \mathcal{P}(\mathbb{R}^n) : A\mathbb{E}_\alpha[X] \in C\}.$$

Moreover,

$$\frac{\mathrm{d}\nu^*}{\mathrm{d}\nu} = \frac{\exp\langle A^t u^*, \, . \rangle}{Z_\nu(A^t u^*)},$$

where u^* minimises the function H defined on $\mathbb{R}^k$ by

$$H(u) = \log Z_\nu(A^t u) - \inf_{y\in C}\langle u, y\rangle.$$

This proposition is proved in [14].

The following lemma gives the convergence of solutions of a sequence of minimisation problems (see Chap. 7 of [23] for more general results).

Lemma A.1. *Let $(H_n)_n$ be a sequence of convex functions on $\mathbb{R}^k$ with values in $\mathbb{R}\cup\{+\infty\}$, and H a convex function on $\mathbb{R}^k$ with values in $\mathbb{R}\cup\{+\infty\}$.*
Suppose that

- *for all n, $\emptyset \neq \overset{\circ}{\mathrm{dom}}\, H \subset \mathrm{dom}\, H_n$;*
- *for all n sufficiently large, the set $\mathrm{Argmin}\, H_n$ of all minimizers of H_n is nonempty;*
- *H has only one minimizer v^* belonging to $\overset{\circ}{\mathrm{dom}}\, H$;*
- *$(H_n)_n$ converges pointly to H on $\overset{\circ}{\mathrm{dom}}\, H$,*

then for all $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\mathrm{Argmin}\, H_n \subset B(v^*, \varepsilon).$$

Proof. As $\mathrm{Argmin}\, H = \{v^*\}$, the convex function H has bounded sublevel sets (see *e.g.* Prop. 3.2.4 p. 107 of [17]). Thus, Lemma A.1 follows from Theorems 7.33, 7.17 and point (c) of exercise 7.32 of [23]. For reader convenience, we give bellow a simple proof of Lemma A.1 relying only on basic convex analysis.

Let us assume that there is an $r > 0$ such that $B(v^*, r) \subset \overset{\circ}{\mathrm{dom}}\, H$ and a sequence $v_n^* \in \mathrm{Argmin}\, H_n$ satisfying $\|v_n^* - v^*\| > r$ for all $n \in \mathbb{N}$.

Let $\overline{v}_n \in B\left(v^*, \frac{r}{3}\right)$ be such that

$$H_n(\overline{v}_n) = \min\left\{H_n(v) : v \in B\left(v^*, \frac{r}{3}\right)\right\}\cdot$$

The sequence $(\overline{v}_n)_n$ is bounded; let $\overline{v}$ be an accumulation point of $\overline{v}_n$ and ϕ such that $\lim\limits_{n\to+\infty} \overline{v}_{\phi(n)} = \overline{v}$.

As $(H_n)_n$ is a sequence of convex functions converging pointwise to H on $\overset{\circ}{\mathrm{dom}}\, H$, it converges uniformly on every compact subset of $\overset{\circ}{\mathrm{dom}}\, H$ (see [17], Th. 3.1.4 p. 105 or Th. 7.17 of [23]). In particular,

$$\left|H_{\phi(n)}(\overline{v}_{\phi(n)}) - H(\overline{v}_{\phi(n)})\right| \leq \sup_{v\in B\left(v^*,\frac{r}{3}\right)} \left|H_{\phi(n)}(v) - H(v)\right| \xrightarrow[n\to+\infty]{} 0.$$

Moreover, H being continuous, $H(\overline{v}_{\phi(n)}) \xrightarrow[n\to+\infty]{} H(\overline{v})$, so $H_{\phi(n)}(\overline{v}_{\phi(n)}) \xrightarrow[n\to+\infty]{} H(\overline{v})$. Now, for all n, one has $H_{\phi(n)}(\overline{v}_{\phi(n)}) \leq H_{\phi(n)}(v^*)$, thus letting $n \to +\infty$, one gets: $H(\overline{v}) \leq H(v^*)$. As v^* is the only one minimizer of H, one has $\overline{v} = v^*$. Thus, v^* is the only one accumulation point of the bounded sequence $(\overline{v}_n)_n$. From this follows that $(\overline{v}_n)_n$ converges to v^*.

For all $n \in \mathbb{N}$, let h_n be defined by

$$h_n : [0,1] \to \mathbb{R} : t \mapsto H_n(v_n^* + t(\overline{v}_n - v_n^*)).$$

The function h_n is convex and attains its minimal value at 0. Consequently, h_n is nondecreasing. Let $t_n \in [0,1]$ be such that $\frac{2r}{3} \leq |v_n^* + t_n(\overline{v}_n - v_n^*) - v^*| \leq r$ and define $z_n = v_n^* + t_n(\overline{v}_n - v_n^*)$. For all n, one has

$$H_n(z_n) \leq H_n(\overline{v}_n) \qquad \text{and} \qquad \frac{2r}{3} \leq |z_n - v^*| \leq r.$$

Thanks to compactness, one can assume that $(z_n)_n$ converges to some z satisfying $\frac{2r}{3} \leq |z - v^*| \leq r$.

As (H_n) converges uniformly to H on $B(v^*, r)$, one easily concludes that $\lim_{n\to+\infty} H_n(z_n) = H(z)$, and letting $n \to +\infty$ in the last inequality, one gets $H(z) \leq H(v^*)$, and so $z = v^*$ - absurd. □

Proof of Theorem 2.2.
Proof of (1) and (2). For all $\nu \in \mathcal{P}(\mathbb{R}^n)$,

$$\begin{aligned}
\mathbb{E}_\nu\left[\int_{\mathcal{X}} \Phi \,\mathrm{d}L_n\right] &= \mathbb{E}_\nu\left[\frac{1}{n}\sum_{i=1}^n z_i\varphi_1(x_i^n), \ldots, \frac{1}{n}\sum_{i=1}^n z_i\varphi_k(x_i^n)\right] \\
&= \frac{1}{n}\begin{pmatrix} \varphi_1(x_1^n) & \ldots & \varphi_1(x_n^n) \\ \vdots & \ldots & \vdots \\ \varphi_k(x_1^n) & \ldots & \varphi_k(x_n^n) \end{pmatrix} \mathbb{E}_\nu[X] \\
&= A_n \mathbb{E}_\nu[X],
\end{aligned}$$

so

$$\Pi_n(C^\varepsilon) = \{\nu \in \mathcal{P}(\mathbb{R}^n) : A_n\mathbb{E}_\nu[X] \in C^\varepsilon\}.$$

Let $\mathcal{S}_{\mu^{\otimes n}}$ denote the support of $\mu^{\otimes n}$ and let us admit that

$$\exists n_0, \quad \forall n \geq n_0, \quad A_n^{-1}(C) \cap \overset{\circ}{\mathrm{co}}\, \mathcal{S}_{\mu^{\otimes n}} \neq \emptyset. \tag{A.1}$$

We will prove (A.1) later; note that for all $\varepsilon \geq 0$, we have also

$$\forall n \geq n_0, \quad A_n^{-1}(C^\varepsilon) \cap \overset{\circ}{\mathrm{co}}\, \mathcal{S}_{\mu^{\otimes n}} \neq \emptyset. \tag{A.2}$$

As $\mathrm{dom}\, Z_{\mu^{\otimes n}} =]-\alpha, \beta[^n$ is open, one can apply Proposition A.1:

- $\mu^{\otimes n}$ has an I-projection $\mu^*_{n,\varepsilon}$ on $\Pi_n(C^\varepsilon)$, which proves (1).
- $\mu^*_{n,\varepsilon}$ satisfies:

$$\frac{\mathrm{d}\mu^*_{n,\varepsilon}}{\mathrm{d}\mu^{\otimes n}} = \frac{\exp\langle A_n^t u^*_{n,\varepsilon}, \,.\,\rangle}{Z_{\mu^{\otimes n}}(A_n^t u^*_{n,\varepsilon})},$$

  where $u^*_{n,\varepsilon} \in \mathbb{R}^k$ is a minimizer of:

$$G_{n,\varepsilon}(u) = \Lambda_{\mu^{\otimes n}}(A_n^t u) - \inf_{y \in C^\varepsilon} \langle u, c\rangle.$$

Now for all $x \in]-\alpha, \beta[^n$

$$\Lambda_{\mu^{\otimes n}}(x) = \Lambda_\mu(x_1) + \cdots + \Lambda_\mu(x_n)$$

and for all $u \in \mathbb{R}^k$,

$$A_n^t u = \begin{bmatrix} \frac{1}{n}\langle \Phi(x_1^n), u\rangle \\ \vdots \\ \frac{1}{n}\langle \Phi(x_n^n), u\rangle \end{bmatrix}.$$

Therefore,

$$\begin{aligned} G_{n,\varepsilon}(u) &= n\left[\frac{1}{n}\sum_{i=1}^n \Lambda_\mu \left\langle \Phi(x_i^n), \frac{u}{n}\right\rangle - \inf_{y\in C^\varepsilon}\left\langle \frac{u}{n}, y\right\rangle\right] \\ &= nH_{n,\varepsilon}\left(\frac{u}{n}\right), \end{aligned}$$

so $u^*_{n,\varepsilon}$ minimises $G_{n,\varepsilon}$ if and only if $\dfrac{u^*_{n,\varepsilon}}{n}$ minimises the functions $H_{n,\varepsilon}$ defined by (2.1).

Letting $v^*_{n,\varepsilon} = \dfrac{u^*_{n,\varepsilon}}{n}$ and $w^*_{n,\varepsilon} = \begin{bmatrix} \langle \Phi(x_1^n), v^*_{n,\varepsilon}\rangle \\ \vdots \\ \langle \Phi(x_n^n), v^*_{n,\varepsilon}\rangle \end{bmatrix}$, one obtains point (2).

Proof of (3).

$$R^*_{n,\varepsilon} = \mathbb{E}_{\mu^*_{n,\varepsilon}}[L_n] = \frac{1}{n}\sum_{i=1}^n \int z_i \,\mathrm{d}\mu^*_{n,\varepsilon}(z)\delta_{x_i^n},$$

but for all $w \in]-\alpha, \beta[$,

$$\int x \,\mathrm{d}\mu_w(x) = \Lambda'_\mu(w),$$

thus for all i:

$$\int z_i \,\mathrm{d}\mu^*_{n,\varepsilon}(z) = \int z \,\mathrm{d}\mu_{(w^*_{n,\varepsilon})_i}(z) = \Lambda'_\mu((w^*_{n,\varepsilon})_i) = \Lambda'_\mu\langle v^*_{n,\varepsilon}, \Phi(x_i^n)\rangle$$

and

$$R^*_{n,\varepsilon} = \frac{1}{n}\sum_{i=1}^n \Lambda'_\mu\langle v^*_{n,\varepsilon}, \Phi(x_i^n)\rangle\delta_{x_i^n}.$$

Proof of (A.1). Let J_μ denote the closed convex hull of the support of μ. It is easily seen that $\operatorname{co}\mathcal{S}_{\mu^{\otimes n}} = (J_\mu)^n$. Let us show that for n sufficiently large, there is $z^n \in (\mathring{J}_\mu)^n$ such that $A_n z^n \in C$.

Let $\mathcal{C}_\mu(\mathcal{X})$ denote the set of continuous functions on $\mathcal{X}$ with values in $\mathring{J}_\mu$. For $g \in \mathcal{C}_\mu(\mathcal{X})$, we define

$$z^n(g) = (g(x_1^n), \ldots, g(x_n^n)) \in (\mathring{J}_\mu)^n.$$

Notice that, for $g \in \mathcal{C}_\mu(\mathcal{X})$

$$A_n z^n(g) = \left[\frac{1}{n}\sum_{i=1}^n g(x_i^n)\varphi_1(x_i^n), \ldots, \frac{1}{n}\sum_{i=1}^n g(x_i^n)\varphi_k(x_i^n)\right].$$

Thus, according to Assumption 2 (2),

$$A_n z^n(g) \xrightarrow[n\to+\infty]{} \int_{\mathcal{X}} g(x)\Phi(x)\,\mathrm{d}R(x).$$

Now Assumption 2 (3) tells us there is $g_0 \in \mathcal{C}_\mu(\mathcal{X})$ such that

$$c_0 := \int_{\mathcal{X}} g_0(x)\Phi(x)\,\mathrm{d}R(x) \in C.$$

Assume that there is an increasing sequence of integers $(n_p)_p$ such that for all p and $g \in \mathcal{C}_\mu(\mathcal{X})$

$$A_{n_p} z_{n_p}(g) \neq c_0.$$

For all p, $\{A_{n_p} z_{n_p}(g) : g \in \mathcal{C}_\mu(\mathcal{X})\} \subset \mathbb{R}^k$ is convex and does not contain c_0. The separation theorem yields $u_{n_p} \in \mathbb{R}^k$ such that $\|u_{n_p}\| = 1$ and

$$\langle u_{n_p}, c_0\rangle \geq \sup_{g\in\mathcal{C}_\mu(\mathcal{X})} \langle u_{n_p}, A_{n_p} z_{n_p}(g)\rangle.$$

Thanks to compactness, one can suppose that u_{n_p} converges to u. For all $g \in \mathcal{C}_\mu(\mathcal{X})$, $\langle u_{n_p}, c_0\rangle \geq \langle u_{n_p}, A_{n_p} z_{n_p}(g)\rangle$, thus letting $n \to +\infty$ in this inequality, one gets

$$\langle u, c_0\rangle \geq \left\langle u, \int_{\mathcal{X}} g(x)\Phi(x)\,\mathrm{d}R(x)\right\rangle.$$

Therefore, for all $g \in \mathcal{C}_\mu(\mathcal{X})$,

$$\left\langle u, \int_{\mathcal{X}} (g - g_0)(x)\Phi(x)\,\mathrm{d}R(x)\right\rangle \leq 0.$$

Let B be the unit ball of $\mathcal{C}(\mathcal{X})$ (the set of continuous real valued functions on $\mathcal{X}$). For sufficiently small $r > 0$, $g_0 + rB \subset \mathcal{C}_\mu(\mathcal{X})$. Hence for all $g \in rB$,

$$\left\langle u, \int_{\mathcal{X}} g(x)\Phi(x)\,\mathrm{d}R(x)\right\rangle \leq 0,$$

and thanks to symmetry and homogeneity,

$$\int_{\mathcal{X}} g(x)\langle u, \Phi(x)\rangle\,\mathrm{d}R(x) = 0$$

holds for all $g \in \mathcal{C}(\mathcal{X})$.

Consequently,

$$R(\langle u, \Phi(x)\rangle = 0) = 1$$

and according to Assumption 1 (2),

$$\langle u, \Phi(x)\rangle = 0 \text{ for all } x \in \mathcal{X}. \tag{A.3}$$

As $u \neq 0$, (A.3) contradicts Assumption 1 (3). Therefore, for all n sufficiently large, there is $z^n \in (\overset{\circ}{J}_\mu)^n$ such that $A_n z^n \in C$.

Proof of (4). The map

$$H(\,.\,) = \int_{\mathcal{X}} \Lambda_\mu\langle\,.\,,\Phi(x)\rangle\,\mathrm{d}R(x) - \inf_{y\in C}\langle\,.\,,y\rangle$$

satisfies

$$\overset{\circ}{\operatorname{dom}} H = \left\{v \in \mathbb{R}^k : \forall x \in \mathcal{X}, \langle v, \Phi(x)\rangle \in]-\alpha, \beta[\right\}$$

and clearly

$$\overset{\circ}{\operatorname{dom}} H \subset \operatorname{dom} H_{n,\varepsilon_n},$$

where $H_{n,\,\varepsilon_n}$ is defined by

$$H_{n,\,\varepsilon_n}(v) = \frac{1}{n}\sum_{i=1}^{n} \Lambda_\mu \langle v, \Phi(x_i^n)\rangle - \inf_{y\in C^{\varepsilon_n}} \langle v, y\rangle.$$

Using the convexity of Λ_μ, one can easily verify that the functions H and $H_{n,\,\varepsilon_n}$ are convex. For all $v \in \overset{\circ}{\operatorname{dom}} H$, the function $\Lambda_\mu\langle v, \Phi(\,.\,)\rangle$ is bounded, so according to Assumption 2 (2), $(H_{n,\,\varepsilon_n})_n$ converges pointwise to H on $\overset{\circ}{\operatorname{dom}} H$. Moreover, according to Assumption 2 (4), H has only one minimizer $v^* \in \overset{\circ}{\operatorname{dom}} H$. Applying Lemma A.1, one concludes that $v^*_{n,\,\varepsilon_n}$ converges to v^*.

Proof of (5).
For all $g \in \mathcal{C}(\mathcal{X})$, one has

$$\int_{\mathcal{X}} g \,\mathrm{d}R^*_{n,\,\varepsilon_n} = \frac{1}{n}\sum_{i=1}^{n} \Lambda'_\mu\langle v^*_{n,\,\varepsilon_n}, \Phi(x_i^n)\rangle g(x_i^n).$$

Lemma 5.1 implies there are a compact interval $K \subset]-\alpha, \beta[$ and m such that, for all $n \geq m$, one has

$$\forall n \geq m, \quad \langle v^*_{n,\,\varepsilon_n}, \Phi(x_i^n)\rangle \in K \qquad \text{and} \qquad \forall x \in \mathcal{X}, \quad \langle v^*, \Phi(x)\rangle \in K.$$

If $M = \sup_{x\in K} \Lambda''_\mu(x)$, it holds:

$$\left| \int_{\mathcal{X}} g\,\mathrm{d}R^*_{n,\,\varepsilon_n} - \frac{1}{n}\sum_{i=1}^{n}\Lambda'_\mu\langle v^*, \Phi(x_i^n)\rangle g(x_i^n)\right| \leq M \sup|g| \,\sup\|\Phi\|\, \|v^* - v^*_{n,\,\varepsilon_n}\| \xrightarrow[n\to+\infty]{} 0.$$

Finally,

$$\frac{1}{n}\sum_{i=1}^{n}\Lambda'_\mu\langle v^*, \Phi(x_i^n)\rangle g(x_i^n) = \int_{\mathcal{X}} \Lambda'_\mu\langle v^*, \Phi(\,.\,)\rangle g(\,.\,)\,\mathrm{d}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i^n}\right)$$

and as $\Lambda'_\mu\langle v^*, \Phi(\,.\,)\rangle g(\,.\,) \in \mathcal{C}(\mathcal{X})$, it follows from Assumption 2 (2) that

$$\int_{\mathcal{X}} g\,\mathrm{d}R^*_{n,\,\varepsilon_n} \xrightarrow[n\to+\infty]{} \int_{\mathcal{X}} \Lambda'_\mu\langle v^*, \Phi(x)\rangle g(x)\,\mathrm{d}R(x),$$

for all $g \in \mathcal{C}(\mathcal{X})$. □

Acknowledgements. I would like to warmly acknowledge Patrick Cattiaux, my Ph.D. advisor, for the many useful advices he gave to me and for his constant support during the preparation of this work. I wish also to acknowledge a very accurate referee in particular for his very interesting questions and comments.

References

[1] S.G. Bobkov and F. Gotze, Exponential integrability and transportation cost related to logarithmic sobolev inequalities. *J. Funct. Anal.* **163** (1999) 1–28.

[2] J.M. Borwein and A.S. Lewis, Duality relationships for entropy-like minimization problems. *SIAM J. Control Optim.* **29** (1991) 325–338.

[3] J.M. Borwein and A.S. Lewis, Partially-finite programming in L_1 and the exitence of maximum entropy estimates. *SIAM J. Optim.* **3** (1993) 248–267.

[4] P. Cattiaux and N. Gozlan, *Deviations lower bounds and conditional principles.* Prépublications de l'Université Paris 10, Nanterre (2002).

[5] I. Csiszar, I-divergence geometry of probability distributions and minimization problems. *Ann. Prob.* **3** (1975) 146–158.

[6] I. Csiszar, Sanov property, generalized I-projection and a conditional limit theorem. *Ann. Prob.* **12** (1984) 768–793.

[7] I. Csiszar, Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Statist.* **19** (1991) 2032–2066.

[8] I. Csiszar, F. Gamboa and E. Gassiat, Mem pixel correlated solutions for generalized moment and interpolation problems. *IEEE Trans. Inform. Theory* **45** (1999) 2253–2270.

[9] D. Dacunha-Castelle and F. Gamboa, Maximum d'entropie et problèmes des moments. *Ann. Inst. Henri Poincaré* **26** (1990) 567–596.

[10] A. Dembo and O. Zeitouni, *Large deviations techniques and applications. Second edition.* Springer-Verlag (1998).

[11] J.D. Deuschel and D.W. Stroock, *Large deviations.* Academic Press (1989).

[12] R.S. Ellis, J. Gough and J.V. Pulé, The large deviation principle for measures with random weights. *Rev. Math. Phys.* **5** (1993) 659–692.

[13] F. Gamboa, *Méthode du maximum d'entropie sur la moyenne et applications.* Thèse Orsay (1989).

[14] F. Gamboa and E. Gassiat, Maximum d'entropie et problèmes des moments: Cas multidimensionnel. *Probab. Math. Statist.* **12** (1991) 67–83.

[15] F. Gamboa and E. Gassiat, Bayesian methods and maximum entropy for ill-posed inverse problems. *Ann. Statist.* **25** (1997) 328–350.

[16] N. Gozlan, *Principe conditionnel de Gibbs pour des contraintes fines approchées et inégalités de transport.* Université Paris 10-Nanterre (2005).

[17] J.B. Hirriart-Urruty and C. Lemaréchal, *Fundamentals of convex analysis.* Springer-Verlag (2001).

[18] C. Léonard, Minimizer of energy functionals. *Acta Math. Hungar.* **93** (2001) 281–325.

[19] C. Léonard, *A convex optimization problem arising from probabilistic questions.* Prépublications de l'Université Paris 10-Nanterre (2004).

[20] C. Léonard, *Dominating points and entropic projections.* Prépublications de l'Université Paris 10-Nanterre (2004).

[21] P. Massart, Saint-Flour Lecture Notes (2003).

[22] J. Najim, A Cramer type theorem for weighted random variables. *Electronic J. Probab.* **7** (2002).

[23] R.T. Rockafellar and R. Wets, *Variational Analysis.* Springer-Verlag (1997).

[24] D.W. Stroock and O. Zeitouni, Microcanonical distributions, Gibbs states and the equivalence of ensembles, R. Durret and H. Kesten Eds., Birkhäuser. *Festschrift in honour of F. Spitzer* (1991) 399–424.

[25] A. Van Der Vaart and J. Wellner, *Weak convergence and empirical processes.* Springer Series in Statistics. Springer (1995).