Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 117 (2007) 221–250
stochastic processes and their applications
www.elsevier.com/locate/spa

Deviations bounds and conditional principles for thin sets

Patrick Cattiaux[a,b,*], Nathael Gozlan[b]

[a] Ecole Polytechnique, CMAP, F-91128 Palaiseau cedex, CNRS 756, France
[b] Université Paris X Nanterre, Équipe MODAL'X, UFR SEGMI, 200 Avenue de la République, F- 92001 Nanterre, Cedex, France

Received 12 October 2005; received in revised form 22 March 2006; accepted 24 July 2006
Available online 17 August 2006

Abstract

The aim of this paper is to use non-asymptotic bounds for the probability of rare events in the Sanov theorem, in order to study the asymptotics in conditional limit theorems (Gibbs conditioning principle for thin sets). Applications to stochastic mechanics and calibration problems for diffusion processes are discussed.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Large deviations; Gibbs conditioning principle

1. Introduction

Let $X_1, X_2, \dots$ be i.i.d. random variables taking their values in some metrizable space (E, d). Set $M_n = \frac{1}{n}\sum_{i=1}^n X_i$ as the empirical mean (assuming here that E is a vector space) and $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ as the empirical measure. In recent years new efforts have been made to understand the asymptotic behavior of laws conditioned by some rare or super-rare event.

The celebrated Gibbs conditioning principle is the corresponding meta-principle for the empirical measure, namely

$$\lim_{n\to+\infty} \mathbb{P}^{\otimes n}((X_1, \dots, X_k) \in B/L_n \in A) = (\mu^*)^{\otimes k}(B),$$

* Corresponding author at: Université Paris X Nanterre, Équipe MODAL'X, UFR SEGMI, 200 Avenue de la République, F- 92001 Nanterre, Cedex, France. Tel.: +33 1 69 33 39 78; fax: +33 1 40 97 47 59.
E-mail addresses: cattiaux@cmapx.polytechnique.fr (P. Cattiaux), nathael.gozlan@u-paris10.fr (N. Gozlan).

0304-4149/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.spa.2006.07.003

where μ^* minimizes the relative entropy $H(\mu^* \mid \mu)$ among the elements in A. When A is thin (i.e. $\mathbb{P}^{\otimes n}(L_n \in A) = 0$), such a statement is meaningless, so one can either try to look at regular disintegration (the so called "thin shell" case) or look at some enlargement of A. The first idea is also meaningless in general (see however the work by Diaconis and Freedman [16]). Therefore we shall focus on the second one.

An enlargement A_ε is then a non-thin set containing A, and the previous statement becomes a double limit one:

$$\lim_{\varepsilon\to 0}\lim_{n\to+\infty} \mathbb{P}^{\otimes n}((X_1,\ldots,X_k) \in B/L_n \in A_\varepsilon) = (\mu^*)^{\otimes k}(B).$$

Precise hypotheses are known for this meta-principle ("thick shell" case) to become a rigorous result, and refinements (namely one can choose some increasing $k(n)$) are known (see e.g. [11] and the references therein). One possible way to prove this result is to identify relative entropy with the rate function in the large deviations principle for empirical measures (Sanov's theorem). In this paper we will introduce an intermediate "approximate thin shell" case, i.e. we will look at a single limit and make the enlargement size depend on n, i.e. $\varepsilon_n \to 0$. We shall also discuss in detail one case of a "super-thin" set, i.e. when relative entropy is infinite for any element in A.

Of course since we are considering conditional probabilities, we are led to get both lower and upper non-asymptotic estimates for the probability of rare events.

The paper is organized as follows.

In Section 2 we shall introduce the notation and recall some results we shall use repeatedly. Then we give the main general result (Theorem 2.7).

When A is some closed subspace (i.e. defined by linear constraints), our program can be carried out by directly using well known inequalities for the sum of independent variables. This will be explained in Section 3. The main result of the section is Theorem 3.4. Roughly speaking, in this situation, the enlargement size can be chosen of order $1/\sqrt{n}$.

The more general case of a general convex constraint is studied in Section 4. In the compact case upper estimates are well known and lower estimates will be derived thanks to a result by Deuschel and Stroock. In both cases one has to compute the metric entropy (i.e. the number of small balls needed to cover A) for some metric compatible with the convergence of measures (see Proposition 4.6). The extension to non-compact convex constraints is done by choosing an adequate rich enough compact subset (see Theorem 4.8). In this situation, the enlargement size heavily depends on the metric entropy.

Section 5 is devoted to some examples, first in a finite dimensional space. We next show that the Schrödinger bridges and the Nelson processes studied in stochastic mechanics are natural "limiting processes" for constraints of marginal type.

Section 6 is devoted to the study of a super-thin example corresponding to the well known problem of volatility calibration in mathematical finance. Our aim is to give a rigorous status to the "relative entropy minimization method" introduced in [2]. The problem here is to choose the diffusion coefficient (volatility) of a diffusion process with a given drift (risk neutral drift), knowing some final moments of the diffusion process. Of course all the possible choices are mutually singular so that the constraint set A does not contain any measure with finite relative entropy, i.e. is super-thin. We shall show that under some conditions, the method of Avellaneda et al. [2] enters our framework, and hence furnishes the natural candidate from a statistical point of view (we shall not discuss any kind of financial related aspect).

Another famous example involving a super-thin set is furnished by statistical mechanics, namely: are the Gibbs measures associated with some Hamiltonian the limiting measures of some conditional law of large numbers? The positive answer gives an interpretation of the famous equivalence of ensembles principle (see [23,15]). It should be interesting to relate the Gibbs variational principle as in [15] to the above Gibbs conditioning principle. This is not done here.

2. Notation and first basic results

Throughout the paper (E, d) will be a Polish space. $M_1(E)$ (resp. $M(E)$) will denote the set of probability measures (resp. bounded signed measures) on E equipped with its Borel σ-field. $M_1(E)$ is equipped with the narrow topology (convergence in law) and its natural Borel σ-field.

In the sequel, we will consider a sequence $X_1, X_2, \ldots$ of i.i.d. E valued random variables. The common law of the X_i's will be denoted by α and their empirical measure by $L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$.

Our aim is to study the asymptotic behavior of the conditional law

$$\alpha^n_{A,k}(B) = \mathbb{P}^{\otimes n}_{\alpha}\left((X_1, \ldots, X_k) \in B / L_n \in A_n\right) \tag{2.1}$$

for some A_n converging to some thin set A when n goes to ∞.

The first tool we need is relative entropy. Recall that for β and γ in $M_1(E)$, the relative entropy $H(\beta \mid \gamma)$ is defined by the two equivalent formulas

(2.2.1) $H(\beta \mid \gamma) = \int \log\left(\frac{\mathrm{d}\beta}{\mathrm{d}\gamma}\right) \mathrm{d}\beta$, if this quantity is well defined and finite, $+\infty$ otherwise,

(2.2.2) $H(\beta \mid \gamma) = \sup\{\int \varphi \mathrm{d}\beta - \log \int \mathrm{e}^{\varphi}\, \mathrm{d}\gamma, \varphi \in C_b(E)\}$.

If B is a measurable set of $M_1(E)$ we will write

$$H(B \mid \gamma) = \inf\{H(\beta \mid \gamma), \beta \in B\}. \tag{2.3}$$

The celebrated Sanov theorem tells us that for any measurable set B

$$-H(\overset{\circ}{B} \mid \alpha) \le \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n \in B) \le \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n \in B) \le -H(\overline{B} \mid \alpha),$$

where the interior $\overset{\circ}{B}$ and the closure $\overline{B}$ of B are for the narrow topology.

Recall that one can reinforce the previous topology by considering the G-topology induced by some subset G of measurable functions containing all the bounded measurable functions. In particular if α satisfies the strong Cramer assumption, i.e. $\forall g \in G, \forall t > 0$,

$$\int \mathrm{e}^{tg(x)}\, \mathrm{d}\alpha(x) < +\infty, \tag{2.4}$$

the previous result is still true for the G-topology (see [18] thm. 1.7). When G is exactly the set of measurable and bounded functions, the G-topology is usually called the τ-topology.

It is thus particularly interesting to have some information on the possible Arginf in (2.3). The result below collects some known facts:

Theorem 2.5. *Let C be a measurable convex subset of $M_1(E)$ such that $H(C \mid \alpha) < +\infty$.*

(1) *There exists a unique probability measure α^* such that any sequence ν_n of C such that $\lim_{n\to+\infty} H(\nu_n \mid \alpha) = H(C \mid \alpha)$ converges in total variation distance to α^*.*
(2) *This probability measure (which we shall call the generalized I-projection of α on C) is characterized by the following Pythagoras inequality*

$$H(\nu \mid \alpha) \le H(\alpha^* \mid \alpha) + H(\nu \mid \alpha^*), \quad \forall \nu \in C.$$

(3) *If α^* belongs to C we shall call it the I-projection (non-generalized). In particular the I-projection on a total variation closed convex subset such that $H(C \mid \alpha) < +\infty$ always exists.*
(4) *Finally if α satisfies the strong Cramer assumption (2.4) one can replace total variation closed by G-closed in the previous statement.*

All these results can be found in [9,18] (see [20], chap. II, for more details).

Before stating our first results on thin constraints, we recall the known results on thick ones.

Theorem 2.6. (2.6.1) *(see [10]). If C is convex, closed for the τ-topology and satisfies $H(C \mid \alpha) = H(\mathring{C} \mid \alpha) < +\infty$ then $\alpha^n_{C,k}$ defined in (2.1) is well defined for n large enough, and converges (when n goes to ∞) in relative entropy to $\alpha^{*\otimes k}$, where α^* is the I-projection of α on C.*

(2.6.2) *(see [23]). If A is a measurable subset such that $H(\bar{A} \mid \alpha) = H(\mathring{A} \mid \alpha) < +\infty$, and if there exists a unique $\alpha^* \in \bar{A}$ such that $H(\bar{A} \mid \alpha) = H(\alpha^* \mid \alpha)$, then $\alpha^n_{A,k}$ converges to $\alpha^{*\otimes k}$ for the narrow convergence.*

When $H(\mathring{A} \mid \alpha) = +\infty$ (in particular if $\mathring{A}$ is empty) but $H(A \mid \alpha) < +\infty$ (thin constraints) we have to face some new problems. The strategy is then to enlarge A, considering some nice A_ε, and to consider limits first in n, next in ε. Here we shall consider enlargements depending on n. Here is a general result in this direction.

Theorem 2.7. *Let C_n be a non-increasing sequence of convex subsets, closed in the G-topology. Define $C = \bigcap_{n=1}^{\infty} C_n$. Assume that*

(2.7.1) $H(C \mid \alpha) < +\infty$,
(2.7.2) *α has an I-projection α^* on C,*
(2.7.3) $\lim_{n\to\infty} H(C_n \mid \alpha) = H(C \mid \alpha)$,
(2.7.4) $\liminf_{n\to\infty} \frac{1}{n} \log \alpha^{\otimes n}(L_n \in C_n) \geq -H(C \mid \alpha)$.

Then, for all $k \in \mathbb{N}^*$, $\alpha^n_{C_n,k}$ converges in total variation distance to $\alpha^{*\otimes k}$.

In particular, when $C_n = C$ the condition (2.7.4) is automatically satisfied if $H(\mathring{C} \mid \alpha) = H(C \mid \alpha)$ and α satisfies the strong Cramer assumption (2.4).

Recall that convergence in relative entropy implies convergence in total variation distance which in turn implies narrow convergence. This will help to compare the statements of Theorems 2.6 and 2.7. Since relative entropy does not satisfy the triangle inequality, the proof of Theorem 2.7 below cannot be extended to relative entropy convergence (see (2.10)).

Remark 2.8. Define

$$\mathbb{L}^a_\tau(\alpha) = \left\{ g \text{ measurable} : \forall s \in \mathbb{R}, \int e^{s|g|}\, d\alpha < +\infty \right\}.$$

If $G \subseteq \mathbb{L}^a_\tau(\alpha)$, we already know (see Theorem 2.5) that α^* exists as soon as $H(C \mid \alpha)$ is finite. In addition, since the relative entropy is a good rate function (according to [18] its level sets are compact) (2.7.3) is also satisfied. Hence, in this case, assuming $H(C \mid \alpha) < +\infty$, the only remaining condition to check is

$$\liminf_{n\to\infty} \frac{1}{n} \log \alpha^{\otimes n}(L_n \in C_n) \geq -H(C \mid \alpha). \tag{2.9}$$

Proof of Theorem 2.7. Let α_n^* be the generalized I-projection of α on C_n. Then

$$\begin{aligned}\|\alpha^n_{C_n,k} - \alpha^{*\otimes k}\|_{TV} &\le \|\alpha^n_{C_n,k} - \alpha_n^{*\otimes k}\|_{TV} + \|\alpha_n^{*\otimes k} - \alpha^{*\otimes k}\|_{TV} \\ &\le \sqrt{2H\left(\alpha^n_{C_n,k} \mid \alpha_n^{*\otimes k}\right)} + \sqrt{2H\left(\alpha^{*\otimes k} \mid \alpha_n^{*\otimes k}\right)} \\ &\le \sqrt{2H\left(\alpha^n_{C_n,k} \mid \alpha_n^{*\otimes k}\right)} + \sqrt{2kH\left(\alpha^* \mid \alpha_n^*\right)} \end{aligned} \tag{2.10}$$

where we have used successively the triangle inequality, the Pinsker inequality and the additivity of relative entropy. Since α^* is the I-projection of α on C, α^* belongs to C and all C_n, so that using Theorem 2.5,

$$H(C \mid \alpha) = H(\alpha^* \mid \alpha) \ge H(\alpha^* \mid \alpha_n^*) + H(C_n \mid \alpha).$$

Thanks to (2.7.3) we thus have $\lim_{n\to\infty} H(\alpha^* \mid \alpha_n^*) = 0$.

To finish the proof (according to (2.10)) it thus remains to show that $\lim_{n\to\infty} H(\alpha^n_{C_n,k} \mid \alpha_n^{*\otimes k}) = 0$. But thanks to (2.7.4), for n large enough, $\alpha^{\otimes n}(L_n \in C_n) > 0$, so that we may apply Lemma 2.11 below with $A = C_n$. This yields

$$\begin{aligned} H\left(\alpha^n_{C_n,k} \mid \alpha_n^{*\otimes k}\right) &\le -\frac{k}{n} \log\left(\alpha^{\otimes n}(L_n \in C_n)\, \mathrm{e}^{nH(C_n|\alpha)}\right) \\ &\le -\frac{k}{n} \log\left(\alpha^{\otimes n}(L_n \in C_n)\, \mathrm{e}^{nH(C|\alpha)}\right) + k(H(C \mid \alpha) - H(C_n \mid \alpha)). \end{aligned}$$

According to (2.7.4) the first term in the right hand side sum has a non-positive lim sup, while the second term goes to 0 thanks to (2.7.3). Since the left hand side is non-negative the result follows. □

We now recall the key lemma due to Csiszar [10] we have just used:

Lemma 2.11. *Let A be a convex G-closed subset, such that $H(A \mid \alpha) < +\infty$. Denote by α^* the generalized I-projection of α on A. Then if $\alpha^{\otimes n}(L_n \in A) > 0$, for all $k \in \mathbb{N}^*$,*

$$H\left(\alpha^n_{A,k} \mid \alpha^{*\otimes k}\right) \le -\frac{k}{n} \log\left(\alpha^{\otimes n}(L_n \in A)\, \mathrm{e}^{nH(A|\alpha)}\right).$$

Under some additional assumption one can improve the convergence in Theorem 2.7. Introduce the usual Orlicz space

$$\mathbb{L}_\tau(\alpha) = \left\{g \text{ measurable } : \exists s \in \mathbb{R}, \int \mathrm{e}^{s|g|(x)}\, \mathrm{d}\alpha(x) < +\infty\right\}.$$

Note the difference from $\mathbb{L}^a_\tau$ (for which $\exists$ is replaced by $\forall$). We equip $\mathbb{L}_\tau$ with the Luxemburg norm

$$\|g\|_\tau = \inf\left\{s > 0, \int \tau(g(x)/s)\, \mathrm{d}\alpha(x) \le 1\right\} \quad \text{where } \tau(u) = \mathrm{e}^{|u|} - |u| - 1.$$

It is well known that the dual space of $\mathbb{L}_\tau(\alpha)$ contains the set of probability measures ν such that $H(\nu \mid \alpha) < +\infty$. We equip this dual space with the dual norm $\|\,\|^*_\tau$.

Proposition 2.12. *In addition to all the assumptions in Theorem 2.7, assume the following: the densities $h_n = \frac{d\alpha_n^*}{d\alpha}$ (α_n^* being the generalized I-projection of α on C_n) define a bounded sequence in $\mathbb{L}^p(\alpha)$ for some $p > 1$. Then*

$$\lim_{n\to+\infty} \|\alpha_{C_n}^n - \alpha^*\|_\tau^* = 0.$$

Proof. The proof is exactly the same as the one of Theorem 2.7 with $k = 1$, just replacing $\|\,\|_{TV}$ by $\|\,\|_\tau^*$ in the first line of (2.10), and then replacing the Pinsker inequality by the following one, available for ν_i's such that $H(\nu_i \mid \alpha) < +\infty$,

$$\|\nu_1 - \nu_2\|_\tau^* \le qC\left(1 + \log\left(4^{1/q}\left\|\frac{d\nu_2}{d\alpha}\right\|_p\right)\right)\left(H(\nu_1 \mid \nu_2) + \sqrt{H(\nu_1 \mid \nu_2)}\right), \tag{2.13}$$

where $q = p/(p-1)$, ν_2 being α_n^* and ν_1 being firstly $\alpha_{C_n}^n$ and secondly α^*.

In order to prove (2.13) we first recall the weighted Pinsker inequality recently shown by Bolley and Villani [3] (also see [20] for another approach): there exists some m such that for all non-negative f and all $\delta > 0$,

$$\|f\nu_1 - f\nu_2\|_{TV} \le (m/\delta)\left(1 + \log\int e^{\delta f} d\nu_2\right)\left(H(\nu_1 \mid \nu_2) + \sqrt{H(\nu_1 \mid \nu_2)}\right),$$

where $f\nu$ denotes the non-negative measure having density f with respect to ν.

For an f such that $\|f\|_\tau \le 1$ and $\delta = 1/q$ it thus holds, first that $\int e^{|f|} d\alpha \le 4$, then thanks to Hölder's inequality that $\int e^{\delta|f|} d\nu_2 \le 4^\delta \|\frac{d\nu_2}{d\alpha}\|_p$. (2.13) immediately follows. $\square$

3. F moment constraints

In this section $G = \mathbb{L}_\tau(\alpha)$ and we consider constraints C in the form

$$C = \left\{\nu \in M_1(E), \int F\, d\nu \in K\right\}$$

where F is a measurable B valued map $((B, \|.\|)$ being a separable Banach space equipped with its cylindrical σ-field) where $\int F d\nu$ denotes the Bochner integral and K is a closed convex set of B. The topological dual space of B will be denoted by B'. We denote by

$$\forall \lambda \in B', \quad Z_F(\lambda) = \int_E \exp(\langle \lambda, F(x)\rangle)\, \alpha(dx), \qquad \Lambda_F(\lambda) = \log Z_F(\lambda)$$

the Laplace transform and cumulant generating function of F.

We always assume that F satisfies the following hypotheses

$$(H\text{–}F) \quad \begin{cases} \|F\| \in \mathbb{L}_\tau(\alpha), \\ \mathrm{dom}\, \Lambda_F = \{\lambda \in B', \Lambda_F(\lambda) < +\infty\} \text{ is a non-empty open set of } B'. \end{cases}$$

The enlargement C_n is defined similarly

$$C_n = \left\{\nu \in M_1(E), \int F d\nu \in K^{\varepsilon_n}\right\}$$

for $K^{\varepsilon_n} = \{x \in B, d(x, C_n) \le \varepsilon_n\}$.

What we have to do is to check all the assumptions of Theorem 2.7. But the situation here is particular since the condition $L_n \in C$ reduces to $\frac{1}{n}\sum_{i=1}^n F(X_i) \in K$. Thanks to the next Lemma 3.1 assumption (2.7.4) reduces to well known estimates:

Lemma 3.1. *Assume that the I-projection α^* of α on C exists and can be written $\alpha^* = \frac{e^{\langle \lambda^*, F\rangle}}{Z_F(\lambda^*)}\alpha$ for some $\lambda^* \in B'$. Then for all $\varepsilon > 0$,*

$$\frac{1}{n}\log\left(\alpha^{\otimes n}(L_n \in C_\varepsilon)\,e^{nH(\alpha^*|\alpha)}\right) \geq \frac{1}{n}\log\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^n F(Y_i) - \int F\mathrm{d}\alpha^*\right\| \leq \varepsilon\right) - \|\lambda^*\|\varepsilon,$$

where the Y_i's are i.i.d. random variables with common law α^*.

Proof. The proof uses the standard centering method in large deviations theory. Denote by $L_n^x = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ the empirical measure of $x = (x_1, \ldots, x_n)$. Then,

$$\begin{aligned}\alpha^{\otimes n}(L_n \in C_\varepsilon) &= \int \mathbb{1}_{C_\varepsilon}(L_n^x)\prod_{i=1}^n \frac{\mathrm{d}\alpha}{\mathrm{d}\alpha^*}(x_i)\,\mathrm{d}\alpha^{*\otimes n}(x)\\ &= \int \mathbb{1}_{C_\varepsilon}(L_n^x)\,\exp\left(-n\left\langle L_n^x, \log\frac{\mathrm{d}\alpha}{\mathrm{d}\alpha^*}\right\rangle\right)\,\mathrm{d}\alpha^{*\otimes n}(x)\\ &= e^{-nH(\alpha^*|\alpha)}\int \mathbb{1}_{C_\varepsilon}(L_n^x)\,\exp\left(-n\left\langle L_n^x - \alpha^*, \log\frac{\mathrm{d}\alpha}{\mathrm{d}\alpha^*}\right\rangle\right)\,\mathrm{d}\alpha^{*\otimes n}(x)\\ &= e^{-nH(\alpha^*|\alpha)}\int \mathbb{1}_{C_\varepsilon}(L_n^x)\,\exp\left(-n\left\langle \lambda^*, \frac{1}{n}\sum_{i=1}^n F(x_i) - \int F\mathrm{d}\alpha^*\right\rangle\right)\\ &\quad\times \mathrm{d}\alpha^{*\otimes n}(x).\end{aligned}$$

Now we may replace C_ε by its subset

$$\widetilde{C}_\varepsilon = \left\{\nu \in M_1(E), \int \|F\|\mathrm{d}\nu < +\infty \text{ and } \left\|\int F\mathrm{d}\nu - \int F\mathrm{d}\alpha^*\right\| \leq \varepsilon\right\}$$

and obtain

$$\begin{aligned}\alpha^{\otimes n}(L_n \in C_\varepsilon)e^{nH(\alpha^*|\alpha)} &\geq \int \mathbb{1}_{\widetilde{C}_\varepsilon}(L_n^x)e^{-n\langle\lambda^*, \frac{1}{n}\sum_{i=1}^n F(x_i) - \int F\mathrm{d}\alpha^*\rangle}\mathrm{d}\alpha^{*\otimes n}(x)\\ &\geq e^{-n\|\lambda^*\|\varepsilon}\int \mathbb{1}_{\widetilde{C}_\varepsilon}(L_n^x)\,\mathrm{d}\alpha^{*\otimes n}(x)\end{aligned}$$

which completes the proof. $\square$

The next Lemma 3.2 is well known in convex analysis. For a complete proof the reader is referred to [20], Lemma II.39,

Lemma 3.2. *Assume that $(H\text{–}F)$ and the following assumption $(H\text{–}K)$ hold*

$$(H\text{–}K)\qquad \begin{array}{c}\textit{The function } H(\lambda) = \Lambda_F(\lambda) - \inf_{y\in K}\langle\lambda, y\rangle\\ \textit{achieves its minimum at (at least one) } \lambda^*.\end{array}$$

Then

$$H(C \mid \alpha) = \sup_{\lambda\in B'}\left\{\inf_{y\in K}\langle\lambda, y\rangle - \Lambda_F(\lambda)\right\} = -\inf_{\lambda\in B'} H(\lambda)$$

is thus finite and the I-projection α^* of α on C exists and can be written as $\alpha^* = \frac{e^{\langle \lambda^*, F\rangle}}{Z_F(\lambda^*)}\alpha$. In particular if $K = \{x_0\}$ with $x_0 = \nabla \Lambda_F(\lambda_0)$ $(H–K)$ is satisfied.

Before to state our first general result let us recall a definition.

Definition 3.3. B is of type 2 if there exists some $a > 0$ such that for all sequences Z_i of $\mathbb{L}^2$ i.i.d. random variables with zero mean and variance equal to 1, the following holds

$$\mathbb{E}\left(\left\|\sum_{i=1}^n Z_i\right\|^2\right) \le a \sum_{i=1}^n \mathbb{E}(\|Z_i\|^2).$$

In particular a Hilbert space is of type 2.

We arrive at

Theorem 3.4. *Assume that B is of type 2 and that $(H–F)$ and $(H–K)$ are satisfied. If $\varepsilon_n > \frac{c}{\sqrt{n}}$ with $c = \sqrt{a\mathrm{Var}_{\alpha^*}(F)}$, then $\alpha^n_{C_n,k}$ converges to $\alpha^{*\otimes k}$ in total variation distance when $n \to \infty$.*

Proof. (2.7.1) and (2.7.2) are satisfied with our hypotheses, according to Lemma 3.2.

In order to prove (2.7.3) introduce the function H_n defined by

$$H_n(\lambda) = \Lambda_F(\lambda) - \inf_{y \in K^{\varepsilon_n}} \langle \lambda, y\rangle.$$

Of course

$$\inf H \le \inf H_n \le H_n(\lambda^*) \xrightarrow[n\to+\infty]{} \inf H = H(\lambda^*)$$

since H_n converges to H pointwise on the domain of H. We already know that $\inf H = -H(C \mid \alpha)$. It is thus enough to prove that $\inf H_n = -H(C_n \mid \alpha)$. But this is a consequence of Csiszar results ([9], thm 3.3, and [10], thms 2 and 3; also see [20], thm II.41, for another proof) since the intersection of the interior of K^{ε_n} and the convex hull of the support of the image measure $F^{-1}\alpha$ is non-empty.

Finally in order to prove (2.7.4), according to Lemmas 3.1 and 3.2 it is enough to check that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^n F(Y_i) - \int F \,\mathrm{d}\alpha^*\right\| \le \varepsilon_n\right) = 0. \tag{3.5}$$

To this end recall the following theorem of Yurinskii (which holds even if B is not of type 2).

Theorem 3.6 (*Yurinskii, [24], Theorem 2.1*). *If Z_i is a B valued sequence of centered independent variables such that there exist b and M both positive, with*

$$\forall i \in \mathbb{N}^*,\ \forall k \ge 2, \quad \mathbb{E}(\|Z_i\|^k) \le \frac{k!}{2} b^2 M^{k-2},$$

then defining $S_n = \sum_{i=1}^n Z_i$ it holds that

$$\forall t > 0, \quad \mathbb{P}(\|S_n\| \ge \mathbb{E}(\|S_n\|) + nt) \le \exp\left(-\frac{1}{8}\frac{nt^2}{b^2 + tM}\right).$$

We may apply Theorem 3.6 with $Z_i = F(Y_i) - \int F\mathrm{d}\alpha^*$, $M = \|F - \int F\mathrm{d}\alpha^*\|_{\mathbb{L}_\tau(\alpha^*)}$ and $b = \sqrt{2}M$ assuming $F \in \mathbb{L}_\tau(\alpha^*)$. Indeed since B is of type 2, $\mathbb{E}(\|S_n\|) \le \sqrt{\mathbb{E}(\|S_n\|^2)} \le \sqrt{an}\sigma$ with $\sigma = \sqrt{\mathbb{E}(\|Z_1\|^2)}$. It follows that

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^n F(Y_i) - \int F\,\mathrm{d}\alpha^*\right\| \le \frac{\sigma\sqrt{a}}{\sqrt{n}}(1+t)\right) \ge 1 - \exp\left(-\frac{1}{8}\frac{a\sigma^2 t^2}{2M^2 + tM}\right).$$

If $\varepsilon_n\sqrt{n} > \sigma\sqrt{a}$, one thus has (3.5).

It remains to prove that $F \in \mathbb{L}_\tau(\alpha^*)$. But thanks to the representation of α^* obtained in Lemma 3.2

$$\begin{aligned}\int \mathrm{e}^{t\|F\|}\mathrm{d}\alpha^* &= \frac{1}{Z_F(\lambda^*)}\int \mathrm{e}^{t\|F\|}\mathrm{e}^{\langle\lambda^*,F\rangle}\,\mathrm{d}\alpha \\ &\le \frac{1}{Z_F(\lambda^*)}\left(\int \mathrm{e}^{tq\|F\|}\mathrm{d}\alpha\right)^{\frac{1}{q}}\left(\int \mathrm{e}^{\langle p\lambda^*,F\rangle}\mathrm{d}\alpha\right)^{\frac{1}{p}}.\end{aligned}$$

Since $\operatorname{dom}\Lambda_F$ is a non-empty open set containing λ^*, there exists some $p > 1$ such that $p\lambda^* \in \operatorname{dom}\Lambda_F$, and the result follows for t small enough since $F \in \mathbb{L}_\tau(\alpha)$. □

Remark 3.7. Note that if for instance F is bounded everything in Theorem 3.4 can be explicitly described with the only parameter n. However (unfortunately) we do not know any explicit bound for the speed of convergence of $\alpha^n_{C_n,k}$, because we do not know in general how to evaluate $H(C_n \mid \alpha) - H(C \mid \alpha)$. Hence from a practical point of view, if we know how to enlarge C, we do not know when a possible algorithm has to be stopped.

It is natural to ask whether $\varepsilon_n \approx 1/\sqrt{n}$ is the optimal order for the enlargement or not. In one dimension the answer is negative as we shall see below.

Theorem 3.8. *If $B = \mathbb{R}$ the conclusion of Theorem 3.4 remains true for $\varepsilon_n > c/n$ for some c large enough.*

Proof. We shall just replace Yurinskii's estimate by the Berry–Esséen bound. Indeed the Berry–Esséen theorem tells us that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n F(Y_i) - \int F\,\mathrm{d}\alpha^*\right| \le \varepsilon_n\right) \ge \Phi\left(\frac{\varepsilon_n\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\varepsilon_n\sqrt{n}}{\sigma}\right) - 20\frac{\kappa}{\sigma^3\sqrt{n}}$$

where $\Phi(u) = \int_{-\infty}^u \mathrm{e}^{-s^2/2}\mathrm{d}s/\sqrt{2\pi}$, $\sigma^2 = \mathrm{Var}_{\alpha^*}F$ and κ is the centered third moment of F with respect to α^*. It easily follows that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n F(Y_i) - \int F\,\mathrm{d}\alpha^*\right| \le \varepsilon_n\right) \ge \frac{2}{\sqrt{n}}\left(\frac{n\varepsilon_n}{\sqrt{2\pi}}\mathrm{e}^{-n\varepsilon_n^2/2\sigma^2} - 10(\kappa/\sigma^3)\right) = \theta_n.$$

The requested $1/n\log\theta_n \to 0$ follows with $\varepsilon_n = c/n$ provided $c > 10\sqrt{2\pi}(\kappa/\sigma^3)$. □

Again one may ask about optimality. Actually it is not difficult to build examples with $\varepsilon_n = c'/n$ for some small c' such that $\mathbb{P}(L_n \in C_n) = 0$ for all n. In a sense this is some proof of optimality. But we do not know how to build examples such that the previous probability is not zero. In higher dimension we do not know whether the optimal ε_n is of order $\frac{1}{n}$ or not but we think it is.

Finally we prove a stronger sense of convergence, still in the finite dimensional case under a slightly more restrictive assumption.

Theorem 3.9. *In Theorem 3.4 assume that $B = \mathbb{R}^d$ and replace the hypothesis $(H\text{–}K)$ by the following: $K \cap \overset{\circ}{S} \neq \emptyset$ where S is the convex hull of the support of the image measure $F^{-1}\alpha$. Then $\alpha^n_{C_n,k}$ converges to $\alpha^{*\otimes k}$ both for the dual norm $\| \|^*_\tau$ and in relative entropy.*

Proof. The first point is that the new hypothesis is stronger than $(H\text{–}K)$. Indeed it is known (see e.g. [12] or [20], Lemma III.65, for complete proofs) that not only does $(H\text{–}K)$ hold (as well as $(H\text{–}K^{\varepsilon_n})$ of course), but also the minimizers λ^* and λ^*_n are unique and $\lambda^*_n \to \lambda^*$ as $n \to \infty$. Hence $H(C_n \mid \alpha) \to H(C \mid \alpha)$ too.

Next $\int \left(\frac{\mathrm{d}\alpha^*_n}{\mathrm{d}\alpha}\right)^p \mathrm{d}\alpha = \frac{Z_F(p\lambda^*_n)}{Z^p_F(\lambda^*_n)}$. Since λ^*_n is a bounded (convergent) sequence, the above quantity can be easily bounded for some $p > 1$ (using again the fact that $\mathrm{dom}\Lambda_F$ is an open set). Convergence for the dual norm $\| \|^*_\tau$ follows from Proposition 2.12.

Finally using exchangeability we have

$$H(\alpha^n_{C_n,k} \mid \alpha^{*\otimes k}_n) = H(\alpha^n_{C_n,k} \mid \alpha^{*\otimes k}) + \int \log \frac{\mathrm{d}\alpha^{*\otimes k}}{\mathrm{d}\alpha^{*\otimes k}_n}\, \mathrm{d}\alpha^n_{C_n}$$
$$= H(\alpha^n_{C_n,k} \mid \alpha^{*\otimes k}) + kH(\alpha^* \mid \alpha^*_n) + k \int \log \frac{\mathrm{d}\alpha^*}{\mathrm{d}\alpha^*_n}\, (\mathrm{d}\alpha^n_{C_n} - \mathrm{d}\alpha^*).$$

We already saw in the proof of Theorem 2.7 that $H(\alpha^* \mid \alpha^*_n)$ and $H(\alpha^n_{C_n,k} \mid \alpha^{*\otimes k}_n)$ go to 0. It remains to prove that $\int \log \frac{\mathrm{d}\alpha^*}{\mathrm{d}\alpha^*_n} (\mathrm{d}\alpha^n_{C_n} - \mathrm{d}\alpha^*)$ goes to 0. But $\log \frac{\mathrm{d}\alpha^*}{\mathrm{d}\alpha^*_n} = \langle \lambda^*_n - \lambda^*, F\rangle$ is bounded in $\mathbb{L}_\tau(\alpha)$ for n large enough since λ^*_n goes to λ^*. Hence convergence to 0 of this last term follows from the convergence for the dual norm $\| \|^*_\tau$ we have just shown. □

Remark 3.10. In Theorem 3.9 one can also replace Yurinskii's bound by the classical Bernstein inequality (see e.g. [13]). This only improves the constants (see [20] for the details).

The results of this section are proved mainly thanks to Lemma 3.1 and the very complete literature on sums of independent variables. The situation is of course more intricate in more delicate situations. We shall study some of them in the following sections.

4. General convex constraints

We start with the key minimization bound we shall use. The following result is stated in [14], Exercise 3.3.23, p. 76. A complete proof is contained in [21] (also see [20]).

Proposition 4.1. *Let $A \subseteq M_1(E)$ be such that $\{x, L^x_n \in A\}$ is measurable. If ν is such that $\nu \ll \alpha$ and $\nu^{\otimes n}(L_n \in A) > 0$, then*

$$\frac{1}{n} \log \left(\alpha^{\otimes n}(L_n \in A)\, \mathrm{e}^{nH(\nu|\alpha)}\right) \geq -H(\nu \mid \alpha) \frac{\nu^{\otimes n}(L_n \in A^c)}{\nu^{\otimes n}(L_n \in A)} + \frac{1}{n} \log \nu^{\otimes n}(L_n \in A)$$
$$- \frac{1}{ne\nu^{\otimes n}(L_n \in A^c)}.$$

Corollary 4.2. *If* (2.7.1), (2.7.2) *and* (2.7.3) *are all satisfied,* (2.7.4) *holds assuming that*

$$\lim_{n\to+\infty} \alpha^{*\otimes n}(L_n \in C_n) = 1.$$

The proof is an immediate application of Proposition 4.1 with $A = C_n$ and $\nu = \alpha^*$ since $H(C \mid \alpha) = H(\alpha^* \mid \alpha)$.

In the remainder of the section we shall assume that $G = C_b(E)$. According to Remark 2.8, it is thus enough to check (2.7.1 and 2.7.4) in order to apply Theorem 2.7. In particular if $H(C \mid \alpha)$ is finite, it just remains to check the condition stated in Corollary 4.2, by choosing appropriate enlargements C_n. To this end we first recall basic facts on metrics on probability measures.

Recall that the narrow topology on $M_1(E)$ is metrizable. Among admissible metrics we shall consider two, namely the Prohorov metric d_P and the Fortet–Mourier metric d_{FM}.

Proposition 4.3. *For two probability measures ν_1 and ν_2 on E the previous metrics are defined as follows*

$$d_P(\nu_1,\nu_2) = \inf\left\{a > 0 : \sup_A(\nu_1(A) - \nu_2(A^a)) \le a \text{ where } A^a = \{x : d(x, A) \le a\}\right\},$$

$$d_{FM}(\nu_1,\nu_2) = \sup\left\{\int f(\mathrm{d}\nu_1 - \mathrm{d}\nu_2) \text{ for } f \in B\mathrm{Lip}(E) \text{ such that } \|f\|_{B\mathrm{Lip}} \le 1\right\},$$

where $B\mathrm{Lip}$ is the set of bounded and Lipschitz functions and $\|f\|_{B\mathrm{Lip}} = \|f\|_\infty + \|f\|_{\mathrm{Lip}}$. For both metrics $M_1(E)$ is Polish. If in addition E is compact then so is $M_1(E)$.

Furthermore the following inequalities are known to hold

$$d_{FM}(\nu_1,\nu_2) \le \|\nu_1 - \nu_2\|_{TV} \quad \text{and} \quad d_P(\nu_1,\nu_2) \le \frac{1}{2}\|\nu_1 - \nu_2\|_{TV},$$

and

$$\varphi(d_P(\nu_1,\nu_2)) \le d_{FM}(\nu_1,\nu_2) \le 2d_P(\nu_1,\nu_2),$$

where $\varphi(u) = \frac{2u^2}{2+u}$.

For these inequalities see e.g. [17], problem 5, p. 312, and Corollary 2.6.5, Chapter 11.

In the sequel

$$C_n = C^{\varepsilon_n} = \{\nu : \bar{d}(C,\nu) \le \varepsilon_n\}$$

where $\bar{d}$ is one of the previous metrics.

Definition 4.4. Let (X, d) be a metric space. If $A \subseteq X$ is totally bounded, we denote by $N_X(A, d, \varepsilon)$ the minimal number of (open) balls with radius ε that cover A. The function N_X is often called the metric entropy. In the sequel we simply write $N(d, \varepsilon)$ for the quantity $N_X(X, d, \varepsilon)$, if X is totally bounded.

Our first result is concerned with compact state spaces.

Theorem 4.5. *Assume that E is compact. Let C be a narrowly closed convex subset of $M_1(E)$ such that $H(C \mid \alpha) < +\infty$, and α^* be the I-projection of α on C. Then for any sequence ε_n going to 0 and such that $N(d_{FM}, \varepsilon_n/4)\,\mathrm{e}^{-n\varepsilon_n^2/8} \to 0$ (resp. $N(d_P, \varepsilon_n/4)\,\mathrm{e}^{-n\varepsilon_n^2/2} \to 0$) as $n \to \infty$, $\alpha^n_{C_n,k} \to \alpha^{*\otimes k}$ in total variation distance.*

Proof. Let $B(\alpha^*, \varepsilon)$ the open ball centered at α^* with radius ε. Then

$$\alpha^{*\otimes n}(L_n \in C^\varepsilon) \geq \alpha^{*\otimes n}(L_n \in B(\alpha^*, \varepsilon)) = 1 - \alpha^{*\otimes n}(L_n \in B^c(\alpha^*, \varepsilon))$$

where B^c is as usual the complement subset of B. But we can cover $B^c(\alpha^*, \varepsilon)$ by

$$N_{M_1(E)}(B^c(\alpha^*, \varepsilon), \bar{d}, \eta) \leq N(\bar{d}, \eta)$$

closed balls with radius η so that

$$\alpha^{*\otimes n}(L_n \in B^c(\alpha^*, \varepsilon)) \leq N(\bar{d}, \eta) \max_j \alpha^{*\otimes n}(L_n \in B_j)$$

for such balls B_j. But a closed ball being closed and convex, Lemma 2.11 shows that

$$\alpha^{*\otimes n}(L_n \in B_j) \leq \mathrm{e}^{-nH(B_j|\alpha^*)}.$$

Since $B_j \subseteq (B^c(\alpha^*, \varepsilon))^{2\eta}$ we have $H(B_j \mid \alpha^*) \geq H((B^c(\alpha^*, \varepsilon))^{2\eta} \mid \alpha^*)$ and finally

$$\alpha^{*\otimes n}(L_n \in B^c(\alpha^*, \varepsilon)) \leq N(\bar{d}, \eta)\, \mathrm{e}^{-nH((B^c(\alpha^*,\varepsilon))^{2\eta}|\alpha^*)}.$$

Choosing $\eta = \varepsilon/4$, and hence $(B^c(\alpha^*, \varepsilon))^{2\eta} = B^c(\alpha^*, \varepsilon/2)$, we may apply the results recalled in Proposition 4.3 to get that for all $\nu \in B^c(\alpha^*, \varepsilon/2)$,

$$H(\nu \mid \alpha^*) \geq \frac{1}{2}\|\nu - \alpha^*\|_{TV}^2 \geq \frac{1}{2}\, d_{FM}^2(\nu, \alpha^*) \geq \varepsilon^2/8.$$

We can replace 8 by 2 when replacing the Fortet–Mourier metric by the Prohorov one.

Hence we may apply Corollary 4.2 and Theorem 2.7. □

The condition on ε_n in the previous theorem is interesting if it can be satisfied by at least one such sequence. The following proposition shows that it is always the case, it also relates the metric entropy on $M_1(E)$ to the metric entropy on E.

Proposition 4.6. *Let (E, d) be a compact metric space. Then for all $\varepsilon > 0$,*

(4.6.1) $N(d_P, \varepsilon) \leq \left(\frac{2e}{\varepsilon}\right)^{N(d,\varepsilon)}$,

(4.6.2) $N(d_{FM}, \varepsilon) \leq \left(\frac{4e}{\varepsilon}\right)^{N(d,\varepsilon/2)}$,

(4.6.3) *there exists at least one sequence ε_n going to 0 and such that*

$$\lim_{n\to\infty} \left(\frac{n\varepsilon_n^2}{8} + (\log \varepsilon_n) N(d, \varepsilon_n/8)\right) = +\infty.$$

Such a sequence fulfills the condition in Theorem 4.5 for both metrics on $M_1(E)$ (but is not sharp).

Proof. The first result is due to Kulkarni and Zeitouni ([22], Lemma 1), the second one follows thanks to Proposition 4.3. Consider

$$f:]0, 1] \to \mathbb{R}^+, \qquad \varepsilon \to -\frac{8(\log \varepsilon) N(d, \varepsilon/8)}{\varepsilon^2},$$

which is clearly decreasing with infinite limit at 0. Let u_n be a $]0, 1]$ valued non-increasing sequence; $w_n = f(u_n)$ is then non-decreasing with infinite limit. Introduce for n large enough $k_n = \max\{k \in \mathbb{N}^*, s.t. w_k \leq \sqrt{n}\}$.

Case 1. If for all n large enough, $k_n \le n$, we choose $\varepsilon_n = u_{k_n}$ for all $n \in [k_n, k_{n+p_n}[$ where $p_n = \inf\{p \ge 1, k_{n+p} > k_n\}$. On one hand $n\varepsilon_n^2 \ge k_n u_{k_n}^2$ goes to infinity. On the other hand,

$$n\varepsilon_n^2 + 8(\log \varepsilon_n)N(d, \varepsilon_n/8) = n\varepsilon_n^2\left(1 - \frac{w_{k_n}}{n}\right) \ge n\varepsilon_n^2\left(1 - \frac{1}{\sqrt{n}}\right) \to +\infty.$$

Case 2. If not, there exists some sequence p_j growing to infinity such that $k_{p_j} \ge p_j$, i.e. $w_{p_j} \le \sqrt{p_j}$. Define $\varphi(n)$ as the unique integer number such that $n \in [p_{\varphi(n)}, p_{\varphi(n)+1}[$, and choose $\varepsilon_n = u_{p_{\varphi(n)}}$. Then $n\varepsilon_n^2 \ge p_{\varphi(n)} u_{p_{\varphi(n)}}^2$ goes to infinity and

$$n\varepsilon_n^2 + 8(\log \varepsilon_n)N(d, \varepsilon_n/8) = n\varepsilon_n^2\left(1 - \frac{w_{p_{\varphi(n)}}}{n}\right) \ge n\varepsilon_n^2\left(1 - \frac{1}{\sqrt{p_{\varphi(n)}}}\right) \to +\infty.$$

The final statement is a consequence of the previous ones. The proof is thus completed. □

Example 4.7. If E is a q dimensional compact Riemannian manifold, it is known that $N(d, \varepsilon) \le C_E/\varepsilon^q$ for some constant C_E. In this case we may thus choose $\varepsilon_n = 1/n^a$ for all $0 < a < \frac{1}{q+2}$. The size of enlargement is thus much greater than for F-moment constraints.

When E is no longer compact, but still Polish, it can be approximated by compact subsets with large probability. Here are the results in this direction.

Theorem 4.8. *Let C be a narrowly closed convex subset of $M_1(E)$ such that $H(C \mid \alpha) < +\infty$, and α^* be the I-projection of α on C. Assume that there exists a sequence $(K_n)_n$ of compact subsets of E and a sequence $(\eta_n)_n$ of non-negative real numbers such that*

$$n\eta_n^2 + 8(\log \eta_n)N_E(K_n, d, \eta_n/8) \to +\infty$$

as $n \to \infty$. Let $\varepsilon_n = \eta_n + 2\alpha^*(K_n^c)$. Suppose one of the following additional assumptions applies:

- $\lim_{n\to\infty} (\alpha^*(K_n))^n = 1$,
- $\log \frac{d\alpha^*}{d\alpha}$ *is continuous and bounded, and* $\lim_{n\to\infty} \alpha^*(K_n) = 1$.

Then $\alpha^n_{C_n,k} \to \alpha^{*\otimes k}$ in total variation distance.

Here again the conditions are not sharp, but they hold for both the Prohorov and the Fortet–Mourier metrics.

Proof. The proof relies on the following lemma.

Lemma 4.9. *For all compact subsets K and all $\eta > 0$,*

$$\alpha^{*\otimes n}\left(\bar{d}(L_n, C) \le \eta + 2\alpha^*(K^c)\right) \ge (\alpha^*(K))^n(1 - (16e/\eta)^{N_E(K,d,\eta/8)}e^{-n\eta^2/8}).$$

Proof. Introduce $\alpha_K^* = \frac{\mathbb{1}_K}{\alpha^*(K)}\alpha^*$. Then

$$\bar{d}(\alpha_K^*, \alpha^*) \le \|\alpha_K^* - \alpha^*\|_{TV} = \int \left|\frac{\mathbb{1}_K}{\alpha^*(K)} - 1\right| d\alpha^* \le 2\alpha^*(K^c),$$

so that according to the triangle inequality $\bar{d}(\nu, \alpha^*) \le \bar{d}(\alpha_K^*, \nu) + 2\alpha^*(K^c)$ for all ν. Hence $B(\alpha_K^*, \eta) \subseteq \{\nu, \bar{d}(\nu, C) \le \eta + 2\alpha^*(K^c)\}$ and

$$\begin{aligned}\alpha^{*\otimes n}\left(\bar{d}(L_n, C) \le \eta + 2\alpha^*(K^c)\right) &\ge \alpha^{*\otimes n}(L_n \in B(\alpha_K^*, \eta))\\ &\ge \alpha^{*\otimes n}(L_n \in B(\alpha_K^*, \eta) \text{ and } x \in K^n)\\ &\ge (\alpha^*(K))^n \alpha_K^{*\otimes n}(L_n \in B(\alpha_K^*, \eta)).\end{aligned}$$

As in the proof of Theorem 4.5 and using (4.6.1 or 4.6.2) we have

$$\alpha_K^{*\otimes n}(L_n \in B(\alpha_K^*, \eta)) \geq 1 - N_{M_1(K)}(\bar{d}, \eta/4)\mathrm{e}^{-n\eta^2/8} \geq 1 - (16e/\eta)^{N_K(d,\eta/8)}\mathrm{e}^{-n\eta^2/8}. \quad \square$$

The first part of the theorem is then immediate.

The second part is a little bit more tricky. Let $h = \log \frac{\mathrm{d}\alpha^*}{\mathrm{d}\alpha}$. For all $\varepsilon > 0$

$$\begin{aligned}\alpha^{\otimes n}(L_n \in C^\varepsilon) &\geq \alpha^{\otimes n}(L_n \in B(\alpha^*, \varepsilon)) = \int \mathbb{1}_{B(\alpha^*,\varepsilon)}(L_n)\mathrm{e}^{-n\langle L_n, h\rangle}\,\mathrm{d}\alpha^{*\otimes n} \\ &\geq \mathrm{e}^{-nH(C|\alpha)} \int \mathbb{1}_{B(\alpha^*,\varepsilon)}(L_n)\mathrm{e}^{-n\langle L_n - \alpha^*, h\rangle}\,\mathrm{d}\alpha^{*\otimes n} \\ &\geq \mathrm{e}^{-nH(C|\alpha)}\mathrm{e}^{-n\Delta(\varepsilon)}\alpha^{*\otimes n}(L_n \in B(\alpha^*, \varepsilon))\end{aligned}$$

where $\Delta(\varepsilon) = \sup_{\nu \in B(\alpha^*,\varepsilon)}\langle \nu - \alpha^*, h\rangle$. Since h is continuous and bounded, it is immediate that $\Delta(\varepsilon)$ goes to 0 as ε goes to 0. Hence if ε_n goes to 0

$$\liminf_{n\to\infty} \log\left(\alpha^{\otimes n}(L_n \in C_n)\mathrm{e}^{nH(C|\alpha)}\right) \geq \liminf_{n\to\infty} \log\left(\alpha^{*\otimes n}(L_n \in B(\alpha^*, \varepsilon_n))\right).$$

Thus if we choose ε_n as in the statement of the theorem, the right hand side of the previous inequality is greater than

$$\liminf\left(\log \alpha^*(K_n) + \frac{1}{n}\log\left(1 - (16e/\eta_n)^{N_E(K_n,d,\eta_n/8)}\mathrm{e}^{-n\eta_n^2/8}\right)\right) = 0$$

and we may apply Theorem 2.7. $\square$

In the next section we shall study some typical examples.

5. Examples

In Section 3 we already discussed the examples of F-moments. In this section we shall first look at the finite dimensional situation, then study examples in relation with stochastic mechanics.

5.1. Finite dimensional convex constraints

Proposition 5.1. *If $E = \mathbb{R}^q$, let C be a narrowly closed convex subset of $M_1(E)$ such that $H(C \mid \alpha) < +\infty$, and α^* be the I-projection of α on C. Then $\alpha^n_{C_n,k} \to \alpha^{*\otimes k}$ in total variation distance with $\varepsilon_n = 2/n^b$ and $0 < b < \frac{1-\frac{q}{a}}{2+q}$ provided there exists $a > q$ such that $\int \|x\|^a\,\mathrm{d}\alpha^* < +\infty$ (that holds in particular if $\int \mathrm{e}^{\lambda\|x\|^a}\mathrm{d}\alpha < +\infty$ for some $\lambda > 0$).*

In addition if either $\int \mathrm{e}^{\lambda\|x\|^a}\mathrm{d}\alpha^* < +\infty$ for some $\lambda > 0$, or $\log \frac{\mathrm{d}\alpha^*}{\mathrm{d}\alpha}$ is bounded and continuous, we may choose $b < \frac{1}{2+q}$.

Of course in general hypotheses on α^* are difficult to check directly. That is why the α exponential integrability is a pleasant sufficient condition.

Proof. Let $M = \int \|x\|^a\,\mathrm{d}\alpha^*$. For $K_n = B(0, n^u)$ we have

$$(\alpha^*(K_n))^n \geq \left(1 - \frac{M}{n^{au}}\right)^n \to 1,$$

provided $au > 1$. In addition, one can find M' such that

$$N_E(K_n, d, \eta/48) \leq M' n^{uq}/\eta^q$$

so that if $\eta_n = 1/n^b$ with $b > 0$

$$n\eta_n^2 + 8(\log \eta_n) N_E(K_n, d, \eta_n/8) \geq n^{1-2b}\left(1 - 8bM'(\log n) n^{uq+b(2+q)-1}\right)$$

goes to $+\infty$ as soon as $b < \frac{1-uq}{2+q}$, i.e. if $b < \frac{1-\frac{q}{a}}{2+q}$ since $au > 1$. We may thus apply Theorem 4.8 with $\varepsilon_n = (1/n^b) + 2(M/n^{ua}) \leq 2(1/n^b)$ for n large enough.

If the α^* exponential integrability condition is satisfied we may choose a as large as we want. If $\log \frac{\mathrm{d}\alpha^*}{\mathrm{d}\alpha}$ is bounded, $\alpha^*(K_n)$ growing to 1, the condition $ua > 1$ is not necessary. $\square$

5.2. *Schrödinger bridges*

In this subsection and the next one $E = C^0([0, 1], M)$ where M is either $\mathbb{R}^q$ or a smooth connected and compact Riemannian manifold of dimension q. E is equipped with the sup-norm and for simplicity with the Wiener measure $\mathcal{W}$ (i.e. the infinitesimal generator is the Laplace Beltrami operator), with initial measure μ_0.

An old question posed by Schrödinger can be described as follows (see [19] for the original sentence in French). Let $(X_j)_{j=1,\ldots,n}$ be an n-sample of $\mathcal{W}$. Assume that the empirical measure at time 1 (i.e. $L_n(1) = \frac{1}{n}\sum_{j=1}^n \delta_{X_j(1)}$) is far from the expected law μ_1 of the Brownian motion at time 1. What is the most likely way to observe such a deviation? Clearly the answer (when the number of Brownian particles grows to infinity) is furnished by the Gibbs conditional principle: the most likely way is to imagine that any block of k particles is made of (almost) independent particles with common law $\mathcal{W}^*$ which minimizes $H(\mathcal{V} \mid \mathcal{W})$ among all probability measures on E such that $\mathcal{W} \circ X^{-1}(0) = \mu_0$ and $\mathcal{W} \circ X^{-1}(1)$ belongs to the observed set of measures. For where the observed set is reduced to a single measure (thin) a double limit formulation of this principle is contained in the first chapter of [1].

To be precise introduce for $\varepsilon \geq 0$

$$C^\varepsilon(\nu_0, \nu_1) = \{\mathcal{V} \in M_1(E)\, s.t.\, \bar{d}(\mathcal{V}_0, \nu_0) \leq \varepsilon, \bar{d}(\mathcal{V}_1, \nu_1) \leq \varepsilon\}, \tag{5.2}$$

where $\mathcal{V}_t$ denotes the law $\mathcal{V} \circ X^{-1}(t)$. When $\varepsilon = 0$ we will not write the superscript 0. We are in the situation studied in the previous section since $C(\nu_0, \nu_1)$ is a narrowly closed convex subset of $M_1(E)$. We shall write as $\mathcal{W}^*$ the I-projection of $\mathcal{W}$ on C (without specifying unless necessary the initial and final measures) when it exists.

Before applying the results in Section 4 we shall recall some known results about C and $\mathcal{W}^*$.

Denote by $\mathcal{V}_{u,v}$ (resp. $\mathcal{W}_{u,v}$) the conditional law of $\mathcal{V}$ knowing that $X(0) = u$ and $X(1) = v$, i.e. the law of the $\mathcal{V}$ bridge from u to v. Also denote by $\nu_{0,1}$ (resp. $\mu_{0,1}$) the $\mathcal{V}$ (resp. $\mathcal{W}$) joint law of $X(0), X(1)$. The decomposition of entropy formula

$$H(\mathcal{V} \mid \mathcal{W}) = H(\nu_{0,1} \mid \mu_{0,1}) + \int H(\mathcal{V}_{u,v} \mid \mathcal{W}_{u,v})\, \mathrm{d}\nu_{0,1}(u, v),$$

immediately shows that, if it exists,

$$\mathcal{W}^* = \int \mathcal{W}_{u,v}\, \mathrm{d}\mu^*_{0,1}(u, v),$$

where $\mu^*_{0,1}$ is the I-projection of $\mu_{0,1}$ on

$$\Pi(\nu_0, \nu_1) = \{\beta \in M_1(M \times M)\, s.t.\, \beta_0 = \nu_0, \beta_1 = \nu_1\},$$

if it exists. In other words the problem reduces to a finite dimensional one, i.e. on $M \times M$. The following theorem collects some results we need.

Theorem 5.3. *Assume that $H(\nu_0 \mid \mu_0)$ and $H(\nu_1 \mid \mu_1)$ are both finite and that $p = \log \frac{\mathrm{d}\mu_{0,1}}{\mathrm{d}(\mu_0 \otimes \mu_1)} \in \mathbb{L}^1(\nu_0 \otimes \nu_1)$. Then $H(\Pi(\nu_0, \nu_1) \mid \mu_{0,1})$ is finite.*

In addition $\frac{\mathrm{d}\mu^*_{0,1}}{\mathrm{d}\mu_{0,1}}(u, v) = f(u)g(v)$ for any pair of functions (f, g) satisfying

$$\begin{cases} \dfrac{\mathrm{d}\nu_0}{\mathrm{d}\mu_0}(u) = f(u) \displaystyle\int p(u, v) g(v) \mathrm{d}\mu_1(v) \\ \dfrac{\mathrm{d}\nu_1}{\mathrm{d}\mu_1}(v) = g(v) \displaystyle\int p(u, v) f(u) \mathrm{d}\mu_0(u). \end{cases} \tag{5.4}$$

The proof is contained in [5], Proposition 6.3, and [19], pp. 161–164.

Finally under the assumptions of Theorem 5.3

$$\frac{\mathrm{d}\mathcal{W}^*}{\mathrm{d}\mathcal{W}} = f(X(0))g(X(1)).$$

We can now state

Theorem 5.5. *Under the assumptions of Theorem 5.3,*

$$\mathcal{W}^n_{\varepsilon_n,k} := \mathcal{L}(X_1, \ldots, X_k / L_n \in C^{\varepsilon_n}(\nu_0, \nu_1)) \to \mathcal{W}^{*\otimes k}$$

in total variation distance for all sequences ε_n going to 0 such that the following holds: for all sequences $(Y_j)_j$ (resp. $(Z_j)_j$) of i.i.d. random variables with law ν_0 (resp. ν_1),

$$\lim_{n\to\infty} \mathbb{P}(\bar{d}(L^Y_n, \nu_0) \le \varepsilon_n) = 1 \quad \textit{and} \quad \lim_{n\to\infty} \mathbb{P}(\bar{d}(L^Z_n, \nu_1) \le \varepsilon_n) = 1.$$

In particular the above convergence holds for instance in the following two cases

- *M is compact and $n\varepsilon_n^2 + 8(\log \varepsilon_n) N_M(d, \varepsilon_n/8) \to +\infty$,*
- *$M = \mathbb{R}^q$, there exists $a > q$ such that for $i = 0, 1$, $\int \|x\|^a \mathrm{d}\nu_i < +\infty$, $\varepsilon_n = 2/n^b$ and $b < \frac{1-\frac{q}{a}}{2+q}$.*

For the proof just apply Corollary 4.2, and for the examples Propositions 4.6 and 5.1.

5.3. Nelson processes

A natural generalization of the framework of Section 5.2 is to impose the full flow of marginal laws instead of only the initial and final ones. Building diffusion processes with a given flow of marginal laws is the first step in Nelson's approach to the Schrödinger equation. The problem was first tackled by Carlen [4]. The relationship with minimization of entropy was first observed by Föllmer [19] and explored in detail in a series of papers by C. Léonard and the first named author [6–8]. This approach and the results below can be viewed as some "statistical mechanics" approach of quantum mechanics. We shall not discuss further the meaning of the previous sentence here. We prefer to insist on the enormous difference between a pair and the flow of all marginal laws.

Hence here

$$C(\nu_t) = \{\mathcal{V} \in M_1(E)\, s.t. \forall t \in [0,1], \mathcal{V}_t = \nu_t\},$$

and for $\varepsilon > 0$

$$C^\varepsilon(\nu_t) = \{\mathcal{V} \in M_1(E)\, s.t. \bar{d}(\mathcal{V}, C(\nu_t)) \leq \varepsilon\}.$$

For simplicity we shall only consider the case $M = \mathbb{R}^q$ (though a similar discussion is possible for a general connected and compact Riemannian manifold). Not to lose sight of our main goal we first state the convergence result we have in mind, and will discuss the hypotheses later on.

Theorem 5.6. *Assume that $C(\nu_t)$ is non-empty and that $\mathcal{W}$ has an I-projection $\mathcal{W}^*$ on $C(\nu_t)$, such that $\log \frac{\mathrm{d}\mathcal{W}^*}{\mathrm{d}\mathcal{W}}$ is bounded and continuous. Assume in addition that the initial law μ_0 has a polynomial concentration rate, i.e. $\mu_0(B(0,R)) \leq C/R^m$ for some $m > 0$ and all $R > 0$. Then if $\varepsilon_n = 1/(\log n)^r$ for some $r < 1/2q$,*

$$\mathcal{W}^n_{\varepsilon_n,k} := \mathcal{L}(X_1, \ldots, X_k / L_n \in C^{\varepsilon_n}(\nu_t)) \to \mathcal{W}^{*\otimes k}$$

in total variation distance.

Proof. According to Theorem 4.8 it is enough to find a sequence K_n of compact subspaces of E and a sequence η_n of positive numbers going to 0 such that

$$\lim_{n\to\infty} \mathcal{W}^*(K_n) = 1 \quad \text{and} \quad \lim_{n\to\infty} \left(n\eta_n^2 + 8(\log \eta_n) N_E(K_n, \|\|_\infty, \eta_n/8)\right) = +\infty.$$

Since $\frac{\mathrm{d}\mathcal{W}^*}{\mathrm{d}\mathcal{W}}$ is bounded by some e^D, we may replace the first condition by $\lim_{n\to\infty} \mathcal{W}(K_n) = 1$ and choose $\varepsilon_n \geq \eta_n + 2\mathrm{e}^D \mathcal{W}(K_n^c)$. The most natural way to choose such compact sets is to use the Kolmogorov regularity criterion. Since the support of $\mathcal{W}$ is included in the set of Hölder paths of order $\beta < 1/2$ introduce

$$K(R, M, \beta) = \left\{ w \in E\, s.t. |w(0)| \leq R \text{ and } \sup_{s \neq t \in [0,1]} \frac{\|w(s) - w(t)\|}{|s-t|^\beta} \leq M \right\},$$

for R, M positive and $\beta < 1/2$. Kolmogorov's criterion tells us that

$$\mathcal{W}(K^c(R, M, \beta)) \leq \mu_0(B(0,R)) + C(p, \beta) M^{-p}$$

for all $p > 1$. In addition, thanks to Theorem 2.7.1, p. 155, in [13]

$$N_E(K(R, M, \beta), \|\|_\infty, \eta/8) \leq c_1(\beta, q)\, (8R/\eta)^q \mathrm{e}^{c_2(\beta,q)(M/\eta)^{q/\beta}}.$$

Choosing $K_n = k(R_n, M_n, \eta_n)$ with

$$R_n = (a \log n)^{\beta/qm} \qquad M_n = (b \log n)^{\beta/q} \qquad \eta_n = (c \log n)^{-\beta/q}$$

we see that $n\eta_n^2 + 8(\log \eta_n)\, N_E(K_n, \|\|_\infty, \eta_n/8)$ is less than

$$n(\log n)^{-2\beta/q} \left(A_1 + A_2 \log(c \log n)\, (\log n)^{q + \frac{2\beta}{q}} n^{c_2(\beta,q)bc - 1} \right)$$

for some A_1 and A_2 independent of n. Choosing b in such a way that $c_2(\beta, q)bc - 1 < 0$ we obtain a leading term going to $+\infty$ as n goes to ∞.

Putting all this together, we get

$$\eta_n + 2\mathrm{e}^D \mathcal{W}(K_n^c) \le (c\log n)^{-\beta/q} + 2C\mathrm{e}^D (a\log n)^{-\beta/q} + 2\mathrm{e}^D C(p,\beta)(b\log n)^{-\beta p/q}$$

which is less than $(\log n)^{-\beta'/q}$ for all $\beta' < 1/2$ and n large enough. $\square$

Remark 5.7. The assumption $\log \frac{\mathrm{d}\mathcal{W}^*}{\mathrm{d}\mathcal{W}}$ bounded and continuous is essential. Indeed without it Theorem 4.8 requires $(\mathcal{W}^*(K_n))^n$ goes to 1, i.e. $\mathcal{W}^*(K_n^c) = o(1/n)$. Assuming that $\frac{\mathrm{d}\mathcal{W}^*}{\mathrm{d}\mathcal{W}}$ belongs to $\mathbb{L}_r(\mathcal{W})$, the Kolmogorov criterion yields M_n of order n^a. It is then easy to see that this is no longer compatible with any choice of η_n such that $\lim_{n\to\infty}\left(n\eta_n^2 + 8(\log\eta_n)N_E(K_n, \|\|_\infty, \eta_n/8)\right) = +\infty$.

To conclude this subsection let us say a few words about our assumptions.

First of all $C(\nu_t)$ is non-empty as soon as ν_t satisfies a Fokker–Planck equation with a drift $B(t, X(t))$ of finite energy (i.e. $\int_0^1\int B^2(t,x)\mathrm{d}\nu_t\mathrm{d}t < +\infty$); see [4,6,7]. In addition Girsanov theory is still available (see [6,7] for the details) so that

$$\frac{\mathrm{d}\mathcal{W}^*}{\mathrm{d}\mathcal{W}} = \frac{\mathrm{d}\nu_0}{\mathrm{d}\mu_0}\exp\left(\int_0^T B(t,w(t))\mathrm{d}w(t) - 1/2\int_0^T |B(t,w(t))|^2\mathrm{d}t\right)$$

where $T = \inf\{s \le 1\ s.t. \int_0^s |B(t,w(t))|^2\mathrm{d}t = +\infty\}$. In general this density (even when $T = 1$) is not continuous.

Nevertheless some interesting cases enter the framework of Theorem 5.6.

Let U be a C_b^2 potential. Then the law $\mathcal{V}_0$ of the unique strong solution of

$$\mathrm{d}X_t = \mathrm{d}W_t - \nabla U(X_t)\mathrm{d}t, \qquad \mathcal{L}(X_0) = \nu_0$$

satisfies

$$\frac{\mathrm{d}\mathcal{V}_0}{\mathrm{d}\mathcal{W}} = \frac{\mathrm{d}\nu_0}{\mathrm{d}\mu_0}\exp\left(U(w(0)) - U(w(1)) - 1/2\int_0^1 (|\nabla U|^2 - \Delta U)(t,w(t))\mathrm{d}t\right).$$

Hence $\log\frac{\mathrm{d}\mathcal{V}_0}{\mathrm{d}\mathcal{W}}$ is bounded and continuous as soon as $\log\frac{\mathrm{d}\nu_0}{\mathrm{d}\mu_0}$ is. In addition $\mathcal{V}_0$ is the I-projection of $\mathcal{W}$ on $C(\nu_t)$ where $\nu_t = \mathcal{L}(X_t)$ (see [6]). The conclusion of Theorem 5.6 is thus available for $\mathcal{V}_0$. If we replace $\mathbb{R}^q$ by a compact manifold we may include the stationary (actually reversible) case, i.e. $\nu_0 = \mathrm{e}^{-2U}\mathrm{d}x/Z_U$.

6. A super-thin case: Volatility calibration

In Sections 5.2 and 5.3 we have studied the laws of some diffusion processes from the point of view of I-projections; hence we only allowed a change of drift. We shall now study the opposite situation: the drift being fixed, how do we choose the diffusion coefficient? We thus immediately lose any kind of absolute continuity, introducing a new difficulty that is super-thin subsets. Let us describe precisely the problem.

Consider a family (indexed by continuous time–space functions σ) of S.D.E.

$$\forall t \in [0,1], \quad \mathrm{d}X(t) = \sigma(t, X(t))\mathrm{d}w(t) + b_0(t, X(t))\,\mathrm{d}t; \quad X(0) = 0, \tag{6.1}$$

where w is a standard Brownian motion. We assume that b_0 is continuous and bounded and

$$0 < \sigma_{\min} \le \sigma \le \sigma_{\max} < +\infty$$

for some real numbers $\sigma_{\min}$ and $\sigma_{\max}$. Under this assumption, it is well known that (6.1) admits weak solutions and that there is uniqueness in law. We will denote in the sequel by $\mathbb{Q}_{\sigma,b_0}$ the probability measure on $\Omega = C([0,1],\mathbb{R})$ thus defined by (6.1).

In [2] the authors addressed the problem of calibrating σ (volatility in mathematical finance) when b_0 is known (a consequence of the "absence of arbitrage") and X satisfies a set of generalized moment constraints

$$\mathbb{E}[f_j(t_j, X(t_j))] = c_j, \quad j \in \Lambda,\ \Lambda \text{ finite}. \tag{6.2}$$

Their strategy is based on the following Bayesian principle: take a prior σ_0; the corresponding prior law of X is $\mathbb{Q}_{\sigma_0,b_0}$. Then the "most probable" $\mathbb{P}$ satisfying (6.2) will be the one which minimizes the relative entropy $H(\mathbb{P} \mid \mathbb{Q}_{\sigma_0,b_0})$. Of course this principle is meaningless here. Indeed, the finiteness of $H(\mathbb{P} \mid \mathbb{Q}_{\sigma_0,b_0})$ implies that $\mathbb{P}$ has the same diffusion coefficient as $\mathbb{Q}_{\sigma_0,b_0}$, and hence there is no such $\mathbb{P}$ satisfying (6.2) unless $\mathbb{Q}_{\sigma_0,b_0}$ does. To bypass this difficulty, the authors propose to approximate $\mathbb{Q}_{\sigma_0,b_0}$ by some well chosen $\mathbb{Q}^{\varepsilon}_{\sigma_0,b_0}$ (actually various time discretizations), in such a way that $\varepsilon H(\mathbb{P}^{\varepsilon} \mid \mathbb{Q}^{\varepsilon}_{\sigma_0,b_0})$ goes to some limit $K(\mathbb{P} \mid \mathbb{Q}_{\sigma_0,b_0})$, and then use K as the cost function to be minimized.

We shall interpret this strategy in the following way.

For simplicity assume that the set of constraints is reduced to a single one, i.e. introduce the set

$$C_F = \{\mathbb{P}, \mathbb{E}_{\mathbb{P}}[F(X(1))] = 1\}$$

where $\mathbb{P}$ describes the set of probability measures on $\Omega = C([0,1],\mathbb{R})$. We will choose as before some ε enlargement of C_F, i.e. define

$$C_F^{\varepsilon} = \left\{\mathbb{P}, \left|\int F(X(1))\,\mathrm{d}\mathbb{P} - 1\right| < \varepsilon\right\}.$$

Again for simplicity, we shall assume that $b(t,x) = b_0$ for some $b_0 > 0$ (extensions to more general cases can be easily made). We also define

$$\Sigma_0 = \{\sigma : [0,1]\times\mathbb{R} \to]\sigma_{\min}, \sigma_{\max}[,\ \text{continuous}\}$$

and for $\varepsilon < b_0$,

$$\mathcal{B}_{\varepsilon} = \{b : [0,1]\times\mathbb{R} \to]b_0 - \varepsilon, b_0 + \varepsilon[,\ \text{continuous}\}.$$

Let us make precise that the space of space–time continuous functions $C([0,1]\times\mathbb{R},\mathbb{R})$ will always be furnished with the topology of uniform convergence on every compact subset of $[0,1]\times\mathbb{R}$. Now we introduce a standard approximation of $\mathbb{Q}_{\sigma,\,b_0}$, namely the trinomial tree.

Choose some $\alpha > \sigma_{\max}$ and $0 < s < b_0$. For $(y,z) \in \mathbb{R}^2$ we define

$$\begin{cases} m^n(y,z) = \dfrac{y^2}{2\alpha^2} + \dfrac{z}{2\alpha\sqrt{n}} \\ d^n(y,z) = \dfrac{y^2}{2\alpha^2} - \dfrac{z}{2\alpha\sqrt{n}} \\ r^n(y,z) = 1 - \dfrac{y^2}{\alpha^2}. \end{cases}$$

For n large enough ($> n_0$), it is easily seen that for all $(y,z) \in [\sigma_{\min}, \sigma_{\max}] \times [b_0 - s, b_0 + s]$ the vector (m^n, d^n, r^n) has all its entries strictly positive (their sum being 1), so that we may define the following transition kernel defined on $\mathbb{R}$ for all $(\sigma, b) \in \overline{\Sigma_0} \times \overline{\mathcal{B}_s}$, $n \geq n_0$ and $(t, x) \in [0, 1] \times \mathbb{R}$,

$$\Pi^n_{\sigma,b}(t, x, .) = m^n(\sigma, b)(t, x).\delta_{x+\frac{\alpha}{\sqrt{n}}} + r^n(\sigma, b)(t, x).\delta_x + d^n(\sigma, b)(t, x).\delta_{x-\frac{\alpha}{\sqrt{n}}}.$$

We thus define the probability measure $\mathbb{Q}^n_{\sigma,b}$

$$\begin{cases} (1)\ \mathbb{Q}^n_{\sigma,b}(X_0 = 0) = 1, \\ (2)\ \mathbb{Q}^n_{\sigma,b}\left(X_t = X_{\frac{k}{n}} + (nt - k)\left[X_{\frac{k+1}{n}} - X_{\frac{k}{n}}\right], \frac{k}{n} \leq t \leq \frac{k+1}{n}\right) = 1, \\ (3)\ \mathbb{Q}^n_{\sigma,b}\left(X_{\frac{k+1}{n}} \in . \,\Big|\, X_{\frac{k}{n}}, \ldots, X_0\right) = \Pi^n_{\sigma,b}\left(\frac{k}{n}, X_{\frac{k}{n}}, .\right). \end{cases} \tag{6.3}$$

In the sequel, we will denote by $\mathbb{E}^n_{\sigma,b}[.]$ the expectation with respect to the trinomial tree $\mathbb{Q}^n_{\sigma,b}$. The support of $\mathbb{Q}^n_{\sigma,b}$ is $\Omega_n \subset \Omega$ defined by

$$\Omega_n = \left\{ \omega \in \Omega : \begin{bmatrix} \omega(0) = 0 \\ \omega\left(\frac{i+1}{n}\right) - \omega\left(\frac{i}{n}\right) \in \left\{-\frac{\alpha}{\sqrt{n}}, 0, \frac{\alpha}{\sqrt{n}}\right\}, \text{ for } i = 0, \ldots, n-1 \\ \omega \text{ affine on } \left[\frac{i}{n}, \frac{i+1}{n}\right], \text{ for } i = 0, \ldots, n-1 \end{bmatrix} \right\}.$$

The set Ω_n is finite with cardinality 3^n.

Finally denoting by $L_m = \frac{1}{m}\sum_{i=1}^m \delta_{\omega_i}$ the empirical measure on Ω, we shall study $\mathbb{R}^{n,m}_\varepsilon$ defined by

$$\mathbb{R}^{n,m}_\varepsilon(B) = (\mathbb{Q}^n_{\sigma_0,b_0})^{\otimes m}(\omega_1 \in B / L_m \in \widetilde{\mathbb{T}}^n_\varepsilon \cap C^\varepsilon_F),$$

where $\widetilde{\mathbb{T}}^n_\varepsilon$ will be defined later. Let us just say for the moment that $\widetilde{\mathbb{T}}^n_\varepsilon$ is an open set of $M_1(\Omega_n)$ which contains all the trinomial trees $\mathbb{Q}^n_{\sigma,b}$ with σ in a totally bounded subset Σ_1 of Σ_0 and $b \in \mathcal{B}_\varepsilon$. Roughly speaking, for each level of approximation (n) we consider a sample of size m of the trinomial tree and look at the conditional law of the first coordinate, knowing that the empirical measure is not too far from being a trinomial tree satisfying the moment constraint.

Our aim is to show that one can find sequences ε_n going to 0 and m_n going to infinity, such that $\mathbb{R}^{n,m_n}_{\varepsilon_n}$ goes towards some $\mathbb{Q}_{\sigma^*,b_0}$, the one proposed in [2], which we will now describe.

First, for fixed n and ε, since all measures are defined on a finite set, it is not difficult to see that the set $\mathcal{M}^n_\varepsilon$ of minimizers of $H(. \mid Q^n_{\sigma_0,b_0})$ on $\widetilde{\mathbb{T}}^n_\varepsilon \cap C^\varepsilon_F$ is non-empty. It can then be shown that the elements of $\mathcal{M}^n_\varepsilon$ are still trinomial trees. Now an easy computation shows that $\sigma \mapsto \frac{1}{n}H(\mathbb{Q}^n_{\sigma,b} \mid \mathbb{Q}^n_{\sigma_0,b_0})$ is converging (in a sense close to the Γ-convergence sense; see Remark 6.10) to

$$\sigma \mapsto I(\sigma \mid \sigma_0) = \mathbb{E}_\sigma\left[\int_0^1 q(\sigma^2(X_t, t), \sigma_0^2(t, X_t))\,\mathrm{d}t\right],$$

with

$$q(x, y) = \log\left(\frac{x}{y}\right)\frac{x}{\alpha^2} + \log\left(\frac{\alpha^2 - x}{\alpha^2 - y}\right)\left[1 - \frac{x}{\alpha^2}\right].$$

One thus expects that the limit $\mathbb{Q}_{\sigma^*,b_0}$ is the one obtained by minimizing I on $\overline{\Sigma_0}$ under the moment constraint.

The remainder of this section will be devoted to giving rigorous statements and proofs. Note that the result gives a rigorous statistical flavor to the method proposed by Avellaneda et al.

6.1. Presentation of the results

We recall that the space $C([0,1]\times\mathbb{R},\mathbb{R})$ is equipped with the topology of uniform convergence on every compact subset of $[0,1]\times\mathbb{R}$. Before presenting our results, let us state the basic convergence property of trinomial trees:

Proposition 6.4. *If $s \geq \varepsilon_n \geq 0$ goes to zero and $\sigma_n \in \overline{\Sigma_0}$ goes to $\sigma \in \overline{\Sigma_0}$ then, for all $b_n \in \overline{\mathcal{B}_{\varepsilon_n}}$, the sequence $\mathbb{Q}^n_{\sigma_n,b_n}$ goes to $\mathbb{Q}_{\sigma,b_0}$.*

From now on, we will make the following assumptions:

- The minimum value of the function $I(.\mid\sigma_0)$ on the set $\left\{\sigma\in\overline{\Sigma_0} : \int F(X_1)\,\mathrm{d}Q_{\sigma,b}=1\right\}$ is attained at a unique point σ^*.
- The minimizer σ^* belongs to Σ_0.

Now let us introduce some notation. For all $\sigma\in\overline{\Sigma_0}$, let $\Delta_{n,\sigma}$ be the continuity modulus of σ on the compact set $[0,1]\times\left[-\alpha\sqrt{n},\alpha\sqrt{n}\right]$, i.e.

$$\Delta_{n,\sigma}(\varepsilon) = \sup\left\{|\sigma(t,x)-\sigma(s,y)| : s,t\in[0,1], x,y\in\left[-\alpha\sqrt{n},\alpha\sqrt{n}\right],\right.$$
$$\left.|t-s|+|x-y|\leq\varepsilon\right\}.$$

Let Σ_1 be defined by

$$\Sigma_1 = \{\sigma\in\Sigma_0 : \forall n\in\mathbb{N}^*, \Delta_{n,\sigma} < 2\Delta_{n,\sigma^*}\}.$$

According to the Ascoli theorem, Σ_1 is easily seen to be totally bounded.

Now let us consider the set $\widetilde{\mathbb{T}}^n_\varepsilon$ of all probability measures $\mathbb{Q}$ on Ω satisfying

$$\begin{cases} (1)\ \mathbb{Q}(X_0=0)=1, \\ (2)\ \mathbb{Q}\left(X_t = X_{\frac{k}{n}} + (nt-k)\left[X_{\frac{k+1}{n}} - X_{\frac{k}{n}}\right], \dfrac{k}{n}\leq t\leq\dfrac{k+1}{n}\right) = 1, \\ (3)\ \exists(\sigma,b)\in\Sigma_1\times\mathcal{B}_\varepsilon \quad \text{such that } \mathbb{Q}\left(X_{\frac{p+1}{n}}\in .\,|X_{\frac{p}{n}}\right) = \Pi^n_{\sigma,b}\left(\dfrac{p}{n}, X_{\frac{p}{n}}, .\right). \end{cases} \tag{6.5}$$

In the sequel we will set $A^n_\varepsilon := \widetilde{\mathbb{T}}^n_\varepsilon \cap C^\varepsilon_F$. Defining (when possible), for all positive integer m,

$$\mathbb{R}^n_{\varepsilon,m} = \mathbb{E}_{(\mathbb{Q}^n_{\sigma_0,b_0})^{\otimes m}}\left[L_m | L_m\in A^n_\varepsilon\right],$$

our main result is the following:

Theorem 6.6. *If $\varepsilon^0_n = \min(|\mathbb{E}^n_{\sigma^*,b_0}[F(X_1)]-1| + 1/n, s)$, then there exists a sequence m_n of positive integers going to $+\infty$, such that $\mathbb{R}^n_{\varepsilon^0_n,m_n}$ converges to $\mathbb{Q}_{\sigma^*,b_0}$.*

In order to prove this theorem, the first step is to study the convergence of $\mathbb{R}^n_{\varepsilon^0_n,m}$ when n is fixed and m goes to $+\infty$. This is done in the two following propositions:

Proposition 6.7. *Recall that d_{FM} denotes the Fortet–Mourier distance, and for all $\varepsilon>0$ let $\mathcal{M}^n_\varepsilon$ be the set of minimizers of $H(.\mid Q^n_{\sigma_0,b_0})$ on $\overline{A^n_\varepsilon}$. Then,*

$$d_{FM}(\mathbb{R}^n_{\varepsilon^0_n,m}, \overline{co}\mathcal{M}^n_{\varepsilon^0_n}) \xrightarrow[m\to+\infty]{} 0,$$

where $\overline{co}\mathcal{M}^n_{\varepsilon^0_n}$ denotes the closed convex hull of $\mathcal{M}^n_{\varepsilon^0_n}$.

Proof. The set $A^n_{\varepsilon^0_n}$ is non-empty (it contains $\mathbb{Q}^n_{\sigma^*,b_0}$) and, according to the proposition below, it is open and satisfies $H(A^n_{\varepsilon^0_n} \mid Q^n_{\sigma_0,b_0}) = H(\overline{A^n_{\varepsilon^0_n}} \mid Q^n_{\sigma_0,b_0})$. The result follows immediately from the classical Gibbs conditioning principle. □

Proposition 6.8. (1) *The set A^n_ε is an open subset of $M_1(\Omega_n)$, and satisfies $H(A^n_\varepsilon \mid Q^n_{\sigma_0,b_0}) = H(\overline{A^n_\varepsilon} \mid Q^n_{\sigma_0,b_0})$.*

(2) *Every element of $\mathcal{M}^n_\varepsilon$ is of the form $\mathbb{Q}^n_{\sigma,b}$ for some $(\sigma, b) \in \overline{Z_1} \times \overline{\mathcal{B}_\varepsilon}$.*

According to Proposition 6.7, we know that for large m, $\mathbb{R}^n_{\varepsilon^0_n,m}$ is close to $\overline{co}\mathcal{M}^n_{\varepsilon^0_n}$. The next step consists in proving that this set is close to $\{\mathbb{Q}_{\sigma^*,b_0}\}$. This will follow from the particular type of convergence of the normalized entropy functions:

Proposition 6.9. (1) *If $0 < \varepsilon_n$ goes to 0, then for every sequence $b_n \in \mathcal{B}_{\varepsilon_n}$, and for every $\sigma \in \overline{\Sigma_0}$, the following holds:*

$$\frac{H(\mathbb{Q}^n_{\sigma,b_n} \mid \mathbb{Q}^n_{\sigma_0,b_0})}{n} \xrightarrow[n\to+\infty]{} I(\sigma \mid \sigma_0).$$

(2) *Furthermore, if $\sigma_n \in \overline{\Sigma_0}$ converges to $\sigma \in \overline{\Sigma_0}$, then*

$$\liminf_{n\to+\infty} \frac{H(\mathbb{Q}^n_{\sigma_n,b_n} \mid \mathbb{Q}^n_{\sigma_0,b_0})}{n} \geq I(\sigma \mid \sigma_0).$$

The proofs of Propositions 6.8 and 6.9 are left to the next section.

Remark 6.10. Recall that a sequence f_n of real valued functions defined on some metric space Γ-converges to some function f, if

- for all x, $\lim_{n\to+\infty} f_n(x) = f(x)$,
- for all sequences x_n converging to some x, $\liminf_{n\to+\infty} f_n(x_n) \geq f(x)$.

The preceding proposition can thus be restated by saying that for every $b_n \in \mathcal{B}_{\varepsilon_n}$ with ε_n going to 0, the sequence of functions $\sigma \mapsto \frac{H(\mathbb{Q}^n_{\sigma,b_n}|\mathbb{Q}^n_{\sigma_0,b_0})}{n}$ Γ-converges to $\sigma \mapsto I(\sigma \mid \sigma_0)$.

It is well known that this kind of convergence is well adapted for deriving the convergence of minimizers. The next proposition illustrates this fact:

Proposition 6.11. *Suppose that for every n, $\mathbb{Q}^n_{\sigma_n,b_n}$ is an element of $\mathcal{M}^n_{\varepsilon^0_n}$; then*

$$\mathbb{Q}^n_{\sigma_n,b_n} \xrightarrow[n\to+\infty]{} \mathbb{Q}_{\sigma^*,b_0}. \tag{6.12}$$

Proof. For all n, $\mathbb{Q}^n_{\sigma^*,b_0}$ belongs to $A^n_{\varepsilon^0_n}$. Thus, using the minimization property of $\mathbb{Q}^n_{\sigma_n,b_n}$, one has $\frac{1}{n}H(\mathbb{Q}^n_{\sigma_n,b_n} \mid \mathbb{Q}^n_{\sigma_0,b_0}) \leq \frac{1}{n}H(\mathbb{Q}^n_{\sigma^*,b_0} \mid \mathbb{Q}^n_{\sigma_0,b_0})$. According to point (1) of Proposition 6.9, this implies that

$$\limsup_{n\to+\infty} \frac{1}{n} H(\mathbb{Q}^n_{\sigma_n,b_n} \mid \mathbb{Q}^n_{\sigma_0,b_0}) \leq I(\sigma^* \mid \sigma_0). \tag{6.13}$$

According to point (2) of Proposition 6.8, $\sigma_n \in \overline{\Sigma_1}$. This set being compact, one can find some converging subsequence σ_{n_p}. Let $\tilde{\sigma}$ be its limit. Point (2) of Proposition 6.9 yields

$$\liminf_{p\to+\infty} \frac{1}{n_p} H(\mathbb{Q}^{n_p}_{\sigma_{n_p},b_{n_p}} \mid \mathbb{Q}^{n_p}_{\sigma_0,b_0}) \geq I(\tilde{\sigma} \mid \sigma_0). \tag{6.14}$$

From (6.13) and (6.14), one deduces that

$$I(\tilde{\sigma} \mid \sigma_0) \leq I(\sigma^* \mid \sigma_0).$$

As σ^* is the unique minimizer of $I(. \mid \sigma_0)$ under the moment constraint, one has $\tilde{\sigma} = \sigma^*$. The point σ^* is thus the unique accumulation point of the compact sequence σ_n. It follows that σ_n converges to σ^*. Now, (6.12) follows immediately from Proposition 6.4. □

We are now ready to prove Theorem 6.6.

Proof of Theorem 6.6. First, we have the following immediate inequality

$$d_{FM}\left(\mathbb{R}^n_{\varepsilon_n^0,m}, \mathbb{Q}_{\sigma^*,b_0}\right) \leq d_{FM}\left(\mathbb{R}^n_{\varepsilon_n^0,m}, \overline{co}\mathcal{M}^n_{\varepsilon_n^0}\right) + \sup_{\mathbb{Q}\in\overline{co}\mathcal{M}^n_{\varepsilon_n^0}} d_{FM}\left(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0}\right).$$

Thus, according to Proposition 6.7, it suffices to prove that

$$\sup_{\mathbb{Q}\in\overline{co}\mathcal{M}^n_{\varepsilon_n^0}} d_{FM}\left(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0}\right) \xrightarrow[n\to+\infty]{} 0.$$

The application $\mathbb{Q} \mapsto d_{FM}\left(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0}\right)$ being convex and continuous, we get

$$\sup_{\mathbb{Q}\in\overline{co}\mathcal{M}^n_{\varepsilon_n^0}} d_{FM}\left(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0}\right) = \sup_{\mathbb{Q}\in\mathcal{M}^n_{\varepsilon_n^0}} d_{FM}\left(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0}\right).$$

But $\mathcal{M}^n_{\varepsilon_n^0}$ is compact. Thus, there exists $\mathbb{Q}^n_{\sigma_n,b_n} \in \mathcal{M}^n_{\varepsilon_n^0}$, such that

$$\sup_{\mathbb{Q}\in\mathcal{M}^n_{\varepsilon_n^0}} d_{FM}\left(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0}\right) = d_{FM}\left(\mathbb{Q}^n_{\sigma_n,b_n}, \mathbb{Q}_{\sigma^*,b_0}\right).$$

Applying Proposition 6.11, we get

$$\mathbb{Q}^n_{\sigma_n,b_n} \xrightarrow[n\to+\infty]{} \mathbb{Q}_{\sigma^*,b_0},$$

which achieves the proof. □

Before giving the proofs of Propositions 6.8 and 6.9, let us make some comments on our result.

Remark 6.15.
- The reason why we work with $\widetilde{\mathbb{T}}^n_\varepsilon$ instead of the more natural set $\mathbb{T}^n_\varepsilon = \{\mathbb{Q}^n_{\sigma,b} : \sigma \in \Sigma_1, b \in \mathcal{B}_\varepsilon\}$ is that $\mathbb{T}^n_\varepsilon$ is of *empty interior*. The set $\mathbb{T}^n_\varepsilon$ was thus a bad candidate for defining a conditioning event in the Gibbs principle. In fact, from the relative entropy point of view, working with $\widetilde{\mathbb{T}}^n_\varepsilon$ does not change anything: point (2) of Proposition 6.8 shows that the entropy minimizers on A^n_ε *are trinomial trees*.
- We introduced the set Σ_1 because some compactness is needed in Proposition 6.8. Note that if we replace Σ_1 by Σ_0 in the definition of $\widetilde{\mathbb{T}}^n_\varepsilon$, this set becomes *convex* (see [20]). In this framework, there is a unique entropy minimizer $\mathbb{Q}^n_{\sigma_n^*,b_n^*}$. But we are not able to prove directly that the sequence σ_n^* is compact. If this was true, Theorem 6.6 would hold with Σ_0 replacing Σ_1.
- The assumption that $I(. \mid \sigma_0)$ admits a unique minimizer under the moment constraint is needed in the proof of Theorem 6.6. Namely, we used in the proof the fact that the function $\mathbb{Q} \mapsto d_{FM}(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0})$ is convex. If we were dealing with a set $\mathcal{M}$ of minimizers containing more than one element, this function would be replaced by the function $\mathbb{Q} \mapsto d_{FM}(\mathbb{Q}, \mathcal{M})$ which is no longer convex.

6.2. Proofs

Proof of (1) of Proposition 6.8. The set $C_F^e p$ being clearly open, it suffices to show that $\widetilde{\mathbb{T}}_\varepsilon^n$ is an open subset of $M_1(\Omega_n)$. First, it is easily seen that there is a constant $c > 0$ depending only on $\sigma_{\min}, \sigma_{\max}, b_0, s$ and α such that

$$\mathbb{Q}\left(X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right) > c,$$

for all $\mathbb{Q} \in \widetilde{\mathbb{T}}_\varepsilon^n$ and all $|j| \leq k \leq n$. For all $|j| \leq k \leq n$ and $\mathbb{Q} \in M_1(\Omega_n)$, let us define

$$F_{k,j}(\mathbb{Q}) = \alpha\sqrt{n}\frac{\mathbb{Q}\left(X_{\frac{k+1}{n}} = \frac{(j+1)\alpha}{\sqrt{n}}, X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right) - \mathbb{Q}\left(X_{\frac{k+1}{n}} = \frac{(j-1)\alpha}{\sqrt{n}}, X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right)}{\mathbb{Q}\left(X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right)} \quad (6.16)$$

and

$$G_{k,j}(\mathbb{Q}) = \alpha^2\frac{\mathbb{Q}\left(X_{\frac{k+1}{n}} = \frac{(j+1)\alpha}{\sqrt{n}}, X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right) + \mathbb{Q}\left(X_{\frac{k+1}{n}} = \frac{(j-1)\alpha}{\sqrt{n}}, X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right)}{\mathbb{Q}\left(X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right)}. \quad (6.17)$$

These applications are continuous on the open set

$$\left\{\mathbb{Q} \in M_1(\Omega_n) : \forall |j| \leq k \leq n, \mathbb{Q}\left(X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right) > c\right\}$$

and the following holds

$$\mathbb{Q} \in \widetilde{\mathbb{T}}_\varepsilon^n \Leftrightarrow \begin{cases} & \mathbb{Q}\left(X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right) > c, \\ \forall |j| \leq k \leq n, & F_{k,j}(Q) \in]b_0 - \varepsilon, b_0 + \varepsilon[, \\ \forall |q| \leq p \leq n, & G_{k,j}(Q) \in]\sigma_{\min}^2, \sigma_{\max}^2[, \\ & \left|\sqrt{G_{k,j}(\mathbb{Q})} - \sqrt{G_{p,q}(\mathbb{Q})}\right| \\ & \quad < 2\Delta_{n,\sigma^*}\left(\left|\frac{k}{n} - \frac{p}{n}\right| + \left|\frac{\alpha j}{\sqrt{n}} - \frac{\alpha q}{\sqrt{n}}\right|\right). \end{cases}$$

One easily concludes from this that $\widetilde{\mathbb{T}}_\varepsilon^n$ is an open subset of $M_1(\Omega_n)$.

Now let us show that $H(A_\varepsilon^n \mid Q_{\sigma_0,b_0}^n) = H(\overline{A_\varepsilon^n} \mid Q_{\sigma_0,b_0}^n)$. As $\mathbb{Q}_{\sigma_0,b_0}^n$ gives a positive mass to every trajectory of Ω_n, the convex function $M_1(\Omega_n) \ni \mathbb{Q} \mapsto H(\mathbb{Q} \mid \mathbb{Q}_{\sigma_0,b_0}^n)$ is everywhere finite thus continuous. As a consequence, $H(O \mid \mathbb{Q}_{\sigma_0,b_0}^n) = H(\overline{O} \mid \mathbb{Q}_{\sigma_0,b_0}^n)$ holds true for all open sets O of $M_1(\Omega_n)$. This is in particular true for A_ε^n. □

In order to prove point (2) of Proposition 6.8, we need the following lemma.

Lemma 6.18. *For all $\sigma \in \overline{\Sigma_0}$, $b \in \mathcal{B}_\varepsilon$, $\varepsilon \leq s$, let us define:*

$$q_{\sigma,b;\sigma_0,b_0}^n(t, x, y) = \frac{\mathrm{d}\Pi_{\sigma,b}^n(t, x, .)}{\mathrm{d}\Pi_{\sigma_0,b_0}^n(t, x, .)}(y)$$

and

$$h_{\sigma,b;\sigma_0,b_0}^n(t, x) = H(\Pi_{\sigma,b}^n(t, x, .) \mid \Pi_{\sigma_0,b_0}^n(t, x, .)).$$

Then it holds that

$$\frac{\mathrm{d}\mathbb{Q}^n_{\sigma,b}}{\mathrm{d}\mathbb{Q}^n_{\sigma_0,b_0}} = \prod_{i=0}^{n-1} q^n_{\sigma,b;\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}, X_{\frac{i+1}{n}}\right) \tag{6.19}$$

$$H(\mathbb{Q}^n_{\sigma,b} \mid \mathbb{Q}^n_{\sigma_0,b_0}) = \sum_{i=0}^{n-1} \mathbb{E}^n_{\sigma,b}\left[h^n_{\sigma,b;\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right]. \tag{6.20}$$

Let $\mathbb{Q}$ *be a probability measure satisfying*

$$\begin{cases} (1) \quad \mathbb{Q}(X_0 = 0) = 1, \\ (2) \quad \mathbb{Q}\left(X_t = X_{\frac{k}{n}} + (nt-k)\left[X_{\frac{k+1}{n}} - X_{\frac{k}{n}}\right], \frac{k}{n} \le t \le \frac{k+1}{n}\right) = 1, \\ (3) \quad \mathbb{Q}\left(X_{\frac{p+1}{n}} \in .|X_{\frac{p}{n}}\right) = \Pi^n_{\sigma,b}\left(\frac{p}{n}, X_{\frac{p}{n}}, .\right) \end{cases} \tag{6.21}$$

for some $\sigma \in \overline{\Sigma_0}$ *and* $b \in \overline{\mathcal{B}_\varepsilon}$. *Then*

$$\forall i = 0, \ldots, n-1, \quad \mathcal{L}_{\mathbb{Q}}\left(X_{\frac{i}{n}}\right) = \mathcal{L}_{\mathbb{Q}^n_{\sigma,b}}\left(X_{\frac{i}{n}}\right). \tag{6.22}$$

Furthermore,

$$H(\mathbb{Q} \mid \mathbb{Q}^n_{\sigma_0,b_0}) = H(\mathbb{Q} \mid \mathbb{Q}^n_{\sigma,b}) + H(\mathbb{Q}^n_{\sigma,b} \mid \mathbb{Q}^n_{\sigma_0,b_0}). \tag{6.23}$$

Proof. The proofs of (6.19), (6.20) and (6.22) rely on very easy computations and are left to the reader. Let us prove (6.23). It is clear that

$$H(\mathbb{Q} \mid \mathbb{Q}^n_{\sigma_0,b_0}) = H(\mathbb{Q} \mid \mathbb{Q}^n_{\sigma,b}) + \int \log\left(\frac{\mathrm{d}\mathbb{Q}^n_{\sigma,b}}{\mathrm{d}\mathbb{Q}^n_{\sigma_0,b_0}}\right)\mathrm{d}\mathbb{Q}. \tag{6.24}$$

Next, we have

$$\begin{aligned} \int \log\left(\frac{\mathrm{d}\mathbb{Q}^n_{\sigma,b}}{\mathrm{d}\mathbb{Q}^n_{\sigma_0,b_0}}\right)\mathrm{d}\mathbb{Q} &\overset{\text{(i)}}{=} \int \sum_{i=0}^{n-1} \log\left[q^n_{\sigma,b;\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}, X_{\frac{i+1}{n}}\right)\right]\mathrm{d}\mathbb{Q} \\ &\overset{\text{(ii)}}{=} \mathbb{E}_{\mathbb{Q}}\left[\sum_{i=0}^{n-1} \int \log\left[q^n_{\sigma,b;\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}, y\right)\right] \Pi^n_{\sigma,b}\left(\frac{i}{n}, X_{\frac{i}{n}}, \mathrm{d}y\right)\right] \\ &= \sum_{i=0}^{n-1} \mathbb{E}_{\mathbb{Q}}\left[h^n_{\sigma,b;\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right] \\ &\overset{\text{(iii)}}{=} \sum_{i=0}^{n-1} \mathbb{E}^n_{\sigma,b}\left[h^n_{\sigma,b;\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right] \\ &\overset{\text{(iv)}}{=} H(\mathbb{Q}^n_{\sigma,b} \mid \mathbb{Q}^n_{\sigma_0,b_0}), \end{aligned}$$

where (i) follows from (6.19), (ii) is obtained by conditioning by X_i, (iii) is a consequence of (6.22) and (iv) of (6.20). Plugging this in (6.24), we obtain (6.23). □

Proof of (2) of Proposition 6.8. Let $\mathbb{Q}$ be in $\mathcal{M}^n_\varepsilon$. As $\mathbb{Q}$ belongs to $\overline{A^n_\varepsilon}$, there exist $\sigma \in \overline{\Sigma_1}$ and $b \in \overline{\mathcal{B}_\varepsilon}$ such that (6.21) is fulfilled. According to (6.23), one has

$$H(\mathbb{Q} \mid \mathbb{Q}^n_{\sigma_0,b_0}) = H(\mathbb{Q} \mid \mathbb{Q}^n_{\sigma,b}) + H(\mathbb{Q}^n_{\sigma,b} \mid \mathbb{Q}^n_{\sigma_0,b_0}).$$

If $\mathbb{Q}^n_{\sigma,b}$ belongs to $\overline{A^n_\varepsilon}$, then we deduce from the preceding equation that $H(\overline{A^n_\varepsilon} \mid \mathbb{Q}^n_{\sigma_0,b_0}) \geq H(\mathbb{Q} \mid \mathbb{Q}^n_{\sigma,b}) + H(\overline{A^n_\varepsilon} \mid \mathbb{Q}^n_{\sigma_0,b_0})$, and consequently $H(\mathbb{Q} \mid \mathbb{Q}^n_{\sigma,b}) = 0$, which implies that $\mathbb{Q} = \mathbb{Q}^n_{\sigma,b}$. Thus, the only thing to do is to prove that $\mathbb{Q}^n_{\sigma,b} \in \overline{A^n_\varepsilon}$.

Let $\left(\mathbb{Q}_p\right)_p$ be a sequence of A^n_ε going to $\mathbb{Q}$. For each p, there is a pair $(\sigma_p, b_p) \in \Sigma_1 \times \mathcal{B}_\varepsilon$ such that (6.21) is fulfilled. For all $|j| \leq k \leq n$, one has

$$b_p\left(\frac{k}{n}, \frac{\alpha j}{\sqrt{n}}\right) = F_{k,j}(\mathbb{Q}_p)$$

and

$$\sigma_p^2\left(\frac{k}{n}, \frac{\alpha j}{\sqrt{n}}\right) = G_{k,j}(\mathbb{Q}_p),$$

where $F_{k,j}$ and $G_{k,j}$ are defined by (6.16) and (6.17). These functions being continuous, we have

$$b_p\left(\frac{k}{n}, \frac{\alpha j}{\sqrt{n}}\right) \xrightarrow[p\to+\infty]{} b\left(\frac{k}{n}, \frac{\alpha j}{\sqrt{n}}\right)$$

and

$$\sigma_p^2\left(\frac{k}{n}, \frac{\alpha j}{\sqrt{n}}\right) \xrightarrow[p\to+\infty]{} (\sigma)^2\left(\frac{k}{n}, \frac{\alpha j}{\sqrt{n}}\right),$$

for all $|j| \leq k \leq n$. It follows easily that

$$\mathbb{Q}^n_{\sigma_p,b_p} \xrightarrow[p\to+\infty]{} \mathbb{Q}^n_{\sigma,b}.$$

But according to (6.22),

$$\mathbb{Q}_p \in A^n_\varepsilon \Rightarrow \mathbb{Q}^n_{\sigma_p,b_p} \in A^n_\varepsilon.$$

Consequently, $\mathbb{Q}^n_{\sigma,b}$ is in the closure of A^n_ε. □

Proof of Proposition 6.9. Recall that for all $\sigma \in \overline{\Sigma_0}$, $I(\sigma \mid \sigma_0)$ is defined by

$$I(\sigma \mid \sigma_0) = \mathbb{E}_{\sigma,b}\left[\int_0^1 q(\sigma^2(t, X_t), \sigma_0^2(t, X_t))\,\mathrm{d}t\right],$$

with

$$q(x, y) = \log\left(\frac{x}{y}\right)\frac{x}{\alpha^2} + \log\left(\frac{\alpha^2 - x}{\alpha^2 - y}\right)\left[1 - \frac{x}{\alpha^2}\right].$$

(1) Let us show that there exists some $K > 0$, depending only on α, $\sigma_{\min}$, $\sigma_{\max}$, b_0 and s, such that

$$|h^n_{\sigma,b;\sigma_0,b_0} - q(\sigma^2, \sigma_0^2)|\left(\frac{k}{n}, x\right) \leq \frac{K}{n} \tag{6.25}$$

for all $(k, x) \in \{0, \ldots, n-1\} \times \frac{\alpha}{\sqrt{n}}\mathbb{Z}$ and $(\sigma, b) \in \overline{\Sigma_0} \times \overline{\mathcal{B}_s}$.

For all $(\sigma, b) \in \overline{\Sigma_0} \times \overline{\mathcal{B}_s}$:

$$\begin{aligned}
\log\left[\frac{m^n(\sigma,b)}{m^n(\sigma_0,b_0)}\right] m^n(\sigma,b) &= \left[\log\left(\frac{\sigma^2}{\sigma_0^2}\right) + \log\left(1+\frac{b\alpha}{\sqrt{n}\sigma^2}\right) - \log\left(1+\frac{b_0\alpha}{\sqrt{n}\sigma_0^2}\right)\right] \\
&\quad \times \left[\frac{\sigma^2}{2\alpha^2} + \frac{b}{2\alpha\sqrt{n}}\right] \\
\log\left[\frac{\mathrm{d}^n(\sigma,b)}{\mathrm{d}^n(\sigma_0,b_0)}\right] \mathrm{d}^n(\sigma,b) &= \left[\log\left(\frac{\sigma^2}{\sigma_0^2}\right) + \log\left(1-\frac{b\alpha}{\sqrt{n}\sigma^2}\right) - \log\left(1-\frac{b_0\alpha}{\sqrt{n}\sigma_0^2}\right)\right] \\
&\quad \times \left[\frac{\sigma^2}{2\alpha^2} - \frac{b}{2\alpha\sqrt{n}}\right] \\
\log\left[\frac{r^n(\sigma,b)}{r^n(\sigma_0,b_0)}\right] r^n(\sigma,b) &= \log\left(\frac{\alpha^2-\sigma^2}{\alpha^2-\sigma_0^2}\right)\left[1-\frac{\sigma^2}{\alpha^2}\right].
\end{aligned}$$

Using Taylor's formula, it is easily seen that for $\varepsilon \in [-1, 1]$,

$$\sup_{\substack{x\in[\sigma_{\min}^2,\sigma_{\max}^2] \\ y\in[b_0-s,b_0+s]}} \left|\log\left(1+\frac{\varepsilon y\alpha}{\sqrt{n}x}\right) - \frac{\varepsilon y\alpha}{\sqrt{n}x} + \frac{1}{2}\left(\frac{\varepsilon y\alpha}{\sqrt{n}x}\right)^2\right| \le \frac{K}{n\sqrt{n}},$$

with K depending only on α, $\sigma_{\max}$, $\sigma_{\min}$, b_0 and s.

After some easy computations, one derives (6.25) from these inequalities.

In the sequel we will use the following notation:

$$\Phi^n = \frac{1}{n}\sum_{i=0}^{n-1} q\left(\sigma^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right), \sigma_0^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right)$$

and

$$\Phi = \int_0^1 q(\sigma^2(t, X_t), \sigma_0^2(t, X_t))\,\mathrm{d}t.$$

The function q is bounded and continuous on $[\sigma_{\min}^2, \sigma_{\max}^2]^2$. Φ^n is thus a sequence of uniformly bounded continuous functions on Ω, which converges pointwise to the bounded continuous function Φ. Let us show that Φ^n converges uniformly to Φ on every compact subset of Ω. The function q is Lipschitz on $[\sigma_{\min}^2, \sigma_{\max}^2]^2$; let $M > 0$ be such that

$$|q(x, y) - q(x', y')| \le M(|x - x'| + |y - y'|).$$

Let Δ be the continuity modulus of σ^2, i.e.

$$\Delta(u) = \sup_{|t-s|+|y-x|\le u} |\sigma^2(s, x) - \sigma^2(t, y)|,$$

and Δ_0 the continuity modulus of σ_0^2.

With this notation, we have

$$\begin{aligned}|\Phi^n - \Phi| &= \left|\frac{1}{n}\sum_{i=0}^{n-1} q\left(\sigma^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right), \sigma_0^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right) - \int_0^1 q(\sigma^2(t, X_t), \sigma_0^2(t, X_t))\,\mathrm{d}t\right| \\
&\le \sum_{i=0}^{n-1}\int_{\frac{i}{n}}^{\frac{i+1}{n}} \left|q\left(\sigma^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right), \sigma_0^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right) - q(\sigma^2(t, X_t), \sigma_0^2(t, X_t))\right| \mathrm{d}t \\
&\le M\sum_{i=0}^{n-1}\int_{\frac{i}{n}}^{\frac{i+1}{n}} \left|\sigma^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right) - \sigma^2(t, X_t)\right| + \left|\sigma_0^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right) - \sigma_0^2(t, X_t)\right| \mathrm{d}t \\
&\le M\left[\sup_{|s-t|\le\frac{1}{n}} |\sigma^2(s, X_s) - \sigma^2(t, X_t)| + \sup_{|s-t|\le\frac{1}{n}} |\sigma_0^2(s, X_s) - \sigma_0^2(t, X_t)|\right] \\
&\le M\left[\sup_{|s-t|\le\frac{1}{n}} \Delta\left(|s-t| + |X_s - X_t|\right) + \sup_{|s-t|\le\frac{1}{n}} \Delta_0\left(|s-t| + |X_s - X_t|\right)\right] \\
&\le M\left[\Delta\left(\frac{1}{n} + \sup_{|s-t|\le\frac{1}{n}} |X_s - X_t|\right) + \Delta_0\left(\frac{1}{n} + \sup_{|s-t|\le\frac{1}{n}} |X_s - X_t|\right)\right].\end{aligned}$$

Let $\mathcal{K}$ be a compact subset of Ω. According to the Ascoli theorem, we have

$$\sup_{\omega\in\mathcal{K}} \sup_{|t-s|\le\frac{1}{n}} |X_s - X_t| \xrightarrow[n\to+\infty]{} 0.$$

Thus

$$\sup_{\omega\in\mathcal{K}} |\Phi^n(\omega) - \Phi(\omega)| \xrightarrow[n\to+\infty]{} 0.$$

According to (6.25):

$$\left|\frac{1}{n}H(\mathbb{Q}^n_{\sigma,b_n} \mid \mathbb{Q}^n_{\sigma_0,b_0}) - \mathbb{E}^n_{\sigma,b_n}\left[\Phi^n\right]\right| \le \frac{K}{n}$$

where K depends only on α, $\sigma_{\max}$, $\sigma_{\min}$, b_0 and s. Using the uniform convergence of $(\Phi^n)_n$ on every compact and the tightness of the sequence $\mathbb{Q}^n_{\sigma,b_n}$, it is now easy to see that

$$\lim_{n\to\infty} \frac{1}{n}H(\mathbb{Q}^n_{\sigma,b_n} \mid \mathbb{Q}^n_{\sigma_0,b_0}) = I(\sigma \mid \sigma_0).$$

(2)

$$\begin{aligned}\frac{1}{n}H(\mathbb{Q}^n_{\sigma_n,b_n} \mid \mathbb{Q}^n_{\sigma_0,b_0}) &= \frac{1}{n}\int \log\left(\frac{\mathrm{d}\mathbb{Q}^n_{\sigma_n,b_n}}{\mathrm{d}\mathbb{Q}^n_{\sigma_0,b_0}}\right)\mathrm{d}\mathbb{Q}^n_{\sigma_n,b_n} \\
&= \frac{1}{n}\int \log\left(\frac{\mathrm{d}\mathbb{Q}^n_{\sigma_n,b_n}}{\mathrm{d}\mathbb{Q}^n_{\sigma,b_n}}\right)\mathrm{d}\mathbb{Q}^n_{\sigma_n,b_n} + \frac{1}{n}\int \log\left(\frac{\mathrm{d}\mathbb{Q}^n_{\sigma,b_n}}{\mathrm{d}\mathbb{Q}^n_{\sigma_0,b_0}}\right)\mathrm{d}\mathbb{Q}^n_{\sigma_n,b_n} \\
&= \frac{1}{n}H(Q^n_{\sigma_n,b_n} \mid Q^n_{\sigma,b_n}) + \frac{1}{n}\int \log\left(\frac{\mathrm{d}\mathbb{Q}^n_{\sigma,b_n}}{\mathrm{d}\mathbb{Q}^n_{\sigma_0,b_0}}\right)\mathrm{d}\mathbb{Q}^n_{\sigma_n,b_n} \\
&\ge \frac{1}{n}\int \log\left(\frac{\mathrm{d}\mathbb{Q}^n_{\sigma,b_n}}{\mathrm{d}\mathbb{Q}^n_{\sigma_0,b_0}}\right)\mathrm{d}\mathbb{Q}^n_{\sigma_n,b_n}.\end{aligned}$$

According to (6.19) of Lemma 6.18

$$\frac{1}{n}\int \log\left(\frac{\mathrm{d}\mathbb{Q}^n_{\sigma,b_n}}{\mathrm{d}\mathbb{Q}^n_{\sigma_0,b_0}}\right)\mathrm{d}\mathbb{Q}^n_{\sigma_n,b_n} = \mathbb{E}^n_{\sigma_n,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1} k^n\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right],$$

where

$$\begin{aligned} k^n &= \log\left(\frac{m^n(\sigma,b_n)}{m^n(\sigma_0,b_0)}\right) m^n(\sigma_n,b_n) + \log\left(\frac{r^n(\sigma,b_n)}{r^n(\sigma_0,b_0)}\right) r^n(\sigma_n,b_n) \\ &\quad + \log\left(\frac{\mathrm{d}^n(\sigma,b_n)}{\mathrm{d}^n(\sigma_0,b_0)}\right)\mathrm{d}^n(\sigma_n,b_n). \end{aligned}$$

It is easily seen that there is a constant K depending only on α, $\sigma_{\min}$, $\sigma_{\max}$, b_0 and s such that

$$\forall R > 0, \qquad \sup_{|x|\le R, t\in[0,1]} |k^n - h^n_{\sigma,b_n;\sigma_0,b_0}|(t,x) \le K \sup_{|x|\le R,t\in[0,1]} |\sigma_n - \sigma|(t,x).$$

The sequence $\mathbb{Q}^n_{\sigma_n,b_n}$ converging to $\mathbb{Q}_{\sigma,b}$ is a tight sequence. As a consequence, for all $\beta > 0$, there is $R > 0$ such that

$$\mathbb{Q}^n_{\sigma_n,b_n}\left(\sup_{t\in[0,1]} |X_t| \le R\right) \ge 1-\beta.$$

One can find $M > 0$ depending on α, $\sigma_{\min}$, $\sigma_{\max}$, b_0 and s, such that $|k^n| \le M$ and $|h^n_{\sigma,b_n;\sigma_0,b_0}| \le M$. Thus,

$$\begin{aligned} &\left|\mathbb{E}^n_{\sigma_n,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1} k^n\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right] - \mathbb{E}^n_{\sigma_n,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1} h^n_{\sigma,b_n;\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right]\right| \\ &\quad \le \mathbb{E}^n_{\sigma_n,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1} |h^n_{\sigma,b_n;\sigma_0,b_0} - k^n| \mathbb{1}_{[0,R]}(\sup_{t\in[0,1]} |X_t|)\right] + 2M(1-\beta) \\ &\quad \le K \sup_{|x|\le R,\, t\in[0,1]} |\sigma_n - \sigma|(t,x) + 2M(1-\beta). \end{aligned}$$

One easily concludes that

$$\mathbb{E}^n_{\sigma_n,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1} k^n\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right] - \mathbb{E}^n_{\sigma_n,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1} h^n_{\sigma,b_n;\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right] \xrightarrow[n\to+\infty]{} 0.$$

A reasoning similar to that in the proof of point (1) shows that

$$\mathbb{E}^n_{\sigma_n,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1} h^n_{\sigma,b_n;\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right] \xrightarrow[n\to+\infty]{} I(\sigma \mid \sigma_0),$$

which achieves the proof. $\square$

Acknowledgements

We want to warmly acknowledge Christian Léonard for so many animated conversations on large deviations problems, and for indicating to us various references on the topic. We also acknowledge an anonymous referee for the very careful reading of the manuscript.

References

[1] R. Aebi, Schrödinger Diffusion Processes, Birkhäuser, Basel, Berlin, Boston, 1996.

[2] M. Avellaneda, C. Friedman, R. Holmes, D. Samperi, Calibrating volatility surfaces via relative-entropy minimization, Appl. Math. Finance 4 (1) (1997) 37–64.

[3] F. Bolley, C. Villani, Weighted Csiszar–Kullback–Pinsker inequalities and applications to transportation inequalities, Ann. Fac. Sci. Toulouse 14 (3) (2005) 331–352.

[4] E. Carlen, Conservative diffusions, Comm. Math. Phys. 94 (1984) 293–316.

[5] P. Cattiaux, F. Gamboa, Large deviations and variational theorems for marginal problems, Bernoulli 5 (1999) 81–108.

[6] P. Cattiaux, C. Léonard, Minimization of the Kullback information of diffusion processes, Ann. Inst. H. Poincaré Probab. Statist. 30 (1) (1994) 83–132 (correction in Ann. Inst. H. Poincaré 31 (1995) 705–707).

[7] P. Cattiaux, C. Léonard, Large deviations and Nelson processes, Forum Math. 7 (1995) 95–115.

[8] P. Cattiaux, C. Léonard, Minimization of the Kullback information for general Markov processes, in: Séminaire de Probas XXX, in: Lect. Notes Math., vol. 1626, 1996, pp. 283–311.

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

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

[11] A. Dembo, O. Zeitouni, Refinements of the Gibbs conditioning principle, Probab. Theory Related Fields 104 (1996) 1–14.

[12] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, second edn, Springer-Verlag, 1998.

[13] A. Van der Vaart, J. Wellner, Weak Convergence and Empirical Processes, in: Springer Series in Statistics, Springer, 1995.

[14] J.D. Deuschel, D.W. Stroock, Large Deviations, Academic Press, 1989.

[15] J.D. Deuschel, D.W. Stroock, H. Zessin, Microcanonical distributions for lattice gases, Comm. Math. Phys. 139 (1991) 83–101.

[16] P. Diaconis, D.A. Freedman, Conditional limit theorems for exponential families and finite versions of de Finetti's theorem, J. Theoret. Probab. 1 (1988) 381–410.

[17] R.M. Dudley, Real Analysis and Probability, Wadsworth and Brooks/Cole, 1989.

[18] P. Eichelsbacher, U. Schmock, Large deviations of u-empirical measures in strong topologies and applications, Ann. Inst. H. Poincaré. Probab. Statist. 38 (5) (2002) 779–797.

[19] H. Föllmer, Random fields and diffusion processes, in: Ecole d'été de probabilités de Saint-Flour, in: Lect. Notes Math., vol. 1362, 1988, pp. 101–204.

[20] N. Gozlan, Principe conditionnel de Gibbs pour des contraintes fines approchées et Inégalités de transport. Ph.D. Thesis, Available at: http://tel.ccsd.cnrs.fr/documents/archives0/00/01/01/73/tel-00010173-00/tel-00010173.pdf (in French).

[21] N. Gozlan, Conditional principles for random weighted measures, ESAIM Probab. Stat. 9 (2005) 283–306.

[22] S.R. Kulkarni, O. Zeitouni, A general classification rule for probability measures, Ann. Statist. 23 (1995) 1393–1407.

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

[24] V.V. Yurinskii, Exponential inequalities for sums of random vectors, J. Multivariate Anal. 6 (1976) 473–499.