Large Deviations and Variational Theorems for Marginal Problems

P. Cattiaux * F. Gamboa†

February 6, 1995

Summary
On a product Probability space $(E \times F, P)$, we give variational characterizations for the existence of a Probability measure Q with given marginals, such that Q is absolutely continuous with respect to P, and its density satisfies some integrability conditions. These characterizations, which are in some sense the dual formulation of a Theorem of Strassen [34], are obtained by using large deviations methods. We also study the minimal realizations of such Q.

Key words: Marginal Problems-Large Deviations

Contents

*Ecole Polytechnique, CMAP, F-91128 Palaiseau Cedex and Université Paris X équipe MODAL'X Nanterre 200 avenue de la république F-92001 Nanterre Cedex. email: cattiaux@paris.polytechnique.fr

†Laboratoire de Statistiques Université Paris Sud F-91405 Orsay and Université Paris-Nord Institut Galilée F-93430 Villetaneuse. email: gamboa@stats.matups.fr

1 Introduction

Let E and F be two topological spaces equipped with their Borel σ-fields, and μ and ν two Probability measures defined on E and F respectively. We take a Probability measure P on $E \times F$, and ask the following question:

$$(*) \begin{cases} \textit{is it possible to find a Probability measure } Q \textit{ on } E \times F, \textit{ with marginals } \mu \textit{ and } \nu, \\ \qquad \textit{such that } Q \ll P \textit{ and } \frac{dQ}{dP} \textit{ satisfies some integrability conditions.} \end{cases}$$

The construction of measures on a product space, with given marginals and satisfying convex constraints is an old problem. A celebrated result of Strassen [34] gives a variational necessary and sufficient condition of existence. In section 2, we explain how to use the Strassen's result in order to answer to our initial question.
Since the constraint is here implicit (before to build Q, one cannot control $\frac{dQ}{dP}$), usual duality results (as in [26]) do not hold. However, as in [6] or [16] (which deal respectively with marginal flows, or moment problems) a kind of dual formulation of Strassen's result can be obtained by using large deviations arguments. This is the aim of section 3, where we derive new variational characterizations. We underline that the method can be extended to more general product spaces (i.e. for instance $C^0([0,1],E)$ considered as a subspace of $E^{[0,1]}$). This will be done elsewhere.
The origin of this work was an original question of Schrödinger [32], linked by Föllmer to a problem of entropy minimization. This point is explained at the beginning of section 5. In this section, the problem of minimal realizations of the problem $(*)$ (minimal for an associated Orlicz norm for instance) is partly studied.

2 Notations and first results

Let E and F be two topological spaces, equipped with their Borel σ-fields $\mathcal{B}(E)$ and $\mathcal{B}(F)$, and two Probability measures μ and ν defined respectively on $(E, \mathcal{B}(E))$ and $(F, \mathcal{B}(F))$. An old problem is to know whether there exists a Probability measure Q on the product space $(E \times F, \mathcal{B}(E) \otimes \mathcal{B}(F))$, belonging to a certain subset Λ and with marginals μ and ν. After several results in particular cases (see e.g. the work of Kellerer [24], [23], [25]), V. Strassen stated in a celebrated paper a nice variational necessary and sufficient condition.

Theorem 2.1 *[[34] Theorem 7] Assume that E and F are Polish spaces, and that Λ is a non empty weakly closed convex subset of $\mathcal{M}_1^+(E \times F)$, the set of Probability measures on $E \times F$. Then, there exists a Q in Λ with marginals μ and ν if and only if*

$$\textit{for all } f \in C_b(E),\ g \in C_b(F),\ \int f d\mu + \int g d\nu \leq \sup_{Q' \in \Lambda} \int f \oplus g \, dQ'$$

where $f \oplus g(x,y) = f(x) + g(y)$ on $E \times F$.

Theorem 2.1 was successively extended to completely regular spaces in [20] and general Hausdorff spaces in [33], assuming in both cases that μ and ν are Radon, and replacing in [33] the classical weak topology on Radon bounded measures by the narrow

topology, and C_b by B_b (the set of Borel bounded function). In the sequel we denote by $\mathcal{M}^b(U)$ (resp. $\mathcal{M}^b_+(U)$; $\mathcal{M}^+_1(U)$) the set of all bounded Radon (resp. positive bounded Radon; Probability Radon) measures on $(U, \mathcal{B}(U))$ where U is a topological Hausdorff space equipped with its Borel σ-field. The weak topology on $\mathcal{M}^b(U)$ is the one induced by the embedding $\mathcal{M}^b(U) \hookrightarrow C_b^*(U)$, where C_b^* is the topological dual space of C_b, the space of real valued bounded functions. The narrow topology is the weakest topology for which the map $\lambda \to \int f d\lambda$ are lower semi continuous whenever f is a bounded lower semi continuous real function.
Recall that for a positive measure P, Radon means

$$P(A) = \sup\{P(K),\ K \text{ compact } K \subset A\}$$

for all Borel set A, and for a signed measure P that P^+ and P^- are Radon (see [11]). The special case we are interested in, in this paper, is the one where

$$\Lambda = \{Q:\ Q \ll P \text{ and } \frac{dQ}{dP} \in \Gamma\}$$

for a given (Radon) Probability measure P defined on $(E \times F, \mathcal{B}(E) \otimes \mathcal{B}(F))$ and Γ a ball of $L^q(P)$, $1 \le q \le +\infty$ or an Orlicz space related to P. We can easily deduce from Theorem 2.1 and its extensions

Corollary 2.2 *Assume that E and F are completely regular, P belongs to $\mathcal{M}^+_1(E \times F)$, K is a real number, and define $\Gamma_{q,K}$ as the closed ball of radius K of $L^q(P)$,*

$$\Gamma^1_{q,K} = \Gamma_{q,K} \cap \{Z \ge 0, \int Z dP = 1\}.$$

Then, for $1 < q \le +\infty$, there exists $Q \in \mathcal{M}^+_1(E \times F)$ such that $Q \ll P$, $\frac{dQ}{dP} \in \Gamma_{q,K}$ and with marginals μ and ν if and only if

$$\text{for all } f \in C_b(E),\ g \in C_b(F)\quad \int f d\mu + \int g d\nu \le \sup_{Z \in \Gamma^1_{q,K}} \int (f \oplus g) Z dP.$$

Proof
The set $\Lambda_{q,K} = \{Q \in \mathcal{M}^+_1(E \times F):\ Q \ll P \text{ and } \frac{dQ}{dP} \in \Gamma_{q,K}\}$ is convex. By Dunford Pettis Theorem, any element Q' of the weak closure of $\Lambda_{q,K}$ is absolutely continuous with respect to P. Furthermore, if $Z = \frac{dQ'}{dP}$ then Z induced a linear form on $(C_b(E \times F), \|\ \|_{q'})$ where q' is the conjugate of q. Since P is inner regular and $E \times F$ completely regular, $C_b(E \times F)$ is dense in $L^{q'}(P)$ (since $1 \le q' < +\infty$), and Z belongs to the strong dual of $L^{q'}(P)$, i.e. $Z \in L^q(P)$ with a norm less or equal than K. This shows that $\Lambda_{q,K}$ is weakly closed (actually weakly compact) and we may apply Theorem 2.1. ∎

Though the fact that $\Lambda_{q,K}$ is weakly closed, is certainly well known to everybody, we included the above proof in order to extend the result to the larger class of Orlicz spaces, which are less well known. Let $L_\theta(P)$ denotes the Orlicz space associated with the Young function θ and $P \in \mathcal{M}^b_+$. Let we recall briefly what we need about such spaces (see [27]).

(2.3.i) θ is a continuous, non decreasing convex function from $\mathbb{R}^+$ into $\mathbb{R}^+$, such that $\theta(0) = 0$ and $\lim_{t\to\infty} \frac{\theta(t)}{t} = +\infty$.

(2.3.ii) $L_\theta(P)$ is the set of functions f such that there exists a constant a with

$$\int \theta\left(\left|\frac{f}{a}\right|\right) dP < +\infty, \text{ and } \|f\|_\theta = \inf\{a : \int \theta\left(\left|\frac{f}{a}\right|\right) dP \leq 1\}.$$

(2.3.iii) θ is called moderate if $\theta(2t) \leq c\,\theta(t)$ for a suitable c and t large enough.

(2.3.iv) $E_\theta(P)$ is the closure of C_b in $(L_\theta(P), \|\;\|_\theta)$, which in general differs from L_θ, but if P is Radon and θ moderate, $E_\theta(P) = L_\theta(P)$.

(2.3.v) If P is Radon, the strong dual of $(E_\theta, \|\;\|_\theta)$ is identified with $(L_{\theta^*}(P), \|\;\|^*_\theta)$, where θ^* is the Legendre-conjugate function of θ, and $\|\;\|^*_\theta$ is a norm on L_{θ^*} satisfying

$$\|\;\|_{\theta^*} \leq \|\;\|^*_\theta \leq 2\|\;\|_{\theta^*}.$$

If we replace $\Gamma_{q,K}$ by the corresponding $\Gamma_{\theta,K}$ in the previous proof, we immediately remark that the only difficulty is the apparition of a factor 2 in the Hölder-Orlicz inequality. Indeed, Z belongs to $(E_{\theta^*})^* = L_\theta$ but $\|Z\|^*_{\theta_*} \leq 2K$ i.e. $\|Z\| \leq 2K$. So $\Lambda_{\theta,K}$ is not clearly weakly closed, but nevertheless we can state

Corollary 2.4 *With the same assumptions as in Corollary 2.2, denote by $\Gamma_{\theta,K}$ the closed ball of $L_\theta(P)$ of radius K. Then*

$$\text{if for all } f \in C_b(E),\ g \in C_b(F) \quad \int f d\mu + \int g d\nu \leq \sup_{Z \in \Gamma^1_{\theta,K}} \int (f \oplus g) Z dP$$

there exists $Q \in \mathcal{M}^+_1(E \times F)$ such that $Q \ll P$ and $\left\|\frac{dQ}{dP}\right\|_\theta \leq 2K$. (The converse statement, without factor 2, is obvious).

Remark that, in fact, we really need to show that

$$\Gamma^1_{\theta,K} = \Gamma_{\theta,K} \cap \{Z : \int Z dP = 1 \text{ and } Z \geq 0\}$$

induced a weakly closed set $\Lambda_{\theta,K}$; and in the above derivation we did not use the fact that Z is a density of Probability. We actually do not know whether this additional condition is enough to show that $\Lambda_{\theta,K}$ is weakly closed in general. But, in the particular (and very important) case of $\theta(t) = (t+1)\log(t+1) - t$; one can modify our request in order to eliminate the factor 2. Indeed, for $Q \in \mathcal{M}^+_1(E \times F)$, introduce the Kullback information of Q (relative to P)

$$\begin{aligned} H(Q,P) &= \int \frac{dQ}{dP} \log\left(\frac{dQ}{dP}\right) dP \text{ if } Q \ll P \text{ and } \tfrac{dQ}{dP}\log\left(\tfrac{dQ}{dP}\right) \in L^1(P) \\ &= +\infty \text{ otherwise.} \end{aligned} \tag{2.5}$$

It is easy to see that

$$\|\frac{dQ}{dP} - 1\|_\theta \leq H(Q,P) + 1.$$

But, since Q and P are inner regular (which implies that $C_b(E\times F)$ is dense in L^1 for each one), the following alternate expression of H is known (see [1] p. 36-37)

$$H(Q,P) = \sup_{f\in C_b(E\times F)} \left(\int f dQ - \log \int \exp(f) dP\right). \tag{2.6}$$

The above form shows that

$$\Lambda_{H,K} = \{Q \in \mathcal{M}_1^+(E\times F):\ H(Q,P) \leq K\} \tag{2.7}$$

is weakly closed (actually weakly compact). The convexity follows from (2.5), and we thus have

Corollary 2.8 *Under the conditions of Corollary 2.2, there exists $Q \in \mathcal{M}_1^+(E\times F)$ such that $H(Q,P) \leq K$, and with marginals μ and ν if and only if*

$$\textit{for all } f \in C_b(E),\ g\in C_b(F),\ \int f d\mu + \int g d\nu \leq \sup_{Q'\in\Lambda_{H,K}} \int (f\oplus g)\, dQ'.$$

In all the above examples, the weak closedness was actually obtained thanks to the weak relative compactness criterion due to Dunford and Pettis (see e.g [11] p. 38). In the L^1 case however, this property is lost, unless we assume some uniform integrability condition. But, la Vallée-Poussin's Theorem (see e.g. [11] p. 38), implies that any uniformaly integrable set of $L^1(P)$ is included in the unit ball of some L_θ. So the next corollary seems to be optimal.

Corollary 2.9 *Under the conditions of Corollary 2.2, there exists $Q \in \mathcal{M}_1^+(E\times F)$ such that $Q \ll P$ and with marginals μ and ν if and only if there exist $K\in \mathbb{R}^+$ and a Young function θ such that*

$$\textit{for all } f \in C_b(E),\ g\in C_b(F),\ \int f d\mu + \int g d\nu \leq \sup_{Q'\in\Lambda_{\theta,K}} \int f\oplus g\, dQ'.$$

As in optimization problems, one now should ask about a *dual* formulation of Strassen's condition. Recently, Kellerer [26] studied in details this kind of problem, but here the constraint is implicit (we want that the density of an unknown Q belongs to some L_θ space) and cannot be treated by the method of [26].
The most natural way would be here to remark that

$$\text{for } Z\in\Gamma_{\theta,K} \quad \int (f\oplus g) Z dP \leq 2K\|f\oplus g\|_{\theta^*} \tag{2.10}$$

and thus to ask whether the following property

$$\text{for all } f\in C_b^+(E),\ g\in C_b^+(F),\ \int f d\mu + \int g d\nu \leq 2K\|f\oplus g\|_{\theta^*} \tag{2.11}$$

is sufficient for the existence of a Q in $\Lambda_{\theta,2K}$ with marginals μ and ν. (One can always add a constant in Strassen's condition, and accordingly check it only for non negative functions).
The answer is clearly negative, and we shall give a basic couter example

(2.12) **Counter example**

Take $E = F = [0,2]$,

$$dP = (\frac{1}{4} 1_{A^c} + \frac{1}{4} h(x)h(y)1_A)dxdy,$$

where A is the unit square and h is any Probability density on $[0,1]$ such that $\forall \alpha > 0$, the function $1/h^\alpha$ is not integrable and $h > 0$ almost surely (for the uniform Probability on $[0,1]$). Now, let μ ν be the uniform distribution on $[0,1]$, and denote by μ_0 and ν_0 the marginals of P. It is clear that $\mu \ll \mu_0$ and

$$\left\| \frac{d\mu}{d\mu_0} \right\|_\infty = \frac{4}{1+h} 1_{[0,1]} \leq 4, \text{ (idem for the pair } (\nu, \nu_0)),$$

so that for all f and g in $C_b^+([0,2])$

$$\int f d\mu + \int g d\nu \leq 4 \int f \oplus g \; dP \leq 4\|f \oplus g\|_q \text{ for all } q \in [1, +\infty].$$

But, any $Q \in \mathcal{M}_1^+(E \times F)$ with marginals μ and ν, has its support in the unit square. Hence, if $Q = Zdxdy \ll P$ and $q \in]1, +\infty]$ we have:

$$1_{[0,1]}(y) = \int_0^1 Z dx = \int_0^1 \frac{Z}{h} h dx \leq \int_0^1 \frac{Z^q}{h^{q-1}} dx.$$

So that, $\frac{dQ}{dP}$ cannot belong to any $L^q(P)$ space.

This counter example also shows that even if $\frac{d\mu}{d\mu_0}$ and $\frac{d\nu}{d\nu_0}$ are bounded and P is equivalent to a product measure on $E \times F$, one cannot necessarily find a Q in $\mathcal{M}_1^+(E \times F)$ such that $\frac{dQ}{dP} \in \mathrm{L}^q(P)$ for some $q > 1$, and with marginals μ and ν. Of course, for $f \in C_b^+(E)$, $g \in C_b^+(F)$ and $1 < q \leq +\infty$

$$\sup_{Z \in \Gamma_{q,K}} \int (f \oplus g) Z dP = K \|f \oplus g\|_{q'} \text{ with } \frac{1}{q} + \frac{1}{q'} = 1$$

but in Strassen's condition, we have to take the supremum over $\Gamma^1_{q,K}$, i.e. with two additional constraints ($Z \geq 0$ and $\int Z dP = 1$), which makes the difference.

However, we shall see in the next section that one can state a necessary and sufficient condition for the existence of Q which is related to (2.11). This will be done by using some large deviations arguments.

3 Large deviations and new variational characterizations

In order to study the large deviations problem for the empirical process associated to a given i.i.d. sample of, say, Brownian motions, Dawson and Gärtner [10], introduced a variational characterization of the infimum of $H(Q,P)$, (for P the Wiener measure) on the set of Q such that $Q \circ X_t^{-1} = \nu_t$ is given (also see [15] for the same problem for bridges). In recent papers, Cattiaux and Léonard have extended the results of [10] to a large class of Markov processes (see [5], [6], [7]). In particular, the problem of finiteness of the infimum (i.e. the existence of such a Q) is tackled in [6] by using a direct large deviations argument. We shall give this argument now, in the context of the present paper.

Definition 3.1 *We say that a sequence (R_n) of Probability measures on a measurable Hausdorff space $(U, \mathcal{B}(U))$ satisfies a large deviations principle (L.D.P.), with rate function I if:*

i) I is lower semi continuous, with values in $\mathbb{R}^+ \cup \{+\infty\}$.

ii) For any measurable set A of U:

$$-I(int(A)) \leq \liminf_{n\to\infty} \frac{1}{n} \log R_n(A) \leq \limsup_{n\to\infty} \frac{1}{n} \log R_n(A) \leq -I(\overline{A}),$$

where $I(A) = \inf_{\xi \in A} I(\xi)$.

We shall frequently use the following properties (see e.g. [12])

(3.2.i) If the level sets of the rate function I are compact, we shall call I a good rate function.

(3.2.ii) Contraction principle: If $T : U \to V$ is a continuous map, and I controls the L.D.P. for a sequence (R_n) on U; then $I'(v) = \inf\{I(U) : u \in T^{-1}\{v\}\}$ controls the L.D.P. for the image measures $R'_n = R_n \circ T^{-1}$. If I is good, I' is also good.

Consider $(W_n)_{n\geq 1}$ an i.i.d. sequence of $E \times F$ valued random variables with common law P and define the empirical measure

$$\overline{W}_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{W_j}$$

whose law is denoted by $\overline{P}_n$.

If E and F are Polish spaces, the L.D.P. for $\overline{P}_n$ is given by Sanov's Theorem ([12] section 6.2) i.e.

(3.3) On $\mathcal{M}_1^+(E \times F)$ equipped with the weak topology, $\overline{P}_n$ satisfies a L.D.P.

with good rate function $I(Q) = H(Q, P)$ (the Kullback information).

Now, write $W_n = (X_n, Y_n)$ and consider the associated pair $(\overline{X}_n, \overline{Y}_n)$ which law is denoted by $\overline{\rho}_n$. $(\overline{X}_n, \overline{Y}_n)$ is the image of $(\overline{W}_n)$ by

$$
\begin{array}{rrcl}
(3.4) \qquad & T:\ \mathcal{M}_1^+(E \times F) & \rightarrow & \mathcal{M}^b(E) \times \mathcal{M}^b(F) \\
& Q & \rightarrow & \left(\int_F Q(\cdot, dy), \int_E Q(dx, \cdot)\right)
\end{array}
$$

which is continuous.

(3.5) Hence, by the contraction principle on $\mathcal{M}^b(E) \times \mathcal{M}^b(F)$ equipped with the product topology, $\overline{\rho}_n$ satisfies a L.D.P. with (good) rate function

$$I(\mu, \nu) = \inf\{H(Q, P) :\ Q \in \mathcal{M}_1^+(E \times F) \text{ and with marginals } \mu \text{ and } \nu\}.$$

(in particular if μ and ν are not Probability measures then $I(\mu, \nu) = +\infty$).

The idea is to obtain directly another expression of the rate function of $\overline{\rho}_n$. This can be done by using the general Dawson-Gärtner projective limit approach, as explained in [12] section 4.6.

(3.6) Indeed, introduce for a given Hausdorff measurable space U, $\mathcal{M}^\sharp(U)$, the algebraic dual of $C_b(U)$, equipped with the $\sigma(C_b(U), \mathcal{M}^\sharp(U))$ topology. On one hand, $\mathcal{M}^b(U)$ is embedded in $\mathcal{M}^\sharp(U)$, and the application T defined in (3.4) can still be viewed as a continuous function from $\mathcal{M}_1^+$ to $\mathcal{M}^\sharp(E) \times \mathcal{M}^\sharp(F)$, so that (3.5) is still true.

On the other hand, one can directly apply Corollary 4.6.11. of [12] in order to get

(3.7) on $\mathcal{M}^\sharp(E) \times \mathcal{M}^\sharp(F)$, $(\overline{\rho}_n)$ satisfies a L.D.P. with good rate function:

$$I(\mu, \nu) = \sup_{f \in C_b(E), g \in C_b(F)} \left(\langle f, \mu \rangle + \langle g, \nu \rangle - \log \int \exp(f \oplus g) dP\right).$$

Finally, recall that on a regular topological space, there exists an unique rate function for a given L.D.P. ([12] Lemma 4.1.4), and that a Hausdorff topological vector space is regular. We thus can conclude

$$
\begin{aligned}
& \inf\{H(Q, P) :\ Q \in \mathcal{M}_1^+(E \times F) \text{ with marginals } (\mu, \nu)\} \\
(3.8) \qquad & = \sup_{f \in C_b(E), g \in C_b(F)} \left(\int f d\mu + \int g d\nu - \log \int \exp(f \oplus g) dP\right)
\end{aligned}
$$

Consequently

Theorem 3.9 *If E and F are Polish spaces, $P \in \mathcal{M}_1^+$, there exists a $Q \in \mathcal{M}_1^+$ such that $H(Q,P) \leq K$ with marginals μ and ν, if and only if for all $f \in C_b(E)$ and all $g \in C_b(F)$*

$$\int f d\mu + \int g d\nu \leq \log \int \exp(f \oplus g) dP + K.$$

(Compare with Corollary 2.8).

Of course, the above method is limited to the entropic case because of the use of Sanov's Theorem. But, a similar idea can be used in all L_θ cases, replacing the empirical measure by a more sophisticated one introduced by Dacunha-Castelle and Gamboa [9], and used by Gamboa and Gassiat in various problems like moments problems [17] or superresolution [18]. The method now known as M.E.M. method (maximum entropy on the mean) is particularly described in terms of large deviations in [16]. We could directly use the results of [16], but they are written in the case of compact Polish spaces, and since we want to minimize our requirements on the topological structure, we shall give all the necessary proofs. Let we come to the method.
Let we consider the random measure on $E \times F$

$$\lambda_n = \frac{1}{n} \sum_{i=1}^{n} Z_i \delta_{(x_i, y_i)} \tag{3.10}$$

where $(Z_n)_{n\geq 1}$ is an i.i.d. sequence of non negative real random variables, with common distribution G, and the sequence $(z_n = (x_n, y_n))_{n\geq 1}$ is chosen such that:

$$P = \text{weak limit of } \frac{1}{n} \sum_{i=1}^{n} \delta_{(x_i, y_i)} \tag{3.11}$$

Thanks to the Glivenko-Cantelli Theorem (z_n) can be chosen, for instance, as almost every all realizations of an infinite sample of P. We then define

$$\psi_G(\tau) = \log \int_{\mathbb{R}_+} \exp(\tau\xi) G(d\xi) \ (\tau \in \mathbb{R}) \tag{3.12}$$

(3.13) $\gamma_G(\xi) = \psi_G^*(\xi) = \sup_{\tau \in \mathbb{R}} (\tau\xi - \psi_G(\tau)),\ (\xi \in \mathbb{R})$ its Legendre conjugate

and we shall assume in the sequel

(3.14) Domain $\psi_G = \mathbb{R}$ and G is not a Dirac mass.

Our aim will be now to prove a L.D.P. for the law R_n of λ_n, and for the laws $L_n = R_n \circ T^{-1}$, and then use the contraction principle in order to identify both rate functions, as we did in (3.8). The first results are obtained by using the projective limit approach as before

Proposition 3.15 *Assume that E and F are Hausdorff spaces and $P \in \mathcal{M}_1^+(E \times F)$.*

i) The laws R_n of λ_n satisfy on $\mathcal{M}^\sharp(E \times F)$ a L.D.P. with good rate function

$$\mathcal{I}_G(Q) = \sup_{l \in C_b(E\times F)} \left(\langle l, Q \rangle - \int \psi_G(l) dP \right)$$

ii) The laws $L_n = R_n \circ T^{-1}$ satisfy on $\mathcal{M}^\sharp(E) \times \mathcal{M}^\sharp(F)$ a L.D.P. with good rate function

$$I_G(\mu, \nu) = \sup_{f \in C_b(E), g \in C_b(F)} \left(\langle f, \mu \rangle + \langle g, \nu \rangle - \int \psi_G(f \oplus g) dP \right).$$

Proof
According to [12] Corollary 4.6.11 we first have to show that

$$J(l) = \lim_{n\to\infty} \frac{1}{n} \log \left(\int \exp(n\langle l, \lambda \rangle) R_n(d\lambda) \right)$$

exists as an extended real number, for $l \in C_b(E \times F)$. Define

$$\begin{aligned} J_n(l) &= \frac{1}{n} \log \left(\int \exp(n\langle l, \lambda \rangle) R_n(d\lambda) \right) \\ &= \frac{1}{n} \sum_{i=1}^{n} \psi_G\left(l(z_i)\right). \end{aligned}$$

By assumption (3.14), ψ_G is bounded on compact sets (since it is continuous on $\mathbb{R}$), and $\psi_G \circ l \in C_b$. So according to (3.11)

$$\lim_{n\to\infty} J_n(l) = \int \psi_G(l) dP < +\infty.$$

Furthermore, for all $l_1, \ldots, l_k \in C_b(E \times F)$

$$t_1, \ldots, t_k \to \int \psi_G \left(\sum_{i=1}^{k} t_i l_i \right) dP$$

is everywhere finite, continuous and everywhere differentiable thanks to (3.14). We thus can apply [12] Corollary 4.6.11 to conclude i). The proof of ii) is exactly the same. ■

Remark: the interested reader will find in [16] a more general result without assuming (3.14).

The expression of the rate function in the Proposition 3.15 are interesting if we are able to study their domains. Indeed, since R_n and L_n are supported by $\mathcal{M}_+^b$ (the set of positive bounded Radon measures), we know that the L.D.P. holds in this space, with the same rate function provided this function is infinite for all the others elements of $\mathcal{M}^\sharp$. We thus study the finiteness of the large deviations functional.

Lemma 3.16 *If $\mathcal{I}_G(Q)$ (resp. $I_G(\mu,\nu)$) is finite, then Q (resp. (μ,ν)) is a positive continuous linear form on $C_b(E\times F)$ (resp. $C_b(E)\times C_b(F)$).*

Proof
Since G is supported by $\mathbb{R}^+$, $\psi_G(\tau)\le 0$ for $\tau\le 0$. If there exists a $l\in C_b(E\times F)$ such that $l\le 0$ and $\langle l,Q\rangle>0$, then for all $t>0$

$$\mathcal{I}_G(Q)\ge t\langle l,Q\rangle-\int\psi_G(tl)dP\ge t\langle l,Q\rangle$$

which proves that Q is positive. Continuity is immediate since ψ_G is locally bounded. ■

In order to identify the positive continuous linear form on $C_b(E\times F)$ we need some topological assumptions.

Proposition 3.17 *If E and F are completely regular, and $\mathcal{I}_G(Q)$ is finite, Q is identified with a regular positive bounded measure $\check{Q}$ on the Stone-Cech compactification $E\check{\times}F$ of $E\times F$, and*

$$\mathcal{I}_G(Q)=\mathcal{I}_G(\check{Q})=\sup_{l\in C(E\check{\times}F)}\left(\int ld\check{Q}-\int\psi_G(l)d\check{P}\right)$$

where $\check{P}$ is the corresponding identification of P.

Proof
Since $E\times F$ is completely regular, $E\times F$ is homeomorphic to a dense subset of $E\check{\times}F$ and $C_b(E\times F)$ is isomorphic to $C(E\check{\times}F)$ (see e.g. [21]). By Riesz representation Theorem, any continuous positive linear form on $E\check{\times}F$ is a regular positive bounded Borel measure. The final equality comes from the identity $\int ldQ=\int\check{l}d\check{Q}$ if $\check{l}$ is the natural extension of $l\in C_b(E\times F)$ to $E\check{\times}F$, and the continuity of ψ_G. ■

We shall now give the key result of our construction.

Proposition 3.18 *Assume that U is a compact Hausdorff space, P a regular bounded positive measure on U. Then, for all regular bounded positive measure Q,*
$\mathcal{I}_G(Q)=\mathcal{J}_G(Q)$ where

$$\mathcal{I}_G(Q)=\sup_{l\in C_b(U)}\left(\langle l,Q\rangle-\int\psi_G(l)dP\right)$$

$$\mathcal{J}_G(Q)=\begin{cases}\displaystyle\int\gamma_G\left(\frac{dQ}{dP}\right)dP & \text{if } Q\ll P \text{ (recall (3.13))}\\ +\infty & \text{otherwise.}\end{cases}$$

Remark: Similar statements are contained in [29] and [30], in a slightly different context, but we prefer give a complete proof since this result is the key point of the construction. The following proof is essentially due to Gamboa and Gassiat (unpublished) with some simplifications due to (3.14).

Proof

1) $\mathcal{J}_G(Q) \geq \mathcal{I}_G(Q)$.
It is enough to check the above equality for $Q \ll P$. But, in this case

$$\mathcal{I}_G(Q) = \sup_{l \in C_b(U)} \left\{ \int \left(l\frac{dQ}{dP} - \psi_G(l) \right) dP \right\} \leq \int \sup_{l \in C_b(U)} \left(l\frac{dQ}{dP} - \psi_G(l) \right) dP = \mathcal{J}_G(Q).$$

2) $\mathcal{J}_G(Q) \leq \mathcal{I}_G(Q)$. Here again we may assume that $\mathcal{I}_G(Q) < +\infty$. Recall the following facts, which are consequences of (3.13), (3.14).

(3.19.i) ψ'_G is everywhere defined, increasing and continuous, with range $]\alpha, \beta[$ such that $\int_{\mathbb{R}^+} \xi dG(\xi) = m \in]\alpha, \beta[$. $\overline{]\alpha, \beta[}$ is the convex hull of the support of G.

(3.19.ii) For $\xi \in]\alpha, \beta[$ $\gamma_G(\xi) < +\infty$, whereas for $\xi \notin \overline{]\alpha, \beta[}$ $\gamma_G(\xi) = +\infty$. Hence, for any $\xi \in]\alpha, \beta[$ there exists $\tau \in \mathbb{R}$, with $\tau = \psi'^{-1}_G(\xi)$, and in this case

$$\gamma_G(\xi) = \tau \psi'_G(\tau) - \psi_G(\tau).$$

Let Q be a regular positive bounded measure on U, with Lebesgue's decomposition $Q = gP + S$ where $g \in L^1(P)$, and S is singular with respect to P. Q, P and S are regular. Denote by (A, A^c) a pair of disjoint Borel subsets of U such that $P(A) = S(A^c) = 0$, $P(A^c) = P(U) = 1$ and $S(A) = S(U)$.
For any $\epsilon > 0$ and $\eta > 0$ define a function h as follows

$$\begin{array}{llcl} & h(x) & = & \delta \text{ if } x \in A \\ (3.20) & h(x) & = & \delta \text{ if } x \in A^c \text{ and } g(x) > \beta \text{ or } g(x) < \alpha \\ & h(x) & = & \psi'^{-1}_G(g(x)) \text{ if } x \in A^c \text{ and } \alpha + \epsilon \leq g(x) \leq \beta - \epsilon \\ & h(x) & = & 0 \text{ otherwise.} \end{array}$$

h is bounded and measurable, and since P, Q, S are regular, one can find a sequence of equibounded continuous function $(h_n)_{n \geq 1}$ such that h_n converges to h, P, Q and S everywhere. Now,

$$\mathcal{I}_G(Q) \geq \int h_n dQ - \int \psi_G(h_n) dP = \theta_n(\delta, \epsilon).$$

We want to identify the limit of θ_n as n goes to infinity. According to (3.19) and Lebesgue's dominated convergence

$$\lim_{n \to \infty} \theta_n(\delta, \epsilon) = \delta S(A) + \int \{\delta g - \psi_G(\delta)\} \left(1_{g < \alpha} + 1_{g > \beta}\right) dP + \int \gamma_G(g) 1_{\alpha + \epsilon \leq g \leq \beta - \epsilon} dP.$$

Now, let $\delta \to \infty$. On the set $\{\xi : \ g(\xi) \notin \overline{]\alpha, \beta[}\}$, $\lim_{\delta \to +\infty}(\delta g(\xi) - \psi_G(\delta)) = +\infty$. We deduce that if $\mathcal{I}_G(Q) < +\infty$, $S(A) = 0$ ($Q \ll P$) and $\frac{dQ}{dP} \in \overline{]\alpha, \beta[}$, P-a.s. Hence, for all $\epsilon > 0$

$$\mathcal{I}_G(Q) \geq \int \gamma_G(g) 1_{\alpha + \epsilon \leq \frac{dQ}{dP} \leq \beta - \epsilon} dP,$$

and $\mathcal{I}_G(Q) \geq \mathcal{J}_G(Q)$ by monotone convergence Theorem. ■

Remark: The same proof shows that the statement of Proposition (3.18) remains true for any Hausdorff (non compact) space, just replacing $C(U)$ by $C_b(U)$.
We finally can state

Theorem 3.21 *Assume that E and F are completely regular and that $P \in \mathcal{M}_1^+(E \times F)$. Then*

i) The laws R_n of λ_n satisfy on $\mathcal{M}_+^b(E \times F)$ a L.D.P. with good rate function $\mathcal{I}_G(Q) = \int \gamma_G\left(\frac{dQ}{dP}\right) dP$ if $Q \ll P$, $+\infty$ otherwise.

ii) The laws $L_n = R_n \circ T^{-1}$ satisfy on $\mathcal{M}_+^b(E) \times \mathcal{M}_+^b(F)$ a L.D.P. with good rate function

$$I_G(\mu, \nu) = \sup_{f \in C_b(E), g \in C_b(F)} \left(\int f d\mu + g d\nu - \int \psi_G(f \oplus g) dP \right).$$

iii) For all $(\mu, \nu) \in \mathcal{M}_+^b(E) \times \mathcal{M}_+^b(F)$

$$I_G(\mu, \nu) = \inf_{Q \in M(\mu,\nu)} \mathcal{I}_G(Q)$$

where $M(\mu, \nu) = \{Q \in \mathcal{M}_+^b(E \times F) :$ with marginals μ and $\nu\}$.

Proof

i) We know (see Proposition 3.15) that R_n satisfy a L.D.P. on $\mathcal{M}^\sharp(E \times F)$ with a rate function $\mathcal{I}_G(Q)$ which is finite if and only if Q can be identified with a regular element $\check{Q}$ of $\mathcal{M}_+^b(E \check{\times} F)$, where $E \check{\times} F$ is the Stone-Cech compactification of $E \times F$ (see Lemma 3.16 and Proposition 3.17), and if $\check{Q}$ satisfies furthermore

$$\check{Q} \ll \check{P} \text{ and } \int \gamma_G \left(\frac{d\check{Q}}{d\check{P}} \right) d\check{P} < +\infty. \text{ (see 3.18).}$$

But we cannot immediately identify Q with $\frac{d\check{Q}}{d\check{P}}|_{E \times F} P$ (where $|_{E\times F}$ stands for the restriction to $E \times F$), because of measurability problems. However, since $\check{P}$ is regular, one can find a sequence $(\check{l}_n)_{n \geq 1}$ of $C(E \check{\times} F)$ which converges both in $L^1(\check{P})$ and $\check{P}$-a.s. to $\frac{d\check{Q}}{d\check{P}}$. If l_n denotes the restriction of $\check{l}_n$ to $E \times F$ (after identification of $C_b(E \times F)$ and $C(E \check{\times} F)$), we also know that for any $h \in C_b(E \times F)$

$$\int h l_n dP = \int \check{h} \check{l}_n d\check{P}$$

which proves that $l_n P$ is weakly convergent (in $\mathcal{M}^b(E \times F)$). But the sequence $(\check{l}_n)$ is uniformly integrable, so by the proof of La Vallée-Poussin's Theorem of [11], there exists a continuous Young function θ such that

$$\sup_n \int \theta(\check{l}_n) d\check{P} < +\infty.$$

The natural property of Stone-Cech Compactification, implies that $\theta(\check{l}_n) = \theta(\check{l}_n)$ and so $\sup_n \int \theta(l_n)dP < +\infty$; which shows that (l_n) is uniformly integrable (thus $\sigma(L^1, L^\infty)$ relatively compact by Dunford-Pettis Theorem), and consequently the weak limit of l_nP is of the form $Q = hP$ with $h \in L^1(P)$. It is now immediate that the initial $\check{Q}$ is associated to the above Q. In order to prove that $\int \gamma_G(\frac{dQ}{dP})dP < +\infty$, it suffices to approximate $\frac{dQ}{dP}1_{\alpha+\epsilon\leq g\leq\beta-\epsilon}$ (with α, β defined in (3.19)) by continuous functions, and use the Lebesgue's bounded convergence Theorem, then to pass to the limit via monotonic convergence as in the proof of Proposition 3.18. Finally, since $\mathcal{I}_G(Q)$ is finite only for $Q \in \mathcal{M}_+^b(E \times F)$, the L.D.P. holds in this space.

ii) and iii) are straightforward applications of the contraction principle (3.2.ii) and uniqueness of the rate function. ■

We shall in the sequel use Theorem 3.21 in the following form

Corollary 3.22 *Let E and F be completely regular topological spaces, $\mu \in \mathcal{M}_1^+(E)$, $\nu \in \mathcal{M}_1^+(F)$ and $P \in \mathcal{M}_1^+(E \times F)$. Then, there exists $Q \in \mathcal{M}_1^+(E \times F)$ such that Q has marginals μ and ν, and $\int \gamma_G(\frac{dQ}{dP})dP \leq K$ if and only if*

$$\text{for all } f \in C_b(E),\ g \in C_b(F),\ \int f d\mu + \int g d\nu \leq \int \psi_G(f \oplus g)dP + K.$$

Application: Let θ be a Young function. We can use the above criterion with $\theta = \gamma_G$, provided θ^* is everywhere defined and $\exp(\theta^*)$ is the Laplace transform of a Probability measure on $\mathbb{R}^+$. Instead of giving a full description of these θ's, we shall give some examples.

3.1 The L^q case $1 < q < +\infty$

Let choose for G the distribution of $Y^{\frac{1}{q}}$ for Y a random variable with gamma distribution $\gamma(\frac{1}{q}, 1)$ i.e.

$$dG(\xi) = \frac{q}{\Gamma\left(\frac{1}{q}\right)} \exp(-\xi^q) 1_{[0,+\infty[}(\xi)d\xi.$$

Then, ψ_G is defined on whole $\mathbb{R}$. Furthermore

Lemma 3.23 *There exists positive constants C_1, C_2 such that:*

1) *For $\tau \to \infty$ we have $\psi_G(\tau) \sim C_1\tau^{q'}$.*

2) *For $\xi \to \infty$ we have $\gamma_G(\xi) \sim C_2\xi^q$, where $\frac{1}{q} + \frac{1}{q'} = 1$.*

The proof of 1) is a straightforward application of Laplace's method, while 2) follows from general results about Legendre conjugacy. According to Corollary 3.22 and Lemma 3.23 we can state

Corollary 3.24 *In the situation of Corollary 3.22, there exists* $Q \in \mathcal{M}_1^+(E \times F)$ *such that* Q *has marginals* μ *and* ν*, which satisfies* $Q \ll P$ *and* $\frac{dQ}{dP} \in L^q(P)$ *if and only if for some* $K > 0$

$$\sup_{f \in C_b(E), g \in C_b(F)} \left(\int f d\mu + \int g d\nu - \int \psi_G(f \oplus g) dP \right) \leq K.$$

3.2 The entropic case

Let G be the Poisson distribution with mean 1. Then

$$\begin{aligned} \psi_G(\tau) &= \exp(\tau) - 1 \quad (\tau \in \mathbb{R}), \\ \gamma_G(\xi) &= \xi \log \xi - \xi + 1 \quad (\xi \geq 0), \\ &= +\infty \quad (\xi < 0). \end{aligned}$$

So that, $\mathcal{I}_G(Q) = H(Q, P)$ for $Q \in \mathcal{M}_1^+(E \times F)$ and

Corollary 3.25 *In the situation of Corollary 3.22, there exists* $Q \in \mathcal{M}_1^+(E \times F)$ *such that* Q *has marginals* μ *and* ν*, which satisfies* $H(Q, P) \leq K$ *(*$K > 0$*) if and only if*

$$\sup_{f \in C_b(E), g \in C_b(F)} \left(\int f d\mu + \int g d\nu - \int \exp(f \oplus g) dP \right) \leq K - 1.$$

(Compare with Theorem 3.9).

3.3 The L^∞ case

Though we cannot realize $\|\frac{dQ}{dP}\|_\infty$ as a $\gamma_G(\frac{dQ}{dP})$, we shall use Corollary 3.22 in the L^∞ framework.
Indeed, for $K > 0$ choose for G the Bernoulli distribution $G = \frac{1}{2}(\delta_0 + \delta_K)$, then

$$\begin{aligned} \psi_G(\tau) &= \log \frac{1 + \exp(K\tau)}{2}, \quad (\xi \in \mathbb{R}) \\ \gamma_G(\xi) &= \frac{\xi}{K} \log \left(\frac{\xi}{K} \right) + (1 - \frac{\xi}{K}) \log \left(1 - \frac{\xi}{K} \right) + \log 2, \text{ if } 0 \leq \xi \leq K, \qquad (3.26) \\ \gamma_G(\xi) &= +\infty \text{ otherwise.} \end{aligned}$$

Hence, if $\xi \in [0, K]$, then $0 \leq \gamma_G(\xi) \leq \log 2$ and $\gamma_G(\xi) = +\infty$ otherwise. In particular

$$\left\| \frac{dQ}{dP} \right\|_\infty \leq K \text{ if and only if } \int \gamma_G \left(\frac{dQ}{dP} \right) dP \leq \log 2.$$

We thus may apply Corollary 3.22 in order to get

Corollary 3.27 *In the situation of Corollary 3.22, there exists* $Q \in \mathcal{M}_1^+(E \times F)$ *with marginals* μ *and* ν*, which satisfies* $Q \ll P$ *and* $\left\| \frac{dQ}{dP} \right\|_\infty \leq K$ *if and only if*

$$\textit{for all } f \in C_b(E),\ g \in C_b(F),\ \int f d\mu + \int g d\nu \leq \int \log\left(1 + \exp K(f \oplus g)\right) dP.$$

The latest condition is equivalent to the following

(3.28) $\forall f \in C_b(E),\ g \in C_b(F),\ \int f d\mu + \int g d\nu \le K \int \log\left(1 + \exp(f \oplus g)\right) dP.$

We conclude this section by two remarks

- (3.29) Remark on the L^1 case:
  If for G we choose an exponential law with parameter 1, i.e

$$\begin{aligned}\psi_G(\tau) &= -\log(1-\tau) \text{ for } \tau < 1\\ &= +\infty \text{ otherwise.}\\ \gamma_G(\xi) &= \xi - 1 - \log(\xi) \text{ for } \xi > 0\\ &= +\infty \text{ otherwise,}\end{aligned}$$

  then, Proposition 3.15 is still available, since ψ_G is essentially smooth. But, all the next results can fail to hold. Actually one can show, that $\mathcal{I}_G(Q)$ is finite for some measures which Lebesgue's decomposition contains a singular part (with respect to P), see [16] for a large discussion on this phenomenon. Thus, criterion of Corollary 3.22 cannot be extented to the L^1 case.

- (3.30) Topological remark
  Assume that E and F are topological Hausdorff spaces, which are Borel isomorphic with some $\tilde{E}$ and $\tilde{F}$. If any bounded measure on $\tilde{E}$ (resp. $\tilde{F}$) is regular, we can apply Corollary 3.22 with the image measures $\tilde{\mu}, \tilde{\nu}, \tilde{P}$ (provided $\tilde{E}$ and $\tilde{F}$ are completely regular). This yields some $\tilde{Q}$ on $\tilde{E} \times \tilde{F}$, which gives us a Q on $E \times F$ satisfying similar requirements (Q is defined as the inverse image measure). In particular, this holds for Lusin spaces where $\tilde{E}$ (resp. $\tilde{F}$) can be chosen as a compact Polish space.

4 Remarks on a set theoretic formulation

Let we go back to Corollary 2.2 with $q = +\infty$ i.e.

(4.1) there exists Q such that $\left\| \frac{dQ}{dP} \right\|_\infty \le K$ and with marginals μ and ν, if and only if

$$\text{for all } f \in C_b(E),\ g \in C_b(F)\ \int f d\mu + \int g d\nu \le \sup_{0 \le Z \le K,\ \int Z dP = 1} \int (f \oplus g) Z dP.$$

The above inequality extends to $f = 1_A$ and $g = 1_B$ for $A \in \mathcal{B}(E)$, $B \in \mathcal{B}(F)$ and since $1_A \oplus 1_B \le 1 + 1_{A \times B}$, we get

(4.2) $$\mu(A) + \nu(B) \le 1 + KP(A \times B).$$

A remarkable fact noticed by Kellerer [24], Strassen ([34] Theorem 6) for Polish spaces, and more recently by Hansel and Troallic ([19] Theorem 4.1) for general measurable spaces is that (4.2) is actually a sufficient condition for the existence of a Probability measure Q with marginals μ and ν satisfying $Q \leq KP$. This latest condition implies $Q \ll P$ and $\frac{dQ}{dP} \leq K$, and (4.2) is thus a necessary and sufficient condition for the existence of $Q \in \Lambda_{\infty,K}$ with marginals μ and ν.

(4.3) **Remark:** A natural question is how to recover (4.2) from our new characterization (3.28). Here again we extend (3.28) to bounded Borel functions, and we choose

$$f = \delta(1_A - 1_{A^c}), \ \ g = \delta(1_B - 1_{B^c}).$$

Hence, we obtain for all $\delta > 0$

$$\mu(A) + \nu(B) \leq 1 + \frac{K}{2\delta}\Big(\log(1+e^{2\delta})P(A\times B) + \log(1+e^{-2\delta})P(A^c \times B^c) \\ + \log 2\,[P(A^c\times B) + P(A\times B^c)]\Big).$$

We thus recover (4.2) by letting δ go to $+\infty$.
The discussion above indicates how to try to get a set theoretic characterization in the general L_θ case. Indeed, if we apply the same idea as the one which leads to (4.2) we get

(4.4) If there exists $Q \in \mathcal{M}_1^+(E\times F)$ such that $\frac{dQ}{dP} \in L_\theta$ and with marginals μ and ν, then

$$\mu(A) + \nu(B) \leq 1 + K\eta[P(A\times B)]$$

for some K, with $\eta(u) = 1/\theta^{*-1}(1/u)$, θ^{*-1} being the reciprocal function of θ.

Unfortunatly, the above set condition is not sufficient to ensure the existence of Q. Here is a classical counter example

(4.5) **Example:** $E = \{0\}$, $F = \mathbb{N}^*$, $P = C\sum_{n=1}^{+\infty} 2^{-2n}\delta_{(0,n)}$, $\mu = \delta_0$ and $\nu = \sum_{n=1}^{+\infty} 2^{-n}\delta_n$, $\theta(x) = x^2$.
(4.4) reduces to $\nu(B) \leq K\sqrt{P(\{0\}\times B)}$ for all $B \in \mathcal{B}(\mathbb{N}^*)$.

Let j be the smallest element in B then

$$\nu(B) \leq 2\cdot 2^{-j} \text{ and } P[(0,B)] \geq C2^{-2j} \text{ i.e. } \nu(B) \leq \frac{\sqrt{P[(0,B)]C}}{2}.$$

But the only Q with marginals μ and ν is $P = \sum_{n=1}^{+\infty} 2^{-n}\delta_{(0,n)}$ which is such that $Q \ll P$ but $\frac{dQ}{dP} \notin L^2(P)$. (Of course we only used the fact that $Q(A) \leq K\sqrt{P(A)}$ does not imply $\frac{dQ}{dP} \in L^2(P)$ in general).
We thus should ask whether the new characterization of section 3 leads to interesting set theoretic inequalities. But

(4.6) If there exists $Q \in \mathcal{M}_1^+(E \times F)$ such that $\int \gamma_G(\frac{dQ}{dP}) < +\infty$ with marginals μ and ν then

$$\text{for all } \delta > 0,\ A \in \mathcal{B}(E),\ B \in \mathcal{B}(F)\ \ \mu(A) + \nu(B) \leq 1 + \frac{K}{2\delta} + \frac{\psi_G(2\delta)}{2\delta} P(A \times B).$$

But a short study of the right hand side of the above inequality, shows that both formulations (4.4) and (4.6) are almost equivalent an the gain is nul. Now, if we look at Strassen's proof (or in a similar way at Hansel and Troallic's one), one can easily see why (4.4) does not furnish a sufficient condition.

Because the computations are tedious in the general Orlicz case, we restrict ourselves to the L^q case ($1 < q < +\infty$) where the set condition in (4.4) becomes

$$\mu(A) + \nu(B) \leq 1 + K[P(A \times B)]^{1-\frac{1}{q}}. \tag{4.7}$$

It easy to prove the following

Lemma 4.8 *For $0 < \delta < 1$, the set function $C \to [P(C)]^\delta$ is a capacity (alternating of order 2 in the Choquet terminology used by Strassen).*

But in general one cannot find a kernel alternating of order 2 (see [34] p. 429) say H such that

$$[P(A \times B)]^{1-\frac{1}{q}} = \int_A H(x, B)\mu_0(dx)$$

where μ_0 denotes the first marginal of P. In the case $q = +\infty$, such a kernel is given by a regular desintegration of P (if for instance E and F are separable metric spaces, see [11] p. 128), thanks to the additivity of P; i.e. in the L^∞ case the situation is linear, and this linearity explains why Strassen's proof goes on.
In the L^q case ($1 < q < +\infty$), we shall however state a set theoretic characterization which is the analogous of (4.2) but is not so beautiful. To this end we first introduce some definitions.

Definition 4.9 *Let $(\Omega, \mathcal{F})$ be a measurable space.*

i) A partition $\mathcal{A}$ of Ω is a finite collection $A_1, \dots, A_n$ of $\mathcal{F}$ such that $\bigcup_{i=1}^n A_i = \Omega$ and $A_i \cap A_j = \emptyset$ if $i \neq j$.

ii) Let $\mathcal{A}$ and $\mathcal{A}'$ be two partitions of Ω. We say that $\mathcal{A}'$ is finer than $\mathcal{A}$ if for any $A' \in \mathcal{A}'$ there exists $A \in \mathcal{A}$ with $A' \subset A$.

iii) We call a partition core, a sequence $(\mathcal{A}_n)_{n\geq 1}$ such that $\mathcal{A}_{n+1}$ is finer than $\mathcal{A}_n$ for all n, and such that $\mathcal{F}$ is generated by $\bigcup_n \mathcal{A}_n$. (in particular if a partition core exists, $\mathcal{F}$ is separable, and conversely if $\mathcal{F}$ is separable a partition core always exists.)

iv) To any partition $\mathcal{A}$ corresponds the field $\sigma(\mathcal{A})$ generated by the elements of $\mathcal{A}$, and if P is a Probability measure on $(\Omega, \mathcal{F})$ we define

$$\begin{aligned} P(U|\mathcal{A})(\omega) &= P(U|A_i)(\omega) \text{ if } \omega \in A_i \ (A_i \in \mathcal{A}) \text{ and } P(A_i) \neq 0 \\ &= 0 \text{ if } \omega \in A_i \text{ and } P(A_i) = 0. \end{aligned}$$

($P(U|V)$ is of course the conditional Probability)

We can now state

Theorem 4.10 *Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be two measurable spaces. μ, ν, P are Probability measures defined respectively on $(E, \mathcal{E})$, $(F, \mathcal{F})$ and $(E \times F, \mathcal{E} \times \mathcal{F})$. μ_0 denotes the first marginal of P. Then, there exists a Probability measure Q on $(E \times F, \mathcal{E} \times \mathcal{F})$ such that $Q \ll P$ $\frac{dQ}{dP} \in L^q(P)$, $\|\frac{dQ}{dP}\|_q \leq K$ and with marginals μ and ν if and only if the following holds:*
Let $\mathcal{E}'$ (resp. $\mathcal{F}'$ be any separable sub σ-field of $\mathcal{E}$ (resp. $\mathcal{F}$). One can find a partition core $(\mathcal{A}_n)$ (resp. $(\mathcal{B}_n)$) of $\mathcal{E}'$ (resp. $\mathcal{F}'$) and a family $(Z_n)_{n\geq 1}$ of non negative random variables such that

i) Z_n is $\sigma(\mathcal{A}_n)$ measurable, $Z_n \in L^q(\mu_0)$ and $\|Z_n\|_q \leq K$.

ii) For all $A \in \sigma(\mathcal{A}_n)$ and $B \in \sigma(\mathcal{B}_n)$

$$\mu(A) + \nu(B) \leq 1 + \int_A Z_n(x)[P(E \times B|\mathcal{A}_n \times F)]^{1-\frac{1}{q}}(x)\mu_0(dx).$$

(where $P(\cdot|\mathcal{A}_n \times F)$ is defined in (4.9.iv)).

Corollary 4.11 *Assume that E and F are Polish spaces, $\mathcal{E} = \mathcal{B}(E)$ $\mathcal{F} = \mathcal{B}(F)$. Denote by $P(x, \cdot)$ a regular desintegration of P with respect to $\mathcal{B}(E)$ (considered as a sub σ-field of $\mathcal{B}(E) \times \mathcal{B}(F)$) i.e.*

$$P(A \times B) = \int_A P(x, B)\mu_0(dx) \text{ for } A \in \mathcal{B}(E),\ B \in \mathcal{B}(F),$$

such a $P(x, \cdot)$ is called a Markov kernel in [34].
Then, the necessary and sufficient condition of Theorem 4.10 is equivalent to the following

$$\mu(A) + \nu(B) \leq 1 + \int_A Z(x)[P(x, B)]^{1-\frac{1}{q}}(x)\mu_0(dx)$$

for some non negative $Z \in L^q(\mu_0)$ with $\|Z\|_q \leq K$.

Remark: If $q = +\infty$, the above condition is equivalent to Strassen's one, and the condition in Theorem 4.10 is also equivalent to (4.2). So we recover as well Theorem 6 of [34] or Theorem 4.1 in [19], but with a different proof for the latter case.

Proof of Theorem 4.10
For the if part we shall closely follow Strassen's method; Indeed, consider $(E, \sigma(\mathcal{A}_n))$,

$(F, \sigma(\mathcal{B}_n))$ and the restrictions of μ, ν, P, μ_0 to the corresponding fields. Actually these spaces are Borel isomorphic to finite discrete topological spaces (choose one point in each A_i (or B_j) of $\mathcal{A}_n$ (or $\mathcal{B}_n$)), which are of course Polish. So, as in the proof of Theorem 6 in [34] we may apply Theorem 4 of [34] in $(E, \sigma(\mathcal{A}_n))$, $(E, \sigma(\mathcal{B}_n))$ and $(E \times F, \sigma(\mathcal{A}_n) \otimes \sigma(\mathcal{B}_n))$. To this end consider

$$\begin{aligned} H_n(A_i, B) &= \min\left(\frac{\mu_0(A_i)}{\mu(A_i)} Z_n(i)[P(E \times B | A_i \times F)]^{1-\frac{1}{q}}, 1\right) \text{ if } \mu(A_i) \neq 0 \\ &= 0 \text{ if } \mu(A_i) = 0, \end{aligned}$$

for $A_i \in \mathcal{A}_n$, $B \in \sigma(\mathcal{B}_n)$ and $Z_n(i) =$ the value of Z_n on A_i.
Let $B \in \sigma(\mathcal{B}_n)$, and A the set where $H_n(\cdot, B) \leq 1$ (we define $H_n(x, B) = H_n(A_i, B)$ if $x \in A_i$). Condition (ii) in Theorem 4.10 yields

$$\begin{aligned} \nu(B) &\leq \mu(E - A) + \int_A Z_n(x)[P(E \times B | \mathcal{A}_n \times F)]^{1-\frac{1}{q}} \mu_0(dx) \\ &\leq \int_{E-A} 1 \mu(dx) + \sum_{i:\, A_i \subset A} Z_n(i)[P(E \times B | \mathcal{A}_n \times F)]^{1-\frac{1}{q}} \frac{\mu_0(A_i)}{\mu(A_i)} \mu(A_i) \\ &\leq \int H_n(x, B) \mu(dx). \end{aligned}$$

But, according to Lemma 4.8, $H_n(x, \cdot)$ is a kernel alternating of order 2 in the sense of [34]. Applying Theorem 4 of [34] as we said before we obtain that there exists a Markov kernel $q_n(\cdot, \cdot)$ defined on $\sigma(\mathcal{A}_n) \otimes \sigma(\mathcal{B}_n)$ such that $\nu = q_n \mu$ and $q_n(x, \cdot) \leq H_n(x, \cdot)$ for all $x \in E$ (we can choose $q_n = 0$ if $x \in A_i$ with $\mu(A_i) = 0$).
Define $Q_n = q_n \times \mu$. Q_n is a Probability measure on $(E \times F, \sigma(\mathcal{A}_n) \otimes \sigma(\mathcal{B}_n))$ with marginals μ and ν, and for $A_i \in \mathcal{A}_n$ and $B_j \in \mathcal{B}_n$;

$$\begin{aligned} Q_n(A_i \times B_j) &\leq Z_n(i)[P(E \times B_j | A_i \times F)]^{1-\frac{1}{q}} \mu_0(A_i) \\ &\leq [P(A_i \times B_j)]^{1-\frac{1}{q}} (\mu_0(A_i))^{\frac{1}{q}}. \end{aligned}$$

Hence, $Q_n \ll P$ in restriction to $(E \times F, \sigma(\mathcal{A}_n) \otimes \sigma(\mathcal{B}_n))$, and

$$\frac{dQ_n}{dP} = Z_n = \sum_{ij} \frac{Q_n(A_i \times B_j)}{P(A_i \times B_j)} 1_{A_i \times B_j} \qquad (\text{ by convention } \frac{0}{0} = 0).$$

It follows that

$$\int Z_n^q dP \leq \sum_i (Z_n(i))^q \mu_0(A_i) \leq K^q, \text{ i.e. } \|Z_n\|_q \leq K.$$

Now, consider the sequence $(Z_n)_{n \geq 1}$ as a sequence of random variables on $(E \times F, \mathcal{E}' \otimes \mathcal{F}')$. Since it is a bounded sequence of $L^q(P)$ (restricted to $\mathcal{E}' \otimes \mathcal{F}'$) one can use again the Dunford-Pettis Theorem (but here in its full power), in order to find a subsequence of Z_n which is $\sigma(L^1, L^\infty)$ convergent to a Z. It follows that $Z \in L^q(P)$, $\|Z\|_q \leq K$ (since L^∞ is dense in $L^{q'}$) and

$$\int_{A \times F} Z dP = \lim_{n \to \infty} \int_{A \times F} Z_n dP = \mu(A) \quad \text{for } A \in \bigcup_{p \geq 1} \mathcal{A}_p,$$

and

$$\int_{E\times B} Z dP = \lim_{n\to\infty} \int_{E\times B} Z_n dP = \nu(B) \quad \text{for } B \in \bigcup_{p\geq 1} \mathcal{B}_p$$

because $(\mathcal{B}_p)_{p\geq 1}$ is a partition core (the above sequence is stationary for n large enough). Now, consider the net of separable sub σ fields ordered by inclusion. To each $\mathcal{E}'$ is associated Z' as above, and again we may apply the Dunford-Pettis Theorem in $(E \times F, \mathcal{E} \times \mathcal{F})$, which says that the set of the Z' (indexed by the previous net) is relatively compact in $\sigma(L^1, L^\infty)$. Take any limit point Z of this net. Then $Z \in L^q(P)$ and $\|Z\|_q \leq K$. The Probability measure $Q = ZP$, of course has marginals μ and ν. The only if part is immediate, with $Z_n(i) = (E[Z^q|A_i \times F])^{\frac{1}{q}}$ for $A_i \in \mathcal{A}_n$ and $Z = \frac{dQ}{dP}$, by using Hölder's conditional inequality. ∎

Proof of Corollary 4.11

The only if part holds with $Z(x) = (\int Z^q(x,y)P(x,dy))^{\frac{1}{q}}$ as above. For the if part, it suffices to mimic the proof of Theorem 4.10 without the final argument since the σ-fields are separable (eventually up to negligible sets which are not relevant). ∎

In order to extend these results to general Orlicz spaces, one essentially needs to check Lemma 4.8 in the situation of (4.4) (i.e. with $1/\theta^{*-1}(1/u)$). Finally, in the L^1 case, one can ask about the following conjecture:

(4.12) There exists $Q \ll P$ with marginals μ and ν if and only if for all $\epsilon > 0$ there exists a $\eta > 0$ such that $\mu(A) + \nu(B) - 1 \geq \epsilon$ implies $P(A \times B) \geq \eta$.

At the present state we do not have any feeling on the exactness of the above conjecture.

5 About particular realizations

In the first three sections we were only interested in finding necessary and sufficient conditions for the existence of Q belonging to $\Lambda(\mu,\nu,\theta)$ where

(5.1) $$\Lambda(\mu,\nu,\theta) = \left\{ Q \in \mathcal{M}_1^+(E\times F) \text{ with marginals } \mu,\nu, \text{ such that } \frac{dQ}{dP} \in L_\theta \right\}.$$

Since $\Lambda(\mu,\nu,\theta)$ intersected with the L_θ ball of radius K is weakly compact (at least if E and F are completely regular), and when θ is strictly convex, the set $\Lambda(\mu,\nu,\theta)$ contains a smallest element i.e.

(5.2) there exists $Q^* \in \Lambda(\mu,\nu,\theta)$ such that $\|\frac{dQ^*}{dP}\|_\theta < \|\frac{dQ}{dP}\|_\theta$ for $Q \neq Q^*$ and $Q \in \Lambda(\mu,\nu,\theta)$.

In this section, we shall ask about particular properties of Q^*. Without further mention, we shall assume that

(5.3) E, F are completely regular, equipped with $\mathcal{B}(E)$, $\mathcal{B}(F)$; μ, ν, P are Radon.

5.1 The entropic case

For reasons we shall explain, the entropic case has already been extensively studied. By entropic case we mean

(5.4) $\Lambda(\mu,\nu,H) = \{Q \in \mathcal{M}_1^+(E\times F)$ with marginals $\mu,\nu,$ and $H(Q,P) < +\infty\}$

(see (2.5) for the definition of H).
But unfortunately some mistakes have been done in many papers on the subject.
The first result concerning the minimal Q^* in $\Lambda(\mu,\nu,H)$ was stated by Csiszár [8] Theorem 3.1 and Corollary 3.1 as follows

(5.5) let Q^* be the minimal element in $\Lambda(\mu,\nu,H)$. Then, there exists $N \in \mathcal{B}(E)\otimes\mathcal{B}(F)$ such that $Q(N^c)=0$ for all $Q\in\Lambda(\mu,\nu,H)$ and

$$\frac{dQ^*}{dP} = \exp\left(a(x)\oplus b(x)\right)1_N(x,y)$$

for some $a \in L^1(E,\mathcal{B}(E),\mu_0)$ and $b\in L^1(F,\mathcal{B}(F),\nu_0)$, μ_0 and ν_0 being the marginals of P. In particular, if one can find a Q in $\Lambda(\mu,\nu,H)$ which is equivalent to P, then N^c is P-negligible and one can supress the indicator function 1_N, i.e. $\frac{dQ^*}{dP}$ splits into $\exp a(x)\otimes\exp b(x)$.

The proof of Csiszár used the following simple idea: to fix the marginals is equivalent to fix an infinite collection of moments (for which the splitting condition is easy to prove) and then use the closedness of $L^1(\mu_0)\oplus L^1(\nu_0)$ in $L^1(P)$. But unfortunately, in general the space $L^1(\mu_0)\oplus L^1(\nu_0)$ is not closed in $L^1(P)$; (a counter example is due to Lindenstauss [28]), and the above result is thus, not completely true.
The above closedness problem is treated in [13], [31] or [2] for instance in a slightly weakened way. These authors obtained sufficient conditions for any element of the $L^1(P)$ closure of $L^1(\mu_0)\oplus L^1(\nu_0)$, to split into $h=f\oplus g$ where f (resp. g) is $\mathcal{B}(E)$ (resp. $\mathcal{B}(F)$) measurable. We shall recall some of the results contained in [31] and [4] (this last paper also contains with a different approach, and in a special case, the variational characterization of section 3 in the entropic case).

Theorem 5.6 *[see [31] Proposition 2] Assume that $P \ll \mu_0\otimes\nu_0$. Then, any h in the closure of $L^1(\mu_0)\oplus L^1(\nu_0)$ splits into $f\oplus g$, where $f\in\mathcal{B}(E)$, $g\in\mathcal{B}(F)$.*

Remark: This Theorem is obtained in a more general context in [31]. For a less general version also see [4] Theorem 4.2, which is based on the results of [2].

Theorem 5.7 *[see [31] Proposition 1] Let $P(x,dy)$ be a regular desintegration of P with respect to $\mathcal{B}(E)$ (which exists if for instance E and F are Polish spaces). If there exists a $\mathcal{B}(E)$ measurable function α with values in $[0,1]$ such that*

$$P(x,dy) \geq \alpha(x)\nu_0(dy) \text{ and } \mu_0(\{x:\ \alpha(x)>0\}) > 0.$$

Then, $L^1(\mu_0) \oplus L^1(\nu_0)$ is $L^1(P)$ closed.

Remark: Also see [4] Theorem 4.8 and Corollary 4.9, where E and F are assumed to be compact and $P = k(x,y)dx \otimes dy$ for some Probability measures dx, dy.

One of the interest of the splitting decomposition of Q^* is the following remark due to Föllmer [15] (we follow here Remark 1 of [31]).

(5.8) Assume that $P = k(x,y)\mu_0 \otimes \nu_0$, and that $\Lambda(\mu,\nu,H)$ is non void. Then (by the contraction property of the Kullback information), $H(\mu,\mu_0)$ and $H(\nu,\nu_0)$ are also finite. In particular, $\frac{d\mu}{d\mu_0} = \alpha$ and $\frac{d\nu}{d\nu_0} = \beta$ for some non negative L^1 functions. Furthermore, $\frac{dQ^*}{dP} = \exp(a(x) \oplus b(x))$ and

$$\begin{aligned} \exp(a(x)) \int_F k(x,y)(\exp b(y)) d\nu_0(y) &= \alpha(x) \\ \exp(b(y)) \int_E k(x,y)(\exp a(x)) d\mu_0(x) &= \beta(y). \end{aligned}$$

The above pair of equations was introduced by Schrödinger [32], and studied by many people (see the references in [4] or [22]). They are strongly related to some probabilistic approaches of the Schrödinger equation (see [22], [35]), and the Markov property of reciprocal processes (see [22] and [14]).

Though this connection is the origin of the present paper, we do not enter further into details. Just remark that one cannot use Strassen's initial Theorem, in order to prove the existence of a solution (a,b) of (5.8), since the closed convex hull of measures $\exp(a(x) \oplus b(y))P$ is too big.

5.2 The logistic case

To end this section we give a Theorem setting the existence and characterizing the Probability Q^* having marginals μ and ν and minimizing the functional $\mathcal{I}_G$ introduced in section 3.3. That is, $\mathcal{I}_G(Q)$ is defined for $P \ll Q$ with $0 \leq \frac{dQ}{dP} \leq K$ by

$$\mathcal{I}_G(Q) = \int \left(\frac{1}{K}\frac{dQ}{dP} \log\left(\frac{1}{K}\frac{dQ}{dP}\right) + \left(1 - \frac{1}{K}\frac{dQ}{dP}\right) \log\left(1 - \frac{1}{K}\frac{dQ}{dP}\right)\right) + \log 2$$

and is infinite otherwise.

Theorem 5.9 *Assume that E and F are compact Polish spaces and that $P \ll \mu_0 \times \nu_0$ with $\frac{dP}{d(\mu_0\times\nu_0)} \geq \xi$ $P-a.s.$ for some $\xi > 0$. Assume further that there exists a continuous function h on $E \times F$ such that:*

(5.10.i) For some $\epsilon > 0$, $\epsilon < h < K - \epsilon$ $P-a.s.$

(5.10.ii) hP has marginals μ, ν.

Then, there exists functions $f^* \in L^1(\mu_0)$, $g^* \in L^1(\nu_0)$ such that setting $Q^* = \frac{KP}{1+\exp[-K(f^*\oplus g^*)]}$ Q^* has marginals μ, ν and

(5.11.i) Any $Q \neq Q^*$ having marginals μ, ν satisfies $\mathcal{I}_G(Q) > \mathcal{I}_G(Q^*)$.

(5.11.ii) $\mathcal{I}_G(Q^*) = I_G(\mu, \nu)$.

Lemma 5.12 *Assume that the assumption made on P in Theorem 5.9 holds. Let $f \in L^1(\mu_0)$ and $g \in L^1(\nu_0)$. Then, there exists a constant C_1 depending on f and g such that*

$$\|f \oplus g\|_1 \geq \xi \|f + C_1\|_1 \text{ and } \|f \oplus g\|_1 \geq \xi \|g - C_1\|_1.$$

Remark
When E and F are compact Polish spaces, Lemma 5.12 implies that any weak accumulation point lying in $L^1(P)$ of a bounded sequence $(f_n \oplus g_n)$ of $L^1(P)$ may be written as $f \oplus g$ (compare with Theorem 5.6 and 5.7).

Proof of Lemma 5.12
Choose $C_1 \in \mathbb{R}$, with $\forall C \in \mathbb{R}$

$$\|f + C_1\|_1 + \|g - C_1\|_1 \leq \|f + C\|_1 + \|g - C\|_1.$$

Then,

$$\begin{aligned}\|f \oplus g\|_1 &= \int |f \oplus g| \frac{dP}{d(\mu_0 \otimes \nu_0)} d(\mu_0 \otimes \nu_0) \\ &= \int [| f \oplus g| \frac{dP}{d(\mu_0 \otimes \nu_0)} d\mu_0] d\nu_0 \geq \xi \|f + C_1\|_1. \blacksquare\end{aligned}$$

Proof of Theorem 5.9
Define for $\sigma \in \mathcal{M}^b(E \times F)$ the functional

$$(5.13) \quad \Xi(\sigma) = \int \log(1 + e^{Kl}) dP + K\tilde{\sigma}^+(E \times F) - \int lh dP - \int h d\tilde{\sigma} - \log 2,$$

where $\sigma = lP + \tilde{\sigma}$ and $\tilde{\sigma} = \tilde{\sigma}^+ - \tilde{\sigma}^-$ are respectively the Lebesgue decomposition of σ and the Jordan decomposition of $\tilde{\sigma}$.
$\Xi(\sigma)$ may be rewritten as

$$(5.14) \qquad \Xi(\sigma) = \int \log(1 + e^{-K|l|}) dP + \int (K - h) d\sigma^+ - \int h d\sigma^- - \log 2,$$

($\sigma = \sigma^+ - \sigma^-$ is the Jordan decomposition of σ).

(5.15) From the last expression of Ξ we deduce easily that $\lim_{\|\sigma\|_V \to +\infty} \Xi(\sigma) = +\infty$

(here $\|\sigma\|_V = \sigma^+(E\times F)+\sigma^-(E\times F)$ denotes the total variation of σ).

(5.16) $\quad\Xi$ has the dual representation (see [3] Theorem 2.7 or [29] Theorem 5)

$$\Xi(\sigma) = \sup_{f\in C_b(E\times F)} \left(\int (f-h)d\sigma - \int \gamma_G(f)dP\right),$$

where γ_G is defined in (3.26). So, we may conclude that Ξ is lower semicontinuous for the weak topology.

Define now for $f\in L^1(\mu_0)$, $g\in L^1(\nu_0)$ and $\tilde\sigma\in\mathcal{M}^b(E\times F)$, $\tilde\sigma\perp P$ the functional

$$M(f,g,\tilde\sigma) = \Xi\left((f\oplus g)P+\tilde\sigma\right).$$

M is actually the lower semicontinuous closure of the dual objective function associated with the computation of $I_G(\mu,\nu)$. From (5.15) we may construct sequences (f_n) of $L^1(\mu_0)$, (g_n) of $L^1(\nu_0)$ and $(\tilde\sigma_n)$ of $\mathcal{M}^b(E\times F)$, (with $\tilde\sigma_n\perp P$) such that the sequence $((f_n\oplus g_n)P+\tilde\sigma_n)$ converges weakly to some σ^* with Lebesgue decomposition $\sigma^* = l^*P+\tilde{\sigma^*}$ and such that

$$\lim_{n\to\infty} M(f_n,g_n,\tilde\sigma_n) = \inf_{f,g,\tilde\sigma} M(f,g,\tilde\sigma).$$

Using Lemma 5.12 we may also assume that $(f_n\mu_0)$ (resp. $g_n\nu_0)$) converges weakly to some $\mu^* = f^*\mu_0+\hat\sigma^*$ with $\hat\sigma^*\perp\mu_0$ (resp. $\nu^* = f^*\mu_0+\check\sigma^*$ with $\check\sigma^*\perp\mu_0$).
Now, from (5.14) $\Xi(\sigma^*)\geq\Xi(l^*P)$ and obviously $l^* = f^*\oplus g^*$. Hence

$$M(f^*,g^*,0) = \inf_{f\in L^1(\mu_0),g\in L^1(\nu_0),\tilde\sigma} M(f,g,\tilde\sigma).$$

Setting now $Q^* = \frac{KP}{1+\exp[-K(f^*\oplus g^*)]}$ a direct computation leads to $\mathcal{I}_G(Q^*) = M(f^*,g^*,0)$, and as $M(f,g,0)$ is continuous on $L^1(\mu_0)\times L^1(\mu_0)$ (for the strong topology) we also have $I_G(\mu,\nu) = \inf_{f\in L^1(\mu_0),g\in L^1(\nu_0),\tilde\sigma} M(f,g,\tilde\sigma)$ so that (5.11.ii) holds. To conclude it remains to show that Q^* has marginals μ,ν this can be easily made by classical variation arguments considering for $f\in C_b(E)$ and $g\in C_b(F)$ the following limits:

$$\lim_{z\to 0^\pm}\frac{1}{z}\left(M(f^*,g^*,0)-M(f^*+zf,g^*,0)\right)\qquad \lim_{z\to 0^\pm}\frac{1}{z}\left(M(f^*,g^*,0)-M(f^*,g^*+zg,0)\right).$$

■

Acknowledgements The first author wishes to thank Hans Föllmer for many fruitful discussions on Schrödinger's problem, and invitation at Bonn where the foundation stone of this work was laid (en français dans le texte).

References

[1] Astérisque 68. *Grandes déviations et applications statistiques.* S.M.F., 1979.

[2] Borwein J M, Lewis A S. Decomposition of multivariate functions. *Canadian J. of Math.*, 44:1–20, 1992.

[3] Borwein J M, Lewis A S. Partially-finite programming in L_1 and the existence of maximum entropy estimates. *S.I.A.M. J. Optimization*, 3:248–267, May 1993.

[4] Borwein J M, Lewis A S, Nussbaum R D. Entropy minimization, *DAD* problems and doubly-stochastic kernels. *J. Func. Anal.*, 123:264–307, 1994.

[5] Cattiaux P, Léonard C. Minimization of Kullback information for diffusion processes. *Annales de l'I.H.P. Proba.*, 30:83–132, 1994.

[6] Cattiaux P, Léonard C. Larges deviations and Nelson processes. *Forum Mathematicum*, to appear, 1995.

[7] Cattiaux P, Léonard C. Minimization of Kullback information of some Markov processes. Preprint, 1995.

[8] Csiszár I. *I*-divergence geometry of probability distributions and minimization problems. *Annals of Proba.*, 3:146–158, 1975.

[9] Dacunha-Castelle D, Gamboa F. Maximum d'entropie et problème des moments. *Annales de l'I.H.P. Proba.*, 26:567–596, 1990.

[10] Dawson D.A., J. Gärtner. Large deviations from the Mac Kean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20:247–308, 1987.

[11] Dellacherie C, Meyer P A. *Probabilités et potentiel.* Herman Paris, 1975.

[12] Dembo A, Zeitouni O. *Large deviations techniques and applications.* Jones and Barblett pub. Boston, 1993.

[13] Donsker M S, Varadhan S R S. Asymptotic evaluation of certain Markov process expectations for lage deviations iii. *Comm. Pure. Appl. Math.*, 29:389–461, 1974.

[14] Föllmer H. Private communication.

[15] Föllmer H. *Random fields and diffusion processes Summer school of St Flour L.N.M. 1362.* Springer Verlag, 1988.

[16] Gamboa F, Gassiat E. Bayesian methods for ill posed problems. Preprint, Orsay, 1994.

[17] Gamboa F, Gassiat E. The maximum entropy method on the mean: Applications to linear programming and superresolution. *Mathematical Programming, Series A*, 66:103–122, 1994.

[18] Gamboa F, Gassiat E. Sets of superresolution and the maximum entropy method on the mean. *S.I.A.M. Journal on Mathematical Analysis*, (to appear) 1995.

[19] Hansel G, Troallic J P. Sur le problème des marges. *Prob. Th. and Rel. Fields*, 71:357–366, 1986.

[20] Hoffmann-Jorgensen J. *Probability in Banach spaces. Summer school of St Flour L.N.M. 598.* Springer Verlag, 1977.

[21] Jameson G J. *Topology and normed spaces.* Chapman and Hall, 1974.

[22] Jamison. Reciprocal processes. *Z. Wahrsch. v. Gebiete*, 153:65–86, 1974.

[23] Kellerer H G. Funktionen auf produkt raümen mit vorgegebenen marginalfunktionen. *Math. Ann.*, 144:323–344, 1961.

[24] Kellerer H G. Masstheoretische marginalprobleme. *Math. Ann.*, 153:168–198, 1964.

[25] Kellerer H G. Schnittmass funktionen in mehrfachen produktraümen. *Math. Ann.*, 155:369–391, 1964.

[26] Kellerer H G. Duality theorems for marginal problems. *Z. Wahrsch. v. Gebiete*, 67:399–432, 1984.

[27] Krasnosel'skii M A, Rutickii Y B. *Convex functions and Orlicz spaces.* Noordhoff Ltd, 1961.

[28] Lindenstrauss J. A remark on extreme doubly stochastic measures. *Amer. Math. Monthly*, 72:379–382, 1965.

[29] Rockafellar R T. Integrals which are convex functionals. *Pacific Jour. of Math.*, 24:525–539, 1968.

[30] Rockafellar R T. Integrals which are convex functionals. *Pacific Jour. of Math.*, 39:439–469, 1971.

[31] Rüschendorff L, Thomsen W. Note on the Schrödinger equation and I-projections. Preprint, 1993.

[32] Schrödinger. Uber die umkehrung das naturgesetze sitzungberichte preuss. *Phys. Math.*, 144:144–153, 1931.

[33] Skala H J. Existence of probability measures with given marginals. *Annals of Prob..*, 21:136–142, 1993.

[34] Strassen V. Existence of probability measures with given marginals. *Annals of Math. Statist.*, 36:423–439, 1965.

[35] Zambrini J.C. *Calculus of variations and quantum probability.* Lecture notes in Control and Inform. Sci, 121 Springer Berlin, 1989.