Bernoulli **5**(1), 1999, 81–108

Large deviations and variational theorems for marginal problems

PATRICK CATTIAUX[1] and FABRICE GAMBOA[2*]

[1]*Ecole Polytechnique, CMAP, F-91128 Palaiseau Cedex, France. e-mail: cattiaux@paris.polytechnique.fr*
[2]*Laboratoire de Statistiques Université Paris Sud, F-91405 Orsay, France. e-mail: Fabrice.Gamboa@math.u-psud.fr*

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 due to Strassen, are obtained by using large-deviations methods. We also study the minimal realizations of such Q.

Keywords: large deviations; marginal problems

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 whether it is *possible to find a probability measure Q on $E \times F$, with marginals μ and ν, such that $Q \ll P$ and* $\mathrm{d}Q/\mathrm{d}P$ *satisfies some integrability conditions*. The construction of measures on a product space, with given marginals and satisfying convex constraints, is an old problem. A celebrated result due to Strassen (1965, Theorem 7; see Theorem 2.1 below) gives a necessary and sufficient variational condition of existence. In Section 2, we explain how to use Strassen's result in order to answer our question.

Since the constraint is here implicit (before building Q, one cannot control $\mathrm{d}Q/\mathrm{d}P$), the usual duality results (as in Kellerer 1984) do not hold. However, following Cattiaux and Léonard (1995a) or Gamboa and Gassiat (1997) – who deal respectively with marginal flows and 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 (Corollary 3.10). We emphasize that the method can be extended to more general product spaces (for instance $C^0([0, 1], E)$ considered as a subspace of $E^{[0,1]}$). This will be done elsewhere.

In Section 4, we give an alternate set-theoretic characterization (see Theorem 4.5) in the

*To whom correspondence should be addressed

spirit of Strassen's result (Strassen 1965, Theorem 6) and many others (see, for example, Hansel and Troalllic 1986, Theorem 4.1).

In Sections 5 and 6 the issue of minimal realizations of our problem (minimal for an Orlicz norm, for instance) is addressed. In Section 5 we show that the minimal element $\mathrm{d}Q^*/\mathrm{d}P$ is suitably approximated by nice functions (belonging to the subgradient of the related log-Laplace transform). In Section 6 we discuss the form of this limit. Applying closedness results of Rüschendorf and Thomsen (1994) it is shown (Proposition 6.2) that $\mathrm{d}Q^*/\mathrm{d}P$ 'almost' belongs to the same set. In the entropic case this leads to a new interpretation of Beurling's (1960) result on an old question posed by Schrödinger (1931). We emphasize that Föllmer was the first to link Schrödinger's question to an entropy minimization problem.

2. Notation and first results

Let E and F be two topological spaces equipped with their Borel σ-fields, $\mathscr{B}(E)$ and $\mathscr{B}(F)$, and two probability measures μ and ν defined respectively on $(E, \mathscr{B}(E))$ and $(F, \mathscr{B}(F))$. An old problem is whether there exists a probability measure Q on the product space $(E \times F, \mathscr{B}(E) \otimes \mathscr{B}(F))$, belonging to a certain subset Λ and with marginals μ and ν. Following on from several results in particular cases (see, for example, Kellerer 1961; 1964a; 1964b), Strassen's (1965) Theorem 7 stated a nice necessary and sufficient variational condition.

Theorem 2.1. *Assume that E and F are Polish spaces, and that Λ is a non-empty weakly closed convex subset of $\mathscr{M}_1^+(E \times F)$, 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, for all $f \in C_\mathrm{b}(E)$, $g \in C_\mathrm{b}(F)$,*

$$\int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\nu \leqslant \sup_{Q' \in \Lambda} \int f \oplus g\,\mathrm{d}Q',$$

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 Hoffmann-Jørgensen (1977) and to general Hausdorff spaces in Skala (1993), assuming in both cases that μ and ν are Radon, and replacing in Skala (1993) the classical weak topology on Radon bounded measures by the narrow topology, and C_b by B_b (the set of Borel bounded functions). In this paper we denote by $\mathscr{M}^\mathrm{b}(U)$ ($\mathscr{M}_+^\mathrm{b}(U)$, $\mathscr{M}_1^+(U)$) the set of all bounded Radon (positive bounded Radon, probability Radon) measures on $(U, \mathscr{B}(U))$, where U is a topological Hausdorff space equipped with its Borel σ-field. The weak topology on $\mathscr{M}^\mathrm{b}(U)$ is the one induced by the embedding $\mathscr{M}^\mathrm{b}(U) \to C_\mathrm{b}^*(U)$, where C_b^* is the topological dual space of C_b, the space of real-valued bounded functions.

Recall, for a positive measure P, that Radon means

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

for all Borel sets A; and, for a signed measure P, that P^+ and P^- are Radon (see Dellacherie and Meyer 1975).

The special case of interest in this paper is the one where

$$\Lambda = \left\{ Q\colon Q \ll P \text{ and } \frac{\mathrm{d}Q}{\mathrm{d}P} \in \Gamma \right\}$$

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

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

$$\Gamma^1_{q,K} = \Gamma_{q,K} \cap \left\{ Z \geqslant 0, \int Z \,\mathrm{d}P = 1 \right\}.$$

Then, for $1 < q \leqslant +\infty$, there exists $Q \in \mathscr{M}_1^+(E \times F)$ such that $Q \ll P$, $\mathrm{d}Q/\mathrm{d}P \in \Gamma_{q,K}$ and with marginals μ and ν if and only if, for all $f \in C_\mathrm{b}(E)$, $g \in C_\mathrm{b}(F)$,

$$\int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\nu \leqslant \sup_{Z \in \Gamma^1_{q,K}} \int (f \oplus g) Z \,\mathrm{d}P.$$

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

Although the fact that $\Lambda_{q,K}$ is weakly closed is certainly well known, we included the above proof in order to extend the result to the larger class of Orlicz spaces, which are less well known. Let $L_\theta(P)$ denote the Orlicz space associated with the Young function θ and $P \in \mathscr{M}_+^\mathrm{b}$. Denote by θ^* the Legendre conjugate function of θ, and by E_θ the (L_θ) closure of C_b – recall that $E_\theta = L_\theta$ once θ is moderate (i.e. satisfies Δ_2-regularity in Orlicz space terminology; see Rao and Ren 1991, pp. 22 and 77).

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 appearance of a factor 2 in the Hölder–Orlicz inequality. Indeed, Z belongs to $(E_{\theta^*})^* = L_\theta$ (see Rao and Ren 1991, p. 110), but $\|Z\|^*_{\theta_*} \leqslant 2K$, i.e. $\|Z\| \leqslant 2K$. So $\Lambda_{\theta,K}$ is not clearly weakly closed, but nevertheless we can state the following corollary.

Corollary 2.3. *With the same assumptions as in Corollary 2.2, denote by $\Gamma_{\theta,K}$ the closed ball in $L_\theta(P)$ of radius K. Then if, for all $f \in C_\mathrm{b}(E)$, $g \in C_\mathrm{b}(F)$,*

$$\int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\nu \leqslant \sup_{Z \in \Gamma^1_{\theta,K}} \int (f \oplus g) Z \,\mathrm{d}P, \tag{2.1}$$

then there exists $Q \in \mathcal{M}_1^+(E \times F)$ *such that* $Q \ll P$ *and* $\|\mathrm{d}Q/\mathrm{d}P\|_\theta \leqslant 2K$. *(The converse statement, without the factor 2, is obvious.)*

Proof. Denote by $\overline{\Lambda_{\theta,K}}$ the weak closure of $\Lambda_{\theta,K}$. Of course, (2.1) implies that

$$\int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\nu \leqslant \sup_{Q \in \overline{\Lambda_{\theta,K}}} \int (f \oplus g) \,\mathrm{d}Q.$$

But, as in the proof of Corollary 2.2, any $Q \in \overline{\Lambda_{\theta,K}}$ satisfies $Q \ll P$. Hence, we may conclude using the remark preceding the corollary. □

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

$$\Gamma^1_{\theta,K} = \Gamma_{\theta,K} \cap \left\{ Z\colon \int Z \,\mathrm{d}P = 1 \text{ and } Z \geqslant 0 \right\}$$

induces a weakly closed set $\Lambda_{\theta,K}$; and in the above derivation we did not use the fact that Z is a probability density. We 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–Leibler information of Q (relative to P),

$$H(Q, P) = \begin{cases} \displaystyle\int \frac{\mathrm{d}Q}{\mathrm{d}P} \log\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\right) \mathrm{d}P & \text{if } Q \ll P \text{ and } \log\left(\dfrac{\mathrm{d}Q}{\mathrm{d}P}\right) \in L^1(Q) \\ +\infty & \text{otherwise.} \end{cases} \tag{2.2}$$

It is easy to see that

$$\left\| \frac{\mathrm{d}Q}{\mathrm{d}P} - 1 \right\|_\theta \leqslant H(Q, P) + 1.$$

But, since Q and P are inner regular (which implies that $C_\mathrm{b}(E \times F)$ is dense in L^1 for each), the following alternative expression of H is known (see Astérisque, 1979, p. 36–37):

$$H(Q, P) = \sup_{f \in C_\mathrm{b}(E \times F)} \left(\int f \,\mathrm{d}Q - \log \int \exp(f) \,\mathrm{d}P \right). \tag{2.3}$$

The above form shows that

$$\Lambda_{H,K} = \{ Q \in \mathcal{M}_1^+(E \times F)\colon H(Q, P) \leqslant K \} \tag{2.4}$$

is weakly closed (actually weakly compact). The convexity follows from (2.2), and we thus have the following corollary.

Corollary 2.4. *Under the conditions of Corollary 2.2, there exists* $Q \in \mathcal{M}_1^+(E \times F)$ *such that* $H(Q, P) \leqslant K$, *and with marginals* μ *and* ν, *if and only if, for all* $f \in C_\mathrm{b}(E)$, $g \in C_\mathrm{b}(F)$,

$$\int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\nu \leqslant \sup_{Q' \in \Lambda_{H,K}} \int (f \oplus g) \,\mathrm{d}Q'.$$

In all the above examples, the weak closure was actually obtained thanks to the weak relative compactness criterion due to Dunford and Pettis (see, for example, Dellacherie and Meyer 1975, 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, for example, Dellacherie and Meyer 1975, p. 38), implies that any uniformly 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 2.5. *Under the conditions of Corollary 2.2, there exists $Q \in \mathscr{M}_1^+(E \times F)$ 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, for all $f \in C_b(E)$, $g \in C_b(F)$,*

$$\int f \, d\mu + \int g \, d\nu \leqslant \sup_{Q' \in \Lambda_{\theta,K}} \int f \oplus g \, dQ'.$$

As in optimization problems, one should now look for a *dual* formulation of Strassen's condition. Recently, Kellerer (1984) studied in detail this kind of problem, but here the constraint is implicit (we want the density of an unknown Q to belong to some L_θ space) and cannot be treated by Kellerer's (1984) method. We shall give such a dual formulation in the next section, by using large-deviations arguments. But let us finish this section with an example showing that there are not sufficient controls on μ and ν alone to obtain a positive answer to our problem.

Example 2.6. Take $E = F = [0, 2]$; then

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

where A is the unit square and h is any probability density on [0, 1] such that, for all $\alpha > 0$, 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 uniformly distributed on [0, 1], and denote by μ_0 and ν_0 the marginals of P. It is clear that $\mu \ll \mu_0$ and

$$\frac{d\mu}{d\mu_0} = \frac{4}{1 + h} 1_{[0,1]} \leqslant 4$$

(and similarly for (ν, ν_0)), so that, for all f and g in $C_b^+([0, 2])$, and for all $q \in [1, +\infty]$,

$$\int f \, d\mu + \int g \, d\nu \leqslant 4 \int f \oplus g \, dP \leqslant 4 \| f \oplus g \|_q.$$

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

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

So dQ/dP cannot belong to any $L^q(P)$ space.

This example shows that even if $d\mu/d\mu_0$ and $d\nu/d\nu_0$ are bounded and P is equivalent to a product measure on $E \times F$, one cannot necessarily find a Q in $\mathscr{M}_1^+(E \times F)$ such that

$\mathrm{d}Q/\mathrm{d}P \in \mathbb{L}^q(P)$ for some $q > 1$, and with marginals μ and ν. Of course, for $f \in C_\mathrm{b}(E)$, $g \in C_\mathrm{b}(F)$ and $1 < q \leqslant +\infty$,

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

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

3. Large deviations and new variational characterizations

In order to study the large-deviations problem for the empirical process associated with a given i.i.d. sample of, say, Brownian motions, Dawson and Gärtner (1987) introduced a variational characterization of the infimum of $H(Q, P)$ (for Wiener measure P) on the set of Q such that $Q \circ X_t^{-1} = \nu_t$ is given; see also Föllmer (1988) for the same problem for bridges. In recent papers, Cattiaux and Léonard (1994; 1995a; 1995b) have extended the results of Dawson and Gärtner (1987) to a large class of Markov processes. In particular, the problem of finiteness of the infimum (i.e. the existence of such a Q) is tackled in Cattiaux and Léonard (1995a) by using a direct large-deviations argument. 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 (1990), and used by Gamboa and Gassiat in various problems such as moments problems (Gamboa and Gassiat 1994) or superresolution (Gamboa and Gassiat 1996). The method now known as the maximum entropy on the mean (MEM) method is described in terms of large deviations in Gamboa and Gassiat (1997). We cannot directly use the results in Gamboa and Gassiat (1997) because our framework is different, but we shall follow the same line of reasoning in Proposition 3.5 below.

Definition 3.1. *We say that a sequence (R_n) of probability measures on a measurable Hausdorff space $(U, \mathscr{B}(U))$ satisfies a large-deviations principle (LDP), with rate function I if:*

(i) I is lower semicontinuous, with values in $\mathbb{R}^+ \cup \{+\infty\}$;
(ii) for any measurable set A in U,

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

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

We shall frequently use the following properties (see, for example, Dembo and Zeitouni 1993).

Definition 3.2. *If the level sets of the rate function I are compact, we shall call I a good rate function.*

Proposition 3.3 (Contraction principle). *If T: $U \to V$ is a continuous map, and I controls* *the LDP for a sequence* (R_n) *on* U *and is good, then* $I'(v) = \inf\{I(U)\colon u \in T^{-1}\{v\}\}$ *controls the LDP for the image measures* $R'_n = R_n \circ T^{-1}$ *and* I' *is also good.*

Let us consider the random measure on $E \times F$,

$$\lambda_n = \frac{1}{n}\sum_{i=1}^{n} Z_i \delta_{(x_i, y_i)}, \tag{3.1}$$

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

$$P = \text{weak limit of } \frac{1}{n}\sum_{i=1}^{n} \delta_{(x_i, y_i)}. \tag{3.2}$$

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

$$\psi_G(\tau) = \log \int_{\mathbb{R}_+} \exp(\tau\xi) G(\mathrm{d}\xi), \qquad \tau \in \mathbb{R}, \tag{3.3}$$

and its Legendre conjugate

$$\gamma_G(\xi) = \psi_G^*(\xi) = \sup_{\tau\in\mathbb{R}}(\tau\xi - \psi_G(\tau)), \qquad \xi \in \mathbb{R}, \tag{3.4}$$

and in what follows we shall make the following assumption:

Assumption 3.4. *Domain* $\psi_G = \mathbb{R}$ *and* G *is not a Dirac mass.*

Next define the natural projection operator T (equipped with the product topology) as

$$\begin{aligned} T\colon \mathscr{M}_1^+(E\times F) &\to \mathscr{M}^{\mathrm{b}}(E) \times \mathscr{M}^{\mathrm{b}}(F). \\ Q &\to \left(\int_F Q(.,\, \mathrm{d}y),\ \int_E Q(\mathrm{d}x,\, .)\right) \end{aligned} \tag{3.5}$$

Our aim will be now to prove an LDP 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 Cattiaux and Léonard (1995a). The first results are obtained by using the projective limit approach of Dawson and Gärtner (1987) as explained in Dembo and Zeitouni (1993, Section 4.6).

For a given Hausdorff measurable space U, $\mathscr{M}^\#(U)$ will denote the algebraic dual of $C_{\mathrm{b}}(U)$, equipped with the $\sigma(\mathscr{M}^\#(U), C_{\mathrm{b}}(U))$ topology. $\mathscr{M}^{\mathrm{b}}(U)$ is embedded in $\mathscr{M}^\#(U)$ and T is still continuous from $\mathscr{M}_1^+$ to $\mathscr{M}^\#(E) \times \mathscr{M}^\#(F)$. Thus we can state

Proposition 3.5. *Assume that* E *and* F *are Hausdorff spaces and* $P \in \mathscr{M}_1^+(E\times F)$.

(i) *The laws* R_n *of* λ_n *satisfy on* $\mathscr{M}^\#(E\times F)$ *an LDP with good rate function*

$$\mathscr{T}_G(Q) = \sup_{l\in C_{\mathrm{b}}(E\times F)} \left(\langle l, Q\rangle - \int \psi_G(l)\,\mathrm{d}P\right).$$

(ii) The laws L_n satisfy on $\mathscr{M}^{\#}(E) \times \mathscr{M}^{\#}(F)$ an LDP with good rate function

$$I_G(\mu, \nu) = \sup_{f \in C_{\mathrm{b}}(E), g \in C_{\mathrm{b}}(F)} \left(\langle f, \mu \rangle + \langle g, \nu \rangle - \int \psi_G(f \oplus g)\, \mathrm{d}P \right).$$

Proof. According to Dembo and Zeitouni (1993, Corollary 4.6.11) we first have to show that

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

exists as an extended real number, for $l \in C_{\mathrm{b}}(E \times F)$. Define

$$J_n(l) = \frac{1}{n} \log \left(\int \exp(n \langle l, \lambda \rangle) R_n(\mathrm{d}\lambda) \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \psi_G(l(z_i)).$$

By Assumption 3.4, ψ_G is bounded on compact sets (since it is continuous on $\mathbb{R}$), and $\psi_G \circ l \in C_{\mathrm{b}}$. So, according to (3.4),

$$\lim_{n \to \infty} J_n(l) = \int \psi_G(l)\, \mathrm{d}P < +\infty.$$

Furthermore, for all $l_1, \ldots, l_k \in C_{\mathrm{b}}(E \times F)$,

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

is everywhere finite, continuous and everywhere differentiable thanks to Assumption 3.4. We thus can apply Dembo and Zeitouni (1993, Corollary 4.6.11) to conclude (i). The proof of (ii) is exactly the same. □

The rate function expressions in the Proposition 3.5 are interesting if we are able to study their domains. Indeed, since R_n and L_n are supported by $\mathscr{M}_+^{\mathrm{b}}$ (the set of positive bounded Radon measures), we know that the LDP holds in this space, with the same rate function provided this function is infinite for all the other elements of $\mathscr{M}^{\#}$. We thus study the finiteness of the large-deviations functional.

Lemma 3.6. *If $\mathscr{T}_G(Q)$ (or $I_G(\mu, \nu)$) is finite, then Q (or (μ, ν)) is a positive continuous linear form on $C_{\mathrm{b}}(E \times F)$ (or $C_{\mathrm{b}}(E) \times C_{\mathrm{b}}(F)$).*

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

$$\mathscr{T}_G(Q) \geqslant t \langle l, Q \rangle - \int \psi_G(tl)\, \mathrm{d}P \geqslant t \langle l, Q \rangle \underset{t \to \infty}{\longrightarrow} \infty,$$

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 3.7. *If E and F are completely regular, and $\mathscr{T}_G(Q)$ 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*

$$\mathscr{T}_G(Q) = \mathscr{T}_G(\check{Q}) = \sup_{l \in C(E \check{\times} F)} \left(\int l \, \mathrm{d}\check{Q} - \int \psi_G(l) \, \mathrm{d}\check{P} \right)$$

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 subject of $E \check{\times} F$ and $C_b(E \times F)$ is isomorphic to $C(E \check{\times} F)$ (see, for example, Jameson 1974). By the 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 $\int l \, \mathrm{d}Q = \int \check{l} \, \mathrm{d}\check{Q}$ 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. □

Remark. Actually, one could directly prove that Q is a positive measure on $E \times F$. The main problem is the regularity of this measure.

We shall now give the key result of our construction.

Proposition 3.8. *Assume that U is a Hausdorff space, P a regular bounded positive measure on U. Then, for any regular bounded positive measure Q, $\mathscr{T}_G(Q) = \mathscr{J}_G(Q)$, where*

$$\mathscr{T}_G(Q) = \sup_{l \in C_b(U)} \left(\langle l, Q \rangle - \int \psi_G(l) \, \mathrm{d}P \right)$$

and

$$\mathscr{J}_G(Q) = \begin{cases} \displaystyle\int \gamma_G \left(\frac{\mathrm{d}Q}{\mathrm{d}P} \right) \mathrm{d}P & \text{if } Q \ll P \text{ (recall (3.6))} \\ +\infty & \text{otherwise.} \end{cases}$$

Remark. Similar statements are contained in Rockafellar (1968; 1971); in particular, Rockafellar (1971, Corollary 4.A) furnishes the above proposition when U is compact (which is actually sufficient for our purpose). Nevertheless, we prefer to give a complete elementary proof (without using compactness). The following proof is essentially due to Gamboa and Gassiat (unpublished).

Proof. For the case of $\mathscr{J}_G(Q) \geqslant \mathscr{T}_G(Q)$, it is enough to check the above equality for $Q \ll P$. But, in this case

$$\mathscr{T}_G(Q) = \sup_{l \in C_b(U)} \left\{ \int \left(l \frac{\mathrm{d}Q}{\mathrm{d}P} - \psi_G(l) \right) \mathrm{d}P \right\} \leqslant \int \sup_{l \in U^{\mathbb{R}}} \left(l \frac{\mathrm{d}Q}{\mathrm{d}P} - \psi_G(l) \right) \mathrm{d}P = \mathscr{J}_G(Q).$$

For the case of $\mathscr{J}_G(Q) \leqslant \mathscr{I}_G(Q)$ again we may assume that $\mathscr{I}_G(Q) < +\infty$. Recall the following facts, which are consequences of (3.4) and Assumption 3.4.

$$\begin{aligned} &\psi'_G \text{ is everywhere defined, increasing and continuous, with range }]\alpha, \beta[\\ &\text{such that } \int_{\mathbb{R}^+} \xi \, \mathrm{d}G(\xi) = m \in]\alpha, \beta[. \overline{]\alpha, \beta[} \text{ is the convex hull of the support of } G. \end{aligned} \tag{3.6}$$

$$\begin{aligned} &\text{For } \xi \in]\alpha, \beta[, \gamma_G(\xi) < +\infty; \text{ whereas for } \xi \notin \overline{]\alpha, \beta[}, \gamma_G(\xi) = +\infty. \\ &\text{Hence, for any } \xi \in]\alpha, \beta[\text{ there exists } \tau \in \mathbb{R}, \text{ with } \tau = \psi_G'^{-1}(\xi), \\ &\text{and in this case } \gamma_G(\xi) = \tau\psi'_G(\tau) - \psi_G(\tau). \end{aligned} \tag{3.7}$$

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:

$$h(x) = \begin{cases} \delta & \text{if } x \in A \\ \delta & \text{if } x \in A^c \text{ and } g(x) > \beta \text{ or } g(x) < \alpha \\ \psi_G'^{-1}(g(x)) & \text{if } x \in A^c \text{ and } \alpha + \epsilon \leqslant g(x) \leqslant \beta - \epsilon \\ 0 & \text{otherwise.} \end{cases} \tag{3.8}$$

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

$$\mathscr{I}_G(Q) \geqslant \int h_n \mathrm{d}Q - \int \psi_G(h_n) \, \mathrm{d}P = \theta_n(\delta, \epsilon).$$

We want to identify the limit of θ_n as n goes to infinity. According to (3.6), (3.7) and Lebesgue's dominated convergence theorem,

$$\lim_{n\to\infty} \theta_n(\delta, \epsilon) = \delta S(A) + \int \{\delta g - \psi_G(\delta)\}(1_{\{g<\alpha\}} + 1_{\{g>\beta\}}) \, \mathrm{d}P + \int \gamma_G(g) 1_{\{\alpha+\epsilon\leqslant g \leqslant \beta-\epsilon\}} \, \mathrm{d}P.$$

Now let $\delta \to \infty$. On the set $\{\xi\colon g(\xi) \notin \overline{]\alpha, \beta[}\}$, $\lim_{\delta\to+\infty}(\delta g(\xi) - \psi_G(\delta)) = +\infty$. We deduce that if $\mathscr{I}_G(Q) < +\infty$, $S(A) = 0$ ($Q \ll P$) and $\mathrm{d}Q/\mathrm{d}P \in \overline{]\alpha, \beta[}$, P-almost surely. Hence, for all $\epsilon > 0$,

$$\mathscr{I}_G(Q) \geqslant \int \gamma_G(g) 1_{\{\alpha+\epsilon\leqslant \frac{\mathrm{d}Q}{\mathrm{d}P} \leqslant \beta-\epsilon\}} \mathrm{d}P,$$

and $\mathscr{I}_G(Q) \geqslant \mathscr{J}_G(Q)$ by the monotone convergence theorem. □

We can finally state the following theorem.

Theorem 3.9. *Assume that E and F are completely regular and that $P \in \mathscr{M}_1^+(E \times F)$. Then:*

(i) the laws R_n of λ_n satisfy on $\mathscr{M}_+^{\mathrm{b}}(E \times F)$ an LDP with good rate function $\mathscr{I}_G(Q) = \int \gamma_G(\mathrm{d}Q/\mathrm{d}P) \, \mathrm{d}P$ if $Q \ll P$, $+\infty$ otherwise;

(ii) the laws $L_n = R_n \circ T^{-1}$ *satisfy on* $\mathscr{M}_+^{\mathrm{b}}(E) \times \mathscr{M}_+^{\mathrm{b}}(F)$ *an LDP with good rate function*

$$I_G(\mu, \nu) = \sup_{f \in C_{\mathrm{b}}(E), g \in C_{\mathrm{b}}(F)} \left(\int f \,\mathrm{d}\mu + g \,\mathrm{d}\nu - \int \psi_G(f \oplus g) \,\mathrm{d}P \right);$$

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

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

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

Proof. (i) We know (see Proposition 3.5) that the laws R_n satisfy an LDP on $\mathscr{M}^{\#}(E \times F)$ with rate function $\mathscr{T}_G(Q)$ which is finite if and only if Q can be identified with a regular element $\check{Q}$ of $\mathscr{M}_+^{\mathrm{b}}(E \check{\times} F)$, where $E \check{\times} F$ is the Stone–Cech compactification of $E \times F$ (see Lemma 3.6 and Proposition 3.7), and if $\check{Q}$ further satisfies

$$\check{Q} \ll \check{P} \text{ and } \int \gamma_G \left(\frac{\mathrm{d}\check{Q}}{\mathrm{d}\check{P}} \right) \mathrm{d}\check{P} < +\infty \text{ (see Proposition 3.8).}$$

But we cannot immediately identify Q with $(\mathrm{d}\check{Q}/\mathrm{d}\check{P})|_{E\times F} P$ (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 $(\check{l}_n)_{n\geqslant 1}$ of $C(E \check{\times} F)$ which converges both in $L^1(\check{P})$ and $\check{P}$-almost surely to $\mathrm{d}\check{Q}/\mathrm{d}\check{P}$. If l_n denotes the restriction of $\check{l}_n$ to $E \times F$ (after identification of $C_{\mathrm{b}}(E \times F)$ and $C(E \check{\times} F)$), we also know that for any $h \in C_{\mathrm{b}}(E \times F)$,

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

which proves that $l_n P$ is weakly convergent (in $\mathscr{M}^{\mathrm{b}}(E \times F)$). But the sequence $(\check{l}_n)$ is uniformly integrable, so by the proof of la Vallée-Poussin's theorem in Dellacherie and Meyer (1975), there exists a continuous Young function θ such that

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

The natural property of Stone–Cech compactification implies that $\sup_n \int \theta(l_n) \,\mathrm{d}P < +\infty$; this shows that (l_n) is uniformly integrable (thus $\sigma(L^1, L^\infty)$ relatively compact by the Dunford–Pettis theorem), and consequently the weak limit of $l_n P$ is of the form $Q = hP$, with $h \in L^1(P)$. It is now immediate that the initial $\check{Q}$ is associated with the above Q. In order to prove that $\int \gamma_G(\mathrm{d}Q/\mathrm{d}P) \,\mathrm{d}P < +\infty$, it suffices to approximate $(\mathrm{d}Q/\mathrm{d}P) 1_{\{\alpha+\epsilon \leqslant g \leqslant \beta-\epsilon\}}$ (with α, β defined in (3.6) and (3.7)) by continuous functions, and use Lebesgue's bounded convergence theorem, then to pass to the limit via monotonic convergence as in the proof of Proposition 3.8. Finally, since $\mathscr{T}_G(Q)$ is finite only for $Q \in \mathscr{M}_+^{\mathrm{b}}(E \times F)$, the LDP holds in this space.

(ii) and (iii) are straightforward applications of the Contraction Principle (Proposition 3.3) and uniqueness of the rate function. □

We shall use Theorem 3.9 in the following form.

Corollary 3.10. *Let E and F be completely regular topological spaces, $\mu \in \mathscr{M}_1^+(E)$, $\nu \in \mathscr{M}_1^+(F)$ and $P \in \mathscr{M}_1^+(E \times F)$. Then there exists $Q \in \mathscr{M}_1^+(E \times F)$ such that Q has marginals μ and ν, and $\int \gamma_G(\mathrm{d}Q/\mathrm{d}P)\,\mathrm{d}P \leqslant K$ if and only if, for all $f \in C_{\mathrm{b}}(E)$, $g \in C_{\mathrm{b}}(F)$,*

$$\int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\nu \leqslant \int \psi_G(f \oplus g)\,\mathrm{d}P + K.$$

By way of an 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 θs, we shall give some examples.

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

Let us 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.

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

Then ψ_G is defined on the whole of $\mathbb{R}$. Furthermore, we have the following lemma.

Lemma 3.11. *There exist positive constants C_1, C_2 such that:*

(i) for $\tau \to +\infty$ we have $\psi_G(\tau) \sim C_1 \tau^{q'}$;
(ii) for $\xi \to +\infty$ we have $\gamma_G(\xi) \sim C_2 \xi^q$, where $1/q + 1/q' = 1$.

The proof of (i) is a straightforward application of Laplace's method, while (ii) follows from general results about Legendre conjugacy. According to Corollary 3.10 and Lemma 3.11 we can state the following corollary.

Corollary 3.12. *In the situation of Corollary 3.10, there exists $Q \in \mathscr{M}_1^+(E \times F)$ such that Q has marginals μ and ν, which satisfies $Q \ll P$ and $\mathrm{d}Q/\mathrm{d}P \in L^q(P)$ if and only if, for some $K > 0$,*

$$\sup_{f \in C_{\mathrm{b}}(E), g \in C_{\mathrm{b}}(F)} \left(\int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\nu - \int \psi_G(f \oplus g)\,\mathrm{d}P \right) \leqslant K.$$

3.2. The entropic case

Let G be the Poisson distribution with mean 1. Then

$$\psi_G(\tau) = \exp(\tau) - 1, \qquad \tau \in \mathbb{R},$$

$$\gamma_G(\xi) = \begin{cases} \xi \log \xi - \xi + 1, & \xi \geqslant 0, \\ +\infty & \xi < 0. \end{cases}$$

Thus $\mathscr{T}_G(Q) = H(Q, P)$ for $Q \in \mathscr{M}_1^+(E \times F)$ and we have the following corollary.

Corollary 3.13. *In the situation of Corollary 3.10, there exists $Q \in \mathscr{M}_1^+(E \times F)$ such that Q has marginals μ and ν, which satisfies $H(Q, P) \leqslant K$ $(K > 0)$ if and only if*

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

3.3. The L^∞ case

Though we cannot realize $\|(\mathrm{d}Q/\mathrm{d}P)\|_\infty$ as a $\gamma_G(\mathrm{d}Q/\mathrm{d}P)$, we shall use Corollary 3.10 in the L^∞ framework. Indeed, for $K > 0$ choose for G the Bernoulli distribution

$$G = \tfrac{1}{2}(\delta_0 + \delta_K) \tag{3.9}$$

then

$$\psi_G(\tau) = \log \frac{1 + \exp(K\tau)}{2}, \qquad \xi \in \mathbb{R},$$

$$\gamma_G(\xi) = \begin{cases} \dfrac{\xi}{K} \log\left(\dfrac{\xi}{K}\right) + \left(1 - \dfrac{\xi}{K}\right) \log\left(1 - \dfrac{\xi}{K}\right) + \log 2, & \text{if } 0 \leqslant \xi \leqslant K, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.10}$$

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

$$\left\| \frac{\mathrm{d}Q}{\mathrm{d}P} \right\|_\infty \leqslant K \text{ if and only if } \int \gamma_G\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\right) \mathrm{d}P \leqslant \log 2.$$

We thus may apply Corollary 3.16 in order to obtain the following corollary.

Corollary 3.14. *In the situation of Corollary 3.10, there exists $Q \in \mathscr{M}_1^+(E \times F)$ with marginals μ and ν, which satisfies $Q \ll P$ and $\|(\mathrm{d}Q/\mathrm{d}P)\|_\infty \leqslant K$ if and only if, for all $f \in C_b(E)$, $g \in C_b(F)$,*

$$\int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\nu \leqslant \int \log(1 + \exp K(f \oplus g)) \,\mathrm{d}P.$$

The last condition is equivalent to the following:

$$\forall f \in C_b(E),\ g \in C_b(F),\ \int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\nu \leqslant K \int \log(1 + \exp(f \oplus g)) \,\mathrm{d}P. \tag{3.11}$$

We conclude this section with two remarks.

Remark 3.15. *Comment on the L^1 case.* If for G we choose an exponential law with parameter 1, i.e

$$\psi_G(\tau) = \begin{cases} -\log(1-\tau), & \text{for } \tau < 1, \\ +\infty, & \text{otherwise,} \end{cases}$$

$$\gamma_G(\xi) = \begin{cases} \xi - 1 - \log \xi, & \text{for } \xi > 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

which does not satisfy Assumption 3.4, then Proposition 3.5 is still available, since ψ_G is essentially smooth. But, all the results later to this proposition can fail to hold. Actually one can show that $\mathscr{T}_G(Q)$ is finite for some measures whose Lebesgue decomposition contains a singular part (with respect to P); see Gamboa and Gassiat (1997) for a long discussion on this phenomenon in another context. The above argument indicates that criterion of Corollary 3.16 cannot be easily extended to the L^1 case.

Remark 3.16. 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 $\tilde{E}$ (or $\tilde{F}$) is regular, we can apply Corollary 3.10 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 $\tilde{E}$ (or $\tilde{F}$) can be chosen as a compact Polish space.

4. Remarks on a set-theoretic formulation

Let us go back to Corollary 2.2 with $q = +\infty$, i.e. there exists Q such that $\|\mathrm{d}Q/\mathrm{d}P\|_\infty \leqslant K$ and with marginals μ and ν, if and only if

$$\text{for all } f \in C_\mathrm{b}(E),\ g \in C_\mathrm{b}(F) \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\nu \leqslant \sup_{0 \leqslant Z \leqslant K, \int Z\,\mathrm{d}P = 1} \int (f \oplus g) Z\,\mathrm{d}P.$$

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

$$\mu(A) + \nu(B) \leqslant 1 + KP(A \times B). \tag{4.1}$$

A remarkable fact noticed by Kellerer (1964a), Strassen (1965, Theorem 6) for Polish spaces, and more recently by Hansel and Troallic (1986, Theorem 4.1) for general measurable spaces, is that (4.1) is actually a sufficient condition for the existence of a probability measure Q with marginals μ and ν satisfying $Q \leqslant KP$. This latest condition implies $Q \ll P$ and $\mathrm{d}Q/\mathrm{d}P \leqslant K$, and (4.1) is thus a necessary and sufficient condition for the existence of $Q \in \Lambda_{\infty,K}$ with marginals μ and ν.

Remark 4.1. Extending (3.11) to bounded Borel functions, and choosing

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

we recover (4.1) by letting δ go to $+\infty$.

The discussion above indicates how to try to obtain a set-theoretic characterization in the general L_θ case. Indeed, if we apply the same idea as that which leads to (4.1) we obtain that if there exists $Q \in \mathscr{M}_1^+(E \times F)$ such that $\mathrm{d}Q/\mathrm{d}P \in L_\theta$ and with marginals μ and ν, then

$$\mu(A) + \nu(B) \leqslant 1 + K\eta[P(A \times B)] \tag{4.2}$$

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

Unfortunately, the above set condition is not sufficient to ensure the existence of Q. Here is a classical counterexample.

Example 4.2. For $E = \{0\}$, $F = \mathbb{N}^*$, $P = C\sum_{n=1}^{+\infty} 2^{-2n}\delta_{(0,n)}$, $\mu = \delta_0$, $\nu = \sum_{n=1}^{+\infty} 2^{-n}\delta_n$, $\theta(x) = x^2$, and (4.2) reduces to $\nu(B) \leqslant K[P(\{0\} \times B)]^{1/2}$ for all $B \in \mathscr{B}(\mathbb{N}^*)$.

Let j be the smallest element in B; then $\nu(B) \leqslant 2 \cdot 2^{-j}$ and $P[(0, B)] \geqslant C2^{-2j}$, i.e. $\nu(B) \leqslant \sqrt{P[(0, B)]C}/2$. 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 $\mathrm{d}Q/\mathrm{d}P \notin L^2(P)$. (Of course, we only used the fact that $Q(A) \leqslant K\sqrt{P(A)}$ does not imply $\mathrm{d}Q/\mathrm{d}P \in L^2(P)$ in general.)

We should therefore ask whether the new characterization of Section 3 leads to interesting set-theoretic inequalities. The answer here again is negative. Now if we look at Strassen's proof (or similarly at Hansel and Troallic's one), one can easily see why (4.2) 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 (4.2) becomes

$$\mu(A) + \nu(B) \leqslant 1 + K[P(A \times B)]^{1-1/q}. \tag{4.3}$$

It easy to prove the following lemma.

Lemma 4.3. *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 Strassen, 1965, p. 429), say H, such that

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

where μ_0 denotes the first marginal of P. In the case $q = +\infty$, such a kernel is given by a regular disintegration of P (if, for instance, E and F are separable metric spaces; see Dellacherie and Meyer 1975, 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 can be used.

In the L^q case ($1 < q < +\infty$), we shall, however, state a set-theoretic characterization which is the analogue of (4.1) but is not so beautiful. To this end we first introduce some definitions.

Definition 4.4. *Let $(\Omega, \mathscr{F})$ be a measurable space.*

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

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

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

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

$$P(U|\mathscr{A})(\omega) = P(U|A_i)(\omega) \text{ if } \omega \in A_i (A_i \in \mathscr{A}) \text{ and } P(A_i) \neq 0$$

$$= 0 \text{ if } \omega \in A_i \text{ and } P(A_i) = 0.$$

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

We can now state the following theorem.

Theorem 4.5. *Let $(E, \mathscr{E})$ and $(F, \mathscr{F})$ be two measurable spaces. μ, ν, P are probability measures defined respectively on $(E, \mathscr{E})$, $(F, \mathscr{F})$ and $(E \times F, \mathscr{E} \times \mathscr{F})$. μ_0 denotes the first marginal of P. Then, there exists a probability measure Q on $(E \times F, \mathscr{E} \times \mathscr{F})$ such that $Q \ll P$, $\mathrm{d}Q/\mathrm{d}P \in L^q(P)$, $\|\mathrm{d}Q/\mathrm{d}P\|_q \leqslant K$ and with marginals μ and ν if and only if the following holds:*

Let $\mathscr{E}'$ ($\mathscr{F}'$) be any separable sub-σ-field of $\mathscr{E}$ ($\mathscr{F}$). One can find a partition core $(\mathscr{A}_n)$ ($(\mathscr{B}_n)$) of $\mathscr{E}'$ ($\mathscr{F}'$) and a family $(Z_n)_{n\geq 1}$ of non-negative random variables such that

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

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

$$\mu(A) + \nu(B) \leqslant 1 + \int_A Z_n(x)[P(E \times B|\mathscr{A}_n \times F)]^{1-1/q}(x)\mu_0(\mathrm{d}x),$$

(where $P(\cdot|\mathscr{A}_n \times F)$ is as per Definition 4.4(iv)).

Corollary 4.6. *Assume that E and F are Polish spaces, $\mathscr{E} = \mathscr{B}(E)$, $\mathscr{F} = \mathscr{B}(F)$. Denote by $P(x, \cdot)$ a regular disintegration of P with respect to $\mathscr{B}(E)$ (considered as a sub-σ-field of $\mathscr{B}(E) \times \mathscr{B}(F)$), i.e.*

$$P(A \times B) = \int_A P(x, B)\mu_0(\mathrm{d}x), \qquad \text{for } A \in \mathscr{B}(E), B \in \mathscr{B}(F);$$

such a $P(x, \cdot)$ is called a Markov kernel in Strassen (1965). Then, the necessary and sufficient condition of Theorem 4.5 is equivalent to

$$\mu(A) + \nu(B) \leqslant 1 + \int_A Z(x)[P(x, B)]^{1-1/q}(x)\mu_0(\mathrm{d}x)$$

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

Remark 4.7. If $q = +\infty$, the above condition is equivalent to that of Strassen, and the condition in Theorem 4.5 is also equivalent to (4.1). So we also recover Theorem 6 of Strassen (1965) or Theorem 4.1 of Hansel and Troallic (1986), but with a different proof for the latter case.

Proof of Theorem 4.5. For the if part we shall closely follow Strassen's method; indeed, consider $(E, \sigma(\mathscr{A}_n))$, $(F, \sigma(\mathscr{B}_n))$ 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 $\mathscr{A}_n$ (or $\mathscr{B}_n$)), which are of course Polish. So, as in the proof of Theorem 6 of Strassen (1965) we may apply Theorem 4 of Strassen (1965) in $(E, \sigma(\mathscr{A}_n))$, $(E, \sigma(\mathscr{B}_n))$ and $(E \times F, \sigma(\mathscr{A}_n) \otimes \sigma(\mathscr{B}_n))$. To this end, consider

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

for $A_i \in \mathscr{A}_n$, $B \in \sigma(\mathscr{B}_n)$ and $Z_n(i)$ equal to the value of Z_n on A_i.

Let $B \in \sigma(\mathscr{B}_n)$, and A the set where $H_n(\cdot, B) < 1$ (we define $H_n(x, B) = H_n(A_i, B)$ if $x \in A_i$). Condition (ii) in Theorem 4.5 yields

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

But, according to Lemma 4.3, $H_n(x, \cdot)$ is a kernel alternating of order 2 in the sense of Strassen (1965). Indeed, we have

$$1 = \nu(F) \leqslant \int H_n(x, F)\mu(\mathrm{d}x) \leqslant 1 \Rightarrow \int H_n(x, F)\mu(\mathrm{d}x) = 1.$$

Applying Theorem 4 of Strassen (1965), as we said before, we obtain that there exists a Markov kernel $q_n(\cdot, \cdot)$ defined on $\sigma(\mathscr{A}_n) \otimes \sigma(\mathscr{B}_n)$ such that $\nu = q_n \mu$ and $q_n(x, \cdot) \leqslant H_n(x, \cdot)$ 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 $(E \times F, \sigma(\mathscr{A}_n) \otimes \sigma(\mathscr{B}_n))$ with marginals μ and ν, and for $A_i \in \mathscr{A}_n$ and $B_j \in \mathscr{B}_n$:

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

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

$$\frac{\mathrm{d}Q_n}{\mathrm{d}P} = \tilde{Z}_n = \sum_{ij} \frac{Q_n(A_i \times B_j)}{P(A_i \times B_j)} 1_{A_i \times B_j} \qquad \left(\text{by convention } \frac{0}{0} = 0\right).$$

It follows that

$$\int \tilde{Z}_n^q \,\mathrm{d}P \leqslant \sum_i (Z_n(i))^q \mu_0(A_i) \leqslant K^q, \qquad \text{i.e. } \|\tilde{Z}_n\|_q \leqslant K.$$

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

$$\int_{A\times F} Z \,\mathrm{d}P = \lim_{n\to\infty} \int_{A\times F} \tilde{Z}_n \mathrm{d}P = \mu(A), \qquad \text{for } A \in \bigcup_{p\geqslant 1} \mathscr{A}_p,$$

and

$$\int_{E\times B} Z \,\mathrm{d}P = \lim_{n\to\infty} \int_{E\times B} \tilde{Z}_n \mathrm{d}P = \nu(B), \qquad \text{for } B \in \bigcup_{p\geqslant 1} \mathscr{B}_p,$$

because $(\mathscr{B}_p)_{p\geqslant 1}$ 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 $\mathscr{E}'$ is associated Z' as above, and again we may apply the Dunford–Pettis theorem in $(E \times F, \mathscr{E} \times \mathscr{F})$, 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 $\|Z\|_q \leqslant K$. The Probability measure $Q = ZP$, of course has marginals μ and ν.

The only if part is immediate, with $Z_n(i) = (\mathrm{E}[Z^q | A_i \times F])^{1/q}$ for $A_i \in \mathscr{A}_n$ and $Z = \mathrm{d}Q/\mathrm{d}P$, by using Hölder's conditional inequality. □

Proof of Corollary 4.6. The only if part holds with $Z(x) = (\int Z^q(x, y) P(x, \mathrm{d}y))^{1/q}$ as above. For the if part, it suffices to mimic the proof of Theorem 4.5 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 4.3 in the situation of (4.2) (i.e. with $1/\theta^{*-1}(1/u)$). Finally, in the L^1 case, one can ask about the following conjecture.

Conjecture 4.8. *There exists $Q \ll P$ with marginals μ and ν if and only if for all $\epsilon > 0$ there exists an $\eta > 0$ such that $\mu(A) + \nu(B) - 1 \geqslant \epsilon$ implies $P(A \times B) \geqslant \eta$.*

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

5. Minimal elements

From now on we assume that E and F are completely regular. Define

$$K^* = \sup_{f \in C_b(E), g \in C_b(F)} \left(\int f \, d\mu + \int g \, d\nu - \int \psi_G(f \oplus g) \, dP \right) \tag{5.1}$$

and assume that

$$K^* < +\infty. \tag{5.2}$$

According to Corollary 3.10 and since γ_G is strictly convex on its domain, there exists $Q^* \in \mathscr{M}_1^+(E \times F)$ with marginals μ and ν such that

$$\mathscr{T}_G(Q^*) = K^* < \mathscr{T}_G(Q) \qquad \text{for all } Q \neq Q^* \text{ with the same marginals.} \tag{5.3}$$

We shall call Q^* the minimal element. Our goal in this section is to describe Q^*. The first main result in this direction is the following

Theorem 5.1. *Assume that (5.2) holds and let $Q^* = Z^* P$ be the minimal element. Then, there exists a sequence $(f_n, g_n) \in C_b(E) \times C_b(F)$ such that $Z_n = \psi'_G(f_n \oplus g_n)$ converges towards Z^* both P-a.s. and in $L^1(P)$.*

Proof. The idea consists in building a good sequence (f_n, g_n) which approximates the supremum in (5.1). Actually, it shall suffice to prove the following lemma.

Lemma 5.2. *There exists a sequence (f_n, g_n) as above such that*

$$\mathscr{T}_G(Z_n P) = i(f_n, g_n) = \int f_n \, d\mu + \int g_n \, d\nu - \int \psi_G(f_n \oplus g_n) \, dP$$

converges towards K^*, and $Z_n = \psi'_G(f_n \oplus g_n)$ converges towards Z^* weakly in $L^1(P)$.

Indeed, according to Pratelli (1992, Theorem 5.1), since $Z_n \to Z^*$ weakly in $L^1(P)$ and $\mathscr{T}_G(Z_n P) = i(f_n, g_n)$ converges towards $\mathscr{T}_G(Z^* P)$, $Z_n \to Z^*$ strongly in $L^1(P)$. Hence, up to a subsequence we may also assume that $Z_n \to Z^*$ P-a.s.

For a given $f \oplus g$, consider the function of two real variables

$$\theta_z(\lambda, \eta) = \lambda + \eta z - \int \psi_G(\lambda + \eta(f \oplus g)) \, dP, \qquad z \in \mathbb{R}. \tag{5.4}$$

When $z = a = \int f \, d\mu + \int g \, d\nu$, (5.1) implies that θ_a is bounded by K^*. θ_z is smooth and strictly concave as soon as $f \oplus g$ is not P-a.s. constant. Furthermore, we have the following lemma.

Lemma 5.3. *We assume that $f \oplus g$ is not P-a.s. constant and that there exists $\tilde{Q} \in \mathscr{M}_1^+(E \times F)$ with marginals μ and ν such that $\mathscr{T}_G(\tilde{Q}) < +\infty$ and*

(i) if $\lim_{\tau \to +\infty} \psi'_G(\tau) = +\infty$, $\tilde{Q}$ is not concentrated on $\{f \oplus g = \mathrm{esssup}_P(f \oplus g)\}$ or on $\{f \oplus g = \mathrm{essinf}_P(f \oplus g)\}$;

(ii) if $\lim_{\tau\to+\infty}\psi'_G(\tau)=M<+\infty$, $\tilde{Q}$ *cannot be written as*

$$(M1_{\{f\oplus g<\xi\}}+M\chi 1_{\{f\oplus g=\xi\}})P \quad \text{or} \quad (M1_{\{f\oplus g>\xi\}}+M\chi 1_{\{f\oplus g=\xi\}})P, \tag{5.5}$$

where $\xi\in\mathbb{R}$ *and* χ *is a measurable function on* $E\times F$.

Then, θ_a *admits a unique maximum* $\theta_a(\lambda_0,\eta_0)$.

Proof. First, let $\lim_{\tau\to+\infty}\psi'_G(\tau)=+\infty$. Without loss of generality we may assume that $z_0=\int(f\oplus g)\,\mathrm{d}P=0$ and $a>0$. Thus, (i) implies that $P(f\oplus g>a+\epsilon)>0$ for some $\epsilon>0$. For such ϵ, let

$$\check{Q}=\left(\xi+(1-\xi)\frac{1_{\{f\oplus g>a+\epsilon\}}}{P(f\oplus g>a+\epsilon)}\right)P, \qquad 0<\xi<1.$$

We may choose ξ such that $c=(1-\xi)(a+\epsilon)>a$, and we have

$$\mathscr{T}_G(\check{Q})\leqslant\gamma_G(\xi)+\gamma_G\left(\xi+\frac{(1-\xi)}{P(f\oplus g>a+\epsilon)}\right)<+\infty.$$

Hence,

$$\exists\check{Q}\in\mathscr{M}_1^+(E\times F)\text{ with }\mathscr{T}_G(\check{Q})<+\infty\text{ and }c=\int(f\oplus g)\,\mathrm{d}\check{Q}>a. \tag{5.6}$$

Now let $\lim_{\tau\to+\infty}\psi'_G(\tau)=M<+\infty$. Here we may assume that $f\oplus g\geqslant 0$ P-a.s. and $z_0=1$. We only consider the case where $a\geqslant 1$ (the case $a<1$ can be treated using the same kind of arguments replacing $f\oplus g$ by $(M-f\oplus g)/(M-1)$ and a by $(M-a)/(M-1)$). Consider the statistical test H_0: P versus H_1: $(f\oplus g)P$ at level $1/M$. Then, the Neyman–Pearson lemma (Lehmann, 1959, Theorem 1, p. 65) says that setting

$$\Phi^*=1_{\{f\oplus g<\xi\}}+\chi 1_{\{f\oplus g=\xi\}}, \tag{5.7}$$

where

$$P(f\oplus g<\xi)+\mathrm{E}(\chi 1_{\{f\oplus g=\xi\}})=\frac{1}{M} \tag{5.8}$$

(ξ is essentially unique and χ satisfies $\mathrm{E}(\chi 1_{\{f\oplus g=\xi\}})=1/M-P(f\oplus g<\xi)$), any test $\Phi\neq\Phi^*$ having the same level satisfies:

$$\int\Phi(f\oplus g)\,\mathrm{d}P<\int\Phi^*(f\oplus g)\,\mathrm{d}P. \tag{5.9}$$

Applied to $\Phi=(1/M)/(\mathrm{d}\tilde{Q}/\mathrm{d}P)$, (5.9) gives $a<\int(f\oplus g)M\Phi^*\,\mathrm{d}P=b$. Let $0<r<1$ with $c=r+(1-r)b>a$ and $\check{Q}=(r+(1-r)M\Phi^*)P$. Since by construction $0\leqslant\chi\leqslant 1$ we have $\gamma_G(r+(1-r)M\Phi^*)\leqslant\max(\gamma_G(r),\gamma_G(r+(1-r)M))<+\infty$ and $\mathscr{T}_G(\check{Q})<+\infty$, so (5.6) holds.

For any $z\in\mathbb{R}$, the Contraction Principle (Proposition 3.3) gives

$$\inf_{Q\in\mathscr{M}_1^+(E\times F),\int(f\oplus g)\,\mathrm{d}Q=z}\mathscr{T}_G(Q)=\sup_{\lambda,\eta}\theta_z(\lambda,\eta). \tag{5.10}$$

Indeed, by the Ellis–Gärtner theorem on $\mathbb{R}^2$ (Dembo and Zeitouni, 1993, Theorem 2.3.6, p. 45), the dual function of $\int \psi_G(\lambda + \eta(f \oplus g))\,\mathrm{d}P$ is the large-deviations functional for the sequence of random vectors $(\lambda_n(E \times F), \int (f \oplus g)\,\mathrm{d}\lambda_n)$. A direct evaluation gives $\sup_{\lambda,\eta}\theta_{z_0}(\lambda, \eta) = \theta_{z_0}(\psi_G'^{-1}(1), 0)$. As θ_z is strictly concave this equality implies

$$\lim_{\|(\lambda,\eta)\|\to+\infty} \theta_{z_0}(\lambda, \eta) = -\infty. \tag{5.11}$$

Now, there exists $0 < \tilde{r} < 1$ with $a = \tilde{r}z_0 + (1 - \tilde{r})c$ so

$$\theta_a(\lambda, \eta) = \tilde{r}\theta_{z_0}(\lambda, \eta) + (1 - \tilde{r})\theta_c(\lambda, \eta).$$

From (5.11) and (5.15), $\theta_c(\lambda, \eta)$ is bounded so that (5.11) implies

$$\lim_{\|(\lambda,\eta)\|\to+\infty} \theta_a(\lambda, \eta) = -\infty$$

which gives the result. □

According to Lemma 5.3, θ_a admits a maximum at (λ_0, η_0) and $\nabla\theta_a(\lambda_0, \eta_0) = 0$. It follows that

$$1 = \int \psi_G'(\lambda_0 + \eta_0(f \oplus g))\,\mathrm{d}P \tag{5.12a}$$

$$\int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\nu = \int (f \oplus g)\psi_G'(\lambda_0 + \eta_0(f \oplus g))\,\mathrm{d}P. \tag{5.12b}$$

So if we replace $f \oplus g$ by $(\lambda_0 + \eta_0 f) \oplus \eta_0 g = \overline{f} \oplus \overline{g}$, we have

$$\begin{cases} 1 = \int \psi_G'(\overline{f} \oplus \overline{g})\,\mathrm{d}P \\ \int (\overline{f} \oplus \overline{g})Z^*\,\mathrm{d}P = \int (\overline{f} \oplus \overline{g})\psi_G'(\overline{f} \oplus \overline{g})\,\mathrm{d}P \\ i(\overline{f}, \overline{g}) \geqslant i(f, g). \end{cases} \tag{5.13}$$

Take a sequence (f_n, g_n) such that $\lim_{n\to\infty} i(f_n, g_n) = K^* = \mathscr{T}_G(Z^*P)$. Without loss of generality, we may assume that f_n, g_n and Z^*P satisfy the assumptions of Lemma 5.3. Indeed, if this is not the case take a small perturbation of f_n, g_n. This means in view of (5.13) that we can assume that $Z_n = \psi_G'(f_n \oplus g_n)$ is a probability density and

$$\begin{aligned} \mathscr{T}_G(Z_nP) &= \int \gamma_G(\psi_G'(f_n \oplus g_n))\,\mathrm{d}P \\ &= \int \{(f_n \oplus g_n)\psi_G'(f_n \oplus g_n) - \psi_G(f_n \oplus g_n)\}\,\mathrm{d}P \\ &= \int (f_n \oplus g_n)Z^*\,\mathrm{d}P - \int \psi_G(f_n \oplus g_n)\,\mathrm{d}P = i(f_n, g_n) \leqslant K^*. \end{aligned} \tag{5.14}$$

Accordingly, thanks again to the Dunford–Pettis theorem, one can find a subsequence of Z_n which converges towards Z weakly in L^1, and

$$\int \gamma_G(Z)\,\mathrm{d}P \leqslant K^*. \tag{5.15}$$

In order to prove that $Z = Z^*$ we have to prove that Z has marginals μ and ν and use the minimality property of Z^*.

Suppose that Z and Z^* do not have same marginals. Since they are both probability measures, one can find a non-negative $f \oplus g (\in C_b(E) \oplus C_b(F))$ such that $\int (f \oplus g)(Z^* - Z)\,\mathrm{d}P = \alpha < 0$. In the following we write $h = f \oplus g$ and $h_n = f_n \oplus g_n$, as well as $i(h)$ instead of $i(f, g)$.

For $\xi \in \mathbb{R}$, consider

$$F_n(\xi) = i(h_n + \xi h) = \int (h_n + \xi h) Z^*\,\mathrm{d}P - \int \psi_G(h_n + \xi h)\,\mathrm{d}P.$$

We may apply the Taylor–Lagrange formula in order to obtain that for $\xi < 0$, there exists $\xi_n \in]\xi, 0[$ such that

$$\begin{aligned} F_n(\xi) &= F_n(0) + \xi F'_n(0) + \xi(F'_n(\xi_n) - F'_n(0)) \\ &= i(h_n) + \xi \int h(Z^* - Z_n)\,\mathrm{d}P + \xi \int h(\psi'_G(h_n) - \psi'_G(h_n + \xi_n h))\,\mathrm{d}P. \end{aligned} \tag{5.16}$$

The key point now is that $\xi_n h \leqslant 0$, hence $0 \leqslant \psi'_G(h_n + \xi_n h) \leqslant \psi'_G(h_n)$. Since $(\psi'_G(h_n))$ is a uniformly integrable sequence, so is $(\psi'_G(h_n + \xi_n h))$. In particular, one can find $a > 0$ such that for all n

$$\int_{h_n > (\psi'_G)^{-1}(a)} \psi'_G(h_n)\,\mathrm{d}P \leqslant \frac{-\alpha}{4\|h\|_\infty} \text{ and } \int_{h_n > (\psi'_G)^{-1}(a)} \psi'_G(h_n + \xi_n h)\,\mathrm{d}P \leqslant \frac{-\alpha}{4\|h\|_\infty}.$$

Finally, we can write

$$F_n(\xi) = i(h_n) + \xi\alpha + \xi \int h(Z - Z_n)\,\mathrm{d}P + \xi(I_1^n + I_2^n).$$

with

$$I_1^n = \xi_n \int_{h_n \leqslant (\psi'_G)^{-1}(a)} h\psi''_G(h_n + \xi'_n h)\,\mathrm{d}P \qquad \text{for some } \xi'_n \in]\xi_n, 0[$$

$$I_2^n = \int_{h_n > (\psi'_G)^{-1}(a)} h(\psi'_G(h_n) - \psi'_G(h_n + \xi_n h))\,\mathrm{d}P.$$

But on the interval $]-\infty, \psi'^{-1}_G(a)]$, ψ''_G is bounded (it is easy to see that $\lim_{\tau \to -\infty} \psi''_G(\tau) = 0$), and so there exists a constant C such that, for all n,

$$|\xi(I_1^n + I_2^n)| \leqslant C\xi^2 + \frac{\xi\alpha}{2}.$$

But $K^* \geqslant F_n(\xi) \geqslant i(h_n) + -C\xi^2 + \xi\alpha/2 + \xi \int h(Z - Z_n)\,\mathrm{d}P$ for all n which yields a contradiction since $-C\xi^2 + \xi\alpha/2$ is strictly positive for $|\xi|$ small enough.

It follows that $Z^* = Z$ and Lemma 5.2 is proved so is Theorem 5.1. $\square$

Remark 5.4. One cannot use the Taylor–Lagrange formula of order 2 directly, because in the case $\lim_{\tau\to+\infty}\psi'_G(\tau)=+\infty$ one cannot, in general, control $\int\psi''_G(h_n+\xi_n h)\,\mathrm{d}P$, even if $\xi_n<0$. Also, remark that it is crucial to know that Z is a probability density in order to choose a non-negative h and obtain a negative α.

In view of the nature of γ_G, one should expect to improve the L^1 strong convergence in Theorem 5.1, and get strong convergence for the Orlicz norm associated with γ_G. Actually, this stronger result is an easy consequence of a Vitali-like theorem in Orlicz space, and we can state the following corollary.

Corollary 5.5. *In addition to the hypotheses of Theorem 5.1, assume that γ_G is moderate (i.e. satisfies Δ_2-regularity in Orlicz space terminology). Then (a subsequence of) Z_n converges towards Z^* strongly in the Orlicz space L_{γ_G} associated with γ_G.*

Proof. According to Theorem 12(b) of Rao and Ren (1991, p. 83), and since (a subsequence) of Z_n almost surely converges towards Z^*, we only need to check that

$$\lim_{n\to\infty}\int\gamma_G(Z_n)\,\mathrm{d}P=\int\gamma_G(Z^*)\,\mathrm{d}P. \tag{5.17}$$

On the one hand, (5.14) proves the upper bound. On the other hand, lower semicontinuity implies the lower bound, which achieves the proof. □

Remark 5.6.

(i) In the entropic case (Section 3.2), setting $Q_n=Z_nP$, we have that

$$H(Q^*,Q_n)=H(Q^*,P)-i(f_n,g_n)$$

goes to 0 as n goes to infinity. (Recall that H denotes the Kullback–Leibler information (see (2.2)).) Indeed, since Q_n and P are equivalent, $Q^*\ll Q_n$ and the following holds:

$$H(Q^*,Q_n)=H(Q^*,P)-E_{Q^*}[\log Z_n].$$

But

$$\mathrm{E}_{Q^*}[\log Z_n]=\int(f_n\oplus g_n)Z^*\,\mathrm{d}P=\int(f_n\oplus g_n)Z^*\,\mathrm{d}P-\int\psi_G(f_n\oplus g_n)\,\mathrm{d}P=i(f_n,g_n)$$

since $\psi_G(\tau)=\mathrm{e}^\tau-1$, $\psi'_G(\tau)=\mathrm{e}^\tau$ and $\psi'_G(f_n\oplus g_n)$ is a probability density thanks to (5.17).

(ii) A similar statement with another approximating sequence $f_n\oplus g_n$ is contained in Borwein *et al*. (1994), Csiszár (1975) and Föllmer (1988). Actually, Csiszár's (1975) *I*-projection yields a sequence $f_n\oplus g_n$, solving a finite number (n) of moment problems, which approximate the marginal problem. The advantage of Remark 5.6(i) is that it gives the exact error $H(Q^*,Q_n)$.

(iii) In the entropic case, one can easily see that $\lim_{\lambda\to\pm\infty}\theta_a(\lambda,1)=-\infty$, so that we can replace in the maximization procedure of Lemma 5.3 the two variables (λ,η) by only one (λ). Easy computations yield the following alternative expression for K^*:

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

This expression is more familiar to aficionados of large deviations, and can be derived by using Sanov's theorem and the contraction principle instead of MEM; see Cattiaux and Léonard (1995a; 1995b) for the method for marginal flows.

(iv) In the general case, Theorem 5.1 is connected with recent results of Csiszár (1995) generalizing the entropic case, with the help of Bregman distances.

6. More on minimal elements and applications

Theorem 5.1 says that the minimal Z^* can be approached by some $\psi'_G(f_n \oplus g_n)$ P-a.s. It follows that $f_n \oplus g_n = \psi'^{-1}_G(Z_n)$ converges P-a.s. to some measurable F^* taking values in $[-\infty, +\infty]$ and $Z^* = \psi'_G(F^*)$. The last question we shall address is the splitting $F^* = f^* \oplus g^*$ and some of its consequences. It is known (see Lindenstrauss, 1965) that this splitting is not always true. Many results, however, are known (see Borwein and Lewis, 1992; Borwein *et al.*, 1994; Donsker and Varadhan, 1974; Föllmer, 1988), but the most satisfactory one for our purpose is the following one due to Rüschendorff and Thomsen (1994). Let $\tilde{\mu}_0$ $(\tilde{\nu}_0)$ be a probability measure on E (F). Observe that these probability measures are not necessarily the marginals of P.

Proposition 6.1 (see Rüschendorff and Thomsen 1994, Proposition 2). *If $P \ll \tilde{\mu}_0 \otimes \tilde{\nu}_0$ and $f_n \oplus g_n$ converges P-a.s. towards F^*, then one can find measurable functions f^* and g^* such that $F^* = f^* \oplus g^*$ on the set $\{-\infty < F^* < +\infty\}$. (Actually, to get this statement just replace A by $A \cap \{-\infty < F^* < +\infty\}$ in the proof of Rüschendorff and Thomsen's proposition.)*

As an immediate consequence we obtain

Proposition 6.2. *If (5.2) holds and $P \ll \tilde{\mu}_0 \otimes \tilde{\nu}_0$, there exists a pair (f^*, g^*) of measurable functions such that*

(i) if $\lim_{\tau \to +\infty} \psi'_G(\tau) = +\infty$, $Z^* = \psi'_G(f^* \oplus g^*) 1_{Z^* > 0}$ P-a.s.;
(ii) if $\lim_{\tau \to +\infty} \psi'_G(\tau) = M < +\infty$, $Z^* = \psi'_G(f^* \oplus g^*) 1_{M > Z^* > 0} + M 1_{Z^* = M}$ P-a.s.

Before we give applications of Proposition 6.2 in the entropic case, we shall say a few words about the L^∞ case. Assume that $Q^* = Z^* P$ has marginals μ and ν and that $\|Z^*\|_\infty = K^*$ is minimal. Then if $P \ll \tilde{\mu}_0 \otimes \tilde{\nu}_0$, a remarkable result due to Kellerer (1984) tells us that one can always find a subset A of $E \times F$ such that $K^* 1_A P$ has the same marginals as Q^* provided $\tilde{\mu}_0$ and $\tilde{\nu}_0$ have no atom. Notice that taking $K = K^*$ in (3.9) we have for the homothetic of a characteristic function of a measurable set A (that is for $K^* 1_A$), $\gamma_G(K^* 1_A) = \log 2$ everywhere, hence as $\gamma_G \leqslant \log 2$, $\mathscr{T}_G(\cdot P)$ hits its maximum on each homothetic of a characteristic function of a measurable set which lies in the convex compact subset M_∞ of probability measures Q with marginals μ and ν such that $\|\mathrm{d}Q/\mathrm{d}P\|_\infty = K^*$ (convexity follows from the minimality of K^*). It is an open question whether all extremal points (in the sense of Krein and Milman) of M_∞ are homothetic of characteristic functions (i.e. maximize γ_G) or not.

Our construction furnishes another candidate (for the minimization of $\|\cdot\|_\infty$), of the form (see Proposition 6.2)

$$Z^{**}P = \left(\frac{K^* 1_{\{K^* > Z^{**} > 0\}}}{1 + \exp(f^* \oplus g^*)} + K^* 1_{\{Z^{**} = K^*\}}\right) P.$$

We next discuss the entropic case. Because of its importance for large deviations theory, the entropic case has been extensively studied. As remarked by Föllmer (Föllmer, 1988; Föllmer and Gantert, 1995) the split decomposition of Z^* is strongly related to an old Schrödinger question as we shall state below. Actually, our approach allows us to improve various results on the subject in the literature.

In the following we assume that

$$P = k\tilde{\mu}_0 \otimes \tilde{\nu}_0, \text{ for some non-negative } k \in L^1(\tilde{\mu}_0 \otimes \tilde{\nu}_0). \tag{6.1}$$

For K^* to be finite it is necessary (but not sufficient) that

$$H(\mu, \tilde{\mu}_0) < +\infty, \qquad H(\nu, \tilde{\nu}_0) < +\infty. \tag{6.2}$$

A particular property of entropy is that $H(\mu \otimes \nu, \tilde{\mu}_0 \otimes \tilde{\nu}_0) = H(\mu, \tilde{\mu}_0) + H(\nu, \tilde{\nu}_0)$. Hence, because

$$H(\mu \otimes \nu, P) = H(\mu \otimes \nu, \tilde{\mu}_0 \otimes \tilde{\nu}_0) - \iint \log k d(\mu \otimes \nu), \tag{6.3}$$

it follows that

$$\begin{aligned} &\text{if } \log k \in Ł^1(\mu \otimes \nu), \text{ then } H(\mu \otimes \nu, P) < +\infty \text{ (i.e. } K^* \text{ is finite) and the} \\ &\text{minimal element } Z^* \text{ satisfies } Z^* = \exp(f^* \oplus g^*) \ P\text{-a.s. on the set } \{Z^* > 0\}. \end{aligned} \tag{6.4}$$

Q^* is supported by the cross product $E' \times F' = \{\mathrm{d}\mu/\mathrm{d}\tilde{\mu}_0 > 0\} \times \{\mathrm{d}\nu/\mathrm{d}\tilde{\nu}_0 > 0\}$. Indeed, $\mu \ll \tilde{\mu}_0$, $\nu \ll \tilde{\nu}_0$ and $\mu \otimes \nu$ is equivalent to $\tilde{\mu}_0 \otimes \tilde{\nu}_0$ on the set $E' \times F'$. But, as Q^* has marginals μ and ν, $Q^*(E' \times F') = 1$. Thus, $Q^* \ll P \ll \mu \otimes \nu$ on $E' \times F'$. Hence, as (6.6) holds, condition (EQ) in Borwein *et al*. (1994) is satisfied. Thus, Theorem 2.7 of Borwein *et al*. (1994) shows that $Z^* > 0$ P-a.s. on $E' \times F'$. We have thus proved the following proposition.

Proposition 6.3. *Assume that* $H(\mu, \tilde{\mu}_0)$ *and* $H(\nu, \tilde{\nu}_0)$ *are finite and that* $\log k \in L^1(\mu \otimes \nu)$. *Then, there exists a pair* (f^*, g^*) *of measurable functions taking values in* $[-\infty, +\infty[$ *such that* $Z^*(x, y) = \exp(f^*(x)) \exp(g^*(y))$ P-a.s.

Indeed, take f^* and g^* as in Proposition 6.2 on $E' \times F'$ and put $f^* = -\infty$ on $E \backslash E'$ ($g^* = -\infty$ on $F \backslash F'$).

Remark 6.4. On the unit square $[0, 1] \times [0, 1]$ take $\mathrm{d}P = \exp(-1/x)\exp(-1/y)\,\mathrm{d}x\,\mathrm{d}y$ up to a normalization constant, μ and ν being Lebesgue measure. It is easily seen as in Example 2.6, that there is no Q with marginals μ and ν such that $H(Q, P) < +\infty$. Of course $\log k \notin L^1(\mu \otimes \nu)$. But, if we replace P by $\mathrm{d}P = \exp\{-1/(x^2 + y^2)^2\}\,\mathrm{d}x\,\mathrm{d}y$ (up to a normalization constant), $\log k \notin L^1(\mu \otimes \nu)$ and it is easy to build a Q with uniform marginals such that $H(Q, P) < +\infty$ (for instance, with support in $[0, \frac{1}{2}] \times [\frac{1}{2}, 1] \cup [\frac{1}{2}, 1] \times [0, \frac{1}{2}]$). The global condition of integrability can thus be improved using a local one. Instead of discussing this point further, we shall now link Proposition 6.3 to Schrödinger's problem.

If we denote by (α, β) the pair $(\mathrm{d}\mu/\mathrm{d}\tilde{\mu}_0, \mathrm{d}\nu/\mathrm{d}\tilde{\nu}_0)$, Proposition 6.3 shows that the pair $(\alpha^*, \beta^*) = (\exp f^*, \exp g^*)$ solves the following system

$$\begin{aligned} \alpha^*(x)\int_F k(x, y)\beta^*(y)\tilde{\nu}_0(\mathrm{d}y) &= \alpha(x) \qquad \tilde{\mu}_0\text{-a.s.} \\ \beta^*(y)\int_E k(x, y)\alpha^*(x)\tilde{\mu}_0(\mathrm{d}x) &= \beta(y) \qquad \tilde{\nu}_0\text{-a.s.} \end{aligned} \tag{6.5}$$

(by convention the left-hand side is equal to 0 whenever α^* (β^*) is equal to 0). This system was introduced by Schrödinger (1931) in the Gaussian real case, as a consequence of a *strange behaviour* of Brownian motion. The strange and *highly improbable* behaviour has a natural explanation in terms of large deviations (see Föllmer, 1988; Cattiaux and Léonard, 1994; 1995a). But the solvability of (6.5) was left open by Schrödinger. Following on from work by Bernstein and Fortet, Beurling (1960) studied this problem in a slightly more general formulation:

> Let k be a non-negative measurable function on $E \times F$. For each pair $(\mu, \nu) \in \mathscr{M}_+^{\mathrm{b}}(E) \times \mathscr{M}_+^{\mathrm{b}}(F)$, does there exist a pair $(\pi_E, \pi_F) \in \mathscr{M}_+^{\mathrm{b}}(E) \times \mathscr{M}_+^{\mathrm{b}}(F)$ such that the marginals of $k(\pi_E \otimes \pi_F)$ are exactly μ and ν? (6.6)

In our notation Beurling's main result is the following (see Beurling 1960, Theorem III, p. 118).

Theorem 6.5 (Beurling's theorem). *Let E and F be locally compact Hausdorff spaces and k be a bounded continuous positive function on $E \times F$ such that $\log k \in L^1(\mu \otimes \nu)$ or, more generally,*

$$\sup_{f \in C_{\mathrm{b}}(E), g \in C_{\mathrm{b}}(F)} \left(\int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\nu - \int \exp(f \oplus g) k \,\mathrm{d}(\mu \otimes \nu) \right) < +\infty.$$

Then, there exists a unique product measure $\alpha^*\mu \otimes \beta^*\nu$ such that the marginals of $k(\alpha^*\mu \otimes \beta^*\nu)$ are μ and ν.

Beurling's proof is variational, but in a different spirit than that of Remark 5.6(i). Remark 5.6(i) and Proposition 6.3 throw light on the probabilistic nature of Beurling's result.

Notice, in particular, that when E and F are compact spaces and k is continuous and positive, the answer to (6.6) is yes, and furthermore the mapping $(\mu, \nu) \to (\pi_E, \pi_F)$ is one to one. Conversely, for (6.6) to hold, it is necessary for k to be positive everywhere.

Problem (6.6) is a key point in the study of the Markov property for reciprocal processes (Jamison, 1974), also called Schrödinger processes (see, for example, Föllmer and Gantert, 1995), which are basic processes in the Euclidean approach of quantum mechanics developed by Zambrini (1989) and others. However, in their recent paper, Föllmer and Gantert (1995) have shown that for infinite-dimensional state spaces, (6.6) is not fully satisfactory for the study of these Schrödinger processes.

Acknowledgements

The first author wishes to thank Hans Föllmer for many fruitful discussions on Schrödinger's problem, and his invitation to Bonn where the foundation stone of this work was laid.

References

Astérisque 68 (1979) *Grandes Déviations et Applications Statistiques*. Paris: S.M.F.

Beurling, A. (1960) An automorphism of product measures. *Ann. Math.*, **72**, 189–200.

Borwein, J.M. and Lewis, A.S. (1992) Decomposition of multivariate functions. *Canad. J. Math.*, **44**, 1–20.

Borwein, J.M., Lewis, A.S. and Nussbaum, R.D. (1994) Entropy minimization, *DAD* problems and doubly-stochastic kernels. *J. Funct. Anal.*, **123**, 264–307.

Cattiaux, P. and Léonard, C. (1994) Minimization of Kullback information for diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **30**, 83–132.

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

Cattiaux, P. and Léonard, C. (1995b) Minimization of Kullback information of some Markov processes. In *Séminaire de Probabilités*, Lecture Notes in Math. 1626, pp. 288–311. New York: Springer-Verlag.

Csiszár, I. (1975) *I*-divergence geometry of probability distributions and minimization problems. *Ann. Probab.*, **3**, 146–158.

Csiszár, I. (1995) Generalized projection for nonnegative functions. *Acta Math. Hungar.*, **68**, 161–185.

Dacunha-Castelle, D. and Gamboa, F. (1990) Maximum d'entropie et problème des moments. *Ann. Inst. H. Poincaré Probab. Statist.*, **26**, 567–596.

Dawson, D.A. and Gärtner, J. (1987) Large deviations from the MacKean–Vlasov limit for weakly interacting diffusions. *Stochastics*, **20**, 247–308.

Dellacherie, C. and Meyer, P.A. (1975) *Probabilités et Potentiel*. Paris: Herman.

Dembo, A. and Zeitouni, O. (1993) *Large Deviations Techniques and Applications*. Boston: Jones and Barblett.

Donsker, M.S. and Varadhan, S.R.S. (1974) Asymptotic evaluation of certain Markov process expectations for large deviations III. *Comm. Pure. Appl. Math.*, **29**, 389–461.

Föllmer, H. (1988) *Random fields and diffusion processes*. In *Lecture Notes in Math. 1362*. New York: Springer-Verlag.

Föllmer, H. and Gantert, N. (1997) Entropy minimization and Schrödinger processes in infinite dimensions. *Ann. Prob.*, **25**, 901–926.

Gamboa, F. and Gassiat, E. (1994) The maximum entropy method on the mean: Applications to linear programming and superresolution. *Math. Programming Ser. A*, **66**, 103–122.

Gamboa, F. and Gassiat, E. (1996) Sets of superresolution and the maximum entropy method on the mean. *SIAM J. Math. Anal.*, **27**, 1129–1152.

Gamboa, F. and Gassiat, E. (1997) Bayesian methods for ill posed problems. *Ann. Statist.*, **25**, 328–350.

Hansel, G. and Troallic, J.P. (1986) Sur le problème des marges. *Probab. Theory Related Fields*, **71**, 357–366.

Hoffmann-Jørgensen, J. (1977) *Probability in Banach spaces*. In *Lecture Notes in Math. 598*. New York: Springer-Verlag.

Jameson, G.J. (1974) *Topology and Normed Spaces*. London: Chapman & Hall.

Jamison, B. (1974) Reciprocal processes. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **30**, 65–86.

Kellerer, H.G. (1961) Funktionen auf Produktraümen mit vorgegebenen Marginalfunktionen. *Math. Ann.*, **144**, 323–344.

Kellerer, H.G. (1964a) Masstheoretische Marginalprobleme. *Math. Ann.*, **153**, 168–198.

Kellerer, H.G. (1964b) Schnittmassfunktionen in mehrfachen Produkträumen. *Math. Ann.*, **155**, 369–391.

Kellerer, H.G. (1984) Duality theorems for marginal problems. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **67**, 399–432.

Lehmann, E.L. (1959) *Testing Statistical Hypotheses*. New York: Wiley.

Lindenstrauss, J. (1965) A remark on extreme doubly stochastic measures. *Amer. Math. Monthly*, **72**, 379–382.

Pratelli, L. (1992) Une caractérisation de la convergence dans L^1. Application aux quasi martingales. In *Séminaire de Probabilités XXVII*. Lecture Notes in Math. New York: Springer-Verlag.

Rao, M.M. and Ren, Z.D. (1991) *Theory of Orlicz Spaces*. New York: Marcel Dekker.

Rockafellar, R.T. (1968) Integrals which are convex functionals. *Pacific J. Math.*, **24**, 525–539.

Rockafellar, R.T. (1971) Integrals which are convex functionals. *Pacific J. Math.*, **39**, 439–469.

Rüschendorff, L. and Thomsen, W. (1994) Note on the Schrödinger equation and *I*-projections. *Statist. Probab. Lett.*, **17**, 369–375.

Schrödinger, E. (1931) Uber die Umkehrung der Naturgesetze. Sonderausgabe aus den Sitzungsberichten der Preussischen Akademie der Wissenschaften. *Phys. Math.*, **144**, 144–153.

Skala, H.J. (1993) Existence of probability measures with given marginals. *Ann. Probab.*, **21**, 136–142.

Strassen, V. (1965) Existence of probability measures with given marginals. *Ann. Math. Statist.*, **36**, 423–439.

Zambrini, J.C. (1989) *Calculus of Variations and Quantum Probability*, Lecture Notes in Control and Inform. Sci. 121. Berlin: Springer-Verlag.

Received May 1996 and revised July 1997.