

Infimum-convolution description of concentration properties of product probability measures, with applications

Paul-Marie Samson

University of Marne-la-Vallée, Laboratoire d'Analyse et de Mathématiques Appliquées, UMR 8050, 5, bd Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée Cedex 2, France

Received 3 October 2005; received in revised form 16 March 2006; accepted 9 May 2006

Available online 18 December 2006

Abstract

This paper is devoted to the concentration properties of product probability measures $\mu = \mu_1 \otimes \cdots \otimes \mu_n$, expressed in term of dimension-free functional inequalities of the form

$$\left(\int e^{\alpha Q_\alpha f} d\mu \right)^{1/\alpha} \left(\int e^{-(1-\alpha)f} d\mu \right)^{1/(1-\alpha)} \leq 1,$$

where α is a parameter, $0 < \alpha < 1$, and $Q_\alpha f$ is an appropriate infimum-convolution operator. This point of view has been introduced by Maurey [B. Maurey, Some deviation inequalities, *Geom. Funct. Anal.* 1 (1991) 188–197]. It has its origins in concentration inequalities by Talagrand where the enlargement of sets is done in accordance with the cost function of the operator $Q_\alpha f$ (see [M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Publ. Math. Inst. Hautes Études Sci.* 81 (1995) 73–205, M. Talagrand, New concentration inequalities in product spaces, *Invent. Math.* 126 (1996) 505–563, M. Talagrand, A new look at independence, *Ann. Probab.* 24 (1996) 1–34]). A main application of the functional inequalities obtained here is optimal deviations inequalities for suprema of sums of independent random variables. As example, we also derive classical deviations bounds for the one-dimensional bin packing problem.

© 2006 Elsevier Masson SAS. All rights reserved.

Résumé

Cet article est consacré à l'étude de propriétés de concentration des probabilités produit $\mu = \mu_1 \otimes \cdots \otimes \mu_n$, en termes d'inégalités fonctionnelles indépendantes de la dimension, de la forme

$$\left(\int e^{\alpha Q_\alpha f} d\mu \right)^{1/\alpha} \left(\int e^{-(1-\alpha)f} d\mu \right)^{1/(1-\alpha)} \leq 1,$$

où α est un paramètre $0 < \alpha < 1$, et $Q_\alpha f$ est un opérateur d'infimum-convolution approprié. Ce point de vue a été introduit par Maurey [B. Maurey, Some deviation inequalities, *Geom. Funct. Anal.* 1 (1991) 188–197]. Il tient ses origines dans des inégalités de concentration de Talagrand, pour lesquelles l'élargissement des ensembles est lié à la fonction de coût de l'opérateur $Q_\alpha f$ (voir [M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Publ. Math. Inst. Hautes Études Sci.* 81 (1995) 73–205, M. Talagrand, New concentration inequalities in product spaces, *Invent. Math.* 126 (1996) 505–563, M. Talagrand, A new look at independence, *Ann. Probab.* 24 (1996) 1–34]). Comme application majeure de ces inégalités fonctionnelles, nous obtenons des inégalités de déviations optimales pour les suprêmes de sommes de variables aléatoires indépendantes. Par ailleurs,

E-mail address: paul-marie.samson@univ-mlv.fr (P.-M. Samson).

à titre d'exemple d'utilisation, nous retrouvons des bornes de déviations pour le problème du rangement de boîtes (bin-packing problem).

© 2006 Elsevier Masson SAS. All rights reserved.

Keywords: Concentration inequalities; Transportation inequalities; Infimum-convolution operator; Empirical processes; Bin packing problem

1. Introduction to the main theorems

The concentration of measure phenomenon on product spaces has been widely investigated by M. Talagrand in [29–31]. In these papers, a large variety of powerful dimension-free concentration inequalities are obtained by induction over the dimension of the product space. Consider a probability space $(E_1, \mathcal{E}_1, \mu_1)$ and its product $E = E_1^n$, $\mu = \mu_1^{\otimes n}$. The basic idea of concentration is that if the measure of a subset $A \subset E$ is not too small ($\mu(A) \geq 1/2$) then most of the points $x \in E$ are “close” to A . Talagrand expressed it in the following form: for a given “distance” $f(x, A)$ between x and A one has

$$\int e^{Kf(x,A)} d\mu(x) \leq \frac{1}{\mu(A)}, \quad (1)$$

where K is a non-negative constant. If $\mu(A)$ is not too small, using Chebyshev inequality, it follows that the set of elements $x \in E$ which are “far” from A is of small measure, since for $t \geq 0$,

$$\mu(\{x \in E, f(x, A) \geq t\}) \leq \frac{1}{\mu(A)} e^{-Kt}.$$

The main issue here is to define an interesting notion of “closeness”. The results of the present paper are connected with the distances associated to the so-called convex hull approach from [29] and [30]. We obtain refinements of Talagrand’s results by extending the so-called infimum-convolution description of concentration introduced by Maurey [21] (see also [5]). One of the main motivation for these investigations is to provide new optimal deviation bounds for suprema of sums of random variables (see Section 3).

Our approach has some of its origins in the so-called “convex property (τ)” of [21] (see also [27]). It is a variant of Maurey’s “property (τ)”. This property was studied by several authors in connection with concentration properties of log-concave measures, as the Gaussian and the exponential measure (see [21,5,4,28]). The “convex property (τ)” is a dimension-free functional inequality which is valid for every product measure $\mu := \mu_1 \otimes \cdots \otimes \mu_n$, when each μ_i is a probability measure on a normed vector space $(F_i, \|\cdot\|)$ supported by a set of diameter less than one. Maurey’s result states: for every convex measurable function f on the product space $F := \prod_{i=1}^n F_i$,

$$\int e^{Qf} d\mu \int e^{-f} d\mu \leq 1.$$

Here, Q is the infimum-convolution operator associated to the quadratic cost function C :

$$Qf(x) := \inf_{y \in F} [f(y) + C(x - y)], \quad x \in F, \quad (2)$$

with $C(z) := \frac{1}{4} \sum_{i=1}^n \|z_i\|^2$, $z = (z_1, \dots, z_n) \in F$.

We define Q_α , a first variant of the operator Q , which is suitable for some abstract probabilistic situation where metric structure is not *a priori* provided. Thus, we extend the “convex property (τ)” from product Banach spaces to any product probability spaces $(E, \mathcal{E}) = \prod_{i=1}^n (E_i, \mathcal{E}_i)$. For further measurability considerations, we assume that every singleton $\{x_i\}$, $x_i \in E_i$, belongs to the σ -field \mathcal{E}_i . Let $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ be a product probability measure on (E, \mathcal{E}) . We establish that for every parameter $\alpha \in (0, 1)$, and for every bounded measurable function f on E one has

$$\left(\int e^{\alpha Q_\alpha f} d\mu \right)^{1/\alpha} \left(\int e^{-(1-\alpha)f} d\mu \right)^{1/(1-\alpha)} \leq 1. \quad (3)$$

This first result is a consequence of a transportation inequality presented further in the introduction (see Theorem 1.2). We observe that (3) will still holds when α equal 0 or 1: when α goes to 1, (3) provides

$$\log \int e^{Q_1 f} d\mu \leq \int f d\mu, \tag{4}$$

and when α goes to 0, it yields

$$\log \int e^{-f} d\mu \leq \int Q_0 f d\mu. \tag{5}$$

In order to define Q_α , we introduce some notations. For a given measurable space (F, \mathcal{F}) , let $\mathcal{P}(F)$ be the set of probability measures on F and $\mathcal{T}(F)$ be the set of all transition probabilities from (F, \mathcal{F}) to (F, \mathcal{F}) . For every $\zeta^{(1)} \in \mathcal{P}(F)$ and $p \in \mathcal{T}(F)$, we define $\zeta^{(1)} \circ p \in \mathcal{P}(F \times F)$ and $\zeta^{(1)} p \in \mathcal{P}(F)$ as follows: for $B \in \mathcal{F} \otimes \mathcal{F}$,

$$\zeta^{(1)} \circ p(B) := \iint \mathbb{1}_B(x, y) p(x, dy) \zeta^{(1)}(dx),$$

and for $A_2 \in \mathcal{F}$,

$$\zeta^{(1)} p(A_2) := \int p(\cdot, A_2) d\zeta^{(1)}.$$

$\zeta^{(1)}$ and $\zeta^{(1)} p$ are the marginals of $\zeta^{(1)} \circ p$. We say that p transports $\zeta^{(1)}$ on $\zeta^{(2)}$ if $\zeta^{(2)} = \zeta^{(1)} p$, and we denote by $\mathcal{T}(\zeta^{(1)}, \zeta^{(2)})$ the set of all transition probabilities that transport $\zeta^{(1)}$ on $\zeta^{(2)}$.

We define a cost function $c_\alpha, \alpha \in (0, 1)$:

$$c_\alpha(\ell) := \frac{\alpha(1 - \ell) \log(1 - \ell) - (1 - \alpha\ell) \log(1 - \alpha\ell)}{\alpha(1 - \alpha)}, \quad \text{for } 0 \leq \ell \leq 1,$$

and $c_\alpha(\ell) := +\infty$ if $\ell > 1$. The family of function $c_\alpha, 0 < \alpha < 1$, dominates the usual quadratic cost function: $c_1(\ell) \geq c_\alpha(\ell) \geq c_0(\ell) \geq \ell^2/2$ for every $\ell \geq 0$. For every $y = (y_1, \dots, y_n) \in E$, let y^i denote the vector $(y_1, \dots, y_i), 1 \leq i \leq n$. For every $x \in E$, we define

$$Q_\alpha f(x) := \inf_{p \in \mathcal{T}(E)} \left\{ \int f(y) p(x, dy) + \int \sum_{i=1}^n c_\alpha \left(\int \mathbb{1}_{x_i \neq y_i} p_i(x_i, dy_i | y^{i-1}) \right) p(x, dy) \right\}, \tag{6}$$

where the infimum is taken over all transition probabilities p in $\mathcal{T}(E)$ that can be written as

$$p(x, dy) = p_1(x_1, dy_1) \circ p_2(x_2, dy_2 | y_1) \circ \dots \circ p_n(x_n, dy_n | y^{n-1}), \tag{7}$$

where $p_i(\cdot, \cdot | y^{i-1}) \in \mathcal{T}(E_i)$ and

$$p_i(x_i, dy_i | y^{i-1}) := \lambda_i(y^{i-1}) \delta_{x_i}(dy_i) + (1 - \lambda_i(y^{i-1})) \nu_i(dy_i), \tag{8}$$

with λ_i being a measurable function on $\prod_{j=1}^{i-1} (E_j, \mathcal{E}_j)$ satisfying $0 \leq \lambda_i(y^{i-1}) \leq 1$, and ν_i being a probability measure on E_i absolutely continuous with respect to μ_i ($\mu_i \ll \nu_i$), for which $\nu_i(\{x_i\}) = 0$. Actually $p_i(x_i, \cdot | y^{i-1})$ is a convex combination of ν_i and the Dirac measure at point x_i . Choosing $p_i(x_i, dy_i | y^{i-1}) = \delta_{x_i}(dy_i)$ for all $1 \leq i \leq n$ in the definition (6), we observe that $Q_\alpha f \leq f$.

For a better understanding, let us consider the one-dimensional definition of Q_α : for $n = 1$,

$$Q_\alpha f(x) := \inf_p \left\{ \int f(y) p(x, dy) + c_\alpha \left(\int \mathbb{1}_{x \neq y} p(x, dy) \right) \right\},$$

with

$$p(x, dy) := \lambda \delta_x + (1 - \lambda) \nu, \quad 0 \leq \lambda \leq 1. \tag{9}$$

Comparing this definition with the usual definition (2), we see that $f(y)$ has been replaced by $\int f(y) p(x, dy)$, and the cost $C(x - y)$ by $c_\alpha(\int \mathbb{1}_{x \neq y} p(x, dy))$, the cost to pay to move from the initial position x . This cost only depends on the probability $1 - \lambda$ to move from x and is independent of the way to move from x given by ν .

A second result of this paper is a functional inequality of the form (3) with a new operator R_α for which the cost term depends on both $(1 - \lambda)$ and the measure ν . More precisely, in dimension one, we set for every $x \in E$,

$$R_\alpha f(x) := \inf_p \left\{ \int f(y) p(x, dy) + \int d_\alpha \left(\mathbb{1}_{x \neq y} \frac{dp(x, y)}{d\mu(y)} \right) d\mu(y) \right\} \tag{10}$$

and in dimension n we set

$$R_\alpha f(x) := \inf_{p \in \mathcal{T}(E)} \left\{ \int f(y) p(x, dy) + \int \sum_{i=1}^n \int d_\alpha \left(\mathbb{1}_{x_i \neq y_i} \frac{dp_i(x_i, y_i | y^{i-1})}{d\mu_i(y_i)} \right) d\mu_i(y_i) p(x, dy) \right\}, \tag{11}$$

where the infimum is taken over all transition probabilities p in $\mathcal{T}(E)$ such that (7) and (8) hold. The function d_α is defined by its convex Legendre-transform d_α^* : $d_\alpha(\ell) := \sup_{h \in \mathbb{R}} [h\ell - d_\alpha^*(h)]$, $\ell \in \mathbb{R}$, where

$$d_\alpha^*(h) := \frac{\alpha e^{(1-\alpha)h} + (1-\alpha)e^{-\alpha h} - 1}{\alpha(1-\alpha)}, \quad h \in \mathbb{R}. \tag{12}$$

We notice that $d_\alpha(\ell)$ is equivalent to $\ell^2/2$ at zero. One has

$$\begin{aligned} d_0(\ell) &:= (1 + \ell) \log(1 + \ell) - \ell, \quad \ell \geq 0, \\ d_{1/2}(\ell) &:= 2\ell \log\left(\frac{\ell}{2} + \sqrt{1 + \frac{\ell^2}{4}}\right) - 4\left(\sqrt{1 + \frac{\ell^2}{4}} - 1\right), \quad \ell \geq 0, \\ d_1(\ell) &:= (1 - \ell) \log(1 - \ell) + \ell, \quad \text{if } 0 \leq \ell \leq 1, \end{aligned}$$

and $d_1(\ell) := +\infty$ if $\ell > 1$.

Theorem 1.1. *For every bounded measurable function f , for every parameter $0 < \alpha < 1$, one has*

$$\left(\int e^{\alpha R_\alpha f} d\mu \right)^{1/\alpha} \left(\int e^{-(1-\alpha)f} d\mu \right)^{1/(1-\alpha)} \leq 1. \tag{13}$$

The proof of Theorem 1.1 is obtained by induction over n . A similar method works for (3). However, as already mentioned, we deduce (3) from the transportation type inequality (14) below. Let us recall the definition of *the relative entropy* of a measure ν with respect to a measure μ with density $d\nu/d\mu$:

$$\text{Ent}_\mu \left(\frac{d\nu}{d\mu} \right) := \int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu.$$

We define a pseudo-distance C_α between probability measures: for $\zeta^{(1)}, \zeta^{(2)} \in \mathcal{P}(E)$:

$$C_\alpha(\zeta^{(1)}, \zeta^{(2)}) := \inf_{p \in \mathcal{T}(\zeta^{(1)}, \zeta^{(2)})} \iint \sum_{i=1}^n c_\alpha \left(\int \mathbb{1}_{x_i \neq y_i} p_i(x_i, dy_i | y^{i-1}) \right) p(x, dy) \zeta^{(1)}(dx),$$

where the infimum runs over all p in $\mathcal{T}(\zeta^{(1)}, \zeta^{(2)})$ satisfying (7) and (8).

Theorem 1.2. *For every $\zeta^{(1)}$ and $\zeta^{(2)}$ in $\mathcal{P}(E)$ absolutely continuous with respect to μ , one has*

$$C_\alpha(\zeta^{(1)}, \zeta^{(2)}) \leq \frac{1}{\alpha} \text{Ent}_\mu \left(\frac{d\zeta^{(1)}}{d\mu} \right) + \frac{1}{1-\alpha} \text{Ent}_\mu \left(\frac{d\zeta^{(2)}}{d\mu} \right). \tag{14}$$

The connection between (3) and (14) is obtained following the lines of [4]. We easily check that

$$\int Q_\alpha f d\zeta^{(1)} - \int f d\zeta^{(2)} \leq C_\alpha(\zeta^{(1)}, \zeta^{(2)}).$$

Then (3) follows by applying Theorem 1.2 and by choosing $\zeta^{(1)}$ and $\zeta^{(2)}$ with respective densities

$$\frac{d\zeta^{(1)}}{d\mu} = \frac{e^{\alpha Q_\alpha f}}{\int e^{\alpha Q_\alpha f} d\mu} \quad \text{and} \quad \frac{d\zeta^{(2)}}{d\mu} = \frac{e^{-(1-\alpha)f}}{\int e^{-(1-\alpha)f} d\mu}.$$

Remarks.

1. Theorem 1.2 improves the transportation inequality of Marton for contracting Markov chains (see [19]), when we restrict the study to product probability measures. Marton introduces a distance \bar{d}_2 defined by

$$n(\bar{d}_2(\zeta^{(1)}, \zeta^{(2)}))^2 := \inf_{p \in \mathcal{T}(\zeta^{(1)}, \zeta^{(2)})} \iint \sum_{i=1}^n \left(\int \mathbb{1}_{x_i \neq y_i} p_i(x_i, dy_i | y^{i-1}) \right)^2 p(x, dy) \zeta^{(1)}(dx).$$

Since $c_\alpha(\ell) \geq \ell^2/2, \ell \geq 0$, one has $n(\bar{d}_2(\zeta^{(1)}, \zeta^{(2)}))^2 \leq 2C_\alpha(\zeta^{(1)}, \zeta^{(2)})$. Then, Marton’s inequality follows by optimizing over α in (14).

2. A first byproduct of (3) and Theorem 1.1 is Talagrand’s concentration inequalities of the form (1). By monotone convergence, we extend (3) to any real measurable function f on E satisfying $\int e^{-(1-\alpha)f} d\mu < \infty$. Then by adopting the convention $(+\infty) \times 0 \leq 1$, (3) is extended to all f taking values in $\mathbb{R} \cup \{+\infty\}$, for which $\int e^{-(1-\alpha)f} d\mu < \infty$ holds. For every measurable set $A \in \mathcal{E}$, let $\phi_A(x) := 0$ if $x \in A$ and $\phi_A(x) := +\infty$ otherwise. Then, for every $x \in E$, one has

$$Q_\alpha \phi_A(x) = \inf_{p, p(x,A)=1} \int \sum_{i=1}^n c_\alpha \left(\int \mathbb{1}_{x_i \neq y_i} p_i(x_i, dy_i | y^{i-1}) \right) p(x, dy).$$

The infimum is taken over all p in $\mathcal{T}(E)$ which transport the mass on A , that is: $p(x, A) = 1$. This defines a pseudo-distance C_α between x and A : $C_\alpha(x, A) := Q_\alpha \phi_A(x)$. Applying (3) with the function ϕ_A , we get Theorem 4.2.4 of [29]:

$$\int e^{\alpha C_\alpha(x,A)} d\mu(x) \leq \frac{1}{\mu(A)^{\alpha/(1-\alpha)}}. \tag{15}$$

The same argument with the functional inequality (13) gives

$$\int e^{\alpha D_\alpha(x,A)} d\mu(x) \leq \frac{1}{\mu(A)^{\alpha/(1-\alpha)}},$$

where

$$D_\alpha(x, A) := \inf_{p, p(x,A)=1} \int \sum_{i=1}^n \int d_\alpha \left(\mathbb{1}_{x_i \neq y_i} \frac{dp_i(x_i, y_i | y^{i-1})}{d\mu_i(y_i)} \right) d\mu_i(y_i) p(x, dy).$$

For the careful reader of [30], this new result improves Theorem 4.2 of [30] and Theorem 1 of [22].

A simple procedure to reach deviation tails for a given function f is the following. If f is sufficiently “regular”, $Q_\alpha f$ or $R_\alpha f$ can be estimated from below, and (3) or (13) provide bounds for the Laplace transform of f . With the operator Q_α , the regularity of f is given by the “best” non-negative measurable functions $h_i : E \rightarrow \mathbb{R}^+ \cup \{+\infty\}$, $1 \leq i \leq n$, such that

$$f(x) - f(y) \leq \sum_{i=1}^n h_i(x) \mathbb{1}_{x_i \neq y_i} \quad \text{for } \mu\text{-almost every } x, y \in E. \tag{16}$$

On the right-hand side, the weighted Hamming distance measures how far $f(x)$ is from $f(y)$. Let c_α^* be the Legendre-transform of c_α . If (16) holds, then

$$Q_\alpha f(x) \geq f(x) - \sum_{i=1}^n c_\alpha^*(h_i(x)) \quad \text{for } \mu\text{-almost every } x \in E. \tag{17}$$

Actually, the “best” functions h_i ’s minimize the quantity $\sum_{i=1}^n c_\alpha^*(h_i(x))$. Similarly, with R_α , the regularity of f is given by the “best” $h_i : E \times E_i \rightarrow \mathbb{R}^+ \cup \{+\infty\}$, $1 \leq i \leq n$, such that

$$f(x) - f(y) \leq \sum_{i=1}^n h_i(x, y_i) \mathbb{1}_{x_i \neq y_i} \quad \text{for } \mu\text{-almost every } x, y \in E. \tag{18}$$

This implies: for μ -almost every x in E

$$R_\alpha f(x) \geq f(x) - \sum_{i=1}^n \int d_\alpha^*(h_i(x, y_i)) d\mu_i(y_i), \tag{19}$$

and the “best” h_i ’s minimize the quantity $\sum_{i=1}^n \int d_\alpha^*(h_i(x, y_i)) d\mu_i(y_i)$.

At the beginning of Section 3, we show how to derive deviation tails from (3) and (17) when f represents a supremum of sum of non-negative random variables (see Corollary 3.3). This is a generic simple example. Then, as a main application of this paper, the same procedure with (13) and (19) provides deviation’s bounds when the variables of the sums are not necessarily non-negative (see Corollaries 3.5 and 3.6). In this matter, the entropy method (developed by [11,1]) has been first used by Ledoux [15] and then by several authors to improve the results by Talagrand, and to reach optimal constants (see [7,8,6,9,10,13,14,16,20,25,26]). The entropy method is a general procedure that yields deviation inequalities from a logarithmic Sobolev type inequality via a differential inequality on Laplace transforms. This is well known as the Herbst argument. Our approach is an alternative to the entropy method. For the suprema of empirical sums, Theorems 3.1 and 3.4 provide exponential deviations tails that improve in some way the ones obtained with the entropy method. To be complete, recall that Panchenko [23] introduced a symmetrization approach that allows to get other concentration results for suprema of empirical processes from (3).

In Section 4.1, we recover from (3) the Talagrand’s deviation tails (around the mean instead of the median) for the one-dimensional bin packing problem.

2. Dimension-free functional inequalities for product measures

The first part of this section is devoted to the proof of Theorem 1.2, and the second part to the proof of Theorem 1.1.

2.1. A transportation type inequality

Theorem 1.2 is obtained by tensorization of the one-dimensional case. Let $\zeta^{(1)} \neq \zeta^{(2)}$ be two probability measures on (E, \mathcal{E}) with respective densities a and b with respect to μ .

For $n = 1$, the proof of (14) is based on the construction of an optimal transition probability $p^* \in \mathcal{T}(\zeta^{(1)}, \zeta^{(2)})$. Let $[c]_+ = \max(c, 0)$, $c \in \mathbb{R}$. For every $x \in E$ with $a(x) \neq 0$, we set

$$p^*(x, \cdot) := \lambda(x)\delta_x(\cdot) + (1 - \lambda(x))\nu(\cdot), \tag{20}$$

with

$$\lambda(x) := \inf\left(1, \frac{b}{a}(x)\right), \quad \text{and} \quad \nu := \frac{[b - a]_+}{\|\zeta^{(1)} - \zeta^{(2)}\|_{\text{TV}}} \mu,$$

where $\|\zeta^{(1)} - \zeta^{(2)}\|_{\text{TV}}$ denotes the total variation distance between $\zeta^{(1)}$ and $\zeta^{(2)}$,

$$\|\zeta^{(1)} - \zeta^{(2)}\|_{\text{TV}} := \frac{1}{2} \int |a - b| d\mu = \int [a - b]_+ d\mu.$$

If $a(x) = 0$, $p^*(x, \cdot)$ is any probability measure in $\mathcal{P}(E)$.

Lemma 2.1. *According to the above definitions, one has*

$$C_\alpha(\zeta^{(1)}, \zeta^{(2)}) = \int c_\alpha \left(\int \mathbb{1}_{x \neq y} p^*(x, dy) \right) \zeta^{(1)}(dx),$$

and

$$\int c_\alpha \left(\int \mathbb{1}_{x \neq y} p^*(x, dy) \right) \zeta^{(1)}(dx) \leq \frac{1}{\alpha} \text{Ent}_\mu \left(\frac{d\zeta^{(1)}}{d\mu} \right) + \frac{1}{1 - \alpha} \text{Ent}_\mu \left(\frac{d\zeta^{(2)}}{d\mu} \right).$$

Proof of Lemma 2.1. For any $p \in \mathcal{T}(\zeta^{(1)}, \zeta^{(2)})$, we consider the measure $\zeta := \zeta^{(1)} \otimes p$, $\zeta \in \mathcal{P}(E \times E)$. Let Δ denote the diagonal of $E \times E$ and ζ_Δ be the restriction of ζ to Δ ,

$$\zeta_\Delta(A) = \zeta(\Delta \cap (A \times A)) := \zeta(\{(x, x), x \in A\}) \quad \text{for every } A \in \mathcal{E}.$$

Since $\zeta \in \mathcal{M}(\zeta^{(1)}, \zeta^{(2)})$, we have $\zeta_\Delta = c\mu$ with $c \leq \inf(a, b)$. Consequently, since c_α is increasing on \mathbb{R}^+ , we get

$$\int c_\alpha \left(1 - \frac{\inf(a, b)}{a}\right) \zeta^{(1)}(dx) \leq \int c_\alpha \left(\int \mathbb{1}_{x \neq y} p(x, dy)\right) \zeta^{(1)}(dx).$$

This lower bound is reached for p^* . Indeed, setting $\zeta^* = \zeta^{(1)} \otimes p^*$, from the definition (20) one has $\zeta_\Delta^* = \inf(a, b)\mu$. The function c_α satisfies for every $u, v > 0$,

$$c_\alpha \left(1 - \frac{v}{u}\right) \leq \frac{1}{\alpha u} c_0(1 - u) + \frac{1}{(1 - \alpha)u} c_0(1 - v).$$

Consequently integrating with respect to $\zeta^{(1)} = a\mu$ and using the identity

$$\int c_0 \left(1 - \frac{d\zeta^{(i)}}{d\mu}\right) d\mu = \text{Ent}_\mu \left(\frac{d\zeta^{(i)}}{d\mu}\right),$$

for $i = 0$ or 1 , we get Lemma 2.1. \square

Let us now consider the n -dimensional case: $(E, \mathcal{E}) := \prod_{i=1}^n (E_i, \mathcal{E}_i)$ and $\mu := \mu_1 \otimes \dots \otimes \mu_n$. Let us first introduce some notations. For every $1 \leq k \leq n$, let $\zeta^{(1)k}$ denote the marginal of $\zeta^{(1)}$ on $(E^k, \mathcal{E}^k) := \prod_{i=1}^k (E_i, \mathcal{E}_i)$. The density of $\zeta^{(1)k}$ with respect to $\mu_1 \otimes \dots \otimes \mu_k$ is

$$a^k(x^k) := \int a(x^k, y_{k+1}, \dots, y_n) \mu_{k+1}(dy_{k+1}) \dots \mu_n(dy_n).$$

One has $\zeta^{(1)n} = \zeta^{(1)}$ and $a^n = a$. Let $\zeta_k^{(1)}(\cdot | x^{k-1})$ denote the probability measure on (E_k, \mathcal{E}_k) with density

$$a^k(x_k | x^{k-1}) := \frac{a^k(x^k)}{a^{k-1}(x^{k-1})},$$

with respect to μ_k . One has $\zeta^{(1)k} = \zeta_1^{(1)} \otimes \dots \otimes \zeta_k^{(1)}$. The same notations are still available with $\zeta^{(2)}$ with its density b : $\zeta^{(2)k} = \zeta_1^{(2)} \otimes \dots \otimes \zeta_k^{(2)}$.

For every vectors x, y in E , let us consider a sequence of transition probabilities $p_k(\cdot, \cdot | x^{k-1}, y^{k-1}) \in \mathcal{T}(\zeta_k^{(1)}(\cdot | x^{k-1}), \zeta_k^{(2)}(\cdot | y^{k-1}))$, $1 \leq k \leq n$. Define

$$\zeta_k(dx_k, dy_k | x^{k-1}, y^{k-1}) := \zeta_k^{(1)}(dx_k | x^{k-1}) \otimes p_k(x_k, dy_k | x^{k-1}, y^{k-1}),$$

one has $\zeta_k(\cdot, \cdot | x^{k-1}, y^{k-1}) \in \mathcal{M}(\zeta_k^{(1)}(\cdot | x^{k-1}), \zeta_k^{(2)}(\cdot | y^{k-1}))$, and ζ_k is a transition probability from $(\prod_{i=1}^{k-1} (E_i, \mathcal{E}_i)) \times (\prod_{i=1}^{k-1} (E_i, \mathcal{E}_i))$ to $(E_k, \mathcal{E}_k) \times (E_k, \mathcal{E}_k)$. The marginals of $\zeta^k := \zeta_1 \otimes \dots \otimes \zeta_k$ are $\zeta^{(1)k}$ and $\zeta^{(2)k}$. Setting

$$p^k(x^k, dy^k) := p_1(x_1, dy_1) \otimes \dots \otimes p_k(x_k, dy_k | x^{k-1}, y^{k-1}),$$

we exactly have $\zeta^k = \nu^{(1)k} \otimes p^k$ and $p^k \nu^{(1)k} = \nu^{(2)k}$. By definition, we say that the transition probability $p(x, dy) := p^n(x^n, dy^n)$ is a well-linked sequence of the transition probabilities p_k , $1 \leq k \leq n$. One has $p \in \mathcal{T}(\zeta^{(1)}, \zeta^{(2)})$.

To get Theorem 1.2, we define p^* as a well-linked sequence of optimal transition probabilities $p_k^*(\cdot, \cdot | x^{k-1}, y^{k-1}) \in \mathcal{T}(\zeta_k^{(1)}(\cdot | x^{k-1}), \zeta_k^{(2)}(\cdot | y^{k-1}))$ defined as in (20). And we show that

$$\iint \sum_{i=1}^n c_\alpha \left(\int \mathbb{1}_{x_i \neq y_i} p_i^*(x_i, dy_i | y^{i-1}, x^{i-1})\right) p^*(x, dy) d\zeta^{(1)}(dx) \leq \frac{1}{\alpha} \text{Ent}_\mu \left(\frac{d\zeta^{(1)}}{d\mu}\right) + \frac{1}{\beta} \text{Ent}_\mu \left(\frac{d\zeta^{(2)}}{d\mu}\right).$$

The left-hand side of this inequality is equal to

$$\sum_{i=1}^n \iint c_\alpha \left(\int \mathbb{1}_{x_i \neq y_i} p_i^*(x_i, dy_i | x^{i-1}, y^{i-1})\right) \zeta_i^{(1)}(dx_i | x^{i-1}) \zeta^{*i-1}(dx^{i-1}, dy^{i-1}).$$

Lemma 2.1 ensures that

$$\int c_\alpha \left(\int \mathbb{1}_{x_i \neq y_i} p_i^*(x_i, dy_i | x^{i-1}, y^{i-1}) \right) \zeta_i^{(1)}(dx_i | x^{i-1}) \leq \frac{1}{\alpha} \text{Ent}_{\mu_i} \left(\frac{d\zeta^{(1)}(\cdot | x^{i-1})}{d\mu_i} \right) + \frac{1}{\beta} \text{Ent}_{\mu_i} \left(\frac{d\zeta^{(2)}(\cdot | y^{i-1})}{d\mu_i} \right).$$

The proof of Theorem 1.2 is ended by integrating this inequality with respect to the coupling measure $\zeta^{*i-1} \in \mathcal{M}(\zeta^{(1)i-1}, \zeta^{(2)i-1})$, and then using the classical tensorization property of entropy

$$\text{Ent}_\mu \left(\frac{d\zeta^{(1)}}{d\mu_i} \right) = \sum_{i=1}^n \int \text{Ent}_{\mu_i} \left(\frac{d\zeta^{(1)}(\cdot | x^{i-1})}{d\mu_i} \right) \zeta^{(1)i-1}(dx^{i-1}).$$

2.2. A dimension-free functional inequality

Theorem 1.1 is obtained by induction over the dimension n using a simple contraction argument given at the end of this section.

For $n = 1$, according to (10), one has for every x in E ,

$$R_\alpha f(x) = f(x) - \sup_{0 \leq \lambda \leq 1} \sup_v \int \left[(f(x) - f(y))(1 - \lambda) \frac{dv}{d\mu}(y) - d_\alpha \left((1 - \lambda) \frac{dv}{d\mu}(y) \right) \right] d\mu(y),$$

where $v \ll \mu$ and $v(\{x\}) = 0$. Since d_α^* denotes the Legendre transform of d_α , it follows

$$R_\alpha f(x) \geq f(x) - \int d_\alpha^*([f(x) - f(y)]_+) d\mu(y).$$

Therefore, Theorem 1.1 implies the next statement.

Lemma 2.2. For every bounded measurable function f ,

$$\left(\int e^{\alpha f(x) - \alpha \int d_\alpha^*([f(x) - f(y)]_+) d\mu(y)} d\mu(x) \right)^{1/\alpha} \left(\int e^{-(1-\alpha)f} d\mu \right)^{1/(1-\alpha)} \leq 1. \tag{21}$$

Actually, for $n = 1$, Theorem 1.1 is equivalent to this lemma. To show it, we need the following observation.

Lemma 2.3. Let f be a bounded measurable function on E . There exists a measurable function \bar{f} on E satisfying $\bar{f}(x) \leq f(x)$ and

$$R_\alpha f(x) \leq \bar{f}(x) - \int d_\alpha^*([\bar{f}(x) - \bar{f}(y)]_+) d\mu(y), \quad \text{for every } x \in E.$$

To simplify the notations, let $\|\cdot\|_\gamma, \gamma \in \mathbb{R}$ denote the $L^\gamma(\mu)$ -norm. For a given function f , Lemma 2.3 ensures that there exists \bar{f} for which

$$\|e^{R_\alpha f}\|_\alpha \leq \left(\int \exp \left[\alpha \bar{f}(x) - \alpha \int d_\alpha^*([\bar{f}(x) - \bar{f}(y)]_+) d\mu(y) \right] d\mu(x) \right)^{1/\alpha}.$$

Then, Lemma 2.2 yields $\|e^{R_\alpha f}\|_\alpha \leq \|e^{\bar{f}}\|_{-\beta}$ with $\beta = 1 - \alpha$, and the proof of Theorem 1.1 for $n = 1$ is finished since $\bar{f} \leq f$.

Proof of Lemma 2.2. Let λ denote the Lebesgue measure on $[0, 1]$. Let h be the increasing repartition function of f under μ : $h(t) := \mu(\{x \in E, f(x) \leq t\}), t \in \mathbb{R}$. Let g denote the inverse repartition function:

$$g(u) := \inf\{t, h(t) \geq u\}, \quad u \in [0, 1].$$

Then, one has

$$\mu(\{x \in E, f(x) \leq t\}) = \lambda(\{u \in [0, 1], g(u) \leq t\}), \quad t \in \mathbb{R},$$

and therefore, (21) is equivalent to

$$\left(\int_0^1 e^{\alpha g(t) - \alpha \int_0^t d_\alpha^*(g(t) - g(s)) ds} dt \right)^{1/\alpha} \left(\int_0^1 e^{-(1-\alpha)g(\omega)} d\omega \right)^{1/(1-\alpha)} \leq 1, \tag{22}$$

where g is an increasing right-continuous function.

We observe that d_α^* is the best non-negative convex function with $d_\alpha^*(0) = 0$, satisfying such an inequality. Applying it to the test functions $g_u(t) := a\mathbb{1}_{(u,1]}(t)$ with $a \geq 0$ and $0 \leq u \leq 1$, we get

$$[u + (1 - u)e^{\alpha a - \alpha d_\alpha^*(a)}]^{1/\alpha} [u + (1 - u)e^{-(1-\alpha)a}]^{1/(1-\alpha)} \leq 1, \quad 0 \leq u \leq 1.$$

When $u \rightarrow 0$, this provides a lower bound for $d_\alpha^*(a)$ which gives exactly its definition (12). The cost function d_α is therefore optimal.

By an elementary approximation argument, we only need to prove (22) for every increasing simple function g with a finite number of values. This is obtained by induction over the number of values of the simple function g . Clearly (22) holds for constant functions g since $d_\alpha^*(0) = 0$. Then, we apply the next induction step.

Proposition 2.4. *Let g be an increasing simple function on $[0, 1]$ that reach its maximum value on $(v, 1]$. For $u \in (v, 1]$ and $a > 0$ let $g_u(t) = a\mathbb{1}_{(u,1]}(t)$, $t \in [0, 1]$. Then, g satisfies (22) implies $g + g_u$ satisfies (22).*

The proof of this proposition is given in Appendix A. \square

Proof of Lemma 2.3. It suffices to find a transition probability p^* satisfying (9) and a function \bar{f} , $\bar{f} \leq f$, for which

$$\int f(y)p^*(x, dy) + \int d_\alpha \left(\mathbb{1}_{x \neq y} \frac{dp^*(x, y)}{d\mu(y)} \right) d\mu(y) = \bar{f}(x) - \int d_\alpha^*([\bar{f}(x) - \bar{f}(y)]_+) d\mu(y). \tag{23}$$

Since d_α^* is convex, the function $\psi_\alpha : u \mapsto \int (d_\alpha^*)'([u - f(y)]_+) d\mu(y)$, is increasing. By differentiating, one has $(d_\alpha^*)'(h) = e^{(1-\alpha)h} - e^{-\alpha h}$, $h \geq 0$. If $u \rightarrow +\infty$, then $\psi_\alpha(u) \rightarrow +\infty$, and if $u \rightarrow -\infty$ then $\psi_\alpha(u) \rightarrow 0$. Therefore, by continuity of ψ_α , there exists a real number f_0 such that

$$\int (d_\alpha^*)'([f_0 - f(y)]_+) d\mu(y) = 1.$$

Let $A := \{x \in E, f(x) \leq f_0\}$. We define \bar{f} by $\bar{f}(x) := f(x)$ if $x \in A$, and $\bar{f}(x) := f_0$ otherwise. Clearly, one has $\bar{f} \leq f$. For a given $x \in E$, we define $\lambda(x)$ by

$$1 - \lambda(x) := \int (d_\alpha^*)'([\bar{f}(x) - \bar{f}(y)]_+) d\mu(y).$$

Since $(d_\alpha^*)'$ is increasing and $(d_\alpha^*)'(0) = 0$, if $x \in A$ then

$$0 \leq 1 - \lambda(x) = \int (d_\alpha^*)'([f(x) - f(y)]_+) d\mu(y) \leq 1,$$

and if $x \notin A$ then $1 - \lambda(x) = \int (d_\alpha^*)'([f_0 - f(y)]_+) d\mu(y) = 1$. Define $p^*(x, dy) = \lambda(x)\delta_x(dy) + (1 - \lambda(x))\nu_x(dy)$, where ν_x is the probability measure with density

$$\frac{d\nu_x}{d\mu}(y) = \frac{(d_\alpha^*)'([\bar{f}(x) - \bar{f}(y)]_+)}{1 - \lambda(x)}.$$

One has $\nu_x(\{x\}) = 0$.

With these definitions, the left-hand side of (23) is

$$\begin{aligned} f(x) - \int \left[(f(x) - f(y))(1 - \lambda(x)) \frac{d\nu_x}{d\mu}(y) - d_\alpha \left((1 - \lambda(x)) \frac{d\nu_x}{d\mu}(y) \right) \right] d\mu(y) \\ = f(x) - \int \left[(f(x) - f(y))(d_\alpha^*)'([\bar{f}(x) - \bar{f}(y)]_+) - d_\alpha((d_\alpha^*)'([\bar{f}(x) - \bar{f}(y)]_+)) \right] d\mu(y). \end{aligned}$$

If $x \in A$, this expression is

$$\bar{f}(x) - \int [(\bar{f}(x) - \bar{f}(y))(d_\alpha^*)'([\bar{f}(x) - \bar{f}(y)]_+) - d_\alpha((d_\alpha^*)'([\bar{f}(x) - \bar{f}(y)]_+))] d\mu(y),$$

that is $\bar{f}(x) - \int d_\alpha^*([\bar{f}(x) - \bar{f}(y)]_+) d\mu(y)$. Next, if $x \notin A$, the same equality holds: from the definition of f_0 , one has

$$\begin{aligned} f(x) - \int (f(x) - f(y))(d_\alpha^*)'([\bar{f}(x) - \bar{f}(y)]_+) d\mu(y) \\ = f(x) - f(x) \int (d_\alpha^*)'([f_0 - f(y)]_+) d\mu(y) + \int f(y)(d_\alpha^*)'([\bar{f}(x) - \bar{f}(y)]_+) d\mu(y) \\ = \int \bar{f}(y)(d_\alpha^*)'([\bar{f}(x) - \bar{f}(y)]_+) d\mu(y) \\ = \bar{f}(x) - \int (\bar{f}(x) - \bar{f}(y))(d_\alpha^*)'([\bar{f}(x) - \bar{f}(y)]_+) d\mu(y). \quad \square \end{aligned}$$

Let us now present the contractivity argument that extends Theorem 1.1 to any dimension n . We sketch the induction step from $n = 1$ to $n = 2$. Then it suffices to repeat the same argument. Let $\mu = \mu_1 \otimes \mu_2$ and $E = E_1 \times E_2$.

We want to show that $\|e^{R_\alpha f}\|_\alpha \leq \|e^f\|_{-\beta}$, with $\beta = 1 - \alpha$. From definition (11), one has for every $(x_1, x_2) \in E$,

$$\begin{aligned} R_\alpha f(x_1, x_2) = \inf_{p_1} \inf_{p_2} \left\{ \int \left[\int f(y_1, y_2) p_2(x_2, dy_2|y_1) \right. \right. \\ \left. \left. + \int d_\alpha \left(\int \mathbb{1}_{x_2 \neq y_2} \frac{dp_2}{d\mu_2}(x_2, y_2|y_1) d\mu_2(y_2) \right) \right] p_1(x_1|dy_1) \right. \\ \left. + \int d_\alpha \left(\int \mathbb{1}_{x_1 \neq y_1} \frac{dp_1}{d\mu_1}(x_1, y_1) \right) d\mu_1(y_1) \right\}. \end{aligned}$$

We easily check that $R_\alpha f = R^{(1)} R^{(2)} f$, with for $g: E_i \rightarrow \mathbb{R}, i = 1, 2$,

$$R^{(i)} g(x_i) := \inf_{p_i} \left\{ \int g(y_i) p_i(x_i, dy_i) + \int d_\alpha \left(\int \mathbb{1}_{x_i \neq y_i} \frac{dp_i}{d\mu_i}(x_i, y_i) \right) d\mu_i(y_i) \right\}.$$

Let $\|\cdot\|_{\gamma(i)}$, denotes the $L^\gamma(\mu_i)$ -norm, $\gamma \in R$. One has

$$\|e^{R_\alpha f}\|_\alpha = \| \|e^{R^{(1)} R^{(2)} f}\|_{\alpha(1)} \|_{\alpha(2)}.$$

Applying Theorem 1.1 with the measure μ_1 and the function $x_1 \mapsto R^{(2)} f(x_1, x_2)$, we get

$$\|e^{R_\alpha f}\|_\alpha \leq \| \|e^{R^{(2)} f}\|_{-\beta(1)} \|_{\alpha(2)} = \left\| \int e^{-\beta R^{(2)} f} d\mu_1 \right\|_{-\frac{\alpha}{\beta}(2)}^{-1/\beta}.$$

For $\gamma < 0$, the function $h \mapsto \|h\|_\gamma$, is concave on the set of positive measurable functions h (see [12]). Therefore, by Jensen it follows that

$$\|e^{R_\alpha f}\|_\alpha \leq \left(\int \|e^{-\beta R^{(2)} f}\|_{-\frac{\alpha}{\beta}(2)} d\mu_1 \right)^{-1/\beta} = \| \|e^{R^{(2)} f}\|_{\alpha(2)} \|_{-\beta(1)}.$$

Then, we apply Theorem 1.1 again with the measure μ_2 and the function $x_2 \mapsto f(x_1, x_2)$, we get $\|e^{R_\alpha f}\|_\alpha \leq \| \|e^f\|_{-\beta(2)} \|_{-\beta(1)} = \|e^f\|_{-\beta}$.

3. Deviation inequalities for suprema of sums of independent random variables

Let \mathcal{F} be a countable set and let $(X_{1,t})_{t \in \mathcal{F}}, \dots, (X_{n,t})_{t \in \mathcal{F}}$ be n independent processes. In this part, we state deviations inequalities for the random variable

$$Z := \sup_{t \in \mathcal{F}} \sum_{i=1}^n X_{i,t}.$$

It is enough to consider a finite set $\mathcal{F} = \{1, \dots, N\}$. The results settled in this section then extend to any countable set by monotone convergence. One has $Z = f(X)$, where $X = ((X_{1,t})_{t \in \mathcal{F}}, \dots, (X_{n,t})_{t \in \mathcal{F}})$ and $f(x) := \sup_{1 \leq t \leq N} \sum_{i=1}^n x_{i,t}$, for $x = (x_1, \dots, x_n) \in (\mathbb{R}^N)^n$ with $x_i = (x_{i,t})_{1 \leq t \leq N}$. For a given x in $(\mathbb{R}^N)^n$, if

$$\tau(x) := \inf \left\{ t \in \mathcal{F}, f(x) = \sum_{i=1}^n x_{i,t} \right\},$$

then $f(x) = \sum_{i=1}^n x_{i,\tau(x)}$. Since $f(y) \geq \sum_{i=1}^n y_{i,\tau(x)}$, $x, y \in (\mathbb{R}^N)^n$, we notice that

$$f(x) - f(y) \leq \sum_{i=1}^n (x_{i,\tau(x)} - y_{i,\tau(x)}) \mathbb{1}_{x_i \neq y_i}.$$

The condition (18) holds setting $h_i(x, y_i) = [x_{i,\tau(x)} - y_{i,\tau(x)}]_+$. If all $x_{i,t}$'s and $y_{i,t}$'s are non-negative, then the condition (16) holds with $h_i(x) = x_{i,\tau(x)}$. These observations are the key point to reach deviation results for Z from (3) and Theorem 1.1 using the estimates (17) and (19).

We first consider the sums of non-negative random variables. Applying (4) and (5) to λf , $\lambda \geq 0$, and using the estimate (17), we get the following basic results.

Theorem 3.1. *Assume $X_{i,t} \geq 0$ for every i and t , then for every $\lambda \geq 0$,*

$$\log \mathbb{E} \left[\prod_{i=1}^n (1 + \lambda X_{i,\tau(X)}) \right] \leq \lambda \mathbb{E}[Z], \tag{24}$$

and

$$\log \mathbb{E}[\exp(-\lambda Z)] \leq \sum_{i=1}^n \mathbb{E}[\exp(-\lambda X_{i,\tau(X)}) - 1]. \tag{25}$$

As first application of this theorem, we recover the well-known bound for the Laplace transform of Z when the $X_{i,t}$'s belongs to $[0, 1]$ (see [20]).

Corollary 3.2. *Assume $X_{i,t} \in [0, 1]$ for every i and t , then for every $\lambda \in \mathbb{R}$*

$$\log \mathbb{E}[\exp(\lambda Z)] \leq \mathbb{E}[Z](e^\lambda - 1).$$

Proof. For $\lambda < 0$, the above inequality is an obvious consequence of (25). Next, for $\lambda > 0$, by convexity of the exponential function, we observe that

$$\log \mathbb{E}[\exp(\lambda Z)] \leq \log \mathbb{E} \left[\prod_{i=1}^n (1 + (e^\lambda - 1) X_{i,\tau(X)}) \right].$$

Then we apply inequality (24). \square

For any $r \in (1, 2]$, define

$$\Sigma_r := \sum_{i=1}^n X_{i,\tau(X)}^r \leq \sup_{t \in \mathcal{F}} \sum_{i=1}^n X_{i,t}^r.$$

When the $X_{i,t}$'s are not upper-bounded, Theorem 3.1 induces the following new results.

Corollary 3.3. *If $\mathbb{E}[\Sigma_r] < \infty$ for some $r \in (1, 2]$, then one has: for any $u \geq 0$,*

$$\mathbb{P} \left[Z \geq \mathbb{E}[Z] + u + \frac{u}{r} \left(\frac{\Sigma_r}{\mathbb{E}[\Sigma_r]} - 1 \right) \right] \leq \exp \left[-\frac{r-1}{r} \left(\frac{u^r}{\mathbb{E}[\Sigma_r]} \right)^{1/(r-1)} \right], \tag{26}$$

and

$$\mathbb{P}[Z \leq \mathbb{E}[Z] - u] \leq \exp\left[-\frac{r-1}{r} \left(\frac{u^r}{\mathbb{E}[\Sigma_r]}\right)^{1/(r-1)}\right]. \tag{27}$$

Proof. Since $e^{-x} + x - 1 \leq (x^r/r)$ for $x \geq 0$ and $r \in (1, 2]$, the inequality (25) implies: for any $\lambda \geq 0$,

$$\log \mathbb{E}[\exp(-\lambda(Z - \mathbb{E}[Z]))] \leq \frac{\lambda^r}{r} \mathbb{E}[\Sigma_r].$$

The inequality (27) then follows from Chebyshev inequality by optimizing over all $\lambda \geq 0$.

From the inequality (24), using $x - \log(1 + x) \leq (x^r/r)$, $x \geq 0$, $r \in]1, 2]$, we also get: for any $\lambda > 0$,

$$\log \mathbb{E}\left[\exp\left(\lambda Z - \frac{\lambda^r}{r} \Sigma_r\right)\right] \leq \lambda \mathbb{E}[Z].$$

Therefore, by Chebyshev inequality, for any $v \geq 0$,

$$\mathbb{P}\left[Z \geq \mathbb{E}[Z] + \frac{\lambda^{r-1}}{r} \Sigma_r + \frac{v}{\lambda}\right] \leq \exp(-v).$$

When Σ_r is close to its mean $\mathbb{E}[\Sigma_r]$, the optimal choice for λ is given by $r v = (r - 1)\lambda^r \mathbb{E}[\Sigma_r]$. Then, the proof of (26) is complete by taking $\lambda^{r-1} = u/\mathbb{E}[\Sigma_r]$. \square

The non-negativity of the $X_{i,t}$'s is a strong restriction that can be relaxed with Theorem 1.1. Using the estimate (19), it gives: for any $\alpha \in (0, 1)$, $\beta = 1 - \alpha$,

$$\mathbb{E}[e^{\alpha\lambda Z - \alpha S_{\alpha,\lambda}}]^{1/\alpha} \mathbb{E}[e^{-\beta\lambda Z}]^{1/\beta} \leq 1, \quad \lambda \geq 0, \tag{28}$$

with

$$S_{\alpha,\lambda} := \sum_{i=1}^n \mathbb{E}[d_{\alpha}^*(\lambda[X_{i,\tau(X)} - X'_{i,\tau(X)}]_+)|X],$$

where $(X'_{i,t})_{t \in \mathcal{F}}$ is an independent copy of $X_i = (X_{i,t})_{t \in \mathcal{F}}$ and $\mathbb{E}[\cdot|X]$ is the conditional expectation given $X = (X_1, \dots, X_n)$. By Chebyshev inequality, it follows that for every $m \in \mathbb{R}$, $v \geq 0$,

$$\mathbb{P}\left[Z \geq m + \frac{v}{\lambda} + \frac{S_{\alpha,\lambda}}{\lambda}\right]^{1-\alpha} \mathbb{P}[Z \leq m]^\alpha \leq e^{-\alpha(1-\alpha)v}. \tag{29}$$

Consequently, it suffices to control the fluctuation of $S_{\alpha,\lambda}^2$ to derive deviation bounds for Z . For $r \in (1, 2]$, we define

$$V_r := \sum_{i=1}^n \mathbb{E}[[X_{i,\tau(X)} - X'_{i,\tau(X)}]_+^r | X].$$

By a Taylor expansion, $S_{\alpha,\lambda}$ is of order $\frac{\lambda^2}{2} V_2$ as $\lambda \rightarrow 0$. V_2 is a variance factor, one has

$$V_2 \leq V := \sup_{t \in \mathcal{F}} \sum_{i=1}^n \mathbb{E}[[X_{i,t} - X'_{i,t}]_+^2 | X_{i,t}].$$

When α or β goes to zero, (28) gives the following main estimates for the Laplace transform of Z .

Theorem 3.4. *For any $\lambda \geq 0$, one has*

$$\log \mathbb{E}[\exp(\lambda Z - S_{1,\lambda})] \leq \lambda \mathbb{E}[Z], \tag{30}$$

and

$$\log \mathbb{E}[\exp(-\lambda(Z - \mathbb{E}[Z]))] \leq \mathbb{E}[S_{0,\lambda}]. \tag{31}$$

For any random variable Y , define the ψ_1 -norm by

$$\|Y\|_{\psi_1} := \inf\{c > 0 \mid \mathbb{E}[\exp(|Y|/c)] \leq 2\}.$$

Corollary 3.5.

(1) If $\mathbb{E}[V_r] < \infty$ for some $r \in (1, 2]$, then for any $u \geq 0$,

$$\mathbb{P}\left[Z \geq \mathbb{E}[Z] + u + \frac{u}{r} \left(\frac{V_r}{\mathbb{E}[V_r]} - 1\right)\right] \leq \exp\left[-\frac{r-1}{r} \left(\frac{u^r}{\mathbb{E}[V_r]}\right)^{1/(r-1)}\right]. \tag{32}$$

(2) If $C_i := \|[X_{i,\tau(X)} - X'_{i,\tau(X)}]_+\|_{\psi_1} < \infty$ for every i , then for any $u \geq 0$,

$$\mathbb{P}[Z \leq \mathbb{E}[Z] - u] \leq \exp\left[-\frac{u^2}{4C^2e^{-1} + Mu}\right], \tag{33}$$

with $C^2 := \sum_{i=1}^n C_i^2$ and $M := \max(C_i)$.

Proof. The inequality (32) is a consequence of (30) using $d_1^*(x) \leq (x^r/r)$, $x \geq 0$, $r \in (1, 2]$, and then following the proof of (26).

Let $Y_i := [X_{i,\tau(x)} - X'_{i,\tau(X)}]_+$, for any $\lambda \geq 0$,

$$\mathbb{E}[d_0^*(\lambda Y_i)] = \int_0^{+\infty} C_i \lambda (e^{C_i \lambda u} - 1) \mathbb{P}[Y_i/C_i \geq u] du.$$

If $0 \leq \lambda < 1/C_i$, then one has

$$\sup_{u \geq 0} \frac{e^{C_i \lambda u} - 1}{e^u} = C_i \lambda \exp \frac{(1 - C_i \lambda) \log(1 - C_i \lambda)}{C_i \lambda} \leq e^{-1} \frac{C_i \lambda}{1 - C_i \lambda},$$

and therefore

$$\mathbb{E}[d_0^*(\lambda Y_i)] \leq e^{-1} C_i^2 \frac{\lambda^2}{1 - C_i \lambda} \int_0^{+\infty} e^u \mathbb{P}[Y_i/C_i \geq u] du = e^{-1} C_i^2 \frac{\lambda^2}{1 - C_i \lambda}.$$

Consequently, for any $0 \leq \lambda < 1/M$,

$$\mathbb{E}[S_{0,\lambda}] \leq e^{-1} C^2 \frac{\lambda^2}{1 - M\lambda}.$$

Finally, (33) follows using (31), applying Chebyshev inequality and then optimizing over all $0 \leq \lambda < 1/M$. \square

The main interest of (29) and Corollary 3.5 is that no boundedness condition are needed on the $X_{i,t}$'s. If they are bounded, the results can be refined as follows.

Corollary 3.6.

(1) Assume that $X_{i,t} \leq M_{i,t}$, and $\mathbb{E}[(M_{i,t} - X_{i,t})^2] \leq 1$, for every i and t , then for any $u \geq 0$,

$$\begin{aligned} \mathbb{P}\left[Z \geq \mathbb{E}[Z] + u\right] &\leq \exp\left[-\frac{u}{2(1 + \varepsilon(u/\mathbb{E}[V]))} \log\left(1 + \frac{u}{\mathbb{E}[V]}\right)\right], \\ &\leq \exp\left[-\frac{u^2}{2\mathbb{E}[V] + 2u}\right], \end{aligned} \tag{34}$$

with $\varepsilon(u) := d_1^*(\log(1 + u))/\log(1 + u)$.

(2) Assume that $m_{i,t} \leq X_{i,t} \leq M_{i,t}$, with $M_{i,t} - m_{i,t} = 1$ for every i and t , then for any $u \geq 0$,

$$\begin{aligned} \mathbb{P}[Z \leq \mathbb{E}[Z] - u] &\leq \exp\left[-\mathbb{E}[V_2]d_0\left(\frac{u}{\mathbb{E}[V_2]}\right)\right], \\ &\leq \exp\left[-\frac{u^2}{2\mathbb{E}[V_2] + \frac{2}{3}u}\right], \end{aligned} \tag{35}$$

with $d_0(u) := (1 + u) \log(1 + u) - u$.

Remarks.

1. These results extend the well-known Bennett’s or Bernstein’s inequalities (see [2,3]). These inequalities apply under the assumption of Theorem 3.6 when $\mathcal{F} = \{t_0\}$, $Z = \sum_{i=1}^n X_{i,t_0}$, and $m_{i,t_0} \leq X_{i,t_0} \leq M_{i,t_0}$, with $M_{i,t_0} - m_{i,t_0} = 1$. The Bennett’s inequality asserts

$$\mathbb{P}[Z \geq \mathbb{E}[Z] + u] \leq \exp\left[-\text{Var}(Z)d_0\left(\frac{u}{\sqrt{\text{Var}(Z)}}\right)\right], \quad u \geq 0,$$

or equivalently

$$\mathbb{P}[Z \leq \mathbb{E}[Z] - u] \leq \exp\left[-\text{Var}(Z)d_0\left(\frac{u}{\sqrt{\text{Var}(Z)}}\right)\right], \quad u \geq 0.$$

If $\mathcal{F} = \{t_0\}$ then $\mathbb{E}[V] = \mathbb{E}[V_2] = \text{Var}(Z)$. Therefore, (35) exactly recover the above Bennett inequality. From the Cramér–Chernoff large deviation theorem, we know that d_0 is an optimal rate function, reached when Z approximates a Poisson distribution (for more details, see [20]). In the right-hand side deviation inequality (34), since $\varepsilon(u) \rightarrow 0$ as $u \rightarrow 0$, the rate of deviation is optimal in the moderate deviations bandwidth, and since $\varepsilon(u) \rightarrow 1$ as $u \rightarrow +\infty$, it is suboptimal in the large deviations bandwidth.

2. If $m_{i,t} \leq X_{i,t} \leq M_{i,t}$ with $M_{i,t} - m_{i,t} = 1$, then by using the classical symmetrization and contraction argument of Ledoux and Talagrand (see [17] Lemma 6.3 and Theorem 4.12, see also Lemma 14 [20]), one can show that

$$\mathbb{E}[V] \leq \sigma^2 + 16\mathbb{E}\left[\sup_{t \in \mathcal{F}} \left| \sum_{i=1}^n (X_{i,t} - \mathbb{E}[X_{i,t}]) \right|\right], \tag{36}$$

with $\sigma^2 := \sup_{t \in \mathcal{F}} \sum_{i=1}^n \text{Var}(X_{i,t})$. Therefore $\mathbb{E}[V]$ is often close to the maximal variance σ^2 .

3. If the $X_{i,t}$ ’s are moreover centered and $Z = \sup_{t \in \mathcal{F}} |\sum_{i=1}^n X_{i,t}|$, then (36) yields $\mathbb{E}[V] \leq \sigma^2 + 16\mathbb{E}[Z]$. Therefore, the results of Corollary 3.6 are of the same nature as the recent results of Klein and Rio [14]. A basic difference is that we do not assume that the $X_{i,t}$ ’s are centered. Moreover the proofs to reach deviation bounds from Theorem 1.1 are simple, especially for the left-hand side deviations, as regard to the entropy method used in [14].

4. We observe that other left-hand side deviation’s bounds for centered $X_{i,t}$ ’s in $(-\infty, 1]$ were given by Klein [13] under additional moment conditions.

Proof of Corollary 3.6. If $m_{i,t} \leq X_{i,t} \leq M_{i,t}$ with $M_{i,t} - m_{i,t} = 1$, then $S_{0,\lambda}^2 \leq d_0^*(\lambda)V_2$, $\lambda \geq 0$. Consequently (31) gives

$$\log \mathbb{E}[\exp(-\lambda(Z - \mathbb{E}[Z]))] \leq d_0^*(\lambda)\mathbb{E}[V_2].$$

Then (35) follows by Chebyshev inequality and optimizing over all $\lambda \geq 0$. Finally, it is well known that (35) implies the Bernstein inequality.

For the right-hand side deviations, by Hölder inequality and using (30), we get for any $q > 1$ and $1/p + 1/q = 1$

$$\mathbb{E}[\exp(\lambda(Z - \mathbb{E}[Z]))] \leq \mathbb{E}\left[\exp\left(\frac{p}{q}S_{1,\lambda,q}^2\right)\right]^{1/p}, \quad \lambda \geq 0. \tag{37}$$

Define $\tilde{S}_{1,\lambda}^2 := \sup_{t \in \mathcal{F}} \sum_{i=1}^n Y_{i,t}(\lambda)$ with $Y_{i,t}(\lambda) := \mathbb{E}[d_1^*(\lambda[X_{i,t} - X'_{i,t}]_+)|X_{i,t}]$. We observe that the function $x \mapsto d_1^*(\sqrt{x})$ is concave, therefore since $X_{i,t} \leq M_{i,t}$ and $\mathbb{E}[(M_{i,t} - X_{i,t})^2] \leq 1$, by Jensen inequality,

$$Y_{i,t}(\lambda) \leq d_1^*(\lambda\|M_{i,t} - X'_{i,t}\|_2) \leq d_1^*(\lambda),$$

where $\|\cdot\|_2$ denotes the L^2 -norm. Consequently, applying Corollary 3.2 with $\tilde{S}_{1,\lambda}^2$, by homogeneity, we obtain

$$\log \mathbb{E} \left[\exp \left(\frac{p}{q} S_{1,\lambda,q}^2 \right) \right] \leq \log \mathbb{E} \left[\exp \left(\frac{p}{q} \tilde{S}_{1,\lambda,q}^2 \right) \right] \leq \frac{e^{\frac{p}{q} d_1^*(\lambda q)} - 1}{d_1^*(\lambda q)} \mathbb{E}[\tilde{S}_{1,\lambda,q}^2].$$

Then from (37) and since $\tilde{S}_{1,\lambda}^2 \leq \frac{\lambda^2}{2} V$, we get

$$\log \mathbb{E}[\exp(\lambda(Z - \mathbb{E}[Z]))] \leq q(q-1)\lambda^2 \frac{e^{d_1^*(\lambda q)/(q-1)} - 1}{2d_1^*(\lambda q)} \mathbb{E}[V]. \tag{38}$$

For each $\lambda > 0$, let $q(\lambda)$ denote the largest $q \geq 1$ so that $H(q) = 0$ with

$$H(q) = d_1^*(\lambda q) - q(q-1)\lambda.$$

Such a $q(\lambda)$ exists since $H(1) \geq 0$ and $H(q) \rightarrow -\infty$ as $q \rightarrow +\infty$. For any $\lambda \geq 0$, (38) ensures that

$$\log \mathbb{E}[\exp(\lambda(Z - \mathbb{E}[Z]))] \leq \frac{\lambda}{2} (e^{q(\lambda)\lambda} - 1) \mathbb{E}[V],$$

and by Chebyshev inequality

$$\mathbb{P}[Z \geq \mathbb{E}[Z] + u] \leq \exp \left[-u\lambda + \frac{\lambda}{2} (e^{q(\lambda)\lambda} - 1) \mathbb{E}[V] \right], \quad u \geq 0. \tag{39}$$

One has $\lambda q(\lambda)(q(\lambda) - 1) = d_1^*(\lambda q(\lambda)) \leq q(\lambda)^2 d_1^*(\lambda)$, and therefore

$$\lambda q(\lambda) \left(1 - \frac{d_1^*(\lambda)}{\lambda} \right) \leq \lambda.$$

Since $\frac{d_1^*(\lambda)}{\lambda} \rightarrow 0$, as $\lambda \rightarrow 0$, it follows that $\lambda q(\lambda) \rightarrow 0$ as $\lambda \rightarrow 0$, and we easily check that $\lambda q(\lambda) \rightarrow +\infty$ as $\lambda \rightarrow +\infty$. Consequently, for any $u \geq 0$ there exists λ_u such that

$$u = (e^{q(\lambda_u)\lambda_u} - 1) \mathbb{E}[V].$$

Hence,

$$q(\lambda_u)\lambda_u = \log \left(1 + \frac{u}{\mathbb{E}[V]} \right) \quad \text{and} \quad H(q(\lambda_u)) = 0$$

gives

$$q(\lambda_u) = 1 + \varepsilon \left(\frac{u}{\mathbb{E}[V]} \right).$$

The proof of (34) ends by choosing $\lambda = \lambda_u$ in (39). Then the Bernstein inequality easily follows. \square

4. The one-dimensional bin packing problem

The one-dimensional bin packing problem can be described as follows: given a finite collection of independent random weights X_1, \dots, X_n and a collection of identical bins with capacity C (which exceeds the largest of the weights), what is the minimum number of bins $N = N(X)$ ($X = \{X_i, 1 \leq i \leq n\}$) into which the weights can be placed without exceeding the bin capacity C . To simplify we choose $C = 1$ so that $0 \leq X_i \leq 1, 1 \leq i \leq n$.

Theorem 4.1. Denote $\Sigma^2 = \sum_{i=1}^n X_i^2$. Then for every $u \geq 0$

$$\mathbb{P}[N \geq \mathbb{E}[N] + 1 + u] \leq \exp \left[-\frac{u^2}{16\mathbb{E}[\Sigma^2] + 3u} \right], \tag{40}$$

and

$$\mathbb{P}[N \leq \mathbb{E}[N] - 1 - u] \leq \exp \left[-\frac{u^2}{4\mathbb{E}[\Sigma^2]} \right]. \tag{41}$$

Observe that for $u > n - 1$, $\mathbb{P}(|N - \mathbb{E}[N]| \geq 1 + u) = 0$ and one early result [24], using martingales approach, is

$$\mathbb{P}[|N - \mathbb{E}[N]| \geq u] \leq 2 \exp\left[-\frac{2u^2}{n}\right], \quad u \geq 0.$$

Therefore, the variance factor $\mathbb{E}[\Sigma^2]$ improves the factor n for $u \leq \mathbb{E}[\Sigma^2]$. Obviously, the sum of the variance of the X_i 's is expected instead of $\mathbb{E}[\Sigma^2]$. Up to constants, our result is Talagrand's Theorem 6.1 [29], where the mean is replaced by a median of N . Talagrand's inequality with the median is a consequence of (15). Lugosi [18] also obtains these inequalities from the entropy method, by first recovering some Talagrand's convex distance inequalities (see [18]).

Proof. We first show that for any $x = \{x_1, \dots, x_n\}$, $y = \{y_1, \dots, y_n\}$,

$$N(x) - N(y) \leq 1 + 2 \sum_{i=1}^n x_i \mathbb{1}_{x_i \neq y_i}.$$

To see this, observe that $N(x) \leq N(y) + N(\{x_i, x_i \neq y_i\})$, then that $N(\{x_i, x_i \neq y_i\}) \leq 1 + \lfloor 2 \sum_{i=1}^n x_i \mathbb{1}_{x_i \neq y_i} \rfloor$ (where $\lfloor \cdot \rfloor$ denotes the integer part), since among $\lfloor 2 \sum_{i=1}^n x_i \mathbb{1}_{x_i \neq y_i} \rfloor$ bins, at least one of them is half-empty. Then, we deduce that for any $x = \{x_1, \dots, x_n\}$, $\lambda \geq 0$,

$$Q_\alpha(\lambda N)(x) \geq \lambda N(x) + \lambda + \sum_{i=1}^n c_\alpha^*(2\lambda x_i) \geq \lambda N(x) + \lambda + 2\lambda^2 \sum_{i=1}^n x_i^2.$$

Following the lines of the proof of (27), we get (41). For the proof of (40), inequality (3) for $\alpha = 1$ combined with Cauchy–Schwarz inequality implies

$$\mathbb{E}[\exp(\lambda(N - \mathbb{E}[N]) - \lambda)] \leq \mathbb{E}[\exp \Sigma_{1,4\lambda}^2]^{1/2}, \quad \lambda \geq 0,$$

where $\Sigma_{1,\lambda}^2 := \sum_{i=1}^n c_1^*(\lambda X_i)$. By convexity and since $0 \leq X_i \leq 1$,

$$\mathbb{E}[\exp c_1^*(4\lambda X_i)] \leq 1 + \mathbb{E}[X_i^2](e^{4\lambda} - 4\lambda - 1) \leq \exp(\mathbb{E}[X_i^2]d_0^*(4\lambda)).$$

Then (40) follows by Chebyshev inequality. \square

Acknowledgements

The author thanks the referee for useful comments that contribute to extend some applications. The author is also grateful to Michel Ledoux, Mathieu Meyer and Krzysztof Oleszkiewicz for helpful comments.

Appendix A

Proof of Proposition 2.4. To simplify the notations, d^* denotes d_α^* . For every measurable increasing functions g on $[0, 1]$, the left-hand side of (22) is

$$\left[\int_0^1 \left(\int_0^1 e^{-\beta(g(\omega) - g(t))} d\omega \right)^{\alpha/\beta} e^{-\alpha \int_0^t d^*(g(t) - g(s)) ds} dt \right]^{1/\alpha}.$$

According to Proposition 2.4, we assume that (22) holds for a given simple increasing function g and we want to show that (22) still holds for $g + g_u$, that is: $M(u, a) + N(u, a) \leq 1$, with

$$M(u, a) := \int_0^u \left(\int_0^u e^{-\beta(g(\omega) - g(t))} d\omega + (1 - u)e^{-\beta a} e^{-\beta(g(1) - g(t))} \right)^{\alpha/\beta} e^{-\alpha \int_0^t d^*(g(t) - g(s)) ds} dt,$$

and

$$N(u, a) := \int_u^1 \left(\int_0^u e^{-\beta(g(\omega) - g(1))} d\omega e^{\beta a} + (1 - u) \right)^{\alpha/\beta} e^{-\alpha \int_0^u d^*(g(t) - g(s) + a) ds} dt.$$

We set $A(u) := \int_0^u e^{\beta(g(1)-g(s))} ds$, and $B(u) := \int e^{-\alpha(g(1)-g(s))} ds$. Since g is increasing, and since α and β are non-negative, we observe that $A(u) \geq u \geq B(u) \geq 0$. One has

$$M(u, a) = \int_0^u (A(u) + (1 - u)e^{-\beta a})^{\alpha/\beta} e^{-\alpha(g(1)-g(t))} e^{-\alpha \int_0^t d^*(g(t)-g(s)) ds} dt,$$

and

$$N(u, a) = (1 - u)(A(u)e^{\beta a} + (1 - u))^{\alpha/\beta} e^{-\alpha \int_0^u d^*(g(1)-g(s)+a) ds}.$$

Since (22) holds with the function g , one has $M(u, 0) + N(u, 0) \leq 1$. Therefore, in order to get Proposition 2.4, it suffices to show that

$$(M(u, a) + N(u, a)) - (M(u, 0) + N(u, 0)) \leq 0. \tag{42}$$

The monotonicity properties of the functions g and d^* ensure that for $0 \leq t \leq u$,

$$\int_0^t d^*(g(t) - g(s)) ds \leq \int_0^u d^*(g(1) - g(s)) ds.$$

Consequently, since β and a are non-negative, one has

$$\begin{aligned} M(u, a) - M(u, 0) &\leq \int_0^u [(A(u) + (1 - u)e^{-\beta a})^{\alpha/\beta} - (A(u) + (1 - u))^{\alpha/\beta}] e^{-\alpha(g(1)-g(t))} e^{-\alpha \int_0^t d^*(g(1)-g(s)) ds} dt \\ &= [(A(u) + (1 - u)e^{-\beta a})^{\alpha/\beta} - (A(u) + (1 - u))^{\alpha/\beta}] B(u) e^{-\alpha \int_0^u d^*(g(1)-g(s)) ds}. \end{aligned} \tag{43}$$

Let

$$\Delta(u, a) := A(u) \frac{e^{\beta a} - 1}{\beta} + B(u) \frac{1 - e^{-\alpha a}}{\alpha}.$$

Observing that

$$\int_0^u d^*(g(1) - g(s) + a) ds = \int_0^u d^*(g(1) - g(s)) ds + \Delta(u, a),$$

and using (43), the proof of (42) is reduced to show that

$$\begin{aligned} &B(u)(A(u) + (1 - u)e^{-\beta a})^{\alpha/\beta} - B(u)(A(u) + (1 - u))^{\alpha/\beta} \\ &+ (1 - u)(A(u)e^{\beta a} + (1 - u))^{\alpha/\beta} e^{-\alpha \Delta(u, a)} - (1 - u)(A(u) + (1 - u))^{\alpha/\beta} \leq 0, \end{aligned}$$

or equivalently after some rearrangements,

$$(A(u) + (1 - u)e^{-\beta a})^{1/\beta} (B(u) + (1 - u)e^{\alpha a - \alpha \Delta(u, a)})^{1/\alpha} \leq (A(u) + (1 - u))^{1/\beta} (B(u) + (1 - u))^{1/\alpha}.$$

Setting $\psi(a) := (A(u) + (1 - u)e^{-\beta a})^{1/\beta} (B(u) + (1 - u)e^{\alpha a - \alpha \Delta(u, a)})^{1/\alpha}$, this inequality is exactly $\psi(a) \leq \psi(0)$. Therefore the proof of Proposition 2.4 is complete if ψ is a non-decreasing function.

By differentiating, we get for every $x \geq 0$,

$$\begin{aligned} \frac{\psi'(x)}{\psi(x)} &= \frac{-(1 - u)e^{-\beta x}}{(1 - u)e^{-\beta x} + A(u)} + \frac{(1 - u)(1 - A(u)e^{\beta x} + B(u)e^{-\alpha x})e^{\alpha x - \alpha \Delta(u, x)}}{B(u) + (1 - u)e^{\alpha x - \alpha \Delta(u, x)}} \\ &= \frac{(1 - u)\phi(x)e^{-\alpha \Delta(u, x)}}{((1 - u)e^{-\beta x} + A(u))(B(u) + (1 - u)e^{\alpha x - \alpha \Delta(u, x)})}, \end{aligned}$$

where

$$\begin{aligned}\phi(x) &:= -e^{-\beta x} (B(u) + (1-u)e^{\alpha x - \alpha \Delta(u,x)}) e^{\alpha \Delta(u,x)} + ((1-u)e^{-\beta x} + A(u))(1 - A(u)e^{\beta x} + B(u)e^{-\alpha x}) e^{\alpha x} \\ &= -B(u)e^{-\beta x} e^{\alpha \Delta(u,x)} + B(u)((1-u)e^{-\beta x} + A(u)) + A(u)e^{(\alpha+\beta)x} (ue^{-\beta x} - A(u)).\end{aligned}$$

Since $\Delta(u, x) \geq 0$ for every $u \geq 0$ and $x \geq 0$, it follows: for every $x \geq 0$,

$$\phi(x) \leq -B(u)(ue^{-\beta x} - A(u)) + A(u)e^{(\alpha+\beta)x} (ue^{-\beta x} - A(u)) = (A(u)e^{(\alpha+\beta)x} - B(u))(ue^{-\beta x} - A(u)).$$

Observing that for every $x \geq 0$,

$$A(u)e^{(\alpha+\beta)x} - B(u) \geq A(u) - B(u) \geq 0,$$

and

$$ue^{-\beta x} - A(u) \leq u - A(u) \leq 0,$$

we conclude that $\phi(x) \leq 0$ and therefore $\psi'(x) \leq 0$ for every $x \geq 0$. This ends the proof of Proposition 2.4 \square

References

- [1] S. Aida, T. Masuda, I. Shigekawa, Logarithmic Sobolev inequalities and exponential integrability, *J. Func. Anal.* 126 (1994) 83–101.
- [2] G. Bennett, Probability inequalities for the sum of independent random variables, *J. Amer. Statist. Assoc.* 57 (297) (1962) 33–45.
- [3] Bernstein, Sur une modification de l'inégalité de Tchebichef, *Annals Science Institute SAV. Ukraine, Sect. Math.* I (1924).
- [4] S. Bobkov, F. Gotze, Exponential integrability and transportation cost related to logarithmic Sobolev inequalities, *J. Func. Anal.* 163 (1999) 1–28.
- [5] S. Bobkov, I. Gentil, M. Ledoux, Hypercontractivity of Hamilton–Jacobi equations, *Geom. Funct. Anal.* 10 (2000) 1028–1052.
- [6] S. Boucheron, O. Bousquet, G. Lugosi, P. Massart, Moment inequalities for functions of independent random variables, *Ann. Probab.* 33 (2005) 514–560.
- [7] S. Boucheron, G. Lugosi, P. Massart, A sharp concentration inequality with applications, *Random Structures Algorithms* 16 (2000) 277–292.
- [8] S. Boucheron, G. Lugosi, P. Massart, Concentration inequalities using the entropy method, *Ann. Probab.* 31 (2003) 1583–1614.
- [9] O. Bousquet, A Bennett concentration inequality and its application to suprema of empirical processes, *C. R. Acad. Sci. Paris, Ser. I* 334 (2002) 495–500.
- [10] O. Bousquet, Concentration inequalities for sub-additive functions using the entropy method, *Stochastic Inequalities and Applications* 56 (2003) 213–247.
- [11] E.B. Davies, B. Simon, Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians, *J. Func. Anal.* 59 (1984) 335–395.
- [12] G.H. Hardy, J.E. Littlewood, G. Polya, *Inequalities*, Cambridge Univ. Press, 1964.
- [13] T. Klein, Une inégalité de concentration à gauche pour les processus empiriques, *C. R. Acad. Sci. Paris, Ser. I* 334 (2002) 500–505.
- [14] T. Klein, E. Rio, Concentration around the mean for maxima of empirical processes, *Ann. Probab.* 33 (2005) 1060–1077.
- [15] M. Ledoux, On Talagrand's deviation inequalities for product measures, *ESAIM: Probab. Statist.* 1 (1996) 63–87.
- [16] M. Ledoux, *The Concentration of Measure Phenomenon*, Mathematical Surveys and Monographs, American Mathematical Society, 2001.
- [17] M. Ledoux, M. Talagrand, *Probability in Banach Spaces*, Springer-Verlag, Berlin, 1991.
- [18] G. Lugosi, Concentration-of-measure inequalities, Private communication.
- [19] K. Marton, A measure concentration inequality for contracting Markov chains, *Geom. Funct. Anal.* 6 (1997) 556–571.
- [20] P. Massart, About the constants in Talagrand's concentration inequalities for empirical processes, *Ann. Probab.* 28 (2000) 863–884.
- [21] B. Maurey, Some deviation inequalities, *Geom. Funct. Anal.* 1 (1991) 188–197.
- [22] D. Panchenko, A note on Talagrand's concentration inequalities, *Electron. Comm. Probab.* 6 (2001) 55–65.
- [23] D. Panchenko, Symmetrization approach to concentration inequalities for empirical processes, *Ann. Probab.* 31 (2003) 2068–2081.
- [24] W. Rhee, M. Talagrand, A sharp deviation inequality for the stochastic traveling salesman problem, *Ann. Probab.* 17 (1989) 1–8.
- [25] E. Rio, Inégalités de concentration pour les processus empiriques de classes de parties, *Probab. Theory Related Fields* 119 (2000) 163–175.
- [26] E. Rio, Une inégalité de Bennett pour les maxima de processus empiriques, *Ann. Inst. H. Poincaré Probab. Statist.* 38 (6) (2002) 1053–1058.
- [27] P.M. Samson, Concentration inequalities for convex functions on Product Spaces, *Progr. Probab.* 56 (2003) 33–52.
- [28] M. Schmuckenschlager, Private communication.
- [29] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Publ. Math. Inst. Hautes Études Sci.* 81 (1995) 73–205.
- [30] M. Talagrand, New concentration inequalities in product spaces, *Invent. Math.* 126 (1996) 505–563.
- [31] M. Talagrand, A new look at independence, *Ann. Probab.* 24 (1996) 1–34.