POLITECNICO DI TORINO
Repository ISTITUZIONALE

Sub-exponentiality in Statistical Exponential Models

Original
Sub-exponentiality in Statistical Exponential Models / Trivellato, B.. - In: JOURNAL OF THEORETICAL PROBABILITY. - ISSN 0894-9840. - ELETTRONICO. - 37:3(2024), pp. 2076-2096. [10.1007/s10959-023-01281-6]

Availability:
This version is available at: 11583/2981096 since: 2023-08-15T09:48:55Z

Publisher:
Springer

Published
DOI:10.1007/s10959-023-01281-6

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

26 November 2026

Journal of Theoretical Probability (2024) 37:2076–2096
https://doi.org/10.1007/s10959-023-01281-6

Sub-exponentiality in Statistical Exponential Models

Barbara Trivellato[1]

Received: 1 May 2023 / Revised: 2 July 2023 / Accepted: 26 July 2023 /
Published online: 12 August 2023
© The Author(s) 2023

Abstract
Improvements in the study of nonparametric maximal exponential models built on Orlicz spaces are proposed. By exploiting the notion of sub-exponential random variable, we give theoretical results which provide a clearer insight into the structure of these models. The explicit constants we obtain when changing the law of Orlicz spaces centered at connected densities allow us to derive uniform bounds with respect to a reference density.

Keywords Sub-exponential random variable · Orlicz space · Exponential model · Concentration inequality

Mathematics Subject Classification (2020) 46E30 · 46N30 · 62B10

1 Introduction

The theory of nonparametric maximal exponential models centered at a given positive density p started with the seminal work by Pistone and Sempi [18] and is a generalization of the statistical theory of exponential families. An infinite-dimensional or nonparametric exponential family is typically defined by the form $\exp(u - K_p(u))p$, where the random variable u varies in an appropriate function space, p is the probability density function of a base probability $\mathbb{P}$, and $K_p(u)$ is the logarithm of the normalization constant, which is also known as the cumulant generating function. In [18], the sufficient statistics u in the exponential family are such that $\exp(\theta u)$ is $p \cdot \mathbb{P}$-integrable for all θ in a real open interval containing 0; in fact, a further restriction is needed to avoid the border of the set $\{u : K_p(u) < +\infty\}$. This integrability condition, in turn, defines the Banach space of exponential Orlicz spaces $L^{\Phi}(p)$, whose various

✉ Barbara Trivellato
barbara.trivellato@polito.it

[1] Dipartimento di Scienze Matematiche "G.L. Lagrange", Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

equivalent characterizations are the starting point from which the results of this paper arise.

The geometric theory of statistical models as manifolds modeled on exponential Orlicz spaces was deeply investigated in [1, 6, 7, 17]. Specifically, two densities p and q belonging to the maximal exponential model are connected by an *open* exponential arc. (By *open*, we essentially mean that the two connected densities are not the extremal points of the arc.) All densities connected by an open exponential arc form a *Banach manifold* $\mathcal{P}$, and the space $L^{\Phi}(p)$, $p \in \mathcal{P}$, is an expression of the tangent space.

Subsequent upgrades and applications of maximal exponential models to Statistics, Information Geometry, Physics and also to Finance have been presented in many works, see, e.g., [8, 12, 14, 16, 20–25]. Mainly, in [20] the authors prove that the exponential connection by arc for two densities is equivalent to the equality of the corresponding Orlicz spaces. The possibility to switch from one Orlicz space to the other was exploited in [23] to improve some duality results in a problem of exponential utility maximization. Furthermore, since connected densities produce the same Orlicz space with equivalent norms, in [25] robust concentration inequalities of Bernstein type have been obtained.

An equivalent formulation given in [20] for defining the maximal exponential model requires that the ratio q/p has to satisfy with respect to both densities p and q an integrability condition, denoted here by (P_α), with $\alpha > 1$. From (P_α), the finiteness of both the Kullback–Leibler divergences $D(q\|p)$ and $D(p\|q)$ immediately follows.

Different authors have generalized the original structure of the maximal exponential model by replacing the exponential function with deformed exponentials (see, e.g., [11, 28, 29]) or by modeling the statistical manifold on other function spaces (see, e.g., [4, 13]).

In this paper, new analytical properties of the maximal exponential model in the topology of the exponential Orlicz space are presented. The notion of sub-exponential random variable, equivalent to belonging to the Orlicz exponential space, leads to investigate its link with the definition of BMO (bounded mean oscillation) function. In this analysis, we use a suitable norm equivalent to the standard Luxemburg norm, which allows us to obtain a sharp estimate for the distance to L^∞ of a sub-exponential random variable. This estimate, in turn, allows us to better understand the structure underlying the maximal exponential model, as it establishes which is the smallest $\alpha > 1$ for which (P_α) holds. A comparison with the BMO theory highlights that condition (P_α), characterizing the maximal exponential model, is a sort of "static" Muckenhoupt condition on $\log(q/p)$ with respect to both densities p and q. This could be a hint on how to introduce time dependence in the context of maximal exponential models, allowing to change the perspective from static to dynamic.

An interesting issue in statistical applications is the possibility of knowing the constants that the change of law produces on the equivalent norms of Orlicz spaces centered at two connected densities. We provide an answer to this matter and exploit our results to derive uniform bounds.

The paper is organized as follows. In Sect. 2, we investigate the notion of sub-exponential random variable and its link with the distance in the Orlicz topology to L^∞. In Sect. 3, we use the results of the previous section to highlight the structure underlying the maximal exponential model. In Sect. 4, we study the relationships

between Orlicz spaces centered at connected densities, in terms of how their equivalent norms transform. We provide uniform bounds for the norms and some concentration inequalities of Bernstein type that are robust in an exponential subfamily with respect to a reference density.

2 Sub-exponentiality in Orlicz Spaces

In this section, we first describe the relevant Banach spaces: the exponential Orlicz space and its conjugate endowed with the Luxemburg norm (2). We then present the exponential space as a space of sub-exponential random variables with a new norm based of moments, equivalent to the Luxemburg norm, and we investigate its connections to BMO spaces. The second part of the section is partly new and is devoted to the study of a bound on the norm-distance of a generic sub-exponential random variable from L^∞. In fact, it is well known that the space of bounded random variables is not dense in the exponential Orlicz space, unless the state space is finite, cf. [19]. Particularly, the sequence of truncations of a generic element of such a space does not converge in general to the original variable. The bound we obtain characterizes in terms of the moment generating function those random variables for which the truncations converge. It also provides a better understanding of the structure of maximal exponential models in the next section.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a fixed probability space. Denote by L^k, $k \geq 1$, the ordinary Lebesgue spaces and by L^0 the set of all random variables u defined on the probability space.

An exponential Orlicz space is a classical Banach space associated with an exponentially growing Young function. Young functions can be seen as generalizations of the power functions, and consequently, Orlicz spaces are generalizations of the Lebesgue spaces. Specifically,

Definition 2.1 A Young function Φ is an even, convex function $\Phi : \mathbb{R} \to [0, +\infty]$ such that

$$i)\ \Phi(0) = 0,\ ii)\ \lim_{x \to +\infty} \Phi(x) = +\infty,\ iii)\ \Phi(x) < \infty \text{ in a neighborhood of } 0.$$

The *conjugate function* Ψ of Φ is defined as $\Psi(y) = \sup_{x \in \mathbb{R}} \{xy - \Phi(x)\}, \forall y \in \mathbb{R}$ and is itself a Young function.

The *Orlicz space* L^Φ, associated with the Young function Φ, is defined as

$$L^\Phi = \left\{ u \in L^0 \ : \ \exists\, \beta > 0 \, s.t. \, \mathbb{E}\left(\Phi(\beta u)\right) < \infty \right\} \tag{1}$$

, and the corresponding subspace of centered random variables is denoted by L_0^Φ.

L^Φ is a complete Banach space when endowed with the *Luxemburg norm*

$$\|u\|_\Phi = \inf \left\{ \alpha > 0 \ : \ \mathbb{E}\left(\Phi\left(\frac{u}{\alpha}\right)\right) \leq 1 \right\}. \tag{2}$$

We refer to [19] for a general background on the subject.

Here, we focalize on the Young function $\Phi_1(x) = \cosh(x) - 1$, whose conjugate function is $\Psi_1(y) = y\log(y+\sqrt{1+y^2}) - \sqrt{1+y^2}+1$. As a consequence, a random variable u belongs to L^{Φ_1} if and only if its moment generating function $M_u(t) = \mathbb{E}(e^{tu})$ is finite in a neighborhood of 0. Moreover, the closed unit ball of L^{Φ_1} is

$$\mathcal{B}_1 = \{u \in L^{\Phi_1} \ \mathbb{E}(\Phi_1(u)) \leq 1\} = \{u \in L^{\Phi_1} \ \mathbb{E}(\cosh(u)) \leq 2\}.$$

It is worth recalling that the same Orlicz space can be related to different *equivalent* Young functions. Specifically, the Young function $\Phi_1(x)$ is equivalent to the more commonly used $\Phi_2(x) = e^{|x|} - |x| - 1$ and its conjugate function $\Psi_1(y)$ is equivalent to $\Psi_2(y) = (1+|y|)\log(1+|y|) - |y|$, whose analytic expression is related to the Kullback–Leibler divergence.

Remark 2.2 We can easily locate the Orlicz spaces L^{Φ_1} and L^{Ψ_1} in the hierarchy of classical L^k spaces:

$$L^\infty \subseteq L^{\Phi_1} \subset L^k \subseteq L^{\Psi_1} \subset L^1, \qquad k > 1;$$

more precisely, every correspondent injection is continuous. In particular, if $u \in L^{\Phi_1}$, then the moment of u of any order is finite.

The following proposition gives equivalent conditions for a random variable to belong to L^{Φ_1} (cf. [30]).

Proposition 2.3 *Let $u \in L^0$. The following conditions are equivalent:*

1. *$u \in L^{\Phi_1}$, i.e., the moment generating function of u is finite in a neighborhood of 0.*
2. *There is $\lambda > 0$ such that*

$$\mathbb{P}(|u| \geq t) \leq \delta e^{-\lambda t}, \ \exists \delta = \delta(\lambda) \geq 1 \text{ and } \forall t > 0. \tag{3}$$

3. $\sup_{k\geq 1} \left(\mathbb{E}\left(|u|^k\right)/k!\right)^{1/k} < \infty.$
4. $\limsup_{k\to\infty} \left(\mathbb{E}\left(|u|^k\right)/k!\right)^{1/k} < \infty.$

For $u \in L^{\Phi_1}$, the mapping

$$\|u\|_\star = \sup_{k\geq 1} \left(\mathbb{E}\left(\frac{|u|^k}{k!}\right)\right)^{1/k} \tag{4}$$

is a norm on L^{Φ_1} equivalent to $\|u\|_{\Phi_1}$ and it holds (see [24])

$$\frac{2}{3}\|u\|_{\Phi_1} \leq \|u\|_\star \leq 4\|u\|_{\Phi_1}. \tag{5}$$

In particular, it can be easily checked that for constant random variables we have $\|c\|_\star = |c|$ and for any bounded random variable u it holds $\|u\|_\star \leq \|u\|_\infty$.

In the sequel, we will prove our results with respect to the norm $\|\cdot\|_\star$, which turns out to be a convenient choice in order to obtain sharp estimates (see Theorem 2.8).

Remark 2.4 A random variable u satisfying condition *2* in the previous proposition is also called *sub-exponential*. If u satisfies the inequality $\mathbb{P}(|u| \geq t) \leq \delta e^{-\lambda t^2},\ \forall t > 0$, it is called *sub-Gaussian*. It is immediate to see that a sub-Gaussian random variable u is also sub-exponential and that u is sub-Gaussian if and only if u^2 is sub-exponential. More generally, u^n is sub-exponential, i.e., $u^n \in L^{\Phi_1}$, if and only if $\mathbb{P}(|u| \geq t) \leq \delta e^{-\lambda t^n},\ \forall t > 0$.

Remark 2.5 Condition *2* of Proposition 2.3 can be rewritten in the form:

2'. There is $\lambda > 0$ such that $\mathbb{P}(|u| \geq t) \leq e^{-\lambda t},\ \forall t > t_0 = t_0(\lambda) \geq 0$,

which results to be a probabilistic version of the John–Nirenberg inequality for BMO-functions in classical analysis (see [9]). Specifically, we recall that a real-valued locally integrable function f on $\mathbb{R}^d$ has bounded mean oscillation, that is, $f \in \mathrm{BMO}$, if

$$\sup_I \frac{1}{|I|} \int_I |f - f_I|\, dx < +\infty, \tag{6}$$

where I denotes any cube in $\mathbb{R}^d$, $|I|$ its volume and f_I the mean of f over I. John and Nirenberg proved in [9] that $f \in \mathrm{BMO}$ if and only if there is $\lambda > 0$ such that

$$\sup_I \frac{1}{|I|} \left|\{x \in I : |f(x) - f_I| \geq t\}\right| \leq e^{-\lambda t},\ \forall t > t_0 = t_0(\lambda) \geq 0.$$

The probabilistic version of (6) for continuous uniformly integrable martingales $M = (M_s)_{0 \leq s \leq \infty}$ with $M_0 = 0$ is

$$\sup_\tau \left\| \mathbb{E}\left(|M_\infty - M_\tau| \,|\, \mathcal{F}_\tau\right) \right\|_\infty < +\infty,$$

where τ denotes any stopping time for the filtration $\mathcal{F}$. The probabilistic analogue of the John–Nirenberg inequality was proved by Emery [3]: $M \in \mathrm{BMO}$ if and only if there is $\lambda > 0$ such that

$$\mathbb{P}\left(\sup_{s \geq 0} |M_{\tau+s} - M_\tau| \geq \epsilon \,|\, \mathcal{F}_\tau\right) \leq e^{-\lambda \epsilon},\ \forall \epsilon > \epsilon_0 = \epsilon_0(\lambda) \geq 0,$$

for any stopping time τ.

The following proposition establishes the equality between the largest range for which the moment generating function of $|u|$ is finite and the largest λ for which condition (3) holds.

Proposition 2.6 *For $u \in L^{\Phi_1}$, define*

$$a(u) = \sup\{a > 0 \,:\, \mathbb{E}e^{a|u|} < +\infty\}$$

and

$$\lambda(u) = \sup\{\lambda > 0 \,:\, u \text{ satisfies (3)}\}.$$

Then, it holds

$$a(u) = \lambda(u).$$

Proof In order to prove $a(u) \leq \lambda(u)$, we show that

$$\{a > 0 \,:\, \mathbb{E}e^{a|u|} < +\infty\} \subseteq \{\lambda > 0 \,:\, u \text{ satisfies (3)}\}.$$

Let us take $a > 0$ such that $\mathbb{E}e^{a|u|} < +\infty$. By Markov inequality, for all $t > 0$ we have

$$\mathbb{P}(|u| \geq t) = \mathbb{P}(e^{a|u|} \geq e^{at}) \leq e^{-at}\mathbb{E}(e^{a|u|}).$$

We then obtain (3) by taking $\lambda = a$ and $\delta = \mathbb{E}(e^{a|u|})$.
To prove $a(u) \geq \lambda(u)$, suppose by contradiction that $a(u) < \lambda(u)$. We can then select $\lambda_1, \lambda_2 > 0$ such that $a(u) < \lambda_1 < \lambda_2 < \lambda(u)$. Since $\lambda_2 < \lambda(u)$, it holds

$$\mathbb{P}(|u| \geq t) \leq \delta e^{-\lambda_2 t}, \tag{7}$$

for a suitable $\delta \geq 1$. Using (7), for any $k = 1, 2...$ we have

$$\begin{aligned}\mathbb{E}(|u|^k) &= \int_0^{+\infty} \mathbb{P}(|u|^k > t)\, \mathrm{d}t = \int_0^{+\infty} \mathbb{P}(|u| > s)\, ks^{k-1}\, \mathrm{d}s \\ &\leq k\delta \int_0^{+\infty} s^{k-1} e^{-\lambda_2 s}\, \mathrm{d}s = k!\frac{\delta}{\lambda_2^k}.\end{aligned}$$

As a consequence, we get

$$\mathbb{E}e^{\lambda_1|u|} = \sum_{k=0}^{+\infty} \frac{\lambda_1^k}{k!}\mathbb{E}(|u|^k) \leq \delta \sum_{k=0}^{+\infty} \left(\frac{\lambda_1}{\lambda_2}\right)^k = \delta \frac{1}{1 - \frac{\lambda_1}{\lambda_2}} < +\infty,$$

which implies $\lambda_1 \leq a(u)$. □

The next result is a technical preliminary to our main results.

Lemma 2.7 *Let $u \in L^{\Phi_1}$. If $\|u\|_\star < 1$, then*

$$\mathbb{E}e^{|u|} \leq \frac{1}{1 - \|u\|_\star}.$$

Proof By the definition (4) of $\|\cdot\|_\star$, we have

$$\mathbb{E}e^{|u|} \le \sum_{k=0}^{\infty} \frac{1}{k!}\mathbb{E}(|u|^k) \le \sum_{k=0}^{\infty} \|u\|_\star^k = \frac{1}{1-\|u\|_\star}.$$

□

The following theorem gives a very nice relation between $a(u)(=\lambda(u))$ and $d_\star(u, L^\infty)$, where

$$d_\star(u, L^\infty) = \inf_{\ell \in L^\infty} \|u-\ell\|_\star.$$

It is in the spirit of the results obtained in classical and stochastic analysis for the distance in BMO to L^∞, where upper and lower bounds for $a(u)$ in terms of $d_{\mathrm{BMO}}(u, L^\infty)$ are given (see, e.g., [5] and, for continuous martingales, [27]). In our setting, we can prove that $a(u)$ is exactly the reciprocal of the $d_\star$-distance of $u \in L^{\Phi_1}$ to L^∞.

Theorem 2.8 *For $u \in L^{\Phi_1}$, we have*

$$a(u) = \frac{1}{d_\star(u, L^\infty)}$$

Proof We first prove that $a(u) \ge \frac{1}{d_\star(u,L^\infty)}$. Let $0 < a < \frac{1}{d_\star(u,L^\infty)}$. Then, $a\|u-\ell\|_\star < 1$ for some $\ell \in L^\infty$. Using Lemma 2.7, we get

$$\mathbb{E}e^{a|u|} \le \mathbb{E}\left(e^{a|u-\ell|}e^{a|\ell|}\right) \le e^{a\|\ell\|_\infty}\mathbb{E}e^{a|u-\ell|} \le \frac{e^{a\|\ell\|_\infty}}{1-a\|u-\ell\|_\star} < +\infty,$$

which implies $a \le a(u)$. We now prove that $a(u) \le \frac{1}{d_\star(u,L^\infty)}$. It suffices to verify that if $0 < a < a(u)$, then $a \le \frac{1}{d_\star(u,L^\infty)}$. Let us fix $0 < a < a(u)$. For any $\ell \in L^\infty$, it holds

$$d_\star(u, L^\infty) \le \|u-\ell\|_\star = \sup_{k\ge 1}\left(\mathbb{E}\left(\frac{|u-\ell|^k}{k!}\right)\right)^{1/k} \le \frac{1}{a}\,\mathbb{E}e^{a|u-\ell|},$$

where the last inequality is due to the fact that for any $a > 0$ and $k \ge 1$ it holds $\frac{\mathbb{E}|v|^k}{k!} \le \frac{1}{a^k}\mathbb{E}e^{a|v|}$ and thus

$$\left(\mathbb{E}\left(\frac{|u-\ell|^k}{k!}\right)\right)^{1/k} \le \frac{1}{a}\left(\mathbb{E}e^{a|u-\ell|}\right)^{1/k} \le \frac{1}{a}\,\mathbb{E}e^{a|u-\ell|}.$$

Now, let us choose the sequence of truncations at n of u, i.e., $u_n = 1_{|u|\le n}u \in L^\infty$; then, the sequence $|u-u_n| = 1_{|u|>n}|u|$ decreases to 0. Since by hypothesis $a < a(u)$, the random variables $e^{a|u-u_n|}$ are bounded by the integrable random variable $e^{a|u|}$. Applying the Lebesgue dominated convergence theorem, we deduce that $\lim_{n\to+\infty}\mathbb{E}e^{a|u-u_n|} = 1$ and thus $d_\star(u, L^\infty) \le \frac{1}{a}$. □

Remark 2.9 From the theorem above, we immediately see that the sequence of truncations u_n converges to the original variable $u \in L^{\Phi_1}$ if and only if the moment generating function of u is defined on the whole real line (cf. [1], Lemma 2).

In the sequel, we investigate the relationships among L^{Φ_1} spaces with respect to different probability measures whose densities with respect to $\mathbb{P}$ belong to an *open* exponential model. For this purpose, we need to recall the notion of open exponential model, as well as its geometric meaning and the corresponding main results within the Orlicz framework.

3 Exponential Models

This section discusses a crucial topic in the geometric theory of statistical models. Here, the main new contributions are represented by Items 6 and 8 of Theorem 3.5 and by Proposition 3.8.

Let $\mathcal{P}$ denote the set of all densities which are positive $\mathbb{P}$-a.s. For each fixed $p \in \mathcal{P}$, we use $\mathbb{E}_p$ to indicate the integral with respect to $p \cdot \mathbb{P}$. Moreover, the corresponding Orlicz space associated with Φ_1 is denoted by $L^{\Phi_1}(p)$.

Let us consider the cumulant generating functional $K_p(u) = \log \mathbb{E}_p(e^u)$ defined on the subspace of centered random variables $L_0^{\Phi_1}(p)$. We recall from [18] that K_p is a positive convex and lower semicontinuous function, vanishing at zero. In addition, the interior of its proper domain, denoted here by $\overset{\circ}{\text{dom}}\, K_p$, is a non-empty convex set containing the open unit ball of $L_0^{\Phi_1}(p)$. This allows us to give the following definition.

Definition 3.1 For every density $p \in \mathcal{P}$, the maximal exponential model at p is

$$\mathcal{E}(p) = \left\{ q = e^{u - K_p(u)} p : u \in \overset{\circ}{\text{dom}}\, K_p \right\} \subseteq \mathcal{P}.$$

Remark 3.2 K_p is defined on the set $L_0^{\Phi_1}(p)$ because centered random variables guarantee the uniqueness of the representation of $q \in \mathcal{E}(p)$.

One of the main results in [18] states that any density belonging to the maximal exponential model centered at p is connected by an *open* exponential arc to p and vice versa. By *open*, we mean that the two densities are not the extremal points of the arc:

Definition 3.3 $p, q \in \mathcal{P}$ are connected by an open exponential arc if there exists an open interval $I \supset [0, 1]$ such that $p(\theta) \propto p^{(1-\theta)} q^{\theta} \in \mathcal{P}, \forall \theta \in I$.

The connection by an open exponential arc is an equivalence relation. An equivalent definition is provided by the following proposition (see [1]).

Proposition 3.4 *$p, q \in \mathcal{P}$ are connected by an open exponential arc iff there exists an open interval $I \supset [0, 1]$ such that $p(\theta) \propto e^{\theta u} p \in \mathcal{P}, \forall \theta \in I$, where $u \in L^{\Phi_1}(p)$ and $p(0) = p,\ p(1) = q$.*

The following theorem gives different equivalent conditions for a density to belong to the maximal exponential model. The proof of assertions 1–4 can be found in [1], while that of assertions 5–7 can be found in [20, 23]. The equivalence between 4 and the new assertion 8 follows from Theorem 2.8 of the previous section.

Theorem 3.5 (Portmanteau Theorem) *Let $p, q \in \mathcal{P}$. The following statements are equivalent.*

1. $q \in \mathcal{E}(p)$;
2. *q is connected to p by an open exponential arc;*
3. $\mathcal{E}(p) = \mathcal{E}(q)$;
4. $\log(q/p) \in L^{\Phi_1}(p) \cap L^{\Phi_1}(q)$;
5. $L^{\Phi_1}(p) = L^{\Phi_1}(q)$;
6. *There exists $\alpha > 1$ such that*

$$(P_\alpha): \qquad q/p \in L^{1/(\alpha-1)}(q) \ \text{ and } \ p/q \in L^{1/(\alpha-1)}(p)$$

7. ${}^m\mathbb{U}_p^q : L^{\Psi_1}(p) \to L^{\Psi_1}(q)$ *s.t.* ${}^m\mathbb{U}_p^q(v) = (p/q)v$ *is an isomorphism of Banach spaces.*
8. $d_{\star,p}(\log(q/p), L^\infty) < +\infty$ *and* $d_{\star,q}(\log(q/p), L^\infty) < +\infty$.

It is worth noting that the maximal exponential model is a good environment when dealing with divergence between densities. In fact, since (P_α) means $q/p \in L^{1+\epsilon}(p)$ and $p/q \in L^{1+\epsilon}(q)$ for some $\epsilon > 0$, it immediately follows that if $q \in \mathcal{E}(p)$, then Kullback–Leibler divergences $D(q\|p)$ and $D(p\|q)$ are both finite.

The equivalence between the equality of the Orlicz spaces $L^{\Phi_1}(p)$ and $L^{\Phi_1}(q)$ and property (P_α) has been exploited in [23] to improve some duality results in the classical problem of exponential utility maximization in incomplete markets. In fact, in many well-known works on relative entropy minimization, the minimal entropy martingale (density) measure q^* satisfies (P_α), allowing to switch from the reference Orlicz space $L^{\Phi_1}(p)$ to $L^{\Phi_1}(q^*)$, and conversely, at convenience.

Remark 3.6 From Hölder inequality, (P_α) implies (P_r) $\forall r > \alpha$.

Lemma 3.7 *The following equivalence holds:*

$$(P_\alpha) \iff \mathbb{E}_q\left(\frac{q}{p}\right)^{\pm\frac{1}{\alpha-1}} < +\infty, \mathbb{E}_p\left(\frac{p}{q}\right)^{\pm\frac{1}{\alpha-1}} < +\infty.$$

Proof It suffices to observe that if $1 < \alpha \leq 2$, then

$$\mathbb{E}_q\left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}} = \mathbb{E}_p\left(\frac{q}{p}\right)^{1+\frac{1}{\alpha-1}} < +\infty \implies \mathbb{E}_p\left(\frac{p}{q}\right)^{-\frac{1}{\alpha-1}} = \mathbb{E}_p\left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}} < +\infty.$$

Similarly, if $1 < \alpha \leq 2$, then

$$\mathbb{E}_p\left(\frac{p}{q}\right)^{\frac{1}{\alpha-1}} = \mathbb{E}_q\left(\frac{p}{q}\right)^{1+\frac{1}{\alpha-1}} < +\infty \implies \mathbb{E}_q\left(\frac{q}{p}\right)^{-\frac{1}{\alpha-1}} = \mathbb{E}_q\left(\frac{p}{q}\right)^{\frac{1}{\alpha-1}} < +\infty.$$

If $\alpha > 2$, the function $x^{\frac{1}{\alpha-1}}$ is concave and, by Jensen inequality, $\mathbb{E}_p\left(\frac{p}{q}\right)^{-\frac{1}{\alpha-1}} = \mathbb{E}_p\left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}} \leq 1$ and $\mathbb{E}_q\left(\frac{q}{p}\right)^{-\frac{1}{\alpha-1}} = \mathbb{E}_q\left(\frac{p}{q}\right)^{\frac{1}{\alpha-1}} \leq 1$. □

The following proposition gives a characterization of the smaller $\alpha > 1$ for which (P_α) holds, in terms of the $d_\star$-distance of $\log(q/p)$ to L^∞.

Proposition 3.8 *Let $q \in \mathcal{E}(p)$ and define*

$$\alpha_{p,q} = \inf\{\alpha > 1 \,:\, (P_\alpha) \text{ holds}\}.$$

Then,

$$\alpha_{p,q} = 1 + \min\left\{d_{\star,p}\left(\log\left(\frac{q}{p}\right), L^\infty\right),\, d_{\star,q}\left(\log\left(\frac{q}{p}\right), L^\infty\right)\right\}. \tag{8}$$

Proof By Lemma (3.7), we get

$$\begin{aligned}\alpha_{p,q} &= \inf\{\alpha > 1 \,:\, \mathbb{E}_q\left(\frac{q}{p}\right)^{\pm\frac{1}{\alpha-1}} < +\infty, \mathbb{E}_p\left(\frac{p}{q}\right)^{\pm\frac{1}{\alpha-1}} < +\infty\}\\ &= \inf\{\alpha > 1 \,:\, \mathbb{E}_q\left(e^{\frac{1}{\alpha-1}|\log\frac{q}{p}|}\right) < +\infty,\ \mathbb{E}_p\left(e^{\frac{1}{\alpha-1}|\log\frac{q}{p}|}\right) < +\infty\}.\end{aligned}$$

By Theorem 2.8, we then deduce

$$\begin{aligned}\alpha_{p,q} &= 1 + \min\left\{\frac{1}{a_q\left(\log\frac{q}{p}\right)}, \frac{1}{a_p\left(\log\frac{q}{p}\right)}\right\}\\ &= 1 + \min\left\{d_{\star,p}\left(\log\left(\frac{q}{p}\right), L^\infty\right),\, d_{\star,q}\left(\log\left(\frac{q}{p}\right), L^\infty\right)\right\}.\end{aligned}$$

□

In the next section, we will see that when $q \in \mathcal{E}(p)$, the distances $d_{\star,p}$ and $d_{\star,q}$ are equivalent. So, from the proposition above we get that $\log\frac{q}{p}$ belongs to the closure in $L^{\Phi_1}(p) = L^{\Phi_1}(q)$ of L^∞ if and only if $\alpha_{p,q} = 1$, which means that (P_α) holds for all $\alpha > 1$.

Remark 3.9 In the classical setting, it was shown (see [5]) that a locally integrable function f on $\mathbb{R}^d$ belongs to the BMO-closure of L^∞ if and only if both e^f and e^{-f} satisfy the Muckenhoupt (A_α) condition for all $\alpha > 1$:

$$\begin{aligned}\sup_I\left(\frac{1}{|I|}\int_I e^f dx\right)\left(\frac{1}{|I|}\int_I e^{-\frac{f}{\alpha-1}} dx\right)^{\alpha-1} &< +\infty,\\ \sup_I\left(\frac{1}{|I|}\int_I e^{-f} dx\right)\left(\frac{1}{|I|}\int_I e^{\frac{f}{\alpha-1}} dx\right)^{\alpha-1} &< +\infty,\end{aligned}$$

where I denotes any cube in $\mathbb{R}^d$ and $|I|$ its volume.

In the stochastic setting, a probabilistic analogue of the above result has been obtained for continuous BMO-martingales (see [10]); specifically, a BMO-martingale $M = (M_t)_{0\le t\le\infty}$ belongs to the BMO-closure of L^∞ if and only if both $\mathbb{E}\left(e^{M_\infty}|\mathcal{F}_\cdot\right)$ and $\mathbb{E}\left(e^{-M_\infty}|\mathcal{F}_\cdot\right)$ satisfy the probabilistic version of the Muckenhoupt (A_α) condition, for all $\alpha > 1$:

$$\sup_\tau \left\| \mathbb{E}\left(e^{M_\infty}|\mathcal{F}_\tau\right) \mathbb{E}\left(e^{-\frac{M_\infty}{\alpha-1}}|\mathcal{F}_\tau\right)^{\alpha-1} \right\|_\infty < +\infty,$$

$$\sup_\tau \left\| \mathbb{E}\left(e^{-M_\infty}|\mathcal{F}_\tau\right) \mathbb{E}\left(e^{\frac{M_\infty}{\alpha-1}}|\mathcal{F}_\tau\right)^{\alpha-1} \right\|_\infty < +\infty,$$

where τ denotes any stopping time for the filtration $\mathcal{F}$. Observe that the class BMO as well as (A_α) condition depends on the underlying probability. For $\tau = 0$, (A_α) implies

$$(A_\alpha^0): \qquad \mathbb{E}\left(e^{\pm M_\infty}\right) < +\infty, \quad \mathbb{E}\left(e^{\pm\frac{M_\infty}{\alpha-1}}\right) < +\infty.$$

Thinking that the role of M_∞ is played by $\log\frac{q}{p}$, we immediately see that requesting joint validity of (P_2) and (P_α) means requesting the validity of (A_α^0) with respect to p and q

$$(P_2) \wedge (P_\alpha) \iff (A_\alpha^0(p)) \wedge (A_\alpha^0(q)).$$

In particular, taking into account Remark 3.6, for $1 < \alpha \le 2$ it holds

$$(P_\alpha) \iff (A_\alpha^0(p)) \wedge (A_\alpha^0(q)).$$

4 Transformation of Norms by a Change of Law

Using essentially new computations with inequalities, this section derives explicit constants in the bounds between norms of equivalent Orlicz spaces. Since these constants depend on the points, they can be viewed as a measure of distance between points in the same maximal exponential model. Two applications of these bounds are proposed in Sects 4.1 and 4.2. Specifically, concentration inequalities frequently appear in high-dimensional probability and statistics ([26, 30]), but studies on the robustness of these inequalities *uniformly* in bounded parts of the statistical model are still limited.

We first start by recalling the following result proved in Cena and Pistone [1].

Proposition 4.1 *If the Orlicz spaces $L^{\Phi_1}(p)$ and $L^{\Phi_1}(q)$ are equal as sets, then their Luxemburg norms $\|\cdot\|_{\Phi_1,p}$ and $\|\cdot\|_{\Phi_1,q}$ are equivalent.*

The proof of the equivalence of norms is based on a standard argument for function spaces, i.e., in proving that the identity map from $L^{\Phi_1}(p)$ to $L^{\Phi_1}(q)$ is an homeomorphism. In many statistical applications, however, it is crucial to know the constants that this change of law yields on the norms. Theorem 4.2 provides an answer to this

issue. It refers to the norm $\|\cdot\|_\star$, equivalent to $\|\cdot\|_{\Phi_1}$ by (5), since it is more suitable for applying the property (P_α) of the Portmanteau theorem.

Theorem 4.2 *Let $v \in L^{\Phi_1}(p) = L^{\Phi_1}(q)$. Then,*

1. *there exists $\alpha > 1$ such that*

$$c_\alpha^{-1}\|v\|_{\star,p} \le \|v\|_{\star,q} \le C_\alpha \|v\|_{\star,p}, \tag{9}$$

where

$$c_\alpha = \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\mathbb{E}_p \left(\frac{p}{q}\right)^{\frac{1}{\alpha-1}}\right)^{\frac{\alpha-1}{\alpha}}, \quad C_\alpha = \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\mathbb{E}_q \left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}}\right)^{\frac{\alpha-1}{\alpha}} \tag{10}$$

are constants independent of v;

2. *it holds*

$$c^{-1}\,\|v\|_{\star,p} \le \|v\|_{\star,q} \le C\,\|v\|_{\star,p}, \tag{11}$$

where

$$c = \inf_{\alpha>1} c_\alpha, \quad C = \inf_{\alpha>1} C_\alpha \tag{12}$$

Proof For any $\gamma > 0$, using Hölder inequality with exponents $\alpha/(\alpha-1)$ and α we have

$$\begin{aligned}\frac{\mathbb{E}_q|v|^k}{k!} &\le \frac{1}{\gamma^k}\mathbb{E}_q e^{\gamma|v|} = \frac{1}{\gamma^k}\mathbb{E}_q\left(\left(\frac{q}{p}\right)^{1/\alpha}\left(\frac{p}{q}\right)^{1/\alpha} e^{\gamma|v|}\right)\\ &\le \frac{1}{\gamma^k}\left(\mathbb{E}_q\left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}}\right)^{\frac{\alpha-1}{\alpha}}\left(\mathbb{E}_p e^{\gamma\alpha|v|}\right)^{\frac{1}{\alpha}}.\end{aligned}$$

Now, choosing $\gamma > 0$ so that $\gamma\alpha\|v\|_{\star,p} < 1$, e.g., $\gamma = \frac{1}{2\alpha\|v\|_{\star,p}}$, by Lemma 2.7 we deduce

$$\mathbb{E}_p e^{\gamma\alpha|v|} \le \frac{1}{1-\gamma\alpha\|v\|_{\star,p}} = 2, \tag{13}$$

and therefore

$$\frac{\mathbb{E}_q|v|^k}{k!} \le 2^{\frac{1}{\alpha}}(2\alpha)^k\|v\|_{\star,p}^k\left(\mathbb{E}_q\left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}}\right)^{\frac{\alpha-1}{\alpha}}. \tag{14}$$

By the first inequality of (P_α), for some $\alpha > 1$ the mean value $\mathbb{E}_q\,(q/p)^{\frac{1}{\alpha-1}}$ is finite. Taking into account the definition (4) of $\|\cdot\|_{\star,q}$, from (14) we deduce

$$\|v\|_{\star,q} = \sup_{k\ge1}\left(\mathbb{E}_q\left(\frac{|v|^k}{k!}\right)\right)^{1/k} \le 2\alpha\|v\|_{\star,p}\sup_{k\ge1}\left\{2^{\frac{1}{\alpha k}}\left(\mathbb{E}_q\left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}}\right)^{\frac{\alpha-1}{\alpha k}}\right\}. \tag{15}$$

Since the function $x^{\frac{\alpha}{\alpha-1}}$ is convex, by Jensen's inequality $\mathbb{E}_q\,(q/p)^{\frac{1}{\alpha-1}} = \mathbb{E}_p\,(q/p)^{\frac{\alpha}{\alpha-1}} \geq 1$. As a consequence, the supremum in (15) is obtained for $k = 1$. The right-hand inequality in (9) then follows.

The left-hand inequality in (9) can be proved by inverting the role of p and q and by exploiting the second inequality of (P_α). Finally, inequality (11) follows immediately from (9). □

Remark 4.3 In the previous theorem, we have chosen $\gamma > 0$ such that $\gamma\alpha\|v\|_{\star,p} = 1/2$, which implies $\frac{1}{1-\gamma\alpha\|v\|_{\star,p}} = 2$. But any other $\gamma > 0$ could be chosen such that $\gamma\alpha\|v\|_{\star,p} < 1$, and this means that in (10) any number strictly greater than 1 can replace the number 2.

The following proposition is a probabilistic version of Gehring's inequality (see [2]) and gives conditions for an upper bound on $\mathbb{E}_q\left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}}$ or (inverting the role of p and q) on $\mathbb{E}_p\left(\frac{p}{q}\right)^{\frac{1}{\alpha-1}}$.

Proposition 4.4 *Let $p, q \in \mathcal{P}$. Assume there exist two constants $\delta > 0$ and $c > 0$ such that*

$$\mathbb{E}_q\left(\mathbb{1}_{\{\frac{q}{p}>\lambda\}}\right) \leq c\lambda\mathbb{E}_p\left(\mathbb{1}_{\{\frac{q}{p}>\delta\lambda\}}\right), \quad \forall\,\lambda > 1. \tag{16}$$

Then, there is $\alpha > 1$ depending only on c, δ such that $\mathbb{E}_q(\frac{q}{p})^{\frac{1}{\alpha-1}} < +\infty$ and it holds

$$\mathbb{E}_q\left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}} \leq \frac{2}{1-\frac{c}{\alpha\delta^{\frac{\alpha}{\alpha-1}}}}. \tag{17}$$

Proof We may assume $0 < \delta < 1$. Define $\Phi_\delta(x) = x\delta^{\frac{x}{x-1}}$ on $(1, +\infty)$. It is a continuous strictly increasing function such that $\lim_{x\to 1^+}\Phi_\delta(x) = 0$ and $\lim_{x\to+\infty}\Phi_\delta(x) = +\infty$. Fix $\alpha > \Phi_\delta^{-1}(c)$. Multiplying both sides of (16) by $\frac{1}{\alpha-1}\lambda^{\frac{1}{\alpha-1}-1}$ and integrating with respect to λ on the interval $[1, +\infty)$, we find

$$\mathbb{E}_q\left(\mathbb{1}_{\{\frac{q}{p}>1\}}\left(\left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}}-1\right)\right) \leq \frac{c}{\alpha}\,\mathbb{E}_p\left(\mathbb{1}_{\{\frac{q}{p}>\delta\}}\left(\left(\frac{q}{p\delta}\right)^{\frac{\alpha}{\alpha-1}}-1\right)\right). \tag{18}$$

Since $\delta < 1$, we have $\mathbb{1}_{\{\frac{q}{p}>\delta\}} = \mathbb{1}_{\{\delta<\frac{q}{p}\leq 1\}} + \mathbb{1}_{\{\frac{q}{p}>1\}}$. Thus, rearranging the terms in (18), we get

$$\left(1-\frac{c}{\alpha\delta^{\frac{\alpha}{\alpha-1}}}\right)\mathbb{E}_q\left(\mathbb{1}_{\{\frac{q}{p}>1\}}\left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}}\right) \leq$$
$$\leq \mathbb{E}_q\left(\mathbb{1}_{\{\frac{q}{p}>1\}}\right) + \frac{c}{\alpha}\,\mathbb{E}_p\left(\mathbb{1}_{\{\delta<\frac{q}{p}\leq 1\}}\left(\left(\frac{q}{p\delta}\right)^{\frac{\alpha}{\alpha-1}}-1\right)\right) \leq 1 + \frac{c}{\alpha\delta^{\frac{\alpha}{\alpha-1}}},$$

from which we deduce

$$\mathbb{E}_q\left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}} \le \frac{1+\frac{c}{\Phi_\delta(\alpha)}}{1-\frac{c}{\Phi_\delta(\alpha)}} + \mathbb{E}_q\left(\mathbb{1}_{\{\frac{q}{p}\le 1\}}\left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}}\right) \le \frac{1+\frac{c}{\Phi_\delta(\alpha)}}{1-\frac{c}{\Phi_\delta(\alpha)}} + 1 = \frac{2}{1-\frac{c}{\Phi_\delta(\alpha)}}.$$

□

We immediately deduce the following result.

Corollary 4.5 *Let $p, q = e^{u-K_p(u)}p \in \mathcal{P}$. Assume (16) holds true for $\delta = \left(\mathbb{E}_p e^u\right)^{-1} \in (0,1)$. We have:*

1. *There is $\alpha > \max\{1, c\}$ which satisfies (17), and it holds $K_p(u) < \frac{\alpha-1}{\alpha}\ln\frac{\alpha}{c}$.*
2. *Any $\alpha > \max\{1, c\}$ such that $K_p(u) < \frac{\alpha-1}{\alpha}\ln\frac{\alpha}{c}$ satisfies (17).*

The proof follows from inequality $\alpha > \Phi_\delta^{-1}(c)$, with $\Phi_\delta(x) = x\delta^{\frac{x}{x-1}}$ and $\delta = \left(\mathbb{E}_p e^u\right)^{-1}$.

Remark 4.6 Note that a sufficient condition for the validity of (16) is

$$\mathbb{E}_p\left(\mathbb{1}_{\{\frac{q}{p}\ge\delta\lambda\}}\right) \ge \frac{1}{c\lambda}, \quad \forall\,\lambda > 1.$$

Letting $\delta = \left(\mathbb{E}_p e^u\right)^{-1}$ as in the corollary above, the inequality rewrites as $\mathbb{E}_p\left(\mathbb{1}_{\{e^u\ge\lambda\}}\right) \ge \frac{1}{c\lambda}, \forall\,\lambda > 1$.

4.1 Robust Transformation of Norms

In the following proposition, we prove some robust transformations of norms when q varies in a subset of $\mathcal{E}(p)$.

Proposition 4.7 *Given $r < 1$, consider the ball $\mathcal{B}_r(p) = \left\{u \in L_0^{\Phi_1}(p) \,:\, \|u\|_{\star,p} \le r\right\}$ and define*

$$\mathcal{E}_r(p) = \left\{q = e^{u-K_p(u)}p \,:\, u \in \mathcal{B}_r(p)\right\} \subseteq \mathcal{E}(p). \tag{19}$$

Then,

1. *for any $v \in L^{\Phi_1}(p)$ and $\alpha > \frac{1}{1-r}$*

$$\sup_{q\in\mathcal{E}_r(p)} \|v\|_{\star,q} \le C_{\alpha,r}\|v\|_{\star,p}, \tag{20}$$

where

$$C_{\alpha,r} = \alpha\, 2^{\frac{\alpha+1}{\alpha}}\left(\frac{\alpha-1}{\alpha(1-r)-1}\right)^{\frac{\alpha-1}{\alpha}}; \tag{21}$$

2. *for any* $v \in L^{\Phi_1}(p)$ *and* $\alpha > 1 + r$

$$c_{\alpha,r}^{-1} \|v\|_{\star,p} \leq \inf_{q \in \mathcal{E}_r(p)} \|v\|_{\star,q}, \tag{22}$$

where

$$c_{\alpha,r} = \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\frac{1}{1-r}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha-1}{\alpha-1-r}\right)^{\frac{\alpha-1}{\alpha}}; \tag{23}$$

3. *for any* $v \in L^{\Phi_1}(p)$

$$c_r^{-1} \|v\|_{\star,p} \leq \inf_{q \in \mathcal{E}_r(p)} \|v\|_{\star,q} \leq \sup_{q \in \mathcal{E}_r(p)} \|v\|_{\star,q} \leq C_r \|v\|_{\star,p}, \tag{24}$$

where

$$c_r = \inf_{\alpha > 1+r} \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\frac{1}{1-r}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha-1}{\alpha-1-r}\right)^{\frac{\alpha-1}{\alpha}},$$

$$C_r = \inf_{\alpha > \frac{1}{1-r}} \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\frac{\alpha-1}{\alpha(1-r)-1}\right)^{\frac{\alpha-1}{\alpha}}.$$

Proof We first observe that $\mathcal{E}_r(p) \subseteq \mathcal{E}(p)$ since $\mathcal{B}_r(p) \subseteq \overset{\circ}{\text{dom}}\, K_p$. In fact, if we take $u \in L_0^{\Phi_1}(p)$ such that $\|u\|_{\star,p} \leq r < 1$, by Lemma 2.7 we get $e^{|u|} \in L^{1+\epsilon}(p)$ for any $0 < \epsilon < 1 - r$. The inclusion then follows since, from the equivalence between *1* and *6* of the Portmanteau theorem, $u \in \overset{\circ}{\text{dom}}\, K_p$ if and only if $u \in L_0^{\Phi_1}(p)$ and $e^u \in L^{1+\epsilon}(p)$ for some $\epsilon > 0$.

Since $K_p(u) \geq 0$, we get

$$\mathbb{E}_q \left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}} = \mathbb{E}_p \left(\frac{q}{p}\right)^{\frac{\alpha}{\alpha-1}} \leq \mathbb{E}_p\, e^{\frac{\alpha}{\alpha-1}u},$$

and, by Lemma 2.7, for $\alpha > 1$ such that $\frac{\alpha}{\alpha-1}\|u\|_{\star,p} < 1$ we deduce

$$\begin{aligned} C_\alpha = \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\mathbb{E}_q \left(\frac{q}{p}\right)^{\frac{1}{\alpha-1}}\right)^{\frac{\alpha-1}{\alpha}} &\leq \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\mathbb{E}_p\, e^{\frac{\alpha}{\alpha-1}u}\right)^{\frac{\alpha-1}{\alpha}} \\ &\leq \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\frac{1}{1-\frac{\alpha}{\alpha-1}\|u\|_{\star,p}}\right)^{\frac{\alpha-1}{\alpha}} \end{aligned} \tag{25}$$

For $u \in \mathcal{B}_r(p)$ and $\alpha > \frac{1}{1-r}$, we then obtain

$$C_\alpha \leq C_{\alpha,r} = \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\frac{1}{1-\frac{\alpha}{\alpha-1}r}\right)^{\frac{\alpha-1}{\alpha}}.$$

From the right-inequality in (9), we finally deduce (20) and, taking the infimum over $\alpha > \frac{1}{1-r}$, we get the right-hand inequality in (24).

In order to prove (22), we start from the left-inequality in (9). Since

$$\mathbb{E}_p \left(\frac{p}{q}\right)^{\frac{1}{\alpha-1}} = \mathbb{E}_p\, e^{\frac{1}{\alpha-1}(-u+K_p(u))} = \left(\mathbb{E}_p\, e^{u}\right)^{\frac{1}{\alpha-1}} \mathbb{E}_p\, e^{\frac{-u}{\alpha-1}},$$

again by Lemma 2.7, for $\alpha > 1$ such that $\frac{1}{\alpha-1}\|u\|_{\star,p} < 1$ we deduce

$$\begin{aligned} c_\alpha = \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\mathbb{E}_p \left(\frac{p}{q}\right)^{\frac{1}{\alpha-1}}\right)^{\frac{\alpha-1}{\alpha}} &= \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\mathbb{E}_p\, e^{u}\right)^{\frac{1}{\alpha}} \left(\mathbb{E}_p\, e^{\frac{-u}{\alpha-1}}\right)^{\frac{\alpha-1}{\alpha}} \\ &\le \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\frac{1}{1-\|u\|_{\star,p}}\right)^{\frac{1}{\alpha}} \left(\frac{1}{1-\frac{1}{\alpha-1}\|u\|_{\star,p}}\right)^{\frac{\alpha-1}{\alpha}}. \end{aligned} \tag{26}$$

For $u \in \mathcal{B}_r(p)$ and $\alpha > 1 + r$, we then obtain

$$c_\alpha \le c_{\alpha,r} = \alpha\, 2^{\frac{\alpha+1}{\alpha}} \left(\frac{1}{1-r}\right)^{\frac{1}{\alpha}} \left(\frac{1}{1-\frac{1}{\alpha-1}r}\right)^{\frac{\alpha-1}{\alpha}}.$$

From the left-inequality in (9) we finally deduce (22) and, taking the infimum over $\alpha > 1 + r$, we get the left-hand inequality in (24). □

Remark 4.8 An equivalent condition to sub-exponentiality that we can add to Proposition 2.3 in the particular case of centered random variables $u \in L_0^{\Phi_1}(p)$ is the following (cf. [26]):

5. *There exist* $a, b > 0$ *such that* $\mathbb{E}_p(e^{\lambda|u|}) \le 2e^{a\lambda^2}, \forall\, 0 < \lambda \le \frac{1}{b}$.

The constants in the bound can be expressed in terms of the norm $\|\cdot\|_{\star,p}$ as follows (cf. [25]):

$$\mathbb{E}_p(e^{\lambda|u|}) \le 2e^{\frac{\beta}{\beta-1}\|u\|^2_{\star,p}\lambda^2}, \quad \forall 0 < \lambda \le \frac{1}{\beta\|u\|_{\star,p}}, \tag{27}$$

where β is any number strictly greater than 1. By (25) and (26), we can then obtain robust bounds for C_α and c_α in exponential form.

Example 4.9 (*The maximal exponential model with Gaussian weight; cf.* [15, 25]) Let us consider the real Borel space $(\mathbb{R}, \mathcal{B})$ endowed with the probability measure $\mathbb{P}$ associated with the standard Gaussian density

$$d\mathbb{P}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

and the Gaussian exponential Orlicz space $L^{\Phi_1} = L^{\Phi_1}(p)$, with $p = 1$.

The Orlicz space L^{Φ_1} trivially contains all second-order polynomials of the form $P(x) = a_2x^2 + a_1x + a_0$.

For a centered polynomial $P(x) \in L_0^{\Phi_1}$, where $a_2 + a_0 = 0$, we get

$$\|P(x)\|_\star \leq |a_2|\|x^2\|_\star + |a_1|\|x\|_\star + |a_2| = 3|a_2| + \sqrt{\frac{2}{\pi}}\,|a_1|. \tag{28}$$

In fact, for $k = 1, 2, \ldots$

$$\mathbb{E}(x^{2k}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k} e^{-\frac{1}{2}x^2}\, dx = 1 \cdot 3 \cdot 5 \cdots (2k-1),$$

and we can write

$$\frac{\mathbb{E}(x^{2k})}{k!} = 1 \cdot (2 - \frac{1}{2}) \cdot (2 - \frac{1}{3}) \cdots (2 - \frac{1}{k}).$$

The sequence $\left(\frac{\mathbb{E}(x^{2k})}{k!}\right)^{1/k}$ is increasing, and it holds

$$\|x^2\|_\star = \sup_{k\geq 1} \left(\mathbb{E}(x^{2k})/k!\right)^{1/k} = 2.$$

In addition, for $k = 1, 2, \ldots$

$$\mathbb{E}(|x|^k) = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^k e^{-\frac{1}{2}x^2}\, dx = \begin{cases} \sqrt{\frac{2}{\pi}} & \text{if } k = 1 \\ \sqrt{\frac{2}{\pi}}(k-1)(k-3)\cdots 2 \cdot 1 & \text{if } k \geq 3, k \text{ odd} \\ (k-1)(k-3)\cdots 3 \cdot 1 & \text{if } k \geq 2, k \text{ even} \end{cases}$$

which implies

$$\|x\|_\star = \sup_{k\geq 1} \left(\mathbb{E}(|x|^k)/k!\right)^{1/k} = \sqrt{\frac{2}{\pi}}.$$

By (28), the ball $\mathcal{B}_r = \left\{u \in L_0^{\Phi_1} : \|u\|_\star \leq r\right\}$ contains all polynomials $P(x) = a_2x^2 + a_1x - a_2$ such that $3|a_2| + \sqrt{\frac{2}{\pi}}\,|a_1| \leq r$.

More generally, $L_0^{\Phi_1}$ also includes the functions $u \in C^2(\mathbb{R}; \mathbb{R})$ having $\mathbb{E}(u) = 0$ and bounded second derivative. In fact, for a suitable $\xi \in \mathbb{R}$ in a neighborhood of 0, these functions can be written in the form

$$u(x) = u(0) + u^{'}(0)x + \frac{1}{2}u^{''}(\xi)x^2, \tag{29}$$

with

$$u(0) = -\frac{1}{2}\mathbb{E}\left(u^{''}(\xi)x^2\right), \quad |u^{''}(x)| \leq c,\ \forall x \in \mathbb{R}, \tag{30}$$

from which we deduce that

$$e^{P_1(x)} \leq e^{u(x)} \leq e^{P_2(x)}, \quad \forall x \in \mathbb{R}, \tag{31}$$

for suitable second-order polynomials P_1 and P_2.

Since

$$\|u\|_\star \leq |u(0)| + |u^{'}(0)|\|x\|_\star + \frac{1}{2}\|u^{''}(\xi)x^2\|_\star,$$

where $|u(0)| \leq \frac{1}{2}c$, $\|x\|_\star = \sqrt{\frac{2}{\pi}}$ and $\|u^{''}(\xi)x^2\|_\star \leq c\|x^2\|_\star = 2c$, we deduce

$$\|u\|_\star \leq \frac{3}{2}c + \sqrt{\frac{2}{\pi}}\,|u^{'}(0)|.$$

This means that $\mathcal{B}_r$ also contains the class of functions u defined by (29), (30) and such that $\frac{3}{2}c + \sqrt{\frac{2}{\pi}}\,|u^{'}(0)| \leq r$, and therefore, our results on the equivalence of norms are robust with respect to all densities $q \propto e^u$ obtained by varying u in this class of functions.

4.2 Robust Concentration Inequalities

In the previous sections, we have seen that densities connected by an open exponential arc produce the same Orlicz space with equivalent norms and we have explicitly stated the related equivalence constants. In addition, a robust equivalence of norms has been obtained when q varies in the subset $\mathcal{E}_r(p)$ of the maximal exponential model. An interesting application that we present in this section is to derive robust concentration inequalities of Bernstein type.

We first recall some concentration inequalities of Bernstein type which hold true for centered random variables belonging to the exponential Orlicz space (see, e.g., [25, 26]).

Proposition 4.10 *Let $v \in L_0^{\Phi_1}$. Then, $\forall t > 0$ and for any $\beta > 1$,*

$$\mathbb{P}(|v| \geq t) \leq 2\exp\left(-\min\left\{\frac{t}{2\beta\|v\|_\star}, \frac{t^2}{4\frac{\beta}{\beta-1}\|v\|_\star^2}\right\}\right), \tag{32}$$

$$\mathbb{P}(|v| \geq t) \leq 2\exp\left(-\frac{t^2}{2\|v\|_\star(t + 2\|v\|_\star)}\right). \tag{33}$$

Since sub-exponentiality is preserved by linear transformations, we can obtain more general concentration inequalities for sums of independent random variables (see [26]).

Due to the form of the concentration bounds, when q varies in $\mathcal{E}_r(p)$ defined by (19), the robust version of the concentration inequalities can be immediately stated as in the following corollary.

Corollary 4.11 *For a fixed $p \in \mathcal{P}$, let $v \in L^{\Phi_1}(p)$. Then, $\forall t > 0$ and for any $\beta > 1$*

$$\sup_{q \in \mathcal{E}_r(p)} \mathbb{E}_q\left(\mathbb{1}_{\{|v-\mathbb{E}_q(v)|\geq t\}}\right) \leq 2\exp\left(-\min\left\{\frac{t}{4\alpha C_r\|v\|_{\star,p}}\ \frac{t^2}{16\frac{\beta}{\beta-1}C_r^2\|v\|_{\star,p}^2}\right\}\right), \tag{34}$$

$$\sup_{q \in \mathcal{E}_r(p)} \mathbb{E}_q\left(\mathbb{1}_{\{|v-\mathbb{E}_q(v)|\geq t\}}\right) \leq 2\exp\left(-\frac{t^2}{4C_r\|v\|_{\star,p}\left(t+4C_r\|v\|_{\star,p}\right)}\right). \tag{35}$$

Proof The proof follows the lines of some concentration inequalities stated in [25], so we only sketch it. For $t > 0$ and for any $\beta > 1$, we can write

$$\sup_{q \in \mathcal{E}_r(p)} \mathbb{E}_q\left(\mathbb{1}_{\{|v-\mathbb{E}_q(v)|\geq t\}}\right) \leq 2\exp\left(-\min\left\{\frac{t}{2\alpha\widetilde{N}_r(v)}\ \frac{t^2}{4\frac{\beta}{\beta-1}\widetilde{N}_r^2(v)}\right\}\right), \tag{36}$$

$$\sup_{q \in \mathcal{E}_r(p)} \mathbb{E}_q\left(\mathbb{1}_{\{|v-\mathbb{E}_q(v)|\geq t\}}\right) \leq 2\exp\left(-\frac{t^2}{2\widetilde{N}_r(v)\left(t+2\widetilde{N}_r(v)\right)}\right), \tag{37}$$

where

$$\widetilde{N}_r(v) = \sup_{q \in \mathcal{E}_r(p)} \|v - \mathbb{E}_q(v)\|_{\star,q} < \infty. \tag{38}$$

Finally, by Proposition 4.7(3) we get

$$\widetilde{N}_r(v) \leq \sup_{q \in \mathcal{E}_r(p)} \|v\|_{\star,q} + \sup_{q \in \mathcal{E}_r(p)} \mathbb{E}_q(|v|) \leq 2C_r\|v\|_{\star,p}.$$

□

5 Conclusion

The theoretical results of this paper first highlighted the link existing between the exponential Orlicz space and the space of BMO functions. As a consequence, it was possible to understand that the property (P_α) characterizing the maximal exponential model is a sort of static Muckenhoupt property which holds simultaneously with respect to two densities connected by an open exponential arc. This could help to understand how to introduce time dependence in the study of maximal exponential models, which is still an open issue. The explicit constants we obtained when changing the law of Orlicz spaces centered at connected densities are one of the main contributions of the paper. As an application, for random variables belonging to the exponential Orlicz space we obtained concentration inequalities of Bernstein type robust with respect to densities in an exponential subfamily. Their extension to sums of sub-exponential random variables can be applied to derive uniform exponential bounds for the law of large numbers.

Acknowledgements The author is grateful to a referee for very helpful comments and suggestions on the initial submission of the paper.

Author Contributions 100% (single author).

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement.

Data Availability Not applicable.

Declarations

Conflict of interest The author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Cena, A., Pistone, G.: Exponential statistical manifold. AISM **59**, 27–56 (2007). https://doi.org/10.1007/s10463-006-0096-y
2. Doléans-Dade, Meyer P.A.: Inégalités de normes avec poids. Séminaire de Probabilités XIII, Université de Strasbourg (Lecture Notes in Math. 721, pp. 313-331). Springer, Berlin (1979)
3. Emery, M.: Une définition faible de BMO. Ann. Inst. Henri Poincaré, Prob. Statist. 21(1), 59-71 (1985)
4. Fukumizu, K.: Exponential manifold by reproducing kernel Hilbert spaces. Algebraic and Geometric methods in Statistics, 291-306. Cambridge University Press (2009). https://doi.org/10.1017/CBO9780511642401.019
5. Garnett, J.B., Jones, P.W.: The distance in BMO to L^∞. Ann. Math. **108**, 373–393 (1978)
6. Gibilisco, P., Pistone, G.: Connections on non-parametric statistical manifolds by Orlicz space geometry. Infin. Dimens. Anal. Quantum Probab. Relat. Top. **1**(02), 325–347 (1998)
7. Grasselli, M.R.: Dual connections in nonparametric classical information geometry. Ann. Inst. Stat. Math. **62**, 873–896 (2010). https://doi.org/10.1007/s10463-008-0191-3
8. Imparato D., Trivellato B.: Geometry of extendend exponential models. In: Algebraic and Geometric Methods in Statistics, pp. 307-326. Cambridge University Press (2009). https://doi.org/10.1017/CBO9780511642401.020
9. John, F., Nirenberg, L.: On functions of bounded mean oscillation. Commun. Pure Appl. Math. **14**, 415–426 (1961)
10. Kazamaki, N.: Continuous exponential martingales and BMO. Springer, Berlin (1994). https://doi.org/10.1007/BFb0073585
11. Montrucchio L., Pistone G.: A Class of non-parametric deformed exponential statistical models. In: Nielsen, F. (eds) Geometric Structures of Information. Signals and Communication Technology, pp. 15-35. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-02520-5_2
12. Naudts, J.: Exponential arcs in manifolds of quantum states. Front. Phys. **11**, 12 (2023). https://doi.org/10.3389/fphy.2023.1042257
13. Newton, N.J.: A class of non-parametric statistical manifolds modelled on Sobolev space. Inf. Geom. **2**, 283–312 (2019). https://doi.org/10.1007/s41884-019-00024-z
14. Pistone, G.: Examples of the application of nonparametric information geometry to statistical physics. Entropy **15**(10), 4042–4065 (2013). https://doi.org/10.3390/e15104042

15. Pistone G.: Information geometry of the gaussian space. In: Ay, N., Gibilisco P., Matus F. (eds) Information Geometry and Its Applications. IGAIA IV 2016. Springer Proceedings in Mathematics & Statistics, vol. 252, pp. 119-155. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-97798-0_5
16. Pistone G.: A Lecture about the Use of Orlicz Spaces in Information Geometry. In: Barbaresco, F., Nielsen, F. (Eds.) SPIGL 2020 PROMS 361, pp. 179-195 (2020). https://doi.org/10.48550/arXiv.2012.03376
17. Pistone, G., Rogantin, M.P.: The exponential statistical manifold: mean parameters, orthogonality and space transformations. Bernpulli **5**, 721–760 (1999). https://doi.org/10.2307/3318699
18. Pistone, G., Sempi, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. Ann. Stat. **23**(5), 1543–1561 (1995). https://doi.org/10.1214/aos/1176324311
19. Rao, M.M., Ren, Z.D.: Theory Of Orlicz spaces. Marcel Dekker Inc., New York (1991)
20. Santacroce, M., Siri, P., Trivellato, B.: New results on mixture and exponential models by Orlicz spaces. Bernoulli **22**(3), 1431–1447 (2016). https://doi.org/10.3150/15-BEJ698
21. Santacroce M., Siri P., Trivellato B.: On mixture and Exponential connection by open arcs. In: Nielsen F., Barbaresco F. (eds) Geometric Science of Information. GSI 2017. Lecture Notes in Computer Science, vol 10589, pp. 577-584. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68445-1_67
22. Santacroce, M., Siri, P., Trivellato, B.: An application of maximal exponential models to duality theory. Entropy **20**(495), 1–9 (2018). https://doi.org/10.3390/e20070495
23. Santacroce, M., Siri, P., Trivellato, B.: Exponential models by Orlicz spaces and applications. J. Appl. Probab. **55**, 682–700 (2018). https://doi.org/10.1017/jpr.2018.45
24. Siri P., Trivellato B.: Minimization of the Kullback-Leibler divergence over a log-normal exponential arc. In: Nielsen F., Barbaresco F. (eds) Geometric Science of Information. GSI 2019. Lecture Notes in Computer Science, vol 11712, pp. 453-461. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-26980-7_47
25. Siri, P., Trivellato, B.: Robust concentration inequalities in maximal exponential models. Stat. Probab. Lett. **170**, 109001 (2021). https://doi.org/10.1016/j.spl.2020.109001
26. Vershynin R.: High-dimentional probability. An introduction with applications in data science. Cambridge University Press (2018). https://doi.org/10.1017/9781108231596
27. Varopoulos, N.T.: A probabilistic proof of the Garnett–Jones theorem on BMO. Proc. J. Math. **90**, 201–221 (1980)
28. Vigelis, R.F., Cavalcante, C.C.: On φ-families of probability distributions. J. Theor. Probab. **26**(3), 870–884 (2013). https://doi.org/10.1007/s10959-011-0400-5
29. Vieira, F.L.J., de Andrade, L.H.F., Vigelis, R.F., Cavalcante, C.C.: A deformed exponential statistical manifold. Entropy **21**(5), 496 (2019). https://doi.org/10.3390/e21050496
30. Wainwright, M.J.: High-dimentional statistics. A non-asymptotic viewpoint. Cambridge University Press (2019). https://doi.org/10.1017/9781108627771

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.