Notes on the speed of entropic convergence in the Central Limit Theorem

Laurent Miclo

Abstract. In the context of the study of convergence speeds in the Central Limit Theorem, we investigate some consequences of a general Lipschitz contraction property of probability transition kernels with respect to relative entropy. This Markovian approach will enable us to discuss examples whose behavior is not covered by results recently obtained by several authors. More precisely, let $X_0, ..., X_n$ be IID real random variables, centered and normalized. It is known that if their law admits a positive spectral gap and a finite relative entropy with respect to ν, the standard Gaussian distribution, then the relative entropy of law of $(X_0 + \cdots + X_n)/\sqrt{n+1}$ with respect to ν goes to zero at least as $\mathcal{O}(1/(n+1))$, for large $n \in \mathbb{N}$. The two goals of this paper are: on one hand, for any fixed $p \in \mathbb{N}^*$, to find conditions insuring an entropic convergence faster than $\mathcal{O}(1/(n+1)^{p/2})$ and on the other hand to relax the spectral gap assumption, even at the cost of slower convergence bounds.

1. Introduction

The objective of this paper is to present a Markovian approach to the entropic convergence in Central Limit Theorem. Even if we are not yet completely satisfied with the results obtained, they enable to exhibit some interesting examples with respect to known behaviors.

We recall the existing results. Let $X_0, X_1, ...$ be a sequence of IID real random variables. We denote by μ their common law and we assume that it is centered and of variance 1. For $n \in \mathbb{N}$, consider m_n the law of $(X_0 + \cdots + X_n)/\sqrt{n+1}$. By Central Limit Theorem, for large $n \in \mathbb{N}$, the distribution m_n weakly converges to ν, the centered and normalized Gaussian law.
Under supplementary conditions, Barron & Johnson [14] and Artstein, Ball, Barthe & Naor [5, 3] studied quantitatively this convergence in a entropic sense. Recall

Received by the editors April 1, 2003.
1991 *Mathematics Subject Classification.* Primary 60F05; Secondary 60J05, 94A17, 49R50, 44A60.
Key words and phrases. Central Limit Theorem, Berry-Essen type bounds, relative entropy, inhomogeneous Markovian chains, moments of Gibbs measures, modified logarithmic Sobolev inequalities, spectral gap, Fisher information, weighted Poincaré inequalities.

that the (relative) entropy of a probability measure m on $\mathbb{R}$ with respect to another one η is defined by

$$\mathrm{Ent}(m|\eta) \quad := \quad \begin{cases} \int \ln\left(\frac{dm}{d\eta}\right) \, dm \;\leq\; +\infty & , \text{if } m \ll \eta \\ +\infty & , \text{otherwise} \end{cases}$$

Furthermore, we say that a law μ on $\mathbb{R}$ admits a spectral gap λ, if the quantity

$$\lambda \quad := \quad \inf_{f \in \mathcal{C}^1_{\mathrm{b}}(\mathbb{R}) \,:\, f \neq \mu[f],\, \mu-\text{a.s.}} \frac{\mu[(f')^2]}{\mu[(f-\mu[f])^2]}$$

is positive, where $\mathcal{C}^1_{\mathrm{b}}(\mathbb{R})$ is the set of $\mathcal{C}^1$ mappings from $\mathbb{R}$ to $\mathbb{R}$ which are bounded and whose derivative is equally bounded.

Theorem 1.1 ([14, 5, 3])**.** *With previous notations, assume that μ is of finite entropy with respect to ν and admits a spectral gap. Then there exists a constant $C_0 \geq 0$ (depending on* $\mathrm{Ent}(\mu|\nu)$ *and λ) such that for any $n \in \mathbb{N}$,*

$$\mathrm{Ent}(m_n|\nu) \quad \leq \quad \frac{C_0}{n+1}$$

and it was furthermore shown in [2] *that the LHS is non-increasing with respect to the time parameter $n \in \mathbb{N}$.*

One outcome of this theorem is that it enables us to strengthen Berry and Essen bounds. Remember that one version (cf [12]) of this result asserts that if μ is supposed to admit a finite third absolute moment (namely $\int |x|^3 \, \mu(dx) < +\infty$), then there exists a universal constant $0 < C_1 \leq 3$ such that for all $n \in \mathbb{N}$,

$$\sup_{x \in \mathbb{R}} |m_n[(-\infty, x]] - \nu[(-\infty, x]]| \quad \leq \quad \frac{C_1}{\sqrt{n+1}} \int |x|^3 \, \mu(dx)$$

This order of convergence in n can be deduced from the above theorem, at least under the assumptions presented there. Indeed, this is a direct consequence of Csiszár-Kullback inequality (see e.g. [1]), saying that for any probabilities m, η, we are assured of

$$\|m - \eta\|_{\mathrm{tv}} \quad \leq \quad \sqrt{2}\sqrt{\mathrm{Ent}(m|\eta)}$$

where our convention for the total variation norm is that

$$\|m - \eta\|_{\mathrm{tv}} \quad := \quad \sup_{f \in \mathcal{B}(\mathbb{R}) \,:\, \|f\|_\infty \leq 1} |(m - \eta)[f]|$$

with $\mathcal{B}(\mathbb{R})$ the set of bounded measurable functions on $\mathbb{R}$ (but the previous bound is valid on any measurable space). In particular, we have

$$\begin{aligned} \sup_{x \in \mathbb{R}} |m_n[(-\infty, x]] - \nu[(-\infty, x]] \quad &\leq \quad \frac{\|m_n - \nu\|_{\mathrm{tv}}}{2} \\ &\leq \quad \frac{1}{\sqrt{2}}\sqrt{\mathrm{Ent}(m_n|\nu)} \end{aligned}$$

Remark nevertheless that the spectral gap hypothesis implies that μ admits moments of all orders (even "small" exponential moments, see e.g. [15]), thus we are far away from the optimal validity condition of Berry and Essen bounds, but it is also true that the total variation norm is much stronger than the supremum of the difference between distribution functions (consider for instance the situation where μ is a weighted sum of Dirac masses, then we have that $\mathrm{Ent}(m_n|\nu) = +\infty$ and $\|m_n - \nu\|_{\mathrm{tv}} = 2$ for all $n \in \mathbb{N}$, while Berry and Essen bounds are valid).

Our purpose in this note is twofold: on one hand for any $p > 1$, we will find probabilities $\mu \neq \nu$ such that the corresponding quantity $\mathrm{Ent}(m_n|\nu)$ goes to zero faster than $1/(n+1)^p$ (Proposition 3.8) and on the other hand we will discuss the spectral gap hypothesis, showing ways to relax it in certain circumstances, in particular if one can content oneself with bad estimates of order $1/\sqrt{n+1}$ for the speed of convergence (Proposition 4.3). Finally let us mention that the ideas put forth here are relatively simple, so we hope numerous improvements are possible, see next section for a challenging direction.

2. A Markovian point of view

Our starting point is the observation that the sequence $((X_0+\cdots+X_n)/\sqrt{n+1})_{n\in\mathbb{N}}$ can be viewed as an inhomogeneous Markov chain. So one can try to apply general Markovian methods for the quantitative study of the relative entropy evolution to this particular context. Indeed, there are several possible choice for the reference measure; one could consider the instantaneous invariant probabilities (as it is customary in simulated annealing see for instance [17]) or the target probability ν (one can even think of intermediate solutions). We will follow here this latter approach, since we already have a lot of informations about the standard Gaussian law.

As in the introduction, let $X_0, ..., X_n, ...$ be a sequence of independent real random variables. For our purpose they no longer need to be identically distributed and for $n \in \mathbb{N}$, we will denote by μ_n the law of X_n, which we assume to be centered and of variance 1.
For fixed $n \in \mathbb{N}$, let K_{n+1} be the Markovian kernel describing the transition from $Y_n \coloneqq (X_0 + \cdots + X_n)/\sqrt{n+1}$ to Y_{n+1}. More precisely, for any $x \in \mathbb{R}$, $K_{n+1}(x, \cdot)$ is the law of $\sqrt{\frac{n+1}{n+2}}x + \frac{X_{n+1}}{\sqrt{n+2}}$, so that this ensures:

$$\forall\ f \in \mathcal{B}(\mathbb{R}), \qquad \mathbb{E}[f(Y_{n+1})|Y_0, Y_1, ..., Y_n] \quad = \quad K_{n+1}[f](Y_n)$$

(as usual this equality is understood $\mathbb{P}$-a.s., with $\mathbb{P}$ the underlying probability).
From general results (cf [10]), we know that for every distribution m (and a priori for every probability ν, even if from now on, the latter will stand for the standard Gaussian law), we have

$$\mathrm{Ent}(mK_{n+1}|\nu K_{n+1}) \quad \leq \quad \mathrm{Ent}(m|\nu)$$

In our applications in this paper, we will only use this easy relation, but let us indicate how it can be quantified, since maybe the reader will be able to go further than us. Following [10], we first construct a “generalized” kernel $K^*_{n+1,\nu}$, namely a Markovian operator from $\mathbb{L}^1(\nu)$ to $\mathbb{L}^1(\nu K_{n+1})$, by requiring that

$$\forall\, f\in\mathbb{L}^1(\nu),\qquad K^*_{n+1,\nu}[f] \ \coloneqq\ \frac{d(f\nu)K_{n+1}}{d\nu K_{n+1}}$$

In order to compute it, let $g\in\mathcal{B}(\mathbb{R})$ be a test function. By definition and by an obvious change of variable, we have

$$\begin{aligned}
&((f\nu)K_{n+1})[g]\\
&\quad= \int f(x)K_{n+1}[g](x)\,\nu(dx)\\
&\quad= \int f(x)g\left(\sqrt{\frac{n+1}{n+2}}x+\frac{y}{\sqrt{n+2}}\right)\mu_{n+1}(dy)\nu(dx)\\
&\quad= \int \mu_{n+1}(dy)\int dx\,\frac{\exp(-x^2/2)}{\sqrt{2\pi}}f(x)g\left(\sqrt{\frac{n+1}{n+2}}x+\frac{y}{\sqrt{n+2}}\right)\\
&\quad= \int \mu_{n+1}(dy)\sqrt{\frac{n+2}{n+1}}\frac{1}{\sqrt{2\pi}}\int dx\exp\left(-\frac{n+2}{n+1}\frac{\left(x-y/\sqrt{n+2}\right)^2}{2}\right)\\
&\qquad\quad f\left(\sqrt{\frac{n+2}{n+1}}\left(x-y/\sqrt{n+2}\right)\right)g(x)\\
&\quad= \int dx\,F_n(x,f)g(x)
\end{aligned}$$

where for any $x\in\mathbb{R}$ and $f\in\mathbb{L}^1(\nu)$,

$$\begin{aligned}
F_n(x,f) \ &\coloneqq\ \sqrt{\frac{n+2}{n+1}}\frac{1}{\sqrt{2\pi}}\int\mu_{n+1}(dy)\exp\left(-\frac{n+2}{n+1}\frac{\left(x-y/\sqrt{n+2}\right)^2}{2}\right)\\
&\qquad f\left(\sqrt{\frac{n+2}{n+1}}\left(x-\frac{y}{\sqrt{n+2}}\right)\right)
\end{aligned}$$

(this expression is dx-a.s. finite), thus it appears that we can take

$$\begin{aligned}
K^*_{n+1,\nu}[f](x) \ &=\ \frac{F_n(x,f)}{F_n(x,\mathbb{1})}\\
&=\ \frac{1}{\varphi_{n+1}(x)}\int\mu_{n+1}(dy)\exp\left(\frac{\sqrt{n+2}}{n+1}xy-\frac{1}{2}\frac{y^2}{n+1}\right)\\
&\qquad f\left(\sqrt{\frac{n+2}{n+1}}\left(x-y/\sqrt{n+2}\right)\right)
\end{aligned}$$

with

$$\forall\, x\in\mathbb{R},\qquad \varphi_{n+1}(x) \;=\; \mathbb{E}\left[\exp\left(\frac{\sqrt{n+2}}{n+1}xX_{n+1}-\frac{1}{2}\frac{X_{n+1}^2}{n+1}\right)\right]$$

In fact our main interest in kernel $K^*_{n+1,\nu}$ comes from the Dirichlet form $\mathcal{E}_{n+1,\nu}$ defined on $\mathbb{L}^2(\nu)$ by the following formula, for any $f,g\in\mathbb{L}^2(\nu)$,

$$\begin{aligned}
&\mathcal{E}_{n+1,\nu}(f,g)\\
&\quad\coloneqq\; \nu[f(\mathrm{Id}-K_{n+1}K^*_{n+1,\nu})[g]]\\
&\quad=\; \int \nu(dx)\, f(x)(g(x)-K_{n+1}K^*_{n+1,\nu}[g](x))\\
&\quad=\; -\int \nu(dx)\, f(x)K_{n+1}K^*_{n+1,\nu}[g-g(x)](x)\\
&\quad=\; -\int \nu(dx)\mu_{n+1}(dy)\, f(x)K^*_{n+1,\nu}[g-g(x)]\left(\sqrt{\frac{n+1}{n+2}}x+\frac{y}{\sqrt{n+2}}\right)
\end{aligned}$$

This leads us to calculate for any $x,y\in\mathbb{R}$,

$$\begin{aligned}
&K^*_{n+1,\nu}[g-g(x)]\left(\sqrt{\frac{n+1}{n+2}}x+\frac{y}{\sqrt{n+2}}\right)\\
&\quad=\; \frac{1}{\varphi_{n+1}\left(\sqrt{\frac{n+1}{n+2}}x+\frac{y}{\sqrt{n+2}}\right)}\int\mu_{n+1}(dz)\exp\left(-\frac{z^2}{2(n+1)}+\frac{xz}{\sqrt{n+1}}+\frac{yz}{n+1}\right)\\
&\qquad\left(g\left(x+y/\sqrt{n+1}-z/\sqrt{n+1}\right)-g(x)\right)
\end{aligned}$$

This suggests to consider the change of variable where x is replaced by $x-y/\sqrt{n+1}$, since then we end up with a rather symmetrical expression:

$$\begin{aligned}
&\mathcal{E}_{n+1,\nu}(f,g)\\
&\quad=\; \int\nu(dx)\mu_{n+1}(dy)\mu_{n+1}(dz)\frac{\exp\left(\frac{x(y+z)}{\sqrt{n+1}}-\frac{y^2}{2(n+1)}-\frac{z^2}{2(n+1)}\right)}{\varphi_{n+1}\left(\sqrt{\frac{n+1}{n+2}}x\right)}f(x-y/\sqrt{n+1})\\
&\qquad(g(x-y/\sqrt{n+1})-g(x-z/\sqrt{n+1}))\\
&\quad=\; \frac{1}{2}\int\nu(dx)\mu_{n+1}(dy)\mu_{n+1}(dz)\frac{\exp\left(\frac{x(y+z)}{\sqrt{n+1}}-\frac{y^2}{2(n+1)}-\frac{z^2}{2(n+1)}\right)}{\varphi_{n+1}\left(\sqrt{\frac{n+1}{n+2}}x\right)}\\
&\qquad(f(x-y/\sqrt{n+1})-f(x-z/\sqrt{n+1}))(g(x-y/\sqrt{n+1})-g(x-z/\sqrt{n+1}))
\end{aligned}$$

So we are led to introduce for $n\in\mathbb{N}$ and $x\in\mathbb{R}$, the distribution $M_{n+1,x}$ defined on $\mathbb{R}$ by

$$M_{n+1,x}[h]$$

$$:= \frac{1}{\varphi_{n+1}\left(\sqrt{\frac{n+1}{n+2}}x\right)}\int \mu_{n+1}(dy)\exp\left(\frac{xy}{\sqrt{n+1}}-\frac{y^2}{2(n+1)}\right)h(x-y/\sqrt{n+1})$$

with $h\in\mathcal{B}(\mathbb{R})$ a test function, because we can rewrite

$$\mathcal{E}_{n+1,\nu}(f,g) \quad = \quad \frac{1}{2}\int \nu(dx)\varphi_{n+1}\left(\sqrt{\frac{n+1}{n+2}}x\right)\mathrm{Cov}(f,g;M_{n+1,x})$$

where $\mathrm{Cov}(f,g;M_{n+1,x})$ denote the covariance of f and g with respect to $M_{n+1,x}$. In particular, the modified logarithmic Sobolev constant α_n associated to ν and $\mathcal{E}_{n+1,\nu}$,

$$\alpha_n \quad := \quad \inf_{f\in\mathbb{L}^2(\nu)\setminus\mathrm{Vect}(\mathbb{1})}\frac{\mathcal{E}_{n+1,\nu}(f^2,\ln(f^2))}{\mathrm{Ent}(f^2;\nu)}$$

where $\mathrm{Ent}(f^2;\nu)$ is an abbreviation for $\mathrm{Ent}(f^2\nu/\nu[f^2]|\nu)$, can be expressed as

$$\begin{aligned}\alpha_n \quad &= \quad \frac{1}{2}\int \nu(dx)\varphi_{n+1}\left(\sqrt{\frac{n+1}{n+2}}x\right)\mathrm{Cov}(f^2,\ln(f^2);M_{n+1,x})\\ &\geq \quad \inf_{f\in\mathbb{L}^2(\nu)\setminus\mathrm{Vect}(\mathbb{1})}\frac{\int \nu(dx)\varphi_{n+1}\left(\sqrt{\frac{n+1}{n+2}}x\right)\mathrm{Ent}(f^2;M_{n+1,x})}{\mathrm{Ent}(f^2;\nu)}\end{aligned}$$

(where we have used Jensen's inequality $\int \ln(f^2)\,dM_{n+1,x}\leq \ln\left(\int f^2\,dM_{n+1,x}\right)$ to traditionally deduce that $\mathrm{Cov}(f^2,\ln(f^2);M_{n+1,x})\geq 2\mathrm{Ent}(f^2;M_{n+1,x})$).
The role of this ergodic coefficient α_n is particularly important, since we have shown in [10] that there exists a universal constant $0<\rho\leq 1$ such that

$$\mathrm{Ent}(mK_{n+1}|\nu K_{n+1}) \quad \leq \quad (1-\rho\alpha_n)\mathrm{Ent}(m|\nu) \tag{2.1}$$

Unfortunately, except in the case where $\mu_{n+1}=\nu$ (or more generally if $d\mu_{n+1}/d\nu$ is bounded above and below by positive constants), we don't know how to estimate α_n! Nevertheless, we remark that if $\mu_{n+1}=\nu$, then this ensures:

$$\mathrm{Ent}(mK_{n+1}|\nu K_{n+1}) \;=\; \mathrm{Ent}(mK_{n+1}|\nu) \;\leq\; \left(1-\frac{1}{n+2}\right)\mathrm{Ent}(m|\nu) \tag{2.2}$$

as it can be proved by directly using a classical Ornstein-Uhlenbeck process, which permits to go from m to mK_{n+1} in a continuous time interval of length $\ln((n+2)/(n+1))$ (cf also [10]).

Coming back to the general situation, we note that

$$\mathrm{Ent}(mK_{n+1}|\nu) \quad = \quad \mathrm{Ent}(mK_{n+1}|\nu K_{n+1})+\int \ln\left(\frac{d\nu K_{n+1}}{d\nu}\right)dm_n$$

and we compute that ν-a.s. for $x\in\mathbb{R}$,

$$\frac{d\nu K_{n+1}}{d\nu}(x) \quad = \quad \sqrt{\frac{n+2}{n+1}}\exp\left(-\frac{x^2}{2(n+1)}\right)\varphi_{n+1}(x)$$

Taking into account that for each $n \in \mathbb{N}$, m_n is of variance 1, we get that

(2.3)

$$\mathrm{Ent}(m_{n+1}|\nu) \quad \leq \quad \mathrm{Ent}(m_n|\nu) + \frac{1}{2}\ln\left(\frac{n+2}{n+1}\right) - \frac{1}{2(n+1)} + \int \ln\left(\varphi_{n+1}\right)\, dm_{n+1}$$

and our task in next section will be to evaluate this last term.

Remarks 2.1. a) The fact that the sequence $(\mathrm{Ent}(m_n|\nu))_{n\geq 0}$ is non-increasing when all the distributions μ_n, $n \in \mathbb{N}$, are equal seems difficult to deduce from (2.3). Of course in general, when these laws are different, this monotonicity property is wrong, consider for instance the cases where $\mu_0 = \nu \neq \mu_1$.
b) The universal constant ρ appears in a very bad place in (2.1). Indeed, let be given a sequence $(E_n)_{n\in\mathbb{N}}$ of nonnegative reals verifying the inequalities

$$\text{(2.4)} \qquad \forall\, n \in \mathbb{N}, \qquad E_{n+1} \leq (1 - a/(n+1))E_n + b/(n+1)^c$$

where $a, b, c > 0$ are fixed. Then by analogy with the corresponding differential inequality, it can be shown that there exists a constant $C > 0$ such that for all $n \in \mathbb{N}$, we are assured of

$$\begin{aligned} c - a < 1 \quad &\Rightarrow \quad E_n \leq \frac{C}{(n+1)^{c-1}} \\ c - a = 1 \quad &\Rightarrow \quad E_n \leq \frac{C\ln(n+2)}{(n+1)^a} \\ c - a > 1 \quad &\Rightarrow \quad E_n \leq \frac{C}{(n+1)^a} \end{aligned}$$

(and starting from the opposite inequalities in (2.4), one has similar reversed bounds, see for instance appendix A of [11]) so the coefficient a is quite crucial for the asymptotic behavior of the sequence $(E_n)_{n\in\mathbb{N}}$.
c) Nevertheless, we note that for fixed $x \in \mathbb{R}$ and $f \in \mathcal{C}^1_{\mathrm{b}}(\mathbb{R})$, we have, if f takes values in some compact subset of $(0, +\infty)$,

$$\begin{aligned} \lim_{n\to\infty} (n+1)\,\mathrm{Cov}(f^2, \ln(f^2); M_{n+1,x}) \quad &= \quad 4(f'(x))^2 \\ \lim_{n\to\infty} \varphi_{n+1}\left(\sqrt{\frac{n+1}{n+2}}x\right) \quad &= \quad 1 \end{aligned}$$

so heuristically, we hope that for large $n \in \mathbb{N}$,

$$\begin{aligned} \alpha_n \quad &\sim \quad \frac{2}{n+1} \inf_{f\in\mathbb{L}^2(\nu)\setminus\mathrm{Vect}(\mathbb{1})} \frac{\int (f'(x))^2\,\nu(dx)}{\mathrm{Ent}(f^2;\nu)} \\ &= \quad \frac{1}{n+1} \end{aligned}$$

where for the last identity we have used the well-known fact that the logarithmic Sobolev constant associated to ν is 1/2 (see for instance [1]).

More precisely, this expected behavior leads us to conjecture that under nice conditions on the family $(\mu_n)_{n\geq 0}$ (a uniform spectral gap assumption for instance?), we have directly for large $n\in\mathbb{N}$,

$$\inf_{f\in\mathbb{L}^2(\nu)\setminus\mathrm{Vect}(\mathbb{1})} \frac{\mathrm{Ent}(f^2;\nu)-\mathrm{Ent}(K_{n+1,\nu}^*[f^2];\nu K_{n+1})}{\mathrm{Ent}(f^2;\nu)} \quad\sim\quad \frac{1}{n+1}$$

(in [10] we have shown that the LHS always belongs to the interval $[\rho\alpha_n,\alpha_n]$ and this the reason of the apparition of the universal constant ρ in (2.1)), namely (2.2) would be asymptotically almost satisfied.
Of course, such a result would greatly help our approach of the subject, since via the difference inequalities of previous remark (b) and the considerations of next sections, we would end up with bounds close to that of Theorem 1.1, under appropriate assumptions (but surprisingly, it would not be possible by this method to get a convergence speed estimate better than $\mathcal{O}(1/(n+1))$). We hope to be able to develop such a study in the future.
Finally, let us remark that it is not really necessary to consider all the functions $f\in\mathbb{L}^2(\nu)\setminus\mathrm{Vect}(\mathbb{1})$ in the above infima, since we only need the corresponding inequalities with $f=f_n:=\sqrt{dm_n/d\nu}$ and one can already have at his disposal some informations on this function (for instance a uniform spectral gap assumption on the family $(\mu_p)_{p\in\mathbb{N}}$ implies the same property for the family $(m_p)_{p\in\mathbb{N}}=(f_p^2\nu)_{p\in\mathbb{N}}$, see e.g. next section).

3. Examples of "fast" convergence

We will be interested here in properties of (subclasses of) $\mathcal{N}_p$, for $p\in\mathbb{N}$, the set of probability measures on $\mathbb{R}$ whose p first moments coincide with those of ν (by convention the moment of order 0 is the total mass, so $\mathcal{N}_0$ is just the set of all probabilities on $\mathbb{R}$).

These sets are quite natural in our setting, for instance our basic assumption in last section was that for every $n\in\mathbb{N}$, $\mu_n\in\mathcal{N}_2$. Then all the distributions m_n, $n\in\mathbb{N}$, also belong to $\mathcal{N}_2$. This kind of stability by appropriately weighted convolution is a general fact for the $\mathcal{N}_p$, $p\in\mathbb{N}$:

Lemma 3.1. *Let $p\in\mathbb{N}$, and $0\leq t\leq 1$ be fixed. If we are given two probabilities $\mu,\mu'\in\mathcal{N}_p$, then we are assured that $m\in\mathcal{N}_p$, where m is the probability defined by*

$$\forall\ f\in\mathcal{B}(\mathbb{R}),\qquad m[f]\quad=\quad\int f(tx+\sqrt{1-t^2}y)\,\mu(dx)\mu'(dy)$$

Proof. Let $k\in\mathbb{N}$, $0\leq k\leq p$, be given. We compute that

$$\int x^k\,m(dx)\quad=\quad\int(tx+\sqrt{1-t^2}y)^k\,\mu(dx)\mu'(dy)$$

$$
\begin{aligned}
&= \int \sum_{0\leq l\leq k} \binom{k}{l} t^l x^l (1-t^2)^{\frac{k-l}{2}} y^{k-l}\,\mu(dx)\mu'(dy)\\
&= \sum_{0\leq l\leq k} \binom{k}{l} t^l (1-t^2)^{\frac{k-l}{2}} \int x^l\,\mu(dx) \int y^{k-l}\,\mu'(dy)\\
&= \sum_{0\leq l\leq k} \binom{k}{l} t^l (1-t^2)^{\frac{k-l}{2}} \int x^l\,\nu(dx) \int y^{k-l}\,\nu(dy)\\
&= \int (tx+\sqrt{1-t^2}y)^k\,\nu(dx)\nu(dy)\\
&= \int x^k\,\nu(dx)
\end{aligned}
$$

□

In particular, if for some $p \in \mathbb{N}$, we assume that for all $n \in \mathbb{N}$, $\mu_n \in \mathcal{N}_p$, then we also end up with $m_n \in \mathcal{N}_p$ for all $n \in \mathbb{N}$.
As mentioned at the end of last section, the spectral gap is also "preserved" by this kind of operation. In a certain manner, this observation (applied to discrete "carrés du champs") was at the heart of celebrated Gross' proof [13] of the logarithmic Sobolev inequality for the standard normal distribution.

Lemma 3.2. *Let $t \geq 0$ and two probabilities μ and μ' be given and define m as in the previous lemma. If we assume that μ and μ' admit respectively as spectral gaps $\lambda > 0$ and $\lambda' > 0$, then m also satisfies such an inequality and its spectral gap is larger than $\lambda \wedge \lambda'$.*

Proof. It is well-known (cf for instance [1]) that $\mu \otimes \mu'$ admits a spectral gap larger than $\lambda \wedge \lambda'$, in the sense that

$$\forall\ g \in \mathcal{C}^1_{\mathrm{b}}(\mathbb{R}^2), \qquad (\lambda\wedge\lambda')\mu\otimes\mu'[(g-\mu\otimes\mu'[g])^2] \quad\leq\quad \mu\otimes\mu'[(\partial_1 g)^2 + (\partial_2 g)^2]$$

(where ∂_1 and ∂_2 designate the partial derivatives with respect to the first and second variables). Let a function $f \in \mathcal{C}^1_{\mathrm{b}}(\mathbb{R})$ be given and consider the mapping $g \in \mathcal{C}^1_{\mathrm{b}}(\mathbb{R}^2)$ defined by

$$\forall\ (x,y) \in \mathbb{R}^2, \qquad g(x,y) \quad\coloneqq\quad f(tx+\sqrt{1-t^2}y)$$

Clearly, we have

$$
\begin{aligned}
\mu\otimes\mu'[g] &= m[f]\\
\mu\otimes\mu'[(g-\mu\otimes\mu'[g])^2] &= m[(f-\mu[f])^2]
\end{aligned}
$$

and since

$$\forall\ (x,y) \in \mathbb{R}^2, \qquad (\partial_1 g)^2(x,y) + (\partial_2 g)^2(x,y) \quad=\quad (f'(tx+\sqrt{1-t^2}y))^2$$

we are also assured of $\mu\otimes\mu'[(\partial_1 g)^2 + (\partial_2 g)^2] = m[(f')^2]$. Thus it appears that

$$(\lambda\wedge\lambda')\ m[(f-\mu[f])^2] \quad\leq\quad m[(f')^2]$$

and the above lemma follows at once. □

These two invariance properties lead us to introduce for any $p \in \mathbb{N}$ and any $\lambda > 0$, the class $\mathcal{N}_p(\lambda)$ of elements from $\mathcal{N}_p$ with a spectral gap larger than λ. Our main task in this section will be to prove the following result which will be fundamental for our future estimations.

Proposition 3.3. *Let $p \in \mathbb{N}$ and $\lambda > 0$ be fixed. With the notations of the previous section, assume that all the distributions μ_n, $n \in \mathbb{N}$, belong to $\mathcal{N}_p(\lambda)$, then there exists a constant $C > 0$ (only depending on p and λ) such that*

$$\forall\; n\in\mathbb{N},\quad \left|\frac{1}{2}\ln\left(\frac{n+2}{n+1}\right)-\frac{1}{2(n+1)}+\int\ln\left(\varphi_{n+1}\right)\,dm_{n+1}\right| \quad\leq\quad \frac{C}{(n+1)^{(p+1)/2}}$$

The proof of this bound is based on classical Taylor expansions (thus in some sense, we are only recycling the idea underlying the simple proof of the Central Limit Theorem via characteristic functions), but we will take some care in justifying them in the next string of technical lemmas. We begin by introducing some notations.
Let a probability μ be fixed, we define for any $0 \leq t \leq 1$ and any $x, y \in \mathbb{R}$,

$$\begin{aligned} h(t) &\coloneqq t\sqrt{1+t^2} \\ U_t(x,y) &= h(t)xy - t^2\frac{y^2}{2} \\ F_x(t) &\coloneqq \ln\left(\int \exp(U_t(x,y))\,d\mu\right) \end{aligned}$$

The parameter t should be think of as $1/\sqrt{n+1}$, since one would have noticed that if $\mu = \mu_{n+1}$, then we get by definition,

$$\forall\; x\in\mathbb{R},\qquad \varphi_{n+1}(x) \quad=\quad F_x\left(\frac{1}{\sqrt{n+1}}\right)$$

So we are interested in differentiating $F_x(t)$ with respect to small t to obtain, for large $n \in \mathbb{N}$, appropriate expansions of the expression considered in Proposition 3.3. Formally it is not very difficult, and the Gibbs probability

$$\mu_{t,x}(dy) \quad\coloneqq\quad \frac{\exp(U_t(x,y))\,\mu(dy)}{\mu[\exp(U_t(x,\cdot))]} \tag{3.1}$$

appears to have a promising role (equally note that the probability $M_{n+1,x}$ introduced in the previous section can then be written $\mu_{\frac{1}{\sqrt{n+1}},\sqrt{\frac{n+1}{n+2}}x}$, if $\mu = \mu_{n+1}$). For instance and at least heuristically, we get

$$\partial_t F_x(t) \quad=\quad \int \partial_t U_t(x,y)\,\mu_{t,x}(dy)$$

Indeed this inequality is correct; the usual rule of differentiation under the integral is fulfilled, since we check that for any $0 \leq t \leq 1$ and $x, y \in \mathbb{R}$,

$$|\partial_t U_t(x,y)\exp(U_t(x,y))| \quad\leq\quad \left|\frac{1+2t^2}{\sqrt{1+t^2}}xy - ty^2\right|\exp((1+t^2)x^2/2)$$

$$\leq \quad (3\,|x|\,|y| + y^2)\exp(x^2)$$

which is integrable in y with respect to μ. Nevertheless, we need better estimations of $\partial_t F_x(t)$ than those deduced from this bound and the Jensen inequality $\mu[\exp(U_t(x,\cdot)] \geq \exp(\mu[U_t(x,\cdot)])$, because we shall rather differentiate in t integrals of $F_x(t)$ with respect to certain distributions of x (which will not necessarily integrate expression like $\exp(\epsilon x^2)$, for any $\epsilon > 0$; recall that the typical example of a probability on $\mathbb{R}$ having a spectral gap is the exponential law on $\mathbb{R}_+$). Before working in this direction, let us recall a general result, in fact valid on any measurable space.

Lemma 3.4. *Let η be a probability and V be a nonnegative measurable function. Then for any $q \geq 0$, we have*

$$\int V^q \frac{\exp(-V)}{Z} d\eta \quad \leq \quad \int V^q \, d\eta$$

where $Z \coloneqq \int \exp(-V)\, d\eta$ is the normalizing constant.

Proof. For $s > 0$, let $Z_s \coloneqq \int \exp(-sV)\, d\eta$ and define η_s as the probability $\exp(-sV)\eta/Z_s$. Without any difficulty, we compute that for $s > 0$,

$$\begin{aligned} \partial_s \int V^q \, d\eta_s &= -\int V^{q+1}\, d\eta_s + \int V^q \, d\eta_s \int V \, d\eta_s \\ &\leq -\int V^{q+1}\, d\eta_s + \left(\int V^{q+1}\, d\eta_s\right)^{q/(q+1)} \left(\int V^{q+1}\, d\eta_s\right)^{1/(q+1)} \\ &= 0 \end{aligned}$$

So for any $0 < u \leq s$, we have

$$\int V^q \, d\eta_s \quad \leq \quad \int V^q \, d\eta_u$$

and the RHS is converging to $\eta[V^q] \leq +\infty$ when u goes to 0_+. □

This simple bound will be quite useful to deduce the next crucial one:

Lemma 3.5. *Assume that μ admits a spectral gap $\lambda > 0$, then for any $q \in \mathbb{N}$, there exists a finite constant $C(\lambda, q)$ such that*

$$\forall\ 0 \leq t \leq 1, \forall\ x \in \mathbb{R}, \qquad \int |y|^q \ \mu_{t,x}(dy) \quad \leq \quad C(\lambda, q)(1 + |x|^{3q})$$

Proof. By classical approximation results, the bound

$$\int f^2 \, d\mu \quad \leq \quad \left(\int f \, d\mu\right)^2 + \lambda^{-1} \int (f')^2 \, d\mu$$

is extended to any function f of class $\mathcal{C}^1$ on $\mathbb{R}$, by allowing that the RHS can be infinite.

For fixed $t \geq 0$, $x \in \mathbb{R}$ and $p \in \mathbb{N}^*$, we apply this inequality with the mapping f defined by

$$f\,:\,\mathbb{R} \ni y \quad \mapsto \quad y^p \exp(U_t(x,y)/2)$$

so, after dividing by $\int \exp(U_t(x,y))\,d\mu$, we obtain that

(3.2)

$$\begin{aligned}
&\int y^{2p}\,\mu_{t,x}(dy)\\
&\quad\leq \frac{\left(\int y^p \exp(U_t(x,y)/2)\,\mu(dy)\right)^2}{\int \exp(U_t(x,y))\,d\mu} + \int (py^{p-1} + y^p(h(t)x - t^2y)/2)^2\,\mu_{t,x}(dy)\\
&\quad\leq \int y^{2p}\,\mu(dy) + 2p^2\int y^{2p-2}\,\mu_{t,x}(dy) + \frac{1}{2}\int y^{2p}(h(t)x - t^2y)^2\,\mu_{t,x}(dy)
\end{aligned}$$

The two first terms of the RHS are quite easy to dispose of: due the spectral gap inequality verified by μ, we know there exists a finite constant $C_1(\lambda, 2p)$ bounding $\int y^{2p}\,\mu(dy)$ independently of such μ. In other respects, there exists a finite constant $C_2(\lambda, 2p)$ such that

$$\forall\, y \in \mathbb{R}, \qquad y^{2p-2} \quad \leq \quad C_2(\lambda, 2p) + \frac{1}{6p^2}y^{2p}$$

relation implying that

$$2p^2\int y^{2p-2}\,\mu_{t,x}(dy) \quad \leq \quad 2p^2C_2(\lambda,2p) + \frac{1}{3}\int y^{2p}\,\mu_{t,x}(dy)$$

To treat the last term of (3.2), we rewrite it as

(3.3)

$$\begin{aligned}
&\frac{t^2}{2}\int y^{2p}(\sqrt{1+t^2}x - ty)^2\,\mu_{t,x}(dy)\\
&\quad\leq \frac{t}{2}\int |y|^{2p-1}\,(\sqrt{1+t^2}\,|x| + |\sqrt{1+t^2}x - ty|)(\sqrt{1+t^2}x - ty)^2\,\mu_{t,x}(dy)\\
&\quad\leq \frac{h(t)}{2}\,|x|\int |y|^{2p-1}\,(\sqrt{1+t^2}x - ty)^2\,\mu_{t,x}(dy)\\
&\qquad + \frac{t}{2}\int |y|^{2p-1}\left|\sqrt{1+t^2}x - ty\right|^3\,\mu_{t,x}(dy)
\end{aligned}$$

Noting that we also have for any $t \geq 0$ and $x \in \mathbb{R}$,

$$\mu_{t,x}(dy) \quad = \quad \frac{\exp(-(\sqrt{1+t^2}x - ty)^2)\mu(dy)}{\int \exp(-(\sqrt{1+t^2}x - tz)^2)\,\mu(dz)}$$

we are led to apply twice Lemma 3.4 with reference probability and potential

$$\begin{aligned}
\eta(dy) \quad &\coloneqq \quad \frac{|y|^{2p-1}\,\mu(dy)}{\int |z|^{2p-1}\,\mu(dz)}\\
V(y) \quad &\coloneqq \quad (\sqrt{1+t^2}x - ty)^2
\end{aligned}$$

and respectively with $q = 1$ and $q = 3/2$. Thus we get for instance that for $0 \le t \le 1$,

$$\begin{aligned}
&\frac{h(t)}{2}|x|\int |y|^{2p-1}(\sqrt{1+t^2}x-ty)^2\,\mu_{t,x}(dy)\\
&\le \frac{1}{\sqrt{2}}|x|\int |y|^{2p-1}(\sqrt{1+t^2}x-ty)^2\,\mu(dy)\int |y|^{2p-1}\,\mu_{t,x}(dy)\\
&\le C_3(\lambda,p)(1+|x|^3)\left(\int |y|^{2p}\,\mu_{t,x}(dy)\right)^{1-1/(2p)}
\end{aligned}$$

where $C_3(\lambda, p)$ is an appropriate constant (once again we have used that the quantity $\int |y|^{2p+3}\,\mu(dy)$ is uniformly bounded over probabilities μ with a spectral gap larger than λ). Now using a Young relation, we can find another finite constant $C_4(\lambda, p)$ such that the last RHS is bounded by

$$C_4(\lambda,p)(1+x^{6p}) + \frac{1}{6}\int |y|^{2p}\,\mu_{t,x}(dy)$$

We can proceed in a similar way with the term $t\int |y|^{2p-1}|\sqrt{1+t^2}x-ty|^3\mu_{t,x}(dy)/2$ and combining all these estimates we end up with the bound stated in the above lemma if $q = 2p$. The general case follows by suitable Hölder inequalities. □

In particular, these computations show that under the hypothesis of previous lemma, one can find a finite constant $C(\lambda)$ such that for any $0 \le t \le 1$ and $x \in \mathbb{R}$,

$$\begin{aligned}
|\partial_t F_x(t)| &= \left|\int \frac{1+2t^2}{\sqrt{1+t^2}}xy - ty^2\,\mu_{t,x}(dy)\right|\\
&\le 3|x|\int |y|\,\mu_{t,x}(dy) + \int y^2\,\mu_{t,x}(dy)\\
&\le C(\lambda)(1+x^6)
\end{aligned}$$

The next result generalizes this kind of bound:

Lemma 3.6. *For any fixed $x \in \mathbb{R}$, the mapping $[0,1] \ni t \mapsto F_x(t)$ belongs to $\mathcal{C}^\infty([0,1])$ and if we assume that μ admits a spectral gap $\lambda > 0$, then for any given $p \in \mathbb{N}^*$, there exists another finite constant $C(\lambda, p)$ such that*

$$\forall\ 0 \le t \le 1,\ \forall\ x \in \mathbb{R}, \qquad |\partial_t^p F_x(t)| \le C(\lambda,p)(1+x^{6p})$$

Proof. If V is a polynomial function in one variable, it is not difficult to justify the following differentiation under the integral, as in the discussion before Lemma 3.4, for any $0 \le t \le 1$ and $x \in \mathbb{R}$,

$$\begin{aligned}
\partial_t \int V(y)\,\mu_{t,x}(dy) &= \int V(y)\partial_t U(x,y)\,\mu_{t,x}(dy)\\
&\quad - \int V(y)\,\mu_{t,x}(dy)\int \partial_t U_t(x,y)\,\mu_{t,x}(dy)
\end{aligned}$$

so taking into account that h belongs to $\mathcal{C}^\infty(\mathbb{R}_+)$, it appears easily that $[0,1] \ni t \mapsto F_x(t)$ is equally of class $\mathcal{C}^\infty$. Indeed, if $p \in \mathbb{N}^*$ is given, $\partial_t^p F_x(t)$ appears as a weighted sum of products of expressions like

$$\int \partial_t^{\alpha_0} U_t(x,y) \partial_t^{\alpha_1} U_t(x,y) \cdots \partial_t^{\alpha_r} U_t(x,y)\, \mu_{t,x}(dy)$$

where $\alpha_0 \geq \alpha_1 \geq \cdots \geq \alpha_r > 0$ are $r+1$ nonnegative integers. More precisely, let us denote $H_\alpha(t,x)$ this integral, where $\alpha = (\alpha_i)_{i\in\mathbb{N}}$ is a multi-index whose entries are non-increasing elements of $\mathbb{N}$ and $\alpha_l = 0$ for $l > r$. Let $\mathcal{A}$ be the set of all such sequences (with varying $r \in \mathbb{N} \sqcup \{-1\}$, $r = -1$ corresponds to the element of $\mathcal{A}$ whose all entries are 0 and which we will also designate by 0, by traditional conventions $H_0 = \mathbb{1}$) and for $\alpha = (\alpha_i)_{i\in\mathbb{N}} \in \mathcal{A}$, we note $|\alpha| = \sum_{i\in\mathbb{N}} \alpha_i \in \mathbb{N}$ (α is then sometimes called a partition of $|\alpha|$).
Let us go one step further and iterate this construction. First we put on $\mathcal{A}$ the lexicographical total order, namely for two given elements $\alpha = (\alpha_i)_{i\in\mathbb{N}}$ and $\alpha' = (\alpha'_i)_{i\in\mathbb{N}}$ of $\mathcal{A}$, we say that $\alpha \geq \alpha'$ if there exists $i_0 \in \mathbb{N} \sqcup \{\infty\}$ such that $\alpha_i = \alpha'_i$ for any $0 \leq i < i_0$ and $\alpha_{i_0} > \alpha'_{i_0}$ (of course this condition is void if $i_0 = \infty$, or equivalently if $\alpha = \alpha'$). Next we consider $\mathbb{A}$ the set of all sequences of non-increasing elements of $\mathcal{A}$ which are null after some rank. As before, the height of an element $A = (A_i)_{i\in\mathbb{N}} \in \mathbb{A}$ is the nonnegative integer $|A| \coloneqq \sum_{i\in\mathbb{N}} |A_i|$ and we associate to A the mapping H_A defined on $[0,1]\times\mathbb{R}$ by

$$\forall\, 0 \leq t \leq 1,\, \forall\, x \in \mathbb{R}, \qquad H_A(t,x) \ \coloneqq\ \prod_{i\in\mathbb{N}} H_{A_i}(t,x)$$

Then it can be shown recursively that for any $A \in \mathbb{A}$, there exists an integer $N(A) \in \mathbb{Z}$, independent from the real numbers $0 \leq t \leq 1$, $x \in \mathbb{R}$, and from the underlying distribution μ, such that for any fixed $p \in \mathbb{N}^*$,

$$\forall\, 0 \leq t \leq 1,\, \forall\, x \in \mathbb{R}, \qquad \partial_t^p F_x(t) \ =\ \sum_{A\in\mathbb{A}\,:\,|A|=p} N(A) H_A(t,x)$$

(one would have noticed there is only a finite number of $A \in \mathbb{A}$ verifying $|A| = p$).

To compute the coefficients $N(A)$, for $A \in \mathbb{A}$, one can apply a sort of tree algorithm: if $A = (A_0, ..., A_r, 0, ...) \in \mathbb{A}$ is given, with $A_r \neq 0$, at height $|A|$, it gives birth to three types of sons, each of them of height $|A|+1$:
• Let us denote for $0 \leq i \leq r$, $A_i = (\alpha_{i,0}, \alpha_{i,1}, ..., \alpha_{i,r_i}, 0, ...)$, with $\alpha_{i,r_i} > 0$. Then for any choice of $0 \leq i \leq r$ and $0 \leq j \leq r_i$, we obtain a son of A by replacing $\alpha_{i,j}$ by $\alpha_{i,j}+1$ (all the other coordinates remaining the same) and by rearranging in a natural way the object thus obtained in order to ensure that it still belongs to $\mathbb{A}$ (i.e. that the monotonicity properties entering the definitions of $\mathcal{A}$ and $\mathbb{A}$ are fulfilled). Thus one has created $\sum_{0\leq i\leq r} r_i$ sons and some of them can be equal.
• Another type of son is obtained by choosing an index $0 \leq i \leq r$, replacing A_i by $(\alpha_{i,0}, \alpha_{i,1}, ..., \alpha_{i,r_i}, 1, 0, 0, ...)$ and proceeding to the necessary rearrangements. This operation creates r sons, as before not necessarily different.
• Finally, we add r new sons, all of them equal to $(A_0, A_1, ..., A_r, 1, 0, 0, ...)$, where 1 designate here the element of $\mathcal{A}$ given by $(1, 0, 0, ...)$.

If B is a son of A, we will write $A \rightharpoonup B$ if it was created by one of the two first procedures and $A \rightharpoondown B$ otherwise. Then we have that for any $B \in \mathbb{A}$,

$$N(B) = \sum_{A\in\mathbb{A}\,:\,A\rightharpoonup B} N(A) - \sum_{A\in\mathbb{A}\,:\,A\rightharpoondown B} N(A)$$

For instance, we compute that at height 3, for any $0 \le t \le 1$ and any $x \in \mathbb{R}$,

$$\begin{aligned}\partial_t^3 F_x(t) &= H_{A^{(0)}}(t,x) + 3H_{A^{(1)}}(t,x) + H_{A^{(2)}}(t,x) - 3H_{A^{(3)}}(t,x) \\ &\quad -3H_{A^{(4)}}(t,x) + 2H_{A^{(5)}}(t,x)\end{aligned} \tag{3.4}$$

where

$$A^{(0)} \coloneqq \begin{array}{|c|}\hline 3 \\ \hline\end{array} \qquad A^{(1)} \coloneqq \begin{array}{|c|c|}\hline 2 & 1 \\ \hline\end{array} \qquad A^{(2)} \coloneqq \begin{array}{|c|c|c|}\hline 1 & 1 & 1 \\ \hline\end{array}$$

$$A^{(3)} \coloneqq \begin{array}{|c|}\hline 2 \\ \hline 1 \\ \hline\end{array} \qquad A^{(4)} \coloneqq \begin{array}{|c|c|}\hline 1 & 1 \\ \hline 1 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array} \qquad A^{(5)} \coloneqq \begin{array}{|c|}\hline 1 \\ \hline 1 \\ \hline 1 \\ \hline\end{array}$$

(with obvious notations, we have represented the elements of $\mathbb{A}$ as arrays, the lines corresponding to elements of $\mathcal{A}^*$, but the fact that above they appear under forms of Young tableaux is accidental).

Now let us come back for some given $\alpha = (\alpha_i)_{i\in\mathbb{N}} \in \mathcal{A}^* \coloneqq \mathcal{A}\setminus\{0\}$ and for $0 \le t \le 1$, $x, y \in \mathbb{R}$, to the expression $\prod_{i\in\mathbb{N}} \partial_t^{\alpha_i} U_t(x,y)$, with the unusual convention that $\partial_t^0 U_t(x,y) \equiv 1$. Taking into account that

$$\begin{aligned} \partial_t U_t(x,y) &= h'(t)xy - ty^2 \\ \partial_t^2 U_t(x,y) &= h''(t)xy - y^2 \\ \forall\, k\in\mathbb{N},\, k\ge 3, \qquad \partial_t^k U_t(x,y) &= h^{(k)}(t)xy \end{aligned}$$

we see there exist coefficients $a(\alpha,k,t)$ depending on $\alpha \in \mathcal{A}^*$, $k \in \mathbb{N}$, $k \le r(\alpha) \coloneqq \max\{i \in \mathbb{N} \,:\, \alpha_i \ne 0\}$ and $0 \le t \le 1$, such that this ensures the following polynomial expansion for any $x, y \in \mathbb{R}$,

$$\prod_{i\in\mathbb{N}} \partial_t^{\alpha_i} U_t(x,y) = \sum_{0\le k\le r(\alpha)} a(\alpha,k,t) x^k y^{2r(\alpha)-k}$$

Furthermore it is quite clear that for any $p \in \mathbb{N}$, the quantity

$$b(p) \coloneqq \sup_{\alpha\in\mathcal{A}\,:\,|\alpha|=p} \left(\sup_{0\le k\le r(\alpha),\, 0\le t\le 1} |a(\alpha,k,t)| \right)$$

is finite, so via Young inequalities, we can find a family of constants $(C(p))_{p\in\mathbb{N}^*}$ such that for any $\alpha = (\alpha_i)_{i\in\mathbb{N}} \in \mathcal{A}^*$,

$$\forall\, 0\le t\le 1,\ \forall\, x,y\in\mathbb{R}, \qquad \left|\prod_{i\in\mathbb{N}} \partial_t^{\alpha_i} U_t(x,y)\right| \le C(|\alpha|)(1 + |x|^{2|\alpha|} + |y|^{2|\alpha|})$$

(one would have noticed the general fact $r(\alpha)\leq|\alpha|$). The bound stated in Lemma 3.6 follows by putting together all these considerations and applying Lemma 3.5 and some more Young inequalities. □

We now have at our disposal all the ingredients needed for the proof of Proposition 3.3.
So let $p\in\mathbb{N}$ and $\lambda>0$ be fixed as in the statement of this result. We make the hypothesis that we are given two probabilities $m,\mu\in\mathcal{N}_p(\lambda)$ and we define for $0\leq t\leq 1$,

$$G(t) \quad := \quad \frac{1}{2}\ln(1+t^2)-\frac{1}{2}t^2-\int F_x(t)\,m(dx) \tag{3.5}$$

which is just the expression we want to bound, if $m=m_{n+1}$, $\mu=\mu_{n+1}$ and $t=1/\sqrt{n+1}$. We also notice that

$$G(t) \quad = \quad \int \ln\left(\frac{d\nu Q_{\mu,t}}{d\nu}\right)dm$$

where $Q_{\mu,t}$ is the Markovian kernel defined as K_{n+1}, but replacing μ_{n+1} by μ and $n+1$ by $1/t^2$ (namely for any $x\in\mathbb{R}$, $Q_{\mu,t}(x\cdot)$ is the law of $\frac{x}{\sqrt{1+t^2}}+\frac{tX}{\sqrt{1+t^2}}$, where X is distributed as μ). In particular G is identically equal to zero if $\mu=\nu$, since in this case the standard Gaussian distribution ν is invariant (even reversible) for $Q_{\nu,t}$, for any $t\geq 0$.
The positive spectral gap of m ensures that it admits moments of all order, thus estimates of Lemma 3.6 enable us to be convinced that $[0,1]\ni t\mapsto G(t)$ belongs to $\mathcal{C}^\infty([0,1])$ and that we can find a finite constant $C_5(\lambda,p)$, depending only on λ and p, bounding $G^{(p+1)}(t)$ uniformly in $t\in[0,1]$. Classical Taylor expansion up to the p-order with remainder term then show that for any $0\leq t\leq 1$,

$$\left|G(t)-G(0)-G'(0)t-\cdots-G^{(p)}(0)\frac{t^p}{p!}\right| \quad \leq \quad C_5(\lambda,p)\frac{t^{p+1}}{(p+1)!} \tag{3.6}$$

But let us return to the considerations developed in the proof of Lemma 3.6. Since we have for any $x,y\in\mathbb{R}$,

$$\begin{array}{rrcl} & \partial_t U_t(x,y)\big|_{t=0} & = & h'(0)xy \quad = \quad xy \\ & \partial_t^2 U_t(x,y)\big|_{t=0} & = & h''(0)xy-y^2 \quad = \quad -y^2 \\ \forall\ k\in\mathbb{N},\ k\geq 3, & \partial_t^k U_t(x,y)\big|_{t=0} & = & h^{(k)}(0)xy \end{array}$$

it appears that for any $\alpha=(\alpha_i)_{i\in\mathbb{N}}\in\mathcal{A}^*$, the expression $\prod_{i\in\mathbb{N}}\partial_t^{\alpha_i}U_t(x,y)\big|_{t=0}$ is a polynomial in x and y where the variable y (respectively x) appears at most at the power $|\alpha|$. Furthermore, for any $x\in\mathbb{R}$, $\mu_{0,x}=\mu$, so for any $k\in\mathbb{N}$, $0<k\leq p$, the quantity $G^{(k)}(0)$ only depends on m and μ through their moments of order less or equal to p. By assumption, the latters coincide with those of the standard normal law and we know that if $\mu=\nu$ then all derivatives of G are null. Thus we conclude that in our case we equally have $G'(0)=\cdots=G^{(p)}(0)=0=G(0)$ (the latter equality is always true), so Proposition 3.3 follows from (3.6), once we allow

the replacements $m = m_{n+1}$, $\mu = \mu_{n+1}$ and $t = 1/\sqrt{n+1}$ (taking into account Lemmas 3.1 and 3.2 to check that $m_{n+1} \in \mathcal{N}_p(\lambda)$).

Remarks 3.7. a) Because the laws we considered have moments of all order, we have not taken much care in exponents of the variable x. Nevertheless, let us indicate that it is possible to improve the $|x|^{3q}$ of Lemma 3.5 into x^{2q} (and by consequence the x^{6p} of Lemma 3.6 can be replaced by x^{4p}). To do so, instead of upper bounding $t\,|y|$ by $\sqrt{1+t^2}\,|x| + |\sqrt{1+t^2}x - ty|$ in (3.3), rather bound t^2y^2 by $(\sqrt{1+t^2}x-ty)^2 + 2\sqrt{1+t^2}\,|x|\,|\sqrt{1+t^2}x-ty| + (1+t^2)x^2$. One can then carry on with manipulations similar to those presented after (3.3) and conclude to the above mentioned improvement. But we are not sure that the power of x thus obtained is the best possible one. Indeed, we conjecture that in Lemma 3.5, it is a term like $|x|^q$ which should appear. Note that it is the case if $\mu = \nu$, situation where exact computations can be conducted.
b) At the end of the proof of Proposition 3.3, the requirement that $m \in \mathcal{N}_p(\lambda)$ was a little too strong. In view of above remark, if $p \in \mathbb{N}^*$ is fixed, it is sufficient that m admits a moment of order $4(p+1)$ to obtain there exists a finite constant C depending on p, $\int x^{4(p+1)}\, m(dx)$ and $\lambda > 0$, such that $\mu \in \mathcal{N}_p(\lambda)$ implies that

$$\forall\, 0 \le t \le 1, \qquad |G(t)| \;\le\; Ct^{p+1}$$

c) In next section we will also discuss about the hypothesis that μ admits a positive spectral gap. But let us already mention that as far as only bounds like (3.6) are concerned, there is an easy condition dispensing us from this assumption. It corresponds to the cases where μ has a compact support, say for instance that $\mu[(-\infty,-M) \sqcup (M,+\infty)] = 0$, where $M > 0$ is finite. In this situation we are assured in Lemma 3.5 of the obvious bound $\int |y|^q\, \mu_{t,x}(dy) \le M^q$, for any $q \in \mathbb{N}$, any $0 \le t \le 1$ and above all any $x \in \mathbb{R}$. Then reexamining the above computations, it appears that for any $p \in \mathbb{N}^*$, we can find a finite constant C depending only on p, M and $\int |x|^{p+1}\, m(dx)$ (assumed to be finite), insuring that for all such $\mu \in \mathcal{N}_p$, we have $|G(t)| \le Ct^{p+1}$. Recall that there exists such probabilities μ which are finite weighted sums of Dirac masses and thus are not admitting a positive spectral gap in the way we have defined it.
d) Using the computation made in (3.4), it appears at height 3 that for $\mu, m \in \mathcal{N}_2(\lambda)$, we have

$$\begin{aligned} G^{(3)}(0) &= \int \partial_t^3 F_x(0)\, m(dx) \\ &= \int H_{A^{(2)}}(0,x)\, m(dx) \\ &= \int x^3\, m(dx) \int y^3\, \mu(dy) \end{aligned}$$

Thus if furthermore $\int y^3\, \mu(dy) \ne 0 \ne \int x^3\, m(dx)$, then the corresponding $G(t)$ is equivalent to $G^{(3)}(0)t^3/6$ for small $t > 0$ (the difference between these terms being at least of order $\mathcal{O}(t^4)$). Nevertheless a little miracle comes to our rescue when we apply this result: if for all $n \in \mathbb{N}$, $\mu_n \in \mathcal{N}_2$ and μ_n admits a moment of order 3, we get that for any $n \in \mathbb{N}$,

$$\begin{aligned}\int x^3\, m_n(dx) &= \mathbb{E}[Y_n^3]\\ &= \frac{1}{(n+1)^{3/2}}\sum_{0\leq i\leq n}\mathbb{E}[X_i^3]\end{aligned}$$

so typically if $\int y^3\,\mu_n(dy)$ is bounded uniformly in $n \in \mathbb{N}$, then $\int x^3\, m_n(dx)$ is of order $1/\sqrt{n+1}$. In particular, assuming that for all $n \in \mathbb{N}$, $\mu_n \in \mathcal{N}_2(\lambda)$ for some fixed $\lambda > 0$, we end up with the existence of a constant $C(\lambda)$ depending only on λ, such that

$$\forall\ n \in \mathbb{N}, \qquad |\mathrm{Ent}(m_{n+1}|\nu) - \mathrm{Ent}(m_{n+1}|\nu K_{n+1})| \ \leq\ \frac{C(\lambda)}{(n+1)^2}$$

Thus taking into account the conjecture given in remark 2.1 (c), this estimate is quite promising, since we would obtain a general bound of order $\ln(e+n)/(n+1)$ (as already mentioned, the considerations of next section will indicate why the spectral gap assumption is not so crucial for the above arguments).

It is time now to present examples where the entropy goes to zero faster than what is predicted by Theorem 1.1. Of course, the (basic, i.e. not taking into account modified logarithmic Sobolev inequalities) Markovian considerations of section 2 and the estimates of Proposition 3.3 are not enough for this kind of result, since they will only offer bounds which are increasing with respect to time. So we need another trick; the convolution with the standard Gaussian law and rearrangements of random variables.

For fixed $p \in \mathbb{N}\setminus\{0,1\}$ and $\lambda > 0$, let us denote by $\widetilde{\mathcal{N}}_p(\lambda)$ the set of laws m which are constructed as in Lemma 3.1, with $\mu \in \mathcal{N}_p(\lambda)$, $\mu' = \nu$ and $t = 1/2$. It follows from Lemmas 3.1 and 3.2 that $\widetilde{\mathcal{N}}_p(\lambda) \subset \mathcal{N}_p(1\wedge\lambda) = \mathcal{N}_p(\lambda)$, since the spectral gap of ν is just 1 and it is the largest possible spectral gap of elements of $\mathcal{N}_2$, as it can be checked by considering the identity as test function. Let us mention that such perturbed measures also lay at the heart of previous analysis of entropic convergence in the Central Limit Theorem by Linnik [16], Brown [8] and Barron [6]. Furthermore, part of the recent progresses of Johnson and Barron [14] and of Ball, Barthe and Naor [5] is to get rid of this necessity.

Proposition 3.8. *In the setting of section 2, if all the distributions μ_n, $n \in \mathbb{N}$, belong to $\widetilde{\mathcal{N}}_p(\lambda)$, for some $p \in \mathbb{N}\setminus\{0,1\}$ and $\lambda > 0$, then there exist a finite constant $C(\lambda,p)$ depending only on those parameters, such that*

$$\forall\ n \in \mathbb{N}, \qquad \mathrm{Ent}(m_n|\nu) \ \leq\ \frac{C(\lambda,p)}{(n+1)^{\frac{p-1}{2}}}$$

Proof. Let us return to probabilist notations. By definition of $\widetilde{\mathcal{N}}_p(\lambda)$, for each $n \in \mathbb{N}$, we can write $X_n = Z_n/\sqrt{2} + W_n/\sqrt{2}$, where Z_n and W_n are independent

and whose respective law belongs to $\mathcal{N}_p(\lambda, p)$ for the former and is equal to ν for the latter. We can also assume that all the random variables Z_n, $n \in \mathbb{N}$, and $W_{n'}$, $n' \in \mathbb{N}$, are mutually independent (at least, these considerations are justified up to a possible modification of the underlying probability space).
Let a time $N \in \mathbb{N}^*$ be temporally fixed. We consider a new set of random variables $(\widetilde{X}_n)_{0\leq n\leq N}$ defined by

$$\widetilde{X}_n \quad \coloneqq \quad \begin{cases} W_{2n}/\sqrt{2} + W_{2n+1}/\sqrt{2} & \text{, if } 0 \leq n < \lfloor (N+1)/2 \rfloor \\ W_{2n}/\sqrt{2} + Z_0/\sqrt{2} & \text{, if } \lfloor (N+1)/2 \rfloor \leq n < (N+1)/2 \\ Z_{2n-N-1}/\sqrt{2} + Z_{2n-N}/\sqrt{2} & \text{, if } (N+1)/2 \leq n \leq N \end{cases}$$

For $0 \leq n \leq N$, let us also denote $\widetilde{m}_n$ the law of $(\widetilde{X}_0 + \cdots + \widetilde{X}_n)/\sqrt{n+1}$. In particular we have $\widetilde{m}_N = m_N$. But up to time $\lfloor (N+1)/2 \rfloor - 1$, we have $\widetilde{m}_n = \nu$ and after this time the difference of entropy $\mathrm{Ent}(\widetilde{m}_{n+1}|\nu) - \mathrm{Ent}(\widetilde{m}_n|\nu)$ is bounded above by $C_6(\lambda, p)/(n+1)^{\frac{p+1}{2}}$, for a certain finite constant $C_6(\lambda, p)$ depending only on λ and p, due to the fact that the law of $\widetilde{X}_{n+1}$ belongs to $\mathcal{N}_p(\lambda)$, according to the remark before the statement of Proposition 3.8. So we end up with the estimate

$$\begin{aligned} \mathrm{Ent}(m_N|\nu) &= \mathrm{Ent}(\widetilde{m}_N|\nu) \\ &= \sum_{\lfloor (N+1)/2 \rfloor - 1 \leq n < N} \mathrm{Ent}(\widetilde{m}_{n+1}|\nu) - \mathrm{Ent}(\widetilde{m}_n|\nu) \\ &\leq C_6(\lambda, p) \sum_{n \geq \lfloor (N+1)/2 \rfloor - 1} \frac{1}{(n+1)^{\frac{p+1}{2}}} \\ &\leq \frac{C(\lambda, p)}{(N+1)^{\frac{p-1}{2}}} \end{aligned}$$

with for instance $C(\lambda, p) \coloneqq \frac{p+1}{p-1} 2^{p-1} C_6(\lambda, p)$.
For $N = 0$, we directly get that $\mathrm{Ent}(m_0|\nu)$ is bounded above by a constant only depending on $\lambda > 0$ and $p \in \mathbb{N} \setminus \{0, 1\}$, by applying the inequality (3.6) with μ the law of Z_0, $m = \mu_0$ and $t = 1/2$ (and resorting to the convention that $\widetilde{m}_{-1} \coloneqq \nu$ is the law of W_0). □

In Lemma 4.4 of next section, we will see how to generalize this result to weaker convolutions.

4. Some bounds without spectral gap assumption

In the previous computations, the spectral gap hypothesis is not as crucial as it may seem at first view and we will discuss here ways to relax it.

At the end of last section, it was necessary to convolve with a Gaussian distribution to obtain our examples of fast convergence. We now recall how, to some extent, it is possible to "deconvolve", via an assumption of finite modified

Fisher information. The latter is the quantity associated to any probability m on $\mathbb{R}$ by the formula

$$I(m) \quad := \quad \begin{cases} \int \left|\nabla \ln\left(\frac{dm}{d\nu}\right)\right|^2 \, dm \;\leq\; +\infty & \text{, if } m \ll \nu \\ +\infty & \text{, otherwise} \end{cases}$$

where ∇ is the weak derivative corresponding to Radon-Nikodym differentiation with respect to Lebesgue measure.
To see its relation with the weighted convolutions under study, let us consider $(P_t)_{t\geq 0}$ the Ornstein-Uhlenbeck semigroup (which has already made a discreet apparition in last section as $(Q_{\nu,\ln(1+t^2)})_{t\geq 0}$), which acts on nonnegative measurable functions f by

$$\forall\, t \geq 0, \forall\, x \in \mathbb{R}, \qquad P_t[f](x) \quad = \quad \int f(\exp(-t/2)x + \sqrt{1-\exp(-t)}y)\, \nu(dy)$$

The next result is so standard in Markovian semigroup theory (see e.g. [4] or [1]), that we will not recall its proof.

Theorem 4.1. *Let $f \in \mathbb{L}^1(\nu)$ be a density of probability with respect to ν. For any $t \geq 0$, we denote $m(t) := P_t[f]\nu$. Then we have*

$$\forall\, t \geq 0, \qquad \partial_t \mathrm{Ent}(m(t)|\nu) \quad = \quad -\frac{I(m(t))}{2}$$

and the mapping

$$\mathbb{R}_+ \ni t \quad \mapsto \quad I(m(t))$$

is non-increasing (i.e.the mapping $\mathbb{R}_+ \ni t \mapsto \mathrm{Ent}(m(t)|\nu)$ is convex).

We are particularly interested in the following consequence, which enables "small" deconvolution:

$$(4.1) \qquad \forall\, t \geq 0, \qquad \mathrm{Ent}(m|\nu) \quad \leq \quad \mathrm{Ent}(m(t)|\nu) + I(m)t/2$$

In order to take advantage of this bound, let us reformalize the results obtained in section 3.

Definition 4.2. For any fixed constants $r, K \geq 0$, we define $\mathcal{M}_r^{(1)}(K)$ the set of probabilities m on $\mathbb{R}$ verifying $\int |x|^r \, m(dx) \leq K$. If furthermore $p \in \mathbb{N}$ is given, then let $\mathcal{M}_{p,r}^{(2)}(K)$ be the subset of $\mathcal{N}_p$ whose elements μ satisfy

$$\forall\, 0 \leq t \leq 1, \forall\, x \in \mathbb{R}, \qquad \int y^{2(p+1)}\, \mu_{t,x}(dy) \quad \leq \quad K(1+|x|^r)$$

where the Gibbs distribution $\mu_{t,x}$ was defined in (3.1) with respect to μ.

The interest of these sets of measures is that if G is defined as in (3.5) with respect to $m \in \mathcal{M}^{(1)}_{r/2+(r/2)\vee(p+1)}(K_1)$ and $\mu \in \mathcal{M}^{(2)}_{p,r}(K_1)$, for some finite constants $p \in \mathbb{N}$ and $r, K_1, K_2 \geq 0$, then we have seen how to obtain a finite constant $C(p, r, K_1, K_2)$ depending only on its parameters, such that

$$(4.2) \qquad \forall\, 0 \leq t \leq 1, \qquad |G(t)| \quad \leq \quad C(p, r, K_1, K_2)t^{p+1}$$

Here is our main statement without any apparent convolution, result which is not very good, since the order $\mathcal{O}(1/n)$ is only asymptotically approached as p goes to infinity!

Proposition 4.3. *Consider once again the setting of section 2. Let assume there exist constants $p \in \mathbb{N} \setminus \{0,1\}$, $r, K_1, K_2, K_3 \geq 0$ such that*

$$\forall\ n \in \mathbb{N}, \qquad \begin{cases} m_n \in \mathcal{M}^{(1)}_{r/2+(r/2)\vee(p+1)}(K_1) \\ \mu_{n+1} \in \mathcal{M}^{(2)}_{p,r}(K_2) \\ I(m_n) \leq K_3 \end{cases}$$

Then there exists another finite constant $C(p, r, K_1, K_2, K_3) \geq 0$ depending only on the previous ones, such that

$$\forall\ n \in \mathbb{N}, \qquad \mathrm{Ent}(m_n|\nu) \quad \leq \quad \frac{C(p, r, K_1, K_2, K_3)}{(n+1)^{(p-1)/(p+1)}}$$

The proof is based on the following extension of Proposition 3.8.

Lemma 4.4. *Assume that the sequence $(\mu_n)_{n\in\mathbb{N}}$ is as in previous proposition, except for the requirement of bounded modified Fisher information, and as usual let $(X_n)_{n\in\mathbb{N}}$ be independent variables distributed according to these laws. Let $0 < s < 1$ be fixed and $(W_n)_{n\in\mathbb{N}}$ be IID standard Gaussian variables, also independent from $(X_n)_{n\in\mathbb{N}}$. We consider for $n \in \mathbb{N}$, $\widetilde{X}_n = \sqrt{1-s^2}X_n + sW_n$ and we denote $\widetilde{m}_n$ the law of*

$$\widetilde{Y}_n \quad \coloneqq \quad \frac{\widetilde{X}_0 + \cdots + \widetilde{X}_n}{\sqrt{n+1}}$$

Then there exists a finite constant $C(p, r, K_1, K_2) \geq 0$, not depending on $0 < s < 1$, such that

$$\forall\ n \in \mathbb{N}, \qquad s^2(n+1) \geq 2 \quad \Longrightarrow \quad \mathrm{Ent}(\widetilde{m}_n|\nu) \quad \leq \quad \frac{C(p, r, K_1, K_2)}{[s^2(n+1)]^{(p-1)/2}}$$

Proof. As in the proof of Lemma 3.8, we begin by fixing a time horizon $N \in \mathbb{N}$, and we consider the new sequence of random variables $(\widehat{Y}_n)_{-1\leq n\leq N}$ defined by the iteration

$$\widehat{Y}_{-1} \quad \coloneqq \quad \frac{W_0 + \cdots + W_N}{\sqrt{N+1}}$$

and for any $-1 \leq n < N$

$$\widehat{Y}_{n+1} \quad \coloneqq \quad \sqrt{\frac{s^2(N+1) + (1-s^2)(n+1)}{s^2(N+1) + (1-s^2)(n+2)}}\widehat{Y}_n + \frac{\sqrt{1-s^2}}{\sqrt{s^2(N+1) + (1-s^2)(n+2)}}X_{n+1}$$

Indeed, we have for any $0 \leq n \leq N$,

$$\widehat{Y}_n \quad = \quad \frac{sW_0 + \cdots + sW_N + \sqrt{1-s^2}X_0 + \cdots + \sqrt{1-s^2}X_n}{\sqrt{s^2(N+1) + (1-s^2)(n+1)}}$$

and we denote by $\widehat{m}_n$ its law. In particular it appears that $\widehat{m}_N = \widetilde{m}_N$. Since we have for any $0 \leq n \leq N$,

$$\forall\ f \in \mathcal{B}(\mathbb{R}), \qquad \widehat{m}_n[f] \;=\; \int f(s_{N,n}x + \sqrt{1-s_{N,n}^2}y)\,\nu(dx)m_n(dy)$$

where $s_{N,n} \coloneqq \sqrt{N+1}s/\sqrt{(N+1)s^2+(1-s^2)(n+1)}$, we show without difficulty that all the $\widehat{m}_n$, $0 \leq n \leq N$, belong to $\mathcal{M}^{(1)}_{r'}(K_1(r'))$, with $r' \coloneqq r/2 + (r/2) \vee (p+1)$ and $K_1(r) \coloneqq 2^{r'/2}\max\{\int |x|^{r'}\,\nu(dx), K_1\}$. Thus we would like to apply bound (4.2) with $0 \leq t \leq 1$ defined by

$$\frac{t}{\sqrt{1+t^2}} \;=\; \widetilde{t} \;\coloneqq\; \frac{\sqrt{1-s^2}}{\sqrt{s^2(N+1)+(1-s^2)(n+2)}}$$

or equivalently $t = \widetilde{t}/\sqrt{1-\widetilde{t}^2}$. That is where the condition $s^2(n+1) \geq 2$ is useful, since it ensures that $\widetilde{t} \leq 1/\sqrt{2}$ and $t \leq 1$, and then that $t \leq \sqrt{2}\,\widetilde{t}$, thus via (4.2) we obtain

$$\begin{aligned}
&\mathrm{Ent}(\widehat{m}_{n+1}|\nu)\\
&\leq\; \mathrm{Ent}(\widehat{m}_n|\nu) + C_1(p,r,K_1,K_2)\left(\frac{\sqrt{1-s^2}}{\sqrt{s^2(N+1)+(1-s^2)(n+2)}}\right)^{p+1}
\end{aligned}$$

for an appropriate finite constant $C_1(p,r,K_1,K_2) \geq 0$ not depending on $0 < s < 1$. So summing these inequalities we end up with

$$\begin{aligned}
\mathrm{Ent}(\widetilde{m}_N|\nu) &\leq C_1(p,r,K_1,K_2)\sum_{-1\leq n\leq N-1}\frac{(1-s^2)^{(p+1)/2}}{(s^2(N+1)+(1-s^2)(n+2))^{\frac{p+1}{2}}}\\
&= C_1(p,r,K_1,K_2)(1-s^2)^{(p+1)/2}\left(\frac{1}{(s^2(N+1)+(1-s^2))^{\frac{p+1}{2}}}\right.\\
&\qquad\left.+\sum_{1\leq n\leq N}\frac{1}{(s^2(N+1)+(1-s^2)(n+1))^{\frac{p+1}{2}}}\right)
\end{aligned}$$

We bound above the last sum by

$$\begin{aligned}
&\int_1^N \frac{1}{(s^2(N+1)+(1-s^2)u)^{\frac{p+1}{2}}}\,du\\
&\leq \int_1^{+\infty}\frac{1}{(s^2(N+1)+(1-s^2)u)^{\frac{p+1}{2}}}\,du\\
&= \frac{2}{(1-s^2)(1-p)}\left[(s^2(N+1)+(1-s^2)u)^{\frac{1-p}{2}}\right]_1^{+\infty}\\
&= \frac{2}{(1-s^2)(p-1)}\frac{1}{(s^2(N+1)+(1-s^2))^{\frac{p-1}{2}}}
\end{aligned}$$

so we get, since $p \geq 2$,

$$\begin{aligned}\mathrm{Ent}(\widetilde{m}_N|\nu) &\leq C_1(p,r,K_1,K_2)\left(\frac{1}{(s^2(N+1)+(1-s^2))^{\frac{p+1}{2}}}\right.\\ &\qquad \left.+\frac{2}{p-1}\frac{1}{(s^2(N+1)+(1-s^2))^{\frac{p-1}{2}}}\right)\\ &\leq 3C_1(p,r,K_1,K_2)\frac{1}{(s^2(N+1)+(1-s^2))^{\frac{p-1}{2}}}\\ &\leq \frac{3C_1(p,r,K_1,K_2)}{(s^2(N+1))^{\frac{p-1}{2}}}\end{aligned}$$

which is the required result. □

We can now come to the

Proof of Proposition 4.3. Using the reversibility of the Ornstein-Uhlenbeck semigroup with respect to ν, we obtain, with the notations of above lemma and of Theorem 4.1,

$$\forall\, n \in \mathbb{N}, \qquad \widetilde{m}_n \;=\; m_n P_{\ln(1/(1-s^2))} \;=\; m_n(\ln(1/(1-s^2)))$$

thus taking into account bound (4.1), our assumptions imply that if $0 < s \leq 1/\sqrt{2}$ and if $(n+1)s^2 \geq 2$, then

$$\begin{aligned}\mathrm{Ent}(m_n|\nu) &\leq \mathrm{Ent}(\widetilde{m}_n|\nu) + \frac{1}{2}I(m_n)\ln\left(\frac{1}{1-s^2}\right)\\ &\leq \frac{C(p,r,K_1,K_2)}{[s^2(n+1)]^{(p-1)/2}} + \frac{K_3}{2}\frac{s^2}{1-s^2}\\ &\leq \frac{C(p,r,K_1,K_2)}{[s^2(n+1)]^{(p-1)/2}} + K_3 s^2\end{aligned}$$

It remains now to optimize in the parameter $0 < s \leq 1/\sqrt{2}$ to be convinced of the validity of the proposition. Indeed, the minimizing s is of order $(n+1)^{\frac{1-p}{2(p+1)}}$, so the condition $2/(n+1) \leq s^2 \leq 1/2$ is satisfied for large enough $n \in \mathbb{N}$. For the other $n \in \mathbb{N}$, note that the logarithmic Sobolev inequality verified by ν directly gives us

$$\forall\, n \in \mathbb{N}, \qquad \mathrm{Ent}(m_n|\nu) \;\leq\; \frac{1}{2}I(m_n) \;\leq \frac{K_3}{2}$$

□

Of course, the problem with the hypotheses of Proposition 4.3 is that they are not immediate to verify, since they use the distributions m_n, for $n \in \mathbb{N}$. So we will give below simple sufficient criteria enabling to check them only in terms of the given family $(\mu_n)_{n\in\mathbb{N}}$.

To treat the bounded modified Fisher information condition, let us come back to the "true" Fisher information of a probability m on $\mathbb{R}$ which is defined as the quantity

$$J(m) \;\coloneqq\; \begin{cases} \int \left|\nabla \ln\left(\frac{dm}{d\lambda}\right)\right|^2 dm \;\leq\; +\infty & \text{, if } m \ll \lambda \\ +\infty & \text{, otherwise} \end{cases}$$

where λ is the traditional Lebesgue measure on $\mathbb{R}$. So if m is a probability absolutely continuous with respect to λ, we have

$$\begin{aligned} I(m) &= \int \left|\nabla \ln\left(\frac{dm}{d\lambda}\right) + \nabla \ln\left(\frac{d\lambda}{d\nu}\right)\right|^2 dm \\ &= \int \left|\nabla \ln\left(\frac{dm}{d\lambda}\right)(x) + x\right|^2 m(dx) \\ &= J(m) + 2\int x\nabla \ln\left(\frac{dm}{d\lambda}\right)(x)\, m(dx) + \int x^2\, m(dx) \\ &= J(m) + 2\int x\nabla \frac{dm}{d\lambda}(x)\, \lambda(dx) + \int x^2\, m(dx) \\ &= J(m) - 2\int (\nabla x)\frac{dm}{d\lambda}(x)\, \lambda(dx) + \int x^2\, m(dx) \\ &= J(m) - 2\int m(dx) + \int x^2\, m(dx) \\ &= J(m) - 2 + \int x^2\, m(dx) \end{aligned}$$

(these computations are justified if $\sqrt{dm/d\lambda}$ is $\mathcal{C}^1$ with compact support, but the relation $I(m) = J(m) - 2 + \int x^2\, m(dx)$ can next be extended to the general case by traditional approximation procedures).
The advantage of J is that it is also an "invariant" quantity in the sense of Lemma 3.1, namely we have the following well-known bound of Blachman and Stam (see for instance [9] or [1]):

Theorem 4.5. *Let m be a distribution constructed on $\mathbb{R}$ as in the statement of Lemma 3.1, starting from $0 \leq t \leq 1$ and from two probabilities μ, μ'. Then we have*

$$J(m) \;\leq\; t^2 J(\mu) + (1-t^2)J(\mu')$$

Remark 4.6. Still with the notations of the above result, we have if μ or μ' are centered and have the same variance,

$$\begin{aligned} I(m) &= J(m) - 2 + \int x^2\, m(dx) \\ &\leq t^2 J(\mu) + (1-t^2)J(\mu') + t^2 \int x^2\, \mu(dx) \\ &\quad +(1-t^2)\int x^2\, \mu'(dx) - 2(t^2 + 1 - t^2) \end{aligned}$$

$$
\begin{aligned}
&= t^2 I(\mu) + (1-t^2) I(\mu') \\
&\leq I(\mu) \vee I(\mu')
\end{aligned}
$$

But this relation is no longer necessarily true if μ and μ' are not centered, consider for instance the case where $\mu = \mu'$ is the normalized Gaussian distribution of mean 1.

Thus to ensure that for every $n \in \mathbb{N}$, we have $I(m_n) \leq K_3$, for some constant $K_3 \geq 0$, it is sufficient to ask that for all $n \in \mathbb{N}$, $\mu_n \in \mathcal{N}_2$ and $J(\mu_n) \leq K_3 + 1$. To deal with the belonging of the m_n, $n \in \mathbb{N}$, to some $\mathcal{M}_r^{(1)}(K_1)$, for fixed constants $r, K_1 \geq 0$, we will use another stability property. More precisely, if $r \in \mathbb{N}$ and $M \geq 1$ are given, let $\widetilde{\mathcal{S}}_r(M)$ be the set of probabilities μ on $\mathbb{R}$ which are symmetrical with respect to the origin and such that

$$
\forall\, 0 \leq k \leq r, \qquad \int x^{2k}\, \mu(dx) \quad \leq \quad M^k \int x^{2k}\, \nu(dx)
$$

Lemma 4.7. *For any r and M as above, the set $\widetilde{\mathcal{S}}_r(M)$ is stable by the operations described in Lemma 3.1, in the sense that if m is constructed in this way from $0 \leq t \leq 1$ and $\mu, \mu' \in \widetilde{\mathcal{S}}_r(M)$, then we also have $m \in \widetilde{\mathcal{S}}_r(M)$.*

Proof. The preservation of the symmetry with respect to zero is clear, so we just compute that for $0 \leq k \leq r$,

$$
\begin{aligned}
&\int x^{2k}\, m(dx) \\
&= \int \sum_{0 \leq l \leq 2k} \binom{2k}{l} t^l x^l (1-t^2)^{\frac{2k-l}{2}} y^{2k-l}\, \mu(dx)\mu'(dy) \\
&= \sum_{0 \leq l \leq k} \int \binom{2k}{2l} t^{2l} x^{2l} (1-t^2)^{k-l} y^{2(k-l)}\, \mu(dx)\mu'(dy) \\
&\leq M^k \sum_{0 \leq l \leq k} \int \binom{2k}{2l} t^{2l} x^{2l} (1-t^2)^{k-l} y^{2(k-l)}\, \nu(dx)\nu(dy) \\
&= M^k \int x^{2k}\, \nu(dx)
\end{aligned}
$$

□

Remark 4.8. This result can be slightly improved by requiring not the symmetry of the elements μ of $\widetilde{\mathcal{S}}_r(M)$, but only that for every $1 \leq k \leq r$,

$$
\int x^{2k-1}\, m(dx) \quad = \quad 0
$$

(note that this property is also stable by the operations described in Lemma 3.1.

For two given as above constants $r \in \mathbb{N}$ and $K \geq 1$, let $\mathcal{S}_r(K)$ be the set of symmetrical distributions μ verifying $\int x^{2r}\,\mu(x) \leq K$. Denoting

$$M_r(K) \;\coloneqq\; \sup_{0\leq k\leq r} \left(\frac{K^{k/r}}{\int x^{2k}\,\nu(dx)}\right)^{1/k} \;<\; +\infty$$

we easily check that $\mathcal{S}_r(K) \subset \widetilde{\mathcal{S}}_r(M_r(K))$. Thus if we assume there exist $r \in \mathbb{N}$ and a finite constant $K_4 \geq 0$ such that for all $n \in \mathbb{N}$, $\mu_n \in \mathcal{S}_r(K_4)$, then we are assured that for every $n \in \mathbb{N}$, $m_n \in \widetilde{\mathcal{S}}_r(M_r(K_4)) \subset \mathcal{M}^{(1)}_{2r}(K_2)$, with $K_2 \coloneqq M_r^r(K_4)\int x^{2r}\,\nu(dx)$.
So for instance we get the following result, where hypotheses are only made on the family $(\mu_n)_{n\in\mathbb{N}}$, but which remains not very good in view of Theorem 1.1 (nevertheless we will check below that the hypothesis of spectral gap is not necessary here, nor in Proposition 4.3).

Corollary 4.9. *Still in the context of section 2, assume there exist $r \in \mathbb{N}$ and a finite constant $K \geq 0$ such that for all $n \in \mathbb{N}$,*

$$\begin{aligned} \int x^2\,\mu_n(dx) &= 1 \\ \mu_n &\in \mathcal{S}_{r+4\vee r}(K) \\ \mu_n &\in \mathcal{M}^{(3)}_{3,2r}(K) \\ J(\mu_n) &\leq K \end{aligned}$$

where for any $p \in \mathbb{N}$ and $r, K \geq 0$, $\mathcal{M}^{(3)}_{p,r}(K)$ is the set of probabilities μ on $\mathbb{R}$ verifying the second point of (4.2), then one can find an appropriate constant $C(r,K)$, depending only on its parameters r and K, such that

$$\forall\, n \in \mathbb{N}, \qquad \mathrm{Ent}(m_n|\nu) \;\leq\; \frac{C(r,K)}{\sqrt{n}}$$

Proof. One would have furthermore noticed that the two first above conditions imply that $\mu_n \in \mathcal{N}_3$, so we can apply Proposition 4.3 with $p = 3$. □

To finish, let us present two kinds of examples without spectral gap satisfying the previous conditions. For simplicity we will only consider constant sequences $(\mu_n)_{n\in\mathbb{N}}$ and denote $\mu \coloneqq \mu_0$ that we assume to be symmetrical. This probability will have moments of all orders and for any $p \in \mathbb{N}$, there will exist $r, K \geq 0$ such that $\mu \in \mathcal{M}^{(3)}_{p,r}(K)$. We will also check that there is no obstruction to the conditions that $\int x^2\,\mu(dx) = 1$ and that μ admits a finite Fisher information.

• The first kind of examples is based on remark 3.7 (c). To ensure that μ has no positive spectral gap, we ask for the existence of two constants $0 < M_1 < 1 < M_2$ such that $\mu[(-M_2,-M_1) \sqcup (M_1,M_2)] = 1$ (indeed, considering a smooth function f on $\mathbb{R}$ verifying $f \equiv 0$ on $(-\infty,0]$ and $f \equiv 1$ on $[M_1,+\infty)$, we get that the spectral gap has to be zero). To obtain a finite Fisher information for μ, we impose that $\mu \ll \lambda$ and that $\sqrt{d\mu/d\lambda}$ is smooth on $\mathbb{R}$ (since one can also write

$J(\mu) = 4\int |\nabla\sqrt{d\mu/d\lambda}|^2\, d\lambda)$, for instance it is sufficient that the density $d\mu/d\lambda$ is positive inside (M_1, M_2) and that in a right neighborhood of M_1 (respectively a left neighborhood of M_2), $d\mu/d\lambda(x)$ is proportional to $\exp(-1/(x-M_1))$ (resp. to $\exp(-1/(M_2-x))$). Then it is quite clear that one find such probabilities μ having furthermore 1 for variance.

• At the opposite of the previous examples, our second type of probabilities μ will have tails heavier that those of exponential distributions and this feature will equally forbid a positive spectral gap. We will nevertheless resort to some weighted Poincaré's inequalities (besides it would be interesting to elaborate more general conditions for the belonging to sets like $\mathcal{M}^{(3)}_{p,r}(K)$, for fixed constants $p, r, K \geq 0$, for instance we are wondering which kind of functional inequalities can serve as criteria).

So let assume there exist two constants $0 < \epsilon < 2$ and $C > 0$ such that the symmetrical probability μ verifies that for any absolutely continuous mapping f on $\mathbb{R}$,

$$\int (f - \mu[f])^2\, d\mu \quad \leq \quad C\int (f'(y))^2(1 + |y|^{2-\epsilon})\, \mu(dy) \tag{4.3}$$

If one considers for test function f the power mapping $\mathbb{R} \ni y \mapsto y^q$, with $q \in \mathbb{N}$, then it appears easily that $\int y^{2q}\, \mu(dy)$ has to be bounded by a quantity depending only on q, ϵ and C, thus μ admits moments of all orders. Then returning to the proof of Lemma 3.5 and in particular to the bound (3.3), where we can directly use a slight variant of the trick mentioned in remark 3.7 (a), $(ty)^2 \leq 2(\sqrt{1+t^2}x - ty)^2 + 2(1+t^2)x^2$, it appears that for any $p \in \mathbb{N}$, μ belongs to $\mathcal{M}^{(3)}_{p,r(p)}(K(\epsilon, C, p))$, with $r(p) \coloneqq 8(p+1)/\epsilon$ and for some appropriate finite constant $K(\epsilon, C, p) > 0$, as usual depending only on ϵ, C and p.

In other respects, the symmetry of μ and Hardy's inequalities (cf [18, 7] or [1]) enable to obtain a simple criterion for the validity of (4.3). More precisely, for fixed $\epsilon > 0$, the best possible constant C in (4.3) satisfies $B/2 \leq C \leq 4B$, with

$$B \quad \coloneqq \quad \sup_{t>0} \int_0^t \frac{1}{(1+|y|^{2-\epsilon})\,\frac{d\mu}{d\lambda}(y)}\, \lambda(dy)\ \mu[[t, +\infty)]$$

(where $d\mu/d\lambda$ is a priori the Radon-Nikodym-Lebesgue derivative of the part of μ which is absolutely continuous with respect to λ). But if we assume that $\mu \ll \lambda$ and that $d\mu/d\lambda(y)$ is proportional to $\exp(-y^\alpha)$ for some given $\alpha > 0$ and for $y > 0$ large enough, then we get that for t large enough, $\mu[[t, +\infty)]$ is proportional to

$$\int_t^{+\infty} \exp(-y^\alpha)\, dy \quad = \quad \frac{1}{\alpha}\int_{t^\alpha}^{+\infty} y^{\frac{1-\alpha}{\alpha}} \exp(-y)\, dy$$

and it is well-known that such an integral is equivalent for large $t > 0$ to the expression $\alpha^{-1}t^{1-\alpha}\exp(-t^\alpha)$. Similar computations show that up to a multiplicative constant, the term $\int_0^t \frac{\exp(-y^\alpha)}{1+|y|^{2-\epsilon}}\, dy$ is equivalent to $t^{\epsilon-\alpha-1}\exp(t^\alpha)$, for large $t > 0$. Thus if $\alpha \geq \epsilon/2$, the inequality (4.3) is satisfied. Of course, if furthermore we impose that $d\mu/d\lambda$ is positive and smooth everywhere, then its above behavior at infinity implies that $J(\mu) < +\infty$. There is also no difficulty in locally deforming this density to ensure that it admits 1 as variance. But if we take $\alpha = \epsilon/2 < 1$, then the tail $\mu[[t, +\infty)]$ does not decrease exponentially fast in large $t > 0$, so the spectral gap has to be null (cf for instance [15]), result which in the above situation can also be deduced by another application of Hardy's inequalities.

Remark 4.10. Let $p \in \mathbb{N} \setminus \{0, 1\}$ be given, it is possible to find examples of the above second type which furthermore belong to $\mathcal{N}_p$, because one has a lot of freedom for the form of the density $d\mu/d\lambda$ on compact subsets. Then performing a convolution with a Gaussian distribution as in Proposition 3.8, we end up with a convergence speed of order $\mathcal{O}(1/n^{(p-1)/2})$. But note that the probabilities thus obtained keep the heavy tail property and thus do not admit a spectral gap. So one can find probabilities μ without spectral gap whose entropic convergence speed is at least of order $\mathcal{O}(1/n^{(p-1)/2})$. This feature leads us to think that the spectral gap assumption in Theorem 1.1 is maybe not very natural.

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*. Société Mathématique de France, Paris, 2000. With a preface by D. Bakry and M. Ledoux.

[2] S. Artstein, K. Ball, F. Barthe, and A. Naor. Entropy growth for sums of independent random variables. Preprint, 2002.

[3] S. Artstein, K. Ball, F. Barthe, and A. Naor. More on entropy production. Preprint in preparation, personal communication, 2002.

[4] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. In P. Bernard, editor, *Lectures on Probability Theory. Ecole d'Eté de Probabilités de Saint-Flour XXII-1992*, Lecture Notes in Mathematics 1581. Springer-Verlag, 1994.

[5] K. Ball, F. Barthe, and A. Naor. Entropy jumps in the presence of a spectral gap. Preprint, 2002.

[6] A. R. Barron. Entropy and the central limit theorem. *Ann. Probab.*, 14(1):336–342, 1986.

[7] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.

[8] L. D. Brown. A proof of the Central Limit Theorem motivated by the Cramér-Rao inequality. In *Statistics and probability: essays in honor of C. R. Rao*, pages 141–148. North-Holland, Amsterdam, 1982.

[9] E. A. Carlen. Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 101(1):194–211, 1991.

[10] P. Del Moral, M. Ledoux, and L. Miclo. About supercontraction properties of Markov kernels. Preprint, Publications du Laboratoire de Statistique et Probabilités, no 2001-01, CNRS and Université Paul Sabatier (Toulouse III), France, 2001.

[11] P. Del Moral and L. Miclo. On convergence of chains with occupational self-interactions. Preprint, to appear in *Royal Society Special Edition on Stochastic Analysis*, 2003.

[12] W. Feller. *An introduction to probability theory and its applications, volume II.* Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, New York, 1966.

[13] L. Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1976.

[14] O. Johnson and A. Barron. Fisher information inequalities and the Central Limit Theorem. Preprint in submission, Statslab Research Report 2001-17, PDF file downloadable at arXiv:math.PR/0111020 v1 2 Nov 2001, 2003.

[15] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXIII*, pages 120–216. Springer, Berlin, 1999.

[16] Ju. V. Linnik. An information-theoretic proof of the central limit theorem with Lindeberg conditions. *Theor. Probability Appl.*, 4:288–299, 1959.

[17] L. Miclo. Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies. In J. Azéma, M. Emery, and M. Yor, editors, *Séminaire de Probabilités XXXI*, Lecture Notes in Mathematics 1655, pages 136–167. Springer-Verlag, Berlin, 1997.

[18] B. Muckenhoupt. Hardy's inequality with weights. *Studia Mathematica*, XLIV:31–38, 1972.

Acknowledgments

This paper was motivated by Franck Barthe's talk given at the conference on Stochastic Inequalities and their Applications (Barcelona, June 17-21, 2002), where he presented some of the results [2, 5, 3], recently obtained with his co-authors. I'm indebted to him for all the supplementary explanations he kindly gave me. I am also very grateful to Michel Ledoux and Paul-Marie Samson for all the discussions we had around this subject and to the referee for complementary informations and references.

Laboratoire de Statistique et Probabilités, UMR 5583,
CNRS et Université Paul Sabatier,
118, route de Narbonne,
31062 Toulouse cedex 4, France,
E-mail address: `miclo@math.ups-tlse.fr`