MOD-ϕ CONVERGENCE, II:
ESTIMATES ON THE SPEED OF CONVERGENCE

VALENTIN FÉRAY, PIERRE-LOÏC MÉLIOT, AND ASHKAN NIKEGHBALI

ABSTRACT. In this paper, we give estimates for the speed of convergence towards a limiting stable law in the recently introduced setting of mod-ϕ convergence. Namely, we define a notion of *zone of control*, closely related to mod-ϕ convergence, and we prove estimates of Berry–Esseen type under this hypothesis. Applications include:
- the winding number of a planar Brownian motion;
- classical approximations of stable laws by compound Poisson laws;
- examples stemming from determinantal point processes (characteristic polynomials of random matrices and zeroes of random analytic functions);
- sums of variables with an underlying dependency graph (for which we recover a result of Rinott, obtained by Stein's method);
- the magnetization in the d-dimensional Ising model;
- and functionals of Markov chains.

CONTENTS

1. INTRODUCTION

1.1. **Mod-ϕ convergence.** Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of real-valued random variables. In many situations, there exists a scale s_n and a limiting law ϕ which is infinitely divisible, such that $(X_n/s_n)_{n \in \mathbb{N}}$ converges in law towards ϕ. For instance, in the classical central limit theorem, if $X_n = \sum_{i=1}^{n} A_i$ is a sum of centered i.i.d. random variables with $\mathbb{E}[(A_1)^2] < \infty$, then

$$s_n = \sqrt{n \, \mathbb{E}[(A_1)^2]}$$

and the limit is the standard Gaussian distribution $\mathcal{N}_{\mathbb{R}}(0, 1)$. In [JKN11] and the subsequent papers [DKN15, FMN16], the notion of mod-ϕ convergence was developed in order to get quantitative estimates on the convergence $\frac{X_n}{s_n} \rightharpoonup \phi$ (throughout the paper, $\rightharpoonup$ denotes convergence in distribution).

Definition 1.1. *Let ϕ be an infinitely divisible probability measure, and $D \subset \mathbb{C}$ be a subset of the complex plane, which we assume to contain 0. We assume that the Laplace transform of ϕ is well defined over D, with Lévy exponent η:*

$$\forall z \in D, \quad \int_{\mathbb{R}} e^{zx} \, \phi(dx) = e^{\eta(z)}.$$

We then say that $(X_n)_{n \in \mathbb{N}}$ converges mod-ϕ over D, with parameters $(t_n)_{n \in \mathbb{N}}$ and limiting function $\psi : D \to \mathbb{C}$, if $t_n \to +\infty$ and if, locally uniformly on D,

$$\lim_{n \to \infty} \mathbb{E}[e^{zX_n}] \, e^{-t_n \eta(z)} = \psi(z).$$

If $D = i\mathbb{R}$, we shall just speak of mod-ϕ convergence; it is then convenient to use the notation

$$\theta_n(\xi) = \mathbb{E}[e^{i\xi X_n}] \, e^{-t_n \eta(i\xi)};$$
$$\theta(\xi) = \psi(i\xi),$$

so that mod-ϕ convergence corresponds to $\lim_{n \to \infty} \theta_n(\xi) = \theta(\xi)$ (uniformly for ξ in compact subsets of $\mathbb{R}$). When nothing is specified, in this paper, we implicitly consider that $D = i\mathbb{R}$. When $D = \mathbb{C}$ we shall speak of *complex* mod-ϕ convergence. In some situations, it is also appropriate to study mod-ϕ convergence on a band $\mathbb{R} \times i[-b, b]$, or $[-c, c] \times i\mathbb{R}$ (see [FMN16, MN15]).

Intuitively, a sequence of random variables $(X_n)_{n \in \mathbb{N}}$ converges mod-ϕ if it can be seen as a large renormalization of the infinitely divisible law ϕ, plus some residue which is asymptotically encoded in the Fourier or Laplace sense by the limiting function ψ. Then, ϕ will typically be:

i) in the case of lattice-valued distributions, a Poisson law or a compound Poisson law (*cf.* [BKN09, KN10, FMN16, CDMN15]);

    ii) or, a *stable distribution*, for instance a Gaussian law.

In this paper, we shall only be interested in the second case. Background on stable distributions is given at the end of this introduction (Section 1.3). In particular we will see that, if ϕ is a stable distribution, then the mod-ϕ convergence of X_n implies the convergence in distribution of a renormalized version Y_n of X_n to ϕ (Proposition 1.3).

    We believe that mod-ϕ is a kind of universality class behind the central limit theorem (or its stable version) in the following sense. For many sequences $(X_n)_{n\in\mathbb{N}}$ of random variables that are proven to be asymptotically normal (or converging to a stable distribution), it is possible to prove mod-ϕ convergence; we refer to our monograph [FMN16] or Sections 3-5 below for such examples. These estimates on the Laplace transform/characteristic function can then be used to derive in an automatic way some *companion theorems*, refining the central limit theorem. In [FMN16], we discuss in details the question of moderate/large deviation estimates and of finding the *normality zone*.

    In the present paper, we shall be interested in the speed of convergence towards the Gaussian (or more generally the stable) distribution of the appropriate renormalization Y_n of X_n. To obtain sharp bounds on this speed of convergence, we do not work with mod-ϕ convergence, but we introduce the notion of *zone of control* for the renormalized characteristic function $\theta_n(\xi)$. In many examples, such a zone of control can be obtained by essentially the same arguments used to prove mod-ϕ convergence, and in most examples, mod-ϕ convergence actually holds.

1.2. **Results and outline of the paper.** We take as reference law a stable distribution ϕ of index $\alpha \in (0,2]$. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of variables that admits a zone of control (this notion will be defined in Definition 2.2; this is closely related to the mod-ϕ convergence of $(X_n)_{n\in\mathbb{N}}$). As we will see in Proposition 2.3, this implies that some renormalization Y_n of X_n converges in distribution towards ϕ and we are interested in the speed of convergence for this convergence. More precisely, we are interested in upper bounds for the *Kolmogorov distance*

$$d_{\text{Kol}}(Y_n,\phi) = \sup_{a\in\mathbb{R}} \left| \mathbb{P}[Y_n \leq a] - \int_{-\infty}^{a} \phi(dx) \right|.$$

The main theorem of Section 2 (Theorem 2.16) shows that this distance is $\mathcal{O}(t_n^{-\gamma-1/\alpha})$, where γ is a parameter describing how large our zone of control is. We also obtain as intermediate result estimates for

$$\left| \mathbb{E}[f(Y_n)] - \int_{\mathbb{R}} f(x)\,\phi(dx) \right|,$$

where f lies in some specific set of tests functions (Proposition 2.12). A detailed discussion on the method of proof of these bounds can be found at the beginning of Section 2.

    Section 3 gives some examples of application of the theoretical results of Section 2. The first one is a toy example, while the other ones are new to the best of our knowledge.

- We first consider sums of i.i.d. random variables with finite third moment. In this case, the classical Berry–Esseen estimate ensures that

$$d_{\text{Kol}}(Y_n, \mathcal{N}_{\mathbb{R}}(0,1)) \leq \frac{3\,\mathbb{E}[|A_1|^3]}{\sigma^3 \sqrt{n}},$$

see [Ber41] or [Fel71, §XVI.5, Theorem 1]. Our general statement for variables with a zone of convergence gives essentially the same result, only the constant factor is not as good.

- We can extend the Berry–Esseen estimates to the case of independent but non identically distributed random variables. As an example, we look at the number of zeroes Z_r of a random analytic series that fall in a disc of radius r; it has the same law as a series of independent Bernoulli variables of parameters r^{2k}, $k \geq 1$. When the radius r of the disc goes to 1, one has a central limit theorem for Z_r, and the theory of zones of control yields an estimate $O((1-r)^{-1/2})$ on the Kolmogorov distance.

- We then look at the winding number φ_t of a planar Brownian motion starting at 1 (see Section 3.2 for a precise definition). This quantity has been proven to converge in the mod-Cauchy sense in [DKN15], based on the computation of the characteristic function done by Spitzer [Spi58]. The same kind of argument easily yields the existence of a zone of control and our general result applies: when t goes to infinity, after renormalization, φ_t converges in distribution towards a Cauchy law and the Kolmogorov distance in this convergence is $\mathcal{O}((\log t)^{-1})$.

- In the third example, we consider *compound Poisson laws* (see [Sat99, Chapter 1, §4]). These laws appear in the proof of the Lévy–Khintchine formula for infinitely divisible laws (*loc. cit.*, Chapter 2, §8, p. 44–45), and we shall be interested in those that approximate the stable distributions $\phi_{c,\alpha,\beta}$. Again, establishing the existence of a zone of control is straight-forward and our general result shows that the speed of convergence is $\mathcal{O}(n^{-1/\min(\alpha,1)})$ (Proposition 3.1), with an additional log factor if $\alpha = 1$ and $\beta \neq 0$ (thus exhibiting an interesting phase transition phenomenon).

- Ratios of Fourier transforms of probability measures appear naturally in the theory of self-decomposable laws and of the corresponding Ornstein–Uhlenbeck processes. Thus, any self-decomposable law ϕ is the limiting distribution of a Markov process $(U_t)_{t \geq 0}$, and when ϕ is a stable law, one has mod-ϕ convergence of an adequate renormalisation of U_t, *with a constant residue*. This leads to an estimate of the speed of convergence which depends on α, on the speed of $(U_t)_{t \geq 0}$ and on its starting point (Proposition 3.3).

- Finally, logarithms of characteristic polynomials of random matrices in a classical compact Lie group are mod-Gaussian convergent (see for instance [FMN16, Section 7.5]), and one can compute a zone of control for this convergence, which yields an estimate of the speed of convergence $O((\log n)^{-3/2})$. For unitary groups, one recovers [BHNY08, Proposition 5.2]. This example shows how one can force the index v of a zone of control of mod-Gaussian convergence to be equal to 3, see Remark 3.4.

The last two sections concentrate on the case where the reference law is Gaussian ($\alpha = 2$). In this case, we show that a sufficient condition for having a zone of control is to have *uniform bounds on cumulants* (see Definition 4.1 and Lemma 4.2). This is not surprising since such bounds are known to imply (with small additional hypotheses) mod-Gaussian convergence [FMN16, Section 5.1]. Combined with our main result, this gives bounds for the Kolmogorov distance for variables having uniform bounds on cumulants – see Corollary 4.3. Note nevertheless that similar results have been

given previously by Statulevičius [Sta66] (see also Saulis and Statulevičius [SS91]). Our Corollary 4.3 coincides up to a constant factor to one of their result. Our contribution here therefore consists in giving a large variety of non-trivial examples where such bounds on cumulants hold:

- The first family of examples relies on a previous result by the authors [FMN16, Chapter 9] (see Theorem 4.7 here), where bounds on cumulants for sums of variables with an underlying *dependency graph* are given. Let us comment a bit. Though introduced originally in the framework of the probabilistic method [AS08, Chapter 5], dependency graphs have been used to prove central limit theorems on various objects: random graphs [Jan88], random permutations [Bón10], probabilistic geometry [PY05], random character values of the symmetric group [FMN16, Chapter 11]. In the context of Stein's method, we can also obtain bounds for the Kolmogorov distance in these central limit theorems [BR89, Rin94].

  The results of this paper give another approach to obtain bounds for this Kolmogorov distance for sums of bounded variables (see Section 4.2). The bounds obtained are, up to a constant, the same as in [Rin94, Theorem 2.2]. Note that our approach is fundamentally different, since it relies on classical Fourier analysis, while Stein's method is based on a functional equation for the Gaussian distribution. We make these bounds explicit in the case of subgraph counts in Erdös–Rényi random graphs and discuss an extension to sum of unbounded variables.

- The next example is the finite volume magnetization in the Ising model on $\mathbb{Z}^d$. The Ising model is one of the most classical models of statistical mechanics, we refer to [FV16] and references therein for an introduction to this vast topic. The magnetization M_Δ (that is the sum of the spins in Δ) is known to have asymptotically normal fluctuations [New80]. Based on a result of Duneau, Iagolnitzer and Souillard [DIS74], we prove that, if the magnetic field is non-zero or if the temperature is sufficiently large, M_Δ has uniform bounds on cumulants. This implies a bound on the Kolmogorov distance (Proposition 5.8):

$$d_{\mathrm{Kol}} \left(\frac{M_\Delta - \mathbb{E}[M_\Delta]}{\sqrt{\mathrm{Var}(M_\Delta)}} , \mathcal{N}_{\mathbb{R}}(0,1) \right) \leq \frac{K}{\sqrt{|\Delta|}}.$$

  It seems that this result improves on what is known so far. In [Bul96], Bulinskii gave a general bound on the Kolmogorov distance for sums of *associated random variables*, which applied to M_Δ, yields a bound with an additional $(\log |\Delta|)^d$ factor comparing to ours. In a slightly different direction, Goldstein and Wiroonsri [GW16] have recently given a bound of order $\mathcal{O}(|\Delta|^{1/(2d+2)})$ for the L^1-distance (the L^1-distance is another distance on distribution functions, which is a priori incomparable with the Kolmogorov distance; note also that their bound is only proved in the special case where $\Delta = \{-m, -m+1, \dots, m\}^d$).

- The last example considers statistics of the form $S_n = \sum_{t=0}^n f_t(X_t)$, where $(X_t)_{t \geq 0}$ is an ergodic discrete time Markov chain on a finite space state. Again we can prove uniform bounds on cumulants and deduce from it bounds for the Kolmogorov distance (Theorem 5.14). The speed of convergence in the central limit theorem for Markov chains has already been studied by Bolthausen [Bol80] (see

also later contributions of Lezaud [Lez96] and Mann [Man96]). These authors study more generally Markov chains on infinite space state, but focus on the case of a statistics f_t independent of the time. Except for these differences, the bounds obtained are of the same order; however our approach and proofs are again quite different.

It is interesting to note that the proofs of the bounds on cumulants in the last two examples are highly non trivial and share some common structure. Each of these statistics decomposes naturally as a sum. In each case, we give an upper bound for joint cumulants of the summands, which writes as a weighted enumeration of spanning trees. Summing terms to get a bound on the cumulant of the sum is then easy.

To formalize this idea, we introduce in Section 5 the notion of uniform weighted dependency graphs. Both proofs for the bounds on cumulants (for magnetization of the Ising model and functional of Markov chains) are presented in this framework. We hope that this will find further applications in the future.

1.3. **Stable distributions and mod-stable convergence.** Let us recall briefly the classification of stable distributions (see [Sat99, Chapter 3]). Fix $c > 0$ (the *scale* parameter), $\alpha \in (0, 2]$ (the *stability* parameter), and $\beta \in [-1, 1]$ (the *skewness* parameter).

Definition 1.2. *The stable distribution of parameters* (c, α, β) *is the infinitely divisible law* $\phi = \phi_{c,\alpha,\beta}$ *whose Fourier transform*

$$\widehat{\phi}(\xi) = \int_{\mathbb{R}} e^{ix\xi}\,\phi(dx) = e^{\eta(i\xi)}$$

has a Lévy exponent $\eta = \eta_{c,\alpha,\beta}$ *given by* $\eta_{c,\alpha,\beta}(i\xi) = -|c\xi|^\alpha\,(1 - i\beta\,h(\alpha, \xi)\,\mathrm{sgn}(\xi))$*, where*

$$h(\alpha, \xi) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \text{if } \alpha \neq 1, \\ -\frac{2}{\pi}\log|\xi| & \text{if } \alpha = 1 \end{cases}$$

and $\mathrm{sgn}(\xi) = \pm 1$ *is the sign of* ξ.

The most usual stable distributions are:
- the standard Gaussian distribution $\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx$ for $c = \frac{1}{\sqrt{2}}$, $\alpha = 2$ and $\beta = 0$;
- the standard Cauchy distribution $\frac{1}{\pi(1+x^2)}\,dx$ for $c = 1$, $\alpha = 1$ and $\beta = 0$;
- the standard Lévy distribution $\frac{1}{\sqrt{2\pi}}\,\frac{e^{-1/2x}}{x^{3/2}}\,\mathbb{1}_{x\geq 0}\,dx$ for $c = 1$, $\alpha = \frac{1}{2}$ and $\beta = 1$.

We recall that mod-ϕ convergence on an open subset D of $\mathbb{C}$ containing 0 can only occur when the characteristic function of ϕ is analytic around 0. Among stable distributions, only Gaussian laws (which correspond to $\alpha = 2$) satisfy this property. Mod-ϕ convergence on $D = i\mathbb{R}$ can however be considered for any stable distribution ϕ.

Since $|e^{\eta(i\xi)}| = e^{-|c\xi|^\alpha}$ is integrable, any stable law $\phi_{c,\alpha,\beta}$ has a density $m_{c,\alpha,\beta}(x)\,dx$ with respect to the Lebesgue measure. Moreover, the corresponding Lévy exponents have the following scaling property: for any $t > 0$,

$$t\,\eta_{c,\alpha,\beta}\left(\frac{i\xi}{t^{1/\alpha}}\right) = \begin{cases} \eta_{c,\alpha,\beta}(i\xi) & \text{if } \alpha \neq 1, \\ \eta_{c,\alpha,\beta}(i\xi) - \left(\frac{2c\beta}{\pi}\log t\right)i\xi & \text{if } \alpha = 1. \end{cases}$$
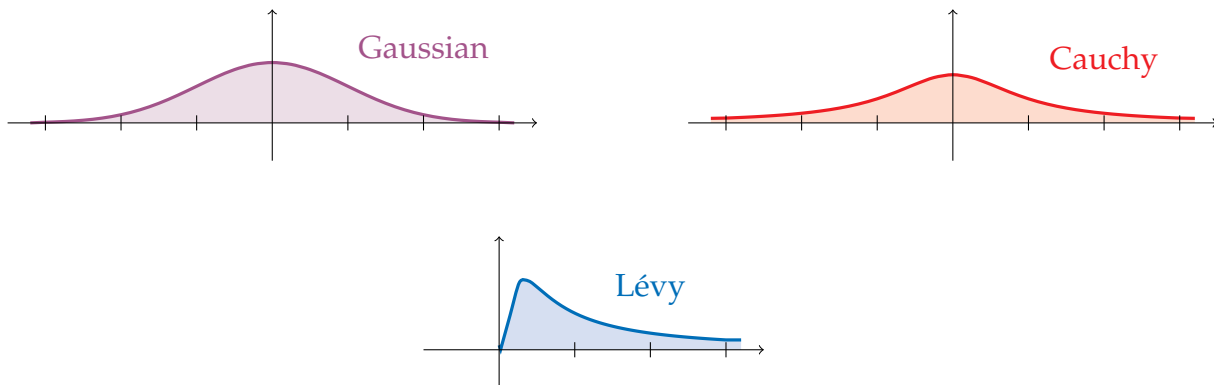
FIGURE 1. Densities of the standard Gaussian, Cauchy and Lévy distribution.

This will be used in the following proposition:

Proposition 1.3. *If $(X_n)_{n\in\mathbb{N}}$ converges in the mod-$\phi_{c,\alpha,\beta}$ sense, then*

$$Y_n = \begin{cases} \frac{X_n}{(t_n)^{1/\alpha}} & \text{if } \alpha \neq 1, \\ \frac{X_n}{t_n} - \frac{2c\beta}{\pi}\log t_n & \text{if } \alpha = 1. \end{cases}$$

converges in law towards $\phi_{c,\alpha,\beta}$.

Proof. In both situations,

$$\mathbb{E}[e^{i\zeta Y_n}] = e^{\eta(i\zeta)}\,\theta_n\left(\frac{\zeta}{(t_n)^{1/\alpha}}\right) = e^{\eta(i\zeta)}\,\theta(0)\,(1+o(1)) = e^{\eta(i\zeta)}\,(1+o(1))$$

thanks to the uniform convergence of θ_n towards θ, and to the scaling property of the Lévy exponent η. $\qquad\square$

2. SPEED OF CONVERGENCE ESTIMATES

The goal of this section is to introduce the notion of zone of control (Section 2.1) and to estimate the speed of convergence in the resulting central limit theorem. More precisely, we take as reference law a stable distribution $\phi_{c,\alpha,\beta}$ and a sequence $(X_n)_{n\in\mathbb{N}}$ that admits a zone of control (with respect to $\phi_{c,\alpha,\beta}$). As for mod-$\phi_{c,\alpha,\beta}$ convergent sequences, it is easy to prove that in this framework, an appropriate renormalization Y_n of X_n converges in distribution towards $\phi_{c,\alpha,\beta}$ (see Proposition 2.3 below).

If Y has distribution $\phi_{c,\alpha,\beta}$, we then want to estimate

$$d_{\mathrm{Kol}}(Y_n, Y) = \sup_{s\in\mathbb{R}}|\mathbb{P}[Y_n \leq s] - \mathbb{P}[Y \leq s]|. \qquad (1)$$

To do this, we follow a strategy proposed by Tao (see [Tao12, Section 2.2]) in the case of sums of i.i.d. random variables with finite third moment. The right-hand side of (1) can be rewritten as

$$\sup_{f\in\mathcal{F}}|\mathbb{E}[f(Y_n)] - \mathbb{E}[f(Y)]|,$$

where $\mathcal{F}$ is the class of measurable functions $y \mapsto 1_{y\leq s}$. Therefore, it is natural to approach the problem of speed of convergence by looking at general estimates on test functions. The basic idea is then to use the Parseval formula to compute the difference

$\mathbb{E}[f(Y_n)] - \mathbb{E}[f(Y)]$, since we have estimates on the Fourier transforms of Y_n and Y. A difficulty comes from the fact that the functions $y \mapsto 1_{y \leq s}$ are not smooth, and in particular, their Fourier transforms are only defined in the sense of distributions. This caveat is dealt with by standard techniques of harmonic analysis (Sections 2.2 to 2.4): namely, we shall work in a space of distributions instead of functions, and use an adequate smoothing kernel in order to be able to work with compactly supported Fourier transforms. Section 2.5 gathers all these tools to give an upper bound for (1). This is the main result of this section and can be found in Theorem 2.16.

Remark 2.1. An alternative way to get an upper bound for (1) from estimates on characteristic functions is to use the following inequality due to Berry (see [Ber41] or [Fel71, Lemma XVI.3.2]). Let X and Y be random variables with characteristic functions $f^*(\zeta)$ and $g^*(\zeta)$. Then, provided that Y has a density bounded by m, we have, for any $s \in \mathbb{R}$,

$$|\mathbb{P}[X \leq s] - \mathbb{P}[Y \leq s]| \leq \frac{1}{\pi} \int_{-T}^{T} \left| \frac{f^*(\zeta) - g^*(\zeta)}{\zeta} \right| d\zeta + \frac{24m}{\pi T}.$$

Using this inequality in our context should lead to similar estimates as the ones we obtain, possibly with different constants. The proof we use here however has the advantage of being more self-contained, and to provide estimates for test functions as intermediate results.

2.1. **The notion of zone of control.**

Definition 2.2. *Let* $(X_n)_{n \in \mathbb{N}}$ *be a sequence of real random variables,* $\phi_{c,\alpha,\beta}$ *a reference stable law, and* $(t_n)_{n \in \mathbb{N}}$ *a sequence growing to infinity. Consider the following assertions:*

(Z1) *Fix* $v > 0$, $w > 0$ *and* $\gamma \in \mathbb{R}$. *There exists a zone* $[-K(t_n)^\gamma, K(t_n)^\gamma]$ *such that, for all* ξ *in this zone, if* $\theta_n(\xi) = \mathbb{E}[e^{i\xi X_n}] e^{-t_n \eta_{c,\alpha,\beta}(i\xi)}$, *then*

$$|\theta_n(\xi) - 1| \leq K_1 |\xi|^v \exp(K_2 |\xi|^w)$$

for some positive constants K_1 *and* K_2 *that are independent of n.*

(Z2) *One has*

$$\alpha \leq w \quad ; \quad -\frac{1}{\alpha} \leq \gamma \leq \frac{1}{w-\alpha} \quad ; \quad 0 < K \leq \left(\frac{c^\alpha}{2K_2} \right)^{\frac{1}{w-\alpha}}.$$

Notice that (Z2) can always be forced by increasing w, and then decreasing K and γ in the bounds of Condition (Z1). If Conditions (Z1) and (Z2) are satisfied, then we say that we have a zone of control $[-K(t_n)^\gamma, K(t_n)^\gamma]$ *with* index (v, w).

Note that although the definition of zone of control depends on the reference law $\phi_{c,\alpha,\beta}$, the latter does not appear in the terminology (throughout the paper, it is considered fixed).

Proposition 2.3. *Let* $(X_n)_{n \in \mathbb{N}}$ *be a sequence of random variables,* $\phi_{c,\alpha,\beta}$ *a reference stable law,* Y *with distribution* $\phi_{c,\alpha,\beta}$ *and* Y_n *as in Proposition 1.3. Assume that* $(X_n)_{n \in \mathbb{N}}$ *has a zone of control* $[-K(t_n)^\gamma, K(t_n)^\gamma]$ *with index* (v, w). *If* $\gamma > -\frac{1}{\alpha}$, *then one has the convergence in law* $Y_n \rightharpoonup Y$.

Proof. Condition (Z1) implies that, if Y_n is the renormalization of X_n and $Y \sim \phi_{c,\alpha,\beta}$, then for fixed ξ,

$$\left| \frac{\mathbb{E}[e^{i\xi Y_n}]}{\mathbb{E}[e^{i\xi Y}]} - 1 \right| = \left| \theta_n \left(\frac{\xi}{(t_n)^{1/\alpha}} \right) - 1 \right| \leq \frac{K_1 |\xi|^v}{(t_n)^{v/\alpha}} \exp \left(\frac{K_2 |\xi|^w}{(t_n)^{w/\alpha}} \right) \quad \text{for } t_n \text{ large enough,}$$

and the right-hand side goes to 0. This proves the convergence in law $Y_n \rightharpoonup Y$. □

The goal of the next few sections will be to get some speed of convergence estimates for this convergence in distribution.

Remark 2.4. In the definition of zone of control, we do not assume the mod-$\phi_{c,\alpha,\beta}$ convergence of $(X_n)_{n \in \mathbb{N}}$ with parameters $(t_n)_{n \in \mathbb{N}}$ and limit $\lim_{n \to \infty} \theta_n(\xi) = \theta(\xi)$. However, in almost all the examples considered, we shall indeed have (complex) mod-ϕ convergence (convergence of the residues θ_n), with the same parameters t_n as for the notion of zone of control. We shall then speak of *mod-ϕ convergence with a zone of convergence* $[-K(t_n)^\gamma, K(t_n)^\gamma]$ *and with index of control* (v, w). Mod-ϕ convergence implies other probabilistic results than estimates of Berry–Esseen type: central limit theorem with a large range of normality, moderate deviations (*cf.* [FMN16]), local limit theorem ([DKN15]), *etc.*

Remark 2.5. If one has mod-$\phi_{c,\alpha,\beta}$ convergence of $(X_n)_{n \in \mathbb{N}}$, then there is at least a zone of convergence $[-K, K]$ of index $(v, w) = (0, 0)$, with $\gamma = 0$; indeed, the residues $\theta_n(\xi)$ stay locally bounded under this hypothesis. Thus, Definition 2.2 is an extension of this statement. However, we allow in the definition the exponent γ to be negative (but not smaller than $-\frac{1}{\alpha}$). Indeed, in the computation of Berry–Esseen type bounds, we shall sometimes need to work with smaller zones than the one given by mod-ϕ convergence, see the hypotheses of Theorem 2.16, and Sections 3.3 and 3.4 for examples.

Remark 2.6. In our definition of zone of control, we ask for a bound on $|\theta_n(\xi) - 1|$ that holds for any $n \in \mathbb{N}$. Of course, if the bound is only valid for $n \geq n_0$ large enough, then the corresponding bound on the Kolmogorov distance (Theorem 2.16) will only hold for $n \geq n_0$.

2.2. **Spaces of test functions.** Until the end of Section 2, all the spaces of functions considered will be spaces of complex valued functions on the real line. If $f \in L^1(\mathbb{R})$, we denote its Fourier transform

$$\widehat{f}(\xi) = \int_{\mathbb{R}} e^{ix\xi} f(x) \, dx.$$

Recall that the Schwartz space $\mathscr{S}(\mathbb{R})$ is by definition the space of infinitely differentiable functions whose derivatives tend to 0 at infinity faster than any power of x. Restricted to $\mathscr{S}(\mathbb{R})$, the Fourier transform is an automorphism, and it satisfies the Parseval formula

$$\forall f, g \in \mathscr{S}(\mathbb{R}), \int_{\mathbb{R}} f(x) \overline{g(x)} \, dx = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(\xi) \overline{\widehat{g}(\xi)} \, d\xi.$$

We refer to [Lan93, Chapter VIII] and [Rud91, Part II] for a proof of this formula, and for the theory of Fourier transforms. The Parseval formula allows to extend by duality and/or density the Fourier transform to other spaces of functions or distributions.

In particular, if $f \in L^2(\mathbb{R})$, then its Fourier transform $\widehat{f}$ is well defined in $L^2(\mathbb{R})$, although in general the integral $\int_{\mathbb{R}} e^{ix\xi} f(x)\, dx$ does not converge; and we have again the Parseval formula

$$\forall f, g \in L^2(\mathbb{R}), \quad \int_{\mathbb{R}} f(x)\, \overline{g(x)}\, dx = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(\xi)\, \overline{\widehat{g}(\xi)}\, d\xi,$$

which amounts to the fact that $f \mapsto \frac{1}{\sqrt{2\pi}} \widehat{f}$ is an isometry of $L^2(\mathbb{R})$ (see [Rud91, §7.9]).

We denote $\mathscr{M}^1(\mathbb{R})$ the set of probability measures on Borel subsets of $\mathbb{R}$. In the sequel, we will need to apply a variant of Parseval's formula, where $\overline{g(x)}\, dx$ is replaced by $\mu(dx)$, with μ in $\mathscr{M}^1(\mathbb{R})$. This is given in the following lemma (see [Str11, Lemma 2.3.3], or [Mal95, p. 134]).

Lemma 2.7. *For any function $f \in L^1(\mathbb{R})$ with $\widehat{f} \in L^1(\mathbb{R})$, and any Borel probability measure $\mu \in \mathscr{M}^1(\mathbb{R})$, the pairing $\langle \mu \mid f \rangle = \int_{\mathbb{R}} f(x)\, \mu(dx)$ is well defined, and the Parseval formula holds:*

$$\int_{\mathbb{R}} f(x)\, \mu(dx) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(\xi)\, \widehat{\mu}(-\xi)\, d\xi,$$

where $\widehat{\mu}(\xi) = \int_{\mathbb{R}} e^{i\xi x}\, \mu(dx)$. The formula also holds for finite signed measures.

Let us now introduce two adequate spaces of test functions, for which we shall be able to prove speed of convergence estimates. We first consider functions $f \in L^1(\mathbb{R})$ with compactly supported Fourier transforms:

Definition 2.8. *We call* smooth test function of order 0, *or simply* smooth test function *an element $f \in L^1(\mathbb{R})$ whose Fourier transform is compactly supported. We denote $\mathscr{T}_0(\mathbb{R})$ the subspace of $L^1(\mathbb{R})$ that consists in smooth test functions; it is an ideal for the convolution product.*

Let us comment a bit this definition. Take f in $\mathscr{T}_0(\mathbb{R})$. Then its Fourier transform $\widehat{f}$ is bounded by $\|f\|_{L^1}$ and vanishes outside an interval $[-C, C]$, so $\widehat{f} \in L^1(\mathbb{R})$. Since f and $\widehat{f}$ are integrable, we can apply Lemma 2.7 with f. Moreover, f is then known to satisfy the Fourier inversion formula (see [Rud91, §7.7]):

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(\xi)\, e^{-i\xi x}\, d\xi.$$

As the integral above is in fact on a compact interval $[-C, C]$, the standard convergence theorems ensure that f is infinitely differentiable in x, hence the term "smooth". Also, by applying the Riemann–Lebesgue lemma to the continuous compactly supported functions $\xi \mapsto (-i\xi)^k \widehat{f}(\xi)$, one sees that $f(x)$ and all its derivatives $f^{(k)}(x)$ go to 0 as x goes to infinity. To conclude, $\mathscr{T}_0(\mathbb{R})$ is included in the space $\mathscr{C}_0^\infty(\mathbb{R})$ of smooth functions whose derivatives all vanish at infinity.

Actually, we will need to work with more general test functions, defined by using the theory of tempered distributions. We endow the Schwartz space $\mathscr{S}(\mathbb{R})$ of smooth

rapidly decreasing functions with its usual topology of metrizable locally convex topological vector space, defined by the family of semi-norms

$$\|f\|_{k,l} = \sum_{a \leq k} \sum_{b \leq l} \sup_{x \in \mathbb{R}} |x^a \, (\partial^b f)(x)|.$$

We recall that a tempered distribution ψ is a continuous linear form $\psi : \mathscr{S}(\mathbb{R}) \to \mathbb{C}$. The value of a tempered distribution ψ on a smooth function f will be denoted $\psi(f)$ or $\langle \psi \mid f \rangle$. The space of all tempered distributions is classically denoted $\mathscr{S}'(\mathbb{R})$, and it is endowed with the $*$-weak topology. The spaces of integrable functions, of square integrable functions and of probability measures can all be embedded in the space $\mathscr{S}'(\mathbb{R})$ as follows: if f is a function in $L^1(\mathbb{R}) \cup L^2(\mathbb{R})$, or if μ is in $\mathscr{M}^1(\mathbb{R})$, then we associate to them the distributions

$$\langle f \mid g \rangle = \int_{\mathbb{R}} f(x)g(x) \, dx \qquad ; \qquad \langle \mu \mid g \rangle = \int_{\mathbb{R}} g(x) \, \mu(dx).$$

We then say that these distributions are represented by the function f and by the measure μ.

The Fourier transform of a tempered distribution ψ is defined by duality: it is the unique tempered distribution $\widehat{\psi}$ such that

$$\langle \widehat{\psi} \mid f \rangle = \left\langle \psi \mid \widehat{f} \right\rangle$$

for any $f \in \mathscr{S}(\mathbb{R})$. This definition agrees with the previous definitions of Fourier transforms for integrable functions, square integrable functions, or probability measures (all these elements can be paired with Schwartz functions). Similarly, if ψ is a tempered distribution, then one can also define by duality its derivative: thus, $\partial \psi$ is the unique tempered distribution such that

$$\langle \partial \psi \mid f \rangle = - \langle \psi \mid \partial f \rangle$$

for any $f \in \mathscr{S}(\mathbb{R})$. The definition agrees with the usual one when ψ comes from a derivable function, by the integration by parts formula. On the other hand, Fourier transform and derivation define linear endomorphisms of $\mathscr{S}'(\mathbb{R})$; also note that the Fourier transform is bijective.

Definition 2.9. *A smooth test function of order 1, or smooth test distribution is a distribution $f \in \mathscr{S}'(\mathbb{R})$, such that ∂f is in $\mathscr{T}_0(\mathbb{R})$, that is to say that the distribution ∂f can be represented by an integrable function with compactly supported Fourier transform. We denote $\mathscr{T}_1(\mathbb{R})$ the space of smooth test distributions.*
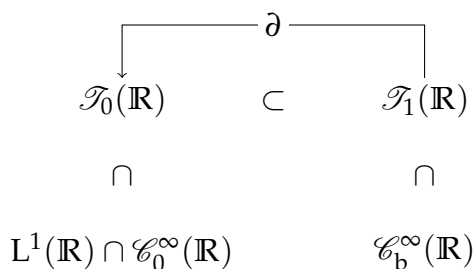

FIGURE 2. The two spaces of test functions $\mathscr{T}_0(\mathbb{R})$ and $\mathscr{T}_1(\mathbb{R})$.

We now discuss Parseval's formula for functions in $\mathscr{T}_1(\mathbb{R})$.

Proposition 2.10. *Any smooth test distribution* $f \in \mathscr{T}_1(\mathbb{R})$ *can be represented by a bounded function in* $\mathscr{C}^\infty(\mathbb{R})$. *Moreover, for any smooth test distribution in* $\mathscr{T}_1(\mathbb{R})$:

(TD1) If μ *is a Borel probability measure, then the pairing* $\langle \mu \mid f \rangle = \int_{\mathbb{R}} f(x) \, \mu(dx)$ *is well defined.*

(TD2) The tempered distribution $\widehat{f}$ *can be paired with the Fourier transform* $\widehat{\mu}$ *of a probability measure with finite first moment, in a way that extends the pairing between* $\mathscr{S}'(\mathbb{R})$ *and* $\mathscr{S}(\mathbb{R})$ *when* μ *(and therefore* $\widehat{\mu}$*) is given by a Schwartz density.*

(TD3) The Parseval formula holds: if $f \in \mathscr{T}_1(\mathbb{R})$ *and* μ *has finite expectation, then*

$$\langle \mu \mid f \rangle = \frac{1}{2\pi} \left\langle \widehat{f} \mid \overline{\widehat{\mu}} \right\rangle.$$

Proof. We start by giving a better description of the tempered distributions f and $\widehat{f}$. Denote $\phi = \partial f$; by assumption, this tempered distribution can be represented by a function $\phi \in \mathscr{T}_0(\mathbb{R})$, which in particular is of class $\mathscr{C}^\infty$ and integrable. Set

$$\widetilde{f}(x) = \int_{y=0}^x \phi(y) \, dy.$$

This is a function of class $\mathscr{C}^\infty$, whose derivative is ϕ, and which is bounded since ϕ is integrable. Therefore, it is a tempered distribution, and for any $g \in \mathscr{S}(\mathbb{R})$,

$$\langle \partial f \mid g \rangle = \langle \phi \mid g \rangle = \left\langle \partial \widetilde{f} \mid g \right\rangle.$$

We conclude that $\partial(f - \widetilde{f})$ is the zero distribution. It is then a standard result that, given a tempered distribution ψ, one has $\partial \psi = 0$ if and only if ψ can be represented a constant. So,

$$f(x) = \int_{y=0}^x \phi(y) \, dy + f(0).$$

This shows in particular that f is a smooth bounded function.

A similar description can be provided for $\widehat{f}$. Recall that the principal value distribution, denoted $\mathrm{pv}(\frac{1}{x})$, is the tempered distribution defined for any $g \in \mathscr{S}(\mathbb{R})$ by

$$\left\langle \mathrm{pv}\left(\frac{1}{x}\right) \mid g \right\rangle = \lim_{\varepsilon \to 0} \left(\int_{|x| \geq \varepsilon} \frac{g(x)}{x} \, dx \right).$$

The existence of the limit is easily proved by making a Taylor expansion of g around 0. Denote

$$\mathscr{S}^{[1]} = \{ g \in \mathscr{S}(\mathbb{R}) \mid g(x) = x \, h(x) \text{ with } h \in \mathscr{S}(\mathbb{R}) \};$$
$$\mathscr{S}_{[1]} = \{ g \in \mathscr{S}(\mathbb{R}) \mid g = \partial h \text{ with } h \in \mathscr{S}(\mathbb{R}) \};$$

the Fourier transform establishes an homeomorphism between $\mathscr{S}^{[1]}$ and $\mathscr{S}_{[1]}$, and the restriction of $\mathrm{pv}(\frac{1}{x})$ to $\mathscr{S}^{[1]}$ is

$$\left\langle \mathrm{pv}\left(\frac{1}{x}\right) \mid g \right\rangle = \int_{\mathbb{R}} \frac{g(x)}{x} \, dx.$$

Let $\widehat{g}(\xi)$ be an element of $\mathscr{S}^{[1]}$, which we write as $\widehat{g}(\xi) = (-\mathrm{i}\xi)\,\widehat{h}(\xi)$ for some $h \in \mathscr{S}(\mathbb{R})$. This is equivalent to $g(x) = (\partial h)(x)$. Let us denote $g_-(x) = g(-x)$, $h_-(x) = h(-x)$, and $\mathrm{pv}(\frac{\mathrm{i}\widehat{\phi}(\xi)}{\xi})$ the tempered distribution defined by

$$\mathrm{pv}\left(\frac{\mathrm{i}\,\widehat{\phi}(\xi)}{\xi}\right) = \mathrm{i}\,\mathrm{pv}\left(\frac{1}{\xi}\right) \circ m_{\widehat{\phi}},$$

with $m_{\widehat{\phi}} : \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\mathbb{R})$ equal to the multiplication by $\widehat{\phi}$. Then we can make the following computation:

$$\left\langle \widehat{f} \,\middle|\, \widehat{g} \right\rangle = \left\langle f \,\middle|\, \widehat{\widehat{g}} \right\rangle = 2\pi \left\langle f \mid g_- \right\rangle = -2\pi \left\langle f \mid \partial h_- \right\rangle = 2\pi \left\langle \phi \mid h_- \right\rangle = \left\langle \phi \,\middle|\, \widehat{\widehat{h}} \right\rangle$$

$$= \left\langle \widehat{\phi} \,\middle|\, \widehat{h} \right\rangle = \left\langle \widehat{\phi} \,\middle|\, \frac{\mathrm{i}\widehat{g}(\xi)}{\xi} \right\rangle = \left\langle \mathrm{pv}\left(\frac{\mathrm{i}\,\widehat{\phi}(\xi)}{\xi}\right) \,\middle|\, \widehat{g} \right\rangle.$$

Thus, the tempered distributions $\widehat{f}$ and $\mathrm{pv}(\frac{\mathrm{i}\widehat{\phi}(\xi)}{\xi})$ agree on the codimension 1 subspace $\mathscr{S}^{[1]}$ of $\mathscr{S}(\mathbb{R})$. However, $\mathscr{S}^{[1]}$ is also the space of functions in $\mathscr{S}(\mathbb{R})$ that vanish at 0, so, if g_0 is any function in $\mathscr{S}(\mathbb{R})$ such that $g_0(0) = 1$, then for $g \in \mathscr{S}(\mathbb{R})$,

$$\left\langle \widehat{f} \,\middle|\, g \right\rangle = \left\langle \widehat{f} \,\middle|\, g - g(0)g_0 \right\rangle + g(0) \left\langle \widehat{f} \,\middle|\, g_0 \right\rangle$$

$$= \left\langle \mathrm{pv}\left(\frac{\mathrm{i}\widehat{\phi}(\xi)}{\xi}\right) \,\middle|\, g - g(0)g_0 \right\rangle + \left\langle \widehat{f} \,\middle|\, g_0 \right\rangle \langle \delta_0 \mid g \rangle$$

$$= \left\langle \mathrm{pv}\left(\frac{\mathrm{i}\widehat{\phi}(\xi)}{\xi}\right) \,\middle|\, g \right\rangle + \left(\left\langle \widehat{f} - \mathrm{pv}\left(\frac{\mathrm{i}\widehat{\phi}(\xi)}{\xi}\right) \,\middle|\, g_0 \right\rangle \right) \langle \delta_0 \mid g \rangle$$

$$= \left\langle \mathrm{pv}\left(\frac{\mathrm{i}\widehat{\phi}(\xi)}{\xi}\right) + L\,\delta_0 \,\middle|\, g \right\rangle$$

where L is some constant. Thus,

$$\widehat{f}(\xi) = \mathrm{pv}\left(\frac{\mathrm{i}\,\widehat{\phi}(\xi)}{\xi}\right) + L\,\delta_0$$

and a computation against test functions shows that $L = 2\pi f(0) - \mathrm{i}\left\langle \mathrm{pv}\left(\frac{1}{\xi}\right) \,\middle|\, \widehat{\phi} \right\rangle$.

The three parts of the proposition are now easily proven. For (TD1), since $f(x)$ is smooth and bounded, we can indeed consider the convergent integral $\int_{\mathbb{R}} f(x)\,\mu(dx)$. For (TD2), assuming that μ has a finite first moment, $\widehat{\mu}$ is a function of class $\mathscr{C}^1$ and with bounded derivative. The same holds for $\widehat{\phi}\widehat{\mu}$, and therefore, one can define

$$\int_{\mathbb{R}} \widehat{f}(\xi)\widehat{\mu}(-\xi)\,d\xi = \mathrm{i}\left\langle \mathrm{pv}\left(\frac{1}{\xi}\right) \,\middle|\, \widehat{\phi}\,\overline{\mu} \right\rangle + L = \left(\lim_{\varepsilon \to \infty} \int_{|x| \geq \varepsilon} \frac{\mathrm{i}\,\widehat{\phi}(\xi)\,\widehat{\mu}(-\xi)}{\xi}\,d\xi \right) + L.$$

Indeed, if $f \in \mathscr{C}^1(\mathbb{R})$, then $\lim_{\varepsilon \to \infty} \int_{1 \geq |x| \geq \varepsilon} \frac{f(x)}{x}\,dx$ always exists, as can be seen by replacing f by its Taylor approximation at 0. Finally, let us prove the Parseval formula

(TD3). The previous calculations show that

$$\frac{1}{2\pi}\int_{\mathbb{R}}\widehat{f}(\xi)\,\widehat{\mu}(-\xi)\,d\xi = \frac{i}{2\pi}\left\langle \mathrm{pv}\left(\frac{1}{\xi}\right)\,\bigg|\,\widehat{\phi}\,\overline{\widehat{\mu}}-\widehat{\phi}\right\rangle + f(0)$$

$$= \lim_{\varepsilon\to 0}\left(\frac{i}{2\pi}\int_{|\xi|\geq\varepsilon}\widehat{\phi}(\xi)\left(\frac{\widehat{\mu}(-\xi)-1}{\xi}\right)d\xi\right) + f(0)$$

$$= \frac{1}{2\pi}\int_{\mathbb{R}}\widehat{\phi}(\xi)\left(\frac{\widehat{\mu}(-\xi)-1}{-i\xi}\right)d\xi + f(0).$$

Indeed, the function $\xi \mapsto \frac{\widehat{\mu}(-\xi)-1}{-i\xi}$ is continuous on $\mathbb{R}$ and bounded, with value $\frac{\widehat{\mu}'(0)}{i} = \int_{\mathbb{R}}x\,\mu(dx)$ at $\xi = 0$; it can therefore be integrated against the function $\widehat{\phi}$ which is integrable (and even with compact support). On the other hand,

$$\int_{\mathbb{R}}f(x)\,\mu(dx) = \int_{x\in\mathbb{R}}\int_{y=0}^{x}\phi(y)\,dy\,\mu(dx) + f(0)$$

$$= \int_{(x,y)\in\mathbb{R}^2}(1_{x>y>0}-1_{x<y<0})\,\phi(y)\,dy\,\mu(dx) + f(0)$$

$$= \int_{y\in\mathbb{R}}\phi(y)\,F(y)\,dy + f(0),\quad \text{with } F(y)=\mu((y,\infty))-1_{y\leq 0}.$$

One has $\int_{\mathbb{R}}|F(y)|\,dy = \int_{y=0}^{\infty}\mu((y,\infty)) + \int_{y=-\infty}^{0}\mu((-\infty,y)) = \int_{\mathbb{R}}|x|\,\mu(dx)$, which is finite. In the integral above, we can therefore consider $F(y)\,dy$ as a finite signed measure, and the Parseval formula applies by Lemma 2.7. One computes readily

$$\widehat{F}(\xi) = \frac{\widehat{\mu}(\xi)-1}{i\xi},$$

which ends the proof. □

Remark 2.11. The Parseval formula of Proposition 2.10 extends readily to finite signed measures μ such that $\int_{\mathbb{R}}|x|\,|\mu|(dx) < +\infty$. Actually, it is sufficient to have a finite signed measure μ such that

$$\frac{\widehat{\mu}(\xi)-\widehat{\mu}(0)}{\xi^v}$$

is bounded in a vicinity of 0, for some $v > 0$. Then, $(\widehat{\mu}(\xi)-\widehat{\mu}(0))/\xi$ is integrable in a neighborhood of 0. This ensures that the distribution $f(x)$ (respectively, the distribution $\widehat{f}(\xi)$) can be evaluated against the measure $\mu(x)$ (respectively, against $\widehat{\mu}(\xi)$), and then, the proof of Parseval's formula is analogous to the previous arguments.

2.3. **Estimates for test functions.** We now give an estimate of $\mathbb{E}[f_n(Y_n)] - \mathbb{E}[f_n(Y)]$, where $(f_n)_{n\in\mathbb{N}}$ is a sequence of test functions in $\mathscr{T}_0(\mathbb{R})$ or $\mathscr{T}_1(\mathbb{R})$, and $(Y_n)_{n\in\mathbb{N}}$ is a sequence of random variables associated to a sequence $(X_n)_{n\in\mathbb{N}}$ which has a zone of control.

Proposition 2.12. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables, $\phi_{c,\alpha,\beta}$ a reference stable law, Y with law $\phi_{c,\alpha,\beta}$ and Y_n as in Proposition 1.3. We assume that:*

(1) $(X_n)_{n\in\mathbb{N}}$ has a zone of control $[-K(t_n)^\gamma, K(t_n)^\gamma]$ with index (v,w);

(2) $(f_n)_{n\in\mathbb{N}}$ is a sequence of smooth test functions in $\mathscr{T}_0(\mathbb{R})$, such that the support of $\widehat{f_n}$ is included into $[-K(t_n)^{\gamma+1/\alpha}, K(t_n)^{\gamma+1/\alpha}]$.

Then,

$$|\mathbb{E}[f_n(Y_n)] - \mathbb{E}[f_n(Y)]| \le C_0(c,\alpha,v)\, K_1 \frac{\|f_n\|_{L^1}}{(t_n)^{v/\alpha}},$$

where $C_0(c,\alpha,v) = \frac{2^{\frac{v+1}{\alpha}}\,\Gamma((v+1)/\alpha)}{\pi\alpha\,c^{v+1}}$.

If instead of (2) we assume:

(2') $(f_n)_{n\in\mathbb{N}}$ is a sequence of smooth test distributions in $\mathscr{T}_1(\mathbb{R})$, such that the support of $\widehat{f_n}$ is included into $[-K(t_n)^{\gamma+1/\alpha}, K(t_n)^{\gamma+1/\alpha}]$,

then

$$|\mathbb{E}[f_n(Y_n)] - \mathbb{E}[f_n(Y)]| \le C_1(c,\alpha,v)\, K_1 \frac{\|\phi_n\|_{L^1}}{(t_n)^{v/\alpha}},$$

where $C_1(c,\alpha,v) = \frac{2^{v/\alpha}\,\Gamma(v/\alpha)}{\pi\alpha\,c^v}$, and $\phi_n = \partial f_n$ is the derivative of the distribution f_n, which is a test function in $\mathscr{T}_0(\mathbb{R})$, hence integrable.

Proof. Consider first a sequence $(f_n)_{n\in\mathbb{N}}$ of test functions in $\mathscr{T}_0(\mathbb{R})$, which satisfies (2). Using Parseval formula and the zone of control assumption, we have

$$\mathbb{E}[f_n(Y_n)] - \mathbb{E}[f_n(Y)] = \frac{1}{2\pi}\int_{-K(t_n)^{\gamma+1/\alpha}}^{K(t_n)^{\gamma+1/\alpha}} \widehat{f_n}(\xi)\, e^{\eta(-i\xi)}\left(\theta_n\left(-\frac{\xi}{(t_n)^{1/\alpha}}\right) - 1\right)d\xi;$$

$$|\mathbb{E}[f_n(Y_n)] - \mathbb{E}[f_n(Y)]| \le \frac{K_1\|\widehat{f_n}\|_\infty}{2\pi(t_n)^{v/\alpha}}\int_{-K(t_n)^{\gamma+1/\alpha}}^{K(t_n)^{\gamma+1/\alpha}} |\xi|^v\, e^{-|c\xi|^\alpha + K_2\left(\frac{|\xi|}{(t_n)^{1/\alpha}}\right)^w}d\xi.$$

For ξ in $[-K(t_n)^{\gamma+1/\alpha}, K(t_n)^{\gamma+1/\alpha}]$, since $(t_n)^{\gamma-1/(w-\alpha)} \le 1$, the second term in the exponent can be bounded as follows:

$$K_2\left(\frac{|\xi|}{(t_n)^{1/\alpha}}\right)^w = K_2|\xi|^\alpha\left(\frac{|\xi|}{(t_n)^{\frac{1}{\alpha}+\frac{1}{w-\alpha}}}\right)^{w-\alpha} \le K_2|\xi|^\alpha\left(K(t_n)^{\gamma-\frac{1}{w-\alpha}}\right)^{w-\alpha} \le \frac{|c\xi|^\alpha}{2}.$$

This is compensated by the term $-|c\xi|^\alpha$ and, therefore,

$$|\mathbb{E}[f_n(Y_n)] - \mathbb{E}[f_n(Y)]| \le \frac{K_1\|\widehat{f_n}\|_\infty}{2\pi(t_n)^{v/\alpha}}\int_{\mathbb{R}} |\xi|^v e^{-\frac{|c\xi|^\alpha}{2}}d\xi$$

$$\le \frac{2^{\frac{v+1}{\alpha}}K_1}{\pi\alpha\,c^{v+1}(t_n)^{v/\alpha}}\,\Gamma\left(\frac{v+1}{\alpha}\right)\|f_n\|_{L^1}.$$

This ends the proof of the first case. For test distributions $f_n \in \mathscr{T}_1(\mathbb{R})$ which satisfies the condition (2'), let us introduce the signed measure $\mu = \mathbb{P}_{Y_n} - \mathbb{P}_Y$. One has $\widehat{\mu}(0) = 0$, and by hypothesis,

$$\left|\frac{\widehat{\mu}(\xi)}{\xi}\right| \le \frac{K_1\,|\xi|^{v-1}}{(t_n)^{v/\alpha}}\, e^{-|c\xi|^\alpha + K_2\left(\frac{|\xi|}{(t_n)^{1/\alpha}}\right)^w}.$$

Remark 2.11 applies, and thus,

$$|\mathbb{E}[f_n(Y_n)] - \mathbb{E}[f_n(Y)]| = |\langle \mu \mid f_n \rangle| = \frac{1}{2\pi} \left| \left\langle \widehat{f_n} \mid \overline{\widehat{\mu}} \right\rangle \right| = \frac{1}{2\pi} \left| \int_{\mathbb{R}} \widehat{\phi_n}(\xi) \, \frac{\widehat{\mu}(-\xi)}{\xi} \, d\xi \right|.$$

From there, the computations are exacly the same as before, with an index $v - 1$ instead of v.   □

2.4. **Smoothing techniques.** We now explain how to relate the estimates on test functions or distributions to estimates on the Kolmogorov distance. The main tool with respect to this problem is the following:

Lemma 2.13. *There exists a function ρ (called kernel) on $\mathbb{R}$ with the following properties.*

  i) *The kernel ρ is non-negative, with $\int_{\mathbb{R}} \rho(x) \, dx = 1$.*

  ii) *The support of $\widehat{\rho}$ is $[-1, 1]$ (hence, ρ is a test function in $\mathscr{T}_0(\mathbb{R})$).*

  iii) *The functions ρ and $\widehat{\rho}$ are even, and*

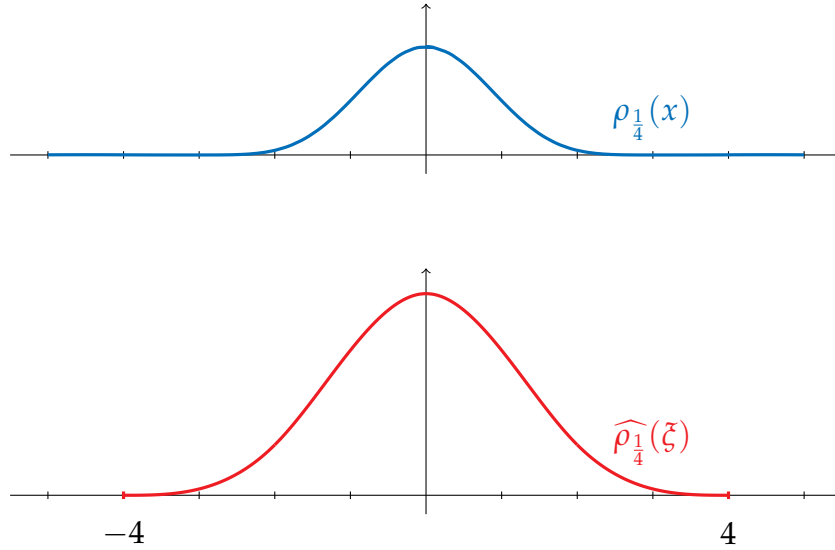$$\rho(K) \leq \min\left(\frac{3}{8\pi}, \frac{96}{\pi \, K^4}\right).$$

FIGURE 3. The smoothing kernel $\rho_{\frac{1}{4}}$, and its Fourier transform which is supported on $[-4, 4]$.

Proof. If $\mathrm{sinc}(x) := \frac{\sin x}{x} = \frac{1}{2} \int_{-1}^{1} e^{ix\xi} \, d\xi$, then by Fourier inversion, $\widehat{\mathrm{sinc}}(\xi) = \pi \, 1_{|\xi| \leq 1}$. As a consequence,

$$\rho(x) = \frac{3}{8\pi} \left(\mathrm{sinc}\left(\frac{x}{4}\right)\right)^4$$

has its Fourier transform supported on $[-\frac{1}{4}, \frac{1}{4}] + [-\frac{1}{4}, \frac{1}{4}] + [-\frac{1}{4}, \frac{1}{4}] + [-\frac{1}{4}, \frac{1}{4}] = [-1, 1]$. On the other hand, an easy computation gives $\int_{\mathbb{R}} \rho(x) \, dx = 1$: use for example the Plancherel formula

$$\int_{\mathbb{R}} |f(x)|^2 \, dx = \frac{1}{2\pi} \int_{\mathbb{R}} |\widehat{f}(\xi)|^2 \, d\xi$$

with $f(x) = \text{sinc}(x)^2$ and thus $\widehat{f}(\xi) = \frac{1}{2\pi}\widehat{\text{sinc}} * \widehat{\text{sinc}}(\xi) = \frac{\pi}{2}(2-|\xi|)_+$. Finally, $\text{sinc}(x) \leq \min(1, \frac{1}{|x|})$, which leads to the inequality stated for $\rho(K)$. $\qquad\square$

In the following, for $\varepsilon > 0$, we set $\rho_\varepsilon(x) = \frac{1}{\varepsilon}\rho(\frac{x}{\varepsilon})$, which has its Fourier transform compactly supported on $[-\frac{1}{\varepsilon}, \frac{1}{\varepsilon}]$; see Figure 3. We also denote $f_{a,\varepsilon}(x) = f_\varepsilon(x-a)$, where f_ε is the function $1_{(-\infty,0]} * \rho_\varepsilon$; cf. Figure 4.
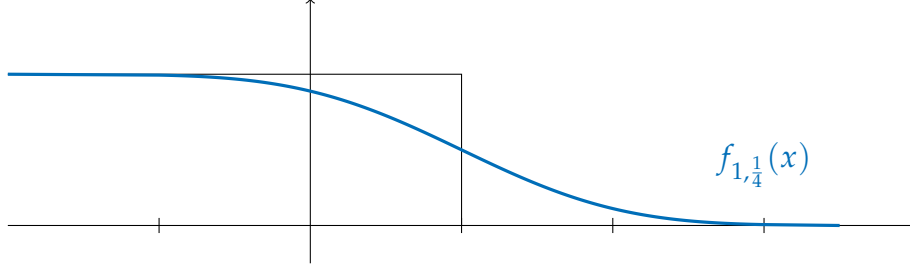

FIGURE 4. The approximation $f_{1,\frac{1}{4}}$ of the Heaviside function $1_{(-\infty,1]}$.

For all a, ε, $f_{a,\varepsilon}$ is an approximation of the Heaviside function $1_{(-\infty,a]}$, and one has the following properties:

Proposition 2.14. *The function $f_{a,\varepsilon}$ is a smooth test distribution in $\mathscr{T}_1(\mathbb{R})$, with Fourier transform compactly supported on $[-\frac{1}{\varepsilon}, \frac{1}{\varepsilon}]$, and $\|\partial f_{a,\varepsilon}\|_{L^1} = 1$. Moreover:*

i) The function $f_{a,\varepsilon}$ has a non-positive derivative, and decreases from 1 to 0.

ii) One has $f_1(x) = 1 - f_1(-x)$, and for all $K \geq 0$,

$$f_1(K) = \int_0^\infty \rho(K+y)\,dy \leq \frac{32}{\pi K^3};$$

$$\int_0^\infty f_1(u-K)\,du \leq K + \int_{w=0}^\infty \min\left(1, \frac{32}{\pi w^3}\right)dw = K + 3\sqrt[3]{\frac{3}{\pi}}.$$

Proof. The derivative of $f_{a,\varepsilon}$ is

$$\partial(f_{a,\varepsilon})(x) = \partial(1_{(-\infty,a]} * \rho_\varepsilon)(x) = \left(\partial(1_{(-\infty,a]}) * \rho_\varepsilon\right)(x)$$
$$= (-\delta_a * \rho_\varepsilon)(x) = -\rho_\varepsilon(x-a),$$

so it is indeed in $\mathscr{T}_0(\mathbb{R})$, and non-positive. Its Fourier transform is supported by $[-\frac{1}{\varepsilon}, \frac{1}{\varepsilon}]$, and $\|\partial f_{a,\varepsilon}\|_{L^1} = \int_\mathbb{R} \rho_\varepsilon(x)\,dx = 1$. Then,

$$\lim_{x\to+\infty} f_{a,\varepsilon}(x) = \lim_{x\to+\infty} f_1\left(\frac{x-a}{\varepsilon}\right) = \lim_{y\to+\infty} f_1(y),$$

so $\lim_{x\to+\infty} f_{a,\varepsilon}(x) = 0$. Since by definition $f_1(y) = \int_y^\infty \rho(u)\,du$, the symmetry relation $f_1(x) = 1 - f_1(-x)$ follows from ρ even; it implies the other limit statement $\lim_{x\to-\infty} f_{a,\varepsilon}(x) = 1$. The inequalities in part ii) are immediate consequences of those of Lemma 2.13. $\qquad\square$

Let us now state a result which converts estimates on smooth test distributions into estimates of Kolmogorov distances. It already appeared in [MN15, Lemma 16], and is inspired by [Tao12, p. 87] and [Fel71, Chapter XVI, §3, Lemma 1]:

Theorem 2.15. *Let X and Y be two random variables with cumulative distribution functions $F_X(a) = \mathbb{P}[X \leq a]$ and $F_Y(a) = \mathbb{P}[Y \leq a]$. Assume that for some $\varepsilon > 0$ and $B > 0$,*

$$\sup_{a \in \mathbb{R}} |\mathbb{E}[f_{a,\varepsilon}(X)] - \mathbb{E}[f_{a,\varepsilon}(Y)]| \leq B\varepsilon.$$

We also suppose that Y has a density w.r.t. Lebesgue measure that is bounded by m. Then, for every $\lambda > 0$,

$$d_{\mathrm{Kol}}(X, Y) = \sup_{a \in \mathbb{R}} |F_X(a) - F_Y(a)| \leq (1 + \lambda) \left(B + \frac{m}{\sqrt[3]{\pi}} \left(4 \sqrt[3]{1 + \frac{1}{\lambda}} + 3 \sqrt[3]{3} \right) \right) \varepsilon.$$

The choice of the parameter λ allows one to optimize constants according to the reference law of Y and to the value of B. A general bound is obtained by choosing $\lambda = \frac{1}{2}$; this gives after some simplifications

$$d_{\mathrm{Kol}}(X, Y) \leq \frac{3}{2}(B + 7m)\,\varepsilon,$$

which is easy to remember and manipulate.

Proof. For the convenience of the reader, we reproduce here the proof given in [MN15]. Fix a positive constant K, and denote $\Delta = \sup_{a \in \mathbb{R}} |F_X(a) - F_Y(a)|$ the Kolmogorov distance between X and Y. One has

$$\begin{aligned} F_X(a) = \mathbb{E}[1_{X \leq a}] &\leq \mathbb{E}[f_{a+K\varepsilon,\varepsilon}(X)] + \mathbb{E}[(1 - f_{a+K\varepsilon,\varepsilon}(X))\,1_{X \leq a}] \\ &\leq \mathbb{E}[f_{a+K\varepsilon,\varepsilon}(Y)] + \mathbb{E}[(1 - f_{a+K\varepsilon,\varepsilon}(X))\,1_{X \leq a}] + B\varepsilon. \end{aligned} \quad (2)$$

The second expectation writes as

$$\begin{aligned} \mathbb{E}[(1 - f_{a+K\varepsilon,\varepsilon}(X))\,1_{X \leq a}] &= \int_{\mathbb{R}} (1 - f_{a+K\varepsilon,\varepsilon}(x))\,1_{(-\infty,a]}(x)\,\mathbb{P}_X(dx) \\ &= -\int_{\mathbb{R}} ((1 - f_{a+K\varepsilon,\varepsilon}(x))\,1_{(-\infty,a]}(x))'\,F_X(x)\,dx = A_1 + A_2, \end{aligned}$$

where $A_1 = (1 - f_{a+K\varepsilon,\varepsilon}(a))\,F_X(a)$ and $A_2 = \int_{\mathbb{R}} f'_{a+K\varepsilon,\varepsilon}(x)\,1_{(-\infty,a]}(x)\,F_X(x)\,dx$. Indeed, in the space of tempered distributions,

$$-((1 - f_{a+K\varepsilon,\varepsilon}(x))\,1_{(-\infty,a]}(x))' = (1 - f_{a+K\varepsilon,\varepsilon}(x))\,\delta_a(x) + f'_{a+K\varepsilon,\varepsilon}(x)\,1_{(-\infty,a]}(x).$$

We evaluate the two terms A_1 and A_2 as follows:

- Since $F_X(a) \leq F_Y(a) + \Delta$,

$$\begin{aligned} A_1 &\leq (1 - f_{a+K\varepsilon,\varepsilon}(a))\,F_Y(a) + (1 - f_{a+K\varepsilon,\varepsilon}(a))\,\Delta \\ &\leq \int_{\mathbb{R}} (1 - f_{a+K\varepsilon,\varepsilon}(x))\,\delta_a(x)\,F_Y(x)\,dx + (1 - f_1(-K))\,\Delta. \end{aligned}$$

- For A_2, since $F_X(x) \geq F_Y(x) - \Delta$ and the derivative of $f_{a+K\varepsilon,\varepsilon}$ is negative, an upper bound is obtained as follows:

$$A_2 \leq \int_{\mathbb{R}} f'_{a+K\varepsilon,\varepsilon}(x)\, 1_{(-\infty,a]}(x)\, F_Y(x)\, dx - \Delta \int_{\mathbb{R}} f'_{a+K\varepsilon,\varepsilon}(x)\, 1_{(-\infty,a]}(x)\, dx$$

$$= \int_{\mathbb{R}} f'_{a+K\varepsilon,\varepsilon}(x)\, 1_{(-\infty,a]}(x)\, F_Y(x)\, dx + (1 - f_{a+K\varepsilon,\varepsilon}(a))\, \Delta$$

$$= \int_{\mathbb{R}} f'_{a+K\varepsilon,\varepsilon}(x)\, 1_{(-\infty,a]}(x)\, F_Y(x)\, dx + (1 - f_1(-K))\, \Delta.$$

Therefore, by gathering the bounds on A_1 and A_2, we get

$$\mathbb{E}[(1 - f_{a+K\varepsilon,\varepsilon}(X))\, 1_{X \leq a}] \leq \mathbb{E}[(1 - f_{a+K\varepsilon,\varepsilon}(Y))\, 1_{Y \leq a}] + 2(1 - f_1(-K))\, \Delta. \tag{3}$$

On the other hand, if m is a bound on the density f_Y of Y, then

$$\mathbb{E}[f_{a+K\varepsilon,\varepsilon}(Y)\, 1_{Y \geq a}] = \int_a^\infty f_{a+K\varepsilon,\varepsilon}(y)\, f_Y(y)\, dy$$

$$\leq m \int_a^\infty f_\varepsilon(y - a - K\varepsilon)\, dy = m \int_0^\infty f_\varepsilon(y - K\varepsilon)\, dy$$

$$\leq m\varepsilon \int_0^\infty f_1(u - K)\, du \leq m\varepsilon \left(K + 3\sqrt[3]{\frac{3}{\pi}} \right);$$

$$\mathbb{E}[f_{a+K\varepsilon,\varepsilon}(Y)] \leq \mathbb{E}[f_{a+K\varepsilon,\varepsilon}(Y)\, 1_{Y \leq a}] + m \left(K + 3\sqrt[3]{\frac{3}{\pi}} \right) \varepsilon. \tag{4}$$

Putting together Eqs. (2), (3) and (4), we get

$$F_X(a) \leq F_Y(a) + \left(B + m \left(K + 3\sqrt[3]{\frac{3}{\pi}} \right) \right) \varepsilon + \frac{64}{\pi K^3}\, \Delta.$$

Similarly, $F_X(a) \geq F_Y(a) - \left(B + m(K + 3\sqrt[3]{\frac{3}{\pi}}) \right) \varepsilon - \frac{64}{\pi K^3} \Delta$, so in the end

$$\Delta = \sup_{a \in \mathbb{R}} |F_X(a) - F_Y(a)| \leq \left(B + m \left(K + 3\sqrt[3]{\frac{3}{\pi}} \right) \right) \varepsilon + \frac{64}{\pi K^3}\, \Delta.$$

As this is true for every K, setting $K = 4\sqrt[3]{\frac{1+\lambda}{\pi\lambda}}$ with $\lambda > 0$, one obtains

$$\Delta \leq (1 + \lambda) \left(B + \frac{m}{\sqrt[3]{\pi}} \left(4\sqrt[3]{1 + \frac{1}{\lambda}} + 3\sqrt[3]{3} \right) \right) \varepsilon. \qquad \square$$

In the next Section 2.5, we shall combine Theorem 2.15 and the estimates on smooth test distributions given by Theorem 2.12 to get a Berry–Esseen type bound on the Kolmogorov distances in the setting of mod-ϕ convergence.

2.5. **Bounds on the Kolmogorov distance.** We are now ready to get an estimate for the Komogorov distance under a zone of control hypothesis.

Theorem 2.16. *Fix a reference stable distribution $\phi_{c,\alpha,\beta}$ and consider a sequence $(X_n)_{n \in \mathbb{N}}$ of random variables with a zone of control $[-K(t_n)^\gamma, K(t_n)^\gamma]$ of index (v, w). Assume in addition that $\gamma \leq \frac{v-1}{\alpha}$. As before, we denote Y a random variable with law $\phi_{c,\alpha,\beta}$, and Y_n the*

renormalization of X_n as in Proposition 1.3. Then, there exists a constant $C(\alpha, c, v, K, K_1)$ such that

$$d_{\mathrm{Kol}}(Y_n, Y) \leq C(\alpha, c, v, K, K_1) \frac{1}{(t_n)^{1/\alpha + \gamma}}.$$

This constant is explicitly given as

$$C(\alpha, c, v, K, K_1) = \min_{\lambda > 0} \left(\frac{1 + \lambda}{\alpha \pi c} \left(\frac{2^{\frac{v}{\alpha}} \Gamma(\frac{v}{\alpha}) K_1}{c^{v-1}} + \frac{\Gamma(\frac{1}{\alpha})}{\sqrt[3]{\pi} K} \left(4 \sqrt[3]{1 + \frac{1}{\lambda}} + 3 \sqrt[3]{3} \right) \right) \right).$$

Note that the additional hypothesis $\gamma \leq \frac{v-1}{\alpha}$ can always be ensured by decreasing γ (but this makes the resulting bound weaker).

Proof. We apply Theorem 2.12 with the smooth test distributions $f_n = f_{a,\varepsilon_n}$, with the value $\varepsilon_n := \frac{1}{K(t_n)^{1/\alpha + \gamma}}$; we know that $\|\partial f_n\|_{\mathrm{L}^1} = 1$ and that the Fourier transform $\widehat{f}_n$ of f_n is supported by the zone $[-K(t_n)^{1/\alpha + \gamma}, K(t_n)^{1/\alpha + \gamma}]$, so that

$$|\mathbb{E}[f_{a,\varepsilon_n}(Y_n)] - \mathbb{E}[f_{a,\varepsilon_n}(Y)]| \leq C_2(c, \alpha, v) \frac{K_1}{(t_n)^{v/\alpha}} \leq \frac{2^{\frac{v}{\alpha}} \Gamma(\frac{v}{\alpha}) K_1 K}{\alpha \pi c^v} \varepsilon_n.$$

This allows to apply Theorem 2.15 with a constant

$$B = \frac{2^{\frac{v}{\alpha}} \Gamma(\frac{v}{\alpha}) K_1 K}{\alpha \pi c^v},$$

and with $\varepsilon = \varepsilon_n = \frac{1}{K(t_n)^{1/\alpha + \gamma}}$. Indeed, note that the density of the law of Y is bounded by

$$m = \frac{1}{2\pi} \|e^{\eta(\mathrm{i}\xi)}\|_{\mathrm{L}^1} = \frac{1}{\alpha \pi c} \Gamma\left(\frac{1}{\alpha}\right). \qquad \square$$

Remark 2.17. Suppose $\alpha = 2$, $c = \frac{1}{\sqrt{2}}$ (mod-Gaussian convergence), and $v = w = 3$. The maximal value allowed for the exponent γ in the size of the zone of control is then $\gamma = 1$, and later we shall encounter many examples of this situation. Then, we obtain

$$d_{\mathrm{Kol}}(Y_n, Y) \leq \frac{1 + \lambda}{\sqrt{2\pi}} \left(2^{\frac{3}{2}} K_1 + \frac{1}{\sqrt[3]{\pi} K} \left(4 \sqrt[3]{1 + \frac{1}{\lambda}} + 3 \sqrt[3]{3} \right) \right) \frac{1}{(t_n)^{3/2}}. \qquad (5)$$

In Section 4, we shall give conditions on cumulants of random variables that lead to mod-Gaussian convergence with a zone of control of size $O(t_n)$ and with index $(3, 3)$, so that (5) holds. We shall then choose K, K_1 and λ to make the constant in the right-hand side as small as possible.

Remark 2.18. In the general case, taking $\lambda = \frac{1}{2}$ in Theorem 2.16 leads to the inequality

$$d_{\mathrm{Kol}}(Y_n, Y) \leq C_3(\alpha, c, v, K_1, K) \frac{1}{(t_n)^{1/\alpha + \gamma}},$$

where $C_3(\alpha, c, v, K_1, K) = \frac{3}{2\pi \alpha c} \left(\frac{2^{\frac{v}{\alpha}} \Gamma(\frac{v}{\alpha}) K_1}{c^{v-1}} + \frac{7 \Gamma(\frac{1}{\alpha})}{K} \right).$

3. Examples with an explicit Fourier transform

3.1. **Sums of independent random variables.** As a direct application of Theorem 2.16, one recovers the classical Berry–Esseen estimates. Let $(A_n)_{n \in \mathbb{N}}$ be a sequence of centered i.i.d. random variables with a third moment. We denote $\mathbb{E}[(A_i)^2] = \sigma^2$ and $\mathbb{E}[|A_i|^3] = \rho$. Set $S_n = \sum_{i=1}^n A_i$, $X_n = S_n/(\sigma n^{1/3})$,

$$t_n = n^{1/3} \quad ; \quad K = \frac{\sigma^3}{\rho} \quad ; \quad v = w = 3 \quad ; \quad \gamma = 1.$$

Notice that $K \le 1$ as a classical application of Hölder inequality. On the zone $\xi \in [-Kn^{1/3}, Kn^{1/3}]$, we have:

$$|\theta_n(\xi) - 1| = \left| \left(\mathbb{E}\left[e^{i\xi \frac{A_1}{\sigma n^{1/3}}} \right] e^{\frac{\xi^2}{2n^{2/3}}} \right)^n - 1 \right|$$

$$\le n \left| \mathbb{E}\left[e^{i\xi \frac{A_1}{\sigma n^{1/3}}} \right] e^{\frac{\xi^2}{2n^{2/3}}} - 1 \right| \left(\max\left(\left| \mathbb{E}\left[e^{i\xi \frac{A_1}{\sigma n^{1/3}}} \right] e^{\frac{\xi^2}{2n^{2/3}}} \right|, 1 \right) \right)^{n-1}.$$

For any t, $\left| e^{it} - 1 - it + \frac{t^2}{2} \right| \le \frac{|t^3|}{6}$, so

$$\left| \mathbb{E}\left[e^{i\xi \frac{A_1}{\sigma n^{1/3}}} \right] e^{\frac{\xi^2}{2n^{2/3}}} - 1 \right| \le \left| \mathbb{E}\left[e^{i\xi \frac{A_1}{\sigma n^{1/3}}} \right] - 1 + \frac{\xi^2}{2n^{2/3}} \right| e^{\frac{\xi^2}{2n^{2/3}}} + \left| e^{-\frac{\xi^2}{2n^{2/3}}} - 1 + \frac{\xi^2}{2n^{2/3}} \right| e^{\frac{\xi^2}{2n^{2/3}}}$$

$$\le \left(\frac{|\xi|^3}{6Kn} + \frac{\xi^4}{8n^{4/3}} \right) e^{\frac{\xi^2}{2n^{2/3}}} \le \frac{7e^{1/2}}{24} \frac{|\xi|^3}{Kn}.$$

For the same reasons,

$$\left| \mathbb{E}\left[e^{i\xi \frac{A_1}{\sigma n^{1/3}}} \right] e^{\frac{\xi^2}{2n^{2/3}}} \right| \le \frac{|\xi|^3}{6Kn} e^{\frac{\xi^2}{2n^{2/3}}} + \left(1 - \frac{\xi^2}{2n^{2/3}} \right) e^{\frac{\xi^2}{2n^{2/3}}}$$

$$\le \frac{|\xi|^3}{6Kn} e^{1/2} + 1 \le e^{\frac{e^{1/2}}{6} \frac{|\xi|^3}{Kn}}$$

We conclude that

$$|\theta_n(\xi) - 1| \le \frac{7e^{1/2}}{24} \frac{|\xi|^3}{K} e^{\frac{e^{1/2}}{6} \frac{|\xi|^3}{K}}$$

on the zone of control $[-Kn^{1/3}, Kn^{1/3}]$. If we want Condition (Z2) to be satisfied, we need to change K and set

$$K = \frac{3}{2e^{1/2}} \frac{\sigma^3}{\rho},$$

which is a little bit smaller than before. We then have a zone of control with constants $K_1 = \frac{7e^{1/2}\rho}{24\,\sigma^3}$ and $K_2 = \frac{e^{1/2}\rho}{6\,\sigma^3}$, and the inequality $K \le \left(\frac{c^\alpha}{2K_2} \right)^{\frac{1}{w-\alpha}}$ is an equality. By Theorem 2.16,

$$d_{\mathrm{Kol}}(Y_n, \mathcal{N}_\mathbb{R}(0,1)) \le \frac{1+\lambda}{\sqrt{2\pi}} \left(\frac{7}{24} 2^{3/2} e^{1/2} + \frac{2e^{1/2}}{3\sqrt[3]{\pi}} \left(4\sqrt[3]{1 + \frac{1}{\lambda}} + 3\sqrt[3]{3} \right) \right) \frac{\rho}{\sigma^3 \sqrt{n}}$$

with $Y_n = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n A_i$. Taking $\lambda = 0.183$, we obtain a bound with a constant $C \le 4.815$, so we recover

$$d_{\mathrm{Kol}}(Y_n, \mathcal{N}_\mathbb{R}(0,1)) \le 4.815 \frac{\rho}{\sigma^3\sqrt{n}},$$

which is almost as good as the statement in the introduction, where a constant $C = 3$ was given (the best constant known today is, as far as we know, $C = 0.4748$, see [KS10]). Of course, the advantage of our method is its large range of applications, as we shall see in the next sections.

Our notion of zone of control allows one to deal with sums of random variables that are independent but not identically distributed. As an example, consider for $r < 1$ a random series

$$Z_r = \sum_{k=1}^{\infty} \mathcal{B}(r^{2k}),$$

with Bernoulli variables of parameters r^{2k} that are independent. The random variable Z_r has the same law as the number of zeroes with module smaller than r of a random analytic series $S(z) = \sum_{n=0}^{\infty} a_n z^n$, where the a_n's are independent standard complex Gaussian variables (see [FMN16, Section 7.1]). If $h = \frac{4\pi r^2}{1-r^2}$ is the hyperbolic area of the disc of radius r and center 0, then we showed in *loc. cit.* that as h goes to infinity and r goes to 1, denoting $Z_r = Z^h$, the sequence

$$X_h = \frac{1}{h^{1/3}} \left(Z^h - \frac{h}{4\pi} \right)$$

is mod-Gaussian convergent with parameters $t_h = \frac{h^{1/3}}{8\pi}$ and limit $\theta(\xi) = \exp(\frac{(\mathrm{i}\xi)^3}{144\pi})$. Let us compute a zone of control for this mod-Gaussian convergence. We change a bit the parameters of the mod-Gaussian convergence and take

$$\widetilde{t}_h = \mathrm{Var}(X_h) = \frac{1}{h^{2/3}} \sum_{k=1}^{\infty} r^{2k}(1 - r^{2k}) = \frac{h^{1/3}(h + 4\pi)}{4\pi(2h + 4\pi)}.$$

The precise reason for this small modification will be given in Remark 3.4. Then,

$$\theta_h(\xi) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\xi X_h}] \, \mathrm{e}^{\frac{\widetilde{t}_h \xi^2}{2}} = \prod_{k=1}^{\infty} \left(1 + r^{2k}(\mathrm{e}^{\frac{\mathrm{i}\xi}{h^{1/3}}} - 1) \right) \mathrm{e}^{-\frac{r^{2k}\mathrm{i}\xi}{h^{1/3}} + \frac{r^{2k}(1-r^{2k})\xi^2}{2h^{2/3}}}.$$

Denote $\theta_{h,k}(\xi)$ the terms of the product on the right-hand side. For $|\xi| \leq \frac{h^{1/3}}{4}$, we are going to compute bounds on $|\theta_{h,k}(\xi)|$ and $|\theta_{h,k}(\xi) - 1|$. The holomorphic function

$$f_\alpha(z) = \log(1 + \alpha(\mathrm{e}^z - 1)) - \alpha z - \frac{\alpha(1 - \alpha) z^2}{2}$$

has its two first derivatives at 0 that vanish, and its third complex derivative is

$$f_\alpha'''(z) = \alpha(1 - \alpha) \, \mathrm{e}^z \, \frac{(1 - \alpha(1 + \mathrm{e}^z))}{(1 + \alpha(\mathrm{e}^z - 1))^3}.$$

If $|\xi| \leq \frac{h^{1/3}}{4}$, then $|\mathrm{e}^{\frac{\mathrm{i}\xi}{h^{1/3}}}| \leq \mathrm{e}^{1/4}$ and $|\mathrm{e}^{\frac{\mathrm{i}\xi}{h^{1/3}}} - 1| \leq \frac{1}{4} \mathrm{e}^{1/4} \leq \frac{1}{2}$, so

$$|\log \theta_{h,k}(\xi)| \leq \frac{|\xi|^3}{6h} \, r^{2k}(1 - r^{2k}) \, \mathrm{e}^{1/4} \, \frac{1 + \frac{r^{2k}}{2}}{(1 - \frac{1}{4}\mathrm{e}^{1/4} \, r^{2k})^3} \leq \frac{|\xi|^3}{4h} \, \frac{\mathrm{e}^{1/4} \, r^{2k}}{(1 - \frac{1}{4}\mathrm{e}^{1/4})^2} \leq \frac{|\xi|^3 r^{2k}}{h}.$$

Therefore,

$$|\theta_{h,k}(\xi)| \leq \exp\left(\frac{|\xi^3| r^{2k}}{h} \right) \qquad ; \qquad |\theta_{h,k}(\xi) - 1| \leq \frac{|\xi^3| r^{2k}}{h} \, \exp\left(\frac{|\xi^3| r^{2k}}{h} \right).$$

We then obtain on the zone $|\xi| \leq \frac{h^{1/3}}{4}$

$$|\theta_h(\xi) - 1| \leq \sum_{k=1}^{\infty} |\theta_{h,k}(\xi) - 1| \prod_{j \neq k} |\theta_{h,j}(\xi)| \leq S \exp S$$

with $S = \sum_{k=1}^{\infty} \frac{|\xi^3| r^{2k}}{h} = \frac{|\xi^3|}{4\pi}$. The inequalities of Condition (Z2) forces us to look at a slightly smaller zone $\xi \in [-\pi \tilde{t}_h, \pi \tilde{t}_h]$; then, this zone of control has index $(3,3)$ and constants $K_1 = K_2 = \frac{1}{4\pi}$. We can then apply Theorem 2.16, and we obtain for h large enough

$$d_{\text{Kol}}\left(\frac{Z^h - \frac{h}{4\pi}}{\sqrt{\text{Var}(Z^h)}}, \mathcal{N}_{\mathbb{R}}(0,1)\right) \leq \frac{C}{\sqrt{h}}$$

with a constant $C \leq 166$.

3.2. Winding number of a planar Brownian motion.

In this section, we consider a standard planar Brownian motion $(Z_t)_{t \in \mathbb{R}_+}$ starting from $z = 1$. It is well known that, a.s., Z_t never touches the origin. One can thus write $Z_t = R_t \, e^{i\varphi_t}$, for continuous functions $t \mapsto R_t$ and $t \mapsto \varphi_t$, where $\varphi_0 = 0$, see Figure 5.
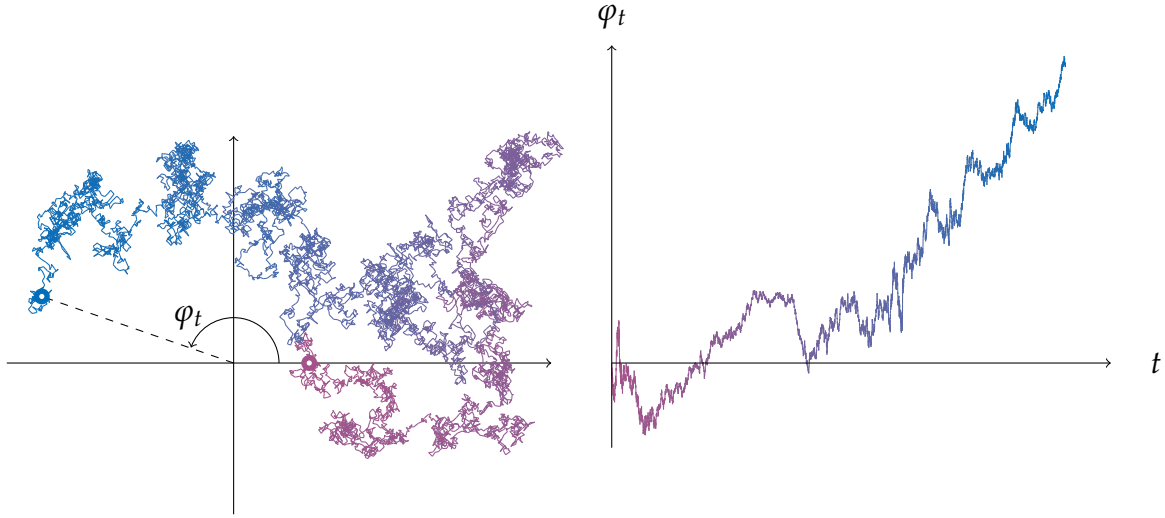

FIGURE 5. Planar Brownian motion and its winding number; we will see that the latter is asymptotically mod-Cauchy.

The Fourier transform of the *winding number* φ_t was computed by Spitzer in [Spi58]:

$$\mathbb{E}[e^{i\xi\varphi_t}] = \sqrt{\frac{\pi}{8t}} \, e^{-\frac{1}{4t}} \left(I_{\frac{|\xi|-1}{2}}\left(\frac{1}{4t}\right) + I_{\frac{|\xi|+1}{2}}\left(\frac{1}{4t}\right) \right),$$

where $I_\nu(z) = \sum_{k=0}^{\infty} \frac{1}{k! \, \Gamma(\nu+k+1)} \left(\frac{z}{2}\right)^{\nu+2k}$ is the modified Bessel function of the first kind. As a consequence, and as was noticed in [DKN15, §3.2], $(\varphi_t)_{t \in \mathbb{R}_+}$ converges mod-Cauchy with parameters $\frac{\log 8t}{2}$ and limiting function $\theta(\xi) = \sqrt{\pi} \, \Gamma\left(\frac{|\xi|+1}{2}\right)^{-1}$.

Indeed,

$$\mathbb{E}[e^{i\xi\varphi_t}]$$

$$= \exp\left(-|\xi|\,\frac{\log 8t}{2}\right)\,\sqrt{\pi}\,e^{-\frac{1}{4t}}\left(\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{1}{8t}\right)^{2k}\left(\frac{1}{\Gamma(k+\frac{|\xi|+1}{2})}+\frac{1}{8t\,\Gamma(k+\frac{|\xi|+3}{2})}\right)\right),$$

and the limit of the power series as t goes to infinity is its constant term $\frac{1}{\Gamma(\frac{|\xi|+1}{2})}$.

Here, $|\theta(\xi)-1|$ is of order $O(|\xi|)$ around 0, since the first derivative of Γ is not zero at $\frac{1}{2}$. Therefore, if the mod-convergence can be given a zone of control, then the index of this control will be $v = 1$, which forces for Berry–Esseen estimates $\gamma \leq 0$ since $\gamma \leq \min(\frac{v-1}{\alpha},\frac{1}{w-\alpha})$. Conversely, for any ξ, notice that the function $x \mapsto \frac{1}{\Gamma(x+\frac{1}{2})}$ has derivative bounded on $\mathbb{R}_+$ by

$$-\frac{\Gamma'(\frac{1}{2})}{\left(\Gamma(\frac{1}{2})\right)^2} = 1.11_- < \frac{2}{\sqrt{\pi}},$$

and therefore that

$$|\theta_t(\xi)-1| \leq 2\,e^{-\frac{1}{4t}}\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{1}{8t}\right)^{2k}\left(1+\frac{1}{8t}\right)\frac{|\xi|}{2} = e^{\left(\frac{1}{8t}\right)^2-\frac{1}{4t}}\left(1+\frac{1}{8t}\right)|\xi| \leq |\xi|$$

for t large enough. So, in particular, one has mod-Cauchy convergence with index of control $(1,1)$, zone of control $[-K, K]$ with K as large as wanted, and constants $K_1 = 1$ and $K_2 = 0$. It follows then from Theorem 2.16 that if $\mathcal{C}$ follows a standard Cauchy law, then

$$d_{\mathrm{Kol}}\left(\frac{2\varphi_t}{\log 8t}, \mathcal{C}\right) \leq \frac{4}{\log 8t}$$

for t large enough. As far as we know, this estimate is new.

3.3. **Approximation of a stable law by compound Poisson laws.** Let $\phi_{c,\alpha,\beta}$ be a stable law; the Lévy–Khintchine formula for its exponent allows one to write

$$\int_{\mathbb{R}} e^{i\xi x}\,\phi_{c,\alpha,\beta}(dx) = \begin{cases} e^{-|c\xi|^2} & \text{if } \alpha = 2, \\ \exp\left(im\xi + \int_{\mathbb{R}}(e^{i\xi x}-1-1_{|x|<1}i\xi x)\,\pi_{c,\alpha,\beta}(dx)\right) & \text{if } \alpha \in (0,2), \end{cases}$$

where $\pi_{c,\alpha,\beta}(dx)$ is the Lévy measure defined for $\alpha \in (0,2)$ by

$$\pi_{c,\alpha,\beta}(dx) = \frac{c_+\,1_{x>0}}{x^{1+\alpha}}+\frac{c_-\,1_{x<0}}{|x|^{1+\alpha}},$$

with $m \in \mathbb{R}$ and $c_+, c_- \in \mathbb{R}_+$ related to (c,α,β) by $\beta = \frac{c_+-c_-}{c_++c_-}$ and

$$m = \begin{cases} \frac{c_+-c_-}{1-\alpha} & \text{if } \alpha \neq 1, \\ \left(\int_0^1 \frac{\sin t-t}{t^2}\,dt+\int_1^{\infty}\frac{\sin t}{t^2}\,dt\right)(c_+-c_-) & \text{if } \alpha = 1; \end{cases}$$

$$c_+ + c_- = \begin{cases} \frac{\alpha\,c^{\alpha}}{\Gamma(1-\alpha)\,\sin\left(\frac{\pi(1-\alpha)}{2}\right)} & \text{if } \alpha \neq 1, \\ \frac{2c}{\pi} & \text{if } \alpha = 1. \end{cases}$$

The proof of the Lévy–Khintchine formula in the general case of an infinitely divisible law involves the following elementary fact (*cf.* [Sat99, Chapter 2]): if μ is infinitely

divisible and $\mu = (\rho_n)^{*n}$ for $n \geq 1$, then the compound Poisson law μ_n of intensity $n\,\rho_n$, which has Fourier transform

$$\widehat{\mu_n}(\xi) = \exp\left(\int_{\mathbb{R}}(e^{ix\xi} - 1)\,n\rho_n(dx)\right),$$

converges in law towards μ; thus, any infinitely divisible law is a limit of compound Poisson laws. In the case of stable laws, this approximation result can be precised in terms of Kolmogorov distances:

Proposition 3.1. *Let Y be a random variable with stable law $\phi_{c,\alpha,\beta}$, and Y_n be a random variable with the following compound Poisson distribution: if μ_n is the law of Y_n, then its Fourier transform is*

$$\widehat{\mu_n}(\xi) = \exp\left(\int_{\mathbb{R}}(e^{ix\xi} - 1)\,n\,\phi_{\frac{c}{n^{1/\alpha}},\alpha,\beta}(dx)\right).$$

The Kolmogorov distance between Y_n and Y is

$$d_{\mathrm{Kol}}(Y_n, Y) \leq \begin{cases} C(\alpha)\,n^{-1/\alpha} & \text{if } \alpha \in (1,2], \\ C(\alpha)\,n^{-1} & \text{if } \alpha \in (0,1) \text{ or } \alpha = 1, \beta = 0, \\ C'\,(\log n)^2\,n^{-1} & \text{if } \alpha = 1, \beta \neq 0, \end{cases}$$

with constants $C(\alpha)$ or C' that depend only on the exponent α.

We thus get a phase transition between the cases $\alpha > 1$ and $\alpha < 1$, with the case $\alpha = 1$ that exhibit distinct transition behaviors according to the value of β.

Proof. Let us distinguish the following cases:

- Suppose first $\alpha \notin \{1, 2\}$. The definition of Y_n implies that

$$\mathbb{E}[e^{i\xi Y_n}] = \widehat{\mu_n}(\xi) = \exp\left(n\left(e^{\frac{\eta_{c,\alpha,\beta}(i\xi)}{n}} - 1\right)\right).$$

Set $X_n = n^{1/(2\alpha)}Y_n$, $t_n = \sqrt{n}$ and $\theta_n(\xi) = \mathbb{E}[e^{i\xi X_n}]\,e^{-t_n\,\eta_{c,\alpha,\xi}(i\xi)}$. We have

$$\theta_n(\xi) = \exp\left(n\left(e^{\frac{\eta_{c,\alpha,\beta}(i\xi)}{n^{1/2}}} - 1 - \frac{\eta_{c,\alpha,\beta}(i\xi)}{n^{1/2}}\right)\right).$$

On the zone $[-K(t_n)^{1/\alpha}, K(t_n)^{1/\alpha}]$ with $K = \frac{|\cos(\frac{\pi\alpha}{2})|^{\frac{2}{\alpha}}}{c}$, we can use a Taylor formula with an integral form remainder:

$$n\left(e^{\frac{\eta_{c,\alpha,\beta}(i\xi)}{n^{1/2}}} - 1 - \frac{\eta_{c,\alpha,\beta}(i\xi)}{n^{1/2}}\right) = (\eta_{c,\alpha,\beta}(i\xi))^2\left(\int_{u=0}^1 (1-u)\,e^{\frac{u\,\eta_{c,\alpha,\beta}(i\xi)}{n^{1/2}}}\,du\right);$$

$$\left|n\left(e^{\frac{\eta_{c,\alpha,\beta}(i\xi)}{n^{1/2}}} - 1 - \frac{\eta_{c,\alpha,\beta}(i\xi)}{n^{1/2}}\right)\right| \leq \frac{1}{2}\,|\eta_{c,\alpha,\beta}(i\xi)|^2 \leq \frac{1}{2}\left(\frac{c^\alpha}{\cos\left(\frac{\pi\alpha}{2}\right)}\right)^2|\xi|^{2\alpha}.$$

We thus obtain a zone of control for $(X_n)_{n\in\mathbb{N}}$ with $v = w = 2\alpha$, $\gamma = \frac{1}{\alpha}$,

$$K_1 = K_2 = \frac{1}{2}\left(\frac{c^\alpha}{\cos\left(\frac{\pi\alpha}{2}\right)}\right)^2,$$

and one checks that

$$\left(\frac{c^\alpha}{2K_2}\right)^{\frac{1}{w-\alpha}} = \frac{|\cos(\frac{\pi\alpha}{2})|^{\frac{2}{\alpha}}}{c} = K.$$

Since we need $\gamma \le \min\left(\frac{1}{w-\alpha}, \frac{v-1}{\alpha}\right)$ to obtain a bound on the Kolmogorov distance (see the hypotheses of Theorem 2.16), this leads to a reduction of γ when $\alpha < 1$:

$$\gamma + \frac{1}{\alpha} = \begin{cases} \frac{2}{\alpha} & \text{if } \alpha > 1, \\ 2 & \text{if } \alpha < 1. \end{cases}$$

Applying Theorem 2.16, we obtain the following upper bound for $d_{\text{Kol}}(Y_n, Y)$:

$$\frac{1+\lambda}{\alpha\pi}\left(\frac{2}{\left(\cos(\frac{\pi\alpha}{2})\right)^2} + \frac{\Gamma(\frac{1}{\alpha})}{\sqrt[3]{\pi}\,|\cos(\frac{\pi\alpha}{2})|^{2/\alpha}}\left(4\sqrt[3]{1+\frac{1}{\lambda}} + 3\sqrt[3]{3}\right)\right)\frac{1}{n^{\frac{\gamma}{2}+\frac{1}{2\alpha}}},$$

Then, any choice of $\lambda > 0$ gives a constant $C(\alpha)$ that depends only on α.

- When $\alpha = 2$, the result follows from the usual Berry–Esseen estimates, since $\sqrt{n}\,Y_n$ has the law of a sum of n independent random variables with same law and finite variance and third moment.

- If $\alpha = 1$ and $\beta = 0$, then the same computations as above can be performed with a constant $K = \frac{1}{c}$, $v = w = 2$, $\gamma = 1$,

$$K_1 = K_2 = \frac{c^2}{2},$$

and this leads to

$$d_{\text{Kol}}(Y_n, Y) \le \frac{1+\lambda}{\pi}\left(2 + \frac{1}{\sqrt[3]{\pi}}\left(4\sqrt[3]{1+\frac{1}{\lambda}} + 3\sqrt[3]{3}\right)\right)\frac{1}{n},$$

and a constant $C = 3.04$ when $\lambda = 0.2$.

- Let us finally treat the case $\alpha = 1$, $\beta \ne 0$. Recall that we then have $\eta_{c,\alpha,\beta}(i\xi) = -|c\xi|\left(1 + \frac{2i\beta}{\pi}\text{sgn}(\xi)\log|\xi|\right)$. We choose t_n such that $t_n \log t_n = \sqrt{n}$, and set

$$X_n = t_n\,Y_n + \frac{2c\beta}{\pi}\sqrt{n}.$$

We then have

$$\theta_n(\xi) = \mathbb{E}[e^{i\xi X_n}]\,e^{-t_n\,\eta_{c,\alpha,\beta}(i\xi)}$$

$$= \exp\left(\frac{2c\beta i\xi}{\pi}t_n\log t_n + n\left(e^{\frac{\eta_{c,\alpha,\beta}(it_n\xi)}{n}} - 1\right) - t_n\,\eta_{c,\alpha,\beta}(i\xi)\right)$$

$$= \exp\left(n\left(e^{\frac{\eta_{c,\alpha,\beta}(it_n\xi)}{n}} - 1 - \frac{\eta_{c,\alpha,\beta}(it_n\xi)}{n}\right)\right),$$

and the Taylor formula with integral remainder yields:

$$\left|n\left(e^{\frac{\eta_{c,\alpha,\beta}(it_n\xi)}{n}} - 1 - \frac{\eta_{c,\alpha,\beta}(it_n\xi)}{n}\right)\right| \le \frac{1}{2n}|\eta_{c,\alpha,\beta}(it_n\xi)|^2$$

$$\le \frac{c^2|\xi|^2}{2}\left(\frac{1 + \frac{4}{\pi^2}(\log|t_n\xi|)^2}{(\log t_n)^2}\right).$$

On the zone $[-\frac{t_n}{2c}, \frac{t_n}{2c}]$, we thus have

$$\left| n \left(e^{\frac{\eta_{c,\alpha,\beta}(it_n\xi)}{n}} - 1 - \frac{\eta_{c,\alpha,\beta}(it_n\xi)}{n} \right) \right| \leq \frac{c^2|\xi|^2}{2} \left(\frac{1 + \frac{4}{\pi^2}(2\log t_n - \log 2c)^2}{(\log t_n)^2} \right)$$

$$\leq c^2|\xi|^2 \quad \text{for } t_n \text{ large enough.}$$

So, there is a zone of control with constants $K_1 = K_2 = c^2$, $v = w = 2$ and $\gamma = 1$, and $K = \frac{1}{2c}$. We thus get as before an estimate of $d_{\mathrm{Kol}}(Y_n, Y)$ of order $\mathcal{O}((t_n)^{-2})$, and since $(t_n \log t_n)^2 = n$, $(t_n)^2$ is equivalent to $\frac{4n}{\log^2 n}$. □

3.4. **Convergence of Ornstein–Uhlenbeck processes to stable laws.** Another way to approximate a stable law $\phi_{c,\alpha,\beta}$ is by using the marginales of a random process of Ornstein–Uhlenbeck type. Consider more generally a self-decomposable law ϕ on $\mathbb{R}$, that is an infinitely divisible distribution such that for any $b \in (0,1)$, there exists a probability measure p_b on $\mathbb{R}$ such that

$$\widehat{\phi}(\xi) = \widehat{\phi}(b\xi)\,\widehat{p_b}(\xi); \tag{6}$$

see [Sat99, Chapter 3, Definition 15.1]. In Equation (6), the laws p_b are the marginale laws of certain Markov processes. Fix a Lévy–Khintchine triplet ($l \in \mathbb{R}$, $v^2 \in \mathbb{R}_+$, ρ) with ρ probability measure on $\mathbb{R} \setminus \{0\}$ that integrates $\min(1, |x|^2)$, and consider the Lévy process $(Z_t)_{t\in\mathbb{R}_+}$ associated to this triplet:

$$\mathbb{E}[e^{i\xi Z_t}] = \exp(t\psi(i\xi)) = \exp\left(t \left(il\xi - \frac{v^2\xi^2}{2} + \int_{\mathbb{R}} (e^{i\xi x} - 1 - 1_{|x|<1} i\xi x)\,\rho(dx) \right) \right).$$

The Ornstein–Uhlenbeck process with triplet (l, v^2, ρ), speed v and starting point x is the solution $(U_t)_{t\geq 0}$ of the stochastic differential equation

$$U_t = e^{-vt}x + \int_0^t e^{-v(t-s)}\,dZ_s.$$

The Ornstein–Uhlenbeck process $(U_t)_{t\geq 0}$ can be shown to be a Markov process whose transition kernel $(P_t(x, dy))_{t\geq 0}$ satisfies:

$$\widehat{P_t(x,\cdot)}(\xi) = \int_{\mathbb{R}} e^{i\xi y}\,P_t(x, dy) = \exp\left(i\xi e^{-vt}x + \int_0^t \psi(i\xi e^{-vs})\,ds \right)$$

see [Sat99, Lemma 17.1]. The connection with self-decomposable laws is provided by:

Theorem 3.2 (Sato–Yamazato, 1983). *For any self-decomposable law ϕ and any fixed speed v, there exists a unique Lévy–Khintchine triplet (l, v^2, ρ) with $\int_{|x|\geq 1} \log|x|\,\rho(dx) < +\infty$, such that the associated Ornstein–Uhlenbeck process $(U_t)_{t\geq 0}$ with speed v satisfies:*

$$\forall x \in \mathbb{R},\ P_t(x, \cdot) \rightharpoonup \phi.$$

If $\psi(i\xi)$ is the exponent associated to (l, v^2, ρ), then

$$\widehat{\phi}(\xi) = \exp\left(\int_{s=0}^{\infty} \psi(i\xi e^{-vs})\,ds \right).$$

We refer to [SY83] and [Sat99, Theorem 17.5]. In the setting of Theorem 3.2, one has the relation

$$\widehat{\phi}(\xi) = \widehat{\phi}(e^{-vt}\xi) \left(\widehat{P_t(x,\cdot)}(\xi)\,\widehat{\delta_{-e^{-vt}x}}(\xi) \right),$$

so if $b \in (0,1)$, setting $b = \mathrm{e}^{-vt}$, one recovers p_b as the law of $U_t - \mathrm{e}^{-vt}x$, where $(U_t)_{t\in\mathbb{R}_+}$ is the Ornstein–Uhlenbeck process that converges in distribution to ϕ and that has speed v and starting point x.

Suppose that $\phi = \phi_{c,\alpha,\beta}$ is a stable law. Then, the previous computations can be reinterpreted in the framework of mod-ϕ convergence. We set

$$\theta(\xi) = \frac{\widehat{\delta_x}(\xi)}{\widehat{\phi}(\xi)} = \exp(\mathrm{i}\xi x - \eta(\mathrm{i}\xi)),$$

and

$$X_t = \begin{cases} \mathrm{e}^{vt}U_t & \text{if } \alpha \neq 1, \\ \mathrm{e}^{vt}\left(U_t - \frac{2c\beta vt}{\pi}\right) & \text{if } \alpha = 1. \end{cases}$$

Then,

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\xi X_t}] = \begin{cases} \widehat{\mu}(\mathrm{e}^{vt}\xi)\,\theta(\xi) & \text{if } \alpha \neq 1, \\ \widehat{\mu}(\mathrm{e}^{vt}\xi)\,\mathrm{e}^{-\mathrm{i}\xi \mathrm{e}^{vt}\frac{2c\beta vt}{\pi}}\,\theta(\xi) & \text{if } \alpha = 1, \end{cases}$$
$$= \mathrm{e}^{\mathrm{e}^{\alpha vt}\eta(\mathrm{i}\xi)}\,\theta(\xi),$$

so $(X_t)_{t\geq 0}$ converges mod-ϕ with parameters $\mathrm{e}^{\alpha vt}$, and with limit equal to the residue $\theta(\xi)$. Note that $\theta_t = \theta$ for any $t \geq 0$, so we are in a special situation where the residues are constant (time-independent). Assuming that $x \neq 0$, one has for any $\xi \in \mathbb{R}$

$$|\theta(\xi) - 1| \leq \begin{cases} K_1\,|\xi|\,\exp(K_2\,|\xi|^\alpha) & \text{if } \alpha \in (1,2], \\ K_1\,|\xi|^\alpha\,\exp(K_2\,|\xi|^\alpha) & \text{if } \alpha \in (0,1) \text{ or } \alpha = 1, \beta = 0, \\ K_1\,|\xi|\,\log|\xi|\,\exp(K_2\,|\xi|) & \text{if } \alpha = 1, \beta \neq 0. \end{cases}$$

For the two first cases, the condition $\gamma \leq \min(\frac{1}{w-\alpha}, \frac{v-1}{\alpha})$ in Theorem 2.16 imposes the following choices of γ when computing Berry–Esseen estimates: $\gamma = \frac{\alpha-1}{\alpha}$ when $\alpha \leq 1$, and $\gamma = 0$ when $\alpha \geq 1$. In these cases, one obtains:

$$d_{\mathrm{Kol}}(U_t, \phi_{c,\alpha,\beta}) = \begin{cases} O(\mathrm{e}^{-vt}) & \text{if } \alpha \in (1,2], \\ O(\mathrm{e}^{-\alpha vt}) & \text{if } \alpha \in (0,1) \text{ or } \alpha = 1, \beta = 0. \end{cases}$$

Because of the term $\log|\xi|$, the last case does not exactly fit the framework of zones of control, but it is easy to adapt the proofs and one gets an estimate $O(vt\,\mathrm{e}^{-vt})$. On the other hand, when $x = 0$, the only difference with the previous discussion is the case $\alpha \in (1,2]$, where we obtain

$$|\theta(\xi) - 1| \leq K_1\,|\xi|^\alpha\,\exp(K_2\,|\xi|^\alpha)$$

and by Theorem 2.16, $d_{\mathrm{Kol}}(U_t, \phi_{c,\alpha,\beta}) = O(\mathrm{e}^{-\alpha vt})$, choosing $\gamma = \frac{\alpha-1}{\alpha}$. So, to summarise:

Proposition 3.3. *Let Y be a random variable with stable law $\phi_{c,\alpha,\beta}$, and $(U_t)_{t\geq 0}$ be the corresponding Ornstein–Uhlenbeck process with starting point x and speed v. We have:*

$$d_{\mathrm{Kol}}(U_t, Y) = \begin{cases} O(\mathrm{e}^{-vt}) & \text{if } \alpha \in (1,2], x \neq 0, \\ O(\mathrm{e}^{-\alpha vt}) & \text{if } \alpha \in (0,1) \text{ or } \alpha = 1, \beta = 0 \text{ or } \alpha \in (1,2], x = 0, \\ O(vt\,\mathrm{e}^{-vt}) & \text{if } \alpha = 1, \beta \neq 0, \end{cases}$$

with constants in the $O(\cdot)$ depending only on x and α.

3.5. **Logarithms of characteristic polynomials of random matrices.** In [KN12, Sections 3 and 4] and [FMN16, Section 7.5], the mod-Gaussian convergence of the following random variables was proven:

random matrix M_n	random variable X_n	parameters t_n	residue $\theta(\xi)$
Haar($U(n)$)	$\mathrm{Re}(\log\det(I_n - M_n))$	$\frac{\log n}{2}$	$\frac{(G(1+\frac{i\xi}{2}))^2}{G(1+i\xi)}$
Haar($USp(n)$)	$\log\det(I_{2n} - M_n) - \frac{1}{2}\log\frac{\pi n}{2}$	$\log\frac{n}{2}$	$\frac{G(\frac{3}{2})}{G(\frac{3}{2}+i\xi)}$
Haar($SO(2n)$)	$\log\det(I_{2n} - M_n) - \frac{1}{2}\log\frac{8\pi}{n}$	$\log\frac{n}{2}$	$\frac{G(\frac{1}{2})}{G(\frac{1}{2}+i\xi)}$

Here, G is Barnes' function, which is the unique entire solution of the equations $G(1) = 1$ and $G(z+1) = G(z)\Gamma(z)$. Moreover, the mod-Gaussian convergence holds in fact on an half-plane $H = \{z \in \mathbb{C} \mid \mathrm{Re}(z) > -\alpha\}$. In the sequel, we denote X_n^A, X_n^C and X_n^D the mod-Gaussian convergent random variables, according to the type of the classical group (A for unitary groups, C for compact symplectic groups and D for even orthogonal groups). Before computing zones of control for these variables, let us make the following essential remark:

Remark 3.4. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables that is mod-Gaussian convergent on a domain $D \subset \mathbb{C}$ which contains a neighborhood of 0 (this ensures that θ_n and all its derivatives converge towards those of θ). We denote $(t_n)_{n\in\mathbb{N}}$ the parameters of mod-Gaussian convergence of $(X_n)_{n\in\mathbb{N}}$. Then, without generality, one can assume $\theta_n'(0) = \theta_n''(0) = 0$ and $\theta'(0) = \theta''(0) = 0$. Indeed, set

$$\widetilde{X}_n = X_n + i\theta_n'(0) \qquad ; \qquad \widetilde{t}_n = t_n - \theta_n''(0).$$

We then have

$$\widetilde{\theta}_n(\xi) := \mathbb{E}[e^{i\xi\widetilde{X}_n}]\, e^{\widetilde{t}_n\frac{\xi^2}{2}} = \theta_n(\xi)\, e^{-\theta_n'(0)\xi - \theta_n''(0)\frac{\xi^2}{2}}$$

and this new residue satisfies $\widetilde{\theta}_n'(0) = \widetilde{\theta}_n''(0) = 0$. For the construction of zones of control, it allows us to force $v = 3$, up to a translation of X_n and of the parameter t_n.

In the following, we only treat the case of unitary groups, the two other ones being totally similar (one could also look at the imaginary part of the log-characteristic polynomial). There is an exact formula for the Fourier transform of X_n^A [KS00, Formula (71)]:

$$\mathbb{E}[e^{i\xi X_n^A}] = \prod_{k=1}^{n} \frac{\Gamma(k)\Gamma(k+i\xi)}{\left(\Gamma(k+\frac{i\xi}{2})\right)^2}.$$

We have $\mathbb{E}[X_n^A] = 0$, and

$$\widetilde{t}_n = \mathbb{E}[(X_n^A)^2] = \frac{1}{2}\sum_{k=1}^{n} \frac{\Gamma''(k)}{\Gamma(k)} - \left(\frac{\Gamma'(k)}{\Gamma(k)}\right)^2 = \frac{1}{2}\sum_{k=1}^{n} \psi_1(k),$$

where $\psi_1(z)$ is the trigamma function $\frac{d^2}{dz^2}(\log\Gamma(z))$, and is given on integers by the remainder of the series $\zeta(2)$:

$$\psi_1(k) = \sum_{m=k}^{\infty} \frac{1}{m^2}.$$

Therefore, $\widetilde{t}_n = \frac{1}{2}\sum_{m=1}^{\infty}\frac{\min(n,m)}{m^2} = \frac{1}{2}(\log n + \gamma + 1 + O(n^{-1}))$. So, $(X_n^A)_{n\in\mathbb{N}}$ is mod-Gaussian convergent with parameters $(\widetilde{t}_n)_{n\in\mathbb{N}}$ and limit

$$\widetilde{\theta}(\xi) = \frac{\left(G(1+\frac{\mathrm{i}\xi}{2})\right)^2}{G(1+\mathrm{i}\xi)}\, \mathrm{e}^{\frac{(\gamma+1)\xi^2}{4}},$$

which satisfies $\widetilde{\theta}'(0) = \widetilde{\theta}''(0) = 0$. With these conventions, we can write the residues $\widetilde{\theta}_n(\xi)$ as

$$\widetilde{\theta}_n(\xi) = \left(\prod_{k=1}^n \frac{\Gamma(k)\Gamma(k+\mathrm{i}\xi)}{\left(\Gamma(k+\frac{\mathrm{i}\xi}{2})\right)^2}\right) \mathrm{e}^{\frac{\widetilde{t}_n\xi^2}{2}} = \prod_{k=1}^n\left(\frac{\Gamma(k)\Gamma(k+\mathrm{i}\xi)}{\left(\Gamma(k+\frac{\mathrm{i}\xi}{2})\right)^2}\, \mathrm{e}^{\frac{\psi_1(k)\xi^2}{4}}\right).$$

Denote $\vartheta_k(\xi)$ the terms of the product on the right-hand side; we use a similar strategy as in Section 3.1 for computing a zone of control. The function $\varphi_k(\xi) = \log\vartheta_k(\xi)$ vanishes at 0, has its two first derivatives that also vanish at 0, and therefore writes as

$$\varphi_k(\xi) = \left(\int_0^1 \varphi_k'''(t\xi)\,(1-t)^2\,dt\right)\frac{\xi^3}{2}.$$

The third derivative of $\varphi_k(\xi)$ is given by

$$\varphi_k'''(\xi) = -\mathrm{i}\,\psi_2(k+\mathrm{i}\xi) + \frac{\mathrm{i}}{2}\,\psi_2\left(k+\frac{\mathrm{i}\xi}{2}\right),$$

with $\psi_2(z) = -2\sum_{j=0}^{\infty}\frac{1}{(j+z)^3}$. As a consequence, $\varphi_k'''(\xi)$ is *uniformly bounded* on $\mathbb{R}$ by

$$3\sum_{j=0}^{\infty}\frac{1}{(j+k)^3} \leq \frac{3\,\zeta(3)}{k^2}.$$

Therefore, $|\varphi_k(\xi)| \leq \frac{\zeta(3)\,|\xi|^3}{2k^2}$, $|\vartheta_k(\xi)| \leq \mathrm{e}^{\frac{\zeta(3)\,|\xi|^3}{2k^2}}$ and $|\vartheta_k(\xi)-1| \leq \frac{\zeta(3)\,|\xi|^3}{2k^2}\, \mathrm{e}^{\frac{\zeta(3)\,|\xi|^3}{2k^2}}$. It follows that for any n and any $\xi\in\mathbb{R}$,

$$|\widetilde{\theta}_n(\xi)-1| \leq S\exp S$$

with $S = \sum_{k=1}^{\infty}\frac{\zeta(3)\,|\xi|^3}{2k^2} = \frac{3\,\zeta(3)\,|\xi|^3}{\pi^2}$. Set $K_1 = K_2 = \frac{3\,\zeta(3)}{\pi^2}$, and $K = \frac{1}{4K_2} = \frac{\pi^2}{12\,\zeta(3)}$. We have a zone of control $[-K t_n, K t_n]$ of index $(3,3)$, with constants K_1 and K_2. We conclude with Theorem 2.16:

Proposition 3.5. *Let M_n be a random unitary matrix taken according to the Haar measure. For n large enough,*

$$d_{\mathrm{Kol}}\left(\frac{\mathrm{Re}(\log\det(I_n - M_n))}{\sqrt{\mathrm{Var}(\mathrm{Re}(\log\det(I_n - M_n)))}}, \mathcal{N}_{\mathbb{R}}(0,1)\right) \leq \frac{C}{(\log n)^{3/2}}$$

with a constant $C \leq 18$. Up to a change of the constant, the same result holds if one replaces $\mathrm{Re}(\log\det(I_n - M_n))$ by $\mathrm{Im}(\log\det(I_n - M_n))$, or by

$$\log\det(I_{2n} - P_n) - \mathbb{E}[\det(I_{2n} - P_n)],$$

with P_n Haar distributed in the unitary compact symplectic group $\mathrm{USp}(n)$ or in the even special orthogonal group $\mathrm{SO}(2n)$.

4. Cumulants and dependency graphs

4.1. **Cumulants, zone of control and Kolmogorov bound.** In this section, we will see that appropriate bounds on the cumulants of a sequence of random variables $(S_n)_{n \in \mathbb{N}}$ imply the existence of a large zone of control for a renormalized version of S_n. In this whole section and in the next one, the reference stable law is the *standard Gaussian law*. We also assume that the random variables S_n are centered.

Let us first recall the definition of cumulants. If X is a real-valued random variable with exponential generating function

$$\mathbb{E}[e^{zX}] = \sum_{r=0}^{\infty} \frac{\mathbb{E}[X^r]}{r!} z^r$$

that is convergent in a neighborhood of 0, then its *cumulants* $\kappa^{(1)}(X), \kappa^{(2)}(X), \dots$ are the coefficients of the series

$$\log \mathbb{E}[e^{zX}] = \sum_{r=1}^{\infty} \frac{\kappa^{(r)}(X)}{r!} z^r,$$

which is also well defined for z in a neighborhood of 0 (see for instance [LS59]). For example, $\kappa^{(1)}(X) = \mathbb{E}[X]$, $\kappa^{(2)}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathrm{Var}(X)$, and

$$\kappa^{(3)}(X) = \mathbb{E}[X^3] - 3\,\mathbb{E}[X^2]\,\mathbb{E}[X] + 2\,\mathbb{E}[X]^3.$$

We are interested in the case where cumulants can be bounded in an appropriate way.

Definition 4.1. *Let $(S_n)_{n \in \mathbb{N}}$ be a sequence of (centered) real-valued random variables. We say that $(S_n)_{n \in \mathbb{N}}$ admits uniform bounds on cumulants with parameters (D_n, N_n, A) if, for any $r \geq 2$, we have*

$$|\kappa^{(r)}(S_n)| \leq N_n\, r^{r-2}\, (2D_n)^{r-1}\, A^r.$$

In the following, it will be convenient to set $(\widetilde{\sigma}_n)^2 = \frac{\mathrm{Var}(S_n)}{N_n D_n}$. The inequality of Definition 4.1 with $r = 2$ gives $(\widetilde{\sigma}_n)^2 \leq 2A^2$.

Lemma 4.2. *Let $(S_n)_{n \in \mathbb{N}}$ be a sequence with uniform bounds on cumulants with parameters (D_n, N_n, A). Set*

$$X_n = \frac{S_n}{(N_n)^{1/3}\,(D_n)^{2/3}} \quad and \quad t_n = (\widetilde{\sigma}_n)^2 \left(\frac{N_n}{D_n}\right)^{1/3} = \mathrm{Var}(X_n).$$

Then, we have for $(X_n)_{n \in \mathbb{N}}$ a zone of control $[-K\,t_n\,,\,K\,t_n]$ of index $(3, 3)$, with the following constants:

$$K = \frac{1}{(8 + 4\mathrm{e})\,A^3}, \qquad K_1 = K_2 = (2 + \mathrm{e})\,A^3.$$

Proof. From the definition of cumulants, since X_n is centered and has variance t_n, we can write

$$\theta_n(\xi) = \mathbb{E}[e^{i\xi X_n}]\, \exp\left(\frac{t_n\,\xi^2}{2}\right) = \exp\left(\sum_{r=3}^{\infty} \frac{\kappa^{(r)}(S_n)}{r!}\, \frac{(i\xi)^r}{(N_n(D_n)^2)^{r/3}}\right) = \exp(z),$$

with

$$|z| \leq \frac{1}{2} \frac{N_n}{D_n} \sum_{r=3}^{\infty} \frac{r^{r-2}}{e^r \, r!} \left(\left(\frac{D_n}{N_n} \right)^{1/3} 2eA \, |\xi| \right)^r .$$

We set $y = \left(\frac{D_n}{N_n} \right)^{1/3} 2eA \, |\xi|$ and suppose that $y \leq 1$, that is to say that ξ is in the zone $\left[-L \left(\frac{N_n}{D_n} \right)^{1/3}, L \left(\frac{N_n}{D_n} \right)^{1/3} \right]$ with $L = \frac{1}{2eA}$.

By Stirling's bound, the series $S(y) = \sum_{r=3}^{\infty} \frac{r^{r-2}}{r! \, e^r} y^r$ is convergent for any $y \in [0,1]$, and we have the inequality $S(y) \leq \frac{y^3}{2e^3 \, (1-y)}$, which implies

$$|z| \leq 2 \frac{(A \, |\xi|)^3}{1 - \left(\frac{D_n}{N_n} \right)^{1/3} 2eA \, |\xi|}.$$

We now consider the zone of control $[-Kt_n, Kt_n]$ with $K = \frac{1}{(4e+8)A^3}$. If ξ is in this zone, then we have indeed

$$|\xi| \leq \frac{(\widetilde{\sigma}_n)^2}{4eA^3} \left(\frac{N_n}{D_n} \right)^{1/3} \leq \frac{1}{2eA} \left(\frac{N_n}{D_n} \right)^{1/3} = L \left(\frac{N_n}{D_n} \right)^{1/3}$$

by using the remark just before the lemma. Then,

$$|z| \leq \frac{2A^3}{1 - \frac{2eA \, (\widetilde{\sigma}_n)^2}{(4e+8)A^3}} |\xi|^3 \leq \frac{2A^3}{1 - \frac{e}{e+2}} |\xi|^3 = (2+e)A^3 \, |\xi|^3.$$

Thus, on the zone of control, $|\theta_n(\xi) - 1| = |e^z - 1| \leq |z| \, e^{|z|}$, whence a control of index $(3,3)$ and with constants

$$K_1 = K_2 = (2+e)A^3.$$

We have chosen K so that $K = \frac{1}{4K_2}$, hence, the inequalities of Condition (Z2) are satisfied. $\qquad \square$

Using the results of Section 2, we obtain:

Corollary 4.3. *Let S_n be a sequence with a uniform bounds on cumulants with parameters (D_n, N_n, A) and let $Y_n = \frac{S_n}{\sqrt{\mathrm{Var}(S_n)}}$. Then we have*

$$d_{\mathrm{Kol}}(Y_n, \mathcal{N}_{\mathbb{R}}(0,1)) \leq \frac{76.36 \, A^3}{(\widetilde{\sigma}_n)^3} \sqrt{\frac{D_n}{N_n}}.$$

Proof. We can apply Theorem 2.16, choosing $\gamma = 1$, and $\lambda = 0.193$. It yields a constant smaller than 77.911. It is however possible to get the better constant given above by redoing some of the computations in this specific setting. With $\rho > 4$, set $K = \frac{1}{(4e+\rho) \, A^3}$, and $\varepsilon_n = \frac{1}{K \, (t_n)^{3/2}}$. On the zone $\xi \in [-\frac{1}{\varepsilon_n}, \frac{1}{\varepsilon_n}]$, we have a bound $|\theta_n(\xi) - 1| \leq$

$M|\xi|^3 \exp(M|\xi|^3)$, this time with $M = (2 + \frac{8e}{\rho})A^3$. Hence,

$$
|\mathbb{E}[f_{a,\varepsilon_n}(Y_n)] - \mathbb{E}[f_{a,\varepsilon_n}(Y)]| \leq \frac{M}{2\pi (t_n)^{3/2}} \int_{-\frac{1}{\varepsilon_n}}^{\frac{1}{\varepsilon_n}} |\xi|^2 e^{-\frac{\xi^2}{2} + M\frac{|\xi|^3}{(t_n)^{3/2}}} d\xi
$$

$$
\leq \frac{M}{2\pi (t_n)^{3/2}} \int_{\frac{1}{\varepsilon_n}}^{\frac{1}{\varepsilon_n}} |\xi|^2 e^{-\xi^2\left(\frac{1}{2} - \frac{2}{\rho}\right)} d\xi
$$

$$
\leq \frac{M}{\sqrt{2\pi} (t_n)^{3/2} (1 - \frac{4}{\rho})^{3/2}} = \frac{1}{\sqrt{2\pi}} \frac{2}{\rho(1 - \frac{4}{\rho})^{3/2}} \varepsilon_n.
$$

By Theorem 2.15, we get, for any $\lambda > 0$:

$$
d_{\mathrm{Kol}}\left(\frac{S_n}{\sqrt{\mathrm{Var}(S_n)}}, \mathcal{N}_{\mathbb{R}}(0,1)\right) \leq \frac{(1+\lambda)}{\sqrt{2\pi}} \left(\frac{2}{\rho(1 - \frac{4}{\rho})^{3/2}} + \frac{1}{\sqrt[3]{\pi}} \left(4\sqrt[3]{1 + \frac{1}{\lambda}} + 3\sqrt[3]{3}\right)\right) \varepsilon_n.
$$

The best constant is then obtained with $\rho = 6.79$ and $\lambda = 0.185$. $\qquad\square$

Remark 4.4. There is a trade-off in the bound of Corollary 4.3 between the *a priori* upper bound on $(\widetilde{\sigma}_n)^2$, and the constant C that one obtains such that the Kolmogorov distance is smaller than

$$
\frac{CA^3}{(\widetilde{\sigma}_n)^3} \sqrt{\frac{D_n}{N_n}}.
$$

This bound gets worse when $(\widetilde{\sigma}_n)^2$ is small, but on the other hand, the knowledge of a better *a priori* upper bound (that is precisely when $(\widetilde{\sigma}_n)^2$ is small) yields a better constant C. So, for instance, if one knowns that $(\widetilde{\sigma}_n)^2 \leq A^2$ (instead of $2A^2$), then one gets a constant $C = 52.52$. A general bound that one can state and that takes into account this trade-off is:

$$
d_{\mathrm{Kol}}(Y_n, \mathcal{N}_{\mathbb{R}}(0,1)) \leq 27.55 \left(\left(\frac{A}{\widetilde{\sigma}_n}\right)^3 + \frac{A}{\widetilde{\sigma}_n}\right) \sqrt{\frac{D_n}{N_n}}.
$$

We are indebted to Martina Dal Borgo for having pointed out this phenomenon. In the sequel, we shall freely use this small improvement of Corollary 4.3.

The above corollary ensures asymptotic normality with a bound on the speed of convergence when $\frac{1}{(\widetilde{\sigma}_n)^3} \sqrt{\frac{D_n}{N_n}}$ tends to 0, *i.e.* when $(\widetilde{\sigma}_n)^2 \left(\frac{N_n}{D_n}\right)^{1/3}$ tends to $+\infty$. Using a theorem of Janson, the asymptotic normality can be obtained under a less restrictive hypothesis, but without bound on the speed of convergence. Even if the main topic of the paper is to find bounds on the speed of convergence, we will recall the result here for the sake of completeness.

Proposition 4.5. *As above, let S_n be a sequence with a uniform bounds on cumulants with parameters (D_n, N_n, A) and assume that*

$$
\lim_{n \to \infty} \frac{\mathrm{Var}(S_n)}{N_n D_n} \left(\frac{N_n}{D_n}\right)^{\varepsilon} = \lim_{n \to \infty} (\widetilde{\sigma}_n)^2 \left(\frac{N_n}{D_n}\right)^{\varepsilon} = +\infty
$$

for some parameter $\varepsilon \in (0, 1)$. Then,

$$\frac{S_n}{\sqrt{\text{Var}(S_n)}} \rightharpoonup \mathcal{N}_{\mathbb{R}}(0, 1).$$

Proof. Note that the bounds on cumulants can be rewritten as

$$|\kappa^{(r)}(Y_n)| \leq C_r \left(\frac{\text{Var}(S_n)}{N_n D_n} \left(\frac{N_n}{D_n} \right)^{1 - \frac{2}{r}} \right)^{-\frac{r}{2}}$$

for some constant C_r. Choosing r large enough so that $1 - \frac{2}{r} \geq \varepsilon$, we conclude that

$$\kappa^{(r)}(Y_n) \to 0 \quad \text{for } r \text{ large enough.}$$

This is a sufficient condition for the convergence in law towards a Gaussian distribution, see [Jan88, Theorem 1] and [Gri92]. □

Remark 4.6. Up to a change of parameters, it would be equivalent to consider bounds of the kind

$$|\kappa^{(r)}(S_n)| \leq (Cr)^r \alpha_n (\beta_n)^r, \quad \text{or} \quad \left| \kappa^{(r)} \left(\frac{S_n}{\sqrt{\text{Var}(S_n)}} \right) \right| \leq \frac{r!}{\Delta_n^{r-2}},$$

as done by Saulis and Statulevičius in [SS91] or by the authors of this paper in [FMN16]. In particular, it was proved in [FMN16, Chapter 5], that under slight additional assumptions on the second and third cumulants, we have the following: the sequence $(X_n)_{n \in \mathbb{N}}$ defined in Lemma 4.2 converges in the mod-Gaussian sense with parameters $(t_n)_{n \in \mathbb{N}}$.

4.2. **Dependency graphs.** In this paragraph, we will see that the uniform bounds on cumulants are satisfied for sums $S_n = \sum_{i=1}^{n} A_{i,n}$ of dependent random variables with specific dependency structure.

More precisely, if $(A_v)_{v \in V}$ is a family of real valued random variables, we call *dependency graph* for this family a graph $G = (V, E)$ with the following property: if V_1 and V_2 are two disjoint subsets of V with no edge $e \in E$ joining a vertex $v_1 \in V_1$ to a vertex $v_2 \in V_2$, then $(A_v)_{v \in V_1}$ and $(A_v)_{v \in V_2}$ are independent random vectors. For instance, let $(A_1, \ldots, A_7)$ be a family of random variables with dependency graph drawn on Figure 6. Then the vectors $(A_1, A_2, A_3, A_4, A_5)$ and (A_6, A_7) corresponding to different connected components must be independent. Moreover, note that (A_1, A_2) and (A_4, A_5) must be independent as well: although they are in the same connected component of the graph G, they are not directly connected by an edge $e \in E$.
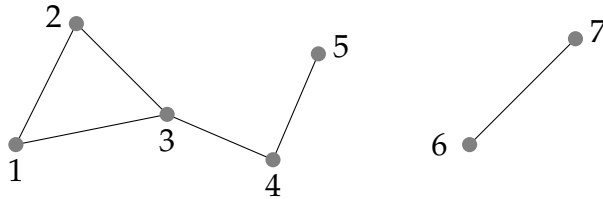

FIGURE 6. A dependency graph for 7 real-valued random variables.

Theorem 4.7 (Féray–Méliot–Nikeghbali, see [FMN16]). *Let $(A_v)_{v \in V}$ be a family of random variables, with $|A_v| \leq A$ a.s., for all $v \in V$. We suppose that $G = (V, E)$ is a dependency graph for the family and denote*

- *$N = \frac{\sum_{v \in V} \mathbb{E}|A_v|}{A} \leq \text{card } V$;*
- *D the maximum degree of a vertex in G plus one.*

If $S = \sum_{v \in V} A_v$, then for all $r \geq 1$,

$$|\kappa^{(r)}(S)| \leq N \, r^{r-2} \, (2D)^{r-1} \, A^r.$$

Consider a sequence $(S_n)_{n \in \mathbb{N}}$, where each S_n writes as $\sum_{v \in V_n} A_{v,n}$, with the $A_{v,n}$ uniformly bounded by A (in a lot of examples, the $A_{v,n}$ are indicator variables, so that we can take $A = 1$). Set

$$N_n = \frac{\sum_{v \in V_n} \mathbb{E}|A_{v,n}|}{A}$$

and assume that, for each n, $(A_{v,n})_{v \in V_n}$ has a dependency graph of maximal degree $D_n - 1$. Then the sequence $(S_n)_{n \in \mathbb{N}}$ admits uniform bounds on cumulants with parameters (D_n, N_n, A) and the result of the previous section applies. Note that in this setting we have the bound $\widetilde{\sigma}_n \leq A$, so the bound of Corollary 4.3 holds with the better constant 52.52.

Remark 4.8. The parameter D is equal to the maximal number of neighbors of a vertex $v \in V$, plus 1. In the following, we shall simply call D the maximal degree, being understood that one always has to add 1. Another way to deal with this convention is to think of dependency graphs as having one loop at each vertex.

Example 4.9. The following example, though quite artificial, shows that one can construct families of random variables with arbitrary parameters N and D for their dependency graphs. Let $(U_k)_{k \in \mathbb{Z}/N\mathbb{Z}}$ be a family of independent Bernoulli random variables with $\mathbb{P}[U_k = 1] = 1 - \mathbb{P}[U_k = 0] = q$; and for $i \in \mathbb{Z}/N\mathbb{Z}$,

$$A_i = 2 \left(\prod_{k=i+1}^{i+D} U_k \right) - 1.$$

Each A_i is a Bernoulli random variable with $\mathbb{P}[A_i = 1] = 1 - \mathbb{P}[A_i = -1] = q^D$, which we denote p (p is considered independent of N). We are interested in the fluctuations of $S = \sum_{i=1}^{N} A_i$. Note that the partial sums $\sum_{i=1}^{k \leq N} A_i$ correspond to random walks with correlated steps: as D increases, the consecutive steps of the random walk have a higher probability to be in the same direction, and therefore, the variance of the sum $S = \sum_{i=1}^{N} A_i$ grows. We refer to Figure 7, where three such random walks are drawn, with parameters $p = \frac{1}{2}$, $N = 1000$ and $D \in \{5, 15, 30\}$.

If $d(i, j) \geq D$ in $\mathbb{Z}/N\mathbb{Z}$, then A_i and A_j do not involve any common variable U_k, and they are independent. It follows that if G is the graph with vertex set $\mathbb{Z}/N\mathbb{Z}$ and with an edge e between i and j if $d(i, j) \leq D$, then G is a dependency graph for the A_i's. This graph has N vertices, and maximal degree $2D - 1$. Moreover, one can compute exactly
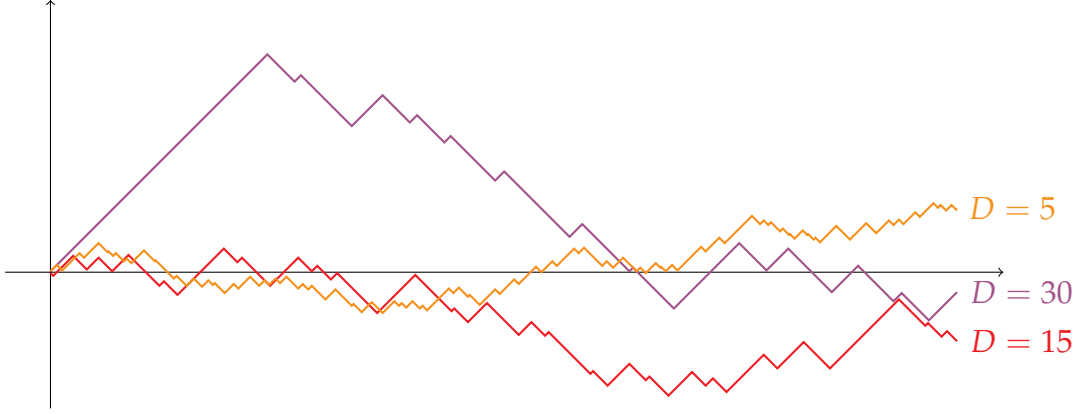
FIGURE 7. Random walks with 1000 steps and correlation lengths $D = 5$, $D = 15$ and $D = 30$.

the expectation and the variance of $S = \sum_{i=1}^{N} A_i$:

$$\mathbb{E}[S] = N(2p - 1);$$

$$\mathrm{Var}(S) = 4 \sum_{i,j=1}^{N} \mathrm{cov}(U_{i+1} \cdots U_{i+D}, U_{j+1} \cdots U_{j+N})$$

$$= 4 \sum_{i,j=1}^{N} \left(q^{D+\min(D,d(i,j))} - p^2 \right) = 4Np \sum_{j=-(D-1)}^{D-1} (q^{|j|} - p)$$

$$= 4Np \left(\frac{1 + p^{\frac{1}{D}} - 2p}{1 - p^{\frac{1}{D}}} - (2D - 1)p \right).$$

If N and D go to infinity with $D = o(N)$, then $q = p^{\frac{1}{D}} = 1 + \frac{\log p}{D} + O(\frac{1}{D^2})$, and

$$\mathbb{E}[S] = N(2p - 1);$$

$$\mathrm{Var}(S) = 8N(D + O(1)) \, p \left(\frac{1 - p}{-\log p} - p \right).$$

So, one can apply Corollary 4.3 to the sum S, and one obtains:

$$d_{\mathrm{Kol}} \left(\frac{S - \mathbb{E}[S]}{\sqrt{\mathrm{Var}(S)}}, \mathcal{N}_{\mathbb{R}}(0,1) \right) \leq \frac{6}{\left(p \left(\frac{1-p}{-\log p} - p \right) \right)^{3/2}} \sqrt{\frac{D}{N}}.$$

Example 4.10. Fix $p \in (0,1)$, and consider a random *Erdös–Rényi graph* $G = G_n = G(n, p)$, which means that one keeps at random each edge of the complete graph K_n with probability p, independently from every other edge. Note that we only consider the case of fixed p here; for $p \to 0$, we would get rather weak bounds, see [FMN16, Section 10.3.3] for a discussion on bounds on cumulants in this framework.

Let $H = (V_H, E_H)$ and $G = (V_G, E_G)$ be two graphs. The number of copies of H in G is the number of injections $i : V_H \to V_G$ such that, if $(h_1, h_2) \in E_H$, then $(i(h_1), i(h_2)) \in E_G$. In random graph theory, this is called the *subgraph count* statistics; we denote it by

$I(H, G)$. We refer to Figure 8 for an example, where $H = K_3$ is the triangle and G is a random Erdös–Rényi graph of parameters $n = 30$ and $p = \frac{1}{10}$.
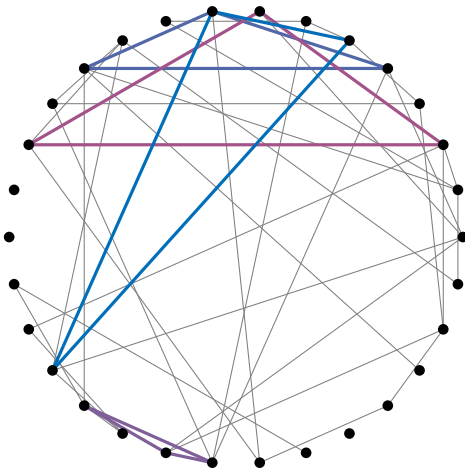

FIGURE 8. Count of triangles in a random Erdös–Rényi graph of parameters $n = 30$ and $p = 0.1$. Here, there are $4 \times 3! = 24$ ways to embed a triangle in the graph.

One can always write $I(H, G)$ as a sum of dependent random variables. Identify V_H with $[\![1, k]\!]$ and V_G with $[\![1, n]\!]$, and denote $\mathfrak{A}(n, k)$ the set of arrangements $(a_1, \ldots, a_k)$ of size k in $[\![1, n]\!]$. Given such an arrangement, the *induced subgraph* $G[a_1, \ldots, a_k]$ is the graph with vertex set $[\![1, k]\!]$, and with an edge between i and j if $(a_i, a_j) \in E_G$. Then,

$$I(H, G) = \sum_{A \in \mathfrak{A}(n,k)} I_A(H, G),$$

where $I_A(H, G) = 1$ if $H \subset G[A]$, and 0 otherwise.

A dependency graph for the random variables $I_A(H, G_n)$ has vertex set $\mathfrak{A}(n, k)$ of cardinality $N_n = n^{\downarrow k} = n(n-1)(n-2) \cdots (n-k+1)$, and an edge between two arrangements $(a_1, \ldots, a_k)$ and $(b_1, \ldots, b_k)$ if they share at least two points (otherwise, the random variables $I_A(H, G_n)$ and $I_B(H, G_n)$ involve disjoint sets of edges and are therefore independent). As a consequence, the maximal degree of the graph is smaller than

$$D_n = \left(\binom{k}{2}^2 2(n-2)(n-3) \cdots (n-k+1) \right),$$

and of order n^{k-2}. Therefore, $\frac{D_n}{N_n} \le \frac{2\binom{k}{2}^2}{n(n-1)} = O(\frac{1}{n^2})$, and on the other hand, if h is the number of edges of H, one can compute the asymptotics of the expectation and of the variance of $I(H, G_n)$:

$$\mathbb{E}[I(H, G_n)] = n^{\downarrow k} p^h \quad ; \quad \text{Var}(I(H, G_n)) = 2h^2 p^{2h-1}(1-p) n^{2k-2} + O(n^{2k-3}),$$

see [FMN16, Section 10] for the details of these computations. In particular,

$$\lim_{n \to \infty} \frac{\text{Var}(S_n)}{N_n D_n} = p^{2h-1}(1-p) \left(\frac{h}{\binom{k}{2}} \right)^2 = \widetilde{\sigma}^2 > 0.$$

Thus, using Corollary 4.3, we get

$$d_{\mathrm{Kol}}\left(\frac{I(H,G_n) - \mathbb{E}[I(H,G_n)]}{\sqrt{\mathrm{Var}(I(H,G_n))}}, \mathcal{N}_{\mathbb{R}}(0,1)\right) \leq \frac{4.65\,(k(k-1))^4}{p^{3h}\,(\frac{1}{p}-1)^{3/2}\,h^3}\frac{1}{n}$$

for n large enough. For instance, if $T_n = I(K_3, G_n)$ is the number of triangles in G_n, then

$$\mathbb{E}[T_n] = n^{\downarrow 3}\,p^3$$
$$\mathrm{Var}[T_n] = 18\,n^{\downarrow 4}\,p^5(1-p) + 6\,n^{\downarrow 3}\,p^3(1-p^3)$$

and

$$d_{\mathrm{Kol}}\left(\frac{T_n - n^{\downarrow 3}p^3}{\sqrt{18\,n^{\downarrow 4}\,p^5(1-p) + 6\,n^{\downarrow 3}\,p^3(1-p^3)}}, \mathcal{N}_{\mathbb{R}}(0,1)\right) \leq \frac{234}{p^9(\frac{1}{p}-1)^{\frac{3}{2}}}\frac{1}{n}(1+O(\tfrac{1}{n})).$$

This result is not new, except maybe the explicit constant. We refer to [BKR89] for an approach of speed of convergence for subgraph counts using Stein's method. More recently, Krokowski, Reichenbachs and Thäle [KRT15] applied Malliavin calculus to the same problem. Our result corresponds to the case where p is constant of their Theorem 1. Similar bounds could be obtained by Stein's method, see [Rin94].

To conclude our presentation of the convergence of sums of bounded random variables with sparse dependency graphs, let us analyse precisely the case of *uncorrelated* random variables.

Corollary 4.11. *Let $S_n = \sum_{i=1}^{N_n} A_{i,n}$ be a sum of centered and bounded random variables, that are uncorrelated and with $\mathbb{E}[(A_{i,n})^2] = 1$ for all i. We suppose that the random variables have a dependency graph of parameters $N_n \to +\infty$ and D_n.*

   i) If $D_n = O((N_n)^{1/2-\varepsilon})$ for $\varepsilon > 0$, then $Y_n = \frac{S_n}{\sqrt{N_n}}$ converges in law towards the Gaussian distribution.

   ii) If $D_n = o((N_n)^{1/4})$, then the Kolmogorov distance between Y_n and $\mathcal{N}_{\mathbb{R}}(0,1)$ is a $O((D_n)^2/(N_n)^{1/2})$.

Proof. It is an immediate consequence of Corollary 4.3 and Proposition 4.5, since S_n admits uniform control on cumulants (see Theorem 4.7) and

$$\mathrm{Var}(S_n) = \sum_{i=1}^{N_n} \mathbb{E}[(A_{i,n})^2] = N_n. \qquad \square$$

4.3. **Unbounded random variables and truncation methods.** A possible generalization regards sums of *unbounded* random variables. In the following, we develop a truncation method that yields a criterion of asymptotic normality similar to Lyapunov's condition (see [Bil95, Chapter 27]). A small modification of this method would similarly yield a Lindeberg type criterion. Let $S_n = \sum_{i=1}^{N_n} A_{i,n}$ be a sum of centered random variables, with

$$\left(\mathbb{E}[|A_{i,n}|^{2+\delta}]\right)^{\frac{1}{2+\delta}} \leq A$$

for some constant A independent of i and n, and some $\delta > 0$. We suppose as before that the family of random variables $(A_{i,n})_{i\in[\![1,N_n]\!]}$ has a (true) dependency graph G_n of parameters N_n and D_n. Note that in this case,

$$\mathrm{Var}(S_n) = \sum_{i,j=1}^{N_n} \mathrm{cov}(A_{i,n}, A_{j,n}) \leq \sum_{i=1}^{N_n}\sum_{j\sim i} \|A_{i,n}\|_2 \|A_{j,n}\|_2$$

$$\leq \sum_{i=1}^{N_n}\sum_{j\sim i} \|A_{i,n}\|_{2+\delta} \|A_{j,n}\|_{2+\delta} \leq A^2 N_n D_n.$$

We set

$$A_{i,n}^- = A_{i,n} 1_{|A_{i,n}|\leq L_n} \qquad ; \qquad A_{i,n}^+ = A_{i,n} 1_{|A_{i,n}|>L_n} ;$$

$$S_n^- = \sum_{i=1}^{N_n} A_{i,n}^- \qquad ; \qquad S_n^+ = \sum_{i=1}^{N_n} A_{i,n}^+ ;$$

where L_n is a truncation level, to be chosen later. Notice that G_n is still a dependency graph for the family of truncated random variables $(A_{i,n}^-)_{i\in[\![1,N_n]\!]}$. Therefore, we can apply the previously developed machinery (Theorem 4.7 and Corollary 4.3) to the scaled sum S_n^-/L_n. On the other hand, by Markov's inequality,

$$d_{\mathrm{Kol}}(S_n, S_n^-) = \sup_{s\in\mathbb{R}} |\mathbb{P}[S_n \geq s] - \mathbb{P}[S_n^- \geq s]| \leq \mathbb{P}[S_n^+ = 0]$$

$$\leq \sum_{i=1}^{N_n} \mathbb{P}[|A_{i,n}| \geq L_n] \leq N_n \left(\frac{A}{L_n}\right)^{2+\delta}.$$

Combining the two arguments leads to the following result (this replaces the previous assumption of boundedness $|A_{i,n}| \leq A$).

Theorem 4.12. *Let $(S_n = \sum_{i=1}^{N_n} A_{i,n})_{n\in\mathbb{N}}$ be a sum of centered random variables, with dependency graphs of parameters $N_n \to +\infty$ and D_n, and with*

$$\|A_{i,n}\|_{2+\delta} = (\mathbb{E}[|A_{i,n}|^{2+\delta}])^{1/(2+\delta)} \leq A$$

for all i, n and for some $\delta > 0$. We set $Y_n = S_n / \sqrt{\mathrm{Var}(S_n)}$. Recall that $(\widetilde{\sigma}_n)^2 = \frac{\mathrm{Var}(S_n)}{N_n D_n}$.

(U1) Set

$$V_n = (\widetilde{\sigma}_n)^2 \left(\frac{N_n}{D_n}\right)^{1/3} \frac{1}{(N_n)^{2/(2+\delta)}}$$

and suppose that $\lim_{n\to\infty} V_n = +\infty$ (which is only possible for $\delta > 4$). Then, for n large enough,

$$d_{\mathrm{Kol}}(Y_n, \mathcal{N}_{\mathbb{R}}(0,1)) \leq 78 \left(\frac{A^2}{V_n}\right)^{\frac{3(\delta+2)}{2(\delta+5)}} = o\left(\frac{1}{V_n}\right).$$

(U2) More generally, for $\varepsilon \in (\frac{2}{2+\delta}, 1)$, set

$$W_n = (\widetilde{\sigma}_n)^2 \left(\frac{N_n}{D_n}\right)^{\varepsilon} \frac{1}{(N_n)^{2/(2+\delta)}}$$

and suppose that $\lim_{n\to\infty} W_n = +\infty$. Then, $Y_n \rightharpoonup \mathcal{N}_{\mathbb{R}}(0,1)$.

Remark 4.13. It should be noticed that if $\delta \to +\infty$, then one essentially recovers the content of Corollary 4.3 (which can be applied because of Theorem 4.7). On the other hand, the inequality $\delta > 4$ amounts to the existence of bounded moments of order strictly higher than 6 for the random variables $A_{i,n}$. In practice, one can for instance ask for bounded moments of order 7 (*i.e.* $\delta = 5$), in which case the first condition (U1) reads

$$\lim_{n\to\infty} \frac{\mathrm{Var}(S_n)}{D_n \, N_n} \frac{(N_n)^{1/9}}{(D_n)^{1/3}} = +\infty.$$

Moreover, we will see in the proof of Theorem 4.12 that in this setting ($\delta = 5$), the constant 78 can be improved to 39, so that, for n large enough:

$$d_{\mathrm{Kol}}(Y_n, \mathcal{N}_{\mathbb{R}}(0,1)) \leq 39 \left(\frac{A^2}{V_n}\right)^{\frac{21}{20}}.$$

Proof. We write as usual $Y_n = \frac{S_n}{\sqrt{\mathrm{Var}(S_n)}}$, and $Y_n^- = \frac{S_n^-}{\sqrt{\mathrm{Var}(S_n^-)}}$. In all cases, we have

$$d_{\mathrm{Kol}}(Y_n, \mathcal{N}_{\mathbb{R}}(0,1)) \leq d_{\mathrm{Kol}}(S_n, S_n^-) + d_{\mathrm{Kol}}\left(Y_n^-, \sqrt{\frac{\mathrm{Var}(S_n)}{\mathrm{Var}(S_n^-)}}\, \mathcal{N}_{\mathbb{R}}(0,1)\right)$$

$$\leq N_n \left(\frac{A}{L_n}\right)^{2+\delta} + d_{\mathrm{Kol}}(Y_n^-, \mathcal{N}_{\mathbb{R}}(0,1)) + d_{\mathrm{Kol}}\left(\mathcal{N}_{\mathbb{R}}(0,1), \mathcal{N}_{\mathbb{R}}\left(0, \frac{\mathrm{Var}(S_n)}{\mathrm{Var}(S_n^-)}\right)\right)$$

by using the invariance of the Kolmogorov distance with respect to multiplication of random variables by a positive constant. In the sequel, we denote a, b and c the three terms on the right-hand side of the inequality. The Kolmogorov distance between two Gaussian distributions is

$$d_{\mathrm{Kol}}(\mathcal{N}_{\mathbb{R}}(0,1), \mathcal{N}_{\mathbb{R}}(0,\lambda^2)) = \frac{1}{\sqrt{2\pi}} \sup_{s\in\mathbb{R}_+} \left(\int_s^{\lambda s} e^{-\frac{u^2}{2}}\, du\right)$$

$$\leq \frac{\lambda - 1}{\sqrt{2\pi}} \sup_{s\in\mathbb{R}_+} \left(s\, e^{-\frac{s^2}{2}}\right) = \sqrt{\frac{1}{2\pi e}}\, |\lambda - 1|$$

if $\lambda \geq 1$. One gets the same result if $\lambda \leq 1$, hence,

$$d_{\mathrm{Kol}}\left(\mathcal{N}_{\mathbb{R}}(0,1), \mathcal{N}_{\mathbb{R}}\left(0, \frac{\mathrm{Var}(S_n)}{\mathrm{Var}(S_n^-)}\right)\right) = d_{\mathrm{Kol}}\left(\mathcal{N}_{\mathbb{R}}(0,1), \mathcal{N}_{\mathbb{R}}\left(0, \frac{\mathrm{Var}(S_n^-)}{\mathrm{Var}(S_n)}\right)\right)$$

$$\leq \sqrt{\frac{1}{2\pi e}} \left|\sqrt{\frac{\mathrm{Var}(S_n^-)}{\mathrm{Var}(S_n)}} - 1\right|$$

$$\leq \sqrt{\frac{1}{2\pi e}} \frac{|\mathrm{Var}(S_n^-) - \mathrm{Var}(S_n)|}{\mathrm{Var}(S_n)}.$$

To evaluate the difference between the variances, notice that

$$\mathrm{Var}(S_n^-) = \mathrm{Var}\left(S_n - \sum_{i=1}^{N_n} A_{i,n}^+\right)$$

$$= \mathrm{Var}(S_n) - 2\sum_{i,j=1}^{N_n} \mathrm{Cov}(A_{i,n}^+, A_{j,n}) + \sum_{i,j=1}^{N_n} \mathrm{Cov}(A_{i,n}^+, A_{j,n}^+)$$

If j is not connected to i in G_n, or equal to i, then $A_{i,n}$ and $A_{j,n}$ are independent, hence, $\mathrm{Cov}(A_{i,n}^+, A_{j,n}) = 0$. Otherwise, using Hölder and Bienaymé-Chebyshev inequalities,

$$|\mathrm{Cov}(A_{i,n}^+, A_{j,n})| = |\mathbb{E}[A_{i,n}^+ A_{j,n}]| \leq \sqrt{\mathbb{E}[(A_{i,n})^2 \mathbb{1}_{|A_{i,n}| \geq L_n}] \, \mathbb{E}[(A_{j,n})^2 \mathbb{1}_{|A_{i,n}| \geq L_n}]}$$

$$\leq \sqrt{\mathbb{E}[(A_{i,n})^{2+\delta}]^{\frac{2}{2+\delta}} \, \mathbb{P}[|A_{i,n}| \geq L_n]^{\frac{2\delta}{2+\delta}} \, \mathbb{E}[(A_{j,n})^{2+\delta}]^{\frac{2}{2+\delta}}}$$

$$\leq A^2 \left(\frac{A}{L_n}\right)^\delta.$$

Similarly,

$$|\mathrm{Cov}(A_{i,n}^+, A_{j,n}^+)| \leq \mathbb{E}[|A_{i,n}^+ A_{j,n}^+|] + \mathbb{E}[|A_{i,n}^+|] \, \mathbb{E}[|A_{j,n}^+|] \leq A^2 \left(\left(\frac{A}{L_n}\right)^\delta + \left(\frac{A}{L_n}\right)^{2+2\delta}\right),$$

hence, assuming that the level of truncation L_n is larger than A,

$$\frac{|\mathrm{Var}(S_n)^- - \mathrm{Var}(S_n)|}{\mathrm{Var}(S_n)} \leq \frac{N_n D_n}{\mathrm{Var}(S_n)} 3A^2 \left(\frac{A}{L_n}\right)^\delta.$$

Let us now place ourselves in the setting of Hypothesis (U1); we set

$$V_n = \frac{\mathrm{Var}(S_n)}{N_n D_n} \left(\frac{N_n}{D_n}\right)^{1/3} \frac{1}{(N_n)^{2/(2+\delta)}}.$$

Suppose that $L_n = K_n (N_n)^{\frac{1}{2+\delta}}$, with K_n going to infinity. We then have $a \leq \frac{A^{2+\delta}}{(K_n)^{2+\delta}}$, and on the other hand,

$$\left|\frac{\mathrm{Var}(S_n^-)}{\mathrm{Var}(S_n)} - 1\right| \leq 3A^{2+\delta} \left(\frac{N_n D_n}{\mathrm{Var}(S_n)} \frac{1}{(N_n)^{\delta/(2+\delta)}} \frac{1}{(K_n)^\delta}\right)$$

$$\leq 3A^{2+\delta} \left(\frac{1}{V_n (D_n)^{1/3} (N_n)^{2/3} (K_n)^\delta}\right) \to 0$$

since by hypothesis, $\lim_{n\to\infty} \frac{1}{V_n} = 0$. So,

$$c \leq \frac{3A^{2+\delta}}{\sqrt{2\pi e}} \frac{1}{V_n (D_n)^{1/3} (N_n)^{2/3} (K_n)^\delta} \leq \frac{3A^{2+\delta}}{\sqrt{2\pi e}} \frac{1}{V_n (K_n)^\delta}.$$

Now, the sequence $(S_n^-/L_n)_{n\in\mathbb{N}}$ is a sequence of sums of centered random variables all bounded by 1, and to apply Corollary 4.3 to this sequence, we need

$$\lim_{n\to\infty} \frac{\mathrm{Var}(S_n^-)}{N_n D_n} \left(\frac{N_n}{D_n}\right)^{1/3} \frac{1}{(L_n)^2} = +\infty.$$

However, the previous computation shows that one can replace $\mathrm{Var}(S_n)$ by $\mathrm{Var}(S_n^-)$ in this expression without changing the asymptotic behavior, so

$$\lim_{n\to\infty} \frac{\mathrm{Var}(S_n^-)}{N_n D_n} \left(\frac{N_n}{D_n}\right)^{1/3} \frac{1}{(L_n)^2} = \lim_{n\to\infty} \frac{V_n}{(K_n)^2},$$

which is $+\infty$ if $(K_n)^2$ is not growing too fast to $+\infty$ (in comparison to the sequence V_n). Then, by Corollary 4.3,

$$b = d_{\mathrm{Kol}}(Y_n^-, \mathcal{N}_\mathbb{R}(0,1)) \leq 77 \frac{(K_n)^3}{(V_n)^{3/2}}$$

for n large enough. Set $K_n = B(V_n)^{\frac{3}{2(5+\delta)}}$. Then,

$$d_{\mathrm{Kol}}(Y_n, \mathcal{N}_\mathbb{R}(0,1)) \leq a + b + c \leq \left(\frac{A^{2+\delta}}{B^{2+\delta}} + 77\,B^3 + o(1) \right) \left(\frac{1}{V_n} \right)^{\frac{3(2+\delta)}{2(5+\delta)}}$$

$$\leq \left(\left(\frac{231}{2+\delta} \right)^{\frac{2+\delta}{5+\delta}} + 77 \left(\frac{2+\delta}{231} \right)^{\frac{3}{5+\delta}} \right) \left(\frac{A^2}{V_n} \right)^{\frac{3(2+\delta)}{2(5+\delta)}}$$

for n large enough, and by choosing B in an optimal way. The term in parenthesis is maximal when $\delta = 229$, and is then equal to 78. This ends the proof of (U1), and one gets a better constant smaller than 39 when $\delta = 5$.

Under the Hypothesis (U2), we set

$$W_n = \frac{\mathrm{Var}(S_n)}{N_n\,D_n} \left(\frac{N_n}{D_n} \right)^\varepsilon \frac{1}{(N_n)^{2/(2+\delta)}}.$$

In order to prove the convergence in law $Y_n \rightharpoonup \mathcal{N}_\mathbb{R}(0,1)$, it suffices to have:

- $S_n - S_n^- = S_n^+$ that converges in probability to 0. This happens as soon as the level L_n is $K_n\,(N_n)^{1/(2+\delta)}$ with $K_n \to +\infty$.
- $\frac{|\mathrm{Var}(S_n) - \mathrm{Var}(S_n^-)|}{\mathrm{Var}(S_n)} \to 0$. With $L_n = K_n\,(N_n)^{1/(2+\delta)}$, the previous computations show that this quantity is a

$$O \left(\frac{1}{W_n\,(N_n)^{1-\varepsilon}\,(D_n)^\varepsilon\,(K_n)^\delta} \right),$$

which goes to 0.

- and by Theorem 4.5,

$$\frac{\mathrm{Var}(S_n)}{N_n\,D_n} \left(\frac{N_n}{D_n} \right)^\varepsilon \frac{1}{(L_n)^2} \to +\infty.$$

This follows from the Hypothesis (U2) if K_n is chosen to grow sufficiently slow.

Thus, the second part of Theorem 4.12 is proven. $\qquad\square$

Example 4.14. Let $(X_i)_{i \in [\![1,N]\!]}$ be a centered Gaussian vector with $\mathbb{E}[(X_i)^2] = 1$ for any i, and with the covariance matrix $(\mathrm{Cov}(X_i, X_j))_{1 \leq i,j \leq N}$ that is sparse in the following sense: for any i, the set of indices j such that $\mathrm{Cov}(X_i, X_j) \neq 0$ has cardinality smaller than D. We set $A_i = \exp(X_i)$; the random variables A_i follow the *log-normal distribution* of density

$$\frac{1}{\sqrt{2\pi}} \frac{1}{u^{1+\frac{\log u}{2}}} \mathbb{1}_{u>0}\,du,$$

and they have moments of all order:

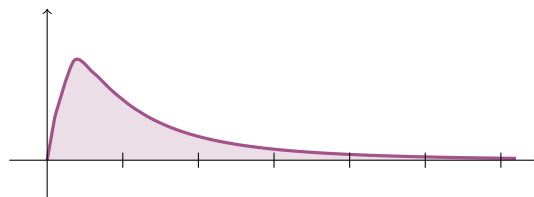$$\mathbb{E}[(A_i)^k] = \mathbb{E}[e^{kX_i}] = e^{\frac{k^2}{2}}.$$

FIGURE 9. The density of the log-normal distribution.

The variables A_i have the same dependency graph as the variables X_i. Moreover, if $\rho_{ij} = \text{Cov}(X_i, X_j)$, then the covariance of two variables A_i and A_j is $e(e^{\rho_{ij}} - 1)$. Using moments of order $2 + \delta$, we see that if

$$Y_N = \frac{\sum_{i=1}^{N}(A_i - e^{\frac{1}{2}})}{\sqrt{e\sum_{1\leq i,j\leq N}(e^{\rho_{ij}} - 1)}} \qquad ; \qquad V_{N,\delta} = \frac{e\sum_{1\leq i,j\leq N}(e^{\rho_{ij}} - 1)}{ND}\frac{N^{\frac{1}{3} - \frac{2}{2+\delta}}}{D^{\frac{1}{3}}} \to +\infty,$$

then $d_{\text{Kol}}(Y_N, \mathcal{N}_{\mathbb{R}}(0,1)) \leq 78\left(\frac{e^{\delta+2}}{V_{N,\delta}}\right)^{\frac{3(\delta+2)}{2(\delta+5)}}$ for N large enough.

To make this result more explicit, let us consider the following dependency structure for the Gaussian vector $X = (X_i)_{i\in[\![1,N]\!]}$:

$$\text{Cov}(X) = \begin{pmatrix} 1 & * & \cdots & * \\ * & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ * & \cdots & * & 1 \end{pmatrix},$$

where the non-diagonal terms $*$ are all smaller than $\frac{\rho}{D}$ in absolute value, and with less than D non-zero terms on each row or column. When $\rho \in [0,1)$, the matrix is diagonally dominant, hence positive-definite, so there exists indeed a Gaussian vector X with these covariances. We then have

$$V_{N,\delta} \geq e\,(1 - D(e^{\frac{\rho}{D}} - 1))\frac{N^{\frac{1}{3} - \frac{2}{2+\delta}}}{D^{\frac{4}{3}}},$$

so if $1 \ll D \ll N^{\frac{1}{4} - \varepsilon}$, then one can apply Theorem 4.12 to get

$$d_{\text{Kol}}(Y_N, \mathcal{N}_{\mathbb{R}}(0,1)) \leq 78\left(\frac{e^{\frac{3}{2\varepsilon}} - 1}{1 - \rho}\right)^{\frac{3}{2}}\left(\frac{D}{N^{\frac{1}{4} - \varepsilon}}\right)^{\frac{2}{2\varepsilon+1}}$$

for N large enough. Moreover, as soon as $1 \ll D \ll N^{\frac{1}{2} - \varepsilon}$, $Y_N \rightharpoonup \mathcal{N}_{\mathbb{R}}(0,1)$.

5. ISING MODEL AND MARKOV CHAINS

In this section, we present examples of random variables that admit uniform bounds on cumulants, which do not come from dependency graphs. Their structure is nevertheless not so different since the variables that we consider write as sums of random variables that are *weakly dependent*. The technique to prove uniform bounds on cumulants relies then on the notion of *uniform weighted dependency graph*, which generalizes the notion of standard dependency graph (see Proposition 5.3).

5.1. **Weighted graphs and spanning trees.** An *edge-weighted graph G*, or *weighted graph* for short, is a graph G in which each edge e is assigned a weight $w_G(e)$. Here we restrict ourselves to weights $w_G(e)$ with $w_G(e) \in \mathbb{R}_+$. Edges not in the graph can be thought of as edges of weight 0, all our definitions being consistent with this convention.

If B is a multiset of vertices of G, we can consider the graph $G[B]$ induced by G on B and defined as follows: the vertices of $G[B]$ correspond to elements of B (if B contains an element with multiplicity m, then m vertices correspond to this element), and there is an edge between two vertices if the corresponding vertices of G are equal or connected by an edge in G. This new graph has a natural weighted graph structure: put on each edge of $G[B]$ the weight of the corresponding edge in G (if the edge connects two copies of the same vertex of G, we put weight 1).

Definition 5.1. *A spanning tree of a graph $G = (V, E)$ is a subset E' of E such that (V, E') is a tree. In other words, it is a subgraph of G that is a tree and covers all vertices.*

The set of spanning trees of T is denoted $\mathrm{ST}(G)$. If G is a weighted graph, we say that the weight $w(T)$ of a spanning tree of G is defined as the *product* of the weights of the edges in T.
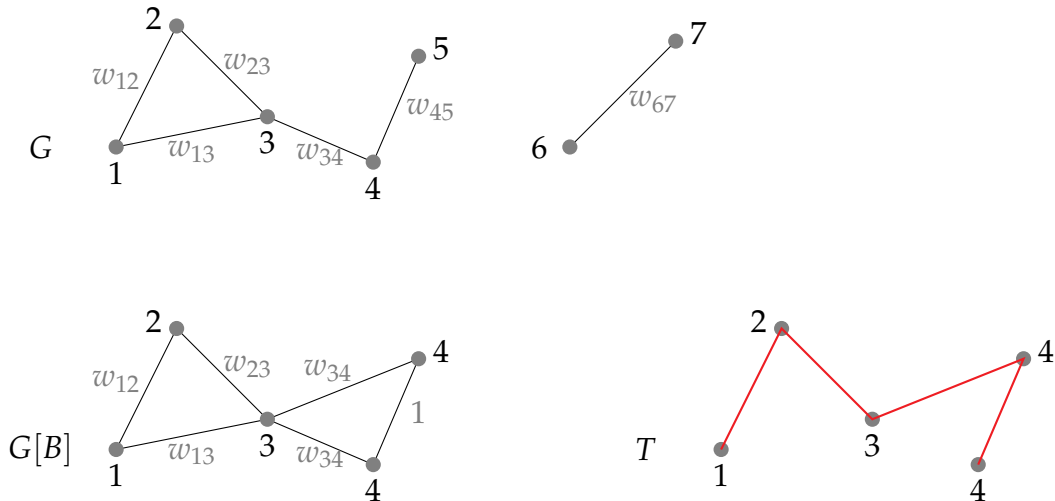

FIGURE 10. A weighted dependency graph G for 7 random variables; the induced graph $G[B]$ with $B = \{1, 2, 3, 4, 4\}$; and a spanning tree T of $G[B]$, with $w(T) = w_{12}w_{23}w_{34}$.

5.2. **Uniform weighted dependency graphs.** If $A_1, \ldots, A_r$ are real-valued random variables, there is a notion of joint cumulant that generalize the cumulants of Section 4:

$$\kappa(A_1, A_2, \ldots, A_r) = [z_1 z_2 \cdots z_r] \left(\log \mathbb{E}[e^{z_1 A_1 + z_2 A_2 + \cdots + z_r A_r}] \right).$$

The joint cumulants are multilinear and symmetric functionals of $A_1, A_2, \ldots, A_r$. On the other hand,

$$\kappa^{(r)}(X) = \kappa(X, X, \ldots, X)$$

with r occurrences of X in the right-hand side. In particular, if $S = \sum_{v \in V} A_v$ is a sum of random variables, then

$$\kappa^{(r)}(S) = \sum_{v_1,\ldots,v_r \in V} \kappa(A_{v_1}, A_{v_2}, \ldots, A_{v_r}).$$

Definition 5.2. *Let $(A_v)_{v \in V}$ be a family of random variables defined on the same probability space. A weighted graph $G = (V, E, w_G)$ is a C-uniform weighted dependency graph for $(A_v)_{v \in V}$ if, for any multiset $B = \{v_1, \ldots, v_r\}$ of elements of V, one has*

$$\left| \kappa(A_v, v \in B) \right| \leq C^{|B|} \sum_{T \in \mathrm{ST}(G[B])} w(T).$$

Proposition 5.3. *Let $(A_v)_{v \in V}$ be a finite family of random variables with a C-uniform weighted dependency graph G. Assume that G has $N = |V|$ vertices, and maximal weighted degree $D - 1$, that is:*

$$\forall v \in V, \quad \sum_{\substack{v': v' \neq v \\ \{v,v'\} \in E_G}} w_G(\{v, v'\}) \leq D - 1.$$

Then, for $r \geq 1$,

$$\left| \kappa^{(r)} \left(\sum_{v \in V} A_v \right) \right| \leq N \, r^{r-2} D^{r-1} \, C^r.$$

Consider a sequence $(S_n)_{n \in \mathbb{N}}$, where each S_n writes as $\sum_{v \in V_n} A_{v,n}$. Set $N_n = |V_n|$ and assume that, for each n, $(A_v)_{v \in V_n}$ has C-uniform weighted dependency graph of maximal degree $D_n - 1$ (by assumption, C does not depend on n). Then the sequence (S_n) admits uniform bounds on cumulants with parameters $(\frac{D_n}{2}, N_n, C)$ and the results of Section 4, in particular Corollary 4.3, apply.

Proof. By multilinearity and definition of a uniform weighted dependency graph, we have

$$\left| \kappa_r \left(\sum_{v \in V} A_v \right) \right| \leq C^r \sum_{v_1,\ldots,v_r} \sum_{T \in \mathrm{ST}(G[v_1,\ldots,v_r])} w(T). \tag{7}$$

By possibly adding edges of weight 0, we may assume that $G[v_1, \ldots, v_r]$ is always the complete graph K_r so that $\mathrm{ST}(G[v_1, \ldots, v_r]) \simeq \mathrm{ST}(K_r)$ as sets. The weight of a tree T however depends on $v_1, \cdots, v_r$, namely

$$w(T) = \prod_{\{i,j\} \in T} w_G(\{v_i, v_j\}),$$

where $w_G(\{v_i, v_j\})$ is the weight of the edge $\{v_i, v_j\}$ in G (or 1 if $v_i = v_j$).

With this viewpoint, we can exchange the order of summation in (7). We claim that the contribution of a fixed tree $T \in \mathrm{ST}(K_r)$ can then be bounded as follows:

$$\sum_{v_1,\ldots,v_r} w(T) = \sum_{v_1,\ldots,v_r} \prod_{\{i,j\} \in T} w_G(\{v_i, v_j\}) \leq ND^{r-1}. \tag{8}$$

Let us prove this claim by induction on r. The case $r = 1$ is trivial. Up to renaming the vertices of T, we may assume that r is a leaf of T so that T is obtained from a spanning

tree $\widetilde{T}$ of K_{r-1} by adding an edge $\{i_0, r\}$ for some $i_0 < r$. Then

$$\sum_{v_1,\dots,v_r} \prod_{\{i,j\}\in T} w_G(\{v_i, v_j\}) = \sum_{v_1,\dots,v_{r-1}} \left(\prod_{\{i,j\}\in\widetilde{T}} w_G(\{v_i, v_j\}) \right) \left[\sum_{v_r\in V} w_G(\{v_{i_0}, v_r\}) \right].$$

The expression in square brackets is by definition smaller than D for all v_{i_0} (the sum for $v_r \neq v_{i_0}$ is smaller than $D-1$ and the term for $v_r = v_{i_0}$ is 1). By induction hypothesis, the sum of the parenthesis is smaller than ND^{r-2}. This concludes the proof of (8) by induction. The lemma now follows immediately, since the number of spanning trees of K_r is well known to be r^{r-2}. □

Remark 5.4. A classical dependency graph with a uniform bound A on all variables A_v can be seen as a C-uniform weighted dependency graph for $C = 2A$ (all edges have weight 1); see [FMN16, Section 9.4]. In this case, Proposition 5.3 reduces to Theorem 4.7. The proof of Proposition 5.3 given here is a simple adaptation of the second part of the proof of Theorem 4.7 (see [FMN16, Chapter 9]) to the weighted context. The first, and probably the hardest part of the proof of Theorem 4.7 consisted in showing that a classical dependency graph is indeed a C-uniform weighted dependency graph.

Remark 5.5. In the case where the set V is a subset of $\mathbb{Z}^d$ and the weight function only depends on the distance, the notion of uniform weighted dependency graph coincides with the notion of *strong cluster properties*, proposed by Duneau, Iagolnitzer and Souillard in [DIS73]. These authors also observed that this implies uniform bounds on cumulants when D is bounded, see [DIS73, Eq. (10)].

Remark 5.6. A weaker notion of weighted dependency graph, where the bound on cumulant is not uniform on r, was recently introduced in [Fér16]. This weaker notion only enables to prove central limit theorem, without normality zone or speed of convergence results. However, it seems to have a larger range of applications.

5.3. **Magnetization of the Ising model.** We consider here the nearest-neighbour Ising model on a finite subset Λ of $\mathbb{Z}^d$ with a quadratic potential, *i.e.* for a *spin configuration* σ in $\{-1, +1\}^\Lambda$, its energy is given by

$$\mathcal{H}^\Lambda_{\beta,h}(w) = -\beta \sum_{\substack{i,j\in\Lambda \\ \{i,j\}\in E_{\mathbb{Z}^d}}} \sigma_i\sigma_j \; - h \sum_{i\in\Lambda} \sigma_i,$$

where $E_{\mathbb{Z}^d}$ is the set of edges of the lattice $\mathbb{Z}^d$ and h and β are two real parameters with $\beta > 0$. The probability $\mu_{\beta,h,\Lambda}[\sigma]$ of taking a configuration σ is then proportional to $\exp(-\mathcal{H}^\Lambda_{\beta,h}(\sigma))$.

We now want to make Λ grow to the whole lattice $\mathbb{Z}^d$. It is well known that for $h \neq 0$ or β smaller than a critical value $\beta_c(d)$ (thus, at high temperature), there is a unique limiting probability measure $\mu_{\beta,h}$ on spin configurations on $\mathbb{Z}^d$, see *e.g.* [FV16, Theorem 3.41]. In the following, we take parameters (β, h) in this domain of uniqueness and consider a random spin configuration σ, whose law is $\mu_{\beta,h}$.

In [DIS74], Duneau, Iagolnitzer and Souillard proved what they call the *strong cluster properties* for spins in the Ising model for $h \neq 0$ or sufficiently small β. Their result is actually more general (it holds for other models than the Ising model) but for simplicity, we stick to the Ising model here. Reformulated with the terminology of the present article, we have:

Theorem 5.7 (Duneau, Iagolnitzer and Souillard, 1974). *Fix the dimension* $d \geq 1$ *and* $h \neq 0$, $\beta > 0$.

> i) *There exist* $C = C(d, \beta, h)$ *and* $\varepsilon = \varepsilon(d, \beta, h) < 1$ *such that under the probability measure* $\mu_{\beta,h}$, *the family* $\{\sigma_i, i \in \mathbb{Z}^d\}$ *has a C-uniform weighted dependency graph G, where the weight of the edge* $\{i, j\}$ *in G is* $\varepsilon^{\|i-j\|_1}$.
>
> ii) *The same holds for* $h = 0$ *and* β *is sufficiently small (i.e.* $\beta < \beta_1(d)$, *for some* $\beta_1(d)$ *depending on the dimension; this is sometimes refered to as the very high temperature regime).*

Note that the maximal weighted degree of this graph is a constant $C' < \infty$.

We now consider the magnetization in a finite box Δ defined as $M_\Delta = \sum_{i \in \Delta} \sigma_i$. We see M_Δ as a sequence of random variables (indexed by the countably many finite subsets of $\mathbb{Z}^d$). Restricting the uniform weighted dependency graph above to $\{\sigma_i, i \in \Delta\}$, each M_Δ is the sum of random variables with a C-uniform weighted dependency graph and maximal weighted degree at most C'. Therefore, using Proposition 5.3, we know that $M_{|\Delta|}$ admits uniform bounds on cumulants with parameters $(\frac{C'}{2}, |\Delta|, C)$. Moreover, since all spins are positively correlated by the FKG inequality (see [FV16, Section 3.6]), we have, using translation invariance

$$\text{Var}(M_\Delta) \geq \sum_{i \in \Delta} \text{Var}(\sigma_i) = \text{Var}(\sigma_0)|\Delta|.$$

Note that $\text{Var}(\sigma_0)$ is independent of Δ. With the notation of Section 4, this inequality ensures that $\widetilde{\sigma}_\Delta$ is bounded from below by a constant. Applying Corollary 4.3, we get:

Proposition 5.8. *Fix the dimension* $d \geq 1$ *and parameters* $h \neq 0$, $\beta > 0$. *The exists a constant* $K = K(d, \beta, h)$ *such that, for all subsets* Δ *of* $\mathbb{Z}^d$, *we have under* $\mu_{\beta,h}$

$$d_{\text{Kol}}\left(\frac{M_\Delta - \mathbb{E}[M_\Delta]}{\sqrt{\text{Var}(M_\Delta)}}, \mathcal{N}_\mathbb{R}(0,1)\right) \leq \frac{K}{\sqrt{|\Delta|}}.$$

The same holds for $h = 0$ *and* β *sufficiently small (very high temperature).*

Remark 5.9. In this remark, we discuss mod-Gaussian convergence in this setting. Consider a sequence Δ_n of subsets of $\mathbb{Z}^d$, tending to $\mathbb{Z}^d$ in the Van Hove sense (*i.e.* the sequence is increasing with union $\mathbb{Z}^d$, and the size of the boundary of Δ_n is asymptotic negligible, compared to the size of Δ_n itself). Then it is known from [Ell85, Lemma 5.7.1] that

$$\lim_{n \to \infty} \frac{1}{|\Delta|} \text{Var}(M_\Delta) = \sum_{k \in \mathbb{Z}^d} \text{Cov}(\sigma_0, \sigma_k),$$

and the right-hand side has a finite value $\widetilde{\sigma}^2(\beta, h)$ for parameters (β, h) for which Theorem 5.7 applies. Similarly, we have

$$\lim_{n \to \infty} \frac{1}{|\Delta|} \kappa^{(3)}(M_\Delta) = \sum_{k,l \in \mathbb{Z}^d} \kappa(\sigma_0, \sigma_k, \sigma_l) < \infty.$$

We call $\rho(\beta, h)$ this quantity, and denote $L = \frac{\rho(\beta, h)}{\widetilde{\sigma}^3(\beta, h)}$.

Let us then consider the rescaled variables

$$X_n = \frac{M_{\Delta_n} - \mathbb{E}[M_{\Delta_n}]}{(\mathrm{Var}(M_{\Delta_n}))^{1/3}}.$$

From [FMN16, Section 5] (with $\alpha_n = \mathrm{Var}(M_\Delta)$ and $\beta_n = 1$), we know that X_n converges in the complex mod-Gaussian sense with parameters $t_n = (\mathrm{Var}(M_{\Delta_n}))^{1/3}$ and limiting function $\psi(z) = \exp(\frac{Lz^3}{6})$. This mod-Gaussian convergence takes place on the whole complex plane. Using the results of [FMN16], this implies a normality zone for $(M_\Delta - \mathbb{E}[M_\Delta])/\sqrt{\mathrm{Var}(M_\Delta)}$ of size $o(|\Delta|^{1/6})$, see Proposition 4.4.1 in *loc. cit.*; and moderate deviation estimates at the edge of this normality zone, see Proposition 5.2.1.

Remark 5.10. For $h = 0$ and $\beta > \beta_c(d)$ (low temperature regime), there is no weighted dependency graph as above. Indeed, this would imply the analycity of the partition function in terms of the magnetic field h, and the latter is known not to be analytic at $h = 0$ for $\beta > \beta_c(d)$; see [MM91, Chapter 6, §5] for details.

5.4. Functionals of Markov chains.

In this section, we consider a discrete time Markov chain $(M_t)_{t \geq 0}$ on a finite state space $\mathfrak{X}$, which is ergodic (irreducible and aperiodic) with invariant measure π. Its transition matrix is denoted P. To simplify the discussion, we shall also assume that the Markov chain is stationary, that is to say that the initial measure (*i.e.* the law of M_0) is π; most results have easy corollaries for any initial measure, using the fast mixing of such Markov chains.

Let us consider a sequence $(f_t)_{t \geq 0}$ of functions on $\mathfrak{X}$ that is uniformly bounded by a constant K. We set $Y_t = f_t(M_t)$. We will show that $\{Y_t\}_{t \in \mathbb{N}}$ admits a uniform weighted dependency graph. The proof roughly follows the one of [Fér16, Section 10], where it was proved that it has a (non-uniform) weighted dependency graph, taking extra care of the dependence in the order r of the cumulant in the bounds. Instead of working directly with classical (joint) cumulants, we start by giving a bound for the so-called *boolean cumulants*. Classical cumulants are then expressed in terms of boolean cumulants thanks to a formula of Saulis and Statulevičius [SS91, Lemma 1.1]; see also a recent article of Arizmendi, Hasebe, Lehner and Vargas [AHLV15] (we warn the reader that, in [SS91], boolean cumulants are called centered moments).

Let $Z_1, \ldots, Z_r$ be random variables with finite moments defined on the same probability space. By definition, their boolean (joint) cumulant is

$$B^{(r)}(Z_1, \ldots, Z_r) = \sum_{l=0}^{r-1} (-1)^l \sum_{1 \leq d_1 < \cdots < d_l \leq r-1} \mathbb{E}[Z_1 \cdots Z_{d_1}] \, \mathbb{E}[Z_{d_1+1} \cdots Z_{d_2}] \, \cdots \, \mathbb{E}[Z_{d_l+1} \cdots Z_r].$$

While not at first sight, this definition is quite similar to the definition of classical (joint) cumulants, replacing the lattice of all set partitions by the lattice of interval set partitions; see [AHLV15, Section 2] for details. Note however that, unlike classical cumulants, boolean cumulants are not symmetric functionals.

Proposition 5.11. *Let $r \geq 1$. With the above notation, there exists a constant θ_P depending on P with the following property. For any integers $t_1 \leq t_2 \leq \cdots \leq t_r$, we have*

$$\left| B^{(r)}(Y_{t_1}, \ldots, Y_{t_r}) \right| \leq M^{\frac{r-1}{2}} K^r (\theta_P)^{t_r - t_1},$$

where $M = |\mathfrak{X}|$.

The proof of this bound relies on arguments due to Diaconis, Stroock and Fill, see [DS91, Fil91]. We also refer to [SS91, Section 4.1] for an alternate approach. Given an ergodic transition matrix P on $\mathfrak{X}$ with invariant measure π, we denote $\widetilde{P}$ the time reversal of P, which is the stochastic matrix defined by the equation

$$\widetilde{P}(x, y) = \frac{\pi(y) \, P(y, x)}{\pi(x)}.$$

This new transition matrix is again ergodic, with stationary measure π. The multiplicative reversiblization of P is the matrix $M(P) = P\widetilde{P}$. It is a stochastic matrix, which is ergodic with stationary measure π, and with all its eigenvalues that are real and belong to $[0, 1]$. Indeed, if D is the diagonal matrix $D = \mathrm{diag}(\pi)$, then $\widetilde{P} = D^{-1}P^t D$, and

$$\mathrm{Spec}(M(P)) = \mathrm{Spec}\left(D^{1/2}P\widetilde{P}D^{-1/2} \right) = \mathrm{Spec}\left((D^{1/2}PD^{-1/2})(D^{-1/2}P^t D^{1/2}) \right)$$

$$= \mathrm{Spec}\left((D^{1/2}PD^{-1/2})(D^{1/2}PD^{-1/2})^t \right).$$

Thus, $M(P)$ has the spectrum of a symmetric positive matrix, so it belongs to $\mathbb{R}_+$, and in fact to $[0, 1]$ since $M(P)$ is also stochastic. We denote

$$(\theta_P)^2 = \max\{|z| \mid z \text{ eigenvalue of } M(P), z \neq 1\}. \tag{9}$$

Notice that if P is reversible, then $\widetilde{P} = P$ and $M(P) = P^2$, so in this case

$$\theta_P = \max\{|z| \mid z \text{ eigenvalue of } P, z \neq 1\}.$$

In general, one can think of θ_P as the analogue of the second largest eigenvalue for non-reversible transition matrices. The following result estimates the rate of convergence of the Markov chain associated to P in terms of θ:

Theorem 5.12 (Fill, 1991). *For any $x \in \mathfrak{X}$,*

$$\sum_{y \in \mathfrak{X}} |P^t(x, y) - \pi(y)| \leq \frac{(\theta_P)^t}{\sqrt{\pi(x)}};$$

$$\sum_{y \in \mathfrak{X}} \frac{|P^t(x, y) - \pi(y)|}{\sqrt{\pi(y)}} \leq \sqrt{M} \, \frac{(\theta_P)^t}{\sqrt{\pi(x)}}$$

where $M = |\mathfrak{X}|$.

Proof. For completeness, we reproduce here the discussion of [Fil91, Section 2], which relies on the following identity due to Mihail. If f is a function on $\mathfrak{X}$, we denote $\mathrm{Var}(f)$

its variance under the stationary probability measure π. We also introduce the Dirichlet form

$$\mathscr{E}(f,g) = \frac{1}{2} \sum_{x,y \in \mathfrak{X}} (f(x) - f(y))(g(x) - g(y))\, \pi(x)\, M(P)(x,y).$$

Then, for any function f,

$$\mathrm{Var}(f) = \mathrm{Var}(\widetilde{P}f) + \mathscr{E}(f,f).$$

Indeed, one can assume without loss of generality that $\pi(f) = \sum_{x \in \mathfrak{X}} \pi(x) f(x) = 0$. If $\langle f \mid g \rangle_\pi = \sum_{x \in \mathfrak{X}} \pi(x)\, f(x)\, g(x)$, then

$$\mathscr{E}(f,f) = \langle f \mid (\mathrm{id} - M(P))f \rangle_\pi = \langle f \mid f \rangle_\pi - \left\langle f \mid P\widetilde{P}f \right\rangle_\pi$$

$$= \langle f \mid f \rangle_\pi - \left\langle \widetilde{P}f \mid \widetilde{P}f \right\rangle_\pi = \mathrm{Var}(f) - \mathrm{Var}(\widetilde{P}f)$$

since $\widetilde{P}$ is the adjoint of P for the action on the left of functions and with respect to the scalar product $\langle \cdot \mid \cdot \rangle_\pi$.

Consider now a Markov chain $(X_t)_{t \in \mathbb{N}}$ on $\mathfrak{X}$ with arbitrary initial distribution π_0, and denote $\pi_t = \pi_0 P^t$ the distribution at time t. We introduce the quantity

$$(\chi_t)^2 = \sum_{y \in \mathfrak{X}} \frac{(\pi_t(y) - \pi(y))^2}{\pi(y)}.$$

This is the variance of $f_t = \frac{\pi_t}{\pi}$ with respect to the probability measure π. By Mihail's identity,

$$(\chi_{t+1})^2 = \mathrm{Var}(f_{t+1}) = \mathrm{Var}(\widetilde{P}f_t) = \mathrm{Var}(f_t) - \mathscr{E}(f_t, f_t) = (\chi_t)^2 - \mathscr{E}(f_t, f_t).$$

By the minimax characterization of eigenvalues of symmetric positive matrices,

$$(\theta_P)^2 = 1 - \inf \left\{ \frac{\mathscr{E}(f,f)}{\mathrm{Var}(f)},\ f \text{ non-constant} \right\}.$$

Therefore, $(\chi_{t+1})^2 \leq (\theta_P)^2 (\chi_t)^2$, and $(\chi_t)^2 \leq (\theta_P)^{2t} (\chi_0)^2$ by induction on t. On the other hand, the Cauchy-Schwarz inequality yields

$$\sum_{y \in \mathfrak{X}} |\pi_t(y) - \pi(y)| \leq \sqrt{\sum_{y \in \mathfrak{X}} \pi(y)} \sqrt{\sum_{y \in \mathfrak{X}} \frac{(\pi_t(y) - \pi(y))^2}{\pi(y)}} = \chi_t.$$

If we choose $\pi_0 = \delta_x$, we finally obtain:

$$\sum_{y \in \mathfrak{X}} |P^t(x,y) - \pi(y)| \leq (\theta_P)^t \chi_0 = (\theta_P)^t \sqrt{\frac{1}{\pi(x)} - 1} \leq \frac{(\theta_P)^t}{\sqrt{\pi(x)}}.$$

Similarly,

$$\sum_{y \in \mathfrak{X}} \frac{|P^t(x,y) - \pi(y)|}{\sqrt{\pi(y)}} \leq \sqrt{M}\, \chi_t \leq \sqrt{M}\, (\theta_P)^t \chi_0 \leq \sqrt{M}\, \frac{(\theta_P)^t}{\sqrt{\pi(x)}}. \qquad \square$$

Proof of Proposition 5.11. If f is a function on $\mathfrak{X}$, denote $D_f = \mathrm{diag}(f(x), x \in \mathfrak{X})$. Then, the boolean cumulant has the following matrix expression:

$$B^{(r)}(Y_{t_1}, \ldots, Y_{t_r}) = \pi \, D_{f_{t_1}} \, (P^{t_2 - t_1} - \mathbf{1} \, \pi) \, D_{f_{t_2}} \cdots D_{f_{t_{r-1}}} \, (P^{t_r - t_{r-1}} - \mathbf{1} \, \pi) \, D_{f_{t_r}} \, \mathbf{1},$$

where $\mathbf{1}$ is the column vector with all its entries equal to 1; see [Fér16, Lemma 10.1]. If we expand this expression as a sum, and denote $Q_t = P^t - \mathbf{1} \, \pi$ and $\delta_i = t_{i+1} - t_i$, then we obtain:

$$B^{(r)}(Y_{t_1}, \ldots, Y_{t_r}) = \sum_{x_1, \ldots, x_r} \pi(x_1) \, f_{t_1}(x_1) \, Q_{t_2 - t_1}(x_1, x_2) \, f_{t_2}(x_2) \cdots Q_{t_r - t_{r-1}}(x_{r-1}, x_r) \, f_{t_r}(x_r)$$

$$\left| B^{(r)}(Y_{t_1}, \ldots, Y_{t_r}) \right| \leq K^r \sum_{x_1, \ldots, x_r} \pi(x_1) \, |Q_{\delta_1}(x_1, x_2)| \cdots |Q_{\delta_{r-1}}(x_{r-1}, x_r)|$$

$$\leq K^r \, (\theta_P)^{\delta_{r-1}} \sum_{x_1, \ldots, x_{r-1}} \pi(x_1) \, |Q_{\delta_1}(x_1, x_2)| \cdots \frac{|Q_{\delta_{r-2}}(x_{r-2}, x_{r-1})|}{\sqrt{\pi(x_{r-1})}}$$

$$\vdots$$

$$\leq K^r \, M^{\frac{r-2}{2}} \, (\theta_P)^{t_r - t_1} \sum_{x_1} \sqrt{\pi(x_1)} \leq K^r M^{\frac{r-1}{2}} \, (\theta_P)^{t_r - t_1}. \qquad \square$$

Proposition 5.13. *The family of random variables $\{Y_t\}_{t \in \mathbb{N}}$ admits a $K\sqrt{M}$-uniform weighted dependency graph, where, for integers $s < t$, the weight between Y_t and Y_s is $2(\theta_P)^{t-s}$.*

Proof. A lemma of Saulis and Statulevičius [SS91, Lemma 1.1] expresses usual cumulants in terms of boolean ones:

$$\kappa^{(r)}(Y_{t_1}, \ldots, Y_{t_r}) = \sum_{\pi \in \mathcal{P}([r])} (-1)^{|\pi| - 1} N(\pi) \prod_{C \in \pi} B^{(|C|)}(Y_{t_j}, j \in C). \tag{10}$$

Here, the sum runs over set-partitions π of $[r] := \{1, \ldots, r\}$; $|\pi|$ is the number of blocks of a set-partition π, the product runs over blocks C in π and $B^{(|C|)}(Y_{t_j}, j \in C)$ is the boolean cumulant of the subfamily (Y_{t_j}) indexed by integers j in the block C, with the times ordered in increasing order (recall that the Boolean are not symmetric functionals). Finally $N(\pi)$ is a combinatorial factor that can be computed as follows. For each block C of π, denote m_C and M_C its smallest and biggest elements; then call n_C the number of blocks $C' \neq C$ such that m_C is in the interval $[m_{C'}; M_{C'}]$. We finally define $N(\pi) = \prod_{C \in \pi; 1 \notin C} n_C$. In other terms, $N(\pi)$ counts the functions g mapping each block C of π (except the one containing 1) to a block $g(C) \neq C$ such that $m_C \in [m_{g(C)}; M_{g(C)}]$.

Let us make an observation. If π is a partition and k an integer such that each block of π either contains only numbers smaller than or equal to k or only numbers bigger than k (π is then said to be *disconnected*), then no function g as above exists (there is no possible image for the block C containing $k + 1$) and $N(\pi) = 0$. On the other hand, for *connected* partitions π, we always have $N(\pi) > 0$, so that the sum in (10) is in effect a sum over connected partitions.

Eq. (10) and Proposition 5.11 imply the bound

$$\left| \kappa^{(r)}(Y_{t_1}, \ldots, Y_{t_r}) \right| \leq \left(K\sqrt{M} \right)^r \sum_{\pi \in \mathcal{P}([r])} N(\pi) \prod_{C \in \pi} (\theta_P)^{t_{M_C} - t_{m_C}}.$$

We would like to prove

$$\left| \kappa^{(r)}(Y_{t_1}, \ldots, Y_{t_r}) \right| \leq 2^{r-1} \left(K\sqrt{M} \right)^r \sum_{T \in \mathrm{ST}(K_r)} w(T),$$

where $w(T) = \prod_{\{j,j'\} \in E_T, j<j'} (\theta_P)^{t_{j'} - t_j}$. Therefore it is sufficient for us to find an injective mapping η from pairs (π, g) as above to edge-bicolored spanning trees $\widetilde{T}$ such that

$$w(\eta(\pi, g)) = \prod_{C \in \pi} (\theta_P)^{t_{M_C} - t_{m_C}}; \tag{11}$$

here, by convention, the weight $w(\widetilde{T})$ of a colored tree is the weight $w(T)$ of its uncolored version. In the following, we describe such a mapping, concluding the proof of the proposition.

Let (π, g) be a pair of objects as above: π is a set-partition of $[r]$ and g is function mapping each block C of π (except the one containing 1) to a block $g(C) \neq C$ such that $m_C \in [m_{g(C)}; M_{g(C)}]$. For each block C of π, we consider the set

$$S(C) = C \cup \{m_{C'}, C' \in g^{-1}(C)\}.$$

Let us call P_C the path with vertex-set $S(C)$, where the vertices are in increasing order along the path. We also color in blue (resp. in red) edges of this path whose extremity with smaller label is in C (resp. in $\{m_{C'}, C' \in g^{-1}(C)\}$).

As an example, take $\pi = \{C_1, \cdots, C_6\}$ with $C_1 = \{1,5,10\}$, $C_2 = \{2,11\}$, $C_3 = \{3,9\}$, $C_4 = \{4,6,13\}$, $C_5 = \{7,12\}$, $C_6 = \{8\}$. As function g, we take $g(C_2) = C_1$, $g(C_3) = C_1$, $g(C_4) = C_2$, $g(C_5) = C_1$ and $g(C_6) = C_4$. In this case, we get $S(C_1) = \{1,2,3,5,7,10\}$, $S(C_2) = \{2,4,11\}$, $S(C_4) = \{4,6,8,13\}$ and $S(C_i) = C_i$ for $i \in \{3,5,6\}$. The associated bicolored paths are drawn on Fig. 11.

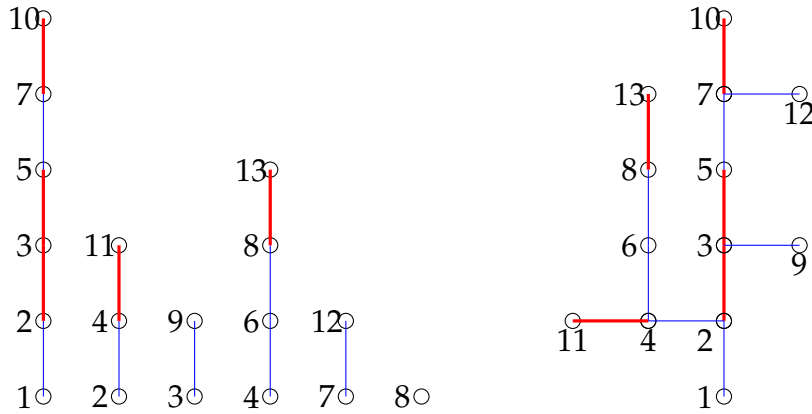FIGURE 11. Illustration of the construction η in the proof of Proposition 5.13: the path P_{C_i} and their gluing $\eta(\pi, g)$. For readers of a black-and-white printed version, red edges are thicker.

As on Fig. 11, we then take the union of the paths P_{C_i}, identifying vertices with the same label (the minimum $m_C \neq 1$ of a block C appears in the path $S(C)$ and in the path $S(g(C))$). Doing that, we get an edge-bicolored graph that we call $\eta(\pi, g)$. Let us first check that $\eta(\pi, g)$ is a tree. To this purpose, we order the blocks $C_1, \ldots, C_{|\pi|}$ of π in

increasing order of their minima (as done in the example). Observe that this implies that $g(C_i) = C_j$ for some $j < i$. We will prove by induction that, for each $i \leq |\pi|$, $P_{C_1} \cup \cdots \cup P_{C_i}$ is a tree. The case $i = 1$ is trivial. For $i > 1$, the graph $P_{C_1} \cup \cdots \cup P_{C_i}$ is obtained by gluing the path P_{C_i} on the graph $P_{C_1} \cup \cdots \cup P_{C_{i-1}}$, identifying m_{C_i} which appears in both. Since $P_{C_1} \cup \cdots \cup P_{C_{i-1}}$ is a tree by induction hypothesis, the resulting graph $P_{C_1} \cup \cdots \cup P_{C_i}$ is a tree as well, concluding the induction. Thus $\eta(\pi, g) = P_{C_1} \cup \cdots \cup P_{C_{|\pi|}}$ is a tree.

The equality (11) is easy: since the edge set of $\eta(\pi, g)$ is the union of the edge sets of the P_{C_i}, we have

$$w\big(\eta(\pi, g)\big) = \prod_{i=1}^{|\pi|} w(P_{C_i}) = \prod_{i=1}^{|\pi|} (\theta_P)^{t_{M_{C_i}} - t_{m_{C_i}}}.$$

We finally need to prove that η is injective, *i.e.* that we can recover (π, g) from $\eta(\pi, g)$. We start by the following easy observation: in $\eta(\pi, g)$, vertices with a red incident edge going to a vertex with bigger label are exactly the vertices with a label which is the minimum $m_C \neq 1$ of a block C of π. By construction, such vertices have at most three incident edges, which are as follow:

(E1) as said above, a first one is red and goes to a vertex to bigger label;

(E2) a second one is either blue or red and goes to a vertex of lower label.

(E3) possibly, a last one is blue and goes to a vertex to bigger label (there is such an edge when m_C is not alone in its block);

Indeed, in the construction, edges (E1) and (E2) comes from $S_{g(C)}$ while edge (E3) comes from S_C. We split the vertex m_C into two, keeping edges (E1) and (E2) in the same component. Doing that for the $|\pi| - 1$ vertices $m_C \neq 1$, we inverse the gluing step of the construction of $\eta(\pi, g)$. It is then straightforward to recover (π, g). □

Theorem 5.14. *Let $(X_t)_{t \in \mathbb{N}}$ be an ergodic Markov chain on a finite state space $\mathfrak{X}$ of size M, and $\theta_P < 1$ the constant associated by (9) with the transition matrix P. We consider a sum $S_n = \sum_{t=1}^n f_t(X_t)$ with $\|f_t\|_\infty \leq K$. Then, for any $r \geq 1$,*

$$\left| \kappa^{(r)}(S_n) \right| \leq n \, r^{r-2} \left(2 \frac{1 + \theta_P}{1 - \theta_P} \right)^{r-1} \left(K\sqrt{M} \right)^r. \tag{12}$$

As a consequence:

i) *When $\frac{\mathrm{Var}(S_n)}{n^{2/3}} \to +\infty$, we have*

$$d_{\mathrm{Kol}} \left(\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}, \mathcal{N}_{\mathbb{R}}(0, 1) \right) \leq 76.36 \left(\frac{K\sqrt{M}}{\sqrt{\frac{\mathrm{Var}(S_n)}{n}}} \right)^3 \left(\frac{1 + \theta_P}{1 - \theta_P} \right)^2 \frac{1}{\sqrt{n}},$$

and in particular, $\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}$ converges in law to a standard Gaussian.

ii) *This convergence in law happens as soon as $\frac{\mathrm{Var}(S_n)}{n^\varepsilon} \to \infty$ for some $\varepsilon > 0$.*

Proof. Combining Propositions 5.13 and 5.3, the sum S_n admits uniform bounds on cumulants with parameters

$$D_n = \left(1 + 2\sum_{t=1}^{\infty}(\theta_P)^t\right) = \frac{1 + \theta_P}{1 - \theta_P},$$

$N_n = n$ and $A = K\sqrt{M}$. Observe that D_n is here independent of n. We can apply Corollary 4.3 to get the first part of the theorem. The second follows from Theorem 4.5. □

Remark 5.15. A bound similar to Eq. (12) is given in [SS91, Theorem 4.19]. We believe however that the proof given there is not correct. Indeed, the proof associates with each partition π such that $N(\pi) \neq 0$ a sequence of number q_j; the authors then claim that "obviously $q_j \leq q_{j+1}$" (p. 93, after eq. (4.53)). This is unfortunately not the case as can be seen on the example of partitions given p. 81 in *loc. cit.*: for this partition $q_3 = 3$, while $q_4 = 2$. As a consequence of this mistake, the authors forget many partitions π when expressing classical cumulants in terms of boolean ones (since they encode partitions with only non-decreasing sequences q_i), which make the resulting bound on classical cumulants too sharp.

We have not found a simpler way around this error, than the use of uniform weighted dependency graphs, presented here. Note nevertheless that our proof still uses several ingredients from [SS91]: the use of boolean cumulants and the relation between boolean and classical cumulants.

Remark 5.16. If the functions f_t are indicators $f_t(x) = 1_{x=s_t}$, then one can remove the size $(\sqrt{M})^3$ in the bound on the Kolmogorov distance. Indeed, in this case, we have

$$B^{(r)}(Y_{t_1}, \ldots, Y_{t_r}) = \pi(s_1)\, Q_{t_2-t_1}(s_1, s_2)Q_{t_3-t_2}(s_2, s_3)\cdots Q_{t_r-t_{r-1}}(s_{r-1}, s_r).$$

On the other hand, the individual terms of the matrix $Q_t(x, y)$ can be bounded by

$$|Q_t(x, y)| \leq \sqrt{\frac{\pi(y)}{\pi(x)}}(\theta_P)^t,$$

by adapting the proof of Theorem 5.12. Therefore,

$$\left|B^{(r)}(Y_{t_1}, \ldots, Y_{t_r})\right| \leq (\theta_P)^{t_r-t_1}\, \pi(s_1)\, \sqrt{\frac{\pi(s_2)}{\pi(s_1)}}\cdots\sqrt{\frac{\pi(s_r)}{\pi(s_{r-1})}}$$

$$\leq (\theta_P)^{t_r-t_1}\sqrt{\pi(s_1)\pi(s_r)} \leq (\theta_P)^{t_r-t_1}.$$

Thus, in this case, one has the bound of Theorem 5.14 without the factor $(\sqrt{M})^3$.

5.5. The case of linear functionals of the empirical measure.

As a particular case of Theorem 5.14, one recovers the central limit theorem for linear functionals of empirical measures of Markov chains, that are random variables

$$Y_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(f(X_t) - \pi(f))$$

with $f : \mathfrak{X} \to \mathbb{R}$ fixed function (independent of the time t). Thus, assuming for instance $\lim_{n\to\infty} \mathrm{Var}(Y_n) = \Sigma^2(f) > 0$, we have

$$d_{\mathrm{Kol}} \left(\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} , \mathcal{N}_{\mathbb{R}}(0,1) \right) \leq 77 \left(\frac{\|f\|_\infty \sqrt{M}}{\Sigma(f)} \right)^3 \left(\frac{1+\theta_P}{1-\theta_P} \right)^2 \frac{1}{\sqrt{n}} \qquad (13)$$

for n large enough. We refer to [Cog72, KV86, Jon04, Häg05] and the references therein for the general background of this Markovian CLT, and to [Bol80, Lez96, Man96] for estimates of the Kolmogorov distance. It seems that we recover some results of [Man96] (see [SC97, §2.1.3]), but we were not able to find and read this paper. In this last paragraph, we discuss the problem of the variance $\mathrm{Var}(Y_n)$, giving sufficient conditions, which are simple to check on examples and ensure $\Sigma^2(f) > 0$. We also remark that, provided that $\Sigma^2(f) > 0$, we can also prove complex mod-Gaussian convergence, which implies a zone of normality result and moderate deviation estimates by [FMN16].

Denote $g = f - \pi(f)$, which has expectation 0 under the stationary measure π. By eventually replacing f with g, we can assume that f is centered. The variance of Y_n tends to

$$\Sigma^2(f) = \mathbb{E}[(f(X_0))^2] + 2 \sum_{t=1}^\infty \mathbb{E}[f(X_0)\,f(X_t)] < +\infty,$$

see [Cog72, Lemma 3.3]. If $\Sigma^2(f) > 0$, then $\frac{\mathrm{Var}(S_n)}{n^{2/3}} = n^{1/3}\,\mathrm{Var}(Y_n) \to +\infty$ and Theorem 5.14 applies. Unfortunately, one can easily construct non-trivial examples with $\Sigma^2(f) = 0$. Thus, consider the Markov chain with 3 states and transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{pmatrix};$$

it admits for invariant measure $\pi(1) = \pi(2) = \frac{2}{5}$ and $\pi(3) = \frac{1}{5}$. Set $f(1) = 1$, $f(2) = -1$ and $f(3) = 0$; one has $\pi(f) = 0$, and one computes

$$\mathbb{E}[f(X_0)f(X_k)] = \frac{1}{5} \left(\frac{2+\mathrm{i}}{(-1-\mathrm{i})^k} + \frac{2-\mathrm{i}}{(-1+\mathrm{i})^k} \right).$$

It follows that $\Sigma^2(f) = 0$, although f is non zero.

In the general case of an ergodic Markov chain, fix an element a of the state space $\mathfrak{X}$, and denote $\tau_a \geq 1$ the first hitting time of a by the Markov chain, which is almost surely finite and with expectation $1/\pi(a)$ when starting from a. Then, the asymptotic variance $\Sigma^2(f)$ can be rewritten as

$$\Sigma^2(f) = \pi(a)\,\mathbb{E}_a \left[\left(\sum_{k=1}^{\tau_a} g(X_k) \right)^2 \right],$$

see [KS76, Chapter 4]. Therefore, a general condition in order to obtain the bound of Eq. (13) is:

Proposition 5.17. *We have $\Sigma^2(f) > 0$ if and only if there exists a cycle $(x_1,\ldots,x_n)$ in the graph of transitions of the Markov chain such that the sum $\sum_{i=1}^n g(x_i)$ of the values of $g = f - \pi(f)$ along this cycle is non-zero. In this case, the bound (13) holds.*

The proposition explains readily why the irreducible aperiodic Markov chain
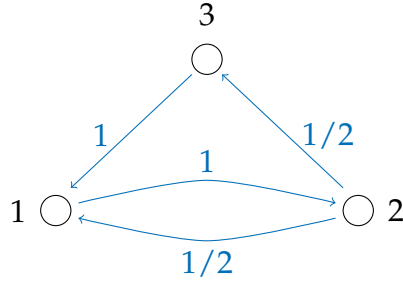
previously studied gives asymptotic variance 0 to the function $f(1) = 1$, $f(2) = -1$ and $f(3) = 0$: the minimal cycles of the chains are $(1,2)$ and $(1,2,3)$, and the sum of the values of f along these cycles is always 0.

Another simple criterion to apply Theorem 5.14 to linear functionals of the empirical measure is:

Proposition 5.18. *Suppose that the ergodic Markov chain is reversible:*

$$\pi(x)\, P(x,y) = \pi(y)\, P(y,x)$$

for any x, y. Then, if f is a non-constant function, $\Sigma^2(f) > 0$ and the bound (13) holds.

Proof. To say that P is reversible is equivalent to the fact that P is a symmetric operator of the Hilbert space $L^2(\frac{1}{\pi})$. In particular, P has only real eigenvalues. Besides, the restriction of the operator $I + 2\sum_{k=1}^{\infty} P^k$ to the space of functions f with $\pi(f) = 0$ is well defined, and it is an auto-adjoint operator on this space with eigenvalues

$$\frac{1+\lambda_2}{1-\lambda_2}, \dots, \frac{1+\lambda_M}{1-\lambda_M},$$

where $\lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_M$ are the real eigenvalues of P different from 1. The quantities above are all positive, and larger than $\frac{1-\theta_P}{1+\theta_P}$ (this value being obtained if $\lambda_M = -\theta_P$). Therefore,

$$\Sigma^2(f) = \left\langle f \,\middle|\, \left(I + 2\sum_{k=1}^{\infty} P^k\right) f \right\rangle_{L^2(\pi)} \geq \frac{1-\theta_P}{1+\theta_P}\, \pi(f^2) > 0.$$

We then obtain

$$d_{\mathrm{Kol}}\left(\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}, \mathcal{N}_{\mathbb{R}}(0,1)\right) \leq 77 \left(\frac{\|g\|_\infty \sqrt{M}}{\sqrt{\pi(g^2)}}\right)^3 \left(\frac{1+\theta_P}{1-\theta_P}\right)^{\frac{7}{2}} \frac{1}{\sqrt{n}}$$

for n large enough. $\qquad\square$

Remark 5.19. In this remark, we discuss mod-Gaussian convergence for linear statistics of Markov chains. We use the above notation and assume that $\Sigma^2(f) > 0$. Consider the third cumulant of S_n. One can easily prove that

$$\rho = \frac{1}{n} \lim_{n\to\infty} \kappa_3(S_n) = \sum_{j,k\in\mathbb{Z}} \kappa(f(X_0, X_j, X_k)),$$

where $(X_t)_{t\in\mathbb{Z}}$ is a bi-infinite stationary Markov chain with transition matrix P. (The sum on the right-hand side is finite, as consequence of Proposition 5.13). Let us call ρ this limit. We then consider the rescaled random variables

$$X_n = \left(\frac{S_n - \mathbb{E}[S_n]}{(\mathrm{Var}(S_n))^{1/3}} \right)_{n\in\mathbb{N}}.$$

As for the magnetization in the Ising model, from [FMN16, Section 5] (with $\alpha_n = \mathrm{Var}(S_n)$ and $\beta_n = 1$), we know that X_n converges in the mod-Gaussian sense with parameters $t_n = (\mathrm{Var}(S_n))^{1/3}$ and limit $\psi(z) = \exp(\frac{Lz^3}{6})$, with $L = \frac{\rho}{\Sigma^3(f)}$. Again, this mod-Gaussian convergence takes place on the whole complex plane and implies a normality zone for $S_n/\sqrt{\mathrm{Var}(S_n)}$ of size $o(n^{1/6})$, see Proposition 4.4.1 in *loc. cit.*; we also have moderate deviation estimates at the edge of this normality zone, see Proposition 5.2.1. This mod-Gaussian convergence could also have been proved by using an argument of the perturbation theory of linear operators, for which we refer to [Kat80]. Indeed, the Laplace transform of X_n writes explicitly as

$$\mathbb{E}[e^{zX_n}] = \pi \left(P_{z,f}\right)^n 1,$$

where 1 is the column vector with all its entries equal to 1, π is the stationary measure of the process, and $P_{z,f}$ is the infinitesimal modification of the transition matrix defined by

$$P_{z,f}(i,j) = P(i,j)\, e^{\frac{z(f(j)-\pi(f))}{(\mathrm{Var}(S_n))^{1/3}}}.$$

For z in a zone of control of size $O(n^{1/3})$, one can give a series expansion of the eigenvalues and eigenvectors of $P_{z,f}$, which allows one to recover the mod-Gaussian convergence. The theory of cumulants and weighted dependency graphs allows one to bypass these difficult analytic arguments.

References

[AHLV15] O. Arizmendi, T. Hasebe, F. Lehner, and C. Vargas. Relations between cumulants in noncommutative probability. *Adv. Math.*, 282:56 – 92, 2015.

[AS08] N. Alon and J. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, 2008. 3rd edn.

[Ber41] A. C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.*, 49(1):122–136, 1941.

[BHNY08] P. Bourgade, C. Hughes, A. Nikeghbali, and M. Yor. The characteristic polynomial of a random unitary matrix: a probabilistic approach. *Duke Math. J.*, 145:45–69, 2008.

[Bil95] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. Wiley, 3rd edition, 1995.

[BKN09] A. D. Barbour, E. Kowalski, and A. Nikeghbali. Mod-discrete expansions. *Probab. Th. Rel. Fields*, 158(3):859–893, 2009.

[BKR89] A. D. Barbour, M. Karoński, and A. Ruciński. A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B*, 47(2):125–145, 1989.

[Bol80] E. Bolthausen. The Berry–Esseen theorem for functionals of discrete Markov chains. *Z. Wahr. verw. Geb.*, 54:59–73, 1980.

[Bón10] M. Bóna. On three different notions of monotone subsequences. In *Permutation Patterns*, volume 376 of *London Math. Soc. Lecture Note Series*, pages 89–113. Cambridge University Press, 2010.

[BR89]   P. Baldi and Y. Rinott. On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.*, 17(4):1646–1650, 1989.

[Bul96]   A. V. Bulinskii. Rate of convergence in the central limit theorem for fields of associated random variables. *Th. Prob. Appl.*, 40(1):136–144, 1996.

[CDMN15]   R. Chhaibi, F. Delbaen, P.-L. Méliot, and A. Nikeghbali. Mod-ϕ convergence: Approximation of discrete measures and harmonic analysis on the torus. `arXiv:1511.03922`, 2015.

[Cog72]   R. Cogburn. The central limit theorem for Markov processes. In *Proceedings of the Sixth Annual Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 485–512. University of California Press, 1972.

[DIS73]   M. Duneau, D. Iagolnitzer, and B. Souillard. Decrease properties of truncated correlation functions and analyticity properties for classical lattices and continuous systems. *Comm. Math. Phys.*, 31(3):191–208, 1973.

[DIS74]   M. Duneau, D. Iagolnitzer, and B. Souillard. Strong cluster properties for classical systems with finite range interaction. *Comm. Math. Phys.*, 35:307–320, 1974.

[DKN15]   F. Delbaen, E. Kowalski, and A. Nikeghbali. Mod-ϕ convergence. *Intern. Math. Res. Not.*, 2015(11):3445–3485, 2015.

[DS91]   P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1(1):36–61, 91.

[Ell85]   R. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Springer, 1985.

[Fel71]   W. Feller. *An Introduction to Probability Theory and Its Applications, Volume II*. Wiley Series in Probability and Mathematical Statistics. Wiley, 2nd edition, 1971.

[Fér16]   V. Féray. Weighted dependency graphs. arXiv preprint 1605.03836, 2016.

[Fil91]   J. A. Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.*, 1(1):62–87, 1991.

[FMN16]   V. Féray, P.-L. Méliot, and A. Nikeghbali. *Mod-ϕ convergence: Normality zones and precise deviations*. SpringerBriefs in Probability and Mathematical Statistics. Springer, 2016.

[FV16]   S. Friedli and Y. Velenik. *Statistical Mechanics of Lattice Systems: a Concrete Mathematical Introduction*. Cambridge University press, 2016. In press.

[Gri92]   G. Grimmett. Weak convergence using higher-order cumulants. *J. Theoretical Probability*, 5(4):767–773, 1992.

[GW16]   L. Goldstein and N. Wiroonsri. Stein's method for positively associated random variables with applications to Ising, percolation and voter models. preprint arXiv:1603.05322, to appear in *Ann. Inst. Henri Poincaré*, 2016.

[Häg05]   O. Häggström. On the central limit theorem for geometrically ergodic Markov chains. *Probab. Th. Rel. Fields*, 132:74–82, 2005.

[Jan88]   S. Janson. Normal convergence by higher semi-invariants with applications to sums of dependent random variables and random graphs. *Ann. Probab.*, 16:305–312, 1988.

[JKN11]   J. Jacod, E. Kowalski, and A. Nikeghbali. Mod-Gaussian convergence: new limit theorems in probability and number theory. *Forum Mathematicum*, 23(4):835–873, 2011.

[Jon04]   G. L. Jones. On the Markov chain central limit theorem. *Probability Surveys*, 1:299–320, 2004.

[Kat80]   T. Kato. *Perturbation Theory for Linear Operators*. Classics in Mathematics. Springer-Verlag, 2nd edition, 1980.

[KN10]   E. Kowalski and A. Nikeghbali. Mod-Poisson convergence in probability and number theory. *Internat. Math. Res. Not.*, 18:3549–3587, 2010.

[KN12]   E. Kowalski and A. Nikeghbali. Mod-Gaussian distribution and the value distribution of $\zeta(\frac{1}{2} + it)$ and related quantities. *J. London Math. Soc.*, 86(2):291–319, 2012.

[KRT15]   K. Krokowski, A. Reichenbachs, and C. Thäle. Discrete Malliavin–Stein method: Berry–Esseen bounds for random graphs, point processes and percolation. `arXiv:1503.01029v1 [math.PR]`, 2015.

[KS76]   J. G. Kemeny and J. L. Snell. *Finite Markov Chains*. Undergraduate Texts in Mathematics. Springer-Verlag, 1976.

[KS00]   J. P. Keating and N. C. Snaith. Random matrix theory and $\zeta(\frac{1}{2} + it)$. *Comm. Math. Phys.*, 214(1):57–89, 2000.

[KS10]   V. Y. Korolev and I. G. Shevtsova. On the upper bound for the absolute constant in the Berry–Esseen inequality. *Theory of Probability and its Applications*, 54(4):638–658, 2010.

[KV86]   C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104:1–19, 1986.

[Lan93]   S. Lang. *Real and Functional Analysis*, volume 142 of *Graduate Texts in Mathematics*. Springer-Verlag, 3rd edition, 1993.

[Lez96]   P. Lezaud. *Chernoff-type bound for finite Markov chains*. PhD thesis, Université Paul Sabatier, Toulouse, 1996.

[LS59]   V. P. Leonov and A. N. Shiryaev. On a method of calculation of semi-invariants. *Theory Prob. Appl.*, 4:319–329, 1959.

[Mal95]   P. Malliavin. *Integration and Probability*, volume 157 of *Graduate Texts in Mathematics*. Springer-Verlag, 1995.

[Man96]   B. Mann. *Berry–Esseen central limit theorems for Markov chains*. PhD thesis, Harvard University, 1996.

[MM91]   V. A. Malyshev and R. A. Minlos. *Gibbs random fields*. Springer, 1991.

[MN15]   P.-L. Méliot and A. Nikeghbali. Mod-Gaussian convergence and its applications for models of statistical mechanics. In *In Memoriam Marc Yor - Séminaire de Probabilités XLVII*, volume 2137 of *Lecture Notes in Mathematics*, pages 369–425. Springer-Verlag, 2015.

[New80]   C. M. Newman. Normal fluctuations and the FKG inequalities. *Comm. Math. Phys.*, 74(2):119–128, 1980.

[PY05]   M. Penrose and J. Yukich. Normal approximation in geometric probability. *Stein's Method and Applications, Lecture Note Series, Institute for Mathematical Sciences, National University of Singapore*, 5:37–58, 2005.

[Rin94]   Y. Rinott. On normal approximation rates for certain sums of dependent random variables. *J. Comp. Appl. Math.*, 55(2):135 – 143, 1994.

[Rud91]   W. Rudin. *Functional Analysis*. McGraw-Hill, 2nd edition, 1991.

[Sat99]   K.-I. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 1999.

[SC97]   L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Mathematics*, pages 301–413. Springer-Verlag, 1997.

[Spi58]   F. Spitzer. Some theorems concerning 2-dimensional Brownian motion. *Trans. Amer. Math. Soc.*, 87:187–197, 1958.

[SS91]   L. Saulis and V. A. Statulevičius. *Limit theorems for large deviations*, volume 73 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers, 1991.

[Sta66]   V. A. Statulevičius. On large deviations. *Prob. Th. Rel. Fields*, 6(2):133–144, 1966.

[Str11]   D. W. Stroock. *Probability Theory: An Analytic View*. Cambridge University Press, 2nd edition, 2011.

[SY83]   K.-I. Sato and M. Yamazato. Stationary processes of ornstein-uhlenbeck type. In K. Itô and J. V. Prokhorov, editors, *Probability Theory and Mathematical Statistics, Fourth USSR-Japan Symp.*, volume 1021 of *Lecture Notes in Mathematics*, pages 541–551. Springer-Verlag, 1983.

[Tao12]   T. Tao. *Topics in Random Matrix Theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012.

(V. Féray) INSTITUT FÜR MATHEMATIK, UNIVERSITÄT ZÜRICH — WINTERTHURERSTRASSE 190, 8057 ZÜRICH, SWITZERLAND
   E-mail address: valentin.feray@math.uzh.ch

(P.-L. Méliot) LABORATOIRE DE MATHÉMATIQUES, BÂTIMENT 425 — FACULTÉ DES SCIENCES D'ORSAY — UNIVERSITÉ PARIS-SUD, 91400 ORSAY, FRANCE
   E-mail address: pierre-loic.meliot@math.u-psud.fr

(A. Nikeghbali) INSTITUT FÜR MATHEMATIK, UNIVERSITÄT ZÜRICH — WINTERTHURERSTRASSE 190, 8057 ZÜRICH, SWITZERLAND
   E-mail address: ashkan.nikeghbali@math.uzh.ch