Effective limit theorems for Markov chains with a spectral gap

Benoît Kloeckner

Effective limit theorems for Markov chains with a spectral gap

Benoît R. Kloeckner *

November 30, 2017

Applying quantitative perturbation theory for linear operators, we prove non-asymptotic limit theorems for Markov chains whose transition kernel has a spectral gap in an arbitrary Banach algebra of functions $\mathscr{X}$. The main results are concentration inequalities and Berry-Esseen bounds, obtained assuming neither reversibility nor "warm start" hypothesis: the law of the first term of the chain can be arbitrary. The spectral gap hypothesis is basically a uniform $\mathscr{X}$-ergodicity hypothesis, and when $\mathscr{X}$ consist in regular functions this is weaker than uniform ergodicity. We show on a few examples how the flexibility in the choice of function space can be used. The constants are completely explicit and reasonable enough to make the results usable in practice, notably in MCMC methods.

1 Introduction

General framework. Let $(X_k)_{k\geq 0}$ be a Markov chain taking value in a general state space Ω, and let $\varphi : \Omega \to \mathbb{R}$ be a function (the "observable"). Under rather general assumptions, there is a unique stationary measure μ_0 and it can be proved that almost surely[1]

$$\frac{1}{n}\sum_{k=1}^{n} \varphi(X_k) \to \mu_0(\varphi) \tag{1}$$

Then a natural question is to ask at what speed this convergence occurs. In many cases, one can prove a Central Limit Theorem, showing that the convergence has the order

*Université Paris-Est, Laboratoire d'Analyse et de Matématiques Appliquées (UMR 8050), UPEM, UPEC, CNRS, F-94010, Créteil, France

[1]Here and in the sequel, we write indifferently $\mu(f)$ or $\int f \,\mathrm{d}\mu$ for the integral of f with respect to the measure μ.

$1/\sqrt{n}$. But this is again an asymptotic result, and one is led to ask for non-asymptotic bounds, both for the Law of Large Numbers (1) ("concentration inequalities") and for the CLT ("Berry-Esseen bounds").

A word on effectivity. In this paper, the emphasis will be on *effective* bounds, i.e. given an explicit sample size n, one should be able to deduce from the bound that the quantity being considered lies in some explicit interval around its limit with at least some explicit probability. In other words, the result should be *non-aymptotic* and all constants should be made explicit. The motivations for this are at least twofold.

First, in practical applications of the Markov chain Monte-Carlo (MCMC) method, where one uses (1) to estimate the integral $\mu_0(\varphi)$, effective results are needed to obtain proven convergence of a given precision. MCMC methods are important when the measure of interest is either unknown, or difficult to sample independently (e.g. uniform in a convex set in large dimension), but happens to be the stationary measure for an easily simulated Markov chain. The Metropolis-Hastings algorithm for example makes it possible to deal with an absolutely continuous measure whose density is only known up to the normalization constant.

A second, more theoretical motivation is that the constants appearing in limit theorem depend on a number of parameters (e.g. the mixing speed of the Markov chain, the law of X_0, etc.). When the constants are not made explicit, one may not be able to deduce from the result how the convergence speed changes when some parameter approaches the limit of the domain where the result is valid (e.g. when the spectral gap tend to 0).

There are very many works proving concentration inequalities and (to a lesser extent) Berry-Esseen bounds for Markoc chains, under a variety of assumptions, and we will only mention a small number of them. To explain the purpose of this article, let us discuss briefly three directions.

Previous works (1): total variation convergence. The first direction is mainly motivated by MCMC; we refer to [RR+04] for a detailed introduction to the topic.

The Markov chains being considered are usually ergodic (either *uniformly*, which in the setting of this paper corresponds to a spectral gap on L^∞, or *geometrically*); one measures difference between probability measure using the *total variation distance*, and the limit theorems are typically obtained for L^∞ observables φ (the emphasis here is not on the boundedness, but on the lack of regularity assumption). Effective concentration inequalities have been obtained in this setting, for example in [GO02] and [KLMM05] which we shall discuss below. Berry-Esseen bounds have been proved in [Bol82], but effective results are less common.

Previous works (2): the spectral method. The second direction grew from the "Nagaev method" [Nag57, Nag61], a functional approach where perturbative spectral theory enables one to adapt the classical Fourier proofs of limit theorems, from independent identically distributed random variable to suitable Markov chains. This approach is described in [HH01] in a quite general setting, and is especially popular in dynamical

systems (the statistical properties of certain dynamical systems can be studied more easily by reversing time, and considering a Markov chain jumping randomly along *backward* orbits).

There, the Markov chain being considered are often *not* ergodic in the total variation sense, but instead their transition kernel has a spectral gap in a space $\mathscr{X}$ made of regular functions (one sometimes say such a Markov chain is $\mathscr{X}$-ergodic). The limit theorems are then restricted to observables $\varphi \in \mathscr{X}$, and the speed of convergence is driven by the regularity of φ as much as by its magnitude. Due to the use of perturbation theory of operator, in most cases this method has not yielded *effective* results.

Note that the spectral method *can* be applied without regularity assumptions, taking e.g. $\mathscr{X} = L^2(\mu_0)$ or $\mathscr{X} = L^\infty(\Omega)$ (or variants, see [KM12]), thus the present direction intersects the previous one.

There are (at least) two exceptions to the aforementioned lack of effectiveness. When $\mathscr{X}$ is a Hilbert space, by symetrization of the transition kernel one can use well-known effective perturbation results. In this way, Lezaud obtains effective concentration inequalities and Berry-Esseen bounds [Lez98, Lez01], see also [Pau15]. Both work in $L^2(\mu_0)$, restricting accordingly the Markov chains that can be considered. Second Dubois [Dub11] gave what seems to be the first effective Berry-Esseen inequality in a dynamical context, and we shall compare the present Berry-Esseen inequality with his.

Previous works (3): optimal transportation. The third direction is quite recent: Joulin and Ollivier [JO10] used ideas from optimal transportation to prove very efficiently effective concentration results under a *positive curvature* hypothesis; this corresponds to a spectral gap with constant 1 on the space $\mathscr{X} = \mathrm{Lip}$ of Lipschitz functions. Paulin [Pau16] extended this method to a spectral gap "with arbitrary constant" in the terminology set up below.

This method is very appealing, but is restricted to a single, pretty restrictive function space constraining both the Markov chains and the observables that can be considered; we will see in examples below that being able to change the function space can be useful to get good constants even when [JO10] can be applied. Moreover, this method seems unable to provide higher-order limit theorem such as the CLT or Berry-Esseen bounds.

Contributions of this work. The goal of this article is to combine recent *effective* perturbation results [Klo17b] with the Nagaev method to obtain effective concentration inequalities and Berry-Esseen bounds for a wealth of Markov chains. Our main hypothesis will be a spectral gap on some function space $\mathscr{X}$, with the sole restriction that we need $\mathscr{X}$ to be a Banach *algebra* (this will in particular restrict us to bounded observables). We obtain three main results:

- a general concentration inequality (Theorem A),
- a variant which, under a bound on the *dynamical variance* of $(\varphi(X_k))_{k\geq 0}$, gives an optimal rate for small enough deviations (Theorem B),
- a general Berry-Esseen bound (Theorem C).

Let us give a few examples where our results apply:

- taking $\mathscr{X} = L^\infty(\Omega)$, our assumptions essentially reduce to uniform ergodicity of the Markov chain and boundedness of the observable, a very classical case. We obtain a convergence rate proportional to the spectral gap, improving on [GO02, KLMM05] where the rate is proportional to the square of the spectral gap (see Section 3.2 and especially Theorem 3.3),

- taking $\mathscr{X} = \mathrm{Lip}(\Omega)$, our assumptions essentially reduce to positively curved Markov chains (in the sense of Ollivier) and bounded Lipschitz observables. This for example applies to contracting Iterated Function Systems and backward random walks of expanding maps. We shall see (Section 3.3) that in the toy case of the discrete hypercube and observables with small Lipschitz constant, Theorem A is less powerful than [JO10] but that for larger Lipschitz constants, Theorem B can improve on [JO10],

- when Ω is a graph, we propose a functional space of functions with small "local total variations" and show on an example that it can improve on [JO10] (also in Section 3.3),

- taking $\mathscr{X} = \mathrm{BV}(I)$ where I is an interval, we show that our results apply to a natural Markov chains related to *Bernoulli convolutions*, and allowing observables of bounded variation makes our result applicable to e.g. characteristic functions of intervals,

- more generally, when Ω is a domain of $\mathbb{R}^d$ some natural Markov chains are $\mathrm{BV}(\Omega)$-ergodic and our results apply to functions of bounded variation, e.g. characteristic functions of sets of finite perimeter – but we will not consider this case here, since it needs a somewhat sophisticated setup,

- Another direction we do not explore here is to take $\mathscr{X} = \mathrm{Hol}_\alpha(\Omega)$, the space of α-Hölder functions, or in case $\Omega = I$ is an interval, $\mathscr{X} = \mathrm{BV}_p(I)$, the space of p-bounded variation functions. These enables one to consider more general functions than $\mathrm{Lip}(\Omega)$ or respectively $\mathrm{BV}(\Omega)$; even for Lipschitz of BV functions, using these spaces can be useful because they tend to give regular observables a much lower norm.

To my knowledge, no effective result was known in the setting of bounded variation functions (and while the usual spectral method could have been used in this case, I do not know of previous asymptotic results either); no effective result could give optimal rate for moderate deviation as Theorem B does; and the effective Berry-Esseen bound seems new in most of the above cases.

Possible follow-ups. While we shall take some time discussing examples, we are far from having exhausted the domain of applicability of our results. Finding other cases to which they apply is a natural direction to pursue.

One limitation of this work is that we ask for a spectral gap *with constant* 1, i.e. the averaging operator defined by the transition kernel should be contracting on the kernel of the stationary measure. Extending to spectral gaps with arbitrary constants (i.e. eventual exponential contractivity) is possible in principle with the methods used here, but would be very technical. It could improve the rate of convergence in many cases, by replacing δ_0 (defined so that $1-\delta_0$ is the contraction constant) by the real spectral gap δ (while δ_0 is a lower bound on the spectral gap). In practical applications, this is not crucial since one can always extract a sub-Markov chain $(X_{\ell k})_{k\geq 0}$: for ℓ large enough one gets δ_0 close to δ (see Remark 2.6).

It appears from the numerical comparison with previous results that our improvement are in several case asymptotically strong but numerically modest (the improvements are large only in relatively extreme ranges of the parameters), and this seems in part due to the restriction to Banach algebra, and the ensuing necessity to combine $\|\cdot\|_\infty$ with a semi-norm (see Section 3). It would be interesting to dispense from this necessity, for example by making effective the Keller-Liverani theorem [KL99].

Structure of the article. In Section 2 we state notation and the main results. Section 3 explains in detail the aforementioned examples and comparison with previous results. In Section 4 we recall how perturbation theory can be used to prove limit theorems, and state the perturbation results we need to carry out this method in a effective manner. In Section 5 we prove the core estimates to be used thereafter, while Section 6 carries out the proof of the concentration inequalities. Section 7 is devoted to the proof of the Berry-Esseen inequality.

2 Assumptions and main results

Let Ω be a polish metric space endowed with its Borel algebra and denote by $\mathcal{P}(\Omega)$ the set of probability measures on Ω. We consider a transition kernel $\mathsf{M} = (m_x)_{x\in\Omega}$ on Ω, i.e. $m_x \in \mathcal{P}(\Omega)$ for each $x \in \Omega$, and a Markov chain $(X_k)_{k\geq 0}$ following the kernel M, i.e. $\mathbb{P}(X_{k+1} \mid X_k = x) = m_x$. We do not ask the Markov chain to be stationary: the law of X_0 is arbitrary ("cold start"); in some cases of interest, the law of each X_k will even be singular with respect to the stationary measure μ_0.

We shall study the behavior of $(X_k)_{k\geq 0}$ by comparing the empirical mean to the stationary mean:

$$\hat{\mu}_n(\varphi) := \frac{1}{n}\sum_{k=1}^{n} \varphi(X_k) \quad \text{vs.} \quad \mu_0(\varphi)$$

for an arbitrary "observable" $\varphi \in \mathscr{X}$, where $\mathscr{X}$ is a space of functions $\Omega \to \mathbb{R}$ (or $\Omega \to \mathbb{C}$).

2.1 Assumptions

Standing assumption 2.1. *In all the paper, we assume $\mathscr{X}$ satisfies the following:*

i. its norm $\|\cdot\|$ *dominates the uniform norm:* $\|\cdot\| \geq \|\cdot\|_\infty$,

ii. $\mathscr{X}$ *is a Banach algebra, i.e. for all* $f, g \in \mathscr{X}$ *we have* $\|fg\| \leq \|f\|\|g\|$,

iii. $\mathscr{X}$ *contain the constant functions and* $\|\mathbf{1}\| = 1$ *(where* $\mathbf{1}$ *denotes the constant function with value* 1*).*

The first hypothesis ensures integrability with respect to arbitrary probability measure, which is important for cold-start Markov chains; it also implies that every probability measure can be seen as a continuous linear form acting on $\mathscr{X}$. The second hypothesis will prove very important in our method where products abound (and can be replaced by the more lenient $\|fg\| \leq C\|f\|\|g\|$ up to multiplying the norm by a constant), and the hypothesis on $\|\mathbf{1}\|$ is a mere matter of convenience and could be removed at the cost of more complicated formulas.

Remark 2.2. This setting may seem restrictive at first: the Banach algebra hypothesis notably excludes L^p spaces, while classically one only makes moment assumptions on the observable. This is quite unavoidable given that we will work with more than one equivalence class of measures, and we want to allow cold start at a given position ($X_0 \sim \delta_{x_0}$). The measures m_x may be singular with respect to the stationary measure μ_0, and as a matter of fact in the dynamical applications m_x will be purely atomic while μ_0 will often be atomless. It may thus happen that for φ an $L^p(\mu_0)$ observable, $\varphi(X_j)$ is undefined with positive probability, or is extremely large even if φ has small moments with respect to μ_0. Our framework ensures enough regularity to prevent such phenomenons.

To the transition kernel M is associated an averaging operator acting on $\mathscr{X}$:

$$\mathrm{L}_0 f(x) = \int_\Omega f(y)\, \mathrm{d}m_x(y).$$

Since each m_x is a probability measure, L_0 has 1 as eigenvalue, with eigenfunction $\mathbf{1}$.

Standing assumption 2.3. *In all the article we assume* M *satisfies the following:*

i. L_0 *acts as a bounded operator from* $\mathscr{X}$ *to itself, and its operator norm* $\|\mathrm{L}_0\|$ *is equal to* 1*.*

ii. L_0 *has a spectral gap with constant* 1 *and size* $\delta_0 > 0$*, i.e. there is an hyperplane* $G_0 \subset \mathscr{X}$ *such that*

$$\|\mathrm{L}_0 f\| \leq (1 - \delta_0)\|f\| \qquad \forall f \in G_0,$$

The first hypothesis could be relaxed, considering operators of arbitrary norm, at the cost of (much) more complicated formulas. The second hypothesis is the main one, and implies in particular that 1 is a simple isolated eigenvalue.

Remark 2.4. This second hypothesis ensures that up to scalar factors there is a unique continuous linear form ϕ_0 acting on $\mathscr{X}$ such that $\phi_0 \circ \mathrm{L}_0 = \phi_0$; since any stationary measure of M satisfy this, all stationary measures coincide on $\mathscr{X}$. They might not be unique (e.g. if $\mathscr{X}$ contains only constants), but since we consider the $\varphi(X_k)$ with $\varphi \in \mathscr{X}$, this will not matter. We will thus denote an arbitrary stationary measure by μ_0, and identify it with ϕ_0 (observe that G_0 is then equal to $\ker \mu_0$). In most cases, $\mathscr{X}$ will be dense in the space of continuous function endowed with the uniform norm, ensuring that two measures coinciding on $\mathscr{X}$ are equal, and then the spectral gap hypothesis ensures the uniqueness of the stationary measure.

Remark 2.5. There are numerous examples where assumptions 2.1 and 2.3 are satisfied; we will discuss a few of them in Section 3. Typically, $\mathscr{X}$ has a norm of the form $\|\cdot\| = \|\cdot\|_\infty + V(\cdot)$ where V is a seminorm measuring the regularity in some sense (e.g. Lipschitz constant, α-Hölder constant, total variation, total p-variation...) and satisfying $V(fg) \leq \|f\|_\infty V(g) + V(f)\|g\|_\infty$. This inequality ensures that $\mathscr{X}$ is a Banach Algebra, and $\|\mathbf{1}\| = 1$ holds as soon as $V(\mathbf{1}) = 0$. Since averaging operators necessarily satisfy $\|\mathrm{L}_0 f\|_\infty \leq \|f\|_\infty$, it is sufficient that L contracts V (i.e. $V(\mathrm{L}_0 f) \leq \theta V(f)$ for some $\theta \in (0,1)$ and all $f \in \mathscr{X}$) to ensure that $\|\mathrm{L}_0\| = 1$. We will prove in Lemma 3.1 that in many cases, the contraction also implies a spectral gap of explicit size and constant 1. In fact, all examples considered here are of this kind, but it seemed better to state our main results in terms of the hypotheses we use directly in the proof. This is done at the expense of some sharpness: indeed we could improve our constants under the hypotheses of Lemma 3.1, by estimating with more precision $\|\pi_0\|$ below. The method is similar to the proof of Lemma 3.1, and is carried out in two examples in [Klo17a].

Remark 2.6. In some cases, one gets a spectral gap with a constant greater than 1, i.e.

$$\|\mathrm{L}_0^n f\| \leq C(1-\delta_0)^n \|f\| \qquad \forall f \in G_0$$

for all $n \in \mathbb{N}$ and some $C > 1$. In this case, all our result apply to the Markov chains $Y_m = X_{n_0+mk}$ where n_0 is arbitrary and k is such that $C(1-\delta_0)^n < 1$. This can be also used when $C = 1$, in cases where the spectral gap is small. In numerical computations, this can be especially useful when the simulation of the random walk is much cheaper than the evaluation of the observable.

2.2 A general concentration inequality

Our first result is a concentration inequality, featuring the expected dichotomy between a Gaussian regime and an exponential regime.

Theorem A. *For all $n \geq 60/\delta_0$ and all $a > 0$, it holds*

$$\mathbb{P}_\mu\Big[|\hat{\mu}_n(\varphi) - \mu_0(\varphi)| \geq a\Big] \leq \begin{cases} 2.5 \exp\Big(-\dfrac{na^2}{\|\varphi\|^2} \dfrac{\delta_0}{13.44\,\delta_0 + 8.324}\Big) & \text{if } \dfrac{a}{\|\varphi\|} \leq \delta_0/3 \\ 2.7 \exp\Big(-\dfrac{na}{\|\varphi\|} \cdot 0.009\,\delta_0^2\Big) & \text{otherwise} \end{cases}$$

Remark 2.7. In Theorem A we made a compromise between precision and simplicity; we in fact obtain slightly better front constants, a slightly relaxed range for n, and a more precise rate in the exponential regime (see Theorems 6.3 and 6.4).

See Section 3 for several applications and comparisons with previous results. Let us stress right away that the main strength of the present result is its broadness: we need no warm-start hypothesis, no reversibility, and we can apply it in many functional spaces. In particular, this makes our results broader than those of [Lez98, Lez01] which assume ergodicity. Lezaud also gets a front constant proportional to the $L^2(\mu_0)$-norm of the density of the distribution of X_0 with respect to the stationary distribution, which would be infinite in many of our cases of applicability; even in the case of a finite state space he then gets a large front constant $X_0 \sim \delta_x$. The approach of Joulin and Ollivier enabled them to get rid of this constant in some test cases, and we compare our results to theirs in Section 3.3.

In some cases, Theorem A can improve on previous results by allowing one to choose the functional space most suited to the situation at hand; an example is treated in detail in Section 3.3.3.

While our constants are certainly not optimal, we get what seems the correct dependence in the spectral gap, at least in the Gaussian regime (rate proportional to δ_0); in the case of Markov chains satisfying the Doeblin minorization condition, this improves on the rate proportional to δ_0^2 in [KLMM05] (see Section 3.2).

2.3 Concentration under a variance bound

The spectral method gives us access to higher-order estimates, enabling us to improve the Gaussian regime bound as soon as we have a good control over the "dynamical variance" (also called "asymptotic variance"). This quantity is defined as

$$\sigma^2(\varphi) = \mu_0(\varphi^2) - (\mu_0\varphi)^2 + 2\sum_{k\geq 1} \mu_0(\varphi \mathrm{L}_0^k \bar{\varphi})$$

where $\bar{\varphi} = \varphi - \mu_0(\varphi)$. The dynamical variance is precisely the variance appearing in the CLT for $(\varphi(X_k))_{k\geq 0}$.

Theorem B. *If $U \in [0, +\infty)$ is a upper bound for $\sigma^2(\varphi)$, then for all $a \leq \dfrac{U\delta_0^2}{26\|\varphi\|}$ and all $n \geq 60/\delta_0$ it holds*

$$\mathbb{P}_\mu\left[|\hat{\mu}_n(\varphi) - \mu_0(\varphi)| \geq a\right] \leq 2.7 \exp\left(-\frac{na^2}{2U} + \frac{na^3\|\varphi\|^3}{U^3} 10(1+\delta_0^{-1})^2\right)$$

For small enough a, the positive term in the exponential is negligible, and the leading term is exactly the best we can expect given the available knowledge: since $(\varphi(X_k))_k$ satisfies a Central Limit Theorem with variance $\sigma^2(\varphi)$, any better value would necessarily imply a better bound on $\sigma^2(\varphi)$.

Remark 2.8. Again we actually prove a slightly more precise result, see Section 6.3.

In Section 3.3.3, we show on an example how Theorem B can improve on Theorem A. However obtaining a good estimation on the dynamical variance can be difficult; in practical applications, one could use other tools to estimate it, and then apply Theorem B.

2.4 A Berry-Esseen bound

Our third main result, proven in section 7, quantifies the speed of convergence in the Central Limit Theorem.

Theorem C. *Assume $\sigma^2(\varphi) > 0$ and let $\tilde{\varphi} := \frac{\varphi - \mu_0(\varphi)}{\sigma(\varphi)}$ be the reduced centered version of φ, and denote by G, F_n the distribution functions of the reduced centered normal law and of $\frac{1}{\sqrt{n}}(\tilde{\varphi}(X_1) + \cdots + \tilde{\varphi}(X_n))$, respectively.*

For all $n \geq (60/\delta_0)^2$ it holds

$$\|F_n - G\|_\infty \leq 177 \frac{(\delta_0^{-1} + 1.13)^2 \max\{\|\tilde{\varphi}\|, \|\tilde{\varphi}\|^3\}}{\sqrt{n}}$$

Remark 2.9. The hypothesis on n is pretty harmless: if $\|\tilde{\varphi}\| \simeq \|\tilde{\varphi}^3\| \simeq 1$ then for $n \leq (60/\delta_0)^2$ the right hand side is much larger than 1, and the inequality is void. As before, a slightly more precise result can be obtained (see Section 7).

Remark 2.10. Note that $\sigma^2(\varphi)$ is always non-negative, as it can be rewritten as

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{Var}_{\mu_0}\Big(\sum_{k=1}^{n} \varphi(X_k)\Big)$$

(where the μ_0 subscript means that the assumption $X_0 \sim \mu_0$ is made). However, $\sigma^2(\varphi)$ can vanish even when φ is not constant modulo μ_0, as is the case in a dynamical setting when m_x is supported on $T^{-1}(x)$ for some map $T : \Omega \to \Omega$, and φ is a coboundary: $\varphi = g - g \circ T$ for some g. One can for example see details [GKLMF15], where σ^2 is interpreted as a semi-norm. Whenever $\sigma^2(\varphi) = 0$, one can use the present method to obtain stronger non-asymptotic concentration inequalities, giving small probability to deviations a such that $a/\|\varphi\| \gg 1/n^{2/3}$ instead of $a/\|\varphi\| \gg 1/\sqrt{n}$.

There are numerous works on Berry-Esseen bounds. In the case of independent identically distributed random variables, the optimal constant is not yet known (the best known constant is, to my knowledge, given by Tyurin [Tyu11]). Berry-Esseen bounds for Markov chains go back to [Bol82], but I know only of two previous *effective* result, by Dubois [Dub11] and by Lezaud [Lez01].

The scope of Dubois' result is quite narrower than ours, as it is only written for uniformly expanding maps of the interval and Lipschitz observables (though the method is expected to have wider application), and our numerical constant is much better: while the dependences on the parameters of the system are stated differently and thus somewhat difficult to compare, Dubois has a front constant of 11460 which is quite large

for practical applications (the order of convergence being $1/\sqrt{n}$, this constant has a squared effect on the number of iterations needed to achieve a given precision).

The scope of Lezaud's Berry-Esseen bound is also restricted, to ergodic reversible Markov chains. Moreover he gets a front constant proportional to the $L^2(\mu_0)$-norm of the density of the distribution of X_0 with respect to the stationary distribution; in comparison, our result is insensitive to the distribution of X_0.

Application to dynamical systems As is well-known, limit theorems for Markov chain also apply in a dynamical setting; let us give some details.

Given a k-to-one map $T : \Omega \to \Omega$, one defines the transfer operator of a potential $A \in \mathscr{X}$ by

$$\mathrm{L}_{T,A} f(x) = \sum_{y \in T^{-1}(x)} e^{A(y)} f(y).$$

One says that A is normalized when $\mathrm{L}_{T,A}\mathbf{1} = \mathbf{1}$. This condition exactly means that $m_x = \sum_{y \in T^{-1}(x)} e^{A(y)} \delta_y$ is a probability measure for all x, making $\mathrm{L}_{T,A}$ the averaging operator of a transition kernel. We could consider more general maps T, considering a transition kernel that is supported on its inverse branches.

If the transfer operator has a spectral gap, then the stationary measure μ_0 is unique, and readily seen to be T-invariant. We shall denote it by μ_A to stress the dependence on the potential. The corresponding stationary Markov chain $(Y_k)_{k \in \mathbb{N}}$ satisfies all results presented above; but for each n, the time-reversed process defined by $X_k = Y_{n-k}$ (where $0 \le k \le n$) satisfies $X_{k+1} = T(X_k)$: all the randomness lies in $X_0 = Y_n$. Having taken Y_n stationary makes the law of Y_n, i.e. X_0, independent of the choice of n. It follows:

Corollary 2.11. *For all normalized $A \in \mathscr{X}$ such that $\mathrm{L}_{T,A}$ has a spectral gap with constant 1 and size δ_0, for all $\varphi \in \mathscr{X}$, Theorems A, B and C hold for the random process $(X_k)_{k \in \mathbb{N}}$ defined by $X_0 \sim \mu_A$ and $X_{k+1} = T(X_k)$.*

In this context, spectral gap was proved in many cases under the impetus of Ruelle, see e.g. the books [Bal00, Rue04], the recent works [BT08, CV13, CS09], and references therein. Let me finally mention [Klo17a] (which is based on the same effective perturbation theory as the present paper) and [Klo17c] (which is my initial motivation to consider the spectral method for limit theorems).

3 Examples

3.1 Preliminary lemma

In each example below we will use the following lemma which, in the spirit of Doeblin-Fortet and Lasota-Yorke inequalities, enables to turn an exponential contraction in the "regularity part" of a functional norm into a spectral gap.

Lemma 3.1. *Consider a normed space $\mathscr{X}$ of (Borel measurable, bounded) functions $\Omega \to \mathbb{R}$, with norm $\|\cdot\| = \|\cdot\|_\infty + V(\cdot)$ where V is a semi-norm (usually quantifying some regularity of the argument, such as* Lip *or* BV*).*

Assume that for some constant $C > 0$, for all probability μ on Ω and for all $f \in \mathscr{X}$ such that $\mu(f) = 0$, $\|f\|_\infty \leq CV(f)$.

Let $\mathrm{L}_0 \in \mathscr{B}(\mathscr{X})$ and assume that for some $\theta \in (0,1)$ and all $f \in \mathscr{X}$:

$$\|\mathrm{L}_0 f\|_\infty \leq \|f\|_\infty \quad \textit{and} \quad V(\mathrm{L}_0 f) \leq \theta V(f)$$

and having eigenvalue 1 with an eigenprobability μ_0, i.e. $\mathrm{L}_0^* \mu_0 = \mu_0$.

Then L_0 has a spectral gap (for the eigenvalue 1, the contraction being on the stable space $\ker \mu_0$) with constant 1, of size

$$\delta_0 = \frac{1-\theta}{1+C\theta}$$

The condition $\|f\|_\infty \leq CV(f)$ is often valid in practice (assuming Ω has finite diameter for spaces such as $\mathrm{Lip}(\Omega)$): the condition that $\mu(f) = 0$ implies that f vanishes (if functions in $\mathscr{X}$ are continuous) or at least takes both non-positive and non-negative values, and $V(f)$ usually bounds the variations of f, implying a bound on its uniform norm.

Proof. Let $f \in \ker \mu_0$; then $\|\mathrm{L}_0 f\|_\infty \leq \|f\|_\infty$ and $\mathrm{L}_0 f \in \ker \mu_0$, so that $\|\mathrm{L}_0 f\|_\infty \leq CV(\mathrm{L}_0 f) \leq C\theta V(f)$.

Denote by $t \in [0,1]$ the number such that $\|f\|_\infty = t\|f\|$ (and therefore $V(f) = (1-t)\|f\|$). The above two controls on $\|\mathrm{L}_0(f)\|_\infty$ can then be written as $\|\mathrm{L}_0(f)\|_\infty \leq \min\big(t, C\theta(1-t)\big)\|f\|$ and using $V(\mathrm{L}_0 f) \leq \theta V(f)$ again we get

$$\begin{aligned}\|\mathrm{L}_0(f)\| &\leq \min\big(t + \theta(1-t), (C+1)\theta(1-t)\big)\|f\| \\ \|(\mathrm{L}_0)_{|\ker \mu_0}\| &\leq \max_{t\in[0,1]} \min\big(t + \theta(1-t), (C+1)\theta(1-t)\big).\end{aligned}$$

The maximum is reached when $t + \theta(1-t) = (C+1)\theta(1-t)$, i.e. when $t = C\theta/(1+C\theta)$, at which point the value in the minimum is $(C+1)\theta/(C\theta+1) \in (0,1)$. Therefore there is a spectral gap with constant 1 and size $1 - (C+1)\theta/(C\theta+1)$, as claimed. □

3.2 Chains with Doeblin's minorization

The simplest example of a Banach Algebra of functions is $L^\infty(\Omega)$, the set of measurable bounded functions.[2] To fit our framework, we will need to endow $L^\infty(\Omega)$ with the following norm:

$$\|f\|_S = \|f\|_\infty + \sup_{x,y\in\Omega} |f(x) - f(y)|$$

Of course, this norm is equivalent to the uniform norm, and it is easily checked what we still get a Banach Algebra. The point of this is that the semi-norm

$$S(f) = \sup_{x,y\in\Omega} |f(x) - f(y)| = \sup f - \inf f$$

[2] We do not have a single reference measure here, which is why we consider genuinely bounded functions rather than essentially bounded functions.

measures how "spread out" f is, wich we need to manage separately from the magnitude of f.

Observe that convergence of measures in duality to $L^\infty(\Omega)$ is convergence in total variation, and the most usual normalization is

$$d_{\mathrm{TV}}(\mu,\nu) := \sup_{S(f)=1} \big|\mu(f)-\nu(f)\big|$$

For a transition kernel M, having an averaging operator L_0 with a spectral gap is a very strong condition, called *uniform ergodicity*.

Glynn and Ormoneit [GO02] and Kontoyiannis, Lastras-Montaño and Meyn [KLMM05] gave explicit concentration results for such chains, using the characterization of uniform ergodicity by the *Doeblin minorization condition*: there exist an integer $\ell \geq 1$, a positive number β and a probability measure ω on Ω such that for all $x \in \Omega$ and all Borel set $B \subset \Omega$:

$$m_x^\ell(B) \geq \beta\omega(B) \tag{2}$$

where m_x^ℓ is the law of X_ℓ conditionally to $X_0 = x$.

We shall look at the case $\ell = 1$, which fits better in our context. For arbitrary value of ℓ, one can in practice apply the result to each extracted chain $(X_{k_0+k\ell})_{k\geq 0}$.

Lemma 3.2. *If* M *satisfies Doeblin's minorization condition* (2) *with* $\ell = 1$*, then its averaging operator* L_0 *has a spectral gap on* $L^\infty(\Omega)$ *with constant* 1 *and size* $\beta/(2-\beta)$*.*

Proof. This is simply the classical maximal coupling method in a functional guise. For each $x \in \Omega$ decompose m_x into $\beta\omega$ and $r_x := m_x - \beta\omega$ (which is a positive measure of mass $1-\beta$). Recall that we denote by μ_0 the stationary measure of M. For all $f \in L^\infty(\Omega)$ we have:

$$\begin{aligned}
\mathrm{L}_0 f(x) &= \beta\omega(f) + r_x(f) \\
\mathrm{L}_0 f(x) - \mathrm{L}_0 f(y) &= \int (r_x(f) - r_y(f))\,\mathrm{d}\mu_0(y) \\
\big|\mathrm{L}_0 f(x) - \mathrm{L}_0 f(y)\big| &\leq \int (1-\beta) S(f)\,\mathrm{d}\mu_0(y) \\
S(\mathrm{L}_0 f) &\leq (1-\beta) S(f).
\end{aligned}$$

We can thus apply Lemma 3.1 with $C = 1$ and $\theta = 1-\beta$, obtaining a spectral gap of size $\beta/(2-\beta)$. □

Theorem 3.3. *If* M *satisfies Doeblin's minorization condition* (2) *with* $\ell = 1$ *and* $\varphi : \Omega \to [-1,1]$*, for all* $n \geq 120/\beta$ *and all* $a \leq \beta/2$ *it holds*

$$\mathbb{P}_\mu\Big[|\hat{\mu}_n(\varphi) - \mu_0(\varphi)| \geq a\Big] \leq 2.5 \exp\Big(-na^2 \cdot \frac{\beta}{150+47\beta}\Big)$$

Proof. We have here $\|\varphi\|_S \leq 2$ and, by Lemma 3.2, $\delta_0 \geq \beta/(2-\beta) \geq \beta/2$. It then suffices to apply Theorem A. □

	Runtime n to ensure error below a with probability ≥ 0.99		
Value of a	Theorem 6.3	Theorem B	[GO02] & [KLMM05]
0.001	2.76×10^{11}	NA	1.18×10^{12}
0.00003	3.07×10^{14}	8.3×10^{12}	1.3×10^{15}

Table 1: Comparison with [GO02, KLMM05] for $\ell = 1$, $\beta = 0.003$, and observable $\varphi : \Omega \to [0, 1]$.

The constant 150 in the rate is not so great, but we obtain a major improvement over [GO02, KLMM05] when β is small (and a in the Gaussian window): their rate is proportional to β^2 instead of the correct order β which we obtain.

Let us give some concrete numerical estimates, sumed up in Table 1. For $a = 0.001$, $\beta = 0.003$ and a 99% certainty, we need to take $n \simeq 2.76 \times 10^{11}$ while [GO02, KLMM05] need $n \simeq 1.18 \times 10^{12}$.

For larger a, our exponential regime has a factor β^2 but then we gain a factor of a. For very small a we can appeal to Theorem B to obtain a much better rate. For this, one only has to observe that $S(\mathrm{L}_0^k \bar{\varphi}) \leq (1-\beta)^k S(\varphi)$ to bound $\sigma^2(\varphi)$ by any $U \geq 1 + \frac{4}{\beta}$. The smaller is U, the best is the leading term in the rate (up to a best of $-na^2\beta/4$) but the smaller is the allowed window (down to $a \lesssim \beta/72$); for say $a = 0.00003$ it suffices to take $n = 8.3 \times 10^{12}$.

As a matter of illustration, for $\varphi : \Omega \to [-1, 1]$, $a = 0.00003$ and $\beta = 0.003$ Theorem B ensures that when $n \simeq 8.3 \times 10^{12}$, the probability to get an error more than a is less than $1/100$. Meanwhile, [GO02, KLMM05] need at least $n \simeq 1.18 \times 10^{15}$ so we gain two orders of magnitude, at the boundary of feasibility for cheaply simulated chains.

3.3 Discrete hypercube

Let us start with the same toy example as Joulin and Ollivier [JO10], the lazy random walk (aka Gibbs sampler, aka Glauber dynamics) on the discrete hypercube $\{0, 1\}^N$: the transition kernel M chooses randomly uniformly a slot $i \in \{1, \ldots, N\}$ and replaces it with the result of a fair coin toss, i.e.

$$m_x = \frac{1}{2}\delta_x + \sum_{y \sim x} \frac{1}{2N}\delta_y.$$

We consider two kind of observables: the "polarization" $\rho : \{0, 1\}^N \to \mathbb{R}$ giving the proportion of 1's in its argument, and the characteristic function $\mathbf{1}_S$ of a subset $S \subset \{0, 1\}^N$. In this second example, we will in particular consider the simple case

$$S = [0] := \{(0, x_2, \ldots, x_N) \colon x_i \in \{0, 1\}\}.$$

We state in Table 2 the estimates only in the level of details that enables to compare with Joulin and Ollivier's result. One sees that we obtain a weaker estimate in the case of ρ, but a better one in the case of $\mathbf{1}_{[0]}$ *if we are careful enough.*

Observable	Runtime to ensure error below a with good probability		
	Theorem A with Lip. norm	Our best result	Joulin-Ollivier
$\frac{1}{N}$-Lip maps such as ρ	$O\left(\frac{N}{a^2}\right)$	$O\left(\frac{N}{a^2}\right)$	$O\left(N+\frac{1}{a^2}\right)$
$\mathbf{1}_{[0]}$	$O\left(\frac{N^2}{a^2}\right)$	$O\left(\frac{N}{a^2}\right)$	$O\left(\frac{N^2}{a^2}\right)$
$\mathbf{1}_S$ where S is "scrambled"	$O\left(\frac{N^2}{a^2}\right)$	$O\left(\frac{1}{a^2}\right)$	$O\left(\frac{N^2}{a^2}\right)$

Table 2: Comparison with [JO10], always assuming a small enough.

The rest of this subsection explains how to get these estimates from our results.

3.3.1 Notation and functional spaces

The discrete hypercube $\{0,1\}^N$ is endowed with the Hamming metric: if $x=(x_1,\dots,x_N)$ and $y=(y_1,\dots,y_N)$, then $d(x,y)$ is the number of indexes i such that $x_i \neq y_i$. Two elements at distance 1 are said to be adjacent, denoted by $x \sim y$.

We denote by E the set of tuples $\epsilon=(\epsilon_i)_{1\leq i\leq N}$ such that exactly one of the ϵ_i is 1. Identifying $\{0,1\}$ with $\mathbb{Z}/2\mathbb{Z}$, an edge thus writes $(x,x+\epsilon)$ for some $x\in\{0,1\}^N$ and some $\epsilon\in E$.

We shall consider several function spaces to showcase the flexibility of the spectral method; since the space $\{0,1\}^N$ is finite, we always consider the space of all functions $\{0,1\}^N\to\mathbb{R}$, and it is the considered norm which will matter. Let us define:

- $\|f\|_L = \|f\|_\infty + \mathrm{Lip}(f)$: this is the standard Lipschitz norm;
- $\|f\|_{dL} = \|f\|_\infty + N\,\mathrm{Lip}(f)$: this is the Lipschitz norm with a weigth to the regularity part equal to the diameter;
- $\|f\|_W = \|f\|_\infty + W(f)$ where

$$W(f) = \sup_{x\in\{0,1\}^N} \sum_{\epsilon\in E} |f(x+\epsilon)-f(x)|;$$

  this norm stays small for functions having large variations only in few directions (small "local total variation").

Proposition 3.4. *Each of the norm $\|\cdot\|_L$, $\|\cdot\|_{dL}$ and $\|\cdot\|_W$ turns the space of all functions $\{0,1\}^N\to\mathbb{R}$ into a Banach algebra where $\mathbf{1}$ has norm 1.*

Moreover the averaging operator L_0 of the transition kernel M has operator norm 1, and spectral gap with constant 1 and respective size $1/N^2$, $1/(2N-1)$ and $1/(4N-1)$ in the norms $\|\cdot\|_L$, $\|\cdot\|_{dL}$ and $\|\cdot\|_W$.

Proof. That the norms define Banach Algebras is proven as indicated in Remark 2.5. All the other properties but the spectral gap are trivial.

To prove the spectral gaps, we simply apply Lemma 3.1. First, it is well-known that for all $\varphi : \{0,1\}^N \to \mathbb{R}$,

$$\mathrm{Lip}(\mathrm{L}_0\varphi) \leq (1-1/N)\,\mathrm{Lip}(\varphi)$$

(in the parlance of [Oll09], M is positively curved with $\kappa = 1/N$).

In the case of $\|\cdot\|_L$, we get $\theta = 1-1/N$ and $C = N$ (since a function of vanishing average must take positive and negative values, and $\mathrm{diam}\{0,1\}^N = N$), hence a spectral gap of size $1/N^2$. In the case of $\|\cdot\|_{dL}$, the normalizing factor gives $C = 1$ (and we still have $\theta = 1-1/N$), hence a spectral gap of size $1/(2N-1)$.

To deal with $\|\cdot\|_W$, we first show that in Lemma 3.1 we can take $\theta = 1-1/(2N)$.

$$\begin{aligned}
W(\mathrm{L}_0\varphi) &= \sup_x \sum_{\epsilon\in E} \left| \frac{1}{2}\varphi(x+\epsilon) + \frac{1}{2N}\sum_{\eta\in E}\varphi(x+\eta+\epsilon) - \frac{1}{2}\varphi(x) - \frac{1}{2N}\sum_{\eta\in E}\varphi(x+\eta)\right| \\
&= \sup_x \sum_{\epsilon\in E} \left| \Big(\frac{1}{2}-\frac{1}{2N}\Big)\varphi(x+\epsilon) + \frac{1}{2N}\sum_{\eta\neq\epsilon}\varphi(x+\eta+\epsilon) \right. \\
&\qquad\qquad \left. - \Big(\frac{1}{2}-\frac{1}{2N}\Big)\varphi(x) - \frac{1}{2N}\sum_{\eta\neq\epsilon}\varphi(x+\eta)\right| \\
&\leq \sup_x \frac{N-1}{2N}\sum_{\epsilon\in E}|\varphi(x+\epsilon)-\varphi(x)| + \frac{1}{2N}\sum_{\epsilon\in E}\sum_{\eta\neq\epsilon}|\varphi(x+\epsilon+\eta)-\varphi(x+\eta)| \\
&\leq \frac{N-1}{2N}W(\varphi) + \frac{1}{2N}\sup_x \sum_{y\sim x}\sum_{\epsilon\in E}|\varphi(y+\epsilon)-\varphi(y)| \\
W(\mathrm{L}_0\varphi) &\leq \Big(1-\frac{1}{2N}\Big)W(\varphi)
\end{aligned}$$

Then Lemma 3.5 below shows that we can take $C = 1$, providing a spectral gap of size $1/(4N-1)$ □

The following optimal estimate and its proof where provided by Fedor Petrov on MathOverflow.

Lemma 3.5 (Fedor Petrov [Pet17])**.** *For all $f : \{0,1\}^N \to \mathbb{R}$ we have*

$$\max f - \min f \leq W(f).$$

Proof. Without lost of generality, we can assume $W(f) \leq 1$ and $f(0,0,\dots,0) = 0$, and reduce to proving $f(1,1,\dots,1) \leq 1$.

Define the *cost* of a path $x^0, x^2, \dots, x^k$ as the number $\sum_{i=0}^{k-1}|f(x^{i+1}) - f(x^i)|$, and let Σ be the sum of the costs of all paths of length N from $(0,0,\dots,0)$ to $(1,1,\dots,1)$. We shall prove that $\Sigma \leq N!$, and since there are $N!$ such paths one of them will have cost at most 1, proving the lemma.

We call "level" of $x \in \{0,1\}^N$ the number of 1s among the coordinates of x, and denote it by $|x|$. For each $i \in \{0,1,\dots,N-1\}$, define $p_i = \frac{i!(N-i)!}{N+1}$. Then all p_i are positive and $p_i + p_{i+1} = i!(N-i-1)!$ is precisely the number of paths that use any given edge from level i to level $i+1$.

The contribution to Σ of an edge $(x, x+\epsilon)$ from level i to level $i+1$ is thus $i!(N-i-1)!|f(x+\epsilon)-f(x)|$, which we split into two parts, one $p_i|f(x+\epsilon)-f(x)|$ attributed to x and the other $p_{i+1}|f(x+\epsilon)-f(x)|$ to $x+\epsilon$. It follows

$$\Sigma \leq \sum_{x\in\{0,1\}^N} p_{|x|}W(f) \leq \sum_{i=0}^{N} p_i \binom{N}{i} = \sum_{i=0}^{N-1} (p_i + p_{i+1}) \binom{N-1}{i} = N(N-1)! = N!$$

as desired. □

3.3.2 Polarization

Consider the "polarization" observable $\rho : \{0,1\}^N \to \mathbb{R}$, where $\rho(x)$ is the proportion of 1's in the word x. We have

$$\|\rho\|_L = 1 + \frac{1}{N}, \qquad \|\rho\|_{dL} = 2, \qquad \|\rho\|_W = 2.$$

To use Theorem A with optimal efficiency, assuming a will be small enough, we need to maximize $\delta_0/\|\rho\|^2$. Here, we shall thus use the norm $\|\cdot\|_{dL}$. For $a \lesssim N$, Theorem A shows that we need at most $O(N/a^2)$ iterations to have a good convergence to the actual mean; meanwhile Joulin and Ollivier only need $O(1/a^2)$, but for concentration around the expectancy of the empiric process, not around the expectancy with respect to the stationary measure. Without burn-in, one also needs to bound the bias, which approaches zero in time $O(N/a)$ according to the bound of Joulin and Ollivier, for a total run time of $O(N/a + 1/a^2)$. With burn-in, they need a run time of $O(N + 1/a^2)$.

For $1/N \lesssim a \lesssim 1$, we enter our exponential regime while staying inside Joulin-Ollivier's Gaussian window; Theorem A shows we need no more than $O(N^2/a)$ iterations, while [JO10] still gives a bound of $O(N + 1/a^2)$.

In this example, Joulin and Ollivier get a sharper result; this seems to be explained in one part by the fact that we do not get to decouple the bias from the convergence of expectancies, and in another part by our need to have a Banach algebra, hence to include the uniform norm in our norm.

3.3.3 Observable with small variance or small local total variation

Consider now the potential $\mathbf{1}_S$, the indicator function for a (non-trivial) set S. This function is only 1-Lipschitz, so that we have $\|\mathbf{1}_S\|_L = 2$ and $\|\mathbf{1}_S\|_{dL} = 1 + N$. If we insist on using a Lipschitz norm, the unormalized one is thus better and with $\delta_0 = 1/N^2$ Theorem A shows that we need (in the Gaussian regime) $O(N^2/a^2)$ iterations to ensure the error is probably less than a, which is the same order of magnitude than given by [JO10] with a worse constant, ~ 34 instead of 8. But here we have two ways to improve on this bound.

The first one is to use Theorem B. When

$$S = [0] := \{0x_2x_3\cdots x_N \in \{0,1\}^N\},$$

the dynamical variance can be computed explicitly (distinguish the cases when the first digit has been changed an odd or even number of times, and observe that at each step the probability of changing the first digit is $1/2N$):

$$\mu_0(\mathbf{1}_{[0]}^2) - (\mu_0 \mathbf{1}_{[0]})^2 = 1/4 \quad \text{and} \quad \sum_{k\geq 1} \mu_0(\mathbf{1}_{[0]} \mathrm{L}_0^k \bar{\mathbf{1}}_{[0]}) = \frac{1}{4} \sum_{k\geq 1} \left(\frac{N-1}{N}\right)^k = \frac{N-1}{4}$$

This gives $\sigma^2(\mathbf{1}_S) \simeq N/2$. Switching back to the norm $\|\cdot\|_{dL}$, when $a \lesssim 1/N^2$ (see Remark 2.8) and $n \geq 60N^2$, in Theorem B the positive term in the exponential is negligible compared to the main term which is $-na^2/N$. In particular $O(N/a^2)$ iterations suffice to get a small probability for a deviation at least a: compared to Joulin and Ollivier, we gain one power of N in this regime (and the optimal constant 1 in the leading term of the rate) but only for very small values of a.[3] This choice of S might seem very specific, but for less regular S the gain should be greater for sufficiently smaller a. For example, if S contains half the vertices and every vertex $x \in \{0,1\}^N$ has exactly $2Np$ neighbors with the same $\mathbf{1}_S$ value, the above computation of variance gives $\sigma^2(\mathbf{1}_S) = \frac{1}{4} + \frac{1-2p}{4p}$. "Scrambled" sets with p independent of N get $\sigma^2(\mathbf{1}_S) \simeq 1$ and taking $n = O(1/a^2)$ is sufficient.

The second way to improve our first estimate is to use the norm $\|\cdot\|_W$ in Theorem A. Then $\|\mathbf{1}_{[0]}\|_W = 2$ and $\delta_0 \simeq 1/N$. For $a \lesssim 1/N$, Theorem A ensures that we need only $O(N/a^2)$ iterations to have a good convergence to the actual mean, which is again the optimal order of magnitude (since it corresponds to the CLT) but obtained on a much larger window than with Theorem B. This extends to all observables with $W(\varphi) \lesssim 1$; observe that this domain of applicability is quite complementary to the domain of applicability of the previous paragraph.

3.4 Bernoulli convolutions and observables of bounded variation

The MCMC method is often used in high dimension, where it can be very difficult to simulate independent random variables of a given law. Let us give an example showing that even in dimension 1, using Markov chains can be efficient.

We consider the "Bernoulli convolution" of parameter $\lambda \in (0,1)$, defined as the law β_λ of the random variable

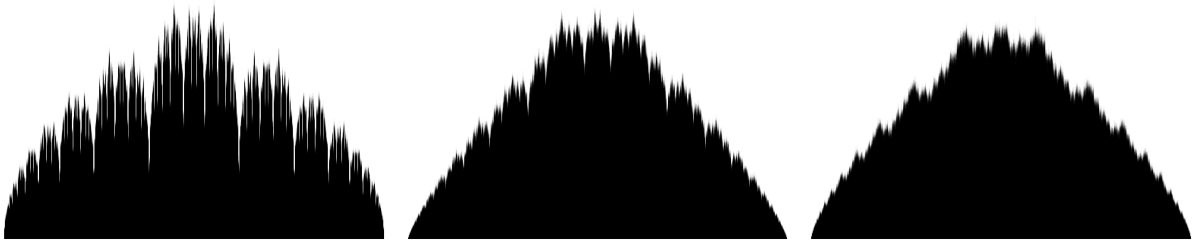
$$\sum_{k\geq 1} \epsilon_k \lambda^k$$

where the ϵ_k are independent Bernoulli variables with parameter $1/2$, i.e. ϵ_k is 1 with probability $1/2$ and -1 with probability $1/2$.

When $\lambda < 1/2$, the support of β_λ is a Cantor set of zero Lebesgue measure, so that β_λ is singular[4]. When $\lambda = 1/2$, β_λ is the uniform measure on $[-1,1]$. But when $\lambda \in (1/2,1)$

[3] If we want to consider a of the order of $1/N$, we can then take $U \simeq N^2$ to enlarge the window, at the cost of a weaker leading term. We get a bound similar to the one of Joulin-Ollivier, possibly with a smaller constant (depending on the value of a).

[4] Unless explicitly mentioned, in this subsection "singular" and "absolutely continuous" will always be meant with respect to Lebesgue measure.



(a) $\lambda = \frac{\sqrt{5}-1}{2} \simeq 0.618$ (b) $\lambda = \frac{e}{4} \simeq 0.680$ (c) $\lambda = \frac{2}{3} \simeq 0.667$

Figure 1: Histogram of the empirical distribution of the Markov chain associated to (T_0, T_1), with $X_0 = 0$, binned in 500 subintervals (averaged image over 30 independent runs of 10^6 points each). Parameter λ is the inverse of a Pisot number on the left, a very well approximable irrational at the center, rational on the right.

(which we assume from now on), the question of the absolute continuity of β_λ is very difficult, and fascinating. It was proved by Erdös [Erd39] that if λ is the inverse of a Pisot number, then β_λ is singular, and a while later Solomyak discovered that for Lebesgue-almost all λ, β_λ is absolutely continuous [Sol95]; see Figure 1 for illustration. Important open questions are to find an explicit λ such that β_λ is absolutely continuous, and to find an explicit λ, not the inverse of a Pisot number, such that β_λ is singular; see [PSS00] for more information on these questions.

Finding an expression for the density of β_λ when it exists seems of course even further from reach; but of course, there is an obvious way to simulate β_λ: draw the ϵ_k up to some rank K. To achieve a precision of p on the result, one needs $K \simeq \frac{\log 1/p}{\log 1/\lambda}$ random bits for *each* independent sample.

One can also realize naturally β_λ as the stationary law of the Markov transition kernel $\mathsf{M} = (m_x)_{x\in\mathbb{R}}$ defined by

$$m_x = \frac{1}{2}\delta_{T_0(x)} + \frac{1}{2}\delta_{T_1(x)}$$

where $T_0(x) = \lambda x - \lambda$ and $T_1(x) = \lambda x + \lambda$ (this is a particular case of an Iterated Function System (IFS)).

In order to evaluate $\beta_\lambda(\varphi)$ by a MCMC method, one cannot use the methods developed for ergodic Markov chains since, conditionally to $X_0 = x$, the law m_x^k of X_k is atomic and thus singular with respect to β_λ: $d_{\mathrm{TV}}(m_x^k, \beta_\lambda) = 1$ for all k. The convergence only holds for observables satisfying some regularity assumption, and it is natural to ask what regularity is needed.

For a Lipschitz observable φ one only need to observe that M has positive curvature in the sense of Ollivier (this is easy using the coupling $\frac{1}{2}\delta_{(T_0(x),T_0(y))} + \frac{1}{2}\delta_{(T_1(x),T_1(y))}$ of m_x and m_y) and apply [JO10]. But what if φ is not Lipschitz (or has large Lipschitz constant)? We shall consider observables of bounded variation, a regularity which has the great advantage over Lipschitz to include the characteristic functions of intervals.

Definition 3.6. Given an interval $I \subset \mathbb{R}$, we consider the Banach space $\mathrm{BV}(I)$ of *bounded variation* functions $I \to \mathbb{R}$, defined by the norm $\|\cdot\|_{\mathrm{BV}} = \|\cdot\|_\infty + \mathrm{var}(\cdot, I)$ where

$$\mathrm{var}(f, I) := \sup_{x_0 < x_1 < \cdots < x_p \in I} \sum_{j=1}^{p} |f(x_j) - f(x_{j-1})|$$

(the uniform norm is usually replaced by the L^1 norm, but when I is bounded our choice is equivalent up to a constant, it does not single out the Lebesgue measure, and most importantly it ensures that $\mathrm{BV}(I)$ is a Banach algebra).

Important features of total variation are:

- its extensiveness: $\mathrm{var}(f, I) \geq \mathrm{var}(f, J) + \mathrm{var}(f, K)$ whenever J, K are disjoint subintervals of I,
- its invariance under monotonic maps: $\mathrm{var}(f \circ T, I) = \mathrm{var}(f, T(I))$ whenever T is monotonic.

The averaging operator L_0 of the transition kernel M has a spectral gap for all λ, but *not* with constant 1 when $\lambda > 1/2$. This only means that we will deal with an extracted Markov chain $(X_{\ell k})_{k \geq 0}$ for some ℓ.

Let I_λ be the attractor of the IFS (T_0, T_1), i.e. the interval whose endpoints are the fixed points of T_0 and T_1:

$$I_\lambda = \big[\frac{-\lambda}{1-\lambda}, \frac{\lambda}{1-\lambda}\big].$$

Given a word $\omega = \omega_1 \omega_2 \dots \omega_k$ in the letters 0 and 1, we define

$$T_\omega = T_{\omega_1} \circ T_{\omega_2} \circ \cdots \circ T_{\omega_k} : I_\lambda \to I_\lambda.$$

Proposition 3.7. *If $\lambda^\ell < \frac{1}{2}$, then L_0^ℓ has a spectral gap on $\mathrm{BV}(I_\lambda)$ of size $1/(2^{\ell+1} - 1)$ and constant 1.*

Proof. Let I_λ^-, I_λ^+ be the left and right halves of I_λ, i.e.

$$I_\lambda^- = \big[\frac{-\lambda}{1-\lambda}, 0\big) \qquad I_\lambda^+ = \big(0, \frac{\lambda}{1-\lambda}\big].$$

Let $f \in \mathrm{BV}(I_\lambda)$ and observe that the condition $\lambda^\ell < \frac{1}{2}$ ensures that $T_{00\dots0}(I_\lambda)$ and $T_{11\dots1}(I_\lambda)$ are disjoint (they have length $< \frac{1}{2}|I_\lambda|$ and each contains an endpoint of I_λ). Then:

$$\begin{aligned} \mathrm{var}(\mathrm{L}_0 f, I_\lambda) &\leq \frac{1}{2^\ell} \sum_{\omega \in \{0,1\}^\ell} \mathrm{var}(f \circ T_\omega, I_\lambda) \\ &\leq \frac{1}{2^\ell} \sum_{\omega \in \{0,1\}^\ell} \mathrm{var}(f, T_\omega(I_\lambda)) \end{aligned}$$

$$\leq \frac{1}{2^\ell}\Big(\operatorname{var}(f, T_{00\dots0}(I_\lambda)) + \operatorname{var}(f, T_{11\dots1}(I_\lambda) + \sum_{\substack{\omega\neq 00\dots0\\ \neq 11\dots1}} \operatorname{var}(f, I_\lambda)\Big)$$

$$\leq \frac{1}{2^\ell}\big(\operatorname{var}(f, I_\lambda) + (2^\ell - 2)\operatorname{var}(f, I_\lambda)\big)$$

$$\operatorname{var}(\mathrm{L}_0 f, I_\lambda) \leq (1 - 2^{-\ell})\operatorname{var}(f, I_\lambda).$$

Applying Lemma 3.1 with $C = 1$ and $\theta = 1 - 2^{-\ell}$ yields the claim. □

This enables us to apply our result to estimate $\beta_\lambda(\varphi)$ for any φ of bounded variation. For example, Theorem A yields the following (where we only state the Gaussian regime, with the slightly better constant of Theorem 6.3).

Theorem 3.8. *Let $\lambda \in (\frac{1}{2}, 1)$ and let ℓ be the smallest integer such that $\lambda^\ell < \frac{1}{2}$. Consider a Markov chain $(X_k)_{k\geq 0}$ with transition probability $2^{-\ell}$ from $x \in I_\lambda$ to $T_\omega(x)$, for each $\omega \in \{0,1\}^\ell$. For any starting distribution $X_0 \sim \mu$, any $\varphi \in \mathrm{BV}(I_\lambda)$, any positive $a < \|\varphi\|_{\mathrm{BV}}/3(2^{\ell+1} - 1)$ and any $n \geq 120 \cdot 2^\ell$ we have*

$$\mathbb{P}_\mu\Big[|\hat{\mu}_n(\varphi) - \mu_0(\varphi)| \geq a\Big] \leq 2.488 \exp\Big(- \frac{na^2}{\|\varphi\|^2_{\mathrm{BV}}(16.65 \cdot 2^\ell + 5.12)}\Big)$$

To the best of our knowledge, this example could not be handled effectively by previously known results. For example [GD12] needs the observable to be at least C^2 to have explicit estimates, and they do not give a concentration inequality.

For the sake of concreteness, fix any $\lambda \in (\frac{1}{2}, \frac{1}{\sqrt{2}})$ so that the above applies with $\ell = 2$. Consider as observable a characteristic function $\mathbf{1}_J$ where $J \subset I_\lambda$ is an (open, say) interval. We have $\|\mathbf{1}_J\|_{\mathrm{BV}} = 3$. The Gaussian window is very large; for all $a \in (0, 1/7)$, to ensure an error at least a occurs with probability less than $1/100$ it is sufficient to take $n = 3561/a^2$, i.e. $7122/a^2$ random bits. The constant is somewhat large, and could probably be improved by taking a larger ℓ, finding more disjoint pairs of intervals in the proof of the spectral gap, enabling one to get a much smaller θ in Lemma 3.1.

Of course, to estimate $\beta_\lambda(\mathbf{1}_J)$ one could be tempted to use Hoeffding's inequality (we would need only $n = 2.65/a^2$ independent samples); but for any given random point Y with distribution β_λ, it is very difficult to determine *a priori* which precision will ensure a correct value for $\mathbf{1}_J(Y)$, and thus how many ϵ_k should be drawn. One can easily construct a stopping time (waiting for the distance between the current value and the boundary of the interval to be larger than $\sum_{k>K} \lambda^k$), but since β_λ might be a very irregular measure, it seems quite difficult to control this stopping time *a priori*.

One could also be tempted to bound $\mathbf{1}_J$ from below and above by Lipschitz functions in order to apply [JO10], but one would need to ensure that these bounding functions are $L^1(\beta_\lambda)$-close one to another. For this one would need them to have very large Lipschitz constants (of the order of $1/a$ *if* one very conservatively assumes that β_λ is absolutely continuous with bounded density), making the total runtime of the order of $1/a^4$ in the best case.

4 Connection with perturbation theory

To any $\varphi \in \mathscr{X}$ (sometimes called a "potential" in this role) is associated a weighted averaging operator, called a transfer operator in the dynamical context:

$$\mathrm{L}_\varphi f(x) = \int_\Omega e^{\varphi(y)} f(y)\, \mathrm{d}m_x(y).$$

The classical guiding idea for the present work combines two observations. First, we have

$$\mathrm{L}_\varphi^2 f(x_0) = \int_\Omega e^{\varphi(x_1)} \mathrm{L}_\varphi f(x_1)\, \mathrm{d}m_{x_0}(x_1) = \int_{\Omega\times\Omega} e^{\varphi(x_1)} e^{\varphi(x_2)} f(x_2)\, \mathrm{d}m_{x_1}(x_2)\, \mathrm{d}m_{x_0}(x_1)$$

and by a direct induction, denoting by $\mathrm{d}m_{x_0}^n(x_1,\dots,x_n)$ the law of n steps of a Markov chain following the transition M and starting at x_0, we have

$$\mathrm{L}_\varphi^n f(x_0) = \int_{\Omega^n} e^{\varphi(x_1)+\cdots+\varphi(x_n)} f(x_n)\, \mathrm{d}m_{x_0}^n(x_1,\dots,x_n).$$

In particular, applying to the function $f = \mathbf{1}$, we get

$$\mathrm{L}_\varphi^n \mathbf{1}(x_0) = \int_{\Omega^n} e^{\varphi(x_1)+\cdots+\varphi(x_n)}\, \mathrm{d}m_{x_0}^n(x_1,\dots,x_n) = \mathbb{E}_{x_0}\left[e^{\varphi(X_1)+\cdots+\varphi(X_n)}\right]$$

where $(X_k)_{k\geq 0}$ is a Markov chain with transitions M and the subscript on expectancy and probabilities specify the initial distribution (x_0 being short for δ_{x_0}).

It follows by linearity that if the Markov chain is started with $X_0 \sim \mu$ where μ is any probability measure, then setting $\hat{\mu}_n\varphi := \frac{1}{n}\varphi(X_1) + \cdots + \frac{1}{n}\varphi(X_n)$ we have

$$\mathbb{E}_\mu\left[\exp(t\hat{\mu}_n\varphi)\right] = \int \mathrm{L}_{\frac{t}{n}\varphi}^n \mathbf{1}(x)\, \mathrm{d}\mu(x) \tag{3}$$

This makes a strong connection between the transfer operators and the behavior of $\hat{\mu}_n\varphi$.

Second, when the potential is small (e.g. $\frac{t}{n}\varphi$ with large n), the transfer operator is a perturbation of L_0, and their spectral properties will be closely related. This is the part that has to be made quantitative to obtain effective limit theorems.

We will state the perturbation results we need after introducing some notation. The letter L will always denote a bounded linear operator, and $\|\cdot\|$ will be used both for the norm in $\mathscr{X}$ and for the operator norm. From now on it is assumed that L_0 has a spectral gap of size δ_0 and constant 1. In [Klo17b] the leading eigenvalue of L_0 is denoted by λ_0, an eigenvector is denoted by u_0, and an eigenform (eigenvector of L_0^*) is denoted by ϕ_0 (similarly the eigenvalue of an operator L close to L_0 is denoted by λ_{L}).

Two quantities appear in the perturbation results below. The first one is the *condition number* $\tau_0 := \frac{\|\phi_0\|\|u_0\|}{|\phi_0(u_0)|}$. To define the second one, we need to introduce π_0, the projection on G_0 along $\langle u_0\rangle$, which here writes $\pi_0(f) = f - \mu_0(f)$, and observe that by the spectral

hypothesis $(\mathrm{L}_0 - \lambda_0)$ is invertible when acting on G_0. Then the *spectral isolation* is defined as

$$\gamma_0 := \|(\mathrm{L}_0 - \lambda_0)^{-1}_{|G_0}\pi_0\|.$$

We shall denote by P_0 the projection on $\langle u_0\rangle$ along G_0, and set $\mathrm{R}_0 = \mathrm{L}_0 \circ \pi_0$. We then have the expression

$$\mathrm{L}_0 = \lambda_0 \mathrm{P}_0 + \mathrm{R}_0$$

with $\mathrm{P}_0\mathrm{R}_0 = \mathrm{R}_0\mathrm{P}_0 = 0$. This decomposition will play a role below, and can be done for all L with a spectral gap: we denote by $\lambda_\mathrm{L}, \pi_\mathrm{L}, \mathrm{P}_\mathrm{L}, \mathrm{R}_\mathrm{L}$ the corresponding objects for L, and by $\lambda, \pi, \mathrm{P}, \mathrm{R}$ we mean the corresponding maps $\mathrm{L} \mapsto \lambda_\mathrm{L}$, etc.

Last, the notation $O_C(\cdot)$ is the Landau notation with an explicit constant C, i.e. $f(x) = O_C(g(x))$ means that for all x, $|f(x)| \leq C|g(x)|$.

Theorem 4.1 (Theorems 2.3 and 2.6 and Proposition 5.1 (viii) of [Klo17b])**.** *All* L *such that* $\|\mathrm{L} - \mathrm{L}_0\| < \dfrac{1}{6\tau_0\gamma_0}$ *have a simple isolated eigenvalue;* $\lambda, \pi, \mathrm{P}, \mathrm{R}$ *are defined and analytic on this ball.*

Given any $K > 1$*, whenever* $\|\mathrm{L} - \mathrm{L}_0\| \leq \dfrac{K-1}{6K\tau_0\gamma_0}$ *we have*

$$\begin{aligned}
\lambda_\mathrm{L} &= \lambda_0 + O_{\tau_0 + \frac{K-1}{3}}\big(\|\mathrm{L} - \mathrm{L}_0\|\big)\\
\lambda_\mathrm{L} &= \lambda_0 + \phi_0(\mathrm{L} - \mathrm{L}_0)u_0 + O_{K\tau_0\gamma_0}\big(\|\mathrm{L} - \mathrm{L}_0\|^2\big)\\
\lambda_\mathrm{L} &= \lambda_0 + \phi_0(\mathrm{L} - \mathrm{L}_0)u_0 + \phi_0(\mathrm{L} - \mathrm{L}_0)\mathrm{S}_0(\mathrm{L} - \mathrm{L}_0)u_0 + O_{2K^2\tau_0^2\gamma_0^2}\Big(\|\mathrm{L} - \mathrm{L}_0\|^3\Big)\\
\mathrm{P}_\mathrm{L} &= \mathrm{P}_0 + O_{2K\tau_0\gamma_0}(\|\mathrm{L} - \mathrm{L}_0\|)\\
\pi_\mathrm{L} &= \pi_0 + O_{\tau_0 + \frac{K-1}{3}}(\|\mathrm{L} - \mathrm{L}_0\|)
\end{aligned}$$

$$\Big\|D\Big[\frac{1}{\lambda}\mathrm{R}\Big]_\mathrm{L}\Big\| \leq \frac{1}{|\lambda_\mathrm{L}|} + \frac{\tau_0 + \frac{K-1}{3}}{|\lambda_\mathrm{L}|^2}\|\mathrm{L}\| + 2K\tau_0\gamma_0.$$

Theorem 4.2 (Corollaire 2.12 from [Klo17b])**.** *In the case* $\lambda_0 = \|\mathrm{L}_0\| = 1$*, all* L *such that*

$$\|\mathrm{L} - \mathrm{L}_0\| \leq \frac{\delta_0(\delta_0 - \delta)}{6(1 + \delta_0 - \delta)\tau_0\|\pi_0\|}$$

have a spectral gap of size δ *below* λ_L*, with constant* 1*.*

Since we will apply these results to the averaging operator L_0, we need to evaluate the parameters in this case.

We have $\lambda_0 = 1$, $u_0 = \mathbf{1}$ and ϕ_0 is identified with the stationary measure μ_0. It first follows that

$$\tau_0 = 1.$$

Indeed $\|u_0\| = 1$ by hypothesis, $\|\phi_0\| = 1$ since $\|\cdot\| \geq \|\cdot\|_\infty$ and ϕ_0 is a probability measure, and $|\phi_0(u_0)| = |\mu_0(\mathbf{1})| = 1$.

Then we have

$$\|\pi_0\| \leq 2$$

since for all $f \in \mathscr{X}$, we have $\pi_0(f) = f - \mu_0(f)$ and $\|\mu_0(f)\mathbf{1}\| = |\mu_0(f)| \le \|f\|_\infty \le \|f\|$. In general this trivial bound can hardly be improved without more information, notably on μ_0: it may be the case that μ_0 is concentrated on a specific region of the space, and then $f - \mu_0(f)$ could have norm close to twice the norm of f.

Last, from the Taylor expansion $(1 - \mathrm{L}_0)^{-1} = \sum_{k\ge 0} \mathrm{L}_0^k$, the spectral gap δ_0, and the upper bound on $\|\pi_0\|$ we deduce

$$\gamma_0 \le 2/\delta_0.$$

5 Main estimates

Standing assumption 2.3 ensures that for all small enough φ we can apply the above perturbation results; recall that μ_0 is the stationary measure, so that for all $f \in \mathscr{X}$ we have $\int \mathrm{L}_0 f \,\mathrm{d}\mu_0 = \int f \,\mathrm{d}\mu_0$.

We will first apply Theorem 4.2 with $\delta = \delta_0/13$; this is somewhat arbitrary, but the exponential decay will be strong enough compared to other quantities that we don't need δ to be large. Taking it quite small allow for a larger radius where the result applies.

As a consequence of this choice, the following smallness assumption will often be needed:

$$\|\varphi\| \le \log\Big(1 + \frac{\delta_0^2}{13 + 12\delta_0}\Big). \tag{4}$$

We will often use φ instead of L_φ in subscripts: for example λ_φ is the main eigenvalue of L_φ and π_φ is linear projection on its eigendirection along the stable complement appearing in the definition of the spectral gap.

Lemma 5.1. *We have*

$$\mathrm{L}_\varphi(\cdot) = \mathrm{L}_0\Big(\sum_{j\ge 0} \frac{\varphi^j}{j!} \cdot\Big) \qquad \textit{and} \qquad \|\mathrm{L}_\varphi - \mathrm{L}_0\| \le e^{\|\varphi\|} - 1.$$

If (4) holds, then we have

$$\begin{aligned}
\|\mathrm{L}_\varphi - \mathrm{L}_0\| &\le \frac{\delta_0^2}{13 + 12\delta_0} \le \frac{1}{25} \\
\mathrm{L}_\varphi &= \mathrm{L}_0 + O_{1.02}(\|\varphi\|) \\
&= \mathrm{L}_0 + \mathrm{L}_0(\varphi\cdot) + O_{0.507}(\|\varphi\|^2) \\
&= \mathrm{L}_0\big((1 + \varphi + \frac{1}{2}\varphi^2)\cdot\big) + O_{0.169}(\|\varphi\|^3), \\
\|\pi_\varphi\| &\le 2.053
\end{aligned}$$

Assumption (4) is in particular sufficient to apply Theorem 4.2 with $\delta = \delta_0/13$ and Theorem 4.1 with $K = 1 + 12\delta_0/13$.

Proof. The first formula is a rephrasing of the definition of L_φ; observe then that thanks to the assumption that $\mathscr{X}$ is a Banach algebra, we have

$$\|\mathrm{L}_\varphi - \mathrm{L}_0\| = \|\mathrm{L}_0\big((e^\varphi - 1)\cdot\big)\|$$

$$\leq \|\mathrm{L}_0\| \Big\| \sum_{j=1}^{\infty} \frac{\varphi^j}{j!} \Big\|$$
$$\leq \sum_{j=1}^{\infty} \frac{\|\varphi\|^j}{j!}$$
$$\|\mathrm{L}_\varphi - \mathrm{L}_0\| \leq e^{\|\varphi\|} - 1$$

Observing that $x \mapsto x^2/(13+12x)$ is increasing from 0 to $1/25$ as x varies from 0 to 1 completes the uniform bound of $\|\mathrm{L}_\varphi - \mathrm{L}_0\|$ and gives $\|\varphi\| \leq \log(1+1/25) := b$. By convexity, we deduce that

$$e^{\|\varphi\|} - 1 \leq (e^b - 1)\frac{\|\varphi\|}{b} \leq 1.02\|\varphi\|$$

and the zeroth order Taylor formula follow.

The higher-order estimates are obtained similarly:

$$\mathrm{L}_\varphi = \mathrm{L}_0\big((\mathbf{1} + \varphi + (e^\varphi - \varphi - 1))\cdot\big) = \mathrm{L}_0 + \mathrm{L}_0(\varphi\cdot) + O_{\|\mathrm{L}_0\|}(e^\varphi - \varphi - 1)$$

and using the triangle inequality, the convexity of $\frac{e^x-x-1}{x}$ and the bound on φ:

$$\|e^\varphi - \varphi - 1\| \leq \frac{e^{\|\varphi\|} - \|\varphi\| - 1}{\|\varphi\|}\|\varphi\| \leq \frac{e^b - b - 1}{b^2}\|\varphi\|^2 \leq 0.507\|\varphi\|^2.$$

The second order remainder is bounded by

$$\|e^\varphi - \frac{1}{2}\varphi^2 - \varphi - 1\| \leq \frac{e^b - \frac{1}{2}b^2 - b - 1}{b^3}\|\varphi\|^3 \leq 0.169\|\varphi\|^3$$

and finally, we have

$$\|\pi_\varphi\| \leq \|\pi_0\| + \big(1 + \frac{4\delta_0}{13}\big)\|\mathrm{L}_\varphi - \mathrm{L}_0\| \leq 2 + \big(1 + \frac{4}{13}\big)\frac{1}{25} \leq 2.053.$$

□

Lemma 5.2. *Under* (4) *we have*

$$|\lambda_\varphi - 1| \leq 0.0524$$
$$\lambda_\varphi = 1 + O_{1.334}(\|\varphi\|)$$
$$\lambda_\varphi = 1 + \mu_0(\varphi) + O_{2.43+2.081\delta_0^{-1}}(\|\varphi\|^2)$$
$$\lambda_\varphi = 1 + \mu_0(\varphi) + \frac{1}{2}\mu_0(\varphi^2) + \sum_{k\geq 1}\mu_0(\varphi \mathrm{L}_0^k(\bar{\varphi})) + O_{7.41+17.75\delta_0^{-1}+8.49\delta_0^{-2}}(\|\varphi\|^3)$$

Proof. With $K = 1 + 12\delta_0/13$ we have $\tau_0 + \frac{K-1}{3} = 1 + 4\delta_0/13$ and by the Theorem 4.1, $\mathrm{L} \mapsto \lambda_\mathrm{L}$ has Lipschitz constant at most $1+4/13 = 17/13$. We get $|\lambda_\varphi - \lambda_0| \leq \frac{17}{13}\|\mathrm{L}_\varphi - \mathrm{L}_0\|$ from which we deduce both

$$|\lambda_\varphi - 1| \leq \frac{17}{13 \times 25} \leq 0.0524$$

$$\text{and} \qquad |\lambda_\varphi - 1| \le \frac{17}{13} 1.02 \|\varphi\| \le 1.334 \|\varphi\|$$

Now we use the first-order Taylor formula for λ, using $K\tau_0\gamma_0 \le 2\delta_0^{-1}(1 + 12\delta_0/13) = \frac{24}{13} + 2\delta_0^{-1}$:

$$\lambda_\varphi = 1 + \mu_0\big((\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1})\big) + O_{\frac{24}{13}+2\delta_0^{-1}}(\|\mathrm{L}_\varphi - \mathrm{L}_0\|^2),$$

then using $\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1} = \mathrm{L}_0(\varphi) + O_{0.507}(\|\varphi\|^2)$ from Lemma 5.1 we get

$$\mu_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1}) = \mu_0(\mathrm{L}_0(\varphi)) + O_{0.507}(\|\varphi\|^2) = \mu_0(\varphi) + O_{0.507}(\|\varphi\|^2).$$

Using $\|\mathrm{L}_\varphi - \mathrm{L}_0\| \le 1.02\|\varphi\|$ gives the following constant in the final $O(\|\varphi\|^2)$ of the first-order formula:

$$0.507 + (1.02)^2(\frac{24}{13} + 2\delta_0^{-1}) \le 2.43 + 2.081\delta_0^{-1}.$$

Then we apply the second-order Taylor formula:

$$\lambda_\varphi = 1 + \mu_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1}) + \mu_0\Big((\mathrm{L}_\varphi - \mathrm{L}_0)\mathrm{S}_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1})\Big) + O_{8K^2\delta_0^{-2}}(\|\mathrm{L}_\varphi - \mathrm{L}_0\|^3).$$

Using $\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1} = \mathrm{L}_0(\varphi + \frac{1}{2}\varphi^2) + O_{0.169}(\|\varphi\|^3)$ from Lemma 5.1 we first get

$$\mu_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1}) = \mu_0(\varphi) + \frac{1}{2}\mu_0(\varphi^2) + O_{0.169}(\|\varphi\|^3).$$

To simplify the second term, we recall that $\mathrm{L}_\varphi - \mathrm{L}_0 = \mathrm{L}_0(\varphi\cdot) + O_{0.507}(\|\varphi\|^2)$ and

$$\mathrm{S}_0 = (1 - \mathrm{L}_0)^{-1}\pi_0 = \big(\sum_{k\ge 0} \mathrm{L}_0^k\big)\pi_0$$

where π_0 is the projection on $\ker \mu_0$ along $\langle \mathbf{1} \rangle$, i.e. $\pi_0(f) = f - \mu_0(f) =: \bar{f}$, and has norm at most 2. We thus have (noticing that in the second line both the main term and the remainder term belong to $\ker \mu_0$):

$$\begin{aligned}
\pi_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1}) &= \pi_0\big(\mathrm{L}_0(\varphi) + O_{0.507}(\|\varphi\|^2)\big) \\
&= \mathrm{L}_0(\bar{\varphi}) + O_{1.014}(\|\varphi\|^2) \\
\mathrm{S}_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1}) &= \sum_{k\ge 1} \mathrm{L}_0^k(\bar{\varphi}) + O_{1.014\delta_0^{-1}}(\|\varphi\|^2).
\end{aligned}$$

We also have

$$\|\mathrm{S}_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1})\| \le \frac{2}{\delta_0}\|\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1}\| \le \frac{2.04}{\delta_0}\|\varphi\|.$$

It then comes

$$\begin{aligned}
(\mathrm{L}_\varphi - \mathrm{L}_0)\mathrm{S}_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1}) = \mathrm{L}_0\big(\varphi \sum_{k\ge 1} \mathrm{L}_0^k(\bar{\varphi})\big) &+ O_{1.014\delta_0^{-1}}(\|\mathrm{L}_\varphi - \mathrm{L}_0\|\|\varphi\|^2) \\
&+ O_{0.507}(\|\varphi\|^2\|\mathrm{S}_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1})\|)
\end{aligned}$$

$$= \mathrm{L}_0\big(\varphi \sum_{k\geq 1} \mathrm{L}_0^k(\bar{\varphi})\big) + O_{2.07\delta_0^{-1}}(\|\varphi\|^3)$$

$$\mu_0(\mathrm{L}_\varphi - \mathrm{L}_0)\mathrm{S}_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0\mathbf{1}) = \sum_{k\geq 1} \mu_0(\varphi \mathrm{L}_0^k(\bar{\varphi})) + O_{2.07\delta_0^{-1}}(\|\varphi\|^3)$$

where the reversal of sum and integral is enabled by normal convergence.

Last, we observe

$$8K^2\delta_0^{-2} = 8(\frac{12}{13} + \delta_0^{-1})^2 \leq 6.82 + 14.77\delta_0^{-1} + 8\delta_0^{-2},$$

and we gather all what precedes:

$$\begin{aligned}
\lambda_\varphi &= 1 + \mu_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0\mathbf{1}) + \mu_0\Big((\mathrm{L}_\varphi - \mathrm{L}_0)\mathrm{S}_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0\mathbf{1})\Big) + O_{8K^2\delta_0^{-2}}(\|\mathrm{L}_\varphi - \mathrm{L}_0\|^3) \\
&= 1 + \mu_0(\varphi) + \frac{1}{2}\mu_0(\varphi^2) + O_{0.169}(\|\varphi\|^3) + \sum_{k\geq 1} \mu_0(\varphi \mathrm{L}_0^k(\bar{\varphi})) + O_{2.07\delta_0^{-1}}(\|\varphi\|^3) \\
&\qquad + O_{(6.82+14.77\delta_0^{-1}+8\delta_0^{-2})1.02^3}(\|\varphi\|^3) \\
&= 1 + \mu_0(\varphi) + \frac{1}{2}\mu_0(\varphi^2) + \sum_{k\geq 1} \mu_0(\varphi \mathrm{L}_0^k(\bar{\varphi})) + O_{7.41+17.75\delta_0^{-1}+8.49\delta_0^{-2}}(\|\varphi\|^3)
\end{aligned}$$

□

Under assumption (4), we know that L_φ has a spectral gap of size $\delta_0/13$ with constant 1, and we can write

$$\mathrm{L}_\varphi = \lambda_\varphi \mathrm{P}_\varphi + \mathrm{R}_\varphi$$

where P_φ is the projection to the eigendirection along the stable complement and $\mathrm{R}_\varphi = \mathrm{L}_\varphi \pi_\varphi$ is the composition of the projection to the stable complement and L_φ. Then it holds $\mathrm{P}_\varphi \mathrm{R}_\varphi = \mathrm{R}_\varphi \mathrm{P}_\varphi = 0$, so that for all $n \in \mathbb{N}$:

$$\mathrm{L}_\varphi^n = \lambda_\varphi^n \mathrm{P}_\varphi + \mathrm{R}_\varphi^n.$$

Lemma 5.3. *Under assumption (4), it holds*

$$\Big\|\Big(\frac{1}{\lambda_\varphi}\mathrm{R}_\varphi\Big)^n \mathbf{1}\Big\| \leq (6.388 + 4.08\delta_0^{-1})(1 - \delta_0/13)^{n-1}\|\varphi\|$$

$$\mathrm{P}_\varphi \mathbf{1} = \mathbf{1} + O_{3.77+4.08\delta_0^{-1}}(\|\varphi\|).$$

Proof. At any $\mathrm{L} = \mathrm{L}_\varphi$ where φ satisfies (4) we have:

$$\begin{aligned}
\Big\|D\Big[\frac{1}{\lambda}\mathrm{R}\Big]_{\mathrm{L}}\Big\| &\leq \frac{1}{|\lambda_{\mathrm{L}}|} + \frac{17/13}{|\lambda_{\mathrm{L}}|^2}|\mathrm{L}| + 2K\tau_0\gamma_0 \\
&\leq \frac{1}{0.9476} + \frac{17}{13 \times 0.9476^2} \times 1.04 + \frac{48}{13} + \frac{4}{\delta_0} \\
&\leq 6.263 + \frac{4}{\delta_0}
\end{aligned}$$

so that

$$\Big\|\frac{1}{\lambda_\varphi}\mathrm{R}_\varphi\mathbf{1}-\frac{1}{\lambda_0}\mathrm{R}_0\mathbf{1}\Big\| \le (6.263+\frac{4}{\delta_0})\|\mathrm{L}_\varphi-\mathrm{L}_0\|\|\mathbf{1}\|$$
$$\Big\|\frac{1}{\lambda_\varphi}\mathrm{R}_\varphi\mathbf{1}-0\Big\| \le 1.02(6.263+\frac{4}{\delta_0})\|\varphi\|$$
$$\Big\|\frac{1}{\lambda_\varphi}\mathrm{R}_\varphi\mathbf{1}\Big\| \le (6.388+4.08\delta_0^{-1})\|\varphi\|.$$

Moreover since $\mathrm{R_L}$ takes its values in G_L where π_L acts as the identity, we have

$$\|\mathrm{R}_\varphi^n\mathbf{1}\| \le \lambda_\varphi^{n-1}(1-\delta_0/13)^{n-1}\|\mathrm{R_L}\mathbf{1}\|$$

from which the first inequality follows.

Then we have $\mathrm{P}_\varphi = \mathrm{P}_0 + O_{2K\tau_0\gamma_0}(\|\mathrm{L}_\varphi-\mathrm{L}_0\|)$, which yields the claimed result using $K = 1+12\delta_0/13$, $\tau_0=1$, $\gamma_0\le 2\delta_0^{-1}$ and $\|\mathrm{L}_\varphi-\mathrm{L}_0\|\le 1.02\|\varphi\|$. □

This control of P_φ and R_φ can be then be used to reduce the estimation of $\mathrm{L}_\varphi^n\mathbf{1}$ to the estimation of λ_φ^n.

Corollary 5.4. *Under assumptions* (4) *and*

$$n \ge 1+\frac{\log 100}{-\log(1-\delta_0/13)} \tag{5}$$

it holds

$$\mathrm{L}_\varphi^n\mathbf{1} = \lambda_\varphi^n\big(1+O_{3.834+4.121\delta_0^{-1}}(\|\varphi\|)\big)$$
$$\lambda_\varphi^n = \exp\big(n\mu_0(\varphi)+O_{3.36+2.081\delta_0^{-1}}(n\|\varphi\|^2)\big)$$
$$\lambda_\varphi^n = \exp\big(n\mu_0(\varphi)+\frac{1}{2}n\sigma^2(\varphi)+O_{10.89+20.04\delta_0^{-1}+8.577\delta_0^{-2}}(n\|\varphi\|^3)\big)$$

Proof. Assuming (4), we first appeal to Lemma 5.3 to write:

$$\begin{aligned}\mathrm{L}_\varphi^n\mathbf{1} &= \lambda_\varphi^n\mathrm{P}_\varphi\mathbf{1}+\mathrm{R}_\varphi^n\mathbf{1}\\ &= \lambda_\varphi^n\Big(\mathbf{1}+O_{3.77+4.08\delta_0^{-1}}(\|\varphi\|)+O_{6.388+4.08\delta_0^{-1}}\big((1-\delta_0/13)^{n-1}\|\varphi\|\big)\Big)\end{aligned} \tag{6}$$

The second factor of (6) is easily controlled if we ask (5), under which we have

$$\begin{aligned}A &:= 1+O_{3.77+4.08\delta_0^{-1}}(\|\varphi\|)+O_{6.388+4.08\delta_0^{-1}}\big((1-\delta_0/13)^{n-1}\|\varphi\|\big)\\ &= 1+O_{3.77+4.08\delta_0^{-1}}(\|\varphi\|)+O_{0.064+0.041\delta_0^{-1}}(\|\varphi\|)\\ &= 1+O_{3.834+4.121\delta_0^{-1}}(\|\varphi\|)\end{aligned}$$

The first estimate for λ_φ^n is obtained through the first-order Taylor formula. We use the monotony and convexity of $x\mapsto(\log(1+x)-x)/x$ and set $x=\lambda_\varphi-1\in[-b,b]$ with $b=0.0524$ to evaluate $\log(\lambda_\varphi)$:

$$\Big|\frac{\log(1+x)-x}{x}\Big| \le \frac{\log(1-b)+b}{-b}\frac{|x|}{b}$$

$$\begin{aligned} &\leq 0.52|x| \\ \log(\lambda_\varphi) &= \log(1+\lambda_\varphi - 1) \\ &= \lambda_\varphi - 1 + O_{0.52}(|\lambda_\varphi - 1|^2) \\ &= \lambda_\varphi - 1 + O_{0.52\times 1.334^2}(\|\varphi\|^2) \\ &= \lambda_\varphi - 1 + O_{0.926}(\|\varphi\|^2). \end{aligned}$$

and then using $\lambda_\varphi = 1 + \mu_0(\varphi) + O_{2.43+2.081\delta_0^{-1}}(\|\varphi\|^2)$ from Lemma 5.2:

$$\begin{aligned} \lambda_\varphi^n &= \exp\big(n\log(\lambda_\varphi)\big) \\ &= \exp\big(n(\lambda_\varphi - 1) + O_{0.926}(n\|\varphi\|^2)\big) \\ &= \exp\big(n\mu_0(\varphi) + O_{3.36+2.081\delta_0^{-1}}(n\|\varphi\|^2)\big). \end{aligned}$$

The second estimate for λ_φ^n is obtained, of course, from the second-order formula given in Lemma 5.2:

$$\lambda_\varphi = 1 + \mu_0(\varphi) + \frac{1}{2}\mu_0(\varphi^2) + \sum_{k\geq 1}\mu_0(\varphi \mathrm{L}_0^k(\bar\varphi)) + O_{7.41+17.75\delta_0^{-1}+8.49\delta_0^{-2}}(\|\varphi\|^3).$$

Here, it is somewhat tedious to use a convexity argument and we instead use the slightly less precise Taylor formula: for $x \in [-b, b]$ (where again $b = 0.0524$) we have

$$\Big|\frac{1}{6}\frac{\mathrm{d}^3}{\mathrm{d}x^3}\log(1+x)\Big| \leq \frac{2}{6(1-0.0524)^3} \leq 0.392$$

so that

$$\log(1+x) = x - \frac{1}{2}x^2 + O_{0.392}(x^3)$$

and therefore (using at one step $|\mu_0(\varphi)| \leq \|\varphi\|$):

$$\begin{aligned} \log(\lambda_\varphi) &= (\lambda_\varphi - 1) - \frac{1}{2}(\lambda_\varphi - 1)^2 + O_{0.392}((\lambda_\varphi - 1)^3) \\ &= \mu_0(\varphi) + \frac{1}{2}\mu_0(\varphi^2) + \sum_{k\geq 1}\mu_0(\varphi \mathrm{L}_0^k\bar\varphi) + O_{7.41+17.75\delta_0^{-1}+8.49\delta_0^{-2}}(\|\varphi\|^3) \\ &\quad - \frac{1}{2}\big(\mu_0(\varphi) + O_{2.43+2.081\delta_0^{-1}}(\|\varphi\|^2)\big)^2 + O_{0.392\times 1.334^3}(\|\varphi\|^3) \\ &= \mu_0(\varphi) + \frac{1}{2}\sigma^2(\varphi) + O_{10.771+19.831\delta_0^{-1}+8.49\delta_0^{-2}}(\|\varphi\|^3) + O_{2.953+5.06\delta_0^{-1}+2.166\delta_0^{-2}}(\|\varphi\|^4) \end{aligned}$$

Now assumption (4) ensures $\|\varphi\| \leq 0.04$, so that we can combine the two error terms into $O_a(\|\varphi\|^3)$ with

$$\begin{aligned} a = 10.771 + 19.831\delta_0^{-1} + 8.49\delta_0^{-2} + 0.04(2.953 + 5.06\delta_0^{-1} + 2.166\delta_0^{-2}) \\ \leq 10.89 + 20.04\delta_0^{-1} + 8.577\delta_0^{-2} \end{aligned}$$

□

6 Concentration inequalities

We will in this section apply Corollary 5.4 to $\frac{t}{n}\varphi$ instead of φ, which we can do as soon as n is large enough with respect to t and $\|\varphi\|$ in the sense that

$$n \geq \frac{\|t\varphi\|}{\log\left(1+\frac{\delta_0^2}{12+13\delta_0}\right)} \quad \text{and} \quad n \geq 1 + \frac{\log 100}{-\log(1-\delta_0/13)}, \tag{7}$$

Remark 6.1. The first condition can be replaced by any of the following stronger but simpler conditions

$$n \geq (13.3\delta_0^{-1} + 12.3\delta_0^{-2})\|t\varphi\| \quad \text{or} \quad n \geq 26\frac{\|t\varphi\|}{\delta_0^2}$$

Similarly, by an elementary function analysis the second condition can be replaced by

$$n \geq \frac{60}{\delta_0}.$$

Under these conditions, we obtain our first control of the moment generating function of the empiric mean

$$\hat{\mu}_n(\varphi) := \frac{1}{n}\varphi(X_1) + \cdots + \frac{1}{n}\varphi(X_n)$$

by plugging the first-order estimate of Corollary 5.4 in (3):

$$\begin{aligned}\frac{\mathbb{E}_\mu\left[\exp(t\hat{\mu}_n(\varphi))\right]}{\exp(t\mu_0(\varphi))} &= e^{-t\mu_0(\varphi)}\int \mathrm{L}^n_{\frac{t}{n}\varphi}\mathbf{1}(x)\,\mathrm{d}\mu(x)\\ &= \left(1+O_{3.834+4.121\delta_0^{-1}}(\frac{t}{n}\|\varphi\|)\right)\exp(O_{3.36+2.081\delta_0^{-1}}(\frac{t^2}{n}\|\varphi\|^2))\end{aligned}$$

By the classical Chernov bound, it follows that for all $a, t > 0$:

$$\begin{aligned}&\mathbb{P}_\mu\left[|\hat{\mu}_n(\varphi) - \mu_0(\varphi)| \geq a\right]\\ &\qquad \leq \left(2 + (7.668 + 8.242\delta_0^{-1})\frac{t}{n}\|\varphi\|\right)\exp\left(-at + (3.36 + 2.081\delta_0^{-1})\frac{t^2}{n}\|\varphi\|^2\right)\end{aligned} \tag{8}$$

6.1 Gaussian regime

Our first concentration inequality is obtained by choosing t to optimize the argument of the exponential in (8), i.e. taking

$$t = \frac{na}{2(3.36 + 2.081\delta_0^{-1})\|\varphi\|^2}.$$

This choice can be made as soon as a is small enough: indeed the first condition on n then reads

$$n \geq \frac{na}{2(3.36 + 2.081\delta_0^{-1})\log\left(1+\frac{\delta_0^2}{12+13\delta_0}\right)\|\varphi\|}$$

i.e.

$$a \leq (6.72 + 4.162\delta_0^{-1}) \log\Big(1 + \frac{\delta_0^2}{12 + 13\delta_0}\Big)\|\varphi\|.$$

Let us find a simpler lower bound for the right-hand side:

$$\begin{aligned}(6.72 + 4.162\delta_0^{-1}) \log\Big(1 + \frac{\delta_0^2}{12 + 13\delta_0}\Big) &\geq (6.72 + 4.162\delta_0^{-1}) \cdot 0.98 \frac{\delta_0^2}{12 + 13\delta_0} \\ &\geq \frac{6.58\delta_0 + 4}{13\delta_0 + 12}\delta_0 \\ &\geq \frac{\delta_0}{3}\end{aligned}$$

so that a sufficient condition to make the above choice for t is

$$a \leq \frac{\delta_0\|\varphi\|}{3}. \tag{9}$$

Then the argument in the exponential becomes

$$-at + (3.36 + 2.081\delta_0^{-1})\frac{t^2}{n}\|\varphi\|^2) \leq -\frac{na^2}{(13.44 + 8.324\delta_0^{-1})\|\varphi\|^2}$$

and the constant in front:

$$\begin{aligned}2 + (7.668 + 8.242\delta_0^{-1})\frac{t}{n}\|\varphi\| &\leq 2 + \frac{(7.668 + 8.242\delta_0^{-1})a}{(6.72 + 4.162\delta_0^{-1})\|\varphi\|} \\ &\leq 2 + \frac{7.668\delta_0^2 + 8.242\delta_0}{20.16\delta_0 + 12.486} \\ &\leq 2 + \frac{7.668 + 8.242}{20.16 + 12.486} \leq 2.488 \leq 2.5\end{aligned}$$

Remark 6.2. We could also have bounded the front constant in a different way to show it can be taken close to 2 for small a:

$$\begin{aligned}2 + (7.668 + 8.424\delta_0^{-1})\frac{t}{n}\|\varphi\| &\leq 2 + \frac{(7.668 + 8.242\delta_0^{-1})a}{(6.72 + 4.162\delta_0^{-1})\|\varphi\|} \\ &\leq 2 + \frac{8.242}{4.162}\frac{a}{\|\varphi\|} \\ &\leq 2 + 2\frac{a}{\|\varphi\|}\end{aligned}$$

We obtain a version of the first part of Theorem A.

Theorem 6.3. *For all n, a such that*

$$n \geq 1 + \frac{\log 100}{-\log(1 - \delta_0/13)} \quad \textit{and} \quad a \leq \frac{\delta_0\|\varphi\|}{3}$$

it holds

$$\mathbb{P}_\mu\Big[|\hat{\mu}_n(\varphi) - \mu_0(\varphi)| \geq a\Big] \leq 2.488 \exp\Big(-n\frac{\delta_0}{13.44\delta_0 + 8.324}\frac{a^2}{\|\varphi\|^2}\Big)$$

A simpler, less precise estimate is

$$\mathbb{P}_\mu\Big[|\hat{\mu}_n(\varphi)-\mu_0(\varphi)|\geq a\Big]\leq 2.5\exp\Big(-n\cdot 0.046\delta_0\frac{a^2}{\|\varphi\|^2}\Big)$$

6.2 Exponential regime

For larger a, we obtain a result with exponential decay by taking t as large as allowed by the first smallness condition (7), i.e.

$$t\simeq\frac{n}{\|\varphi\|}\log\Big(1+\frac{\delta_0^2}{12+13\delta_0}\Big).$$

To simplify, we precisely take the slightly smaller

$$t=\frac{n}{\|\varphi\|}\times\frac{0.98\delta_0^2}{12+13\delta_0}$$

Then the argument in the exponential becomes

$$\begin{aligned}-at+(3.36+2.081\delta_0^{-1})\frac{t^2}{n}\|\varphi\|^2)&=n\frac{0.98\delta_0^2}{12+13\delta_0}\Big(-\frac{a}{\|\varphi\|}+\frac{0.98(3.36\delta_0^2+2.081\delta_0)}{12+13\delta_0}\Big)\\&\leq -n\frac{0.98\delta_0^2}{12+13\delta_0}\Big(\frac{a}{\|\varphi\|}-0.254\delta_0\Big)\end{aligned}$$

and the constant in front:

$$\begin{aligned}2+(7.668+8.242\delta_0^{-1})\frac{t}{n}\|\varphi\|&=2+(7.668+8.242\delta_0^{-1})\frac{0.98\delta_0^2}{12+13\delta_0}\\&=2+\frac{7.515\delta_0^2+8.078\delta_0}{12+13\delta_0}\\&\leq 2+\frac{15.593}{25}\leq 2.624\end{aligned}$$

We obtain a version of the second part of Theorem A.

Theorem 6.4. *For all n, a such that*

$$n\geq 1+\frac{\log 100}{-\log(1-\delta_0/13)}\quad\text{and}\quad a\geq\frac{\delta_0\|\varphi\|}{3}$$

it holds

$$\mathbb{P}_\mu\Big[|\hat{\mu}_n(\varphi)-\mu_0(\varphi)|\geq a\Big]\leq 2.624\exp\Big(-n\frac{0.98\delta_0^2}{12+13\delta_0}\Big(\frac{a}{\|\varphi\|}-0.254\delta_0\Big)\Big).$$

A simpler, less precise estimate is:

$$\mathbb{P}_\mu\Big[|\hat{\mu}_n(\varphi)-\mu_0(\varphi)|\geq a\Big]\leq 2.7\exp\Big(-n\cdot 0.009\delta_0^2\frac{a}{\|\varphi\|}\Big).$$

6.3 Second-order concentration

In the case one has a good upper bound for the variance

$$\sigma^2(\varphi) = \mu_0(\varphi^2) - (\mu_0\varphi)^2 + 2\sum_{k\geq 1} \mu_0(\varphi \mathrm{L}_0^k \bar{\varphi})$$

then the previous concentration results can be improved by using the second-order formula in Corollary 5.4, which yields

$$\frac{\mathbb{E}_\mu\left[\exp(t\hat{\mu}_n(\varphi))\right]}{\exp(t\mu_0(\varphi))} = \exp\big(\frac{t^2}{2n}\sigma^2(\varphi) + O_{10.89+20.04\delta_0^{-1}+8.577\delta_0^{-2}}(\frac{t^3}{n^2}\|\varphi\|^3)\big) \times \big(1 + O_{3.834+4.121\delta_0^{-1}}(\frac{t}{n}\|\varphi\|)\big)$$

so that, if we know $\sigma^2(\varphi) \leq U$:

$$\mathbb{P}_\mu\left[|\hat{\mu}_n(\varphi) - \mu_0(\varphi)| \geq a\right] \leq \big(2 + \frac{(7.668 + 8.242\delta_0^{-1})t}{n}\|\varphi\|\big) \exp\big(-at + \frac{t^2}{2n}U + C\frac{t^3}{n^2}\|\varphi\|^3\big)$$

where C can be any number above $10.89 + 20.04\delta_0^{-1} + 8.577\delta_0^{-2}$. To get a compact expression, we observe that $0.89 + 0.04\delta_0^{-1} \leq 0.93\delta_0^{-2}$ so that

$$10.89 + 20.04\delta_0^{-1} + 8.577\delta_0^{-2} \leq 10 + 20\delta_0^{-1} + 9.507\delta_0^{-2} \leq 10(1+\delta_0^{-1})^2 =: C.$$

The choice of t can then be adapted to the circumstances; we will only explore the choice $t = an/U$ which is nearly optimal when a is small.

This choice can be made as soon as

$$a \leq \frac{U}{\|\varphi\|}\log\Big(1 + \frac{\delta_0^2}{12 + 13\delta_0}\Big)$$

and entails the following upper bound for the front constant:

$$2 + (7.668 + 8.242\delta_0^{-1})\frac{\delta_0^2}{12 + 13\delta_0} \leq 2 + \frac{7.668 + 8.242}{12 + 13} \leq 2.637$$

Meanwhile, the exponent becomes

$$-at + \frac{t^2}{2n}U + C\frac{t^3}{n^2}\|\varphi\|^3 = -\frac{a^2 n}{2U} + \frac{C\|\varphi\|^3 a^3 n}{U^3}$$

which, given $\hat{\mu}_n(\varphi)$ satisfies the Central Limit Theorem, is nearly optimal if $a \ll \frac{U^2}{2C\|\varphi\|^3}$ and U is close to $\sigma^2(\varphi)$. We obtain Theorem B, in the following version.

Theorem 6.5. *For all $n \geq 60/\delta_0$, if $\sigma^2(\varphi) \leq U$ and*

$$a \leq \frac{U}{\|\varphi\|}\log\Big(1 + \frac{\delta_0^2}{12 + 13\delta_0}\Big)$$

then it holds

$$\mathbb{P}_\mu\left[|\hat{\mu}_n(\varphi) - \mu_0(\varphi)| \geq a\right] \leq 2.637\exp\Big(-n\cdot\big(\frac{a^2}{2U} - 10(1+\delta_0^{-1})^2\frac{\|\varphi\|^3 a^3}{U^3}\big)\Big)$$

7 Berry-Esseen bounds

In this section, we use the second-order Taylor formula for the leading eigenvalue to prove effective Berry-Esseen bounds. The method we use is the one proposed by Feller [Fel66], which does not yield the best constant in the IID case, but is quite easily adapted to the Markov or dynamical case as observed in [CP90].

The starting point is a "smoothing" argument that allows to translate the proximity of characteristic functions into a proximity of distribution functions.

Proposition 7.1 ([Fel66]). *Let F, G be the distribution functions and ϕ, γ be the characteristic functions of real random variables with vanishing expectation. Assume G is derivable and $\|G'\|_\infty \leq m$; then for all $T > 0$:*

$$\|F - G\|_\infty \leq \frac{1}{\pi} \int_{-T}^{T} \Big| \frac{\phi(t) - \gamma(t)}{t} \Big| \, \mathrm{d}t + \frac{24m}{\pi T}.$$

We set $G(T) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{T} e^{-\frac{t^2}{2}} \, \mathrm{d}t$ the reduced normal distribution function, so that $\|G'\|_\infty = (2\pi)^{-\frac{1}{2}}$ and $\gamma(t) = e^{-\frac{t^2}{2}}$, and apply the above estimate to the distribution function F_n of the random variable $\frac{1}{\sqrt{n}}(\tilde{\varphi}(X_1) + \cdots + \tilde{\varphi}(X_n))$, where here $\tilde{\varphi}$ is the fully normalized version of φ:

$$\tilde{\varphi} = \frac{\varphi - \mu_0(\varphi)}{\sigma(\varphi)} \qquad \text{where } \sigma^2(\varphi) = \mu_0(\varphi^2) - (\mu_0\varphi)^2 + 2\sum_{k\geq 1} \mu_0\big(\varphi \mathrm{L}_0^k(\bar{\varphi})\big),$$

assuming $\sigma^2(\varphi) > 0$ and with $\bar{\varphi} := \varphi - \mu_0(\varphi)$. We save for later the following observation:

$$\begin{aligned} \sigma^2(\varphi) &= \sigma^2(\bar{\varphi}) \\ &\leq \|\bar{\varphi}^2\|_\infty + 2\sum_{k\geq 1} \|\bar{\varphi}\|_\infty (1-\delta_0)^k \|\bar{\varphi}\| \\ &\leq \|\bar{\varphi}\|^2 \big(1 + \frac{2}{\delta_0}\big) \end{aligned}$$

so that

$$\begin{aligned} \|\tilde{\varphi}\| &\geq \big(1 + \frac{2}{\delta_0}\big)^{-\frac{1}{2}} \\ &\geq \sqrt{\delta_0/3} \end{aligned}$$

Applying formula (3) to $\frac{it}{\sqrt{n}}\tilde{\varphi}$, we obtain an expression for the characteristic function

$$\begin{aligned} \phi_n(t) &= \int \mathrm{L}_{\frac{it}{\sqrt{n}}\tilde{\varphi}} \mathbf{1}(x) \, \mathrm{d}\mu(x) \\ &= \lambda^n_{\frac{it}{\sqrt{n}}\tilde{\varphi}} \underbrace{\Big(\int \mathrm{P}_{\frac{it}{\sqrt{n}}\tilde{\varphi}} \mathbf{1} \, \mathrm{d}\mu + \int \big[\mathrm{R}/\lambda\big]^n_{\frac{it}{\sqrt{n}}\tilde{\varphi}} \mathbf{1} \, \mathrm{d}\mu \Big)}_{=:A} \end{aligned}$$

where μ is the law of X_0. From now on, we assume

$$\sqrt{n} \geq \frac{\|t\tilde{\varphi}\|}{\log\left(1+\frac{\delta_0^2}{13+12\delta_0}\right)} \quad \text{and} \quad \sqrt{n} \geq 1 + \frac{\log 100}{-\log(1-\delta_0/13)}$$

to apply the estimates from Section 5. We will use later that this condition, considering the extremal case $\delta_0 = 1$, implies $n \geq 3311$.

We then get (Corollary 5.4):

$$A = \int \mathrm{P}_{\frac{it}{\sqrt{n}}\tilde{\varphi}}\mathbf{1}\,\mathrm{d}\mu + \lambda_{\frac{it}{\sqrt{n}}\tilde{\varphi}}^{-n}\int \mathrm{R}_{\frac{it}{\sqrt{n}}\tilde{\varphi}}^{n}\mathbf{1}\,\mathrm{d}\mu = 1 + O_{3.668+4.121\delta_0^{-1}}(\|\frac{t}{\sqrt{n}}\tilde{\varphi}\|).$$

We also have from Corollary 5.4

$$\lambda_{\frac{it}{\sqrt{n}}\tilde{\varphi}}^{n} = \exp\Big(-\frac{t^2}{2} + O_{10.89+20.04\delta_0^{-1}+8.577\delta_0^{-2}}(\frac{1}{\sqrt{n}}\|t\tilde{\varphi}\|^3)\Big).$$

In order to bound $|\phi_n(t) - \gamma(t)|$, following Feller [Fel66] we use that for all a, b, c such that $|a|, |b| \leq c$ and all $n \in \mathbb{N}$ it holds

$$|a^n - b^n| \leq n|a-b|c^{n-1}.$$

We take $a = \phi_n(t)^{\frac{1}{n}}$, $b = \gamma(t)^{\frac{1}{n}}$ and c an upper bound which we will now choose. Feller takes $c = e^{-\frac{t^2}{4n}}$, but we need two adaptations and take $c = 1.32^{\frac{1}{n}}e^{-\alpha\frac{t^2}{n}}$ where $\alpha \in (0, 0.5)$ will be optimized later on.

We have $\gamma(t)^{\frac{1}{n}} = e^{-\frac{t^2}{2n}} \leq c$ and

$$\phi_n(t)^{\frac{1}{n}} \leq e^{-\frac{t^2}{2n}}\exp\big((10.89 + 20.04\delta_0^{-1} + 8.577\delta_0^{-2})(\frac{1}{n^{3/2}}\|t\tilde{\varphi}\|^3)\big)A^{\frac{1}{n}}$$

where

$$\begin{aligned} A &\leq 1 + (3.834 + 4.121\delta_0^{-1})\|\frac{t}{\sqrt{n}}\tilde{\varphi}\| \\ &\leq 1 + (3.834 + 4.121\delta_0^{-1})\frac{\delta_0^2}{13+12\delta_0} \\ &\leq 1.32 \end{aligned}$$

To ensure $\phi_n(t)^{\frac{1}{n}} \leq c$, it is therefore sufficient that

$$(10.89 + 20.04\delta_0^{-1} + 8.577\delta_0^{-2})(\frac{1}{\sqrt{n}}\|t\tilde{\varphi}\|^3) \leq (0.5-\alpha)t^2$$

i.e. it is sufficient to ask

$$\sqrt{n} \geq \frac{10.89 + 20.04\delta_0^{-1} + 8.577\delta_0^{-2}}{0.5-\alpha}|t|\|\tilde{\varphi}\|^3 \tag{10}$$

Under this assumption, we have (using $n \geq 3311$ to bound $(n-1)/n$ by 0.9996):

$$|\phi_n(t) - \gamma(t)| \leq 1.32ne^{-0.9996\alpha t^2}\big|\phi_n(t)^{\frac{1}{n}} - \gamma(t)^{\frac{1}{n}}\big|. \tag{11}$$

Now we will bound $|\phi_n(t)^{\frac{1}{n}} - \gamma(t)^{\frac{1}{n}}|$, starting by a finer evaluation of A:

$$\begin{aligned} A^{\frac{1}{n}} &= (1 + O_{3.834+4.121\delta_0^{-1}}(\|\frac{t}{\sqrt{n}}\tilde{\varphi}\|))^{\frac{1}{n}} \\ &\leq \exp\big(\frac{1}{n^{3/2}}(3.834 + 4.121\delta_0^{-1})\|t\tilde{\varphi}\|\big)\big) \end{aligned}$$

By our assumptions the argument of the exponential is not greater than

$$\frac{1}{n}(3.834 + 4.121\delta_0^{-1})\log\Big(1 + \frac{\delta_0^2}{13 + 12\delta_0}\Big) \leq \frac{1}{3311}\frac{3.834\delta_0^2 + 4.121\delta_0}{13 + 12\delta_0} \leq 0.0001$$

and using $e^{0.0001} \leq 1.0002$, for all $\varepsilon \in [0, 0.0001]$ we have $\exp(\varepsilon) \leq 1 + 1.0002\varepsilon$ and therefore:

$$A^{\frac{1}{n}} \leq 1 + \frac{3.835 + 4.122\delta_0^{-1}}{n^{3/2}}\|t\tilde{\varphi}\|$$

Now we have

$$\big|\phi_n(t)^{\frac{1}{n}} - \gamma(t)^{\frac{1}{n}}\big| \leq \Big|\lambda_{\frac{it}{\sqrt{n}}\tilde{\varphi}}(1 + \frac{3.835 + 4.122\delta_0^{-1}}{n^{3/2}}\|t\tilde{\varphi}\|) - 1 + \frac{t^2}{2n}\Big| + \Big|e^{-\frac{t^2}{2n}} - 1 + \frac{t^2}{2n}\Big|$$

Since for all $x \in [0, +\infty[$ we have $0 \leq e^{-x} - 1 + x \leq \frac{1}{2}x^2$, the second summand is bounded above by $\frac{t^4}{8n^2}$. In the first summand we use (Lemma 5.2, definition of σ^2 and normalization of $\tilde{\varphi}$)

$$\lambda_{\frac{it}{\sqrt{n}}\tilde{\varphi}} = 1 - \frac{t^2}{2n} + O_{7.41+17.75\delta_0^{-1}+8.49\delta_0^{-2}}\big(\|\frac{t}{\sqrt{n}}\tilde{\varphi}\|^3\big).$$

The lower order terms simplify and we obtain

$$\begin{aligned} \big|\phi_n(t)^{\frac{1}{n}} - \gamma(t)^{\frac{1}{n}}\big| &\leq \Big|O_{7.41+17.75\delta_0^{-1}+8.49\delta_0^{-2}}(\|\frac{t}{\sqrt{n}}\tilde{\varphi}\|^3) + \lambda_{\frac{it}{\sqrt{n}}\tilde{\varphi}}\frac{3.835 + 4.122\delta_0^{-1}}{n^{3/2}}\|t\tilde{\varphi}\|\Big| + \frac{t^4}{8n^2} \\ &\leq \frac{1}{n^{3/2}}\Big((7.41 + 17.75\delta_0^{-1} + 8.49\delta_0^{-2})\|t\tilde{\varphi}\|^3 \\ &\qquad + 1.0524(3.835 + 4.122\delta_0^{-1})\|t\tilde{\varphi}\|\Big) + \frac{t^4}{8n^2} \\ &\leq \frac{(7.41 + 17.75\delta_0^{-1} + 8.49\delta_0^{-2})\|t\tilde{\varphi}\|^3 + (4.036 + 4.338\delta_0^{-1})\|t\tilde{\varphi}\|}{n^{3/2}} + \frac{t^4}{8n^2} \end{aligned}$$

For all $T > 0$ such that the above conditions on n, t hold for all $t \in [-T, T]$, we have by Proposition 7.1:

$$\|F_n - G\|_\infty \leq \frac{1}{\pi}\int_{-T}^{T}\Big|\frac{\phi(t) - \gamma(t)}{t}\Big|\,\mathrm{d}t + \frac{24m}{\pi T}$$

$$\leq \frac{2.64}{\pi}\int_0^T ne^{-0.9996\alpha t^2}\left|\phi_n(t)^{\frac{1}{n}}-\gamma(t)^{\frac{1}{n}}\right|\,\mathrm{d}t+\frac{3.048}{T}$$
$$\leq \frac{2.64}{\pi\sqrt{n}}\int_0^\infty e^{-0.9996\alpha t^2}\left(d\|t\tilde{\varphi}\|^3+f\|t\tilde{\varphi}\|+gt^4\right)\,\mathrm{d}t+\frac{3.048}{T}$$

where $d = 7.41 + 17.75\delta_0^{-1} + 8.49\delta_0^{-2}$, $f = 4.036 + 4.338\delta_0^{-1}$ and, using $n \geq 3311$, $g = 0.0022$. We want to take T as large as possible to lower the last term, but we need to ensure two conditions:

$$T \leq \frac{\sqrt{n}}{\|\tilde{\varphi}\|}\log\left(1+\frac{\delta_0^2}{13+12\delta_0}\right) \quad \text{and} \quad T \leq \frac{\sqrt{n}}{\|\tilde{\varphi}\|^3}\frac{(0.5-\alpha)}{10.89+20.04\delta_0^{-1}+8.577\delta_0^{-2}}$$

We could use here the lower bound on $\|\tilde{\varphi}\|$ to replace the left condition by a condition of the same form as the right one, but this would be too strong when $\|\tilde{\varphi}\|$ is far from the bound. We will make a choice which will be better when $\|\tilde{\varphi}\|$ is of the order of 1 (recall $\tilde{\varphi}$ is normalized, and therefore insensitive to scaling φ), by replacing the above conditions by the more stringent

$$T \leq \frac{\sqrt{n}}{\max\{\|\tilde{\varphi}\|,\|\tilde{\varphi}\|^3\}}\min\left\{\log\left(1+\frac{\delta_0^2}{13+12\delta_0}\right),\frac{(0.5-\alpha)}{10.89+20.04\delta_0^{-1}+8.577\delta_0^{-2}}\right\}$$

In the min, the first term is larger than $0.98\delta_0^2/(12+13\delta_0)$ which is easily seen to be larger than the second term for all δ_0. We thus take

$$T = \frac{\sqrt{n}}{\max\{\|\tilde{\varphi}\|,\|\tilde{\varphi}\|^3\}}\frac{(0.5-\alpha)}{10.89+20.04\delta_0^{-1}+8.577\delta_0^{-2}}$$

and we obtain

$$\|F_n-G\|_\infty \leq \frac{2.64}{\pi\sqrt{n}}\int_0^{+\infty} e^{-0.9996\alpha t^2}\left(d\|\tilde{\varphi}\|^3t^3+f\|\tilde{\varphi}\|t+gt^4\right)\,\mathrm{d}t$$
$$+\frac{(33.193+61.082\delta_0^{-1}+26.082\delta_0^{-2})\max\{\|\tilde{\varphi}\|,\|\tilde{\varphi}\|^3\}}{(0.5-\alpha)\sqrt{n}}$$

We have for $d = 1, 3, 4$:

$$\int_0^{+\infty} e^{-0.9996\alpha t^2}t^d\,\mathrm{d}t = (0.9996\alpha)^{-\frac{d+1}{2}}\int_0^{+\infty} e^{-t^2}t^d\,\mathrm{d}t.$$

Since $\int_0^{+\infty} e^{-t^2}t^d\,\mathrm{d}t = \frac{1}{2}\Gamma(\frac{d+1}{2})$ we thus have

$$\|F_n-G\|_\infty \leq \frac{1.32}{\pi\sqrt{n}}\left(d(0.9996\alpha)^{-2}\|\tilde{\varphi}\|^3+f(0.9996\alpha)^{-1}\|\tilde{\varphi}\|+g(0.9996\alpha)^{-\frac{5}{2}}\Gamma(\frac{5}{2})\right)$$
$$+\frac{(33.193+61.082\delta_0^{-1}+26.082\delta_0^{-2})\max\{\|\tilde{\varphi}\|,\|\tilde{\varphi}\|^3\}}{(0.5-\alpha)\sqrt{n}}$$

We will now choose α, by comparing the two most troublesome coefficients $\frac{1.32a}{\pi(0.9996\alpha)^2}$, which is close to $\frac{0.42d}{\alpha^2}$ (and makes us want to take α large), and $\frac{(33.193+61.082\delta_0^{-1}+26.082\delta_0^{-2})}{0.5-\alpha}$ which is somewhat close to $\frac{2.9a}{0.5-\alpha}$ when δ_0 is small (and makes us want to take α small). This leads us to take $\alpha = 0.2$. We then get

$$\begin{aligned}\|F_n - G\|_\infty &\le \frac{1}{\sqrt{n}}\Big((77.9 + 186.6\delta_0^{-1} + 89.26\delta_0^{-2})\|\tilde\varphi\|^3 + (8.49 + 9.12\delta_0^{-1})\|\tilde\varphi\| + 0.069 \\ &\qquad + (110.65 + 203.61\delta_0^{-1} + 86.94\delta_0^{-2})\max\{\|\tilde\varphi\|, \|\tilde\varphi\|^3\}\Big) \\ &\le \frac{1}{\sqrt{n}}\Big((197.04 + 399.33\delta_0^{-1} + 176.2\delta_0^{-2})\max\{\|\tilde\varphi\|, \|\tilde\varphi\|^3\} + 0.069\Big)\end{aligned}$$

Using $\|\tilde\varphi\| \ge \sqrt{\delta_0/3}$ and since $\delta_0 \le 1$, we obtain $0.069 \le 0.36\max\{\|\tilde\varphi\|, \|\tilde\varphi\|^3\}\delta_0^{-2}$ so that

$$\|F_n - G\|_\infty \le \frac{\max\{\|\tilde\varphi\|, \|\tilde\varphi\|^3\}}{\sqrt{n}}\big(197.04 + 399.33\delta_0^{-1} + 176.56\delta_0^{-2}\big)$$

and finally

$$\|F_n - G\|_\infty \le 177(\delta_0^{-2} + 2.26\delta_0^{-1} + 1.1)\frac{\max\{\|\tilde\varphi\|, \|\tilde\varphi\|^3\}}{\sqrt{n}} \le 177(\delta_0^{-1} + 1.13)^2\frac{\max\{\|\tilde\varphi\|, \|\tilde\varphi\|^3\}}{\sqrt{n}}$$

which is Theorem C.

References

[Bal00] Viviane Baladi, *Positive transfer operators and decay of correlations*, Advanced Series in Nonlinear Dynamics, vol. 16, World Scientific Publishing Co., Inc., River Edge, NJ, 2000. MR 1793194 (2001k:37035) 2.4

[Bol82] Erwin Bolthausen, *The berry-esseen theorem for strongly mixing harris recurrent markov chains*, Probability Theory and Related Fields **60** (1982), no. 3, 283–289. 1, 2.4

[BT08] Henk Bruin and Mike Todd, *Equilibrium states for interval maps: potentials with* $\sup\phi - \inf\phi < h_{top}(f)$, Comm. Math. Phys. **283** (2008), no. 3, 579–611. MR 2434739 2.4

[CP90] Zaqueu Coelho and William Parry, *Central limit asymptotics for shifts of finite type*, Israel J. Math. **69** (1990), no. 2, 235–249. MR 1045376 7

[CS09] Van Cyr and Omri Sarig, *Spectral gap and transience for Ruelle operators on countable Markov shifts*, Comm. Math. Phys. **292** (2009), no. 3, 637–666. MR 2551790 2.4

[CV13] A. Castro and P. Varandas, *Equilibrium states for non-uniformly expanding maps: decay of correlations and strong stability*, Ann. Inst. H. Poincaré Anal. Non Linéaire **30** (2013), no. 2, 225–249. MR 3035975 2.4

[Dub11] Loïc Dubois, *An explicit berry-esséen bound for uniformly expanding maps on the interval*, Israel Journal of Mathematics **186** (2011), no. 1, 221–250. 1, 2.4

[Erd39] Paul Erdös, *On a family of symmetric bernoulli convolutions*, American Journal of Mathematics **61** (1939), no. 4, 974–976. 3.4

[Fel66] William Feller, *An introduction to probability theory and its applications. Vol. II*, John Wiley & Sons, Inc., New York-London-Sydney, 1966. MR 0210154 7, 7.1, 7

[GD12] David M Gómez and Pablo Dartnell, *Simple monte carlo integration with respect to bernoulli convolutions*, Applications of Mathematics **57** (2012), no. 6, 617–626. 3.4

[GKLMF15] Paolo Giulietti, Benoît R. Kloeckner, Artur O. Lopes, and Diego Marcon Farias, *The calculus of thermodynamical formalism*, arXiv:1508.01297, to appear in *J. Eur. Math. Soc.*, 2015. 2.10

[GO02] Peter W Glynn and Dirk Ormoneit, *Hoeffding's inequality for uniformly ergodic markov chains*, Statistics & probability letters **56** (2002), no. 2, 143–146. 1, 1, 3.2, **??**, 1, 3.2

[HH01] Hubert Hennion and Loïc Hervé, *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*, Lecture Notes in Mathematics, vol. 1766, Springer-Verlag, Berlin, 2001. 1

[JO10] Aldéric Joulin and Yann Ollivier, *Curvature, concentration and error estimates for Markov chain Monte Carlo*, Ann. Probab. **38** (2010), no. 6, 2418–2442. MR 2683634 1, 1, 3.3, 2, 3.3.2, 3.3.3, 3.4, 3.4

[KL99] Gerhard Keller and Carlangelo Liverani, *Stability of the spectrum for transfer operators*, Annali della Scuola Normale Superiore di Pisa-Classe di Scienze **28** (1999), no. 1, 141–152. 1

[KLMM05] Ioannis Kontoyiannis, Luis A Lastras-Montano, and Sean P Meyn, *Relative entropy and exponential deviation bounds for general markov chains*, Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on, IEEE, 2005, pp. 1563–1567. 1, 1, 2.2, 3.2, **??**, 1, 3.2

[Klo17a] Benoît R. Kloeckner, *Effective high-temperature estimates for intermittent maps*, To appear in Ergodic Theory Dynam. Systems, arXiv:1704.00586, 2017. 2.5, 2.4

[Klo17b] ———, *Effective perturbation theory for linear operators*, arXiv:1703.09425, 2017. 1, 4, 4.1, 4.2

[Klo17c] ———, *An optimal transportation approach to the decay of correlations for non-uniformly expanding maps*, arXiv:1711.08052, 2017. 2.4

[KM12] Ioannis Kontoyiannis and Sean P Meyn, *Geometric ergodicity and the spectral gap of non-reversible markov chains*, Probability Theory and Related Fields (2012), 1–13. 1

[Lez98] Pascal Lezaud, *Chernoff-type bound for finite Markov chains*, Ann. Appl. Probab. **8** (1998), no. 3, 849–867. MR 1627795 1, 2.2

[Lez01] ———, *Chernoff and berry–esséen inequalities for markov processes*, ESAIM: Probability and Statistics **5** (2001), 183–201. 1, 2.2, 2.4

[Nag57] S. V. Nagaev, *Some limit theorems for stationary Markov chains*, Teor. Veroyatnost. i Primenen. **2** (1957), 389–416. MR 0094846 1

[Nag61] ———, *More exact limit theorems for homogeneous Markov chains*, Teor. Verojatnost. i Primenen. **6** (1961), 67–86. MR 0131291 1

[Oll09] Yann Ollivier, *Ricci curvature of Markov chains on metric spaces*, J. Funct. Anal. **256** (2009), no. 3, 810–864. MR 2484937 3.3.1

[Pau15] Daniel Paulin, *Concentration inequalities for markov chains by marton couplings and spectral methods*, Electronic Journal of Probability **20** (2015). 1

[Pau16] ———, *Mixing and concentration by ricci curvature*, Journal of Functional Analysis **270** (2016), no. 5, 1623–1662. 1

[Pet17] Fedor Petrov, *Answer to "diameter of a weighted hamming cube"*, MathOverflow, 2017, https://mathoverflow.net/a/286346/4961. 3.5

[PSS00] Yuval Peres, Wilhelm Schlag, and Boris Solomyak, *Sixty years of bernoulli convolutions*, Progress in probability (2000), 39–68. 3.4

[RR+04] Gareth O Roberts, Jeffrey S Rosenthal, et al., *General state space markov chains and mcmc algorithms*, Probability Surveys **1** (2004), 20–71. 1

[Rue04] David Ruelle, *Thermodynamic formalism*, second ed., Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2004, The mathematical structures of equilibrium statistical mechanics. MR 2129258 (2006a:82008) 2.4

[Sol95] Boris Solomyak, *On the random series* $\sum \pm\lambda^n$ *(an erdos problem)*, Annals of Mathematics (1995), 611–625. 3.4

[Tyu11] I. S. Tyurin, *Improvement of the remainder in the Lyapunov theorem*, Teor. Veroyatn. Primen. **56** (2011), no. 4, 808–811. MR 3137072 2.4