HAL open science

Effective limit theorems for Markov chains with a spectral gap

Benoît Kloeckner

To cite this version:

Benoît Kloeckner. Effective limit theorems for Markov chains with a spectral gap. 2017. hal-01497377v1

HAL Id: hal-01497377

https://hal.science/hal-01497377v1

Preprint submitted on 28 Mar 2017 (v1), last revised 2 Mar 2019 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Effective limit theorems for Markov chains with a spectral gap

Benoît R. Kloeckner *

March 28, 2017

Applying quantitative perturbation theory for linear operators, we prove non-asymptotic limit theorems for Markov chains whose averaging operator has a spectral gap in a suitable function space. The main results are concentration inequalities and Berry-Esséen bounds, with some flexibility in the choice of the function space and no "warm start" hypothesis nor burn-in: the law of the first term of the chain can be arbitrary. The constants are completely explicit and reasonable enough to make the results usable in practice, notably in MCMC methods.

1 Introduction

The goal of this article is to prove *effective* limit theorems for some Markov chains. By effective, it is meant that given an explicit sample size n, one should be able to deduce from the result that the quantity being considered lies in some explicit interval 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 estimate an integral using samples from a Markov chain, effective results are needed to obtain proved 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 is a prominent example of such an approach.

It is well-known that assuming a *spectral gap* (which will be our main assumption) is sufficient to obtain most classical limit theorems, notably the Law of Large Number and the Central Limit Theorem, and their quantitative counterparts: concentration

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

inequalities and Berry-Esséen bounds. While the question of improving the constant in the Berry-Esséen theorem for independent identically distributed random variable has been the object of many works,[1] these quantitative results are rarely made effective for Markov chains. Since the constants tend to depend on a number of parameters, it is not usually clear how much uniformity holds, nor what is the behavior of the constants when parameters approach the limits of the domain of their admissible values (e.g. when the spectral gap goes to zero). This is a second, more theoretical motivation to look for effective results: one then obtain precise information on the dependence of the convergence speed on the parameters.

Although effective results are scarce, many previous works have considered related questions, and it would be a daunting task to give all relevant references. Let us consider a few works that make good entry points in the literature. Hennion and Hervé [HH01] give a precise and general account of the main method we shall use below, in the more general setting of *quasi-compact* operators. They credit Nagaev [Nag57, Nag61] for giving birth to this method. Joulin and Ollivier [JO10] seem to have proved the first truly effective concentration inequalities for Markov chains, and we shall compare some of the present results to theirs. Among the references they give, Lezaud [Lez98] is especially relevant, as he also relies on perturbation theory (but he works with finitely many states). Joulin and Ollivier observe that they can lift three assumptions needed by Lezaud: reversibility, warm start (asking the law of X_0 to have a density with respect to the stationary measure), and boundedness of the observable. Below it will be shown that the perturbation method can in fact dispense at least from the first two conditions. Last, Dubois [Dub11] gave what seems to be the first effective Berry-Esséen inequality in a dynamical or Markov chain context, and we shall compare the present Berry-Esséen inequality with his.

Structure of the article. In Section 2 we state notation and the main results. Section 3 is devoted to a few examples, for which we compare our concentration inequalities with Joulin-Ollivier's result. 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 and Section 7 proves the Berry-Esséen 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

[1] To my knowledge the currently known best bound in the IID case is due to Tyurin [Tyu11].

singular with respect to the stationary measure.

We shall study the behavior of $(X_k)_{k\geq 0}$ through the empirical mean

$$\hat{\mu}_n(\varphi) := \frac{1}{n}\sum_{k=1}^{n} \varphi(X_k)$$

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

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

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

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

iii. $\mathcal{X}(\Omega)$ contain the constant functions and $\|\mathbf{1}\| = 1$.

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 $\mathcal{X}(\Omega)$. 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 naturally associated an averaging operator acting on $\mathcal{X}(\Omega)$, defined by

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

Since the (m_x) are probability measures, 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 $\mathcal{X}(\Omega)$ 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 *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 that 1 is a simple isolated eigenvalue. This ensures that up to scalar factors there is a unique continuous linear form ϕ_0 acting on $\mathcal{X}(\Omega)$ such that $\phi_0 \circ \mathrm{L}_0 = \phi_0$; since any stationary measure of M satisfy this, all stationary measures coincide on $\mathcal{X}(\Omega)$. They might not be unique (e.g. if $\mathcal{X}(\Omega)$ contains only constants), but since we consider the $\varphi(X_k)$ with $\varphi \in \mathcal{X}(\Omega)$, 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, $\mathcal{X}(\Omega)$ will be dense in the space of continuous function endowed with the uniform norm, ensuring that two measures coinciding on $\mathcal{X}(\Omega)$ are equal, and then the spectral gap hypothesis ensures the uniqueness of the stationary measure.

Remark 2.4. There are numerous examples where assumptions 2.1 and 2.3 are satisfied; we will discuss a few of them in Section 3. Typically, $\mathcal{X}(\Omega)$ 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 ensure $\mathcal{X}(\Omega)$ 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 $f \in \mathcal{X}(\omega)$) 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.

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

Theorem 2.5. *For all* $n \geq 60/\delta_0$ *and all* $a > 0$*, it holds*

$$\mathbb{P}_\mu\Big[|\hat{\mu}_n(\varphi) - \mu_0(\varphi)| \geq a\Big] \leq \begin{cases} 2.5\exp\big(-\dfrac{na^2}{\|\varphi\|^2}\cdot C_1\delta_0\big) & \text{if } \dfrac{a}{\|\varphi\|} \leq \delta_0/3 \\ 2.7\exp\big(-\dfrac{na}{\|\varphi\|}\cdot C_2\delta_0^2\big) & \text{otherwise} \end{cases}$$

and one can take $C_1 = 0.046$ *and* $C_2 = 0.009$*.*

Remark 2.6. In Theorem 2.5 we tried to get the cleaner and simpler possible statement, but in fact we obtain more precise estimates, which are slightly more complicated to state:

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

in the Gaussian regime ($a \leq \frac{\delta_0\|\varphi\|}{3}$) and

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

in the exponential regime ($a > \frac{\delta_0\|\varphi\|}{3}$). Moreover the condition on n can be relaxed to

$$n \geq 1 + \frac{\log 100}{-\log(1-\delta_0/13)}.$$

We apply this result and compare it to [JO10] on a few examples in Section 3. It turns out that our constants are sometimes a bit disappointing, but 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". This quantity is defined as

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

and is precisely the variance appearing in the CLT for $(\varphi(X_k))_{k\geq 0}$.

Theorem 2.7. *If $\sigma^2(\varphi) \leq V$, then for all $a \leq \dfrac{V\delta_0^2}{26\|\varphi\|}$ and all $n \geq 60/\delta_0$ it holds*

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

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 more precise result: the assumption can be replaced by

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

and a slightly more precise conclusion can be found in Section 6.3.

We end with a Berry-Esséen bound, proved in section 7, quantifying the speed of convergence in the Central Limit Theorem.

Theorem 2.9. *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 2.10. 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 2.11. 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 $a/\|\varphi\| \gg 1/\sqrt{n}$.

We know of only one previous result in the same flavor of Theorem 2.9, by Dubois [Dub11]. However the scope of Dubois' result is somewhat 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).

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 $A \in \mathcal{X}(\Omega)$ 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 L_φ 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 $0 \leq k \leq n$) 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 2.12. *For all normalized $A \in \mathcal{X}(\Omega)$ such that $\mathrm{L}_{T,A}$ has a spectral gap with constant 1 and size δ_0, for all $\varphi \in \mathcal{X}(\Omega)$, Theorems 2.5, 2.7 and 2.9 hold for the random process $(X_k)_{k\in\mathbb{N}}$ defined by $X_0 \sim \mu_A$ and $X_{k+1} = T(X_k)$.*

3 Examples

Let us consider some examples to apply our estimates to. We will use several times the following lemma which, in the spirit of Doeblin-Fortet and Lasota-Yorke, 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 $\mathcal{X}(\Omega)$ 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 $f \in \mathcal{X}(\Omega)$ such that $\mu(f) = 0$, $\|f\|_\infty \le CV(f)$.

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

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

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 $\|f\|_\infty \le CV(f)$ is often valid in practice when Ω has finite diameter: the condition that $\mu(f) = 0$ implies that f vanishes (if functions in $\mathcal{X}$ 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 $\|\mathrm{L}_0 f\|_\infty \le \|f\|_\infty$ and $\mathrm{L}_0 f \in \ker \mu_0$, so that $\|\mathrm{L}_0 f\|_\infty \le CV(\mathrm{L}_0 f) \le C\theta V(f)$.

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 $\|\mathrm{L}_0(f)\|_\infty \le \min\big(t, C\theta(1-t)\big)\|f\|$ and using $V(\mathrm{L}_0 f) \le \theta V(f)$ again we get

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

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. □

3.1 Discrete hypercube

Let us start with the same base example as Joulin and Ollivier [JO10], the discrete hypercube $\{0,1\}^N$ 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 $x \sim y$. Consider the random walk M which chooses randomly uniformly a slot $i \in \{1, \dots, N\}$ and replace it with the result of a fair coin toss, i.e.

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

Then, in the parlance of [Oll09], M is positively curved with $\kappa = 1/N$, i.e. $\mathrm{Lip}(\mathrm{L}_0 f) \leq (1 - 1/N)\,\mathrm{Lip}(f)$ for all Lipschitz function $f : \{0,1\}^N \to \mathbb{R}$.

Example 3.2. Consider the observable $\varphi_1 : \{0,1\}^N \to \mathbb{R}$ where $\varphi_1(x)$ is the proportion of 1's in the word x.

Endowing the space of Lipschitz potentials with the Lipschitz norm $\|\cdot\|_{\mathrm{Lip}} = \|\cdot\|_\infty + N\,\mathrm{Lip}(\cdot)$, Lemma 3.1 applies with $C = 1$ so that L_0 has a spectral gap on this space with constant 1 and size $\delta_0 = 1/(2N-1)$. We have $\mathrm{Lip}(\varphi_1) = 1/N$ and $\|\varphi_1\|_\infty = 1$, so that this choice yields $\|\varphi_1\|_{\mathrm{Lip}} = 2$.

For $a \leq \frac{2}{6N-3}$, as soon as $n \geq 120N - 60$ Theorem 6.3 (see Remark 2.6) yields

$$\mathbb{P}_\mu\left[|\hat{\mu}_n(\varphi_1) - \mu_0(\varphi_1)| \geq a\right] \leq 2.5 \exp\Big(-\frac{na^2}{67N+21}\Big)$$

We thus need $O(N/a^2)$ 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 $O(N^2/a)$ according to the bound of Joulin and Ollivier, for a total run time of $O(N^2/a + 1/a^2)$. This bound has the same order as ours if a is close to the boundary of our Gaussian window (in $1/N$), but is better for smaller a.

For $1/N \lesssim a \lesssim 1$, we enter our exponential regime while staying inside Joulin-Ollivier's Gaussian window; Theorem 6.4 (Remark 2.6) shows we need no more than $O(N^2/a)$ iterations, while [JO10] still gives a bound of $O(N^2/a + 1/a^2) = O(N^2/a)$. In this regime, the concentration around the expectancy of the empirical process appears to be faster than the decay of bias, and our bound approximately match the bound in [JO10].

Example 3.3. Consider now the potential $\mathbf{1}_S$, the indicator function for a (non-trivial) set S. We have $\mathrm{Lip}(\mathbf{1}_S) = 1$ and $\|\mathbf{1}_S\| = 1$, and in this case it is more efficient to use the norm $\|\cdot\|_\infty + \mathrm{Lip}(\cdot)$: then Lemma 3.1 holds with $C = N$, we get a spectral gap of size at least $\delta_0 = 1/N^2$ and $\mathbf{1}_S$ has norm 2. For $a \leq 2/3N^2$ we get from Theorem 6.3 (Remark 2.6)

$$\mathbb{P}_\mu\left[|\hat{\mu}_n(\mathbf{1}_S) - \mu_0(\mathbf{1}_S)| \geq a\right] \leq 2.5 \exp\Big(-\frac{na^2}{34N^2+54}\Big)$$

so that we need $O(N^2/a^2)$ iterations to ensure a small probability of an error larger than a. This is the same order than the concentration time given by [JO10], but with a worse constant (34 instead of 8).

For $1 \gg a \gg 2/3N^2$ we get from Theorem 6.4 (Remark 2.6) a bound of the order of $\exp(-na/(25N^4))$ (neglecting the terms going to zero with N in the exponent) and thus a need for $O(N^4/a)$ iterations. In [JO10], this regime still belongs to the Gaussian window and they need $O(N^2/a^2)$ iterations: we trade two powers of N for a power of a, which is not a good deal.

Now, for sufficiently small a we can get a much better bound from Theorem 2.7. We consider the case when

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

where the variance can be computed explicitly[2] (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 $1/2N$):

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

This gives $\sigma^2(\mathbf{1}_S) \simeq N/2$.

If we use the norm $\|\cdot\|_\infty + \mathrm{Lip}(\cdot)$, then condition to apply Theorem 2.7 is $a \lesssim 1/48N^3$ (see Remark 2.8) on top of $n \geq 60N^2$. Under these conditions the positive term in the exponential is of the order of na^3N, 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 exponent) but only for very small values of a. Here we can play on two parameters to extend the result: the norm and the choice of V. First, switching back to the norm $\|\cdot\|_\infty + N\,\mathrm{Lip}(\cdot)$ is better here: we get $\delta_0 \simeq 1/2N$ and $\|\mathbf{1}_S\| \simeq N$, and for $V = \sigma^2(\mathbf{1}_S)$ the allowed window is $a \lesssim 1/96N^2$, one power of N larger than before. The remainder term in the exponential is still negligible compared to the leading term, which is left unchanged. We thus keep the bound $O(N/a^2)$ on iterations, but on a larger window. If we want to consider a of the order of $1/N$, we can then take $V \simeq N^2$ to enlarge the window, at the cost of a weaker leading term. We still get a bound similar to the one of Joulin-Ollivier, possibly with a smaller constant (depending on the value of a).

This choice of S might seem very specific, but for less regular S the gain should be greater for sufficiently smaller a.

[2] Among sets containing half the elements, this case should be the worst one since it corresponds to the less scrambled S, and a more scrambled S should produce smaller autocorrelations. I currently do not know how to prove such a result, which would be a kind of discrete Levy-Gromov isoperimetric inequality.

3.2 Iterated Function Systems

A number of fractals are defined as the attractor of an IFS (Iterated Function System), and natural probability measures can be constructed which are supported on the attractor. These measures are the stationary measures of certain random walks, the support of whose empirical process can be used to get a good approximation of the attractor.

More precisely, consider a system $\mathcal{F}$ of K maps $f_1, \dots, f_K : \mathbb{R}^N \to \mathbb{R}^N$. Then $\mathcal{F}$ acts on non-empty compact sets by $\mathcal{F}(C) := \cup_k f_k(C)$, and an attracting fixed non-empty compact set A is called an attractor of $\mathcal{F}$.

Fix a probability vector $P = (p_1, \dots, p_K) \in [0,1]^K$ (i.e. $\sum_k p_k = 1$) and consider the Markov chain $\mathsf{M} = (m_x)_{x \in \mathbb{R}^N}$ where

$$m_x = \sum_{k=0}^{K} p_k \delta_{f_k(x)},$$

i.e. from a point x we jump to its image by one of the maps f_k, chosen randomly according to the chosen probability vector, and let $X_0, X_1, \dots$ be a Markov chain following this transition kernel, with arbitrary start X_0.

Example 3.4. Assume that $\mathcal{F}$ is contracting, i.e. for some $\theta \in (0,1)$, all $x, y \in \mathbb{R}^N$ and all $k \in \{1, \dots, K\}$ it holds

$$|f_k(x) - f_k(y)| \le \theta |x - y|$$

where $|\cdot - \cdot|$ denotes the usual Euclidean distance (or another norm, for that matter). Then the action of $\mathcal{F}$ on compact sets is a θ-contraction in the Hausdorff metric, and it follows that there exists a unique attractor A.

It is also easily seen that the Markov chain has positive Ollivier-Ricci curvature at least $\kappa = 1-\theta$, i.e. is θ-contracting in the Wasserstein metric (it is sufficient to check this on Dirac measures, in which case one can use the obvious coupling $\sum_k p_k \delta_{f_k(x)} \otimes \delta_{f_k(y)}$ between m_x and m_y). In particular there is a unique stationary measure μ_P, which must be concentrated on the attractor A.

If $\varphi : \mathbb{R}^N \to \mathbb{R}$ is a Lipschitz observable, we can therefore use our results to estimate the behavior of $|\hat{\mu}_n(\varphi) - \mu_P(\varphi)|$.

Denote by c_k the unique fixed point of f_k and set $S = \max_{k,k'} |c_k - c_{k'}|$. Each ball of radius S/κ centered at one of the c_k is then sent into itself by each $f_{k'}$, so that $A \subset \bigcap_k B(c_k, S/\kappa)$. We can thus restrict to a domain Ω of some finite diameter D (not greater than $2S/\kappa$). Assume for simplicity that φ vanishes somewhere in Ω and is 1-Lipschitz, so that $\|\varphi\|_\infty \le D$. We consider here the norm $\|\cdot\|_\infty + D\,\mathrm{Lip}(\cdot)$, since it gives φ a norm of the same order ($\le 2D$) and makes it possible to apply Lemma 3.1 with $C = 1$, so that $\delta_0 = (1-\theta)/(1+\theta)$.

In the notation of [JO10], using the estimates $\sigma(x)^2 \le D/2$ (which could arguably be improved), $n_x \ge 1$, $\sigma_\infty \le D/2$, we get $V^2 = \frac{D}{2(1-\theta)^2 n}$, $r_{\max} = \frac{4}{3(1-\theta)}$ and in the Gaussian window $a \le r_{\max}$ [JO10] tells us

$$\mathbb{P}_\mu \Big[|\hat{\mu}_n(\mathbf{1}_S) - \mu_0(\mathbf{1}_S)| \ge a \Big] \le 2 \exp\Big(- n \frac{a^2 (1-\theta)^2}{8D} \Big).$$

Meanwhile, Theorem 6.3 (2.6) gives for $a \leq \frac{2D(1-\theta)}{3(1+\theta)}$ and $n \geq 60(1+\theta)/(1-\theta)$:

$$\mathbb{P}_\mu \Big[|\hat{\mu}_n(\mathbf{1}_S) - \mu_0(\mathbf{1}_S)| \geq a \Big] \leq 2.5 \exp\Big(- n \frac{a^2(1-\theta)}{(88-20\theta)D^2} \Big).$$

For D fixed and θ close to 1, we gain a factor $(1-\theta)$, but in some other regimes, such as D large and θ away from 1, the comparison is not in our favor.

Example 3.5. In some cases we are able to dispense from the contraction hypothesis above; for simplicity we consider only the dimension $N = 1$. Assume now that the family $f_1, \dots, f_K : [0,1] \to [0,1]$ is such that:

- each f_k is monotone (but they may be discontinuous),
- the convex hulls of the sets $f_k([0,1])$ have pairwise disjoint interiors.

We also assume that the probability vector P is non-degenerate, i.e. $\max P := \max\{p_1, \dots, p_K\} < 1$ (otherwise the Markov chain is deterministic).

We consider the Banach space $\mathrm{BV}([0,1])$ of *bounded variation* functions, defined by the norm $\|\cdot\|_{\mathrm{BV}} = \|\cdot\|_\infty + \mathrm{var}(\cdot, [0,1])$ where

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

(the uniform norm is usually replaced by the L^1 norm, but our choice is equivalent up to a constant, does not single out the Lebesgue measure, and ensures $\mathrm{BV}([0,1])$ is a Banach algebra).

Then, denoting as usual by L_0 the operator associated with M, we have

$$\begin{aligned}
\mathrm{var}(\mathrm{L}_0\varphi) &= \sup_{0 \leq x_0 < x_1 < \dots < x_p \leq 1} \sum_{j=1}^{p} |\mathrm{L}_0\varphi(x_j) - \mathrm{L}_0\varphi(x_{j-1})| \\
&\leq \sup_{0 \leq x_0 < x_1 < \dots < x_p \leq 1} \sum_{j=1}^{p} \sum_{k=1}^{K} p_k |\varphi(f_k(x_j)) - \varphi(f_k(x_{j-1}))| \\
&\leq \max(P) \sum_{k=1}^{K} \mathrm{var}(\varphi, f_k([0,1])) \\
&\leq \max(P)\, \mathrm{var}(\varphi).
\end{aligned}$$

By classical compactness arguments, there exists at least one stationary measure μ_0. Then $\ker \mu_0 \subset \mathrm{BV}([0,1])$ is an invariant hyperplane, and from Lemma 3.1 it follows that L_0 has a spectral gap of size $\delta_0 = \frac{1-\max P}{1+\max P}$. Since L_0^* has a unique eigenform and $\mathrm{BV}([0,1])$ contains the characteristic functions of the intervals, μ_0 is the *unique* stationary measure.

All our results above thus apply to all $\varphi \in \mathrm{BV}([0,1])$; note that since $\mathrm{var}(\varphi) \leq \mathrm{Lip}(\varphi)$, they also apply to Lipschitz φ, even though we had to pass through another norm to

obtain them. If the f_k where additionally θ-Lipschitz for some θ very close to 1, [JO10] could have been used but would have given a very weak concentration speed; when the f_k are not contracting (e.g. not continuous) no concentration at all could have been deduced from their results.

Remark 3.6. In Example 3.4, one can also consider Hölder observable by observing that the IFS is still contracting (with ratio θ^α) in the distance $|\cdot - \cdot|^\alpha$. Similarly, one can consider $p-$BV observables in Example 3.5, which in particular include $1/p$-Hölder observables.

4 Connection with perturbation theory

To any $\varphi \in \mathcal{X}(\Omega)$ (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

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

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

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 $\mathcal{X}$ 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 [Klo17] 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, $|f(x)| \le C|g(x)|$.

Theorem 4.1 (Theorems 2.3 and 2.6 and Proposition 5.1 (viii) of [Klo17])**.** *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* $\|\mathrm{L} - \mathrm{L}_0\| \le \dfrac{K-1}{6K\tau_0\gamma_0}$ *we have*

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

Theorem 4.2 (Corollaire 2.12 from [Klo17])**.** *In the case* $\lambda_0 = \|\mathrm{L}_0\| = 1$*, all* L *such that*

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

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 $f \in \mathcal{X}(\Omega)$, we have $\pi_0(f) = f - \mu_0(f)$ and $\|\mu_0(f)\mathbf{1}\| = |\mu_0(f)| \leq \|f\|_\infty \leq \|f\|$. 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 $(1 - \mathrm{L}_0)^{-1} = \sum_{k\geq 0} \mathrm{L}_0^k$, the spectral gap δ_0, and the upper bound on $\|\pi_0\|$ we deduce

$$\gamma_0 \leq 2/\delta_0.$$

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 $f \in \mathcal{X}(\Omega)$ 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:

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

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*

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

If (2) *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 (2) *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 $\mathcal{X}(\Omega)$ is a Banach algebra, we have

$$\begin{aligned}
\|\mathrm{L}_\varphi - \mathrm{L}_0\| &= \|\mathrm{L}_0\big((e^\varphi - 1)\cdot\big)\| \\
&\le \|\mathrm{L}_0\|\Big\|\sum_{j=1}^{\infty}\frac{\varphi^j}{j!}\Big\| \\
&\le \sum_{j=1}^{\infty}\frac{\|\varphi\|^j}{j!} \\
\|\mathrm{L}_\varphi - \mathrm{L}_0\| &\le e^{\|\varphi\|} - 1
\end{aligned}$$

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 $\|\varphi\| \le \log(1+1/25) := b$. By convexity, we deduce that

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

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:

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

The second order remainder is bounded by

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

and finally, we have

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

□

Lemma 5.2. *Under* (2) *we have*

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

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 $|\lambda_\varphi - \lambda_0| \le \frac{17}{13}\|\mathrm{L}_\varphi - \mathrm{L}_0\|$ from which we deduce both

$$
\begin{aligned}
&|\lambda_\varphi - 1| \le \frac{17}{13\times 25} \le 0.0524\\
\text{and}\qquad &|\lambda_\varphi - 1| \le \frac{17}{13}1.02\|\varphi\| \le 1.334\|\varphi\|
\end{aligned}
$$

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 $\ker \mu_0$):

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

We also have

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

It then comes

$$\begin{aligned}
(\mathrm{L}_\varphi - \mathrm{L}_0)\mathrm{S}_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1}) &= \mathrm{L}_0\big(\varphi \sum_{k \geq 1} \mathrm{L}_0^k(\bar{\varphi})\big) + O_{1.014\delta_0^{-1}}(\|\mathrm{L}_\varphi - \mathrm{L}_0\| \|\varphi\|^2) \\
&\quad + O_{0.507}(\|\varphi\|^2 \|\mathrm{S}_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1})\|) \\
&= \mathrm{L}_0\big(\varphi \sum_{k \geq 1} \mathrm{L}_0^k(\bar{\varphi})\big) + O_{2.07\delta_0^{-1}}(\|\varphi\|^3) \\
\mu_0(\mathrm{L}_\varphi - \mathrm{L}_0)\mathrm{S}_0(\mathrm{L}_\varphi \mathbf{1} - \mathrm{L}_0 \mathbf{1}) &= \sum_{k \geq 1} \mu_0(\varphi \mathrm{L}_0^k(\bar{\varphi})) + O_{2.07\delta_0^{-1}}(\|\varphi\|^3)
\end{aligned}$$

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 (2), 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 (2), it holds*

$$\Big\|\Big(\frac{1}{\lambda_\varphi}\mathrm{R}_\varphi\Big)^n\mathbf{1}\Big\| \le (6.388+4.08\delta_0^{-1})(1-\delta_0/13)^{n-1}\|\varphi\|$$
$$\mathrm{P}_\varphi\mathbf{1} = \mathbf{1} + O_{3.77+4.08\delta_0^{-1}}(\|\varphi\|).$$

Proof. At any $\mathrm{L}=\mathrm{L}_\varphi$ where φ satisfies (2) we have:

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

so that

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

Moreover since R_{L} takes its values in G_{L} where π_{L} acts as the identity, we have

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

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 (2) and*

$$n \ge 1+\frac{\log 100}{-\log(1-\delta_0/13)} \tag{3}$$

it holds

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

Proof. Assuming (2), we first appeal to Lemma 5.3 to write:

$$\begin{aligned}\mathrm{L}_\varphi^n\mathbf{1} &= \lambda_\varphi^n \mathrm{P}_\varphi\mathbf{1} + \mathrm{R}_\varphi^n\mathbf{1}\\ &= \lambda_\varphi^n\Big(\mathbf{1} + O_{3.77+4.08\delta_0^{-1}}(\|\varphi\|) + O_{6.388+4.08\delta_0^{-1}}\big((1-\delta_0/13)^{n-1}\|\varphi\|\big)\Big)\end{aligned} \tag{4}$$

The second factor of (4) is easily controlled if we ask (3), 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 $\log(\lambda_\varphi)$:

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

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:

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

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

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

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 (2) 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

$$n \geq \frac{\|t\varphi\|}{\log\left(1 + \frac{\delta_0^2}{12+13\delta_0}\right)} \quad \text{and} \quad n \geq 1 + \frac{\log 100}{-\log(1-\delta_0/13)}, \tag{5}$$

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 (1):

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

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

$$\mathbb{P}_\mu \left[|\hat{\mu}_n(\varphi) - \mu_0(\varphi)| \geq a\right]$$
$$\leq \big(2 + (7.668 + 8.242\delta_0^{-1})\frac{t}{n}\|\varphi\|\big) \exp\big(- at + (3.36 + 2.081\delta_0^{-1})\frac{t^2}{n}\|\varphi\|^2)\big) \quad (6)$$

6.1 Gaussian regime

Our first concentration inequality is obtained by choosing t to optimize the argument of the exponential in (6), 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.

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

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

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

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

$$a \leq \frac{\delta_0\|\varphi\|}{3}. \quad (7)$$

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 2.5.

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 (5), 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:

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

We obtain a version of the second part of Theorem 2.5.

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

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

it holds

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

A simpler, less precise estimate is:

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

6.3 Second-order concentration

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

$$\sigma^2(\varphi)=\mu_0(\varphi^2)-(\mu_0\varphi)^2+2\sum_{k\ge1}\mu_0(\varphi \mathrm{L}_0^k\bar\varphi)$$

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

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

so that, if we know $\sigma^2(\varphi)\le V$:

$$\mathbb{P}_\mu\big[|\hat\mu_n(\varphi)-\mu_0(\varphi)|\ge a\big]\le\big(2+\frac{(7.668+8.242\delta_0^{-1})t}{n}\|\varphi\|\big)\exp\big(-at+\frac{t^2}{2n}V+C\frac{t^3}{n^2}\|\varphi\|^3\big)$$

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 $0.89+0.04\delta_0^{-1}\le 0.93\delta_0^{-2}$ so that

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

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

This choice can be made as soon as

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

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

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

which, given $\hat{\mu}_n(\varphi)$ satisfies the Central Limit Theorem, is nearly optimal if $a \ll \frac{V^2}{2C\|\varphi\|^3}$ and V is close to $\sigma^2(\varphi)$. We obtain Theorem 2.7, in the following version.

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

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

then it holds

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

7 Berry-Esseen bounds

In this section, we use the second-order Taylor formula for the leading eigenvalue to prove effective Berry-Esséen 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 (1) 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

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

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):

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

We also have from Corollary 5.4

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

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

$$\sqrt{n} \geq \frac{10.89 + 20.04\delta_0^{-1} + 8.577\delta_0^{-2}}{0.5 - \alpha}|t|\|\tilde{\varphi}\|^3 \tag{8}$$

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

$$|\phi_n(t) - \gamma(t)| \leq 1.32ne^{-0.9996\alpha t^2}\big|\phi_n(t)^{\frac{1}{n}} - \gamma(t)^{\frac{1}{n}}\big|. \tag{9}$$

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

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

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

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

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

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

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:

$$\begin{aligned}
\|F_n - G\|_\infty &\leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\phi(t) - \gamma(t)}{t}\right| \mathrm{d}t + \frac{24m}{\pi T} \\
&\leq \frac{2.64}{\pi}\int_0^T ne^{-0.9996\alpha t^2}\left|\phi_n(t)^{\frac{1}{n}} - \gamma(t)^{\frac{1}{n}}\right| \mathrm{d}t + \frac{3.048}{T} \\
&\leq \frac{2.64}{\pi\sqrt{n}}\int_0^\infty e^{-0.9996\alpha t^2}\left(d\|t\tilde{\varphi}\|^3 + f\|t\tilde{\varphi}\| + gt^4\right) \mathrm{d}t + \frac{3.048}{T}
\end{aligned}$$

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

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

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

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

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

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

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

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

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

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

and finally

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

which is Theorem 2.9.

References

[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

[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

[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

[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. 2.11

[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, 2, 3.1, 3.2, 3.3, 3.4, 3.5

[Klo17] Benoît R. Kloeckner, *Effective perturbation theory for linear operators*, 2017. 4, 4.1, 4.2

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

[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 3.1

[Tyu11] I. S. Tyurin, *Improvement of the remainder in the Lyapunov theorem*, Teor. Veroyatn. Primen. **56** (2011), no. 4, 808–811. MR 3137072 1