1

Exponential inequalities and estimation of conditional probabilities

V. Maume-Deschamps

Université de Bourgogne B.P. 47870 21078 Dijon Cedex FRANCE
vmaume@u-bourgogne.fr

Summary. We prove exponential inequalities inspired from [DP] to obtain estimators of conditional probabilities for weak dependant sequences. This generalize results from Csiszár ([Cs]). For Gibbs measures and dynamical systems, these results lead to construct estimators of the potential function.

This paper deals with the problems of typicality and conditional typicality of "empirical probabilities" for stochastic process and the estimation of potential functions for Gibbs measures and dynamical systems. The questions of typicality have been studied in [FKT] for independent sequences, in [BRY, R] for Markov chains. In order to prove the consistency of estimators of transition probability for Markov chains of unknown order, results on typicality and conditional typicality for some (Ψ)-mixing process where obtained in [CsS, Cs]. Unfortunately, lots of natural mixing process do not satisfy this Ψ-mixing condition (see [DP]). We consider a class of mixing process inspired from [DP]. For this class, we prove strong typicality and strong conditional typicality. In the particular case of Gibbs measures (or complete connexions chains) and for certain dynamical systems, from the typicality results we derive an estimation of the potential as well as a procedure to test the nullity of the asymptotic variance of the process.

More formally, we consider X_0, ..., X_n, ... a stochastic process taking values on an complete set Σ and a sequence of countable partitions of Σ, $(\mathcal{P}_k)_{k\in\mathbb{N}}$

Thanks : Many ideas of the paper have been discussed with Bernard Schmitt, it is a pleasure to thank him. I am grateful to Jérôme Dedecker who patiently answered my questions on weak-dependence coefficients. I whish also to thank the referees of this paper for many interesting remarks and for the suggestion of the proof of Corollary 1.2.9.
Keywords : Exponential inequalities, weak dependent sequences, dynamical systems.
AMS clasification : 37A50, 60E15, 37D20

such that if $P \in \mathcal{P}_k$ then there exists a unique $\widetilde{P} \in \mathcal{P}_{k-1}$ such that almost surely, $X_j \in P \;\Rightarrow\; X_{j-1} \in \widetilde{P}$. Our aim is to obtain empirical estimates on the probabilities :

$$\mathbb{P}(X_j \in P),\ P \in \mathcal{P}_k,$$

the conditional probabilities :

$$\mathbb{P}(X_j \in P \mid X_{j-1} \in \widetilde{P}),\ P \in \mathcal{P}_k$$

and the limit when $k \to \infty$ when it makes sense.
We shall define a notion of mixing with respect to a class of functions. Let $\mathcal{C}$ be a Banach space of real **bounded** functions endowed with a norm of the form :

$$\|f\|_{\mathcal{C}} = C(f) + \|f\|,$$

where $C(f)$ is a semi-norm (i.e. $\forall f \in \mathcal{C}$, $C(f) \geq 0$, $C(\lambda f) = |\lambda| C(f)$ for $\lambda \in \mathbb{R}$, $C(f+g) \leq C(f) + C(g)$) and $\|\ \|$ is a norm on $\mathcal{C}$. We will denote by $\mathcal{C}_1$ the subset of functions in $\mathcal{C}$ such that $C(f) \leq 1$.
Particular choices of $\mathcal{C}$ may be the space BV of functions of bounded variation on Σ if it is totally ordered or the space of Hölder (or piecewise Hölder) functions. Recall that a function f on Σ is of bounded variation if it is bounded and

$$\bigvee f := \sup \sum_{i=0}^{n} |f(x_i) - f(x_{i+1})| < \infty,$$

where the sup is taken over all finite sequences $x_1 < \cdots < x_n$ of elements of Σ. The space BV endowed with the norm $\|f\| = \bigvee f + \|f\|_\infty$ is a Banach space.
Inspired from [DP], we define the $\Phi_{\mathcal{C}}$-mixing coefficients.

Definition 1 *For $i \in \mathbb{N}$, let $\mathcal{M}_i$ be the sigma algebra generated by X_1, ..., X_i. For $k \in \mathbb{N}$,*

$$\begin{aligned}\Phi_{\mathcal{C}}(k) = &\sup\{|\mathbb{E}(Y f(X_{i+k})) - \mathbb{E}(Y)\mathbb{E}(f(X_{i+k}))|\ i \in \mathbb{N}\ Y\ \textit{is}\\ &\mathcal{M}_i - \textit{measurable with } \|Y\|_1 \leq 1,\ f \in \mathcal{C}_1\}.\end{aligned} \tag{*}$$

Our main assumption on the process is the following.

Assumption 1

$$\sum_{k=0}^{n-1} (n-k) \Phi_{\mathcal{C}}(k) = O(n).$$

Remarks 1 *Assumption 1 is equivalent to $(\Phi_{\mathcal{C}}(k))_{k\in\mathbb{N}}$ summable. We prefer to formulate it in the above form because it appears more naturally in our context.*
Our definition is inspired from Csiszár's (which is Ψ-mixing for variables taking values in a finite alphabet) and Dedecker-Prieur. It covers lots of natural systems (see Section 1.2 for an example with dynamical systems and [DP] for further examples). Our definition extends Csiszár's which was for random variables on a finite alphabet.

We consider a sequence $(\mathcal{P}_k)_{k\in\mathbb{N}}$ of countable partitions of Σ such that : almost surely, for all $j,k\in\mathbb{N}$, we have

$$\text{for any } P\in\mathcal{P}_k \text{ there exists } \widetilde{P}\in\mathcal{P}_{k-1}\ ,\ X_j\in P\ \Rightarrow X_{j-1}\in\widetilde{P}. \qquad (**)$$

For $i,\ell\in\mathbb{N}$, for $P\in\mathcal{P}_k$, consider the random variable :

$$N_i^\ell(P)=\sum_{j=i}^{\ell+i-1}\mathbf{1}_P(X_j).$$

Our aim is to have quantitative informations on how close are the empirical probabilities $\frac{N_i^{i+n}(P)}{n}$ to the expected value $Q_i^{i+n}(P):=\mathbb{E}\left(\frac{N_i^{i+n}(P)}{n}\right)$. We are especially interested in "large scale typicality": k will grow with n. We wonder also about "conditional typicality", for $P\in\mathcal{P}_k$, let

$$\hat{g}_n(P)=\frac{N_1^{n+1}(P)}{N_0^{n-1}(\widetilde{P})}\frac{n-1}{n}.$$

Our main result is that $\hat{g}_n(P)$ is a consistent estimator of the conditional probabilities $Q_n(P|\widetilde{P}):=\dfrac{Q_1^{n+1}(P)}{Q_0^{n-1}(\widetilde{P})}$. This follows from an exponential inequality (see Theorem 1.1.5). If the conditional probabilities $Q_n(P|\widetilde{P})$ converge when $k\to\infty$, we may obtain an estimator of the limit function. This is the case for certain dynamical systems (see Section 1.2) and g-measures (see Section 1.3). In these settings, we obtain a consistent estimator of the potential function. This may leads to a way of testing the nullity of the asymptotic variance of the system (see Section 1.4 for details).

Section 1.1 contains general results on typicality and conditional typicality for some weak-dependant sequences. In Section 1.2, we apply these results to expanding dynamical systems of the interval. Section 1.3 is devoted to Gibbs measures and chains with complete connections. Finally, in Section 1.4 we sketch an attempt to test the nullity of the asymptotic variance of the system.

1.1 Typicality and conditional typicality via exponential inequalities

Following Csiszár, we wonder about typicality that is : how close are the "empirical probabilities" $\frac{N_i^{n+i}(P)}{n}$ to the expected probability $Q_i^{n+i}(P)$? This is done via a "Hoeffding-type" inequality for partial sums.

The following Proposition has been obtained in [DP], we sketch here the proof because our context is a bit different.

Proposition 1.1.1 *Let (X_i) be a sequence a random variables. Let the coefficients $\Phi_{\mathcal{C}}(k)$ be defined by (*). For $\varphi \in \mathcal{C}$, $p \geq 2$, define*

$$S_n(\varphi) = \sum_{i=1}^{n} \varphi(X_i)$$

and

$$b_{i,n} = \left(\sum_{k=0}^{n-i} \Phi(k) \right) \|\varphi(X_i) - \mathbb{E}(\varphi(X_i))\|_{\frac{p}{2}} C(\varphi).$$

For any $p \geq 2$, we have the inequality :

$$\begin{aligned} \|S_n(\varphi) - \mathbb{E}(S_n(\varphi))\|_p &\leq \left(2p \sum_{i=1}^{n} b_{i,n} \right)^{\frac{1}{2}} \\ &\leq C(\varphi) \left(2p \sum_{k=0}^{n-1} (n-k) \Phi_{\mathcal{C}}(k) \right)^{\frac{1}{2}} . \end{aligned} \tag{1.1}$$

As a consequence, we obtain

$$\begin{aligned} &\mathbb{P}\left(|S_n(\varphi) - \mathbb{E}(S_n(\varphi))| > t\right) \\ &\leq e^{\frac{1}{e}} \exp\left(\frac{-t^2}{2e(C(\varphi))^2 \sum_{k=0}^{n-1} (n-k) \Phi_{\mathcal{C}}(k)} \right) . \end{aligned} \tag{1.2}$$

Proof (Sketch of proof). There are two ingredients to get (1.1). Firstly we need a counterpart to Lemma 4 in [DP].

Lemma 1.1.2

$$\Phi_{\mathcal{C}}(k) = \sup \left\{ \|\mathbb{E}(\varphi(X_{i+k})|\mathcal{M}_i) - \mathbb{E}(\varphi(X_{i+k}))\|_\infty \quad , \ \varphi \in \mathcal{C}_1 \right\} .$$

We postpone the proof of Lemma 1.1.2 to the end of the proof of the proposition.
Secondly, we apply Proposition 4 in [DD] to get : (let $Y_i = \varphi(X_i) - \mathbb{E}(\varphi(X_i))$)

$$\begin{aligned} \|S_n(\varphi) - \mathbb{E}(S_n(\varphi))\|_p &\leq \left(2p \sum_{i=1}^{n} \max_{i \leq \ell \leq n} \|Y_i \sum_{k=i}^{\ell} \mathbb{E}(Y_k|\mathcal{M}_i)\|_{\frac{p}{2}} \right)^{\frac{1}{2}} \\ &\leq \left(2p \sum_{i=1}^{n} \|Y_i\|_{\frac{p}{2}} \sum_{k=i}^{n} \|\mathbb{E}(Y_k|\mathcal{M}_i)\|_\infty \right)^{\frac{1}{2}} \\ &\leq \left(2p \sum_{i=1}^{n} b_{i,n} \right)^{\frac{1}{2}} \end{aligned}$$

(we have used that by Lemma 1.1.2, $\|\mathbb{E}(Y_{k+i}|\mathcal{M}_i)\|_\infty \leq C(\varphi)\Phi_{\mathcal{C}}(k)$). To obtain the second part of inequality (1.2), use $\|Y_i\|_{\frac{p}{2}} \leq \|Y_i\|_\infty \leq C(\varphi)\Phi_{\mathcal{C}}(0)$). The second inequality (1.2) follows from (1.1) as in [DP].

Proof (Proof of Lemma 1.1.2). We write

$$\begin{aligned}\mathbb{E}(Yf(X_{i+k})) - \mathbb{E}(Y)\mathbb{E}(f(X_{i+k})) &= \mathbb{E}(Y[\mathbb{E}(f(X_{i+k})|\mathcal{M}_i) - \mathbb{E}(f(X_{i+k}))]) \\ &\le \|\mathbb{E}(f(X_{i+k})|\mathcal{M}_i) - \mathbb{E}(f(X_{i+k}))\|_\infty .\end{aligned}$$

To prove the converse inequality, for $\varepsilon > 0$, consider an event A_ε such that for $\omega \in A_\varepsilon$,

$$|\mathbb{E}(f(X_{i+k})|\mathcal{M}_i)(\omega) - \mathbb{E}(f(X_{i+k}))| \ge \|\mathbb{E}(f(X_{i+k})|\mathcal{M}_i) - \mathbb{E}(f(X_{i+k}))\|_\infty - \varepsilon,$$

and consider the random variable

$$Y_\varepsilon = \frac{\mathbf{1}_{A_\varepsilon}}{\mathbb{P}(A_\varepsilon)} \operatorname{sign}(\mathbb{E}(h(X_{i+k})|\mathcal{M}_i)(\omega) - \mathbb{E}(f(X_{i+k}))).$$

Y_ε is $\mathcal{M}_i$-measurable, $\|Y_\varepsilon\|_1 \le 1$ and

$$\mathbb{E}(Y_\varepsilon f(X_{i+k})) - \mathbb{E}(Y_\varepsilon)\mathbb{E}(f(X_{i+k})) \ge \|\mathbb{E}(f(X_{i+k})|\mathcal{M}_i) - \mathbb{E}(f(X_{i+k}))\|_\infty - \varepsilon.$$

Thus, the lemma is proven.

We shall apply inequality (1.2) to the function $\varphi = \mathbf{1}_P$, $P \in \mathcal{P}_k$.

Corollary 1.1.3 *If the process $(X_1, \ldots, X_n, \ldots)$ satisfies Assumption 1, if the sequence of partitions $(\mathcal{P}_k)_{k\in\mathbb{N}}$ satisfies (**) and for all $P \in \mathcal{P}_k$, $\mathbf{1}_P \in \mathcal{C}$, then, there exists a constant $C > 0$ such that for all $k \in \mathbb{N}$, for all $P \in \mathcal{P}_k$, for any $t \in \mathbb{R}$, for all $i, n \in \mathbb{N}$,*

$$\mathbb{P}\left(\left|\frac{N_i^{n+i}(P)}{n} - Q_i^{n+i}(P)\right| > t\right) \le e^{\frac{1}{e}} e^{\left(-\frac{Ct^2 n}{C(\mathbf{1}_P)^2}\right)}. \tag{1.3}$$

Proof. It follows directly from (1.2) applied to $\varphi = \mathbf{1}_P$ and Assumption 1.

Let us denote by $\hat{\mathbb{P}}_i^{n+i}(P) = \frac{N_i^{n+i}(P)}{n}$. The following corollary is a counterpart to Csiszár's result (Theorem 1 in [Cs]) in our context.

Corollary 1.1.4 *There exists $C > 0$ such that for all $P \in \mathcal{P}_k$ for which $\left(\frac{Q_i^{n+i}(P)}{C(\mathbf{1}_P)}\right)^2 n \ge \ln^2 n$, we have :*

$$\mathbb{P}\left(\left|\frac{\hat{\mathbb{P}}_i^{n+i}(P)}{Q_i^{n+i}(P)} - 1\right| > t\right) \le e^{\frac{1}{e}} e^{\left(-Ct^2 \ln^2 n\right)}.$$

Proof. We apply Corollary 1.1.3 with $t \cdot Q_i^{n+i}(P)$ instead of t. We get :

$$\mathbb{P}\left(\left|\frac{\hat{\mathbb{P}}_i^{n+i}(P)}{Q_i^{n+i}(P)} - 1\right| > t\right) \le e^{\frac{1}{e}} \exp\left(-\frac{Ct^2(Q_i^{n+i}(P))^2 n}{(C(\mathbf{1}_P))^2}\right).$$

The result follows.

Remark 1 *Let us consider the case where $\mathcal{C} = BV$. If the partition $\mathcal{P}_k$ is a partition into interval, then for all $P \in \mathcal{P}_k$, $C(\mathbf{1}_P) = 2$.*

We are now in position to prove our theorem on conditional typicality. Recall that

$$\hat{g}_n(P) = \frac{n-1}{n} \frac{N_1^{n+1}(P)}{N_0^{n-1}(\tilde{P})}.$$

Theorem 1.1.5 *Let the process $(X_p)_{p\in\mathbb{N}}$ satisfy Assumption 1, let the sequence of partitions $(\mathcal{P}_k)_{k\in\mathbb{N}}$ satisfy (**) and assume that if $P \in \mathcal{P}_k$ then $\mathbf{1}_P \in \mathcal{C}$. There exists $K > 0$ such that for all $\varepsilon < 1$, for all $P \in \mathcal{P}_k$ for which*

$$\frac{Q_0^{n-1}(\widetilde{P})}{C(\mathbf{1}_P)} \text{ and } \frac{Q_0^{n-1}(\widetilde{P})}{C(\mathbf{1}_{\widetilde{P}})} \geq n^{-\frac{\varepsilon}{2}},$$

we have

$$\mathbb{P}\left(\left|\hat{g}_n(P) - Q_n(P|\widetilde{P})\right| > t\right) \leq 4e^{-Kt^2n^{1-\varepsilon}} + 2e^{-Kn^{1-\varepsilon}}.$$

If the sequence is stationary, the result may be rewritten as :

$$\mathbb{P}\left(\left|\hat{g}_n(P) - \mathbb{P}(X_1 \in P \mid X_0 \in \widetilde{P})\right| > t\right) \leq 4e^{-Kt^2n^{1-\varepsilon}} + 2e^{-Kn^{1-\varepsilon}}.$$

Proof. Fix $R > 0$, let us bound the probability

$$\mathbb{P}\left(\left|\hat{g}_n(P) - Q_n(P|\widetilde{P})\right| > t\right)$$

with the sum of the probabilities :

$$(1) = \mathbb{P}\left(\left|\hat{\mathbb{P}}_1^{n+1}(P) - Q_1^{n+1}(P)\right| > \frac{t \cdot Q_0^{n-1}(\widetilde{P})}{2}\right),$$

$$(2) = \mathbb{P}\left(\left|\hat{\mathbb{P}}_0^{n-1}(\widetilde{P}) - Q_0^{n-1}(\widetilde{P})\right| > \frac{tQ_0^{n-1}(\widetilde{P})R}{2}\right),$$

$$(3) = \mathbb{P}\left(\frac{\hat{\mathbb{P}}_0^{n-1}(\widetilde{P})}{\hat{\mathbb{P}}_1^{n+1}(P)} < R\right).$$

The terms (1) and (2) are easily bounded using Corollary 1.1.3 : we get

$$(1) \leq e^{\frac{1}{e}} \exp\left(-\frac{Ct^2n^{1-\varepsilon}}{4}\right) \quad (2) \leq e^{\frac{1}{e}} \exp\left(-\frac{Ct^2R^2(n-1)^{1-\varepsilon}}{4}\right).$$

It remains to bound the term (3). We have (recall that almost surely, $X_j \in P \Rightarrow X_{j-1} \in \widetilde{P}$) :

$$\frac{\hat{\mathbb{P}}_1^{n+1}(P)}{\hat{\mathbb{P}}_0^{n-1}(\widetilde{P})} \leq \frac{n-1}{n}\left(1+\frac{\mathbf{1}_{\{X_n\in P\}}}{N_0^{n-1}(\widetilde{P})}\right).$$

So we have that $\dfrac{\hat{\mathbb{P}}_1^{n+1}(P)}{\hat{\mathbb{P}}_0^{n-1}(\widetilde{P})} < 2$ unless if $N_0^{n-1}(\widetilde{P}) = 0$. Take $R = \frac{1}{2}$, we have :

$$(3) \leq \mathbb{P}(N_0^{n-1}(\widetilde{P}) = 0) \text{ and}$$

$$\mathbb{P}(N_0^{n-1}(\widetilde{P}) = 0) \leq \mathbb{P}\left(\hat{\mathbb{P}}_0^{n-1}(\widetilde{P}) \leq \frac{Q_0^{n-1}(\widetilde{P})}{2}\right).$$

Apply Corollary 1.1.3 with $t = \frac{Q_0^{n-1}(\widetilde{P})}{2}$ (of course our hypothesis imply that $Q_0^{n-1}(\widetilde{P}) > 0$) to get

$$(3) \leq e^{\frac{1}{e}} e^{-Cn^{1-\varepsilon}}.$$

These three bounds give the result (we have bounded $e^{\frac{1}{e}}$ by 2).

1.2 Applications to dynamical systems

We turn now to our main motivation : dynamical systems. Consider a dynamical system (Σ, T, μ). Σ is a complete space, $T : \Sigma \to \Sigma$ is a measurable map, μ is a T-invariant probability measure on Σ. As before, $\mathcal{C}$ is a Banach space of bounded functions on Σ (typically, $\mathcal{C}$ will be the space of function of bounded variations or a space of piecewise Hölder functions, see examples in Section 1.2.1). It is endowed with a norm of the form :

$$\|f\|_{\mathcal{C}} = C(f) + \|f\|,$$

where $C(f)$ is a semi-norm (i.e. $\forall f \in \mathcal{C}$, $C(f) \geq 0$, $C(\lambda f) = |\lambda| C(f)$ for $\lambda \in \mathbb{R}$, $C(f+g) \leq C(f) + C(g)$) and $\| \ \|$ is a norm on $\mathcal{C}$. In addition, we assume that the norm $\| \ \|$ on $\mathcal{C}$ is such that for any $\varphi \in \mathcal{C}$, there exists a real number $R(\varphi)$ such that $\|\varphi + R(\varphi)\| \leq C(\varphi)$ (for example, this is the case if $\| \ \| = \| \ \|_\infty$ and $C(\varphi) = \bigvee(\varphi)$ or $\| \ \| = \| \ \|_\infty$ and $C(\varphi)$ is the Hölder constant). We assume that the dynamical system satisfy the following mixing property : for all $\varphi \in L^1(\mu)$, $\psi \in \mathcal{C}$,

$$\left| \int_\Sigma \psi \cdot \varphi \circ T^n d\mu - \int_\Sigma \psi d\mu \int_\Sigma \varphi d\mu \right| \leq \Phi(n) \|\varphi\|_1 \|\psi\|_{\mathcal{C}}, \tag{1.1}$$

with $\Phi(n)$ summable.

Consider a countable partition $A_1, \ldots, A_p, \ldots$ of Σ. Denote by $\mathcal{P}_k$ the countable partition of Σ whose atoms are defined by : for $i_0, \ldots, i_{k-1}$, denote

$$A_{i_0,\ldots,i_{k-1}} = \{x \in \Sigma \ / \text{ for } j = 0, \ldots, k-1, \ T^j(x) \in A_{i_j}.\}$$

We assume that for all $i_0,\dots,i_{k-1}$, $f=\mathbf{1}_{A_{i_0,\dots,i_{k-1}}}\in\mathcal{C}$ and note $C(i_{i_0},\dots,i_{k-1})=C(f)$. Consider the process taking values into Σ : $X_j(x)=T^j(x)$, $j\in\mathbb{N}$, $x\in\Sigma$. Clearly if $X_j\in A_{i_0,\dots,i_{k-1}}$ then $X_{j+1}\in A_{i_1,\dots,i_{k-1}}$. That is for any $P\in\mathcal{P}_k$, there exists a unique $\widetilde{P}\in\mathcal{P}_{k-1}$ such that $X_j\in P\ \Rightarrow\ X_{j+1}\in\widetilde{P}$. Condition (1.1) may be rewritten as : for all $\varphi\in L^1(\mu)$, $\psi\in\mathcal{C}$,

$$|\mathrm{Cov}(\psi(X_0),\varphi(X_n))|\le\Phi(n)\|\varphi\|_1\|\psi\|_{\mathcal{C}}.$$

Moreover, we assume that for $\psi\in\mathcal{C}$, there exists a real number $R(\psi)$ such that $\|\psi+R(\psi)\|\le C(\psi)$. We have :

$$\begin{aligned}|\mathrm{Cov}(\psi(X_0),\varphi(X_n))|&=|\mathrm{Cov}([\psi(X_0)+R(\psi)],\varphi(X_n))|\\&\le\Phi(n)\|\varphi\|_1\|\psi+R(\psi)\|_{\mathcal{C}}\\&\le 2\Phi(n)\|\varphi\|_1C(\psi).\end{aligned}\tag{1.2}$$

Using the stationarity of the sequence (X_j), we have for all $i\in\mathbb{N}$, for $\psi\in\mathcal{C}_1$, $\varphi\in L^1$, $\|\varphi\|_1\le 1$,

$$|\mathrm{Cov}(\psi(X_i),\varphi(X_{n+i}))|\le 2\Phi(n).\tag{1.3}$$

So, our Assumptions 1 and (**) are satisfied for a "time reversed" process: consider a process $(Y_n)_{n\in\mathbb{N}}$ such that $(Y_n,\cdots,Y_0)$ as the same law as $(X_0,\cdots,X_n)$, then $\mathrm{Cov}(\psi(X_i),\varphi(X_{n+i}))=\mathrm{Cov}(\psi(Y_{i+n}),\varphi(Y_i))$ and the process $(Y_n)_{n\in\mathbb{N}}$ satisfies our Assumptions 1. Using the stationarity, it satisfies also(**), see [BGR] and [DP] for more developments on this "trick". Applying Theorem 1.1.5 to the process $(Y_n)_{n\in\mathbb{N}}$ and using that

$$\sum_{j=1}^{n}\mathbf{1}_P(Y_j)\overset{\text{Law}}{=}\sum_{0}^{n-1}\mathbf{1}_P(X_j)$$

and

$$\sum_{j=0}^{n-2}\mathbf{1}_{\widetilde{P}}(Y_j)\overset{\text{Law}}{=}\sum_{j=0}^{n-2}\mathbf{1}_{\widetilde{P}}(X_j),$$

we obtain the following result.

Theorem 1.2.1 *There exists a constant $C>0$, such that for all $k,n\in\mathbb{N}$, for any sequence $i_0,\dots,i_{k-1}$, for all $t\in\mathbb{R}$,*

$$\mathbb{P}\left(\left|\frac{N_0^n(A_{i_0,\dots,i_{k-1}})}{n}-\mu(A_{i_0,\dots,i_{k-1}})\right|>t\right)\le e^{\frac{1}{e}}e^{-\frac{Ct^2n}{C(i_0,\dots,i_{k-1})^2}}.$$

Let $\hat{g}_n(A_{i_0,\dots,i_{k-1}})=\frac{N_0^n(A_{i_0,\dots,i_{k-1}})}{N_0^{n-1}(A_{i_1,\dots,i_{k-1}})}\frac{n-1}{n}$, there exists $K>0$ such that for all $\varepsilon<1$, if

$$\frac{\mu(A_{i_1,\dots,i_{k-1}})}{C(i_0,\dots,i_{k-1})}\ \textit{and}\ \frac{\mu(A_{i_1,\dots,i_{k-1}})}{C(i_1,\dots,i_{k-1})}\ge n^{-\frac{\varepsilon}{2}},$$

then we have :

$$\mathbb{P}\left(\left|\hat{g}_n(A_{i_0,\dots,i_{k-1}}) - \mathbb{P}(X_0 \in A_{i_0} | X_1 \in A_{i_1}, \dots, X_{k-1} \in A_{i_{k-1}})\right| > t\right) \\ \leq 4e^{-Kt^2 n^{1-\varepsilon}} + 2e^{-Kn^{1-\varepsilon}}.$$

Let us terminate this section with a lemma stating that the elements $P \in \mathcal{P}_k$ are exponentially small. It indicates that we might not expect to take k of order greater than $\ln n$ in the above theorem.

Lemma 1.2.2 *Assume that* $C_{\max} = \max\limits_{j=1,\dots,} C(\mathbf{1}_{A_j}) < \infty$. *There exists* $0 < \gamma < 1$ *such that for all* $P \in \mathcal{P}_k$*, we have*

$$\mu(P) \leq \gamma^k.$$

Proof. The proof of Lemma 1.2.2 follows from the mixing property. It is inspired from [Pa]. Let $n_0 \in \mathbb{N}$ to be fixed later. Let $P \in \mathcal{P}_k$, for some indices $i_0, \dots, i_{k-1}$, we have that

$$P = \{x \in A_{i_0}, \dots, T^{k-1}x \in A_{i_{k-1}}\}.$$

Then, let $\ell = [\frac{k}{n_0}]$,

$$\begin{aligned} \mu(P) &= \mathbb{P}(X_0 \in A_{i_0}, \dots, X_{k-1} \in A_{i_{k-1}}) \\ &\leq \mathbb{P}(X_0 \in A_{i_0}, X_{n_0} \in A_{i_{n_0}}, \dots,\ X_{\ell n_0} \in A_{i_{\ell n_0}}) \end{aligned}$$

Now, the random variable

$$Y = \frac{\mathbf{1}_{A_{i_{n_0}}}(X_{n_0}) \cdots \mathbf{1}_{A_{i_{\ell n_0}}}(X_{\ell n_0})}{\mathbb{P}(X_{n_0} \in A_{i_{n_0}}, \dots,\ X_{\ell n_0} \in A_{i_{\ell n_0}})}$$

is $\mathcal{M}_{\ell n_0}$-measurable with L^1 norm less than 1 and $\frac{\mathbf{1}_{A_{i_0}}}{C_{\max}}$ is in $\mathcal{C}_1$. From the mixing property (1.3), we get : (let $s = \sup_{j=1,\dots} \mu(A_j) < 1$)

$$\begin{aligned} &\mathbb{P}(X_0 \in A_{i_0}, X_{n_0} \in A_{i_{n_0}}, \dots,\ X_{\ell n_0} \in A_{i_{\ell n_0}}) \\ &\leq \mathbb{P}(X_{n_0} \in A_{i_{n_0}}, \dots,\ X_{\ell n_0} \in A_{i_{\ell n_0}}) \cdot (\Phi_C(n_0) C_{\max} + s). \end{aligned}$$

Choosing n_0 such that $\Phi_C(n_0)C_{\max} + s < 1$, we obtain the result by induction.

1.2.1 Expanding maps of the interval

In this section, we consider piecewise expanding maps on the interval $I = [0, 1]$. That is, T is a piecewise expanding map, defined on a finite partition into intervals $A_1, \dots, A_\ell$. $\mathcal{P}_k$ is the partition of I with atoms : $A_{i_0} \cap T^{-1}A_{i_1} \cap \dots \cap T^{-(k-1)}A_{i_{k-1}}$. If for all $j = 1, \dots, \ell$, $T(A_j)$ is a union of the A_p's, T is said to be a *Markov map*. For $x \in I$, let $C_k(x)$ be the atom of the partition $\mathcal{P}_k$ containing x. Under an assumption of aperiodicity in the Markov case or covering in general, the map T admits a unique invariant measure absolutely continuous with respect to the Lebesgue measure m. Let h be the invariant density. The potential of the system is $g = \frac{h}{|T'| \cdot h \circ T}$, we have also that g^{-1} is the Radon-Nikodym derivative of $\mu \circ T$ with respect to μ (if $\mu = hm$). We shall prove that $g(x)$ may be estimated by $\hat{g}_{n,k}(x) := \hat{g}_n(C_k(x))$ for $k = \Theta(\ln n)$. Formally the assumptions on the system are the following.

Assumption 2 *1. the restriction of T to each $\overline{A_j}$ is a C^2 one-to-one map from $\overline{A_j}$ to $T(\overline{A_j}) =: B_j$.*

2. T is expanding: there exists $1 < \theta^{-1}$ such that for all $x \in I$, $\theta^{-1} \leq |T'(x)|$.

3. If T is a Markov map, we assume that it is aperiodic: there exists $N \in \mathbb{N}$ such that for all i, $j = 1, \dots, \ell$, for all $n \geq N$,

$$T^{-n} A_i \cap A_j \neq \emptyset.$$

4. If T is not Markov, we assume that it satisfies the covering property : for all $k \in \mathbb{N}$, there exists $N(k)$ such that for all $P \in \mathcal{P}_k$,

$$T^{N(k)} P = [0, 1].$$

The above conditions are sufficient to ensure existence and uniqueness of an absolutely continuous invariant measure as well as an estimation of the speed of mixing (see for example [Sc] for the Markov case and [Co], [Li] for the general case). Under more technical assumptions, these results on existence and uniqueness of an absolutely continuous invariant measure as well as an estimation of the speed of mixing remain valid, with an infinite countable partition ([Br], [L,S,V], [Ma1]).

Theorem 1.2.3 *([Sc], [Co], [Li]) Let $\mathcal{C}$ be the space of functions on $[0, 1]$ of bounded variations. Let T satisfy the assumptions 2. Then we have the following mixing property : there exists $C > 0$, $0 < \xi < 1$ such that for all $\varphi \in L^1(\mu)$, $\psi \in \mathcal{C}$,*

$$\left| \int_\Sigma \psi \cdot \varphi \circ T^n d\mu - \int_\Sigma \psi d\mu \int_\Sigma \varphi d\mu \right| \leq C \xi^n \|\varphi\|_1 \|\psi\|_{\mathcal{C}}.$$

Moreover, we have that the invariant density h belongs to BV and $0 < \inf h \leq \sup h < \infty$. If the map is Markov, then h is C^1 on each B_j.

In other words, our system satisfy (1.1) for bounded variation functions. Moreover, for any $k \in \mathbb{N}$, the element P of $\mathcal{P}_k$ are subintervals, so the indicators $\mathbf{1}_P$ belong to BV and $C(\mathbf{1}_P) = \bigvee(\mathbf{1}_P) = 2$. So, we shall apply Theorem 1.2.1, this will lead to the announced estimation of the potential g.

Let us also introduce a very useful tool in dynamical systems : the transfer operator. For $f \in BV$, let

$$\mathcal{L}(f)(x) = \sum_{y/T(y)=x} g(y)f(y).$$

We have $\mathcal{L}(\mathbf{1}) = \mathbf{1}$, for all $f_1 \in BV$, $f_2 \in L^1(\mu)$,

$$\int_I \mathcal{L}(f_1) \cdot f_2 d\mu = \int_I f_1 \cdot f_2 \circ T d\mu.$$

The process $(Y_n)_{n\in\mathbb{N}}$ introduced after Lemma 1.2.2 is a Markov process with kernel $\mathcal{L}$ (see [BGR]). The following three lemmas are the last needed bricks between Theorem 1.2.1 and the estimation of the potential g.

Lemma 1.2.4 *Assume that T satisfies Assumption 2 and is a Markov map, let γ be given by Lemma 1.2.2. There exists $K > 0$ such that for all $k \in \mathbb{N}$, for all $x \in I$,*

$$(1 - K\gamma^k)g(x) \leq \frac{\mu(C_k(x))}{\mu(C_{k-1}(Tx))} \leq (1 + K\gamma^k)g(x) \tag{1.4}$$

Proof. Because the map is Markov, for all $x \in I$, $T(C_k(x)) = C_{k-1}(Tx)$. We have :

$$\mu(T(C_k(x))) = \int \frac{1}{g}\mathbf{1}_{C_k(x)} d\mu,$$

$$\min_{y\in C_k(x)} \frac{1}{g(y)} \int \mathbf{1}_{C_k(x)} d\mu \leq \int \frac{1}{g}\mathbf{1}_{C_k(x)} d\mu \leq \max_{y\in C_k(x)} \frac{1}{g(y)} \int \mathbf{1}_{C_k(x)} d\mu.$$

Since the map is Markov, h and $h \circ T$ are C^1 on each $C_k(x)$, so g is C^1 on $C_k(x)$ and since T is expanding, we conclude that

$$\max_{y\in C_k(x)} \frac{1}{g(y)} \leq (1 + K\gamma^k)\frac{1}{g(x)}$$

$$\text{and} \quad \min_{y\in C_k(x)} \frac{1}{g(y)} \geq (1 - K\gamma^k)\frac{1}{g(x)}.$$

The result follows.

If the map T is not Markov, we shall prove a result not so strong (but sufficient for our purpose). To deal with non Markov maps, we have to modify the above proof at two points : firstly, we have not $T(C_k(x)) = C_{k-1}(Tx)$ for all x (but for lots of them) ; secondly, $g = \frac{h}{|T'|h\circ T}$ is not smooth (due to h). The following lemma shows that we control the irregularity of h.

Lemma 1.2.5 *Let $a = \bigvee h$, for any interval P, let $\bigvee_P h$ be the variation of h on P. For all $k \geq 1$, for all $u_k > 0$,*

$$\mu\{x \in [0,1] \ / \ \bigvee_{C_k(x)} h \geq u_k\} \leq \frac{\gamma^k}{u_k a}.$$

Proof. We have :

$$\mu\{x \in [0,1] \ / \ \bigvee_{C_k(x)} h \geq u_k\} = \sum_{\substack{P \in \mathcal{P}_k \\ \bigvee_P h \geq u_k}} \mu(P).$$

$$\begin{aligned} a = \bigvee h &\geq \sum_{P \in \mathcal{P}_k} \bigvee_P h \\ &\geq \#\{P \in \mathcal{P}_k \ / \ \bigvee_P h \geq u_k\} u_k. \end{aligned}$$

In other words, $\#\{P \in \mathcal{P}_k \ / \ \bigvee_P h \geq u_k\} \leq \dfrac{a}{u_k}$. Using Lemma 1.2.2, we get :

$$\begin{aligned} \mu\{x \in [0,1] \ / \ \bigvee_{C_k(x)} h \geq u_k\} &\leq \#\{P \in \mathcal{P}_k \ / \ \bigvee_P h \geq u_k\} \gamma^k \\ &\leq \frac{\gamma^k}{u_k a}. \end{aligned}$$

Corollary 1.2.6 *For all $\kappa > \gamma$, there exists a constant $K > 0$ and for all $k \in \mathbb{N}^*$, a set B_k such that $\mu(B_k) \leq \frac{\gamma^k}{\kappa^k a}$ and if $x \notin B_k$, $y \in C_k(x)$,*

$$(1 - K\kappa^k) \leq \frac{g(x)}{g(y)} \leq (1 + K\kappa^k). \tag{1.5}$$

Proof. Recall that $g = \frac{h}{|T'| h \circ T}$. Because T is piecewise C^2 and expanding, $\frac{1}{|T'|}$ satisfies an equation of the type (1.5) for all $x \in [0,1]$, for $\kappa = \gamma$. We just have to prove that h satisfies such an inequality. Fix $\kappa > \gamma$, let

$$B_k = \{x \in [0,1] \ / \ \bigvee_{C_k(x)} h \geq \kappa^k\}.$$

Let $x \notin B_k$ and $y \in C_k(x)$.

$$|h(x) - h(y)| \leq \bigvee_{C_k(x)} h \leq \kappa^k.$$

Now, $\frac{h(x)}{h(y)} = 1 + \frac{h(x) - h(y)}{h(y)}$, thus

$$1 - \frac{1}{\sup h} \kappa^k \leq \frac{h(x)}{h(y)} \leq 1 + \frac{1}{\inf h} \kappa^k.$$

Of course, the same equation holds for $h \circ T$ by replacing k with $k - 1$, combining this equations (for h, $h \circ T$ and $|T'|$) gives the result.

Lemma 1.2.7 *Assume that T satisfies Assumption 2 and is not necessary a Markov map. There exists $K > 0$ such that for all $k \in \mathbb{N}$, for all $\kappa > \gamma$,*

$$\mu\left\{x \in I \;/\; (1 - K\kappa^k)g(x) \leq \frac{\mu(C_k(x))}{\mu(C_{k-1}(Tx))} \leq (1 + K\kappa^k)g(x)\right\}$$

$$\geq 1 - (2\ell\gamma^k + a\left(\frac{\gamma}{\kappa}\right)^k).$$

Proof. We begin with a simple remark. Let us denote $\partial\mathcal{P}$ the union of the boundaries of the A_j's. For $x \in [0,1]$, if $\overline{C_k(x)} \cap \partial\mathcal{P} = \emptyset$ then $T(C_k(x)) = C_{k-1}(Tx)$, otherwise, $T(C_k(x))$ is strictly included into $C_{k-1}(Tx)$. This elementary remark is very useful in the study of non Markov maps. The points x such that $T(C_k(x)) = C_{k-1}(Tx)$ will be called k-Markov points. If the map is Markov then all points are k-Markov for all $k \in \mathbb{N}$. For k-Markov points, we may rewrite the proof of Lemma 1.2.4 to get the inequalities :

$$\min_{y \in C_k(x)} \frac{1}{g(y)}\mu(C_k(x)) \leq \mu(C_{k-1}(Tx)) \leq \max_{y \in C_k(x)} \frac{1}{g(y)}\mu(C_k(x)).$$

Now, we use Corollary 1.2.6 and we have that if x is a k-Markov point that do not belong to B_k then

$$(1 - K\kappa^k)g(x) \leq \frac{\mu(C_k(x))}{\mu(C_{k-1}(Tx))} \leq (1 + K\kappa^k)g(x). \tag{1.6}$$

So, we have that the set D_k of points not satisfying 1.6 for one k is included into the set of points x such that $\overline{C_k(x)} \cap \partial\mathcal{P} \neq \emptyset$ or in B_k (given by Corollary 1.2.6). Clearly, there are at most 2ℓ elements P of $\mathcal{P}_k$ such that $\overline{P} \cap \partial\mathcal{P} \neq \emptyset$, moreover, by Lemma 1.2.2, we have for $P \in \mathcal{P}_k$, $\mu(P) \leq \gamma^k$. We have proven that $\mu(D_k) \leq 2\ell\gamma^k + \frac{\gamma^k}{\kappa^k a}$.

We are now in position to prove that $\hat{g}_{n,k}(x)$ is a consistent estimator of the potential $g(x)$.

Theorem 1.2.8 *For all $\kappa > \gamma$, there exists D_k and E_k finite union of elements of $\mathcal{P}_k$ satisfying $\mu(D_k) \leq 2\ell\gamma^k + a\left(\frac{\gamma}{\kappa}\right)^k$, $\mu(E_k) \leq \gamma^k$ and there exists $L > 0$ such that if*

- $x \notin D_k \cup E_k$,
- $\dfrac{\ln(\frac{t}{2K})}{\ln(\kappa)} \leq k \leq \dfrac{\varepsilon}{2}\dfrac{\ln 2n}{\ln(\frac{\ell}{\gamma})}$

then

$$\mathbb{P}(|\hat{g}_{n,k}(x) - g(x)| > t) \leq 4e^{-Lt^2n^{1-\varepsilon}} + 2e^{-Ln^{1-\varepsilon}}$$

Proof. Fix $\kappa > \gamma$, let D_k be given by Lemma 1.2.7 : if $x \notin D_k$ then

$$(1 - K\kappa^k)g(x) \leq \frac{\mu(C_k(x))}{\mu(C_{k-1}(Tx))} \leq (1 + K\kappa^k)g(x),$$

let E_k be the set of points x such that $\mu(C_k(x)) \le \frac{\gamma^k}{\ell^k}$. Clearly, if $x \in D_k$ then $C_k(x) \subset D_k$ and if $x \in E_k$ then $C_k(x) \subset E_k$, so D_k and E_k are finite union of elements of $\mathcal{P}_k$.
Let $x \notin D_k \cup E_k$, then $\mu(C_k(x)) > \frac{\gamma^k}{\ell^k}$. If $k \le \frac{\varepsilon}{2}\frac{\ln 2n}{\ln(\frac{\ell}{\gamma})}$ then $\mu(C_k(x)) \ge 2n^{-\frac{\varepsilon}{2}}$. Since $C_k(x)$ is an interval, we have $C(\mathbf{1}_{C_k(x)}) = \bigvee(\mathbf{1}_{C_k(x)}) = 2$ and then

$$\frac{\mu(C_{k-1}(Tx))}{C(\mathbf{1}_{C_k(x)})} = \frac{\mu(C_{k-1}(Tx))}{C(\mathbf{1}_{C_{k-1}(Tx)})} \ge \frac{\mu(C_k(x))}{2} \ge n^{-\frac{\varepsilon}{2}}.$$

We shall use Theorem 1.2.1.

$$\begin{aligned}
&\mathbb{P}(|\hat{g}_{n,k}(x) - g(x)| > t) \\
&\le \mathbb{P}(|\hat{g}_{n,k}(x) - \frac{\mu(C_k(x))}{\mu(C_{k-1}(Tx))}| > t - |\frac{\mu(C_k(x))}{\mu(C_{k-1}(Tx))} - g(x)|) \\
&\le \mathbb{P}(|\hat{g}_{n,k}(x) - \frac{\mu(C_k(x))}{\mu(C_{k-1}(Tx))}| > t - K\kappa^k) \text{ because } x \notin D_k \\
&\le 4e^{-L(t-K\kappa^k)^2 n^{1-\varepsilon}} + 2e^{-Ln^{1-\varepsilon}} \text{ we have used Theorem 1.2.1.}
\end{aligned}$$

If $\dfrac{\ln(\frac{t}{2})}{\ln(\frac{1}{\kappa})} \le k$, we conclude

$$\mathbb{P}(|\hat{g}_{n,k}(x) - g(x)| > t) \le 4e^{-Lt^2 n^{1-\varepsilon}} + 2e^{-Ln^{1-\varepsilon}}.$$

We derive the following corollary. Fix $\kappa > \gamma$.

Corollary 1.2.9 *Let $\alpha = \frac{c}{2(1+c)}$ with $c = \frac{\ln(1/\kappa)}{\ln(l/\gamma)}$ and $k(n)$ be an increasing sequence such that*

$$\frac{\ln\left(\frac{1}{2Kn^\alpha}\right)}{\ln(\kappa)} \le k(n) \le \frac{\varepsilon}{2}\frac{\ln 2n}{\ln(\frac{\ell}{\gamma})},$$

let $\hat{g}_n = \hat{g}_{n,k(n)}$, then $|\hat{g}_n(x) - g(x)| = O_{\mathbb{P}}(n^{-\alpha})$.

Proof. It suffices to prove that :

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}(n^\alpha|\hat{g}_n(x) - g(x)| > M) = 0.$$

We chose $t = n^{-\alpha}$ in Theorem 1.2.8 and obtain :

$$\mathbb{P}(n^\alpha|\hat{g}_n(x) - g(x)| > M) \le \mathbb{P}(|\hat{g}_n(x) - g(x)| > \frac{1}{n^\alpha}) \le 4e^{-Ln^{1-\varepsilon-2\alpha}} + o(1).$$

The best rate is obtained for $\alpha = \frac{c}{2(1+c)}$ with $c = \frac{\ln(1/\kappa)}{\ln(l/\gamma)}$.

Remark 2 *In [CMS], an exponential inequality is proven for Lipschitz functions of several variables for expanding dynamical systems of the interval. We can not use such a result here because characteristic functions of intervals are* *not Lipschitz, the result could maybe be improved to take into consideration piecewise Lipschitz functions. The Lipchitz constant enter in the bound of the exponential inequality and any kind of piecewise Lipschitz constant would be exponentially big for* $\mathbf{1}_P$*,* $P \in \mathcal{P}_k$*. Nevertheless, such a result for functions of several variables could be interesting to estimate the conditional probabilities and potential* g *: we could construct an estimator by replacing* $N_j^\ell(A_{i_0,\ldots,i_{k-1}})$ *with*

$$\widetilde{N_j^n}(A_{i_0,\ldots,i_{k-1}}) = \left|\left\{p \in \{j,...,n+j-k\} \ / \ X_j \in A_{i_0}, \ldots, X_{j+k-1} \in A_{i_{k-1}}\right\}\right|.$$

1.3 Gibbs measures and chains with complete connections

In this section, we state our results in the particular setting of Gibbs measures or chains with complete connections. Gibbs measures and chains with complete connections are two different point of view of the same thing - consider a stationary process $(X_i)_{i\in\mathbb{N} \text{ or } \mathbb{Z}}$ taking values into a finite set A satisfying : for all a_0, ..., a_k, ... in A. If $\mathbb{P}(X_0 = a_0, \ldots, X_k = a_k) \neq 0$ for all k, then

$$\lim_{k\to\infty} \mathbb{P}(X_0 = a_0|X_1 = a_1, \ldots, X_{k-1} = a_{k-1}) = \mathbb{P}(X_0 = a_0|X_i = a_i, \ i \geq 1),$$

exists. Moreover, there exists a summable sequence $\gamma_k > 0$ such that if $a_0 = b_0$, ..., $a_k = b_k$,

$$\left|\frac{\mathbb{P}(X_0 = a_0|X_i = a_i, \ i \geq 1)}{\mathbb{P}(X_0 = b_0|X_i = b_i, \ i \geq 1)} - 1\right| \leq \gamma_k. \tag{1.1}$$

Define $\Sigma \subset A^{\mathbb{N}}$ be the set of admissible sequences :

$$\Sigma = \{x = (x_0, \ldots, x_k, \ldots,) \in A^{\mathbb{N}} \ / \\ \text{for all } k \geq 0, \ \mathbb{P}(X_0 = x_0, \ldots, X_k = x_k) \neq 0\}.$$

Σ is compact for the product topology and is invariant by the shift map $\sigma : \sigma(x_0, x_1, \ldots) = (x_1, \ldots)$. We denote by μ the image measure of the X_i's. We assume that the process is mixing : there exists $N > 0$ such that for all $i, j \in A$, for all $n > N$,

$$\mathbb{P}(X_0 = i \text{ and } X_n = j) \neq 0.$$

We shall denote by

$$A_j = \{x \in \Sigma \ / \ x_0 = j\} \text{ and } A_{i_0,\ldots,i_{k-1}} = \{x \in \Sigma \ / \ x_j = i_j \ j = 0, \ldots k-1\}.$$

As before, $\mathcal{P}_k$ is the partition of Σ whose atoms are the $A_{i_0,\ldots,i_{k-1}}$'s and $C_k(x)$ is the atom of $\mathcal{P}_k$ containing x.
We assume also that the process has a *Markov structure* : for $x = (x_0, \ldots,) \in \Sigma$, $ax = (a, x_0, \ldots) \in \Sigma$ if and only if $ay \in \Sigma$ for all $y \in A_{x_0}$.

For $x \in \Sigma$, let $g(x) = \mathbb{P}(X_0 = x_0 | X_i = x_i,\ i \geq 1)$. We shall prove that $\hat{g}_{n,k}$ is a consistent estimator of g.
It is known (see [KMS], [Ma2], [BGF], [Po]) that such a process is mixing for suitable functions.
Let $\gamma_n^\star = \sum_{k \geq n} \gamma_k$, define a distance on Σ by $d(x,y) = \gamma_n^\star$ if and only if $x_j = y_j$ for $j = 0, \ldots, n-1$ and $x_n \neq y_n$. Let L be the space of Lipschitz functions for this distance, endowed with the norm $\|\psi\| = \sup |\psi| + L(\psi)$ where $L(\psi)$ is the Lipschitz constant of ψ.

Theorem 1.3.1 *([KMS], [Ma2], [BGF], [Po]) A process satisfying (1.1), being mixing and having a Markov structure is mixing for functions in L in the sense that equation (1.1) is verified for $\varphi \in L^1(\mu)$ and $\psi \in L$ with $\Phi(n) \overset{n\to\infty}{\longrightarrow} 0$. If $\gamma_n^\star$ is summable, so is $\Phi(n)$.*

In what follows, we assume that $\gamma_n^\star$ is summable. For any $\psi \in L$, let $R = -\inf \psi$ then $\sup |\psi + R| \leq L(\psi)$, then we have (1.3) for the process $(X_i)_{i\in\mathbb{N}}$ and $\psi \in L$ such that $L(\psi) \leq 1$ and Theorem 1.2.1 is satisfied.
We have that

$$L(\mathbf{1}_{A_j}) \leq \frac{1}{\gamma_0^\star} \text{ and } L\left(\mathbf{1}_{A_{i_0,\ldots,i_{k-1}}}\right) \leq \frac{1}{\gamma_k^\star}.$$

Equation (1.1) gives the following lemma which will be used instead of Corollary 1.2.6.

Lemma 1.3.2 *For all $x \in \Sigma$, for all $k \in \mathbb{N}$, $y \in C_k(x)$,*

$$1 - \gamma_k \leq \frac{g(x)}{g(y)} \leq 1 + \gamma_k.$$

Following the proof of Lemma 1.2.4, we get : for all $x \in \Sigma$, for $k \in \mathbb{N}$,

$$(1-\gamma_k) g(x) \leq \frac{\mu(C_k(x))}{\mu(C_{k-1}(T(x)))} \leq (1+\gamma_k) g(x). \tag{1.2}$$

Let $\gamma < 1$ be given by Lemma 1.2.2, let $\ell = |A|$.

Theorem 1.3.3 *Assume that $\gamma_k^\star$ is summable, and that the process satisfy (1.1), is mixing and has a Markov structure. Then there exists $L > 0$ such that if :*

1. $\mu(C_k(x)) \geq \frac{\gamma^k}{\ell^k}$,
2. $\left(\frac{\gamma}{\ell}\right)^k \gamma_k^\star \geq n^{-\frac{\varepsilon}{2}}$,
3. $\gamma_k \leq \frac{t}{2}$.

we have

$$\mathbb{P}(|\hat{g}_{n,k}(x) - g(x)| > t) \leq 4e^{-Lt^2 n^{1-\varepsilon}} + 2e^{-Ln^{1-\varepsilon}}.$$

Moreover,

1. $\mu\{x \in \Sigma \ / \ \mu(C_k(x)) < \frac{\gamma^k}{\ell^k}\} \leq \gamma^k$,

2. $\left(\frac{\gamma}{\ell}\right)^k \gamma_k^\star \geq n^{-\frac{\varepsilon}{2}}$ *if* $k \leq a \ln n$ *for suitable* $a > 0$,
3. $\gamma_k \leq \frac{t}{2}$ *if* $k \geq bt^{-\frac{1}{2}}$ *for suitable* $b > 0$.

Proof. The proof follows the proof of Theorem 1.2.8 using Lemma 1.3.2 instead of Lemma 1.2.7. The estimates on k are obtained by noting that since $\gamma_k^\star$ is summable then $\gamma_k = o(\frac{1}{k^2})$ and $\gamma_k^\star = o(\frac{1}{k})$. Of course, better estimates may be obtained if γ_k decreases faster.

As in Section 1.2.1, we derive the following corollary.

Corollary 1.3.4 *For* $k = \Theta(\ln n)$, *there exists* $\alpha > 0$ *such that* $\hat{g}_{n,k}$ *goes to* $g(x)$ *in probability at rate* $\frac{1}{n^\alpha}$.

1.4 Testing if the asymptotic variance is zero : the complete case

In this section, we study the problem of testing whether the asymptotic variance of the process is zero. This is motivated by the fact that for the process studied in the previous sections, we may prove a central limit theorem provided the asymptotic variance is not zero (see [Br], [V] for examples). We are concerned with a process $(X_j)_{j\in\mathbb{N}}$ satisfying Conditions of Section 1.2.1 or Section 1.3. We assume moreover that the system is complete: $T(A_i) = I$ for all i if we are in the context of Section 1.2.1 or $\sigma(A_i) = \Sigma$ if we are in the context of Section 1.3. Our arguments should probably be generalized to non complete situations. In what follows, we shall denote T for $T\ : I \to I$ as well as $\sigma\ : \Sigma \to \Sigma$.

Definition 2 *([Br]) Let*

$$S_n = \sum_{j=0}^{n-1} (X_j - \mathbb{E}(X_0)) \text{ and } M_n = \int \left(\frac{S_n}{\sqrt{n}}\right)^2 d\mathbb{P}.$$

The sequence M_n *converges to* V *which we shall call the asymptotic variance.*

Proposition 1.4.1 *([Br], [CM]) The asymptotic variance* V *is zero if and only if the potential* $\log g$ *is a cohomologous to a constant :* $\log g = \log a + u - u \circ T$, *with* $a > 0$, $u \in BV$ *or* $u \in L$.

Because we are in a stationary setting, we have that the asymptotic variance is zero if and only if g is indeed constant (the fact that the system is complete is here very important). We deduce a way of testing if the asymptotic variance is zero. Using Theorem 1.2.8 or Theorem 1.3.3, we have that if g is constant,

$$\mathbb{P}(|\sup \hat{g}_{n,k} - \inf \hat{g}_{n,k}| > t) \leq 2 \cdot (4e^{-Lt^2 n^{1-\varepsilon}} + 2e^{-Ln^{1-\varepsilon}}) + \gamma^k.$$

To use such a result, we have to compute $\sup \hat{g}_{n,k}$ and $\inf \hat{g}_{n,k}$, so we have ℓ^k computations to make with $k = \Omega(\ln n)$. A priori, all the constants in the above inequality, may be specified. In theory, for $t > 0$, we may find k, n satisfying the hypothesis of Theorem 1.2.8 or Theorem 1.3.3 so that $\mathbb{P}(|\sup \hat{g}_{n,k} - \inf \hat{g}_{n,k}| > t)$ is smaller than a specified value. If the computed values of $\sup \hat{g}_{n,k}$ and $\inf \hat{g}_{n,k}$ agree with this estimation this will indicates that g is probably constant so that the asymptotic variance is probably 0.

References

[BGR] A.D. Barbour, R. Gerrard, G. Reinert, *Iterates of expanding maps.* Probab. Theory Related Fields 116 (2000), no. 2, 151–180.

[BRY] A. Barron, J. Rissanen, B. Yu, *The minimum description length principle in coding and modeling.* Information theory: 1948–1998. IEEE Trans. Inform. Theory 44 (1998), no. 6, 2743–2760.

[BGF] X Bressaud, R. Fernàndez, A. Galves, *Decay of correlations for non-Hölderian dynamics. A coupling approach.* Electron. J. Probab. 4 (1999), no. 3, 19 pp.

[Br] A. Broise, *Transformations dilatantes de l'intervalle et théorèmes limites.* Études spectrales d'opérateurs de transfert et applications. Astérisque 1996, no. 238, 1–109.

[CM] F. Chazal, V. Maume-Deschamps, *Statistical properties of Markov dynamical sources: applications to information theory.* Discrete Math. Theor. Comput. Sci. 6 (2004), no. 2, 283–314

[Co] P. Collet, *Some ergodic properties of maps of the interval.* Dynamical systems (Temuco, 1991/1992), 55–91, Travaux en Cours, 52, Hermann, Paris, 1996.

[CMS] P. Collet,S. Martinez, B. Schmitt, *Exponential inequalities for dynamical measures of expanding maps of the interval.* Probab. Theory Related Fields 123 (2002), no. 3, 301–322.

[Cs] I. Csiszár, *Large-scale typicality of Markov sample paths and consistency of MDL order estimators.* Special issue on Shannon theory: perspective, trends, and applications. IEEE Trans. Inform. Theory 48 (2002), no. 6, 1616–1628.

[CsS] I. Csiszár, P. C. Shields, *The consistency of the BIC Markov order estimator.* Ann. Statist. 28 (2000), no. 6, 1601–1619.

[DD] J. Dedecker, P. Doukhan, *A new covariance inequality and applications.* Stochastic Process. Appl. 106 (2003), no. 1, 63–80.

[DP] J. Dedecker, C. Prieur, *New dependence coefficients. Examples and applications to statistics.* Probab. Theory and Relat. Fields 132 (2005), 203–236.

[FKT] P. Flajolet, P. Kirschenhofer, R.F. Tichy, *Deviations from uniformity in random strings.* Probab. Theory Related Fields 80 (1988), no. 1, 139–150.

[KMS] A. Kondah, V. Maume, B. Schmitt, *Vitesse de convergence vers l'état d'équlibre pour des dynamiques markoviennes non holdériennes.* Ann. Inst. H. Poincaré Probab. Statist. 33 (1997), no. 6, 675–695.

[Li] C. Liverani, *Decay of correlations for piecewise expanding maps.* J. Statist. Phys. 78 (1995), no. 3-4, 1111–1129.

[L,S,V] C. Liverani, B. Saussol, S. Vaienti, *Conformal measure and decay of correlation for covering weighted systems.* Ergodic Theory Dynam. Systems 18 (1998), no. 6, 1399–1420.

[Ma1] V. Maume-Deschamps, *Correlation decay for Markov maps on a countable state space.* Ergodic Theory Dynam. Systems 21 (2001), no. 1, 165–196.

[Ma2] V Maume-Deschamps, *Propriétés de mélange pour des sytèmes dynamiques markoviens.* Thèse de l'Université de Bourgogne, available at `http://math.u-bourgogne.fr/IMB/maume`.

[Pa] F. Paccaut *Propriétés statistiques de systèmes dynamiques non markoviens* Ph D Thesis Université de Bourgogne (2000). Available at `http://www.lamfa.u-picardie.fr/paccaut/publi.html`

[Po] M. Policott *Rates of mixing for potentials of summable variation.* Trans. Amer. Math. Soc. 352 (2000), no. 2, 843–853.

[R] J. Rissanen, non k-Markov points.*Stochastic complexity in statistical inquiry.* World Scientific Series in Computer Science, 15. World Scientific Publishing Co., Inc., Teaneck, NJ, 1989. vi+178 pp.

[Sc] B. Schmitt, *Ergodic theory and thermodynamic of one-dimensional Markov expanding endomorphisms.* Dynamical systems (Temuco, 1991/1992), 93–123, Travaux en Cours, 52, Hermann, Paris, 1996.

[V] B. Vallée *Dynamical sources in information theory: fundamental intervals and word prefixes* Algorithmica, **29**, 262-306, (2001).