CENTRAL LIMIT THEOREM FOR DETERMINISTIC SYSTEMS

Carlangelo Liverani

University of Rome *Tor Vergata*

July 27, 1995

Abstract. A unified approach to obtaining the central limit theorem for hyperbolic dynamical systems is presented. It builds on previous results for one dimensional maps but it applies to the multidimensional case as well.

CONTENT

This paper originated out of discussions with D. Szasz and A. Kramli, and was made possible by D.Szasz key suggestion to use K-partitions. I wish to thank E. Olivieri, E. Presutti, B. Tot, and L. Triolo for helpful discussions. In addition, I am indebted to S.Olla for explaining me the subtleties of the Kipnis-Varadhan approach. This work has been partially supported by grant CIPA-CT92-4016 of the Commission of the European Community.

Typeset by $\mathcal{AMS}$-TEX

§0 Introduction.

A discrete time dynamical system consists of a measurable space X together with a σ-algebra $\mathcal{F}$, a measurable map $T : X \to X$ which describes the dynamics, and a probability measure P invariant with respect to T. This setting is particularly well suited to study problems involving statistical properties of the motion of deterministic systems.

Typically the properties of interest are ergodicity, mixing, bounds on the decay of correlations, Central Limit Theorems (CLT) and so on. Several approaches have been developed to tackle such problems at various levels. Given a system, one first explores the weaker statistical properties and then tries to investigate the stronger ones using the already obtained results plus some extra properties.

The position of this paper in the above mentioned hierarchy is between obtaining bounds on the decay of correlations and CLT. In other words we discuss a general approach that gives checkable conditions under which, in a mixing system, an observable enjoys the CLT. Such general approaches already exist but they are either limited to one dimensional systems [Ke] or relay on the existence of special partitions of the phase space [Ch], partitions which concrete construction may be far from trivial [BSC1], [BSC2]; for a very nice review of the state of affairs up to 1989 (but still actual) see [De].

Here, I want to put forward the following point of view: the above described dynamical systems are most naturally viewed as giving rise to a (deterministic) Markov process. It is therefore tempting to think that there should exists some general probabilistic theorem that states abstract conditions for the validity of the CLT, and that all the concrete cases can simply be obtained by the direct application of such a theorem to the system under consideration (without having to code the system in some symbolic type dynamics). General theorems of this type are well known in probability theory but they are normally not well suited for applications to the case at hand. Two such general theorems, tailored for dynamical systems, can be found in this paper.

Attempts in this directions already exists for some time [Go], [IL], but they are satisfactory only for the one dimensional case (the equivalent of Theorem 1.1 in this paper). Particular mention must be given to [DG], the results obtained there are essentially comparable to the one presented here in section 1 and could be applied to the multidimensional case. Unfortunately, not much attention is given there to applications, so that the possibility to bypass a symbolic representation of the system is completely overlooked.

The approach used here is a martingale approximation inspired by [KV]. Since this is a typical probabilistic technique, I think it underlines very well the purely probabilistic nature of the result hereby clarifying which characteristics of a deterministic system yield such a drastic statistical behavior.

As we will see, a major difference with the analogous type results in probability is that the CLT holds for a much smaller class of observables than the square summable ones. This is not an artifact of the proof: it is an inevitable consequence of the deterministic nature of the systems under consideration so that only observables that operates some "coarse graining" (and therefore enjoy some degree of smoothness) can yield strong statistical behavior. Here no particular attempt is made to find the most general class of observables to which the Theorems apply; nonetheless, the technique put forward lends itself to an extension in such a direction.

The paper includes some concrete examples as well. Their aim is to show how the general theorems can be applied in special cases. The cases discussed belong to quite general classes (expanding one dimensional maps, area preserving piecewise smooth uniformly hyperbolic maps in two and more dimensions), yet no real new result is contained in such examples. This reflects the spirit of the paper of presenting an approach to the problem rather than new implementations. Nevertheless, the application of the present results in technically complex situations (e.g., hyperbolic billiards) greatly simplifies the proof of the validity of the CLT. In addition, it is conceivable that some new results can be obtained by this approach since the two above mentioned theorems hold in more general cases than the ones already present in the literature (a brief comparison with previously known results is inserted after the proof of each theorem).

The plan of the paper is as follows. Section 1 contains two probabilistic theorems that are well suited for the study of dynamical systems. In fact, they may seem a bit unnatural from the pure probabilistic point of view. On the one hand, both theorems deal only with functions in L^∞ instead than L^2. The reason is that normally the decay of correlations in dynamical systems can be obtained only for classes of functions with some amount of smoothness, which makes them automatically bounded. The issue is not purely a matter of taste: a look at the proofs will show that such an hypothesis has really been used and that many key estimates would not hold in L^2. On the other hand, in Theorem 2 are introduced σ-algebras $\mathcal{F}_i$ that behave nicely with respect to the dynamics. This may make little sense from the purely probabilistic point of view but it is instead a cornerstone in the treatment of hyperbolic dynamical systems.

In the above sense the results of section 1, although purely probabilistic in nature, are expressly developed for applications to dynamical systems.

Section 2 describe how the technique applies to non-invertible maps. The case of piecewise smooth expanding maps of the interval is discussed in detail.

Section 3 deals with the most interesting applications: the multidimensional case. As an example I treat a large subclass of piecewise smooth symplectic maps. Such maps are well studied in the literature for some relevant physical models (e.g. billiards) are naturally described in their terms. It is shown that very general considerations imply the applicability of the results developed in section one.

§1 A general probabilistic result.

Let X be a complete separable metric space, $\mathcal{F}$ a σ-algebra, P a probability measure ($P(X) = 1$) and $T : X \to X$ a measurable map.[1] We will call $\mathbb{E}$ the

[1] Actually, we assume that, for each $A \in \mathcal{F}$, not only $T^{-1}A \in \mathcal{F}$ but also $TA \in \mathcal{F}$.

expectation with respect to P. In addition, we require that P is invariant with respect to T (i.e., for all $A \in \mathcal{F}$ holds $P(T^{-1}A) = P(A)$), and that the dynamical system (T, X, P) be ergodic.

For each $\phi \in L^2(X)$ define $\widehat{T} : L^2(X) \to L^2(X)$ by

$$\widehat{T}\phi = \phi \circ T,$$

and let $\widehat{T}^* : L^2(X) \to L^2(X)$ be the dual of $\widehat{T}$.

If $\mathbb{E}(f) = 0$, then by ergodicity $\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \widehat{T}^n f = \mathbb{E}(f) = 0$. The CLT gives us informations on the speed of convergence; namely the conditions under which there exists $\sigma \in \mathbb{R}^+$: for each interval $I \subset \mathbb{R}$

$$\lim_{n\to\infty} P\left(\left\{\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} \widehat{T}^n f \in I\right\}\right) = \frac{1}{\sqrt{2\pi}\sigma}\int_I e^{-\frac{x^2}{2\sigma^2}}\,dx;$$

this is called "convergence in law (or distribution)" to a Gaussian random variable of zero mean and variance σ.

Consider a sub-σ-algebra $\mathcal{F}_0$ of $\mathcal{F}$ and define $\mathcal{F}_i = T^{-i}\mathcal{F}_0$, $i \in \mathbb{Z}$, then the following holds.

Theorem 1.1. *If $\mathcal{F}_i$ is coarser than $\mathcal{F}_{i-1}$ and, for each $\phi \in L^\infty(X)$, we have*

$$\mathbb{E}(\widehat{T}\widehat{T}^*\phi|\mathcal{F}_1) = \mathbb{E}(\phi|\mathcal{F}_1),$$

then, for each $f \in L^\infty(X)$, $\mathbb{E}(f) = 0$ and $\mathbb{E}(f|\mathcal{F}_0) = f$, such that

(1) $\sum_{n=0}^\infty |\mathbb{E}(f\widehat{T}^n f)| < \infty$,
(2) *the series $\sum_{n=0}^\infty \mathbb{E}(\widehat{T}^{*n} f|\mathcal{F}_0)$ converges absolutely almost surely,*[2]

the sequence

$$\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} \widehat{T}^i f$$

converges in law to a Gaussian random variable of zero mean and finite variance σ, $\sigma^2 \le -\mathbb{E}(f^2) + 2\sum_{n=0}^\infty \mathbb{E}(f\widehat{T}^n f)$.

In addition, $\sigma = 0$ if and only if there exists a $\mathcal{F}_0$–mesurable function g such that

$$\widehat{T}f = \widehat{T}g - g.$$

Finally, if (2) converges in $L^1(X)$, then $\sigma^2 = -\mathbb{E}(f^2) + 2\sum_{n=0}^\infty \mathbb{E}(f\widehat{T}^n f)$.

Proof. The key idea is to use a Martingale approximation. That is, to find $Y_i \in L^2(X)$ and g $\mathcal{F}_0$–measurable, and almost everywhere finite, such that

$$\mathbb{E}(Y_{i-1}|\mathcal{F}_i) = Y_{i-1}\ ; \quad \mathbb{E}(Y_i|\mathcal{F}_i) = 0, \tag{1.1}$$

[2] As we will see in the proof, this implies that there exists an almost everywhere finite $\mathcal{F}_0$-meausurable function g, such that $f = g - \mathbb{E}(\widehat{T}^* g|\mathcal{F}_0)$.

(i.e., Y_i is a reverse Martingale difference with respect to the filtration $\{\mathcal{F}_i\}_{i=0}^{\infty}$), and

$$\widehat{T}^i f = Y_i + \widehat{T}^i g - \widehat{T}^{i-1} g \quad \forall i > 0. \tag{1.2}$$

Accordingly,

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \widehat{T}^i f = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} Y_i + \frac{1}{\sqrt{n}} [\widehat{T}^n g - g]. \tag{1.3}$$

Equation (1.3) shows that we can obtain the central limit theorem for our random variable provided we have the central limit theorem for the martingale difference Y_i. In fact, $\frac{1}{\sqrt{n}}[\widehat{T}^n g - g]$ converges to zero in probability when $n \to \infty$.

Note that (1.1) and (1.2) are equivalent to

$$\mathbb{E}(\widehat{T}^i f | \mathcal{F}_i) = \mathbb{E}(\widehat{T}^i g | \mathcal{F}_i) - \mathbb{E}(\widehat{T}^{i-1} g | \mathcal{F}_i) \quad \forall i > 0.$$

Since by the definition of $\mathcal{F}_i$ follows that, for each $\phi \in L^1(X)$,

$$\mathbb{E}(\widehat{T}^i \phi | \mathcal{F}_i) = \widehat{T}^i \mathbb{E}(\phi | \mathcal{F}_0) \quad \forall i > 0,$$

and because the invariance of $\mathbb{E}$ with respect to T implies $\widehat{T}^* \widehat{T} = \mathbb{1}$, we have

$$\begin{aligned} f &= \mathbb{E}(g|\mathcal{F}_0) - \widehat{T}^* \mathbb{E}(g|\mathcal{F}_1) = g - \widehat{T}^* \mathbb{E}(\widehat{T}\widehat{T}^* g | \mathcal{F}_1) \\ &= g - \mathbb{E}(\widehat{T}^* g | \mathcal{F}_0). \end{aligned} \tag{1.4}$$

It is immediate to see that $g = \sum\limits_{n=0}^{\infty} \mathbb{E}(\widehat{T}^{*n} f | \mathcal{F}_0)$ (the convergence of the series is the hypothesis (2) in the statement of the theorem) is a solution of the above equation, and therefore of (1.2), (clearly, $Y_i = \widehat{T}^{i-1} Y_1$).[3]

In fact, setting $T_0 \phi = \mathbb{E}(\widehat{T}^* \phi | \mathcal{F}_0)$, the solution of (1.4) is given by the Neumann series $\sum_{n=0}^{\infty} T_0^n f$. But $T_0^n f = \mathbb{E}(\widehat{T}^{*n} f | \mathcal{F}_0)$ since

$$\begin{aligned} \mathbb{E}(\widehat{T}^* \mathbb{E}(\widehat{T}^{*n} f | \mathcal{F}_0) | \mathcal{F}_0) &= \widehat{T}^* \widehat{T} \mathbb{E}(\widehat{T}^* \mathbb{E}(\widehat{T}^{*n} f | \mathcal{F}_0) | \mathcal{F}_0) = \widehat{T}^* \mathbb{E}(\widehat{T}\widehat{T}^* \mathbb{E}(\widehat{T}^{*n} f | \mathcal{F}_0) | \mathcal{F}_1) \\ &= \widehat{T}^* \mathbb{E}(\mathbb{E}(\widehat{T}^{*n} f | \mathcal{F}_0) | \mathcal{F}_1) = \widehat{T}^* \mathbb{E}(\widehat{T}^{*n} f | \mathcal{F}_1) \\ &= \widehat{T}^* \mathbb{E}(\widehat{T}\widehat{T}^{*(n+1)} f | \mathcal{F}_1) = \mathbb{E}(\widehat{T}^{*(n+1)} f | \mathcal{F}_0). \end{aligned}$$

To insure that the central limit theorem for Y_i holds, we need only to show that Y_i is square summable due to the following [Ne]:

[3] It is remarkable that, once we have g, the Y_i are defined by (1.2) itself, and will automatically satisfy (1.1).

Theorem. *Let $(Y_n)_{n\geq 1}$ be a stationary, ergodic, martingale difference (or reversed martingale difference) with respect to the filtration $\{\mathcal{F}_n\}_{n\geq 1}$. If $Y_1 \in L^2(X)$, then $\sigma^2 = \mathbb{E}(Y_1^2)$ and the CLT holds.*

The above theorem applies to our case since the stationarity of (Y_n) is implied by the invariance of the measure with respect to T, while the ergodicity follows from the ergodicity of the dynamical system $(X,\, T,\, P)$.

If the series $\sum_{n=0}^{\infty} \mathbb{E}(\widehat{T}^{*n} f | \mathcal{F}_0)$ would converge in $L^2(X)$, then $Y_1 \in L^2(X)$ would hold and the Theorem would be proven. It is however a remarkable fact that Y_i can be in $L^2(X)$ without g being even integrable [KV]. Unfortunately, the road to this result is a bit indirect and consists in carrying out an argument similar to the one above but producing a sequence of martingale differences $Y_i(\lambda)$ that approximate Y_i.

Let us look for $Y_i(\lambda)$, $\lambda > 1$, such that

$$\mathbb{E}(Y_{i-1}(\lambda)|\mathcal{F}_i) = Y_{i-1}(\lambda)\ ; \quad \mathbb{E}(Y_i(\lambda)|\mathcal{F}_i) = 0; \tag{1.5}$$

and

$$\widehat{T}^i f = Y_i(\lambda) + \widehat{T}^i g(\lambda) - \lambda^{-1}\widehat{T}^{i-1} g(\lambda) \quad \forall i > 0,\ \lambda > 1. \tag{1.6}$$

In analogy with what we have seen before $g(\lambda) = \sum\limits_{n=0}^{\infty} \lambda^{-n}\mathbb{E}(\widehat{T}^{*n} f|\mathcal{F}_0)$, only now $g(\lambda) \in L^2(X)$ for each $\lambda > 1$. Since $\lim\limits_{\lambda\to 1} g(\lambda) = g(1) = g$ almost surely, it follows that $\lim\limits_{\lambda\to 1} Y_i(\lambda) = Y_i$ almost surely. In addition,

$$\begin{aligned}
\mathbb{E}(Y_i(\lambda)^2) =& \mathbb{E}(Y_1(\lambda)^2) = \mathbb{E}([\widehat{T} f - \widehat{T} g(\lambda) + \lambda^{-1} g(\lambda)]^2)\\
=& \mathbb{E}(\widehat{T} f[\widehat{T} f - \widehat{T} g(\lambda) + \lambda^{-1} g(\lambda)])\\
&- \mathbb{E}([\widehat{T} g(\lambda) - \lambda^{-1} g(\lambda)][\widehat{T} f - \widehat{T} g(\lambda) + \lambda^{-1} g(\lambda)]),
\end{aligned}$$

since $\mathbb{E}(\widehat{T} f - \widehat{T} g(\lambda) + \lambda^{-1} g(\lambda)|\mathcal{F}_1) = \mathbb{E}(Y_1|\mathcal{F}_1) = 0$. Hence,

$$\begin{aligned}
\mathbb{E}(Y_i(\lambda)^2) =& - \mathbb{E}((\widehat{T} f)^2) + \mathbb{E}([\widehat{T} g(\lambda) - \lambda^{-1} g(\lambda)]^2)\\
=& - E(f^2) + \mathbb{E}(\widehat{T} g(\lambda)[\widehat{T} g(\lambda) - \lambda^{-1} g(\lambda)])\\
&- \lambda^{-1}\mathbb{E}(g(\lambda)\widehat{T} g(\lambda)) + \lambda^{-2}\mathbb{E}(\widehat{T} g(\lambda)^2)\\
=& - \mathbb{E}(f^2) + 2\mathbb{E}(\widehat{T} g(\lambda)[\widehat{T} g(\lambda) - \lambda^{-1} g(\lambda)]) - (1-\lambda^{-2})\mathbb{E}(g(\lambda)^2)\\
=& - \mathbb{E}(f^2) + 2\mathbb{E}(\widehat{T} g(\lambda)\widehat{T} f) - (1-\lambda^{-2})\mathbb{E}(g(\lambda)^2)\\
=& - \mathbb{E}(f^2) + 2\mathbb{E}(g(\lambda) f) - (1-\lambda^{-2})\mathbb{E}(g(\lambda)^2)\\
\leq& - \mathbb{E}(f^2) + 2\sum_{n=0}^{\infty} \lambda^{-n}\mathbb{E}(f\widehat{T}^n f) \leq -\mathbb{E}(f^2) + 2\sum_{n=0}^{\infty} |\mathbb{E}(f\widehat{T}^n f)|.
\end{aligned}$$

The wanted estimates follows from

$$\mathbb{E}(Y_1^2) = \mathbb{E}(\liminf_{\lambda\to 1} Y_1(\lambda)^2) \leq \liminf_{\lambda\to 1} \mathbb{E}(Y_1(\lambda)^2) \leq -\mathbb{E}(f^2) + 2\sum_{n=0}^{\infty} \mathbb{E}(f\widehat{T}^n f).$$

In conclusion, we have seen that the random variable under consideration converges in law to a Gaussian of variance $\sigma^2 = \mathbb{E}(Y_1^2) < \infty$. If $\sigma = 0$ then the second assertion of the statement follows since

$$\mathbb{E}(Y_1^2) = \mathbb{E}([\widehat{T}f - \widehat{T}g + g]^2).$$

If we assume that the series in (2) converges in $L^1(X)$, then it is possible to obtain the much sharper result

$$\lim_{\lambda\to 1} \mathbb{E}(Y_1(\lambda)^2) = -\mathbb{E}(f^2) + 2\sum_{n=0}^{\infty} \lambda^{-n}\mathbb{E}(f\widehat{T}^n f).$$

In fact, for each $\varepsilon > 0$

$$\begin{aligned}\left|\mathbb{E}(Y_i(\lambda)^2) - \mathbb{E}(f^2) + 2\sum_{n=0}^{\infty}\lambda^{-n}\mathbb{E}(f\widehat{T}^n f)\right| &\le \sum_{n=0}^{\infty}(1-\lambda^{-n})\mathbb{E}(f\widehat{T}^n f) \\ + (1-\lambda^{-2})\mathbb{E}(g(\lambda)^2) &\le (1-\lambda^{-M})\sum_{n=0}^{\infty}\mathbb{E}(f\widehat{T}^n f) + \sum_{n=M}^{\infty}\mathbb{E}(f\widehat{T}^n f) \\ + (1-\lambda^{-2})\mathbb{E}(g(\lambda)^2) &\le \varepsilon + (1-\lambda^{-2})\mathbb{E}(g(\lambda)^2)\end{aligned}$$

where M has been chosen sufficiently large and λ sufficiently close to one. In order to continue we need to estimate the last term in the above expression. For further use we will deal with a more general estimate: for each λ, $\mu \in (1, \infty)$ holds

$$\begin{aligned}\mathbb{E}(g(\lambda)g(\mu)) &= \sum_{n,m=0}^{\infty}\lambda^{-n}\mu^{-m}\mathbb{E}(\widehat{T}^{*n} f\mathbb{E}(\widehat{T}^{*m} f|\mathcal{F}_0)) \\ &\le \sum_{n=0}^{\infty}\lambda^{-n}\sum_{m=0}^{M-1}\|f\|_\infty\mathbb{E}(|\mathbb{E}(\widehat{T}^{*n} f|\mathcal{F}_0)|) + \sum_{n=0}^{\infty}\lambda^{-n}\sum_{m=M}^{\infty}\mathbb{E}(|\mathbb{E}(\widehat{T}^{*m} f|\mathcal{F}_0)|) \qquad (1.7) \\ &\le M\|f\|_\infty\sum_{n=0}^{\infty}\mathbb{E}(|\mathbb{E}(\widehat{T}^{*n} f|\mathcal{F}_0)|) + \frac{\|f\|_\infty}{1-\lambda^{-1}}\sum_{m=M}^{\infty}\mathbb{E}(|\mathbb{E}(\widehat{T}^{*m} f|\mathcal{F}_0)|).\end{aligned}$$

That is, choosing again M large and λ sufficiently close to 1,

$$(1-\lambda^{-1})\mathbb{E}(g(\lambda)^2) \le 2\varepsilon.$$

This is not the end of the story: it is possible to prove that Y_1 is the limit of $Y_1(\lambda)$ in $L^2(X)$. To see this it suffices to estimate

$$\begin{aligned}\mathbb{E}([Y_1(\lambda) - Y_1(\mu)]^2) &= \mathbb{E}([\lambda^{-1}g(\lambda) - \mu^{-1}g(\mu)][Y_1(\lambda) - Y_1(\mu)]) \\ &= \mathbb{E}([\lambda^{-1}g(\lambda) - \mu^{-1}g(\mu)]^2) + \mathbb{E}([g(\lambda) - g(\mu)]^2) \\ &\le (1-\mu^{-1}\lambda^{-1})\mathbb{E}(g(\lambda)g(\mu)),\end{aligned}$$

since no generality is lost by choosing $\lambda \ge \mu > 1$, the result follows thanks to the estimate (1.7). $\square$

Let us discuss briefly how the above result compares with the ones present in the literature. In the work of Gordin [Go], used by Keller [Ke], a very similar theorem is present. The main difference is that condition (1) and (2) are replaced by the much stronger condition

$$\sum_{n=0}^{\infty} \mathbb{E}(\mathbb{E}(\widehat{T}^{*n} f|\mathcal{F}_0)^2) < \infty.$$

A similar comment applies to [DG], where moreover there is no discussion of the case $\sigma = 0$.

Theorem 1.1 often is applicable in cases in which T is not invertible, where sometime it is possible to choose $\mathcal{F}_0 = \mathcal{F}$ (see §2).

When T is invertible the choice $\mathcal{F}_0 = \mathcal{F}$ is likely to yield $\mathcal{F}_i = \mathcal{F}$ for each $i \in \mathbb{Z}$, this would undermine the possibility of capturing any type of dynamical coarse graining effect, whereby nullifying the hope of obtaining an interesting statistical behavior. In such a case, there are situations in which a natural choice for $\mathcal{F}_0$ exists (see §3), but it would be too restrictive to require f to be $\mathcal{F}_0$–measurable.

The above difficulties can be dealt with by the following Theorem.

Theorem 1.2. *Suppose T one to one and onto. If $\mathcal{F}_i$ is coarser than $\mathcal{F}_{i-1}$, then, for each $f \in L^\infty(X)$, $\mathbb{E}(f) = 0$ such that*

(1) $\sum_{n=0}^{\infty} |\mathbb{E}(f\widehat{T}^n f)| < \infty$,
(2) *the series* $\sum_{n=0}^{\infty} |\mathbb{E}(\widehat{T}^{*n} f|\mathcal{F}_0)|$ *converges in* L^1,
(3) $\exists\ \alpha > 1$: $\sup_{k\in\mathbb{N}} k^\alpha \mathbb{E}(|\mathbb{E}(f|\mathcal{F}_{-k}) - f|) < \infty$,[4]

the sequence

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \widehat{T}^i f$$

converges in law to a Gaussian random variable of zero mean and finite variance σ, $\sigma^2 = -\mathbb{E}(f^2) + 2\sum_{n=0}^{\infty} \mathbb{E}(f\widehat{T}^n f)$.

In addition, if $\sum_{n=0}^{\infty} n|\mathbb{E}(f\widehat{T}^n f)| < \infty$, *then* $\sigma = 0$ *if and only if there exists* $g \in L^2(X)$ *such that*

$$\widehat{T} f = \widehat{T} g - g.$$

Proof. The key idea is to first approximate f by $\mathbb{E}(f|\mathcal{F}_{-k})$ and then use the same type of Martingale approximation introduced in Theorem 1.1. That is, to find $Y_i(k,\lambda) \in L^2(X)$ and $g(k,\lambda) \in L^2(X)$ such that, given $k > 0$, for each $i > 0$ and $\lambda > 1$

$$\mathbb{E}(Y_{i-1}(k,\lambda)|\mathcal{F}_{i-k}) = Y_{i-1}(k,\lambda)\ ; \quad \mathbb{E}(Y_i(k,\lambda)|\mathcal{F}_{i-k}) = 0, \tag{1.8}$$

(i.e., $Y_i(k,\lambda)$ is a reverse Martingale difference with respect to the filtration $\{\mathcal{F}_i\}_{i=-k}^{\infty}$) and

$$\widehat{T}^i \mathbb{E}(f|\mathcal{F}_{-k}) = Y_i(k,\lambda) + \widehat{T}^i g(k,\lambda) - \lambda^{-1}\widehat{T}^{i-1} g(k,\lambda) \quad \forall i > 0,\ \lambda \geq 1. \tag{1.9}$$

[4] This condition it is not optimal, as it can be seen by looking at the proof, yet I do not know of any application in which a weaker condition could be of interest.

Note that (1.8) and (1.9) are equivalent to

$$\mathbb{E}(f|\mathcal{F}_{-k}) = g(k,\,\lambda) - \lambda^{-1}\mathbb{E}(\widehat{T}^* g(k,\,\lambda)|\mathcal{F}_{-k}).$$

It is immediate to see that $g(k,\,\lambda) = \sum_{n=0}^{\infty} \lambda^{-n}\mathbb{E}(\widehat{T}^{*n} f|\mathcal{F}_{-k}) \in L^2(X)$ for each $\lambda > 1$ and in $L^1(X)$ for $\lambda = 1$ (this is a consequence of hypothesis (2) in the statement of the Theorem) is a solution of the above equation (see the analogous discussion in Theorem 1.1).

Again we want to show that the $Y_i(k,\,1)$ are square summable, actually, in this case, we need a uniform estimate in k. In partial analogy with Theorem 1.1, we have

$$\begin{aligned}\mathbb{E}(Y_i(k,\,\lambda)^2) =&\mathbb{E}(Y_1(k,\,\lambda)^2) = -\mathbb{E}(\mathbb{E}(f|\mathcal{F}_{-k})^2) + \mathbb{E}([\widehat{T}g(k,\,\lambda) - \lambda^{-1}g(k,\,\lambda)]^2)\\ =&-\mathbb{E}(\mathbb{E}(f|\mathcal{F}_{-k})^2) + 2\mathbb{E}(g(k,\,\lambda)\mathbb{E}(f|\mathcal{F}_{-k})) - (1-\lambda^{-2})\mathbb{E}(g(k,\,\lambda)^2).\end{aligned}$$

In addition, for each $\lambda > 1$,

$$\begin{aligned}\mathbb{E}(g(k,\,\lambda)\mathbb{E}(f|\mathcal{F}_{-k})) \le& \sum_{n=0}^{\infty} |\mathbb{E}(f\mathbb{E}(\widehat{T}^{*n} f|\mathcal{F}_{-k}))| = \sum_{n=0}^{\infty} |\mathbb{E}(\widehat{T}^{*n} f\mathbb{E}(f|\mathcal{F}_{-k}))|\\ \le& 2\|f\|_\infty k\mathbb{E}(|\mathbb{E}(f|\mathcal{F}_{-k}) - f|) + \sum_{n=k}^{\infty}\mathbb{E}(\widehat{T}^k f\mathbb{E}(\widehat{T}^{*n} f|\mathcal{F}_0))\\ &+ \sum_{n=0}^{2k-1}\mathbb{E}(\widehat{T}^n ff) < \infty,\end{aligned}$$

where the uniform bound follows from the hypotheses (1), (2), (3) of the Theorem. The previous estimates show that $Y_i(k,\,1)$ are uniformly square integrable martingale differences. Moreover,

$$\lim_{k\to\infty}\lim_{\lambda\to 1}\mathbb{E}(Y_1(k,\,\lambda)^2) = -\mathbb{E}(f^2) + 2\sum_{n=0}^{\infty}\mathbb{E}(f\widehat{T}^n f) = \sigma^2.$$

To see this it, it is enough to compute

$$\begin{aligned}\mathbb{E}(g(k,\,\lambda)g(k,\,\mu)) =& \sum_{n,m=0}^{\infty} \lambda^{-n}\mu^{-m}\mathbb{E}(\widehat{T}^{*n} f\mathbb{E}(\widehat{T}^{*m} f|\mathcal{F}_{-k}))\\ \le& \sum_{n=0}^{\infty}\lambda^{-n}M\|f\|_\infty\mathbb{E}(|\mathbb{E}(\widehat{T}^{*n} f|\mathcal{F}_{-k})|)\\ &+ \sum_{n=0}^{\infty}\lambda^{-n}\|f\|_\infty\sum_{m=M}^{\infty}\mathbb{E}(|\mathbb{E}(\widehat{T}^{*m} f|\mathcal{F}_{-k})|)\\ \le& M\|f\|_\infty^2 k + M\|f\|_\infty\sum_{n=0}^{\infty}\mathbb{E}(|\mathbb{E}(\widehat{T}^{*n} f|\mathcal{F}_0)|)\\ &+ (1-\lambda^{-1})^{-1}\|f\|_\infty\sum_{m=M-k}^{\infty}\mathbb{E}(|\mathbb{E}(\widehat{T}^{*m} f|\mathcal{F}_0)|),\end{aligned}$$

so, since M can be chosen arbitrarily large, $\lim_{\lambda\to 1}(1-\lambda)\mathbb{E}(g(k,\,\lambda)^2)=0$. Furthermore, in analogy with Theorem 1.1, easily follows that $Y_1(k,\,\lambda)$ converges to $Y_1(k,\,1)$ in $L^2(X)$.

This implies that, defining

$$S_n=\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\widehat{T}^i f\ ;\quad S_n^k=\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}\widehat{T}^i\mathbb{E}(f|\mathcal{F}_k),$$

the S_n^k converges in law to a gaussian with zero means and variance $\mathbb{E}(Y_1(k,\,1)^2)$.

The next step is to obtain the needed convergence as k goes to infinity.

$$\begin{aligned}
\mathbb{E}([S_n^k-S_n]^2)=&\frac{1}{n}\sum_{i,\,j=0}^{n-1}\mathbb{E}(\widehat{T}^i[f-\mathbb{E}(f|\mathcal{F}_{-k})]\widehat{T}^j[f-\mathbb{E}(f|\mathcal{F}_{-k})])\\
\le&\mathbb{E}([f-\mathbb{E}(f|\mathcal{F}_{-k})]^2)+2\sum_{i=1}^{n-1}|\mathbb{E}([f-\mathbb{E}(f|\mathcal{F}_{-k})]\widehat{T}^i[f-\mathbb{E}(f|\mathcal{F}_{-k})])|\\
\le&2\|f\|_\infty\mathbb{E}(|f-\mathbb{E}(f|\mathcal{F}_{-k})|)+2\sum_{i=1}^{n-1}|\mathbb{E}(\widehat{T}^i f[f-\mathbb{E}(f|\mathcal{F}_{-k})])|\\
=&2\|f\|_\infty\mathbb{E}(|f-\mathbb{E}(f|\mathcal{F}_{-k})|)+2\sum_{i=1}^{n-1}|\mathbb{E}(\widehat{T}^{*i} f[f-\mathbb{E}(f|\mathcal{F}_{-k-i})])|\\
\le&\ 2\|f\|_\infty\sum_{i=k}^{\infty}\mathbb{E}(|f-\mathbb{E}(f|\mathcal{F}_{-i})|),
\end{aligned}$$

which it is smaller than ε uniformly in n, since (3) implies the convergence of the series $\sum_{i=0}^{\infty}\mathbb{E}(|f-\mathbb{E}(f|\mathcal{F}_{-i})|)$.

Collecting the previous estimates follows that S_n converges to a Gaussian of zero mean and variance σ^2.

Next, suppose that $\sigma^2=0$ and $\sum_{n=0}^{\infty}n|\mathbb{E}(f\widehat{T}^n f)|\le\infty$, then

$$\begin{aligned}
(1-\lambda^{-2})\mathbb{E}(g(k,\,\lambda)^2)+\mathbb{E}(Y_1(k,\,\lambda)^2)&\le-\mathbb{E}(\mathbb{E}(f|\mathcal{F}_{-k})^2)+2\mathbb{E}(g(k,\,\lambda)f)\\
&\le\mathbb{E}(f^2)-\mathbb{E}(\mathbb{E}(f|\mathcal{F}_{-k})^2)+2\sum_{n=1}^{\infty}(1-\lambda^{-n})\mathbb{E}(f\widehat{T}^n f)\\
&+2\sum_{n=0}^{\infty}\lambda^{-n}\left[\mathbb{E}(\mathbb{E}(\widehat{T}^{*n}f|\mathcal{F}_{-k})f)-\mathbb{E}(f\widehat{T}^n f)\right]\\
&\le\|f\|_\infty\mathbb{E}(|f-\mathbb{E}(f|\mathcal{F}_{-k})|)+2(1-\lambda^{-1})\sum_{n=0}^{\infty}n|\mathbb{E}(f\widehat{T}^n f)|\\
&+2\|f\|_\infty\left(\sum_{n=0}^{2k-1}\mathbb{E}(|\mathbb{E}(f|\mathcal{F}_{-k})-f|)+\sum_{n=k}^{\infty}\mathbb{E}(|\mathbb{E}(\widehat{T}^{*n}f|\mathcal{F}_0)|)\right)\\
&+2\sum_{n=2k}^{\infty}\mathbb{E}(f\widehat{T}^n f).
\end{aligned}$$

Accordingly, it is possible to define $\phi : (0, 1) \to \mathbb{N}$, $\lim_{\lambda\to 1} \phi(\lambda) = \infty$, such that

$$\mathbb{E}(g(\phi(\lambda), \lambda)^2) \le M \quad \forall\lambda > 1$$
$$\lim_{\lambda\to 1} \mathbb{E}(Y_1(\phi(\lambda), \lambda)^2) = 0,$$

where M is some fixed positive number.

Since $L^2(X)$ is a Hilbert space, and therefore reflexive, the unit ball is compact in the weak topology, so $\{g(\phi(\lambda), \lambda\}_{\lambda>1}$ is a weakly compact set and we can extract a subsequence $\{\lambda_j\}$, $\lim_{j\to\infty} \lambda_j = 1$, such that $\{g(\phi(\lambda_j), \lambda_j\}$ converges weakly to a function $g \in L^2(X)$. In addition, (1.9) implies, for each $\varphi \in L^2(X)$,

$$\mathbb{E}(\widehat{T}^*\varphi\mathbb{E}(f|\mathcal{F}_{-k})) = \mathbb{E}(Y_1(\phi(\lambda_j), \lambda_j)\varphi) + \mathbb{E}(\widehat{T}^*\varphi g(\phi(\lambda_j), \lambda_j)) - \lambda_j^{-1}\mathbb{E}(\varphi g(\phi(\lambda_j), \lambda_j)),$$

and taking the limit $j \to \infty$ yields

$$\mathbb{E}(\widehat{T}^*\varphi f) = \mathbb{E}(\widehat{T}^*\varphi g) - \mathbb{E}(\varphi g) \quad \forall\varphi \in L^2(X).$$

That is

$$\widehat{T}f = \widehat{T}g - g.$$

□

This theorem is rather similar to Theorem 4.4 in [DG], the main difference is the absence, in [DG], of a discussion of the degenerate case $\sigma = 0$. The only other results known to the author that have a breath similar to Theorem 2.1 are contained in [Ch]. The comparison it is not so easy because the results in [Ch] are stated directly in the language of special families of finite partitions. This language it is well suited for applications to the case in which the system is studied by the type of coding called Markov sieves, but it is not so transparent in an abstract contest. At any rate an evident different is that Chernov's result requires the existence of the first moment of the correlations (i.e., $\sum_{n=0}^{\infty} n\mathbb{E}(ff \circ T^n) < \infty$) in order to obtain the CLT while in Theorem 1.2 such a condition is not necessary, unless one wants the coboundary characterization of the functions that yields to a degenerate limit.

§2 Non invertible maps.

In this section we will see how the results of the previous section apply to the case in which T is onto but not one to one.

We choose $\mathcal{F} = \mathcal{F}_0$, so $\mathcal{F}_i = \mathcal{F}$ for all $i \le 0$. Note that if $\mathbb{E}(\phi|\mathcal{F}_1) = \phi$, then $g(x) = \phi(T^{-1}x)$ is well defined, hence Range(T) is exactly the $\mathcal{F}_1$-measurable functions. Moreover, $\widehat{T}\widehat{T}^*$ is an orthogonal projection onto Range(T), while $\mathbb{E}(\cdot|\mathcal{F}_1)$ is an orthogonal projection onto the $\mathcal{F}_1$-mesuarable functions. That is, for each $\phi \in L^1(X)$

$$\widehat{T}\widehat{T}^*\phi = \mathbb{E}(\phi|\mathcal{F}_1).$$

The first condition of Theorem 1.1 is then satisfied quite generally. To see how the theorem works let us apply it to the case of one dimensional maps (i.e. $X = [0, 1]$).

Let us consider a partition of $[0, 1]$ into finitely many intervals $\{I_k\}_{k=1}^p$. And $T : [0, 1] \to [0, 1]$ such that

(1) $T\big|_{\overline{I}_k} \in \mathcal{C}^{(2)}$ for each $k \in \{1, ..., p\}$
(2) $\inf_{x\in[0,1]} |D_xT| \geq \lambda > 1$.

That is a piecewise smooth expanding map. If the reader wants to consider a concrete example, here is a very simple one: the piecewise linear map $T : [0, 1] \to [0, 1]$ define by

$$T(x) = \begin{cases} \frac{9}{2}\left(\frac{1}{9} - x\right) & x \in \left(0, \frac{1}{9}\right) \\ \frac{9}{2}\left(x - \frac{1}{9}\right) & x \in \left(\frac{1}{9}, \frac{3}{9}\right) \\ \frac{9}{2}\left(\frac{5}{9} - x\right) & x \in \left(\frac{3}{9}, \frac{5}{9}\right) \\ \frac{9}{2}\left(x - \frac{5}{9}\right) & x \in \left(\frac{5}{9}, \frac{7}{9}\right) \\ \frac{9}{2}(1 - x) & x \in \left(\frac{7}{9}, 1\right) \end{cases}$$

The map satisfies our assumptions since $|DT| = \frac{9}{2} > 1$.

The following result is well known [HK]:

Theorem 2.1. *There exists a unique probability measure μ, absolutely continuous with respect to Lebesgue, which is invariant with respect to the map T. In addition, there exist $\Lambda \in (0, 1)$ and $K > 0$ such that, for each $f \in BV([0, 1])$ (the space of functions of bounded variation), and $g \in L^1([0, 1], \mu)$*

$$\left|\int_0^1 fg \circ T^n d\mu - \int_0^1 fd\mu \int_0^1 gd\mu\right| \leq K\Lambda^n \|f\|_{\mathrm{BV}} \|g\|_1$$

Since μ is absolutely continuous with respect to the Lebesgue measure m the Radon–Nicodým derivative $h = \frac{d\mu}{dm}$ is in $L^1([0, 1], m)$. For simplicity assume $h \geq \varepsilon > 0$,[5] then it follows

$$\widehat{T}^* f(x) = h(x)^{-1} \sum_{y\in T^{-1}(x)} h(y)f(y)|D_yT|^{-1}.$$

Such a representation implies that the last statement of Theorem 2.1 can be rephrased as follows: for each $f \in BV([0, 1])$, $\int_0^1 fd\mu = 0$

$$\|\widehat{T}^{*n} f\|_\infty \leq K\Lambda^n \|f\|_{\mathrm{BV}}.$$

[5]This is always verified if T is continuous, like in our example; but see [L2] for a discussion of the general case.

It is then immediate to see that Theorem 1.1 applies to this situation yielding the central limit theorem for all functions of bounded variation. The reader can easily see that such a result can be improved obtaining the central limit theorem for functions with less regularity (e.g., by an approximation argument) but this is not the main focus here. In addition, similar results can be obtained for several cases in which the map T consists of infinitely many smooth pieces.

It is also immediate to verify that the theorem will yield the CLT for BV functions also for some non-hyperbolic maps (such as the quadratic family [Y]) or maps that are non-uniformly hyperbolic ([LSV]).

§3 Invertible maps.

In this case it would be useless to choose $\mathcal{F}_0 = \mathcal{F}$: typically this would yield $\mathcal{F}_i = \mathcal{F}$ for each $i \in \mathbb{Z}$. So the choice of $\mathcal{F}_0$ must be motivated by dynamical considerations. Here we will discuss a general class of systems for which such a choice is quite natural: the hyperbolic systems.[6]

For simplicity I will confine the discussion to the case in which X is as compact symplectic manifold with a Riemannian structure that yields a volume form equivalent to the symplectic one and T a piecewise $\mathcal{C}^2$ symplectic map, but see [KS] and [LW] for more general possibilities. By hypothesis the symplectic (or Riemannian) volume μ is invariant. (The more general case of dissipative systems can also be treated with the same arguments, again the details are left to the reader).

We will assume T uniformly hyperbolic, since almost nothing is known on the decay of correlations for non-uniformly hyperbolic systems. By this we mean that at each point $x \in X$ there exists two subspaces $E^u(x)$, $E^s(x) \in \mathcal{T}_x X$, $E^u(x) \cap E^s(x) = \{0\}$ and $E^u(x) \oplus E^s(x) = \mathcal{T}_x X$, invariant (i.e., $D_x T E^{u,s}(x) = E^{u,s}(Tx)$), and there exists $\lambda > 1$ such that for each $x \in X$,

$$\begin{aligned} &\|D_x T v\| \geq \lambda \|v\| \quad \forall v \in E^u(x) \\ &\|D_x T v\| \leq \lambda^{-1} \|v\| \quad \forall v \in E^s(x). \end{aligned}$$

Also, we assume that $E^{u,s}(x)$ depends continuously with respect to x (the above systems are called Anosov, in the smooth case). In the smooth case such systems are known to be ergodic (in fact, Bernoulli), one can see [LW] for sufficient conditions that insure ergodicity also in the non smooth case. To help the reader in better visualizing the following discussion let us consider the simplest possible non-trivial example.

We consider a family of linear maps of the plane defined by

$$\begin{aligned} x_1' &= x_1 + a x_2 \\ x_2' &= x_2, \end{aligned}$$

where a is a real parameter. We use these linear maps to define (discontinuous if $a \notin \mathbb{N}$) maps of the torus by restricting the formulas to the strip $\{0 \leq x_2 \leq 1\}$ and

[6] More generally this strategy can be applied to K-systems.

further taking them modulo 1. In this way we define a mapping T_1 of the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ which is discontinuous on the circle $\{x_2 \in \mathbb{Z}\}$ (except when a is equal to an integer) and preserves the Lebesgue measure μ.

Similarly we define another family of maps depending on the same parameter a by restricting the formulas

$$\begin{aligned} x_1' &= x_1 \\ x_2' &= a x_1 + x_2 \end{aligned}$$

to the strip $\{0 \le x_1 \le 1\}$ and then taking them modulo 1. Thus for each a we get a mapping T_2 of the torus which is discontinuous on the circle $\{x_1 \in \mathbb{Z}\}$ (except when a is equal to an integer) and preserves the Lebesgue measure μ.

Finally we introduce the composition of these maps $T = T_2 T_1$ which depends on one real parameter a. An alternative way of describing the map T is by introducing two fundamental domains for the torus $\mathcal{M}^+ = \{0 \le x_1 + a x_2 \le 1,\ 0 \le x_2 \le 1\}$ and $\mathcal{M}^- = \{0 \le x_1 \le 1,\ 0 \le -a x_1 + x_2 \le 1,\ \}$.

The linear map defined by the matrix

$$\begin{pmatrix} 1 & a \\ a & 1 + a^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix} \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$$

takes $\mathcal{M}^+$ onto $\mathcal{M}^-$ thus defining a map of the torus which is discontinuous at most on the boundary of $\mathcal{M}^+$ and preserves the Lebesgue measure. This is the map T that constitute our toy model.

Let us go back to the more general case, according to [KS] such systems have a natural measurable partition (in fact a K-partition): the partition into stable manifolds.

Such a partition $\mathcal{P}$ can be constructed as to satisfy the following requirements:

(1) there exists a finite number of codimension one smooth manifolds $\{S_i\}_{i=1}^{m_0}$, transversal to the stable direction, such that each $p \in \mathcal{P}$ has the boundaries points belonging to the set $\cup_{j=1}^{m_0} \cup_{n=0}^{\infty} T^{-n} S_i$;[7]

(2) for each $p \in \mathcal{P}$ $\operatorname{diam}(p) \le 2\delta$;[8]

(3) for each $p \in \mathcal{P}$ there exists $\{p_i\}_{i=1}^{k} \subset \mathcal{P}$ such that $T^{-1} p = \cup_{i=1}^{k} p_i$.

The above properties imply that choosing as $\mathcal{F}_0$ the σ-algebra generated by the partition $\mathcal{P}$, then $\{\mathcal{F}_i\}_{i=0}^{\infty}$ has the dynamical properties requested in the hypotheses of Theorem 1.2.

To make the previous statement more clear let us see how such a partition looks like in the concrete example mentioned above.

The map T is piecewise linear and it has constant contracting direction v. Let us call S the discontinuity set of T^{-1} and $S_\infty = \cap_{n=0}^{\infty} T^{-n} S$. Then the stable partition is made of segments along the direction v with the endpoints belonging to S_∞.[9] Since S_∞ is an invariant set, properties (1)-(3) are readily verified.

Further, we will assume that the manifolds $\{S_i\}$ satisfy the following property:

[7] In the discontinuous case such manifolds can be simply chosen as the set of points at which T is not $\mathcal{C}^{(2)}$.

[8] δ is some previously fixed number.

[9] See [LW] for the details of such a construction and the proof that almost every point belongs to one such segment.

Property 0. *For each $i \neq j$ $\overline{S}_i \cap \overline{S}_j$ is either empty or consists of smooth submanifolds I_{ij} of codimension at least two. Moreover, setting*[10]

$$M \equiv \sup_{ij} \sharp\{k \in \{1, ..., m_0\} \mid \overline{S}_k \cap I_{ij} \neq \emptyset\},$$

we require

$$\nu \equiv \lambda^{-1} M < 1.$$

Note that Property 0 may not be satisfied by T but may be enjoyed by T^q, for some $q > 1$. In fact, it is not so hard to see that "generically" this will be the case (i.e., Property 0 will hold for some iterate of the map). In such a situation we can apply all the following to the dynamical system (X, μ, T^q) obtaining the same conclusions as far as the CLT is concerned. Here, for simplicity, we restrict ourselves to the case $q = 1$.

If we think to our model example we see that $M = 2$, so that $\lambda^{-1}M < 1$ if $|a| > \frac{1}{\sqrt{2}}$. The reader can easily compute M for powers of T and see that Property 0 is satisfied for smaller and smaller values of a. Of course, $a = 0$ corresponds to the identity, for which no hyperbolicity is present.

For the systems under consideration holds the following (see [KS] for details)

Property 1. *For each $p \in \mathcal{P}$ define the measure μ_p by*

$$\mathbb{E}(g|\mathcal{F}_0)(x) \equiv \int_p g d\mu_p,$$

for $g \in \mathcal{C}^{(0)}(X)$, and $x \in p$.

Then, calling m_p the measure induced by the Riemannian structure on p, and $\phi_p = \frac{d\mu_p}{dm_p}$ the Radon–Nicodým derivative, there exist c_0 such that $\sup_p \|\phi_p\|_\infty \leq c_0$.

For our simple example we see that $\mu_p = \frac{1}{m_p(p)} m_p$.

The map is invertible, thus $\widehat{T}^* f = f \circ T^{-1}$. A very important consequence of Property 1 is that, if $p \in \mathcal{P}$ and $\mathcal{P}' \subset \mathcal{P}$ is such that $\bigcup_{q \in \mathcal{P}'} q = T^{-n}p$, then for each $f \in L^1(X, \mu)$

$$\int_p f \circ T^{-n} d\mu_p = \sum_{q \in \mathcal{P}'} \mu_p(T^n q) \int_q f d\mu_p.$$

In addition, one can prove the following (see [L1] for a complete discussion of the two-dimensional case).

Property 2. *There exists $K \in \mathbb{R}$ and $\Lambda > 1$, such that, for each $x \in X$ that belongs to $p \in \mathcal{P}$ with diam$(p) \geq \delta$, and for each $g, f \in \mathcal{C}^\alpha(X)$ (Hölder continuous of class $\alpha > 0$), $\int_X g = 0$,*[11]

$$\mathbb{E}(f \widehat{T}^{*n} g|\mathcal{F}_0)(x) \leq K \Lambda^{-n} \|g\|_\alpha^s \|f\|_\alpha^u$$

In the rest of the section we will see that Properties 0-2 imply, for the systems under consideration, the hypotheses of Theorem 1.2.

[10] By $\sharp B$ I mean the cardinality of the set B.

[11] By $\|f\|_\alpha$ we mean the usual $\mathcal{C}^{(\alpha)}$ norm, while $\|f\|_\alpha^s = \sup_{p \in \mathcal{P}} \sup_{x, y \in p} \frac{|f(x) - f(y)|}{\|x - y\|^\alpha} + \|f\|_\infty$; and $\|f\|_\alpha^u$ is defined analogously by using the unstable partition. Essentially, This norms measure the Hölder derivative in the stable (or unstable) direction only.

Lemma 3.1. *Calling $\mathcal{A}_\varepsilon = \{x \in X \mid diam(p(x)) \leq \varepsilon\}$ we have*[12]

$$m(\mathcal{A}_\varepsilon) \leq C\varepsilon$$

for some fixed $C \in \mathbb{R}^+$.

Proof. Since ∂p is made up of points belonging to the preimages of the manifolds S_i, it follows that if $\mathrm{diam}(p) \leq \varepsilon$ then there exists $z \in \partial p$ and $n \in \mathbb{N}$, $i \in \{1, ..., m_0\}$ such that $T^n z \in S_i$. Accordingly, $T^n p$ must lie in a $\lambda^{-n}\varepsilon$ neighborhood of S_i. Such a neighborhood has measure $c_1\lambda^{-n}\varepsilon$, for some fixed c_1. It is then clear that

$$m(\mathcal{A}_\varepsilon) \leq \sum_{n=0}^{\infty} m_0 c_1 \lambda^{-n}\varepsilon = \frac{m_0 c_1}{1-\lambda}\varepsilon.$$

□

The problem in applying our theorem comes from the possible presence in $\mathcal{P}$ of very small elements. On such elements Property 2 does not provide any direct control. To our advantage instead works Lemma 3.1 that informs us that the total measure of the very small pieces is small.

Yet, small pieces may be present.[13] The idea to deal with them consists in iterating them: if $T^{-n}|_p$ is smooth, then $\mathrm{diam}(T^{-n}p) \geq \lambda^n \mathrm{diam}(p)$. Unfortunately, in general T is not smooth so we have to handle the iteration with more care.

Fix $p \in \mathcal{P}$, by construction there exists $\mathcal{P}_1 \subset \mathcal{P}$ such that $T^{-1}p = \bigcup_{q\in\mathcal{P}_1} q$. Call $\mathcal{P}_1^- = \{q \in \mathcal{P}_1 \mid \mathrm{diam}(q) \leq \delta\}$ and $p_1 = \bigcup_{q\in\mathcal{P}_1^-} Tq \subset p$. In other words p_1 consists in the part of p that, under the actions of T^{-1}, does not give rise to sufficiently large elements of the partition. The process can obviously be iterated: let $\mathcal{P}_2$ be the collection such that $T^{-2}p_1 = \bigcap_{q\in\mathcal{P}_2} T^{-2}q$, $\mathcal{P}_2^- = \{q \in \mathcal{P}_2 \mid \mathrm{diam}(q) \leq \delta\}$, $p_2 = \bigcup_{q\in\mathcal{P}_2^-} T^2 q \subset p_1$ and so on.

Lemma 3.2. *If δ is chosen sufficiently small and $p \in \mathcal{A}_\varepsilon$, then for $n \geq \frac{\log \varepsilon^{-1}\delta}{\log \nu^{-1}} + m$,*

$$m_p(p_n) \leq \varepsilon \nu^m.$$

Proof. By choosing δ sufficiently small we can insure, thanks to Property 0, that each element with diameter less than δ can intersect at most M manifolds S_i. Since the S_i describe all the possible discontinuities in our system, it follows that $\sharp\mathcal{P}_1 \leq M$. But the same argument applies to each connected piece of p_j: since the diameter of $T^{-l}p_j$ is, by definition, less than δ, for $l < j$, it follows that $T^{-l-1}p_j$

[12]By $m(\cdot)$ we mean the symplectic or Riemannian metric that, according to our hypotheses, is the invariant measure of the system.

[13]In fact, this is certainly the case in the non-smooth case. If T is smooth, then it is possible to construct $\mathcal{P}$ in such a way that $\mathrm{diam}(\mathcal{P}) \geq \delta$ for some fixed δ, by using Markov partitions. When finite Markov partitions are available the present method boils down to a repackaging of well known facts.

can consists of, at most, M elements of the partition. In conclusion, $\sharp\mathcal{P}_n^- \leq M^n$, and

$$m_p(p_n) \leq M^n\lambda^{-n}\delta \leq \nu^m\varepsilon.$$

□

Using the above estimates, for each $n,\, k \in \mathbb{N}$, $k < n$, and calling $\chi_{\mathcal{B}}$ the characteristic function of the set $\mathcal{B}$, we have

$$\begin{aligned}
\mathbb{E}(|\mathbb{E}(g\widehat{T}^{*n}f|\mathcal{F}_0)|) &\leq \sum_{m=0}^{\infty}\mathbb{E}(\chi_{\mathcal{A}_{\nu^m\delta}}|\mathbb{E}(g\widehat{T}^{*n}f|\mathcal{F}_0)|)\\
&= \sum_{m=0}^{n-k}\mathbb{E}(\chi_{\mathcal{A}_{\nu^m\delta}}|\mathbb{E}(\widehat{T}^{*k}[\widehat{T}^k g\widehat{T}^{*(n-k)}f]|\mathcal{F}_0)|)\\
&+ \sum_{m=n-k+1}^{\infty}\|f\|_\infty\|g\|_\infty\mu(\mathcal{A}_{\nu^m\delta}).
\end{aligned}$$

Next, for each $x \in p \subset \mathcal{A}_{\nu^m\delta}$, using the notations of the Lemma 3.2 and setting $\phi = \widehat{T}^k g\widehat{T}^{*(n-k)}f$,

$$\begin{aligned}
|\mathbb{E}(g\widehat{T}^{*n}f|\mathcal{F}_0)| &= \left|\int_p \phi\circ T^{-k}d\mu_p\right| \leq \left|\int_{p_k}\phi\circ T^{-k}d\mu_p\right| + \left|\int_{p\setminus p_k} f\circ T^{-k}d\mu_p\right|\\
&\leq c_0 m_p(p_k)\|\phi\|_\infty + \sum_{\substack{q\in\mathcal{P}\\ T^k q\subset p\setminus p_k}}\mu_p(T^k q)\left|\int_q\phi d\mu_p\right|\\
&\leq c_0\|f\|_\infty\|g\|_\infty\delta\nu^k + \sum_{\substack{q\in\mathcal{P}\\ T^k q\subset p\setminus p_k}}\mu_p(T^k q)|\mathbb{E}(\phi|\mathcal{F}_0)|\\
&\leq c_0\|f\|_\infty\|g\|_\infty\delta\nu^k + K\Lambda^{-n+k}\|\widehat{T}^k g\|_\alpha^s\|f\|_\alpha^u\\
&\leq \|f\|_\alpha\|g\|_\alpha[c_0\delta\nu^k + K\Lambda^{-n+k}\lambda^{-k\alpha}].
\end{aligned}$$

Using the above estimates yield

$$\mathbb{E}(|\mathbb{E}(g\widehat{T}^{*n}f|\mathcal{F}_0)|) \leq \|f\|_\alpha\|g\|_\alpha\left[(n-k)(c_0\nu^k\delta + K\Lambda^{-n+k}) + \frac{C\delta}{1-\nu}\nu^{n-k+1}\right].$$

Hence, by choosing $k = \left[\frac{n}{2}\right]$, it follows that there exists $c_1 \in \mathbb{R}^+$ such that

$$\mathbb{E}(|\mathbb{E}(gT^{*n}f|\mathcal{F}_0)| \leq c_1 n\max\{\nu,\, \Lambda^{-1}\}^{\frac{n}{2}}.$$

This verifies the hypotheses 1–2 of Theorem 1.2. Hypothesis 3 follows trivially from the assumption $f \in \mathcal{C}^{(\alpha)}$ (or f piecewise Hölder). To see this, consider that $\mathcal{F}_{-k}$ is the σ-algebra associated to the partition $T^k\mathcal{P}$. But, if $p \in T^k\mathcal{P}$, then

$\text{diam}(p) \leq 2\lambda^{-k}\delta$. Thus, for each $x \in p \in T^k\mathcal{P}$, let $q \in \mathcal{P}$ be the unique element such that $p \subset q$, then

$$|\mathbb{E}(f|\mathcal{F}_{-k})(x) - f(x)| = |\mu_q(p)^{-1}\int_p f d\mu_q - f(x)| \leq \sup_{y\in p}|f(y) - f(x)|$$
$$\leq \|f\|_\alpha \text{diam}(p)^\alpha \leq \|f\|_\alpha 2^\alpha \lambda^{-\alpha k}\delta^\alpha.$$

Hence Theorem 1.2 applies to the class of systems under consideration and yields the CLT for all Hölder continuous functions (in fact, the above considerations can easily be pushed to obtain the CLT for much larger classes of functions; e.g. piecewise continuous functions with a logarithmic modulus of continuity).

References

[BSC1] L.A.Bunimovich, Ya.G. Sinai, N.I. Chernov, *Markov partitions for two dimensional billiards*, Rus. Math. Surv. **45** (1990), 105–152.

[BSC2] L.A.Bunimovich, Ya.G. Sinai, N.I. Chernov, *Statistical properties of two dimensional hyperbolic billiards*, Rus. Math. Surv. **46** (1991), 47–106.

[Ch] N.I.Chernov, *Limit Theorems and Markov approximations for chaotic dynamical systems*, preprint (1994).

[De] M. Denker, *The Central Limit theorem for dynamical systems*, Dynamical Systems and Ergodic Theory, Banach Center Publications, **23**, Czeslaw Olech editor, PWN–Polish Scientific Publisher, Warsaw, 1989.

[DG] D.Dürr, S.Goldstein, *Remarks on the Central Limit Theorem for weakly dependent random variables*, Stochastic Processes–Mathematics and Physics, S.Albeverio, Ph.Blanchard and L.Streit editors, Lecture Notes in Mathematics, vol. **1158**, Springer-Verlag, 1986.

[Go] M.I.Gordin, *The Central Limit Theorem for stationary processes*, Soviet.Math.Dokl. **10** n.5 (1986), 1174–1176.

[HK] F. Hofbauer, G. Keller, *Ergodic properties of invariant measures for piecewise monotone transformations*, Math. Zeit. **180** (1982), 119–140.

[IL] I.A.Ibragimov, Y.V.Linnik, *Independent and stationary sequences of random variables*, Wolters-Noordhoff, Groningen, 1971.

[Ke] G. Keller, *Un théorem de la limite centrale pour une classe de transformations monotones par morceaux*, Comptes Rendus de l'Académie des Sciences, Série A **291** (1980), 155–158.

[KS] A.Katok, J.M. Strelcyn with collaboration of F. Ledrappier and F. Przytycki, *Invariant Manifolds, Entropy and Billiards, Smooth Maps with Singularities*, Lectures Notes in Mathematics **1222** (1986).

[KV] C.Kipnis, S.R.S.Varadhan, *Central Limit Theorem for addictive functions of reversible Markov process and applications to simple exclusions*, Commun.Math.Phys. **104** (1986), 1–19.

[L1] C. Liverani, *Decay of correlations*, to appear in Annals Of Mathematics.

[L2] C. Liverani, *Decay of Correlations in Piecewise Expanding maps*, Journal of Statistical Physics **78,** 3/4 (1995), 1111–1129.

[LW] C. Liverani, M.P.Wojtokowski, *Ergodicity in Hamiltonian Systems*, Dynamics Reported **4** (1995), 130–202.

[LSV] A.Lambert, S.Siboni, V.Vaienti, *Statistical properties of a non-uniformly hyperbolic map of the interval*, Journal of Stat.Phys. **72** (1993), 1305.

[Ne] J.Neveu, *Mathematical foundations of the calculus of probability*, Holden-Day, San Francisco, 1965.

[Y] L.-S.Young, *Decay of correlations for certain quadratic maps*, Comm.Math.Phys. **146** (1992), 123–138.

Liverani Carlangelo, Mathematics Department, University of Rome II, Tor Vergata, 00133 Rome, Italy.

E-mail address: liverani@mat.utovrm.it