SPECTRAL METHODS FOR STATISTICAL LIMIT THEOREMS IN DYNAMICS

ANSHUL ADVE

ABSTRACT. We present a functional analytic approach to proving the central limit theorem (CLT) for sequences of random variables which are weakly dependent. In particular, we focus on sequences which arise naturally from dynamical systems. The method is standard, but in Sections 4 and 5 we see that it provides significantly more information than is necessary to prove the CLT. We characterize the consequences of the method via decay of correlations of the random variables, and discuss how to use the extra information to construct Edgeworth series, which are higher order versions of the CLT.

CONTENTS

1. INTRODUCTION

Some of the most fundamental theorems in both pure and applied probability theory concern the average behavior of independent and identically distributed (iid) random variables.

For the remainder of this paper, fix a probability space (X, μ). By default, $L^p = L^p(X, \mu)$.

Theorem 1.1 (Law of Large Numbers). *Let $\psi^n \in L^1$ be iid random variables with $\int \psi^n \,\mathrm{d}\mu = 0$. Denote $\psi_n = \psi^1 + \cdots + \psi^n$. Then $\frac{1}{n}\psi_n \to 0$ in L^1 and pointwise a.e.*

As we shall see later, this is in some sense a first order result. If the ψ^n lie in L^2, then for most points in X we get a "second order" refinement.

Date: September 11, 2019.

Theorem 1.2 (Classical Central Limit Theorem)**.** *Let $\psi^n \in L^2$ be real-valued iid random variables with mean $\int \psi^n \,\mathrm{d}\mu = 0$ and standard deviation $\sigma = \|\psi^n\|_{L^2}$. Denote $\psi_n = \psi^1 + \cdots + \psi^n$. Then $\frac{1}{\sqrt{n}}\psi_n$ converges in distribution to the normal distribution $N(0,\sigma^2)$ with mean 0 and variance σ^2. Explicitly, one has*

$$\mu\left\{x : \frac{1}{\sqrt{n}}\psi_n(x) \le t\right\} \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{t} e^{-y^2/2\sigma^2}\,\mathrm{d}y$$

for each $t \in \mathbf{R}$.

See [4] for proofs.

In ergodic theory, a similar setup arises, and it is natural to ask whether or not there exist analogs of Theorems 1.1 and 1.2.

Definition 1.3. A *measure preserving system* (MPS) is a triple (X,μ,T) consisting of a probability space (X,μ) and a map $T\colon X \to X$ which preserves μ, in the sense that for each $f \in L^1$, one has $f \circ T \in L^1$ and

$$\int f \circ T \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu.$$

Measure preserving systems are the central objects of study in ergodic theory. Henceforth, fix a map $T\colon X \to X$ which makes (X,μ) into an MPS. This map should be thought of as describing how X evolves with each timestep.

A measurable map $\psi\colon X \to \mathbf{R}$ is called an *observable*. One can think of ψ as measuring some property of the system X. The measurement of this property at time n is then given by $\psi^n = \psi \circ T^n$. Since T preserves μ, the ψ^n are identically distributed, though likely not independent. However, if we replace independence with an indecomposability assumption, then we get an analog of Theorem 1.1.

Definition 1.4. The system (X,μ,T) is *ergodic* if for observables f, one has that $f = f \circ T$ a.e. implies f is constant a.e.

Taking f to be the indicator function of some set, this implies (and in fact is equivalent to) that X cannot be decomposed into two invariant subspaces of positive measure.

Theorem 1.5 (Ergodic Theorem)**.** *Suppose (X,μ,T) is ergodic. Let $\psi \in L^1$ with $\int \psi \,\mathrm{d}\mu = 0$, and set $\psi^n = \psi \circ T^n$. Denote $\psi_n = \psi^1 + \cdots + \psi^n$. Then $\frac{1}{n}\psi_n \to 0$ in L^1 and pointwise a.e.*

A proof may be found in [5].

This is evidently analogous to Theorem 1.1. Theorem 1.2 is more difficult.

Question 1.6. *Let $\psi \in L^2$ be a real-valued observable with $\int \psi \,\mathrm{d}\mu = 0$, and set $\psi^n = \psi \circ T^n$. Denote $\psi_n = \psi^1 + \cdots + \psi^n$. Does there exist some σ such that $\frac{1}{\sqrt{n}}\psi_n \to N(0,\sigma^2)$ in distribution?*

If for some ψ the answer to Question 1.6 is positive, then we say that ψ *satisfies the CLT*, or some such phrase.

In general, the answer to Question 1.6 is negative. The difficulty relative to Theorem 1.5 arises from the fact that $\frac{1}{\sqrt{n}}\psi_n$ is much more sensitive to the dependence between the ψ^n than $\frac{1}{n}\psi_n$. One might hope, therefore, that if the pairwise correlation between the ψ^n decays sufficiently fast, then the CLT will hold for ψ. This is not strictly true. In fact, in Theorem 1.2, we cannot replace independence with pairwise independence. Nevertheless, Remark 1.8 below suggests a general principle along these lines which is helpful to keep in mind.

Definition 1.7. One defines the *covariance* of $f, g \in L^2$ by

$$\operatorname{cov}(f, g) = \int fg \, \mathrm{d}\mu - \left(\int f \, \mathrm{d}\mu\right)\left(\int g \, \mathrm{d}\mu\right).$$

If f, g are independent, then $\operatorname{cov}(f, g) = 0$. Thus covariance provides one way to quantify how much correlation there is between f and g.

Remark 1.8. If one has exponential decay of covariances $\operatorname{cov}(f, g \circ T^n) = O(r^n)$, $r < 1$, for some large class $\mathcal{C} \subseteq L^2$ of observables f, g, then one should expect any $f \in \mathcal{C}$ to satisfy the CLT (see, e.g., [1, 2, 8]). In practice, X is often a compact metric space, and $\mathcal{C}$ consists of the Hölder continuous functions $X \to \mathbf{R}$ (e.g., in Section 3.1). We emphasize that this is purely a heuristic — for most natural notions of correlation, neither exponential decay of correlations nor the CLT imply the other.

Decay of correlations is discussed in depth in [1].

The same philosophy as in Remark 1.8 suggests that if T is "sufficiently chaotic," then the ψ^n should quickly become less and less correlated, and the CLT should hold. Indeed, we will see manifestations of this intuition in Section 3.

The structure of the paper is as follows. In Section 2, we present a general method to prove the CLT for observables ψ under suitable conditions. The key to this method is to find a Banach space which is somehow adapted both to ψ and to the dynamics of the system (X, μ, T). In Section 3, we will study two concrete classes of systems for which Banach spaces of the form described in Section 2 may be constructed (for sufficiently regular observables ψ). Many more examples can be found for instance in [2, 3, 7, 8]. In Sections 4 and 5, we switch our focus to studying more general consequences of the existence of such a Banach space. Section 4 introduces a notion of covariance for more than two functions, and provides a condition equivalent to the existence of such Banach spaces in terms of the decay of these "higher" covariances. This gives a precise statement in the same vein as the heuristic given by Remark 1.8. Finally, Section 5 shows how one can compute "higher order information" from these Banach spaces, and discusses how to use this information to compute arbitrarily precise error terms in the CLT under some extra conditions.

2. From Spectral Gap to CLT

Throughout this section, fix an observable $\psi \in L^\infty$ with $\int \psi \, \mathrm{d}\mu = 0$. Denote

$$\psi_n = \psi + \psi \circ T + \cdots + \psi \circ T^{n-1}.$$

2.1. **The Spectral Method.**

Definition 2.1. The *characteristic function* of an observable $f \colon X \to \mathbf{R}$ is the function $\varphi \colon \mathbf{R} \to \mathbf{C}$ given by

$$\varphi(t) = \int e^{itf} \, \mathrm{d}\mu.$$

Our proof of CLT for ψ will go through the characteristic functions of the ψ_n and the normal distribution. In hindsight, this is very natural. The CLT concerns the distribution of a sum of independent random variables, which is the convolution of the distributions of the summands. The Fourier transform turns convolution into multiplication, so one might expect that studying the Fourier transform of the distribution might be helpful. This is precisely the characteristic function.

Theorem 2.2 (Lévy Continuity Theorem). *Let f_n, f be observables and φ_n, φ their characteristic functions. Then $f_n \to f$ in distribution if and only if $\varphi_n \to \varphi$ pointwise.*

This is Theorem 3.3.17 in [4].

The characteristic function of the normal distribution $N(0, \sigma^2)$ with mean 0 and variance σ^2 is $e^{-\frac{1}{2}\sigma^2 t^2}$. Thus by the continuity theorem, it is enough to show that

$$\int e^{i \frac{t}{\sqrt{n}} \psi_n} \, \mathrm{d}\mu \to e^{-\frac{1}{2}\sigma^2 t^2} \tag{2.3}$$

for each $t \in \mathbf{R}$.

We will study the characteristic functions of the ψ_n from a functional analytic viewpoint.

Definition 2.4. The *Koopman operator* $\mathcal{U} : L^2 \to L^2$ is defined by $f \mapsto f \circ T$.

Since T preserves μ, this is an isometry on L^2. It turns out, however, that the most useful operator to study in our context is the adjoint to $\mathcal{U}$.

Definition 2.5. The *transfer operator* $\mathcal{P} : L^2 \to L^2$ is defined to be the adjoint of the Koopman operator. That is, it is the unique bounded linear map satisfying

$$\int (f \circ T) g \, \mathrm{d}\mu = \int f(\mathcal{P} g) \, \mathrm{d}\mu.$$

for all $f, g \in L^2$.

This definition is already natural in view of Remark 1.8, because we can write

$$\begin{aligned} \operatorname{cov}(f, g \circ T^n) &= \int f(g \circ T^n) \, \mathrm{d}\mu - \left(\int f \, \mathrm{d}\mu\right)\left(\int g \circ T^n \, \mathrm{d}\mu\right) \\ &= \int (\mathcal{P}^n f) g \, \mathrm{d}\mu - \left(\int f \, \mathrm{d}\mu\right)\left(\int g \, \mathrm{d}\mu\right). \end{aligned}$$

Moreover, we will see in Lemma 2.10 and (2.12) that we can express the characteristic functions of the ψ_n in a very simple way in terms of a twisted version of the transfer operator (see Definition 2.9). The advantage of studying the transfer operator over the Koopman operator is that the transfer operator tends to increase regularity. This is explicated from an intuitive standpoint in the remark below, and we will see this explicitly in Section 3.1 and a little more subtly in Section 3.2.

Remark 2.6. One should think of $\mathcal{P}$ as the natural action of T when functions are viewed as (absolutely continuous) mass densities. More precisely, given $f \in L^2$, identify f with the measure $\mu_f = f \, \mathrm{d}\mu$. Maps act on measures via pushforward, so from T we get a new measure $T_*(f \, \mathrm{d}\mu)$ which gives a set $E \subset X$ the measure $\mu_f(T^{-1}E)$. Then one can check that $\mathcal{P} f$ is the density of $T_*(f \, \mathrm{d}\mu)$ with respect to μ. For instance, suppose $X = \mathbf{R}/\mathbf{Z}$, μ is Haar measure (i.e. Lebesgue measure when $\mathbf{R}/\mathbf{Z}$ is identified with $[0, 1]$), and $T(x) = 2x$. Take f to be a "spiky" function supported on $[-\varepsilon, \varepsilon]$. Then the graph of $\mathcal{P}f$ will look like the graph of f, but it will be spread out to be supported on $[-2\varepsilon, 2\varepsilon]$ and will be half as tall. So $\mathcal{P}$ "smooths out spikes," and thus increases regularity. In general, we have already remarked that we expect a CLT when T is "chaotic." Intuitively, what it means for T to be chaotic is that it pulls points which are close together apart. In our example where $T(x) = 2x$, we see that indeed $\operatorname{dist}(T^n x, T^n y) = 2^n \operatorname{dist}(x, y)$ when x and y are very close. So if T is chaotic, then as we saw, $\mathcal{P}$ should spread spiky functions out in the horizontal direction, thereby increasing regularity.

The takeaway from this is that if we restrict the transfer operator $\mathcal{P}$ to act on a Banach space of functions whose norm controls regularity, then $\mathcal{P}$ should have nice functional analytic properties. We will be more precise about what properties we want later in this section. For now, we need to collect some facts about $\mathcal{P}$ and relate it to the characteristic functions of the ψ_n.

Proposition 2.7. *The transfer operator preserves constant functions and integrals.*

Proof. If g is constant and $f \in L^2$, then

$$\int (f \circ T) g \, \mathrm{d}\mu = \int f g \, \mathrm{d}\mu$$

since T preserves μ. Thus $\mathcal{P}g = g$. In addition,

$$\int f \, \mathrm{d}\mu = \int (\mathbf{1} \circ T) f \, \mathrm{d}\mu = \int \mathcal{P}f \, \mathrm{d}\mu,$$

where $\mathbf{1}$ is the constant function $x \mapsto 1$. □

Proposition 2.8. *Let $f \in L^\infty$. Then $\|\mathcal{P}f\|_\infty \leq \|f\|_\infty$.*

Proof. Viewing the L^∞ norm as the operator norm on the dual of L^1, we can write

$$\|\mathcal{P}f\|_\infty = \sup_{\|g\|_1 \leq 1} \left| \int (\mathcal{P}f) g \, \mathrm{d}\mu \right| = \sup_{\|g\|_1 \leq 1} \left| \int f (g \circ T) \, \mathrm{d}\mu \right| \leq \sup_{\|h\|_1 \leq 1} \left| \int f h \, \mathrm{d}\mu \right| = \|f\|_\infty.$$

The inequality is due to the fact that T preserves μ, so $\|g \circ T\|_1 = \|g\|_1$. The second equality follows from the same facts, since it holds for $g \in L^2$, and thus extends to $g \in L^1$ by continuity. □

The key to relating $\mathcal{P}$ with the characteristic functions of the ψ_n is the following.

Definition 2.9. The *twisted transfer operator* $\mathcal{P}_z \colon L^2 \to L^2$, indexed by a complex number z, is given by $f \mapsto \mathcal{P}(e^{iz\psi} f)$.

Note that $\mathcal{P}_0 = \mathcal{P}$.

Lemma 2.10. *For any $f \in L^2$ and $n \in \mathbf{N}$,*

$$\mathcal{P}_z^n f = \mathcal{P}^n (e^{iz\psi_n} f).$$

Proof. This is the definition of $\mathcal{P}_z$ when $n = 1$. For $n > 1$, we wish to show that

$$\mathcal{P}_z \mathcal{P}^n (e^{iz\psi_n} f) = \mathcal{P}(e^{iz\psi} \mathcal{P}^n (e^{iz\psi_n} f)) = \mathcal{P}^{n+1} (e^{iz\psi_{n+1}} f). \tag{2.11}$$

Indeed, for all $g \in L^2$,

$$\begin{aligned}\int \mathcal{P}(e^{iz\psi}\mathcal{P}^n(e^{iz\psi_n}f))g \,\mathrm{d}\mu &= \int e^{iz\psi}\mathcal{P}^n(e^{iz\psi_n}f)(g\circ T)\,\mathrm{d}\mu \\ &= \int e^{iz\psi_n} f e^{iz\psi\circ T^n}(g\circ T^{n+1})\,\mathrm{d}\mu \\ &= \int e^{iz\psi_{n+1}} f(g\circ T^{n+1})\,\mathrm{d}\mu \\ &= \int \mathcal{P}^{n+1}(e^{iz\psi_{n+1}}f)g\,\mathrm{d}\mu.\end{aligned}$$

Thus (2.11) holds and we conclude by induction. □

By Lemma 2.10,

$$\int e^{i\frac{t}{\sqrt{n}}\psi_n}\,\mathrm{d}\mu = \int \mathcal{P}^n_{\frac{t}{\sqrt{n}}}(\mathbf{1})\,\mathrm{d}\mu. \tag{2.12}$$

This reduces the CLT to understanding the twisted transfer operators $\mathcal{P}_z$ for small real z. As discussed above, one can often find a Banach space on which $\mathcal{P} = \mathcal{P}_0$ acts with nice properties. We will then use classical perturbation theory of operators to extend these properties to $\mathcal{P}_z$.

Definition 2.13. Let L be a bounded linear operator on a Banach space. The *spectrum* of L, denoted $\sigma(S)$, is the set of complex numbers λ such that $\lambda I - L$ does not have a bounded inverse.

The spectrum is always a nonempty compact subset of $\mathbf{C}$ (see, e.g., [10]).

Definition 2.14. The *spectral radius* $\rho(L)$ of a bounded linear operator L on a Banach space is the maximum modulus of a point in $\sigma(L)$.

Gelfand gave the following formula for the spectral radius (again, see [10]).

Proposition 2.15 (Spectral Radius Formula)**.** *Let L be a bounded linear operator on a Banach space. Then the spectral radius is given by*

$$\rho(L) = \inf_n \|L^n\|^{\frac{1}{n}} = \lim_{n\to\infty} \|L^n\|^{\frac{1}{n}}.$$

Now we make the assumptions on ψ and the system (X, μ, T) from which we will deduce the CLT. Suppose there exists a Banach space $\mathcal{B}$ satisfying (B1)-(B6) below. Here and throughout, M_g denotes the multiplication operator $f \mapsto fg$.

- (B1) $\mathcal{B}$ contains the constant functions.
- (B2) $\mathcal{B} \cap L^\infty$ is dense in $\mathcal{B}$.
- (B3) Integration with respect to μ extends from $\mathcal{B} \cap L^\infty$ to a bounded linear functional on $\mathcal{B}$.
- (B4) The multiplication map $M_\psi : f \mapsto \psi f$ extends from $\mathcal{B} \cap L^\infty$ to a bounded linear operator on $\mathcal{B}$.
- (B5) The transfer operator $\mathcal{P}$ maps $\mathcal{B} \cap L^\infty$ into itself, and extends to a bounded linear operator on $\mathcal{B}$.
- (B6) The spectrum of the restriction of $\mathcal{P}$ to $\mathcal{B}_0 = \{f \in \mathcal{B} : \int f \,\mathrm{d}\mu = 0\}$ (where integration is defined via (B2)) is contained in the disc $\{\lambda \in \mathbf{C} : |\lambda| < r\}$ for some $r < 1$.

Remark 2.16. We are purposely sloppy with notation in (B2) — it will be clear from context how to interpret $\mathcal{B} \cap L^\infty$. We only need (B2) to define integration, M_ψ, and $\mathcal{P}$ in a unique and natural way.

Remark 2.17. By Proposition 2.8, the condition in (B5) that $\mathcal{P}$ maps $\mathcal{B} \cap L^\infty$ into itself is equivalent to $\mathcal{P}$ mapping $\mathcal{B} \cap L^\infty$ into $\mathcal{B}$.

Remark 2.18. The most important properties are (B4) and (B6). Note that (B4) is the only property that depends on ψ. Combined with (B1), it ensures in particular that $\psi \in \mathcal{B}$. When (B6) holds, we say that $\mathcal{P}$ has "spectral gap." We have already seen that the space of constant functions is an eigenspace of $\mathcal{P}$ with eigenvalue 1. In fact, we will see later that the spectrum of $\mathcal{P}$ is exactly $\sigma(\mathcal{P}|_{\mathcal{B}_0}) \cup \{1\}$. Thus (B6) does indeed tell us that $\sigma(P)$ has a "gap." Via perturbation theory, this is the property that controls the behavior of $\mathcal{P}_z$ for small z.

The property (B5) allows us to define the transfer operator on $\mathcal{B}$. We wish to define the twisted transfer operators on $\mathcal{B}$ as well. By (B4), it makes sense to put

$$M_{e^{iz\psi}} := \sum_{j=0}^{\infty} \frac{(iz)^j}{j!} M_\psi^j$$

as an operator on $\mathcal{B}$. Thus we can define the twisted transfer operators by

$$\mathcal{P}_z := \mathcal{P} M_{e^{iz\psi}} = \sum_{j=0}^{\infty} \frac{(iz)^j}{j!} \mathcal{P} M_\psi^j. \tag{2.19}$$

On $\mathcal{B} \cap L^\infty$, this agrees with Definition 2.9.

Definition 2.20. A function f on an open subset of $\mathbf{C}$ taking values in a Banach space $\mathcal{C}$ is *holomorphic* at z if

$$\frac{f(z+h) - f(z)}{h}$$

converges to a limit in $\mathcal{C}$ as $h \to 0$. If f is defined and holomorphic on all of $\mathbf{C}$, then f is *entire*.

Most of the results from elementary complex analysis (Morera's theorem, equivalence of holomorphicity and analyticity, etc.) translate *mutatis mutandis* to Banach space valued functions, with analogous proofs.

By (2.19), $\mathcal{P}_z$ is an entire function of z, and

$$\mathcal{P}_0^{(j)} := \frac{\mathrm{d}^j}{\mathrm{d}z^j}\Big|_{z=0} \mathcal{P}_z = \frac{i^j}{j!} \mathcal{P} M_\psi^j.$$

The final tool we need to analyze the twisted transfer operators is the following theorem, which extends (B6) from $\mathcal{P}$ to $\mathcal{P}_z$ and provides a sort of eigenspace decomposition. The proof of this theorem requires some functional analytic machinery, so we defer it to Section 2.2.

Theorem 2.21 (Analytic Perturbation Theorem)**.** *For sufficiently small z, there is a decomposition*

$$\mathcal{P}_z = \lambda(z)\Pi_z + N_z, \tag{2.22}$$

where $\lambda(z) \in \mathbf{C}$ and Π_z, N_z are bounded operators on $\mathcal{B}$, satisfying (D1)-(D4) below.

(D1) $\lambda(z)$, Π_z, and N_z vary holomorphically with z.

(D2) Π_z *is a projection, i.e.,* $\Pi_z^2 = \Pi_z$.
(D3) $\Pi_z N_z = N_z \Pi_z = 0$.
(D4) The spectral radius $\rho(N_z) < r < 1$, *where* r *is as in (B6).*

From now on, let $\lambda(z)$, Π_z, and N_z be as in the theorem. It will be important that (D4), combined with the spectral radius formula, implies

$$\|N_z^n\| = O(r^n). \tag{2.23}$$

For reference, we isolate a lemma that appears in the proof of Theorem 2.21.

Lemma 2.24. *We have* $\lambda(0) = 1$ *and* $\Pi_0 f = \left(\int f \, \mathrm{d}\mu\right) \mathbf{1}$.

Proof. This is the content of Lemma 2.30. □

Combining the decomposition of Theorem 2.21 with (2.12), we get

$$\int e^{i\frac{t}{\sqrt{n}}\psi_n} \, \mathrm{d}\mu = \int \mathcal{P}^n_{\frac{t}{\sqrt{n}}}(\mathbf{1}) \, \mathrm{d}\mu = \lambda\left(\frac{t}{\sqrt{n}}\right)^n \int \Pi_{\frac{t}{\sqrt{n}}}(\mathbf{1}) \, \mathrm{d}\mu + \int N^n_{\frac{t}{\sqrt{n}}}(\mathbf{1}) \, \mathrm{d}\mu. \tag{2.25}$$

By continuity, $\lambda(t/\sqrt{n})$ is close to $\lambda(0) = 1$ for large n, so by (2.23),

$$\lambda\left(\frac{t}{\sqrt{n}}\right)^{-n} \int N^n_{\frac{t}{\sqrt{n}}}(\mathbf{1}) \, \mathrm{d}\mu \xrightarrow{n\to\infty} 0.$$

Similarly,

$$\int \Pi_{\frac{t}{\sqrt{n}}}(\mathbf{1}) \, \mathrm{d}\mu \xrightarrow{n\to\infty} \int \Pi_0(\mathbf{1}) \, \mathrm{d}\mu = 1.$$

Thus (2.25) becomes

$$\int e^{i\frac{t}{\sqrt{n}}\psi_n} \, \mathrm{d}\mu = \lambda\left(\frac{t}{\sqrt{n}}\right)^n [1 + o(1)].$$

Taylor expanding λ around 0, we can rewrite this as

$$\int e^{i\frac{t}{\sqrt{n}}\psi_n} \, \mathrm{d}\mu = \left(1 + \lambda'(0)\frac{t}{\sqrt{n}} + \lambda''(0)\frac{t^2}{2n} + O\left(\frac{t^3}{n\sqrt{n}}\right)\right)^n [1 + o(1)]. \tag{2.26}$$

Lemma 2.27. *We have* $\lambda'(0) = 0$ *and*

$$\lambda''(0) = -\lim_{n\to\infty} \frac{1}{n} \int \psi_n^2 \, \mathrm{d}\mu.$$

Existence of the limit is part of the statement.

Proof. This computation is carried out in Section 5.1. The statement of the lemma is finally deduced in Example 5.5. □

Plugging Lemma 2.27 into (2.26) gives

$$\int e^{i\frac{t}{\sqrt{n}}\psi_n} \, \mathrm{d}\mu = e^{\frac{1}{2}\lambda''(0)t^2}[1 + o(1)].$$

Thus if we put

$$\sigma^2 = -\lambda''(0) = \lim_{n\to\infty} \frac{1}{n} \int \psi_n^2 \, \mathrm{d}\mu,$$

then we conclude (2.3), and can apply the Lévy continuity theorem.

Hence whenever a Banach space satisfying (B1)-(B6) exists, ψ satisfies the CLT.

2.2. **Analytic Perturbation Theory.** This section generally follows [11]. A more complete treatment may be found in [9].

The proof of Theorem 2.21 rests on the holomorphic functional calculus, which gives a meaningful way to evaluate holomorphic functions at operators (or indeed at any element of a Banach algebra).

Let L be a bounded linear operator on a Banach space $\mathcal{B}$. If p is a polynomial, then it makes sense to write $p(L)$. If $p(t) = t^2+1$, for example, then $p(L) = L \circ L + I$. Now suppose $g : U \to \mathbf{C}$ is holomorphic on an open set $U \subseteq \mathbf{C}$. When can we make sense of $g(L)$? If $g(\zeta) = (\lambda - \zeta)^{-1}$ for $\lambda \in \mathbf{C} \setminus \sigma(L)$, then $g(L) = (\lambda I - L)^{-1}$. However, if $\lambda \in \sigma(L)$, then this cannot be defined. This suggests that one should be able to define $g(L)$ whenever g is holomorphic on (a neighborhood of) $\sigma(L)$. Indeed, this can be done via Cauchy's integral formula.

Theorem 2.28 (Holomorphic Functional Calculus)**.** *There exists a unique continuous algebra homomorphism $g \mapsto g(L)$ from holomorphic functions on $\sigma(L)$ (with the compact-open topology) to bounded linear operators on $\mathcal{B}$ (with the norm topology), such that if $g(\zeta) = \zeta$ for all ζ in a neighborhood of $\sigma(L)$, then $g(L) = L$. Moreover, if γ is a curve in the domain of g which has winding number 1 around each point in $\sigma(L)$, then*

$$g(L) = \frac{1}{2\pi i} \int_\gamma g(\zeta)(\zeta I - L)^{-1} \, \mathrm{d}\zeta.$$

Here the contour integral may be interpreted as a Riemann integral where the Riemann sums converge in operator norm.

In large part, the utility of this calculus is due to the following theorem.

Theorem 2.29 (Spectral Mapping Theorem)**.** *Let g be holomorphic in a neighborhood of $\sigma(L)$. Then $\sigma(g(L)) = g(\sigma(L))$.*

Proofs of the previous two theorems may be found in Chapter 10 of [10].

Now we begin the proof of Theorem 2.21. One has $\mathcal{B} = \mathcal{B}_0 \oplus \mathbf{C}$, where $\mathcal{B}_0 = \{f \in \mathcal{B} : \int f \, \mathrm{d}\mu = 0\}$ and $\mathbf{C}$ is identified with the space of constant functions. By Proposition 2.7, these subspaces are $\mathcal{P}$-invariant. Moreover, $\mathcal{P}$ is the identity on $\mathbf{C}$, so we have the decomposition $\mathcal{P} = Q \oplus I$ where $Q = \mathcal{P}|_{\mathcal{B}_0}$. Thus $\sigma(P) = \sigma(Q) \cup \{1\}$. Let D_0 be the open disc of radius r centered at the origin in the complex plane, where r is as in (B6). Let D_1 be a disjoint open disc centered at 1, and set $D = D_0 \cup D_1$. Then $\sigma(\mathcal{P}) \subset D$. Since the condition of having a bounded inverse is open with respect to the operator norm (see [10]), $\sigma(\mathcal{P}_z) \subset D$ whenever z is sufficiently small.

Let $g = \mathbf{1}_{D_1}$. Then g is holomorphic on D, so we can define $\Pi_z = g(\mathcal{P}_z)$. Since $g^2 = g$, Π_z is a projection. Differentiating under the integral sign in Cauchy's integral formula, we see that Π_z is holomorphic in z.

Lemma 2.30. $\Pi_0 f = (\int f \, \mathrm{d}\mu)\mathbf{1}$ *for each* $f \in \mathcal{B}$.

Proof. Since $\sigma(Q) \subset \sigma(\mathcal{P})$, any function h which is holomorphic in a neighborhood of $\sigma(\mathcal{P})$ is holomorphic in a neighborhood of $\sigma(Q)$. Thus if we set $h(\mathcal{P}) := h(Q) \oplus I$, then $h \mapsto h(\mathcal{P})$ is a continuous algebra homomorphism in the sense of Theorem 2.28. It follows from the uniqueness of the holomorphic functional calculus that $\Pi_0 = g(\mathcal{P}) = g(Q) \oplus I$. But by (B6), $\sigma(Q) \subset D_0$. Thus $g = 0$ in a neighborhood of $\sigma(Q)$, and consequently $g(Q) = 0$. The proposition follows. $\square$

This result shows in particular that Π_0 has rank 1. This can be extended to Π_z.

Lemma 2.31. *Let P_1 and P_2 be bounded projections on a Banach space satisfying $\|P_1 - P_2\| < 1$. Then their images are isomorphic as vector spaces.*

A computational proof is given in [9].

Let $h(\zeta) = \zeta$. Then $\mathcal{P}_z\Pi_z = (hg)(\mathcal{P}_z) = \Pi_z P_z$, so the image of Π_z is $\mathcal{P}_z$-invariant. But by Lemma 2.31, the image of Π_z is one-dimensional, so there is a constant $\lambda(z)$ such that $\mathcal{P}_z\Pi_z = \lambda(z)\Pi_z$. Explicitly,

$$\lambda(z) = \frac{\int \mathcal{P}_z\Pi_z(\mathbf{1})\,\mathrm{d}\mu}{\int \Pi_z(\mathbf{1})\,\mathrm{d}\mu}.$$

The denominator is equal to 1 when $z = 0$, so by continuity it is nonzero for z sufficiently small. This shows that $\lambda(z)$ varies holomorphically in z. Note in particular that $\lambda(0) = 1$.

Finally, let $h_0(\zeta) = \zeta\mathbf{1}_{D_0}$ and $N_z = h_0(\mathcal{P}_z)$. Since $h(\zeta)g(\zeta) + h_0(\zeta) = \zeta$ and $h(\mathcal{P}_z)g(\mathcal{P}_z) = \mathcal{P}_z\Pi_z = \lambda_z\Pi_z$, we get that $\mathcal{P}_z = \lambda_z\Pi_z + N_z$. It remains to check the properties (D1)-(D4) in the statement of Theorem 2.21. (D1) and (D2) have already been discussed. (D3) follows from the fact that $h_0 g = g h_0 = 0$, and (D4) from the spectral mapping theorem.

3. Two Examples

In this section we use the methods of Section 2 to prove two instances of the CLT.

3.1. Bernoulli Shifts.

Let (X, d) be a compact metric space and μ a Borel probability measure on X. Form the countable product $X^{\mathbf{N}}$ of all sequences $\mathbf{x} = (x_1, x_2, \dots)$ of points in X, and equip $X^{\mathbf{N}}$ with the product measure $\mu^{\mathbf{N}}$ and the product metric

$$d^{\mathbf{N}}(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{\infty} 2^{-j} d(x_j, y_j).$$

The left shift $T(x_1, x_2, \dots) = (x_2, x_3, \dots)$ preserves $\mu^{\mathbf{N}}$, so $(X^{\mathbf{N}}, \mu^{\mathbf{N}}, T)$ is a measure preserving system.

If $\mathbf{x} = (x_1, x_2, \dots)$ and $x \in X$, denote

$$(x, \mathbf{x}) = (x, x_1, x_2, \dots).$$

Proposition 3.1. *The transfer operator $\mathcal{P}$ is the "averaged right shift"*

$$\mathcal{P}f(\mathbf{x}) = \int f(x, \mathbf{x})\,\mathrm{d}\mu(x).$$

Proof. We must check adjointness. Let $f, g \in L^2(X^{\mathbf{N}})$. Then

$$\begin{aligned}\int f(\mathbf{x})g(T(\mathbf{x}))\,\mathrm{d}\mu^{\mathbf{N}}(\mathbf{x}) &= \int f(x, \mathbf{x})g(\mathbf{x})\,\mathrm{d}\mu^{\mathbf{N}}(x, \mathbf{x}) \\ &= \int \left[\int f(x, \mathbf{x})\,\mathrm{d}\mu(x)\right] g(\mathbf{x})\,\mathrm{d}\mu^{\mathbf{N}}(\mathbf{x}),\end{aligned}$$

where the second equality comes from $\mu^{\mathbf{N}} = \mu \times \mu^{\mathbf{N}}$ and Fubini's theorem. Thus $\mathcal{P}f$ must be the expression in brackets. □

Definition 3.2. Fix $\alpha \in (0,1)$. A function $f : X^{\mathbf{N}} \to \mathbf{R}$ is *α-Hölder continuous* if the *α-Hölder seminorm*

$$|f|_\alpha = \sup_{\mathbf{x}\neq\mathbf{y}} \frac{|f(\mathbf{x}) - f(\mathbf{y})|}{d^{\mathbf{N}}(\mathbf{x},\mathbf{y})^\alpha}$$

is finite. Let $C^\alpha(X^{\mathbf{N}})$ denote the vector space of such functions. This becomes a Banach space with respect to the norm

$$\|f\|_\alpha := \left|\int f \, \mathrm{d}\mu^{\mathbf{N}}\right| + |f|_\alpha.$$

Proposition 3.3. *Let $f \in C^\alpha(X^{\mathbf{N}})$. Then $|\mathcal{P}f|_\alpha \leq 2^{-\alpha}|f|_\alpha$.*

Proof. Let $\mathbf{x},\mathbf{y} \in X^{\mathbf{N}}$ with $\mathbf{x} \neq \mathbf{y}$ arbitrary. Then

$$\begin{aligned}\frac{|\mathcal{P}f(\mathbf{x}) - \mathcal{P}f(\mathbf{y})|}{d^{\mathbf{N}}(\mathbf{x},\mathbf{y})^\alpha} &\leq \frac{1}{d^{\mathbf{N}}(\mathbf{x},\mathbf{y})^\alpha}\int |f(u,\mathbf{x}) - f(u,\mathbf{y})| \, \mathrm{d}\mu(u) \\ &= \frac{1}{2^\alpha}\int \frac{|f(u,\mathbf{x}) - f(u,\mathbf{y})|}{d^{\mathbf{N}}((u,\mathbf{x}),(u,\mathbf{y}))^\alpha} \, \mathrm{d}\mu \leq \frac{|f|_\alpha}{2^\alpha}.\end{aligned}$$

Taking the supremum over the left hand side gives the desired result. □

Here we see Remark 2.6 in action. Indeed, T is chaotic in the sense of the remark, and $\mathcal{P}$ correspondingly increases regularity (i.e. decreases the Hölder seminorm).

Theorem 3.4. *Let $\mathcal{B} = C^\alpha(X^{\mathbf{N}})$ and $\psi \in \mathcal{B}$ with $\int \psi \, \mathrm{d}\mu^{\mathbf{N}} = 0$. Then the conditions (B1)-(B6) from Section 2.1 hold.*

Proof. (B1)-(B3) are immediate. Let $f \in \mathcal{B}$ be arbitrary. Observe that

$$\|f\|_{L^\infty} \leq \operatorname{diam}(X^{\mathbf{N}})^\alpha \|f\|_\alpha.$$

Thus for any $\mathbf{x},\mathbf{y} \in X^{\mathbf{N}}$,

$$|\psi(\mathbf{x})f(\mathbf{x}) - \psi(\mathbf{y})f(\mathbf{y})| \leq \|\psi\|_{L^\infty}|f(\mathbf{x}) - f(\mathbf{y})| + \|f\|_{L^\infty}|\psi(\mathbf{x}) - \psi(\mathbf{y})|.$$

Consequently

$$\begin{aligned}\|\psi f\|_\alpha &= \left|\int \psi f \, \mathrm{d}\mu^{\mathbf{N}}\right| + |\psi f|_\alpha \leq \|\psi\|_{L^\infty}\|f\|_{L^\infty} + \|\psi\|_{L^\infty}|f|_\alpha + \|f\|_{L^\infty}|\psi|_\alpha \\ &\leq (\operatorname{diam}(X^{\mathbf{N}})^\alpha\|\psi\|_{L^\infty} + \|\psi\|_{L^\infty} + \operatorname{diam}(X^{\mathbf{N}})^\alpha|\psi|_\alpha)\|f\|_\alpha.\end{aligned}$$

This proves (B4). (B5) follows from Proposition 3.3. Finally, (B6) follows from Proposition 3.3 combined with the spectral radius formula. □

Corollary 3.5. *Let $\psi\colon X^{\mathbf{N}} \to \mathbf{R}$ be Hölder continuous with $\int \psi \, \mathrm{d}\mu^{\mathbf{N}} = 0$. Then ψ satisfies the CLT.*

3.2. Hyperbolic Toral Endomorphisms.

Let $\mathbf{T}^d = (\mathbf{R}/\mathbf{Z})^d$ denote the d-torus. Let μ be Haar measure on $\mathbf{T}^d$, normalized so that $\mu(\mathbf{T}^d) = 1$. If we identify $\mathbf{T}^d$ with $[0,1]^d$, then μ is d-dimensional Lebesgue measure.

Definition 3.6. A matrix is *hyperbolic* if it has no eigenvalues on the unit circle in the complex plane.

Let A be a hyperbolic $d \times d$ matrix with integer coefficients and nonzero determinant. Then A preserves $\mathbb{Z}^d$, so the map $T(x) = Ax$ is a well-defined endomorphism of $\mathbf{T}^d$.

Proposition 3.7. *$(\mathbf{T}^d, \mu, T)$ is a measure-preserving system.*

This is a standard result, and the proof presented below is essentially the same as the one in [5].

Proof. Let $f \in L^2$ and let $\widehat{f} : \mathbf{Z}^d \to \mathbf{C}$ be its Fourier transform. Then

$$f(T(x)) = \sum_{k \in \mathbf{Z}^d} \widehat{f}(k) e^{2\pi i k \cdot Ax} = \sum_{k \in \mathbf{Z}^d} \widehat{f}(k) e^{2\pi i A^t k \cdot x},$$

where the sum converges in L^2. Thus

$$\int f \circ T \, \mathrm{d}\mu = \sum_{\substack{k \in \mathbf{Z}^d \\ A^t k = 0}} \widehat{f}(k) = \widehat{f}(0) = \int f \, \mathrm{d}\mu,$$

as desired. Here the second equality follows from our assumption that A has nonzero determinant. □

Remark 3.8. For this system, as opposed to the Bernoulli shift in Section 3.1, we will have to carefully adapt our Banach space to the map T. We saw in Remark 2.6 that $\mathcal{P}$ should increase regularity as long as T pulls nearby points apart. But here A may have eigenvalues of modulus less than 1, in which case it pushes points along the corresponding eigenvector together. These directions are "bad" in the sense that $\mathcal{P}$ will make functions more irregular along them. Thus the norm on the Banach space we construct must measure regularity along the "good" directions (those which are eigenvectors with eigenvalues of modulus greater than 1) while simultaneously not penalizing irregularity along the bad directions.

The proof of Proposition 3.7 suggests that it may be fruitful to look for Banach spaces with some algebraic structure. Indeed, we will construct a space of trigonometric polynomials with a sort of weighted ℓ^2 norm. Our construction is inspired by [3].

Let $\mathcal{V}$ denote the complex vector space of trigonometric polynomials

$$f(x) = \sum_{k \in \mathbf{Z}^d} \widehat{f}(k) e^{2\pi i k \cdot x},$$

where $\widehat{f} : \mathbf{Z}^d \to \mathbf{C}$ is zero at all but finitely many points.

Proposition 3.9. *Let $f \in \mathcal{V}$. The transfer operator acts on f by*

$$\mathcal{P} f(x) = \sum_{m \in \mathbf{Z}^d} \widehat{f}(A^t m) e^{2\pi i m \cdot x}.$$

Proof. We need to check adjointness. Let $g \in L^2$ be arbitrary. Then

$$g(T(x)) = \sum_{k \in \mathbf{Z}} \widehat{g}(k) e^{2\pi i k \cdot Ax} = \sum_{k \in \mathbf{Z}} \widehat{g}(k) e^{2\pi i A^t k \cdot x}.$$

Thus

$$\int f (g \circ T) \, \mathrm{d}\mu = \sum_{k \in \mathbf{Z}^d} \widehat{f}(-A^t k) \widehat{g}(k) = \int \left[\sum_{m \in \mathbf{Z}^d} \widehat{f}(A^t m) e^{2\pi i m \cdot x} \right] g(x) \, \mathrm{d}\mu$$

by Parseval's theorem. □

Now place the norm

$$\|f\|_w^2 = \sum_{m\in\mathbf{Z}^d} |\widehat{f}(m)|^2 w(m)$$

on $\mathcal{V}$, where $w\colon \mathbf{Z}^d \to (0,\infty)$ is a weight which we will specify later.

Let $\mathcal{B}$ be the completion of $\mathcal{V}$ with respect to the norm $\|\cdot\|_w$.

At this point we need a technical lemma whose proof we defer to the end of this section.

Lemma 3.10. *There exists a choice of w with the following properties.*

(W1) There is a constant $c < 1$ such that $w(m) \le cw(A^t m)$ for all $m \in \mathbf{Z}^d \setminus \{0\}$.
(W2) $\log w$ is Lipschitz on $\mathbf{Z}^d$.

From now on, fix w satisfying (W1) and (W2) in the lemma. Let $\psi \in \mathcal{V}$ be a real-valued trigonometric polynomial (so $\widehat{\psi}(-k) = \overline{\widehat{\psi}(k)}$) with $\int \psi \, \mathrm{d}\mu = 0$.

Theorem 3.11. *$\mathcal{B}$ satisfies (B1)-(B6), where ψ is as above.*

Proof. (B1) and (B2) are immediate by construction. Since the integral of any trigonometric polynomial f is $\widehat{f}(0)$, integration is a bounded linear functional on $\mathcal{V}$, and thus extends to $\mathcal{B}$. So (B3) holds. For (B4), we may assume $\psi(x) = e^{2\pi i k\cdot x}$ for some $k \in \mathbf{Z}^d$ by linearity. If $f \in \mathcal{V}$, then $\widehat{\psi f}(m) = \widehat{f}(m-k)$. Thus

$$\|\psi f\|_w^2 = \sum_{m\in\mathbf{Z}^d} |\widehat{f}(m-k)|^2 w(m) = \sum_{m\in\mathbf{Z}^d} |\widehat{f}(m)|^2 w(m+k) = O_k(1)\|f\|_w^2,$$

where the last equality follows from (W2). This proves (B4). By Proposition 3.9, $\widehat{\mathcal{P}f}(m) = \widehat{f}(A^t m)$, so by (W1),

$$\begin{aligned}\|\mathcal{P}f\|_w^2 &= \sum_{m\in\mathbf{Z}^d} |\widehat{f}(A^t m)|^2 w(m) \\ &\le |\widehat{f}(0)|^2 w(0) + c \sum_{m\in\mathbf{Z}^d\setminus\{0\}} |\widehat{f}(A^t m)|^2 w(A^t m) \le \|f\|_w^2.\end{aligned}$$

Thus (B5) holds. If in particular $\int f \, \mathrm{d}\mu = 0$, i.e., $\widehat{f}(0) = 0$, then the above calculation shows $\|\mathcal{P}f\|_w^2 \le c\|f\|_w^2$. Thus by the spectral radius formula, (B6) holds. ☐

Corollary 3.12. *If ψ is a real-valued trigonometric polynomial as above, then ψ satisfies the CLT.*

Proof of Lemma 3.10. The idea is to give $m \in \mathbf{Z}^d$ a lot of weight if it lies along an eigenvector whose corresponding eigenvalue has modulus greater than 1, but very little weight if it lies along an eigenvector whose corresponding eigenvalue has modulus less than 1 (cf. Remark 3.8).

Viewing A^t as a complex matrix, we put A^t into Jordan canonical form. So we can write

$$\mathbf{C}^d = E_1 \oplus \cdots \oplus E_r \qquad \text{and} \qquad A^t = A_1 \oplus \cdots \oplus A_r,$$

where $A_i \colon E_i \to E_i$ and the matrix representation of A_i is

$$\begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}$$

in some basis of E_i. Scaling the jth vector in this basis by ε^j, we get a new basis in which A_i is represented by

$$\begin{pmatrix} \lambda_i & \varepsilon & & \\ & \lambda_i & \ddots & \\ & & \ddots & \varepsilon \\ & & & \lambda_i \end{pmatrix}.$$

Let B be the union of each of these new bases, so B is a basis of $\mathbf{C}^d$. Define a map $\sigma : B \to \{-1, 1\}$ as follows. If $e \in B$, then $e \in E_i$ for some unique i. Set $\sigma(e) = 1$ if $|\lambda_i| > 1$ and $\sigma(e) = -1$ if $|\lambda_i| < 1$ (by assumption $|\lambda_i| \neq 1$). Now given $m \in \mathbf{C}^d$, write $m = c_1 e_1 + \cdots + c_d e_d$ in the basis B, and let

$$w(m) = \exp\left(\sigma(e_1)|c_1| + \cdots + \sigma(e_d)|c_d|\right)$$

(we will restrict m to $\mathbf{Z}^d$ later). Pick ε above so that $||\lambda_i| - 1| > \varepsilon$ for each i. Then choose δ so that $||\lambda_i| - 1| - \varepsilon > \delta$ for each i. By construction, $\sigma(e)(|\lambda_i| - 1) = ||\lambda_i| - 1|$ if $e \in E_i$. It follows that

$$\frac{w(A^t m)}{w(m)} \geq \exp(\delta(|c_1| + \cdots + |c_d|)).$$

Since the map $m \mapsto |c_1| + \cdots + |c_d|$ is a norm on $\mathbf{C}^d$, and all norms on finite dimensional spaces are equivalent, there is some constant C such that $|c_1| + \cdots + |c_d| \geq C$ whenever $m \in \mathbf{Z}^d \setminus \{0\}$. Thus we can take $c = \exp(-\delta C)$ in (W1). For (W2), we use again the equivalence of norms on $\mathbf{C}^d$ to see that $\log w$ is Lipschitz. □

4. Spectral Gap versus Decay of Correlations

Our goal in this section is to give a precise characterization of when a Banach space satisfying (B1)-(B6) can be constructed in terms of how fast correlations between the $\psi \circ T^n$ decay. This in some sense makes Remark 1.8 meaningful.

We begin by constructing a notion of correlation for any number of random variables. Define $\mathcal{C}_k \colon (L^\infty)^k \to L^\infty$ recursively by $\mathcal{C}_1(f) = f$, and

$$\mathcal{C}_{k+1}(f_1, \ldots, f_{k+1}) = \left[\mathcal{C}_k(f_1, \ldots, f_k) - \int \mathcal{C}_k(f_1, \ldots, f_k)\, \mathrm{d}\mu\right] \left[f_{k+1} - \int f_{k+1}\, \mathrm{d}\mu\right].$$

Denote

$$\mathrm{cov}_k(f_1, \ldots, f_k) = \int \mathcal{C}_k(f_1, \ldots, f_k)\, \mathrm{d}\mu.$$

Remark 4.1. Observe that $\mathrm{cov}_2(f, g)$ agrees with the usual notion of covariance. However, cov_k is not a standard or particularly nice generalization of covariance. In fact, for $k > 2$, $\mathrm{cov}_k(f_1, \ldots, f_k)$ depends on the order of the f_i, in contrast with any intuitive notion of correlation. For our purposes though, cov_k is the right quantity to study.

Proposition 4.2. *The $\mathcal{C}_k$ and cov_k satisfy the following properties.*

(C1) Adding a constant to any of the $f_2, \dots, f_k$ (but not f_1) preserves $\mathcal{C}_k(f_1, \dots, f_k)$.
(C2) $\mathcal{C}_k$ is linear in each coordinate.
(C3) $\mathcal{C}_k(f_1, \dots, f_k) \circ T = \mathcal{C}_k(f_1 \circ T, \dots, f_k \circ T)$.
(C4) $\operatorname{cov}_{k+1}(f_1, \dots, f_k, g) = \operatorname{cov}_k(f_1, \dots, f_k g)$ whenever $\int f_k \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu = 0$.

Proof. (C1) and (C2) are evident from the definition. (C3) follows by induction and the fact that T preserves μ. (C4) is a computation. □

Theorem 4.3. *Fix a (real-valued) observable $\psi \in L^\infty(X)$ with $\int \psi \, \mathrm{d}\mu = 0$. Then there is a Banach space $\mathcal{B}$ satisfying (B1)-(B6) if and only if there exist constants $0 < r < 1 < R < \infty$ such that*

$$|\operatorname{cov}_{k+1}(\psi \circ T^{n_1}, \dots, \psi \circ T^{n_k}, \psi)| \le R^k r^{n_1} \tag{4.4}$$

whenever $k \in \mathbf{N}$ and $n_1 \ge \cdots \ge n_k \ge 0$.

Proof. ($\Rightarrow$) Suppose there is a Banach space $\mathcal{B}$ satisfying (B1)-(B6). We show by induction on k that if $f \in \mathcal{B} \cap L^\infty$, then

$$|\operatorname{cov}_{k+1}(\psi \circ T^{n_1}, \dots, \psi \circ T^{n_k}, f)| = O(R^k r^{n_1} \|f\|), \tag{4.5}$$

where $0 < r < 1 < R < \infty$ are constants we will choose later, and the asymptotic constant is absolute. In the base case $k = 0$,

$$|\operatorname{cov}_1(f)| = \left| \int f \, \mathrm{d}\mu \right| = O(\|f\|).$$

If $k > 0$, then by (C1) in Proposition 4.2 we may assume $\int f \, \mathrm{d}\mu = 0$. By (C3)-(C4),

$$\begin{aligned} |\operatorname{cov}_{k+1}(\psi \circ T^{n_1}, \dots, \psi \circ T^{n_k}, f)| &= \left| \int \mathcal{C}_{k+1}(\psi \circ T^{n_1}, \dots, \psi \circ T^{n_k}, f) \, \mathrm{d}\mu \right| \\ &= \left| \int \mathcal{C}_k(\psi \circ T^{n_1}, \dots, \psi \circ T^{n_k}) f \, \mathrm{d}\mu \right| \\ &= \left| \int \mathcal{C}_k(\psi \circ T^{n_1 - n_k}, \dots, \psi)(\mathcal{P}^{n_k} f) \, \mathrm{d}\mu \right| \\ &= |\operatorname{cov}_k(\psi \circ T^{n_1 - n_k}, \dots, \psi \circ T^{n_{k-1} - n_k}, \psi(\mathcal{P}^{n_k} f))|. \end{aligned}$$

By (B6) and the spectral radius formula, there are constants $C > 0$ and $r < 1$ such that the operator norm $\|(\mathcal{P}|_{\mathcal{B}_0})^n\| \le C r^n$. Then since $\int f \, \mathrm{d}\mu = 0$,

$$\|\psi(\mathcal{P}^{n_k} f)\| \le C r^{n_k} \|M_\psi\| \|f\|.$$

Thus if we take $R \ge C\|M_\psi\|$ (which doesn't depend on k), then we can conclude (4.5) by induction.

Finally, specializing (4.5) to the case $f = \psi$ and increasing R to account for the asymptotic constant, we obtain (4.4).

($\Leftarrow$) Suppose (4.4) holds for some $0 < r < 1 < R < \infty$. For $f \in L^\infty$, define

$$\|f\| = \left| \int f \, \mathrm{d}\mu \right| + \sup_{n_1 \ge \cdots \ge n_k} R^{-k} r^{-n_1} |\operatorname{cov}_{k+1}(\psi \circ T^{n_1}, \dots, \psi \circ T^{n_k}, f)|, \tag{4.6}$$

where the supremum is taken over all finite nonincreasing sequences of nonnegative integers. By (C2), $\| \cdot \|$ is a seminorm on the space $\mathcal{V}$ of functions $f \in L^\infty$ with $\|f\|$ finite. Observe that for constant functions f, the covariance term in (4.6) is zero, so we can write $\mathcal{V} = \mathbf{C} \oplus \mathcal{V}_0$, where $\mathcal{V}_0$ is the space of functions $f \in \mathcal{V}$ with $\int f \, \mathrm{d}\mu = 0$. If $f \in \mathbf{C}$ is a constant function, then $\mathcal{P}f = f$, and by (C2) and (4.4),

$$\|M_\psi f\| = \left| f \int \psi \, \mathrm{d}\mu \right| + \sup_{n_1 \geq \cdots \geq n_k} R^{-k} r^{-n_1} |f \operatorname{cov}_{k+1}(\psi \circ T^{n_1}, \ldots, \psi \circ T^{n_k}, \psi)| \leq \|f\|.$$

If $f \in \mathcal{V}_0$, then $\int \mathcal{P}f \, \mathrm{d}\mu = 0$, so by (C3),

$$\begin{aligned} \|\mathcal{P}f\| &= \sup_{n_1 \geq \cdots \geq n_k} R^{-k} r^{-n_1} |\operatorname{cov}_{k+1}(\psi \circ T^{n_1}, \ldots, \psi \circ T^{n_k}, \mathcal{P}f)| \\ &= \sup_{n_1 \geq \cdots \geq n_k} R^{-k} r^{-n_1} \left| \int \mathcal{C}_k(\psi \circ T^{n_1}, \ldots, \psi \circ T^{n_k}) \mathcal{P}f \, \mathrm{d}\mu \right| \\ &= \sup_{n_1 \geq \cdots \geq n_k} R^{-k} r^{-n_1} \left| \int \mathcal{C}_k(\psi \circ T^{n_1+1}, \ldots, \psi \circ T^{n_k+1}) f \, \mathrm{d}\mu \right| \\ &= r \sup_{n_1 \geq \cdots \geq n_k} R^{-k} r^{-n_1-1} \left| \operatorname{cov}_{k+1}(\psi \circ T^{n_1+1}, \ldots, \psi \circ T^{n_k+1}, f) \right| \leq r\|f\|. \end{aligned}$$

In addition, if $f \in \mathcal{V}_0$, then by (C4),

$$\begin{aligned} \|M_\psi f\| &= \left| \int \psi f \, \mathrm{d}\mu \right| + \sup_{n_1 \geq \cdots \geq n_k} R^{-k} r^{-n_1} |\operatorname{cov}_{k+1}(\psi \circ T^{n_1}, \ldots, \psi \circ T^{n_k}, \psi f)| \\ &= |\operatorname{cov}_2(\psi, f)| + R \sup_{n_1 \geq \cdots \geq n_k} R^{-k-1} r^{-n_1} |\operatorname{cov}_{k+2}(\psi \circ T^{n_1}, \ldots, \psi \circ T^{n_k}, \psi, f)| \\ &\leq 2R\|f\|. \end{aligned}$$

These computations demonstrate (B1)-(B5) with $\mathcal{V}$ in place of $\mathcal{B}$. Since $\|\cdot\|$ is only a seminorm and $\mathcal{V}$ is not complete with respect to $\|\cdot\|$, it doesn't make sense to speak of the spectrum of $\mathcal{P}$. However, if $\mathcal{V}$ were a Banach space, then the estimate $\|\mathcal{P}f\| \leq r\|f\|$ combined with the spectral radius formula would give (B6). With this in mind, we can quotient $\mathcal{V}$ to get a normed vector space and then complete it to get a Banach space $\mathcal{B}$. Since integration, M_ψ, and $\mathcal{P}$ are all bounded with respect to $\|\cdot\|$, they can be defined on $\mathcal{B}$ in the obvious way, and they will satisfy the same norm bounds. If we identify $f \in \mathcal{V}$ with its equivalence class in $\mathcal{B}$, then we can think of $\mathcal{B} \cap L^\infty$ as $\mathcal{V}$ (see Remark 2.16). Thus (B1)-(B6) hold. □

5. Higher Order Expansions

As usual, fix in this section a (real-valued) observable $\psi \in L^\infty$ with $\int \psi \, \mathrm{d}\mu = 0$.

In the previous section, we saw that the properties (B1)-(B6) tell us more than just decay of pairwise correlations, but rather encode information about correlations of arbitrarily high order. Similarly, we will see in this section that these properties tell us more than just the CLT. In fact, they often imply "higher order" central limit type theorems.

5.1. Computing the Spectral Decomposition.

Assume there is a Banach space satisfying (B1)-(B6), and write $\mathcal{P}_z = \lambda(z)\Pi_z + N_z$ as in Theorem 2.21. In the proof of the CLT in Section 2.1, we only used a second order approximation of $\lambda(z)$ and that $\int \Pi_0(\mathbf{1}) \, \mathrm{d}\mu = 1$. Whenever this decomposition exists, however, we can get more information by looking at more terms of the Taylor expansions of $\lambda(z)$ and $\int \Pi_z(\mathbf{1}) \, \mathrm{d}\mu$. Here we demonstrate how to compute all the derivatives of $\lambda(z)$ and $\int \Pi_z(\mathbf{1}) \, \mathrm{d}\mu$ at $z = 0$, and later we provide an application of this computation.

Write

$$\Pi_z = \frac{\mathcal{P}_z^n}{\lambda(z)^n} - \widetilde{N}_z^n,$$

where $\widetilde{N}_z = \lambda(z)^{-1} N_z$. For small enough z, $\lambda(z)$ is close to 1, so $\widetilde{N}_z^n$ has operator norm $O(r^n)$ for some $r < 1$. Thus from the Leibniz rule for differentiating products, we deduce the following.

Lemma 5.1. *There exists $r < 1$ such that*

$$\left\| \frac{\mathrm{d}^k}{\mathrm{d}z^k}\Big|_{z=0} \widetilde{N}_z^n \right\| = O_k(r^n).$$

Notation 5.2. Write $f(n) \approx g(n)$ if $f(n) - g(n)$ is bounded in norm by $O(r^n)$ for some $r < 1$ (where r and the implied constant may depend on anything except n).

By Lemma 5.1,

$$\frac{\mathrm{d}^k}{\mathrm{d}z^k}\Big|_{z=0} \Pi_z \approx \frac{\mathrm{d}^k}{\mathrm{d}z^k}\Big|_{z=0} \left[\frac{\mathcal{P}_z^n}{\lambda(z)^n}\right] = \sum_{k_1+k_2=k} \binom{k}{k_1} \left[(\lambda^{-n})^{(k_1)}(0)\right] \left[\frac{\mathrm{d}^{k_2}}{\mathrm{d}z^{k_2}}\Big|_{z=0} \mathcal{P}_z^n\right].$$

The difference quotients defining the derivatives in the equation above converge in norm, so by (B3) we can integrate both sides and swap the derivatives with the integral. This gives

$$\begin{aligned}
\frac{\mathrm{d}^k}{\mathrm{d}z^k}\Big|_{z=0} \int \Pi_z(\mathbf{1})\,\mathrm{d}\mu &\approx \sum_{k_1+k_2=k} \binom{k}{k_1} \left[(\lambda^{-n})^{(k_1)}(0)\right] \left[\frac{\mathrm{d}^{k_2}}{\mathrm{d}z^{k_2}}\Big|_{z=0} \int \mathcal{P}_z^n(\mathbf{1})\,\mathrm{d}\mu\right] \\
&= \sum_{k_1+k_2=k} \binom{k}{k_1} \left[(\lambda^{-n})^{(k_1)}(0)\right] \left[\frac{\mathrm{d}^{k_2}}{\mathrm{d}z^{k_2}}\Big|_{z=0} \int e^{iz\psi_n}\,\mathrm{d}\mu\right] \\
&= \sum_{k_1+k_2=k} \binom{k}{k_1} \left[(\lambda^{-n})^{(k_1)}(0)\right] \left[i^{k_2} \int \psi_n^{k_2}\,\mathrm{d}\mu\right].
\end{aligned}$$

Thus we can express the derivatives of $\int \Pi_z(\mathbf{1})\,\mathrm{d}\mu$ in terms of the derivatives of $\lambda(z)$. Now, notice that the left hand side is independent of n, so the difference between the right hand side for $n-1$ and for n is

$$\sum_{k_1+k_2=k} i^{k_2} \binom{k}{k_1} \left[(\lambda^{1-n})^{(k_1)}(0) \int \psi_{n-1}^{k_2}\,\mathrm{d}\mu - (\lambda^{-n})^{(k_1)}(0) \int \psi_n^{k_2}\,\mathrm{d}\mu\right] \approx 0. \tag{5.3}$$

The Leibniz rule states that

$$(\lambda^{1-n})^{(k_1)} = \sum_{j_1+j_2=k_1} \binom{k_1}{j_1} \lambda^{(j_1)} (\lambda^{-n})^{(j_2)},$$

and the binomial theorem says that

$$\psi_{n-1}^{k_2} = (\psi_n - \psi \circ T^{n-1})^{k_2} = \sum_{j_3+j_4=k_2} (-1)^{j_4} \binom{k_2}{j_3} \psi_n^{j_3} (\psi^{j_4} \circ T^{n-1}).$$

Plugging these equations into (5.3) and using the identity of multinomial coefficients

$$\binom{k}{j_1, j_2, j_3, j_4} := \frac{k!}{j_1! j_2! j_3! j_4!} = \binom{k}{k_1}\binom{k_1}{j_1}\binom{k_2}{j_3},$$

we obtain

$$\sum_{\substack{j_1+j_2+j_3+j_4=k \\ j_1+j_4\neq 0}} i^{j_3+j_4}\binom{k}{j_1,j_2,j_3,j_4}\lambda^{(j_1)}(0)(\lambda^{-n})^{(j_2)}(0)\int (-1)^{j_4}\psi_n^{j_3}(\psi^{j_4}\circ T^{n-1})\,\mathrm{d}\mu \approx 0.$$

The reason the condition $j_1 + j_4 \neq 0$ is imposed in the sum is that the terms $j_1 + j_4 = 0$ cancel with the second term in brackets in (5.3). Now, subtracting the $j_1 = k$ term from both sides above gives us that $-\lambda^{(k)}(0) \approx$ the expression below.

$$\sum_{\substack{j_1+j_2+j_3+j_4=k \\ j_1+j_4\neq 0 \\ j_1\neq k}} i^{j_3+3j_4}\binom{k}{j_1,j_2,j_3,j_4}\lambda^{(j_1)}(0)(\lambda^{-n})^{(j_2)}(0)\int \psi_n^{j_3}(\psi^{j_4}\circ T^{n-1})\,\mathrm{d}\mu \tag{5.4}$$

In this sum, one can never have $j_1 = k$ or $j_2 = k$. Thus this gives us a recurrence for $\lambda^{(k)}(0)$, at least up to an error which decays exponentially in n. If we want to use this recurrence to compute $\lambda^{(k)}(0)$, we need to make sure this error remains exponential when we plug in our approximations of $\lambda^{(j)}(0)$ for $j < k$. By Faà di Bruno's formula, $(\lambda^{-n})^{(j_2)}(0)$ can be written as a sum of $O_k(1)$ many monomials in the $\lambda^{(j)}(0)$, each of which grows polynomially with n. The L^∞ norm of the integrand in (5.4) also grows at most polynomially in n. Thus any exponential error we obtain by plugging in approximations of $\lambda^{(j)}(0)$ is multiplied by $n^{O_k(1)}$, and is therefore still exponential.

Example 5.5. Here we compute $\lambda^{(k)}(0)$ for $k = 1, 2, 3$.

$\underline{k=1}$: The only term in (5.4) corresponds to $(j_1, j_2, j_3, j_4) = (0, 0, 0, 1)$. Thus

$$\lambda'(0) \approx i\int \psi\circ T^{n-1}\,\mathrm{d}\mu = 0.$$

This implies that for general k, any term in (5.4) in which $j_1 = 1$ or $j_2 = 1$ vanishes. Similarly, any term in which $j_3 + j_4 = 1$ vanishes.

$\underline{k=2}$: The only nonzero terms in (5.4) correspond to $(j_1, j_2, j_3, j_4) = (0, 0, 1, 1)$ and $(0, 0, 0, 2)$. Thus

$$\begin{aligned}\lambda''(0) &\approx -2\int \psi_n(\psi\circ T^{n-1})\,\mathrm{d}\mu + \int \psi^2\circ T^{n-1}\,\mathrm{d}\mu \\ &\xrightarrow{n\to\infty} -\sum\int(\psi\circ T^i)(\psi\circ T^j)\,\mathrm{d}\mu,\end{aligned}$$

where the sum is over all pairs (i, j) with at least one of $i, j = 0$. If T is invertible, the sum may be written in the slightly neater form

$$\lambda''(0) = -\sum_{n\in\mathbf{Z}}\int \psi(\psi\circ T^n)\,\mathrm{d}\mu.$$

This is known as the *Green–Kubo formula*. With some additional algebra, this implies

$$\lambda''(0) = -\lim_{n\to\infty}\frac{1}{n}\int \psi_n^2\,\mathrm{d}\mu,$$

which is the formula given in Lemma 2.27. Recall that the variance $\sigma^2 = -\lambda''(0)$, so this form establishes the nonnegativity of the variance.

$\underline{k=3:}$ The only nonzero terms in (5.4) correspond to $(j_1, j_2, j_3, j_4) = (0,0,0,3)$, $(0,0,1,2)$, or $(0,0,2,1)$. Thus

$$\lambda'''(0) \approx -i \int \psi^3 \circ T^{n-1} \,\mathrm{d}\mu + 3i \int \psi_n(\psi^2 \circ T^{n-1}) \,\mathrm{d}\mu - 3i \int \psi_n^2(\psi \circ T^{n-1}) \,\mathrm{d}\mu.$$

The case $k = 4$ already has many nonzero terms, so we will not state the full formula for $\lambda^{(4)}(0)$ here. This is the first case where one has to recurse, because for instance $(j_1, j_2, j_3, j_4) = (2, 0, 2, 0)$ corresponds to the term

$$-6\lambda''(0) \int \psi_n^2 \,\mathrm{d}\mu.$$

We have a formula for $\lambda''(0)$ of the form $f(n) + O(r^n)$ with $r < 1$. By the discussion before the beginning of this example, we can write

$$-6[f(n) + O(r^n)] \int \psi_n^2 \,\mathrm{d}\mu = -6f(n) \int \psi_n^2 \,\mathrm{d}\mu + O(s^n)$$

for any $s > r$. Thus the error does not accumulate too much upon recursion.

5.2. **Edgeworth Series.** If the CLT corresponds to a second-order expansion of the $\lambda(z)$, it is natural to ask if higher order expansions of $\lambda(z)$ can give error terms on the CLT, and whether or not we can get formulas for the error terms similar to the Green–Kubo formula for the variance. Under the additional technical assumptions (B7) and (B8) below, [7] shows that these error terms exist, and the methods of the previous section give formulas analogous to Green–Kubo.

- (B7) For all $z \in \mathbf{R} \setminus \{0\}$, the spectrum of $\mathcal{P}_z$ (as an operator on $\mathcal{B}$) is contained in the open unit disc in the complex plane.
- (B8) For any fixed $r > 0$, there are constants K, N_0 large enough that $\|\mathcal{P}_z^N\| \le \frac{1}{N}$ for all $N > N_0$ and real z satisfying $K \le |z| \le N^r$.

Note that the perturbation-theoretic methods in Section 2 only give us information about $\mathcal{P}_z$ for small z, whereas (B7) and (B8) both depend on $\mathcal{P}_z$ for large z. Thus it takes some effort to check these conditions, and we will not give any examples here.

The theorem below is a slightly weaker version of Theorem 2.1 in [7].

Theorem 5.6. *If (B1)-(B8) hold, and $\sigma^2 = -\lambda''(0)$, then there exist polynomials $P_1, P_2, \ldots$ with real coefficients such that for any $k \in \mathbf{N}$,*

$$\mu\left\{x : \frac{1}{\sqrt{n}}\psi_n(x) \le t\right\} - \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^t e^{-y^2/2\sigma^2} \,\mathrm{d}y = \sum_{j=1}^k \frac{P_j(t)}{n^{j/2}} e^{-t^2/2\sigma^2} + o_k(n^{-k/2}),$$

where the implied constant does not depend on $t \in \mathbf{R}$.

This is known as an *Edgeworth expansion of order* $k+2$. When $k = 0$, this is the usual CLT.

In particular, the proof given in [7] constructs the polynomials P_j in terms of the derivatives $\lambda^{(k)}(0)$ and $\frac{\mathrm{d}^k}{\mathrm{d}z^k}|_{z=0} \int \Pi_z(\mathbf{1}) \,\mathrm{d}\mu$. Thus the results of Section 5.1 allow us to explicitly compute Edgeworth expansions of arbitrarily high order, and to do so very precisely (i.e., up to exponential error).

The reason that these derivatives get involved is that the proof relies critically on characteristic functions via the following fundamental estimate.

Theorem 5.7 (Berry–Esseen Inequality)**.** *Let $f\colon X \to \mathbf{R}$ be a real-valued random variable with mean 0 and characteristic function φ. Let $g\colon \mathbf{R} \to [-M, M]$ be an L^1 function with continuously differentiable Fourier transform γ, satisfying $\gamma(0) = 1$ and $\gamma'(0) = 0$. Then*

$$\left|\mu\{x : f(x) \leq t\} - \int_{-\infty}^{t} g(y)\,\mathrm{d}y\right| \leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\varphi(t)-\gamma(t)}{t}\right|\,\mathrm{d}t + O(1/T),$$

where the constant depends only on M (in fact we can take it to be $24M/\pi$).

This is Lemma XVI.3.2 in [6].

In particular, we are interested in the case where $f = \frac{1}{\sqrt{n}}\psi_n$ and $g(y)$ is $e^{-y^2/2\sigma^2}$ multiplied by a polynomial in y and $\sqrt{n}$.

Here we begin to see why the assumptions (B7) and (B8) are necessary. Indeed, the Taylor expansion from the previous section allows us to control the characteristic function $\varphi(t)$ for small t, but in the Berry–Esseen inequality we would like to take T large. Thus we need tail estimates on $\varphi(t)$, and these follow from (B7) and (B8).

For details, as well as examples of systems in which (B1)-(B8) hold, we refer the reader to [7].

Acknowledgments

First, I'm very grateful to my mentor, Jon DeWitt, for his many insightful comments, helpful suggestions, and his constant cheerful and supportive attitude. I'd like to thank Peter May for all the work he did so that this REU could be as productive as it was. Finally, I'd like to acknowledge my friends at the REU, who provided a fun and exciting environment in which to do mathematics.

References

[1] V. Baladi, *Positive Transfer Operators and Decay of Correlations.* Advanced Series in Nonlinear Dynamics. World Scientific Publishing Co., 2000.
[2] V. Baladi, *The Quest for the Ultimate Anisotropic Banach Space*, J. Stat. Phys. Vol. 166, p. 525-557, 2017.
[3] O. Bandtlow, W. Just, J. Slipantschuk, *Complete Spectral Data for Analytic Anosov Maps of the Torus*, Nonlinearity. Vol. 30, p. 2667–2686, 2017.
[4] R. Durrett, *Probability: Theory and Examples.* 5th ed. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2019.
[5] M. Einsiedler, T. Ward, *Ergodic Theory with a View Towards Number Theory.* Graduate Texts in Mathematics. London: Springer, 2011.
[6] W. Feller, *An Introduction to Probability Theory and Its Applications*, Volume 2. 2nd ed. Wiley Series in Probability and Statistics. New York: John Wiley & Sons, Inc, 1970.
[7] K. Fernando, C. Liverani, *Edgeworth Expansions for Weakly Dependent Random Variables.* 2017. arXiv:1803.07667.
[8] S. Gouëzel, *Limit Theorems in Dynamical Systems Using the Spectral Method.* Proceedings of Symposia in Pure Mathematics. Vol. 89, p. 161–193, 2015.
[9] T. Kato, *Perturbation Theory for Linear Operators.* Classics in Mathematics. Berlin: Springer, 1995.
[10] W. Rudin, *Functional Analysis.* International Series in Pure and Applied Mathematics. New York: McGraw-Hill, Inc., 1991.
[11] O. Sarig, *Introduction to the Transfer Operator Method.* 2012. http://www.weizmann.ac.il/math/sarigo/sites/math.sarigo/files/uploads/transferoperatorcourse.pdf.