

STATISTICAL PROPERTIES FOR CERTAIN DYNAMICAL SYSTEMS

CONTENTS

1. Introduction to the main topics of the course	1
1.1. Additive random processes	1
1.2. Multiplicative random processes	3
1.3. Observed dynamical systems	5
1.4. The moment method and Bernstein's trick	6
2. Stochastic dynamical systems	12
2.1. Strongly mixing Markov chains	12
2.2. Large deviations for strongly mixing Markov chains	15
2.3. Applications of the abstract LDT	19
2.4. Large deviations for random linear cocycles	22
2.5. Stationary measures	25
2.6. Conditional expectation	33
3. Limit laws for multiplicative random systems	33
4. Limit laws for hyperbolic systems	33
5. Partially hyperbolic systems	33

1. INTRODUCTION TO THE MAIN TOPICS OF THE COURSE

1.1. Additive random processes. Let $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ be a sequence of independent and identically distributed (i.i.d.) real random variables. Let

$$S_n := \xi_0 + \xi_1 + \dots + \xi_{n-1}$$

be the partial sum process and let

$$\frac{1}{n}S_n = \frac{1}{n} (\xi_0 + \xi_1 + \dots + \xi_{n-1})$$

be the average partial sum process.

Question. What is the behavior of these averages when $n \rightarrow \infty$?

Remark 1.1. Recall that two random variables ξ_1 and ξ_2 are identically distributed if $\mathbb{P}\{\xi_1 \in E\} = \mathbb{P}\{\xi_2 \in E\}$ for any Borel measurable set $E \subset \mathbb{R}$. In this case $\mathbb{E}\xi_1 = \mathbb{E}\xi_2$ and in fact $\mathbb{E}\phi(\xi_1) = \mathbb{E}\phi(\xi_2)$ for any integrable function $\phi: \mathbb{R} \rightarrow \mathbb{R}$.

Recall also that the random variables ξ_1, \dots, ξ_n are independent if for any Borel measurable sets $E_1, \dots, E_n \subset \mathbb{R}$,

$$\mathbb{P}\{\xi_1 \in E_1 \wedge \dots \wedge \xi_n \in E_n\} = \mathbb{P}\{\xi_1 \in E_1\} \cdots \mathbb{P}\{\xi_n \in E_n\}.$$

Theorem 1.1 (The law of large numbers - LLN). *Given i.i.d. sequence $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ of real random variables, if $\mathbb{E}\xi_0 < \infty$ then*

$$\frac{1}{n}S_n \rightarrow \mathbb{E}\xi_0 \quad a.s.$$

In particular, convergence in probability also holds. That is, $\forall \epsilon > 0$,

$$\mathbb{P}\left\{\left|\frac{1}{n}S_n - \mathbb{E}\xi_0\right| > \epsilon\right\} \rightarrow 0 \quad as \quad n \rightarrow \infty.$$

Question. It is natural to ask if there is a rate of convergence to 0 of the probability of the tail event above. It turns out that there is, as shown by the large deviations principle (LDP) below.

Theorem 1.2 (LDP of Cramér). *Assume that the common distribution of the i.i.d. sequence of real random variables $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ satisfies a certain growth condition and is non-trivial. Then $\forall \epsilon > 0$,*

$$\mathbb{P}\left\{\left|\frac{1}{n}S_n - \mathbb{E}\xi_0\right| > \epsilon\right\} \asymp e^{-c(\epsilon)n} \quad as \quad n \rightarrow \infty,$$

where $c(\epsilon) \asymp c_0\epsilon^2$ for some $c_0 > 0$.

More precisely, assuming that the common distribution has finite exponential moments:

$$M(t) := \mathbb{E}(e^{t\xi_0}) < \infty \quad \forall t \in \mathbb{R},$$

it follows that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left\{\left|\frac{1}{n}S_n - \mathbb{E}\xi_0\right| > \epsilon\right\} = -c(\epsilon)$$

where

$$c(\epsilon) = \sup_{t \in \mathbb{R}} (t\epsilon - \log M(t))$$

is the Legendre transform of $\log M(t)$.

This rate function $c(\epsilon)$ is strictly convex near $\epsilon = 0$, with $c(0) = 0$, $c'(0) = 0$ and $c''(0) > 0$, so that $c(\epsilon) \asymp c_0\epsilon^2$.

Remark 1.2. The LDP is a very precise but asymptotic result. We are usually more interested in *finitary*, albeit less precise results, which will be referred to as large deviations type (LDT) estimates. A typical such result is the following.

Theorem 1.3 (Hoeffding's Inequality). *Assume the much stronger growth condition $|\xi_0| \leq C$ a.s. Then $\forall \epsilon > 0$ the following holds for all $n \in \mathbb{N}$:*

$$\mathbb{P} \left\{ \left| \frac{1}{n} S_n - \mathbb{E} \xi_0 \right| > \epsilon \right\} \leq 2e^{-(2C)^{-2} \epsilon^2 n}.$$

Question. What is the typical size of the sum process $S_n - n\mathbb{E}\xi_0$? Note that by the LLN, almost surely we have

$$\frac{S_n - n\mathbb{E}\xi_0}{n} \rightarrow 0,$$

which implies that $S_n - n\mathbb{E}\xi_0 \ll n$. It turns out that from a certain point of view, $S_n - n\mathbb{E}\xi_0 \asymp \sqrt{n}$. More precisely, the following central limit theorem (CLT) holds.

Theorem 1.4 (CLT of Lindeberg-Lévy). *Consider an i.i.d. sequence $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ of real random variables and assume that the variance $\sigma^2 = \mathbb{E}\xi_0^2 - (\mathbb{E}\xi_0)^2 \in (0, \infty)$. Then for all $[a, b] \subset \mathbb{R}$,*

$$\mathbb{P} \left\{ \frac{S_n - n\mathbb{E}\xi_0}{\sigma\sqrt{n}} \in [a, b] \right\} \rightarrow \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \quad \text{as } n \rightarrow \infty.$$

In other words, with the appropriate scaling we have the convergence in distribution to the standard normal distribution

$$\frac{S_n - n\mathbb{E}\xi_0}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

1.2. Multiplicative random processes. Let μ be a probability measure on the group of matrices $\text{GL}_2(\mathbb{R})$. Given $g_0, g_1, \dots, g_{n-1}, g_n, \dots$ an i.i.d. sequence of random matrices chosen according to the probability μ , consider

$$\Pi_n := g_{n-1} \cdots g_1 g_0$$

the corresponding multiplicative process.

Recall that for a matrix $g \in \text{GL}_2(\mathbb{R})$, the norm is its maximal expansion

$$\|g\| = \max_{\|v\|=1} \|gv\|$$

while the co-norm is its minimal expansion

$$m(g) = \min_{\|v\|=1} \|gv\| = \|g^{-1}\|^{-1}.$$

The LLN for additive random processes has the following analog for multiplicative random processes.

Theorem 1.5 (Furstenberg-Kesten). *Assuming the integrability condition $\mathbb{E}(\log \|g\|) d\mu(g) < \infty$, there are two numbers $L^+(\mu) \geq L^-(\mu)$ called the maximal respectively the minimal Lyapunov exponents of μ such that*

$$\frac{1}{n} \log \|\Pi_n\| \rightarrow L^+(\mu), \quad a.s.$$

and

$$\frac{1}{n} \log \|\Pi_n^{-1}\|^{-1} \rightarrow L^-(\mu), \quad a.s.$$

In particular we also have convergence in probability: $\forall \epsilon > 0$,

$$\mathbb{P} \left\{ \left| \frac{1}{n} \log \|\Pi_n\| - L^+(\mu) \right| > \epsilon \right\} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Instead of the maximal (or minimal) expansion of the random matrix products, we may consider the expansion of any vector. That is, given $v \in \mathbb{R}^2$, $v \neq 0$ consider the random walk $\{g_{n-1} \cdots g_1 g_0 v : n \geq 0\}$.

Theorem 1.6 (Furstenberg-Kifer's non-random filtration). *For any given vector $v \in \mathbb{R}^2$, $v \neq 0$, either*

$$\frac{1}{n} \log \|\Pi_n v\| \rightarrow L^+(\mu) \quad \text{as } n \rightarrow \infty,$$

or

$$\frac{1}{n} \log \|\Pi_n v\| \rightarrow L^-(\mu) \quad \text{as } n \rightarrow \infty.$$

Remark 1.3. It turns out that under certain generic conditions to be defined in the future (namely the irreducibility of the measure μ), we have that $\forall v \in \mathbb{R}^2, v \neq 0$ the almost sure limit is the maximal Lyapunov exponent:

$$\frac{1}{n} \log \|\Pi_n v\| \rightarrow L^+(\mu) \quad a.s.$$

Moreover, if $L^+(\mu) > L^-(\mu)$ then

$$\mathbb{E} \left(\frac{1}{n} \log \|\Pi_n v\| \right) \rightarrow L^+(\mu)$$

uniformly in v .

Question. It is natural to ask if in this multiplicative random setting there are analogues of the LDP, LDT and CLT from the additive setting. As shown below, the answer is affirmative, at least in the generic setting. The precise statements will be provided later.

Theorem 1.7 (LDP - Le Page). *Under generic assumptions, if $L^+(\mu) > L^-(\mu)$, then $\forall v \in \mathbb{R}^2$, $v \neq 0$ and $\forall \epsilon > 0$,*

$$\mathbb{P} \left\{ \left| \frac{1}{n} \log \|\Pi_n\| - L^+(\mu) \right| > \epsilon \right\} \asymp e^{-c(\epsilon)n} \quad \text{as } n \rightarrow \infty.$$

Theorem 1.8 (LDT - Duarte, Klein). *Under generic assumptions, if $L^+(\mu) > L^-(\mu)$, then $\forall v \in \mathbb{R}^2$, $v \neq 0$, $\forall \epsilon > 0$ and $\forall n \in \mathbb{N}$,*

$$\mathbb{P} \left\{ \left| \frac{1}{n} \log \|\Pi_n\| - L^+(\mu) \right| > \epsilon \right\} \leq C e^{-c(\epsilon)n}$$

for some constant $C < \infty$ and $c(\epsilon) > 0$.

Theorem 1.9 (CLT - Le Page). *Under generic assumptions, there is $\sigma \in (0, \infty)$ such that $\forall v \in \mathbb{R}^2$, $v \neq 0$,*

$$\mathbb{P} \left\{ \frac{\log \|\Pi_n v\| - nL^+(\mu)}{\sigma \sqrt{n}} \in [a, b] \right\} \rightarrow \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \quad \text{as } n \rightarrow \infty.$$

1.3. Observed dynamical systems. Let (M, f) be a dynamical system where M is a compact metric space and $f : M \rightarrow M$ is continuous. Consider an appropriate f -invariant measure $\nu \in \text{Prob}(M)$.

Remark 1.4. Recall that ν is called f -invariant if $f_*\nu = \nu$, which is equivalent to saying that $\nu(E) = \nu(f^{-1}(E))$ for all Borel measurable $E \subset M$.

Moreover, ν is called ergodic w.r.t. f if all f -invariant sets (i.e. E such that $E = f^{-1}(E)$) are of ν measure 0 or 1. Note that ergodic measures are extremal points in the space of f -invariant measures (this space is convex and weak-* compact).

The triple (M, f, ν) is called a measure-preserving dynamical system (MPDS). Given an observable $\xi : M \rightarrow \mathbb{R}$ in an appropriate space of functions, the quadruple (M, f, ν, ξ) is called an observed MPDS.

For all iterates j , consider the real-valued random variable on M

$$\xi_j := \xi \circ f^j.$$

Since ν is f -invariant, and hence f^j -invariant for all j , the sequence $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ is identically distributed. However, in general this sequence is *not* independent.

Consider the sum process, that is, the Birkhoff sums

$$S_n \xi := \xi + \xi \circ f + \dots + \xi \circ f^{n-1} = \xi_0 + \xi_1 + \dots + \xi_{n-1}.$$

Birkhoff's ergodic theorem is a generalization of the LLN in this setting.

Theorem 1.10 (Birkhoff's ergodic theorem). *Assume that ν is ergodic w.r.t. f and that $\int_M |\xi| d\nu < \infty$. Then*

$$\frac{1}{n} S_n \xi \rightarrow \int \xi d\nu \quad \nu\text{-a.e.}$$

In particular the convergence in measure also holds: $\forall \epsilon > 0$,

$$\nu \left\{ x \in M : \left| \frac{1}{n} S_n \xi(x) - \int_M \xi d\nu \right| > \epsilon \right\} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Question. A fundamental problem in ergodic theory is to establish statistical properties like LDP, LDT, CLT for various kinds of observed dynamical systems.

In other words, the question is to determine for which dynamical system (M, f) , for which appropriate choice of f -invariant measures ν and for which kinds of observables ξ one has an LDT estimate

$$\nu \left\{ x \in M : \left| \frac{1}{n} S_n \xi(x) - \int_M \xi d\nu \right| > \epsilon \right\} \leq C e^{-c(\epsilon)n}$$

or a CLT

$$\nu \left\{ x \in M : \frac{S_n \xi(x) - n \int \xi d\nu}{\sigma \sqrt{n}} \in [a, b] \right\} \rightarrow \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}.$$

A short but vague answer is that systems with *some hyperbolicity* should satisfy such statistical properties. The question is extremely far reaching, and for now it only has a very incomplete rigorous answer.

Some of the main tools used to address it, which will make their entry in this course in due time, are the transition (or Markov) operator and the transfer (or Ruelle) operator.

1.4. The moment method and Bernstein's trick. [You decide where to put this aula 2.](#)

Let ξ be a random variable on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution (or law) of ξ is the probability measure μ_ξ on \mathbb{R} given by

$$\mu_\xi(E) = \mathbb{P} \{ \xi \in E \} = \mathbb{P} \{ \xi^{-1} E \}$$

where $E \subset \mathbb{R}$ is Borel measurable. In other words, $\mu_\xi = \xi_* \mathbb{P}$. Given a random variable ξ and $\mu \in \text{Prob}(\mathbb{R})$, we write $\xi \sim \mu$ when $\mu_\xi = \mu$.

Example 1. The continuous uniform distribution on some interval $[a, b] \subset \mathbb{R}$ is

$$\mu_{unif} = \frac{1}{b-a} \mathbb{1}_{[a,b]} dm$$

which is absolutely continuous to the Lebesgue measure m on \mathbb{R} .

Example 2. The standard normal distribution

$$\mathcal{N}(0, 1) = G(t)dm$$

where $G(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$ is the Gaussian.

Remark 1.5. The distribution of a random variable ξ determines its expectation, standard deviation, moments, etc. For example, its expectation satisfies

$$\mathbb{E}\xi = \int_{\mathbb{R}} x d\mu_{\xi}(x).$$

More generally, if $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is Lebesgue integrable, then

$$\mathbb{E}\varphi(\xi) = \int_{\mathbb{R}} \varphi(x) d\mu_{\xi}(x).$$

In fact, by the change of variables formula we have

$$\begin{aligned} \mathbb{E}\varphi(\xi) &= \int_{\Omega} \varphi(\xi(\omega)) d\mathbb{P}(\omega) \\ &= \int_{\mathbb{R}} \varphi(x) d\xi_*\mathbb{P}(x) \\ &= \int_{\mathbb{R}} \varphi(x) d\mu_{\xi}(x). \end{aligned}$$

We recall the meaning of random variables being identically distributed and independent in the following.

Definition 1.1. ξ_1 and ξ_2 are identically distributed if $\mu_{\xi_1} = \mu_{\xi_2}$. $\xi_1, \xi_2, \dots, \xi_n$ are independent if

$$\mu_{(\xi_1 \dots \xi_n)} = \mu_{\xi_1} \times \dots \times \mu_{\xi_n}.$$

Namely, the joint distribution is precisely the product measure.

From now on, let us fix some notations as follows.

ξ is the real random variable.

$\mu = \mathbb{E}\xi$ is the expectation of ξ .

$\sigma^2 = \mathbb{E}(\xi - \mu)^2 = \mathbb{E}\xi^2 - \mu^2 \in [0, \infty]$ is the variance of ξ .

$\mathbb{E}\xi^n$ is called the n -th moment of ξ . By the Hölder inequality we have $\mathbb{E}\xi \lesssim (\mathbb{E}\xi^2)^{\frac{1}{2}}$ and $\mathbb{E}\xi^2 \lesssim (\mathbb{E}\xi^4)^{\frac{1}{2}}$ etc. Note that working with even moments avoids negativity.

The following lemma is trivial but extremely useful throughout probability theory.

Lemma 1.1 (Markov's inequality). *If $X \geq 0$ and $\lambda > 0$ then*

$$\mathbb{P}\{X \geq \lambda\} \leq \frac{\mathbb{E}X}{\lambda}.$$

Proof. Denote $E = \{X \geq \lambda\}$, then we have $\mathbb{E}X \geq \int_E X d\mathbb{P} \geq \lambda\mathbb{P}(E)$. \square

We will use Markov's inequality to prove weak LLN and strong LLN respectively under some minor additional conditions.

Theorem 1.11 (Weak LLN). *Given i.i.d. sequence $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ of real random variables, if $\mathbb{E}\xi_0^2 < \infty$ then $\forall \epsilon > 0$,*

$$\mathbb{P} \left\{ \left| \frac{1}{n} S_n - \mathbb{E}\xi_0 \right| > \epsilon \right\} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Proof. Without loss of generality, we may assume that $\mu = 0$. Then it is enough to show $\mathbb{P} \left\{ \frac{S_n^2}{n^2} > \epsilon^2 \right\} = \mathbb{P} \{S_n^2 > n^2 \epsilon^2\} \rightarrow 0$ as $n \rightarrow \infty$.

By Markov's inequality, we have

$$\mathbb{P} \{S_n^2 > n^2 \epsilon^2\} \leq \frac{\mathbb{E}S_n^2}{n^2 \epsilon^2}.$$

Note that $S_n^2 = (\sum_{j=0}^{n-1} \xi_j)^2 = \sum_{j=0}^{n-1} \xi_j^2 + \sum_{j \neq k} \xi_j \xi_k$. Taking expectations on both sides, we obtain

$$\mathbb{E}S_n^2 = \sum_{j=0}^{n-1} \mathbb{E}\xi_j^2 + \sum_{j \neq k} \mathbb{E}(\xi_j \xi_k) = \sum_{j=0}^{n-1} \mathbb{E}\xi_j^2 = n\mathbb{E}\xi_0^2.$$

Here the second equality uses the independence of the random variables.

This shows that

$$\mathbb{P} \{S_n^2 > n^2 \epsilon^2\} \leq \frac{\mathbb{E}S_n^2}{n^2 \epsilon^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This finishes the proof of weak LLN. \square

Remark 1.6. If $X_n \rightarrow X$ a.s. then $X_n \rightarrow X$ in probability. In general, the converse is not true. However, if $\forall \epsilon > 0$ we have

$$\sum_{n=0}^{\infty} \mathbb{P} \{|X_n - X| > \epsilon\} < \infty,$$

then $X_n \rightarrow X$ a.s. This is ensured by Borel-Cantelli Lemma.

Theorem 1.12 (Strong LLN). *Given i.i.d. sequence $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ of real random variables, if $\mathbb{E}\xi_0^4 < \infty$ then $\frac{S_n}{n} \rightarrow \mu$ a.s.*

Proof. Without loss of generality, we may again assume $\mu = 0$. By Remark 1.6, it is enough to show $\mathbb{P} \{S_n^4 > n^4 \epsilon^4\} \leq \frac{c}{n^2}$ where c is a constant.

By Markov's inequality, we have

$$\mathbb{P} \{S_n^4 > n^4 \epsilon^4\} \leq \frac{\mathbb{E} S_n^4}{n^4 \epsilon^4}.$$

By direct computations and use the independence condition we get $\mathbb{E} S_n^4 = O(n^2)$. Therefore,

$$\mathbb{P} \{S_n^4 > n^4 \epsilon^4\} \lesssim \frac{1}{n^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

□

In the following, we are going to prove the following LDT estimates.

Theorem 1.13 (Cramér's inequality). *Assume that the common distribution of the i.i.d. sequence of real random variables $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ satisfies a certain growth condition and $\sigma^2 > 0$. Then $\forall \epsilon > 0$,*

$$\mathbb{P} \left\{ \left| \frac{1}{n} S_n - \mu \right| > \epsilon \right\} \leq 2e^{-\hat{C}(\epsilon)n} \quad \text{as } n \rightarrow \infty,$$

where $\hat{C}(\epsilon) \approx C_0 \epsilon^2 > 0$ with constant $C_0 > 0$.

We introduce the Bernstein's trick first.

Let X be a random variable and $\lambda \in \mathbb{R}$. Then

$$X \geq \lambda \Leftrightarrow e^{tX} \geq e^{t\lambda}, \forall t > 0.$$

By Markov's inequality,

$$\mathbb{P} \{X \geq \lambda\} = \mathbb{P} \{e^{tX} \geq e^{t\lambda}\} \leq \frac{\mathbb{E}(e^{tX})}{e^{t\lambda}},$$

which gives $\mathbb{P} \{X \geq \lambda\} \leq e^{-t\lambda} \mathbb{E}(e^{tX})$.

Definition 1.2. The function $M : \mathbb{R} \rightarrow (0, \infty)$ defined by $M(t) = \mathbb{E}(e^{tX})$ is called the moment generating function of X while $c(t) = \log M(t)$ is called the cumulant generating function of X .

Proof of Theorem 1.13. Without loss of generality, assume $\mu = 0$. Note that it is enough to estimate $\mathbb{P} \{S_n > n\epsilon\}$, the other part $\mathbb{P} \{-S_n > n\epsilon\}$ is the same. This is why the coefficient 2 appears in the r.h.s. of the inequality.

By Bernstein's trick, we have

$$\mathbb{P} \{S_n > n\epsilon\} \leq e^{-tn\epsilon} \mathbb{E}(e^{tS_n}).$$

Typically, $\mathbb{E}(e^{tS_n})$ can be exponentially large. But if we can prove something like

$$\mathbb{E}(e^{tS_n}) \leq e^{nLt^2},$$

then we would have

$$\mathbb{P}\{S_n > n\epsilon\} \leq e^{-nt\epsilon} e^{nLt^2} = e^{-n(t\epsilon - Lt^2)} = e^{-nc(\epsilon)}$$

It is easy to check that $c(\epsilon) = \frac{1}{4L}\epsilon^2$ is the maximum value of $t\epsilon - Lt^2$. Thus it is enough to estimate $\mathbb{E}(e^{tS_n})$.

Using the independence condition, we have

$$\mathbb{E}(e^{tS_n}) = \mathbb{E}(e^{t\xi_0}) \cdots \mathbb{E}(e^{t\xi_{n-1}}) = (\mathbb{E}(e^{t\xi_0}))^n = e^{nC_{\xi_0}(t)}.$$

Therefore,

$$\mathbb{P}\{S_n > n\epsilon\} \leq e^{-nt\epsilon} e^{nC_{\xi_0}(t)} = e^{-n(t\epsilon - C_{\xi_0}(t))}.$$

Let $\hat{C}_{\xi_0}(\epsilon) := \sup_{t \in \mathbb{R}} (t\epsilon - C_{\xi_0}(t))$. This is called the Legendre transform of $C_{\xi_0}(t)$. Thus we have

$$\mathbb{P}\{S_n > n\epsilon\} \leq e^{-n\hat{C}_{\xi_0}(\epsilon)}.$$

Since $\mu = 0$ and $\sigma^2 > 0$, it is straightforward to check that $C_{\xi_0}(t)$ satisfies $C_{\xi_0}(0) = 0$, $C'_{\xi_0}(0) = 0$ and $C''_{\xi_0}(0) = \sigma^2 > 0$. So $C_{\xi_0}(t) \approx \frac{\sigma^2}{2}t^2$ when $|t| \ll 1$. This gives us $\hat{C}_{\xi_0}(\epsilon) \approx C_0\epsilon^2 > 0$ with constant $C_0 > 0$.

This finishes the proof. \square

In the rest of this section, we are going to prove the CLT of Lindeberg-Lévy.

We first recall some definitions.

Definition 1.3 (Convergence in distribution). $X_n \xrightarrow{d} X$ if $\mu_{X_n} \rightarrow \mu_X$ in the weak* topology. More precisely, $\int_{\mathbb{R}} g d\mu_{X_n} \rightarrow \int_{\mathbb{R}} g d\mu_X$, $\forall g \in C_c(\mathbb{R})$.

Remark 1.7. Almost sure convergence implies convergence in probability, which further implies convergence in distribution. In general, the inverse directions are not true.

Definition 1.4. The cumulative distribution function (CDF) of a random variable X is

$$F_X(t) = \mathbb{P}(X \leq t), F_X : \mathbb{R} \rightarrow [0, 1]$$

which is non-decreasing. This implies that F_X is continuous almost everywhere.

We list a useful Proposition below without proof.

Proposition 1.2. $X_n \xrightarrow{d} X \Leftrightarrow F_{X_n}(t) \rightarrow F_X(t)$ for all t which is a continuous point of F_X .

For convenience, we recall the CLT below

Theorem 1.14 (CLT of Lindeberg-Lévy). *Consider an i.i.d. sequence $\xi_0, \xi_1, \dots, \xi_{n-1}, \xi_n, \dots$ of real random variables and assume that the variance $\sigma^2 = \mathbb{E}\xi_0^2 - (\mathbb{E}\xi_0)^2 \in (0, \infty)$. Then for all $[a, b] \subset \mathbb{R}$,*

$$\mathbb{P} \left\{ \frac{S_n - n\mathbb{E}\xi_0}{\sigma\sqrt{n}} \in [a, b] \right\} \rightarrow \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \quad \text{as } n \rightarrow \infty.$$

In other words, with the appropriate scaling we have the convergence in distribution to the standard normal distribution

$$\frac{S_n - n\mathbb{E}\xi_0}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Proof. Without loss of generality, we can assume that $\mu = 0$ and $\sigma^2 = 1$ so that we just need to prove

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

The proof follows from Lévy and we will use Fourier analysis.

Define the characteristic function of a random variable X by

$$\varphi_X : \mathbb{R} \rightarrow \mathbb{C}, \quad \varphi_X(t) = \mathbb{E}(e^{itX}) = \int_{\mathbb{R}} e^{itx} d\mu_X(x).$$

This is the Fourier transform of μ_X .

Recall that Lévy's continuity theorem says the following:

$$X_n \xrightarrow{d} X \iff \varphi_{X_n}(t) \rightarrow \varphi_X(t), \forall t \in \mathbb{R}.$$

This indicates the phenomenon that μ_{X_n} converges in the weak* topology if and only if its Fourier transform $\hat{\mu}_{X_n}$ converges for all t .

Moreover, we list some properties of $\varphi_X(t) = \mathbb{E}(e^{itX})$.

- $\varphi_X(0) = 1$,
- If X subjects to $\mathcal{N}(0, 1)$, then $\mu_X = G(t)dm$ and $\varphi_X(t) = \hat{G}(t) = e^{-\frac{t^2}{2}}$,
- $\varphi_{cX} = \varphi_X(ct)$,
- If X, Y are independent, then $\varphi_{X+Y} = \varphi_X \cdot \varphi_Y$.

By the properties, we get

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right) = \prod_{j=0}^{n-1} \varphi_{\xi_j}\left(\frac{t}{\sqrt{n}}\right) = [\varphi_{\xi_0}\left(\frac{t}{\sqrt{n}}\right)]^n.$$

By direct computations, we have $\varphi_{\xi_0}(0) = 1, \varphi'_{\xi_0}(0) = i\mu = 0$ and $\varphi''_{\xi_0}(0) = -\sigma^2 = -1$.

Therefore, by Taylor expansion we obtain

$$\varphi_{\xi_0}\left(\frac{t}{\sqrt{n}}\right) = e^{-(\frac{t}{\sqrt{n}})^2/2 + o((\frac{t}{\sqrt{n}})^3)}.$$

This proves that

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = e^{-\frac{t^2}{2} + o(\frac{t^3}{\sqrt{n}})},$$

which implies

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) \rightarrow e^{-\frac{t^2}{2}} \quad \text{as } n \rightarrow \infty.$$

The proof is finished by using Lévy's continuity theorem. \square

2. STOCHASTIC DYNAMICAL SYSTEMS

2.1. Strongly mixing Markov chains. To prove LDT and CLT for dynamical systems, we have to work with certain types of Markov chains (which are non-independent processes in general).

Example. LDT for multiplicative random processes.

Given $\mu \in \text{Prob}_c(\text{GL}_2(\mathbb{R}))$ and assume some generic condition, for an i.i.d. sequence $\{g_n\}_{n \geq 0}$ and $\Pi_n := g_{n-1} \cdots g_1 g_0$, we have that $\forall v \in \mathbb{R}^2, v \neq 0$

$$\mathbb{P} \left\{ \left| \frac{1}{n} \log \|\Pi_n v\| - L^+(\mu) \right| > \epsilon \right\} \leq C e^{-c(\epsilon)n}$$

for some $C < \infty$ and $c(\epsilon) > 0$.

In fact, we may relate the multiplicative process to a Markov chain in the following sense. For simplicity, let us try $n = 3$ first. For any $v \in \mathbb{S}^1$, we have

$$\frac{1}{3} \log \|g_2 g_1 g_0 v\| = \frac{1}{3} [\log \frac{\|g_2 g_1 g_0 v\|}{\|g_1 g_0 v\|} + \log \frac{\|g_1 g_0 v\|}{\|g_0 v\|} + \log \|g_0 v\|].$$

Denote $\Sigma = \text{supp}(\mu) \subset \text{GL}_2(\mathbb{R})$. Define $\varphi : \Sigma \times \mathbb{S}^1 \rightarrow \mathbb{R}$ by $\varphi(g, v) = \log \|gv\|$. Let $\omega \in \Omega = \Sigma^{\mathbb{N}}$ and $\omega = \{g_i\}_{i \geq 0}$. Define $Z_j^v : \Omega \rightarrow \Sigma \times \mathbb{S}^1$ by $Z_j^v(\omega) = (g_j, \frac{g_{j-1} \cdots g_0 v}{\|g_{j-1} \cdots g_0 v\|})$, $j \geq 1$ and $Z_0^v(\omega) = (g_0, v)$. Then we obtain

$$\frac{1}{3} \log \|g_2 g_1 g_0 v\| = \frac{1}{3} [\varphi(Z_2^v(\omega)) + \varphi(Z_1^v(\omega)) + \varphi(Z_0^v(\omega))].$$

In general,

$$\frac{1}{n} \log \|\Pi_n v\| = \frac{1}{n} \sum_{j=0}^{n-1} \varphi(Z_j^v(\omega)).$$

where $\varphi(g, v) = \log \|gv\|$ and $\{Z_n\}_{n \geq 0}$ is a Markov chain with values in $\Sigma \times \mathbb{S}^1$ and transition $(g_0, v) \rightarrow (g_1, \frac{g_0 v}{\|g_0 v\|})$ which is precisely the underlying fiber projective dynamics of the multiplicative random process.

Therefore, in order to prove LDT and CLT for multiplicative processes or other types of dynamical systems, we need to study appropriate Markov chains. Let us begin with a simple model.

Model: subshift of finite type.

Let $\Sigma = \{1, \dots, n\}$ be a finite space of symbols and let $P = \{p_{ij}\}_{1 \leq i, j \leq n}$ be a stochastic matrix. Namely,

$$\forall 1 \leq i \leq n, \sum_{j=1}^n p_{ij} = 1; \quad p_{ij} \geq 0, \forall 1 \leq i, j \leq n.$$

P can be seen as a transition matrix giving the transition probability from i to j by p_{ij} .

Let $q = (q_1, \dots, q_n)$ be a probability vector satisfying $q_i \geq 0, \forall 1 \leq i \leq n$ and $\sum_{i=1}^n q_i = 1$.

Definition 2.1. q is P -stationary if $qP = q$. That is

$$q_j = \sum_{i=1}^n q_i p_{ij}, \forall 1 \leq j \leq n.$$

Remark 2.1. Every stochastic matrix P has at least one stationary measure. Moreover, if P is primitive which means that $\exists m \in \mathbb{Z}^+$ such that $P_{ij}^m > 0, \forall 1 \leq i, j \leq n$, then $\exists!$ stationary vector q and $P_{ij}^n \rightarrow q_j$ exponentially fast for any $1 \leq i \leq n$.

In the following, we are going to define the Markov measure. Let us begin with some notations.

$$X^+ = \Sigma^{\mathbb{N}} = \{\{x_n\}_{n \geq 0} : x_n \in \Sigma\}.$$

\mathcal{B}^+ = σ -algebra generated by cylinders of the form

$$C[i_0, \dots, i_n] = \{x \in X^+ : x_0 = i_0, \dots, x_n = i_n\}.$$

Given q a probability vector and P a stochastic matrix, define

$$\mathbb{P}_{(q,P)}(C[i_0, \dots, i_n]) := q_{i_0} P_{i_0 i_1} \cdots P_{i_{n-1} i_n}.$$

This is a pre-measure. By Carathéodory's extension theorem, this pre-measure has a unique extension to a measure on \mathcal{B}^+ called Markov measure.

Let $\sigma : X^+ \rightarrow X^+$ be the forward shift. Note that if q is P -stationary, then $\mathbb{P}_{(q,P)}$ is σ -invariant. Therefore, $(X^+, \sigma, \mathbb{P}_{(q,P)})$ is an MPDS called a subshift of finite type. Moreover, if P is primitive, then $(X^+, \sigma, \mathbb{P}_{(q,P)})$ is exponentially mixing (hence ergodic).

A Markov chain with values in Σ is a sequence of random variables $\{Z_n\}_{n \geq 0}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $Z_n : \Omega \rightarrow \Sigma$ satisfying the Markov property.

$$\mathbb{P}\{Z_{n+1} = j | Z_n = i_n, \dots, Z_0 = i_0\} = \mathbb{P}\{Z_{n+1} = j | Z_n = i_n\}.$$

A Markov chain $\{Z_n\}_{n \geq 0}$ is said to have an initial distribution q and a transition P if

$$\begin{aligned}\mathbb{P}\{Z_0 = i\} &= q_i, \\ \mathbb{P}\{Z_{n+1} = j | Z_n = i\} &= p_{ij}.\end{aligned}$$

By Kolmogorov, there are such Markov chains on $(X^+, \mathcal{B}^+, \mathbb{P}_{(q,P)})$. In fact, the Markov chain $\{Z_j\}_{j \geq 0}$ is precisely the projection $Z_j : X^+ \rightarrow \Sigma$ defined by $Z_j(x) = x_j$ for any $j \geq 0$.

Now let us consider a more general setting.

Let (M, \mathcal{F}) be a measurable space. M is a compact metric space and \mathcal{F} is the Borel σ -algebra on M .

Definition 2.2 (Markov kernel). A Markov kernel $K(x, E)$ (which can be interpreted as the probability of x transitioning to E) is a function $K : M \times \mathcal{F} \rightarrow [0, 1]$ such that

- (1) $\forall x \in M, E \mapsto K_x(E)$ is a probability measure on \mathcal{F} ,
- (2) $\forall E \in \mathcal{F}, K(\cdot, E)$ is \mathcal{F} -measurable.

Remark 2.2. In practice, we may assume that $x \mapsto K_x = K(x, \cdot) \in \text{Prob}(M)$ is continuous which in particular implies (2). In other words, we can think of a Markov kernel as a continuous function $K : M \rightarrow \text{Prob}(M)$ where we interpret K_x as the probability of transitioning from x to somewhere.

Definition 2.3 (Stationary measure). $\mu \in \text{Prob}(M)$ is called K -stationary if $\mu = \int_M K_x d\mu(x)$ in the sense that $\mu(E) = \int_M K_x(E) d\mu(x), \forall E \in \mathcal{F}$.

Now we can define the Markov measure.

Given $\pi \in \text{Prob}(M)$ and K a Markov kernel, by Kolmogorov there exists a unique probability measure $\mathbb{P}_\pi = \mathbb{P}_{(\pi, K)}$ on $X^+ = M^\mathbb{N}$, $\mathcal{B}^+ = \sigma$ -algebra generated by the cylinders of the form:

$$C[A_0, \dots, A_n] = \{x = \{x_n\}_{n \geq 0} \in X^+ : x_j \in A_j, \forall 0 \leq j \leq n\},$$

where all $A_j \in \mathcal{F}$. It is easy to check that

$$\mathbb{P}_{(\pi, K)}(C[A_0, \dots, A_n]) = \int_{A_0} \int_{A_1} \dots \int_{A_n} 1 dK_{x_0}(x_1) \dots dK_{x_{n-1}}(x_n) d\pi(x_0).$$

Note that If $\varphi : X^+ \rightarrow \mathbb{R}$, then $\mathbb{E}_\pi(\varphi) = \int_{X^+} \varphi d\mathbb{P}_\pi$.

A Markov chain $\{Z_n : \Omega \rightarrow M\}$ is a sequence of random variables with values in M on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying the following Markov property.

$$\mathbb{P}\{Z_{n+1} \in E | Z_n, \dots, Z_0\} = \mathbb{P}\{Z_{n+1} \in E | Z_n\}.$$

The Markov chain $\{Z_n\}_{n \geq 0}$ is said to have initial distribution π and transition K if

$$\begin{aligned}\mathbb{P}\{Z_0 \in E\} &= \pi(E), \\ \mathbb{P}\{Z_{n+1} \in E | Z_n = x\} &= K_x(E).\end{aligned}$$

Example: $\Omega = X^+ = M^{\mathbb{N}}$, $\mathcal{F} = \mathcal{B}^+ = \sigma$ -algebra generated by cylinders. $\mathbb{P} = \mathbb{P}_{(\pi, K)}$, $Z_n : X^+ \rightarrow M$, $Z_n(x) = x_n \forall n \geq 0$.

Note that if $\pi = \delta_x$ then we write $\mathbb{P}_{\delta_x} := \mathbb{P}_x$ and $\mathbb{E}_{\delta_x} := \mathbb{E}_x$.

If μ is a K -stationary measure, then a (μ, K) Markov chain $\{Z_n\}_{n \geq 0}$ is stationary. More precisely,

$$\mathbb{P}\{Z_0 \in E_0, \dots, Z_n \in E_n\} = \mathbb{P}\{Z_j \in E_0, \dots, Z_{j+n} \in E_n\}, \forall j, n \geq 0$$

where $E_j \in \mathcal{F}$ is arbitrary. Moreover, the Markov shift (X^+, σ) is \mathbb{P}_μ -invariant so $(X^+, \sigma, \mathbb{P}_\mu)$ is an MPDS.

The powers of the Markov kernel can be derived inductively:

$$K^{n+1}(x, E) = \int_M K(y, E) dK_x^n(y).$$

Definition 2.4 (Markov system and strongly mixing). If μ is K -stationary, then (M, K, μ) is called a Markov system. It is strongly mixing if $K_x^n \rightarrow \mu$ exponentially fast $\forall x \in M$ in the weak* topology. More precisely, $\forall \varphi \in L^\infty(M)$,

$$\left\| \int_M \varphi(y) dK_x^n(y) - \int_M \varphi(y) d\mu(y) \right\|_\infty \leq C \rho^n \|\varphi\|_\infty \quad (2.1)$$

holds $\forall n \in \mathbb{N}$ where $C < \infty$ and $\rho \in (0, 1)$.

We may also consider the same concept from a different perspective, as we shall see below.

Definition 2.5 (Markov operator). Given a Markov system (M, K, μ) , the Markov operator $Q = Q_K : L^\infty(M) \rightarrow L^\infty(M)$ defined by

$$(Q\varphi)(x) = \int_M \varphi(y) dK_x(y).$$

The n -th iterates are

$$(Q^n \varphi)(x_0) = \int_M \dots \int_M \varphi(x_n) dK_{x_{n-1}}(x_n) \dots dK_{x_0}(x_1) = \int_M \varphi dK_{x_0}^n.$$

Therefore, (2.1) is equivalent to

$$\left\| (Q^n \varphi)(x) - \int_M \varphi d\mu \right\|_\infty \leq C \rho^n \|\varphi\|_\infty, \forall n \in \mathbb{N}.$$

2.2. Large deviations for strongly mixing Markov chains. We first recall some definitions. We begin with

Deterministic dynamical systems (DDS) (M, f) .

Let M be a metric space and let $f : M \rightarrow M$ be a continuous map. Once the initial state of the system $x_0 = x$ is fixed, then $x_n = f^n(x_0)$, $n \geq 0$ are all determined.

A probability measure $\mu \in \text{Prob}(M)$ is f -invariant if $f_*\mu = \mu$. Equivalently, $\int_M \delta_{f(x)} d\mu(x) = \mu$ or $\forall \varphi \in C_c(M), \int_M \varphi(f(x)) d\mu(x) = \int \varphi d\mu$. The triple (M, f, μ) is called a measure preserving dynamical system (as a convention, we omit the Borel σ -algebra on M).

A subset $E \subset M$ is called f -invariant if $f^{-1}(E) = E$. Equivalently, $x \in E \Leftrightarrow f(x) \in E$ or $x \in E \Leftrightarrow \delta_{f(x)}(E) = 1$. μ is f -ergodic if E is f -invariant $\Rightarrow \mu(E) = 0$ or 1 . Given an observable $\varphi : M \rightarrow \mathbb{R}$, then $\varphi(f^n(x))$ is the observed n -th state of the system which is to be considered.

Stochastic dynamical system (SDS) (M, K) .

Let M be a metric space and let $K : M \rightarrow \text{Prob}(M), x \rightarrow K_x$ be a continuous kernel. If $x_0 = x$ is the initial state of the system, the next state x_1 is not determined like in the DDS case by a transition law f . It is known only with a certain probability: $\mathbb{P}\{x_1 \in E\} = K_{x_0}(E)$. The iterates of K are $K_x^n = \int_M K_y^{n-1} dK_x(y)$.

$\mu \in \text{Prob}(M)$ is called K -stationary if $K * \mu = \mu$ in the sense that $\int_M K_x d\mu(x) = \mu$. The triple (M, K, μ) is a Markov system.

$E \subset M$ is K -invariant if $x \in E \Leftrightarrow K_x(E) = 1$. A K -stationary measure μ is ergodic if whenever E is K -invariant, we have $\mu(E) = 0$ or 1 . If $\varphi : M \rightarrow \mathbb{R}$ is an observable, we will consider

$$(Q^n \varphi)(x) = \int_M \varphi(y) dK_x^n(y) = \int_M \cdots \int_M \varphi(y) dK_{x_{n-1}}(y) \cdots dK_{x_0}(x_1).$$

Example 1. Any DDS (M, f) is itself an SDS. That is, $M \rightarrow \text{Prob}(M), x \rightarrow \delta_{f(x)}$.

Example 2. $\mu \in \text{Prob}_c(\text{GL}_2(\mathbb{R})), \Sigma = \text{supp}(\mu)$ and $\{g_n\}_{n \geq 0}$ is a sequence of i.i.d. matrices chosen with law μ . We may consider the kernel K on $\Sigma \times S^1$ as follows $K : \Sigma \times S^1 \rightarrow \text{Prob}(\Sigma \times S^1)$ such that $K_{(g_0, \hat{v})} = \mu \times \delta_{g_0 \hat{v}}$. Then $(\Sigma \times S^1, K)$ is an SDS.

Let us formally talk about the Kolmogorov extension. Let (M, K, μ) be a Markov system. Denote $X^+ = M^{\mathbb{N}} = \{x = \{x_n\}_{n \geq 0} : x_n \in M\}$. If $\pi \in \text{Prob}(M)$, then $\exists! \mathbb{P}_\pi \in \text{Prob}(X^+)$ s.t.

$$\mathbb{P}_\pi(C[E_0]) = \pi(E_0).$$

$$\mathbb{P}_\pi(C[E_0, E_1]) = \int_{E_0} \int_{E_1} 1 dK_{x_0}(x_1) d\pi(x_0).$$

If $f : X^+ \rightarrow \mathbb{R}$, then

$$\mathbb{E}_\pi(f) = \int_{X^+} f(x_0, \dots, x_n, \dots) dK_{x_0}(x_1) \cdots K_{x_{n-1}}(x_n) \cdots d\pi(x_0).$$

When $\pi = \delta_x$, we simply write \mathbb{E}_x and \mathbb{P}_x . When $\pi = \mu$ which is K -stationary, we write $\mathbb{E}_\mu = \mathbb{E}$ and $\mathbb{P}_\mu = \mathbb{P}$.

Things here are repeated, see Aula 3. You decide which to be deleted.

Recall that a Markov chain $\{Z_n : \Omega \rightarrow M\}$ is a sequence of random variables with values in M on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying the following Markov property.

$$\mathbb{P}\{Z_{n+1} \in E | Z_n, \dots, Z_0\} = \mathbb{P}\{Z_{n+1} \in E | Z_n\}.$$

The Markov chain $\{Z_n\}_{n \geq 0}$ is said to have initial distribution π and transition K if

$$\begin{aligned} \mathbb{P}\{Z_0 \in E\} &= \pi(E), \\ \mathbb{P}\{Z_{n+1} \in E | Z_n = x\} &= K_x(E). \end{aligned}$$

Example: $\Omega = X^+ = M^{\mathbb{N}}$, $\mathcal{F} = \mathcal{B}^+ = \sigma$ -algebra generated by cylinders. $\mathbb{P} = \mathbb{P}_{(\pi, K)}$, $Z_n : X^+ \rightarrow M$, $Z_n(x) = x_n \forall n \geq 0$.

Note that any K -stationary Markov chain can be realized as the example because $Z_n : \Omega \rightarrow M$ can always be written as $Z_n = e_n \circ Z$, when $Z(\omega) = \{Z_n(\omega)\}_n \in X^+$ and $e_n(\{Z_n(\omega)\}_{n \geq 0}) = Z_n(\omega)$ is the standard projection on to the n -th coordinates.

We have already defined the Markov operator in Definition 2.5. Here we consider Q defined on the space of continuous functions on M .

Example 1. Let (M, f) be a DDS, then $K_x = \delta_{f(x)}$. The corresponding Markov operator

$$(Q\varphi)(x) = \int \varphi(y) d\delta_{f(x)}(y) = \varphi(f(x)).$$

Namely $Q\varphi = \varphi \circ f$ is the Koopman operator.

Example 2. $K_{(g_0, \hat{v})} = \mu \times \delta_{\widehat{g_0 v}}$, then the corresponding Markov operator is

$$(Q\varphi)(g_0, \hat{v}) = \int \varphi(g_1, \widehat{g_0 v}) d\mu(g_1).$$

There are some basic properties of the Markov operator Q . Let us give two examples.

- (1) $Q1 = 1$,
- (2) $\|Q\varphi\|_{\infty} \leq \|\varphi\|_{\infty}$,
- (3) If $\varphi \geq 0$, then $Q\varphi \geq 0$.

The dual of Q , denoted by Q^* , acts on the space of probabilities $\text{Prob}(M)$. By definition, we have $\langle \varphi, Q^*\nu \rangle = \langle Q\varphi, \nu \rangle$ for any $\varphi \in C(M)$ and $\nu \in \text{Prob}(M)$. In other words, $Q^*\nu$ is the probability on M s.t.

$$\int_M \varphi dQ^*\nu = \int_M Q\varphi d\nu, \quad \forall \varphi \in C(M).$$

Note that μ is K -stationary $\Leftrightarrow Q^*\mu = \mu$.

In practice, the assumptions in Definition 2.5 is unreasonably strong, primarily because of $\varphi \in C(M)$. We are going to replace it by something weaker.

let $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$ be a Banach space where $\mathcal{E} \subset C^0(M)$ is Q -invariant in the sense that $\varphi \in \mathcal{E} \Leftrightarrow Q\varphi \in \mathcal{E}$. Moreover, we assume the constant function $1 \in \mathcal{E}$ and the inclusion of $\mathcal{E} \subset C^0(M)$ is continuous, namely $\|\varphi\|_{\infty} \leq C_1 \|\varphi\|_{\mathcal{E}}$ for some constant $C_1 < \infty$. We also assume that Q is bounded (or continuous) on $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$, i.e. $\|Q\varphi\|_{\mathcal{E}} \leq C_2 \|\varphi\|_{\mathcal{E}}$ with $C_2 < \infty$.

Definition 2.6 (Weaker version of strongly mixing). The system (M, K, μ, \mathcal{E}) is strongly mixing if $\forall n \geq 0$,

$$\left\| (Q^n \varphi)(x) - \int_M \varphi d\mu \right\|_{\infty} \leq C \|\varphi\|_{\mathcal{E}} r_n, \forall \varphi \in \mathcal{E}$$

for some $C < \infty$ and for some mixing rate $\{r_n\}_{n \geq 0}$ (e.g. $r_n = \rho^n$ with $\rho \in (0, 1)$ or $r_n = \frac{1}{(1+n)^p}$ with $p > 0$).

Let $\{Z_n\}_{n \geq 0}$ be the K -Markov chain, $Z_n : X^+ \rightarrow M, Z_n(x) = x_n$, for $\varphi : M \rightarrow \mathbb{R}$, we denote

$$S_n \varphi = \varphi(Z_0) + \cdots + \varphi(Z_{n-1}) := \varphi_0 + \cdots + \varphi_{n-1}.$$

Theorem 2.1 (Cai, Duarte, Klein 2022). *If (M, K, μ, \mathcal{E}) is a strongly mixing Markov system with mixing rate $r_n = \frac{1}{(1+n)^p}, p > 0$, then $\forall x \in M$ and $\forall \epsilon > 0$*

$$\mathbb{P}_x \left\{ \left| \frac{1}{n} S_n \varphi - \int_M \varphi d\mu \right| > \epsilon \right\} \leq e^{-c(\epsilon)n}$$

holds for all $n \geq 0$, for all $\varphi \in \mathcal{E}$ and for some $c(\epsilon) > 0$.

Note that \mathbb{P}_x can be replaced by \mathbb{P}_{μ} simply because $\mathbb{P}_{\mu} = \int_M \mathbb{P}_x d\mu(x)$. One should be careful with the notation x because it is both used as the element in M and the element in X^+ respectively.

Proof. Without loss of generality, we assume $\mathbb{E}\varphi = 0$, otherwise we consider $\varphi - \mathbb{E}\varphi$. Moreover, it is enough to consider $\mathbb{P}_x\{S_n \varphi \geq n\epsilon\}$. Using Bernstein's trick, for any $t > 0$ we have

$$\mathbb{P}_x\{S_n \varphi \geq n\epsilon\} = \mathbb{P}_x\{e^{tS_n \varphi} \geq e^{tn\epsilon}\} \leq e^{-tn\epsilon} \mathbb{E}_x(e^{tS_n \varphi}).$$

So our goal in the following is to estimate $\mathbb{E}_x(e^{tS_n \varphi})$ by relating it to $Q^{n_0}(\varphi)$ for some suitable choice of $n_0 \leq n$.

Note that

$$e^{tS_n \varphi} = \prod_{j=0}^{n-1} e^{t\varphi_j} := \prod_{j=0}^{n-1} f_j = f_0 \cdots f_{n-1}.$$

Take $n_0 \leq n$ such that $n = n_0 m + r$ and we first treat the case when $r = 0$. The key trick that we use is the following. We rewrite $f_0 \cdots f_{n-1}$ as

$$(f_0 f_{n_0} \cdots f_{(m-1)n_0})(f_1 f_{n_0+1} \cdots f_{(m-1)n_0+1}) \cdots (f_{n_0-1} f_{2n_0-1} \cdots f_{mn_0-1}).$$

We denote $F_j = f_j f_{n_0+j} \cdots f_{(m-1)n_0+j}$. By using Hölder inequality, we have

$$\mathbb{E}_x\left(\prod_{j=0}^{n-1} f_j\right) = \mathbb{E}_x(F_0 F_1 \cdots F_{n_0-1}) \leq \prod_{r=0}^{n_0-1} [\mathbb{E}_x(F_r^{n_0})]^{\frac{1}{n_0}}.$$

Thus, it is enough to estimate each $\mathbb{E}_x(F_r^{n_0})$.

Aula 5 is basically the complete proof of this theorem, you told me to leave to you so I will jump to Aula 6 \square

We also state an abstract CLT here, the proof will be given later.

Theorem 2.2 (Abstract CLT, Gordin-Livšic). *Let $\psi \in L^2(\mu)$, $\int \psi d\mu = 0$. Assume that*

$$\sum_{n=0}^{\infty} \|Q^n \psi\|_2 < \infty.$$

Then, denoting $\varphi = \sum_{n=0}^{\infty} Q^n \psi$, we have $\varphi \in L^2(\mu)$ and $\psi = \varphi - Q\varphi$. If $\sigma^2 = \|\varphi\|_2^2 - \|Q\varphi\|_2^2 > 0$, then the following CLT holds:

$$\frac{S_n \psi}{\sigma \sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Corollary 2.1. *If (M, K, μ) is strongly mixing on \mathcal{E} with mixing rate $r_n = \frac{1}{(1+n)^p}$, $p > 1$, then CLT holds for any observable $\psi \in \mathcal{E}$ provided $\sigma(\psi) > 0$.*

In the next subsections, we will introduce examples of dynamical systems that fit this abstract framework.

2.3. Applications of the abstract LDT. We will mainly study two certain skew-products.

Mixed random-quasiperiodic systems. Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the one dimensional torus. Assume $\alpha \notin \mathbb{Q}$, let $\tau_\alpha : \mathbb{T} \rightarrow \mathbb{T}$ be the torus translation by α such that $\tau_\alpha(\theta) = \theta + \alpha \pmod{1}$. Therefore, $(\mathbb{T}, \tau_\alpha, m)$ is an ergodic MPDS.

Remark 2.3. The (Markov) Koopman operator of this system is not strongly mixing, so the torus translation cannot be studied in this abstract framework. This is simply because $Q^n \varphi(\theta) = \varphi(\theta + n\alpha) \not\rightarrow \int \varphi d\mu$ as $n \rightarrow \infty$ for any non-constant $\varphi \in C^0(\mathbb{T})$.

Therefore, instead of torus translation, we are going to consider an iterated functions system (IFS) of rotations.

Let $\mu \in \text{Prob}(\mathbb{T})$, denote $\{\alpha_n\}_{n \geq 0}$ an i.i.d. sequence of translations with distribution μ . We consider the iterates

$$\theta \mapsto \theta + \alpha_0 \mapsto \theta + \alpha_0 + \alpha_1 \mapsto \cdots$$

Then given $\varphi : \mathbb{T} \rightarrow \mathbb{R}$ an observable, we may consider

$$Q\varphi(\theta) = \int \varphi(\theta + \alpha) d\mu(\alpha).$$

Obviously, the corresponding kernel $K : \mathbb{T} \rightarrow \text{Prob}(\mathbb{T})$ is $K_\theta = \mu * \delta_\theta$.

It turns out that the system (\mathbb{T}, K, m) is strongly mixing with a certain rate r_n having either polynomial or exponential decay, provided μ satisfies some general arithmetic properties (to be defined later) and φ is Hölder continuous. The proof will use some Fourier Analysis.

Note that the observable φ above only depends on one variable. In fact, we will consider a more complex system which allows φ to depend on infinite coordinates.

Regard $\Sigma := \mathbb{T}$ as the space of symbols with the measure μ . Let $X := \Sigma^{\mathbb{Z}}$ and consider the shift system $(X, \sigma, \mu^{\mathbb{Z}})$ where σ is the two sided Bernoulli shift. Then the skew product dynamical system is defined by

$$f : X \times \mathbb{T} \rightarrow X \times \mathbb{T}, f(\alpha, \theta) = (\sigma\alpha, \theta + \alpha_0).$$

The n -iterates are $f^n(\alpha, \theta) = (\sigma^n\alpha, \theta + \alpha_0 + \cdots + \alpha_{n-1})$.

The triple $X \times \mathbb{T}, f, \mu^{\mathbb{Z}} \times m$ is called a mixed random-quasiperiodic system. Under certain general assumptions on μ , it is ergodic and it satisfied LDT and CLT for certain types of observables.

Certain types of linear cocycles. Examples are Random, Markov, Fiber-bunched and Mixed cocycles.

We first recall the definition of linear cocycles. For more details, see [Viana] and [DK-CBM].

Let (X, f, μ) be an ergodic MPDS. A linear cocycle over (X, f, μ) is a skew-product map

$$F : X \times \mathbb{R}^2 \rightarrow X \times \mathbb{R}^2, F(x, v) = (f(x), A(x)v),$$

where $A : X \rightarrow \text{GL}_2(\mathbb{R})$ is a measurable function. We usually call f the base dynamics and A the fiber dynamics. We may also consider the projective cocycle.

$$\hat{F} : X \times \mathbb{P} \rightarrow X \times \mathbb{P}, \hat{F}(x, \hat{v}) = (f(x), \widehat{A(x)v}).$$

The n -th iterates of the cocycle are $F^n(x, v) = (f^n(x), A^n(x)v)$ where

$$A^n(x) = A(f^{n-1}(x)) \cdots A(f(x))A(x)$$

are called transfer matrices in Mathematical Physics.

We will always assume a mild integrability condition:

$$\int_X \log \|A(x)\| d\mu(x) < \infty.$$

Denote by $\varphi_x(x) := \log \|A^n(x)\|$, then the sequence $\{\varphi_n\}_{n \geq 0}$ is f -subadditive in the sense that

$$\varphi_{n+m} \leq \varphi_n \circ f^m + \varphi_m, \quad \forall m, n \in \mathbb{N}$$

By Kingman's subadditive ergodic theorem,

$$\frac{1}{n} \varphi \rightarrow L, \quad \mu\text{-a.e.}$$

That is

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|A^n(x)\| = L^+(A), \quad \mu\text{-a.e. } x \in X$$

where $L^+(A)$ is called the maximal Lyapunov exponent of A . Moreover, for μ -a.e. $x \in X$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|A^n(x)\| = \lim_{n \rightarrow \infty} \int \frac{1}{n} \log \|A^n\| d\mu = \inf_{n \geq 1} \int \frac{1}{n} \log \|A^n\| d\mu.$$

By a similar argument, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|A^n(x)^{-1}\|^{-1} = L^-(A), \quad \mu\text{-a.e. } x \in X$$

Note that $\forall g \in \text{GL}_2(\mathbb{R})$, $\|g^{-1}\|^{-1} \leq \|g\|$, so $L^-(A) \leq L^+(A)$.

We recall the Oseledets multiplicative ergodic theorem.

Theorem 2.3. *Let $F_A : X \times \mathbb{R}^2 \rightarrow X \times \mathbb{R}^2$ be a μ -integrable cocycle given by $A : X \rightarrow \text{GL}_2(\mathbb{R})$ over an ergodic MPDS (X, f, μ) , then*

(1) *If $L^+(A) = L^-(A)$, then $\forall v \in \mathbb{R}^2$ non-zero,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|A^n(x)v\| = L^+(A), \quad \mu\text{-a.e. } x \in X$$

(2) *If $L^+(A) > L^-(A)$, then there is a measurable map*

$$x \mapsto V_x \subset \mathbb{R}^2$$

where V_x is a one dimensional subspace of \mathbb{R}^2 , such that

$$A(x)V_x = V_{f(x)}$$

i.e. V_x is an F -invariant section. Moreover, if $v \notin V_x$, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|A^n(x)v\| = L^-(A), \quad \mu\text{-a.e. } x \in X.$$

Otherwise, if $v \in V_x$, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|A^n(x)v\| = L^+(A), \quad \mu\text{-a.e. } x \in X.$$

Moreover, if f is invertible then there exists a measurable splitting of the fiber: for μ -almost every $x \in X$, $\mathbb{R}^2 = E_x^+ \oplus E_x^-$ such that

- (1) $A(x)E_x^\pm = E_{f(x)}^\pm$.
- (2) $\lim_{n \rightarrow \infty} \frac{1}{n} \log \|A^n(x)v\| = L^\pm(A)$, $v \in E_x^\pm, v \neq 0$.
- (3) $\lim_{n \rightarrow \infty} \frac{1}{n} \log \left| \sin \angle(E_{f^n(x)}^+, E_{f^n(x)}^-) \right| = 0$

Examples of linear cocycles are quasi-periodic cocycles over a torus translation τ_α (which does not fit our framework) and random cocycles over a Bernoulli shift σ .

2.4. Large deviations for random linear cocycles. We begin with the definition of a random linear cocycle.

Setup. Let (Σ, μ) be a probability space. Denote $X := \Sigma^{\mathbb{Z}}$ and let σ be the two sided (Bernoulli) shift. $\mu^{\mathbb{Z}}$ is the product measure on the infinite product space X . The triple $(X, \sigma, \mu^{\mathbb{Z}})$ is called a Bernoulli shift. This is the base dynamics.

Let $A : X \rightarrow \text{GL}_2(\mathbb{R})$ be a continuous random cocycle. Moreover, assume that A is locally constant, namely, $A(\omega) = A(\omega_0)$ where $\omega = \{\omega_n\}_{n \in \mathbb{Z}}$. Given the Bernoulli shift, A determines a random linear cocycle

$$F = F_A : X \times \mathbb{R}^2 \rightarrow X \times \mathbb{R}^2, \quad F(\omega, v) = (\sigma\omega, A(\omega_0)v).$$

The n -th iterates of the cocycle are

$$F^n(\omega, v) = (\sigma^n\omega, A^n(\omega)v)$$

where $A^n(\omega) = A(\omega_{n-1}) \cdots A(\omega_1)A(\omega_0)$. As before, we may also consider the projective cocycle \hat{F} that is similarly defined.

We say that F satisfies a fiber LDT (or A satisfies an LDT) if $\forall v \neq 0, v \in \mathbb{R}^2, \forall \epsilon > 0$

$$\mu^{\mathbb{Z}} \left\{ \omega \in X : \left| \frac{1}{n} \log \|A^n(\omega)v\| - L^+(A) \right| > \epsilon \right\} < e^{-c(\epsilon)n}$$

for all $n \geq n(\epsilon, A)$ and for some $c(\epsilon) > 0$.

We will prove this LDT under certain “generic assumptions” on A and μ . Under the same assumption, we will also get a CLT:

$$\frac{\log \|A^n(\omega)v\| - nL^+(A)}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Remark 2.4. Note that since $A : \Sigma \rightarrow \text{GL}_2(\mathbb{R})$ is continuous, then $\nu = A_*\mu \in \text{Prob}_c(\text{GL}_2(\mathbb{R}))$. Therefore, we can start with a compactly supported probability measure ν in $\text{Prob}_c(\text{GL}_2(\mathbb{R}))$ and consider the multiplicative process associated to an i.i.d. sequence of random matrices $\{g_n\}_{n \in \mathbb{Z}}, g_n \in \text{GL}_2(\mathbb{R})$ with distribution ν . These two settings are essentially equivalent.

Generic assumptions. Let (Σ, μ) be a probability space, $A \in \Sigma \rightarrow \text{GL}_2(\mathbb{R})$.

Definition 2.7. A line $l \subset \mathbb{R}^2$ is A -invariant if $A(x)l = l$ for μ -a.e. $x \in \Sigma$.

Let \mathcal{H}_A be the group generated by the support of $A_*\mu$. Note that if l is A -invariant, then l is \mathcal{H}_A -invariant.

Definition 2.8. A cocycle A is called irreducible if there is no A -invariant line.

Definition 2.9. A cocycle A is called strongly irreducible if there is no finite union of lines which is A -invariant. Namely, $\forall n \in \mathbb{Z}^+$, there exist no lines $\{l_j\}_{1 \leq j \leq n}$ such that $A(x) \bigcup_{j=1}^n l_j = \bigcup_{j=1}^n l_j$ for a.e. $x \in \Sigma$.

Definition 2.10. A (or $A_*\mu$) is called non-compact if there exists a sequence of matrices $\{h_n\}_{n \geq 1} \subset \mathcal{H}_A$ such that $\|h_n\| \cdot \|h_n^{-1}\| \rightarrow \infty$.

We introduce a profound theorem of Furstenberg.

Theorem 2.4 (Furstenberg's Theorem). *If A is non-compact and strongly irreducible, then $L^+(A) > L^-(A)$.*

Example 1. Triangular matrices $A : \Sigma \rightarrow \text{GL}_2(\mathbb{R})$:

$$A(x) = \begin{pmatrix} a(x) & b(x) \\ 0 & c(x) \end{pmatrix}$$

is reducible because the line l of the direction $(1, 0)$ is A -invariant.

Example 2. Random Schrödinger cocycles. Let $\Sigma \subset \mathbb{R}$ be compact and $\mu \in \text{Prob}(\Sigma)$. Then $S : \Sigma \rightarrow \text{GL}_2(\mathbb{R})$ is defined by

$$S(a) = \begin{pmatrix} a & -1 \\ 1 & 0 \end{pmatrix} \in \text{SL}_2(\mathbb{R}).$$

Assume that $\#\text{supp}(\mu) \geq 2$ (μ is not a single Dirac), then S is non-compact and strongly irreducible. We leave the proof to the readers. Hint: play with $S(a)S(b)^{-1}$ and $S(a)^{-1}S(b)$, show that $(1, 0) \cup (0, 1)$ is the only potential candidate for S -invariance and prove it is actually not S -invariant. Thus strongly irreducible condition is fulfilled. As for non-compactness, take n -th power of either $S(a)S(b)^{-1}$ or $S(a)^{-1}S(b)$.

Let $l \subset \mathbb{R}^2$ be an A -invariant line, namely $A(x)l = l$ for μ -a.e. $x \in \Sigma$, then we can restrict A to l . We denote it by $A|_l$. Fix a unit vector $v \in l$, then $A(\omega)v = \lambda(\omega)v$ for some $\lambda : X \rightarrow \mathbb{R}$ which is in fact also locally constant. Let $A^n(\omega)v = A(\sigma^{n-1}\omega) \cdots A(\sigma\omega)A(\omega)v$, then by Birkhoff ergodic theorem

$$\frac{1}{n} \log \|A^n(\omega)v\| = \frac{1}{n} \sum_{j=0}^{n-1} \log |\lambda(\sigma^j\omega)| = \int_X \log |\lambda(\omega)| d\mu^{\mathbb{Z}}(\omega) = L(A|_l).$$

Definition 2.11. A is called quasi-irreducible if either there is no A -invariant line or $L(A|_l) = L^+(A)$.

We will prove the following theorem.

Theorem 2.5 (Le-Page, Duarte-Klein). *If A is quasi-irreducible and $L^+(A) > L^-(A)$, then A satisfies LDT: $\forall \epsilon > 0$*

$$\mu^{\mathbb{Z}} \left\{ \omega \in X : \left| \frac{1}{n} \log \|A^n(\omega)v\| - L^+(A) \right| > \epsilon \right\} < e^{-c(\epsilon)n}$$

holds $\forall v \neq 0, v \in \mathbb{R}^2, \forall n \geq n(\epsilon, A)$ and for some $c(\epsilon) > 0$.

General strategy for the proof. Consider the projective cocycle

$$\hat{F}_A : X \times \mathbb{P} \rightarrow X \times \mathbb{P}, \hat{F}_A(\omega, \hat{v}) = (\sigma\omega, \hat{A}(\omega_0)\hat{v}).$$

The corresponding Markov chain on $M := \Sigma \times \mathbb{P}$ is

$$(\omega_0, \hat{v}) \rightarrow (\omega_1, \hat{A}(\omega_0)\hat{v}) \rightarrow (\omega_2, \hat{A}(\omega_1)\hat{A}(\omega_0)\hat{v}) \rightarrow \cdots$$

where we denote $(\omega_n, \hat{A}(\omega_{n-1}) \cdots \hat{A}(\omega_0)\hat{v}) =: x_n$.

The associated SDS is

$$\bar{K} : \Sigma \times \mathbb{P} \rightarrow \text{Prob}(\Sigma \times \mathbb{P}), \bar{K}_{(\omega_0, \hat{v})} = \mu \times \delta_{\hat{A}(\omega_0)\hat{v}}.$$

This kernel \bar{K} defines a Markov operator

$$\bar{Q} : C^0(\Sigma \times \mathbb{P}) \rightarrow C^0(\Sigma \times \mathbb{P}),$$

$$\bar{Q}\varphi(\omega_0, \hat{v}) = \int_{\Sigma} \varphi(\omega_1, \hat{A}(\omega_0)\hat{v}) d\mu(\omega_1).$$

We will consider a special observable $\xi = \xi_A : \Sigma \times \mathbb{P} \rightarrow \mathbb{R}$ such that

$$\xi_A(\omega_0, \hat{v}) = \log \|A(\omega_0)v\|$$

where v is a unit representative of \hat{v} .

Recall that $m \in \text{Prob}(\Sigma \times \mathbb{P})$ is \bar{K} -stationary if and only if $\bar{Q}^*m = m$ where \bar{Q}^* is the dual of \bar{Q} . Then by Furstenberg's Formula, we have

$$L^+(A) = \max_{m \in \text{Prob}_{\bar{K}}(\Sigma \times \mathbb{P})} \left\{ \int_{\Sigma \times \mathbb{P}} \xi_A(\omega_0, \hat{v}) dm(\omega_0, \hat{v}) \right\}$$

Let $x = \{x_n\}_{n \geq 0} \in M^{\mathbb{N}}$, then if we start with an initial \hat{v} which is a unit vector. we have

$$S_n \xi_A(x) = \xi_A(x_0) + \xi_A(x_1) + \cdots + \xi_A(x_{n-1}) = \log \|A^n(\omega)v\|.$$

Thus

$$\frac{1}{n} S_n \xi_A(x) = \frac{1}{n} \log \|A^n(\omega)v\|. \quad (2.2)$$

Note that the l.h.s. will converge to $\int \xi_A dm = L^+(A)$ by Birkhoff, so the LDT follows.

2.5. Stationary measures. Equation (2.2) shows that in order to prove fiber-LDT for A , it would be enough to prove the corresponding Markov chain with observable ξ_A . For this purpose, it would be enough to show (because of the abstract LDT) that the Markov operator \bar{Q} is strongly mixing on some appropriate space $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$ which contains the special observable ξ_A .

A priori, this space will be $\mathcal{H}^{\alpha}(\Sigma \times \mathbb{P})$ of α -Hölder continuous functions in \mathbb{P} with some appropriate norm with some $\alpha > 0$. On this space, \bar{Q} will be shown to be quasi-compact and simple in which case $r_n = \sigma^n$ with $\sigma \in (0, 1)$. In fact, it will be convenient to work with a simpler kernel and the associated Markov operator.

Let $Q : C(\mathbb{P}) \rightarrow C(\mathbb{P})$ such that

$$Q\psi(\hat{v}) = \int \psi(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0).$$

Then Q is the Markov operator corresponding to the kernel $K : \mathbb{P} \rightarrow \text{Prob}(\mathbb{P})$:

$$K_{\hat{v}} = \int \delta_{\hat{A}(\omega_0)\hat{v}} d\mu(\omega_0).$$

Consider the projection $\Pi : C(\Sigma \times \mathbb{P}) \rightarrow C(\mathbb{P})$ defined by

$$\Pi\varphi(\hat{v}) = \int \varphi(\omega_0, \hat{v}) d\mu(\omega_0).$$

Lemma 2.2. *The following diagram is commutative.*

$$\begin{array}{ccc} C^0(\Sigma \times \mathbb{P}) & \xrightarrow{\bar{Q}} & C^0(\Sigma \times \mathbb{P}) \\ \Pi \downarrow & & \downarrow \Pi \\ C^0(\mathbb{P}) & \xrightarrow{Q} & C^0(\mathbb{P}) \end{array}$$

Namely, $\Pi \circ \bar{Q} = Q \circ \Pi$.

Proof. A simple calculation. □

Lemma 2.3. $\forall \varphi \in C^0(\Sigma \times \mathbb{P}), \forall n \geq 1$, we have

$$\bar{Q}^n \varphi(\omega_0, \hat{v}) = Q^{n-1}(\Pi \varphi)(\hat{A}(\omega_0) \hat{v}).$$

This shows that in order to prove that \bar{Q} is strongly mixing on \mathcal{E} , it is enough to show that Q is strongly mixing on $\Pi(\mathcal{E})$.

Proof.

$$\begin{aligned} \bar{Q}^n \varphi(\omega_0, \hat{v}) &= \int_{\Sigma^n} \varphi(\omega_n, \hat{A}^n(\omega) \hat{v}) d\mu(\omega_n) \cdots d\mu(\omega_1) \\ &= \int_{\Sigma^n} \varphi(\omega_n, \hat{A}(\omega_{n-1}) \cdots \hat{A}(\omega_0) \hat{v}) d\mu(\omega_n) \cdots d\mu(\omega_1). \end{aligned}$$

On the other hand, for any $\psi \in C(\mathbb{P})$ and $\hat{p} \in \mathbb{P}$, we have

$$Q^{n-1} \psi(\hat{p}) = \int_{\Sigma^{n-1}} \psi(\hat{A}(\omega_{n-1}) \cdots \hat{A}(\omega_1) \hat{p}) d\mu(\omega_{n-1}) \cdots d\mu(\omega_1). \quad (2.3)$$

If we take $\hat{p} = \hat{A}(\omega_0) \hat{v}$ and $\psi = \Pi \varphi$, then (2.3) equals to

$$\int_{\Sigma^{n-1}} \int_{\Sigma} \varphi(\omega_n, \hat{A}(\omega_{n-1}) \cdots \hat{A}(\omega_0) \hat{v}) d\mu(\omega_{n-1}) \cdots d\mu(\omega_1).$$

This finished the proof. \square

Recall that given a Markov kernel $K : M \rightarrow \text{Prob}(M)$, a measure $\eta \in \text{Prob}(M)$ is called K -stationary if $Q^* \eta = \eta$ where Q^* is the dual of the Markov operator Q associated with K . In this case, we will denote $\eta \in \text{Prob}_K(M)$. In fact, there are several equivalent definitions as follows:

- (1) $Q^* \eta = \eta$
- (2) $\forall \varphi \in C^0(M), \int_M Q \varphi d\eta = \int_M \varphi d\eta$.
- (3) $\forall \varphi \in C^0(M), \int_M \int_M \varphi(y) dK_x(y) d\eta(x) = \int_M \varphi d\eta$.
- (4) $K * \eta = \eta$ where $K * \eta = \int K_x d\eta(x)$.

Proposition 2.4. *Given $\eta \in \text{Prob}(\mathbb{P})$, the following are equivalent (TFAE):*

- (1) η is K -stationary.
- (2) $\mu \times \eta$ is \bar{K} -stationary.
- (3) $\mu^{\mathbb{N}} \times \eta$ is \hat{F}^+ -invariant where $\hat{F}^+ : X^+ \times \mathbb{P} \rightarrow X^+ \times \mathbb{P}$. Namely, $(X^+ \times \mathbb{P}, \hat{F}^+, \mu^{\mathbb{N}} \times \eta)$ is an MPDS.

Proof. We will first prove (1) \Leftrightarrow (2). It is enough to show that $K * \eta = \eta \Leftrightarrow \bar{K} * (\mu \times \eta) = \mu \times \eta$. In fact, we will show that

$$\bar{K} * (\mu \times \eta) = \mu \times (K * \eta). \quad (2.4)$$

This will conclude the proof because if $K * \eta = \eta$, then $\bar{K} * (\mu \times \eta) = \mu \times \eta$. If $\bar{K} * (\mu \times \eta) = \mu \times \eta$, then $\mu \times (K * \eta) = \mu \times \eta$ which gives $K * \eta = \eta$.

Note that (2.4) is equivalent to saying that: $\forall \varphi \in C^0(\Sigma \times \mathbb{P})$,

$$\int \varphi d[\bar{K} * (\mu \times \eta)] = \int \varphi d\mu d(K * \eta).$$

We first look at the left hand side. By the definition of convolution,

$$\bar{K} * (\mu \times \eta) = \int \bar{K}_{(\omega_0, \hat{v})} d\mu(\omega_0) d\eta(\hat{v}) = \int \mu \times \delta_{\hat{A}(\omega_0)\hat{v}} d\mu(\omega_0) d\eta(\hat{v}).$$

Thus we have

$$\int \varphi d[\bar{K} * (\mu \times \eta)] = \int \varphi(\omega_1, \hat{A}(\omega_0)\hat{v}) d\mu(\omega_1) d\mu(\omega_0) d\eta(\hat{v}).$$

Now let us focus on the r.h.s.

$$K * \eta = \int K_{\hat{v}} d\eta(\hat{v}) = \int \int \delta_{\hat{A}(\omega_0)\hat{v}} d\mu(\omega_0) d\eta(\hat{v}).$$

Thus we have

$$\int \varphi d\mu d(K * \eta) = \int \int \int \varphi(\omega_1, \hat{A}(\omega_0)\hat{v}) d\mu(\omega_1) d\mu(\omega_0) d\eta(\hat{v}).$$

This proves (1) \Leftrightarrow (2).

In the following, we are going to prove (1) \Leftrightarrow (3). Recall that $\mu^{\mathbb{N}} \times \eta$ is \hat{F}^+ -invariant if and only if $\forall \varphi \in C^0(X^+ \times \mathbb{P})$,

$$\int \varphi d\mu^{\mathbb{N}} \times \eta = \int \varphi \circ \hat{F}^+ d\mu^{\mathbb{N}} \times \eta.$$

More precisely,

$$\int \varphi(\omega, \hat{v}) d\mu^{\mathbb{N}}(\omega_0, \omega_1, \dots) d\eta(\hat{v}) = \int \varphi(\sigma\omega, \hat{A}(\omega_0)\hat{v}) d\mu^{\mathbb{N}}(\omega_1, \omega_2, \dots) d\eta(\hat{v}).$$

If we denote $\psi := \int \varphi(\omega, \hat{v}) d\mu^{\mathbb{N}}(\omega) \in C^0(\mathbb{P})$ which is arbitrary since φ is arbitrary, then the l.h.s. becomes $\int \psi d\eta$ and the r.h.s. becomes

$$\begin{aligned} \int \varphi(\sigma\omega, \hat{A}(\omega_0)\hat{v}) d\mu^{\mathbb{N}}(\omega_1, \omega_2, \dots) d\eta(\hat{v}) &= \int \int \psi(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0) d\eta(\hat{v}) \\ &= \int Q\psi(\hat{v}) d\eta(\hat{v}). \end{aligned}$$

Thus it is clear that (1) \Leftrightarrow (3). This finishes the proof. \square

Before we proceed, we recall some convex analysis concepts.

Let \mathfrak{X} be a topological vector space that is Hausdorff and locally convex. Given $D \subset \mathfrak{X}$, $p \in D$ is an extremal point of D if it is not

between any two different points in D . That is, there are no $x, y \in D$ with $x \neq y$ such that for some $t \in (0, 1)$,

$$p = tx + (1 - t)y.$$

Theorem 2.6 (Krein-Milman). *If $D \subset \mathfrak{X}$ is compacy, convex and non-empty, then D has at least one extremal point, i.e. $\text{extreme}(D) \neq \emptyset$. Moreover, the closed convex hull of $\text{extreme}(D)$ is D .*

Here the closed convex hull $\overline{\text{Co}}(S)$ is the smallest closed convex set containing S .

Example 1. Let M be a compact metric space. $D := \text{Prob}(M)$ with the weak* topology is compact convex and non-empty. So Krein-Milman applies. In this case, \mathfrak{X} is the space of signed measures which is metrizable with the weak* topology.

Example 2. Under the same settings as in Ex 1, let $D := \text{Prob}_K(M)$ which is closed. Thus D is compact, convex and non-empty. So Krein-Milman also applies.

Stationary measures, continuation.

To be more precise, let us rewrite the three levels of objects.

(1) DDS, projective linear cocycle:

$$\hat{F}^+ : X^+ \times \mathbb{P} \rightarrow X^+ \times \mathbb{P}, \quad \hat{F}^+(\omega, \hat{v}) = (\sigma\omega, \hat{A}(\omega_0)\hat{v}).$$

(2) SDS on $\Sigma \times \mathbb{P}$:

$$\bar{K}_{(\omega, \hat{v})} = \mu \times \delta_{\hat{A}(\omega_0)\hat{v}},$$

with the corresponding Markov operator \bar{Q} :

$$\bar{Q} : C^0(\Sigma \times \mathbb{P}) \rightarrow C^0(\Sigma \times \mathbb{P}),$$

$$\bar{Q}\varphi(\omega_0, \hat{v}) = \int_{\Sigma} \varphi(\omega_1, \hat{A}(\omega_0)\hat{v}) d\mu(\omega_1).$$

\bar{K} -Markov chain $\{Z_n\}_{n \geq 0}$ where $Z_n : X^+ \times \mathbb{P} \rightarrow \Sigma \times \mathbb{P}$ such that

$$Z_0(\omega, \hat{v}) = (\omega_0, \hat{v}), \quad Z_n(\omega, \hat{v}) = (\omega_n, \hat{A}^n(\omega)\hat{v})$$

with initial distribution $\mu \times \delta_{\hat{v}}$ (non-stationary case) or $\mu \times \eta$ (stationary case).

(3) SDS on \mathbb{P} :

$$K_{\hat{v}} = \int_{\Sigma} \delta_{\hat{A}(\omega_0)\hat{v}} d\mu(\omega_0),$$

with the corresponding Markov operator Q :

$$Q : C^0(\mathbb{P}) \rightarrow C^0(\mathbb{P}), \quad Q\varphi(\hat{v}) = \int_{\mathbb{P}} \varphi(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0).$$

We will mainly consider the special observable on $\Sigma \times \mathbb{P}$:

$$\bar{\xi} : \Sigma \times \mathbb{P} \rightarrow \mathbb{R}, \quad \bar{\xi}(\omega_0, \hat{v}) = \log \|A(\omega_0)v\|$$

where $v \in \hat{v}$ with $\|v\| = 1$.

The corresponding observable on \mathbb{P} is $\xi = \Pi \bar{\xi} : \mathbb{P} \rightarrow \mathbb{R}$ where $\Pi \varphi(\hat{v}) = \int_{\Sigma} \varphi(\omega_0, \hat{v}) d\mu(\omega_0)$ for any $\varphi \in C^0(\Sigma \times \mathbb{P})$.

The corresponding observable on $X^+ \times \mathbb{P}$ is

$$\Phi : X^+ \times \mathbb{P} \rightarrow \mathbb{R}, \quad \Phi(\omega, \hat{v}) = \bar{\xi}(\omega_0, \hat{v}).$$

Remark 2.5. We emphasize that all the places where $\varphi, \psi \in C^0$ above can be replaced by $\varphi, \psi \in L^\infty$ simply because we can define the Markov operator not only on the continuous function space, but also on the space of essentially bounded functions.

In the following, we will prove that η is an extremal point of $\text{Prob}_K(\mathbb{P})$ if and only if $\mu^{\mathbb{N}} \times \eta$ is \hat{F}^+ -invariant.

Definition 2.12. An observable $\varphi \in L^\infty(\mathbb{P})$ is called η -stationary if $Q\varphi(\hat{v}) = \varphi(\hat{v})$ for η -a.e. $\hat{v} \in \mathbb{P}$. A Borel set $E \subset \mathbb{P}$ is called η -stationary if $\mathbb{1}_E$ is η -stationary. Or equivalently, η -a.e. $\hat{v} \in E \Leftrightarrow \hat{A}(\omega_0)\hat{v} \in E$ for μ -a.e. $\omega_0 \in \Sigma$.

The equivalence statement is due to the condition:

$$Q\mathbb{1}_E(\hat{v}) = \int_{\Sigma} \mathbb{1}_E(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0) = \mathbb{1}_E(\hat{v}), \quad \eta\text{-a.e. } \hat{v} \in \mathbb{P}.$$

Proposition 2.5. *Let $\eta \in \text{Prob}_K(\mathbb{P})$, the following are equivalent:*

- (1) η is an extremal point of $\text{Prob}_K(\mathbb{P})$.
- (2) If $F \subset \mathbb{P}$ is η -stationary, then $\eta(F) = 0$ or 1.
- (3) If $\varphi \in L^\infty(\mathbb{P})$ is η -stationary, then $\varphi \equiv \text{const}$, η -a.e.
- (4) $(X^+ \times \mathbb{P}, \hat{F}^+, \mu^{\mathbb{N}} \times \eta)$ is an ergodic MPDS.

Proof. We prove by this order: (1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (1).

(1) \Rightarrow (2). Assume by contradiction the existence of $F \subset \mathbb{P}$ which is η -stationary with $t = \eta(F) \in (0, 1)$. The same holds for F^c . Namely, F^c is also η -stationary and $\eta(F^c) = 1 - t \in (0, 1)$.

Let

$$\eta_F \in \text{Prob}(\mathbb{P}), \quad \eta_F(E) = \frac{\eta(E \cap F)}{\eta(F)}$$

which is the conditional probability. Then by the Law of total probability,

$$\eta = t\eta_F + (1 - t)\eta_{F^c}.$$

Moreover, since $\eta_F(F) = 1$ and $\eta_{F^c}(F) = 0$, we have $\eta_F \neq \eta_{F^c}$.

If we can show that $\eta_F \in \text{Prob}_K(\mathbb{P})$ (then so does η_{F^c}), we will get a contradiction because η is assumed to be an extremal point of $\text{Prob}_K(\mathbb{P})$.

Let $\varphi \in L^\infty(\mathbb{P})$, direct computation shows

$$\begin{aligned}
\int_{\mathbb{P}} Q\varphi d\eta_F &= \frac{1}{\eta(F)} \int_F Q\varphi d\eta \\
&= \frac{1}{\eta(F)} \int_F \int_{\Sigma} \varphi(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}} \int_{\Sigma} \varphi(\hat{A}(\omega_0)\hat{v}) \mathbf{1}_F(\hat{v}) d\mu(\omega_0) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}} \int_{\Sigma} \varphi(\hat{A}(\omega_0)\hat{v}) \mathbf{1}_{F(\hat{A}(\omega_0)\hat{v})} d\mu(\omega_0) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}} \int_{\Sigma} (\varphi|_F)(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}} \int_{\Sigma} (\varphi|_F)(\hat{v}) d\mu(\omega_0) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}} Q(\varphi|_F)(\hat{v}) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_{\mathbb{P}} (\varphi|_F)(\hat{v}) d\eta(\hat{v}) \\
&= \frac{1}{\eta(F)} \int_F \varphi(\hat{v}) d\eta(\hat{v}) \\
&= \int_{\mathbb{P}} \varphi d\eta_F.
\end{aligned}$$

This proves that η_F is K -stationary, so is η_{F^c} . This contradicts that η is extremal, so (2) holds.

(2) \Rightarrow (3). Let $\varphi \in L^\infty(\mathbb{P})$ be η -stationary. We will use the following useful fact from measure theory.

Exercise. φ is constant η -a.e. iff the sub-level sets $\{\varphi > c\} = \{\hat{v} : \varphi(\hat{v}) < c\}$ have η measure either 1 or 0, $\forall c \in \mathbb{R}$.

Fix $c \in \mathbb{R}$, let $E = \{\hat{v} : \varphi(\hat{v}) < c\}$. We will show that $\mathbf{1}_E$ is η -stationary. Namely, E is η -stationary and by (2) we obtain that $\eta(E)$ is either 0 or 1. Since c is arbitrary, by the Exercise above, we get φ is constant η -a.e.

Let $\mathcal{S} := \{\varphi \in L^\infty(\mathbb{P}) : \varphi \text{ is } \eta\text{-stationary}\}$. We will show that $\mathbf{1}_E \in \mathcal{S}$. We list two properties of \mathcal{S} below:

- (1) \mathcal{S} is a linear space,
- (2) \mathcal{S} is a lattice.

Item (1) is obvious. For Item (2), being a lattice means

- $\phi \in \mathcal{S} \Rightarrow |\phi| \in \mathcal{S}$.
- If $\varphi, \psi \in \mathcal{S}$, then $\min\{\varphi, \psi\}, \max\{\varphi, \psi\} \in \mathcal{S}$.

We prove the first item. Since $\eta \in \text{Prob}_K(\mathbb{P})$, we have

$$\int_{\mathbb{P}} Q|\varphi| - |\varphi| d\eta = \int_{\mathbb{P}} Q|\varphi| d\eta - \int_{\mathbb{P}} |\varphi| d\eta = 0.$$

$\varphi \in \mathcal{S} \Rightarrow Q\varphi = \varphi$, η -a.e. This implies

$$\begin{aligned} |\varphi(\hat{v})| &= |Q\varphi(\hat{v})| \\ &= \left| \int \varphi(\hat{A}(\omega_0)\hat{v}) d\mu(\omega_0) \right| \\ &\leq \int \left| \varphi(\hat{A}(\omega_0)\hat{v}) \right| d\mu(\omega_0) \\ &= Q|\varphi|(\hat{v}). \end{aligned}$$

That is, $|\varphi| \leq Q|\varphi|$, η -a.e. Therefore, $|\varphi| = Q|\varphi|$, η -a.e. which shows $|\varphi| \in \mathcal{S}$.

The second item follows simply from the first item and the linearity because

$$\min\{\varphi, \psi\} = \frac{\varphi + \psi}{2} - \frac{|\varphi - \psi|}{2},$$

$$\max\{\varphi, \psi\} = \frac{\varphi + \psi}{2} + \frac{|\varphi - \psi|}{2}.$$

Now, let $\varphi_n(\hat{v}) = \min\{1, n \cdot \{c - \varphi(\hat{v}), 0\}\}$. Clearly, $\varphi_n \rightarrow \mathbf{1}_E$ as $n \rightarrow \infty$. Moreover, by the properties of \mathcal{S} and the definition of φ_n , we have $\varphi_n \in \mathcal{S}$. Thus $Q\varphi_n = \varphi_n$, η -a.e. Then we have $Q\varphi_n = \varphi_n \rightarrow \mathbf{1}_E$ and also $Q\varphi_n \rightarrow Q\mathbf{1}_E$, η -a.e. Finally, by the uniqueness of limit, we have $Q\mathbf{1}_E = \mathbf{1}_E$, η -a.e. which proves $\mathbf{1}_E \in \mathcal{S}$. This gives (3).

(3) \Rightarrow (4). To prove that $\mu^{\mathbb{N}} \times \eta$ is \hat{F}^+ -ergodic, it is equivalent to showing that if $\psi \in L^\infty(X^+ \times \mathbb{P})$ satisfies $\psi \circ \hat{F}^+ = \psi$, $\mu^{\mathbb{N}} \times \eta$ -a.e., then $\psi \equiv \text{const}$, $\mu^{\mathbb{N}} \times \eta$ -a.e.

Let $\varphi : \mathbb{P} \rightarrow \mathbb{R}$, $\varphi(\hat{v}) = \int_{X^+} \psi(\omega, \hat{v}) d\mu^{\mathbb{N}}(\omega)$. We will first show that $\varphi \equiv \text{const}$, η -a.e. For this, it is enough to show that φ is η -stationary

because of (3). By direct computation,

$$\begin{aligned}
Q\varphi(\hat{v}) &= \int_{\Sigma} \varphi(\hat{A}(\omega_0)\hat{v})d\mu(\omega_0) \\
&= \int_{\Sigma} \int_{X^+} \Psi(\omega', \hat{A}(\omega_0)\hat{v})d\mu^{\mathbb{N}}(\omega')d\mu(\omega_0) \\
&= \int_{X^+} \psi(\sigma\omega, \hat{A}(\omega_0)\hat{v})d\mu^{\mathbb{N}}(\omega) \\
&= \int_{X^+} \psi \circ \hat{F}^+(\omega, \hat{v})d\mu^{\mathbb{N}}(\omega) \\
&= \int_{X^+} \psi(\omega, \hat{v})d\mu^{\mathbb{N}}(\omega) \\
&= \varphi(\hat{v})
\end{aligned}$$

for η -a.e. $\hat{v} \in \mathbb{P}$.

It is left to show that ψ does not depend on $\omega = (\omega_0, \dots, \omega_{k-1}, \dots)$. Fix $k \geq 1$, it is enough to show ψ does not depend on $(\omega_0, \dots, \omega_{k-1})$. By assumption, we have

$$\psi = \psi \circ \hat{F}^+ = \dots = \psi \circ (\hat{F}^+)^k, \quad \mu^{\mathbb{N}} \times \eta\text{-a.e.}$$

Namely,

$$\psi(\omega, \hat{v}) = \psi(\sigma^k \omega, \hat{A}^k(\omega)\hat{v}).$$

Therefore, we have

$$\begin{aligned}
\int_{X^+} \psi(\omega, \hat{v})d\mu^{\mathbb{N}}(\omega_k, \omega_{k+1}, \dots) &= \int_{X^+} \psi(\sigma^k \omega, \hat{A}^k(\omega)\hat{v})d\mu^{\mathbb{N}}(\omega_k, \omega_{k+1}, \dots) \\
&= \varphi(\hat{A}^k(\omega)\hat{v}) \\
&= \text{const}
\end{aligned}$$

for η -a.e. $\hat{v} \in \mathbb{P}$.

So ψ is constant in $(\omega_0, \dots, \omega_{k-1})$, $\forall k \geq 1$. Thus ψ is constant in ω (one can also consider in terms of conditional expectation w.r.t. sub-algebras generated by cylinders). This proves (4).

(4) \Rightarrow (1). Assume by contradiction that η is not extremal, then $\exists t \in (0, 1), \eta_1 \neq \eta_2 \in \text{Prob}_K(\mathbb{P})$ such that $\eta = t\eta_1 + (1-t)\eta_2$. In particular,

$$\mu^{\mathbb{N}} \times \eta = t\mu^{\mathbb{N}} \times \eta_1 + (1-t)\mu^{\mathbb{N}} \times \eta_2.$$

Since $\mu^{\mathbb{N}} \times \eta_i, i = 1, 2$ are \hat{F}^+ -invariant, then $\mu^{\mathbb{N}} \times \eta$ is not ergodic because ergodic measures are extremal points of the space of invariant measures. This contradicts (4), thus (1) holds.

The proof is thus finished. □

As a corollary, we have

Corollary 2.6. *If $\eta \in \text{Prob}_K(\mathbb{P})$ is extremal, then*

$$\frac{1}{n} \log \|A^n(\omega)v\| \rightarrow \int_{\mathbb{P}} \int_{\Sigma} \log \|A(\omega_0)\hat{v}\| d\mu(\omega_0)d\eta(\hat{v}), \text{ as } n \rightarrow \infty$$

for $\mu^{\mathbb{N}} \times \eta$ -a.e. (ω, \hat{v}) , where $v \in \hat{v}$ with $\|v\| = 1$.

2.6. Conditional expectation. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\xi \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and denote $\mathbb{E}\xi = \int_{\Omega} \xi d\mathbb{P}$. Take $\mathcal{F}_0 \subset \mathcal{F}$ a sub- σ -algebra.

$\mathbb{E}(\xi|\mathcal{F}_0)$ is the conditional expectation of ξ w.r.t. \mathcal{F}_0 . Intuitively, it is the best prediction of ξ given the information \mathcal{F}_0 . Formally, we have

Definition 2.13. $\mathbb{E}(\xi|\mathcal{F}_0)$ is the “unique” random variable $\tilde{\xi} : \Omega \rightarrow \mathbb{R}$ such that

- (1) $\tilde{\xi}$ is \mathcal{F}_0 -measurable.
- (2) $\forall E \in \mathcal{F}_0, \int_E \tilde{\xi} d\mathbb{P} = \int_E \xi d\mathbb{P}$.

The existence and uniqueness of $\mathbb{E}(\xi|\mathcal{F}_0)$ are given by Lebesgue-Radon-Nykodim. More precisely, consider the map

$$\mathcal{F}_0 \ni E \mapsto \int_E \xi d\mathbb{P} \in \mathbb{R},$$

which is a signed measure. We denote it by ν . Moreover, it is clear that $\nu \ll \mathbb{P}|_{\mathcal{F}_0}$ since if $\mathbb{P}(E) = 0, E \in \mathcal{F}_0$ then $\nu(E) = \int_E \xi d\mathbb{P} = 0$. Then by Lebesgue-Radon-Nykodim, $\exists \tilde{\xi} \in L^1(\Omega, \mathcal{F}_0, \mathbb{P})$ s.t. $\frac{d\nu}{d\mathbb{P}|_{\mathcal{F}_0}} = \tilde{\xi}$, which gives (2).

3. LIMIT LAWS FOR MULTIPLICATIVE RANDOM SYSTEMS

4. LIMIT LAWS FOR HYPERBOLIC SYSTEMS

5. PARTIALLY HYPERBOLIC SYSTEMS