Central Limit Theorems in Ergodic Theory

Attila Herczegh

Thesis for M.Sc. degree in Mathematics

Advisor: Vilmos Prokaj
Eötvös Loránd University

Faculty of Science,
Eötvös Loránd University, Budapest

Budapest, 2009.

Contents

Acknowledgements

I would like to express my gratitude to my advisor in Budapest, Vilmos Prokaj, and my former supervisor, Karma Dajani, for all the advice, support and encouragement they gave me throughout the time I spent working on this thesis. It has been a pleasure to work with both of them.

I also thank for my universities, the Eötvös Loránd University in Budapest and the VU University Amsterdam in Amsterdam, to make everything possible.

Chapter 1

Introduction

1.1 Basic definitions and examples

First of all, let us start with a little introduction about ergodic theory. Ergodic theory is the study of the long term average behaviour of systems evolving in time. As a starting point we have a probability space $(X, \mathcal{B}, \mu)$. The collection of all states of the system form a space X and the evolution, in our case, is represented by a measurable transformation $T\colon X \to X$, so that $T^{-1}A \in \mathcal{B}$ for all $A \in \mathcal{B}$. In most cases we would like our transformation to be measure preserving and ergodic, i.e.:

Definition 1.1. *Let $(X, \mathcal{B}, \mu)$ be a probability space and $T\colon X \to X$ measurable. The map T is said to be measure preserving with respect to μ if $\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{B}$. We call the quadruple $(X, \mathcal{B}, \mu, T)$ a measure preserving or dynamical system.*

Definition 1.2. *Let T be a measure preserving trasformation on a probability space $(X, \mathcal{B}, \mu)$. T is called ergodic with respect to μ if for every measurable set A satisfying $A = T^{-1}A$, we have $\mu(A) = 0$ or 1.*

Example 1.1. *Consider $([0,1), \mathcal{B}, \mu)$, where $\mathcal{B}$ is the Borel σ-algebra and μ is the Gauss measure given by the density $\frac{1}{\log 2}\frac{1}{1+x}$ with respect to the Lebesgue measure. Let T be the Gauss transformation given by $T(x) = \frac{1}{x} \pmod 1$. It is well known that T preserves the Gauss measure. Moreover, T is ergodic with respect to μ.*

Example 1.2. *Consider $([0,1), \mathcal{B}, \lambda)$, where $\mathcal{B}$ is the Lebesgue σ-algebra and λ is the Lebesgue-measure. Let $T\colon [0,1) \to [0,1)$ be given by $Tx = rx \bmod 1$, where r is a positive integer. Then T is ergodic with respect to λ.*

Example 1.3. *Consider* $([0,1),\mathcal{B},\lambda)$ *as above. Let* $\beta > 1$ *a non-integer, and consider* $T_\beta\colon [0,1) \to [0,1)$ *given by* $T_\beta x = \beta x \bmod 1$. *Then* T_β *is ergodic with respect to* λ, *i.e. if* $T_\beta^{-1}A = A$, *then* $\lambda(A) = 0$ *or* 1.

The following lemma provides, for example in the cases mentioned above, a useful tool to verify ergodicity of a measure preserving transformation defined on $([0,1),\mathcal{B},\mu)$, where μ is equivalent to the Lebesgue measure.

Lemma 1.1. *(Knopp's lemma) If* B *is a Lebesgue-set and* $\mathcal{C}$ *is a class of subintervals of* $[0,1)$ *satisfying*

(a) every open subintervals of $[0,1)$ *is at most a countable union of disjoint elements from* $\mathcal{C}$,

(b) $\forall A \in \mathcal{C}$, $\lambda(A \cap B) \geq \gamma\lambda(A)$, *where* $\gamma > 0$ *is independent of* A,

then $\lambda(B) = 1$.

An important theorem is the Ergodic Theorem also known as Birkhoff's Ergodic Theorem, which is in fact a generalization of the Strong Law of Large Numbers. The theorem goes as follows:

Theorem 1.1. *(Ergodic Theorem) Let* $(X,\mathcal{B},\mu)$ *be a probability space and let* $T\colon X \to X$ *be a measure preserving transformation. Then* $\forall f \in L^1(X,\mathcal{B},\mu)$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) = f^*(x)$$

exists μ*-a.e.* $x \in X$, *is* T*-invariant and* $\int_X f d\mu = \int_X f^* d\mu$. *If moreover* T *is ergodic, then* f^* *is a constant a.e. and* $f^* = \int_X f d\mu$.

This is widely used theorem, for example, it is crucial in the proof of the Shannon-McMillan-Breiman theorem which we will state later. Using the Ergodic Theorem one can give another characterization of ergodicity:

Corollary 1.1. *Let* $(X,\mathcal{B},\mu)$ *be a probability space and* $T\colon X \to X$ *a measure preserving transformation. Then* T *is ergodic if and only if for all* $A, B \in \mathcal{B}$ *one has*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mu\left(T^{-i}B \cap A\right) = \mu(A)\mu(B).$$

This corollary gives a new definition for ergodicity, namely, the asymptotic average independence. We can define other notions like this which are stronger than ergodicity, called mixing properties:

Definition 1.3. *Let* $(X,\mathcal{B},\mu)$ *be a probability space and* $T\colon X \to X$ *a measure preserving transformation. Then,*

(i) T *is weakly mixing if for all* $A, B \in \mathcal{B}$ *one has*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left|\mu\left(T^{-i}B \cap A\right) - \mu(A)\mu(B)\right| = 0.$$

(ii) T is strongly mixing if for all $A, B \in \mathcal{B}$ one has

$$\lim_{n\to\infty} \mu\left(T^{-i}B \cap A\right) = \mu(A)\mu(B).$$

It is not hard to see that strongly mixing implies weakly mixing and weakly mixing implies ergodicity. This follows from the simple fact that if $\{a_n\}$ such that $\lim_{n\to\infty} a_n = 0$ then $\lim_{n\to\infty} \frac{1}{n}\sum |a_n| = 0$ and hence $\lim_{n\to\infty} \frac{1}{n}\sum a_n = 0$. The following proposition helps us check whether a transformation is mixing:

Proposition 1.1. *Let $(X, \mathcal{B}, \mu)$ be a probability space and $T\colon X \to X$ a measure preserving transformation. Let $\mathcal{C}$ be a generating semi-algebra of $\mathcal{B}$. Then,*

(a) If for all $A, B \in \mathcal{C}$ one has

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left|\mu\left(T^{-i}B \cap A\right) - \mu(A)\mu(B)\right| = 0,$$

then T is weakly mixing.

(b) If for all $A, B \in \mathcal{C}$ one has

$$\lim_{n\to\infty} \mu\left(T^{-i}B \cap A\right) = \mu(A)\mu(B),$$

then T is strongly mixing.

Example 1.4. *(Bernoulli shifts) Let $X = \{0, 1, \ldots, k-1\}^{\mathbb{N}}$, $\mathcal{B}$ generated by the cylinders. Let $p = (p_0, p_1, \ldots, p_{k-1})$ be a positive probability vector. We define the measure μ on $\mathcal{B}$ by specifying it on the cylinder sets as follows:*

$$\mu(\{x\colon x_0 = a_0, \ldots, x_n = a_n\}) = p_{a_0}p_{a_1}\cdots p_{a_n}.$$

Let $T\colon X \to X$ be defined by $Tx = y$ where $y_n = x_{n+1}$. This map T, called the left shift, is measure preserving and even strongly mixing with respect to μ. The measure preservingness follows easily from the fact that

$$T^{-1}\{x\colon x_0 = a_0, \ldots, x_n = a_n\} = \bigcup_{j=0}^{k-1}\{x\colon x_0 = j, x_1 = a_0, \ldots, x_{n+1} = a_n\}.$$

The cylinder sets form a semi-algebra, so we can use Proposition 1.1. Take A, B cylinders: $A = \{x\colon x_0 = a_0, x_1 = a_1, \ldots, x_k = a_k\}$ and $B = \{x\colon x_0 = b_0, x_1 = b_1, \ldots, x_m = b_m\}$. Now if we take $n \geq m+1$, then $T^{-n}A$ and B specify different coordinates, thus

$$\mu(T^{-n}A \cap B) = \mu(\{x\colon x_0 = b_0, \ldots, x_m = b_m, x_n = a_0, \ldots, x_{n+k} = a_k\}) =$$

$$= \mu(\{x\colon x_0 = b_0, \ldots, x_m = b_m\})\mu(\{x\colon x_n = a_0, \ldots, x_{n+k} = a_k\}) =$$

$$= \mu(B)\mu(T^{-n}A) = \mu(A)\mu(B),$$

which implies that T is strongly mixing.

Example 1.5. *(Markov shifts) Let* $(X, \mathcal{B}, T)$ *be as above. We define a measure* ν *on* $\mathcal{B}$ *by specifying it on the cylinder sets as follows: Let* $P = (p_{ij})$ *be a stochastic* $k \times k$ *matrix, and* $q = (q_0, q_1, \ldots, q_{k-1})$ *a positive probability vector such that* $qP = q$. *Define* ν *the following way:*

$$\nu(\{x \colon x_0 = a_0, \ldots, x_n = a_n\}) = q_{a_0} p_{a_0 a_1} \cdots p_{a_{n-1} a_n}.$$

Just as in the previous example, one can easily see that T *is measure preserving with respect to* ν. *Note that the Bernoulli shifts are Markov shifts with* $q = p$ *and with* $P = (p_0 \mathbb{1}, p_1 \mathbb{1}, \ldots, p_{k-1} \mathbb{1})$ *where* $\mathbb{1}$ *denotes the the vector for which all the* k *coordinates are 1.*

1.2 Entropy

There is a very important notion in ergodic theory called entropy. To define it, we need a few steps. First, let α be a finite or countable partition, i.e. X is a disjoint union (up to measure 0) of $A \in \alpha$. We define the entropy of the partition α by

$$H(\alpha) = H_\mu(\alpha) := -\sum_{A \in \alpha} \mu(A) \log \mu(A).$$

Here and from now on everywhere log represents logarithm with base 2. If α and β are partitions, then we define

$$\alpha \vee \beta := \{A \cap B \colon A \in \alpha, B \in \beta\}$$

and under $T^{-1}\alpha$ we consider the partition

$$T^{-1}\alpha = \{T^{-1}A \colon A \in \alpha\}.$$

Now consider the partition $\bigvee_{i=0}^{n-1} T^{-i}\alpha$, whose atoms, i.e. the members of the partition, are of the form $A_{i_0} \cap T^{-1}A_{i_1} \cap \cdots \cap T^{-(n-1)}A_{i_{n-1}}$. Then the following proposition can be proven:

Proposition 1.2. *Let* α *be a finite or a countable partition of the dynamical system* $(X, \mathcal{B}, \mu, T)$ *with* T *measure preserving transformation. Assume that* $H(\alpha) < \infty$. *Then* $\lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$ *exists.*

The proof is based on the fact that the above sequence is subadditive, as it can be shown that

$$H(\alpha \vee \beta) \le H(\alpha) + H(\beta)$$

and it is obvious that if T is μ-invariant, then $H(T^{-1}\alpha) = H(\alpha)$. This proposition allows to define the following:

Definition 1.4. *The entropy of the measure preserving transformation* T *with respect to the partition* α *is given by*

$$h(\alpha, T) = h_\mu(\alpha, T) := \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha).$$

For each $x \in X$, let us introduce the notation $\alpha(x)$ for the element of α to which x belongs. We close the introduction by stating one of the most important theorems of ergodic theory, which we are going to use quite often:

Theorem 1.2. *(The Shannon-McMillan-Breiman Theorem) Suppose T is an ergodic measure preserving transformation on a probability space $(X, \mathcal{B}, \mu)$, and let α be a countable partition with $H(\alpha) < \infty$. Denote $\alpha_n = \bigvee_{i=0}^{n-1} T^{-i}\alpha$, then*

$$\lim_{n\to\infty} -\frac{\log \mu(\alpha_n(x))}{n} = h(\alpha, T) \quad \mu - a.e.$$

For proofs and more detailed introduction, see([2]).

Chapter 2

Equipartitions

In this chapter we will investigate the rate at which the digits of one number-theoretic expansion determine those of another. First of all, let us define what a number-theoretic expansion is:

Definition 2.1. *A surjective map $T : [0,1) \to [0,1)$ is called a number-theoretic fibered map (NTFM) if it satisfies the following conditions:*

(a) there exists a finite or countable partition of intervals $P = \{P_i : i \in D\}$ such that T restricted to each atom of P (cylinder set of order 0) is monotone, continuous and injective,

(b) T is ergodic with respect to Lebesgue measure λ, and there exists a T-invariant probability measure μ equivalent to λ with bounded density. (Both $\frac{d\mu}{d\lambda}$ and $\frac{d\lambda}{d\mu}$ are bounded.)

Iterations of T generate expansions of points $x \in [0, 1)$ with digits in D. We refer to the resulting expansion as the T-expansion of x. Throughout the chapter we are going to assume the followings: let T and S be number-theoretic fibered maps on [0,1) with probability measures μ_1 and μ_2 respectively, each boundedly equivalent to Lebesgue measure and with generating partitions (cylinders of order 0) P and Q respectively. Denote by P_n and Q_n the interval partitions of [0,1) into cylinder sets of order n, namely, $P_n = \bigvee_{i=0}^{n-1} T^{-i}P$ and $Q_n = \bigvee_{i=0}^{n-1} S^{-i}Q$. Denote by $P_n(x)$ the element of P_n containing x (similarly for $Q_n(x)$), and introduce

$$m(n, x) = \sup\{m \geq 0 | P_n(x) \subset Q_m(x)\}.$$

Suppose that $H_{\mu_1}(P)$ and $H_{\mu_2}(Q)$ are finite and $h(T) = h_{\mu_1}(T)$ and $h(S) = h_{\mu_2}(S)$ are positive.

2.1 Central limit theorem

Our starting point in this section is the following theorem by Dajani and Fieldsteel[3]:

Theorem 2.1.

$$\lim_{n\to\infty} \frac{m(n,x)}{n} = \frac{h(T)}{h(S)} \quad a.e.$$

The same holds true if we look at the smallest k for which $P_k(x)$ is still contained in $Q_m(x)$. So let us define

$$k(m,x) = \inf\{k \geq 0 | P_k(x) \subset Q_m(x)\}.$$

Then under the same assumptions, the following is true:

Theorem 2.2.

$$\lim_{n\to\infty} \frac{k(m,x)}{m} = \frac{h(S)}{h(T)} \quad a.e.$$

Proof. Let us take x such that the convergence in Theorem 2.1 holds. Then for all $m > m(1,x)$ we can find $n \in \mathbb{N}$ such that $m(n,x) < m \leq m(n+1,x)$. From the definition of $m(n,x)$, this means that $P_{n+1}(x) \subset Q_{m(n+1,x)}(x) \subset Q_m(x)$ and that $P_n(x) \nsubseteq Q_m(x)$, i.e. $k(m,x) = n+1$. Thus

$$\frac{n+1}{m(n+1,x)} \leq \frac{n+1}{m} = \frac{k(m,x)}{m} < \frac{n+1}{m(n,x)}.$$

Here both the first and the last terms converge to $\frac{h(S)}{h(T)}$ from the previous theorem, which means that $\frac{k(m,x)}{m}$ converges and the limit is what we wanted. Since the convergence in Theorem 2.1 holds almost everywhere, we have completed the proof. $\square$

In the sequel the following two properties are going to play an important role:

Property 2.1. *We say that the triplet (P, μ_1, T) or in short the transformation T satisfies the 0-property if*

$$\frac{-\log \mu_1(P_n(x)) - nh(T)}{\sqrt{n}} \to 0$$

almost everywhere.

Property 2.2. *We say that the triplet (Q, μ_2, S) or in short the transformation S satisfies the CLT-property if*

$$\lim_{n\to\infty} \mu_2\left(\frac{-\log \mu_2(Q_n(x)) - nh(S)}{\sigma\sqrt{n}} \leq u\right) = \int\limits_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx,$$

for every $u \in \mathbb{R}$ and for some $\sigma > 0$.

Remark 2.1. *Notice that both properties can be stated with the Lebesgue measure since we only look at measures equivalent to it. In the case of the CLT-property, we mean that* $\log \mu_2(Q_n(x))$ *can be replaced with* $\log \lambda(Q_n(x))$.

Lemma 2.1. *If the transformation* T *satisfies the 0-property, then*

$$\log\left(\frac{\lambda(Q_m(x))}{\lambda(P_{k(m,x)}(x))}\right) = o\left(\sqrt{m}\right) \quad \textit{in probability} \tag{2.1}$$

Proof. We know that $\lim_{m\to\infty} \frac{k(m,x)}{m} = \frac{h(S)}{h(T)} > 0$ from Theorem 2.2. From the zero property we also know that for large n and for most of the x points $\log(\lambda(P_{n+1}(x))) - \log(\lambda(P_n(x))) \approx h(T) + o(\sqrt{n})$. We combine these two properties to get the result.

Indeed, since by definition $Q_m \supset P_{k(m,x)}$, we see that the left hand side of (2.1) is non negative. So we only have to prove that for any positive ε there is an m_0 such that for $m > m_0$

$$\lambda\left(\left\{x \in [0,1]\colon \log\left(\lambda(Q_m(x))\right) - \log\left(\lambda(P_{k(m,x)}(x))\right) > \varepsilon\sqrt{m}\right\}\right) < \varepsilon$$

Fix $\varepsilon > 0$ and put

$$A_n = \left\{x \in [0,1]\colon |\log(\lambda(P_n(x))) + nh(T)| > \frac{1}{2}\varepsilon\sqrt{n}\right\}.$$

By the zero property there is an n_0 such that $\lambda(\cup_{n\ge n_0} A_n) \le \varepsilon$.

Put

$$B_n = \left\{x\colon \forall m > n,\ \frac{1}{2} < \frac{k(m,x)}{m}\cdot\frac{h(T)}{h(S)} < 2\right\}$$

Since $\frac{k(m,x)}{m}$ has a finite positive limit $\frac{h(S)}{h(T)}$ for almost every x, we can find $m_1 > 2n_0\frac{h(S)}{h(T)}$ such that

$$\lambda\left(B_{m_1}\right) > 1 - \varepsilon.$$

Now if $x \notin \cup_{n\ge n_0} A_n$ then for $n > n_0$ we have that

$$\log(\lambda(P_{n-1}(x))) - \log(\lambda(P_n(x))) \le h(T) + \sqrt{n}\varepsilon$$

So if $m > m_1$ and $x \in B_{m_1} \setminus \bigcup_{n\ge n_0} A_n$ then

$$\begin{aligned} &d(x, \partial Q_m(x)) \le \lambda(P_{k(m,x)-1}(x)) \le \\ &\qquad 2^{h(T)+\varepsilon\sqrt{k(m,x)}}\lambda(P_{k(m,x)}(x)) \le 2^{h(T)+\varepsilon\sqrt{2\frac{h(S)}{h(T)}m}}\lambda(P_{k(m,x)}(x)) \end{aligned}$$

Hence

$$\begin{aligned} &\frac{\log\left(\lambda(Q_m(x))\right) - \log\left(\lambda(P_{k(m,x)}(x))\right)}{\sqrt{m}} \le \\ &\qquad \frac{\log\left(\lambda(Q_m(x))\right) - \log(d(x,\partial Q_m(x))) + h(T)}{\sqrt{m}} + \varepsilon\sqrt{2\frac{h(S)}{h(T)}}. \end{aligned} \tag{2.2}$$

For each m and $I \in Q_m$ let $I' \subset I$ a concentric interval such that $\lambda(I') = (1-\varepsilon)\lambda(I)$. Then $\lambda(\cup_{I\in Q_m} I') = 1-\varepsilon$ and for each $x \in \cup_{I\in Q_m} I'$ the right hand side of (2.2) is not greater than

$$\frac{\left|\log\left(\frac{2}{\varepsilon}\right)\right| + h(T)}{\sqrt{m}} + \varepsilon\sqrt{2\frac{h(S)}{h(T)}}.$$

Now we can put the above pieces together. Let $m_0 > m_1$ so large that

$$\frac{\left|\log\left(\frac{2}{\varepsilon}\right)\right| + h(T)}{\sqrt{m_0}} + \varepsilon\sqrt{2\frac{h(S)}{h(T)}} < \varepsilon\sqrt{3\frac{h(S)}{h(T)}}.$$

For $m > m_0$ and $x \in (B_{m_1} \cap \bigcup_{I\in Q_m} I') \setminus \bigcup_{n\geq n_0} A_n$ all the above estimation are valid, hence

$$\lambda\left(\left\{x \in [0,1]\colon\ \log\left(\lambda(Q_m(x))\right) - \log\left(\lambda(P_{k(m,x)}(x))\right) > \varepsilon\sqrt{3\frac{h(S)}{h(T)}m}\right\}\right) < 3\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the proof is complete. □

Now we can prove the following theorem which says that the speed of convergence in Theorem 2.2 is in fact of order $\frac{1}{\sqrt{m}}$:

Theorem 2.3. *Let us suppose that the transformation S satisfies the CLT-property and that T satisfies the 0-property. Then the following holds:*

$$\frac{k(m,x) - m\frac{h(S)}{h(T)}}{\sigma_1\sqrt{m}} \Rightarrow \mathcal{N}(0,1),$$

where $\sigma_1 = \frac{\sigma}{h(T)}$ and $\Rightarrow$ is the convergence in law with respect to the probability measure μ_2.

Proof. We can divide the expression above into four parts the following way:

$$\frac{k(m,x) - m\frac{h(S)}{h(T)}}{\sigma_1\sqrt{m}} =$$

$$= \frac{\log\lambda(P_{k(m,x)}(x)) + k(m,x)h(T)}{h(T)\sigma_1\sqrt{m}} + \frac{-\log\lambda(P_{k(m,x)}(x)) + \log\lambda(Q_m(x))}{h(T)\sigma_1\sqrt{m}} +$$

$$+ \frac{-\log\lambda(Q_m(x)) + \log\mu_2(Q_m(x))}{h(T)\sigma_1\sqrt{m}} + \frac{-\log\mu_2(Q_m(x)) - mh(S)}{h(T)\sigma_1\sqrt{m}}.$$

Here the first term goes to 0 for almost every x because of the conditions made on the transformation T. To see this, let us take $x \in [0,1)$ such that the convergence in $\frac{k(m,x)}{m} \to \frac{h(S)}{h(T)}$ is satisfied. It is enough to see that for such an x the first term goes to 0. We can prove this easily:

$$\frac{\log\lambda(P_{k(m,x)}(x)) + k(m,x)h(T)}{h(T)\sigma_1\sqrt{m}} =$$

$$= \frac{1}{h(T)\sigma_1}\sqrt{\frac{k(m,x)}{m}}\frac{\log\lambda(P_{k(m,x)}(x)) + k(m,x)h(T)}{\sqrt{k(m,x)}},$$

and since the multiplier $\frac{1}{h(T)\sigma_1}\sqrt{\frac{k(m,x)}{m}} \to \frac{1}{h(T)\sigma_1}\sqrt{\frac{h(S)}{h(T)}} < \infty$ and the rest is a subsequence of the opposite of the sequence appearing in the assumption of the 0-property since $k(m,x) \to \infty$ as $m \to \infty$, the whole thing goes to 0.

The second term equals to $\frac{1}{h(T)\sigma_1\sqrt{m}} \log \frac{\lambda(Q_m(x))}{\lambda(P_{k(m,x)}(x))}$ and thus goes to 0 in probability by Lemma 2.1.

The third term will go to 0 for every x because of the equivalency of the measures λ and μ_2. That is to say there $\exists$ K_1 and K_2 positive and finite constants such that for every Borel-set B: $K_1\mu_2(B) \leq \lambda(B) \leq K_2\mu_2(B)$. Hence

$$\frac{-\log K_1}{h(T)\sigma_1\sqrt{m}} = \frac{-\log K_1 - \log\mu_2(Q_m(x)) + \log\mu_2(Q_m(x))}{h(T)\sigma_1\sqrt{m}} \geq$$

$$\geq \frac{-\log\lambda(Q_m(x)) + \log\mu_2(Q_m(x))}{h(T)\sigma_1\sqrt{m}} \geq$$

$$\geq \frac{-\log K_2 - \log\mu_2(Q_m(x)) + \log\mu_2(Q_m(x))}{h(T)\sigma_1\sqrt{m}} = \frac{-\log K_2}{h(T)\sigma_1\sqrt{m}}.$$

Here both the first and the last expression go to 0 for every x and thus so does the the third term above.

So all is left is the fourth term, and that one has a limiting standard normal distribution since we have assumed that for the transformation S and we have chosen $\sigma_1 = \frac{\sigma}{h(T)}$. This means that altogether $k(m,x)$ satisfies a central limit theorem. □

Remark 2.2. *From now on there will be times when we would like to show that something converges in probability using that for some other expression we already know this. Here is how we are going to proceed in these cases. We would like to use the expression - even in connection with a sequence of random variables that only converges in probability - that we 'take a point where the convergence holds'. By this we are going to mean the following. It is well-known that a sequence of random variables converges in probability if and only if for every subsequence there is a sub-subsequence which converges almost surely. So if we want to show that a sequence converges in probability, we can take a subsequence and we only need to show that there is a sub-subsequence which converges almost surely. And vice versa, if we have a sequence that converges in probability, then for every subsequence of it, we can take a sub-subsequence that converges almost everywhere. So when we say for a sequence of random variables which converges in probability that let us take a point where the convergence holds, we mean that there has been a subsequence for which we can take a sub-subsequence which converges almost everywhere and thus we take a point where this almost sure convergence holds. But we do not want to bother with subsequences and sub-subsequences which would cause some problems with the notations. However, it is very important to bear in mind that this is what we mean by that.*

2.2 Weak invariance principle

The central limit theorem can be improved to what is called the weak invariance principle if we assume more for the transformation S.

For each $x \in [0,1)$ and $m \in \mathbb{N}$ let us take the following random variables:

$$W_{m,x}\left(\frac{l}{m}\right) = \frac{-\log \mu_2(Q_l(x)) - lh(S)}{\sigma\sqrt{m}}$$

and we extend this linearly on each subintervals $[\frac{l}{m}, \frac{l+1}{m}]$. This way for each x $W_{m,x}(t)$ is an element of the space $C[0,1]$ of the continuous functions on $[0,1]$ topologised with the supremum norm.

Definition 2.2. *We say that S satisfies the weak invariance principle if the process $W_m(t)$ converges in law to the Brownian motion on $[0,1]$.*

In this section we are going to suppose for the transformation S that it satisfies this weak invariance principle and then, as we did it previously with the CLT-property, we are going to prove that so does $k(m,x)$. Our first theorem is the following:

Theorem 2.4. *Suppose that the transformation T satisfies the 0-property and that S satiesfies the weak invariance principle. Let us take*

$$K_{m,x}\left(\frac{l}{m}\right) = \frac{k(l,x) - l\frac{h(S)}{h(T)}}{\sigma_1\sqrt{m}}$$

and let us extend it linearly on each subintervals $[\frac{l}{m}, \frac{l+1}{m}]$ in $[0,1]$, where $\sigma_1 = \frac{\sigma}{h(T)}$. Then the process $K_m(t)$ for $t \in [0,1]$ converges in law to the Brownian motion on $[0,1]$.

Proof. Since S satiesfies the weak invariance principle, it is enough to show that

$$\sup_{t\in[0,1]} |K_m(t) - W_m(t)| \to 0$$

in probability. Let us notice that both $K_m(t)$ and $W_m(t)$ are constructed the same way that is we take their value at points $\frac{l}{m}$ and then extend them linearly in between. Hence it is sufficient to take the supremum for $t = \frac{l}{m}$ for $l = 0, 1, ..., m$. This is easy to see from the fact that the difference of two linear functions is linear and that the supremum of a linear function over a closed interval is taken at one of the endpoints. So let us take $0 \le l \le m$ and $x \in [0,1)$. Then we can divide the difference into three parts:

$$K_{m,x}\left(\frac{l}{m}\right) - W_{m,x}\left(\frac{l}{m}\right) =$$

$$= \frac{k(l,x) - l\frac{h(S)}{h(T)}}{\sigma_1\sqrt{m}} - \frac{-\log \mu_2(Q_l(x)) - lh(S)}{\sigma\sqrt{m}} =$$

$$= \frac{k(l,x)h(T) - lh(S) + \log \mu_2(Q_l(x)) + lh(S)}{h(T)\sigma_1\sqrt{m}} =$$

$$= \frac{k(l,x)h(T) + \log \lambda(P_{k(l,x)}(x))}{h(T)\sigma_1\sqrt{m}} + \frac{\log \frac{\lambda(Q_l(x))}{\lambda(P_{k(l,x)}(x))}}{h(T)\sigma_1\sqrt{m}} + \frac{\log \frac{\mu_2(Q_l(x))}{\lambda(Q_l(x))}}{h(T)\sigma_1\sqrt{m}}.$$

Let us denote this three parts with $R_1\left(\frac{l}{m}, x\right)$, $R_2\left(\frac{l}{m}, x\right)$ and $R_3\left(\frac{l}{m}, x\right)$ respectively. We are going to prove that

$$\max_{l=0,1,\ldots,m} \left| R_i\left(\frac{l}{m}, x\right) \right| \to 0,$$

in probability, for i=1,2,3. The third term is the easiest: because of the equivalence of the measures there exists K_1 and K_2 that $K_1 \leq \frac{\mu_2(B)}{\lambda(B)} \leq K_2$ for every B Borel-set and thus

$$0 \leq \max_{l=0,1,\ldots,m} \left| R_3\left(\frac{l}{m}, x\right) \right| \leq \frac{\max\{|\log K_1|, |\log K_2|\}}{h(T)\sigma_1\sqrt{m}}$$

and this goes to 0. For the other two terms we are going to need a little lemma:

Lemma 2.2. *Let $g_m(l) = \frac{1}{\sqrt{m}} f(l)$ be a real valued function for positive integer m, where $0 \leq l \leq m$ is also an integer. Let us suppose that $g_m(m) \to 0$ as $m \to \infty$. Then*

$$\max_{0 \leq l \leq m} |g_m(l)| \to 0$$

as $m \to \infty$.

Proof. First we are going to prove that

$$\max_{0 \leq l \leq m} g_m(l) \to 0$$

as $m \to \infty$. Let $l(m)$ be the largest integer not bigger than m, where the maximum occurs, i.e.

$$g_m(l(m)) = \max_{0 \leq l \leq m} g_m(l).$$

Let us notice the following:

$$g_{m+1}(l(m+1)) = \max\left\{ \frac{\sqrt{m}}{\sqrt{m+1}} g_m(l(m)), g_{m+1}(m+1) \right\}.$$

From this it is easy to see that either $l(m+1) = l(m)$ or $l(m+1) = m+1$. Hence whenever $l(m+1) > l(m)$ then $l(m+1) = m+1$. Thus $l(m)$ is monotone increasing and if it is not bounded, then there are infinitely many m integers for which $l(m) = m$. This way we can distinguish two cases: the first is when $l(m)$ is bounded: this can only happen if there exists N integer that for every $m \geq N$: $l(m) = l(N)$. So let us take in this case $m \geq N$:

$$g_m(m) \leq \max_{0 \leq l \leq m} g_m(l) = g_m(l(m)) = g_m(l(N)) = \frac{\sqrt{N}}{\sqrt{m}} g_N(l(N)).$$

Here both the first and the last term goes to 0 as $m \to \infty$ and so does the maximum.

The second case is when $l(m)$ is not bounded: this can only happen if there is a strictly increasing (and thus going to infinity) sequence m_n such that $l(m_n) = m_n$ and if $m_n \leq m < m_{n+1}$, then $l(m) = l(m_n) = m_n$. This way:

$$g_m(m) \leq \max_{0 \leq l \leq m} g_m(l) = g_m(l(m)) = g_m(m_n) = \frac{\sqrt{m_n}}{\sqrt{m}} g_{m_n}(m_n) \leq g_{m_n}(m_n)$$

here again the first and the last term goes to 0 as $m \to \infty$.

Now let us notice the following:

$$\max_{0 \leq l \leq m} |g_m(l)| = \max\{\max_{0 \leq l \leq m} g_m(l), -\min_{0 \leq l \leq m} g_m(l)\}.$$

We can use the first part of the proof for the function $r = -g$ and conclude that

$$\max_{0 \leq l \leq m} r_m(l) = -\min_{0 \leq l \leq m} g_m(l) \to 0,$$

which finishes the proof. $\square$

We can use this lemma to conclude the proof of our theorem. First we show that

$$\max_{l=0,1,\ldots,m} \left| R_1\left(\frac{l}{m}, x\right) \right| \to 0$$

almost everywhere. Let us take $x \in [0, 1)$ where the convergence of the 0-property and the convergence of Theorem 2.2 hold. Both happen almost everywhere, so it is sufficient to prove the above convergence for these x. Now we can take

$$f(l) = \frac{k(l,x)h(T) + \log \lambda(P_{k(l,x)}(x))}{h(T)\sigma_1}.$$

We can use the lemma with this f because we assumed that at the point x the 0-property holds, i.e.:

$$\frac{-\log \mu_1(P_n(x)) - nh(T)}{\sqrt{n}} \to 0.$$

As usual, because of the equivalence of the measures λ and μ_1 the same holds if we write λ instead of μ_1 in the convergence of the 0-property. We need to check if the assumption of the lemma holds:

$$g_m(m) = \frac{1}{\sqrt{m}} f(m) = \frac{k(m,x)h(T) + \log \lambda(P_{k(m,x)}(x))}{h(T)\sigma_1\sqrt{m}} =$$

$$= \sqrt{\frac{k(m,x)}{m}} \frac{k(m,x)h(T) + \log \lambda(P_{k(m,x)}(x))}{h(T)\sigma_1\sqrt{k(m,x)}},$$

which goes to 0 as $m \to \infty$.

Next we show that

$$\max_{l=0,1,\ldots,m}\left|R_2\left(\frac{l}{m},x\right)\right|\to 0$$

in probability. Here we take $x\in[0,1)$ - bearing in mind Remark 2.2 - such that the convergence in Lemma 2.1 holds. Then we take

$$f(l)=\frac{1}{h(T)\sigma_1}\log\frac{\lambda(Q_l(x))}{\lambda(P_{k(l,x)}(x))}.$$

Now we can again use the previous lemma if we show that its assumption holds:

$$g_m(m)=\frac{1}{\sqrt{m}}f(m)=\frac{1}{h(T)\sigma_1\sqrt{m}}\log\frac{\lambda(Q_m(x))}{\lambda(P_{k(m,x)}(x))},$$

which goes to 0 because of Lemma 2.1. At last we can put all the pieces together:

$$0\le\sup_{t\in[0,1]}|K_{m,x}(t)-W_{m,x}(t)|=\max_{l=0,1,\ldots,m}\left|K_{m,x}\left(\frac{l}{m}\right)-W_{m,x}\left(\frac{l}{m}\right)\right|\le$$

$$\le\max_{l=0,1,\ldots,m}\left|R_1\left(\frac{l}{m},x\right)\right|+\max_{l=0,1,\ldots,m}\left|R_2\left(\frac{l}{m},x\right)\right|+\max_{l=0,1,\ldots,m}\left|R_3\left(\frac{l}{m},x\right)\right|\to 0$$

in probability. This proves the theorem. □

We are going to use the previous theorem to get a similar result for $m(n,x)$. To get there, first we need a few propositions as we are going step by step.

Proposition 2.1. *Let us suppose that we have a stochastic process S_t for $t\ge 0$, with continuous realizations. Let us take*

$$\mathcal{S}_n(t)=\frac{1}{\sqrt{n}}S_{nt}.$$

If $\mathcal{S}_n(t)$ converges in law to the Brownian motion on $[0,1]$, then it also converges in law to it on $[0,k]$ for every $k\in\mathbb{N}$.

Proof. Let us take $\phi\colon C([0,1])\to C([0,k])$ such that $\phi(\omega(t))=\sqrt{k}\omega\left(\frac{t}{k}\right)$. Then this ϕ is a continuous mapping. Let us notice that $\phi\left(\mathcal{S}_{kn|_{[0,1]}}\right)=\mathcal{S}_{n|_{[0,k]}}$. It is well-known that if $B(t)$ is a Brownian motion, then so is $aB\left(\frac{t}{a^2}\right)$ for every $a\neq 0$. This means that $\phi\left(B_{|_{[0,1]}}\right)=B_{|_{[0,k]}}$ in law, and thus it is a Brownian motion on $[0,k]$. Hence as ϕ is continuous and $\mathcal{S}_{kn|_{[0,1]}}$ converges in law to the Brownian motion on $[0,1]$, $\mathcal{S}_{n|_{[0,k]}}$ converges in law to the Brownian motion on $[0,k]$. This concludes the proof. □

This proposition tells us that if we extend the definition of K_n, then it follows that it converges in law to the Brownian motion on $[0,k]$. This observation will be of great help in the sequel. But first of all, let us introduce the following two processes:

$$K'_{n,x}\left(t_l^n(x)\right) = \frac{l\frac{h(S)}{h(T)} - k(l,x)}{\sigma_1\sqrt{n}}$$

and

$$M'_{n,x}\left(t_l^n(x)\right) = \frac{\frac{h(S)}{h(T)}m(k(l,x),x) - k(l,x)}{\sigma_1\sqrt{n}}$$

for $t_l^n(x) = \frac{h(T)}{h(S)}\frac{k(l,x)}{n}$ and for $l = 0, 1, \ldots, n_1(x)+1$, where $n_1(x)$ is the largest positive integer for which $t^n_{n_1(x)}(x) \leq 1$. Then as before we extend these linearly on each subintervals $[t_l^n(x), t_{l+1}^n(x)]$ for $l = 0, 1, \ldots, n_1(x)$ and we take the parts for $t \in [0,1]$.

Proposition 2.2. *Suppose that the transformation T satisfies the 0-property and that S satiesfies the weak invariance principle. Then the process $K'_n(t)$ converge in law to the Brownian motion on* $[0,1]$.

Proof. We will show that $K'_n(t)$ has the same limit as $-K_n(t)$. As both of them are piecewise linear, the maximum of the difference can only occur at points $\frac{l}{n}$ for $l \in \{0, \ldots, n\}$ or at $t_l^n(x)$ for $l \in \{0, \ldots, n_1(x)\}$. Let us also notice that because of the definitions of the processes

$$K'_{n,x}\left(t_l^n(x)\right) = -K_{n,x}\left(\frac{l}{n}\right).$$

Thus we can write the difference between K'_n and $-K_n$ at points $t_l^n(x)$ for $l = 0, \ldots, n_1(x)$ as the difference of the value of K_n at two certain points:

$$K'_{n,x}\left(t_l^n(x)\right) + K_{n,x}\left(t_l^n(x)\right) = K_{n,x}\left(t_l^n(x)\right) - K_{n,x}\left(\frac{l}{n}\right).$$

We would like to do the same for the points $\frac{l}{n}$ too. Let us introduce $l(n,x) = \max\{l \leq n \colon t^n_{n_1(x)}(x) > \frac{l}{n}\}$. For every $l \leq l(n,x)$ there exists an $l' < n_1(x)$ such that $t^n_{l'}(x) \leq \frac{l}{n} < t^n_{l'+1}(x) \leq 1$. For $l(n,x) < l \leq n$ we can take $l' = n_1(x)$ as $t^n_{n_1(x)}(x) \leq \frac{l}{n} \leq 1 < t^n_{n_1(x)+1}(x)$. From the construction of the process K'_n it follows that

$$K'_{n,x}(t^n_{l'}(x)) \leq K'_{n,x}\left(\frac{l}{n}\right) \leq K'_{n,x}(t^n_{l'+1}(x))$$

or

$$K'_{n,x}(t^n_{l'}(x)) \geq K'_{n,x}\left(\frac{l}{n}\right) \geq K'_{n,x}(t^n_{l'+1}(x)).$$

Then as $K'_{n,x}(t_l^n(x)) = -K_{n,x}\left(\frac{l}{n}\right)$ we get that

$$-K_{n,x}\left(\frac{l'}{n}\right) \leq K'_{n,x}\left(\frac{l}{n}\right) \leq -K_{n,x}\left(\frac{l'+1}{n}\right)$$

or

$$-K_{n,x}\left(\frac{l'}{n}\right) \geq K'_{n,x}\left(\frac{l}{n}\right) \geq -K_{n,x}\left(\frac{l'+1}{n}\right).$$

Now we can put the above pieces together to get

$$\sup_{t\in[0,1]} \left|K'_{n,x}(t) + K_{n,x}(t)\right| \leq$$

$$\max_{l=0,\dots,n}\left\{\left|K'_{n,x}\left(\frac{l}{n}\right) + K_{n,x}\left(\frac{l}{n}\right)\right|\right\} + \max_{l=0,\dots,n_1(x)}\left|K'_{n,x}\left(t_l^n(x)\right) + K_{n,x}\left(t_l^n(x)\right)\right|$$

$$\leq \max_{l=0,\dots,n}\{R_1^n(l,l',x)\} + \max_{l=0,\dots,n}\{R_2^n(l,l',x)\} + \max_{l=0,\dots,n_1(x)}\{R_3^n(l,x)\},$$

where

$$R_1^n(l,l',x) = \left|K_{n,x}\left(\frac{l}{n}\right) - K_{n,x}\left(\frac{l'}{n}\right)\right|,$$

$$R_2^n(l,l',x) = \left|K_{n,x}\left(\frac{l}{n}\right) - K_{n,x}\left(\frac{l'+1}{n}\right)\right|,$$

and

$$R_3^n(l,x) = \left|K_{n,x}\left(t_l^n(x)\right) - K_{n,x}\left(\frac{l}{n}\right)\right|.$$

From the previous theorem we know that $K_n(t)$ converges in law. From the tightness of the sequence it follows that for every $\eta > 0$ there exists a compact set $K_\eta \subset C[0,1]$ that

$$\mu_2(x \in [0,1]\colon K_{n,x}(t) \in K_\eta) > 1-\eta, \quad \text{for each } n.$$

From the compactness of a set K it follows that for $\varepsilon > 0$ exists $\delta > 0$ such that $w \in K$ implies that $|w(t) - w(s)| \leq \varepsilon$ if $|t-s| < \delta$. Hence in order to prove that the sum above goes to 0 in probability, it is enough to show that

$$\max_{l=0,\dots,n}\left\{\left|\frac{l-l'}{n}\right|\right\} \to 0 \quad \text{in probability}$$

and

$$\max_{l=0,\dots,n_1(x)}\left\{\left|t_l^n(x) - \frac{l}{n}\right|\right\} \to 0 \quad \text{in probability.}$$

First we are going to need a little lemma. Since we will use it later in a more general form that is how we state it. Let us introduce - analogously to $n_1(x)$ - the notation $n_k(x) = \max\{l \geq 0\colon t_l^n(x) \leq k\}$.

Lemma 2.3.

$$\lim_{n\to\infty} \lambda(x \in [0,1]\colon n_k(x) < 2kn) = 1.$$

Proof. We need to show that if we take $\varepsilon > 0$, $\exists\, N$ such that for every $n \geq N$:

$$\lambda(x \in [0,1]\colon n_k(x) < 2kn) > 1 - \varepsilon.$$

From Theorem 2.1 we know that

$$\lim_{n\to\infty} \frac{h(T)}{h(S)} \frac{k(n,x)}{n} = 1 \quad a.e.$$

Then from Jegorov's theorem it follows that the convergence is also true almost uniformly, meaning that for every $\varepsilon > 0$ there exists $A_k \subset [0,1]$ such that $\lambda(A_k) < \varepsilon$ and for $\eta > 0$ $\exists M$ such that for $m \geq M$ and $x \in [0,1] \setminus A_k$:

$$\frac{h(T)}{h(S)} \frac{k(m,x)}{m} > 1 - \eta.$$

Let us take $\eta = \frac{1}{4k}$ and $m = 2kn$. Then we have

$$\frac{h(T)}{h(S)} \frac{k(2kn,x)}{n} > 2k - \frac{1}{2} > k.$$

As $t^n_{n_k(x)}(x) \leq k$, we get for $n \geq N = \lceil \frac{M}{2k} \rceil$ and $x \in [0,1] \setminus A_k$

$$n_k(x) < 2kn.$$

□

With this lemma, it is sufficient to show that on the set $[0,1] \setminus A_1$

$$\max_{l=0,\ldots,n_1(x)} \left\{ \left| t^n_l(x) - \frac{l}{n} \right| \right\} \leq \max_{l=0,\ldots,2n} \left\{ \left| t^n_l(x) - \frac{l}{n} \right| \right\} \to 0$$

in probability. This follows easily from Lemma 2.2. We can take the function f as

$$f(l) = \frac{h(T)k(l,x) - h(S)l}{h(S)}$$

and we shall remark that the lemma is also true and can be easily proven if we multiply f with $\frac{1}{m}$ instead of $\frac{1}{\sqrt{m}}$. Thus we have that

$$\max_{l=0,\ldots,2n} \left\{ \left| t^n_l(x) - \frac{l}{n} \right| \right\} \to 0$$

if $\left| \frac{f(2n)}{n} \right| \to 0$. This is easy to see:

$$\left| \frac{f(2n)}{n} \right| = \left| \frac{h(T)k(2n,x) - h(S)2n}{h(S)n} \right| =$$

$$= \frac{h(T)}{h(S)} \left| \frac{k(2n,x)}{n} - 2\frac{h(S)}{h(T)} \right| \to 0 \quad a.e.$$

from Theorem 2.1.

Now from the definition of l' we have that $\frac{h(T)k(l',x)}{h(S)} \leq l < \frac{h(T)k(l'+1,x)}{h(S)}$. Hence we can write:

$$\frac{|l-l'|}{n} \leq \frac{l - \frac{h(T)}{h(S)}k(l',x) + \left|\frac{h(T)}{h(S)}k(l',x) - l'\right|}{n} \leq$$

$$\leq \frac{\frac{h(T)}{h(S)}k(l'+1,x) - \frac{h(T)}{h(S)}k(l',x)}{n} + \frac{\left|\frac{h(T)}{h(S)}k(l',x) - l'\right|}{n}.$$

Thus for $x \in [0,1] \setminus A_1$

$$\max_{l=0,\dots,n} \left\{ \left| \frac{l-l'}{n} \right| \right\} \leq$$

$$\leq \max_{l'=0,\dots,n_1(x)} \left\{ \frac{\left|\frac{h(T)}{h(S)}k(l'+1,x) - \frac{h(T)}{h(S)}k(l',x)\right|}{n} + \frac{\left|\frac{h(T)}{h(S)}k(l',x) - l'\right|}{n} \right\} \leq$$

$$\leq \max_{l'=0,\dots,2n} \left\{ \frac{h(T)}{h(S)} \frac{k(l'+1,x) - k(l',x)}{n} + \frac{\left|\frac{h(T)}{h(S)}k(l',x) - l'\right|}{n} \right\}.$$

Here again we can use the modification of Lemma 2.2 - meaning that we divide by n and not by $\sqrt{n}$ - with

$$f_1(l) = k(l+1,x) - k(l,x) \quad \text{ and } f_2(l) = \left| \frac{h(T)}{h(S)} k(l,x) - l \right|.$$

We only have to check that

$$\frac{f_1(2n)}{n} = \frac{k(2n+1,x) - k(2n,x)}{n} \to 0 \quad a.e.$$

and

$$\frac{f_2(2n)}{n} = \frac{\frac{h(T)}{h(S)}k(2n,x) - 2n}{n} \to 0 \quad a.e.$$

Both of them follow easily from Theorem 2.1.

With this we have just finished proving that

$$\sup_{t\in[0,1]} \left| K'_{n,x}(t) + K_{n,x}(t) \right| \to 0 \quad a.e.$$

which means that K'_n also converges weakly to the Brownian motion on $[0,1]$. $\square$

To take the next step we need a proposition which will help to understand the strong connection between $k(m,x)$ and $m(n,x)$.

Proposition 2.3. *Assume that T satisfies the 0-property and that S satisfies the weak invariance principle. Then,*

$$0 \le m(k(n,x),x) - n \le o\left(\sqrt{n}\right), \quad \textit{in probability.}$$

Proof. It is clear from the defintions that $m'(n,x) = m(k(n,x),x) \ge n$ for all $n \in \mathbb{N}$. From Lemma 2.1 it is also clear that

$$\log \lambda(Q_n(x)) - \log \lambda(Q_{m'(n,x)}(x)) \le \log\left(\frac{\lambda(Q_n(x))}{\lambda(P_{k(n,x)}(x))}\right) = o\left(\sqrt{n}\right)$$

in probability. Since μ_2 is equivalent to λ, we have that

$$\lim_{n\to\infty} \log \mu_2(Q_n(x)) - \log \lambda(Q_n(x)) = \log \frac{d\mu_2}{d\lambda}(x).$$

Thus

$$\lim_{n\to\infty} \frac{\log \mu_2(Q_n(x)) - \log \mu_2(Q_{m'(n,x)}(x))}{\sqrt{n}} = 0 \quad \text{in measure (both } \lambda \text{ and } \mu_2\text{).} \tag{2.3}$$

Let us take $W_n(t)$ as in Definition 2.2. From the weak invariance principle

$$W_n \Rightarrow W, \quad \text{as } n \to \infty$$

where W is a Brownian motion on $[0,1]$.

As $\frac{m'(n,x)}{n} \to 1$, for n large enough $m'(n,x) < 2n$ on a large set of x-es, i.e. for $\varepsilon > 0$ exists n_0 such that $\mu_2\left(x\colon m'(n,x) < 2n\right) > 1-\varepsilon$ if $n > n_0$. From the tightness of the sequence (W_n), for any $\eta > 0$ there is a compact set $K_\eta \subset C([0,1])$ such that

$$\mu_2\left(x\colon W_n(.,x) \in K_\eta\right) > 1-\eta, \quad \text{for each } n. \tag{2.4}$$

From the compactness of K_η it follows that for $\varepsilon > 0$ exists $\delta > 0$ such that $w \in K_\eta$ implies that $|w(t) - w(s)| \le \varepsilon$ if $|t-s| < \delta$. Thus,

$$\left| W_n\left(\frac{1}{2}, x\right) - W_n\left(\frac{m'(n,x)}{2n}, x\right)\right| \le \varepsilon$$

provided that n so large that $m'(n,x) < n(1+2\delta)$, and $W_n(.,x) \in K_\eta$.

As

$$h(S)\frac{m'(n,x) - n}{\sqrt{2n}} =$$

$$= \frac{-\log \mu_2(Q_n(x)) - nh(S) + \log \mu_2(Q_{m'(n,x)}(x)) + m'(n,x)h(S)}{\sqrt{2n}} +$$

$$+\frac{\log \mu_2(Q_n(x)) - \log \mu_2(Q_{m'(n,x)}(x))}{\sqrt{2n}} =$$

$$= \left(W_{2n}\left(\frac{1}{2}, x\right) - W_{2n}\left(\frac{m'(n,x)}{2n}, x\right)\right) + \frac{\log \mu_2(Q_n(x)) - \log \mu_2(Q_{m'(n,x)}(x))}{\sqrt{2n}},$$

and here both terms converges to 0 in probability, which we can see from (2.3) and (2.4). Hence so does $\frac{m'(n,x)-n}{\sqrt{n}}$. This completes the proof.

□

With this the next step becomes quite easy:

Proposition 2.4. *Suppose that the transformation T satisfies the 0-property and that S satiesfies the weak invariance principle. Then the process $M'_n(t)$ converges in law to the Brownian motion on* $[0,1]$.

Proof. We see that difference between K'_n and M'_n, and we notice that the supremum of the difference can only occur at points $t^n_l(x)$. So for $x \in [0,1] \setminus A_1$

$$\sup_{t\in[0,1]} \left|M'_{n,x}(t) - K'_{n,x}(t)\right| \le \max_{l=0,\dots,n_1(x)} \left|\frac{m(k(l,x),x)\frac{h(S)}{h(T)} - l\frac{h(S)}{h(T)}}{\sigma_1\sqrt{n}}\right| \le$$

$$\le \max_{l=0,\dots,2n} \frac{h(S)}{h(T)\sigma_1}\left|\frac{m(k(l,x),x) - l}{\sqrt{n}}\right|.$$

We can again use Lemma 2.2 by putting $f(l) = m(k(l,x),x) - l$ and then from Proposition 2.3 we have that $M'_{n,x}(t)$ also converges to the Brownian motion on $[0,1]$, since

$$0 \le \frac{m(k(2n,x),x) - 2n}{\sqrt{n}} \to 0 \quad \text{in probability.}$$

□

Now we can state the last theorem, which claims that under the same assumptions as before $m_n(x)$ also satisfies the weak invariance principle:

Theorem 2.5. *Suppose that the transformation T satisfies the 0-property and that S satiesfies the weak invariance principle. Let us take $k \in \mathbb{N}$ such that $\frac{h(T)}{h(S)} < k$, then let us take*

$$M_{n,x}\left(\frac{h(T)}{h(S)}\frac{l}{n}\right) = \frac{\frac{h(S)}{h(T)}m(l,x) - l}{\sigma_1\sqrt{n}}$$

and let us extend it linearly on each subintervals $[\frac{h(T)}{h(S)}\frac{l}{n}, \frac{h(T)}{h(S)}\frac{l+1}{n}]$. Then the process $M_n(t)$ converge in law to the Brownian motion on $[0,k]$.

Proof. We just need to show that the difference between $M_{n,x}(t)$ and $M'_{n,x}(t)$ goes to 0 in probability. Let us introduce the notation $L = \left\lfloor \frac{h(S)}{h(T)} nk \right\rfloor$. We notice in this case that the supremum of the difference can only occur at points $\frac{h(T)}{h(S)} \frac{l}{n}$ for $l = 0, \ldots, L$ or at point k. We take $n_k(x) = \max \left\{ l \geq 0 \colon t_l^n(x) \leq k < t_{l+1}^n(x) \right\}$. Denote $l_k(n,x) = \max\{l \leq L \colon t^n_{n_k(x)}(x) > \frac{h(T)}{h(S)} \frac{l}{n}\}$. For every $l \leq l_k(n,x)$ there exists an $l' < n_k(x)$ such that $k(l',x) \leq l < k(l'+1,x)$. For $L \geq l > l_k(n,x)$ there is $n_k(x)$ such that $k(n_k(x),x) \leq l < k(n_k(x)+1,x)$, i.e. for these points we can take $l' = n_k(x)$. So we get that

$$\frac{\frac{h(S)}{h(T)} m(k(l',x),x) - k(l'+1,x)}{\sigma_1 \sqrt{n}} \leq \frac{\frac{h(S)}{h(T)} m(l,x) - l}{\sigma_1 \sqrt{n}} \leq$$

$$\leq \frac{\frac{h(S)}{h(T)} m(k(l'+1,x),x) - k(l',x)}{\sigma_1 \sqrt{n}}.$$

Hence we have

$$M'_{n,x}(t^n_{l'}(x)) - \frac{k(l'+1,x) - k(l',x)}{\sigma_1 \sqrt{n}} \leq M_{n,x}\left(\frac{h(T)}{h(S)} \frac{l}{n}\right) \leq$$

$$\leq M'_{n,x}(t^n_{l'+1}(x)) + \frac{k(l'+1,x) - k(l',x)}{\sigma_1 \sqrt{n}}.$$

First let us show that the correcting terms above go to 0:

$$\max_{l=0,\ldots,L} \left\{ \frac{k(l'+1,x) - k(l',x)}{\sqrt{n}} \right\} \leq$$

$$\leq \max_{l=0,\ldots,n_k(x)} \left\{ \frac{k(l+1,x) - k(l,x)}{\sqrt{n}} \right\} \to 0 \quad \text{in probability.}$$

Because of Lemma 2.3, it is enough to show that

$$\max_{l=0,\ldots,2kn-1} \left\{ \frac{k(l+1,x) - k(l,x)}{\sqrt{n}} \right\} \to 0 \quad \text{in probability.}$$

Again we can use Lemma 2.2 and we bear in mind Remark 2.2 and thus we only need to show that

$$\frac{k(2kn,x) - k(2kn-1,x)}{\sqrt{n}} \to 0 \quad \text{in probability.}$$

Notice that this term equals to $K_{n,x}(2k) - K_{n,x}\left(\frac{2kn-1}{n}\right) + \frac{h(S)}{h(T)\sigma_1\sqrt{n}}$. This converges to 0 in probability, which follows from the fact that K_n converges weakly to the Brownian motion. Hence we got that

$$\max_{l=0,\ldots,L} \left| M'_{n,x}\left(\frac{h(T)}{h(S)} \frac{l}{n}\right) - M_{n,x}\left(\frac{h(T)}{h(S)} \frac{l}{n}\right) \right| \leq$$

$$\leq \max_{l=0,\ldots,L} \left\{ R_1^n(l,l',x), R_2^n(l,l',x) \right\},$$

where

$$R_1^n(l,l',x) = \left| M'_{n,x}\left(\frac{h(T)}{h(S)}\frac{l}{n}\right) - M'_{n,x}\left(t^n_{l'}(x)\right) \right|$$

$$R_2^n(l,l',x) = \left| M'_{n,x}\left(\frac{h(T)}{h(S)}\frac{l}{n}\right) - M'_{n,x}\left(t^n_{l'+1}(x)\right) \right|.$$

As before, because of the tightness of the sequence $M'_{n,x}$, we only need to show that

$$\max_{l=0,\ldots,L} \left\{ \frac{h(T)}{h(S)} \frac{l - k(l',x)}{n} \right\} \to 0 \quad \text{in probability}$$

and that

$$\max_{l=0,\ldots,L} \left\{ \frac{h(T)}{h(S)} \frac{k(l'+1,x) - l}{n} \right\} \to 0 \quad \text{in probability.}$$

Both follow if we show that

$$\max_{l=0,\ldots,L} \left\{ \frac{h(T)}{h(S)} \frac{k(l'+1,x) - k(l',x)}{n} \right\} \leq$$

$$\leq \max_{l=0,\ldots,n_k(x)} \left\{ \frac{h(T)}{h(S)} \frac{k(l+1,x) - k(l,x)}{n} \right\} \to 0 \quad \text{in probability.}$$

Again it is enough to prove this for $x \in [0,1] \setminus A_k$. And for these x's we can use Lemma 2.2. And as before it is sufficient to show that

$$\frac{k(2kn,x) - k(2kn-1,x)}{n} \to 0 \quad a.e.$$

which follows from Theorem 2.1 since both terms converge to $\frac{h(S)}{h(T)}2k$. We only have to check the difference at point k now. Because of the definition of L, $\frac{h(T)}{h(S)}\frac{L+1}{n} > k$, so the difference at point k cannot be bigger than the maximum of the differences at points $\frac{h(T)}{h(S)}\frac{L}{n}$ and $\frac{h(T)}{h(S)}\frac{L+1}{n}$. We have already covered the point $\frac{h(T)}{h(S)}\frac{L}{n}$, so let us take a look at the other one. If there exists l such that $k(l,x) = L+1$, then the difference is 0. Let us also notice that $k(n_k(x)+1,x) \geq L+1 > L \geq k(n_k(x),x)$, thus if such an l exists, then $l = n_k(x)+1$ else $k(n_k(x)+1,x) > L+1$. In this case

$$\frac{\frac{h(S)}{h(T)} m(k(n_k(x),x),x) - k(n_k(x)+1,x)}{\sigma_1\sqrt{n}} \leq \frac{\frac{h(T)}{h(S)} m(L+1,x) - L+1}{\sigma_1\sqrt{n}} \leq$$

$$\leq \frac{\frac{h(T)}{h(S)} m(k(n_k(x)+1,x),x) - k(n_k(x),x)}{\sigma_1\sqrt{n}}.$$

Hence we have

$$M'_{n,x}(t^n_{n_k(x)}(x)) - \frac{k(n_k(x)+1,x) - k(n_k(x),x)}{\sigma_1\sqrt{n}} \leq M_{n,x}\left(\frac{h(T)}{h(S)}\frac{L+1}{n}\right) \leq$$

$$\leq M'_{n,x}(t^n_{n_k(x)+1}(x)) + \frac{k(n_k(x)+1,x) - k(n_k(x),x)}{\sigma_1\sqrt{n}}.$$

First again let us see that the correcting terms above go to 0. From Lemma 2.3 it follows that for $x \in [0,1] \setminus A_k$ and for $n \geq N$

$$\lambda(x \in [0,1] \colon n_k(x) < 2kn) > 1 - \varepsilon.$$

Then again it is enough to show that the convergence in probability holds on this set. Thus

$$\frac{k(n_k(x)+1,x) - k(n_k(x),x)}{\sqrt{n}} \leq \max_{l=0,\dots,2kn-1} \left\{ \frac{k(l+1,x) - k(l,x)}{\sqrt{n}} \right\} \to 0$$

in probability, which we have already shown. Hence, as before, because of the tightness of the sequence $M'_n(t)$, we only have to prove that

$$\max\{k - t^n_{n_k(x)}(x), t^n_{n_k(x)+1}(x) - k\} \leq t^n_{n_k(x)+1}(x) - t^n_{n_k(x)}(x) \to 0$$

in probability. This is also easy to see:

$$t^n_{n_k(x)+1}(x) - t^n_{n_k(x)}(x) \leq \max_{l=0,\dots,2nk-1} \left\{ \frac{h(T)}{h(S)} \frac{k(l+1,x) - k(l,x)}{n} \right\} \to 0$$

in probability, which we have also already seen. Hence the difference at point k also converges to 0 in probability. This finishes the proof. $\square$

Finally, we present a corollary, which states the central limit theorem for $m(n,x)$, and thus in a way it is a generalization of Faivre's theorem, see ([4]).

Corollary 2.1. *Under the assumptions of Theorem 2.5*

$$\frac{m(n,x) - n\frac{h(T)}{h(S)}}{\sigma_2\sqrt{n}} \Rightarrow \mathcal{N}(0,1),$$

where $\sigma_2 = \sqrt{\left(\frac{h(T)}{h(S)}\right)^3}\sigma_1 = \sqrt{\frac{h(T)}{h(S)^3}}\sigma$ *and* $\Rightarrow$ *is the convergence in law with respect to the probability measure* μ_1*.*

Proof. From Theorem 2.5, we have that

$$M_{n,x}\left(\frac{h(T)}{h(S)}\right) = \frac{\frac{h(S)}{h(T)}m(n,x) - n}{\sigma_1\sqrt{n}} \Rightarrow \mathcal{N}\left(0, \frac{h(T)}{h(S)}\right).$$

Hence if we divide by $\sqrt{\frac{h(T)}{h(S)}}$, we get that

$$\frac{\frac{h(S)}{h(T)}m(n,x) - n}{\sigma_1\sqrt{\frac{h(T)}{h(S)}n}} = \frac{m(n,x) - n\frac{h(T)}{h(S)}}{\sigma_2\sqrt{n}} \Rightarrow \mathcal{N}(0,1).$$

$\square$

Chapter 3

Examples

3.1 CLT-property

We start the section with a theorem which gives the CLT-property for a wide class of transformations (see [1]):

Theorem 3.1. *Let T be a strongly mixing Markov shift, and put*

$$\sigma^2 = \lim_{n\to\infty} \frac{Var[-\log \mu(P_n(x))]}{n},$$

where 'Var' stands for the variance. If $\sigma^2 \neq 0$, then

$$\lim_{n\to\infty} \mu\left(\left\{x\colon \frac{-\log \mu(P_n(x)) - nh(T)}{\sigma\sqrt{n}} \leq u\right\}\right) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx.$$

Now let us take Example 1.4 with $k = 2$ and $p_0 = p$, $p_1 = 1 - p$ and suppose that $p \neq \frac{1}{2}$. We have seen that this is a Markov shift and we have proven that it is strongly mixing. We take the partition with atoms $A_0 = \{y\colon y_0 = 0\}$ and $A_1 = \{y\colon y_0 = 1\}$. The only thing we have to check is that $\lim_{n\to\infty} \frac{Var[-\log \mu(P_n(x))]}{n} \neq 0$. So let us check it. Take $x \in X$, then $P_n(x) = \{y\colon y_0 = x_0, \ldots, y_{n-1} = x_{n-1}\}$, since

$$P_n = \{A\colon A = \{y\colon y_0 = a_0, \ldots, y_{n-1} = a_{n-1}\} \text{ for } (a_0, \ldots, a_{n-1}) \in \{0,1\}^n\}.$$

This means that $\mu(P_n(x)) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$. Then by introducing the random variable Y with $Y = \sum_{i=1}^n x_i$, we have

$$-\log \mu(P_n(x)) = -\log pY - \log{(1-p)}(n - Y),$$

where Y is binomial with parameters n and p. Hence we can easily compute $Var[-\log\mu(P_n(x))]$:

$$Var[-\log\mu(P_n(x))] =$$

$$= \log^2 p Var[Y] + \log^2(1-p)Var[n-Y] + 2\log p\log(1-p)Cov(Y, n-Y).$$

Here we know that $Var[Y] = Var[n-Y] = np(1-p)$, and we can easily compute $Cov(Y, n-Y)$ from the definition:

$$Cov(Y, n-Y) = E(Y(n-Y)) - E(Y)E(n-Y) =$$

$$= E(nY) - E(Y^2) - nE(Y) + E^2(Y) = -Var[Y] = -np(1-p).$$

Thus we have

$$\frac{Var[-\log\mu(P_n(x))]}{n} = p(1-p)(\log(1-p) - \log p)^2 = p(1-p)\log^2\left(\frac{1-p}{p}\right).$$

Hence $\lim_{n\to\infty}\frac{Var[-\log\mu(P_n(x))]}{n} = p(1-p)\log^2\left(\frac{1-p}{p}\right) \neq 0$ if $p \neq \frac{1}{2}$. Therefore the conditions of Theorem 3.1 are satisfied, hence the transformation has the CLT-property.

3.2 0-property

Our first example is Example 1.2. For these transformations even more is true. Take $P = \left\{\left[\frac{k}{r}, \frac{k+1}{r}\right) : k = 0, 1, \ldots, r-1\right\}$, then $-\log\lambda(P_n(x)) = nh(T)$ for every n and for every $x \in [0, 1)$. To see this, first take $x \in [0, 1)$ which is not in the form of $\frac{k}{r}$ for some $k \in \{0, 1, \ldots, r-1\}$. Then T generates the digits of x in the number system with base r: $\lfloor rT^{k-1}x \rfloor$ gives the k-th digit. This means that for every $n \geq 1$

$$\sum_{k=1}^{n}\frac{\lfloor rT^{k-1}x\rfloor}{r^k} \leq x < \sum_{k=1}^{n-1}\frac{\lfloor rT^{k-1}x\rfloor}{r^k} + \frac{\lfloor rT^{n-1}x\rfloor + 1}{r^n}.$$

Since $\left[\sum_{k=1}^{n}\frac{\lfloor rT^{k-1}x\rfloor}{r^k}, \sum_{k=1}^{n-1}\frac{\lfloor rT^{k-1}x\rfloor}{r^k} + \frac{\lfloor rT^{n-1}x\rfloor+1}{r^n}\right) \in P_n$, we have that

$$P_n(x) = \left[\sum_{k=1}^{n}\frac{\lfloor rT^{k-1}x\rfloor}{r^k}, \sum_{k=1}^{n-1}\frac{\lfloor rT^{k-1}x\rfloor}{r^k} + \frac{\lfloor rT^{n-1}x\rfloor + 1}{r^n}\right).$$

Hence $-\log\lambda(P_n(x)) = -\log\frac{1}{r^n} = n\log r$ for almost every $x \in [0, 1)$, which means from the Shannon-McMillan-Breiman theorem that $h(T) = \log r$. Thus T satisfies the 0-property.

Some of the β-transformation as in Example 1.3 also satisfy the property. We take

$$P = \left\{\left[\frac{k}{\beta}, \frac{k+1}{\beta}\right) : k = 0, 1, \ldots, \lfloor\beta\rfloor - 1\right\} \cup \left\{\left[\frac{\lfloor\beta\rfloor}{\beta}, 1\right)\right\}.$$

The big difference between the two types of examples in this section is that in this case the last interval is what is called non-full, meaning that its measure is $1-\frac{\lfloor\beta\rfloor}{\beta}$ which is smaller than that of the others which is $\frac{1}{\beta}$. Nevertheless, we will prove the 0-property for some β-transformation which have something similar to our first example. But first of all, let us look into the partitions generated by the β-transformation, in general.

We will see that what plays a very important role in the study of the length of the elements of the partitions is the error of the expansions of 1. Let us define $T_\beta 1$ with $T_\beta 1 = \beta - \lfloor\beta\rfloor \in [0,1)$, so then we can talk about $T_\beta^i 1$ for $i \geq 0$. This way we can introduce the following notations:

$$a_n = 1 - \sum_{k=1}^{n} \frac{\left\lfloor \beta T_\beta^{k-1} 1 \right\rfloor}{\beta^k}$$

for $n \geq 0$. The proposition below explains why these are so important:

Proposition 3.1. *Let us denote the set of possible lenghts of n-cylinders: $A_n = \bigcup_{A\in P_n} \lambda(A)$. Then*

$$A_n = \left\{ \frac{a_i}{\beta^{n-i}} : i = 0, 1, \ldots, n \right\}$$

.

Proof. We will prove the statement by induction. First for $n = 1$ we have $A_1 = \left\{\frac{1}{\beta}, 1 - \frac{\lfloor\beta\rfloor}{\beta}\right\} = \left\{\frac{a_0}{\beta}, a_1\right\}$. Now let us suppose that the statement is true for some $n \geq 1$, and let us try to prove for $n+1$. First take an arbitrary $A \in P_n$ and look at $T_\beta^{-1}A \cap \left[0, \frac{1}{\beta}\right)$. Since for $x \in \left[0, \frac{1}{\beta}\right)$ $T_\beta x$ is simply βx, this means that $\lambda\left(T_\beta^{-1}A \cap \left[0, \frac{1}{\beta}\right)\right) = \frac{1}{\beta}\lambda(A)$, by using the fact the every $A \in P_n$ is an interval (which is easy to see by induction). Thus $\frac{1}{\beta}A_n = \left\{\frac{a}{\beta} : a \in A_n\right\} \subset A_{n+1}$. We get the same result if we look at $T_\beta^{-1}A \cap \left[\frac{k}{\beta}, \frac{k+1}{\beta}\right)$ for every $k \in \{0, 1, \ldots, \lfloor\beta\rfloor - 1\}$. The only $(n+1)$-cylinders that we still have to check are $T_\beta^{-1}A \cap \left[\frac{\lfloor\beta\rfloor}{\beta}, 1\right)$. Let $A = [a, b)$. Then if $b \leq \beta - \lfloor\beta\rfloor$, then $T_\beta^{-1}A \cap \left[\frac{\lfloor\beta\rfloor}{\beta}, 1\right) = \left[\frac{\lfloor\beta\rfloor}{\beta} + \frac{a}{\beta}, \frac{\lfloor\beta\rfloor}{\beta} + \frac{b}{\beta}\right)$. In this case $\lambda\left(T_\beta^{-1}A \cap \left[\frac{\lfloor\beta\rfloor}{\beta}, 1\right)\right) \in \frac{1}{\beta}A_n$. If $a \geq \beta - \lfloor\beta\rfloor$, then $T_\beta^{-1}A \cap \left[\frac{\lfloor\beta\rfloor}{\beta}, 1\right) = \emptyset$. The interesting case is when $a < \beta - \lfloor\beta\rfloor < b$. Then $T_\beta^{-1}A \cap \left[\frac{\lfloor\beta\rfloor}{\beta}, 1\right) = \left[\frac{\lfloor\beta\rfloor}{\beta} + \frac{a}{\beta}, 1\right)$. So this leads us to the question: What points can be endpoints of an n-cylinder?

Lemma 3.1. *The endpoints of n-cylinders are elements of the following set, denoted by E_n:*

$$E_n = \left\{ 1, \sum_{k=1}^{n} \frac{i_k}{\beta^k} : i_k \in \{0, 1, \ldots, \lfloor\beta\rfloor\} \right\} \cap [0, 1].$$

Proof. $E_1 = \left\{0, \frac{1}{\beta}, \ldots, \frac{\lfloor\beta\rfloor}{\beta}, 1\right\}$. Then it is enough to show that $T_\beta^{-1}(E_n \backslash 1) \subset E_{n+1}$. Let then $e = \sum_{k=1}^{n} \frac{i_k}{\beta^k} < 1$ for some $i_k \in \{0, 1, \ldots, \lfloor\beta\rfloor\}$ for $k = 1, \ldots, n$. In this case, take $j \in \{0, 1, \ldots, \lfloor\beta\rfloor - 1\}$: $T_\beta^{-1} e \cap \left[\frac{j}{\beta}, \frac{j+1}{\beta}\right) = \frac{j}{\beta} + \frac{e}{\beta}$. (It is easy to see that $\frac{j}{\beta} + \frac{e}{\beta}$ is an inverse image of e and it is in the interval $\left[\frac{j}{\beta}, \frac{j+1}{\beta}\right)$.) Now if $e \geq \beta - \lfloor\beta\rfloor$, then $T_\beta^{-1} e \cap \left[\frac{\lfloor\beta\rfloor}{\beta}, 1\right) = \emptyset$ else $T_\beta^{-1} e \cap \left[\frac{\lfloor\beta\rfloor}{\beta}, 1\right) = \frac{\lfloor\beta\rfloor}{\beta} + \frac{e}{\beta}$. In every case $T_\beta^{-1} e = \left\{\frac{j}{\beta} + \frac{e}{\beta} : j \in \{0, 1, \ldots, \lfloor\beta\rfloor\}\right\} \cap [0, 1) \subset \left\{\sum_{k=1}^{n+1} \frac{i_k}{\beta^k} : i_k \in \{0, 1, \ldots, \lfloor\beta\rfloor\}\right\} \cap [0, 1) \subset E_{n+1}$. $\square$

From the proof it follows that any $e < 1$ endpoint of an n-cylinder in the form $e = \sum_{k=1}^{n} \frac{i_k}{\beta^k}$ for some $i_k \in \{0, 1, \ldots, \lfloor\beta\rfloor\}$ is in the inverse image $T^{-j}(\sum_{k=j+1}^{n} \frac{i_k}{\beta^{k-j}})$ for any $n - 1 \geq j \geq 0$.

Now we can finish the proof of Proposition 3.1. The only thing left that we need to check is when we have a cylinder in the form $[a, b)$ where $a < \beta - \lfloor\beta\rfloor < b$. It means that a is the biggest of the possible endpoints of n-cylinders which is smaller than $\beta - \lfloor\beta\rfloor$. We know from the previous lemma that a can be written in the form $\sum_{k=1}^{n} \frac{i_k}{\beta^k}$ for some $i_k \in \{0, 1, \ldots, \lfloor\beta\rfloor\}$. Now we also know that $\frac{\lfloor\beta\rfloor}{\beta} + \frac{a}{\beta}$ is the biggest number that is smaller than 1 amongst those who are in the form $\sum_{k=1}^{n+1} \frac{i_k}{\beta^k}$ for some $i_k \in \{0, 1, \ldots, \lfloor\beta\rfloor\}$. Let us write then

$$\frac{\lfloor\beta\rfloor}{\beta} + \frac{a}{\beta} = \sum_{k=1}^{n+1} \frac{j_k}{\beta^k}.$$

Introduce the notation $b_k = \lfloor\beta T_\beta^{k-1} 1\rfloor$ for the moment. We are going to prove by induction that $j_k = b_k$ for $k = 1, 2, \ldots, n$ for every $n \geq 1$. First let $n = 1$. Then of course $j_1 = \lfloor\beta\rfloor = \lfloor\beta T_\beta^0 1\rfloor = b_1$. Now suppose that we have that $j_k = b_k$ for every $k = 1, 2, \ldots, j$ and for every $n \geq j \geq 1$, and let us try to prove it for $n + 1$. We are going to prove that by contradiction. Suppose that there are some $i_1, i_2, \ldots, i_{n+1}$ such that not all $i_k = b_k$ and that

$$\sum_{k=1}^{n+1} \frac{b_k}{\beta^k} < \sum_{k=1}^{n+1} \frac{i_k}{\beta^k} < 1.$$

Let $j \geq 1$ the smallest integer for which $i_j \neq b_j$ and suppose that $j \leq n$. Then it is also true that

$$\sum_{k=j}^{n+1} \frac{b_k}{\beta^{k-j+1}} < \sum_{k=j}^{n+1} \frac{i_k}{\beta^{k-j+1}}.$$

It follows from the inductive assumptions that $b_j > i_j$. If the opposite were true, then if we look at the case when $n = j$, then we would have a bigger choice then $\sum_{k=1}^{j} \frac{b_k}{\beta^k}$, namely $\sum_{k=1}^{j} \frac{i_k}{\beta^k} < 1$. That is a contradiction. So we have that $b_j \geq i_j + 1$. Thus we have that

$$1 \le b_j - i_j < \sum_{k=j+1}^{n+1} \frac{i_k - b_k}{\beta^{k-j}} \le \sum_{k=j+1}^{n+1} \frac{i_k}{\beta^{k-j}}.$$

From the proof of Lemma 3.1 and the remark after that, if $\sum_{k=1}^{n+1} \frac{i_k}{\beta^k}$ is an endpoint of a cylinder, then it is an inverse image of $T^{-j}(\sum_{k=j+1}^{n+1} \frac{i_k}{\beta^{k-j}})$ for any $n \ge j \ge 0$. But this contradicts the fact that $\sum_{k=j+1}^{n+1} \frac{i_k}{\beta^{k-j}} > 1$.

The only possibility left is that $j = n+1$. Then we cannot use the inductive assumptions but we can use the following lemma, which will be useful later on too:

Lemma 3.2. *For every $x \in [0, 1]$*

$$x = \sum_{k=1}^{n} \frac{\left\lfloor \beta T_\beta^{k-1} x \right\rfloor}{\beta^k} + \frac{T_\beta^n x}{\beta^n}.$$

Proof. Again by induction. For $n = 1$ it follows from the definition of the transformation. Now suppose that it is true for some $n \ge 1$. Take $n+1$. Then $T_\beta^{n+1} x = \beta T_\beta^n x - \lfloor \beta T_\beta^n x \rfloor$. Thus

$$\frac{T_\beta^n x}{\beta^n} = \frac{\left\lfloor \beta T_\beta^n x \right\rfloor}{\beta^{n+1}} + \frac{T_\beta^{n+1} x}{\beta^{n+1}}.$$

□

This lemma shows us that it cannot be that $b_{n+1} < i_{n+1}$ because then if we use the lemma for $x = 1$ we get that

$$1 = \sum_{k=1}^{n+1} \frac{b_k}{\beta^k} + \frac{T^{n+1}1}{\beta^{n+1}} < \sum_{k=1}^{n} \frac{b_k}{\beta^k} + \frac{b_{n+1}+1}{\beta^{n+1}} = \sum_{k=1}^{n+1} \frac{i_k}{\beta^k} < 1,$$

which is again a contradiction.

This proves that $j_k = b_k$ for every $k \ge 1$. And we needed all this to get that

$$\frac{\lfloor \beta \rfloor}{\beta} + \frac{a}{\beta} = \sum_{k=1}^{n+1} \frac{b_k}{\beta^k}$$

and from this we have that the measure of the $(n+1)$-cylinder is: $1 - (\frac{\lfloor \beta \rfloor}{\beta} + \frac{a}{\beta}) = 1 - \sum_{k=1}^{n+1} \frac{b_k}{\beta^k} = a_{n+1}$, and this proves Proposition 3.1. □

From the precedings, we can prove the 0-property for the following class of β-transformation: those for which the digits of the expansion of 1 is periodic, meaning that $\left\lfloor \beta T_\beta^j 1 \right\rfloor = \left\lfloor \beta T_\beta^{j+n} 1 \right\rfloor$ for every $j \ge k \ge 1$ and with $n \ge 1$, i.e.

$$1 = \sum_{i=1}^{k-1} \frac{\left\lfloor \beta T_\beta^{i-1} 1 \right\rfloor}{\beta^i} + \sum_{i=k}^{k+n-1} \frac{\left\lfloor \beta T_\beta^{i-1} 1 \right\rfloor}{\beta^i} \left(\sum_{l=1}^{\infty} \frac{1}{\beta^{ln}} \right).$$

This follows from the fact that for every $x \in [0,1]$: $x = \sum_{k=1}^{\infty} \frac{\lfloor \beta T_\beta^{k-1} x \rfloor}{\beta^k}$.

It is apparent from Proposition 3.1 that the important thing to look at is a_m. Let $m \geq k-1$. Then from the definition of a_m and from the special form of the expansion of 1:

$$a_m = 1 - \sum_{i=1}^{m} \frac{\lfloor \beta T_\beta^{i-1} 1 \rfloor}{\beta^i} = \sum_{i=m+1}^{\infty} \frac{\lfloor \beta T_\beta^{i-1} 1 \rfloor}{\beta^i} = \beta^n \sum_{i=m+n+1}^{\infty} \frac{\lfloor \beta T_\beta^{i-1} 1 \rfloor}{\beta^i} = \beta^n a_{n+m}.$$

This means that the possible $-\log P_m(x) - mh(T_\beta)$ values can be:

$$-\log \frac{a_i}{\beta^{m-i}} - m \log \beta = -\log a_i + (m-i)\log \beta - m \log \beta = -\log a_i - i \log \beta,$$

for $i \in \{0, 1, \ldots, m\}$. But this gives only at most $k+n$ different values: for $i = 0, 1, \ldots, k+n-1$, as for $i \geq k+n$:

$$-\log a_i - i \log \beta = -\log a_{i-n} \frac{1}{\beta^n} - i \log \beta = -\log a_{i-n} - (i-n) \log \beta.$$

So let us take

$$M = \max_{i \in \{0,1,\ldots,k+n-1\}} |-\log a_i - i \log \beta| = \max_{i \in \mathbb{N}} |-\log a_i - i \log \beta|.$$

Now let $\varepsilon > 0$. We can take $N = \left\lceil \frac{M^2}{\varepsilon^2} \right\rceil$. Then for every $m \geq N$ we have $\frac{M}{\sqrt{m}} \leq \varepsilon$. From this it follows that for every $x \in [0,1)$:

$$\left| \frac{-\log \lambda(P_m(x)) - mh(T_\beta)}{\sqrt{m}} \right| \leq \varepsilon,$$

hence the 0-property is satisfied.

Just to have some specific examples: we can take β the golden mean, i.e. $\beta^2 = \beta + 1$. Then $1 = \frac{1}{\beta} + \frac{1}{\beta^2}$, so we can already suspect that this is going to satisfy the condition. In fact, $T_\beta^2 1 = \beta(\beta - \lfloor \beta \rfloor) \bmod 1$, and since $\beta - \lfloor \beta \rfloor = \beta - 1 = \frac{1}{\beta}$, so $\beta(\beta - \lfloor \beta \rfloor) = 1$, hence $T_\beta^2 1 = 0$ and as the 0 is a fixpoint for the transformation, this has the property in question.

Bibliography

[1] G. H. Choe: Computational Ergodic Theory, page 253, Springer (2004).

[2] K. Dajani: Ergodic Theory, Lecture notes, http://www.math.uu.nl/people/dajani/lecturenotes2006.pdf, 2006.

[3] K. Dajani, A. Fieldsteel: Equipartition of interval partitions and an application to number theory, Proc. Amer. Math. Soc. 129 (2001), no. 12, 3453-3460 (electronic).

[4] C. Faivre: A central limit theorem related to decimal and continued fraction expansion, Arch. Math. (Basel) 70 (1998), no. 6, 455-463.