

A constructive law of large numbers with application to countable Markov chains

Peter Gács
Boston University
gacs@bu.edu

December 31, 2010

Abstract Let X_1, X_2, \dots be a sequence of identically distributed, pairwise independent random variables with distribution P . Let the expected value be $\mu < \infty$. Let $S_n = \sum_{i=1}^n X_i$. It is well-known that S_n/n converges to μ almost surely. We show that this convergence is effective in (P, μ) . In particular, if P, μ are computable then the convergence is effective. On the other hand, if the convergence is effective in P then μ is computable from P . The effectiveness of convergence is *detached* in the sense that nothing can be inferred about the speed of convergence in the law of large numbers from the speed of convergence in computing P and μ .

This theorem can be used to show an effective renewal theorem, which then can be used to prove an effective ergodic theorem for countable Markov chains. The last result is a special case of effective ergodic theorems proven by Avigad-Gerhardy-Towsner and Galatolo-Hoyrup-Rojas, but we hope that the direct constructivization of the probability-theory proofs is still useful.

1 Introduction

The paper [7] gives an example of a countable Markov chain with a computable distribution, for which the convergence of relative frequencies to their limit, guaranteed by the ergodic theorem, is provably non-effective. The author has been intrigued by this example, since the Markov chain given there is not ergodic. The present note is devoted to showing that for ergodic Markov chains,

the convergence is constructive. This is a special case of effective ergodic theorems proven in [1] and simplified in [3], but we hope that the direct constructivization of the probability-theory proofs is still useful. Along the way, we explore constructive content of the law of large numbers and some related theorems of probability theory.

In the most frequently used laws of large numbers, in which some higher moments are assumed to exist, speed of convergence is simple, and is an important part of the statement and the proof. But in case of identically distributed, (pairwise) independent variables, the law of large numbers follows already from the existence of the expected value. This is the version we need for countable Markov chains, and in this case, the question of speed of convergence is more subtle.

Let X_1, X_2, \dots be a sequence of identically distributed, pairwise independent random variables with distribution P . Let the expected value be $\mu < \infty$. Let $S_n = \sum_{i=1}^n X_i$. It is well-known that S_n/n converges to μ almost surely. We show that this convergence is effective in (P, μ) . In particular, if P, μ are computable then the convergence is effective. On the other hand, if the convergence is effective in P then μ is computable from P .

To a probabilist, the theorem should sound unnatural, and with good reason. Fixing the computability of P , it said that the existence of a computable speed of convergence in the law of large numbers depends on whether μ is computable. This suggests that a faster way of computing μ will result in faster convergence in the law of large numbers. But this is not so. Even with $P(x)$ concentrated on positive integers, taking binary rational values computable in linear time from x , and even with $\mu = 1$, the convergence in the law of large numbers can be arbitrarily slow.

2 Computable random variables

We assume familiarity with computable probability spaces, see [5].

Notation 2.1. We will denote the expected value with respect to distribution P by \mathbf{E}_P , but if (as in most cases) the distribution P is obvious from the context then we will omit the subscript. The same convention is used for the variance Var_P . ┘

Definition 2.2. A random variable is *computable* if its distribution over \mathbb{R} is a computable measure. A sequence X_1, X_2, \dots of random variables is computable

if all finite joint distributions are uniformly computable. ┘

The following observation seems useful.

Proposition 2.1. *A random variable X is computable if and only if both the distribution function $F(x) = P[X < x]$ of X and the distribution function $G(x)$ of $-X$ are lower semicomputable.*

Proof. The “only if” part is immediate. For the “if” part, by a theorem of Hoyrup-Rojas in [5], the random variable X is computable if and only if all probabilities of the form $P[a < X < b]$ for rational $a < b$ are lower semicomputable. Now

$$P[a < X < b] = P[X < b] - P[X \leq a] = F(b) + G(-a) - 1.$$

□

The expected value of a probability distribution is necessarily lower semicomputable as a function of the distribution P . The following proposition is frequently used.

Proposition 2.2. *Let X be a nonnegative random variable with distribution function $F(x) = P[X < x]$, then $\mathbf{E}_P X = \int_0^\infty (1 - F(x)) dx$.*

In particular, if X has natural number values, then $\mathbf{E}_P X = \sum_{n \geq 1} (1 - F(n))$.

Computability of the distribution does not guarantee computability of the expected value.

Example 2.3 (Computable probability distribution with finite non-computable expected value). It is easy to see that there is a nonnegative sequence $\alpha_1, \alpha_2, \dots$ whose sum is a non-computable number less than 1, each member α_i of which is either 0 or is of the form 2^{-k} for some $k \leq i$. Now let $p_i(\cdot)$ be a nondecreasing sequence of functions defined by the following recursion.

1. $p_0(n) = 0$.
2. If $i > 0$, $\alpha_i = 0$ then $p_i(0) = p_{i-1}(0) + 2^{-i}$, and $p_i(n) = p_{i-1}(n)$ for $n > 0$.
3. If $i > 0$, $\alpha_i = 2^{-k}$ then $p_i(2^{n-k}) = p_{i-1}(2^{i-k}) + 2^{-i}$, and $p_i(n) = p_{i-1}(n)$ for all $n \neq 2^{i-k}$.

Now $\sum_n p_i(n) = \sum_{j=1}^i 2^{-j}$, and $\sum_n n p_i(n) = \sum_{j=1}^i \alpha_j$. In the limit, the desired property is obtained. ┘

3 Effective convergence

Let us define effective versions of the standard convergence notions.

Definition 3.1. Let T, U be computable metric spaces. Let $x_1(t), x_2(t), \dots$ be a sequence of functions on T , with values in U . We say that it *converges effectively* to the function $y(t)$ if there is a function $m(\varepsilon, t)$, $m : [0, 1] \times T \rightarrow [0, \infty]$ *upper semicomputable* on T such that for all t , all $\varepsilon > 0$ and every $n \geq m(\varepsilon, t)$ we have $d(x_n(t), y(t)) \leq \varepsilon$. The function $m(\varepsilon, t)$ will be called the *threshold function*.

If there is a threshold function $m(\varepsilon)$ not dependent on t then we say that $x_n(t)$ converges *effectively uniformly* in t . \lrcorner

Remark 3.1. We could require $m(\varepsilon, t)$ to be computable, instead of upper semicomputable. We could also require it to be integer-valued. However, we could not require it to be both, since computable functions are continuous, and so for example a computable function $\mathbb{R} \rightarrow \mathbb{Z}_+$ would have to be constant. But upper semicomputability is sufficient and comes handy anyway. \lrcorner

It is sometimes more convenient to work with the inverse of $m(\varepsilon, t)$, therefore we introduce the following reformulation, inspired by the notion of “order function” in Schnorr’s text [6].

Definition 3.2. Let T, U be metric spaces. A function $b : \mathbb{Z}_+ \times T \rightarrow \overline{\mathbb{R}}_+$, $(n, t) \mapsto b(n, t)$ is called a *shrinkorder function* if it has the following properties:

- Upper semicomputable.
- $b(n, t) \searrow 0$ for all t .

If the parameter t is a tuple $t = (t_1, t_2)$ then we may write $b(n, t) = b(n, t_1, t_2)$. \lrcorner

The upper semicomputability property in the definition can essentially be replaced with computability:

Proposition 3.2. *Let $b(n, t)$ be a shrinkorder function. Then there is an upper semicomputable function $m : T \rightarrow \mathbb{Z}_+$ and a computable shrinkorder function $b'(n, t)$ with $b(n, t) \leq b'(n, t)$ for all $n \geq m(t)$.*

Proof. The function $\bar{b}(n, t) = 1 \wedge b(n, t)$ is upper semicomputable and uniformly bounded. Let $m(t) = \inf\{n : b(n, t) \leq 1\}$, then $\bar{b}(n, t) = b(n, t)$ for all $n \geq m(t)$. There is a uniformly computable sequence of functions $b_i(n, t) \leq 1$ with $b_i(n, t) \searrow \bar{b}(n, t)$ as $i \rightarrow \infty$, and we can even require $b_i(n, t)$ to be monotonically decreasing in n . Choose $b'(n, t) = b_n(n, t)$. Then $b'(n, t)$ is decreasing in

n by definition. Let us show it converges to 0. For every ε there is an n with $b(n,t) < \varepsilon$. But then also there is an i with $b_i(n,t) < 2\varepsilon$. With $k = \max(i,n)$ we have $b'(k,t) = b_k(k,t) < 2\varepsilon$. \square

The following characterization of effective convergence seems to make the notion more intuitive:

Proposition 3.3. *Let T, U be computable metric spaces. The sequence $x_1(t), x_2(t), \dots$ of functions on T , with values in U converges effectively to the function $y(t)$ if and only if there is a shrinkorder function $b(n,t)$ with $d(x_n(t), y(t)) \leq b(n,t)$ for all n, t .*

Proof. Suppose that there is a shrinkorder function with the desired property. Then the function

$$m(\varepsilon, t) = \inf\{n : b(n, t) < \varepsilon\}$$

is the desired threshold function.

Suppose now that $x_n(t)$ converges to $y(t)$ effectively, with a threshold function $m(\varepsilon, t)$. Then

$$b(n, t) = \inf\{\varepsilon : m(\varepsilon, t) \leq n\}$$

is the desired shrinkorder function. \square

The following observation is useful:

Proposition 3.4. *Let $x_1(t) \leq x_2(t) \leq \dots$ be a sequence of functions from T to \mathbb{R}_+ whose elements are uniformly lower semicomputable, in parameter $t \in T$. Then this sequence converges effectively in (t, y) , in the set*

$$S = \{(t, y) : \lim_n x_n(t) = y\} \subseteq T \times \mathbb{R}.$$

As a special case, if the series z_n is uniformly lower semicomputable and its sum $\sum_n z_n$ is computable then the series converges effectively.

Proof. The function $b(n, t, y) = y - x_n(t)$ satisfies the requirements of a shrinkorder function on the set S . \square

Remark 3.5 (Detachment). The result should not suggest that faster computability of x_n and y implies faster convergence of x_n to y . Given an arbitrary shrinkorder function $b(n)$ there is a sequence x_n of binary rational numbers computable in linear time and increasing monotonically, such that $\lim_n x_n = 1$ but

the $x_n \leq 1 - b(n)$. Indeed, just set $x_n = 1 - b(n)$. Proposition 3.2 shows that we can make $b(n)$ computable, and the proof shows that can even be required to be computable in linear time.

This remark can be extended to a number of other effective convergence results in this paper, including the main result on the law of large numbers. \lrcorner

Definition 3.3 (Effective stochastic convergence). Let X_1, X_2, \dots be a sequence of random variables, with joint probability distribution P (about which we do not assume anything for the moment). We say that X_n *effectively converges to Y in probability, or stochastically*, if there is an upper semicomputable function $m(\delta, \varepsilon)$ with the following property: for all rational $\delta, \varepsilon > 0$ and $n \geq m(\delta, \varepsilon)$ we have $P[|X_n - Y| > \delta] < \varepsilon$.

We say that $X_n \rightarrow Y$ *almost surely, effectively*, if there is an upper semicomputable function $m(\delta, \varepsilon)$ with the following property: for all rational $\delta, \varepsilon > 0$ we have $P[\sup_{n \geq m(\delta, \varepsilon)} |X_n - Y| > \delta] < \varepsilon$.

The function $m(\delta, \varepsilon)$ will be called a *threshold function*. We will also use threshold functions $m(\delta, \varepsilon, t)$ that are computable in some parameter $t \in T$, for example the distribution P . \lrcorner

There is again an intuitive characterization via shrinkorder functions. The following characterization of effective convergence in probability and almost everywhere seems to make the notion more intuitive:

Proposition 3.6. *Let X_1, X_2, \dots be a sequence of random variables. X_n converges to Y in probability, effectively, if and only if there is shrinkorder function $b(n, \varepsilon)$ with the following property for all n :*

$$P[|X_n - Y| > b(n, \varepsilon)] \leq \varepsilon. \quad (1)$$

X_n converges to Y almost surely, effectively, if and only if there is a shrinkorder function $b(n, \varepsilon)$ with the property

$$P[\sup_{k \geq n} |X_k - Y| > b(n, \varepsilon)] \leq \varepsilon.$$

Proof. The proof is the same for both the convergence in probability and the almost sure convergence.

Suppose that there is a shrinkorder function $b(n, \varepsilon)$ with the desired property. Then the function $m(\delta, \varepsilon) = \inf\{n : b(n, \varepsilon) < \delta\}$ is the desired threshold function.

Suppose now that there is a threshold function $m(\delta, \varepsilon)$. Then $b(n, \varepsilon) = \inf\{\delta : m(\delta, \varepsilon) \leq n\}$ is the desired shrinkorder function. \square

The above characterization is asymmetric: it is not clear why in **1**, we did not require $P[|X_n - Y| > \varepsilon] \leq b(n, \varepsilon)$ instead. The following characterization is, on the other hand, symmetric:

Proposition 3.7. *The sequence X_n converges to Y in probability effectively if and only if there are two shrinkorder functions $b_1(n), b_2(n)$ with the property*

$$P[|X_n - Y| > b_1(n)] \leq b_2(n)$$

for all n . Similar characterization holds for almost sure convergence.

Of course, we could add the requirement $b_1(n) = b_2(n)$.

Proof. We will prove the statement for convergence in probability, the proof is the same for the almost sure case. Suppose that the shrinkorder functions $b_1(n), b_2(n)$ exist with the desired properties. Then we can define

$$b(n, \varepsilon) = \begin{cases} b_1(n) & \text{if } b_2(n) < \varepsilon, \\ \infty & \text{otherwise.} \end{cases}$$

Conversely, assume that a shrinkorder function $b(n, \varepsilon)$ exists. Then $b_1(n) = b_2(n) = \inf\{\varepsilon : b(n, \varepsilon) < \varepsilon\}$. Of course, here the infimum of the empty set is ∞ . (If we want we can set $b_2(n) = 1$ instead of ∞ in this case, since $b_2(n)$ is a probability bound.) \square

It is routine to see that, for random variables X_n, Y , if $X_n \rightarrow Y$ almost surely, effectively, then in the product space of all the random variables involved, there is a P -Martin-Löf test such that there is convergence for all elements of the space passing the test. (As shown in [4], there is even a P -test that is a generalization of a Schnorr test (Schnorr tests were only defined for symbol sequences).) On the other hand, the paper [7] shows that the almost sure convergence in Birkhoff's theorem does give rise to a Martin-Löf test even when there is no effective convergence.

Of course, effective almost sure convergence implies effective convergence in probability, but not vice versa. Many simple properties of convergence will also hold for effective almost sure convergence.

Here is a way to imply effective almost sure convergence:

Proposition 3.8. *Let $X_1 \geq X_2 \geq \dots$ be a sequence of nonnegative random variables, with possible values ∞ , too. If $\mathbf{E}X_n \rightarrow 0$ effectively then $X_n \rightarrow 0$ effectively almost surely.*

Proof. There is a shrinkorder function $b_1(n) \geq \mathbf{E}X_n$. The Markov inequality gives $P[X_n > b_1(n)/\varepsilon] \leq \varepsilon$. Using the monotonicity of the sequence X_n , the choice $b(n, \varepsilon) = b_1(n)/\varepsilon$ for a shrinkorder function satisfies the requirement of effective almost sure convergence. \square

The above proposition also holds if the effectivity of the convergence depends uniformly on some parameter, for example the distribution P .

4 The law of large numbers

In what follows we consider effectiveness in the law of large numbers. Effective convergence in the law of large numbers gives a way to compute the expected value:

Proposition 4.1. *Let X_1, X_2, \dots be a sequence of identically distributed random variables with distribution P and expected value μ . Let $S_n = \sum_{i=1}^n X_i$. If S_n/n converges to μ effectively in probability, then μ is computable from P .*

Proof. We postpone writing down the routine proof. \square

Corollary 4.2. *If μ is not computable then even if the distribution P is computable, the convergence to μ in the weak law of large numbers is not effective.*

Our goal is to show the converse:

Theorem 1 (Constructive strong law). *Let X_1, X_2, \dots be a sequence of identically distributed, pairwise independent random variables with distribution P . Let $\mathbf{E}|X| = \mu < \infty$. Let $S_n = \sum_{i=1}^n X_i$.*

It is well-known that S_n/n converges to $\mathbf{E}X$ almost surely. We claim that this convergence is effective in (P, μ) . In particular, if P, μ are computable then the convergence is effective.

Remark 3.5 on the detachment of convergence speeds holds also in this case. Independently of the speed of computability of P and μ , the speed of convergence of S_n/n can be arbitrarily low.

The proof follows [2], constructivizing each step as necessary.

Lemma 4.3. *Suppose that for a sequence Y_1, Y_2, \dots of real-valued random variables with joint distribution P the sequence $\mathbf{E}Y_n$ converges to c and $\sum_n \text{Var} Y_n < \infty$, both effectively in (P, c) (in other words $\sum_{n>k} \text{Var} Y_n \rightarrow 0$ effectively in (P, c) , as $k \rightarrow \infty$). Then $Y_n \rightarrow c$ almost surely, effectively in (P, c) .*

Proof. Let $Z_n = Y_n - \mathbf{E}Y_n$ for each n . It is sufficient to prove $Z_n \rightarrow 0$ almost surely, effectively in (P, c) .

Define the random variable $W = \sum_n Z_n^2$ which can also take the value ∞ . By the Monotone Convergence Theorem

$$\mathbf{E}W = \sum_n \mathbf{E}Z_n^2 = \sum_n \text{Var}Y_n < \infty,$$

where the convergence is effective in (P, c) . The decreasing nonnegative sequence $W_n = \sum_{i \geq n} Z_i^2$ satisfies therefore the requirements of Proposition 3.8, hence $W_n \rightarrow 0$ almost surely, effectively in (P, c) . Since $|Z_n| \leq W_n^{1/2}$ this implies $Z_n \rightarrow 0$ almost surely, effectively in (P, c) . \square

Since we do not assume the existence of higher moments, we will consider truncated versions of our random variables. For this, the following lemmas are used in preparation.

First we elaborate on Proposition 2.2.

Lemma 4.4. *Let X be a nonnegative random variable with distribution P and expected value μ . The series $\sum_n P[X \geq n]$ converges effectively in (P, μ) .*

Proof. Proposition 2.2 can be written as

$$\mathbf{E}X = \sum_{n=1}^{\infty} \int_{n-1}^n P[X \geq t] dx.$$

Proposition 3.4 implies that this series converges effectively. It majorizes the series $\sum_n P[X \geq n]$, term-for-term, so this latter series converges effectively, too. \square

For the rest of the proof, let

$$Y_n = X_n 1_{[|X_n| < n]}, \quad S_n = \sum_{i=1}^n X_i, \quad T_n = \sum_{i=1}^n Y_i.$$

Lemma 4.5. *Let X_1, X_2, \dots be a sequence of identically distributed nonnegative random variables with distribution P and $\mu = \mathbf{E}X < \infty$. Then $P[(\exists k \geq n) Y_k \neq X_k] \rightarrow 0$ effectively in (P, μ) as $n \rightarrow \infty$.*

Proof. This follows from the upper bound

$$P[(\exists k \geq n) Y_k \neq X_k] \leq \sum_{k \geq n} P[X_k \geq k] = \sum_{k \geq n} P[X_1 \geq k]$$

and Lemma 4.4. \square

The following lemma allows us to concentrate on the sequence Y_n :

Lemma 4.6. *With the assumptions of Lemma 4.5, but not requiring nonnegativity:*

$$\begin{aligned} \frac{S_n}{n} - \frac{T_n}{n} &\rightarrow 0 \text{ almost surely,} \\ \frac{\mathbf{E}S_n}{n} - \frac{\mathbf{E}T_n}{n} &\rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$, effectively in (P, μ) .

Proof. Let us prove the first assertion first. For any m let \mathcal{F}_m be the event $[(\exists n \geq m) X_n \neq Y_n]$. By Lemma 4.5 there is a shrinkorder function $b_1(n)$ with $P(\mathcal{F}_m) \leq b_1(m)$. For any m, δ , let us define the event

$$\mathcal{E}_m(\delta) = [|S_m/m - T_m/m| > \delta].$$

The Markov inequality gives $P(\mathcal{E}_m(\delta)) < \delta \mathbf{E}S_m/m = \delta\mu$. For $n \geq m$, if neither $\mathcal{E}_m(\delta)$ nor \mathcal{F}_m hold then

$$\begin{aligned} |S_n - T_n| &= |S_m - T_m| \leq m\delta, \\ \frac{|S_n - T_n|}{n} &\leq m\delta/n \leq \delta. \end{aligned}$$

As we computed, on the other hand,

$$P(\mathcal{E}_m(\delta) \cup \mathcal{F}_m) \leq b_1(m) + \delta\mu.$$

Choosing $\delta = b_1(m)$ gives

$$P[(\exists n \geq m) \frac{|S_n - T_n|}{n} > b_1(m)] \leq b_1(m)(1 + \mu).$$

So we can define the shrinkorder function $b_2(m) = b_1(m)(1 + \mu)$ and finish using Proposition 3.7.

Now for the second assertion. With $\mu_n = \mu - \int_{-n+0}^{n-0} x dP$,

$$\begin{aligned} \mathbf{E}X_n - \mathbf{E}Y_n &= \mu_n, \\ \mathbf{E}S_n - \mathbf{E}T_n &= \frac{\sum_{i=1}^n \mu_i}{n}. \end{aligned}$$

The convergence $\mu_n \rightarrow 0$ is effective in (P, μ) , since μ_n is a shrinkorder function. From here it is routine to show that the convergence $\frac{\sum_{i=1}^n \mu_i}{n} \rightarrow 0$ is also effective in (P, μ) . \square

The following estimate on variances will be used.

Lemma 4.7. *With the assumptions of Lemma 4.5, the sum $\sum_{j>0} \text{Var} Y_j / j^2$ is finite, effectively.*

This lemma exploits identical distribution in a subtle way: its result, along with a uniform effective convergence condition of tails, could replace the condition of identical distribution.

Proof. We will use the estimate for $n > 1$:

$$\sum_{i=n}^{\infty} \frac{1}{i^2} < \sum_{i=n}^{\infty} \frac{1}{(i-1)i} = \sum_{i=n}^{\infty} \frac{1}{i-1} - \frac{1}{i} = \frac{1}{n-1}. \quad (2)$$

Let $F(x) = \mathbf{P}[X < x]$ be the distribution function of X_1 . For $m \geq 1$, using (2):

$$\begin{aligned} \sum_{j \geq m} \text{Var} Y_j / j^2 &\leq \sum_{j \geq m} \mathbf{E} Y_j^2 / j^2 = \sum_{j \geq m} \int_0^j x^2 j^{-2} dF \\ &\leq \int_0^{\infty} x^2 \sum_{j \geq m \vee x} j^{-2} dF = \int_0^m x^2 dF \sum_{j \geq m} j^{-2} + \int_m^{\infty} x^2 \sum_{j \geq x} j^{-2} dF \\ &\leq \frac{1}{m-1} \int_0^m x^2 dF + \int_m^{\infty} \frac{x^2}{x-1} dF \\ &\leq \frac{1}{m-1} \int_0^m x^2 dF + \frac{m}{m-1} \int_m^{\infty} x dF. \end{aligned}$$

The second term converges to 0 effectively in (P, μ) , as $m \rightarrow \infty$ via Proposition 3.4. To bound the first term, we estimate it as

$$\frac{1}{m-1} \int_0^{\sqrt{m}} x^2 dF + \frac{1}{m-1} \int_{\sqrt{m}}^m x^2 dF \leq \frac{\sqrt{m}}{m-1} \mu + \frac{m}{m-1} \int_{\sqrt{m}}^{\infty} i dF.$$

Both terms converge to 0 effectively in (P, μ) as $m \rightarrow \infty$. \square

Now we prove convergence for a subsequence, assuming pairwise independence.

Lemma 4.8. *With the assumptions of Lemma 4.5, assume in addition that the variables X_n are pairwise independent. Fix $c > 1$ and for $m \geq 0$ define $b_m = \lceil c^m \rceil$. Then $T_{b_m} / b_m \rightarrow \mu$ almost surely, effectively in (P, μ) .*

Proof. In order to apply Lemma 4.3, we write for some $k > 1$:

$$\begin{aligned} \sum_{m \geq k} \text{Var} \frac{T_{b_m}}{b_m} &\leq \sum_{m \geq k} c^{-2m} \sum_{j=1}^{b_m} \text{Var} Y_j = \sum_{j \geq 1} \text{Var} Y_j \sum_{m \geq k: c^m > j-1} c^{-2m} \\ &= \sum_{j=1}^k + \sum_{j > k} \leq \frac{c^{-2k}}{1-c^{-2}} \sum_{j=1}^k \text{Var} Y_j + \frac{1}{1-c^{-2}} \sum_{j > k} \frac{\text{Var} Y_j}{(j-1)^2} \\ &\leq \frac{k^2 c^{-2k}}{1-c^{-2}} \sum_{j=1}^{\infty} \frac{\text{Var} Y_j}{j^2} + \frac{1}{1-c^{-2}} \sum_{j > k} \frac{\text{Var} Y_j}{(j-1)^2}, \end{aligned}$$

which converges effectively to 0 with $k \rightarrow \infty$ according to Lemma 4.7. \square

The following lemma completes the proof of Theorem 1.

Lemma 4.9. *Under the assumptions of the preceding lemma, T_n/n converges to μ almost surely, effectively in (P, μ) .*

Proof. Let $c > 1$ and define b_m as in Lemma 4.8. Let $M(n)$ be the smallest m with $n \leq c^m$. Then

$$\frac{T_n}{n} \leq \frac{T_{b_{M(n)}}}{n} \leq c \frac{T_{b_{M(n)}}}{b_{M(n)}}.$$

By Lemma 4.8, the right-hand side converges almost surely, to $c\mu$, effectively in (P, μ) . We similarly get a lower bound converging effectively almost surely to μ/c . Since c can be chosen arbitrarily close to 1, we are done. \square

5 Markov chains

5.1 Renewal theory

Definition 5.1. Let T_0, J_1, J_2, \dots be independent integer random variables in $[0, \infty]$, that is taking possibly also the value ∞ , where T_0 has distribution Q over $\mathbb{Z} \cap [0, \infty)$, and for $i > 0$ the variables $J_i > 0$ are identically distributed with distribution R , and $\mathbf{E}J_i = \mu < \infty$. Define

$$\begin{aligned} T_n &= T_0 + \sum_{i=1}^n J_i, \\ X_m &= \begin{cases} 1 & \text{if } (\exists i) m = T_i, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

The sequence X_0, X_1, \dots will be called a *delayed renewal sequence*. It is called simply a *renewal sequence* if $T_0 = 0$. The sequence is called *recurrent* if $P[J_i < \infty] = 1$, and *positive recurrent* if $\mu < \infty$. \lrcorner

The following lemma is a consequence of the strong law of large numbers:

Lemma 5.1 (Constructive strong law for renewal sequences). *Let X_0, X_1, \dots be a positive recurrent renewal sequence, with $S_n = \sum_{i=0}^n X_i$. Then S_n/n converges to $1/\mu$ almost surely, effectively in (R, μ) .*

Proof. Recall $\mu \geq 1$. For any $c < 1$, and $k = \lfloor cn/\mu \rfloor$:

$$\begin{aligned} S_n/n \leq c/\mu &\Rightarrow S_n \leq k \Leftrightarrow n < T_k \\ &\Leftrightarrow n/k < T_k/k \Rightarrow \mu/c < T_k/k. \end{aligned} \quad (3)$$

Since T_m/m converges to μ almost surely, effectively in (R, μ) , there is a shrink-order sequence $b(m, \varepsilon)$ with $P[(\exists k \geq m) T_k/k > \mu + b(m, \varepsilon)] \leq \varepsilon$. Using this and (3), we can find a shrinkorder sequence for showing effectively almost surely $\liminf S_n/n \geq 1/\mu$. The constructive upper bound on $\limsup_n S_n/n$ is similar. \square

Theorem 2 (Constructive strong law for delayed renewal sequences). *Let X_0, X_1, \dots be a positive recurrent delayed renewal sequence with distribution Q of T_0 and distribution R of $T_{i+1} - T_i$, with $P[T_0 < \infty] = 1$ and $S_n = \sum_{i=0}^n X_i$. Then S_n/n converges to $1/\mu$ almost surely, effectively in (Q, R, μ) .*

Proof. Let P' be the probability distribution of the renewal sequence with $T_0 = 0$. Let $q_n = P[T_0 = n]$. Then the probability of any event \mathcal{E} can be written as

$$P(E) = \sum_{k \geq 0} q_k P(E \mid T_0 = k).$$

In particular,

$$P[S_n = s] = \sum_{k \geq 0} q_k P[S_n = s \mid T_0 = k] = \sum_{k \geq 0} q_k P'[S_{n-k} = s].$$

From the constructive convergence theorem for renewal sequences we have a shrinkorder function $b'(n, \varepsilon)$ with

$$P'[(\exists n \geq m) |S_n/n - \mu| > b(m, \varepsilon)] \leq \varepsilon.$$

Choose $K(\varepsilon) > 0$ such that $\sum_{k>K} q_k < \varepsilon/2$ (this can be done in an upper semi-computable way, since $\sum_k q_k$ converges effectively). Now

$$\frac{n-k}{n} \frac{S_{n-k}}{n-k} = \frac{S_n}{n}.$$

Since for each k the sequence $\frac{S_{n-k}}{n-k}$ converges almost surely effectively to $1/\mu$, so does the sequence S_n/n for each k , under the distribution $P[\cdot | T_0 = k]$, with a uniformly computable shrinkorder function $b_k(n, \varepsilon)$. Let $b(n, \varepsilon) = \bigvee_{k=1}^K b_k(\varepsilon/2, n)$. Then

$$\begin{aligned} P[(\exists n) |S_n/n - 1/\mu| > b(n, \varepsilon)] \\ &\leq P[X_0 > K] + \sum_{k=1}^K q_k P'[(\exists n \geq k) |S_n/n - 1/\mu| > b_k(\varepsilon/2, n)] \\ &\leq \varepsilon/2 + \varepsilon/2. \end{aligned}$$

□

5.2 Application to Markov chains

Let $T(x, y) = \mathbf{P}[X_{n+1} = y | X_n = x]$ be the the transition matrix of a countable Markov chain X_0, X_1, \dots with a countable state space \mathcal{X} . Let P^x be the conditional distribution when starting from x (this is determined by $T(x, y)$), and let $T^n(x, y)$ be the n th power of the matrix T , that is the n -step transition function.

Definition 5.2. For $x, y \in \mathcal{X}$ and a set $B \subseteq \mathcal{X}$, we define

$$\begin{aligned} \pi_{x,B} &= P^x[(\exists n) X_n \in B | X_0 = x], \\ \pi_{x,y} &= \pi_{x,\{y\}}. \end{aligned}$$

We say that state y is *accessible* from state x if $\pi_{x,y} > 0$. A Markov chain is *irreducible* if all states are accessible from each other.

Let γ_x denote the smallest period of the return time distribution, when starting from x . It is known that for an irreducible chain, γ_x is independent of x . The chain is called *aperiodic* if it is 1.

A state x is *transient* if $\pi_{x,x} < 1$, and *recurrent* otherwise. In case of a state x , let

$$m_x$$

denote the expected return time. If it is finite the state is called *positive recurrent*.

┘

The following theorem is known (see for example [2]):

Theorem 3. *Assume that the chain with transition matrix T is irreducible, aperiodic and positive recurrent. Then $(p_x = 1/m_x : x \in \mathcal{X})$ is a probability distribution, and*

$$\lim_{k \rightarrow \infty} T^k(x, y) = p_y.$$

The computability of $T(x, y)$ does not imply the computability of the equilibrium distribution p_x , as the following example shows.

Example 5.2 (Computable transition, non-computable equilibrium). Let us define a Markov chain. The set of states is $\{b_0, b_1, b_2, \dots\}$. Let $T(b_0, b_i) = 2^{-i}$. For $i > 0$ let $T(b_i, b_0) = \alpha_i$, $T(b_i, b_i) = 1 - \alpha_i$ for some $0 < \alpha_i < 1$ to be determined. For $i > 0$ the expected return time from b_0 is $1/\alpha_i$, so the expected return time to b_0 is

$$m_0 = \sum_{i>0} 2^{-i}/\alpha_i.$$

Now we can choose computable α_i in a way that m_0 would still not be computable, similarly to Example 2.3. \square

From now on, we assume that the chain is irreducible, aperiodic and positive recurrent. In order to prove a law of large numbers, fix a state y , and consider the sequence T_0, T_1, \dots of return times when starting from y , and also

$$Y_n = 1_{\{y\}} \circ X_n,$$

the sequence that is 1 if $X_n = y$ and 0 otherwise. This is a delayed renewal sequence, therefore Theorem 2 gives that $\sum_{i=1}^n Y_n/n$ converges to $p_y = 1/m_y$ effectively almost surely relative to $(T(\cdot, \cdot), p_y)$. From this, it is routine to conclude:

Theorem 4 (Computable ergodic theorem for bounded functions). *Let the sequence of random variables X_0, X_1, X_2, \dots be a stationary Markov chain that is irreducible, aperiodic and positive recurrent, with distribution P (this includes not only the transition function $T(x, y)$ but also the equilibrium distribution, that is the distribution of X_0). Then for an arbitrary bounded computable function $f : \mathcal{X} \rightarrow \mathbb{R}$,*

$$\sum_{i=1}^n f(X_n)/n \rightarrow \mathbf{E}f(X_0)$$

almost surely, effectively in P .

Remark 3.5 on the *detachment* of convergence speeds from speed of computability of P is valid in this case, too.

In the case when $f(x)$ is not bounded but $c = \mathbf{E}f(X_0)$ exists, the computability would be in (P, c) , it cannot be just in P , as we have already seen in the case of a sequence of independent random variables.

6 Conclusion

The paper answers some elementary questions about the effectivity of convergence in some limit theorems of probability theory. In the law of large numbers, for random variables with no assumption about second moments, we found that the effectivity of convergence is directly related to the computability of the expected value. Similar results were found already in [1, 3] in the more general setting of stationary processes. But our direct computations may permit a more concrete understanding of the relation between the nature of computability of the expected value and the convergence speed in the law of large numbers.

Bibliography

- [1] Jeremy Avigad, Philipp Gerhardy, and Henry Towsner. Local stability of ergodic averages. *Transactions of the American Mathematical Society*, 362(1):261–288, 2010. 1, 6
- [2] Bert Fristedt and Lawrence Gray. *A Modern Approach to Probability Theory*. Birkhäuser, Boston, 1997. 4, 5.2
- [3] Stefano Galatolo, Mathieu Hoyrup, and Cristobal Rojas. Computing the speed of convergence of ergodic averages and pseudorandom points in computable dynamical systems. In *CCA2010*, pages 1–12, 2010. 1, 6
- [4] Mathieu Hoyrup and Cristóbal Rojas. Doctoral dissertations. Technical report, Ecole Normal Supérieure, Paris, 2007. 3
- [5] Mathieu Hoyrup and Cristóbal Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. *Information and Computation*, 207(7):830–847, 2009. 2, 2
- [6] Claus Peter Schnorr. *Zufälligkeit und Wahrscheinlichkeit*. Springer Verlag, New York, 1971. Lecture Notes in Mathematics. 3

- [7] V. V. V'yugin. Ergodic theorems for individual random sequences. *Theoretical Computer Science*, 207(2):343–361, 1998. [1](#), [3](#)