Effective ergodic theory

Tomasz Steifer [1,2]

[1]Pontificia Universidad Católica de Chile

[2]Institute of Fundamental Technological Research, Polish Academy of Sciences

1 czerwca 2022

The effectivization program

Fix your favorite notion of algorithmic randomness and let $\boldsymbol{P}$ be a probability measure (on the Cantor space).

Theorem

We have $\varphi(\boldsymbol{x})$ for $\boldsymbol{P}$-almost all $\boldsymbol{x}$.

$$\Downarrow \Downarrow \Downarrow$$

Theorem

We have $\varphi(\boldsymbol{x})$ for all $\boldsymbol{P}$-random $\boldsymbol{x}$.

For example:

- Borel-Cantelli lemma.
- Doob's martingale convergence and Levy's law .
- Law of large numbers

Let's consider a specialized case.

Take a sequence of uniformly computable functions $\boldsymbol{f_n} : \{0,1\}^{\mathbb{N}} \mapsto \mathbb{R}$ such that $\boldsymbol{f_n}$ converge almost surely. Does it follows that $\mathbf{lim}_{n\to\infty} \boldsymbol{f_n(\omega)}$ exists for every Martin-Löf random ω?

Short answer: no. By the low basis theorem, there are $\mathbf{\Delta}^0_2$ 1-random points. One way of looking at $\mathbf{\Delta}^0_2$ is as follows. If a point ω is $\mathbf{\Delta}^0_2$ then there exists a sequence of uniformly computable $\boldsymbol{g_n} : \{0,1\}^{\mathbb{N}} \mapsto \{0,1\}$ such that ω is the only point for which $\boldsymbol{g_n(\omega) = 1}$ for infinitely many $\boldsymbol{n}$ (essentially Gold).

Let's consider a specialized case.

Take a sequence of uniformly computable functions $\boldsymbol{f_n} : \{0,1\}^{\mathbb{N}} \mapsto \mathbb{R}$ such that $\boldsymbol{f_n}$ converge almost surely. Does it follows that $\mathbf{lim}_{n\to\infty} \boldsymbol{f_n(\omega)}$ exists for every Martin-Löf random $\boldsymbol{\omega}$?

When we can have 'yes'? A simple case: when $\boldsymbol{f_n}$ converges to some $\boldsymbol{f}$ in an effective way, i.e., there exists a computable function $\boldsymbol{g(n,\delta)}$ such that $\boldsymbol{P(|f_n - f| > \delta) < g(n,\delta)}$ and $\boldsymbol{lim_{n\to\infty} g(n,\delta) =} 0$ fastly for each rational $\boldsymbol{\delta}$. In that case, the effective bounds on the probabilities give us automatically a Martin-Löf test.

As an example, consider the law of large numbers (LLN) for Bernoulli (i.i.d) measures.

Theorem (LLN)

Let P be a Bernoulli measure with parameter p. Then for P-almost every ω we have

$$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{1}{n}\omega_i = p.$$

The effective convergence for LLN is given by the Hoeffding inequality:

Theorem (Hoeffding's inequality)

Let $X_1, \ldots, X_n$ be independent binary random variables with the probability distribution μ. Then

$$\mu\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) > t\right) \leq e^{-2nt^2}.$$

Consequently, we get the effective version of the law of large numbers:

Theorem (LLN)

Let $\boldsymbol{P}$ be a Bernoulli measure with parameter $\boldsymbol{p}$. Then every Martin-Löf random $\boldsymbol{\omega}$ we have

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{n}\omega_i = p.$$

Note that this is really just a toy example as for LLN we don't need that much! LLN is already satisfied by every Church random (by definition).

A natural generalization of the LLN is given by several ergodic theorems. Let $\boldsymbol{T}$ be a transformation from $\{0,1\}^{\mathbb{N}}$ to $\{0,1\}^{\mathbb{N}}$.

We say that $\boldsymbol{T}$ is $\boldsymbol{P}$-measure preserving if $\boldsymbol{T}$ is measurable and if $\boldsymbol{P(A) = P(T^{-1}(A))}$ for each Borel sets $\boldsymbol{A}$ (in which case we also say that $\boldsymbol{P}$ is invariant or stationary with respect to $\boldsymbol{T}$). A $\boldsymbol{P}$-measure preserving transformation is called ergodic if for each Borel set $\boldsymbol{A}$ such that $\boldsymbol{A = T^{-1}(A)}$ we have either $\boldsymbol{P(A)} = 1$ or $\boldsymbol{P(A)} = 0$ (in which case we sometimes say that $\boldsymbol{P}$ is ergodic)

Theorem (Birkhoff ergodic theorem)

For a stationary measure $\boldsymbol{P}$ and a random variable $\boldsymbol{G}$ such that $\mathrm{E}\,|\boldsymbol{G}| < \infty$, $\boldsymbol{P}$-almost surely

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} G \circ T^{i} \tag{1}$$

*exists and is equal $\int$ **GdP** if T is ergodic.*

Is the convergence in Birkhoff's ergodic theorem effective? Vyugin showed that it is not.

Theorem (Vyugin)

There exists a computable probability measure $\boldsymbol{P}$ and computable $\boldsymbol{P}$-measure preserving transformation $\boldsymbol{T}$ such that the convergence of $\sum_{\boldsymbol{i}}^{\boldsymbol{n}} \omega_{\boldsymbol{i}}/\boldsymbol{n}$ is not effective.

This already tells that effective theorems might be tricky to prove.

A constructive proof of Birkhoff's ergodic theorem was given by Bishop and then used by Vyugin to prove the following effective version:

Theorem (effective Birkhoff ergodic theorem I)

Let $\boldsymbol{P}$ be a computable probability measure stationary and $\boldsymbol{P}$-measure preserving computable transformation $\boldsymbol{T}$ and let $\boldsymbol{f}$ be an integrable computable function from $\{0,1\}^{\mathbb{N}}$ to $\mathbb{R}$. For every Martin-Löf random point $\boldsymbol{\omega}$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(\omega)) \tag{2}$$

exists and is equal $\int \boldsymbol{fdP}$ if $\boldsymbol{T}$ is also ergodic.

Can we get more, for instance, lower semicomputable functions $\boldsymbol{f}$?

One can also ask if some version of Birkhoff's ergodic theorem holds for Schnorr random and the answer was already provided by Gacs, Hoyrup and Rojas.

Theorem (effective Birkhoff ergodic theorem II)

Let P be a computable probability measure stationary and ergodic with respect to computable transformation T and let f be an integrable computable function from $\{0,1\}^{\mathbb{N}}$ to $\mathbb{R}$. For every Schnorr random point ω

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(\omega)) = \int f dP. \quad (3)$$

The following strengthening was proven independently by Bienvenu-Day-Hoyrup-Mezhirov-Shen and Franklin-Greenberg-Miller-Ng.

Theorem (effective Birkhoff ergodic theorem III)

Let P be a computable probability measure stationary and ergodic with respect to computable transformation T and let f be an integrable lower semicomputable function from $\{0,1\}^{\mathbb{N}}$ to $\mathbb{R}$. For every Martin-Löf random point ω

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(\omega)) = \int f dP. \tag{4}$$

Seeing that Martin-Löf random sequences satisfy a version of Birkhoff's ergodic theorem, we might wonder if we can get a converse result.

Theorem (Bienvenu-Day-Hoyrup-Mezhirov-Shen)

If ω is not Martin-Löf random, then there exists a computable ergodic transformation T and a lower semicomputable integrable function f such that $\sum_{i=1}^{n} f(T^i(\omega))/n$ diverges.

Theorem (Franklin-Towsner)

If ω is not Martin-Löf random, then there exists a computable P-measure preserving transformation T and a computable integrable function f such that $\sum_{i=1}^{n} f(T^i(\omega))/n$ diverges.

By a result of Gacs-Hoyrup-Rojas the analogous fact holds for Schnorr randomness and ergodic transformations (historically, this result is even older than the ones for Martin-Löf randomness).

Theorem

If ω is not Schnorr random, then there exists a computable $\boldsymbol{P}$-measure preserving ergodic transformation $\boldsymbol{T}$ and a computable integrable function $\boldsymbol{f}$ such that $\sum_{i=1}^{n} f(T^i(\omega))/n$ diverges.

This seems to be an open problem:

What notion of randomness is characterized via convergence of ergodic averages for measure-preserving (possibly non-ergodic) computable transformations and lower semicomputable functions?

Source coding

Let's take a short break from ergodic theorems. In what follows we restrict our attention to a single computable transformation, i.e., the shift. Stationary = shift-invariant. Recall the entropy rate of the canonical process $\boldsymbol{X}$ (where $\boldsymbol{X_i(\omega) = \omega_i}$).

$$\boldsymbol{h_P := \lim_{n\to\infty} \frac{1}{n} \operatorname{E}\left[-\log P(X_1^n)\right] = \lim_{k\to\infty} \operatorname{E}\left[-\log P(X_{k+1}|X_1^k)\right]}. \tag{5}$$

Theorem (Shannon-McMillan-Breiman theorem)

For a stationary ergodic probability measure $\boldsymbol{P}$, almost surely we have

$$\boldsymbol{\lim_{n\to\infty} \frac{1}{n}\left[-\log P(X_1^n)\right] = \lim_{n\to\infty} \frac{1}{n} \operatorname{E}\left[-\log P(X_1^n)\right]}. \tag{6}$$

Source coding

In coding theory, one is interesting in functions acting on prefixes, namely, codes, whose lengths converge to the entropy rate almost surely on every ergodic measure (hence, the name: universal codes). As an example, one can consider prefix-free Kolmogorov complexity K (computable alternatives include Lempel-Ziv, PPM measure etc.).

Theorem (Levin)

For a stationary ergodic probability measure P *we have almost surely*

$$\lim_{n\to\infty} \frac{K(X_1^n)}{n} = h_P.$$

Note that the left side is both the effective Hausdorff and effective packing dimension.

Hoyrup proved an effective theorem combining both SMB theorem and Levin's theorem. It is worth noting that this theorem was also implicit in the paper by Hochman, who proved a very nice upcrossing inequality for subadditive processes.

Theorem (Hoyrup, independently Hochman)

For a stationary ergodic probability measure $\boldsymbol{P}$ and every Martin-Löf random point ω

$$\lim_{n\to\infty} \frac{1}{n}\left[-\log P(\omega_1^n)\right] = \lim_{n\to\infty} \frac{K(\omega_1^n)}{n} = h_P.$$

Ergodic theorems

Hoyrup's proof uses an effective version of Breiman's ergodic theorem, which is a generalization of Birkhoff's. Firstly, recall that a stationary probability measure on $\{0,1\}^{\mathbb{N}}$ extends uniquely to a stationary probability measure on $\{0,1\}^{\mathbb{Z}}$. Definition of Martin-Löf randomness extends to this case in a straigthforward way. For simplicity, we state the effective Breiman's theorem for the shift only, but the theorem extends to arbitrary computable transformations.

Ergodic theorems

Theorem (Dębowski-S.)

Let $g : \{0,1\}^{<\mathbb{N}} \to \mathbb{R}^+$ *be a computable function with* $\lim_{n\to\infty} g(X_0^n)$ *existing almost surely and* $\int (\sup_i |g(X_0^i)|) dP < \infty$. *Then for every Martin-Löf random sequence* $\omega \in \{0,1\}^{\mathbb{Z}}$,

$$\limsup_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} g(\omega_0^i) \le \int (\limsup_{k\to\infty} g(X_{-k}^{-1})) dP,$$

$$\liminf_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} g(\omega_0^i) \ge \int (\liminf_{k\to\infty} g(X_{-k}^{-1})) dP.$$

Note that one need to have an effective Birkhoff's ergodic theorem for lower semicomputable functions to get Breiman's theorem.

An application

Having an effective Breiman's theorem we can now say a bit about universal prediction for ergodic random points. The following result is known in the non-effective setting (but an effective version holds as well).

Theorem (Ornstein)

There exist uniformly computable $\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots$ such that for every two-sided infinite process $\boldsymbol{X} = \ldots \boldsymbol{X}_{-1}, \boldsymbol{X}_0, \boldsymbol{X}_1 \ldots$ with the stationary probability distribution $\boldsymbol{\mu}$,

$$\lim_{n\to\infty} g_n(\omega_{-n}^{-1}) = \mu(X_0 = 1 | X_{-\infty}^{-1} = \omega_{-\infty}^{-1})$$

for $\boldsymbol{\mu}$-almost every sequence $\boldsymbol{\omega}$.

Universal prediction

Using the effective version of Breiman's theorem we get:

Theorem

There exists a computable function $f : \{0,1\}^{<\mathbb{N}} \mapsto \mathbb{R}$ *such that for every stationary ergodic measure* P *and for every Martin-Löf random point* ω

$$\lim_{n\to\infty} \sum_{i=0}^{n} |f(\omega_0^i) - P(\omega_0^i)| = 0.$$

Convergence in Cesaro averages is as good as we can hope for. Furthermore, since the proofs of ergodic theorems relativize, we can get the above result for arbitrary uncomputable measures (with a single computable estimator!).

Ergodic decomposition

A fundamental theorem in ergodic theory tells us that every stationary probability measure is actually a barycenter of a probability measure on ergodic measures. To put it more intuitively, sampling a stationary measure may be seen as choosing randomly an ergodic measure, a then sampling the selected ergodic measure.

Theorem (ergodic decomposition)

For every stationary measure $\boldsymbol{P}$ there exists a unique probability measure $\boldsymbol{m_P}$ on probability measures such that $\boldsymbol{m_P}$ is concentrated on ergodic measures and $\boldsymbol{P}$ is a barycenter of $\boldsymbol{m}$, namely, for a measurable $\boldsymbol{A}$ we have $\boldsymbol{P(A) = \int Q(A) dm_P(Q).}$ We call $\boldsymbol{m_P}$ the Choquet measure associated with $\boldsymbol{P}$.

Following Hoyrup, we may ask, how effective is the ergodic decomposition?

Ergodic decomposition

If $\boldsymbol{m_P}$ is computable then we obtain

Theorem (Hoyrup)

Let $\boldsymbol{m_p}$ be a computable measure and $\boldsymbol{P}$ be its barycenter. Then $\boldsymbol{x} \in \{0,1\}^{\mathbb{N}}$ is $\boldsymbol{P}$ random if and only if it is $\boldsymbol{Q}$-random for some $\boldsymbol{m_p}$-random $\boldsymbol{Q}$.

However it already follows from a counter-example by Vyugin that

Theorem (Vyugin)

There exists a computable measure $\boldsymbol{P}$ with uncomputable Choquet measure.

Ergodic decomposition

More recently (2021), Coronel, Frank, Hoyrup and Rojas proved the following.

Theorem

There exists a computable dynamical system with exactly two ergodic measures, none of which is computable.