The Ergodic Theorem and randomness

Peter Gács

Department of Computer Science
Boston University

March 19, 2008

Introduction

I will present some results of V. V. V'yugin that I find interesting. The motivation is that I wanted to present a topic here that is connected with the theory of randomness but at the same time relates to interests of the audience. I assume that ergodic theory falls into the latter category.

The train of thought

1. In the theory of randomness, random sequences are defined as those that do not fall into any “constructive null sets” (see later).
2. Such sequences can be expected to satisfy the “almost every”-type laws of probability theory. Why? Since the proofs of those laws generally enclose the violating sequences into a constructive null set.
3. Example: the strong law of large numbers. The proofs also yields a speed of convergence.
4. In the proof of the Ergodic Theorem, no speed of convergence is obtained. V’yugin’s Theorem 1: sometimes, there is no computable speed of convergence here, even “in probability”.
5. V’yugin’s Theorem 2: the proof of the Ergodic Theorem can still be constructivized, showing that all random sequences obey the Ergodic Theorem.

Content

- The non-effective convergence.
- Notions of randomness.
- Constructive content of the Ergodic Theorem.
- A characterization of randomness via complexity.

I assume that the audience is not expert in the theory of computability, but the extent to which it is used here should be understandable. Let Σ be an alphabet and Σ^* the set of strings in this alphabet.
We will work with an intuitive notion of algorithms. The following is known. Fix some computer (a "Turing machine", but is OK to think of an ordinary computer). It defines a function

$$A(i,x)$$

as follows: we start the computer with program i and input x. If it stops with some output y we set $A(i,x) = y$, else $A(i,x)$ is undefined.
We call $A(i,x)$ a universal computable function since it is known that for every function $f : \Sigma^* \to \Sigma^*$ computable with the help of any algorithm there is a computer program i such that $f(x) = A(i,x)$ for all x.

Consider a measure μ over the set $\Sigma^{\mathbb{N}}$ of sequences in the alphabet Σ. It is determined by the values

$$\mu(\boldsymbol{x}) := \mu(\boldsymbol{x}\Sigma^{\mathbb{N}})$$

for all segments $\boldsymbol{x} \in \Sigma^*$.

Definition

Such a measure μ is called computable if there is a computable function $\mu(\boldsymbol{x}, \varepsilon)$ with rational values such that for all rational ε we have

$$|\mu(\boldsymbol{x}) - \mu(\boldsymbol{x}, \varepsilon)| < \varepsilon.$$

Let $X_1, X_2, \dots$ be a sequence of random variables and Y a random variable. As usual, we say $X_n \to Y$ in probability if $\mathbf{P}\left[\, |X_n - Y| > \delta \,\right] \to 0$ for all δ.

Definition

The convergence above is effective if there is a computable integer-valued function $m(\delta, \varepsilon)$ such that for all rational $\varepsilon, \delta > 0$ and all $n, n' > m(\delta, \varepsilon)$ we have

$$\mathbf{P}\left[\, |X_n - X_{n'}| > \delta \,\right] < \varepsilon.$$

It is easy to check that, for example, Chebyshev's inequality provides algorithmically effective convergence in the Law of Large Numbers (where applicable).

The non-effectiveness theorem

Consider a measure over the set of infinite 0-1 sequences giving rise to the stationary sequence of random variables $X_1, X_2, \dots$, with $S_n = \sum_{i=1}^{n} X_i$. By the (weak) Ergodic Theorem there is a random variable Y with the property that $S_n/n \to Y$ in probability.

Theorem

There is a computable measure of the above form such that the above convergence in probability is not effective.

We will do a little better, showing that there is not even a computable function $m(\varepsilon)$ with $\mathbf{P}\left[\,|S_n/n - S_{n'}/n'| > \frac{1}{4}\,\right] < \varepsilon$ for all $n, n' \geqslant m(\varepsilon)$.

We will define the stationary measure μ as the mixture

$$\mu = \sum_{i=1}^{\infty} 2^{-i} P_i,$$

with the help of the auxiliary processes P_i as follows.

- Let $K(i, \varepsilon)$ be a universal computable function (for rational argument ε).
- Let $t(i, \varepsilon)$ be the number of steps for the algorithm to compute $K(i, \varepsilon)$ (∞ if the algorithm does not terminate). We can assume $t(i, \varepsilon) \geqslant \max(K(i, \varepsilon), 10)$. Let $k(i) = t(i, 2^{-(i+1)})$.
- Let P_i be a simple two-state Markov process with states 0,1, with $P_i[X_1 = 0] = P_i[X_1 = 1] = \frac{1}{2}$, and with the following transition probabilities:

$$\mathbf{P}\left[\, X_{s+1} \neq X_s \,\right] = \alpha_i = 2^{-k(i)}.$$

The process is designed to thwart every possible computable convergence bound $m(\varepsilon)$. Assume $m(\varepsilon) = K(i, \varepsilon)$, then $\alpha_i = 2^{k(i)} > 0$, hence the Markov chain P_i satisfies

$$P_i\left[|S_i/n - \tfrac{1}{2}| < 0.1\right] \to 1.$$

Further, we have

$$\begin{aligned} P_i(0^{k(i)}) = P_i(1^{k(i)}) &= \tfrac{1}{2}(1 - \alpha_i)^{k(i)-1} > \tfrac{1}{2}(1 - (k(i) - 1)\alpha_i) \\ &= \tfrac{1}{2}(1 - (k(i) - 1)2^{-k(i)}) > \tfrac{2}{5} \end{aligned}$$

since $k(i) > 10$. Hence for all sufficiently large n we have

$$\begin{aligned} P_i\left[|S_{k(i)}/k(i) - S_n/n| > \tfrac{1}{4}\right] &\geqslant P_i\left[S_{k(i)}/k(i) \in \{0, 1\}\right] > \tfrac{4}{5}, \\ \mu\left[|S_{k(i)}/k(i) - S_n/n| > \tfrac{1}{4}\right] &\geqslant 2^{-i} P_i\left[|S_{k(i)}/k(i) - S_n/n| > \tfrac{1}{4}\right] > 2^{-(i+1)}. \end{aligned}$$

But $k(i) = t(i, 2^{-(i+1)}) \geqslant m(2^{-(i+1)})$, so $m(2^{-(i+1)})$ fails.

Open question

- This example is not ergodic. Is there an ergodic example?
- This example is a countable Markov chain. Is there an example that is an ergodic Markov chain? (If not, this would mean that every computable ergodic Markov chain has a computable convergence speed in the law of large numbers.)

Randomness

Let $\Omega = \Sigma^{\mathbb{N}}$ be the space of infinite sequences.

Definition

A cylinder is of the form $\Gamma_{\boldsymbol{x}} = \boldsymbol{x}\Sigma^{\mathbb{N}}$ for some $\boldsymbol{x} \in \Sigma^*$.

- An open set is the union of cylinders.
- It is constructively open if it is $\bigcup_i \Gamma_{\boldsymbol{x}_i}$ for a computable sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots$.
- A sequence of sets $G_1, G_2, \dots$ is uniformly constructively open if there is a recursive function $\boldsymbol{x}(i,j)$ such that $G_i = \bigcup_j \Gamma_{\boldsymbol{x}(i,j)}$.

Let $\mathscr{B}$ be the set of Borel sets generated by these open sets. Then $(\Omega, \mathscr{B})$ is a measureable space. Let μ be a computable probability measure on $(\Omega, \mathscr{B})$, then we have a measure space $(\Omega, \mathscr{B}, \mu)$.

If N is a set of measure 0 then for all $\varepsilon > 0$, it can be covered by an open set G_ε with $\mu(G_\varepsilon) < \varepsilon$.

Definition

The set N is a constructive nullset if for rational ε the above sets G_ε can be chosen to be a sequence of uniformly constructively open sets.

There is only a countable number of constructive nullsets.

Definition

An element ω is random if it is not in any constructive set of measure 0.

Let us introduce a useful equivalent definition.

Definition

A function $f : \Sigma^* \to \mathbb{R}$ is lower semicomputable if the set

$$\{\, \omega : f(\omega) > r \,\}$$

is a constructive open set, uniformly in the rational number r.

A function $t : \Omega \to \mathbb{R}_+ \cup \infty$ is a payoff test with respect to measure μ if

1. It is lower semicomputable.
2. We have $\int t(\omega) d\mu \leqslant 1$.

Proposition

An element ω is random if and only if for every payoff test t we have $t(\omega) < \infty$.

We can imagine the function $t(\omega)$ as the payoff function of a fair bet against ω. We pay 1 dollar for the game, and ω pays us $t(\omega)$. The bet is fair since the expected bet is $\int t(\omega) d\mu \leqslant 1$. The bet is lower semicomputable, allowing to increase our win as we discover new and new "regularities" in ω.
The element is random if we cannot win an infinite amount against it.

The ergodic theorem

Definition

A function $f : \Sigma^{\mathbb{N}} \to \Sigma^{\mathbb{N}}$ is a computable transformation if there is a program i for our machine that, reading sequentially the symbols of any input sequence $s_1 s_2 \dots$, computes and outputs gradually more-and-more symbols of an output sequence $t_1 t_2 \dots = f(s_1 s_2 \dots)$.

Note that a computable transformation is always continuous in the topology of $\Sigma^{\mathbb{N}}$.

Let T be computable measure-preserving transformation over $(\Omega, \mathscr{B}, \mu)$ with $\Omega = \Sigma^{\mathbb{N}}$, and let $f(\omega)$ be an integrable function with $S_n(\omega) = \sum_{i=0}^{n-1} f(T^i\omega)$.

Theorem (Constructive Ergodic Theorem)

For all random ω we have

$$S_n(\omega)/n \to \tilde{f}(\omega)$$

for a certain integrable $\tilde{f}$.

For the proof, we define a payoff test for the set of points ω for which $S_n(\omega)/n$ does not converge. Note that in this case there are rational α, β with

$$\liminf_n S_n(\omega)/n < \alpha < \beta < \limsup_n S_n(\omega)/n.$$

So there are infinitely many u, v with

$$S_u(\omega) - u\alpha < 0 < S_v(\omega) - v\beta.$$

Let

$$\begin{aligned} u \overset{\omega}{\rightarrow} v \Leftrightarrow v \overset{\omega}{\leftarrow} u &\Leftrightarrow S_u(\omega) - u\alpha < S_v(\omega) - v\beta, \\ (-1) \overset{\omega}{\rightarrow} v \qquad &\Leftrightarrow 0 < S_v(\omega) - v\beta. \end{aligned}$$

A sequence $s = (u_1, v_1, \dots, u_N, v_N)$ is n-admissible if

$$-1 \leqslant u_1 < v_1 \leqslant u_2 < v_2 \leqslant \dots \leqslant u_N < v_N \leqslant n.$$

We denote $|s| = N$. Let
$\sigma_n(\omega, \alpha, \beta) = \max\{N : u_1 \overset{\omega}{\rightarrow} v_1 \overset{\omega}{\leftarrow} u_2 \overset{\omega}{\rightarrow} v_2 \overset{\omega}{\leftarrow} \cdots \overset{\omega}{\rightarrow} v_N\}$ where $(u_1, v_1, \dots, u_N, v_N)$ is running over all n-admissible sequences.

It is easy to check that $\sigma_n(\omega, \alpha, \beta)$ is lower semicomputable, and so is

$$\sigma(\omega, \alpha, \beta) = \sup_n \sigma_n(\omega, \alpha, \beta).$$

The tough part of the proof is to show

$$\int \sigma(\omega, \alpha, \beta) d\mu < C(\alpha, \beta)$$

for a constant $C(\alpha, \beta)$. Then we can construct a payoff test $t(\omega)$ combined from all the $\sigma(\omega, \alpha, \beta)$ (with constant weights).
If $S_n(\omega)/n$ does not converge then $\sigma(\omega, \alpha, \beta) = \infty$ for an appropriate α, β, so $t(\omega) = \infty$.
Summarizing: we do not get a speed of convergence, but we get an effective probabilistic bound on the the maximum number of oscillations through any interval $[\alpha, \beta]$.

To estimate $\sigma_n(\omega, \alpha, \beta)$ we introduce a quantity:

Definition

The oscillation cost of an admissible sequence $d = (s_1, t_1, \dots, s_N, t_N)$ is

$$S(d, \omega) = \sum_{j=1}^{N} \big[(S_{v_i} - v_i\beta) - (S_{u_i} - u_i\alpha)\big].$$

The following lemma brings us close:

Lemma (Cost shift)

For every n-admissible sequence q there is an n-admissible sequence r with

$$S(q, \omega) \leqslant S(r, T\omega) + |f(\omega) - \alpha|^{+} - (\beta - \alpha)\sigma_n(\omega, \alpha, \beta).$$

Let us apply this lemma. With $\lambda_n(\omega) = \sup\{\, S(d,\omega) : d \text{ is } n\text{-admissible}\,\}$, we get:

$$\begin{aligned}
\lambda_n(\omega) &\leqslant \lambda_n(T\omega) + |f(\omega) - \alpha|^+ - (\beta - \alpha)\sigma_n(\omega, \alpha, \beta),\\
(\beta - \alpha)\sigma_n(\omega, \alpha, \beta) &\leqslant \lambda_n(T\omega) - \lambda_n(T\omega) + |f(\omega) - \alpha|^+,\\
(\beta - \alpha)\int \sigma_n(\omega, \alpha, \beta) d\mu &\leqslant \int |f(\omega) - \alpha|^+ d\mu,\\
\int \sigma(\omega, \alpha, \beta) d\mu &\leqslant (\beta - \alpha)^{-1} \int |f(\omega) - \alpha|^+ d\mu,
\end{aligned}$$

and we will be done.

To prove the cost shift lemma, the following lemma is used:

Lemma (Combinatorial)

For a given ω and any n-admissible sequence q there is an n-admissible sequence d with $S(d,\omega) \geqslant S(q,\omega)$ and $|d| = \sigma_n(\omega,\alpha,\beta)$.

Its proof would take some work. Also, define the shift for an admissible sequence $d = (s_1, t_1, \dots, s_m, t_m)$:

$$d' = \begin{cases} (s_1 - 1, t_1 - 1, \dots, s_m - 1, t_m - 1) & \text{if } s_1 \geqslant 0, \\ (-1, t_1 - 1, \dots, s_m - 1, t_m - 1) & \text{if } s_1 = -1, t_1 > 0, \\ (s_2 - 1, t_2 - 1, \dots, s_m - 1, t_m - 1) & \text{otherwise.} \end{cases}$$

The following can be verified directly:

$$S(d,\omega) \leqslant S(d', T\omega) + |f(\omega) - \alpha|^+ - (\beta - \alpha)m.$$

Using the combinatorial lemma, for all n-admissible q there is an n-admissible d with with $S(d,\omega) \geqslant S(q,\omega)$ and $|d| = \sigma_n(\omega,\alpha,\beta)$. Applying it:

$$S(q,\omega) \leqslant S(d,\omega) \leqslant S(d',T\omega) + |f(\omega)-\alpha|^+ - (\beta-\alpha)\sigma_n(\omega,\alpha,\beta),$$

which proves the cost shift lemma.

History

Birkhoff → Bishop → V'yugin. (With some credit to Lambalgen.)

Randomness and complexity

It turns out that there is one payoff test function encompassing all:

Proposition (Universal test)

Let us be given a computable measure over $(\Omega, \mathcal{B})$ where $\Omega = \Sigma^{\mathbb{N}}$. Then there is a universal payoff test $u(\omega)$ for μ in the sense that for every other payoff test $t(\omega)$ there is a c_t such that for all ω we have

$$c_t u(\omega) \geqslant t(\omega).$$

From now on, we fix a universal payoff test $u(\omega)$. In order to understand what it measures, we define the notion of description complexity.

Definition

A (partial) computable function $A : \{0,1\}^* \to \Sigma^*$ is called an interpreter if its domain of definition is prefix-free: if $A(\boldsymbol{p})$ and $A(\boldsymbol{q})$ are both defined then $\boldsymbol{p}$ is not the prefix of $\boldsymbol{q}$. If $A(\cdot)$ is an interpreter then we define the description complexity:

$$K_A(\boldsymbol{x}) = \min\{\, |\boldsymbol{p}| : A(\boldsymbol{p}) = \boldsymbol{x}\,\}.$$

A theorem similar to the existence of a universal test is the existence of an optimal interpreter:

Proposition (Invariance)

There is an optimal interpreter U in the sense that for every other interpreter A there is a constant c_A such that for all $\boldsymbol{x}$ we have

$$K_U(\boldsymbol{x}) \leqslant K_A(\boldsymbol{x}) + c_A.$$

From now on we fix a universal interpreter U and write $K(\boldsymbol{x}) = K_U(\boldsymbol{x})$.

The following theorem characterizes a universal payoff test in terms of complexity:

Proposition (Test characterization)

Let μ be a computable measure over the measureable space $(\Omega, \mathcal{B})$ with $\Omega = \Sigma^{\mathbb{N}}$. Let ω^n denote the prefix of length n of the infinite sequence ω. Then we have

$$\log u(\omega) = \sup_n(-\log \mu(\omega^n) - K(\omega^n)) + O(1).$$

Thus, ω is random if and only if the complexity of each of its prefixes $\boldsymbol{x}$ is never much smaller than $-\log \mu(\boldsymbol{x})$.
In a sense, such a nice characterization justifies the appropriateness of both the randomness notion and the complexity notion.