On Empirical Meaning of Randomness with Respect to a Real Parameter

Vladimir V'yugin

Institute for Information Transmission Problems, Russian Academy of Sciences,
Bol'shoi Karetnyi per. 19, Moscow GSP-4, 127994, Russia

Abstract. We study the empirical meaning of randomness with respect to a family of probability distributions P_θ, where θ is a real parameter, using algorithmic randomness theory. In the case when for a computable probability distribution P_θ an effectively strongly consistent estimate exists, we show that the Levin's a priory semicomputable semimeasure of the set of all P_θ-random sequences is positive if and only if the parameter θ is a computable real number. The different methods for generating "meaningful" P_θ-random sequences with noncomputable θ are discussed.

1 Introduction

We use algorithmic randomness theory to analyze the empirical meaning of random data generated by a parametric family of probability distributions when the parameter value is noncomputable. More correctly, let a parametric family of probability distributions P_θ (θ is a real number) be given such that an effectively strongly consistent estimate exists for this family. The Bernoulli family with a real parameter θ is an example of such family. We show that in this case the Levin's a priory semicomputable semimeasure of the set of all P_θ-random sequences is positive if and only if the parameter value θ is a computable real number.

We say that a property of infinite sequences have an empirical meaning if the Levin's a priory semimeasure of the set of sequences possessing this property is positive. In this respect, the model of the biased coin with "a prespecified" probability θ of the head is invalid if θ is a noncomputable real number; noncomputable parameters θ can have empirical meaning only in their totality, i.e., as elements of some uncountable sets. For example, P_θ-random sequences with noncomputable θ can be generated by a Bayesian mixture of these P_θ using a computable prior. In this case, evidently, the semicomputable semimeasure of the set of all sequences random with respect to this mixture is positive.

We also show that the Bayesian statistical approach is insufficient to cover all possible "meaningful" cases: a probabilistic machine can be constructed, which with probability close to one outputs a random θ-Bernoulli sequence, where the parameter θ is not random with respect to each computable probability distribution.

V. Diekert, M. Volkov, and A. Voronkov (Eds.): CSR 2007, LNCS 4649, pp. 387–396, 2007.

2 Preliminaries

Let Ξ be the set of all finite binary sequences, Λ be the empty sequence, and Ω be the set of all infinite binary sequences. We write $x \subseteq y$ if a sequence y is an extension of a sequence x, $l(x)$ is the length of x. A real-valued function $P(x)$, where $x \in \Xi$, is called semimeasure if

$$P(\Lambda) \leq 1,$$
$$P(x0) + P(x1) \leq P(x) \tag{1}$$

for all x, and the function P is semicomputable from below; this means that the set $\{(r, x) : r < P(x)\}$, where r is a rational number, is recursively enumerable. A definition of upper semicomputability is analogous.

Solomonoff proposed ideas for defining the a priori probability distribution on the basis of the general theory of algorithms. Levin [13], [3] gave a precise form of Solomonoff's ideas in a concept of a maximal semimeasure semicomputable from below (see also Li and Vitányi [7], Section 4.5, Shen et al. [9]). Levin proved that there exists a maximal to within a multiplicative positive constant factor semimeasure M semicomputable from below, i.e. for every semimeasure P semicomputable from below a positive constant c exists such that the inequality

$$cM(x) \geq P(x) \tag{2}$$

holds for all x. The semimeasure M is called *the a priory* or universal semimeasure.

A function P is a measure if (1) holds, where both inequality signs $\leq$ are replaced on $=$. Any function P satisfying (1) (with equalities) can be extended on all Borel subsets of Ω if we define $P(\Gamma_x) = P(x)$ in Ω, where $x \in \Xi$ and $\Gamma_x = \{\omega \in \Omega : x \subseteq \omega\}$; after that, we use the standard method for extending P to all Borel subsets of Ω. By simple set in Ω we mean a union of intervals Γ_x from a finite set.

A measure P is computable if it is, at one time, lower and upper semicomputable.

For technical reazons, for any semimeasure P, we consider the maximal measure $\bar{P}$ such that $\bar{P} \leq P$. This measure satisfies

$$\bar{P}(x) = \inf_n \sum_{l(y)=n, x \subseteq y} P(y).$$

In general, the measure $\bar{P}$ is noncomputable (and it is not a probability measure). By (2), for each lower semicomputable semimeasure P, the inequality $c\bar{M}(A) \geq \bar{P}(A)$ holds for every Borel set A, where c is a positive constant.

In the manner of Levin's papers [4,5,6,13] (see also [12]), we consider combinations of probabilistic and deterministic processes as the most general class of processes for generating data. With any probabilistic process some computable probability distribution can be assigned. Any deterministic process is realized by means of an algorithm. Algorithmic processes transform sequences generated by probabilistic processes into new sequences. More precise, a probabilistic computer is a pair (P, F), where P is a computable probability distribution, and F is a Turing machine supplied with an additional input tape. In the process of computation this machine reads on this tape a sequence ω distributed according to P and produces a sequence $\omega' = F(\omega)$ (A correct definition see in [4],[7], [9],[12]). So, we can compute the probability

$$Q(x) = P\{\omega \in \Omega : x \subseteq F(\omega)\}$$

that the result $F(\omega)$ of the computation begins with a finite sequence x. It is easy to see that $Q(x)$ is a semimeasure semicomputable from below.

Generally, the semimeasure Q can be not a probability distribution in Ω, since $F(\omega)$ may be finite for some infinite ω.

The converse result is proved in Zvonkin and Levin [13]: for every semimeasure $Q(x)$ semicomputable from below a probabilistic computer (L, F) exists such that

$$Q(x) = L\{\omega | x \subseteq F(\omega)\},$$

for all x, where $L(x) = 2^{-l(x)}$ is the uniform probability distribution in the set of all binary sequences.

Therefore, by (2) $M(x)$ defines an universal upper bound of the probability of generating x by probabilistic computers.

We refer readers to Li and Vitányi [7] and to Shen et al. [9] for the theory of algorithmic randomness. We use definition of a random sequence in terms of universal probability. Let P be some computable measure in Ω. The deficiency of randomness of a sequence $\omega \in \Omega$ with respect to P is defined as

$$d(\omega|P) = \sup_n \frac{M(\omega^n)}{P(\omega^n)}, \tag{3}$$

where $\omega^n = \omega_1 \omega_2 \dots \omega_n$. This definition leads to the same class of random sequences as the original Martin-Löf [8] definition. Let R_P be the set of all infinite binary sequences random with respect to a measure P

$$R_P = \{\omega \in \Omega : d(\omega|P) < \infty\}.$$

We also consider *parametric families* of probability distributions $P_\theta(x)$, where θ is a real number; we suppose that $\theta \in [0, 1]$. An example of such family is the Bernoulli family $B_\theta(x) = \theta^k (1-\theta)^{n-k}$, where n is the length of x and k is the number of ones in it.

We associate with a binary sequence $\theta_1 \theta_2 \dots$ a real number with the binary expansion $0.\theta_1 \theta_2 \dots$. When the sequence $\theta_1 \theta_2 \dots$ is computable or random with respect to some measure we say that the number $0.\theta_1 \theta_2 \dots$ is computable or random with respect to the corresponding measure in $[0, 1]$.

We consider probability distributions P_θ computable with respect to a parameter θ. Informally, this means that there exists an algorithm enumerating all pairs of rational numbers (r_1, r_2) such that $r_1 < P_\theta(x) < r_2$. This algorithm uses an infinite sequence θ as an additional input; if some pair (r_1, r_2) was enumerated by this algorithm then only a finite initial fragment of θ was used in the process of computation (for correct definition, see also Shen et al. [9] and Vovk and V'yugin [10]).

Analogously, we consider parametric lower semicomputable semimeasures. It can be proved that there exist a universal parametric lower semicomputable semimeasure M_θ. This means that for any parametric lower semicomputable semimeasure R_θ there exists a positive constant C such that $CM_\theta(x) \geq R_\theta(x)$ for all x and θ.

The corresponding definition of randomness with respect to a family P_θ is obtained by relativization of (3) with respect to θ

$$d_\theta(\omega) = \sup_n \frac{M_\theta(\omega^n)}{P_\theta(\omega^n)}$$

(see also [3]). This definition leads to the same class of random sequences as the original Martin-Löf [8] definition relitivized with respect to a parameter θ.

For any θ, let

$$I_\theta = \{\omega \in \Omega : d_\theta(\omega) < \infty\}$$

be the set of all infinite binary sequences random with respect to the measure P_θ. In case of Bernoulli family, we call elements of this set θ-*Bernoulli sequences.*

3 Results

We need some statistical notions (see Cox and Hinkley [2]). Let P_θ be some computable parametric family of probability distributions. A function $\hat{\theta}(x)$ from Ξ to $[0,1]$ is called *an estimate.* An estimate $\hat{\theta}$ is called *strongly consistent* if for any parameter value θ

$$\hat{\theta}(\omega^n) \to \theta$$

for P_θ-almost all ω. We suppose that ϵ and δ are rational numbers. An estimate $\hat{\theta}$ is called *effectively strongly consistent* if there exists a computable function $N(\epsilon, \delta)$ such that for any θ for all ϵ and δ

$$P_\theta\{\omega \in \Omega : \sup_{n \geq N(\epsilon,\delta)} |\hat{\theta}(\omega^n) - \theta| > \epsilon\} \leq \delta \tag{4}$$

The strong law of large numbers Borovkov [1] (Chapter 5)

$$B_\theta \left\{ \sup_{k \geq n} \left| \frac{1}{k} \sum_{i=1}^{k} \omega_i - \theta \right| \geq \epsilon \right\} < \frac{1}{\epsilon^4 n}$$

shows that the function $\hat{\theta}(\omega^n) = \frac{1}{n} \sum_{i=1}^{n} \omega_i$ is a computable strongly consistent estimate for the Bernoulli family B_θ.

Proposition 1. *For any effectively strongly consistent estimate* $\hat{\theta}$*,*

$$\lim_{n\to\infty} \hat{\theta}(\omega^n) = \theta$$

for each $\omega \in I_\theta$*.*

Proof. Let, for some θ, an infinite sequence ω be Martin-Löf random with respect to P_θ.

At first, we prove that $\lim\limits_{n\to\infty} \hat{\theta}(\omega^n)$ exists. Let for $j = 1, 2, \ldots$,

$$W_j = \{\alpha \in \Omega : (\exists n, k \geq N(1/j, 2^{-(j+1)}))|\hat{\theta}(\alpha^n) - \hat{\theta}(\alpha^k)| > 1/j\}.$$

By (4) for any θ, $P_\theta(W_j) < 2^{-j}$ for all j. Define $V_i = \cup_{j>i} W_j$ for all i. By definition, for any θ, $P_\theta(V_i) < 2^{-i}$ for all i. Also, any set V_i can be represented as a recursively enumerable union of intervals of type Γ_x. To reduce this definition of Martin-Löf test to the definition of the test (3) define a sequence of uniform lower semicomputable parametric semimeasures

$$R_{\theta,i}(x) = \begin{cases} 2^i P_\theta(x) \text{ if } \Gamma_x \subseteq V_i \\ 0 \text{ otherwise} \end{cases}$$

and consider the mixture $R_\theta(x) = \sum\limits_{i=1}^{\infty} \frac{1}{i(i+1)} R_{\theta,i}(x)$.

Suppose that the limit $\lim\limits_{n\to\infty} \hat{\theta}(\omega^n)$ does not exist. Then for each sufficiently large j, $|\hat{\theta}(\omega^n) - \hat{\theta}(\omega^k)| > 1/j$ for infinitely many n, k. This implies that $\omega \in V_i$ for all i, and then for some positive constant c,

$$d_\theta(\omega) = \sup_n \frac{M_\theta(\omega^n)}{P_\theta(\omega^n)} \geq \sup_n \frac{R_\theta(\omega^n)}{cP_\theta(\omega^n)} = \infty,$$

i.e., ω is not Martin-Löf random with respect to P_θ.

Suppose that $\lim\limits_{n\to\infty} \hat{\theta}(\omega^n) \neq \theta$. Then the rational numbers r_1, r_2 exist such that $r_1 < \lim\limits_{n\to\infty} \hat{\theta}(\omega^n) < r_2$ and $\theta \notin [r_1, r_2]$. Since the estimate $\hat{\theta}$ is consistent, $P_\theta\{\alpha : r_1 < \lim\limits_{n\to\infty} \hat{\theta}(\alpha^n) < r_2\} = 0$, and we can effectively (using θ) enumerate an infinite sequence of positive integer numbers $n_1 < n_2 < \ldots$ such that for

$$W'_j = \cup\{\Gamma_x : l(x) \geq n_j, r_1 < \hat{\theta}(x) < r_2\},$$

we have $P_\theta(W'_j) < 2^{-j}$ for all j. Define $V'_i = \cup_{j>i} W'_j$ for all i. We have $P_\theta(V'_i) \leq 2^{-i}$ and $\omega \in V'_i$ for all i. Then ω can not be Martin-Löf random with respect to P_θ. These two contradictions obtained above prove the proposition. △

The following theorem shows that, from the point of view of the philosophy explained above, P_θ-random sequences with "a prespecified" noncomputable parameter θ can not be obtained in any combinations of stochastic and deterministic processes.

Theorem 1. *Let a computable parametric family P_θ of probability distributions has an effectively strongly consistent estimate. Then for any θ, $\bar{M}(I_\theta) > 0$ if and only if θ is computable.*

Proof. If θ is computable then the probability distribution P_θ is also computable and by (2) $c\bar{M}(I_\theta) \geq P_\theta(I_\theta) = 1$, where c is a positive constant.

The proof of the converse assertion is more complicated. Let $\bar{M}(I_\theta) > 0$. There exists a simple set V (a union of a finite set of intervals) and a rational number r such that $\frac{1}{2}\bar{M}(V) < r < \bar{M}(I_\theta \cup V)$. For any finite set $X \subseteq \Xi$, let $\bar{X} = \cup_{x\in X}\Gamma_x$.

Let n be a positive integer number. Let us compute a rational approximation θ_n of θ up to $\frac{1}{2n}$. Using the exhaustive search, we find a finite set X_n of pairwise incomparable finite sequences of lengths $\geq N(1/n, 2^{-n})$ such that

$$\bar{X}_n \subseteq V, \sum_{x\in X_n} M(x) > r,$$

$$|\hat{\theta}(x) - \hat{\theta}(x')| \leq \frac{1}{2n} \tag{5}$$

for all $x, x' \in X_n$. If any such set X_n will be found, we put $\theta_n = \hat{\theta}(x)$, where $x \in X_n$ is minimal with respect to some natural (lexicographic) ordering of all finite binary sequences.

Now we prove that for any n the set X_n exists. Since $\bar{M}(I_\theta \cap V) > r$, there exists a closed (in the topology defined by intervals Γ_x) set $E \subseteq I_\theta \cap V$ such that $\bar{M}(E) > r$. Consider the function

$$f_k(\omega) = \inf\{n : n \geq k, |\hat{\theta}(\omega^n) - \theta| \leq \frac{1}{4n}\}.$$

By Proposition 1 this function is continuous on Ω and, since the set E is compact, it is bounded on E. Hence, for any k, a finite set $X \subseteq \Xi$ exists consisting of pairwise incomparable sequences of length $\geq k$ such that $E \subseteq \bar{X}$ and $|\hat{\theta}(x) - \hat{\theta}(x')| \leq \frac{1}{2n}$ for any $x, x' \in X$. Since $E \subseteq \bar{X}$, we have $\sum_{x\in X} M(x) > r$. Therefore, the set X_n can be found by exhaustive search.

Lemma 1. *For any Borel set $V \subseteq \Omega$, $\bar{M}(V) > 0$ and $V \subseteq I_\theta$ imply $P_\theta(V) > 0$.*

Proof. By definition of M_θ any computable parametric measure P_θ is absolutely continuous with respect to the measure $\bar{M}_\theta$, and so, we have representation

$$P_\theta(X) = \int_X \frac{dP_\theta}{d\bar{M}_\theta}(\omega) d\bar{M}_\theta(\omega), \tag{6}$$

where $\frac{dP_\theta}{d\bar{M}_\theta}(\omega)$ is the Radon-Nicodim derivative; it exists for $\bar{M}_\theta$-almost all ω.

By definition we have for $\bar{M}_\theta$-almost all $\omega \in I_\theta$

$$\frac{dP_\theta}{d\bar{M}_\theta}(\omega) = \lim_{n\to\infty} \frac{P_\theta}{\bar{M}_\theta}(\omega^n) \geq \liminf_{n\to\infty} \frac{P_\theta}{\bar{M}_\theta}(\omega^n) \geq C_{\theta,\omega} > 0. \tag{7}$$

By definition $c_\theta \bar{M}_\theta(X) \geq \bar{M}(X)$ for all Borel sets X, where c_θ is some positive constant (depending on θ). Then by (6) and (7) the inequality $\bar{M}(X) > 0$ implies $P_\theta(X) > 0$ for each Borel set X. △

We rewrite (4) in a form

$$E_n = \{\omega \in \Omega : \sup_{N \geq N(1/(2n), 2^{-n})} |\hat{\theta}(\omega^N) - \theta| \geq \frac{1}{2n}\} \tag{8}$$

and $P_\theta(E_n) \leq 2^{-n}$ for all n. We prove that $X_n \not\subseteq E_n$ for almost all n. Suppose that the opposite assertion holds. Then there exists an increasing infinite sequence of positive integer numbers $n_1, n_2 \ldots$ such that $X_{n_i} \subseteq E_{n_i}$ for all $i = 1, 2, \ldots$. This implies $P_\theta(X_{n_i}) \leq 2^{-n_i}$ for all i. For any k, define $U_k = \cup_{i \geq k} X_{n_i}$. Clearly, we have for all k, $\bar{M}(\bar{U}_k) > r$ and $P_\theta(\bar{U}_k) \leq \sum\limits_{i \geq k} 2^{-n_i} \leq 2^{-n_k+1}$. Let $U = \cap U_k$. Then $P_\theta(U) = 0$ and $\bar{M}(U) \geq r > \frac{1}{2}\bar{M}(V)$. From $U \subseteq V$ and $\bar{M}(I_\theta \cap V) > \frac{1}{2}\bar{M}(V)$ the inequality $\bar{M}(I_\theta \cap U) > 0$ follows. Then the set $I_\theta \cap U$ consists of P_θ-random sequences, $P_\theta(I_\theta \cap U) = 0$ and $\bar{M}(I_\theta \cap U) > 0$. This is a contradiction with Lemma 1.

Let $X_n \not\subseteq E_n$ for all $n \geq n_0$. Let also, a finite sequence $x_n \in X_n$ is defined such that

$$\Gamma_{x_n} \cap (\Omega \setminus E_n) \neq \emptyset.$$

Then from $l(x_n) \geq N(\frac{1}{2n}, 2^{-n})$ the inequality

$$\left| \frac{1}{l(x_n)} \sum_{i=1}^{l(x_n)} (x_n)_i - \theta \right| < \frac{1}{2n}$$

follows. By (5) we obtain $|\theta_n - \theta| < \frac{1}{n}$. This means that the real number θ is computable. Theorem is proved. △

Bayesian mixture of computable (with respect to θ) probability measures P_θ using a computable prior on θ gives to P_θ-random sequences "the empirical meaning". Let Q be a computable probability distribution on θ (i.e., in the set Ω). Then the Bayesian mixture

$$P(x) = \int P_\theta(x) dQ(\theta)$$

is also computable.

Recall that R_Q is the set of all infinite sequences Martin-Löf random with respect to a computable probability measure Q. Obviously, $P(\cup_{\theta \in R_Q} I_\theta) = 1$, and then $\bar{M}(\cup_{\theta \in R_Q} I_\theta)) > 0$. Moreover, it follows from Corollary 4 of Vovk and V'yugin [10]

Theorem 2. *For any computable measure Q a sequence ω is random with respect to the Bayesian mixture P if and only if ω is random with respect to a measure P_θ for some θ random with respect to the measure Q; in other words,*

$$R_P = \cup_{\theta \in R_Q} I_\theta.$$

Notice, that any computable θ is Martin-Löf random with respect to the computable probability distribution concentrated on this sequence.

The following Theorem 3 shows that the Bayesian approach is insufficient to cover all possible "meaningful" cases: a probabilistic machine can be constructed, which with probability close to one outputs a random θ-Bernoulli sequence, where the parameter θ is not random with respect to each computable probability distribution.

Let $\mathcal{P}$ be the set of all computable probability measures in Ω, and let

$$St = \cup_{P\in\mathcal{P}} R_P$$

be the set of all sequences Martin-Löf random with respect to allcomputable probability measures. We call these sequences - *stochastic*. Its complement $NSt = \Omega \setminus St$ consists of all sequences nonrandom with respect to all computable probability measures. We call them *nonstochastic.*

We proved in V'yugin [11], [12] that $\bar{M}(NSt) > 0$. Namely, the following proposition holds[1] .

Proposition 2. *For any ϵ, $0 < \epsilon < 1$, a lower semicomputable semimeasure Q can be constructed such that*

$$\bar{Q}(\Omega) > 1 - \epsilon, \tag{9}$$

$$NSt = \cup_{Q(x)>0} \Gamma_x. \tag{10}$$

We show that this result can be extended to parameters of the Bernoulli sequences.

Theorem 3. *Let I_θ be the set of all θ-Bernoulli sequences. Then*

$$\bar{M}(\cup_{\theta\in NSt} I_\theta) > 0.$$

In terms of probabilistic computers, for any ϵ, $0 < \epsilon < 1$, a probabilistic machine (L, F) can be constructed, which with probability $\geq 1-\epsilon$ generates an θ-Bernoulli sequence, where $\theta \in NSt$ (i.e., θ is nonstochastic).

Proof. For any $\epsilon > 0$, $0 < \epsilon < 1$, we define a lower semicomputable semimeasure P such that

$$\bar{P}(\cup_{\theta\in NSt} I_\theta) > 1 - \epsilon.$$

The proof of the theorem is based on Proposition 2. The property (10) can be rewritten as: $Q(\omega^n) = 0$ for all sufficiently large n if and only if $\omega \in St$ (i.e., ω is Martin-Löf random with respect to some computable probability measure).

For the measure

$$R^-(x) = \int B_\theta(x) d\bar{Q}(\theta), \tag{11}$$

[1] We also prove in these papers that $M(\Omega\setminus\bar{R}_L) > 0$, where $\bar{R}_L$ is the set of all sequences Turing reducible to sequences from R_L random with respect to the uniform measure L. By [13] it holds $St \subseteq \bar{R}_L$. The corresponding strengthening of the Theorem 3 is: $\bar{M}(\cup_{\theta\in\Omega\setminus\bar{R}_L} I_\theta) > 0$.

where B_θ is the Bernoulli measure, by (9) we have $R^-(\Omega) > 1-\epsilon$, and by (10) we have $R^-(\cup_{\theta\in St} I_\theta) = 0$.

Unfortunately, we can not conclude that $c\bar{M} \geq R^-$ for some constant c, since the measure R^- is not represented in the form $R^- = \bar{P}$ for some lower semicomputable semimeasure P. To overcome this problem, we consider some semicomputable approximation of this measure.

For any finite binary sequences α and x, let $B^-_\alpha(x) = (\theta^-)^K(1-\theta^+)^{N-K}$, where N is the length of x and K is the number of ones in it, θ^- is the left side of the subinterval corresponding to the sequence α and θ^+ is its right side. By definition, $B^-_\alpha(x) \leq B_\theta(x)$ for all $\theta^- \leq \theta \leq \theta^+$.

Suppose that ϵ is a rational number. Let $Q^s(x)$ be equal to the maximal rational number $r < Q(x)$ computed in s steps of enumeration of $Q(x)$ from below. Using (9), we can define for $n = 1, 2, \ldots$ and for any x of length n a computable sequence of positive integer numbers $s_x \geq n$ and a sequence of finite binary sequences $\alpha_{x,1}, \alpha_{x,2}, \ldots \alpha_{x,k_x}$ of length $\geq n$ such that the function $P(x)$ defined by

$$P(x) = \sum_{i=1}^{k_x} B^-_{\alpha_{x,i}}(x) Q^{s_x}(\alpha_{x,i}) \quad (12)$$

is a semimeasure, i.e., such that condition (1) holds for all x, and such that

$$\sum_{l(x)=n} P(x) > 1-\epsilon \quad (13)$$

holds for all n. These sequences exist, since the limit function R^- defined by (11) is a measure satisfying $R^-(\Omega) > 1-\epsilon$.

By definition the semimeasure $P(x)$ is lower semicomputable. Then $cM(x) \geq P(x)$ holds for all $x \in \Xi$, where c is a positive constant.

To prove that $\bar{P}(\Omega \setminus \cup_\theta I_\theta) = 0$ we consider some probability measure $Q^+ \geq Q$. Since (1) holds, it is possible to define some noncomputable measure Q^+ satisfying these properties in many different ways. Define the mixture of the Bernoulli measures with respect to Q^+

$$R^+(x) = \int B_\theta(x) dQ^+(\theta). \quad (14)$$

By definition $R^+(\Omega \setminus \cup_\theta I_\theta) = 0$. Using definitions (12) and (14), it can be easily proved that $\bar{P} \leq R^+$. Then $\bar{P}(\Omega \setminus \cup_\theta I_\theta) = 0$. By (10) we have $\bar{P}(\cup_{\theta\in St} I_\theta) = 0$. By (13) we have $\bar{P}(\Omega) > 1-\epsilon$. Then $\bar{P}(\cup_{\theta\in NSt} I_\theta) > 0$. Therefore, $\bar{M}(\cup_{\theta\in NSt} I_\theta) > 0$. △

Acknowledgements

This work was partially supported by Russian foundation for fundamental research: 06-01-00122.

References

1. Borovkov, A.A.: Theory of Probability. Nauka. (Probability Theory. Gordon and Breach 1998) (1999)
2. Cox, D.R., Hinkley, D.V.: Theoretical Statistics. Chapman and Hall, London (1974)
3. Levin, L.A.: On the notion of random sequence. Soviet Math. Dokl. 14, 1413–1416 (1973)
4. Levin, L.A.: Laws of information conservation (non-growth) and aspects of the foundation of probability theory. Problems Inform. Transmission. 10, 206–210 (1974)
5. Levin, L.A., V'jugin, V.V.: Invariant Properties of Informational Bulks. In: Gruska, J. (ed.) Mathematical Foundations of Computer Science 1977. LNCS, vol. 53, pp. 359–364. Springer, Heidelberg (1977)
6. Levin, L.A.: Randomness conservation inequalities; information and independence in mathematical theories. Inform. and Control. 61, 15–37 (1984)
7. Li, M., Vitányi, P.: An Introduction to Kolmogorov Complexity and Its Applications, 2nd edn. Springer, New York (1997)
8. Martin-Löf, P.: The definition of random sequences. Inform. and Control. 9, 602–619 (1966)
9. Shen, A., Uspensky, V.A., Vereshchagin, N.K.: Lecture Notes on Kolmogorov Complexity (2007), `http://lpes.math.msu.su/ver/kolm-book`
10. Vovk, V.G., V'yugin, V.V.: On the empirical validity of the Bayesian rule. J. R. Statist. Soc. B. 55, 317–351 (1993)
11. V'yugin, V.V.: On Turing invariant sets. Soviet Math. Dokl. 17, 1090–1094 (1976)
12. V'yugin, V.V.: The algebra of invariant properties of binary sequences. Problems Inform. Transmission. 18, 147–161 (1982)
13. Zvonkin, A.K., Levin, L.A.: The complexity of finite objects and the algorithmic concepts of information and randomness. Russ. Math. Surv. 25, 83–124 (1973)