Cover

MIT/LCS/TR-235

A CONCEPT OF INDEPENDENCE WITH APPLICATIONS IN VARIOUS FIELDS OF MATHEMATICS

Leonid A. Levin

A CONCEPT OF INDEPENDENCE
WITH APPLICATIONS IN VARIOUS FIELDS OF MATHEMATICS

Leonid A. Levin

April 1980

This report was prepared with the support of a National Science Foundation Grant No. MCS77-19754.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LABORATORY FOR COMPUTER SCIENCE

CAMBRIDGE MASSACHUSETTS 02139

A CONCEPT OF INDEPENDENCE WITH APPLICATIONS IN VARIOUS FIELDS OF MATHEMATICS

by Leonid A. Levin

ABSTRACT

We use Kolmogorov's algorithmic approach to information theory to define a concept of independence of sequences, or equivalently, the boundedness of their mutual information. This concept is applied to probability theory, intuitionistic logic, and the theory of algorithms. For each case, we study the advantage of accepting the postulate that the objects studied by the theory are independent of any sequence determined by a mathematical property.

Key words: Complexity, Information, Randomness, Independence.

CONTENTS

ACKNOWLEDGMENTS

The author feels obliged to A.R. Meyer more than to anybody else for major contribution in the improvement of the text and the financial, administrative and moral support. P. Elias and G.E. Sacks accepted the trouble of reading the work and L. Adleman discussed its content.

It was suggested at the start, that this work would be more original in the field of English, than in mathematics. H. Abelson, L. Blum, P. Gacs, L. Lastovetsky, I. Melchuk and R. Winklman promoted a diminution of this effect. It was difficult because of the author's insistence on his awkward expressions. Computer programs Doc, Spell, Tex formatted text, and Ellen Golden was a guard angel preventing troubles between them and the author.

The author is deeply grateful to all these persons and programs and apologizes to them that their mission in this case turned out to be a lot more difficult than usual.

0. PRELIMINARIES

0.1 INTRODUCTION

The following attempt to define precisely the concept of independence may seem frivolous, for there are probably as many different concepts of "independence" in science as there are concepts of "freedom" in the humanities. To justify our efforts, we try to demonstrate that this definition can be applied in various fields of mathematics.

The main idea of this work is to formalize, justify and apply the following physical postulate: If α is a sequence generated by a process of the physical world, and β is a sequence determined by a property formulated with no reference to events of the physical world, then α and β are independent. This agrees with Church's Thesis, which asserts the recursiveness of any sequence which is both mathematically defined and physically realizable, since a recursive sequence is by our definition independent even of itself. Our Independence Postulate expresses "autonomy of the physical world", its independence of anything outside itself. This corresponds to the idea of causality in physics.

The sequence of papers in a mathematical journal, or the sequence of oil prices, are examples of α. The sequence of all true assertions in number theory is an example of β. Below we take as examples of β sequences defined by mathematical properties. In Chapters two through four the random sequences, the free-choice sequences of intuitionistic theories, and the representatives of "regular" Turing degrees respectively are considered as α. In each of these three cases we show that accepting the Independence Postulate allows us to radically simplify the corresponding theories.

The theory is developed in the simplest version which is sufficient for the applications considered below. Proofs are presented rather formally and may be omitted at the first reading.

0.2 AN EXAMPLE

Recursive function theory allows us to construct analogues of many concepts of classical analysis by presuming recursive enumerability of the sets considered. The analogy obtained is quite good because intrinsically non-algorithmic methods are the exception rather than the rule in classical mathematics. Moreover, the general theory of algorithms is very similar to descriptive set theory. (This explains why the main attention of "constructive analysis" has been directed toward the search for exotic counter-examples to some theorems of classical analysis. These investigations have always remained of narrow special interest.) However, there is an important distinction between the constructive and classical theories: the existence of a universal algorithm. The set of r.e. sets is r.e. while the set of countable sets is uncountable. As was discovered in [S64, K65], this rather abstract difference implies "more concrete" differences and opens new analytical possibilities which have no analogies in "non-algorithmic" analysis. Let us explain this with a simple but important example.

The space l_1 of all absolutely convergent number series p, ($p\in l_1$, iff $p\in\mathbf{R}^{\mathbb{N}}\ \&\sum|p(x)| < \infty$) is well studied in analysis. Its recursive analogue $\bar{l}_1\subset l_1$ consists of r.e. elements of l_1 (i.e. of series p whose subgraph $\{(r,x)\colon p(x)>r\in\mathbf{Q}\}$ is r.e.). It is known in calculus that l_1, has no maximal to within a constant factor element: $\forall p\in l_1 \exists q\in l_1 \lim(q(x)/p(x)) = \infty$. In contrast to this, $\bar{l}_1$ contains an "absorbing" element m such that $\forall q\in\bar{l}_1 \sup(q(x)/m(x)) < \infty$. This will follow from Theorem 1. This fact is closely connected to the discovery by R.J. Solomonoff and A.N. Kolmogorov of optimal coding of finite objects which originated a new approach to information theory, the foundations of probability theory, inductive inference and a number of other fields.

All of these results, which we combine under the name "algorithmic information theory", are based on purely analytical features which distinguish the recursive analogues of some spaces of analysis from their classical prototypes. The preceding example illustrates this distinction.

0.3 NOTATION AND ASSUMPTIONS

There are several natural general contexts for the formulation of this work. They differ in the space Ω of objects considered to be carriers of information. In all cases we need a countable family of functions on Ω, which extract parts of this information. Declaring these functions continuous and identifying any two objects on which the values of all these functions coincide, we may regard Ω as a topological space with a countable basis and with the Kolmogorov property: for any two points an open set exist containing only one of them (assuming this property for the functions' range). Among such spaces there is a universal one, i.e., a space containing the homeomorphic image of any other. Formulating the theory for this space is the most general possibility. If the range of the family of functions is a metric space, the Kolmogorov property of Ω is strengthened to complete regularity. Among completely regular spaces with a countable base, there is also a universal one, $\mathbb{R}^{\mathbb{N}}$. As usual, considerations look much simpler for a regular space. We even introduce a further, unessential simplification, namely, that Ω is totally disconnected (i.e., any two points can be distinguished by a continuous mapping into a discrete space). Cantor's perfect set $\overline{\mathbb{N}}^{\mathbb{N}}$ (or $\{0,1\}^{\mathbb{N}}$) which we do in fact consider, is universal among such spaces. Moreover, what we discuss in most detail is an even more special case: the space $\mathbb{N}$ of non-negative integers.

We compactify $\mathbb{N}$ to $\overline{\mathbb{N}}$ by adding the symbol "∞". The number $m+((m+n)(m+n+1)/2)$ is called the pair (m,n) of the numbers $m, n\in\overline{\mathbb{N}}$. This enumeration of the pairs is bijective on $\mathbb{N}^2\subset\overline{\mathbb{N}}^2$. The projections π_1 and π_2 are the functions on $\mathbb{N}$ such that $n = (\pi_1(n), \pi_2(n))$. Henceforth, Ω denote Cantor's perfect set, represented in the form of $\overline{\mathbb{N}}^{\mathbb{N}}$. This form is more convenient than $\{0,1\}^{\mathbb{N}}$ since the pairs (α,β) and the projections π_1 and π_2 are simpler defined on it: $(\alpha,\beta)(i) = (\alpha(i), \beta(i))$, where $\alpha(i), \beta(i)\in\overline{\mathbb{N}}$ are the i-th terms of α and β.

Let $S_k = \{0, 1, ...k, \infty\}^k$ and $S = \bigcup S_k$. Λ is the empty sequence, $S_0 = \{\Lambda\}$; x*is the number of $x\in S$ in a natural effective enumeration of S. $l(x)$ is the length of $x\in S$, i.e. the number k such that $x\in S_k$. If $a\in\Omega$ or $a\in S_n$ and $k\leq n$, then $a_k\in S_k$ is the initial segment of a of length k in which all the terms larger than k are replaced by ∞. $x\subset y$ means $l(x)\leq l(y)$ and $x = y_{l(x)}$; likewise for $x\subset a$. Γ_x is the set of $a\in\Omega$, such that $x\subset a$. The sets Γ_x form a countable basis consisting of the clopen (i.e. closed and open) subsets of Ω. It is easy to see that if $\Gamma_x\bigcap\Gamma_y\neq\emptyset$ then $\Gamma_x\subset\Gamma_y$ or $\Gamma_y\subset\Gamma_x$ (i.e. $y\subset x$ or $x\subset y$). B is the set of finite binary sequences; $\mathbb{Q}, \mathbb{Q}^+, \mathbb{R}, \mathbb{R}^+, \overline{\mathbb{Q}}^+, \overline{\mathbb{R}}^+, \overline{\mathbb{R}}, \overline{\mathbb{Q}}$ are the sets of rational, nonnegative rational, real numbers and so on respectively. $\lfloor x \rfloor$ is the integer part of **x**.

While considering topological spaces with natural countable bases we call an open set **recursively** enumerable (r.e.) if it equals the union of an r.e. family of basis sets. The function F with values in $\overline{\mathbb{R}}$ we call r.e. if its subgraph, i.e., the set $\{(x,r)\colon r<F(x)\}$ is r.e.. We call F recursive if F and -F are r.e.. We shall systematically assert the recursive enumerability of sets without giving the formal tedious constructions. Similarly, in Chapter 3, we assert the expressibility or provability of predicates of formal arithmetic without writing out the corresponding lengthy formulas or proofs. These assertions can be checked routinely.

The symbols $\prec$, $\succ$ and $\sim$ denote inequality and equality to within an additive constant; $\preceq$, $\succeq$ and $\simeq$ denote these relations to within a constant factor; $\lesssim$, $\gtrsim$ and $\approx$ denote asymptotic relations, (i.e. $f \lesssim g \Leftrightarrow \forall\epsilon>0\exists a(f>a \Rightarrow g>(1-\epsilon)f)$). Such expressions as $\sum f$, $\sup f$, $\min f$, etc. denote the corresponding operations, taken over the values of all free variables of the term f.

1. ALGORITHMIC INFORMATION THEORY

1.1 UNIVERSAL SEMIMEASURE.

Let $\sigma(x) = \{y\colon x\subset y, l(y)=l(x)+1\}$. Then $\Gamma_x = \bigcup\Gamma_y$, where $y\in\sigma(x)$. Each finite positive Borel measure on Ω is uniquely determined by a function $\mu\colon S\rightarrow\mathbf{R}^+$ such that $\forall x\sum\mu(y) = \mu(x)$, where $y\in\sigma(x)$. We identify this function (giving the measure of the set Γ_x) with the measure itself. Let us introduce a somewhat more general concept.

Definition 1: *A* semimeasure *on Ω is a function $\mu\colon S\rightarrow\mathbf{R}^+$ such that $\forall x\sum\mu(y)\leq\mu(x)$ $(y\in\sigma(x))$. Unless otherwise stipulated, we assume that semimeasures are normalized, that is,* $\mu(\Lambda)=1$.

A semimeasure μ corresponds to the measure μ' on $\Omega\bigcup S$, where for $x\in S$, $\mu'(\{x\}) = \mu(x)-\sum\mu(y)$, where $y\in\sigma(x)$. Any r.e. measure μ is also recursive, because $\mu(x) = 1-\sum\mu(y)$, where $l(y)=l(x), y\neq x$.

Theorem 1:

There exists a largest to within a constant factor r.e. semimeasure: $\exists M\forall\mu\exists c\forall x\,\mu(x)\leq cM(x)$.

Proof: We prove, first, that the set of all r.e. semimeasures is r.e. For each finite set $A\subset S\times\mathbf{Q}^+$ a minimal semimeasure μ_A exists the closure of whose subgraph contains A. The norm of this μ_A (i.e. $\mu_A(\Lambda)$) is evidently computable on a finite A. We denote it by $|A|$. Let $f(i,n)$ be a total recursive function such that for each i the set of values of f is the i-th r.e. subset of $S\times\mathbf{Q}^+$ containing $(\Lambda,1)$. Let $A(i,n)$ be the set of values of f on the pairs (i,m), where $m<n$. Let $\bar{f}(i,n)=f(i,n)$ if $|A(i,n)|\leq1$, otherwise $\bar{f}(i,n) = (\Lambda,1)$. Let $A(i)$ be the set of values of $\bar{f}$ on the pairs (i,n). Obviously $\mu_{A(i)} = \mu_{A(i,\infty)}$ iff $|A(i,\infty)|\leq1$. Thus, the family $\mu_{A(i)}$ enumerates all the normalized r.e. semimeasures (and only them). The semimeasure $\sum(1/i^2)\mu_{A(i)}$ is obviously r.e. and finite, and exceeds any other such semimeasure to within a constant factor $1/i^2$. Q.E.D.

This semimeasure is called universal and denoted as M. It is the central technical concept of this work. Being the largest (to within a constant factor) among all r.e. semimeasures, it determines the broadest class of sets $A\subset\Omega$ of positive measure.

In mathematical statistics one tries, given α, to get a probability distribution μ for which it would be reasonable to assert that "α is random with respect to μ". This usually means that some properties of α (i.e. sets $A\in\Omega$ containing α), are of positive probability with respect to μ. But the latter assertion is the weakest in the case $\mu = M$. So, we can take M *a priori*, before studying what the properties of α really are. For this reason we call M the *a priori* probability distribution.

Let us express this in other words. Suppose we want to predict the properties of some unknown sequence α. The assumption that α occurs randomly with probability distribution μ allows us to conclude that α will have a property A, when the probability $\mu(\neg A)$ of the opposite property equals 0. The class of such properties is the narrowest in the case where $\mu = M$ and therefore these properties can be presumed before clearing up what μ really is. Therefore, before determining the nature of a random process, one may assume *a priori* such properties of an outcome which are certain to hold for the random process with distribution M. This justifies calling M the *a priori* probability.

M has all the properties necessary for the construction of an inductive inference theory in accordance with the ideas of R.J. Solomonoff [S64], but we cannot go into this question here. In further accounts we will consider M as the *a priori* probability according to the use of this concept in statistics. Let us note that if μ is an r.e. semimeasure, then with probability 1 (by μ) a sequence α is such that values $\mu(a_n)$ and $M(a_n)$ agree to within a factor independent of n. This property of α can be used as a definition of the concept of "a sequence random with respect to the probability distribution μ". We do not attempt to explore fully the properties of M as the *a priori* probability; our main application of M is to algorithmic information theory.

1.2 DISCRETE CASE: COMPLEXITY AND INFORMATION

Before introducing our concepts for the Cantor set Ω let us consider the simpler space $\mathbf{N}$. Let m be obtained by applying M to $\mathbf{N}$, namely, define $m(k) = M\{a{:}a(0)=k\}$. The definitions of M and m imply trivially that $m\in\bar{l}_1$ and is, in fact, the largest element in $\bar{l}_1$ to within a constant factor.

The complexity of $n\in\mathbf{N}$ is $K(n) = -\lfloor \log_2 m(n)\rfloor$. This function, as it turns out, defines the length of the shortest code for n, using an optimal self-delimiting coding:

An algorithm $A{:}B\rightarrow\mathbf{N}$ is called self-delimiting if $A(p_1)=A(p_2)$ for any p_1, p_2 such that $p_2\subset p_1$ and $A(p_1)$ and $A(p_2)$ are defined. This means that if A has produced a result on some input, it cannot produce anything different on any continuation of this input. Informally, the algorithm A recognizes the end of the "essential part of the input" and pays no attention to further symbols. Such an algorithm needs no special symbol, to distinguish the end of input, and its input alphabet is consequently "authentically binary".

Proposition 2 (about coding): *There exists a self- delimiting algorithm A capable of generating any* $n\in\mathbf{N}$ *from an input of length* $K(n)$*; more precisely* $\exists A\forall n\exists p\,(A(p)=n$ *and* $l(p)=K(n)+1)$.
No self- delimiting algorithm A' *can be better by more than an additive constant i.e.* $\forall A'\exists c\forall n\forall p(A'(p)=n) \Rightarrow l(p)\geq K(n)-c$.

Therefore, the value $K(n)$ defines, to an additive constant, the minimal amount of information necessary to determine n. This corresponds to Shannon's idea that the amount of information in an event equals the negative logarithm of its probability. $K(n)$ is the full amount of information in n.

Proof of Proposition 2: The proof is almost obvious. We need to show for an arbitrary function $K'{:}\mathbf{N}\rightarrow\bar{\mathbf{N}}$ that $2^{-K'}\in\bar{l}_1$ (see 0.3), iff a constant C and a self-delimiting algorithm A exist such that $K'(x)+C = \min\{l(q){:}\ A(q)=x\}$. Being self-delimiting, A cannot take different values on segments of the same sequence, and can be considered as a function from Ω_2. The natural measure B_2 on Ω_2 is $B_2(\Gamma_q) = 2^{-l(q)}$. If $K'(x)+C = \min l(q){:}A(q)=x$ then, obviously, $2^{-(K'(x)+C)}\leq B_2\{a{:}A(a)=x\}$. Then $\sum 2^{-(K'(x)+C)}\leq 1$ and $2^{-K'}\in l_1$. Also $2^{-K'}$ is r.e. because the graph of A is.
Vice versa, if $2^{-\overline{K'}}\in\bar{l}_1$ then the set $\overline{\mathrm{A}} = \{(x,n){:}\ n>K'(x)\}$ is r.e. Let $\mu(x,n)=2^{-n}$, if $n>K'(x)$, and $\mu(x,n)=0$ otherwise. Then $C = 2+\lfloor \log_2\sum\mu(x,n)\rfloor\leq 2+\log_2\sum 2^{-K'(x)} < \infty$. It is easy to find a recursive bijection $(x,n)\rightarrow q_{x,n}$ of $\overline{\mathrm{A}}$ to a self-delimiting set $\overline{\mathrm{A}}'\subset B$ such that $B_2(q_{x,n}) = 2^{-(n+C)}$, and thus $l(q_{x,n}) = n+C$. The desired algorithm A maps $q_{x,n}$ to x, Q.E.D.

The code of a pair (n,m) of numbers can be shorter than K(n)+K(m) because n and m may contain mutual information coded only once.

Definition 2: *The value* $I(n{:}m) = K(n) + K(m) - K(n,m)$ *is called* **the amount of mutual information in n and m.**

Remark: The self-delimitedness of the coding algorithms A implies that $I(n{:}m)\succ 0$, since the pair (n,m) can be encoded by p_1p_2, where p_1 and p_2 are the shortest codes for n and m respectively.

1.3 DISCRETE CASE: RANDOMNESS AND INDEPENDENCE

In order to arrive at the expression I(x:y) from another point of view, let us consider some problems connected with the concept of randomness. In 1.1 we mentioned the possibility of characterizing the properties of sequences occurring randomly with a probability distribution μ by the boundedness of the ratio of M to μ on their segments. Now we touch upon this matter in the simplest case: $x \in \mathbf{N}$. While considering a probability distribution on a countable set, we usually cannot talk about "properties random objects *must* have" since usually only the whole space is of probability 1. Then we have to consider quantities which must be small on random objects. (This means that a given "test" takes large values only with a small probability).

Let μ be a recursive measure on $\mathbf{N}$, $\mu{:}\mathbf{N}\rightarrow\mathbf{R}^{+}, \sum\mu(x)=1$. Let us call a randomness test with respect to μ or a μ-test any r.e. function $\delta{:}\mathbf{N}\rightarrow\overline{\mathbf{N}}$, which satisfies the Martin-Lof condition: $\forall n \log_2 \mu\{x{:}\ \delta(x)>n\} \leq -n$. Let m be the universal measure on $\mathbf{N}$, defined in 1.2.

Proposition 3:
a) For any recursive measure μ the function $d(x/\mu) = \lfloor \log_2(m(x)/\mu(x))\rfloor$ is a μ-test.
b) For any recursive measure μ and μ-test δ, $\delta(x) \lesssim d(x/\mu)$.
c) m is maximal to within a constant factor among all functions for which a) holds.

Proposition 3 indicates that d(x/μ) is, in a sense, a universal characteristic of "non-randomness" and we call it the randomness deficiency of x with respect to μ. Motivations of the concept of randomness are discussed in Chapter 2.

Proof: a) Obviously d is r.e. Let $\mu\{x{:}\log_2(m(x)/\mu(x))>n\} > 2^{-n}$. Then $\mu\{x{:}m(x)>2^n\mu(x)\} > 2^{-n}$. Then $m\{x{:}m(x)>2^n\mu(x)\} > 2^n\mu\{x{:}m(x)>2^n\mu(x)\} > 1$, which contradicts the normality of m.
b) Let $\mu'(x)=\mu(x)2^{\delta(x)}/\delta^2(x)$. Then, $\sum\mu'(x)\leq\sum_{n=\delta(x)}\mu(x)2^n/n^2=\sum(2^n/n^2)\mu\{x{:}\delta(x)=n\} \leq \sum(2^n/n^2)2^{-n} = \sum 1/n^2 < \infty$. Thus $\mu'(x)$ is the r.e. semimeasure. Then $\mu'(x)\preceq m(x)$, which implies the required inequality.
c) Let $\overline{\mathrm{m}}$ satisfy a) as well as m. Obviously $\overline{\mathrm{m}}$ is r.e.. It remained to show that $\overline{\mathrm{m}}$ is a semimeasure, i.e. $\sum\overline{\mathrm{m}}(x) < \infty$. Let $\sum\overline{\mathrm{m}}(x)=\infty$. Then, obviously a recursive function $m'(x)\leq\overline{\mathrm{m}}(x)$ exists such that $\sum m'(x) = \infty$. Let $s(x) = \lfloor \log_2\sum_{y\leq x} m'(y)\rfloor$. Let $\mu(x) = m'(x)2^{-s(x)}/s^2(x)$. Then $\mu(x)$ is a measure since $\sum\mu(x)\leq\sum 1/n^2 < \infty$. Obviously $\mu\{x{:}\overline{\mathrm{m}}(x)/\mu(x)>2^n\} \geq \mu\{x{:}m'(x)/\mu(x)>2^n\} \geq \mu\{x{:}s(x)\geq n\} \geq 1/n^2$, which contradicts the Martin-Lof condition. Q.E.D.

Let two random variables, defined on the same probability space, be independent and have the same distribution μ. This is equivalent to the fact that their joint distribution is $\mu\otimes\mu$ where $\mu\otimes\mu(a,b) = \mu(a)\mu(b)$. Suppose the properties of the pair $(x,y)\in\mathbf{N}^2$ correspond to the results of a random process with distribution $i=m\otimes m$, i.e. $d((x,y)/i)$ is small. What is the intuitive meaning of this? The same as of the assertion that "(x,y) has the properties of the results of the pair of two independent random processes, and each of x and y has the properties of the results of a random process with distribution m". The second part of this assertion is vacuously true, since all numbers have the properties of the results of a random process with the *a priori* distribution m: $d(x/m)\equiv\mathbf{0}$.

Therefore, the smallness of d((x,y)/i) means only that (x,y) has the properties of the pair of objects generated in an arbitrary way but independently of each other. It is natural to consider the value d((x,y)/i) as the deficiency of independence. Obviously $d((x,y)/i) = I(x{:}y)$! This is consistent with the theorem of classical probabilistic information theory stating that two random variables are independent if and only if the mutual information between them equals 0. The difference is that the concepts given above are applicable to the individual values themselves, and not only to probability distributions (i.e. random variables) on the set of values.

1.4 CONSERVATION OF INDEPENDENCE

The information I(x:y) has a remarkable property. It increases in no random or algorithmic (deterministic) processing of x or y and hence in none of their combinations. On the one hand this is natural, since if x contains no information about y then hope is little to find out something about y by subjecting x to various kinds of processing. (Torturing an uninformed witness cannot give information about the crime!) This may conflict with the common experience that the Monte-Carlo method solves many problems which are intractable without using a random number generator. The clue here is that one can always solve these problems by computing the probability distribution of the results of random input processing. For this one needs to consider all possible inputs (instead of a single random one) which is an unrealistic volume of work. Even so, theoretically, the Monte-Carlo method produces no "absolutely new" possibilities in this respect.

Theorem 4 (Independence Conservation):
Let $A{:}\mathbb{N}\rightarrow\mathbb{N}$ *be a recursive function, and* φ *be an r.e. measure on* $\mathbb{N}$. *Then*
1) $I(x{:}y)\succ I(A(x){:}y)$,
2) $\exp(I(x{:}y))\succeq E_{\varphi(z)}\exp(I((x,z){:}y))$, *where E means mathematical expectation.*

Proof: $I(x{:}y)\sim I((x,A(x)){:}y)$ since x and (x,A(x)) are computable from each other. It remains to prove that $I((z,x){:}y)\succ I(x{:}y)$ for $x{=}A(z)$. This reduces to $K(x,y,z)\prec K(x,y)+K(x,z)-K(x)$.
We need an elegant lemma of Peter Gacs:

Lemma 1: $K(t,K(t))\sim K(t)$.

Indeed, let p be the shortest code for t. Obviously, K(t)=l(p) is computable from p as well as t is. Therefore, the complexity of (t,K(t)) equals l(p)=K(t).

Definition: An r.e. function $m(/){:}\mathbb{N}\times\mathbb{N}\rightarrow\mathbb{R}^+$, largest to within a constant factor among such ones that $\sup_y(\sum m(x/y)) < \infty$ is called the universal conditional measure. $K(x/y) = -\lfloor \log_2 m(x/y)\rfloor$.

Lemma 2: $K(x,y)\sim K(y/(x,K(x)))+K(x)$.

Let $m_\infty(y/x,n) = m(x,y)2^n$. A nondecreasing by k, recursive sequence $m_k(y/x,n){:}\ A_k\rightarrow\mathbb{Q}^+$ exists, such that $m_\infty = \sup m_k$, where A_k are finite subsets of $\mathbb{N}^3$. Let $\overline{\mathrm{m}}(y/x,n) = \sup_k\{m_k(y/x,n){:}\text{-}\sum m_k(z/x,n)\leq 1\}$. Obviously $\forall x,n\sum\overline{\mathrm{m}}(y/x,n)\leq 1$ (thus $\overline{\mathrm{m}}(/)\prec m(/)$) and $\forall x,n$ if $\sum m(x,y)\leq 2^{-n}$, then $m_\infty(y/x,n) = \overline{\mathrm{m}}(y/x,n)$. Therefore $\forall x,n$ if $\sum m(x,y)\leq 2^{-n}$, (i.e. if $m(x)\preceq 2^{-n}$, or $n\succ K(x)$) then $m(y/x,n)\succeq\overline{\mathrm{m}}(y/x,n) = m_\infty(y/x,n) = 2^n m(x,y)$. Thus $K(y/x,K(x))\prec K(x,y)-K(x)$.

It remains to prove that $K(y/(x,K(x)))\succ K(x,y)-K(x)\sim K(x,y)-K(x,K(x))$. This follows from the facts that $K(x,y)\prec K(y,x,K(x))$, $K(x)\sim K(x,K(x))$ and $K(y,t)\prec K(t)+K(y/t)$. The latter inequality holds since $m'(y,t) = m(t)m(y/t)$ is obviously an r.e. semimeasure and then $m'(y,t)\preceq m(y,t)$. Analogously can be obtained $K(x,y,z) \prec K(x,K(X)) + K(y/(x,K(x))) + K(z/(x,K(x)))$.

Now, item 1 follows from the note that

$$K(x,y) + K(x,z) - K(x) \sim K(y/(x,K(x))) + K(x,K(x)) + K(z/(x,K(x))) + K(x,K(x)) - K(x,K(x))$$

For the proof of 2) one needs to show that: $m(x,y)/(m(x)m(y))\succeq E_{\varphi(z)}m(x,y,z)/(m(y)m(x,z))$ which can be reduced to $E_{m(z)}m(x,y,z)/m(x,z)\preceq m(x,y)/m(x)$ since $m(z)\succeq\varphi(z)$. Let us transform it: $\sum_z m(z)m(x,y,z)/m(x,z)\preceq m(x,y)/m(x)$; $\sum_z m(z)m(x)m(x,y,z)/m(x,z)\preceq m(x,y)$. The latter inequality follows from the obvious ones: $m(z)m(x)\preceq m(x,z)$ and $\sum m(x,y,z)\preceq m(x,y)$. Q.E.D.

Returning from the logarithmic scale to the linear one strengthened item 2) of Theorem 4. This scale is natural to use with such linear operations as mathematical expectation. Theorem 4 is formulated to within additive constants independent of x,y, but dependent on A or φ (bounded by K(A) and K(φ) respectively). Item 2) is related not to z but only to the mathematical expectation on it. I.e. the information may increase in a random process, but only with negligible probability (by n bits with probability 2^{-n}). Both reservations are unremovable, since one can increase information by randomly guessing n symbols of y or by means of an algorithm A already having these symbols in its program. This does not diminish the meaning of the Theorem; since one should consider this additional information as having been present originally in the program of A (or "in our luck" in the case of guessing) rather than arising from the processing of x. Theorem 4 excludes any more efficient possibilities. Processes more complex than those in Theorem 4 can be obtained by combining its items (e.g. a generalization of 2) by substitution $\varphi_x(z)$, dependent on x, for $\varphi(z)$). Theorem 4 also implies non-increasing information in any combination of random and deterministic (recursive) processes. This supports the Principle below about the conservation of independence in any physically realizable process of information transformation.

The following formulation and discussion of this Principle is a deviation from the formal account, given for motivation of the formal results. To confirm the Independence Principle one may say that it is usually possible to "explain" known physical processes. To "explain" means to reduce them to simpler ones in combination with recursive and random transformations. General ideas about the development of the physical universe, on the whole, also assume that it was originally in a state of random movement of a hot plasma and then was transformed according to the (recursive) equations of quantum mechanics (additional randomness appears in the observation processes). It is clear that, not being a mathematical assertion (the physical world is not defined mathematically), the Independence Principle (like, for example, Church's thesis) cannot be proved.

The Principle will be used in further chapters for the case of infinite sequences, for which independence means finiteness of the mutual information. It is clear that such an understanding is not suitable for finite objects. What we mean here by the independence of $x, y \in \mathbf{N}$ is the smallness of I(x:y). Thus it is not an absolute property (as in the case of infinite sequences), but rather a quantitative characteristic. (I(x:y) is the "deficiency of independence", see Section 1.3). The prediction that x and y are independent means that for any n the degree of our certainty that I(x:y)<n is the same as that the first n results of the "honest" toss game will not consist of total zeros. So,

Independence Principle:

If x is a sequence generated by a process in the physical world, and y is one determined (ineffectively) by a property $P(y)$, formulated with no reference to events of the physical world, then x and y are independent to within the formulation length of P, i.e. $I(x{:}y) - l(P)$ is small.

This Principle is not trivial only for those (ineffective) properties P which determine sequences with complexity essentially bigger then length of P. As an example, x might be the collection of publications of the American Mathematical Society and y might be the list of all true arithmetical assertions of a length less than 1,000,000,000.

1.5 RANDOMNESS AND INFORMATION FOR INFINITE SEQUENCES

For the perfect extension of our concepts to the space Ω one would need a non-intuitive technique of functional analysis. The following notions are not perfect, but clearly connected to the preceding sections. The next definition is a version of a definition from [L73]. In the special case of the uniform measure it is equivalent to a definition from [Ch75].

Definition 3: *The value* $D(\alpha/\mu) = \lfloor \log_2 \sup(M((\alpha_n)^*)/\mu(\alpha_n)) \rfloor$ *is called the* **deficiency of randomness of $\alpha \in \Omega$ with respect to a semimeasure μ.**

Definition 4: *The value* $I(\alpha{:}\beta) = D((\alpha,\beta)/M \otimes M)$ *is called* **the amount of information in α about β** *or the* **deficiency of their independence.**

Any r.e. measure is recursive and thus computable with any accuracy, but r.e. semimeasures of a segment of a sequence can in general be effectively approached only from below. In particular, any r.e. set of recursive upper bounds to M is bounded from below. But it may be known about some $\alpha \in \Omega$ that on its segments the r.e. semimeasure M agrees with some r.e measure μ to within a constant factor. Then, computing μ we can find $M(\alpha_n)$ to within a factor (or $K(\alpha_n) = -\lfloor \log_2 M(\alpha_n) \rfloor$ to within an additive constant). Such α we call complete, denoted $\alpha \in C$ or $C(\alpha) \Leftrightarrow \exists \mu \sup(M(\alpha_n)/\mu(\alpha_n)) < \infty$. This means that α contains all information necessary for computation of complexity of its segments. By Proposition 5, C is very extensive. By virtue of its item 2), any sequence α satisfying the Independence Principle (as x) has a completion $(\alpha,\beta) \in C$, satisfying this Principle as well.

Proposition 5:

1) The set C is closed under the application of any total recursive operators ($A(C) \subset C$) and the complement of C is of measure 0 in any recursive measure.

2) Let γ be a sequence to which a universal r.e. set is Turing reducible and α be independent of γ. Then $\beta \in \mathbf{N}^{\mathbf{N}}$ exists such that (α,β) is complete and independent of γ.

Proof: Let $\delta_m(\alpha) = \log_2 \sup(M(\alpha_n)/\mu(\alpha_n))$. Analogously with Proposition 3, δ_m is a Martin-Lof μ-test (Definition 5). Let $\alpha \in C$. Then $\exists \mu{:}\delta_m(\alpha) < \infty$. Let $\mu'(x) = \mu\{\alpha{:}A(\alpha) \supset x\}$. Then μ' is also an r.e. measure, and $\delta_{m'}(A(\alpha))$ is a Martin-Lof μ-test. Then by virtue of Proposition 6, $\delta_{m'}(A(\alpha)) \lesssim D(\alpha/\mu)$. Obviously $D(\alpha/\mu) \prec \delta_m(\alpha)$. By our assumption $\delta_m(\alpha) < \infty$. Therefore $\delta_{m'}(A(\alpha)) < \infty$ and $A(\alpha) \in C$. Obviously $\mu(C)=1$ because $\delta_m(\alpha)$ is a Martin-Lof test.

It remained to prove 2). In section 3.2 of [L70] it is shown that M (like any other r.e. semimeasure) can be obtained by means of a partial recursive operator A from a recursive measure μ: $M(x) = \mu\{\alpha{:}A(\alpha) \supset x\}$. Let $A'(\alpha) = (A(\alpha), t_A(\alpha))$, where $t_A(\alpha)$ is the sequence of values of the time of $A(\alpha)$ terms calculation. The operator A$'$ is total and, hence, $\mu'(x) = \mu\{\alpha{:}A'(\alpha) \supset x\}$ is a recursive measure. M is generated from μ' by the projector $(\alpha,t) \rightarrow \alpha$. By Proposition 7, $\mu'\{(\alpha,t){:}\ I((\alpha,t){:}\gamma)=\infty\} = 0$. Also $\mu'(\Omega - C) = 0$. Therefore, $M\{\alpha{:}\forall t \in \Omega(((\alpha,t) \notin C) \bigvee I((\alpha,t){:}\gamma)=\infty)\} = 0$. By Lemma 3, for any set A such that $M(A) = 0$, a sequence β exists on which all elements of A depend. The same is fit for any sequence to which β is reducible. Using reducibility to γ of the universal r.e. set, one can routinely check that the necessary β is computable with respect to γ. Thus γ depend on all sequences α not completable to a complete, independent of γ sequence (α,t). Q.E.D.

The α comes from (α,β) by a partial recursive (but not total) projection operator. As V'jugin [L77] has shown, partial operators can lead out of C if the time of their work is bounded by no total recursive operator. In Chapter 3 we postulate, along with a version of the Principle of independence conservation, an axiom that means intuitively that every sequence in the physical world comes from a complete one as a result of the application of a partial recursive operator.

1.6 TIME OF COMPUTATION

In this work the computational resources necessary for enumeration of various r.e. sets are, as a rule, ignored. Now we touch this question briefly. Let $t_{A(p)}$ mean time of the A(p) computation. If $K(s) \leq n$, then $\exists q\colon l(q) \leq n, A(q)=s$, where A is the optimal algorithm from Proposition 2. This q can be found by searching through all words shorter then n. This requires very large (exponential) time, even if $t_{A(q)}$ is linear. Now we give the optimal (by time) algorithm for searching for q. We assume that algorithms are realized by storage modification machines of Kolmogorov-Uspenskii.

Let $Kt_A(x/y) = \min\{(l(p) + \log_2 t_{A(p,y)})\colon A(p,y)=x\}$, where p is a binary sequence without termination mark: the algorithm A can receive, by request, the symbols of p in order until p is ended; in case of further requests A gets no reply and gives no output. $Kt_A(x) = Kt_A(x/\Lambda)$. Analogously with Theorem 1, an algorithm A exists such that Kt_A is minimal up to an additive constant, and we will denote Kt_A by Kt. There exists an algorithm G(n,y) generating the list $\{x\colon Kt(x/y)=n\}$ in time 2^n; and up to a constant factor, Kt is a minimal function with this property. (The asymptotically minimal one is $kt(x) = \min\{Kt(a)\colon x \in a \subset \mathbb{N}\}, |a| < \infty$.)

Let R(s,q) be a predicate, recognizable in time $t_{R(s,q)} < P(l(q))$, where P is a polynomial. The problems of finding q (if it exists) satisfying R(s,q) are called search problems, and the problems of discerning $\exists q(R(s,q) \& l(q) < P(l(s)))$ are called the NP-problems. Without loss of generality one can consider P linear by adding zeros to s and q. Searching through all q in the order of increasing $Kt(q/s)$ (instead of $l(q)$) gives the fastest algorithm (up to a constant factor) for solving any search problem (see [L73a]; related ideas also have been expressed by L. Adleman). In particular, it is optimal for finding q such that $A'(q)=s, l(q) \leq n$, $t_{A'(q)} \simeq l(s)$. For the case of large $t_{A'(q)}$ a similar algorithm works: namely in the definition of Kt, replace the expression $t_{A(p,y)}$ by $t_{A'(A(p,y))} = t_{A(p,y)} + t_{A'(x)}$.

Functions of the type Kt are of a particular interest for the case of algorithms with random number generators. For $f\colon \mathbb{N} \to \overline{\mathbf{R}}^+$ let $C(f) = -\log_2 \int d\omega/(t_{A(\omega)} + f(A(\omega))$, where ω is the random variable and, like above, A is the optimal algorithm minimizing C. For $F \subset \mathbb{N}$, $C(F)$ means $C(f)$, where $f(x)=0$, if $x \in F$, else $f(x)=\infty$. The above algorithm G(n,ω), generating numbers x randomly, hits any $F \subset \mathbb{N}$ in time $\leq 2^n$ with the probability $p > 1\text{–}2^{-a}$, where $\log a \gtrsim n\text{–}C(F)$. Obviously, for any such algorithm: $p < 2^{n-C(F)}$. Thus $C(F)$ determines the time which is necessary, and essentially guaranteed, for "hitting" F. A function f, with range other then just $\{\infty, 0\}$, can be interpreted as a "price" (for instance, the time) necessary for establishing $x \in F = f^{-1}(\mathbf{R}^+)$. Everything is analogous for $C(f/y) = -\log_2 \int d\omega/(t_{A(\omega,y)} + f(A(\omega,y),y))$.

A number of search (NP) problems is known which have no proof of quick solvability by deterministic algorithms, but are very quickly solvable by probabilistic ones. E.g. integer compositness and constructing "non-compressible" words q, i.e. ones equivalent to no essentially (say twice) shorter word p (q and p are equivalent if they are transformable into each other by a simple and quick algorithm). The complexity of the search problem R(s,q) for probabilistic algorithms is characterized by $C(f_s/s)$, where $f_s(q) = t_{R(s,q)}$, if R holds, and $f_s(q)=\infty$ otherwise. The relationship of this complexity with l(s) is a "randomized" version of the P=NP problem. But the problem of its relationship with the "complexity of obtaining s" looks much more interesting. More accurately: how does the function of n, $C(\{s\colon \infty > C(f_s/s) > n\})$ grow, polynomially or exponentially? A short s may exist for which it is very difficult to find q such that R(s,q), but to find such an s may be even more difficult.

Other results about the computation time may be obtained by diagonal methods. E.g. the results of [L73b] for Turing machine space remain valid for many other types of complexity: time of storage modification machines, exponent of Turing machines space etc.

2. APPLICATION TO THE FOUNDATIONS OF PROBABILITY THEORY

2.1 FOUNDATIONAL DIFFICULTIES (A historical digression)

This section is not formally related to the work. Its purpose is to clarify the context in which the problems to be studied in Chapter 2 arise (in particular the problem of the introduction of the concept of "randomness").

Hilbert's sixth problem suggests "**To treat in the same manner** (as geometry), **by means of axioms, those physical sciences, in which mathematics plays an important part; in the first rank are the theory of probabilities and mechanics.**" (see [H02])

It was generally considered that this problem is completely solved in A.N. Kolmogorov's 1933 book [K33]. However, this is only partially so. Kolmogorov's work opened great possibilities for the development of techniques of probabilistic methods and their applications. At the same time certain foundational difficulties were left unresolved. Kolmogorov noted this in the foreword to the second Russian edition of the book, where he refers to works by Kolmogorov, Zvonkin and Levin [K65, L70] for his new approach. The well-known previous attempts to overcome these difficulties by J. von Mises [vM64] and A. Church [C40] turned out to be imperfect.

The difficulties lie in the gap between intuitive probabilistic ideas and those methods which are justifiable theoretically. The premise for the use of probabilistic methods is the assumption that the result x of a physical process arises randomly with probability distribution μ. This μ is discovered or hypothesized e.g. by analogy with other processes and statistical data about them, considerations of symmetry, etc. Then, according to the naive ideas, those properties of x are indicated as probabilistic laws whose μ-probability is 1 (approximately, in the finite case). E.g. when $x = x_1, ..., x_n$, where $x_1, ..., x_n$ are independent and identically distributed (i.e. $\mu(x_1, ...x_n) = \mu'(x_1)\mu'(x_2)...\mu'(x_n)$) the law of large numbers plays an important role. For each property B it asserts that with μ-probability close to 1, the frequency of $B(x_i)$ realization is close to the probability $\mu'(B)$. In any case, subjecting x to such laws is predicted, i.e. having properties whose probability is 1 (approximately, in the finite case).

The problem is that **jointly the properties of probability 1 have probability 0 !** We cannot predict the realization of all of them simultaneously, but we should choose one or a few of them for prediction. Thus if the result had arisen before we managed to make a prediction, we could not expect to subject this result to any statistical tests. For example, classical theory provides no rigorous basis to doubt the honesty of the lottery director after his son wins the first prize in ten consecutive years, if we discover this "post factum"! We cannot subject an election to criticism when the share of votes for the ruling party in a series of consecutive years formed a sequence $0.99k_i$, even if k_i turns out to be the digits of the decimal expansion of the number π ! Of course, one can select a few "standard laws" and presume their predictions if before the beginning of the experiment this selection was not changed. **However, standard probability theory contains no principles which would allow the distinction of such standard laws from others.** Besides, it would not solve the problem of applying probability theory to events which had occurred before such a standardization (for example, to cosmology, history, geology, etc.).

The idea of solving this paradox consists in considering as "standard" those properties of probability close to 1 (in the finite case), which are "simply expressible". The objects not satisfying such a property form a simply expressible set of small measure and correspondingly small cardinality. Thus any such set element is simple itself, being determinable by its number (smaller then set cardinality) with the simple set description. This allows us, instead of indicating many simple "standard" properties, to consider a single one: "not to be a simple object". Kolmogorov's algorithmic information theory was a surprising discovery, provided a rigorous basis for the obscure notion of simplicity. In the infinite case the corresponding property is "to be random with respect to distribution μ". Then only this property is postulated to follow from the assumption about the random occurrence of an object in a process with distribution μ. This property is of μ-measure 1 and implies all other "good" properties of μ-measure 1. Attempts to introduce such a concept were also undertaken by von Mises and continued by Church for distributions μ of the Bernoulli type [vM64,C40]. However, it was found [V39] that even such standard properties as the law of the iterated logarithm do not follow from their notion of randomness (i.e. from the property of being a collective).

2.2 THE LAWS OF RANDOMNESS AND INDEPENDENCE

We consider below two types of properties of sequences $\alpha \in \Omega$. These properties are the laws of probability theory in the usual classical sense (i.e. the probability of their violation equals 0). For simplicity we restrict ourselves to the case of recursive probability distributions, though these results can be generalized to non-recursive cases as well.

The Law of Randomness: *Let us say that $\alpha \in \Omega$ satisfies this law (is random) with respect to a semimeasure μ, if $D(\alpha/\mu) < \infty$ (D is given in Definition 3).*

This means that the values of μ on segments of α are not much smaller than ones of the *a priori* probability M. (i.e. the hypothesis that α has occurred randomly with probability distribution μ is at least as consistent with reality as the *a priori* idea about occurring it with the distribution M). As we will see below, this property implies fulfillment of all effective probabilistic laws.

Definition 5: *An r.e. function $\delta{:}\Omega \to \mathbb{N}$ is called a* Martin-Lof test with respect to a recursive measure μ *(or a μ-test) iff $\forall n \ \log_2 \mu\{\alpha{:}\ \delta(\alpha) > n\} \lesssim -n$.*

It is said that a sequence withstands the test if $\delta(\alpha) < \infty$. Definition 5 is a formalization of the concept of a "good" law of probability theory. The value δ means the degree of deviation from such a law. Complete deviation occurs at $\delta(\alpha) = \infty$, the probability of which is 0. The deviations can be effectively discovered since δ is r.e. The logarithmic scale of deviations is chosen for convenience (the definition serves equally well with any other recursive scale).

Proposition 6: *Let μ be a recursive measure, then*
1) $D(\alpha/\mu)$ is a Martin- Lof test with respect to μ.
2) For any Martin- Lof test δ we have $\delta(\alpha) \lesssim D(\alpha/\mu)$.

(The proof is analogous to the proof of Proposition 3.) We see that if α withstands test D with respect to measure μ then it withstands all conceivable μ-tests. These tests correspond to the "good" laws of probability theory. What is the situation with the bad ones? This is interesting because it clarifies the relation between the algorithmic and classical approaches to probability theory. Let us give an important example of non-recursive laws.

The Law of Independence: *Let $\gamma \in \Omega$. We say that $\alpha \in \Omega$ satisfies this law if α and γ are independent, i.e. $I(\alpha{:}\gamma) < \infty$. Then $I(\alpha{:}\gamma)$ is "the degree of deviation" from this law.*

Proposition 7: *For any $\gamma \in \Omega$ and r.e. semimeasure μ, the value $I(\alpha{:}\gamma)$ satisfies the relation* $\log_2 \mu\{\alpha{:}I(\alpha{:}\gamma) \geq n\} \lesssim -n$ (Compare this inequality with one in Definition 5).

Proof: It is sufficient to show that $\log_2 M\{\alpha{:}I(\alpha{:}\ \beta) > n\} \lesssim -n$. Let $DN(\alpha/\mu) = \log_2 \sup_{x \subset \alpha}$-$\inf_{x' \supset x}(M(x')/\mu(x'))$, and $IN(\alpha{:}\beta) = DN((\alpha,\beta)/M \otimes M)$. Let us prove first for any r.e. semimeasure μ that $DN(\alpha/\mu) \gtrsim D(\alpha/\mu)$. Let $\mu_{t,x}$ be a recursive non-decreasing by t sequence of semimeasures such that $\Gamma_{x'} \bigcap \Gamma_x = \infty \Rightarrow \mu_{t,x}(x') = 0; l(x') \geq t \Rightarrow \mu_{t,x}(x') = 0; \Gamma_{x'} \bigcap \Gamma_x \neq \infty \Rightarrow \sup \mu_{t,x}(x') = \mu(x')$. Let $t(x,n) = \sup\{t'{:}\mu_{t',x}(\Lambda) \leq 2^{-n} m(x^*)\}$; $\mu'_{n,x} = \mu_{t(x,n),x}$ and $\mu' = \sum (2^n/n^2)\mu_{n,x}$. Obviously, μ' is an r.e. semimeasure and, hence, $\mu' \prec M$. Besides, $\forall x, x', n((x \subset x';\ m(x)/\mu(x) > 2^n) \Rightarrow (\mu'(x')/\mu(x') > 2^n/n^2) \Rightarrow (M(x')/\mu(x') \succeq 2^n/n^2))$. Then, by the definitions of D and DN, $D(\alpha/\mu) > n \Rightarrow DN(\alpha/\mu) \succ n - 2\log_2 n$.

It remained to show that $\log_2 M\{\alpha{:}IN(\alpha{:}\beta) > n\} \prec -n$. Let $A_{n,\beta} = \{\alpha{:}IN(\alpha{:}\beta) > n\}$ and $M(A_{n,\beta}) > 2^{-n+c}$. Being an open set, $A_{n,\beta}$ has a clopen subset A′ such that $M(A') > 2^{-n+c}$. Then k and $T \subset S_k$ exist, such that $A' = \bigcup \Gamma_x{:}x \in T$ and thus, $\forall x \in T$: $\log_2(M(x,\beta_k)/M(x)M(\beta_k)) > n$; $2^{-n+c} < \sum M(x){:}x \in T$. Then $\sum M(x,\beta_k) > 2^n M(\beta_k) \sum M(x) > 2^n M(\beta_k) 2^{-n+c}$, and therefore $\sum M(x,\beta_k) > 2^c M(\beta_k)$, which is impossible for c large enough. Q.E.D.

2.3 COVERING OF THE CLASSICAL FORMULATION OF PROBABILITY THEORY

The law of independence (as well as of randomness) is violated only with probability 0, and thus it is a law of probability theory in the customary "classical" sense. Its meaning is that a randomly generated sequence must be independent of a sequence given beforehand. This law depends only on the parameter γ, and nothing is suggested relative to γ except that it is present in advance (e.g. is uniquely defined by some property in the language of our law formulation, e.g. in formal arithmetic). Note that in the formulation of this law ($I(a{:}\gamma) < \infty$) the probability μ is not mentioned at all. This property of a can be prescribed before specifying the probability distribution of the process generating a (i.e. this property holds independently of the parameters of probability theory).

In section 1.4 we have seen that this law is more general than the usual laws of probability theory. The Independence Principle in 1.4 asserts that this law is realized not only in the usual random processes, but in any process of the physical world as well. This encourages one to bring this law outside the limits of probability theory and to consider other probabilistic laws only for those sequences which satisfy the law of independence. It turns out that this makes other laws unnecessary!

Theorem 8: *Let μ be an r.e. measure. For any set A such that $\mu(A)=0$, such a γ exists that $A \subset \{a{:}\, D(a/\mu)=\infty\} \bigcup \{a{:}\, I(a{:}\gamma)=\infty\}$, i.e. any probabilistic law (in the classical sense) is reduced to the two laws considered above.*

If we confine ourselves to sequences satisfying the law of independence (this includes, in accordance with the Independence Principle, any sequence in the physical world) then any law (recursive or not) of classical probability theory is reducible to the "law of randomness".

Proof of Theorem 8: It is easy to see that $\delta_m(a) = \log_2 \sup(M(a_n)/\mu(a_n))$ is a Martin-Lof test. Therefore if $D(a/\mu) < \infty$, then $\exists c \forall n\, M(a_n) < c\mu(a_n)$. Let $A' = A \bigcap \{a{:}\, D(a/\mu) < \infty\}$. It is clear that $M(A')=0$ follows from $\mu(A')=0$. Thus it is sufficient to prove the next

Lemma 3:
For each set A' such that $M(A')=0$ a sequence β exists on which all elements of A' depend.

Obviously a sequence $A_m \supset A'$ of open sets exists such that $M(A_m) \leq 2^{-m}$. Let $A'_m \subset S \times \mathbf{Q}^+$ be such that $((x, r_1), (y, r_2) \in A'_m, x \subset y) \Rightarrow (x=y, r_1=r_2)$; $A_m = \{a{:} \exists (x,r) \in A'_m, x \subset a\}$; $(x, r) \in A'_m \Rightarrow M(x) < r < 2M(x)$. Let β be a sequence with respect to which A'_m is r.e. Then a recursive set T exists such that $(x, r) \in A'_m \Leftrightarrow \exists y \subset \beta (x, y, m, r) \in T$; $\forall m \forall a \in \Omega \sum \pi_4(s) < 2^{-m+1}$, where $s \in T$, $\pi_2(s) \subset a$, $p_3(s)=m$ (if $s=(a_1, a_2, a_3, a_4)$ then $\pi_i(s)=a_i$). Now we shall replace T in such a way as to make $(x, y, m, r) \in T' \Rightarrow l(x)=l(y)$ fulfilled. First we replace each quadruple (x, y, m, r), where $l(x)>l(y)$ by the set of all quadruples (x, y', m, r), where $l(y')=l(x), y' \supset y$. Then we replace each quadruple (x, y, m, r) where $l(y)>l(x)$, by the set of the quadruples (x', y, m, r'), where $l(x')=l(y)$, $x' \supset x$, and $r' < M'_{y,x,r}(x')$. $M'_{y,x,r}(x')$ is given in the following way. We generate evaluations from below of numbers $M(x'){:}(x' \supset x, l(x')=l(y))$ until their sum exceeds r. If this happens we stop the process on the previous step and the result will be $M'_{y,x,r}(x')$. Otherwise $M'_{y,x,r}(x') = M(x')$. Let T″ be the set of triples (x, y, m) such that $\exists r (x, y, m, r) \in T'$. If $s \in T''$, then $r(s) = \sup\{r'{:}(s, r') \in T'\}$. The obtained T″ and r satisfy the following conditions:

1) $(x, y, m) \in X'' \Rightarrow l(x)=l(y)$
2) $A_m = \{a{:}\, \exists (x, y, m) \in T'', y \subset \beta, x \subset a\}$
3) $y \subset \beta \Rightarrow r(x, y, m) \geq M(x)$
4) $\forall m \forall \beta\ 2^{-m+1} \geq \sum r(x, y, m), y \subset \beta$

It follows from 4) that $\sum r(x, y) < \infty$, where $r(x, y) = \sum (2^m/m^2) r(x, y, m) M(y)$.
Therefore, $r(x, y) \prec M((x, y)^*)$. Obviously $\forall a \in A_m \exists x, y{:}\, l(x)=l(y), x \subset a, y \subset \beta, r(x, y, m) \geq M(x)$.
Hence $\forall a \in A_m I(a{:}\beta) \succ 2^m/m^2$. Then $\forall a \in A\, I(a{:}\beta)=\infty$. Q.E.D.

3. APPLICATIONS TO INTUITIONISTIC MATHEMATICS

3.1 A DIGRESSION

It is known that second order theories are much more complicated logically than those of the first order. The theories permitting free handling of elements of a continuum (in particular, quantification over them) belong to the second order. First order theories admit quantification only over constructive objects.

At the beginning of this century a number of mathematicians (the intuitionists) suggested that these complications are artificial and even dangerous in the sense of the possibility of paradoxes. In particular, they asserted that elements of a continuum (unlimitedly elongated sequences, unlimitedly small points, etc.) do not make sense as logically defined formal objects, but are taken from the physical world. Therefore, the applicability of usual logical operations to them is not *a priori* obvious when these operations have no analogies in the physical world. For example, in order "to apply" a classical universal quantification, one would need the ability to scan all conceivable sequences; this is not, of course, physically implementable. It was suggested that, having restricted our logical means only to such formal procedures and postulates that have closer connection with "physical intuition", we would obtain a mathematics whose proof power is more mensurable and less suspicious. The evident difficulty is in the obscurity of our physical intuition. This brings up difficulties in the choice of the formal principles which would reflect adequately the nature of sequences being generated as a result of events of the physical world. Brouwer's original idea of sequences generating by "free choice" of their terms clarifies the situation not enough, since the concept of "freedom" is itself obscure. A result is a great variety of intuitionistic principles and theories that strengthen, weaken or contradict each other.

As a rule, these theories are too strong, on the one hand but too weak on the other. They are strong to the extent that the connection of their principles with physical intuition ceases to be obvious. This is aggravated by the fact that often with respect to the possibility of the inconsistency occurrence, these theories turn out to be equivalent to the corresponding classical ones (which kills the hope for the increased "reliability" of intuitionism). They are weak to the extent that they leave unsolved many natural questions about the validity of other principles of intuitionistic reasoning. The latter fact generates multiform possibilities of extending these theories, and provides abundant material for research. However, this eliminates the possibility of obtaining a theory which gives us some feeling of completeness and is suitable for "canonization" as the universal foundation of intuitionistic mathematics.

In this section we will try partly to overcome these difficulties by using an axiom schema which corresponds to the Independence Principle (see 1.4). On the one hand, this Principle seems to have more tangible (physically clear) foundations than many arguments about the nature of "free choice". It turns out that with respect to consistency and mensurability of the proof power, the theory obtained is equivalent to the classical *first* order arithmetic. More accurately, the intuitionistic *second* order arithmetic (analysis) considered below is a *con ser vative* extension of the classical *first* order arithmetic, formulated without disjunction and existential quantification. On the other hand, it is in a sense complete. More accurately, it has no essential extension which would retain the indicated property of conservativeness (i.e. an extension gotten by adding an essentially new principle which is "purely logical" i.e. implies no new theorems of classical number theory). All these "virtues" of the theory below are connected with the fact that the Independence Postulate excludes the existence of sequences containing unbounded information about the truth of mathematical statements. It is natural to attribute the usual troubles of second order theories to such fancy "logical" sequences which in fact do not exist in the physical world.

3.2 THE PRELIMINARY CALCULUS A

Our theory AI will be constructed in Section 3.3 by adding a group of axioms to the basic calculus A, described below. A is formulated in the usual language of second order arithmetic. This language is obtained from the one of first-order arithmetic (see [KL67] section 38), by adding a countable list of second order variables denoting sequences (functions) of natural numbers and adopting the term $\alpha(t)$ and formulas $\forall\alpha F$ and $\exists\alpha F$ for any second order variable α, term t, and formula F. A formula is called absolute if it is constructed from equalities between terms with the aid of conjunction, implication, negation and universal quantification of first order variables. Absolute formulas have the identical meaning and equivalent provability in intuitionistic and classical theories. In section 0.2 a definition of the pair of numbers is given. In the same fashion we give meanings to the notion of the pair of terms, expressions $(a_1, ..., a_n)$, $\alpha(a_1, ..., a_n)$, (α, β), etc. Allowing liberties with the language we use the notation $n = Pr_t\tau$ (n equals the projection on variable t of the term τ) for the fact: $\exists s(\tau(s){=}n{+}1\&(\forall s'{<}s\colon \tau(s'){=}0))$, (i.e. (n+1) is the first non-zero term of sequence $\tau(t)$). Handling the expression $Pr_t\tau$ like a term will never cause any misunderstanding, in particular thanks to (3.2.2).

The postulates of A consist of the postulates of first order arithmetic (see [KL67], p.387, List of Postulates, Schema 8 is taken in the intuitionistic version 8′) and three second order postulates:

Schema of Choice: $(\forall n(\neg A \Rightarrow \exists x B(x))) \Rightarrow \exists\alpha\forall n(\neg A \Rightarrow B(Pr_t\alpha(n,t)))$ (3.2.1)
Markov Principle: $(\neg\forall n\, \alpha(n){=}0) \Rightarrow \exists n\alpha(n){\neq}0$ (3.2.2)
Axiom of Countability: $\exists\alpha\forall\beta\exists k\forall n\, \beta(n){=}pr_t\alpha(k,n,t)$ (3.2.3)

Axiom (3.2.3) asserts that the set of intuitionistic sequences is countable. Under the interpretation of intuitionistic sequences as sequences of results of real macro-events in the physical world, this axiom corresponds to the customary statement on the existence of a countable basis of open sets in the space-time. We do not discuss the axioms of A in detail, since they are not original. We observe only that for the construction of any complete calculus (one that satisfies Theorem 10), it is necessary to adopt either these axioms (at least under double negation) or their negation or their equivalence to some undecidable absolute statements of number theory. The last two variants seem less natural. It is known that (3.2.1 - 3.2.3) are inconsistent with the principles of continuity and bar-induction. In this respect the calculus A more resembles Kleene's theory of recursive realizability. Of course, the calculus A is still too weak. Nonetheless, we have

Proposition 9: *For any formula F an absolute P exists such that $A \vdash F \Leftrightarrow \forall\alpha\exists\beta\, P$.*

Proof of Proposition 9: The proof is based on the fact that the axioms of A allow us to introduce a concept analogous to Kleene's recursive realizability, by using the universal sequence α from axiom (3.2.3). Namely, the concept "a number x realizes a formula F with respect to a sequence α" is defined in the same way as in Kleene's book (c.f. *Introduction to metamathematics,* Chapter 2), but recursiveness of all the functions used is replaced by recursiveness with respect to α. It is easy to prove in A the equivalence of any formula F to the existence of a realization of F with respect to a universal α. The latter assertion is equivalent to the fact, that for any β there exist a sequence α and a number x realizing F with respect to (α, β). It is easy (though bulky) to check that A contains all the axioms necessary for formalizing these arguments, i.e. the deduction of $F \Leftrightarrow \forall\beta\exists\alpha\, P$, where P is the absolute formula expressing that $\alpha(0)$ is the realization of F with respect to (α, β). Q.E.D.

3.3 THE CALCULUS AI; ITS RELATIVE CONSISTENCY AND COMPLETENESS

Let P(n) be an absolute formula with a single free variable n. A finite binary sequence p is called compatible with P (denoted $p\sqsubset P$) if, $\forall n\leq l(p)$: $(p(n)=0 \Leftrightarrow P(n))$. The abbreviation $I(\alpha{:}P)$ means $\sup\{I(\alpha{:}p)\colon p\sqsubset P\}$. For a given P, the statement $I(\alpha{:}P)\leq c$ can be easily expressed by an absolute formula with free variables α, c. Using that we introduce an axiom schema with a parameter P:

$$\text{Independence Postulate: } \forall\alpha\exists c I(\alpha{:}P)\leq c \qquad (3.3.1)$$

One more informational statement must be valid in AI. The property of completeness (defined in (1.5)) is expressible by an absolute formula C(α). Our last axiom asserts feasibility of sequences completion mentioned in Proposition 5 within the bounds of the theory:

$$\forall\alpha\exists\gamma C(\alpha,\gamma) \qquad (3.3.2)$$

The double negation of this axiom follows from the weaker statement $\neg\exists\alpha\forall\gamma\neg C(\alpha,\gamma)$ inasmuch as we can use the existence of a "universal" sequence by axiom (3.2.3). Analogously, the double negation of (3.3.1) follows from the statement $\neg\exists\alpha\forall c\, I(\alpha{:}P)>c$. These weaker versions are sufficient for our purposes, but we chose the formulations (3.3.1) and (3.3.2) because they are simpler.

Definition 6. *A theory is called* absolute *if for every closed formula F an absolute (see 3.2) formula P exists such that $\neg F\Leftrightarrow P$ is provable in this theory.*

The theory of recursive realizability of S.C. Kleene is an example of a theory known to be absolute. This is the theory obtained from A by replacing (3.2.3) with

$$\text{Church's thesis (CT): } \forall\beta\exists k\forall n\colon \beta(n)=U(k,n) \qquad (3.3.3)$$

where U(k,n) is a universal partial recursive function. Condition (3.3.3) is obtainable from (3.2.3) by imposing the condition of general recursiveness on α. Our theory AI is, of course, not absolute, inasmuch as the formula $\neg\neg(CT)$ is not deducible in it, nor is it refutable, nor can it be reduced to any absolute formula. This formula however, is the only one of this sort; namely,

Lemma 4. *For any closed formula F four absolute formulas P_1, P_2, P_3, P_4 exist such that these statements are deducible in AI:*

$$\neg\neg(P_1\vee P_2\vee P_3\vee P_4);\quad \neg\neg(P_1\Rightarrow F);\quad \neg\neg(P_2\Rightarrow\neg F);\quad \neg\neg(P_3\Rightarrow(F\Leftrightarrow(CT)));\quad \neg\neg(P_4\Rightarrow(F\Leftrightarrow\neg(CT)))$$

Then to get an absolute theory an axiom is necessary implying the truth or the falsity of (CT). It turns out that this is sufficient as well. The theory AI+(CT) is equivalent to the theory of recursive realizability of S.C. Kleene and is consequently absolute. It is of little interest for our purposes since by admitting Church's thesis (3.3.3) we would exclude from consideration all non-recursive sequences (for instance the random ones). To the degree that (CT) is a very strong axiom, the axiom $\neg(CT)$ is, inversely, very weak. Thus one might have not expected that the theory AI+$\neg(CT)$ is also absolute. This fact follows from the following Theorem.

Theorem 10: *The class of absolute closed formulas deducible in $AI+\neg(CT)$ coincides with the class of absolute theorems of the classical first order arithmetic. No essential extension (i.e. one containing new theorems of the form $\neg F$) of the theory $AI+\neg(CT)$ has this property.*

Thus the theory $AI+\neg(CT)$ is a maximal conservative extension of classical arithmetic. This property is in a sense consistency and completeness relative to classical arithmetic. The basic goal of the construction of this theory was the study of the possibilities given by the axiom schema (3.3.1).

Proof of Theorem 10: It is sufficient for each closed formula F to establish a corresponding absolute formula $\overline{\mathrm{F}}$ such that: $(AI+\neg(CT))\vdash\neg F\Leftrightarrow\overline{\mathrm{F}}$, and if F itself is absolute, then $\neg F\Leftrightarrow\overline{\mathrm{F}}$ is deducible in first order arithmetic. Besides, one needs to show that every axiom F of $(AI+\neg(CT))$ will be converted into theorem $\neg\overline{\mathrm{F}}$ of first order arithmetic and the rules of deduction will be converted into derivative first order deduction rules. We shall indicate the transformation F into $\overline{\mathrm{F}}$ and explain its meaning without writing out all routine formal deductions. Due to Proposition 9 it is sufficient to restrict ourselves to formulas of the kind $F=\forall a\exists\beta P(a,\beta)$, where P is absolute. We say that F is rejected on $\gamma\in\Omega$ if for any recursive function $r:\mathbb{N}\to\mathbb{N}$ it is false that for any recursive operator $k:\Omega\to\Omega$, applicable to $\gamma, k'=r(k)$ is also applicable to γ and $P(a,\beta)$ holds, where $a=k(\gamma)$, and $\beta=k'(\gamma)$. Let μ be a recursive continuous measure. It turns out that the equivalence between $\neg F$ and the formula "F is rejected for μ-almost all γ" is deducible in $AI+\neg(CT)$.

The latter formula can be written in an absolute form and chosen as $\overline{\mathrm{F}}$. The point is that the quantifier "for almost all γ" in contrast to the quantifier "for all γ" is expressible in the first order language. Obviously the formula "F is rejected on γ", being absolute, can be presented in the form of $\forall n_k\neg\forall n_{k-1}\neg...\forall n_0\neg R(\gamma,n_0,n_1...n_k)$, where R is a recursive predicate, monotonic on each of the arguments n_i (up – for the even i and down – for the odd ones). Let us show by means of induction on i, how the predicate $\mu\{\gamma:\forall n_{i-1}\neg\forall n_{i-2}\neg...\forall n_0\neg R(\gamma,n_0...n_k)\}\geq r$ is expressed by an absolute formula. For i=0 it is trivial. Now let, at the given i, our predicate be expressed in the form of $S_i(r,n_k...n_i)$. Then $\forall n_i\forall r'>(1-r)\neg S_i(r',n_k...n_i)$ can serve as $S_{i+1}(r,n_k...n_{i+1})$. Thus, it remains to show that $\neg$F is equivalent in $AI+\neg(CT)$ to the assertion "F is rejected for μ-almost all γ".

Lemma 5: *Let μ and μ' be r.e. measures and μ be continuous. Then recursive operators P and P′ on Ω exist such that:*

1) $\forall A\subset\Omega\; m'(A)=\mu(P^{-1}(A))$

2) $\forall\omega, a\neq\omega(P(P'(\omega))\neq a\neq P'(P(\omega)))$

3) P′ (respectively P) is defined on μ' (resp. μ)-almost all non-recursive sequences.

The proof of this lemma follows from Theorem 3.1 b) in [L70]. Since the property "F is rejected on γ" is invariant with respect to any recursive reversible transformation of γ, it is sufficient to prove the equivalence of $\neg$F to "F is rejected for μ-almost all γ" just for $\mu=B_2$ (the uniform measure on Ω_2). By virtue of the same invariance and Kolmogorov's 0-1 law (see [k33]), the set A of all γ, on which F is rejected, can be only of measure 0 or 1 with respect to B_2. Hence if R is the set of all recursive sequences, the measure of $(A\cap\neg R)$ or of $(\neg A\cap\neg R)$ equals 0 with respect to any other recursive μ as well. Then by virtue of Theorem 8 a sequence exists (and it can easily be defined by an absolute formula), on which all complete γ from this set depend. The axioms of $AI+\neg(CT)$ imply that any universal sequence (from axiom 3.2.3) is non-recursive, equivalent to a complete one, and independent from sequences, defined by absolute formulas. Therefore in the case $\mu(A)=0$, F is not rejected on a universal γ and $\neg\neg F$ holds. In the opposite case $\neg$F holds by analogous reasons. These reasonings can be easily transformed to formal proofs in $AI+\neg(CT)$. Each of the two cases gives implication in one of the directions between $\neg$F and "F is rejected for μ-almost all γ". Q.E.D.

4. APPLICATION TO THE THEORY OF TURING DEGREES

4.1 INDEPENDENCE AND NEGLIGIBLE SETS

One of the natural fields for application of algorithmic information theory is the theory of Turing degrees. It is natural to interpret the recursive reducibility of α to β as that β contains all information about α, more accurately, all information except a finite amount equal to the complexity of the reducing algorithm. However, the informational concepts are subtler and less awkward than reducibility degrees. In particular, the first concepts unlike the latter ones, are invariants applicable to finite objects as well (Theorem 4 shows that I(x,y) is invariant to within a constant by all recursive reversible transformations of $\mathbb{N}$). Algorithmic information theory gives new interesting possibilities. One of them is the introduction of the concept of independence in addition to the concept of reducibility. In the language of reducibility degrees it would be possible also to say that α and β are independent if any sequence reducible to both of them is trivial (recursive). But a simple example shows that this definition is not adequate to intuition. Let α and γ be 0,1-sequences, obtained in random processes of independent trials, where the probability of $\alpha_n=0$ is 1/2, and the probability of $\gamma_n=0$ is 0.99. Let $\beta_n = \alpha_n \oplus \gamma_n$. Then, α and β are almost always such that no nontrivial sequence reduces simultaneously to both of them, though 99 percent of the contents of α and β coincide (in view of which it is hard to consider them as independent).

We shall use the concept of independence from Chapter 1 for the definition of the concept of "negligible sets" of sequences. This give us the possibility of studying properties of Turing degrees "to within this negligibility". Many exotic types of Turing degrees are known. Such are, for example, "minimal" degrees containing indivisible information (any part of the information of such a degree β, i.e. a degree $\alpha < \beta$, is equivalent to 0 or β). The existence of such degrees is proved by diagonal methods and the reality of the respective sequences would be strange. In particular, (see [Г67]) the impossibility of the appearance of such sequences in any combination of random and recursive processes was proved. One may hope that many complications of the theory of Turing degrees are caused by exotic examples of this kind, and the theory of "real degrees" is simpler. We shall see below that this is partially so, but only partially. We call a set $A \subset \Omega$ inaccessible, if its complement is closed with respect to the use of every recursive operator F (i.e. $\forall\alpha(a \not\in A \Rightarrow F(\alpha) \not\in A)$).

Proposition 11: *The following four properties of the set $A \subset \Omega$ are equivalent:*

1) A sequence $\alpha \in \Omega$ exists on which all $\beta \in A$ are dependent (i.e. $\exists\alpha\forall\beta\in A$: $I(\alpha:\beta)=\infty$).

2) A is a subset of some inaccessible set A_1, any r.e. measure of which is 0.

3) A is a subset of some inaccessible set A_1 of measure 0 in some r.e. measure μ not concentrated on a countable set ($\mu(\neg B)>0$, if B is countable).

4) M(A)=0.

Proof: 1)⇒4) and 4)⇒1) follows from Theorems 7 and Lemma 3 respectively. It is obvious that F(M), the image of M at an arbitrary recursive mapping $F:\Omega\to\Omega$, is an r.e. semimeasure and hence $F(M)\preceq M$. Therefore, if $M(A)=0$, then $A_1=\bigcup F^{-1}(A)$ is inaccessible and $A_1, M(A_1)=0$. This gives 4)⇒2)⇒3). Lemma 5 implies that 3)⇒2). Any r.e. semimeasure is the image of an r.e. measure at a recursive mapping $\Omega\Rightarrow\Omega$ (see [L70], section 3.2). This gives 2)⇒4). Q.E.D.

We call negligible the sets having any of these four properties (this neglect is, of course, based on our belief in the Independence Principle. We call i-equivalent two sets A and B if their symmetric difference is negligible. "Property of Turing degrees" means a set $A \subset \Omega$ invariant with respect to Turing equivalence. Studying them to within i-equivalence, we can exclude from consideration some properties of "exotic" "unreal" degrees which will simplify the study. We denote by K the Boolean algebra of Borel sets of Turing degrees, and by L – its factor-algebra with respect to the i-equivalence. If $A\in K$, then $\bar{A}\in L$ is the element generated by A, i.e. the i-equivalence class containing A.

4.2 TYPES OF TURING DEGREES

In (1.5) the concept of "sequence completeness" was considered. The set of incomplete sequences has a property very close to negligibility. Namely, Item 2) in Proposition 11 is obtained from Item 1) of Proposition 5 by omitting the word "total". Thus, incomplete sequences cannot arise in a process completable to a total one (in particular, in a process with the working time bounded by a total recursive function). It is natural to consider properties of the Turing degrees generated by complete sequences. It turns out that they are organized quite simply. Only four of them are not equivalent.

Theorem 12: *Let $A \subset \Omega$ be the closure with respect to Turing- equivalence of a Borel set of complete sequences. Then A is i-equivalent to one of the four sets:*
a) The empty set;
b) The set of recursive sequences;
c) The set of all complete sequences;
d) The set of all complete non- recursive sequences.

Thus, the properties of a complete sequence (to within i-negligible sets) depend only on its recursiveness, and these sequences form the two most natural elements (atoms) of the algebra L.

Proof: As it follows from Lemma 5, any set A of non-recursive sequences, invariant with respect to Turing equivalence, either is of measure 0 at any recursive measure μ, or (for any μ) contains μ-almost all non-recursive sequences. Then, by virtue of Theorem 8, a γ exists such that all the complete non-recursive sequences either from A, or from the complement of A, respectively, depend on γ. Taking into account that the invariant set A contains either all recursive sequences, or none, we obtain that A is i-equivalent to one of the four sets, mentioned in Theorem 12. Q.E.D.

Let us make a few notes about the rest of Turing degree types (containing no complete sequence). Even the proof that their union is not a negligible set turns out to be very non-trivial. It has been given by V. V'jugin [L77] who proved that L contains an infinitely divisible element and a countable number of atoms. Only two of them (namely, b) and d) of Theorem 12) contain complete sequences. V'jugin's constructions are very complicated. A portion of his proofs has not been published yet.

5. BRIEF REFERENCES AND THE BIBLIOGRAPHY

The following remarks do not claim to present the history of the question and concern mainly the works directly used above. The algorithmic information theory originated with A.N. Kolmogorov's and R.J. Solomonoff's algorithmic approach to the concepts of information, randomness and *a priori* probability. This idea was based on the fundamental discovery of an optimal (up to an additive constant) coding method for constructive objects and a recursively invariant concept of complexity arising from it (cf. [K65, S64]). "Uspekhi Mat. Nauk" reports about Kolmogorov's talks on this subject for the Moscow Mathematical Society in 1961 and consecutive years. Some R.J. Solomonoff's ideas in the field were also given in preprints and in [M62]. See also [Ma64] and [Ch66].

However, in spite of the depth of the main idea, the accuracy of the mathematical expression of the basic quantities was not perfect. Many important relationships hold only with an error degree such as the logarithm of complexity. This error rate is of course negligible in comparison to the complexity itself, but it can exceed such derived quantities as mutual information, deficiency of randomness, or conditional *a priori* probability. This is connected with the fact that subtraction and division are used in the expression of the latter quantities. Thus, the main terms of the degree of complexity can be annihilated and only terms smaller than the logarithm of the main ones remain. Therefore, these errors distorted the picture very much and hindered the development of a transparent theory.

With respect to the concept of randomness, these difficulties were overcome in very important work [ML66]. But the concept of random sequences proposed there was related only to recursive measures and did not cover other important cases. Som e other difficulties were overcome in [L70] where we introduced the concepts of the universal measure as the *a priori* probability and complexity as its logarithm. Very interesting studies of randomness concept were made by C.P. Schnorr [Sc71].

For the concept of information, the problem of giving a precise definition proved to be more difficult. The first non-trivial results were obtained by A.N. Kolmogorov and L.A. Levin in 1967. The initial definition of the mutual information [K65] was non-symmetric and had monotonicity only over one of the arguments. Kolmogorov and Levin [K68, L70] demonstrated that this value coincides approximately (up to a logarithm of the complexity) with a symmetric expression and therefore is approximately monotonic over its second argument as well. This yields the intuitive, and theoretically desirable property that a given text contains not less information about any given pair of texts than about either of them.

In [L70] the universal measure was introduced. Its logarithm (equal to the length of the shortest code over the optimal self-delimiting algorithm) turned out to be a more satisfactory complexity measure on $\mathbb{N}$ than the original proposal from [K65]. It allowed improvement of the definitions of randomness ([L73]) and information ([L74]). The new definition of information was monotonic with a constant (instead of logarithmic) error and can be extended to the case of infinite sequences. This work is connected with the very subtle and non-trivial results of P. Gacs [G74] concerning the differences between the symmetric and the asymmetric expressions for information. A number of the results of [K68, L70, L73, L74, G74] were rediscovered independently by G. Chaitin in his famous work [Ch75]. Versions of some results of the present work were reported in [L74, L76, L77].

BIBLIOGRAPHY

[C40] A. Church, On the Concept of Random Sequence *Bull. Amer. Math. Soc.* 46 (1940) 254-260

[Ch66] G.J. Chaitin, On the Length of Programs for Computing Finite Binary Sequences, I, II *J. Assoc. Comput. Math.* 13 (1966) 547-570; 15 (1968)

[Ch75] G.J. Chaitin, A Theory of Program-Size Formally Identical to Information Theory *Journal ACM* 22 (1975) 329-340

[G74] P. Gacs, On the Symmetry of Algorithmic Information *Soviet Math. Dokl.* 15 (1974) 1477-1480

[H 02] D. Hilbert, Mathematical Problems *Bull. Amer. Math. Soc.* Ser. 2, 8 (1902) 437-479

[Kl65] S.C. Kleene, R.E. Vesley *The Foundations of Intuitionistic Mathematics.* N.-Holland Publish. Co., Amsterdam (1965)

[Kl67] S.C. Kleene *Mathematical Logic* J. Wiley & Sons, Inc., New York (1967)

[K 33] A.N. Kolmogorov *Grundbegriffe der Wahrscheinlichkeitrechung* Berlin (1933) (The 2nd Russian Edition – *Osnovnye Poniatija Teorii Verojatnostej* Moscow, Nauka, 1974)

[K 65] A.N. Kolmogorov, Three Approaches to the Concept of "The Amount of Information" *Probl. Pered. Inf.=Probl. of Inf. Transm.* 1/1 (1965)

[K 68] A.N. Kolmogorov, Talk Resume *Uspekhi Mat. Nauk* 2 (1968) 201

[L70] A.K. Zvonkin, L.A. Levin, The Complexity of Finite Objects and the Development of the Concepts of Information and Randomness by Means of the Theory of Algorithms *Uspekhi Mat. Nauk = Russian Math. Surveys* 25/6 (1970) 83-124

[L73] L.A. Levin, On the notion of a Random Sequence *Soviet Math. Dokl.* 14/5 (1973) 1413

[L73a] L.A. Levin, Universal Sequential Search Problems *Probl.Pered.Inf. = Probl. Inf. Transm.* 9/3 (1973) 265-266

[L73b] L.A. Levin, On Storage Capacity for Algorithms *Soviet Math. Dokl.* 14/5 (1973) 1464-1466

[L74] L.A. Levin, Laws of Information Conservation (Non-growth) and Aspects of the Foundations of Probability Theory *Probl.Pered.Inf. = Probl. of Inf. Transm.* 10/3 (1974) 206-210

[L76] L.A. Levin, On the Principle of Conservation of Information in Intuitionistic Mathematics *Soviet Math. Dokl.* 17 (1976) 601-605

[L76a] L.A. Levin, Various Measures of Complexity for Finite Objects (Axiomatic Descriptions) *Soviet Math. Dokl.* 17/2 (1976) 522-526

[L77] L.A. Levin, V.V. V'jugin, Invariant Properties of Informational Bulks. Springer *Lecture Notes on Computer Science* 53 (1977) 359-364

[Ma64] A.A. Markov, On Normal Algorithms Which Compute Boolean Functions *Soviet Math. Dokl.* 5 (1964) 922-924

[ML66] P. Martin-Lof, The Definition of Random Sequences *Inf. and Control* 9 (1966) 602-619

[M 62] M.L. Minsky, Problems of Formulation for Artificial Intelligence *Proc. of Symp. in Applied Math.* 14 (1962) Am. Math. Soc.

[vM64] R. von Mises and H. Geiringer *The Mathematical Theory of Probability and Statistics* Academic Press N.Y. (1964)

[R 67] H. Rogers *Theory of Recursive Functions and Effective Computability* New York (1967)

[Sc71] C.P. Schnorr *Zufaelligkeit und Wahrscheinlichkeit* Springer, Lecture Notes in Math., vol. 218 (1971)

[S64] R.J. Solomonoff, A Formal Theory of Inductive Inference *Inf. and Control* 7/1 (1964) 1-22

[V 39] J. Ville *Etude critique de la concept de Collectif* Gauthier-Villars, Paris (1939)