where

$$o(N)=\beta^{-1}\ln A(N^{-1}\ln N)^{1/2}-N^{-1}\ln\left(1-\frac{d_0^2}{\ln N}\right). \quad \text{(A-14)}$$

Note that $o(N)$ is independent of l and that

$$\lim_{N\to\infty}\left[o(N)/(N^{-1}\ln N)^{1/2}\right]=\beta^{-1}\ln A. \quad \text{(A-15)}$$

That is, $o(N)$ goes to zero as fast as $(N^{-1}\ln N)^{1/2}$.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[2] T. Berger, *Rate Distortion Theory—A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[3] J. Ziv, "Coding of sources with unknown statistics—Part II: Distortion relative to a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 389–394, May 1972.

[4] D. L. Neuhoff, R. M. Gray, and L. D. Davisson, "Fixed rate universal block source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 511–523, Sept. 1975.

[5] J. Ziv, "Coding theorems for individual sequences," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 405–412, July 1978.

[6] J. K. Omura, "A coding theorem for discrete-time sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 490–498, July 1973.

[7] S. Singh and N. S. Kambo, "Source code error bound in the excess rate region," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 65–70, Jan. 1977.

[8] J. K. Omura, "Universal source coding of finite alphabet sources," in *Proc. Inter. Telemetering Conf.*, vol. 12, pp. 146–153, 1976.

[9] R. M. Gray and L. D. Davisson, "The ergodic decomposition of discrete stationary sources," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 625–636, Sept. 1974.

[10] R. M. Gray and L. D. Davisson, "Source coding theorems without the ergodic assumption," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 502–516, July 1974.

[11] V. A. Rohklin, "Lectures on the entropy theory of measure-preserving transformations," *Russian Math. Surveys*, vol. 22, no. 5, pp. 1–52, 1967.

[12] M. B. Pursley and L. D. Davisson, "Variable rate coding for nonergodic sources and classes of ergodic sources subject to a fidelity constraint," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 324–337, May 1976.

[13] D. R. Martin, "Robust source coding of finite alphabet sources via composition classes," Ph.D. dissertation, UCLA., Los Angeles, CA, 1976.

[14] J. Nedoma, "Über die Ergodizität und r-Ergodizität Stationärer Wahrscheinlichkeitsmase," *Z. Wahr.*, vol. 2, pp. 90–97, 1963.

On the Algorithmic Foundation of Information Theory

ROLAND HEIM

***Abstract*—The information content of binary sequences is defined by minimal program complexity measures and is related to computable martingales. The equivalence of the complexity approach and the martingale approach after restriction to effective random tests is used to establish generalized source coding theorems and converses. Finite state complexity and decomposable martingales are related to classical block codes and the relative frequency behavior of sequences.**

I. Introduction

INFORMATION theory, usually viewed as a branch of classical probability theory, has inherently computational concepts. Block coding schemes are examples of how to apply an algorithmic procedure when some regularities of an information source or channel are known. On the other hand, most coding algorithms thus far considered in the classical theory are of the finite state type and are, from a computational point of view, not the most general way to handle data.

Manuscript received November 5, 1976; revised January 8, 1979. This work was supported in part by the Deutsche Forschungsgemeinschaft under Grant Pf 74/7,8 and is part of the author's doctoral dissertation.

The author is with the Institute for Information Sciences, University of Tübingen, 7400 Tübingen, West Germany.

This paper is devoted to a generalization of universal source coding using Turing-computable procedures. Initiated by Kolmogoroff [5], Martin-Löf [6], and Chaitin [8]–[10], the algorithmic approach to a measure of information content and the closely related concept of randomness is now an intriguing alternative to the well established statistical and measure theoretic one. The following discussion is based on the important contributions of Schnorr [1], [2], [4], who reintroduced the martingales of Ville [17] and proposed a uniform definition of a random sequence. Whereas Schnorr's concern was to construct a consistent algorithmic theory of probability, the aim of this paper is to indicate how algorithmic information theory could be formulated. One attractive feature of this approach is its complete independence of any probabilistic concepts. Minimal program complexity measures are a quite natural and intuitively appealing way to quantify the information content of a message.

In the next section we restate the basic definitions of the so-called effective random tests introduced by Schnorr. We restrict our attention to Kolmogoroff's program complexity measure and to the martingales first

0018-9448/79/0900-0557$00.75

investigated by Ville [17] in his criticism of early, tentative definitions of random sequences [15], [16]. In Section III we formulate the familiar block codes of classical information theory in terms of martingales instead of probabilities and demonstrate the relationship between finite-state complexity and relative frequency. In Section IV we establish a generalized source coding theorem by showing that Schnorr's effective random tests can be translated into each other by a constructive procedure. We show that sequential coding schemes also can be derived from martingales. In Section V, starting with Fine's theorem about apparently convergent relative frequencies [13], we state the conditions under which an infinite sequence has a maximal compression factor $H(p)$.

II. Effective Tests and Random Sequences

We start with some basic definitions of effective random tests and random sequences. We omit Martin-Löf's recursive sequential tests [6], based on constructive measure theory, which form the bridge between classical probability theory and the algorithmic approach. Instead, we will concentrate on the computational aspects of any coding procedure. For a thorough discussion of random tests and their interrelations see [1].

For simplicity we restrict ourselves to the binary alphabet $A=\{0,1\}$. We denote the set of all finite (infinite) binary sequences by A^* (A^∞). $\Lambda\in A^*$ is the empty sequence. For $z\in A^*\cup A^\infty$ we write $z(n)=z_1z_2\cdots z_n$, $z(m,n)=z_mz_{m+1}\cdots z_n$ for $m\leqslant n$. Also $\mathbb{N}$, $\mathbb{Q}$ and $\mathbb{R}$ denote the natural, rational, and real numbers, respectively. The cardinality of a set is denoted by $\#$.

Definition 2.1: A *computable martingale* is a computable function $V: A^*\to\mathbb{R}^+$ with the following properties:

$$V(\Lambda)=1,$$
$$V(x)=\tfrac{1}{2}V(x0)+\tfrac{1}{2}V(x1), \qquad x\in A^*. \qquad (M)$$

We can think of $V(z(n))$ as our fortune after n plays in a fair binary game when we bet according to a recursive strategy, start with an initial capital of one, and the sequence $z\in A^\infty$ denotes the outcomes of the game. We assume that debts are forbidden and the payoff is according to an assumed equal distribution of the zeros and ones in z. It is straightforward to generalize this martingale concept to arbitrary distributions.

Definition 2.2: A *computable probability function* (cpf) is a computable mapping $p: A^*\to[0,1]$ with the property

$$p(\Lambda)=1,$$
$$p(x)=p(x0)+p(x1), \qquad x\in A^*. \qquad (2.1)$$

Definition 2.3: A *computable p-martingale* is a computable mapping $V^p: A^*\to\mathbb{R}^+\cup\{\infty\}$ with the property

$$V^p(\Lambda)=1$$
$$p(x)V^p(x)=p(x0)V^p(x0)+p(x1)V^p(x1), \qquad x\in A^*, \qquad (M_p)$$

for a cpf p. If $p(x)=0$, we set $V^p(x)=\infty$ and $p(x)V^p(x)=0$.

Corollary 2.1: For a cpf p and a computable martingale V^p, the product $p\cdot V^p$ is again a cpf.

Comparing (M) and (M_p), we see that (M) is the special case where $p(xa)/p(x)=p(a|x)=\frac{1}{2}$, $a\in A$, for all $x\in A^*$.

We next establish an interesting relation between p-martingales.

Lemma 2.1: Let V^p be a p-martingale and p,q cpf's with the property that $p(x)=0$ iff $q(x)=0$ for all $x\in A^*$. Then the function V^q with

$$p(x)V^p(x)=q(x)V^q(x), \qquad x\in A^*, \qquad (2.2)$$

is a q-martingale.

Proof: This is evident from (M_p).

As an instructive example, let us consider the Bernoulli cpf $p(x)=r_0^{s_0(x)}r_1^{s_1(x)}$, where $s_w(x)$ denotes the number of occurrences of $w\in A^*$ in the sequence x, and r_0,r_1 are computable positive reals with $r_0+r_1=1$. As a p-martingale, we choose the constant function $V^p(x)=1$ for all $x\in A^*$, the result of taking no risk. For the special Bernoulli cpf $p(x)=2^{-l(x)}$, $l(x)$ being the length of x, the transformation rule (2.2) reads $p(x)V^p(x)=2^{-l(x)}V(x)$ or $V(x)=2^{l(x)}r_0^{s_0(x)}r_1^{s_1(x)}$. Now let $z\in A^\infty$ have the property $\operatorname{clim}_{n\to\infty}n^{-1}s_a(z(n))=q_a$, $a\in A$ (where clim denotes constructive limit or equivalently that q_a is a computable real). Then we can write (with $\log=\log_2$, $\exp=\exp_2$)

$$V(z(n))=\exp\left[n+q_0\log r_0+q_1\log r_1+o(n)\right]$$
$$=\exp\left[n(1-H(r,q))+o(n)\right] \qquad (2.3)$$

where $H(r,q)$ is the "subjective" entropy which reflects the uncertainty of an observer who estimates the distribution to be r when the true distribution is q.

The role of a p-martingale as a random test is based on the intuitive idea that if the tested sequence has distribution p but no further regularities, then there is no strategy resulting in an unlimited growth of the gambler's capital. In the above example the selection of a constant p-martingale is equivalent to the assumption of having a p-distributed random sequence, since taking no risk is always the best we can do in this case. On the other hand, a martingale ($p(x)=2^{-l(x)}$) is an absolute random test since an equal distribution of zeros and ones is a necessary condition for maximal irregularity. In (2.3) we have an exponentially growing martingale V unless $H(r,q)=1$. The speed of growth is essentially determined by the redundancy $1-H(r,q)$.

Later we will show the intimate relationship between exponentially growing martingales and relative frequencies of words $w\in A^*$. As a consequence, martingales are finer random tests than statistical ones, as the growing speed may be polynomial, logarithmic, etc.

Before giving a precise definition of randomness and p-randomness in the algorithmic framework we introduce the set $\mathcal{G}$ of growth functions and a partial ordering $\leqslant_{\mathcal{G}}$ with respect to growth rate.

Definition 2.4: A *growth function* is a recursive, nondecreasing, unbounded function f: $\mathbb{N}\to\mathbb{N}$.

Definition 2.5: Let g,h be arbitrary functions g,h: $\mathbb{N}\to\mathbb{R}$. The relation $\leqslant_{\mathcal{G}}$ is given by

$$g\leqslant_{\mathcal{G}}h\Leftrightarrow\forall f\in\mathcal{G}:\overset{\infty}{\forall}n\in\mathbb{N}:\ g(n)\leqslant h(n)+f(n),$$

where $\forall$ means for all and $\overset{\infty}{\forall}$ means for all but finitely many. For a set of functions $\mathcal{F}$ we call $g\in\mathcal{F}$ $\mathcal{G}$-maximal iff $h\leqslant_{\mathcal{G}}g$ for all $h\in\mathcal{F}$.

It is clear how to define the relations $<_{\mathcal{G}}$, $=_{\mathcal{G}}$, $\geqslant_{\mathcal{G}}$, $>_{\mathcal{G}}$. For instance, $g=_{\mathcal{G}}h$ iff $g\leqslant_{\mathcal{G}}h$ and $h\leqslant_{\mathcal{G}}g$. Now we are ready for the martingale definition of randomness and p-randomness.

Definition 2.6: A sequence $z\in A^{\infty}$ is *random* (*p-random*) iff $V(z(n))=_{\mathcal{G}}1$ ($V^{p}(z(n))=_{\mathcal{G}}1$) for all martingales ($p$-martingales).

Note that in Definition 2.6 we do not require that the martingale be bounded. The essential point is that the growth rate of the martingale is effective only if the underlying sequence has regularities. A justification of the twofold effective Definition 2.6 is given at the end of this section.

Now we give an important theorem of Schnorr and Fuchs [23].

Theorem 2.1: A sequence $z\in A^{\infty}$ is p-random iff there is a martingale V with the property $\log V(z(n))\geqslant_{\mathcal{G}}\log V'(z(n))$ for all martingales V'.

Proof:

a) Let z be p-random. Then by definition we have $V^{p}(z(n))=_{\mathcal{G}}1$ for all p-martingales. By applying the transformation rule (2.2) we see that $V(z(n))=2^{n}p(z(n))$ is $\mathcal{G}$-maximal.

b) Let V be a $\mathcal{G}$-maximal martingale on z. By Corollary 2.1, $p(x)=2^{-l(x)}V(x)$ is a cpf. Then we have with (2.2)

$$V^{p}(z(n))=2^{-n}V(z(n))/p(z(n))=1\,.\qquad\square$$

The significance of Theorem 2.1 lies in the fact that, as we will see in Section IV, the existence of a $\mathcal{G}$-maximal martingale is equivalent to the existence of an optimal coding scheme. Since there is a maximal amount of data compression using effective methods, we may say that p-random sequences have a definite information content.

Now let us turn to the minimal program complexity measures. Let $\mathcal{P}$ be the set of partial recursive mappings Ψ: $A^{*}\to A^{*}$.

Definition 2.7: Let $x\in A^{*}$ and $\Psi\in\mathcal{P}$. The *program complexity* $K_{\Psi}(x)$ is

$$K_{\Psi}(x)=\begin{cases}\min\limits_{p\in A^{*}}\{l(p)\mid\Psi(p)=x\}\,, & \text{if } x \text{ is in the range of } \Psi,\\ l(x), & \text{otherwise.}\end{cases}$$

(There is no loss of generality in defining $K_{\psi}(x)=l(x)$ when x is not in the range of Ψ, instead of the more usual $K_{\psi}(x)=\infty$.) It is also possible to define the universal program complexity with respect to all partial recursive functions, but we do not need the definition here.

In general, the program complexity K is not recursive. Otherwise we could construct a sequence of high complexity by a recursive procedure, a contradiction. More formally, the problem of the recursiveness of K can be reduced to the undecidable halting problem of Turing machines. To get a counterpart to effective martingales, it is necessary to restrict the class of complexity measures to effective complexity measures.

Definition 2.8: A program complexity K is *effective* iff K: $A^{*}\to\mathbb{N}$ is a recursive function.

Unless otherwise stated we assume complexity measures to be effective. We omit the subscript Ψ if a specification is not necessary and write $R(x)$ for the program redundancy $R(x)=l(x)-K(x)$.

As an example of an important class of effective complexity measures, take the block coding schemes of classical information theory, e.g., Huffman coding. The encoded sequence plays the role of the program p, the algorithm Ψ is the mapping codeword $\to$ block. We will treat this in full detail in Section III.

We can characterize the same class of random sequences by effective complexity measures as we can by martingales. The corresponding definition reads as follows.

Definition 2.9: A sequence $z\in A^{\infty}$ is *random* iff

$$\left\{\inf_{i>n}R(z(i))\right\}_{n\in\mathbb{N}}=_{\mathcal{G}}1$$

for all effective complexity measures.

It is not possible to characterize p-randomness for arbitrary cpf's by complexity measures. We will return to this problem in Section V.

The reason why we use $\inf_{i\geqslant n}R(z(i))$ rather than $R(z(n))$ itself as a test for randomness is the oscillation of some complexity measures, as first noted by Martin-Löf [7]. Let us consider the following example. We select $z\in A^{\infty}$ and construct an enumeration τ: $\mathbb{N}\to A^{*}$ of A^{*} in lexicographical order. Then we define the algorithm Ψ by

$$\Psi(x)=\begin{cases}\tau(l(x))x, & \text{if } \exists i:\ \tau(l(x))x=z(i),\\ x, & \text{otherwise,}\end{cases}$$

where $\exists$ means exists. For any z there are infinitely many integers i such that $z(i)=\tau(l(x))x$. Because $l(\tau(n))\leqslant\log n$, we have, for infinitely many n, $n-K_{\Psi}(z(n))=R_{\Psi}(z(n))\geqslant\log n$. In fact the following theorem holds [7].

Theorem 2.2: Select $f\in\mathcal{G}$ such that $\Sigma_{n\in\mathbb{N}}2^{-f(n)}$ diverges. Then for all sequences $z\in A^{\infty}$, there is a complexity measure Ψ such that $\limsup_{n\to\infty}(R_{\Psi}(z(n))-f(n))>0$.

When $f(n)=[\log n]$ the series indeed diverges. As an application of martingales, we will prove in Appendix I that Martin-Löf's theorem cannot be improved.

After this digression, we now introduce some special classes of algorithms. In contrast to classical source and

channel coding procedures, successive representation of initial segments of a $z \in A^\infty$ by programs is in general not a sequential procedure. That is, from $\Psi(p)=x$ we cannot conclude that there are $q,y \neq \Lambda$ such that $\Psi(pq)=xy$. Schnorr proposed the following modification [2]:

Definition 2.10: An algorithm $\Psi \in \mathcal{P}$ is a *process* iff $\Psi(xy) \in \Psi(x)A^*$ for all $x, xy \in \text{domain}(\Psi)$.

Process complexity measures are much easier to handle and do not exhibit oscillatory behavior. In the next section we will prove that $\limsup_{n\to\infty}(R(z(n))-f(n))>0$ implies the existence of a martingale V with $\limsup_n V(z(n))/g(n)>0$, $f,g \in \mathcal{G}$, for any process redundancy R.

While decoding from a process representation is sequential, encoding in general is not. If Ψ is a process and $\Psi(p)=x$, $\Psi(q)=xy$, $y \neq \Lambda$, q is not necessarily an extension of p. In analogy to classical coding schemes, we now define coding procedures which are sequential in both directions.

Definition 2.11: A *sequential coding scheme* is a set (ψ, Ψ, g) of processes ψ, Ψ and a growth function $g \in \mathcal{G}$ such that $\psi(z(g(n)))=\bar{z}(h(n))$ and $\Psi(\bar{z}(h(n)))=z(g(n))$ for $z, \bar{z} \in A^\infty$ with $z(g(n)) \in \text{domain}(\psi)$, $\bar{z}(h(n)) \in \text{domain}(\Psi)$, $n \in \mathbb{N}$ and $h(n)$ nondecreasing.

Thus any sequential coding scheme produces an infinite sequence as a code of another infinite sequence, whereas a program representation by general algorithms is an infinite collection of finite sequences. A different class of program-representing algorithms is investigated by Chaitin [12], who restricted the admissible domains to prefix-free subsets of A^*. In classical information theory such prefix-free sets are known as instantanous codes. As a justification for Chaitin's point of view, suppose we are representing successive initial segments of a $z \in A^\infty$ by, e.g., Fortran programs. No valid Fortran program can be the prefix of another.

The reason why we have defined random sequences by twofold effective (i.e., effective complexity tests and effective growth rates) random tests is the following. Of course, it is possible to characterize $z \in A^\infty$ as random by requiring any computable martingale, respectively, the infimum of any program redundancy measure, to be bounded. However it turns out that this definition yields a different class of random sequences, and one that also differs from the class of sequences passing Martin-Löf's recursive sequential tests [1], [4]. Schnorr's modification overcomes this difficulty in a quite natural way. For instance, we get an effective program complexity measure by circumventing the halting problem and limiting the number of computation steps in any procedure. Since all "real" algorithms are step-limited, this is as irrelevant in practice as the restriction to observable growing speeds.

The other important reason is that the equivalence of the different approaches can be demonstrated constructively through constructive translations between the complexity measures. These translations will enable us to establish generalized universal source coding theorems.

III. Finite-State Complexity and Relative Frequency

In this section we will study the algorithmically simplest class of sequential coding schemes, the well-known block codes of classical information theory. There is an intimate relationship between the relative frequency $\lim_{n\to\infty} n^{-1} s_w(z(n))$ of words $w \in A^*$ in an infinite z and the possibility of data compression by a coding procedure realizable by a finite-state automaton. From our algorithmic point of view, the Shannon entropy H is the finite-state complexity per symbol. The only tribute to our twofold constructive approach is a sometimes necessary restriction to computable reals and computable limits (clim). But this is also irrelevant, as computable numbers are the only ones encountered in practice.

For a subset $C \subseteq A^*$, we use the notation C^* for the set of all finite concatenations of elements of C. A^k is the set of the 2^k finite sequences of length k. To indicate that $x \in A^k$, we write x^k.

A classical block coding scheme is a pair of homomorphisms (ω_k, Ω_k) and a codeword set $C_k \subseteq A^*$ such that

$$\omega_k(pq)=\omega_k(p)\omega_k(q), \qquad \text{for } p,q \in C_k^*,$$

$$\Omega_k(xy)=\Omega_k(x)\Omega_k(y), \qquad \text{for } x,y \in [\omega_k(C_k)]^* \subseteq (A^k)^*.$$

Thus ω_k and Ω_k are finite state processes and form a sequential coding scheme (Definition 2.11) with growth function $g(n)=kn$. Without loss of generality we assume C_k to be a complete code, that is, $\sum_{p \in C_k} 2^{-l(p)}=1$.

Now let us formulate noiseless coding in terms of martingales instead of probabilities. We define the martingale $V_k: A^* \to \mathbb{Q}^+$ by

$$V_k(x^k)=\begin{cases} 2^{k-l(\Omega_k(x^k))}, & \text{if } x^k \in \omega_k(C_k) \\ 0, & \text{otherwise} \end{cases}$$

and

$$V_k(x^{kn})=\prod_{i=0}^{n-1} V_k(x^{kn}(ik+1,(i+1)k)), \qquad \text{for } n \in \mathbb{N}. \tag{3.1}$$

Because $\sum_{x \in \omega_k(C_k)} V_k(x)=2^k$, V_k agrees with (M) on $(A^k)^*$ and can be extended to a martingale on A^* via this functional equation. Assume there is a $z \in A^\infty$ with

$$\operatorname*{clim}_{n\to\infty} (kn)^{-1} l(\Omega_k(z(kn)))=\operatorname*{clim}_{n\to\infty} (kn)^{-1} K_{\omega_k}(z(kn))=H.$$

Then from (3.1) it follows immediately that

$$V_k(z(kn))=\exp(R_{\omega_k}(z(kn)))$$

or

$$V_k(z(n))=\exp[n(1-H)+o(n)].$$

This is an example of an effective translation of a complexity measure into a martingale. Given a complexity measure K, we can construct a martingale V such that the growth rate of the redundancy R corresponds to the growth rate of V by an exponential relationship. Conversely, assume there is a martingale V_k decomposable

into factors as in (3.1), that is,

$$V_k(x^{kn+j}) = V_k(x^k_{(1)}) \cdot V_k(x^k_{(2)}) \cdot \cdots \cdot V_k(x^k_{(n)}) \cdot V_k(x^j), \quad j<k,$$

where $x^k_{(i)}$ means the ith block of length k. The construction of a codeword set C_k and the pair of homomorphic processes (ω_k, Ω_k) is immediate if there are integers l_i such that $2^{k-l_i} = V_k(x^k_{(i)})$, $V_k(x^k_{(i)}) > 0$. Note that $\Sigma_i 2^{-l_i} = \Sigma_{x\in A^k} 2^{-k} V_k(x) = 1$. If $k - \log V_k(x^k)$ is not integer valued, then we can proceed as in the well-known classical constructions: take a coarser decomposition $V_{jk}(x^{jkn}) = \Pi_{i=1}^n V_{jk}(x^{jk}_{(i)})$ and order the values $V_{jk}(x^{jk})$ according to their magnitude. Then by applying one of the well-known procedures, e.g., the Huffman coding scheme, to the table of the martingale values rather than the probabilities $p(x^{jk})$, we get a process ω_{jk} in the limit $j\to\infty$, where the relation $R_{\omega_k}(z(jkn)) = \log V_{jk}(z(jkn))$ holds asymptotically.

The decomposable martingale of (3.1) is of course also of the finite state type, since we only have to know the $2^{k+1}-1$ values $V_k(x)$, for $x \in \cup_{i=0}^k A^i$. In order to establish the announced relationship between finite state complexity and relative frequency, we state the following theorem.

Theorem 3.1:

a) Assume for a sequence $z \in A^\infty$ there exists $\operatorname{clim}_{n\to\infty} n^{-1} s_w(z(n)) = r_w$ for all $w \in A^*$. Then there is a sequence of codes C_k and homomorphisms ω_k, $k \dot\in \mathbb{N}$, such that

$$\lim_{k\to\infty} \operatorname{clim}_{n\to\infty} (kn)^{-1} K_{\omega_k}(z(kn)) = \lim_{j\to\infty} H(A|A^j) = H_\infty. \tag{3.2}$$

b) For a sequence $z \in A^\infty$ there is an injective homomorphism ω: $(C)^* \to (A^k)^*$ with the property 1)

$$\limsup_{n\to\infty} (kn)^{-1} R_\omega(z(kn)) > 0 \tag{3.3}$$

iff 2) there is a $w \in AA^*$ such that $\lim_{n\to\infty} n^{-1} s_w(z(n)) \neq 2^{-l(w)}$ or the limit does not exist.

Proof: For any martingale V, we define the factorization $\Pi^k V$ by $\Pi^k V(x^{kn+j}) = V(x^k_{(1)}) \cdot \cdots \cdot V(x^k_{(n)}) \cdot V(x^j)$. Clearly $\Pi^k V$ is again a martingale.

To prove part a), we have to show the existence of a martingale V whose factorization $\Pi^k V$ has, in the limit $k\to\infty$, the same growth rate as V on z. We do not know the values r_w; therefore we adopt an adaptive strategy by defining

$$V_j(xa) = \begin{cases} V_j(x)(1+q(w,x)), & \text{if } xa \in A^* w1 \\ V_j(x)(1-q(w,x)), & \text{if } xa \in A^* w0 \\ 1, & \text{otherwise} \end{cases} \tag{3.4}$$

where $q(w,x) = (s_{w1}(x) - s_{w0}(x))/s_w(x)$, $w \in A^j$ with the convention $s_\Lambda(x) = 1$ for all x and $q(w,x) = 0$ iff $s_w(x) = 0$.

From (3.4) we have

$$\log V_j(z(kn)) = kn \left[1 + \sum_{a\in A} \sum_{w\in A^j} \frac{s_{wa}(z(kn))}{kn} \cdot \log \frac{s_{wa}(z(kn))}{s_w(z(kn))} \right] \tag{3.5}$$

$$\log \Pi^k V_j(z(kn)) = kn \left[1 + \frac{1}{n} \cdot \sum_{i=1}^n \sum_{a\in A} \sum_{w\in X^j} \frac{s_{wa}(z^k_{(i)})}{k} \log \frac{s_{wa}(z^k_{(i)})}{s_w(z^k_{(i)})} \right]$$

where $z^k_{(i)}$ abbreviates $z((i-1)k+1, ik)$. For simplicity we only treat the case $w = \Lambda$; the generalization is obvious. Without loss of generality we assume $k \geqslant l(w) = j$.

Inspection of (3.5) reveals that we have to relate the expression $n^{-1}\Sigma_{i=1}^n F[k^{-1} s_a(z^k_{(i)})]$ to $F[(kn)^{-1} s_a(z(kn))]$, $a \in A$, where F: $[0,1] \to \mathbb{R}$ is computable and continuous. We omit the subscript a and write $\rho(z(n)) = n^{-1} s(z(n))$. By definition we have

$$\rho(z^k_{(i)}) = \rho(z(ki)) + (i-1)[\rho(z(ki)) - \rho(z(k(i-1)))].$$

Because of the existence of $\operatorname{clim}_{n\to\infty} n^{-1} s(z(n))$, there is an $h \in \mathcal{G}$ such that, for all $n \in \mathbb{N}$, $|\rho(z(n)) - r| \leqslant h(n)^{-1}$. From $[s(z(ki)) - s(z^k_{(i)})]/k(i-1) = \rho(z(k(i-1)))$, it follows that

$$\rho(z(ki)) - \rho(z(k(i-1))) = (i-1)^{-1} \left[\frac{s(z^k_{(i)})}{k} - \rho(z(ki)) \right].$$

Assuming constructively convergent relative frequencies, there are functions $g_i \in \mathcal{G}$ such that for all $k \in \mathbb{N}$

$$\left| \frac{s(z^k_{(i)})}{k} - \rho(z(ki)) \right| \leqslant g_i(k)^{-1}.$$

Combining the above relations we obtain

$$|\rho(z^k_{(i)}) - r| \leqslant h(ki)^{-1} + g_i(k)^{-1}. \tag{3.6}$$

Furthermore, $\operatorname{clim}_{n\to\infty} n^{-1} \Sigma_{i=1}^n \delta(t, \rho(z^k_{(i)}))$ exists for $t = 0, 1/k, 2/k, \cdots 1$ where $\delta(t,t') = 1$ if $t = t'$, 0 otherwise. This is because all relative frequency limits exist. From (3.6) we have the desired result

$$\lim_{k\to\infty} \operatorname{clim}_{n\to\infty} n^{-1} \sum_{i=1}^n F(\rho(z^k_{(i)})) = F(r). \tag{3.7}$$

In the case $w \in AA^*$ we have to take into account the boundary effect that the expression $\Sigma_{i=1}^n s_w(z^k_{(i)})$, $l(w) \leqslant k$, does not count the w-blocks overlapping the boundary between $z^k_{(i)}$, $z^k_{(i+1)}$, $i = 0, n-1$. We have the inequality

$$\rho_w(z(kn)) - \frac{(n-1)(l(w)-1)}{kn} \leqslant \sum_{i=1}^n \rho_w(z^k_{(i)}).$$

Implementing this modification, we have, in the limit $k\to\infty$, the same result (3.7) as for $w = \Lambda$.

Now we construct a complete code C_{kj} and injective homomorphism ω_{kj}: $(C_{kj})^* \to (A^k)^*$ for any factorization $\Pi^k V_j$. A suitably chosen diagonal sequence of code sets and homomorphisms satisfies (3.2).

To prove part b), assume $\lim_{n\to\infty} n^{-1}s_w(z(n)) = 2^{-l(w)}$ for all $w \in A^*$. Let ω: $(C)^* \to (A^k)^*$ be an arbitrary bijective homomorphism. Then

$$(kn)^{-1}K_\omega(z(kn)) = k^{-1}\sum_{w\in A^k} n^{-1}\sum_{i=1}^{n} s_w(z_{(i)}^k)\cdot l(\Omega(w)) = k^{-1}\sum_{w\in A^k}\left[2^{-l(w)} + o(1)\right]\cdot l(\Omega(w)). \tag{3.8}$$

Applying Kraft's inequality and the standard inequality $\log_e t \leqslant t-1$, we obtain

$$\sum_{w\in A^k} 2^{-l(w)}\log_e \frac{2^{-l(\Omega(w))}}{2^{-l(w)}} \leqslant \sum_{w\in A^k}\left[2^{-l(\Omega(w))} - 2^{-l(w)}\right] \leqslant 0$$

or

$$\sum_{w\in A^k} 2^{-l(w)}\log_e 2^{l(\Omega(w))} \geqslant \sum_{w\in A^k} 2^{-l(w)}\cdot\log_e 2^{l(w)}$$

and, after passing to $\log_2$,

$$\sum_{w\in A^k} 2^{-l(w)} l(\Omega(w)) \geqslant H(A^k) = k. \tag{3.9}$$

From (3.8) and (3.9) it follows that $\lim_{n\to\infty}(kn)^{-1}K_\omega \cdot(z(kn)) \geqslant 1$, that is, 1) implies 2).

Now assume 2) holds. Then we can construct a martingale whose factorization grows exponentially. If $w \in A^j$, then inspection of (3.4) reveals that, for a suitably choosen $k > j$, the expressions enclosed in brackets in (3.5) take values $\geqslant c > 0$ infinitely many times since either $\liminf_{n\to\infty} n^{-1}s_w(z(n)) \neq 2^{-l(w)}$ or $\limsup_{n\to\infty} n^{-1}s_w(z(n)) \neq 2^{-l(w)}$. Then 1) holds for an injective homomorphism associated with $\Pi^k V_j$. □

Part a) of Theorem 3.1 is the noiseless coding theorem of classical information theory stated algorithmically. Note that the only assumption was the existence of the relative frequencies, not their concrete values. This is in analogy with the work of Davisson [24], which establishes universal coding theorems under the assumption of stationarity but not ergodicity.

From part b) we learn that the possibility of compressing a sequence $z \in A^\infty$ by a factor $H < 1$ using a finite state procedure depends on the relative frequency behavior. Any deviation from the Bernouilli property $\lim_{n\to\infty} n^{-1}s_w(z(n)) = 2^{-l(w)}$ can be detected by, e.g., an exponential growth rate of a finite state martingale. An exponential growth rate of a martingale implies a linear growth rate of the corresponding program redundancy, a necessary condition for a compression factor less than one. There are two possibilities. a) The relative frequency limits exist but are different from $2^{-l(w)}$. Then the limit $\lim_{k\to\infty}\lim_{n\to\infty}(kn)^{-1}K_\omega(z(kn)) = H_\infty < 1$ exists. H_∞ is computable if the clim's exist. In this sense H_∞ is the finite state complexity per letter. b) The relative frequency does not exist. Then the finite state martingale as well as the finite state redundancy oscillate. We show in Appendix II that we can always replace a martingale by a slightly slower growing but strictly increasing martingale. This replacement, however, generally does not preserve the martingale's finite-state character. But there is a sequential coding scheme such that $\lim_{k\to\infty}\lim_{n\to\infty}(kn)^{-1}K_\Psi(z(kn)) < 1$. This follows from the generalized coding theorem which we will prove in the next section. After establishing the generalized coding theorem we will return to the relations between relative frequencies, p-martingales and complexity measures. We close this section with a martingale construction exhibiting a technical advantage of martingales. Let $\{t_n\}_{n\in I}$ be a sequence of computable and nonnegative reals summing to one. Then for a sequence $\{V_n\}_{n\in I}$ of martingales, $\Sigma_{n\in I} t_n V_n$ is again a martingale.

The martingale (3.4) has an exponential growth rate essentially determined by the relative-frequency-based redundancy $1 - H(A|A^j)$. We now construct a martingale as a superposition of various random tests having a growth rate determined by, for simplicity, the first order redundancy $1 - H(A)$.

Let A_p^n denote the subset of A^n containing all sequences with exactly p zeros. We define the martingale $V_{n,p}$ by

$$V_{n,p}(x) = \begin{cases} 2^n\cdot\binom{n}{p}^{-1}, & \text{if } x \in A_p^n A^*, \\ 0, & \text{if } x \notin A_p^n A^* \text{ and } l(x) \geqslant n. \end{cases}$$

Because $\Sigma_{x\in A^n A^k} V_{n,p}(x) = 2^{n+k}$, $V_{n,p}$ agrees with (M) on $A^n A^*$ and can be extended to A^* by this functional equation. Now we define the martingale $V_n(x) = (n+1)^{-1}\Sigma_{p=0}^n V_{n,p}(x)$ and the desired martingale $V(x) = \Sigma_{n\in\mathbb{N}}[n(n+1)]^{-1}V_n(x)$. By observing the fact that $\binom{n}{nr}$, nr integer valued, grows as $2^{nH(r)}$ for large n, it is easy to verify that $V(z(n)) = \exp[n(1-H(r)) + o(n)]$ for any p-random sequence z with Bernoulli measure $p(x) = r^{s_0(x)}(1-r)^{s_1(x)}$.

See [3] for a detailed discussion of finite-state regularities using a different approach.

IV. Generalized Source Coding

Applying a finite-state block coding procedure to $z \in A^\infty$, one starts with a partition of z into blocks of length k according to a partition function $g(n) = kn$. Then one constructs an invertible mapping between A^k and the codeword set $C_k \subseteq A^*$. As a finite-state procedure, the encoding (respectively decoding) is performed independently on previous segments of z (respectively the encoded sequence $\bar{z} \in A^\infty$). The asymptotically best result is reached, in general, in the limit $k \to \infty$.

For Turing-computable martingales we have to adopt the following modifications. Instead of $g(n)$ dividing z into blocks of equal length, we have to use growth functions of linear, polynomial, exponential, etc. order. The mapping block $\leftrightarrow$ codeword is generally performed by using the knowledge of all initial segments already coded.

Theorem 4.1: Let V: $A^* \to \mathbb{R}^+$ be a computable martingale and $z \in A^\infty$ be such that $\limsup_{n\to\infty} V(z(n))/$

$f(n)>0$ for some $f\in\mathcal{G}$. Then there exists a sequential coding scheme (ψ,Ψ,g) with $g(n+1)-g(n)=h(n)\in\mathcal{G}$ such that

$$\limsup_{n\to\infty}\left[R_\Psi(z(g(n)))-(\log f(g(n))-n)\right]>-\infty. \tag{4.1}$$

Proof: Any martingale V can be written as a pseudofactorization $V(xyz\cdots)=V(x)\cdot V_x(y)\cdot V_{xy}(z)\cdot\cdots$, $x,y,z\cdots\in A^*$. Given the pseudofactorization

$$V(z(g(n)))=\prod_{i=0}^{n-1} V_{z(g(i))}(z(g(i))+1,g(i+1)))$$

where $g(0)=0$, we construct (e.g. by the Huffman procedure) an optimal and complete code $\omega_{z(g(i))}$: $C_{z(g(i))}\to A^{h(i)}$ for all i. The code for the initial segment $z(g(n))$ is simply the concatenation

$$\psi(z(g(n)))=\prod_{i=0}^{n-1}\omega^{-1}_{z(g(i))}(z(g(i))+1,g(i+1))).$$

The decoding Ψ works as follows. Using the martingale V and $g\in\mathcal{G}$, we construct ω_Λ, C_Λ and identify the uniquely determined prefix $\bar{z}(n_1)\in C_\Lambda$ of the encoded $\bar{z}\in A^\infty$. Then $\Psi(\bar{z}(n_1))=z(g(1))$. Assume we have reconstructed the initial segment $z(g(i))$. Then from V and g we construct $\omega_{z(g(i))}$, $C_{z(g(i))}$ and identify the uniquely determined segment $\bar{z}(n_i+1,n_{i+1})\in C_{z(g(i))}$. This gives $\Psi(\bar{z}(n_{i+1}))=z(g(i))\omega_{z(g(i))}(\bar{z}(n_i+1,i+1)$.

Now we estimate the process complexity K_Ψ. In general, $\log V_{z(g(i))}(x^{h(i)})$ is not an integer. Therefore we can determine the codeword p of $x^{h(i)}$ satisfying only the inequality ([18, p. 50])

$$l(p)-\left[h(i)-\log V_{z(g(i))}(x^{h(i)})\right]<1.$$

Summing this inequality gives

$$\begin{aligned}K_\Psi(z(g(n)))&=\sum_{i=0}^{n-1}\left[h(i)-\log V_{z(g(i))}(z(g(i))+1,g(i+1)))\right]+\epsilon(n)\\&=g(n)-\log V(z(g(n)))+\epsilon(n),\end{aligned} \tag{4.2}$$

where $\epsilon(n)$ is an error term with $\epsilon(n)<n$. By assumption $\limsup_{n\to\infty}[\log V(z(n))-\log f(n)]>-\infty$ and this together with (4.2) gives

$$\limsup_{n\to\infty}\left[R_\Psi(z(g(n))-(\log f(g(n))-\epsilon(n))\right]>-\infty.$$

For a sufficiently rapidly growing g we obtain (4.1). □

The relation (4.1) shows the significance of the partition function g. To get the asymptotically best result, the term $\log f$ must outgrow the term n. If the martingale V has an exponential growth rate, a sequence of functions kn, $k=1,2,\cdots$ is sufficient. If V grows logarithmically, then g has to be selected as doubly exponential, and so on.

Theorem 4.1 has the following converse.

Theorem 4.2: Let K be an effective process complexity measure and $z\in A^\infty$ be such that $\limsup_{n\to\infty}[R(z(n))-f(n)]>0$. Then we can effectively construct a martingale V such that for a suitably chosen and arbitrarily rapidly growing $h\in\mathcal{G}$

$$\limsup_{n\to\infty}V(z(h(n)))\cdot\exp\left[-(f(h(n))-\log f(n))\right]>0. \tag{4.3}$$

Proof: We define the recursive sets $Y=\{x\in A^*\mid R(x)\geqslant f(l(x))\}$ and $Y=Y\cap A^nA^*$. Y^{pf} denotes the largest prefix-free subset of Y; that is, no element of Y^{pf} is prefix of another. Then we define the following function V_n: $A^*\to\mathbb{Q}^+$:

$$V_n(x)=\sum_{xy\in Y^{pf}}2^{f(l(xy))-l(y)}+\sum_{k<l(x)}\mathrm{I}_{Y^{pf}}(x(k))\cdot 2^{f(l(x))}, \tag{4.4}$$

where I is an indicator function. We show that V_n has the martingale property (M). To this end we consider the following four cases.

1) x is the prefix of an $xau\in Y^{pf}$, $a\in A$, $u\in A^*$. Then in (4.4) xau contributes the following ($\bar{0}=1$, $\bar{1}=0$)

$$\left.\begin{array}{ll}\text{to } V_n(x): & \frac{1}{2}\cdot 2^{f(l(xau))-l(u)}\\ \text{to } V_n(xa): & 2^{f(l(xau))-l(u)}\\ \text{to } V_n(x\bar{a}): & 0\end{array}\right\}\ \text{(first sum)}.$$

2) x is an element of Y^{pf}. Then we have $V_n(x)=2^{f(l(x))}$ (first sum), $V_n(xa)=V_n(x\bar{a})=2^{f(l(x))}$ (second sum).

3) A prefix of x is an element of Y^{pf}. Then $V_n(x)=V_n(xa)=V_n(x\bar{a})=2^{f(l(x))}$ (second sum).

4) In the remaining case all sums are zero: $V_n(x)=V_n(xa)=V_n(x\bar{a})=0$.

Cases 1)–4) cover all possibilities, so obviously V_n has the property (M). By assumption there are infinitely many prefixes of z such that $z(n)\in Y^{pf}_n$. The recursiveness of Y_n implies the existence of a $g\in\mathcal{G}$ such that $z(g(n))\in Y^{pf}_{g(n)}$. We construct a speed-up function $h\in\mathcal{G}$ such that for any $k\in\mathbb{N}$ there is an $l\in\mathbb{N}$ with $h(g(k))=g(l)$. Furthermore, since f is monotone, we can construct g in such a way that $\mathrm{clim}_{n\to\infty}\Sigma_{i=1}^n f(g(i))^{-1}<\infty$. Define

$$V(x)=\sum_{n\in\mathbb{N}}V_{h(g(n))}(x)\cdot f(g(n))^{-1}.$$

From case 2) and the special construction of g and h, we conclude that $V(z(h(g(n))))\geqslant 2^{f(h(g(n)))}/f(g(n))$. Therefore there are infinitely many n such that $V(z(h(n))\geqslant\exp[-(f(h(n))-\log f(n))]$.

Finally, we prove the finiteness of V. For the set $U_n=\Psi^{-1}(Y^{pf}_n)$, the process property of Ψ implies the inequality

$$V_n(\Lambda)=\sum_{y\in Y^{pf}_n}2^{f(l(y))-l(y)}\leqslant\sum_{u\in U^{pf}_n}2^{-l(u)}\leqslant 1. \tag{4.5}$$

Thus we can normalize $\bar{V}(x)=V(x)/V(\Lambda)$. The computability follows from the recursiveness of Y and Y_n and the construction. Because $V_n(\Lambda)\leqslant 1$, $\bar{V}$ also satisfies (4.3). □

More general coding theorems between martingales and arbitrary complexity measures can be found in [25].

From Theorem 4.1 we can derive an interesting generalization of Shannon's definition of the information content of a random variable X: $I(X)=-\log p(X)$. In (4.2) we may assign the term $g(n)-\log V(z(g(n)))$ to the true complexity of the prefix $z(g(n))$, whereas the term $\epsilon(n)$

comes into play for technical reasons. This is a parallel to classical block codes, where the average codeword length is generally not the entropy H but can only be brought to within one digit above H. In analogy to the relative frequency-based definition of information content, we define a martingale-based function $I: A^+ \to \mathbb{R}^+$ by $I(x) = l(x) - \log V(x)$ for all $x \in A^*$. Now select a cpf p and the constant p-martingale $V^p(x) = 1$. By the transformation rule (2.2), we have a corresponding martingale V given by $V(x) = 2^{-l(x)}p(x)$. This leads to the relation

$$I(x) = -\log p(x), \qquad \text{for } x \in A^*. \tag{4.6}$$

Equation (4.6) is a definition of the information content of a finite sequence x based on a cpf which is equivalent to a martingale characterization since a constant p-martingale, necessary for the transformation leading to (4.6), trivially always exists. Thus we may define:

Definition 4.1: The *information content* of a sequence $x \in A^*$ relative to a cpf p is given by $I(x) = -\log p(x)$.

It is an interesting outcome of our constructive approach that a meaningful measure of information content formally identical to Shannon's definition can be derived and that this measure applies to arbitrary, particularly nonstationary cpf's.

Theorem 4.1 justifies our introductory remark about p-random sequences as being optimally compressible. A $\mathcal{G}$-maximal martingale on a sequence $z \in A^\infty$ implies the existence of an asymptotically $\mathcal{G}$-maximal compression by a sequential coding scheme.

V. Complexity and Relative Frequency

Theorems 4.1 and 4.2 demonstrate the equivalence of the martingale and the complexity definition of a random sequence. We shall see that it is not possible to characterize p-randomness for arbitrary cpf's p via complexity measures. In what follows we investigate what is possible in this direction. We first quote a theorem of Fine [13].

Theorem 5.1: $\exists c$: $\forall m \in \mathbb{N}$, $\forall \epsilon > 0$, *and* $\forall x \in A^*$:

$$K_{\mathcal{P}}(x \mid l(x), s_1(x)) > \log \binom{l(x)}{s_1(x)} - \log(m\epsilon^4) + c$$

implies

$$\max_{m < j \le l(x)} \left| \frac{s_1(x(j))}{j} - \frac{s_1(x)}{l(x)} \right| < \epsilon.$$

Here $K_{\mathcal{P}}$ is the noneffective, universal program complexity in a modification of Loveland [21] (see [26] where it is shown that Loveland's modification does not affect our coding theorems).

Theorem 5.1 is a "good mixture" condition as it asserts that the higher the complexity, the smaller the fluctuations of the initial relative frequencies must be. Conversely, the failure of convergence forces universal complexity to be less than maximal. On the other hand, we know that maximal effective complexity (of order n) for all effective measures implies randomness which is obviously much more than simply well-mixed initial relative frequencies and convergence to $\frac{1}{2}$. What about a maximal complexity of order $nH(r)$ according to the assumption in Fine's theorem?

In this section, we only deal with the Bernoulli cpf

$$p(x) = p(0)^{s_0(x)} p(1)^{s_1(x)}, \qquad x \in A^*. \tag{5.1}$$

Theorem 5.2: Let $z \in A^\infty$ be p-random with $p(a) \neq \frac{1}{2}$, $a \in A^*$, and V a $\mathcal{G}$-maximal martingale on z. Then there exists $\bar{z} \in A^\infty$ such that $\operatorname{clim}_{n\to\infty} n^{-1} s_a(\bar{z}(n)) = p(a)$ and any $\mathcal{G}$-maximal martingale $\bar{V}$ on $\bar{z}$ is $\mathcal{G}$-equivalent to V, but $\bar{z}$ is not p-random.

Proof: We first consider the nontrivial case $p(0) \neq 0,1$. Let $x \in A^\infty$ be a random and $y \in A^\infty$ a recursive sequence. With $f, g \in \mathcal{G}$ strictly increasing and $f(n) \neq g(n)$ for all n, we define

$$\bar{z}_n = \begin{cases} x_i, & \text{if } \exists i: f(i) = n, \\ y_i, & \text{if } \exists i: g(i) = n, \\ z_n, & \text{otherwise.} \end{cases} \qquad i, n \in \mathbb{N},$$

From the $\mathcal{G}$-maximal V we derive a $\mathcal{G}$-maximal $\bar{V}$ on $\bar{z}$ as follows:

$$\bar{V}(xa) = \begin{cases} 2 \cdot \bar{V}(x) \cdot \delta(a, x_i), & \text{if } \exists i: f(i) = l(xa), \\ \bar{V}(x), & \text{if } \exists i: g(i) = l(xa), \\ 2 \cdot p(a) \cdot \bar{V}(x), & \text{otherwise.} \end{cases}$$

The best strategy on a random sequence is to take no risk. The best strategy on a recursive (that is, perfectly predictable) sequence is to bet all the capital, which results in a doubling of the capital at each step. Therefore $\bar{V}$ is $\mathcal{G}$-maximal on $\bar{z}$. If f and g are sufficiently fast growing, we have $\operatorname{clim}_{n\to\infty} n^{-1} s_a(\bar{z}(n)) = p(a)$. With the abbreviation $R(p) = 1 - H(p)$, we can write

$$\log V(z(n)) = nR(p) + \Delta(z(n)),$$

$$\log \bar{V}(\bar{z}(n)) = \left[(n - (F(n) + G(n))) \right] R(p) + F(n) + \bar{\Delta}(z(\bar{n})),$$

where $\Delta(z(n)) = \sum_{a \in A} [s_a(z(n)) - np(a)] \log p(a)$, $F(n) = \#\{i \in \mathbb{N} \mid f(i) \le n\}$, and $\bar{\Delta}(z(\bar{n}))$ and $G(n)$ are defined similarly. Since by assumption $0 < R(p) < 1$, we can construct f and g such that $(F(n) + G(n))R(p) =_{\mathcal{G}} F(n)$. By construction the term $\bar{\Delta}(\bar{z}(n))$ depends only on the unaltered rest of z in $\bar{z}$. A recursive deletion of elements cannot change the p-random character of a sequence; therefore, we have $\Delta(z(n)) =_{\mathcal{G}} \bar{\Delta}(\bar{z}(n))$ and $V(z(n) =_{\mathcal{G}} \bar{V}(\bar{z}(n))$, but because of its recursive part y, $\bar{z}$ is not p-random.

In the singular case $p(0) = 0$ or 1 and a p-random z has the form 1^∞ or 0^∞, respectively. Such a z is recursive, and on a recursive sequence the function 2^n is always a $\mathcal{G}$-maximal growth rate. Thus any recursive $\bar{z} \neq 0^\infty$, 1^∞ has the desired property. □

In the construction of $\bar{z}$ the elements of x and y are inserted so sparsely, in order to maintain $\operatorname{clim}_{n\to\infty} n^{-1} \cdot s_a(z(n)) = p(a)$, that no p-martingale on $\bar{z}$ can be of exponential order; that is, $\limsup_{n\to\infty} V^p(\bar{z}(n)) \cdot 2^{-rn} = 0$ for all $r > 0$. This follows from the next theorem.

Theorem 5.3: Let p be a Bernoulli cpf with $p(a)\neq 0,1$. Then assertions 1) and 2) are equivalent:

1) $\operatorname{clim}_{n\to\infty} n^{-1}s_a(z(n))=p(a)$ and $\lim_{n\to\infty} n^{-1}K(z(n)) \geqslant H(p)$ for all complexity measures K where this limit exists.
2) $\limsup_{n\to\infty} V^p(z(n))\cdot 2^{-rn}=0$ for all p-martingales and all $r>0$.

Proof: From (2.2) we have

$$V(z(n))=2^{-n}p(0)^{s_0(z(n))}p(1)^{s_1(z(n))}V^p(z(n)). \quad (5.2)$$

To prove 1)⇒2), we observe that (5.2) can be written as

$$V^p(z(n))=V(z(n))\cdot 2^{-nR(p)+o(n)}. \quad (5.3)$$

By assumption $V(z(n))\cdot 2^{-nR(p)}=2^{o(n)}$, otherwise there is a complexity measure giving a compression factor less than $H(p)$. This follows from Theorem 4.1 and Appendix II. Then 2) follows immediately from (5.3).

2)⇒1): From 2) we conclude $\operatorname{clim}_{n\to\infty} n^{-1}s_a(z(n))=p(a)$; otherwise we could construct an exponential p-martingale (see (2.3)). From $\limsup_n V(z(n))\cdot 2^{-nR(p)-nr+o(n)}=0$ for all $r>0$ we see that $V(z(n))$ may be written as $V(z(n))=2^{-nR'+o(n)}$ where $R'\leqslant R(p)$. This implies the second part of 1). □

In the singular case $p(0)=0$ or 1, we have either $V^0(z(n))(V^1(z(n)))=$ constant if $z=1^\infty(0^\infty)$, or $V^0(1^k0x)=\infty=V^1(0^k1x)$ for all $k\in\mathbb{N}$ and $x\in A^*$. Thus for singular Bernoulli cpf's Theorem 5.3 does not hold. See [1, ch. 5] for a detailed discussion of singular distributions.

We conclude that, except for $p(a)=\frac{1}{2}$, a maximal compression factor $H(p)$ together with converging relative frequencies implies only that any p-martingale will grow with slower than exponential rate.

An example of a probability law not of exponential order is Khintchin's law of the iterated logarithm. It is possible that a sequence admits no exponentially growing martingale but does not pass the iterated-logarithm test. The nature of this law in terms of martingale growth rates is an open problem [1].

Acknowledgment

The author wishes to thank Prof. Pfaffelhuber, Prof. Braitenberg and Mr. Sommer for many valuable suggestions. Also the author is in debt to one of the referees for exceptionally thorough reviewing and helpful comments.

Appendix I

From Theorem 4.1 we know that the constructive divergence of $\{\sup_{i\leqslant n}R(z(i))\}_{n\in\mathbb{N}}$ for any process complexity measure implies the nonrandomness of z. From Martin-Löf's Theorem 2.2 we see that this is in general not true for arbitrary complexity measures. For all sequences $z\in A^\infty$, there is a complexity measure with the property $\limsup_{n\to\infty}(R(x(n))-f(n))>0$ if $f(n)$ has the property that $\Sigma_{n\in\mathbb{N}}2^{-f(n)}$ is divergent. We will see that this condition cannot be improved.

We select $f\in\mathcal{G}$ such that $\operatorname{clim}_{n\to\infty}\Sigma_{i=1}^n 2^{-f(n)}<\infty$. Clearly, it is possible to construct $g\in\mathcal{G}$ having a sufficiently slow growth rate such that $\operatorname{clim}_{n\to\infty}\Sigma_{i=1}^n 2^{g(n)-f(n)}<\infty$ also. Now assume that

$$\limsup_{n\to\infty}(R(z(n))-f(n))>0 \quad (A.1)$$

for $z\in A^\infty$ and a complexity measure K. We define the functions V_n, which can be extended to A^* by (M), by

$$V_n(x)=\begin{cases}2^{g(n)}, & \text{if } R(x(n))>f(n),\ l(x)\geqslant n,\\ 0, & \text{otherwise.}\end{cases}$$

There are less than $2^{n-f(n)}$ elements in A^n with the property $R(x)>f(n)$. By iteration of (M), we have $V_n(\Lambda)=2^{-n}\Sigma_{x\in A^n}V_n(x)\leqslant 2^{-n}2^{n-f(n)}2^{g(n)}=2^{g(n)-f(n)}$. Thus $V(x)=\Sigma_{n\in\mathbb{N}}V_n(x)$ is finite and by construction computable. By assumption (A.1), there are infinitely many $n\in\mathbb{N}$ such that $V(z(n))\geqslant 2^{g(n)}$. For the martingale $\bar{V}(x)=V(x)/V(\Lambda)$ we have

$$\limsup_{n\to\infty}\bar{V}(z(n))\cdot 2^{-g(n)}>0, \quad (A.2)$$

that is, z is a regular sequence.

Appendix II

We show that, without loss of generality, we need consider only strictly increasing martingales. Given any martingale we can construct a slightly slower growing but nonoscillating martingale.

Given a martingale $\bar{V}$ we define the set $Y_i=\{x\in A^*\mid \bar{V}(x)\geqslant 2^i\}$. Without loss of generality we can assume Y_i to be recursive, e.g., by requiring $\bar{V}$ to be rational. Since we can approximate a computable real-valued function by a rational-valued one as closely as desired, this does not affect our definitions and theorems concerning martingale growth rates. Now we construct functions V_i by

$$V_i(x)=\sum_{xy\in Y_i^{pf}}2^{i-l(y)}+\mathrm{I}_{Y_i^{pf}AA^*}(x)\cdot 2^i. \quad (A.3)$$

Applying the arguments in the proof of Theorem 4.2 we verify that V_i has the property (M). If $z(n)\in Y_i^{pf}$, then $V_i(x(k))$ is strictly increasing for $k\leqslant n$ and remains constant for $k>n$. An upper bound for $V_i(\Lambda)$ follows from the definition of Y_i and (M):

$$V_i(\Lambda)=\sum_{y\in Y_i^{pf}}2^i2^{-l(y)}\leqslant\sum_{y\in Y_i^{pf}}2^{-l(y)}\bar{V}(y)\leqslant\bar{V}(\Lambda)=1.$$

Select a sequence $\{a_i\}_{i\in\mathbb{N}}$ such that $\operatorname{clim}_{n\to\infty}\Sigma_{i=1}^n a_i<\infty$ and an arbitrarily rapidly growing $g\in\mathcal{G}$. Define the function $V'(x)=\Sigma_{i\in\mathbb{N}}a_iV_{g(i)}(x)$ and the martingale $V(x)=V'(x)/V'(\Lambda)$. Suppose $\bar{V}(z(n))$ is unbounded for some $z\in A^\infty$. Because $V(xa)\leqslant 2V(x)$, $a\in A$, $x\in A^*$, for any martingale, there is for any $i\in\mathbb{N}$ a $k\in\mathbb{N}$ such that $2^{g(i)}\leqslant\bar{V}(z(k))<2^{g(i)+1}$. This implies the inequality $V(z(k))\geqslant V'(z(k))\geqslant a_i2^{g(i)}$ (we can assume $z(k)\in Y_{g(i)}^{pf}$). By construction, $V(z(n))$ is strictly increasing and we have the following assertion: to any $\bar{f}\in\mathcal{G}$ there is an $f\in\mathcal{G}$ with $f<_{\mathcal{G}}\bar{f}$ but arbitrarily close to $\bar{f}$ in the $\mathcal{G}$-ordering, such that $\limsup_{n\to\infty}\bar{V}(z(n))/\bar{f}(n)>0$ implies $\lim_{n\to\infty}V(z(n))/f(n)>0$.

References

[1] C. P. Schnorr, "Zufälligkeit und Wahrscheinlichkeit," *Lecture Notes in Mathematics* 218. New York, Berlin, Heidelberg: Springer Verlag, 1971.
[2] ——, "Process complexity and effective random tests," *J. Comp. and Syst. Sci.*, vol. 4, pp. 376–388, 1973.
[3] C. P. Schnorr and H. Stimm, "Endliche Automaten und Zufallsfolgen," *Acta Informatica*, vol. 1, pp. 345–359, 1972.

[4] C. P. Schnorr, "A uniform approach to the definition of randomness," *Math. System Theory*, vol. 5, pp. 9–28, 1971.
[5] A. N. Kolmogorov, "Three approaches to the definition of the concept 'quantity of information'," *Probl. Pederachi Inform.*, vol. 1, pp. 3–11, 1965.
[6] P. Martin-Löf, "The definition of random sequences," *Inform. Contr.*, vol. 6, pp. 602–619, 1966.
[7] ——, "Complexity oscillations in binary sequences," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 19, pp. 225–230, 1971.
[8] G. Chaitin, "On the length of programs for computing finite binary sequences," *J. ACM*, vol. 13, pp. 547–569, 1966.
[9] ——, "On the length of programs for computing finite binary sequences: statistical considerations," *J. ACM*, vol. 16, pp. 407–422, 1969.
[10] ——, "On the difficulty of computations," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 5–9, 1970.
[11] ——, "Information-theoretic computational complexity," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 10–15, 1974.
[12] ——, "A theory of program size formally identical to information theory," *J. ACM*, vol. 22, pp. 329–340, 1975.
[13] T. L. Fine, "On the apparent convergence of relative frequencies and its implications," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 251–257, 1970.
[14] H. Rogers, *Theory of Recursive Functions and Effective Computability*. New York: McGraw-Hill, 1967.
[15] R. von Mises, "Grundlagen der Wahrscheinlichkeitstheorie," *Math. Z.* vol. 5, pp. 52–99, 1919.
[16] A. Church, "On the concept of a random sequence," *Bull. Amer. math. Soc.*, vol. 46, pp. 130–135, 1940.
[17] J. Ville, *Etude critique de la notion de Collectif*. Paris: Gauthier–Villars, 1939.
[18] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[19] M. Blum, "A machine-independent theory of recursive functions," *J. Assoc. Comp. Mach.*, vol. 14, pp. 322–336, 1967.
[20] R. P. Daley, "Minimal-program complexity of pseudo-recursive and pseudo-random sequences," *Math. Systems Theory*, vol. 9, pp. 83–94, 1975.
[21] D. W. Loveland, "A variant of the Kolmogorov concept of complexity," *Inform. Control*, vol. 15, pp. 510–526, 1969.
[22] L. A. Levin, "On the notion of a random sequence," *Soviet. Math. Dokl.*, vol. 14, 1973.
[23] C. P. Schnorr and P. H. Fuchs, "General random sequences and the concept of learnable sequences," Univ. of Frankfurt, preprint, 1975; see also P. H. Fuchs, "Programmkomplerität, Zufälligkeit, Lernbarkeit," Univ. of Frankfurt, doctoral dissertation, 1975.
[24] L. D. Davisson, "Universal noiseless coding," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 783–795, 1973.
[25] R. Heim, "Zufall und Information," Univ. of Tübingen, doctoral dissertation, 1976.

Higher Dimensional Hadamard Matrices

PAUL J. SHLICHTA

***Abstract*—The concept of a Hadamard matrix as a binary orthogonal matrix is extended to higher dimensions. An n-dimensional Hadamard matrix $[h_{ijk\cdots n}]$ is defined as one in which all parallel $(n-1)$-dimensional layers, in any axis-normal orientation, are uncorrelated. This is equivalent to the requirements that $h_{ijk\cdots n}=\pm 1$ and that**

$$\sum_p \sum_q \sum_r \cdots \sum_y h_{pqr\cdots ya} h_{pqr\cdots yb} = m^{(n-1)}\delta_{ab}$$

where $(pqr\cdots yz)$ represents all permutations of $(ijk\cdots n)$. A "proper" n-dimensional Hadamard matrix is defined as a special case of the above in which all two-dimensional layers, in all axis-normal orientations, are Hadamard matrices, as a consequence of which all intermediate-dimensional layers are also Hadamard matrices. Procedures are described for deriving three- and four-dimensional Hadamard matrices of varying propriety from two-dimensional Hadamard matrices. A formula is given for a fully proper n-dimensional matrix of order two, which can be expanded by direct multiplication to yield proper $(2^r)^n$ Hadamard matrices. It is suggested that proper higher dimensional Hadamard matrices may find application in error-correcting codes, where their hierarchy of orthogonalities permit a variety of checking procedures. Other types of Hadamard matrices may be of use in security codes on the basis of their resemblance to random binary matrices.

Manuscript received November 20, 1978; revised February 12, 1979. This work was presented in part at a meeting of the American Physical Society, August 26, 1971, Seattle, WA.

The author is with the Technical Staff, Jet Propulsion Laboratory, California Institute of Technology, Building 77, Pasadena, CA 91103.

I. Introduction

HADAMARD matrices [1], [2] may be defined as binary orthogonal matrices or, equivalently, as binary matrices in which all parallel rows or columns are uncorrelated; i.e., the m^2 matrix $[h_{ij}]$ is a Hadamard matrix if all $h_{ij}=\pm 1$ and

$$\sum_i h_{ia}h_{ib} = \sum_j h_{aj}h_{bj} = m\delta_{ab}. \tag{1}$$

Aside from the trivial cases of $m=1$ and $m=2$, these conditions can be satisfied only if m is a multiple of four. Following the work of Paley [3], recent research has focused on demonstrating the existence of at least one Hadamard matrix for all values of $m=4t$ [4], [5] or on generating families of special (e.g., skew) Hadamard matrices [6], [7].

In recent years, Hadamard matrices have been used successfully for a variety of practical applications such as switching networks [8], error-correcting codes and signal processing [9], [10], and high-speed multiplex spectrometry [11]. Several other applications are foreseeable or currently under investigation [12], [17].

In the light of this interest, it is surprising that no attempt has been made to generate or utilize higher

0018-9448/79/0900-0566$00.75