UNIVERSAL GAMBLING SCHEMES
AND THE
COMPLEXITY MEASURES OF KOLMOGOROV AND CHAITIN

BY

THOMAS M. COVER

TECHNICAL REPORT NO. 12
OCTOBER 14, 1974

PREPARED UNDER THE AUSPICES
OF
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
CONTRACT AFOSR #F44620-74-C-0068

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

UNIVERSAL GAMBLING SCHEMES
AND THE
COMPLEXITY MEASURES OF KOLMOGOROV AND CHAITIN

by

THOMAS M. COVER

TECHNICAL REPORT NO. 12
OCTOBER 14, 1974

PREPARED UNDER THE AUSPICES
OF
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
CONTRACT AFOSR #F44620-74-C-0068

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Universal Gambling Schemes
and the
Complexity Measures of Kolmogorov and Chaitin

Thomas M. Cover†
Stanford University

Abstract

Let $x \in \{0,1\}^\infty$ and $x(n) = (x_1, x_2, \ldots, x_n)$. Consider a sequential gambling scheme on binary sequences defined informally as follows:

1) The initial capital $S(\Lambda)$ is one unit;

2) The amount wagered on the outcome x_{k+1} is a function solely of the observations $x_1, x_2, \ldots, x_k$;

3) The amount wagered is never more than the current amount of capital $S(x(k))$;

4) Wagers are paid at even odds.

The gambling capital $S(\cdot)$ will be shown to be achievable by a sequential gambling scheme iff $\forall n$, $\forall x(n)$, $S(x(n))2^{-n}$ are the marginal distributions for some stochastic process $\{X_i\}_1^\infty$. As a consequence, it will be shown that if $\{X_i\}_1^\infty$ is a Bernoulli sequence with parameter 1/2 and S is a r.v., there exists a sequential gambling scheme achieving $S_n \to S$ wp 1 iff $S \geq 0$, $ES \leq 1$.

Let $C(x_1, x_2, \ldots, x_n)$ denote the length of the shortest binary program which computes the sequence $x(n) \in \{0,1\}^n$, where no program is allowed to be the prefix of another. This is the algorithmic entropy of Chaitin, which is nearly equivalent to the Kolmogorov complexity. A

† Portions of this paper were delivered at the International Symposium on Information Theory, Tallin, Estonia, U.S.S.R., June 1973 and at the IMS annual meeting, New York, N.Y., December 1973. The research was supported by contract AFOSR #F44620-74-C-0068.
Submitted September, 1974 to Annals of Statistics.

universal gambling scheme will be exhibited achieving $S(x(n)) \geq 2^{n-C(x(n))}$, $\forall n$, $\forall x \in \{0,1\}^{\infty}$. The gambling scheme yields what we feel to be a satisfactory inference procedure achieving the goals of Solomonoff.

Thus the initial capital is doubled $n - C$ times where C is the length of the shortest "law" of the sequence, and $n - C$ represents the remaining determinism within the sequence. It is also shown that if $\{X_n, n \geq 1\}$ is an ergodic binary stochastic process with entropy H, this same gambling scheme will achieve capital that grows as $\exp_2(n(1-H(X))$ with probability one. The latter result generalizes Breiman's results on gambling in favorable independent games.

1. Introduction

In this paper we investigate sequential gambling schemes in as concrete detail as possible. We shall find that any distribution of final capital is achievable subject to a certain simple constraint. This provides the underlying reason for the exponential growth of capital in the work of Kelley [6] and Breiman [1] and yields immediate extension of their results to arbitrary ergodic processes. This work also extends the investigations in Cover [4].

In particular, when we apply these results to a gambling goal intimately related to the Kolmogorov complexity of sequences (due to Kolmogorov [7] and Chaitin [2]), we show the existence of a sequential gambling scheme which doubles the gambler's money approximately $n - K$ times on a binary sequence of length n, where K is the length of the shortest binary description of the sequence. Thus the Kolmogorov complexity is invested with a quantitative operational significance in terms of gambling and prediction. The gambling approach to complexity has also been investigated by Schnorr [10,11], who applies his results to the definition of random sequences.

2. Preliminaries

Let $\{0,1\}^*$ denote the set of all binary sequences of finite length, including the empty sequence. For any $x = (x_1,x_2,\ldots) \in \{0,1\}^\infty \cup \{0,1\}^*$, let $x(n) = (x_1,x_2,\ldots x_n)$ denote the first n terms of x. Let $\ell(x)$ denote the length of x for $x \in \{0,1\}^*$.

<u>Definition</u>: Let $b(\cdot|\cdot) : \{0,1\} \times \{0,1\}^* \to \mathbb{R}$ be called a <u>sequential gambling system</u> if the following conditions are satisfied:

$$b \ge 0$$

$$\sum_{x_{k+1}} b\left(x_{k+1} \middle| x(k)\right) = 1 \ , \qquad (1)$$

for all $k = 0,1,2,\ldots,$ and all $x(k) \in \{0,1\}^k$.

The interpretation is that $b(x_{k+1}|x(k))$ is the proportion of the current capital $S(x(k))$ which is bet that the next outcome will be x_{k+1}. The winning bet is paid off at even odds.

The gambling system induces the following sequence of capitals on the sequence $x \in \{0,1\}^\infty$:

$$S(\Lambda) = 1$$
$$S\left(x(n+1)\right) = 2b\left(x_{n+1} \middle| x(n)\right) \ S\left(x(n)\right) \ . \qquad (2)$$

Here Λ denotes the empty string, and $S(\Lambda) = 1$ is the initial capital. Thus $b(1|x(n))$ indicates a gamble of all the current capital on $x_{n+1} = 1$, thus resulting in $S = 0$ (now and for all future time) if $x_{n+1} = 0$; while $b(1|x(n)) = 1/2$ corresponds to keeping all the current capital in the gambler's pocket until the next bet.

Definition: We shall say $S : \{0,1\}^* \to \mathbb{R}$ is achievable if there exists a sequential gambling scheme with initial capital $S(\Lambda) = 1$ achieving $S(x)$, for all $x \in \{0,1\}^*$.

The following simple theorem is the key to the results obtained in this paper.

Theorem 1.

a) The capital function $S : \{0,1\}^* \to \mathbb{R}$ is achievable by a sequential gambling system if and only if

$$\forall n,\ \forall x(n),\ S\big(x(n)\big)\, 2^{-n} = p\big(x(n)\big) \tag{3}$$

are the marginal distributions for some stochastic process $\{X_i\}_1^\infty$.

b) Equivalently, S is achievable if and only if

$$\begin{aligned} &S(x) \geq 0 \\ &S(x) = \frac{1}{2}\, S(x1) + \frac{1}{2}\, S(x0) \\ &\text{for all } x \in \{0,1\}^* . \end{aligned} \tag{4}$$

i.e., iff S is a martingale with respect to Bernoulli measure.

Remark: The extension of this result to finite valued alphabets with $S : \{1,2,\ldots,m\}^* \to R$ with wagers paid at $m - 1$ to 1 odds yields the condition that $S(x(n))m^{-n}$ must be the marginals of some stochastic process.

Proof: It is well known that any sequential gambling system yields a martingale. The converse, that any martingale-consistent S can be achieved by a sequential gambling system, yields all the results in this paper.

To prove that b yields S satisfying Eq. (4), observe from Eq. (2), that

$$S(x1) = 2b(1|x)\ S(x)$$
$$S(x0) = 2b(0|x)\ S(x) \tag{5}$$

and hence that

$$\frac{1}{2}\Big(S(x1) + S(x0)\Big) = \Big(b(1|x) + b(0|x)\Big)\ S(x) = S(x),\ \forall\ x \in \{0,1\}^* . \tag{6}$$

Also, $b \geq 0$ and Eq. (2) imply that

$$S\big(x(n)\big) = 2^n \prod_{k=1}^{n} b\big(x_k|x(k-1)\big)\ S(\Lambda) \geq 0,\ \forall n,\ \forall x(n) . \tag{7}$$

Conversely, to achieve a given $S(\cdot)$ satisfying Eq. (4), set

$$b\big(x_{k+1}|x(k)\big) = S\big(x(k)x_{k+1}\big)\big/2S\big(x(k)\big) . \tag{8}$$

This b precisely satisfies Eq. (2) and thus qualifies as a sequential gambling system. That b achieves $S(\cdot)$ follows from

$$\begin{aligned} S'\big(x(n)\big) &= 2^n \prod_{k=1}^{n} b\big(x_k|x(k-1)\big) \\ &= 2^n \prod_{k=1}^{n} \Big(S\big(x(k)x_{k+1}\big)\big/2S\big(x(k)\big)\Big) \\ &= S\big(x(n)\big)\big/S(\Lambda) = S\big(x(n)\big) . \end{aligned} \tag{9}$$

Finally, to prove part (a), note that $p : \{0,1\}^* \to \mathbb{R}$ is the set of marginals for a stochastic process if and only if

$$p(\Lambda) = 1, \; p(x) \geq 0$$
$$p\big(x(k)\big) = p\big(x(k)1\big) + p\big(x(k)0\big) \;, \tag{10}$$
$$\text{for all } k, \quad \text{for all } x(k) \in \{0,1\}^k \;.$$

Substituting $p(x) = S(x)\, 2^{-\ell(x)}$ from Eq. (3), we see that $S(\cdot)$ satisfies Eq. (10) and conversely. Thus (a) and (b) are equivalent.

Corollary: Equivalent expressions for the betting scheme achieving $S(\cdot)$ are given by

$$\text{(i)} \quad b_1\big(x(k)\big) = S\big(x(k)1\big) \Big/ \Big(S\big(x(k)0\big) + S\big(x(k)1\big)\Big)$$
$$\text{(ii)} \quad b_1\big(x(k)\big) = S\big(x(k)1\big) \Big/ \Big(2S\big(x(k)\big)\Big) \tag{11}$$
$$\text{(iii)} \quad b_1\big(x(k)\big) = p\big(x_{k+1}=1 \,|\, x(k)\big) \;,$$

where p is the probability assignment induced by S given in Eq. (3).

3. Special Cases

Theorem 2.

For a given integer n, the function $S : \{0,1\}^n \to \mathbb{R}$ is achievable by a sequential gambling system if and only if

$$\text{i)}\quad S\big(x(n)\big) \geq 0, \ \forall x(n) \in \{0,1\}^n ,$$

$$\text{ii)}\quad (1/2)^n \sum_{x(n)} S\big(x(n)\big) = 1 . \qquad (12)$$

Proof: Let $p(x(n)) = (1/2)^n S(x(n))$, $\forall x(n) \in \{0,1\}^n$. From conditions i) and ii) we see that p satisfies the conditions of Theorem 1. Thus S is achievable. The betting that achieves S is seen to be

$$b_1\big(x(k)\big) = p\big(x(k)1\big)\Big/p\big(x(k)\big)$$

$$= \sum_{x(k)1 \sqsubseteq x(n)} S\big(x(n)\big) \Big/ \sum_{x(k) \sqsubseteq x(n)} S\big(x(n)\big) , \qquad (13)$$

where $x(k) \sqsubseteq x(n)$ means that $x(k)$ is the prefix of the sequence $x(n)$.

For a given subset $T \subseteq \{0,1\}^n$, define

$$S^* = \max_{b} \min_{x(n)\in T} S\big(x(n)\big) , \qquad (14)$$

where $S(\Lambda) = 1$ and the maximum is taken over all sequential gambling systems. Thus S^* is the most that can be guaranteed, given knowledge only of the subset T in which $x(n)$ lies.

Theorem 3.

$S^* = 2^n/|T|$, where $|T|$ denotes the number of elements in T.

Proof: Clearly $S(x(n)) = 2^n/|T|$, $x(n) \in T$; $S(x(n)) = 0$, $x(n) \notin T$, is minimax and satisfies Theorem 1.

Example 1 (Bar Bet): Suppose one has a deck of $n = 52$ cards, 26 of them red and 26 of them black. Then a minimax gambling scheme will achieve

$$S^* = 2^n / |T| = 2^{52} \Big/ \binom{52}{26} = 9.081 \tag{15}$$

units uniformly in all $\binom{52}{26}$ sequences of red and black. This is equivalent to knowing three cards perfectly. The betting scheme that achieves this goal is to bet a proportion $B/(B+R)$ of the current capital on black and $R/(B+R)$ of the current capital on red, where B and R are the respective numbers of black and red cards remaining in the deck.

Example 2 (Random Time): Suppose it is known a priori that $\sum_1^n x_i = 1$, i.e.,

$$T = \left\{ x(n) \ : \ \sum_{i=1}^{n} x_i = 1 \right\} . \tag{16}$$

Here x_i can be considered as the indicator function for a random time in $\{1,2,\ldots,n\}$. Then $|T| = n$ and $S^* = 2^n/n$. S^* is achieved by betting

$$b\big(0 \,|\, x(k)\big) = \begin{cases} 1 & , \ \sum_1^k x_i = 1 \\ \frac{n-k-1}{n-k} & , \ \sum_1^k x_i = 0 \ , \end{cases} \tag{17}$$

$$b\big(1 \,|\, x(k)\big) = 1 - b\big(0 \,|\, x(k)\big) \ .$$

Thus, for example, $S^* = 102.4$ units can be won gambling on the time of occurence of a 1 in a string of length 10.

Example 3 (Hypothesis Testing): Let T be a rejection region for a test of the hypothesis $H_0 : X_1, X_2, \ldots, X_n$ are Bernoulli random variables with parameter 1/2. Let $\Pr\{T|H_0\} = \alpha$. Thus $\Pr\{T|H_0\} = \sum_{x \in T} p(x|H_0) = |T| 2^{-n} = \alpha$, so $|T| = \alpha 2^n$, and there must exist a gambling scheme b^* achieving $S^* = 1/\alpha$. For example, if $\alpha = .05$, then $S^*(x(n)) = 20$, $\forall x(n) \in T$. Thus the level of the test and the amount of money that can be made by a gambling scheme naturally associated with that test are inversely related. See also Martin-Lof [9] and C.P. Schnorr [11].

Example 4 (Degrees of Freedom): Let T be the set of all binary n-sequences generated by k^{th} degree polynomials by examination of the sign. Specifically, $x(n) \in T$ if $\exists\, a_0, a_1, \ldots, a_k \in R$,

$$f(m) = \sum_{j=0}^{k} a_j m^j ,$$

$$x_i = \begin{cases} 1 , & f(i) \geq 0 \\ 0 , & f(i) < 0 \end{cases} , \quad i = 1,2,\ldots,n . \tag{18}$$

Then, since a k^{th} degree polynomial can change sign at most k times, generating $\leq k + 1$ runs of arbitrary lengths $n_1, n_2, \ldots, n_{k+1}$, we have

$$|T| = 2 \sum_{i=0}^{k} \binom{n-1}{i} , \tag{19}$$

and

$$S^* = 2^{n-1} \Big/ \sum_{i=0}^{k} \binom{n-1}{i} . \qquad (20)$$

Thus the number of degrees of freedom k of the polynomial is related to the amount of money that can be earned. Note $S^* \approx 2^{n-k} \ln n$, for $k \ll n$. Fuller development of the relation of degrees of freedom, complexity, and gambling predictability will be given in Cover [5].

4. Achieving an arbitrary distribution of capital

A gambling system, the kind sold in and around casinos, can be considered as a scheme for exchanging one unit of capital for a random variable S on some probability space such that its (mathematical) expectation is less than or equal to one. Suppose that one is allowed to bet at even odds on a sequence of fair coin tosses. We shall show that any nonnegative r.v. S with mean ≤ 1 is achievable. This theorem characterizes all gambling systems on fair coin tosses. The proof involves finding sequentially refined partitions of the positive real line compatible with the gambling constraints. Let F be the distribution function for S. Let S_n denote the random amount of capital after the n^{th} outcome.

Theorem 4.

Let F be a distribution function. There exists a gambling scheme such that $S_n \to S$ wp 1, where $S \sim F$, if and only if $S \ge 0$, $ES \le 1$, i.e., iff $F(0) = 0$, $\int_0^\infty s dF(s) \le 1$.

Proof: Let $X_1, X_2, \ldots$ be the Bernoulli sequence with parameter $p = 1/2$, on which we shall place the bets. Because of the broad conditions for the achievability of $S : \{0,1\}^* \to \mathbb{R}$ given in Theorem 1, there are many possible assignments of $S(\cdot)$ to $\{0,1\}^*$ achieving the desired goal. We shall consider a certain natural assignment.

Let $F(s) = \Pr\{S \le s\}$ be the d.f. for S, the desired limiting r.v. capital. For $x = (x_1, x_2, \ldots) \in \{0,1\}^\infty$, let $.x = .x_1 x_2 \ldots$ and $.x(n) = .x_1 x_2 \ldots x_n$ denote the dyadic representations of points in the unit interval. Thus $.X$ is uniformly distributed on $[0,1]$. The idea will be to win the amount $S(x) = F^{-1}(.x)$ if the observed binary sequence is x, where $F^{-1}(t) = \inf\{s : F(s) = t\}$. Then

$$\begin{aligned} \Pr\left\{S \le s\right\} &= \Pr\left\{F^{-1}(.X) \le s\right\} \\ &= \Pr\left\{.X \le F(s)\right\} = F(s) , \end{aligned} \tag{21}$$

and S has the desired distribution.

The existence of a sequential betting scheme achieving $S_n \to S$ must now be established. Define

$$S_n(x) = S\left(x(n)\right) = 2^n \int_{I(x(n))} s\,dF(s) , \tag{22}$$

where

$$I\left(x(n)\right) = \left\{s : F(s) \in \left[.x(n),\ .x(n) + 2^{-n}\right)\right\} . \tag{23}$$

This assignment has the readily verified property

$$S\left(x(n)\right) = \frac{1}{2}\left(S\left(x(n)0\right) + S\left(x(n)1\right)\right) , \tag{24}$$

i.e.,

$$S_n = E\left\{S_{n+1} \mid \mathcal{F}_n\right\} . \tag{25}$$

Thus $(S_n, \mathcal{F}_n)$ is a Martingale on $\{[0,1], \mathcal{B}, \mu\}$, where $\mathcal{B}$ is the Borel σ-field for $[0,1]$, μ is Lebesque measure, and $\mathcal{F}_n$ is the σ-field generated by the partition $\{[k2^{-n}, (k+1)2^{-n})\}_{k=0}^{2^n-1}$. Proceeding directly, we observe from Eqs. (22,23) that

$$\begin{aligned} 2^n \int_{I(x(n))} F^{-1}\left(.x(n)\right) \, dF(s) \le S\left(x(n)\right) &= \int_{I(x(n))} s dF(s) \\ &\le 2^n \int_{I(x(n))} F^{-1}\left(.x(n) + 2^{-n}\right) \, dF(s) , \end{aligned} \tag{26}$$

or

$$F^{-1}\left(.x(n)\right) \le S\left(x(n)\right) \le F^{-1}\left(.x(n) + 2^{-n}\right) . \tag{27}$$

Now S_n converges wp 1 since

$$\begin{aligned} S_n(x) = S\left(x(n)\right) &\in \left[F^{-1}\left(.x(n)\right), \; F^{-1}\left(.x(n)\right) + 2^{-n}\right)\right) \\ &\to F^{-1}(.x) = S(.x) , \end{aligned} \tag{28}$$

for all $.x \in [0,1]$ with the exception of the at most countable number of jump points for F^{-1}. As noted in Eq. (21), S has the desired distribution F.

Next we note from Theorem 1 and Eq. (24), that the assignment $S(x(n))$ is achievable by a sequential gambling scheme. Specifically, S_n, $n = 1,2,\ldots,$ is achieved by a scheme b that bets the total amount (not proportion) of capital

$$b\left(x_{n+1} \middle| x(n)\right) S\left(x(n)\right) = \frac{1}{2} S\left(x(n+1)\right) = 2^n \int_{I(x(n+1))} s\,dF(s) \tag{29}$$

on the outcome x_{n+1} given the past sequence $x(n)$. Thus b yields a r.v. S_n at time n that takes values $S(x(n))$, $x(n) \in \{0,1\}^n$, with equal probability 2^{-n}. Thus $S \geq 0$, $ES \leq 1$ is achievable by the above sequential gambling scheme.

Conversely, any sequential gambling scheme b generates a Martingale $(S_n, \mathcal{F}_n)$ on $\{[0,1], \mathcal{B}, \mu\}$. Clearly $\sup ES_n \equiv 1 < \infty$. Thus by the Martingale convergence theorem, $S_n \to S$ a.s., with $ES = \int_0^1 s(t)dt \leq 1$. The possibility $ES < 1$ occurs, for example, for betting schemes placing a certain proportion α of $S_0 = 1$ on a given sequence x and "letting it ride."

Corollary: Let S_n, $n = 1,2,\ldots,$ denote the sequence of r.v.'s arising from the betting scheme achieving S in Theorem 4, and let F_n, $n = 1,2,\ldots,$ denote the associated d.f.'s. Then

$$\sup_{0 \leq s < \infty} \left| F_n(s) - F(s) \right| \leq 2^{-n}, \quad \text{for } n = 0,1,2,\ldots . \tag{30}$$

Proof: From Eq. (27),

$$.x(n) \leq F\left(S\left(x(n)\right)\right) \leq .x(n) + 2^{-n} . \tag{31}$$

Also, from Eq. (22),

$$F_n\Big(S\big(x(n)\big)\Big) = \Pr\Big\{.X \le .x(n) + 2^{-n}\Big\} = .x(n) + 2^{-n} \tag{32}$$

Thus

$$\Big| F_n\Big(S\big(x(n)\big)\Big) - F\Big(S\big(x(n)\big)\Big) \Big| \le 2^{-n}, \ \forall x(n) \in \{0,1\}^n . \tag{33}$$

Since F_n takes on all values in $\{k2^{-n}\}_{k=0}^{2^n-1}$, the Corollary follows.

<u>Example</u>: Suppose we wish to achieve a uniform distribution of capital on the interval $[0,2]$. Thus $F(s) = s/2$, $0 \le s \le 2$. We find explicitly that

$$S\big(x(n)\big) = 2\big(.x(n)1\big) = 2 \sum_{i=1}^{n} x_i (\tfrac{1}{2})^i + (\tfrac{1}{2})^n \tag{34}$$

and

$$b\big(1|x(n)\big)\ S\big(x(n)\big) - b\big(0|x(n)\big)\ S\big(x(n)\big) = \tfrac{1}{2}\ S\big(x(n)1\big) - \tfrac{1}{2}\ S\big(x(n)0\big) = .x(n)11 - .x(n)01 = 2^{-n-1} . \tag{35}$$

Thus the gambling scheme achieving this uniform distribution is independent of the past in the sense that a bet at time n of $(1/2)^{n+1}$ total units more on $x_{n+1} = 1$ than on $x_{n+1} = 0$ will achieve the desired goal.

5. Stochastic Application

Now consider a probability distribution

$$p\big(x(n)\big) \geq 0, \quad \sum_{x(n)\in\{0,1\}^n} p\big(x(n)\big) = 1 \ , \tag{36}$$

on $\{0,1\}^n$. Consider the goal

$$\max_b E \log S\big(x(n)\big) \ . \tag{37}$$

Thus, we desire to maximize

$$\sum p\big(x(n)\big) \log S\big(x(n)\big) + \lambda \sum S\big(x(n)\big) \tag{38}$$

subject to the constraints (from Theorem 1)

$$\sum S\big(x(n)\big) = 2^n, \ S\big(x(n)\big) \geq 0 \ . \tag{39}$$

Differentiation yields

$$S^*\big(x(n)\big) = 2^n \ p\big(x(n)\big) \ . \tag{40}$$

Thus, from Theorem 1,

$$b^*\big(x_{k+1} \big| x(k)\big) = p\big(x_{k+1} \big| x(k)\big), \ k = 0,1,\ldots \ . \tag{41}$$

This results in

$$\max_b E \log S\big(X(n)\big) = H(X_1, X_2, \ldots, X_n) \tag{42}$$
$$\triangleq - \sum p\big(x(n)\big) \log p\big(x(n)\big) .$$

Note that b^* is independent of the horizon n and is thus optimal for all n. Hence we have,

Theorem 5.

For any sequential gambling scheme b',

$$E \log S_n \le H(X_1, X_2, \ldots, X_n) , \tag{43}$$

for all n, with equality iff b' is the proportional gambling scheme b^* in Eq. (41).

This gambling scheme has a nice property for ergodic binary (and, in general, r-ary) processes $\{X_k, k \ge 1\}$. We recall from the Shannon-MacMillan-Breiman Theorem, that if $\{X_k, k \ge 1\}$ is ergodic then $-(1/n) \log_2 p(X_1, X_2, \ldots X_n) \to H(X)$ with probability one, where

$$\begin{aligned} H(X) &= \lim_{n \to \infty} H(X_n | X_1, \ldots, X_{n-1}) \\ &= \lim_{n \to \infty} \underset{X_1, X_2, \ldots, X_{n-1}}{E} \Big(- P_n \log_2 P_n - (1-P_n) \log_2 (1-P_n) \Big) , \end{aligned} \tag{44}$$

and the r.v. P_n is given by $P_n = \Pr\{X_n = 1 | X_1, \ldots, X_{n-1}\}$. Thus, b^* achieves

$$S^*\big(x(n)\big) = 2^n \, p\big(x(n)\big), \ \forall x(n) \in \{0,1\}^n . \tag{45}$$

But

$$\frac{1}{2} \log S^*\left(x(n)\right) = 1 + (1/n) \log p\left(x(n)\right) \to 1 - H(X) \text{ wp } 1, \text{ as } n \to \infty . \quad (46)$$

This represents a generalization of Breiman's gambling scheme on independent r.v.'s to ergodic processes. One simply bets now in proportion to the conditional probability that the next term will be 1 or 0. The rate of exponential growth of capital will be $1 - H(X)$.

Remark: From the asymptotic equipartition theorem, it can be seen by counting sequences that no scheme can achieve a higher rate of growth for any subset of sequences $x \in \{0,1\}^\infty$ of non-zero probability measure. Collecting this we have

Theorem 6.

If $\{X_n\}_{n=1}^\infty$ is an ergodic binary process, there exists a gambling scheme achieving

$$\frac{1}{n} \log S_n \to 1 - H(X), \text{ wp } 1 . \quad (47)$$

Moreover, for any sequential gambling system,

$$\Pr\left\{\overline{\lim}\ \frac{1}{n} \log S_n \le 1 - H(X)\right\} = 1 . \quad (48)$$

6. Program complexity and a universal gambling scheme

Let $A : \{0,1\}^* \times N \to \{0,1\}^*$ be a partial recursive function.

Definition: Let the program complexity of $x(n) \in \{0,1\}^n$, given the length n of the string $x(n)$, be defined by

$$K_A\left(x(n)\middle| n\right) = \min_{A(p,n)=x(n)} \ell(p) \ , \tag{49}$$

where $\ell(\cdot)$ denotes the length of the (program) string.

This is the program complexity put forth by Kolmogorov [7] and Chaitin [2]. The idea is that the intrinsic complexity of a string is the shortest computer program that describes it. Recall the following theorem from Kolmogorov [7] (see also Chaitin [2]).

Theorem 7.

There exists a universal partial recursive function U such that

$$\forall A,\ \exists c,\ \forall n,\ \forall x(n),\ K_U\left(x(n)\middle| n\right) \le K_A\left(x(n)\middle| n\right) + c \ . \tag{50}$$

The constant c depends only on A and U.

Thus U has the same (or simpler) notion of complexity as any other A, up to a constant. For example,

i) a finite string of 0's has K_U = constant;

ii) the first n bits of π has $K_U(x(n)|n) \le$ constant, $\forall n$;

iii) a string of length n having k 1's has $K_U(x(n)|n) \le c + nh(k/n) + \log n$, where h is the Shannon entropy function $h(p) = - p \log p - q \log q$.

We need the following well known result for program complexity:

Theorem 8.

There are at most 2^k sequences $x \in \{0,1\}^*$ having program complexity $K(x|\ell(x)) < k$. Also, there exists a constant c such that

$$\forall\, x \in \{0,1\}^* \;,\; K\big(x|\ell(x)\big) \le \ell(x) + c \;. \tag{51}$$

Proof: The first part is proved by simply summing the number of programs 2^j having length j, $j = 0,1,2,\ldots,k-1$. The second part is proved by simulating with U a function C that merely copies the sequence x.

We now prove a gambling result with respect to this algorithmic complexity measure.

Theorem 9.

For a given n, $S(x(n)) \ge \exp_2(n - K(x(n)|n) - \log(n+1))$ is achievable by a sequential gambling system. The betting scheme that achieves this is

$$b\big(x_{k+1}|x(k)\big) = \sum_{z\in\{0,1\}^{n-k-1}} 2^{-K(x(k)x_{k+1}z)} \Big/ \sum_{z\in\{0,1\}^{n-k}} 2^{-K(x(k)z)} \;. \tag{52}$$

Proof: Let

$$\tilde{S}\big(x(n)\big) = 2^{n-K(x(n)|n)-\sigma(K)} \;, \tag{53}$$

where

$$\sigma(K) = \log_2\left(\sum_{x(n)\in\{0,1\}^n} 2^{-K(x(n)|n)}\right) . \tag{54}$$

Clearly, by the definition of the normalizing factor $\sigma(K)$, $2^{-n} \sum \tilde{S} = 1$, and $\tilde{S} \geq 0$, so $\tilde{S}$ is achievable. The summation in $\sigma(K)$ is maximized when K takes small values on $x(n)$. There are at most 2^k sequences $x(n)$ of program complexity k. Thus

$$\sum_{x(n)} 2^{-K(x(n)|n)} \leq 1 + 2 \cdot 2^{-1} + 2^2 \cdot 2^{-2} + 2^{n-1} \cdot 2^{-(n-1)} + 1 \cdot 2^{-n} = n + 2^{-n} \leq n + 1 \, . \tag{55}$$

Thus $\sigma(K) \leq \log(n+1)$ and the theorem is proved.

<u>Remark</u>: The gambling scheme in Theorem 9 can be trivially modified to guarantee that the capital $S(x(n)) \geq 1/2$ for all sequences $x \in \{0,1\}^*$. This is achieved by the simple expedient of witholding $1/2$ unit from the betting and using the other $1/2$ unit to gamble with. This subtracts at most one unit from the exponent.

Let $b^{(n)}$ be the sequential betting scheme, which achieves $S(x(n))$ of the previous theorem. Then

$$b = (6/\pi^2) \sum_{n=1}^{\infty} b^{(n)}/n^2 \tag{56}$$

is also a sequential betting scheme with $S(\Lambda) = 1$. But $(6/\pi^2 n^2)\, b^{(n)}$ achieves

$$S\left(x(n)\right) \geq 2^{n-K(x(n)|n)-3\log(n+1)-1} \, . \tag{57}$$

Thus b achieves this goal for all n, $\forall\, x \in \{0,1\}^{\infty}$. This idea is closely related to Schnorr's [6] universal gambling scheme in which $b^{(1)}, b^{(2)}, \ldots$ are all partial recursive gambling schemes, i.e., all sequential gambling schemes for which there exists a finite description that leads through a finite mechanical computation to the actual amount to be bet as a function of the past. Let $\alpha_i \geq 0$, $\Sigma\, \alpha_i = 1$. Then $\Sigma\, \alpha_i b^{(i)}$ is a mixture of all gambling schemes.

The program complexity measure $K(x|\ell(x))$ has two difficulties from our point of view. First, it is conditioned on the sequence length $\ell(x)$. Second, it can be shown that

$$\sum_{x \in \{0,1\}^*} 2^{-K(x|\ell(x))} = \infty \quad , \tag{58}$$

thus preventing a natural normalization of $2^{-K(x)}$, $x \in X^*$, as a probability distribution or a betting scheme. Both of these properties are remedied by the following natural complexity measure recently introduced by Chaitin [9]. (See also the "dual" notion of process complexity given by Schnorr [10].)

<u>Definition</u>: A subset S of X^* is said to have the <u>prefix property</u> if no string in S is the proper prefix of another. That is,

$$S \cap S \times X^* = \emptyset \quad . \tag{59}$$

Thus, $S = \{0, 10, 11\}$ has the prefix property, but $S = \{0, 01, 11\}$ does not.

<u>Theorem 10 (Chaitin)</u>.

There exists a partial recursive function $U_0 : \{0,1\}^* \rightarrow \{0,1\}^*$ such that the domain of U_0 has the prefix property and such that, for any other partial recursive function $U' : \{0,1\}^* \to \{0,1\}^*$ with the prefix property, there is a constant c such that

$$U'(p') \text{ is defined } \Rightarrow \exists\, p\left[U_0(p) = U'(p') \text{ and } \ell(p) \le \ell(p') + c\right] . \quad (60)$$

Such a function U_0 will be called a <u>universal prefix function</u>. We think of $p \in \{0,1\}^*$ as the program and $U_0(p)$ as the output. The existence of a universal prefix function U_0 is apparent from inspection of Turing machines and ordinary computers. Most modern computers are universal (in the sense that they can model the action of any other computer, by appending a subroutine of appropriate length), and their programming languages (like Fortran) have the prefix property (because of the END statements). Henceforth, let U_0 be a fixed universal prefix function.

<u>Definition</u>: The Chaitin complexity $C(\cdot)$ is given by

$$C(x) = \min_{U_0(p)=x} \ell(p) \quad . \quad (61)$$

The complexity $C(x)$ can be thought of as the length of the shortest computer program which yields x when the program is provided to the computer U_0. Note, from the universality of U_0 that, for any other prefix function U,

$$\exists c,\ C(x) \le C_U(x) + c,\ \forall x \in \{0,1\}^* \quad . \quad (62)$$

Thus the Chaitin complexity measure $C(x)$ is universal.

The complexity measures C and K can be related as follows:

Theorem 11.

$$\exists c,\ \forall x \in \{0,1\}^*,\ K\big(x \big| \ell(x)\big) + c \le C(x) \le K\big(x \big| \ell(x)\big) + 2 \log K\big(x \big| \ell(x)\big) + c \quad . \tag{63}$$

Proof: Let $\tilde{p}$ be a program for the universal partial recursive function $\tilde{U}$ used in defining K. The desired program p for use by the universal prefix function U_0 is written $p = \tilde{\ell}_1\tilde{\ell}_1\tilde{\ell}_2\tilde{\ell}_2 \ldots \tilde{\ell}_\ell\tilde{\ell}_\ell\ 0\ 1\ \tilde{p}$, where $\tilde{\ell}_1\tilde{\ell}_2\ldots\tilde{\ell}_\ell$ is the binary encoding of the length of $\tilde{p}$ and $\ell = \lceil \log \ell(\tilde{p}) \rceil$ is the length of the encoding. U_0 then reads $\ell(\tilde{p})$ to determine where p ends and simulates $\tilde{U}$ on $\tilde{p}$ to obtain $\tilde{U}(\tilde{p})$. Finally, $\ell(p) = \ell(\tilde{p}) + 2 \log \ell(\tilde{p}) + c$. A more sophisticated encoding can probably reduce the bound to $C \le K + \log K + \log(\log K) + \log(\log(\log K)) + \ldots + c$.

Theorem 12.

There exists a universal gambling system b^* such that

$$S(x) \ge 2^{\ell(x) - C(x)},\ \forall x \in \{0,1\}^* \quad . \tag{64}$$

A scheme b^* achieving this goal is given by

$$b^*\big(x_{k+1} \big| x(k)\big) = \sum_{z \in \{0,1\}^*} 2^{-C(x(k)x_{k+1}z)} \bigg/ \sum_{z \in \{0,1\}^*} 2^{-C(x(k)z)} \quad . \tag{65}$$

Proof: Consider the following gambling system. Of the initial capital $S(\Lambda) = 1$, let the amount $2^{-C(x)}$ be placed on the finite sequence x, and let it ride. We observe that

$$\sum_{x \in \{0,1\}^*} 2^{-C(x)} \le \sum_{U_0(p) \text{ defined}} 2^{-\ell(p)} \le 1 \quad , \tag{66}$$

follows from the Kraft inequality of information theory, because no program is the prefix of another. Thus the sum total of initial bets is ≤ 1, as desired. The amount of money $2^{-C(x)}$ bet on x will be doubled precisely $\ell(x)$ times if x occurs and will be lost otherwise, thus resulting in $2^{\ell(x)-C(x)}$ units of capital. In fact some other bets will be surviving at time $\ell(x)$, all bets on finite sequences x' that either extend or are extended by x. But survival of this one bet is enough to prove the theorem. Summarizing, for this betting scheme,

$$\begin{aligned} S(x) &= 2^{\ell(x)} \sum_{x \sqsubset x'} 2^{-C(x')} + \sum_{x' \sqsubseteq x} 2^{\ell(x')-C(x')} \\ &\ge 2^{\ell(x)-C(x)}, \ \forall x \in \{0,1\}^* \quad . \end{aligned} \tag{67}$$

The above betting scheme achieves the desired goal even though money is thrown away, e.g., $\sum_{x_1 \in \{0,1\}} b(x_1)$ may be strictly less than 1. By apportioning the unallocated capital at each stage in the same proportions as the capital that is being utilized, we obtain a true sequential betting scheme satisfying Eq. (1), with a performance $S(\cdot)$ that is everywhere at least as large as before. This is the scheme b^* given in Eq. (65).

<u>Remark</u>: Examination of Eq. (65) shows how very close Solomonoff was in his original papers [11, 12] to the natural correct definition of a universal inductive inference procedure. Solomonoff's formula, which embodies the notion of weighting explanations (or programs) by their

exponentiated length, can be interpreted as the conditional probability that the next term in the sequence will be a 1. The theorem above gives an operational gambling significance to these conditional probabilities and also provides a satisfactory definition of them. Thus $b^*(x_{k+1}|x(k))$ seems to be a satisfactory notion of the conditional probability of the next term given the past.

It should be mentioned that b^* is not recursive. That is, although b^* may be calculated to arbitrary accuracy by sufficiently long calculations, the degree of accuracy cannot be known. Actually, b^* is not recursive solely because the goal $S(x) = 2^{\ell(x)-C(x)}$ is not. However, there exists a sequence of total recursive functions C_n such that

$$C_n(x) \searrow C(x), \ \forall x \in \{0,1\}^* \quad . \tag{68}$$

For example, $C_n(x)$ might be the least length of a program among the first n programs in some enumeration such that a prefix computer U_0 running for n steps prints x. Then if C_n is substituted for C in Eq. (65), the resulting gambling scheme b_n^* would achieve

$$S(x) \geq 2^{\ell(x)-C_n(x)}, \ \forall x \quad . \tag{69}$$

One may wonder whether b^* in Eq. (65) may be dominated by some superior gambling scheme. This is not possible, since all sequential gambling systems are admissible; because if b generates S, and b' generates S', with $S'(x) \geq S(x)$, $\forall x$, with strict inequality for some x, then $p(x) = 2^{-\ell(x)}S(x)$ and $p'(x) = 2^{-\ell(x)}S'(x)$ cannot both be the marginals of a stochastic process, violating Theorem 1.

Randomness Interpretation of K(or C):

Suppose that a sequence of length n begins with K random 0's and 1's and terminates with $n - K$ deterministic outcomes. Such a sequence will have Kolmogorov complexity K with high probability. Any gambler who would be asked to bet on such a sequence would clearly be able to double his money $n - K$ times on the final $n - K$ terms of the sequence, but would be unable to gain any money on the average from bets on the initial K random terms. Thus, doubling his money $n - K$ times would be an ultimate goal of a reasonable gambler. Now we have shown that for any sequence of Kolmogorov complexity K the money will be doubled $n - K - \log(n+1)$ times, and thus we interpret our results as saying that even if the redundancy of the sequence is distributed throughout the sequence, we can do as well as if all of the randomness occurs at the beginning of the sequence, and $n - K$ deterministic terms follow. Thus K has a natural gambling interpretation as the number of random bits of a sequence, and $n - K$ the number of deterministic terms. Of course, not all sequences have their randomness at the beginning, so the result is somewhat surprising.

In Theorem 6, we showed that, for an ergodic process with known statistics and entropy H, there exists a b achieving an optimal exponential growth rate of capital with rate $1 - H$. Next we argue that the universal sequential gambling scheme b^* achieves the optimal growth rate of capital against all ergodic processes of unknown statistical description.

Theorem 13.

Let $\{X_n\}_{n=1}^{\infty}$ be a binary ergodic process with unknown entropy $H(X)$ and unknown statistics. Then b^* of Theorem 12 achieves

$$\frac{1}{n} \log S\left(x(n)\right) \to 1 - H(X) \text{ wp } 1 \quad . \tag{70}$$

Proof: From Levin and Zhvonkin [14], we know

$$\frac{1}{n} K(X_1, X_2, \ldots, X_n) \to H(X), \text{ wp } 1 \quad . \tag{71}$$

Thus, from Theorem 11,

$$\frac{1}{n} C(X_1, X_2, \ldots, X_n) \to H(X), \text{ wp } 1 \quad . \tag{72}$$

But, from Theorem 12, b^* achieves

$$\frac{1}{n} \log S\left(x(n)\right) \geq \frac{n - C(x(n))}{n} \to 1 - H(X), \text{ wp } 1 \quad . \tag{73}$$

Thus we find that b^* simultaneously causes money to grow at the rate $2^{n(1+p \log p + q \log q)}$ if X is the Bernoulli process with parameter p; at rate 2^n if X is a (recursive) deterministic process; at rate $2^{n(1-H)}$, for X any ergodic process. In fact, b^* is achieving the maximal rate of growth of capital for all of these processes, while of course, simultaneously achieving the goals of Theorem 12. This helps to substantiate the naturalness of the goal $S(x) = 2^{\ell(x) - C(x)}$.

7. Conclusions

In this paper we have shown that if a sequence of length n can be compressed into a sequence of length K with respect to some universal algorithm U, then it follows that there exists a gambling system b^*

(depending only on U) that will guarantee that the amount of capital achieved at time n will be doubled n - K times. Thus K may be considered to be the amount of randomness in the sequence and n - K the amount of determinism. Any order that exists will reflect itself in an increase in capital for the gambler.

Acknowledgement

I would like to thank G. Chaitin, J. Gill, and Andre Tchen for their comments.

References

1. Breiman, L., "Optimal Gambling Systems for Favorable Games," Fourth Berkeley Symposium, Vol. 1, pp. 65-78, 1961.

2. Chaitin, G.J., "On the Length of Programs for Computing Finite Binary Sequences," JACM, 13, 4, pp. 547-569, 1966.

3. Chaitin, G.J., "A Theory of Program Size Formally Identical to Information Theory," Research Report RC4805, IBM Research Center, Yorktown Heights, N.Y., 1974, submitted to JACM.

4. Cover, T.M., "Behavior of Sequential Predictors of Binary Sequences," Transactions of the Fourth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Pub. House of Chechoslovak Academy of Sciences, Prague, pp. 263-272, 1967.

5. Cover, T.M., "A Complexity Approach to Pattern Recognition, " in preparation, to be submitted to Annals of Statistics.

6. Kelly, J.L., Jr., "A New Interpretation of Information Rate," Bell Systems Technical Journal, Vol. 35, pp. 917-926, 1956.

7. Kolmogorov, A.N., "Three Approaches to the Concept of the 'Amount of Information,'" Problemy Peredachi Informatsii, 1, 1(1965), pp. 3-11.

8. Zhvonkin, A.K., and Levin, L.A., "The Complexity of Finite Objects and the Development of the Concepts of Information and Randomness by Means of the Theory of Algorithms," Russian Mathematical Surveys 25, 1970, pp. 83-124.

9. Martin-Lof, P., "The Definition of Random Sequences," Information and Control, 9, pp. 602-619, 1966.

10. Schnorr, C.P., "A Unified Approach to the Definition of Random Sequences," Math. Systems Theory, 5, No. 3, pp. 246-258, 1971.

11. Schnorr, C.P., "Zufalligkeit und Wahrscheinlichkeit," Lecture Notes in Mathematics, Vol. 218, Springer, 1971.

12. Schnorr, C.P., "The Process Complexity and Effective Random Tests," Journal Computer and Syst. Sci, 7, pp. 376-388, 1973.

13. Solomonoff, R.J., "A Formal Theory of Inductive Inference. Part I," Information and Control, 7, pp. 1-22, 1964.

14. Solomonoff, R.J., "A Formal Theory of Inductive Inference. Part II," Information and Control, 7, pp. 224-254, 1964.