A CLASS OF MATHEMATICAL MODELS FOR EVOLUTION AND HIERARCHICAL INFORMATION THEORY

By

Jonathan D.H. Smith

IMA Preprint Series # 396

March 1988

A CLASS OF MATHEMATICAL MODELS FOR EVOLUTION AND HIERARCHICAL INFORMATION THEORY

JONATHAN D. H. SMITH†

Abstract. The paper introduces a class of mathematical models for evolution according to the Brooks-Wiley unified theory. These models, partitioned Lebesgue spaces with automorphism, display increasing, concave entropy simultaneously with increasing organisation. In addition, a rigorous foundation for hierarchical information theory is laid. The theory is used to analyse the distribution of information in evolution models and Markov chains.

Introduction. At first glance, the process of evolution appears to be incompatible with the Second Law of Thermodynamics. As a creationist pamphlet [GC] expresses the dilemma:

> By definition, evolution is a process of continuous change from a lower, simpler ... state to a higher, more complex ... state. ... [T]he Second Law of Thermodynamics, one of the most fundamental laws of nature, states that every change which takes place spontaneously tends to go from a state of order to one of disorder. ... Evolution requires a steady overall increase in order. Clearly, evolution dramatically conflicts with the Second Law of Thermodynamics.

A vitalist might dodge the issue by claiming that living organisms are not subject to all the physical laws governing inanimate matter. The creationists, on the other hand, deny that the diversity of biological form is the product of a process undergoing changes in time to which the Second Law might apply. If one finds such arguments unacceptable, one is faced with the problem of reconciling evolutionary theory with physical laws. This problem is the first of "four major items of unfinished business" in evolutionary theory [BW, p. 3]. In their seminal book **Evolution as Entropy** [BW], Brooks and Wiley lay the foundations for a unified theory of evolution to address these questions.

The general outline proposed by Brooks and Wiley may be represented schematically as in Fig. 1 (*cf.* [Br, Fig. 3] and [BW, Fig. 2.5]).

†Department of Mathematics, Iowa State University, Ames, Iowa 50011.

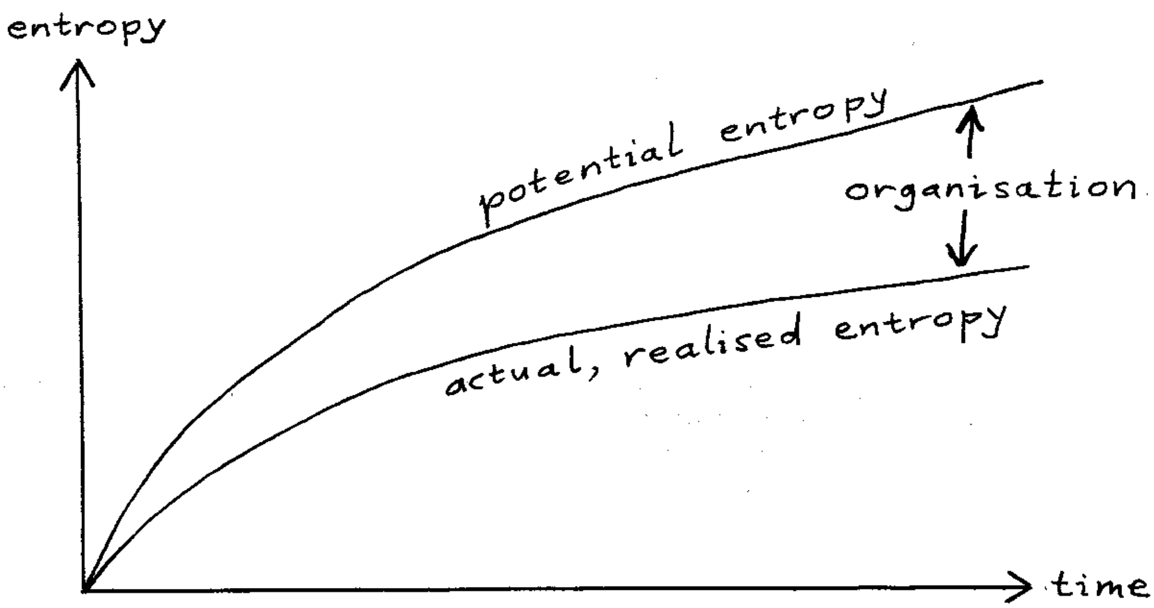


FIGURE 1 *Entropy against time.*

As a system evolves with time, becoming more complex, the phase space needed to describe all its potential states at a given time expands. Thus its potential entropy, measured as proportional to the logarithm of the number of potential states, increases with time, along the upper of the concave curves of Fig. 1. The actual entropy of the system is plotted by the lower of the concave curves of Fig. 1. In accordance with the Second Law of Thermodynamics, this actual entropy is an increasing function of time. However, the development of the system subsequent to a given time is constrained by its actual state at that time. At the given time, the system is not completely random. Its departure from complete randomness is measured by the difference between its potential and actual entropies at that time. This difference may be characterised as the information embodied in the organization of the system at that time. This organisation furnishes initial conditions under which the subsequent development takes place according to physical laws. The constraints set by the initial conditions slow the growth of randomness in subsequent stages. As Brooks and Wiley express it [BW, p. 69]:

> The amount of increase in entropy declines as the historical burden of constraint grows.

Or [BW, p. 107]:

> The amount of entropy increase from one stage to another through time always decreases.

The salient features of the general outline may thus be summarised as follows:

(1.1) the actual entropy is an increasing function of time, in accordance with the Second Law of Thermodynamics;

(1.2) the actual entropy is a concave function of time, as historical constraints retard the rate of entropy increase;

(1.3) the difference between the potential and actual entropies is an increasing function of time, commensurate with the increasing organisation taking place.

One of the two main aims of this paper is to propose a general class of mathematical models exhibiting these features, for a discrete time parameter. The models are known as partitioned Lebesgue spaces with automorphsm. They are introduced in Definition 2.1 of Section 2, and shown to have the properties (1.1)–(1.3) in Theorem 3.9 of Section 3.

An important step towards making the general outline more specific is to give a more detailed analysis of the information density in a system that is hierarchically organised, for example by the time-ordered hierarchy of its successive stages of development. Such an analysis was adumbrated by Gatlin [Ga, Ch. 3], in order to attempt an answer to

> the primitive question: "How much information is stored in the base sequence of a given DNA molecule?"

[Ga, p. 20]. The hierarchy implicit in a base sequence is its family of sets of subsequences of given length, ranked according to increasing length. Brooks, Cumming, and LeBlond [Bs, CB] have laid the foundations for a general theory of such analyses, called *Hierarchical Information Theory*. The second main aim of this paper is to give a rigorous treatment of Hierarchical Information Theory for discrete, linear hierarchies that are finite or ordered like the set of natural numbers. These include the time-ordered hierarchies in the partitioned Lebesgue spaces with automorphism. The general theory is presented in Section 4, while Section 5 discusses its application to Markov chains and Bernoulli shifts.

2. Automorphsms of partitioned Lebesgue spaces. The basic mathematical model for evolution according to the general outline of Brooks and Wiley (Fig. 1), with a discrete time parameter, and enjoying the properties (1.1)–(1.3), consists of a partitioned Lebesgue space with automorphism.

DEFINITION 2.1. *A partitioned Lebesgue space with automorphism* (M, S, μ, T, ξ) *comprises:*

(a) *a Lebesgue space* (M, S, μ) *with* $\mu M = 1$ [CF, App. 1, Def. 4];

(b) *an automorphism* T *of* (M, S, μ) [CF, §1.1, Def. 1];

(c) *a finite partition* ξ *of* (M, S, μ) *modulo* O [CF, §10.6] [CF, Appl. 1, Section 2].

Intuitively, the space M may be thought of as the set of all possible histories of particles. The σ-algebra S consists of the set of events to which probabilities may be assigned: the measure of a subset A of M lying in S is the probability that event A occurs. The automorphism T represents the passage of (discrete) time: if an element s of M describes the history of a certain particle, then the element Tx of M describes the history of a second particle which is always one time unit ahead of the first. Thus if the membership relationships $x \in A \in S$ mean that the first particle has a certain property at time O, the membership $Tx \in A$ means that the second particle has the property at time O, or equivalently that the first particle has the property at time 1. Finally, the finite partition $\xi = \{C_1, C_2, \ldots, C_r\}$ represents a complete set of possible initial states. There are r initial states: the membership $x \in C_i \in S$ means that the particle described by x is in the i-th state at time O. If $\eta = \{D_1, \ldots, D_s\}$ is a second finite partition modulo O, then the common refinement $\xi \vee \eta = \{C_i \cap D_j \mid 1 \leq i \leq r, 1 \leq j \leq s\}$ of ξ and η represents an enlarged state space with rs elements. In particular, the common refinement $\xi \vee T^{-1}\xi$ describes evolution over one time unit from an initial phase space with only r states to a new phase space with r^2 states. The refinement $\xi \vee T^{-1}\xi \vee \cdots \vee T^{1-n}\xi$ describes evolution over a period of $n-1$ time units from an initial phase space with r states to a new phase space with r^n states.

Perhaps the simplest significant example illustrating Definition 2.1 is given by a (first order) Markov chain having r states $1, 2, \ldots, r$ distributed with stationary probabilities $p(1), p(2), \ldots, p(r)$ and transition matrix $\big(p(i,j)\big)$. The space M is the set of all functions

$$x : \mathbf{Z} \longrightarrow \{1, 2, \ldots, r\}; j \longmapsto x(j). \tag{2.2}$$

For each ordered pair $s < t$ of distinct integers and function $f : \{s, s+1, \ldots, t\} \longrightarrow \{1, 2, \ldots, r\};\ j \longmapsto f(j)$, the "cylinder" $C_f = \{x \in M \mid \forall s \leq j \leq t, x(j) = f(j)\}$ has measure

$$\mu\, C_f = p\big(f(s)\big)p\big(f(s), f(s+1)\big) \ldots p\big(f(t-1), f(t)\big). \tag{2.3}$$

If S is the Borel σ-algebra for the product topology on $M = \{1, 2, \ldots, r\}^{\mathbf{Z}}$ induced by the discrete topology on $\{1, 2, \ldots, r\}$, then by Kolmogorov's Theorem formula (2.3) extends to give a uniquely defined measure μ on S, making $(M,\ S,\ \mu)$ a Lebesgue space [CF, pp. 4, 182, 450]. The "shift"

$$T : M \longrightarrow M; \big(i \longmapsto x(i)\big) \longmapsto \big(i \longmapsto x(i+1)\big) \tag{2.4}$$

is an automorphism of $(M,\ S,\ \mu)$ [CF, §8.1 Subsection 2]. Let $\xi = \{C_1, \ldots, C_r\}$ be the partition with

$$C_i = \{x \in M \mid x(O) = i\}. \tag{2.5}$$

Then $(M,\ S,\ \mu,\ T,\ \xi)$ as defined by (2.2)–(2.5) is a partitioned Lebesgue space with automorphism, called the *first-order Markov chain model.* The partition ξ represents a phase space with r elements describing the distribution of single letters in the Markov chains, while the refinement $\xi \vee T^{-1}\xi \vee \ldots T^{1-n}\xi$ represents a phase space with r^n elements describing the distribution of n-tuplets in the chains.

Remark 2.6. From a naive point of view, the use of automorphisms to represent the passage of (discrete) time may appear unsuitable, since automorphisms are reversible. By the same token, it may seem unnatural for the histories embodied in the space M to stretch infinitely far backward in time as well as forward. However, as discussed in [CF, §10.4], spaces with automorphisms are equivalent to spaces with endomorphisms, where histories and the passage of time only stretch into the future. On the other hand, spaces with automorphisms are technically more convenient to deal with, for example in the use made of equation (3.10) below.

3. Entropy in partitioned Lebesgue spaces. If $\xi = \{C_1, \ldots, C_r\}$ is a finite partition of (M, S, μ) modulo O, then the *entropy* $H(\xi)$ of ξ is defined as

$$H(\xi) = -\sum_{i=1}^{r} \mu(C_i) \log \mu(C_i) \tag{3.1}$$

[CF, §10.6 Def. 1], where $\mu(C_i)\log(C_i)$ is interpreted as O if $\mu(C_i) = O$. Logarithms are taken to base 2, so that the units of entropy used are bits. Intuitively, $H(\xi)$ may be thought of as the expected surprise experienced or information gained as a result of the following "experiment ξ": pick an x from M, and observe which of the states $C_1, C_2, \ldots, C_r$ it lies in. The entropy $H(\xi)$ satisfies the inequality

$$O \leq H(\xi) \leq \log r \tag{3.2}$$

[CF, §10.6]. Equality obtains on the left if and only if $\mu C_i = 1$ for some i: if you already know in advance that the outcome of the experiment ξ is C_i, then the experiment holds no surprise for you. Equality obtains on the right if and only if $\mu C_i = 1/r$ for each i: the most informative experiments are those designed so that all their different outcomes are equally likely. In other words, a partition with r states offers a potential entropy of $\log r$.

Given a second finite partition $\eta = \{D_1, \ldots, D_s\}$, the *conditional entropy $H(\xi \mid \eta)$ of ξ given η* is defined as

$$H(\xi|\eta) = -\sum_{j=1}^{s} \mu(D_j) \sum_{i=1}^{r} \mu(C_i|D_j) \log \mu(C_i|D_j) \tag{3.3}$$

[CF, §10.6 Def. 2]. The conditional entropy satisfies the equality

$$H(\xi|\eta) = H(\xi \vee \eta) - H(\eta) \tag{3.4}$$

[CF, §10.6]. Thus $H(\xi|\eta)$ may be interpreted as the additional information gained as a result of experiment ξ if the result of experiment η is already known. It is sometimes useful to be able to consider the entropy $H(\xi)$ of (3.1) as a conditional entropy $H(\xi \mid \eta)$ as in

(3.3). This may be achieved by taking η to be the singleton partition $\{M\}$. The sum (3.3) then reduces to (3.1), so that

$$H(\xi \mid \{M\}) = H(\xi). \tag{3.5}$$

Given a third finite partition $\zeta = \{E_1, \ldots, E_t\}$, the fundamental inequality

$$H(\xi \mid \eta \vee \zeta) \leq H(\xi \mid \eta) \tag{3.6}$$

[CF,§10.6] holds. it may be interpreted as saying that the more you know already in advance, the less new information any subsequent experiment can bring you.

Now let $(M,\ S,\ \mu,\ T,\ \xi)$ be a partitioned measure space with automorphism. Set $H_O = 0$, and for each positive integer n, define

$$H_n = H(\xi \vee T^{-1}\xi \vee \cdots \vee T^{1-n}\xi). \tag{3.7}$$

Thus H_n represents the *actual entropy* at time n of the evolving system described by $(M,\ S,\ \mu,\ T,\ \xi)$. At time n, there are r^n states potentially available in the partition $\xi \vee T^{-1}\xi \vee \cdots \vee T^{-1}\xi$, so the system has a *potential entropy* of

$$\log r^n = n \log r \tag{3.8}$$

at time n.

THEOREM 3.9. *A partitioned measure space $(M,\ S,\ \mu,\ T,\ \xi)$ with automorphism, having actual entropies (3.7) and potential entropies (3.8), enjoys the properties (1.1)–(1.3).*

Proof. (a) To verify (1.1), observe that

$$\begin{aligned} H_{n+1} - H_{n+1} =& H(\xi \vee \cdots \vee T^{-n}\xi \mid (\xi \vee \cdots \vee T^{1-n}\xi) \vee T^{-n}\xi) \\ &\leq H(\xi \vee \cdots \vee T^{-n}\xi \mid \xi \vee \cdots \vee T^{1-n}\xi) = H_{n+1} - H_n, \end{aligned}$$

the equalities holding by (3.4) and the inequality by (3.6). Subtracting H_{n+1} from each side and multiplying by -1 gives the desired inequality $H_n \leq H_{n+1}$.

(b) To verify (1.2), observe that $H_{n+1} - H_n = H(T^{-n}\xi \mid \xi \vee (T^{-1}\xi \vee \cdots \vee T^{1-n}\xi)) \leq H(T^{-n}\xi \mid T^{-1}\xi \vee \cdots \vee T^{1-n}\xi) = H(T^{-1}\xi \vee \cdots \vee T^{1-n}\xi \vee T^{-n}\xi) - H(T^{-1}\xi \vee \cdots \vee T^{1-n}\xi) = H(\xi \vee \cdots \vee T^{1-n}\xi) - H(\xi \vee \cdots \vee T^{2-n}\xi) = H_n - H_{n-1}$, as required. Here the first two equalities hold by (3.4) and the inequality by (3.6). The penultimate equality holds since T is an automorphism of $(M,\ S,\ \mu)$, indeed

$$H(\eta) = H(T\eta) \tag{3.10}$$

holds for any finite partition $\eta\{D_1, \ldots, D_s\}$ since $\mu(D_j) = \mu(TD_j)$ for each j implies

$$-\sum_{j=1}^{s} \mu(D_j) \log \mu(D_j) = -\sum_{j=1}^{s} \mu(TD_j) \log \mu(tD_j).$$

(c) To verify (1.3), observe that $H_{n+1} - H_n = H(T^{-n}\xi \mid \xi \vee \cdots \vee T^{1-n}\xi) = H(T^{-n}\xi \mid \{M\} \vee (\xi \vee \cdots \vee T^{1-n}\xi)) \leq H(T^{-n}\xi \mid \{M\}) = H(T^{-n}\xi) \leq \log r$, so that $n \log r - H_n \leq (n+1)\log r - H_{n+1}$ as required. Here the first equality holds by (3.4) and the third by (3.5). The first inequality holds by (3.6) and the second by (3.2). □

4. Hierarchical information theory. Underlying each partitioned Lebesgue space (M, S, μ, T, ξ) is the hierarchy $(\mathbf{N}, \leq)$ of successive discrete times $0 < 1 < 2 < \cdots < n < \ldots$. At each stage n, (3.7) assigns an entropy H_n to the system described. As shown by Theorem 3.9, the (non-negative) real-valued entropy function

(4.1) $$H : \mathbf{N} \longrightarrow \mathbf{R}^+; n \longmapsto H_n$$

is increasing and concave. In any context yielding such an entropy function, *e.g.*, partitioned Lebesgue spaces with automorphism, hierarchical information theory may be interpreted as the numerical analysis of this function $H : \mathbf{N} \longrightarrow \mathbf{R}^+$.

Since $H : \mathbf{N} \to \mathbf{R}^+$ is increasing and concave, the first difference function

(4.2) $$\Delta H : \mathbf{N} \longrightarrow \mathbf{R}^+; n \longmapsto \Delta H_n = H_{n+1} - H_n$$

is non-negative and decreasing. The negative

(4.3) $$-\Delta H^2 : \mathbf{N} \longrightarrow \mathbf{R}^+; n \longmapsto -\Delta^2 H_n = \Delta H_n - \Delta H_{n+1}$$

of the second difference function is then non-negative. According to (3.4) and (3.7), the first difference function (4.2) may be interpreted as giving the conditional entropies

(4.4) $$\Delta H_n = H(T^{-n}\xi \mid \xi \vee T^{-1}\xi \vee \cdots \vee T^{1-n}\xi).$$

The negated second difference function (4.3) may be interpreted as giving the quanities D_{n+2} that Gatlin called "divergences" [Ga, pp. 36, 38, 68] and Brooks and Wiley call "information measures" [BW, p. 142]:

(4.5) $$-\Delta^2 H_n = D_{n+2}.$$

Gatlin's D_1 may be defined by (4.5) in the context of a partitioned Lebesgue space (M, S, μ, T, ξ), where the partition ξ has r parts, on setting

(4.6) $$H_{-1} = -\log r$$

so that

$$\Delta H_{-1} = \log r \tag{4.7}$$

and

$$-\Delta^2 H_{-1} = D_1 = \log r - H_1 \quad . \tag{4.8}$$

Fig. 2 displays a typical entropy function along with its first differences and divergences. This graph gives a different way of presenting Gatlin's "detailed entropy scale" [Ga, Fig. 7]. Alternatively, one may use a finite difference scheme such as that given in Fig. 3 of Section 5 below for the first-order Markov chain.

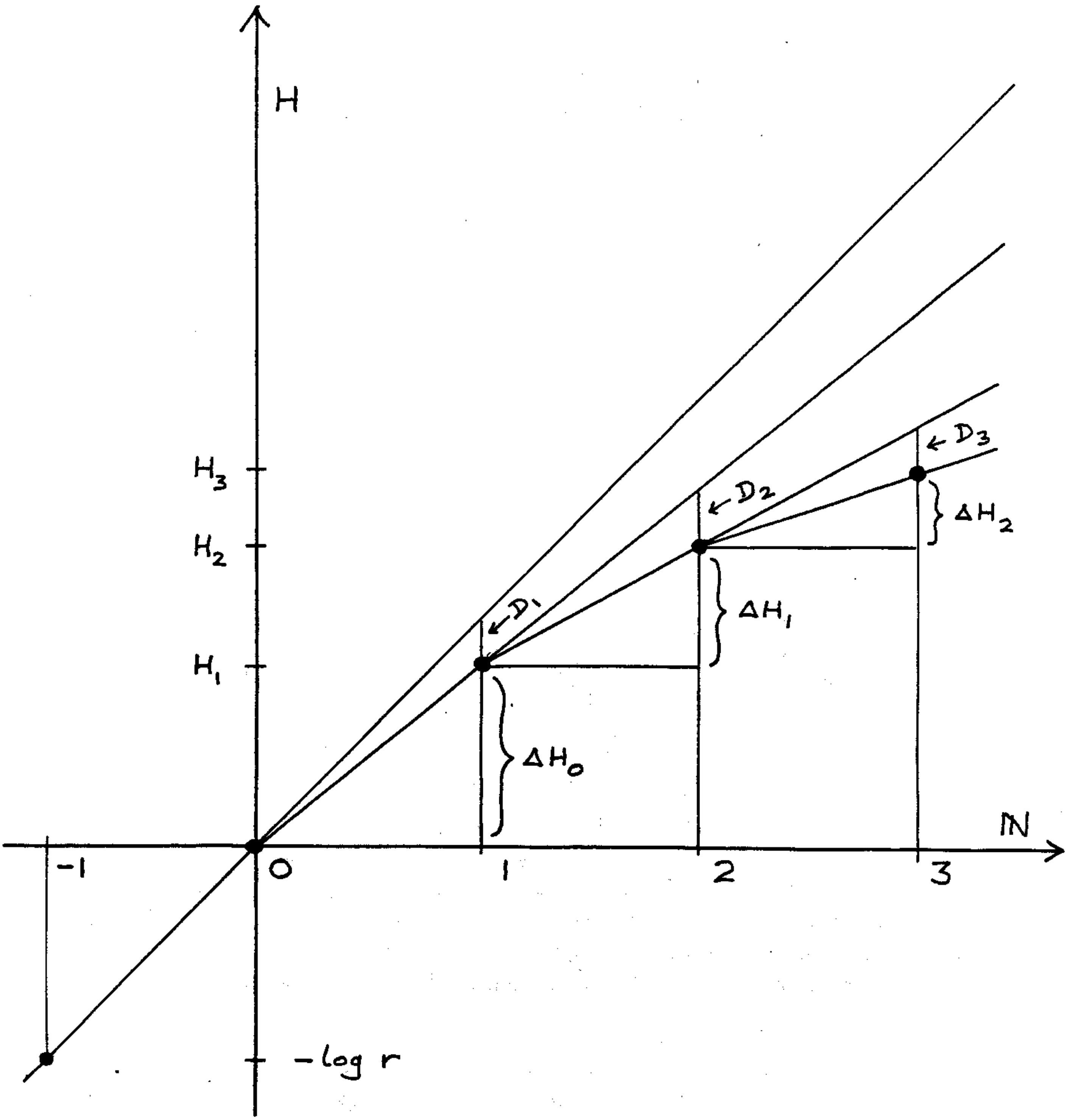


FIGURE 2. *Graph of a typical entropy function*

In Hierarchical Information Theory, an important quantity that Gatlin called the "information density" is obtained by summing all the divergences [Ga, (42a)]. Since this sum is infinite, the question of convergence arises. In what follows it will be shown that the sum does converge for partitioned Lebesgue spaces with automorphism.

Let $(M,\ S,\ \mu,\ T,\ \xi)$ be a partitioned Lebesgue space with automorphism. Then the limit

$$h(T,\xi) = \lim_{n\to\infty} H(\xi \mid \bigvee_{k=1}^{n} T^{-k}\xi) \tag{4.9}$$

exists [CF, §10.6]. Its value is called the *conditional entropy of T given ξ*. Write

$$I(T,\xi) = \sum_{n=1}^{\infty} D_n. \tag{4.10}$$

THEOREM 4.11. *In a partitioned Lebesgue space $(M,\ S,\ \mu,\ T,\ \xi)$, the limit (4.10) exists and takes the value*

$$I(T,\xi) = \log r - h(T,\xi). \tag{4.12}$$

Proof. For a positive integer n, one has $D_1 + D_2 + \cdots + D_{n+1} = -\Delta^2 H_{-1} - \Delta^2 H_0 - \cdots - \Delta^2 H_{n-1} = (\Delta H_{-1} - \Delta H_0) + (\Delta H_0 - \Delta H_1) + \cdots + (\Delta H_{n-1} - \Delta H_n) = \Delta H_{-1} - \Delta H_n = \log r - (H_{n+1} - H_n) = \log r - \left[H(\xi \vee T^{-1}\xi \vee \cdots \vee T^{-n}\xi) - H(\xi \vee \cdots \vee T^{1-n}\xi)\right] = \log r - \left[H(\xi \vee T^{-1}\xi \vee \cdots \vee T^{-n}\xi) - H(T^{-1}\xi \vee \cdots \vee T^{-n}\xi)\right] = \log r - H(\xi | \bigvee_{k=1}^{n} T^{-k}\xi)$, where the penultimate equality holds by applying (3.10) to the last terms on each side. Taking the limit as $n \to \infty$ gives the required result. □

One may thus formally define the *information density* of the partitioned Lebesgue space with automorphism to be $I(T,\xi) = \log r - h(T,\xi)$. The interpretation of equation (4.12) is that the density of information stored in the system $(M,\ S,\ \mu,\ T,\ \xi)$ is measured by the difference between its maximum possible density $\log r$ and its actual (ultimate) entropy $h(T,\xi)$. The interpretation of equation (4.10) is that the various divergences D_n measure how much of the total information density $I(T,\xi)$ is stored at the n-th level of the hierarchy.

5. Markov chains and Bernoulli shifts. As an initial illustration of hierarchical information theory as presented in the previous section, consider the first-order Markov chain model from the end of Section 2. For each positive integer n one has here that $-\Delta^2 H_n = \Delta H_n - \Delta H_{n+1} = H(T^{-n}\xi \mid \xi \vee \cdots \vee T^{1-n}\xi) - H(T^{-n-1}\xi \mid \xi \vee \cdots \vee T^{-n}\xi) = H(T^{-n}\xi \mid T^{1-n}\xi) - H(T^{-n-1}\xi \mid T^{-n}\xi) = H(T^{n}\xi \vee T^{1-n}\xi) - H(T^{1-n}\xi) - H(T^{-n-1}\xi \vee T^{-n}\xi) + H(T^{-n}\xi) = 0$, the third equality holding by the first-order Markov property. Thus D_{n+2} vanishes for positive n, and the decomposition (4.10) of the information density becomes

$$I(T,\xi) = D_1 + D_2 \tag{5.1}$$

($cf.$ [Ga, (49)]). Knowing

$$-\Delta^2 H_n = 0 \tag{5.2}$$

for positive n from above, along with (4.6), the conventional $H_0 = 0$ and the two values H_1, H_2, the entire funciton $H : \mathbf{N} \to \mathbf{R}^+$ may be reconstructed from the difference scheme of Fig. 3. Note in particular that no further calculations with (2.3) or (3.1) are required, beyond those used to obtain H_1, and H_2.

FIGURE 3. *Difference scheme for the entropy of a first-order Markov chain.*

$$H_{-1} = -\log r$$

$$\Delta H_{-1} = \log r$$

$$H_0 = 0 \qquad\qquad -\Delta^2 H_{-1} = \log r - H_1 = D_1$$

$$\Delta H_0 = H_1$$

$$H_1 \qquad\qquad -\Delta^2 H_0 = 2H_1 - H_2 = D_2$$

$$\Delta H_1 = H_2 - H_1$$

$$H_2 \qquad\qquad -\Delta^2 H_1 = 0 = D_3$$

$$\Delta H_2 = H_2 - H_1$$

$$H_3 = 2H_2 - H_1 \qquad\qquad -\Delta^2 H_2 = 0 = D_4$$

$$\Delta H_3 = H_2 - H_1$$

$$H_4 = 3H_2 - H_1$$

Remark 5.3. Hierarchical information theory may be used in this way to simplify the calculation of entropies of automorphisms in ergodic theory. Consider the example of the Bernoulli shift [CF, §10.6, Ex. 3]. By the mutual independence of successive states, $-\Delta^2 H_n = D_{n+2} = 0$ for all natural numbers n. Thus $\log r - H_1 = D_1 = I(T,\xi) = \log r - h(T,\xi)$, whence $h(T,\xi) = H_1 = -\sum_{i=1}^{r} p(i)\log p(i)$ directly.

Now consider the case of an m-th order Markov chain with r states $\{1,\ldots,r\}$. Reduce to the equivalent *derived* first-order Markov chain with set $Y = \{y = (y_1,\ldots,y_m) | 1 \le i \le r\}$ of r^m states [Bi,§6]. Let these states be distributed with stationary probabilities $p(y)$ and transition matrix $(p(y,z)|y,z \in Y)$, where

$$p(y,z) = \begin{cases} p(y_2,\ldots,y_m;z_m) \text{ if } y_2 = z_1,\ldots,y_m = z_{m-1} & ; \\ 0 \quad \text{otherwise} . \end{cases} \tag{5.4}$$

One may then set up an appropriate Lebesgue space (M,S,μ) as in Section 2 for the first-order case. Now, of course, M becomes the space $Y^{\mathbf{Z}}$ of all functions

$$x : \mathbf{Z} \longrightarrow Y;\ j \longmapsto x(j) = (x(j)_1,\ldots,x(j)_m). \tag{5.5}$$

The shift automorphsm T is again defined by (2.4). The partition $\xi = \{C_1,\ldots,C_r\}$ has

$$C_i = \{x \in M | x(0)_m = i\}. \tag{5.6}$$

The partitioned Lebesgue space (M,S,μ,T,ξ) obtained in this way is called the *m-th order Markov chain model.*

Let n be an integer not less than m. Then $-\Delta^2 H_n = \Delta H_n - \Delta H_{n+1} = H(T^{-n}\xi|\xi \vee \cdots \vee T^{1+m-n}\,\xi \vee T^{m-n}\xi \vee \cdots \vee T^{1-n}\xi) - H(T^{-n-1}\xi|\xi \vee \cdots \vee T^{m-n}\xi \vee T^{m-n-1}\xi \vee \cdots \vee T^{-n}\xi) = H(T^{-n}\xi|T^{m-n}\xi \vee \cdots \vee T^{1-n}\xi) - H(T^{-n-1}\,\xi|T^{m-n-1}\xi \vee \cdots \vee T^{-n}\xi) = 0$, i.e.

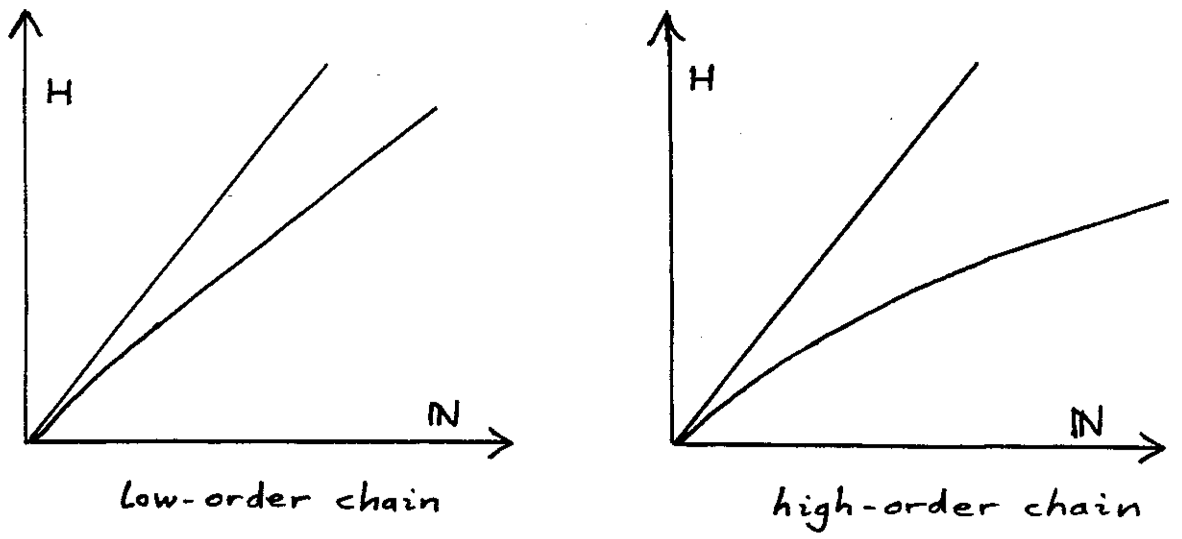
$$-\Delta^2 H_n = 0 \qquad \text{for} \qquad n \ge m \tag{5.7}$$

in the m-th order Markov source. The decomposition (4.10) of the information density reduces to

$$I(T,\xi) = D_1 + D_2 + \cdots + D_m + D_{m+1} \tag{5.8}$$

in this case (cf. [Ga, (42a)]). Fig. 4 contrasts the graphs of the entropy functions $H : \mathbf{N} \longrightarrow \mathbf{R}^+$ (as in Fig. 2) for low-order and high-order Markov chains. According to (5.7), the graph for the low-order chain becomes linear sooner than the graph for the high-order chain. It is instructive to relate these graphs to Fig. 1.

Fig. 4 *Entropy functions of Markov chains*



Acknowledgments. This paper was written at the Institute for Mathematics and its Applications during the 1987–88 Program in Applied Combinatorics. The author gratefully acknowledges a stimulating and encouraging correspondence with D.R. Brooks.

REFERENCES

[Bi] P. BILLINGSLEY, *Statistical Inference for Markov Processes*, University of Chicago Press, Chicago 1961.

[Br] D.R. BROOKS, *Incorporating origins into evolutionary theory*, in Understanding Origins.

[Bs] D.R. BROOKS, D.D. CUMMING, AND P.H. LEBLOND, *Dollo's Law and the Second Law of Thermodynamics: analogy or extension?*, in Entropy, Information and Evolution: New Perspectives on Physical and Biological Evolution (B.H. Weber, D.J. Depew and J.D. Smith eds.), pp 189–224, M.I.T. Press, Cambridge 1988.

[BW] D.R. BROOKS AND E.O. WILEY, *Evolution as Entropy*, University of Chicago Press, Chicago 1986.

[CB] D.D. CUMMING, D.R. BROOKS, AND P.H. LEBLOND, *Hierarchical information theory*, preprint.

[CF] I.P. CORNFELD, S.W. FOMIN, AND YA.G. SINAI, *Ergodic Theory*, Springer, New York 1982.

[Ga] L.L. GATLIN, *Information Theory and the Living System*, Columbia University Press, New York 1972.

[GC] GREAT COMMISSION STUDENTS, *You may have heard evolution is a fact*, Great Commission Church, Ames 1987.