NASA CONTRACTOR REPORT

NASA CR-478



# ON CERTAIN INFORMATION-THEORETIC CONCEPTS IN THE THEORY OF GRAPHS

*by M. Christine Wilson*

*Prepared by*
KENT STATE UNIVERSITY
Kent, Ohio
*for*

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • MAY 1966

NASA CR-478

# ON CERTAIN INFORMATION-THEORETIC CONCEPTS IN THE THEORY OF GRAPHS

By M. Christine Wilson

Distribution of this report is provided in the interest of information exchange. Responsibility for the contents resides in the author or organization that prepared it.

Prepared under Grant No. NsG-568 by
KENT STATE UNIVERSITY
Kent, Ohio

for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – Price $2.00

# ABSTRACT

In this thesis we present the basic information-theoretic concepts as utilized in measuring the amount of information contained in a given experiment. Our main object is the study of the concept of entropy which is defined as a measure of the amount of uncertainty. An extensive but concise review of the vàrious seemingly different approaches to the notion of entropy is made, and precise formulations are given for computing the entropy function for (i) the discrete probability distribution, (ii) the continuous probability distribution, (iii) the generalized probability distribution, and (iv) Markov chains.

Next, an investigation of the evolution of random graphs is made from the viewpoint of information theory. A graph, and especially a digraph, has been shown by Bhargava, among others, to be a reasonable probabilistic model in many applied situations such as group dynamics and communication theory. Time changes in such a graph are described by means of the evolution of a random graph, which in turn is formulated in terms of a stochastic process. In this thesis we evaluate the entropy function for such random graphs in a few special cases.

Finally, a specific group dynamics problem is considered, and the entropy is computed for the exact and approximate probability distributions of two particular cases. A brief empirical examination of these numerical computations is also made.

TABLE OF CONTENTS

# INTRODUCTION AND SUMMARY

The purpose of this thesis is twofold: (i) to present a review of basic information-theoretic concepts as utilized in measuring the amount of information contained in a given experiment, and (ii) to investigate the evolution of random graphs from the viewpoint of information theory.

Information theory is essentially a branch of the mathematical theory of probability and statistics and has applications in such fields as statistical physics, quantum mechanics, and biological sciences. However, the notion of information as we know it today has its basic foundations in the field of communication theory. A simple communication system consists of an information source, a transmitter, a channel, a receiver, and a destination. The information source contains a set of possible messages from which a desired message is chosen. This message is then changed into some sort of signal, by means of coding, to be transmitted along the channel to the receiver where it is decoded and sent to its destination. Any distortion of the original signal is called noise.

The amount of information obtained from such a system is not contained in the single message received, that is, information has no relationship to the meaning of the message. Instead, the concept of information applies to the situation as a whole and is a measure of the freedom of choice in selecting the message. If there is complete choice in selection, that is, if all messages are equally likely to be chosen, then maximum information is obtained. However, if the probabilities of being chosen are different for various messages of the set, then the measure of information becomes an obvious function of these probabilities.

The statistical concept of information as introduced by Fisher consists in measuring the amount of information directly from the given experiment. More specifically, in large samples, a numerical measure of information, which the sample contains and which is relevant to the parameter, is obtained as the reciprocal of the variance of the estimator. However, later it was found more fruitful to define a measure of information by determining the amount of uncertainty involved in the experiment. That this should be so follows from the observation that, a priori, just before an experiment is carried out our interest lies in the amount of uncertainty involved in the outcome of the experiment, while, posteriori, after the experiment has been carried out our concern is the amount of information gained from the experiment. It is not very surprising then that the measure of information and the measure of uncertainty are found to be closely related to each other.

Our main concern in this thesis is the information-theoretic concept of entropy which is defined to be a measure of the amount of uncertainty. The first systematic study of entropy and its properties was made by Shannon approximately twenty years ago. Since that time many well-known mathematicians such as Khinchin, Feinstein, Wolfowitz, Rényi and Pinsker have made important contributions to the mathematical soundness and theoretical development of the field of information theory, mainly from the viewpoint of entropy.

Today there exist two basic approaches for measuring information: (i) the axiomatic, or postulational, approach, and (ii) the pragmatic approach. The axiomatic approach has been found to be more useful in practice than the pragmatic approach; however, these two points of view are not really opposed to each other. We discuss these approaches at some length later in this thesis.

A graph, especially a digraph, which consists of a set of points, and a set of edges defined between some or all pairs of points in the set, has been shown by Bhargava, among others, to be a reasonable probabilistic model, under suitable assumptions, in certain practical situations. Time changes in such a graph (or digraph) can be described by means of the evolution of a random graph, which in turn can be formulated in terms of a stochastic process. In this thesis we evaluate the entropy function for such models in a few special cases.

The first chapter of this thesis is entirely devoted to various approaches to the notion of information and entropy, and whenever possible or relevant, to a comparative study of these seemingly different approaches. In section 1.1 we give definitions of statistical information as given by Fisher, Shannon and Kullback. In section 1.2 axiomatic approaches to the definition of entropy as expounded by Shannon, Khinchin and Rényi are presented. First, we define the entropy for the simplest case, namely for the finite discrete probability distribution, then for the continuous probability distribution, and finally for the generalized probability distribution. The pragmatic approach of Wolfowitz is presented in section 1.3, with a few short remarks comparing it with the axiomatic approach as given in section 1.2.

A unified approach to the mathematical definition of entropy directly in terms of information, as given by Dobrushin and expounded by Pinsker, is given in section 1.4, while a definition of entropy for Markov chains as developed by Ambarcumjan is given in section 1.5. Finally, we mention an axiomatic characterization of entropy, without presupposing probability, by means of finite Boolean rings, as given by Ingarden and Urbanik.

In the second chapter, first of all, some basic relevant definitions from the theory of graphs, directed graphs, and random graphs as studied by Erdös, Rényi, Gilbert, and Bhargava are given in sections 2.1 and 2.2. A measure of entropy for different kinds of random graphs and digraphs is derived in section 2.3. Finally, computations of the entropy for a specific probability model are made in section 2.4.

CHAPTER I

# ENTROPY AND INFORMATION

In this chapter we present the notion of statistical information, entropy as related to information, and a unified treatment of information and entropy. For details we refer to Shannon [22], Khinchin [16], Rényi [20] and Pinsker [19].

## 1.1 Statistical Information

Let x be a random variable with the probability density function given by $f(x|\theta)$. Let the density function f satisfy the well-known Cramer-Rao conditions, and $I(\theta)$ denote the information function.

Definition 1.1.1: (Fisher)

$$I(\theta) = - \int (\partial^2/\partial\theta^2) [\ln f(x|\theta)]f(x|\theta)dx$$

Definition 1.1.2: (Shannon)

$$I(\theta) = \int [\ln f(x|\theta)]f(x|\theta)dx$$

The following defines the information for "discriminating in favor of $H_1(\theta_1)$ against $H_2(\theta_2)$":

Definition 1.1.3: (Kullback)

$$I(\theta_1,\theta_2) = \int \ln[f(x|\theta_1)/f(x|\theta_2)] f(x|\theta_1)dx$$

For an extensive historical review of statistical information theory we refer to the excellent papers by Cherry [4], Fraser [10], Green [13], and Gnedenko [12].

## 1.2 *Axiomatic Approach to Entropy*

*Entropy of finite discrete distributions*: Let $A_1, A_2, \ldots, A_n$ be a complete system of events, that is, a set of events which are mutually exclusive and totally exhaustive.

*Definition 1.2.1* The complete system $\{A_i, 1 \leq i \leq n\}$, together with its probability set $\{p_i = P(A_i), 1 \leq i \leq n$: $p_i \overset{\geq}{\leq} 0, \sum_{i=1}^{n} p_i = 1\}$, is called a *finite scheme*, and is denoted by

$$A = \begin{pmatrix} A_1, & A_2, & \ldots, & A_n \\ p_1, & p_2, & \ldots, & p_n \end{pmatrix}$$

Every finite scheme describes a state of uncertainty and for many applications it is desirable to introduce a quantity which in a reasonable way measures the amount of uncertainty.

*Definition 1.2.2* The *entropy of the finite scheme A* is a measure of uncertainty of the scheme and is denoted by $H(p_1, p_2, \ldots, p_n)$; it is characterized by the following postulates:

*Postulate 1*: $H(p, 1-p)$ is continuous for $0 \leq p \leq 1$, and $H(1/2, 1/2) = 1$.

*Postulate 2*: $H(p_1, p_2, \ldots, p_n)$ is a symmetrical function of its variables.

Postulate 3: If all the $p_i$'s are equal, that is, $p_i = 1/n$ $(i=1,\ldots,n)$, then H is a monotonic increasing function of n.

Postulate 4: If $0 \leq t \leq 1$, then

$$H(p_1, \ldots, p_{n-1}, tp_n, (1-t)p_n) = H(p_1, \ldots, p_n) + p_n H(t,1-t).$$

Theorem 1.2.1 (Shannon) The only function satisfying the above postulates, viz. 1, 2, 3, and 4, is of the form

$$H(p_1, p_2, \ldots, p_n) = -K \sum_{i=1}^{n} p_i \log p_i \qquad (1.2.1)$$

where K is a positive constant which depends upon the choice of a unit of measure, and all logarithms are taken to an arbitrary but fixed base, with $p_i \log p_i = 0$ if $p_i = 0$.

Proof. Let $H(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}) = A(n)$. From postulate (4) we can decompose a choice from $s^m$ equally likely possibilities into a series of m choices each from s equally likely possibilities and obtain

$$A(s^m) = mA(s)$$

Similarly

$$A(t^n) = nA(t)$$

We can choose n arbitrarily large and find an m to satisfy

$$s^m \leq t^n < s^{(m+1)}$$

Thus, taking logarithms and dividing by n log s,

$$(1) \quad \frac{m}{n} \leq \frac{\log t}{\log s} \leq \frac{m}{n} + \frac{1}{n} \quad \text{or} \quad \left| \frac{m}{n} - \frac{\log t}{\log s} \right| < \varepsilon$$

where $\varepsilon$ is arbitrarily small. Now from the monotonic property of A(n), $A(s^m) \leq A(t^n) \leq A(s^{m+1})$

$$\text{or,} \quad mA(s) \leq nA(t) \leq (m+1)A(s).$$

Hence, dividing by nA(s),

$$(2) \qquad \frac{m}{n} \leq \frac{A(t)}{A(s)} \leq \frac{m}{n} + \frac{1}{n} \qquad \text{or} \qquad \left| \frac{m}{n} - \frac{A(t)}{A(s)} \right| < \varepsilon$$

Using (1) and (2) above

$$\left| \frac{A(t)}{A(s)} - \frac{\log t}{\log s} \right| \leq 2\varepsilon, \quad \text{which yields } A(t) = -K \log t$$

where K must be positive to satisfy postulate (3).

Now suppose we have a choice from n possibilities with commeasurable probabilities $p_i = \frac{n_i}{\sum n_i}$ where the $n_i$ are integers. We can break down a choice from $\sum n_i$ possibilities into a choice from n possibilities with probabilities $p_1, \ldots, p_n$ and then, if the ith was chosen, a choice from $n_i$ with equal probabilities. Using postulate (4) again, we equate the total choice from $\sum n_i$ as computed by two methods

$$K \log \sum n_i = H(p_1, \ldots, p_n) + K \sum p_i \log n_i$$

Hence

$$H(p_1, \ldots, p_n) = K \left[ \sum p_i \log \sum n_i - \sum p_i \log n_i \right]$$

$$= -K \sum p_i \log \frac{n_i}{\sum n_i} = -K \sum p_i \log p_i$$

If the $p_i$ are incommeasurable, they may be approximated by rationals and the same expression must hold by our continuity assumption. Thus the expression holds in general. The choice of coefficient K is a matter of convenience and amounts to the choice of a unit of measure.

It can be shown (see [16]) that the entropy function $H(p_1, p_2, \ldots, p_n)$ has the following properties, all of which we would intuitively expect of a reasonable measure of uncertainty.

(i) For fixed n the scheme with the most uncertainty is the one with equally likely outcomes, i.e., $p_i = 1/n$, $i = 1, \ldots, n$.

(ii) For two finite schemes

$$A = \begin{pmatrix} A_1, & A_2, \ldots, & A_n \\ p_1, & p_2, \ldots, & p_n \end{pmatrix} \qquad B = \begin{pmatrix} B_1, & B_2, \ldots, B_m \\ q_1, & q_2, \ldots, q_m \end{pmatrix}$$

such that A and B are mutually independent, the probability $\pi_{ij}$ of the joint occurrence of the events $A_i$ and $B_j$ is $p_i q_j$. The set of events $A_i B_j$, $(1 \leq i \leq n,\ 1 \leq j \leq m)$, with probabilities $\pi_{ij}$ represents another finite scheme, which is called the product of the schemes A and B and is designated by AB.

If H(A), H(B), and H(AB) are the corresponding entropies of the schemes A, B, and AB, then $H(AB) = H(A) + H(B)$.

(iii) For the case where the schemes A and B are mutually dependent, we denote by $q_{ij}$ the probability that the event

$B_j$ of the scheme B occurs, given that the event $A_i$ of the scheme A has already occurred, so that

$$\pi_{ij} = p_i q_{ij}, \qquad 1 \le i \le n, \quad 1 \le j \le m.$$

If $H_A(B)$ denotes the conditional mathematical expectation of the quantity $H(B)$ in the scheme A, then

$$H(AB) = H(A) + H_A(B)$$

(iv) In all cases $H_A(B) \le H(B)$, that is, on the average the knowledge of the outcome of scheme A can only decrease the uncertainty of scheme B.

(v) $H(p_1, p_2, \ldots, p_n, 0) = H(p_1, p_2, \ldots, p_n)$, that is, adding the impossible event or any number of impossible events to a scheme does not change its entropy.

We note that while (ii) above gives one of the most important properties of entropy, namely "additivity," we cannot replace postulate (4) in the characterization of entropy by property (ii), because (ii) is much weaker. We also remark that there are functions other than $H(P) = -\sum_{i=1}^{n} p_i \log p_i$ which satisfy postulates (1), (2), (3) and (ii). For example, Rényi gives the following definition of an entropy:

<u>Definition 1.2.3</u> The function

$$H_\alpha(p_1, p_2, \ldots, p_n) = \frac{1}{1-\alpha} \log_2\left(\sum_{i=1}^{n} p_i^\alpha\right), \quad \alpha > 0, \ \alpha \neq 1,$$

is called the entropy of order $\alpha$ of the distribution $P = (p_1, p_2, \ldots, p_n)$ and is denoted by $H_\alpha(P)$.

However, Shannon's measure of entropy is the limiting case, as $\alpha \to 1$, of Rényi's measure of entropy, that is,

$$\lim_{\alpha \to 1} H_\alpha(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i = H(P).$$

Hence Shannon's measure of entropy, which is denoted by $H_1(p_1, p_2, \ldots, p_n)$, may be called the entropy of order 1 of the distribution P.

Entropy of a Continuous Distribution

Definition 1.2.4 The entropy of a continuous distribution with the density function p(x) is

$$H = -\int_{-\infty}^{\infty} p(x) \log p(x)\, dx.$$

For an n-dimensional distribution with the density function $p(x_1, \ldots, x_n)$

$$H = \int \ldots \int p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n.$$

Given random variables x and y (which may themselves be multi-dimensional), we have

Definition 1.2.5 The joint entropy of p(x,y) is

$$H = -\int\int p(x,y) \log p(x,y)\, dx\, dy$$

Definition 1.2.6 The conditional entropies for p(x,y) are

$$H_x(y) = -\int\int p(x,y) \log \frac{p(x,y)}{p(x)} dx\, dy$$

$$H_y(x) = -\int\int p(x,y) \log \frac{p(x,y)}{p(y)} dx\, dy$$

where

$$p(x) = \int p(x,y)\, dy$$

$$p(y) = \int p(x,y)\, dx.$$

The entropies of continuous distributions have most (but not all) of the properties of the discrete case. For example, we have

(i) for any two variables x, y

$$H(x,y) = H(x) + H(y)$$

if and only if x and y are independent.(Random variables x and y are independent if $p(x,y) = p(x)\, p(y)$),

(ii) $H(x,y) = H(x) + H_x(y) = H(y) + H_y(x)$, and

$H_x(y) \leq H(y)$, $H_y(x) \leq H(x)$.

There is, however, one important difference between the discrete and continuous entropies. In the discrete case the entropy measures in an absolute way the randomness of the chance variable. In the continuous case the measurement is relative to the coordinate system; if the coordinates are changed, the entropy will, in general, also change. In the continuous case the entropy can be considered a measure of randomness relative to an assumed standard, namely the coordinate system chosen.

In spite of this dependence on the coordinate system, the entropy concept is as important in the continuous case as in the discrete case. This is due to the fact that the derived concepts of information rate and channel capacity depend on the difference of two entropies and this difference does not depend on the coordinate frame, each of the two terms being changed by the same amount.

## Entropy of Generalized Probability Distributions

The characterization of measures of entropy (and information) becomes much simpler if we consider these quantities as defined on the set of generalized probability distributions. Let $(\Omega, \beta, P)$ be a probability space in which $\Omega$ is an arbitrary nonempty set called the set of elementary events. $\beta$ is a $\sigma$-algebra of subsets of $\Omega$ which contains $\Omega$ itself. And P is a probability measure, which is a nonnegative and additive set function for which $P(\Omega) = 1$, defined on $\beta$. First we give some relevant standard definitions from the probability theory.

Definition 1.2.7 A function $\xi = \xi(\omega)$ which is defined for $\omega \in \Omega_1$, where $\Omega_1 \in \beta$ and $P(\Omega_1) > 0$, and which is measurable with respect to $\beta$ is called a generalized random variable.

Definition 1.2.8 If $P(\Omega_1) = 1$ then $\xi$ is called an ordinary (or complete) random variable.

Definition 1.2.9 If $0 < P(\Omega_1) < 1$, then $\xi$ is called an incomplete random variable. An incomplete random variable can be interpreted as a quantity describing the result of an experiment, depending on chance which is not always observable, only with probability $P(\Omega_1) < 1$.

Definition 1.2.10 The distribution of a generalized random variable is called a generalized probability distribution.

In particular, when $\xi$ takes on only a finite number of different values $x_1, x_2, \ldots, x_n$, the distribution of $\xi$ consists of the set of numbers $p_i = P(\xi = x_i)$ for $i = 1, 2, \ldots, n$. Thus,

Definition 1.2.11 A finite discrete generalized probability distribution is a sequence $p_1, p_2, \ldots, p_n$ of nonnegative numbers such that for $P = (p_1, p_2, \ldots, p_n)$ and $W(P) = \sum_{i=1}^{n} p_i$, we have $0 < W(P) \leq 1$. $W(P)$ is called the weight of the distribution. Thus the weight of an ordinary distribution is equal to 1.

Definition 1.2.12 An incomplete distribution is a distribution for which $W(P)$ is strictly less than 1.

Let $\Delta$ denote the set of all finite discrete generalized probability distributions, that is, $\Delta$ is the set of all sequences $P = (p_1, p_2, \ldots, p_n)$ of nonnegative numbers such that $0 < \sum_{i=1}^{n} p_i \leq 1$.

Definition 1.2.13 The entropy $H_1(P)$ of a generalized probability distribution $P = (p_1, p_2, \ldots, p_n)$ is characterized by the following five postulates:

Postulate 1: $H(P)$ is a symmetric function of the elements of $\Delta$.

Postulate 2: If $\{p\}$ denotes the generalized probability distribution consisting of the single probability p, then $H(\{p\})$ is a continuous function of p in the interval $0 < p \leq 1$.

Postulate 3: $H(\{1/2\}) = 1$.

Postulate 4: For $P \in \Delta$ and $Q \in \Delta$, $H(P * Q) = H(P) + H(Q)$.

If $P = (p_1, p_2, \ldots, p_m)$ and $Q = (q_1, q_2, \ldots, q_n)$ define two generalized distributions such that $W(P) + W(Q) \leq 1$, then we define $P \cup Q = (p_1, p_2, \ldots, p_m, q_1, q_2, \ldots, q_n)$; if $W(P) + W(Q) > 1$, $P \cup Q$ is not defined.

Postulate 5: If $P \in \Delta$, $Q \in \Delta$, and $W(P) + W(Q) \leq 1$,

$$\text{then } H(P \cup Q) = \frac{W(P)\, H(P) + W(Q)\, H(Q)}{W(P) + W(Q)} .$$

For a more detailed presentation see [20].

## 1.3 Pragmatic Approach to Entropy

In the preceding section we investigated the axiomatic approach to the concept of entropy. In other words, starting from the intuitive notion of information, we described the properties which a reasonable measure of information must possess and then it was necessary to find those mathematical expressions which satisfy the postulated properties.

On the other hand, it is valid to consider the pragmatic approach to the same problem. We may consider certain specific problems in information theory and accept as a measure of information the actual solution obtained. This approach has been emphasized by Wolfowitz in his book [23]. We present now, as an example, a simple coding problem and the formulation of its entropy as described by Rényi in [21].

Let $\xi_1, \xi_2, \ldots, \xi_n$ be a sequence of independent identically distributed random variables, each of which takes on the different values $x_1, x_2, \ldots, x_a$ with corresponding probabilities $p_1, p_2, \ldots, p_a$, that is,

$$\{p_k = P(\xi_j = x_k),\ 1 \leq k \leq a,\quad 1 \leq j \leq n:\quad p_k \geq 0,\ \sum_{k=1}^{a} p_k = 1\}.$$

The sequence $\xi_1, \xi_2, \ldots, \xi_n$ may be interpreted as produced by an information source emitting stationary and independent signals. Let $\Omega_n$ be the set of all ordered sequences of length $n$ of the symbols $x_1, x_2, \ldots, x_a$. Let a fixed number $\varepsilon$ $(0< \varepsilon <1)$ be given and consider those subsets E of $\Omega_n$ for which $P_n(E) \geq 1 - \varepsilon$ where $P_n(E)$ is the probability that the observed sequence $\xi_1, \xi_2, \ldots \xi_n$ belongs to the set E. Let $b(n,\varepsilon)$ denote the minimum of the number of elements of such sets. In other words, if $N(E)$ denotes the number of elements of such sets E, then $b(n,\varepsilon) = \min N(E)$, where $P_n(E) \geq 1 - \varepsilon$.

Now it can be shown that the limit

$$\lim_{n\to +\infty} \frac{\log_2 b(n,\varepsilon)}{n} = H(P)$$

exists, is independent of E, and it depends only on the distribution P.

That is,

$$H(P) = \sum_{k=1}^{n} p_k \log_2 \left(\frac{1}{p_k}\right).$$

If the numbers $p_k$ are all equal then

$$\log_2 b(n, \varepsilon) = nH(P) + \sqrt{n}\,\lambda \quad D - \frac{\log n}{2 \log 2} + O(1)$$

where

$$D = \left[\sum_{k=1}^{a} \left( \log_2 \frac{1}{p_k} - H(P)\right)^2 p_k\right]^{\frac{1}{2}}$$

and $\lambda$ is defined by $\Phi(\lambda) = 1 - \varepsilon$, where $\Phi(x)$ denotes the standard normal distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$

and $O(1)$ denotes a remainder term which remains bounded for $n \to +\infty$.

Thus we may consider as a pragmatic definition of the entropy $H(P)$ the following interpretation. If we want to express the sequence of signals $\xi_1, \xi_2, \ldots, \xi_n$ by a sequence of 0's and 1's so that the correspondence should be one-to-one then this can be accomplished by using sequences of 0's and 1's of length $nH(P) + O(\sqrt{n})$.

Thus if we accept as the unity of the amount of information the maximal amount of information which a signal capable of only two values (0 and 1) can carry, the H(P) can be interpreted as the amount of information per signal produced by a stationary source of independent signals, if the probability distribution of the possible values of the signals is

$P = (p_1, p_2, \ldots, p_a)$. In which case

$$H(P) \leq \log_2 a$$

with equality if $p_1 = p_2 = \ldots = p_a = 1/a$.

Comparison of the axiomatic approach to the pragmatic approach reveals that these are not really opposing points of view, but instead are complementary to each other. One provides a necessary check on the other. The axiomatic approach may give a theoretically sound characterization of entropy but it must also provide a mathematical expression which can be utilized to advantage in practical problems. On the other hand, solutions to specific problems are not of special consequence if they are basically unrelated to each other. However, if these solutions repeatedly follow a definite form having similar properties, then the approach is significant. Such is the case with the measure of information.

## 1.4 Entropy in Terms of Information

Let $\xi$ and $\eta$ be random variables with values in the measurable spaces $(X, S_x)$ and $(Y, S_y)$ respectively.

Definition 1.4.1 The value

$$I(\xi, \eta) = \sup \sum_{i,j} P_{\xi\eta}(E_i \times F_j) \log \frac{P_{\xi\eta}(E_i \times F_j)}{P_\xi(E_i) P_\eta(F_j)},$$

where the supremum is taken over all partitions $\{E_i\}$ of X and $\{F_j\}$ of Y, is called the information of one of the variables with respect to the other.

Definition 1.4.2 The quantity $I(\xi, \xi) = H(\xi)$ is called the entropy of the random variable $\xi$.

Definition 1.4.3 The function $i_{\xi\eta}(x,y) = \log a_{\xi\eta}(x,y)$ is called the information density of the random variables $\xi$ and $\eta$.

The following basic properties of information are due to Pinsker and proofs may be found in [19].

(i) $I(\xi, \eta) \geq 0$.

$I(\xi, \eta) = 0$ if and only if $\xi$ and $\eta$ are independent.

(ii) $I(\xi, \eta) = I(\eta, \xi)$.

(iii) If the random variables $\xi = (\xi_1, \xi_2)$ and $\eta = (\eta_1, \eta_2)$ are independent, then

$$I((\xi_1, \eta_1), (\xi_2, \eta_2)) = I(\xi_1, \xi_2) + I(\eta_1, r_2),$$

and with probability one

$$i((\xi_1,\xi_2),\ (\eta_1,\eta_2)) = i(\xi_1,\xi_2) + i(\eta_1,\eta_2).$$

(iv) If the random variable $\eta = f(\zeta)$ is a measurable function of the random variable $\zeta$, then

$$I(\xi,\eta) \leq I(\zeta,\eta).$$

(v) If $\xi = (\xi_1,\xi_2,\ldots)$, then

$$I(\xi,\eta) = \lim_{n \to \infty} I((\xi_1,\ldots,\ \xi_n),\ \eta).$$

Definition 1.4.4 The function $h_\xi(x) = \log a_{\xi\xi}(x,x) = \log 1/P_\xi(x) = -\log P_\xi(x)$ is called the entropy density of $\xi$, and is different from zero only on the countable set of points $x_1,\ x_2,\ \ldots$ for which $P(x_1),\ P(x_2),\ \ldots \neq 0$.

For our purposes, it will be sufficient to generalize the concept of entropy so that information becomes a special case of entropy.

Let $P_1$ and $P_2$ be two probability measures defined on the same measurable space $(\Omega,S_\omega)$ and let $\{E_i\}$ be a partition of $\Omega$.

Definition 1.4.5 The entropy $H_{P_2}(P_1)$ of $P_1$ with respect to $P_2$ is

$$H_{P_2}(P_1) = \sup \sum_i P_1\{E_i\} \log \left( P_1\{E_i\} / P_2\{E_i\}\right)$$

where the supremum is taken over all partitions of $\Omega$. Then

$$I(\xi,\eta) = H_{P_{\xi\times\eta}}(P_{\xi\eta}).$$

In general the properties of the entropy $H_{P_2}(P_1)$ are simply repetitions of the corresponding properties of information (see Pinsker [19]).

## 1.5 Entropy of Markov Chains

Let $A_1, A_2, \ldots, A_k, \ldots$ be events joining a stationary Markov chain of order r with a finite number of states $S_1, S_2, \ldots, S_n$ for every event. Let p(j), j=1,2,...,n denote the probability of the state $S_j$, and $\sum_{j=1}^{n} p_j = 1$. Let $p_{j_1 j_2 \ldots j_m}(j_{m+1})$ denote the probability of the state $S_{j_{m+1}}$ for the event $A_{m+1}$ given that the first m events were in the states $S_{j_1}, S_{j_2}, \ldots, S_{j_m}$.

In general, for chains of order r, we have

$$p_{j_1 j_2 \cdots j_m}(j_{m+1}) = p_{j_{m-r+1}\, j_{m-r+2} \cdots j_m}(j_{m+1}) \text{ where } m \geq r.$$

Definition 1.5.1 One-step entropy of a Markov chain of order r, we have denoted by $H_r$, and is defined by the quantity

$$H_r = -\sum_{j_1}\sum_{j_2} \cdots \sum_{j_r}\sum_{j_{r+1}} [p(j_1)\, p_{j_1}(j_2) \cdots p_{j_1 j_2 \cdots j_r}(j_{r+1})$$

$$\cdot \log \{p_{j_1 j_2 \cdots j_r}(j_{r+1})\}] .$$

We notice that

(i) for independent events

$$p_{j_1 j_2 \ldots j_m}(j_{m+1}) = p(j_{m+1}).$$

(ii) for simple chains (order one),

$$p_{j_1 j_2 \ldots j_m}(j_{m+1}) = p_{j_m}(j_{m+1}).$$

(iii) for chains of order two

$$p_{j_1 j_2 \ldots j_m}(j_{m+1}) = p_{j_{m-1} j_m}(j_{m+1}), \text{ etc.}$$

From this it is obvious that

(i) for independent events we have

$$H_r = H_{r-1} = \ldots = H_0$$

(ii) for simple chains (order one)

$$H_r = H_{r-1} = \ldots = H_1,$$

(iii) for chains of order two

$$H_r = H_{r-1} = \ldots = H_2, \text{ etc.}$$

Let $p(j_1 j_2 \ldots j_k)$ denote the probability of the joint occurrence of k events in states $j_1, j_2, \ldots j_k$ respectively.

Definition 1.5.2 The mean one-step entropy (denoted by $\overline{H}_k$) is defined by

$$\overline{H}_k = -\frac{1}{k} \sum_{j_1} \sum_{j_2} \ldots \sum_{j_k} p(j_1, j_2, \ldots, j_k) \log p(j_1, j_2, \ldots, j_k).$$

It can be shown that for Markov chains of order r,

$k > r$, we have $\lim_{k\to\infty} \overline{H}_k = H_r$.

For proof of the above, and a detailed treatment of various definitions of the Markov chains we refer to a paper of Ambarcumjan [1], translated from Russian by Dowker [6].

## 1.6 Remarks

For the sake of completeness, we also mention that a definition of entropy, without presupposing the notion of probability has been given by Ingarden and Urbanik [15]. This definition makes use of Boolean rings with elements of a ring being considered as events. Hence, the whole algebraic structure is interpreted as an experiment. The information is defined as a real-valued function on a set of finite Boolean rings. However, it is far too involved to be presented here. It must be pointed out here that Rényi [21] has shown that this approach which first introduces information, without using probabilities, does finally involve probabilities, and that the information thus defined can be expressed in terms of Shannon's entropy measure by means of a uniquely defined conditional probability measure. This incidentally also supports the view that the notion of information cannot be separated from that of probability.

# CHAPTER II

# ENTROPY AND GRAPH THEORY

In this chapter we present the fundamental notions and definitions for graphs and digraphs, and the two basic definitions of a random graph as given by Erodös and Rényi [7], and Gilbert [11]. The entropy function is formulated for the most common random graphs and digraphs, and actual values of the entropy are computed for a specific problem of group dynamics taken from Bhargava [3].

## 2.1 Definitions From Graph Theory

In this section we state in simple terms some of the definitions, from the theory of graphs and directed graphs, which are relevant to our work. Detailed accounts may be found in Berge [2], Bhargava [3], and Ore [18].

Let $A = \{P_1, P_2, \ldots, P_n\}$ be a finite collection of n points.

Definition 2.1.1 A *graph* G(A) of *order* n, on the set A consists of all the points in the set A (called *vertices* of the graph), and a set or collection of lines (called *edges* of the graph) joining pairs of points in the set A.

Definition 2.1.2 A *path* from vertex $P_i$ of G(A) to another vertex $P_j$ consists of a chain of edges from $P_i$ to $P_j$, and the number of lines in the path denotes the <u>length</u> of the path.

<u>Definition 2.1.3</u> A <u>cycle</u> is a closed path.

<u>Definition 2.1.4</u> A graph is <u>connected</u> if there exists a path between every pair of its points; otherwise <u>disconnected</u>.

<u>Definition 2.1.5</u> A <u>digraph</u> (directed graph) $\Gamma(A)$, of <u>order</u> n on the set A consists of all the points in the set A, and a set of <u>directed edges</u> joining ordered pairs of points in A.

<u>Definition 2.1.6</u> A <u>directed path</u> from $P_i$ to $P_j$ is a chain of directed edges of the form $\overrightarrow{P_iP_{i1}}$, $\overrightarrow{P_{i1}P_{i2}}$, ..., $\overrightarrow{P_{iL}P_j}$; the <u>length of the path</u> is L, the number of directed edges in the directed path.

<u>Definition 2.1.7</u> $\Gamma(A)$ is a <u>labelled digraph</u> if each vertex of $\Gamma(A)$ is distinguishable from every other vertex.

<u>Definition 2.1.8</u> A point $P_j$ is said to be <u>accessible</u> from a point $P_i$ if there is a directed path of some length greater than zero from $P_i$ to $P_j$.

<u>Definition 2.1.9</u> $\Gamma(A)$ is <u>strongly connected</u> if each point of A is accessible from every other point; $\Gamma(A)$ is <u>unilaterally connected</u> if, for every pair of points belonging to A, there is a directed path from at least one of them to the other; $\Gamma(A)$ is <u>weakly connected</u> if there is a chain of connections, ignoring all directions, from each point of A to every other point; $\Gamma(A)$ is <u>disconnected</u> if it is not even weakly connected.

## 2.2 Evolution of Random Graphs

There are two known different approaches to the theory of random graphs; one, due to Erdös and Rényi [7], and the other due to Gilbert [11]. We describe these briefly in this section.

Erdös and Rényi Approach: Let $E_{n,N}$ denote the set of all graphs having n given labelled vertices $P_1, P_2, \ldots, P_n$ and N edges. These graphs are not oriented, without parallel edges and without loops. Thus a graph belonging to the set $E_{n,N}$ is obtained by choosing N out of the possible $\binom{n}{2}$ edges between the points $P_1, P_2, \ldots, P_n$, and therefore the number of elements of $E_{n,N}$ is equal to $\binom{\binom{n}{2}}{N}$. We define a random graph, denoted by $\Gamma_{n,N}$, to be an element of $E_{n,N}$ chosen at random, so that each of the elements of $E_{n,N}$ have the same probability to be chosen, namely $1/\binom{\binom{n}{2}}{N}$.

Another point of view describes the formulation of a random graph as a stochastic process, but is equivalent to the one given above. At time $t = 1$ we choose one of the $\binom{n}{2}$ possible edges connecting the points $P_1, P_2, \ldots, P_n$, each of these edges having the same probability to be chosen; let this edge be denoted by $e_1$. At time $t = 2$ we choose one of the possible $\binom{n}{2} - 1$ edges, different from $e_1$, all these being equiprobable. Continuing this process at time $t = k + 1$ we choose one of the $\binom{n}{2} - k$ possible edges different from edges $e_1, e_2, \ldots, e_k$ already chosen, each of the remaining edges having the probability $1/\binom{n}{2} - k$ . We denote by $\Gamma_{n,N}$ the graph consisting of the vertices $P_1, P_2, \ldots, P_n$ and the edges $e_1, e_2, \ldots, e_N$.

The following theorem on connectedness of random graphs is due to Erdös and Rényi:

Theorem 2.2.1 Let $P_0(n, N_c)$ denote the probability of $\Gamma_{n,N}$ being completely connected for $N = N_c$, where $N_c = (1/2\ n \log n + cn)$. Then

$$\lim_{n\to+\infty} P_0(n,N_c) = e^{-e^{-2c}} .$$

For proof, we refer to [7].

Gilbert Approach: Let $P_1, P_2, \ldots, P_n$ be a set of n points. There are $n(n-1)/2$ lines which can be drawn joining pairs of these points. Any subset of these lines is a graph and there are $2^{n(n-1)/2}$ possible graphs in all. One of these graphs is chosen by the following random process, and is called a random graph: for all pairs of points make random choices, independent of each other, whether or not to join the points of the pair by a line. In such a random graph a point $P_i$ is connected to a point $P_j$ if there is a path of edges from $P_i$ to $P_j$, and the random graph is said to be connected if for every pair of points $(P_i,P_j)$, $P_i$ is connected to $P_j$.

Let the probability that the graph is connected be $P_n$ and the probability that two specific points are connected be $R_n$. We have (see Gilbert [11]):

Theorem 2.2.2

$$P_n = 1 - nq^{n-1} + 0(n^2q^{3n/2})$$

$$R_n = 1 - 2q^{n-1} + 0(nq^{3n/2})$$

or, asymptotically for n,

$$P_n \sim 1 - nq^{n-1}$$

$$R_n \sim 1 - 2q^{n-1}$$

where p is the probability of adding an edge to the random graph, and $q = 1 - p$ the probability of erasing an edge from the complete graph.

## 2.3 Entropy of Random Graphs

We derive, in this section, the form of the entropy function for a few special cases of the evolution of a random graph and digraph.

Special Cases:

(i) Let $A_1$ be the event that there is an edge from $P_i$ to $P_j$ in the random graph $\Gamma_n$ consisting of n vertices, and let $A_2$ be the event that there is no edge from $P_i$ to $P_j$. For all $i,j = 1, 2, \ldots, n$, we take $P(A_1) = 1/(n^2/2)$, and $P(A_2) = 1 - 1/(n^2/2)$, that is, the probability of

CR-478

finding any edge out of the $n^2/2$ possible edges (including loops; a loop is an edge from a point to itself) is the same.

We have a finite scheme

$$\begin{pmatrix} A_1 & A_2 \\ 1/(n^2/2) & 1 - 1/(n^2/2) \end{pmatrix}$$

for which the entropy H is given by

$$\begin{aligned} H &= -K\sum_{j=1}^{2} P(A_j)\log P(A_j) \\ &= -K[1/(n^2/2)\log 1/(n^2/2) + (1-1/(n^2/2)\log(1-1/(n^2/2))] \\ &= -K[2/n^2(\log 2/n^2 - \log(1-2/n^2)) + \log(1-2/n^2)] \\ H &= -K[2/n^2\log 2/(n^2-2) + \log(1-2/n^2)] \qquad (2.3.1) \end{aligned}$$

(ii) We note that if loops are not allowed then the probabilities $P(A_1)$ and $P(A_2)$ are $1/\frac{n(n-1)}{2}$, and $1 - 1/\frac{n(n-1)}{2}$ respectively, and the entropy H is given by

$$\begin{aligned} H &= -K\sum_{j=1}^{2} P(A_j)\log P(A_j) \\ &= -K\left(\frac{2}{n(n-1)}\log \frac{2}{n(n-1)} + [1-\frac{2}{n(n-1)}]\log[1 - \frac{2}{n(n-1)}]\right) \\ &= -K\left(\frac{2}{n(n-1)}[\log\frac{2}{n(n-1)} - \log\frac{n^2-n-2}{n(n-1)}] + \log[1-\frac{2}{n(n-1)}]\right) \\ H &= -K\left(\frac{2}{n(n-1)}\log\frac{2}{n^2-n-2} + \log[1 - \frac{2}{n(n-1)}]\right) \qquad (2.3.2) \end{aligned}$$

(iii) In general, let $P(A_1) = p$, $P(A_2) = q$, $q = 1 - p$, so that

$$H = -K[p \log p + (1-p)\log(1-p)]$$

$$H = -K[p \log \frac{p}{1-p} + \log(1-p)] \qquad (2.3.3)$$

(iv) In a random digraph, with proper modifications of terminology, $P(A_1) = 1/n^2$, $P(A_2) = 1 - 1/n^2$ for the case when loops are allowed; and $P(A_1) = 1/n(n-1)$, $P(A_2) = 1-1/n(n-1)$, for the case when loops are not allowed. For these two cases, we find:

$$H = -K[1/n^2 \log 1/n^2 + (1 - 1/n^2)\log(1 - 1/n^2)]$$

$$H = -K[1/n^2 \log 1/(n^2-1) + \log(1 - 1/n^2)] \qquad (2.3.4)$$

and,

$$H = -K[\frac{1}{n(n-1)} \log \frac{1}{n(n-1)} + (1 - \frac{1}{n(n-1)} \log(1- \frac{1}{n(n-1)})]$$

$$H = -K[\frac{1}{n(n-1)} \log \frac{1}{n^2-n-1} + \log(1 - \frac{1}{n(n-1)})] \qquad (2.3.4')$$

(v) Letting $p = 1/2$, we find, for the general finite scheme

$$\begin{pmatrix} A_1 & A_2 \\ 1/2 & 1/2 \end{pmatrix}$$

the maximum value, $H_{max}$, of the entropy $H$:

$$\begin{aligned} H_{max} &= -K[1/2 \log 1/2 + 1/2 \log 1/2] \\ &= -K[(1/2 + 1/2)\log 1/2] \\ &= -K[-\log 2] \end{aligned}$$

$$H_{max} = K \log 2 \qquad (2.3.5)$$

General Cases:

(i) Consider now the general scheme in a random graph (with loops), with events $A_j$ taking place with probabilities $P(A_j) = 1/\binom{n^2/2}{j}$, $j = 1, 2, \ldots, n^2/2$. The entropy function is

$$H = -K \sum_{j=1}^{n^2/2} \left( \frac{(\frac{n^2}{2} - j)!\ j!}{(\frac{n^2}{2})!} \right) \log \left( \frac{(\frac{n^2}{2} - j)!\ j!}{(\frac{n^2}{2})!} \right)$$

$$H = -K \sum_{j=1}^{n^2/2} \left( \frac{(\frac{n^2}{2} - j)!\ j!}{(\frac{n^2}{2})!} \right) \left[ \sum_{k=1}^{\frac{n^2}{2}-j-2} \log \left(\frac{n^2}{2} - j - k\right) + \sum_{k=1}^{j-2} \log(j-k) - \sum_{k=1}^{\frac{n^2}{2}-2} \log\left(\frac{n^2}{2} - k\right) \right] \tag{2.3.6}$$

(ii) The other interesting case of (i) consists in taking $P(A_j) = 1/(\frac{n^2}{2} - j + 1)$ in which case we get

$$H = -K \sum_{j=1}^{n^2/2} \frac{1}{(\frac{n^2}{2} - j + 1)} \log \frac{1}{(\frac{n^2}{2} - j + 1)}$$

$$H = K \sum_{j=1}^{n^2/2} \frac{1}{(\frac{n^2}{2} - j + 1)} \log \left(\frac{n^2}{2} - j + 1\right) \tag{2.3.7}$$

(iii) Finally, taking $P(A_j)$ to be all equal to $1/(n^2/2)$, we get the maximum entropy

$$H_{max} = -K \sum_{j=1}^{n^2/2} P(A_j) \log P(A_j)$$

$$H_{max} = -K[(1/(n^2/2) + \dots + 1/(n^2/2)) \log 1/(n^2/2)]$$

$$H_{max} = K \log(n^2/2) \qquad (2.3.8)$$

(iv) We may similarly derive the entropy function H for the three cases in a random digraph (with loops), giving

$$H = -K \sum_{j=1}^{n^2} P(A_j) \log P(A_j) \quad ; \quad P(A_j) = 1/\binom{n^2}{2}$$

$$H = -K \sum_{j=1}^{n^2} \left( \frac{(n^2-j)!\ j!}{(n^2)!} \right) \left[ \sum_{k=1}^{n^2-j-2} \log(n^2-j-k) \right. +$$

$$\left. \sum_{k=1}^{j-2} \log(j-k) - \sum_{k=1}^{n^2-2} \log(n^2-k) \right] \qquad (2.3.9)$$

$$H = -K \sum_{j=1}^{n^2} P(A_j) \log P(A_j) \quad ; \quad P(A_j) = 1/(n^2-j+1)$$

$$H = K[\frac{1}{(n^2-j+1)} \log (n^2-j+1)] \qquad (2.3.10)$$

$$H_{max} = -K \sum_{j=1}^{n^2} P(A_j) \log P(A_j) \quad ; \quad P(A_j) = 1/n^2$$

$$H_{max} = K \log n^2 \qquad (2.3.11)$$

Erdős and Rényi case: Let $A_1$ be the event that a random graph $\Gamma_{n,N_c}$ is completely connected, so that $P(A_1) = e^{-e^{-2c}}$ (Theorem 2.2.1); let $A_2$ be the event that $\Gamma_{n,N_c}$ is not completely connected, then $P(A_2) = 1 - e^{-e^{-2c}}$. In this case

$$H = -K[e^{-e^{-2c}} \log e^{-e^{-2c}} + (1-e^{-e^{-2c}})\log(1-e^{-e^{-2c}})]$$

$$H = -K[-e^{-(2c+e^{-2c})} + \log(1 - e^{-e^{-2c}}) - e^{-e^{-2c}}\log(1 - e^{-e^{-2c}})] \qquad (2.3.12)$$

Other cases may be handled in a more or less similar manner.

## 2.4 Entropy for a Probability Model

We compute now the value of the entropy function for an actual set of data taken from an example given in Bhargava [3]. In [3] Bhargava considers a digraph type model for a group dynamics situation which consists of 25 "members," each making three "choices" of "association" from among the remaining 24 "members". The "group configuration" thus obtained can be reasonably represented by a digraph with n=25 vertices, and N=75 directed edges, such that there are exactly three outgoing edges from every vertex to any of the remaining 24 vertices. In his probabilistic model, Bhargava's approach consists in viewing the total "group configuration" as an aggregate of the "subgroup configurations of order k, $k \geq 2$," where each of the "k-order subgroup configurations" is itself an aggregate of the "binary dyadic relations" between pairs of "members" of the group. In this fashion, a stochastic process is described, and exact and apprcximate probabilities are derived for various cases. We investigate below two of these cases, for k=2 and k=3, from the viewpoint of the amount of information contained in these experiments.

<u>Case k = 2</u>: The events $A_1$, $A_2$, $A_3$ correspond, for a digraph of order two, to the connectedness properties---strongly connected, unilaterally (but not strongly) connected, and disconnected. (There are no weakly connected digraphs of order two.) The exact and approximate probabilities are

denoted by $P(A_j)$ and $\overline{P}(A_j)$ respectively, and are given by

(i) $P(A_1) = \theta^2$, $P(A_2) = 2\theta(1 - \theta)$, $P(A_3) = (1 - \theta)^2$

where $\theta$ is the ratio of the number of outgoing edges from a vertex to the total number of vertices minus one ($\theta$ is also called the "rate of valency").

(ii) For $\theta = 3/24$, we have

$$P(A_1) = 5/300,\ P(A_2) = 68/300,\ P(A_3) = 227/300.$$

$$\overline{P}(A_1) = 5/300,\ \overline{P}(A_2) = 66/300,\ \overline{P}(A_3) = 229/300.$$

For case (i) the entropy function H is given by

$$\begin{aligned}
H &= -K \sum_{j=1}^{3} P(A_j) \log P(A_j) \\
&= -K[\theta^2 \log \theta^2 + 2\theta(1-\theta) \log 2\theta(1-\theta) + \\
&\qquad (1-2\theta+\theta^2)\log(1-2\theta+\theta^2)] \\
&= -K[2\theta^2 \log\theta/2(1-\theta) + 2\theta\log 2\theta(1-\theta)/(1-\theta)^2 + \\
&\qquad \theta^2\log(1-\theta)^2 + \log (1-\theta)^2] \\
&= -K[\theta^2\log 1/4 + 2\theta\log 2\theta/(1-\theta) + \log (1-\theta)^2] \\
H &= -K[-.6021\theta^2 + 2\theta\log 2\theta/(1-\theta) + 2 \log (1-\theta)] \quad (2.4.1)
\end{aligned}$$

where $K \doteq 2.103$ using common logarithms to base 10.

For case (ii) the entropy functions $H_{exact}$ and $H_{approx.}$ are found to be (using logarithms to base 10)

$$H_{exact} = -K \sum_{j=1}^{3} P(A_j) \log P(A_j)$$

$$= -K[5/300 \log 5/300 + 68/300 \log 68/300 + 227/300 \log 227/300]$$

$$= -K[.0167(-1.77815) + .2267(-.64461) + .7567(-.12109)]$$

$$= -K[-.2674]$$

$$H_{exact} = .5623 \qquad (2.4.2)$$

$$H_{approx} = -K \sum_{j=1}^{3} \overline{P}(A_j) \log \overline{P}(A_j)$$

$$= -K[5/300 \log 5/300 + 66/300 \log 66/300 + 229/300 \log 229/300]$$

$$= -K[-.02970 -.1447 -.08974]$$

$$= -K[-.2641]$$

$$H_{approx} = .5554 \qquad (2.4.3)$$

We remark that the values of $H_{exact}$ and $H_{approx}$, as given by (2.4.2) and (2.4.3) respectively are not much different from each other, and from the viewpoint of the entropy, the approximation seems to be quite good.

Case k = 3: The events $A_1$, $A_2$, $A_3$, $A_4$ correspond, for a digraph of order three, to the connectedness properties---

strongly connected, unilaterally (but not strongly) connected, weakly (but not unilaterally) connected, and disconnected. The exact and approximate probabilities are again denoted by $P(A_j)$ and $\bar{P}(A_j)$ respectively, and are given by

(i) $P(A_1) = \theta^3(2 + 3\theta - 3\theta^2 + 2\theta^3)$

$P(A_2) = 6\theta^2(1 - \theta)(1 - 2\theta^2 + \theta^3)$

$P(A_3) = 6\theta^2(1 - \theta)^4$

$P(A_4) = (1 - \theta)^4(1 + 4\theta - 2\theta^2)$

where $\theta$ is the rate of valency as previously described.

(ii) For $\theta = 3/24$, we have

| | |
|---|---|
| $P(A_1) = 10/2300$ | $\bar{P}(A_1) = 10/2300$ |
| $P(A_2) = 186/2300$ | $\bar{P}(A_2) = 183/2300$ |
| $P(A_3) = 112/2300$ | $\bar{P}(A_3) = 126/2300$ |
| $P(A_4) = 1992/2300$ | $\bar{P}(A_4) = 1981/2300$ |

For case (ii) the entropy functions $H_{exact}$ and $H_{approx.}$ are found to be

$$H_{exact} = -K \sum_{j=1}^{4} P(A_j) \log P(A_j)$$

$$= -K[10/2300 \log 10/2300 + 186/2300 \log 186/2300 + 112/2300 \log 112/2300 + 1992/2300 \log 1992/2300]$$

$$= -K[.004348\ (-2.36173) + .08087(-1.09222) + .4870\ (-1.31251) + .8661\ (-.06244)]$$

$$= -K[-.21660]$$

$$H_{exact} = .3598 \qquad (2.4.4)$$

where K $\doteq$ 1.661.

$$H_{approx.} = -K \sum_{j=1}^{4} P(A_j) \log P(A_j)$$

$$= -K[10/2300 \log 10/2300 + 183/2300 \log 183/2300 + 126/2300 \log 126/2300 + 1981/2300 \log 1981/2300]$$

$$= -K[.004348(-2.36173) + .07956(-1.09928) + .06897(-1.26136) + .8618(-.06485)]$$

$$= -K[-.22265]$$

$$H_{approx.} = .3698 \qquad (2.4.5)$$

By putting $P(A_j) = 1/4$, $j = 1, 2, 3, 4$, we get

$$H_{max.} = -K \sum_{j=1}^{4} P(A_j) \log P(A_j)$$

$$= -K[1/4 \log 1/4 + 1/4 \log 1/4 + 1/4 \log 1/4 + 1/4 \log 1/4]$$

$$= -K[-\log 4]$$

$$= -K[-.60206]$$

$$H_{max.} = 1.00 \tag{2.4.6}$$

We notice that for case k = 2 the entropy is larger for the exact probabilities than for the approximate probabilities, as we would expect. And for case k = 3 the reverse situation is true; the entropy for the approximate probabilities is greater than that for the exact probabilities. However, in both cases, the absolute difference between the exact case and the approximate case is relatively very small, and, hence, from the viewpoint of entropy the approximations seem very good.

## REFERENCES


[1] Ambarcumjan, G. A. "Entropy of Markov Chains." Izv. Akad. Nauk Armjan. SSR Ser. Fiz.-Mat. Nauk, 11 (1958), 31-40.

[2] Berge, C. The Theory of Graphs and Its Applications. Translated by A. Doig. New York: Wiley and Sons, Inc., 1962.

[3] Bhargava, T. N. "A Stochastic Model for Time Changes in a Binary Dyadic Relation, with Application to Group Dynamics." Unpublished Ph.D. dissertation, Michigan State University, 1962.

[4] Cherry, C. (ed.) Information Theory, Papers Read at a Symposium on 'Information Theory'. Royal Institution, London, Sept., 1955; New York: Academic Press, 1956.

[5] Dobrushin, R. L. "General Formulation of Shannon's Basic Theorems of the Theory of Information." Dokl. Akad. Nauk SSSR, vol. 126 (1959), 474.

[6] Dowker, Y. N. Selected Translations in Mathematical Statistics and Probability, 4 (1963), 1-11.

[7] Erdös, P., and Rényi, A. "On Random Graphs." Publications Mathematicae, 6 (1959), 290-297.

[8] Feinstein, A. Foundations of Information Theory. New York: McGraw-Hill, Inc., 1958.

[9] Fisher, R. A. "Theory of Statistical Estimation." Proc. Cambridge Philos. Soc., 22 (1925), 700-725.

[10] Fraser, D. A. S. "On Information in Statistics." Annals of Mathematical Statistics, 36 (1965), 890-896.

[11] Gilbert, E. N. "Random Graphs." Annals of Mathematical Statistics, 30 (1959), 1141-1144.

[12] Gnedenko, B. V. On Some Soviet Papers on Information Theory. ("Transactions of the First Prague Conference on Information Theory; Statistical Decision Functions and Stochastic Processes," 21-28), Prague: Publishing House of the Czechoslovac Academy of Science, 1957.

[13] Green, P. E., Jr. "Information Theory in the U.S.S.R." IRE WESCON Convention Record, part 2 (1957), 67-83.

[14] Ingarden, R. S., and Urbanik, K. "Information as a Fundamental Notion of Statistical Physics." Bull. de l'Acad. Sci. Polonaise, Ser. Sci. Math, 9 (1961), 313-316.

[15] Ingarden, R. S., and Urbanik, K. "Information without Probability." Colloquium Mathematicum, 9 (1962), 131-150.

[16] Khinchin, A. I. Mathematical Foundations of Information Theory, New York: Dover Publications, Inc., 1957.

[17] Kullback, S. Information Theory and Statistics. New York: Wiley and Sons, Inc., 1959.

[18] Ore, O. Theory of Graphs. American Math. Soc. Coll. Publ., 38, Providence, 1962.

[19] Pinsker, M. S. Information and Information Stability of Random Variables and Processes. London: Holden-Day, Inc., 1964.

[20] Rényi, A. "On Measures of Entropy and Information." Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, I (1961) 547-553.

[21] ______________. "On the Foundations of Information Theory." Unpublished paper. Math. Institute of the Hungarian Acad. of Sci.

[22] Shannon, C. E. A Mathematical Theory of Communication. Urbana: University of Illinois Press, 1963.

[23] Wolfowitz, J. Coding Theorems of Information Theory. Englewood Cliffs, N. J.: Prentice Hall, Inc. 1961.



NASA-Langley, 1966