Calculation of information flow rate from mutual information

1988 J. Phys. A: Math. Gen. 21 1405

(http://iopscience.iop.org/0305-4470/21/6/016)

J. Phys. A: Math. Gen. 21 (1988) 1405–1414. Printed in the UK

Calculation of information flow rate from mutual information

Kenji Matsumoto† and Ichiro Tsuda‡

† Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo 060, Japan
‡ Bioholonics Project, Research Development Corporation of Japan, Nissho Building, 5F, 14–24 Koishikawa 4-chome, Bunkyo-ku, Tokyo 112, Japan

Received 31 March 1987, in final form 9 November 1987

Abstract. The information flow rates in a time series and between two time series are defined. These can be calculated from mutual information curves. These flow rates enable us to trace the origin of information flowing in an element of a system. This is actually done in a coupled system of one-dimensional maps, in which a flow of information is quantitatively measured.

1. Introduction

In understanding complex chaotic dynamics, a quantity called the KS entropy has been proved to be useful and essential (Kolmogorov 1958). Its origin is in information theory and it measures the rate of information production in dynamical systems. Since a high information production rate means a high degree of randomness, KS entropy is considered to characterise the randomness of chaotic dynamics.

Recently, KS entropy has been reinterpreted as the rate of information flow in dynamical systems (Shaw 1981). It is a flow of information about initial conditions from lower figures of values of the variables in the system to higher figures. This flow is a different expression for a property of chaotic dynamics called 'the sensitive dependence on initial conditions'. The existence of this type of information flow is a salient characteristic of chaotic dynamics.

Although the concept of information flow is fascinating, the application of this concept is limited due to the fact that the generalisation of KS entropy as a flow rate is not obvious. This is because the definition of KS entropy is not based on the concept of 'flow'.

In this paper, we present a theoretical framework for the concept of information flow, and define various types of information flow rates which can be calculated by computers. They may be considered as the natural extensions of KS entropy.

The term 'flow rate' means the amount of change in unit time. The amount of new information at a certain time t may be taken as a definition of information flow rate. But it is not enough for the information flow rates to simply compare the status of the system at times t and $t-1$. This is because the same information which appears and disappears in an element of the system more than once must be counted only once. To obtain the net information flow rates, we must define the amount of new information at time t which has not appeared previously.

To define a rate in this way, we must know which information is new at any particular time. A formal decomposition of information is useful here. In this decomposition, each term represents the amount of information at a particular time; we know

0305-4470/88/061405+10\$02.50

exactly when the information represented by each term exists in the system. Using this decomposition, to obtain the information flow rates we simply have to collect appropriate terms in the decomposition. Moreover, we can derive the relation between them comparing the terms in them.

The decomposition is explained in § 2. The definition of information flow rates and the derivations of the relations between them are in § 3. In § 4, we show that we can approximate the information flow rates defined in § 3 by calculable quantities. Section 5 contains an example of the calculation of information flow rates, thus demonstrating the feasibility of the formalism.

2. Inductive definition of conditional mutual information

In the decomposition, a central role is played by the conditional mutual information, which is defined below inductively.

In this section, a variable like i_k denotes a random variable which takes one of a number of finite states according to a particular probability. In the next section, the notation for variables has a special meaning. We assume that the joint probabilities necessary for the calculation of information are known.

Let us begin with the fundamental quantities in information theory (Shannon and Weaver 1949). The Shannon entropy of a random variable i with probability $p(i)$ is defined as

$$H(i)=\sum_i p(i)\log\frac{1}{p(i)}. \tag{2.1}$$

The definition of the mutual information $I(i;j)$ between random variables i and j can be rewritten in the following form:

$$H(i)=H_j(i)+I(i;j). \tag{2.2}$$

The first term $H_j(i)$ is the conditional entropy defined as

$$H_j(i)=\sum_{i,j} p(j)p(i|j)\log\frac{1}{p(i|j)} \tag{2.3}$$

where $p(i|j)$ is the conditional probability of i when the value of j is known. This represents the amount of information in variable i which is independent of variable j. $I(i;j)$ represents the information common to both variables.

Equation (2.2) is the first step of our decomposition. In this equation the information of the variable i is decomposed into two terms where one is independent of the variable j and the other is not. The same type of formal decomposition can be carried out on the terms on the right-hand side of (2.2). In this way, we can inductively define a series of new quantities which we will call conditional mutual information. Note that these quantities are defined formally according to whether they are independent of a particular set of variables or not.

The first inductive relation decomposes a series of conditional entropies and defines the conditional mutual information with two arguments,

$$H_{j_1,\ldots,j_n}(i)=H_{j_1,\ldots,j_n,k}(i)+I_{j_1,\ldots,j_n}(i;k). \tag{2.4}$$

With the definition of the conditional entropy,

$$H_{j_1,\ldots,j_n}(i)=\sum_{i,j_1,\ldots,j_n} p(j_1,\ldots,j_n)p(i|j_1,\ldots,j_n)\log\frac{1}{p(i|j_1,\ldots,j_n)} \tag{2.5}$$

where $p(i|j_1,\ldots,j_n)$ is the conditional probability of variable i when the values of variables $j_1,\ldots,j_n$ are known. Note that in (2.4) the conditional entropy on the left-hand side is decomposed into two terms on the right-hand side. The first represents the part of information independent of k, and the other represents the part common to k and i. The latter is the conditional mutual information of two arguments.

The second inductive relation below defines the conditional mutual information with more than two arguments starting with those with two arguments,

$$I_{j_1,\ldots,j_m}(i_1;\ldots;i_n)=I_{j_1,\ldots,j_m,k}(i_1;\ldots;i_n)+I_{j_1,\ldots,j_m}(i_1;\ldots;i_n;k). \tag{2.6}$$

By definition, the conditional mutual information represents the amount of information independent of the variables in its subscripts and common to the variables in its arguments. In appendix 1, we show that the conditional mutual information is invariant with respect to any permutations in its arguments or its subscripts.

3. Definitions of information flow rates

We consider a time series $i(n)$, $-\infty<n<+\infty$, of finite states. In the following calculations of information, we need the probabilities of type $p(i_0,\ldots,i_n)$ of the occurrence of a particular sequence of states $i_0,\ldots,i_n$ in the time series. In this notation, each i_k is a random variable taking one of the states and the subscript denotes the relative time ordering among the random variables in the arguments. Since this numbering is relative, we have $p(i_0,\ldots,i_n)=p(i_{0+l},\ldots,i_{n+l})$ for any integer l. When we consider two time series $i(n)$ and $j(n)$, we write i_k and j_k for the corresponding random variables. There are an infinite number of random variables i_n, $-\infty<n<+\infty$, for a time series.

It is convenient for collection of terms in the decomposition to consider conditional mutual information with an infinite number of arguments and/or subscripts. We use the following abbreviation for the conditional mutual information:

$$J(i_{k_1},\ldots,i_{k_n})=I_{\text{all the other variables}}(i_{k_1};\ldots;i_{k_n}). \tag{3.1}$$

We call this type of quantity the fundamental information. The whole set of random variables is i_n, $-\infty<n<+\infty$, for one time series, and i_n and j_n, $-\infty<n<+\infty$, for two time series. According to the definition of conditional mutual information, these represent the amount of information common to the variables in the arguments and independent of all the other variables. In other words, $J(i_{k_1},\ldots,i_{k_n})$ means that the same information of this amount appears at time $k_1,\ldots,k_n$ in the time series. From the definition of the random variables we have time translational symmetry $J(i_{k_1},\ldots,i_{k_n})=J(i_{k_1+l},\ldots,i_{k_n+l})$ for any integer l. We do not discuss their convergence, since they appear only in the form of summations, which are obviously not divergent.

Now we define the information flow rate in a single time series. We define the amount of new information that first appears in the time series at a particular time as the rate. This type of information is represented by fundamental information that have i_0 and do not have i_k for $k<0$ as their arguments. The rate is the summation of all the fundamental information. If we denote by s a set of arguments, the information flow rate $K(i)$ in the time series i is defined as follows:

$$K(i)=\sum_{s\ni i_0\text{ and }s\not\ni i_k\text{ for }k<0}J(s). \tag{3.2}$$

We will see in § 4 that this quantity is equivalent to KS entropy when the partition is appropriately taken.

In some cases such as high-dimensional dynamical systems or systems consisting of many elements, the analysis of more than one time series is desirable. This type of analysis is one of the aims of the present formalism. In the following, we consider two time series i and j. We will define nine types of information flow rate and derive the relations between them.

The rate of information flow in two time series is the amount of new information that appears first in one or both of the two time series at a certain time. This type of information is represented by the fundamental information that does not have i_k, j_k for $k<0$ and has at least i_0 or j_0. The summation of all this fundamental information is the total information flow rate $K(i,j)$,

$$K(i,j)=\sum_{s\ni i_0,j_0 \text{ and } s\not\ni i_k,j_k \text{ for } k<0} J(s). \tag{3.3}$$

The other types of flow rates are defined as partial summations of this total summation.

Among the fundamental information in $K(i,j)$, some information does not have the random variables of time series j as arguments. We denote by $K_j(i)$ the summation of all this fundamental information,

$$K_j(i)=\sum_{s\ni i_0 \text{ and } s\not\ni i_k \text{ for } k<0 \text{ and } s\not\ni j_l \text{ for any } l} J(s). \tag{3.4}$$

This represents the rate of information flow in the time series i independent of the time series j. Likewise, we can define $K_i(j)$ which represents the rate of information flow in j independent of i. The summation of the other fundamental information in $K(i,j)$ is the cross information flow rate $K(i;j)$. This is the rate of information flowing between two time series. By definition, we have the following relation between the above four rates:

$$K(i,j)=K_i(j)+K_j(i)+K(i;j). \tag{3.5}$$

The rate $K(i)$ of information flow in the time series i is the amount of new information that first appears in the time series at time 0. Using the time translational symmetry of the random variables, we can show that this is equal to the following summation:

$$K(i)=K_j(i)+K(i;j). \tag{3.6}$$

We have a corresponding relation for the time series j:

$$K(j)=K_i(j)+K(i;j). \tag{3.7}$$

Among the fundamental information in the summation for $K(i;j)$, we can recognise three types according to whether they have both i_0 and j_0 or only one of them. The fundamental information containing both represents the information that appears in both time series at the same time. We denote by $K(i \text{ and } j)$ the summation of all this information. The fundamental information which has i_0 and not j_0 represents the information that appears in the time series i first and then moves to j. We denote by $K(i \text{ to } j)$ the summation of all this information. In the same way, we can define the information flow rate $K(j \text{ to } i)$ from the time series j to i. By definition, we have the following relation:

$$K(i;j)=K(i \text{ to } j)+K(j \text{ to } i)+K(i \text{ and } j). \tag{3.8}$$

4. Approximate expression for information flow rates

We can approximate the various summations for flow rates by certain calculable quantities.

The information flow rate in a single time series can be approximated by a series of conditional entropy,

$$K(i)=\lim_{n\to\infty} H_{i_{-n},\ldots,i_{-1}}(i_0). \tag{4.1}$$

The random variables have the same meaning as in § 3.

To see that this limit approximates the information flow rate (3.2), let us write the conditional entropies in the form of a summation of fundamental information. This is done by applying the rules (2.4) and (2.6) to the conditional entropies repeatedly with respect to the other random variables in the time series. We see that the conditional entropy $H_{i_{-n},\ldots,i_{-1}}(i_0)$ is the summation of all the fundamental information that has i_0 and not $i_{-n},\ldots,i_{-1}$ as arguments. In the limit of infinite n, this is the precisely the summation for $K(i)$.

The same expression as in the right-hand side of (4.1) appears when we attempt to obtain KS entropy by approximating the time series of a dynamical system by a Markov process. In this procedure, a string of symbols of infinite length is obtained from an orbit of the dynamical system in terms of a partition of the phase space (Alekseev and Yakobson 1981). Then the rule to produce the string is approximated by a sequence of Markov processes. The Shannon entropy of the string is calculated by (4.1), as a limit of Shannon entropies of the approximating Markov processes. The maximum Shannon entropy of the string over various partitions is the KS entropy, and the partition which gives the maximum value is called the generator. So the information flow rate $K(i)$ is equivalent to KS entropy if the partition is a generator.

The information flow rate $K(i,j)$ in two time series is approximated by the following expression:

$$K(i,j)=\lim_{n\to\infty} H_{i_{-n},j_{-n},\ldots,i_{-1},j_{-1}}(i_0,j_0). \tag{4.2}$$

To show that this is the correct approximation, we have to obtain the expression for the conditional entropies in the above limit in the form of a summation of fundamental information. Since they have two arguments, we cannot directly apply the rules (2.4) and (2.6) to them. The following relation is derived in appendix 2:

$$\begin{aligned} H_{i_{-n},j_{-n},\ldots,i_{-1},j_{-1}}(i_0,j_0) &= I_{i_{-n},j_{-n},\ldots,i_{-1},j_{-1}}(i_0;j_0) \\ &\quad + H_{i_{-n},j_{-n},\ldots,i_{-1},j_{-1},j_0}(i_0)+H_{i_{-n},j_{-n},\ldots,i_{-1},j_{-1},i_0}(j_0). \end{aligned} \tag{4.3}$$

With this relation, we see that $H_{i_{-n},j_{-n},\ldots,i_{-1},j_{-1}}(i_0,j_0)$ is the summation of all the fundamental information which has at least i_0 or j_0 and does not have $i_{-n},j_{-n},\ldots,i_{-1},j_{-1}$ as arguments. When we take the above limit, this is precisely the definition of $K(i;j)$.

The approximate expression for the cross information flow rate $K(i;j)$ is obtained through the following relation which is obtained from (3.5)-(3.7):

$$K(i;j)=K(i)+K(j)-K(i,j). \tag{4.4}$$

Substituting the approximation for $K(i)$, $K(j)$ and $K(i,j)$, we have

$$K(i;j)=\lim_{n\to\infty}\{H_{i_{-n},\ldots,i_{-1}}(i_0)+H_{j_{-n},\ldots,j_{-1}}(j_0)-H_{i_{-n},j_{-n},\ldots,i_{-1},j_{-1}}(i_0,j_0)\}. \tag{4.5}$$

By a simple calculation, we have

$$K(i;j)=\lim_{n\to\infty}\{I(i_{-n},\ldots,i_0;j_{-n},\ldots,j_0)-I(i_{-n},\ldots,i_{-1};j_{-n},\ldots,j_{-1})\}. \tag{4.6}$$

Due to the time translational symmetry of the random variables, (4.6) may be rewritten as follows for any integer l:

$$K(i;j)=\lim_{n\to\infty}\{I(i_{-n},\ldots,i_0;j_{l-n},\ldots,j_l)-I(i_{-n},\ldots,i_{-1};j_{l-n},\ldots,j_{l-1})\}. \tag{4.7}$$

We can choose l so that the convergence is rapid. This expression is used in the calculation of information flow rates in § 5.

In the same way, we have the following expressions for the directed information flow rates:

$$K(i \text{ and } j)=\lim_{n\to\infty} I_{i_{-n},j_{-n},\ldots,i_{-1},j_{-1}}(i_0;j_0) \tag{4.8}$$

$$K(i \text{ to } j)+K(i \text{ and } j)=\lim_{n\to\infty} I_{j_{-n},\ldots,j_{-1}}(i_{-n},\ldots,i_0;j_0) \tag{4.9}$$

$$K(j \text{ to } i)+K(i \text{ and } j)=\lim_{n\to\infty} I_{i_{-n},\ldots,i_{-1}}(j_{-n},\ldots,j_0;i_0). \tag{4.10}$$

5. Numerical results on the BZ chain

In this section, we calculate the various information flow rates by computer simulation using the expressions in § 4. The model is a system of one-dimensional maps coupled as a linear chain. The one-dimensional map is obtained from the experimental results of the Belousov-Zhabotinsky chemical reaction.

The BZ map is defined as follows:

$$f_b(x)=\frac{\text{constant}\times\{\tan^{-1}[200(x-0.2)]+\tan^{-1}(40)\}}{1+(2x)^{19}}+b \tag{5.1}$$

where

$$\text{constant}=\frac{0.8[1+(0.7)^{19}]}{\tan^{-1}(30)+\tan^{-1}(40)}$$

is defined so that $f_0(0.35)=0.8$. The time evolution equation for kth element of the chain $x^{(k)}$ is given by

$$x_{n+1}^{(k)}=f_{b_k}(x_n^{(k)})+D(x_n^{(k-1)}-x_n^{(k)}) \tag{5.2}$$

where the coupling constant $D=0.12$. The coupling between two adjacent maps is uneven. The information is made to flow only from the kth element to the $(k+1)$th element of the chain. In the following calculation, we use a chain of length 10 $(1\leqslant k\leqslant 10)$. At the lower end of the chain ((5.2) for $k=1$), we simply put $x_n^{(0)}=0$.

To this chain, we connect an extra one-dimensional map,

$$g_b(x)=\begin{cases}[(x-0.125)^{1/3}+C_1]\,\mathrm{e}^{-x}+b & x<0.3\\ C_2(10x\,\mathrm{e}^{-10x/3})^{19}+b & x\geqslant 0.3\end{cases} \tag{5.3}$$

where $C_1=0.506\,073\,57$ and $C_2=0.121\,205\,692$ are empirically determined constants. This is the original BZ map. Equation (5.1) is a simpler version of this original. At $b=0.023\,288\,5279$, the dynamics of this map is observed to be chaotic in computer simulation. We insert the coupling term $D'x_n$ to the third element of the chain, where

$D' = 0.2$. Thus the information generated in map g is made to flow into the third element of the chain.

In figure 1, a typical time series of the model is shown. Note that chaotic bursts or modulations of periods seem to transmit through the chain.

By calculating the information flow rates, we attempt to describe this transmission of modulations of periods quantitatively. Since the direction of the information flow is obvious from the definition of the model, the cross information flow rate $K(i; j)$ is enough to describe the movement of information in the model.

To calculate the information, we have to choose a partition to translate a time series of real number into that of finite states. For our purposes, we must choose a partition so as to extract information contained in the modulations of periods of each element. To do this, the Lorenz plot of each element is helpful. In figure 2, we depict the Lorentz plot of an element (seventh element) of the chain which is typical of elements 3-10. It consists of islands corresponding to the periodic motion and detailed structure within each island which corresponds to other small fluctuations. It preserves substantially the shape of the original BZ map (5.1). For the original map, it is sufficient to divide the phase space into two at the peak point of the map to calculate the KS entropy. So to extract the information in modulations of periods, we simply have to divide the phase space of each element into two at a point between islands near the peak point of the original map. The dividing points thus chosen are 0.5, 0.5, 0.36, 0.5, 0.34, 0.5, 0.34 and 0.5 for elements 3-10, respectively. The small differences in the positions of the dividing points do not affect the value of the information, as long as they are between the islands. Thus we neglect the information contained in the small fluctuation. The dividing point for the extra map (5.3) is the peak point of the map (0.3).

We calculate the information flow rate (4.1) for each element and the cross information flow rate (4.7) between every two elements. The maximum value of n is 12 for (4.1), and 8 for (4.7). The convergence of the results as n approaches the maximum value is confirmed.

The results are shown in table 1. The information generated in the extra chaotic map is certainly transmitted to all the element of the chain as random modulations

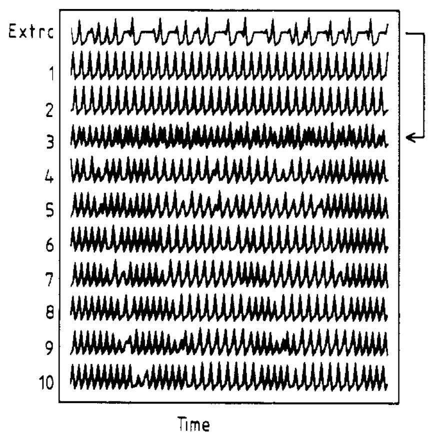

Figure 1. Typical time series of all the elements of the BZ chain. The abscissa is time and the ordinate is phase space. The uppermost time series is that of the extra map, and time series of elements 1-10 are depicted below in order. Note that chaotic bursts and modulations of periods transmit through the chain.

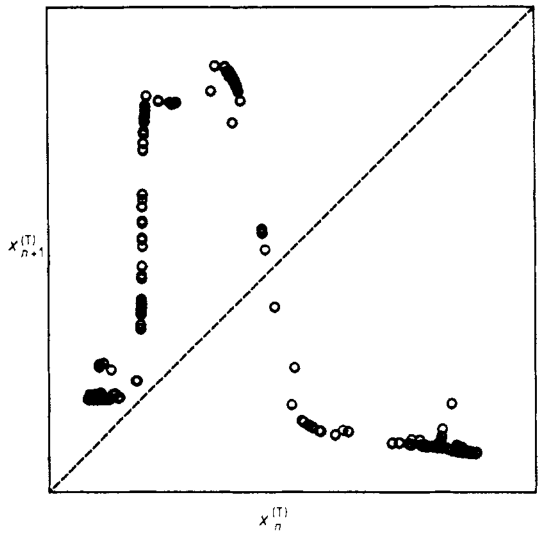

Figure 2. The Lorenz plot for the seventh element of the BZ chain. Reflecting the non-one-dimensional character of the dynamics, the plot is one to many.

Table 1. The information flow rates obtained by computer simulation. The parameters of the BZ chain are the same as those in figure 1. A figure at row M and column N is the information flow rate from element M to N. The information flow rate in an element M is found in row M and column M. A blank means zero by definition. The KS entropy of the extra map is 0.298. Since the dynamics of elements 1 and 2 are periodic, these two elements are omitted from this table. The rates are calculated using (4.7) where $n = 8$ and l is chosen so as to give the maximum value of $I(i_{-4}, \ldots, i_0; j_{l-4}, \ldots, j_l)$. This gives the most rapid convergence of (4.7). Each figure is the average of the results of three different runs with different initial conditions. The largest deviation in these three runs is ≈ 0.002. The coupled maps are iterated 300 000 times in each run.

	3	4	5	6	7	8	9	10
Extra	0.119	0.095	0.090	0.055	0.048	0.039	0.040	0.033
3	0.195	0.165	0.143	0.106	0.084	0.073	0.069	0.062
4		0.218	0.177	0.134	0.105	0.093	0.089	0.079
5			0.243	0.176	0.144	0.125	0.116	0.102
6				0.195	0.149	0.131	0.117	0.105
7					0.204	0.164	0.153	0.130
8						0.178	0.158	0.138
9							0.199	0.163
10								0.173

of periods. But this information alone cannot account for all the information flowing in each element.

6. Summary and discussion

We have defined the information flow rates of time series based on the concept of 'flow'. In the course of analysis of the movement of information, the decomposition of information is useful. In terms of this decomposition, the rate is an appropriate summation of terms in the decomposition. The information flow rate in a time series defined here is shown to be equivalent to the KS entropy. Therefore we can consider the other various information flow rates defined here as natural extensions of KS entropy.

With these types of rate quantities we can argue how much information which is observed to flow in an element is due to another element of the system. This is actually done for coupled BZ maps where the transmission of information is suspected from the previous calculation of mutual information (Matsumoto and Tsuda 1987). (Use of mutual information in a similar context is also found in Martien *et al* (1985), and Herzel and Ebeling (1985).)

Calculation of information flow rates between every pair of elements reveals quantitatively the movements of information in the BZ chain. Although the calculation is limited to the information contained in the modulations of periods, the information generated in the external map is observed to flow throughout the system.

Appendix 1

In this appendix, we show inductively that the conditional mutual information is invariant with respect to any permutations in arguments or in subscripts.

First, the conditional entropy $H_{j_1,\ldots,j_n}(i)$ is invariant to any permutations in its subscripts by definition (2.5).

Second, the conditional mutual information with two arguments is invariant to any permutations in its subscripts and in arguments by definition (2.4).

Assuming the conditional mutual information with n and $n-1$ arguments to be invariant to any permutations in its subscripts or in its arguments, we show that this is valid for information with $n+1$ arguments. From the definition (2.6) of the conditional mutual information with $n+1$ arguments, we have

$$I_{j_1,\ldots,j_m}(i_1;\ldots;i_n;i_{n+1}) = I_{j_1,\ldots,j_m}(i_1;\ldots;i_n) - I_{j_1,\ldots,j_m,i_{n+1}}(i_1;\ldots;i_n). \tag{A1.1}$$

From the assumption, we see that the term on the left-hand side is invariant with respect to any permutations in the subscripts or in the arguments $i_1,\ldots,i_n$. For our purpose, it is enough to show that i_{n+1} can be exchanged with i_1. Using the definition (2.6) twice, we have

$$\begin{aligned} I_{j_1,\ldots,j_m}(i_1;\ldots;i_{n+1}) &= I_{j_1,\ldots,j_m}(i_1;\ldots;i_n) - I_{j_1,\ldots,j_m,i_{n+1}}(i_1;\ldots;i_n) \\ &= I_{j_1,\ldots,j_m}(i_2;\ldots;i_n) - I_{j_1,\ldots,j_m,i_1}(i_2;\ldots;i_n) \\ &\quad - I_{j_1,\ldots,j_m,i_{n+1}}(i_2;\ldots;i_n) + I_{j_1,\ldots,j_m,i_{n+1},i_1}(i_2;\ldots;i_n). \end{aligned} \tag{A1.2}$$

(The last two lines are equal to

$$H_{j_1,\ldots,j_m}(i_2) - H_{j_1,\ldots,j_m,i_1}(i_2) - H_{j_1,\ldots,j_m,i_{n+1}}(i_2) + H_{j_1,\ldots,j_m,i_{n+1},i_1}(i_2)$$

for $n=2$.) Since the last expression is symmetric with respect to i_1 and i_{n+1}, we can exchange them in the original expression. So the conditional mutual information with $n+1$ arguments is invariant with respect to any permutations in its subscripts or in its arguments. This completes the inductive steps.

Appendix 2

In this appendix we derive the following relation:

$$H_k(i_0, j_0) = I_k(i_0; j_0) + H_{k,j_0}(i_0) + H_{k,i_0}(j_0). \tag{A2.1}$$

The conditional entropy can be written as a difference of two Shannon entropies,

$$\begin{aligned} H_k(i_0, j_0) &= H(k, i_0, j_0) - H(k) \\ &= H(k, i_0, j_0) - H(k, i_0) + H(k, i_0) - H(k) \\ &= H_{k,i_0}(j_0) + H_k(i_0). \end{aligned} \tag{A2.2}$$

Applying the decomposition rule (2.4) to the second term of the last line, we have

$$H_k(i_0) = I_k(i_0; j_0) + H_{k,j_0}(i_0). \tag{A2.3}$$

From (A2.2) and (A2.3), we have (A2.1). The above demonstration is correct if we replace k by a set of random variables.

References

Alekseev V M and Yakobson M V 1981 *Phys. Rep.* **75** 287
Herzel H P and Ebeling W 1985 *Phys. Lett.* **111A** 1
Kolmogorov A N 1958 *Dokl. Akad. Nauk* **119** 861
Martien P, Pope S C, Scott P L and Shaw R S 1985 *Phys. Lett.* **110A** 399
Matsumoto K and Tsuda I 1987 *Physica* **26D** 347
Shannon C E and Weaver W 1949 *The Mathematical Theory of Communication* (Urbana, IL: University of Illinois Press)
Shaw R 1981 *Z. Naturf.* **36a** 80