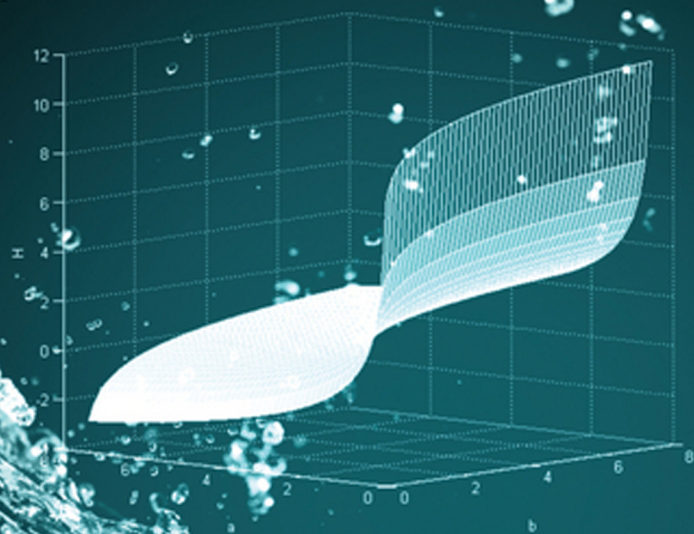# ENTROPY THEORY and its APPLICATION in ENVIRONMENTAL and WATER ENGINEERING

Vijay P. Singh

Entropy Theory and its Application in Environmental and Water Engineering

Entropy Theory and its Application in Environmental and Water Engineering

Vijay P. Singh

Department of Biological and Agricultural Engineering &
Department of Civil and Environmental Engineering
Texas A & M University
Texas, USA

WILEY-BLACKWELL

A John Wiley & Sons, Ltd., Publication

Dedicated to
My wife Anita,
son Vinay,
daughter-in-law Sonali
daughter Arti, and
grandson Ronin

Contents

Preface

Since the pioneering work of Shannon in 1948 on the development of informational entropy theory and the landmark contributions of Kullback and Leibler in 1951 leading to the development of the principle of minimum cross-entropy, of Lindley in 1956 leading to the development of mutual information, and of Jaynes in 1957–8 leading to the development of the principle of maximum entropy and theorem of concentration, the entropy theory has been widely applied to a wide spectrum of areas, including biology, genetics, chemistry, physics and quantum mechanics, statistical mechanics, thermodynamics, electronics and communication engineering, image processing, photogrammetry, map construction, management sciences, operations research, pattern recognition and identification, topology, economics, psychology, social sciences, ecology, data acquisition and storage and retrieval, fluid mechanics, turbulence modeling, geology and geomorphology, geophysics, geography, geotechnical engineering, hydraulics, hydrology, reliability analysis, reservoir engineering, transportation engineering, and so on. New areas finding application of entropy have since continued to unfold. The entropy theory is indeed versatile and its application is widespread.

In the area of hydrologic and environmental sciences and water engineering, a range of applications of entropy have been reported during the past four and half decades, and new topics applying entropy are emerging each year. There are many books on entropy written in the fields of statistics, communication engineering, economics, biology and reliability analysis. These books have been written with different objectives in mind and for addressing different kinds of problems. Application of entropy concepts and techniques discussed in these books to hydrologic science and water engineering problems is not always straightforward. Therefore, there exists a need for a book that deals with basic concepts of entropy theory from a hydrologic and water engineering perspective and then for a book that deals with applications of these concepts to a range of water engineering problems. Currently there is no book devoted to covering basic aspects of the entropy theory and its application in hydrologic and environmental sciences and water engineering. This book attempts to fill this need.

Much of the material in the book is derived from lecture notes prepared for a course on entropy theory and its application in water engineering taught to graduate students in biological and agricultural engineering, civil and environmental engineering, and hydrologic science and water management at Texas, A & M University, College Station, Texas. Comments, critics and discussions offered by the students have, to some extent, influenced the style of presentation in the book.

The book is divided into 16 chapters. The first chapter introduces the concept of entropy. Providing a short discussion of systems and their characteristics, the chapter goes on to

discuss different types of entropies; and connection between information, uncertainty and entropy; and concludes with a brief treatment of entropy-related concepts. Chapter 2 presents the entropy theory, including formulation of entropy and connotations of information and entropy. It then describes discrete entropy for univariate, bivariate and multidimensional cases. The discussion is extended to continuous entropy for univariate, bivariate and multivariate cases. It also includes a treatment of different aspects that influence entropy. Reflecting on the various interpretations of entropy, the chapter provides hints of different types of applications.

The principle of maximum entropy (POME) is the subject matter of Chapter 3, including the formulation of POME and the development of the POME formalism for discrete variables, continuous variables, and two variables. The chapter concludes with a discussion of the effect of constraints on entropy and invariance of entropy. The derivation of POME-based discrete and continuous probability distributions under different constraints constitutes the discussion in Chapter 4. The discussion is extended to multivariate distributions in Chapter 5. First, the discussion is restricted to normal and exponential distributions and then extended to multivariate distributions by combining the entropy theory with the copula method.

Chapter 6 deals with the principle of minimum cross-entropy (POMCE). Beginning with the formulation of POMCE, it discusses properties and formalism of POMCE for discrete and continuous variables and relation to POME, mutual information and variational distance. The discussion on POMCE is extended to deriving discrete and continuous probability distributions under different constraints and priors in Chapter 7. Chapter 8 presents entropy-based methods for parameter estimation, including the ordinary entropy-based method, the parameter-space expansion method, and a numerical method.

Spatial entropy is the subject matter of Chapter 9. Beginning with a discussion of the organization of spatial data and spatial entropy statistics, it goes on to discussing one-dimensional and two-dimensional aggregation, entropy maximizing for modeling spatial phenomena, cluster analysis, spatial visualization and mapping, scale and entropy and spatial probability distributions. Inverse spatial entropy is dealt with in Chapter 10. It includes the principle of entropy decomposition, measures of information gain, aggregate properties, spatial interpretations, hierarchical decomposition, and comparative measures of spatial decomposition.

Maximum entropy-based spectral analysis is presented in Chapter 11. It first presents the characteristics of time series, and then discusses spectral analyses using the Burg entropy, configurational entropy, and mutual information principle. Chapter 12 discusses minimum cross-entropy spectral analysis. Presenting the power spectrum probability density function first, it discusses minimum cross-entropy-based power spectrum given autocorrelation, and cross-entropy between input and output of linear filter, and concludes with a general method for minimum cross-entropy spectral estimation.

Chapter 13 presents the evaluation and design of sampling and measurement networks. It first discusses design considerations and information-related approaches, and then goes on to discussing entropy measures and their application, directional information transfer index, total correlation, and maximum information minimum redundancy (MIMR).

Selection of variables and models constitutes the subject matter of Chapter 14. It presents the methods of selection, the Kullback–Leibler (KL) distance, variable selection, transitivity, logit model, and risk and vulnerability assessment. Chapter 15 is on neural networks comprising neural network training, principle of maximum information preservation, redundancy and

diversity, and decision trees and entropy nets. Model complexity is treated in Chapter 16. The complexity measures discussed include Ferdinand's measure of complexity, Kapur's complexity measure, Cornacchio's generalized complexity measure and other complexity measures.

Vijay P. Singh
College Station, Texas

Acknowledgments

My family members – brothers and sisters in India – have been a continuous source of inspiration. My wife Anita, son Vinay, daughter-in-law Sonali, grandson Ronin, and daughter Arti have been most supportive and allowed me to work during nights, weekends, and holidays, often away from them. They provided encouragement, showed patience, and helped in myriad ways. Most importantly, they were always there whenever I needed them, and I am deeply grateful. Without their support and affection, this book would not have come to fruition.

Vijay P. Singh
College Station, Texas

1   **Introduction**

Beginning with a short introduction of systems and system states, this chapter presents concepts of thermodynamic entropy and statistical-mechanical entropy, and definitions of informational entropies, including the Shannon entropy, exponential entropy, Tsallis entropy, and Renyi entropy. Then, it provides a short discussion of entropy-related concepts and potential for their application.

1.1 Systems and their characteristics

1.1.1 Classes of systems

In thermodynamics a system is defined to be any part of the universe that is made up of a large number of particles. The remainder of the universe then is referred to as surroundings. Thermodynamics distinguishes four classes of systems, depending on the constraints imposed on them. The classification of systems is based on the transfer of (i) matter, (ii) heat, and/or (iii) energy across the system boundaries (Denbigh, 1989). The four classes of systems, as shown in Figure 1.1, are: (1) Isolated systems: These systems do not permit exchange of matter or energy across their boundaries. (2) Adiabatically isolated systems: These systems do not permit transfer of heat (also of matter) but permit transfer of energy across the boundaries. (3) Closed systems: These systems do not permit transfer of matter but permit transfer of energy as work or transfer of heat. (4) Open systems: These systems are defined by their geometrical boundaries which permit exchange of energy and heat together with the molecules of some chemical substances.

The second law of thermodynamics states that the entropy of a system can only increase or remain constant; this law applies to only isolated or adiabatically isolated systems. The vast majority of systems belong to class (4). Isolation and closedness are not rampant in nature.

1.1.2 System states

There are two states of a system: microstate and macrostate. A system and its surroundings can be isolated from each other, and for such a system there is no interchange of heat or matter with its surroundings. Such a system eventually reaches a state of equilibrium in a thermodynamic sense, meaning no significant change in the state of the system will occur. The state of the system here refers to the macrostate, not microstate at the atomic scale, because the
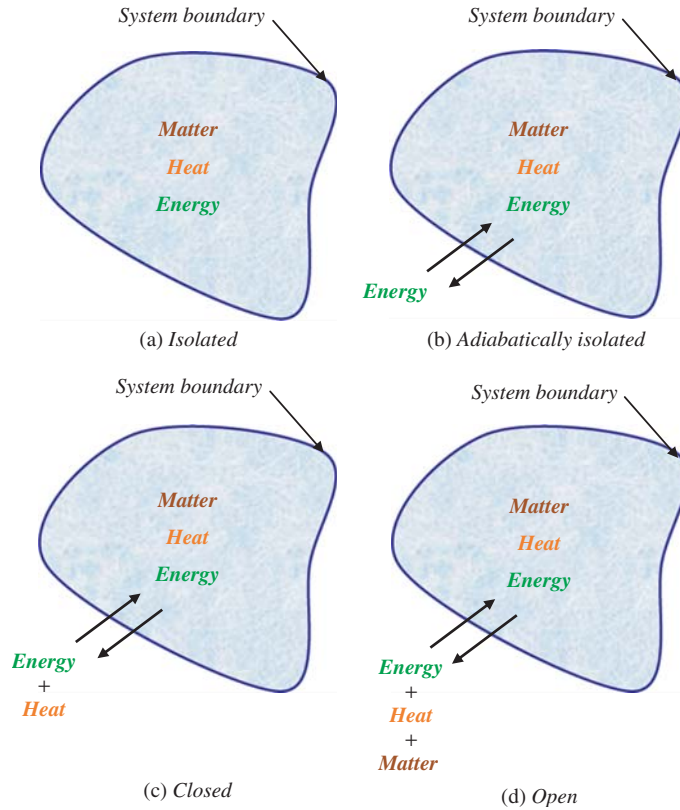
Figure 1.1 Classification of systems.

microstate of such a system will continuously change. The macrostate is a thermodynamic state which can be completely described by observing thermodynamic variables, such as pressure, volume, temperature, and so on. Thus, in classical thermodynamics, a system is described by its macroscopic state entailing experimentally observable properties and the effects of heat and work on the interaction between the system and its surroundings. Thermodynamics does not distinguish between various microstates in which the system can exist, and hence does not deal with the mechanisms operating at the atomic scale (Fast, 1968). For a given thermodynamic state there can be many microstates. Thermodynamic states are distinguished when there are measurable changes in thermodynamic variables.

1.1.3 Change of state

Whenever a system is undergoing a change because of introduction of heat or extraction of heat or any other reason, changes of state of the system can be of two types: reversible and irreversible. As the name suggests, reversible means that any kind of change occurring during a reversible process in the system and its surroundings can be restored by reversing the process. For example, changes in the system state caused by the addition of heat can be restored by the extraction of heat. On the contrary, this is not true in the case of irreversible change of state in which the original state of the system cannot be regained without making changes in the surroundings. Natural processes are irreversible processes. For processes to be reversible, they must occur infinitely slowly.

It may be worthwhile to visit the first law of thermodynamics, also called the law of conservation of energy, which was based on the transformation of work and heat into one another. Consider a system which is not isolated from its surroundings, and let a quantity of heat dQ be introduced to the system. This heat performs work denoted as dW. If the internal energy of the system is denoted by U, then dQ and dW will lead to an increase in U: $dU = dQ + dW$. The work performed may be of mechanical, electrical, chemical, or magnetic nature, and the internal energy is the sum of kinetic energy and potential energy of all particles that the system is made up of. If the system passes from an initial state 1 to a final state 2, then, $\int_1^2 dU = \int_1^2 dQ + \int_1^2 dW$. It should be noted that the integral $\int_1^2 dU$ depends on the initial and final states but the integrals $\int_1^2 dQ$ and $\int_1^2 dW$ also depend on the path followed. Since the system is not isolated and is interactive, there will be exchanges of heat and work with the surroundings. If the system finally returns to its original state, then the sum of integral of heat and integral of work will be zero, meaning the integral of internal energy will also be zero, that is, $\int_1^2 dU + \int_2^1 dU = 0$, or $-\int_1^2 dU = -\int_2^1 dU$. Were it not the case, the energy would either be created or destroyed. The internal energy of a system depends on pressure, temperature, volume, chemical composition, and structure which define the system state and does not depend on the prior history.

1.1.4 Thermodynamic entropy

Let Q denote the quantity of heat. For a system to transition from state 1 to state 2, the amount of heat, $\int_1^2 dQ$, required is not uniquely defined, but depends on the path that is followed for transition from state 1 to state 2, as shown in Figures 1.2a and b. There can be two paths: (i) reversible path: transition from state 1 to state 2 and back to state 1 following the same path, and (ii) irreversible path: transition from state 1 to state 2 and back to state 1 following a different path. The second path leads to what is known in environmental and water engineering as hysteresis. The amount of heat contained in the system under a given condition is not meaningful here. On the other hand, if T is the absolute temperature (degrees kelvin or simply kelvin) (i.e., $T = 273.15 + $ temperature in $°C$), then a closely related quantity, $\int_1^2 dQ/T$, is uniquely defined and is therefore independent of the path the system takes to transition from state 1 to state 2, provided the path is reversible (see Figure 1.2a). Note that when integrating, each elementary amount of heat is divided by the temperature at which it is introduced. The system must expend this heat in order to accomplish the transition and this heat expenditure is referred to as heat loss. When calculated from the zero point of absolute temperature, the integral:

$$S = \int_0^T \frac{dQ_{rev}}{T} \tag{1.1}$$

is called entropy of the system, denoted by S. Subscript of Q, rev, indicates that the path is reversible. Actually, the quantity S in equation (1.1) is the change of entropy $\Delta S (= S - S_0)$
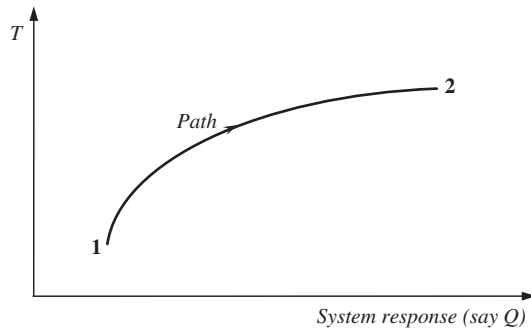
(a) One path
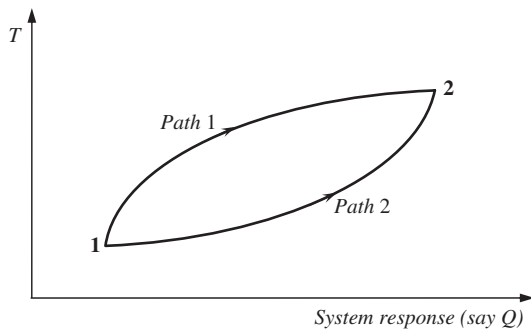
(b) Two paths

Figure 1.2 (a) Single path: transition from state 1 to state 2, and (b) two paths: transition from state 1 to state 2.

occurring in the transition from state 1 (corresponding to zero absolute temperature) to state 2. Equation (1.1) defines what Clausius termed thermodynamic entropy; it defines the second law of thermodynamics as the entropy increase law, and shows that the measurement of entropy of the system depends on the measurement of the quantities of heat, that is, calorimetry.

Equation (1.1) defines the experimental entropy given by Clausius in 1850. In this manner it is expressed as a function of macroscopic variables, such as temperature and pressure, and its numerical value can be measured up to a certain constant which is derived from the third law. Entropy S vanishes at the absolute zero of temperature. In 1865, while studying heat engines, Clausius discovered that although the total energy of an isolated system was conserved, some of the energy was being converted continuously to a form, such as heat, friction, and so on, and that this conversion was irrecoverable and was not available for any useful purpose; this part of the energy can be construed as energy loss, and can be interpreted in terms of entropy. Clausius remarked that the energy of the world was constant and the entropy of the world was increasing. Eddington called entropy the arrow of time.

The second law states that the entropy of a closed system always either increases or remains constant. A system can be as small as the piston, cylinder of a car (if one is trying to design a better car) or as big as the entire sky above an area (if one is attempting to predict weather). A closed system is thermally isolated from the rest of the environment and hence is a special kind of system. As an example of a closed system, consider a perfectly insulated cup of water in which a sugar cube is dissolved. As the sugar cube melts away into water, it would be

logical to say that the water-sugar system has become more disordered, meaning its entropy has increased. The sugar cube will never reform to its original form at the bottom of the cup. However, that does not mean that the entropy of the water-sugar will never decrease. Indeed, if the system is made open and if enough heat is added to boil off the water, the sugar will recrystallize and the entropy will decrease. The entropy of open systems is decreased all the time, as for example, in the case of making ice in the freezer. It also occurs naturally in the case where rain occurs when disordered water vapor transforms to more ordered liquid. The same applies when it snows wherein one witnesses pictures of beautiful order in ice crystals or snowflakes. Indeed, sun shines by converting simple atoms (hydrogen) into more complex ones (helium, carbon, oxygen, etc.).

1.1.5 Evolutive connotation of entropy

Explaining entropy in the macroscopic world, Prigogine (1989) emphasized the evolutive connotation of entropy and laid out three conditions that must be satisfied in the evolutionary world: irreversibility, probability and coherence.

Irreversibility: Past and present cannot be the same in evolution. Irreversibility is related to entropy. For any system with irreversible processes, entropy can be considered as the sum of two components: one dealing with the entropy exchange with the external environment and the other dealing with internal entropy production which is always positive. For an isolated system, the first component is zero, as there is no entropy exchange, and the second term may only increase, reaching a maximum. There are many processes in nature that occur in one direction only, as for example, a house afire goes in the direction of ashes, a man going from the state of being a baby to being an old man, a gas leaking from a tank or air leaking from a car tire, food being eaten and getting transformed into different elements, and so on. Such events are associated with entropy which has a tendency to increase and are irreversible.

Entropy production is related to irreversible processes which are ubiquitous in water and environmental engineering. Following Prigogine (1989), entropy production plays a dual role. It does not necessarily lead to disorder, but may often be a mechanism for producing order. In the case of thermal diffusion, for example, entropy production is associated with heat flow which yields disorder, but it is also associated with anti-diffusion which leads to order. The law of increase of entropy and production of a structure are not necessarily opposed to each other. Irreversibility leads to a structure as is seen in a case of the development of a town or crop growth.

Probability: Away from equilibrium, systems are nonlinear and hence have multiple solutions to equations describing their evolution. The transition from instability to probability also leads to irreversibility. Entropy states that the world is characterized by unstable dynamical systems. According to Prigogine (1989), the study of entropy must occur on three levels: The first is the phenomenological level in thermodynamics where irreversible processes have a constructive role. The second is embedding of irreversibility in classical dynamics in which instability incorporates irreversibility. The third level is quantum theory and general relativity and their modification to include the second law of thermodynamics.

Coherence: There exists some mechanism of coherence that would permit an account of evolutionary universe wherein new, organized phenomena occur.

1.1.6 Statistical mechanical entropy

Statistical mechanics deals with the behavior of a system at the atomic scale and is therefore concerned with microstates of the system. Because the number of particles in the system is so huge, it is impractical to deal with the microstate of each particle, statistical methods are

therefore resorted to; in other words, it is more important to characterize the distribution function of the microstates. There can be many microstates at the atomic scale which may be indistinguishable at the level of a thermodynamic state. In other words, there can be many possibilities of the realization of a thermodynamic state. If the number of these microstates is denoted by N, then statistical entropy is defined as

$$S = k \ln N \tag{1.2}$$

where k is Boltzmann constant (1.3806×10^{-16} erg/K or 1.3806×10^{-23} J/K (kg-m^2/s^2-K)), that is, the gas constant per molecule

$$k = \frac{R}{N_0} \tag{1.3}$$

where R is gas constant per mole (1.9872 cal/K), and N_0 is Avogadro's number (6.0221×10^{23} per mole). Equation (1.2) is also called Boltzmann entropy, and assumes that all microstates have the same probability of occurrence. In other words, in statistical mechanics the Boltzmann entropy is for the canonical ensemble. Clearly, S increases as N increases and its maximum represents the most probable state, that is, maximum number of possibilities of realization. Thus, this can be considered as a direct measure of the probability of the thermodynamic state. Entropy defined by equation (1.2) exhibits all the properties attributed to the thermodynamic entropy defined by equation (1.1).

Equation (1.2) can be generalized by considering an ensemble of systems. The systems will be in different microstates. If the number of systems in the i-th microstate is denoted by n_i then the statistical entropy of the i-th microstate is $S_i = k \log n_i$. For the ensemble the entropy is expressed as a weighted sum:

$$S = k \sum_{i=1}^{N} n_i \log n_i \tag{1.4a}$$

where N is the total number of microstates in which all systems are organized. Dividing by N, and expressing the fraction of systems by $p_i = n_i/N$, the result is the statistical entropy of the ensemble expressed as

$$S = -k \sum_{i=1}^{N} p_i \ln p_i \tag{1.4b}$$

where k is again Boltzmann's constant. The measurement of S here depends on counting the number of microstates. Equation (1.2) can be obtained from equation (1.4b), assuming the ensemble of systems is distributed over N states. Then $p_i = 1/N$, and equation (1.4b) becomes

$$S = -kN \frac{1}{N} \ln \frac{1}{N} = k \ln N \tag{1.5}$$

which is equation (1.2).

Entropy of a system is an extensive thermodynamic property, such as mass, energy, volume, momentum, charge, or number of atoms of chemical species, but unlike these quantities, entropy does not obey the conservation law. Extensive thermodynamic quantities are those that are halved when a system in equilibrium containing these quantities is

partitioned into two equal parts, but intensive quantities remain unchanged. Examples of extensive variables include volume, mass, number of molecules, and entropy; and examples of intensive variables include temperature and pressure. The total entropy of a system equals the sum of entropies of individual parts. The most probable distribution of energy in a system is the one that corresponds to the maximum entropy of the system. This occurs under the condition of dynamic equilibrium. During evolution toward a stationary state, the rate of entropy production per unit mass should be minimum, compatible with external constraints. In thermodynamics entropy has been employed as a measure of the degree of disorderliness of the state of a system.

The entropy of a closed and isolated system always tends to increase to its maximum value. In a hydraulic system, if there were no energy loss the system would be orderly and organized. It is the energy loss and its causes that make the system disorderly and chaotic. Thus, entropy can be interpreted as a measure of the amount of chaos or disorder within a system. In hydraulics, a portion of flow energy (or mechanical energy) is expended by the hydraulic system to overcome friction, which then is dissipated to the external environment. The energy so converted is frequently referred to as energy loss. The conversion is only in one direction, that is, from available energy to nonavailable energy or energy loss. A measure of the amount of irrecoverable flow energy is entropy which is not conserved and which always increases, that is, the entropy change is irreversible. Entropy increase implies increase of disorder. Thus, the process equation in hydraulics expressing the energy (or head) loss can be argued to originate in the entropy concept.

1.2 Informational entropies

Before describing different types of entropies, let us further develop an intuitive feel about entropy. Since disorder, chaos, uncertainty, or surprise can be considered as different shades of information, entropy comes in handy as a measure thereof. Consider a random experiment with outcomes $x_1, x_2, \ldots, x_N$ with probabilities $p_1, p_2, \ldots, p_N$, respectively; one can say that these outcomes are the values that a discrete random variable X takes on. Each value of X, x_i, represents an event with a corresponding probability of occurrence, p_i. The probability p_i of event x_i can be interpreted as a measure of uncertainty about the occurrence of event x_i. One can also state that the occurrence of an event x_i provides a measure of information about the likelihood of that probability p_i being correct (Batty, 2010). If p_i is very low, say, 0.01, then it is reasonable to be certain that event x_i will not occur and if x_i actually occurred then there would be a great deal of surprise as to the occurrence of x_i with $p_i = 0.01$, because our anticipation of it was highly uncertain. On the other hand, if p_i is very high, say, 0.99, then it is reasonable to be certain that event x_i will occur and if x_i did actually occur then there would hardly be any surprise about the occurrence of x_i with $p_i = 0.99$, because our anticipation of it was quite certain.

Uncertainty about the occurrence of an event suggests that the random variable may take on different values. Information is gained by observing it only if there is uncertainty about the event. If an event occurs with a high probability, it conveys less information and vice versa. On the other hand, more information will be needed to characterize less probable or more uncertain events or reduce uncertainty about the occurrence of such an event. In a similar vein, if an event is more certain to occur, its occurrence or observation conveys less information and less information will be needed to characterize it. This suggests that the more uncertain an event the more information its occurrence transmits or the more information

needed to characterize it. This means that there is a connection between entropy, information, uncertainty, and surprise.

It seems intuitive that one can scale uncertainty or its complement certainty or information, depending on the probability of occurrence. If $p(x_i) = 0.5$, the uncertainty about the occurrence would be maximum. It should be noted that the assignment of a measure of uncertainty should be based not on the occurrence of a single event of the experiment but of any event from the collection of mutually exclusive events whose union equals the experiment or collection of all outcomes. The measure of uncertainty about the collection of events is called entropy. Thus, entropy can be interpreted as a measure of uncertainty about the event prior to the experimentation. Once the experiment is conducted and the results about the events are known, the uncertainty is removed. This means that the experiment yields information about events equal to the entropy of the collection of events, implying uncertainty equaling information.

Now the question arises: What can be said about the information when two independent events x and y occur with probability p_x and p_y? The probability of the joint occurrence of x and y is $p_x p_y$. It would seem logical that the information to be gained from their joint occurrence would be the inverse of the probability of their occurrence, that is, $1/(p_x p_y)$. This shows that this information does not equal the sum of information gained from the occurrence of event x, $1/p_x$, and the information gained from the occurrence of event y, $1/p_y$, that is,

$$\frac{1}{p_x p_y} \neq \frac{1}{p_x} + \frac{1}{p_y} \tag{1.6}$$

This inequality can be mathematically expressed as a function $g(.)$ as

$$g\left(\frac{1}{p_x p_y}\right) = g\left(\frac{1}{p_x}\right) + g\left(\frac{1}{p_y}\right) \tag{1.7}$$

Taking g as a logarithmic function which seems to be the only solution, then one can express

$$-\log\left(\frac{1}{p_x p_y}\right) = -\log\left(\frac{1}{p_x}\right) - \log\left(\frac{1}{p_y}\right) \tag{1.8}$$

Thus, one can summarize that the information gained from the occurrence of any event with probability p is $\log(1/p) = -\log p$. Tribus (1969) regarded $-\log p$ as a measure of uncertainty of the event occurring with probability p or a measure of surprise about the occurrence of that event. This concept can be extended to a series of N events occurring with probabilities $p_1, p_2, \ldots, p_N$, which then leads to the Shannon entropy to be described in what follows.

1.2.1 Types of entropies

There are several types of informational entropies (Kapur, 1989), such as Shannon entropy (Shannon, 1948), Tsallis entropy (Tsallis, 1988), exponential entropy (Pal and Pal, 1991a, b), epsilon entropy (Rosenthal and Binia, 1988), algorithmic entropy (Zurek, 1989), Hartley entropy (Hartley, 1928), Renyi's entropy (1961), Kapur entropy (Kapur, 1989), and so on. Of these the most important are the Shannon entropy, the Tsallis entropy, the Renyi entropy, and the exponential entropy. These four types of entropies are briefly introduced in this chapter and the first will be detailed in the remainder of the book.