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 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

Entropy Theory and its Application in Environmental and Water Engineering, First Edition. Vijay P. Singh.

© 2013 John Wiley & Sons, Ltd. Published 2013 by John Wiley & Sons, Ltd.

2 Entropy Theory and its Application in Environmental and Water Engineering



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.

CHAPTER 1 Introduction 3

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 $\int_{1}^{1} dQ$ and $\int_{1}^{1} dW$ 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, $\int_{1}^{2} dU + \int_{1}^{1} dU = 0$, or $-\int_{1}^{2} dU = -\int_{1}^{1} dU$. 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 take the system takes the takes the state 1 to state 2 and the path the system takes to take the system takes the takes the state 1 to state 2 and the path the system takes to take the system takes the takes the system independent of the path the system takes to take takes the state 1 to state 2 and the system takes the takes the state 1 to state 2 and the system takes the system takes the takes the state 1 to state 2 and the system takes takes the system takes takes the system takes t

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:

$$S = \int_{0}^{T} \frac{dQ_{rev}}{T}$$
(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 $\Delta S (= S - S_0)$



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

6 Entropy Theory and its Application in Environmental and Water Engineering

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-m²/s²-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 × 10²³ 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

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

1.2.2 Shannon entropy

In 1948, Shannon introduced what is now referred to as information-theoretic or simply informational entropy. It is now more frequently referred to as Shannon entropy. Realizing that when information was specified, uncertainty was reduced or removed, he sought a measure of uncertainty. For a probability distribution $P = \{p_1, p_2, ..., p_N\}$, where $p_1, p_2, ..., p_N$ are probabilities of *N* outcomes $(x_i, i = 1, 2, ..., N)$ of a random variable *X* or a random experiment, that is, each value corresponds to an event, one can write

$$-\log\left(\frac{1}{p_1p_2\dots p_N}\right) = -\log\left(\frac{1}{p_1}\right) - \log\left(\frac{1}{p_2}\right) - \dots - \log\left(\frac{1}{p_N}\right)$$
(1.9)

Equation (1.9) states the information gained by observing the joint occurrence of N events. One can write the average information as the expected value (or weighted average) of this series as

$$H = -\sum_{i=1}^{N} p_i \log p_i$$
 (1.10)

where *H* is termed as entropy, defined by Shannon (1948).

The informational entropy of Shannon (1948) given by equation (1.10) has a form similar to that of the thermodynamic entropy given by equation (1.4b) whose development can be attributed to Boltzmann and Gibbs. Some investigators therefore designate H as Shannon-Boltzmann-Gibbs entropy (see Papalexiou and Koutsyiannis, 2012). In this text, we will call it the Shannon entropy. Equation (1.4b) or (1.10) defining entropy, H, can be re-written as

$$H(X) = H(P) = -K \sum_{i=1}^{N} p(x_i) \log[p(x_i)], \ \sum_{i=1}^{N} p(x_i) = 1$$
(1.11)

where H(X) is the entropy of random variable $X: \{x_1, x_2, ..., x_N\}$, $P: \{p_1, p_2, ..., p_N\}$ is the probability distribution of X, N is the sample size, and K is a parameter whose value depends on the base of the logarithm used. If different units of entropy are used, then the base of the logarithm changes. For example, one uses bits for base 2, Napier or nat or nit for base e, and decibels or logit or docit for base 10.

In general, K can be taken as unity, and equation (1.11), therefore, becomes

$$H(X) = H(P) = -\sum_{i=1}^{N} p(x_i) \log[p(x_i)]$$
(1.12)

H(X), given by equation (1.12), represents the information content of random variable *X* or its probability distribution P(x). It is a measure of the amount of uncertainty or indirectly the average amount of information content of a single value of *X*. Equation (1.12) satisfies a number of desiderata, such as continuity, symmetry, additivity, expansibility, recursivity, and others (Shannon and Weaver, 1949), and has the same form of expression as the thermodynamic entropy and hence the designation of *H* as entropy.

Equation (1.12) states that *H* is a measure of uncertainty of an experimental outcome or a measure of the information obtained in the experiment which reduces uncertainty. It also states the expected value of the amount of information transmitted by a source with

10 Entropy Theory and its Application in Environmental and Water Engineering

probability distribution $(p_1, p_2, ..., p_N)$. The Shannon entropy may be viewed as the indecision of an observer who guesses the nature of one outcome, or as the disorder of a system in which different arrangements can be found. This measure considers only the possibility of occurrence of an event, not its meaning or value. This is the main limitation of the entropy concept (Marchand, 1972). Thus, *H* is sometimes referred to as the information index or the information content.

If *X* is a deterministic variable, then the probability that it will take on a certain value is one, and the probabilities of all other alternative values are zero. Then, equation (1.12) shows that H(x) = 0 which can be viewed as the lower limit of the values the entropy function may assume. This corresponds to the absolute certainty, that is, there is no uncertainty and the system is completely ordered. On the other hand, when all x_i s are equally likely, that is, the variable is uniformly distributed ($p_i = 1/N$, i = 1, 2, ..., N), that is, if all probabilities are equal, $p_i = p$, i = 1, 2, ..., N, then equation (1.12) yields

$$H(X) = H_{\max}(X) = \log N \tag{1.13}$$

This shows that the entropy function attains a maximum, and equation (1.13) thus defines the upper limit or would lead to the maximum entropy. This also reveals that the outcome has the maximum uncertainty. Equation (1.10) and in turn equation (1.13) show that the larger the number of events the larger the entropy measure. This is intuitively appealing because more information is gained from the occurrence of more events, unless, of course, events have zero probability of occurrence. The maximum entropy occurs when the uncertainty is maximum or the disorder is maximum.

One can now state that entropy of any variable always assumes positive values within the limits defined as:

$$0 \le H(x) \le \log N \tag{1.14}$$

It is logical to say that many probability distributions lie between these two extremes and their entropies between these two limits. As an example, consider a random variable *X* which takes on a value of 1 with a probability *p* and 0 with a probability q = 1 - p. Taking different values of *p*, one can plot H(p) as a function of *p*. It is seen that for p = 1/2, H(p) = 1 bit is the maximum.

When entropy is minimum, $H_{\min} = 0$, the system is completely ordered and there is no uncertainty about its structure. This extreme case would correspond to the situation where $p_i = 1$, $p_j = 0$, $\forall j \neq i$. On the other hand, the maximum entropy H_{\max} can be considered as a measure of maximum uncertainty and the disorder would be maximum which would occur if all events occur with the same probability, that is, there are no constraints on the system. This suggests that there is order-disorder continuum with respect to H; that is, more constraints on the form of the distribution lead to reduced entropy. The statistically most probable state corresponds to the maximum entropy. One can extend this interpretation further.

If there are two probability distributions with equiprobable outcomes, one given as above (i.e., $p_i = p, i = 1, 2, ..., N$), and the other as $q_i = q, i = 1, 2, ..., M$, then one can determine the difference in the information contents of the two distributions as $\Delta H = H_p - H_q = \log_2 p - \log_2 q = \log_2(p/q)$ bits, where H_p is the information content or entropy of $\{p_i, i = 1, 2, ..., N\}$ and H_q is the information content or entropy of $\{q_i, i = 1, 2, ..., N\}$ and H_q is the information content or entropy of $\{q_i, i = 1, 2, ..., M\}$. One can observe that if q > p or (M < N), $\Delta H > 0$. In this case the entropy increases or information is lost because of the increase in the number of possible outcomes or outcome uncertainty. On the other hand, if q < p or (M > N), then $\Delta H < 0$. This case corresponds to

the gain in information because of the decrease in the number of possible outcomes or in uncertainty.

Comparing with H_{max} , a measure of information can be constructed as

$$I = H_{\max} - H = \log n + \sum_{i=1}^{n} p_i \, \log p_i$$

= $\sum_{i=1}^{n} p_i \, \log\left(\frac{p_i}{1/n}\right) = \sum_{i=1}^{n} p_i \, \log\left(\frac{p_i}{q_i}\right)$ (1.15)

where $q_i = 1/n$. In equation (1.15), $\{q_i\}$ can be considered as a prior distribution and $\{p_i\}$ as a posterior distribution. Normalization of *I* by H_{max} leads to

$$R = \frac{I}{H_{\text{max}}} = 1 - \frac{H}{H_{\text{max}}}$$
(1.16)

where *R* is called the relative redundancy varying between 0 and 1.

In equation (1.12), the logarithm is to the base of 2, because it is more convenient to use logarithms to the base of 2, rather than logarithms to the base e or 10. Therefore, the entropy is measured in bits (short for binary digits). A bit can be physically interpreted in terms of the fraction of alternatives that are reduced by knowledge of some kind. These alternatives are equally likely. Thus, the amount of information depends on the fraction, not the absolute number of alternatives. This means that each time the number of alternatives is reduced to half based on some knowledge or one message, there will be a gain of one bit of information or the message has one bit of information. Consider there are four alternatives and this number is reduced to two, then one bit of information is transmitted. In the case of two alternative messages the amount of information $= \log_2 2 = 1$. This unit of information is called bit (as in binary system). The same amount of information is transmitted if 100 alternatives are reduced to 50, that is, $\log_2(100/50) = \log_2 2 = 1$. In general, one can express that $\log_2 x$ is bits of information transmitted or the message has if N alternatives are reduced to N/x. If 1000 alternatives are reduced to 500 (one bit of information is transmitted) and then 500 alternatives to 250 (another bit of information is transmitted), then x = 4, and $\log_2 4 = 2$ bits. Further, if one message reduces the number of alternatives N to N/x and another message reduces N to N/2x then the former message has one bit less information than the latter. On the other hand, if one has eight alternative messages to choose from, then $\log_2 8 = \log_2 2^3 = 3$ bits, that is, this case is associated with three bits of information or this defines the amount of information that can be determined from the number of alternatives to choose from. If one has 128 alternatives the amount of information is $\log_2(2)^7 = 7$ bits.

The measurement of entropy is in nits (nats) in the case of natural logarithm (to the base *e*) and in logits (or decibles) with common logarithm. It may be noted that if $n^x = y$, then $x \log n = \log y$, meaning *x* is the logarithm of *y* to the base *n*, that is, $x \log_n n = \log_n y$. To be specific, the amount of information is measured by the logarithm of the number of choices. One can go from base *b* to base *a* as: $\log_h N = \log_h a \times \log_a N$.

From the above discussion it is clear that the value of *H* being one or unity depends on the base of the logarithm: bit (binary digit) for \log_2 and dit (decimal digit) for \log_{10} . Then one dit expresses the uncertainty of an experiment having ten equiprobable outcomes. Likewise, one bit corresponds to the uncertainty of an experiment having two equiprobable outcomes. If p = 1, then the entropy is zero, because the occurrence of the event is certain and there is

no uncertainty as to the outcome of the experiment. The same applies when p = 0 and the entropy is zero.

In communication, each representation of random variable X can be regarded as a message. If X is a continuous variable (say, amplitude), then it would carry an infinite amount of information. In practice X is uniformly quantized into a finite number of discrete levels, and then X may be regarded as a discrete variable:

$$X = \{x_i, i = 0, \pm 1, \dots, \pm N\}$$
(1.17)

where x_i is a discrete number, and (2N + 1) is the total number of discrete levels. Then, random variable *X*, taking on discrete values, produces a finite amount of information.

1.2.3 Information gain function

From the above discussion it would intuitively seem that the gain in information from an event is inversely proportional to its probability of occurrence. Let this gain be represented by G(p) or ΔI . Following Shannon (1948),

$$G(p) = \Delta I = \log\left(\frac{1}{p_i}\right) = -\log(p_i) \tag{1.18}$$

where G(p) is the gain function. Equation (1.18) is a measure of that gain in information or can be called as gain function (Pal and Pal, 1991a). Put another way, the uncertainty removed by the message that the event *i* occurred or the information transmitted by it is measured by equation (1.18). The use of logarithm is convenient, since the combination of the probabilities of independent events is a multiplicative relation. Thus, logarithms allow for expressing the combination of their entropies as a simple additive relation. For example, if $P(A \cap B) = P_A P_B$, then $H(AB) = -\log P_A - \log P_B = H(A) + H(B)$. If the probability of an event is very small, say $p_i = 0.01$, then the partial information transmitted by this event is very large $\Delta I = 2$ dits if the base of the logarithm is taken as 10; such an outcome will not occur in the long run. If there are *N* events, one can compute the total gain in information as

$$I = \sum_{i=1}^{N} \Delta I_i = -\sum_{i=1}^{N} \log(p_i)$$
(1.19)

Each event occurs with a different probability.

The entropy or global information of an event *i* is expressed as a weighted value:

$$H(p_i) = -p_i \log p_i \tag{1.20}$$

Since $0 \le p_i \le 1$, *H* is always positive. Therefore, the average or expected gain in information can be obtained by taking the weighted average of individual gains of information:

$$H = E(\Delta I) = -\sum_{i=1}^{N} p_i(\Delta I_i) = -\sum_{i=1}^{N} p_i \log p_i$$
(1.21)

which is the same as equation (1.10) or (1.12). What is interesting to note here is that one can define different types of entropy by simply defining the gain function or uncertainty differently. Three other types of entropies are defined in this chapter.

Equation (1.21) can be viewed in another way. Probabilities of outcomes of an experiment correspond to the partitioning of space among outcomes. Because the intersection of outcomes

is empty, the global entropy of the experiment is the sum of elementary entropies of the *N* outcomes:

$$H = H_1 + H_2 + \dots + H_N = \sum_{i=1}^N H_i$$
(1.22a)

$$= -p_1 \log p_1 - p_2 \log p_2 - \dots - p_N \log p_N = -\sum_{i=1}^N p_i \log p_i$$
(1.22b)

which is the same as equation (1.21). Clearly, *H* is maximum when all outcomes are equiprobable, that is, $p_i = 1/N$. This has an important application in hydrology, geography, meteorology, and socio-economic and political sciences. If a topology of data measured on nominal scales has classes possessing the same number of observations then it will transmit the maximum amount of information (entropy). This condition is not entirely true if by computing distances between elements one can minimize intra-class variance and maximize inter-class variance. This would lead to distributions with a smaller entropy but a higher variance value (Marchand, 1972).

Example 1.1: Plot the gain function defined by equation (1.18) for different values of probability: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. Take the base of logarithm as 2 as well as *e*. What do you conclude from this plot?

Solution: The gain function is plotted in Figure 1.3. It is seen that the gain function decreases as the probability of occurrence increases. Indeed the gain function becomes zero when the probability of occurrence is one. For lower logarithmic base, the gain function is higher, that is, the gain function with logarithmic base of 2 is higher than that with logarithmic base *e*.



Figure 1.3 Plot of Shannon's gain function.

Example 1.2: Consider a two-state variable taking on values x_1 or x_2 . Assume that $p(x_1) = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$. Note that $p(x_2) = 1 - p(x_1) = 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1$, and 0.0. Compute and plot the Shannon entropy. Take the base of the logarithm as 2 as well as *e*. What do you conclude from the plot?

Solution: The Shannon entropy for a two-state variable is plotted as a function of probability in Figure 1.4. It is seen that entropy increases with increasing probability up to the point where the probability becomes 0.5 and then decreases with increasing probability, reaching zero when the probability becomes one. A higher logarithmic base produces lower entropy and vice versa, that is, the Shannon entropy is greater for logarithmic base 2 than it is for logarithmic base *e*. Because of symmetry, $H(X_1) = H(X_2)$ and therefore graphs will be the same.

1.2.4 Boltzmann, Gibbs and Shannon entropies

Using theoretical arguments Gull (1991) has explained that the Gibbs entropy is based on the ensemble which represents the probability that an *N*-particle system is in a particular microstate and making inferences given incomplete information. The Boltzmann entropy is based on systems each with one particle. The Gibbs entropy, when maximized (i.e., for the canonical ensemble), results numerically in the thermodynamic entropy defined by Clausius. The Gibbs entropy is defined for all probability distributions, not just for the canonical ensemble. Therefore,

$$S_G \leq S_E$$

where S_G is the Gibbs entropy, and S_E is the experimental entropy. Because the Boltzmann entropy is defined in terms of the single particle distribution, it ignores both the internal energy and the effect of inter-particle forces on the pressure. The Boltzmann entropy becomes



Figure 1.4 Shannon entropy for two-state variables.

the same as the Clausius entropy only for a perfect gas, when it also equals the maximized Gibbs entropy.

It may be interesting to compare the Shannon entropy with the thermodynamic entropy. The Shannon entropy provides a measurement of information of a system, and increasing of this information implies that the system has more information. In the canonical ensemble case, the Shannon entropy and the thermodynamic entropy are approximately equal to each other. Ng (1996) distinguished between these two entropies and the entropy for the second law of thermodynamics, and expressed the total entropy *S* of a system at a given state as

$$S = S_1 + S_2$$
 (1.23)

where S_1 is the Shannon entropy and S_2 is the entropy for the second law. The increasing of S_2 implies that the entropy of an isolated system increases as regarded by the second law of thermodynamics, and that the system is in decay. S_2 increases when the total energy of the system is constant, the dissipated energy increases and the absolute temperature is constant or decreases. From the point of view of living systems, the Shannon entropy (or thermodynamic entropy) is the entropy for maintaining the complex structure of living systems and their evolution. The entropy for the second law is not the Shannon entropy. Zurek (1989) defined physical entropy as the sum of missing information (Shannon entropy) and of the length of the most concise record expressing the information already available (algorithmic entropy), which is similar to equation (1.23). Physical entropy can be reduced by a gain of information or as a result of measurement.

1.2.5 Negentropy

The Shannon entropy is a statistical measure of dispersion in a set organized through an equivalent relation, whereas the thermodynamic entropy in a system is proportional to its ability to work, as discussed earlier. The second law of thermodynamics or Carnot's second principle is the degradation of energy from a superior level (electrical and mechanical energy) to a midlevel (chemical energy) and to an inferior level (heat energy). The difference in the nature and repartition of energy is measured by the physical energy. For example, if a system experiences an increase in heat, dQ, the corresponding increase in entropy dS can be expressed as

$$dS = \frac{dQ}{T} \tag{1.24}$$

where *T* is the absolute temperature, and *S* is the thermodynamic entropy.

Carnot's first principle of energy, conservation of energy, is

$$W - Q = 0 \tag{1.25}$$

and the second principle states

$$dS \ge 0 \tag{1.26}$$

where W is the work produced or output. This shows that entropy must always increase. Any system in time tends towards a state of perfect homogeneity (perfect disorder) where it is incapable of producing any more work, providing there are no internal constraints. The Shannon entropy in this case attains the maximum value. However, this is exactly the opposite of that in physics in that it is defined by Maxwell (1872) as follows: "Entropy of a system is the

16 Entropy Theory and its Application in Environmental and Water Engineering

mechanical work it can perform without communication of heat or change of volume. When the temperature and pressure have become constant, the entropy of the system is exhausted."

Brillouin (1956) reintroduced the Maxwell entropy while conserving the Shannon entropy as negentropy: "An isolated system contains negentropy if it reveals a possibility for doing a mechanical or electrical work. If a system is not at a uniform temperature, it contains a certain amount of negentropy." Thus, Marchand (1972) reasoned that entropy means homogeneity and disorder, and negentropy means heterogeneity and order in a system:

Negentropy = -entropy

Entropy is always positive and attains a maximum value, and therefore negentropy is always negative or zero, and its maximum value is zero. Note that the ability of a system to perform work is not measured by its energy, since energy is constant, but by its negentropy. For example, a perfectly disordered system, with a uniform temperature contains a certain amount of energy but is incapable of producing any work because its entropy is maximum and its negentropy is minimum. It may be concluded that information (disorder) and negentropy (order) are interchangeable. Acquisition of information translates into an increase of entropy and decrease of negentropy; likewise decrease of entropy translates into increase of negentropy. One cannot observe a phenomenon without altering it and the information acquired through an observation is always slightly smaller than the disorder it introduces into the system. This implies that a system cannot be exactly reconstructed as it was before the observation was made. Thus, the relation between the information and entropy *S* in thermodynamics is: $S = k \log N$, k = Boltzmann's constant (1.3806 × 10⁻¹⁶ erg/K), and <math>N = number of microscopic configurations of the system. The very small value of *k* means that a very small change in entropy corresponds to a huge change in information and vice versa.

Sugawara (1971) used negentropy as a measure of order in discussing problems in water resources. For example, in the case of hydropower generation, the water falls down and its potential energy is converted into heat energy and then into electrical energy. The hydropower station utilizes the negentropy of water. Another example is river discharge, which, with large fluctuations, has low negentropy or the smaller the fluctuation the higher the negentropy. In the case of a water treatment plant, input water is dirty and output water is clear or clean, meaning an increase in negentropy. Consider an example of rainwater distributed in time and space. The rainwater is in a state of low negentropy. Then, rainwater infiltrates and becomes groundwater and runoff from this groundwater becomes baseflow. This is in a state of high negentropy achieved in exchange of lost potential energy. The negentropy of a system can conserve entropy of water resources.

1.2.6 Exponential entropy

If the gain in information from an event occurring with probability p_i is defined as

$$G(p) = \Delta I = \exp[(1 - p_i)] \tag{1.27a}$$

then the exponential entropy, defined by Pal and Pal (1991a), can be expressed as

$$H = E(\Delta I) = \sum_{i=1}^{N} p_i \exp[(1 - p_i)]$$
(1.27b)

The entropy, defined by equation (1.27b), possesses some interesting properties. For example, following Pal and Pal (1991a), equation (1.27b) is defined for all p_i between 0 and 1,

is continuous in this interval, and possesses a finite value. As p_i increases, ΔI decreases exponentially. Indeed, *H* given by equation (1.27b) is maximum when all p_i 's are equal. Pal and Pal (1992) have mathematically proved these and other properties. If one were to plot the exponential entropy, the plot would be almost identical to the Shannon entropy. Pal and Pal (1991b) and Pal and Bezdek (1994) have used the exponential entropy in pattern recognition, image extraction, feature evaluation, and image enhancement and thresholding.

Example 1.3: Plot the gain function defined by equation (1.27a) for different values of probability: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. What do you conclude from this plot? Compare this plot with that in Example 1.1. How do the two gain functions differ?

Solution: The gain function is plotted as a function of probability in Figure 1.5. It is seen that as the probability increases, the gain function decreases, reaching the lowest value of one when the probability becomes unity. Comparing Figure 1.5 with Figure 1.3, it is observed that the exponential gain function changes more slowly and has a smaller range of variability than does the Shannon gain function.

Example 1.4: Consider a two-state variable taking on values x_1 or x_2 . Assume that $p(x_1) = 0.0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. Note that $p(x_2) = 1 - p(x_1) = 1.0$, 0.9, 0.8, 0.7, 0.6, 0.5, 0.5, 0.4, 0.3, 0.2, 0.1, and 0.0. Compute and plot the exponential entropy. What do you conclude from the plot? Compare the exponential entropy with the Shannon entropy.

Solution: The exponential entropy is plotted in Figure 1.6. It increases with increasing probability, reaching a maximum value when the probability becomes 0.5 and then declines, reaching a minimum value of one when the probability becomes 1.0. The pattern of variation of the exponential entropy is similar to that of the Shannon entropy. For any given probability



Figure 1.5 Plot of gain function of exponential entropy as defined by equation (1.27a).

18 Entropy Theory and its Application in Environmental and Water Engineering

value, the exponential entropy is higher than the Shannon entropy. Note that $H(X_1) = H(X_2)$; therefore graphs will be identical for X_1 and X_2 .

1.2.7 Tsallis entropy

Tsallis (1988) proposed another formulation for the gain in information from an event occurring with probability p_i as

$$G(p) = \Delta I = \frac{k}{q-1} [(1-p_i^{q-1})]$$
(1.28)

where k is a conventional positive constant, and q is any number. Then the Tsallis entropy can be defined as the expectation of the gain function in equation (1.28):

$$H = E(\Delta I) = \frac{k}{q-1} \sum_{i=1}^{N} p_i[(1-p_i^{q-1})]$$
(1.29)

Equation (1.29) shows that *H* is greater than or equal to zero in all cases. This can be considered as a generalization of the Shannon entropy or Boltzmann–Gibbs entropy. The Tsallis entropy has some interesting properties. Equation (1.29) achieves its maximum when all probabilities are equal. It vanishes when N = 1; as well as when there is only one event with probability one and others have vanishing probabilities. It converges to the Shannon entropy when *q* tends to unity.

Example 1.5: Plot the gain function defined by equation (1.18) for different values of probability: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. Take k as 1, and q as -1, 0,



Figure 1.6 Plot of exponential and Shannon entropy for two-state variables.

1.1, and 2. What do you conclude from this plot? Compare the gain function with the gain functions obtained in Examples 1.1 and 1.3.

Solution: The Tsallis gain function is plotted in Figure 1.7. It is seen that the gain function is highly sensitive to the value of q. For q = 1.1, and q = 2, the gain function is almost zero; for q = -1, and 0, it declines rapidly with increasing probability – indeed it reaches a very small value when the probability is about 0.5 or higher. Its variation is significantly steeper than the Shannon and exponential gain functions, and its pattern of variation is also quite different.

Example 1.6: Consider a two-state variable taking on values x_1 or x_2 . Assume that $p(x_1) = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$. Note that $p(x_2) = 1 - p(x_1) = 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1$, and 0.0. Compute and plot the Tsallis entropy. Take *q* as 1.5 and 2.0. What do you conclude from the plot?

Solution: The Tsallis entropy is plotted in Figure 1.8. It increases with increasing probability reaching a maximum value at the probability of about 0.6 and then declines with increasing probability. The Tsallis entropy is higher for q = 1.5 than it is for q = 2.0.

1.2.8 Renyi entropy

Renyi (1961) defined a generalized form of entropy called Renyi entropy which specializes into the Shannon entropy, Kapur entropy, and others. Recall that the amount of uncertainty

or the entropy of a probability distribution $P = (p_1, p_2, \dots, p_n)$, where $p_i \ge 0$ and $\sum_{i=1}^{n} p_i = 1$, denotes the amount of uncertainty as regards the outcome of an experiment whose values have probabilities p_1, p_2, \dots, p_n , measured by the quantity $H(p) = H(p_1, p_2, \dots, p_n)$. H(p, 1-p) is a continuous function of $p, 0 \le p \le 1$. Following Renyi (1961), one can also write: $H(wp_1, (1-w)p_1, p_2, \dots, p_n) = H(p_1, p_2, \dots, p_n) + p_1H(w, 1-w)$ for $0 \le w \ge 1$.



Figure 1.7 Plot of gain function for k = 1, and q = -1, 0, 1.1, and 2.



Figure 1.8 Plot of Tsallis entropy for two state variables with q = 1.5 and 2.

Renyi (1961) expressed

$$H_{\alpha}(p_1, p_2, \dots, p_n) = \frac{1}{1 - \alpha} \log_2 \left(\sum_{i=1}^n p_i^{\alpha} \right)$$
(1.30)

where $\alpha > 0$ and $\alpha \neq 1$. Equation (1.30) also is a measure of entropy and can be called the entropy of order α of distribution *P*. It can be shown from equation (1.30) that

$$\lim_{\alpha \to 1} H_{\alpha}(p_1, p_2, \dots, p_n) = \sum_{i=1}^n p_i \log \frac{1}{p_i}$$
(1.31)

which is the same as equation (1.12). Thus, the Shannon entropy is a limiting case of the Renyi entropy given by equation (1.30) for $\alpha \rightarrow 1$.

Let W(P) be the weight of the distribution P, 0 < W(P) < 1. The weight of an ordinary distribution is 1. A distribution which has weight less than 1 is called an incomplete distribution:

$$W(P) = \sum_{i=1}^{n} p_i$$
(1.32)

For two generalized distributions *P* and *Q*, such that $W(P) + W(Q) \le 1$,

$$H(P \cup Q) = \frac{W(P)H(P) + W(Q)H(Q)}{W(P) + W(Q)}$$
(1.33)

This is called the mean value property of entropy; the entropy of the union of two incomplete distributions is the weighted mean value of the entropies of the two distributions, where the entropy of each component is weighted by its own weight. This can be generalized as

$$H(P_1 \cup P_2 \cup P_3 \dots \cup P_n) = \frac{W(P_1)H(P_1) + W(P_2)H(P_2) + \dots W(P_n)H(P_n)}{W(P_1) + W(P_2) + \dots W(P_n)}$$
(1.34)

Any generalized $P\{p_1, p_2, \dots, p_n\}$ can be written as

$$P = \{p_1\} \cup \{p_2\} \cup \ldots \cup \{p_n\}$$
(1.35)

Thus, Renyi (1961) defined an entropy as

$$H(X) = \frac{1}{1-a} \log \frac{\sum_{i=1}^{N} p_i^a}{\sum_{i=1}^{N} p_i}, \ a \neq 1, \ a > 0$$
(1.36)

This is an entropy of order *a* of the generalized distribution *P*. As $a \rightarrow 1$, equation (1.36) converges to the Shannon entropy. Thus, the Shannon entropy can be considered as a limiting case of Reny's entropy. The Kapur entropy is a further generalization of the Renyi entropy as

$$H(X) = \frac{1}{1-a} \ln \frac{\sum_{i=1}^{N} p_i^a}{\sum_{i=1}^{N} p_i^b}, \ a \neq 1, \ b > 0, \ a+b-1 > 0$$
(1.37)

If b = 1, equation (1.37) reduces to the Renyi entropy. For b = 1, and a = 0, equation (1.37) reduces to log *N*, if $p_i = 1/N$, which is Hartley's measure (Hartley, 1928).

1.3 Entropy, information, and uncertainty

Consider a discrete random variable $X : \{x_1, x_2, ..., x_N\}$ with a probability distribution $P(x) = \{p_1, p_2, ..., p_N\}$. When the variable is observed to have a value x_i , the information is gained; the amount of information I_i so gained is defined as the magnitude of the logarithm of the probability:

$$I_i = -\log p_i = |\log p_i|$$

One may ask the question: How much uncertainty was there about the variable before observation? The question is answered by linking uncertainty to information. The amount of uncertainty can be defined as the average amount of information expected to be gained by observation. This expected amount of information is referred to as the entropy of the distribution

$$H = \sum_{i=1}^{N} p_i I_i = -\sum_{i=1}^{N} p_i \log p_i = \sum_{i=1}^{N} p_i \log p_i$$

This entropy of a discrete probability distribution denotes the average amount of information expected to be gained from observation. Once a value of the random variable X has been observed, the variable has this observed value with probability one. Then, the entropy of the new conditional distribution is zero. However, this will not be true if the variable is continuous.

1.3.1 Information

The term "information" is variously defined. In Webster's International Dictionary, definitions of "information" encompass a broad spectrum from semantic to technical, including "the communication or reception of knowledge and intelligence," "knowledge communicated by others and/or obtained from investigation, study, or instruction," "facts and figures ready for communication or use as distinguished from those incorporated in a formally organized branch of knowledge, data," "the process by which the form of an object of knowledge is impressed upon the apprehending mind so as to bring about the state of knowing," and "a numerical quantity that measures the uncertainty in outcome of an experiment to be performed." The last definition is an objective one and indeed corresponds to the informational entropy. Semantically, information is used intuitively, that is, it does not correspond to a well-defined numerical quantity which can quantify the change in uncertainty with change in the state of the system. Technically, information corresponds to a well-defined function which can quantify the change in uncertainty. This technical aspect is pursued in this book. In particular, the entropy of a probability distribution can be considered as a measure of uncertainty and also a measure of information. The amount of information obtained when observing the result of an experiment can be considered numerically equal to the amount of uncertainty as regards the outcome of the experiment before performing it. Perhaps the earliest definition of information was provided by Fisher (1921) who used the inverse of the variance as a measure of information contained in a distribution about the outcome of a random draw from that distribution.

Following Renyi (1961), another amount of information can be expressed as follows. Consider a random variable *X*. An event E is observed which in some way is related to *X*. The question arises: What is the amount of information concerning *X*? To answer this question, let *P* be the probability (original, unconditional) distribution of *X*, and *Q* be the conditional distribution of *X*, subject to the condition that event E has taken place. A measure of the amount of information concerning the random variable *X* contained in the observation of event E can be denoted by I(Q|P), where *Q* is absolutely continuous with respect to *P*. If h = dQ/dP, the Radon-Nikodym derivative of *Q* with respect to *P*, then a possible measure of the amount of information in question can be written as:

$$I_i(Q|P) = \int h \log_2 dQ = \int h \log_2 h dP$$
(1.38)

Assume *X* takes on a finite number of values: $X : \{x_1, x_2, ..., x_n\}$ If $P(X = x_i) = p_i$ and $P(X = x_i|E) = q_i$, for i = 1, 2, ..., n, then equation (1.38) becomes

$$I_1(Q|P) = \sum_{i=1}^n q_i \log_2 \frac{q_i}{p_i}$$
(1.39)

Also,

$$I_{\alpha}(Q|P) = \frac{1}{\alpha - 1} \log_2 \left(\sum_{i=1}^n \frac{q_i^{\alpha}}{p_i^{\alpha - 1}} \right)$$
(1.40)

For $\alpha \rightarrow 1$,

$$\lim_{\alpha \to 1} I_{\alpha}(Q|P) = I_1(Q|P)$$
(1.41)

This measures the amount of information contained in the observation of event E with respect to the random variable *X*, or the information of order α obtained if the distribution *P* is replaced by distribution *Q*.

If $I(Q_1|P_1)$ and $I(Q_2|P_2)$ are defined, and $P = P_1P_2$ and $Q = Q_1Q_2$ and the correspondence between the elements of *P* and *Q* is that introduced by the correspondence between the elements of P_1 and Q_1 and those of P_2 and Q_2 , then

$$I(Q|P) = I(Q_1|P_1) + I(Q_2|P_2)$$
(1.42)

If

$$W(P_1) + W(P_2) \le 1$$
 and $W(Q_1) + W(Q_2) \le 1$ (1.43)

then

$$I(Q_1 U Q_2 | P_1 U P_2) = \frac{W(Q_1) I(Q_1 | P_1) + W(Q_2) I(Q_2 | P_2)}{W(Q_1) + W(Q_2)}$$
(1.44)

The entropies can be generalized as

$$I_1(Q|P) = \frac{\sum_{i=1}^n q_i \log_2 \frac{q_i}{p_i}}{\sum_{i=1}^n q_i}$$
(1.45)

Likewise,

$$I_{\alpha}(Q|P) = \frac{1}{\alpha - 1} \log_2 \left[\frac{\sum_{i=1}^{n} \frac{q_i^{\alpha}}{p_i^{\alpha - 1}}}{\sum_{i=1}^{n} q_i} \right]$$
(1.46)

If *P* and *Q* are complete distributions then equation (1.45) will reduce to equation (1.39) and equation (1.46) to equation (1.40).

Information is a measure of one's freedom of choice when selecting an alternative or a message. Thus, it should not be confused with the meaning of the message. For example, two messages, one filled with meaning and the other with nonsense can be equivalent. Information relates not so much to what one does say as to what one could say. If there are two alternative messages and one has to choose one message then it is arbitrarily stated that the information associated with this case is unity which indicates the amount of freedom one has in selecting a message. Thus, the concept of information applies to the whole situation, not to individual messages. The messages can be anything one likes.

The measure of information is entropy. Entropy is a measure of randomness or shuffledness. Physical systems tend to become more and more shuffled, less and less organized. If a system is highly organized and it is not characterized by a large degree of randomness, then its information (entropy) is low.

If *H* is zero ($p_i = 1$, certainty) and ($p_j = 0$, $j \neq i$ impossibility) then information is zero and there is no freedom of choice. When one is completely free, *H* is maximum and reduces to zero when the freedom of choice is gone. Thus, *H* increases with the increasing number of alternatives or by equiprobability of alternatives if the number of alternatives is fixed. There is more information if the number of alternatives to choose from is more.

Entropy H(X) permits to measure information and for that reason it is also referred to as informational entropy. Intuitively, information reduces uncertainty which is a measure of surprise. Thus, information *I* is a reduction in uncertainty H(X) and can be defined as

$$I = H_I - H_0 \tag{1.47}$$

where H_I is the entropy (or uncertainty) of input (or message sent through a channel), and H_O is the entropy (or uncertainty) of output (or message received). Equation (1.47) defines a reduction in uncertainty. Consider an input-output channel or transmission conduit. Were there no noise in the conduit, the output (the message received by the receiver or receptor) would be certain as soon as the input (message sent by the emitter) was known. This means that the uncertainty in output H_O would be 0 and $I = H_I$.

1.3.2 Uncertainty and surprise

The concept of information is closely linked with the concept of uncertainty or surprise. The quantity $-\log (1/p_i)$ can be used to denote surprise or unexpectedness (Watanabe, 1969). When all probabilities are equal, it is impossible to state that one possibility is more likely than another. This means there is complete uncertainty. Any information about the nature of an event under such conditions can be expected to shed more light than in any other condition. Maximum entropy is therefore a measure of complete uncertainty. Maximum uncertainty can be equated with a condition in which the expected information from actual events is also maximized. Now assume that $X = x_i$ and it occurs with probability one, $p_i = 1$; that is, the event occurs with certainty and hence there is no uncertainty. This means that $p_j = 0$, $j \neq i$. In this case, there is no surprise and therefore the occurrence of event $X = x_i$ conveys no information, since it is known what the event must be. One can state that the information content of observing x_i or the anticipatory uncertainty of x_i prior to the observation is a decreasing function of the probability $p(x_i)$. The more likely the occurrence of x_i , the less information its actual observation contains.

If x_i 's occur with probabilities p_i 's, $p_i \neq p_j$, $i = j = 0, \pm 1, \ldots, \pm N$, then there is more surprise and therefore more information that $X = x_i$ occurs with probability p_i than does $X = x_j$ with probability p_j where $p_j > p_i$. Thus, information, uncertainty and surprise are all related. Information is gained only if there is uncertainty about an event. Uncertainty suggests that the event may take on different values. The value that occurs with a higher probability conveys less information and vice versa. The probability of occurrence of a certain value is the measure of uncertainty or the degree of expectedness and hence of information. Shannon (1948) argued that entropy is the expected value of the probabilities of alternative values that an event may take on. The information gained is indirectly measured as the amount of reduction of uncertainty or of entropy.

The above discussion suggests that uncertainty can be understood to be a form of information deficiency or reflects information reduction, which may be because information is unreliable, biased, contradictory, vague, incomplete, imprecise, erroneous, fragmentary, or unfounded.

In many cases, information deficiency can be reduced and hence uncertainty. Consider, for example, prediction of a 100-year flood from a 20-year record. This prediction has uncertainty, say, u_1 (it can be referred to as a priori uncertainty). If the record length is increased to 50 years, the prediction will have less uncertainty, say u_2 (it can be referred to as posteriori uncertainty). The reduction in uncertainty due to a more complete record (or an action) is $u_1 - u_2$ which is equal to the information gain, that is, this is the amount of information realized as a result of uncertainty reduction. Klir (2006) refers to this uncertainty as uncertainty-based information, and reasons that this type of information does not encompass the concept of uncertainty in its entirety and is hence restricted somewhat. On the other hand, information is understood to reduce uncertainty or reflects uncertainty reduction. Klir (2006) calls this an information-based uncertainty.

1.4 Types of uncertainty

Uncertainty can appear in different forms. It can appear in both probabilistic and deterministic phenomena. In deterministic phenomena, it appears as a result of fuzziness about the phenomena, in data or in relations about the variables, and can be dealt with using the fuzzy set theory (plausibility, possibility, and feasibility). Probabilistic uncertainty is associated with the probability of outcomes and is entropy. This is also linked with arrow of time, meaning that it increases from past to present to future.

In environmental and water resources engineering models which express relations among states of given variables are constructed for a variety of purposes, including prediction, retrodiction, forecasting, diagnosis, prescription, planning, scheduling, control, simulation, detection, estimation, extrapolation, and design. Each of these purposes is subject to uncertainty. Depending on the purpose, unknown states of some variables are determined from the known states of other variables, using appropriate relation(s). If the relation is unique, the model is deterministic; otherwise it is nondeterministic and involves uncertainty. The uncertainty relates to the purpose for the construction of the model, and can thus be distinguished as predictive uncertainty, retrodictive uncertainty, forecasting uncertainty, diagnostic uncertainty, simulation uncertainty, detection uncertainty, estimation uncertainty, extrapolation uncertainty, and design uncertainty. It is logical that this uncertainty is incorporated into the model description. A decision is an action from a set of actions, based on the consequences of individual actions. Clearly, these actions are subject to anticipated uncertainty due to the uncertainty associated with consequences.

For probabilistic uncertainty, the value of p(x) represents the degree of evidential support that x is the true alternative, $x \in X : \rightarrow [0, 1]$ set. Then the Shannon entropy measures the amount of uncertainty in evidence expressed by the probability distribution P on the finite set:

$$-c\sum p(x)\log_b p(x) \tag{1.48}$$

where *c* and *b* are constant, and $b \neq 1$. The choice of *b* and *c* determines the unit in which the uncertainty is measured. The most common measurement unit is a bit. If

$$-c\log_b \frac{1}{2} = 1$$
 (1.49)

26 Entropy Theory and its Application in Environmental and Water Engineering

then b = 2, c = 1, and it would imply $X = [x_1, x_2]$ and $p(x_1) = p(x_2) = 0.5$. This is often referred to as a normalization requirement. Thus, one bit is the amount of information gained or uncertainty removed when one learns the answer to a question whose two possible outcomes are equally likely. Thus, H(p) is called the Shannon measure of uncertainty or Shannon entropy.

To gain further insight about the type of uncertainty measured by the Shannon entropy, one can write Shannon entropy as

$$H(p) = -\sum_{x \in X} p(x) \log_2 \left[1 - \sum_{y \neq x} p(y) \right]$$
(1.50)

Now consider the term

$$con(x) = \sum_{y \neq x} p(y) \tag{1.51}$$

which expresses the total evidence (sum) as a result of the alternatives that are different from *x*, that is, $y \neq x$. This evidence is in conflict with the one focusing on *x*. It is seen that $con(x) \in [0, 1]$ for each $x \in X$. The term $-\log_2 [1 - con(x)]$ in equation (1.51) increases monotonically with con(x) and its range is extended from [0, 1] to $[0, \infty]$. Thus, the Shannon entropy is the mean (expected value) of the conflict among evidences expressed by each probability distribution *P*.

Example 1.7: One way to gain further insight into the Shannon uncertainty is from equation $s(a) = c \log_b a$, where *c* and *b* are constants, and $b \neq 1$. The Shannon uncertainty here is analogous to the gain function defined by equation (1.18). Taking c = -1, and b = 2, $s(a) = c \log_2 a$. Plot this function taking a = 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0. What do you conclude from this graph?

Solution: The function is plotted as a function of *a* in Figure 1.9. The function declines with increasing *a* and reaches zero when a = 1.



Figure 1.9 Plot of function $s(a) = -\log_2 a$ for different values of a.

Example 1.8: Consider $X = \{x_1, x_2\}$ with, $p(x_1) = a$, and $p(x_2) = 1 - a$, $a \in [0, 1]$; x_1 and x_2 represent two alternatives. The Shannon entropy depends only on *a* and is comprised of two components $S_1 = -a \log_2 a$ and $S_2 = -(1 - a) \log_2(1 - a)$; each component is analogous to the gain function. Compute the Shannon entropy as well as each of the two components, taking a = 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0, and graph them. What do you conclude from these graphs?

Solution: The Shannon entropy and each component thereof are plotted in Figure 1.10. The Shannon entropy graph is as before. The two components are mirror images of each other, as shown in Figure 1.10a. Graphs of S_1 and S_2 are shown in Figure 1.10a. Graph S_1 and S_2 are the same except for the change of scale.

1.5 Entropy and related concepts

Hydrologic and environmental systems are inherently spatial and complex, and our understanding of these systems is less than complete. Many of the systems are either fully stochastic or part-stochastic and part-deterministic. Their stochastic nature can be attributed to the randomness in one or more of the following components that constitute them: 1) system structure (geometry), 2) system dynamics, 3) forcing functions (sources and sinks), and 4) initial and boundary conditions. As a result, a stochastic description of these systems is needed, and the entropy theory enables the development of such a description.

Fundamental to the planning, design, development, operation, and management of environmental and water resources projects is the data that are observed either in field or experimentally and the information they convey. If this information can be determined, it can also serve as a basis for design and evaluation of data collection networks, design of sampling schemes, choosing between models, testing the goodness-of-fit of a model, and so on.

Engineering decisions concerning hydrologic systems are frequently made with less than adequate information. Such decisions may often be based on experience, professional judgment, thumb rules, crude analyses, safety factors, or probabilistic methods. Usually, decision making under uncertainty tends to be relatively conservative. Quite often, sufficient data are not available to describe the random behavior of such systems. Although probabilistic methods allow for a more explicit and quantitative accounting of uncertainty, their major difficulty occurs due to the lack of sufficient or complete data. Small sample sizes and limited information render the estimation of probability distributions of system variables with conventional methods difficult. This problem can be alleviated by the use of entropy theory which enables to determine the least-biased probability distributions with limited knowledge and data. Where the shortage of data is widely rampant as is normally the case in many countries, the entropy theory is particularly appealing.

1.5.1 Information content of data

One frequently encounters a situation in which to exercise freedom of choice, evaluate uncertainty or measure the information gain or loss. The freedom of choice, uncertainty, disorder, information content, or information gain or loss has been variously measured by relative entropy, redundancy, and conditional and joint entropies employing conditional and joint probabilities. As an example, in the analysis of empirical data, the variance has often been interpreted as a measure of uncertainty and as revealing gain or loss in information. However, entropy is another measure of dispersion – an alternative to variance. This suggests that it is possible to determine the variance whenever it is possible to determine the entropy



Figure 1.10 Shannon entropy.

measure, but the reverse is not necessarily true. However, variance is not the appropriate measure if the sample size is small.

1.5.2 Criteria for model selection

Usually there are more models than one needs and a choice has to be made as to which model to choose. Akaike (1973) formulated a criterion, called Akaike Information Criterion (*AIC*),

for selecting the best model from amongst several models as

$$AIC = -2\log(maximized \ likelihood) + 2k \tag{1.52}$$

where *k* is the number of parameters of the model. *AIC* provides a method for model identification and can be expressed as minus twice the logarithm of the maximum likelihood plus twice the number of parameters used to find the best model. The maximum likelihood and entropy are uniquely related. When there are several models, the model giving the minimum value of *AIC* should be selected. When the maximum likelihood is identical for two models, the model with the smaller number of parameters should be selected, for that will lead to smaller *AIC* and comply with the principle of parsimony.

1.5.3 Hypothesis testing

Another important application of the entropy theory is in the testing of hypotheses (Tribus, 1969). With use of Bayes' theorem in logarithmic form, an evidence function can be defined for comparing two hypotheses. The evidence in favor of a hypothesis over its competitor is the difference between the respective entropies of the competition and the hypothesis under testing. Defining surprisal as the negative of the logarithm of the probability, the mean surprisal for a set of observations is expressed. Therefore, the evidence function for two hypotheses is obtained as the difference between the two values of the mean surprisal multiplied by the number of observations.

1.5.4 Risk assessment

There are different types of risk, such as business risk, social risk, economic risk, safety risk, investment risk, occupational risk, and so on. In common language, risk is the possibility of loss or injury and the degree of probability of such loss. Rational decision making requires a clear and quantitative way of expressing risk. In general, risk cannot be avoided and a choice has to be made between risks. To put risk in proper perspective, it is useful to clarify the distinction between risk, uncertainty, and hazard.

The notion of risk involves both uncertainty and some kind of loss or damage. Uncertainty reflects the variability of our state of knowledge or state of confidence in a prior evaluation. Thus, risk is the sum of uncertainty plus damage. Hazard is commonly defined as a source of danger and involves a scenario identification (e.g., failure of a dam) and a measure of the consequence of that scenario or a measure of the ensuing damage. Risk encompasses the likelihood of conversion of that source into the actual delivery of loss, injury, or some form of damage. Thus, risk is the ratio of hazard to safeguards. By increasing safeguards, risk can be reduced but it is never zero. Since awareness of risk reduces risk, awareness is part of safeguards. Qualitatively, risk is subjective and is relative to the observer. Risk involves the probability of scenario and its consequence resulting from the happening of the scenario. Thus, one can say that risk is probability and consequence. Kaplan and Garrick (1981) have analyzed risk using entropy. Luce (1960) has reasoned that entropy should be described as an average measure of risk, not of uncertainty.

Questions

Q.1.1 Assume that there are 256 possibilities in a particular case. These possibilities are arranged in such a way that each time an appropriate piece of information becomes

available, the number of possibilities reduces to half. What is the information gain in bits if the number of possibilities is reduced to 128, to 64, to 32, to 16, to 8, to 4, and to 2?

- **Q.1.2** Assume that there are 10,000 possibilities in a particular case. These possibilities are arranged in such a way that each time an appropriate piece of information becomes available, the number of possibilities reduces to one tenth. What is the gain in information in decibels or dits if the number of possibilities is reduced to 1,000, to 100, and to 10?
- **Q.1.3** Consider that a random variable *X* takes on values x_i , i = 1, 2, 3, 4, 5, with probabilities $p(x_i) = 0.10, 0.20, 0.30, 0.25$, and 0.15. Compute the gain in information for each value using the Shannon entropy, exponential entropy and Tsallis entropy with q = 0.5. Which entropy provides a larger gain?
- **Q.1.4** Consider the probabilities in Q.1.3. Order them in order of increasing surprise and relate the surprise to the gain in information computed in Q.1.3.
- **Q.1.5** Consider two distributions $P_i = p = 0.1$, i = 1, 2, ..., 10; $q_j = q = 0.05$, j = 1, 2, 3, ..., 20 having equiprobable outcomes. Compute the maximum entropy of each distribution in bits. Compare these two distributions by determining the difference in the information contents of these distributions. Is there a loss of information with the increase in the number of possible outcomes?
- **Q.1.6** Consider two distributions $P_i = p = 0.05$, i = 1, 2, ..., 20; $q_i = q = 0.10$, j = 1, 2, 3, ..., 10 having equiprobable outcomes. Compute the maximum entropy of each distribution in bits. Compare these two distributions by determining the difference in the information contents of these distributions. Is there a gain of information with the decrease in the number of possible outcomes?
- **Q.1.7** Consider that a discrete random variable *X* takes on 10 values with probability distribution $P : P : \{p_1, p_2, ..., p_{10}\}$ corresponding to $X : \{x_i, i = 1, 2, ..., 10\}$. What distribution *P* will yield the maximum and minimum values of the Shannon entropy?
- **Q.1.8** Consider an event A. The probability of the occurrence of event A can be regarded as a measure of uncertainty about its occurrence or non-occurrence. For what value of the probability will the uncertainty be maximum and why?
- **Q.1.9** Consider a coin tossing experiment. Let the probability of the occurrence of head be denoted as *p* and that of tail as *q*. Express the Shannon entropy of this experiment. Note q = l p or p = l q. Plot a graph of entropy by taking different values of *p*. For what value of *p* does the entropy attain a maximum?
- **Q.1.10** Consider a six-faced dice throwing experiment. The dice is unbiased so the probability of the occurrence of any face is the same. In this case there are six possible events and each event is equally likely. Express the Shannon entropy of this experiment and compute its value. Now consider that the concern is whether an even-numbered or an odd-numbered face shows upon throw. In this case there are only two possible events: (even, odd). Express the Shannon entropy of this experiment and compute its value. Which of these two cases has higher entropy? Which case is more uncertain? Is there any reduction in uncertainty in going from case one to case two?

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceedings, 2nd International Symposium on Information Theory*, B.N. Petrov and F. Csaki, eds., Publishing House of the Hungarian Academy of Sciences, Budapest, Hungary.
- Batty, M. (2010). Space, scale, and scaling in entropy maximizing. *Geographical Analysis*, Vol. 42, pp. 395–421.

Brillouin, L. (1956). Science and Information Theory. Academic Press, New York.

- Denbigh, K.G. (1989). Note on entropy, disorder and disorganization. *British Journal of Philosophical Science*, Vol. 40, pp. 323–32.
- Fast, J.D. (1968). *Entropy: The Significance of the Concept of Entropy and its Application in Science and Technology*. Gordon and Breach, Science Publishers Inc., New York.
- Fisher, R.A. (1921). On mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, Series A, Vol. 222, pp. 30–368.
- Gull, S.F. (1991). Some misconceptions about entropy. Chapter 7 in: *Maximum Entropy in Action*, edited by B. Buck and V.A. McCauley, Oxford Science Publishers, pp. 171–86.
- Hartley, R.V.L. (1928). Transmission of information. *The Bell System Technical Journal*, Vol. 7, No. 3, pp. 535–63.
- Kaplan, S. and Garrick, B.J. (1981). On the quantitative definition of risk. *Risk Analysis*, Vol. 1, No. 1, pp. 11–27.
- Kapur, J.N. (1989). *Maximum Entropy Models in Science and Engineering*. Wiley Eastern Ltd., New Delhi, India.
- Klir, G. J. (2006). Uncertainty and Information: Foundations of Generalized Information Theory. John Wiley & Sons, New York.
- Luce, R.D., (1960). The theory of selective information and source of its behavioral applications. In: *Developments in Mathematical Psychology*, edited by R.D. Luce, The Free Press, Glencoe.
- Marchand, B., (1972). Information theory and geography. Geographical Analysis, Vol. 4, pp. 234-57.
- Maxwell, J.C. (1872). The Theory of Heat. D. Appleton, New York (reproduced by University Microfilms).
- Ng, S.K. (1996). Information and system modeling. *Mathematical and Computer Modeling*, Vol. 23, No. 5, pp. 1–15.
- Pal, N.R. and Pal, S.K. (1991a). Entropy: A new definition and its applications. *IEEE: Transactions on Systems, Man, and Cybernetics,* Vol. 21, No. 5, pp. 1260–70.
- Pal, N.R. and Pal, S.K. (1991b). Image model, Poisson distribution and object extraction. *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 5, No. 3, pp. 459–83.
- Pal, N.R. and Pal, S.K. (1992). Some properties of the exponential entropy. *Informational Sciences*, Vol. 66, pp. 119–37.
- Pal, N.R. and Bezdek, J.C. (1994). Measuring fuzzy uncertainty. *IEEE Transactions on Fuzzy Systems*, Vol. 2, No. 2, pp. 107–18.
- Papalexiou, S.M. and Koutsoyiannis, D. (2012). Entropy based derivation of probability distributions: A case study to daily rainfall. *Advances in Water Resources*, Vol. 45, pp. 51–57.
- Prigogine, I. (1989). What is entropy? Naturwissenschaften, Vol. 76, pp. 1-8.
- Renyi, A. (1961). On measures of entropy and information. Proceedings, 4th Berkeley Symposium on Mathematics, *Statistics and Probability*, Vol. 1, pp. 547–561.
- Rosenthal, H. and Binia, J. (1988). On the epsilon entropy of mixed random variables. *IEEE Transactions* on *Information Theory*, Vol. 34, No. 5, pp. 1110–14.
- Shannon, C.E. (1948). A mathematical theory of communications, I and II. *Bell System Technical Journal*, Vol. 27, pp. 379–443.

- Shannon, C.E. and Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, Illinois.
- Sugawara, M. (1971). Water resources and negentropy. Proceedings of the Warsaw Symposium on Mathematical Models in Hydrology, Vol. 2, pp. 876–8, IAHS-UNESCO-WM, Warsaw, Poland.
- Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, Vol. 52, No. 1/2, pp. 479–87.
- Tribus, M. (1969). Rational Description: Decision and Designs. Pergamon Press, New York.
- Watanabe, S. (1969). *Knowing and Guessing: A Quantitative Study of Inference and Information*. John Wiley & Sons, New York.
- Zurek, W.H., 1989. Algorithmic randomness and physical entropy. *Physical Review A*, Vol. 40, No. 8, pp. 4731–51.

Additional References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Vol. AC-19, No. 6, pp. 710–23.
- Akaike, H. (1985). Prediction and entropy. Chapter 1 in: *A Celebration of Statistics*, edited by A.C. Atkinson and S.E. Fienberg, pp. 1–24, Springer-Verlag, Heidelberg, Germany.
- Carnap, R. (1977). Two Essays on Entropy. University of California Press, Berkeley, California.
- Cover, T.M. and Thomas, J.A. (1991). Elements of Information Theory. John Wiley & Sons, New York.
- Harmancioglu, N.B. and Singh, V.P. (1998). Entropy in Environmental and Water Resources. pp. 225–241, Chapter in *Encyclopedia of Hydrology and Water Resources*, edited by D.R. Herschy, Kluwer Acdemic Publishers, Dordrecht.
- Jaynes, E.T. (1958). Probability Theory in Science and Engineering. Colloquium Lectures in Pure and Applied Science, Vol. 4, Field Research Laboratory, Socony Mobil Oil Company, Inc., USA.
- Khinchin, A.I. (1957). The Entropy Concept in Probability Theory. Translated by R.A. Silverman and M.D. Friedman, Dover Publications, New York.
- Jaynes, E.T. (1957). Information theory and statistical mechanics, I. Physical Review, Vol. 106, pp. 620-30.
- Jaynes, E.T. (1982). On the rationale of maximum entropy methods. *Proceedings of the IEEE*, Vol. 70, pp. 939–52.
- Levine, R.D. and Tribus, M., eds., (1978). *The Maximum Entropy Formalism*, 498 p., The MIT Press, Cambridge, Massachusetts.
- Singh, V.P. (1998). *Entropy-Based Parameter Estimation in Hydrology*. Kluwer Academic Publishers, Boston, Massachusetts.
- Singh, V.P. (1998). The use of entropy in hydrology and water resources. *Hydrological Processes*, Vol. 11, pp. 587–626.
- Singh, V.P. (1998). Entropy as a decision tool in environmental and water resources. *Hydrology Journal*, Vol. XXI, No. 1–4, pp. 1–12.
- Singh, V.P. (2000). The entropy theory as tool for modeling and decision making in environmental and water resources. *Water SA*, Vol. 26, No. 1, pp. 1–11.
- Singh, V.P. (2003). The entropy theory as a decision making tool in environmental and water resources. In: *Entropy Measures, Maximum Entropy and Emerging Applications*, edited by Karmeshu, Springer-Verlag, Bonn, Germany, pp. 261–97.
- Singh, V.P. (2005). Entropy theory for hydrologic modeling. in: Water Encyclopedia: Oceanography; Meteorology; Physics and Chemistry; Water Law; and Water History, Art, and Culture, edited by J. H. Lehr, Jack Keeley, Janet Lehr and Thomas B. Kingery, John Wiley & Sons, Hoboken, New Jersey, pp. 217–23.
- Singh, V.P. (2011). Hydrologic synthesis using entropy theory: Review. *Journal of Hydrologic Engineering*, Vol. 16, No. 5, pp. 421–33.
- Singh, V.P. and Fiorentino, M., editors (1992). *Entropy and Energy Dissipation in Hydrology*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 595 p.
- White, H. (1965). The entropy of a continuous distribution. *Bulletin of Mathematical Biophysics*, Vol. 27, pp. 135–43.