On Generalized Information Measures and Their Applications

Article *in* Advances in Imaging and Electron Physics · December 1989

Impact Factor: 0.34 · DOI: 10.1016/S0065-2539(08)60580-6

1 author:

Inder J. Taneja
Federal University of Santa Catarina

ADVANCES IN ELECTRONICS AND ELECTRON PHYSICS, VOL. 76

On Generalized Information Measures and Their Applications

INDER JEET TANEJA

Departamento de Matemática
Universidade Federal de Santa Catarina
Florianópolis, Brazil

0-12-014676-2

I. Introduction

Information theory is a relatively new branch of mathematics that was made mathematically rigorous only in 1940s. The term *information theory* does not possess a unique definition. Broadly speaking, information theory deals with the study of problems concerning any system. This includes information processing, information storage, information retrieval, and decision making. In a narrow sense, information theory studies all theoretical problems connected with the transmission of information over communication channels. This includes the study of uncertainty (information) measure and various practical and economical methods of coding information for transmission.

The first studies in this direction were undertaken by Nyquist in 1924 and 1928 and by Hartley in 1928, who recognized the logarithmic nature of the measure of information. In 1948, Shannon published a remarkable paper on the properties of information sources and of the communication channels used to transmit the outputs of these sources. Around the same time, Wiener (1948) also considered the communication situation and came up, independently, with results similar to those of Shannon.

In the last 40 years, the literature on information theory has become quite voluminous. Apart from communication theory, it has found deep applications in many social, physical, and biological sciences, e.g., economics, accounting, language, statistics, physics, ecology, psychology, pattern recognition, fuzzy sets theory, computer sciences, etc.

A key feature of Shannon information theory is that the term *information* can often be given a mathematical meaning as a numerically measurable quantity, on the basis of a probabilistic model, in such a way that the solutions of many important problems of information storage and transmission can be formulated in terms of this measure of the amount of information. This important measure has a very concrete operational interpretation: It roughly equals the minimum number of binary digits needed, on the average, to encode the message in question. The coding theorems of information theory provide such an overwhelming evidence for the adequacy of the Shannon information measure that to look for essentially different measures of information might appear to make no sense at all. Moreover, it has been shown by several authors, starting with Shannon (1948), that the measure of the amount of information is uniquely determined by some rather natural postulates. Still, all the evidence that the Shannon information measure is the only possible one is valid only within the restricted scope of coding problems considered by Shannon. As pointed out by Rényi (1961) in his fundamental paper on generalized information measures, in other sorts of problems other quantities may serve just as well, or even better, as measures of information. This should be supported either by their operational significance or by a set of natural postulates characterizing them, or, preferably, by both. Thus the idea of generalized entropies arises in the literature. It found its birth in Rényi (1961), who characterized a scalar parametric entropy as an entropy of order r, which includes Shannon entropy as a limiting case.

In these notes we propose to discuss various generalized entropies and generalized divergence measures. Mainly, we have taken generalizations of Shannon's entropy, directed divergence, J-divergence, and information radius. Applications of these generalized information measures to statistical pattern recognition are discussed. These studies have been conducted on the generalized information measures written in unified expressions.

Most of the results presented are the author's or joint contributions with the author. In part, these results were presented as seminars at the Università di Salerno, Italy during 1983/1984; as a summer course at Universidade Federal do Rio de Janeiro, Brazil during January and February, 1987, and as a short course at the International Symposium on Information and Coding Theory, Campinas, Brazil during July 1987.

II. Shannon's Entropy and Its Generalizations

This section deals with the Shannon's entropy and its generalizations. The generalizations considered involve one and two scalar parameters. Some of these generalizations are written in a unified way. Some properties of unified entropy are studied. Some results connected with inequalities and entropy series are also specified.

A. Shannon's Entropy

Let

$$\Delta_n = \{P = (p_1, p_2, \ldots, p_n) : p_i > 0, i = 1, 2, \ldots, n, \sum_{i=1}^{n} p_i = 1\},$$

and

$$\Delta_n^0 = \{P = (p_1, p_2, \ldots, p_n) : p_i \geqq 0, i = 1, 2, \ldots, n, \sum_{i=1}^{n} p_i = 1\}, \qquad n \geq 2$$

be two sets of discrete finite (n-ary) probability distributions.

Shannon (1948), first investigated and characterized, through certain postulates, a measure of information given by

$$H(P) = -\sum_{i=1}^{n} p_i \log p_i, \tag{1}$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$.

For Equation (1) $H(P)$ is known in the literature as *Shannon's entropy*.

Remarks. Throughout these notes it is understood that all the logarithms are in base 2 and $0.\log 0 = 0$. Also, we shall take $p_{\max} = \max\{p_1, p_2, \ldots, p_n\}$.

Equation (1) is also known as a measure of uncertainty. It measures the amount of information contained in a distribution, i.e., the amount of uncertainty concerning the outcome of an experiment. It has been shown by several autors, starting with Shannon (1948), that the measure of the amount of information [Eq. (1)] is uniquely determined by rather natural postulates. In other words, it arises naturally from statistical concepts (Ash, 1965). For a brief review refer to Aczél and Daróczy (1975), Mathai and Rathie (1975), and Taneja (1979). During its characterization, mainly three properties or postulates are considered by several authors, together or separately. These properties are given by:

1. *Additivity.* We can write

$$H(P*Q) = H(P) + H(Q),$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, $Q = (q_1, q_2, \ldots, q_m) \in \Delta_m^0$, and $P*Q = (p_1q_1, p_1q_2, \ldots, p_1q_m, \ldots, p_nq_1, p_nq_2, \ldots, p_nq_m) \in \Delta_{nm}^0$.

2. *Recursivity* or *branching.* We can write

$$\begin{aligned} H(p_1, p_2, \ldots, p_n) = {} & H(p_1 + p_2, p_3, \ldots, p_n) \\ & + (p_1 + p_2) H\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right), \qquad p_1 + p_2 > 0, \end{aligned}$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$.

3. *Sum representation.* We can write

$$H(P) = \sum_{i=1}^{n} f(p_i),$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, where $f(p) = -p \log p$, $p \in [0, 1]$.

B. Entropy of Order r

A systematic attempt to develop a generalization of Shannon's entropy was carried out by Rényi (1961), who characterized an *entropy of order r* given by

$$H_r(P) = (1 - r)^{-1} \log\left(\sum_{i=1}^{n} p_i^r\right), \qquad r \neq 1, r > 0 \tag{2}$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, where r is a real parameter. We can easily verify that $\lim_{r \to 1} H_r(P) = H(P)$.

Campbell (1965), for the first time, has shown that the variable length version of the elementary coding theorem carries over to entropy of order r if one considers exponential averaging in the definition of average codeword length instead of the standard arithmetic procedure. Parker (1980) proved that a simple generalization of Huffman algorithm solves the problem of minimizing generalized exponential length and its increasing functions, which includes, in particular, a generalized length in terms of entropy of order r. Blumer (1982) considered the problem of minimizing redundancy of order r defined in terms of entropy of order r and obtained bounds sharper than that of Gallager (1978). Taneja (1984a; 1985b) extended the concept of exponentiated average codeword length of order r to the best one-to-one codes. For some other applications of entropy of order r, refer to Jelinek (1968a, 1968b), Jelinek and Schneider (1972), Csiszár (1974), Nath (1975), Arimoto (1975, 1976), Ben-Bassat and Raviv (1978), Kieffer (1979), Campbell (1985, 1987), and Kapur (1983, 1986).

Entropy of order r satisfies *additivity* but lacks *recursivity* as well as *sum property*. Rényi (1961) considered an additional axiom generalizing sum property that is generally known as *quasi-linearity*. Based on the same motivations of Rényi, later researchers (Aczél and Daróczy, 1963; Varma, 1966; Kapur, 1967; and Rathie, 1970) generalized the entropy of order r by changing some of its postulates. These generalizations are given in Section II.F.

C. Entropy of Degree s

For operational purposes, it seems more natural to consider, the slar expression $\Sigma_{i=1}^n p_i^s$ as an information measure instead of Rényi's entropy of order r. So, Havrda and Charvát (1967) proposed the following *entropy of degree s*

$$H^s(P) = (2^{1-s} - 1)^{-1}\left[\sum_{i=1}^{n} p_i^s - 1\right], \qquad s \neq 1, s > 0, \tag{3}$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$. In this case, we can also easily verify that $\lim_{s \to 1} H^s(P) = H(P)$.

This quantity permits a simpler characterization (Havrda and Charvát, 1967). It lacks the additivity property but satisfies the *recursivity of degree s* and the *sum representation* given by

$$H^s(P) = H^s(p_1 + p_2, p_3, \ldots, p_n) + (p_1 + p_2)^s H^s\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right),$$

$$p_1 + p_2 > 0,$$

and

$$H^s(P) = \sum_{i=1}^{n} f_s(p_i),$$

respectively, for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, where $f_s(p) = (2^{1-s} - 1)^{-1}[p^s - p]$, $s \neq 1, s > 0$ for all $p \in [0, 1]$.

Taneja (1975) (refer also Sharma and Taneja, 1975, 1977) studied a generalization of $H^s(P)$ involving two scalar parameters. Its exact expression, along with other entropies, is given in Section II.F. Some results on redundancy of degree s can be seen in Taneja (1986a).

D. *Entropy of Kind t*

Arimoto (1971) considered a generalized f-entropy involving a real function f with some conditions. Being an example of this generalized f-entropy, Arimoto came up to a generalized entropy involving a real parameter, here we call it *entropy of kind t*, given by

$$_tH(P) = (2^{t-1} - 1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^{1/t}\right)^t - 1\right], \qquad t \neq 1, t > 0 \tag{4}$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$. In this case, also, we can easily verify that $\lim_{t\to 1} {}_tH(P) = H(P)$.

Arimoto's main motivation in considering generalized f-entropy was to prove some important results on decision theory connected to the Bayesian probability of error.

This entropy of kind t is neither *additive*, nor *recursive*, nor does it satisfy the *sum representation.*

Thus, we see that out of the three generalized entropies, i.e., the entropy of order r, the entropy of degree s, and the entropy of kind t, the first one is additive, the second one is recursive and satisfies sum representation, while the third one does not satisfy any of these properties. However, all contained the common function $\Sigma_{i=1}^n p_i^r$, which makes them related as follows:

$$\begin{aligned} H_r(P) &= (1 - r)^{-1} \log[(2^{1-r} - 1)^{-1} H^r(P) + 1] \\ &= r(1 - r)^{-1} \log[(2^{r^{-1}(1-r)} - 1)_{1/r}H(P) + 1], \qquad r \neq 1, r > 0 \end{aligned}$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$.

E. Entropy of Order r and Degree s

Sharma and Mittal (1975) introduced and characterized two entropies called *entropy of order 1 and degree s* and *entropy of order r and degree s* given by

$$H_1^s(P) = (2^{1-s} - 1)^{-1}\left[\exp_2\left((s-1)\sum_{i=1}^{n} p_i \log p_i\right) - 1\right], \qquad s \neq 1, s > 0, \tag{5}$$

and

$$H_r^s(P) = (2^{1-s} - 1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^r\right)^{s-1/r-1} - 1\right],$$

$$r \neq 1, s \neq 1, r > 0, s > 0, \tag{6}$$

respectively, for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$.

Sharma and Mittal's main motivation was to generalize the three entropies, $H_r(P)$, $H^s(P)$, and ${}_tH(P)$. With this aim, they arrived at $H_r^s(P)$. $H_r^s(P)$ reduces to $H^s(P)$ and ${}_tH(P)$ when $r = s$ and $r^{-1} = t = (2 - s)$, respectively. $H_r^s(P)$ reduces to $H_1^s(P)$ and $H_r(P)$ when $r \to 1$ and $s \to 1$, respectively. Also, $H_1^s(P)$ reduces to Shannon's entropy, $H(P)$, when $s \to 1$.

Thus, we can see that the entropy of order r and degree s contain, either as a limiting case or as a particular case, Shannon's entropy, the entropy of order r, the entropy of degree s, the entropy of kind t, and the entropy of order 1 and degree s.

The entropies $H_1^s(P)$ and $H_r^s(P)$ are not additive, not recursive, and do not have sum representations.

Before proceeding further, we shall present in the next section a list of most of the generalized entropies known in the literature. For convenient reference, the entropies listed above are also written again.

F. List of Generalized Entropies

For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, the following generalized entropies are known in the literature by their respective authors, starting with Shannon (1948). In some cases it is understood that $P \in \Delta_n$. By no means can we say that the list is complete. At the end of this chapter it is shown in a graphic way, how these entropies reduce to Shannon's case either in the limiting or in the particular case.

Shannon (1948)

$$\phi_1(P) = -\sum_{i=1}^{n} p_i \log p_i$$

Rényi (1961)

$$\phi_2(P) = (1-r)^{-1}\log\left(\sum_{i=1}^{n} p_i^r\right), \qquad r \neq 1, r > 0$$

Aczél and Daróczy (1963)

$$\phi_3(P) = -\frac{\sum_{i=1}^{n} p_i^r \log p_i}{\sum_{i=1}^{n} p_i^r}, \qquad r > 0$$

$$\phi_4(P) = (s-r)^{-1}\log\left\{\frac{\sum_{i=1}^{n} p_i^r}{\sum_{i=1}^{n} p_i^s}\right\}, \qquad r \neq s, r > 0, s > 0$$

$$\phi_5(P) = \frac{1}{s}\arctan\left\{\frac{\sum_{i=1}^{n} p_i^r \sin(s\log p_i)}{\sum_{i=1}^{n} p_i^r \cos(s\log p_i)}\right\}, \qquad s \neq 0, r > 0$$

Varma (1966)

$$\phi_6(P) = \frac{1}{m-r}\log\left(\sum_{i=1}^{n} p_i^{r-m+1}\right), \qquad m-1 < r < m, m \geqq 1$$

$$\phi_7(P) = \frac{1}{m(m-r)}\log\left(\sum_{i=1}^{n} p_i^{r/m}\right), \qquad 0 < r < m, m \geqq 1$$

Kapur (1967)

$$\phi_8(P) = (1-t)^{-1}\log\left(\frac{\sum_{i=1}^{n} p_i^{t+s-1}}{\sum_{i=1}^{n} p_i^s}\right), \qquad t \neq 1, t > 0, s \geqq 1$$

Havrda and Charvát (1967)

$$\phi_9(P) = (2^{1-s}-1)^{-1}\left[\sum_{i=1}^{n} p_i^s - 1\right], \qquad s \neq 1, s > 0$$

Belis and Guiasu (1968)

$$\phi_{10}(P) = -\frac{\sum_{i=1}^{n} p_i w_i \log p_i}{\sum_{i=1}^{n} p_i w_i}, \qquad w_i > 0, i = 1, 2, \ldots, n$$

Rathie (1970)

$$\phi_{11}(P) = (1-r)^{-1}\log\left(\frac{\sum_{i=1}^{n} p_i^{r+s_i-1}}{\sum_{i=1}^{n} p_i^{s_i}}\right),$$

$$s_i \geqq 1,\ i = 1,2,\ldots,n,\ r \neq 1, r > 0$$

Arimoto (1971)

$$\phi_{12}(P) = (2^{t-1}-1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^{1/t}\right)^{t} - 1\right], \qquad t \neq 1, t > 0$$

Sharma and Mittal (1975)

$$\phi_{13}(P) = (2^{1-s}-1)^{-1}\left[\exp_2\left((s-1)\sum_{i=1}^{n} p_i \log p_i\right) - 1\right], \qquad s \neq 1, s > 0$$

$$\phi_{14}(P) = (2^{1-s}-1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^{r}\right)^{s-1/r-1} - 1\right],$$

$$r \neq 1, s \neq 1, r > 0, s > 0$$

Taneja (1975) (refer also Sharma and Taneja, 1975, 1977)

$$\phi_{15}(P) = -2^{r-1}\sum_{i=1}^{n} p_i^{r}\log p_i, \qquad r > 0,$$

$$\phi_{16}(P) = (2^{1-r}-2^{1-s})^{-1}\sum_{i=1}^{n}(p_i^{r}-p_i^{s}), \qquad r \neq s, r > 0, \qquad s > 0$$

$$\phi_{17}(P) = -\frac{2^{r-1}}{\sin s}\sum_{i=1}^{n} p_i^{r}\sin(s\log p_i),$$

$$s \neq k\pi,\ k = 0,1,2,\ldots, r > 0$$

Picard (1979)

$$\phi_{18}(P) = -\frac{\sum_{i=1}^{n} v_i \log p_i}{\sum_{i=1}^{n} v_i},$$

$$\phi_{19}(P) = (1-r)^{-1}\log\left(\frac{\sum_{i=1}^{n} p_i^{r-1}v_i}{\sum_{i=1}^{n} v_i}\right), \qquad r \neq 1, r > 0$$

$$\phi_{20}(P) = (2^{1-s} - 1)^{-1}\left[\left(\frac{\exp_2\left((s-1)\sum_{i=1}^{n} v_i \log p_i\right)}{\sum_{i=1}^{n} v_i}\right) - 1\right], \qquad s \neq 1, s > 0$$

$$\phi_{21}(P) = (2^{1-s} - 1)^{-1}\left[\left(\frac{\sum_{i=1}^{n} p_i^{r-1} v_i}{\sum_{i=1}^{n} v_i}\right)^{s-1/r-1} - 1\right],$$

$$r \neq 1, s \neq 1, r > 0, s > 0$$

where $v_i > 0$, $i = 1, 2, \ldots, n$ and $P = (p_1, p_2, \ldots, p_n) \in \Delta_n$.

Ferreri (1980)

$$\phi_{22}(P) = \left(1 + \frac{1}{\lambda}\right)\log(1+\lambda) - \frac{1}{\lambda}\sum_{i=1}^{n}(1+\lambda p_i)\log(1+\lambda p_i), \qquad \lambda > 0$$

Sant'anna and Taneja (1983)

$$\phi_{23}(P) = -\sum_{i=1}^{n} p_i \log\left(\frac{\sin sp_i}{2\sin\frac{s}{2}}\right), \qquad 0 < s < \pi$$

$$\phi_{24}(P) = -\sum_{i=1}^{n}\left(\frac{\sin sp_i}{2\sin\frac{s}{2}}\right)\log\left(\frac{\sin sp_i}{2\sin\frac{s}{2}}\right), \qquad 0 < s < \pi$$

and

$$\phi_{25}(P) = \sum_{i=1}^{n}\left(\frac{\sin sp_i}{2\sin\frac{s}{2}}\right), \qquad 0 < s < \pi.$$

The last entropy, $\phi_{25}(P)$ does not reduce to Shannon's entropy but has some properties similar to Shannon's entropy (Sant'anna and Taneja, 1985).

G. *Unified (r, s)-Entropy*

The reduction of $H_r^s(P)$ to ${}_tH(P)$ is obtained by substituting $r = 1/t$ and $s = 2 - t$. As $s > 0$, this implies that $0 < t < 2$. This means that ${}_tH(P)$ is contained in $H_r^s(P)$ only for $0 < t < 2$. To avoid this problem, let us relax the condition of the positivity of s, i.e., we shall take any s. Instead of studying the properties of the generalized entropies given in Eqs. (2)–(6) individually, our aim here is to study them in a unified way. This unification is as follows:

$$\mathscr{E}_r^s(P) = \begin{cases} H_r^s(P), & r \neq 1, s \neq 1, r > 0, \\ H_1^s(P), & r = 1, s \neq 1, \\ H_r^1(P), & r \neq 1, s = 1, r > 0, \\ H(P), & r = 1, s = 1, \end{cases} \tag{7}$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, where $H_r^1(P)$ is the same as $H_r(P)$ given in Eq. (2). From now onwards we shall use the notation $H_r^1(P)$ instead of $H_r(P)$.

The entropies $H^s(P)$ and ${}_tH(P)$ do not appear in the unified expression given in Eq. (7) because they are a particular case of $H_r^s(P)$ and hence are already contained in it. According to the notations above, $H^s(P)$ given in Eq. (3) means $H_s^s(P)$ for $r = s$ in $H_r^s(P)$. Hence forth, we shall consider this notation too.

We call the unified expression (7), i.e., $\mathscr{E}_r^s(P)$, as *unified* (r, s)*-entropy*.

As specified before, our aim here is to study some important properties of the unified (r, s)-entropy. Before proceeding further, we shall first present some definitions and composition relations used in the subsequent sections.

Definition 1. A numerical function $\theta: \Delta_n^0 \to \mathscr{R}$ (reals) is *concave* in Δ_n^0 if for all $P, U \in \Delta_n^0$, we have

$$\theta(\lambda P + \mu U) \geqq \lambda\theta(P) + \mu\theta(U),$$

where $\lambda + \mu = 1, \lambda > 0, \mu > 0$.

For the *convex* functions the above inequality is reversed.

Definition 2. A numerical differentiable function $\theta: \Delta_n^0 \to \mathscr{R}$ (reals) is *pseudoconcave* in Δ_n^0 if for all $P, U \in \Delta_n^0$, we have

$$\nabla\theta(P)(U - P) \leqq 0 \text{ implies } \theta(U) \leqq \theta(P),$$

where ∇ represents the gradient operator.

For the *pseudoconvexity*, we have

$$\nabla\theta(P)(U - P) \geqq \text{ implies } \theta(U) \geqq \theta(P),$$

for all $P, U \in \Delta_n^0$.

Definition 3. A numerical differentiable function $\theta: \Delta_n^0 \to \mathscr{R}$ (reals) is *quasiconcave* in Δ_n^0 if for all $P, U \in \Delta_n^0$, we have

$$\theta(U) \geqq \theta(P) \text{ implies } \nabla\theta(P)(U - P) \geqq 0.$$

For the *quasiconvexity* we have

$$\theta(U) \leqq \theta(P) \text{ implies } \nabla\theta(P)(U - P) \leqq 0,$$

for all $P, U \in \Delta_n^0$.

Definition 4. Schur Concavity. Before giving the definition of Schur concavity, first we shall define the concept of *majorization.*

Majorization. For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$ and $U = (u_1, u_2, \ldots, u_n) \in \Delta_n^0$, we say that *P is majorized by U*, i.e., $P \underset{m}{\prec} U$ if

$$p_1 \geqq p_2 \geqq \cdots \geqq p_n; \qquad u_1 \geqq u_2 \geqq \cdots \geqq u_n,$$

with

$$\sum_{i=1}^{\sigma} p_i \leqq \sum_{i=1}^{\sigma} u_i, \qquad 1 \leqq \sigma \leqq n,$$

or equivalently, there exists a doubly stochastic matrix $\{c_{ik}\}$, $c_{ik} \geqq 0$, $i, k = 1, 2, \ldots, n$ with $\Sigma_{i=1}^n c_{ik} = \Sigma_{k=1}^n c_{ik} = 1$ such that

$$p_i = \sum_{k=1}^{n} c_{ik} u_k, \qquad i = 1, 2, \ldots, n.$$

Schur Concavity. A numerical function θ: $\Delta_n^0 \to \mathscr{R}$ (reals) is *Schur concave* in Δ_n^0 if $P \underset{m}{\prec} U$, i.e., P *is majorized by* U in Δ_n^0 implies $\theta(P) \geq \theta(U)$ for all $P, U \in \Delta_n^0$.

For the *Schur convexity*, for all $P, U \in \Delta_n^0$, we have

$$P \underset{m}{\prec} U \text{ implies } \theta(P) \leq \theta(U).$$

Definition 1 is already known in the literature. For definitions 2 and 3 refer to Mangasarian (1961). For definition 4 refer to Marshall and Olkin (1979) or to Hardy et al. (1934).

Let

$$g_s(x) = (2^{1-s} - 1)^{-1}[2^{(1-s)x} - 1], \qquad s \neq 1 \tag{8}$$

be a function defined for all $x \geqq 0$. Then we can write

$$H_r^s(P) = g_s[H_r^1(P)], \tag{9}$$

and

$$H_1^s(P) = g_s[H(P)], \tag{10}$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$.

The following proposition holds:

Proposition 2.1. We have

(i) $\lim_{s \to 1} g_s(x) = x$.

(ii) $g_s(x) \geqq 0$ for all $x \geqq 0$ and any s.

(iii) g_s is an increasing function of x.

(iv) g_s is a convex function of x for $s < 1$.

(v) g_s is a concave function of x for $s > 1$.

(vi) $g_s(x)\begin{cases} \geqq N(s) \cdot x, & s < 1, \\ \leqq N(s) \cdot x, & s > 1, \end{cases}$

where

$$N(s) = \frac{(1-s)\ln 2}{2^{1-s} - 1}\begin{cases} < 1, & s < 1, \\ = 1, & s = 1, \\ > 1, & s > 1, \end{cases} \tag{11}$$

(vii) $g_s(x)\begin{cases} \geqq x, & (x \geqq 1, s < 1) \quad \text{or} \quad (0 \leqq x \leqq 1, s > 1), \\ \leqq x, & (0 \leqq x \leqq 1, s < 1) \quad \text{or} \quad (x \geqq 1, s > 1). \end{cases}$

Proof. Parts (i)–(v) are easy verifications.

(vi) It follows from the known result (Hardy et al., 1934, pp. 106, Theorem 150),

$$\ln v \leqq v - 1, \qquad v > 0,$$

where we substitute $v = 2^{(1-s)x}$.

(vii) It follows from the result (Hardy et al., 1934, pp. 40, Theorem 42),

$$v^\gamma - 1\begin{cases} \geqq \gamma(v-1), & \gamma \geqq 1, \\ \leqq \gamma(v-1), & 0 \leqq \gamma \leqq 1, \end{cases}$$

$v \geqq 0$, where we substitute $v = 2^{1-s}$ and $\gamma = x$.

H. Analytic and Algebraic Properties of Unified (r, s)-Entropy

In this subsection, we shall study some properties of unified (r, s)-entropy given in Eq. (7). Some of these properties can be seen in Capocelli and Taneja (1985). Unless otherwise specified, it is understood that the results given below are true for all $r > 0$ and any s.

Property 1. Nonnegativity. For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, $\mathscr{E}_r^s(P) \geqq 0$.

Property 2. Continuity. For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, $\mathscr{E}_r^s(P)$ is a continuous function of P.

Property 3. *Symmetry.* For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, $\mathscr{E}_r^s(P)$ is a symmetric function of its arguments, i.e.,

$$\mathscr{E}_r^s(p_1, p_2, \ldots, p_n) = \mathscr{E}_r^s(p_{\tau(1)}, p_{\tau(2)}, \ldots, p_{\tau(n)}),$$

where τ is an arbitrary permutation of $\{1, 2, \ldots, n\}$.

Property 4. *Normality.* $\mathscr{E}_r^s(\frac{1}{2}, \frac{1}{2}) = 1$.

Property 5. *Nonadditivity.* For all

$$P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0, \; Q = (q_1, q_2, \ldots, q_m) \in \Delta_m^0, \quad \text{and}$$

$$P*Q = (p_1q_1, p_1q_2, \ldots, p_1q_m, p_2q_1, p_2q_2, \ldots,$$
$$p_2q_m, \ldots, p_nq_1, p_nq_2, \ldots, p_nq_m) \in \Delta_{nm}^0,$$

we have

$$\mathscr{E}_r^s(P*Q) = \mathscr{E}_r^s(P) + \mathscr{E}_r^s(Q) + (2^{1-s} - 1) \; \mathscr{E}_r^s(P) \cdot \mathscr{E}_r^s(Q).$$

These properties are easily verified.

Property 6. *Limiting case.* For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, we have

$$\lim_{r \to \infty} \mathscr{E}_r^s(P) = \begin{cases} (2^{1-s} - 1)^{-1}[p_{\max}^{s-1} - 1], & s \neq 1, \\ -\log p_{\max}, & s = 1. \end{cases}$$

For $s = 1$, i.e., $\lim_{r \to \infty} H_r^1(P) = -\log p_{\max}$ can be seen in Shiva et al. (1973). For $s \neq 1$, the result can be proved by the composition relation given in Eq. (9).

Property 7. *Monotonicity.* For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, $\mathscr{E}_r^s(P)$ is a decreasing function of r (s fixed).

For $s = 1$, refer to Shiva et al. (1973). For $s \neq 1, s > 0$, refer to Sharma and Mittal (1975). While the extension to $s \leqq 0$ is an easy verification.

Property 8. *Concavity.* For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, $\mathscr{E}_r^s(P)$ is a concave function of P for all $(r, s) \in \Gamma_1$, where

$$\Gamma_1 = \left\{(r, s)\colon r > 0 \text{ with } s \geqq r \text{ or } s \geqq 2 - \frac{1}{r}\right\}. \tag{12}$$

It is already known that $H_s^s(P)$ and ${}_tH(P)$ are concave functions of P for all the values of the parameters. $H_r^1(P)$ is a concave function of P for $0 < r < 1$. The concavity of $H(P)$ is already known. The concavity of $H_1^s(P)$ for $s > 1$ is a direct consequence of the concavity of $H(P)$ because of relation (10) and propositions 2.1(iii) and 2.1(v). The concavity of $H_r^s(P)$ for $s \geqq 2 - 1/r, r \neq 1, s \neq 1, r > 0$ can be seen in Van der Pyl (1977). The concavity of $H_r^s(P)$ for $s \geqq r > 0$ follows on the lines of proposition 4.2(ii), and it can be seen in Taneja (1988c). Thus combining all these we get the required result.

Property 9. Pseudoconcavity. For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, $\mathscr{E}_r^s(P)$ is a pseudoconcave function of P for all $r > 0$ and for any s.

Pseudoconcavity of $H(P)$ follows from the concavity of $H(P)$, because every concave function is pseudoconcave. For pseudoconcavity of $H_r^1(P)$, $r \neq 1, r > 0$ refer to Ben-Bassat and Raviv (1978). The pseudoconcavity of $H_r^s(P)\,(r \neq 1, s \neq 1, r > 0)$ and $H_1^s(P)$ follows from proposition 2.1(iii) with the composition relations (9) and (10), respectively.

Property 10. Quasiconcavity. For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, $\mathscr{E}_r^s(P)$ is a quasiconcave function of P for all $r > 0$ and any s.

The proof follows from property 9, because every pseudoconcave function is quasiconcave (Mangasarian, 1961, pp. 143, Theorem 5).

Property 11. Schur concavity. For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, $\mathscr{E}_r^s(P)$ is a Schur-concave function of P for all $r > 0$ and any s.

Proof. In case of Shannon's entropy, $H(P)$, the result is already known (Csiszár and Körner, 1961). In view of relations (9), (10), and proposition 2.1(iii), it is sufficient to prove the result only for $H_r^1(P)$ $(r \neq 1, r > 0)$.

Let $P, U \in \Delta_n^0$ such that

$$p_i = \sum_{k=1}^{n} c_{ik} u_k,$$

and

$$\sum_{i=1}^{n} c_{ik} = \sum_{k=1}^{n} c_{ik} = 1, \qquad c_{ik} \geqq 0,\ i, k = 1, 2, \ldots, n.$$

We can write

$$\begin{aligned} H_r^1(P) &= (1-r)^{-1} \log\left(\sum_{i=1}^{n} p_i^r\right) \\ &= (1-r)^{-1} \log\left[\sum_{i=1}^{n} \left(\sum_{k=1}^{n} c_{ik} u_k\right)^r\right]. \end{aligned} \tag{13}$$

We know that (Gallager, 1968, pp. 523)

$$\left(\sum_{k=1}^{n} c_{ik}u_k\right)^r \begin{cases} \leqq \sum_{k=1}^{n} c_{ik}u_k^r, & r > 1, \\ \geqq \sum_{k=1}^{n} c_{ik}u_k^r, & 0 < r < 1 \end{cases} \tag{14}$$

for all $i = 1, 2, \ldots, n$. Taking log(.) on both sides of Eqs. (14), multiplying by $(1-r)^{-1}(r \neq 1)$, and using the expression (13), we get

$$H_r^1(P) \geqq H_r^1(U), \qquad r \neq 1, r > 0.$$

This completes the proof.

Property 12. Maximality. For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, $\mathscr{E}_r^s(P)$ is maximum when the probability distribution is uniform, i.e.,

$$0 \leqq \mathscr{E}_r^s(P) \leqq \mathscr{E}_r^s\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right), \qquad n \geqq 2$$

for all $r > 0$ and any s.

The proof follows from the property 11 (Marshall and Olkin, 1979, pp. 7).

I. Inequalities and Bounds on Generalized Entropies

In this subsection we shall provide some inequalities involving generalized entropies. Upper and lower bounds on the unified (r, s)-entropy in terms of maximum probability are given. Some bounds on the entropy series in the case of Shannon's entropy are also given.

Inequality 1. Inequalities among entropies. For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, we have

$$\text{(i)}\ H_r^s(P) \begin{cases} \leqq H_1^s(P), & r > 1, s \neq 1, \\ \geqq H_1^s(P), & 0 < r < 1, s \neq 1, \end{cases}$$

$$\text{(ii)}\ H_r^1(P) \begin{cases} \leqq H(P), & r > 1, \\ \geqq H(P), & 0 < r < 1, \end{cases}$$

$$\text{(iii)}\ H_r^s(P) \begin{cases} \geqq N(s) \cdot H_r^1(P), & s < 1, \\ \leqq N(s) \cdot H_r^1(P), & s > 1, \end{cases}$$

for all $r \neq 1, r > 0$.

$$\text{(iv)}\quad H_1^s(P)\begin{cases}\geqq N(s)\cdot H(P), & s>1,\\ \leqq N(s)\cdot H(P), & s<1,\end{cases}$$

where $N(s)$ appearing in (iii) and (iv) is given by Eq. (11).

$$\text{(v)}\quad H_r^s(P)\begin{cases}\geqq H_r^1(P), & (H_r^1(P)\geqq 1, s<1) \quad \text{or} \quad (H_r^1(P)\leqq 1, s>1),\\ \leqq H_r^1(P), & (H_r^1(P)\leqq 1, s<1) \quad \text{or} \quad (H_r^1(P)\geqq 1, s>1)\end{cases}$$

for all $r \neq 1, r > 0$.

$$\text{(vi)}\quad H_1^s(P)\begin{cases}\geqq H(P), & (H(P)\geqq 1, s<1) \quad \text{or} \quad (H(P)\leqq 1, s>1),\\ \leqq H(P), & (H(P)\leqq 1, s<1) \quad \text{or} \quad (H(P)\geqq 1, s>1).\end{cases}$$

The proof of parts (i) and (ii) follows from property 7. Parts (iii) and (iv) follows from proposition 2.1(vi). Parts (v) and (vi) follow from proposition 2.1(vii).

Inequality 2. Order preserving. For all $P=(p_1,p_2,\ldots,p_n)\in\Delta_n^0$, $Q=(q_1,q_2,\ldots,q_m)\in\Delta_m^0$, if

$$H(P) > H(Q),$$

then

$$\mathscr{E}_r^s(P) > \mathscr{E}_r^s(Q),$$

under the following conditions:

$$\text{(i)}\quad H(Q) < -\log p_{\max},$$

and

$$\text{(ii)}\quad \log m < H(P).$$

For the proof for $s = 1$, i.e., in case of entropy of order r, $H_r^1(P)(r \neq 1, r > 0)$ refer to Shiva et al. (1973). The other cases follow from relations (9) and (10), and from proposition 2.1(iii).

Inequality 3. Bounds on $\mathscr{E}_r^s$(P). For all $P=(p_1,p_2,\ldots,p_n)\in\Delta_n^0$, we have

$$\text{(i)}\quad \mathscr{E}_r^s\left(\sum_{i=1}^{\sigma} p_i, 1-\sum_{i=1}^{\sigma} p_i\right) \leqq \mathscr{E}_r^s(P)$$

$$\leqq \mathscr{E}_r^s\Bigg(p_1,p_2,\ldots,p_\sigma,\underbrace{\frac{1-\sum_{i=1}^{\sigma}p_i}{n-\sigma},\ldots,\frac{1-\sum_{i=1}^{\sigma}p_i}{n-\sigma}}_{(n-\sigma)\text{ times}}\Bigg), \qquad 1\leqq\sigma\leqq n.$$

(ii) $\mathscr{E}_r^s(p_{\max}, 1 - p_{\max}) \leqq \mathscr{E}_r^s(P)$

$$\leqq \mathscr{E}_r^s\left(\frac{1-p_{\max}}{n-1}, \ldots, \frac{1-p_{\max}}{n-1}, p_{\max}\right).$$

(iii) $1 - p_{\max} \leqq \frac{1}{2}\mathscr{E}_r^s(P)$. (15)

Inequalities 3.(i) and 3.(ii) are true for all $r > 0$ and any s, while inequality 3.(iii) is true for $(r, s) \in \Gamma_2$, where

$$\Gamma_2 = \left\{(r,s)\colon \left(\frac{1}{n} < p_{\max} \leqq \frac{1}{2}, s \leqq 2\right) \text{ or } p_{\max} \geqq \frac{1}{2}, \left(s \geqq 2 \text{ or } (r,s) \in \Gamma_1\right)\right\}, \quad (16)$$

for all $r > 0$.

Proof.

(i) In case of Shannon's entropy, the left-hand side of the inequality can be proved by recursivity property and the right-hand side follows from Jensen's inequality (McEliece, 1977). In view of relations (9) and (10), and proposition 2.1(iii), it is sufficient to prove the result only for the entropy of order r, i.e., we need to show that

$$H_r^1\left(\sum_{i=1}^{\sigma} p_i, 1 - \sum_{i=1}^{\sigma} p_i\right) \leqq H_r^1(P)$$

$$\leqq H_r^1\Bigg(p_1, p_2, \ldots, p_\sigma, \underbrace{\frac{1-\sum_{i=1}^{\sigma} p_i}{n-\sigma}, \ldots, \frac{1-\sum_{i=1}^{\sigma} p_i}{n-\sigma}}_{(n-\sigma)\text{ times}}\Bigg), \qquad 1 \leqq \sigma \leqq n. \quad (17)$$

We know that (Gallager, 1968, pp. 523)

$$\sum_{i=\sigma+1}^{n} \frac{1}{n-\sigma} p_i^r \begin{cases} \leqq \left(\sum_{i=\sigma+1}^{n} \frac{1}{n-\sigma} p_i\right)^r, & 0 < r < 1, \\ \geqq \left(\sum_{i=\sigma+1}^{n} \frac{1}{n-\sigma} p_i\right)^r, & r > 1, \end{cases}$$

i.e.,

$$\sum_{i=\sigma+1}^{n} p_i^r \begin{cases} \leqq (n-\sigma)\left(\sum_{i=\sigma+1}^{n} \frac{1}{n-\sigma} p_i\right)^r, & 0 < r < 1, \\ \geqq (n-\sigma)\left(\sum_{i=\sigma+1}^{n} \frac{1}{n-\sigma} p_i\right)^r, & r > 1, \end{cases}$$

i.e.,

$$\sum_{i=1}^{n} p_i^r = \sum_{i=1}^{\sigma} p_i^r + \sum_{i=\sigma+1}^{n} p_i^r$$

$$\begin{cases} \leqq \sum_{i=1}^{\sigma} p_i^r + (n-\sigma)\left(\dfrac{1-\sum_{i=1}^{\sigma} p_i}{n-\sigma}\right)^r, & 0<r<1, \\ \geqq \sum_{i=1}^{\sigma} p_i^r + (n-\sigma)\left(\dfrac{1-\sum_{i=1}^{\sigma} p_i}{n-\sigma}\right)^r, & r>1, \end{cases} \tag{18}$$

where $1 \leqq \sigma \leqq n$. Taking $\log(\cdot)$ on both sides of Eq. (18), multiplying by $(1-r)^{-1}(r \neq 1)$ and simplifying, we get the right-hand side of inequality (17). Let us now prove the left-hand side. Again, we know that (Gallager, 1968, pp. 523)

$$\sum_{i=1}^{\sigma} p_i^r \begin{cases} \leqq \left(\sum_{i=1}^{\sigma} p_i\right)^r, & r>1, \\ \geqq \left(\sum_{i=1}^{\sigma} p_i\right)^r, & 0<r<1. \end{cases} \tag{19}$$

Similarly,

$$\sum_{i=\sigma+1}^{n} p_i^r \begin{cases} \leqq \left(\sum_{i=\sigma+1}^{n} p_i\right)^r, & r>1, \\ \geqq \left(\sum_{i=\sigma+1}^{n} p_i\right)^r, & 0<r<1, \end{cases} \tag{20}$$

where $1 \leqq \sigma \leqq n$ in Eqs. (29) and (20). Adding Eqs. (19) and (20), we get

$$\sum_{i=1}^{\sigma} p_i^r + \sum_{i=\sigma+1}^{n} p_i^r \begin{cases} \leqq \left(\sum_{i=1}^{\sigma} p_i\right)^r + \left(\sum_{i=\sigma+1}^{n} p_i\right)^r, & r>1, \\ \geqq \left(\sum_{i=1}^{\sigma} p_i\right)^r + \left(\sum_{i=\sigma+1}^{n} p_i\right)^r, & 0<r<1. \end{cases} \tag{21}$$

Taking $\log(\cdot)$ on both sides of Eq. (21), multiplying by $(1-r)^{-1}(r \neq 1)$, and simplifying, we get the left-hand side of Eq. (17).

(ii) Without loss of generality we can suppose that $p_n = p_{\max}$. Let $\sigma = n-1$, then $1 - p_{\max} = 1 - p_n = 1 - \Sigma_{i=1}^{n-1} p_i = 1 - \Sigma_{i=1}^{\sigma} p_i$. Making these substitutions in part (i), we get the required result.

(iii) The proof of this part is divided into two parts.

First part. In this part we will show that

$$1 - p_{\max} \leqq \frac{1}{2}\mathscr{E}^s_r(P), \qquad \left(\frac{1}{n} < p_{\max} \leqq \frac{1}{2}, s \leqq 2\right)$$

$$\text{or} \quad \left(p_{\max} \geqq \frac{1}{2}, s \geqq 2\right) \tag{22}$$

for all $r > 0$.

Consider a function

$$\xi_s(p) = 1 - p - \frac{1}{2}h_s(p), \qquad 0 < p \leqq 1,$$

where

$$h_s(p) = \begin{cases} (2^{1-s} - 1)^{-1}[p^{s-1} - 1], & s \neq 1, \\ -\log p, & s = 1. \end{cases}$$

Then

$$\xi'_s(p) = -1 - \frac{1}{2}h'_s(p),$$

and

$$\xi''_s(p) = -\frac{1}{2}h''_s(p),$$

where

$$h'_s(p) = \begin{cases} (2^{1-s} - 1)^{-1}(s - 1)p^{s-2}, & s \neq 1, \\ -\dfrac{1}{\ln 2}\dfrac{1}{p}, & s = 1, \end{cases}$$

and

$$h''_s(p) = \begin{cases} (2^{1-s} - 1)^{-1}(s - 1)(s - 2)p^{s-3}, & s \neq 1, \\ \dfrac{1}{\ln 2}\dfrac{1}{p^2}, & s = 1. \end{cases}$$

Also,

$$\xi_s\left(\frac{1}{2}\right) = \xi_s(1) = 0.$$

If $s < 2$, $\xi_s''(p) < 0$ for all $p \in (0, 1]$. This implies that the function $\xi_s(p)$ is strictly concave and attains its single maximum at $\xi_s'(p) = 0$, i.e., when

$$p = \begin{cases} [2(1-s)^{-1}(2^{1-s}-1)]^{1/s-2}, & s \neq 1, \\ \dfrac{1}{2\ln 2}, & s = 1. \end{cases}$$

Thus the only zeros of $\xi_s(p)$ are when $p = 1/2$ or $p = 1$. As $p \to 0$, $\xi_s(p) \to -\infty (s \leqq 1)$. Thus for $s < 2$, we have

$$\xi_s(p) = \begin{cases} \leqq 0, & \text{if } 0 < p \leqq \dfrac{1}{2}, \\ \geqq 0, & \text{if } \dfrac{1}{2} \leqq p \leqq 1. \end{cases}$$

Similarly for $s > 2$, we have

$$\xi_s(p) = \begin{cases} \geqq 0, & \text{if } 0 < p \leqq \dfrac{1}{2}, \\ \leqq 0, & \text{if } \dfrac{1}{2} \leqq p \leqq 1. \end{cases}$$

For $s = 2$, we have

$$\xi_s(p) = 0, \qquad 0 < p \leqq 1.$$

Finally,

$$\xi_s(p) \leqq 0, \qquad \text{if} \left(0 < p \leqq \frac{1}{2}, s \leqq 2\right) \quad \text{or} \quad \left(\frac{1}{2} \leqq p \leqq 1, s \geqq 2\right). \tag{23}$$

We have an equality sign in Eq. (23) when $s = 2$, $p = 1/2$, or $p = 1$. If we replace p by $p_{\max} = \max\{p_1, p_2, \ldots, p_n\}$ in Eq. (23), we have

$$2(1 - p_{\max}) \leqq \begin{cases} (2^{1-s}-1)^{-1}[p_{\max}^{s-1} - 1], & s \neq 1, \\ -\log p_{\max}, & s = 1 \end{cases} \tag{24}$$

for all $s \leqq 2$, $1/n \leqq p_{\max} \leqq 1/2$ or $s \geqq 2$, $p_{\max} \geqq 1/2$. Expression (24) and property 6 together give

$$1 - p_{\max} \leqq \frac{1}{2}\mathscr{E}_\infty^s(P), \left(s \leqq 2, \frac{1}{n} < p_{\max} \leqq \frac{1}{2}\right) \quad \text{or} \quad \left(s \geqq 2, \frac{1}{2} \leqq p_{\max} \leqq 1\right), \tag{25}$$

where

$$\mathscr{E}_\infty^s(P) = \lim_{r \to \infty} \mathscr{E}_r^s(P).$$

By property 7, we can write

$$\mathscr{E}^s_\infty(P) \leqq \mathscr{E}^s_r(P), \qquad 0 < r < \infty. \tag{26}$$

Expressions (25) and (26) together give (22).

Second part. In this part we shall show that

$$1 - p_{\max} \leqq \frac{1}{2}\mathscr{E}^s_r(P), \qquad p_{\max} \geqq \frac{1}{2}, \qquad (r, s) \in \Gamma_1, \tag{27}$$

where Γ_1 is as given in Eq. (12).

In order to prove this we shall use the following properties:

(e_1) $\mathscr{E}^s_r(p, 1-p)$ is a continuous function of $p \in [0, 1]$.

(e_2) $\mathscr{E}^s_r(p, 1-p)$ is a concave function of $p \in [0, 1]$ for $(r, s) \in \Gamma_1$.

(e_3) $\frac{1}{2}\mathscr{E}^s_r\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2}$.

(e_4) $\mathscr{E}^s_r(1, 0) = \mathscr{E}^s_r(0, 1) = 0$.

(e_5) $\mathscr{E}^s_r(1 - p_{\max}, p_{\max}) \leqq \mathscr{E}^s_r(P), \qquad P \in \Delta^0_n$.

We know that the graph of $1 - \max\{p, 1-p\}$, $0 \leqq p \leqq 1$ contains two straight lines between $(0,0)$ and $(\frac{1}{2}, \frac{1}{2})$ and between $(\frac{1}{2}, \frac{1}{2})$ and $(0,0)$. Thus using (e_1)–(e_4), we have

$$1 - \max\{p, 1-p\} \leqq \frac{1}{2}\mathscr{E}^s_r(p, 1-p), \qquad \max\{p, 1-p\} \geqq \frac{1}{2}, \qquad (r, s) \in \Gamma_1. \tag{28}$$

Substituting $\max\{p, 1-p\}$ with $\max\{p_1, p_2, \ldots, p_n\} = p_{\max}$ in Eq. (28) and using (e_5) we get Eq. (27).

Thus, combining Eqs. (22) and (27) we get the required result.

Inequality 4. Generalized Shannon's or Gibb's inequalities. For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n$ and $U = (u_1, u_2, \ldots, u_n) \in \Delta_n$, we have

$$\mathscr{E}^s_r(P) \leqq {}^\alpha\mathscr{E}^s_r(P||U), \qquad \alpha = 1 \text{ and } 2, \tag{29}$$

where

$${}^\alpha\mathscr{E}^s_r(P||U) = \begin{cases} (2^{1-s} - 1)^{-1}[{}^\alpha M_r(P||U)^{r-1/s-1} - 1], & r \neq 1, s \neq 1, r > 0, \\ (2^{1-s} - 1)^{-1}[2^{(s-1)H(P||U)} - 1], & r = 1, s \neq 1, \\ (1-r)^{-1}\log[{}^\alpha M_r(P||U)], & r \neq 1, s = 1, r > 0, \\ H(P||U), & r = 1, s = 1 \end{cases}$$

for $\alpha = 1$ and 2, with

$$^1M_r(P||U) = \frac{\sum_{i=1}^{n} p_i^r}{\sum_{i=1}^{n} p_i^r u_i^{1-r}}, \qquad r > 0,$$

$$^2M_r(P||U) = \left(\sum_{i=1}^{n} p_i u_i^{r-1/r}\right)^r, \qquad r > 0,$$

and

$$H(P||U) = -\sum_{i=1}^{n} p_i \log u_i. \tag{30}$$

Proof. Nath (1975) and Van der Lubbe (1978) proved the following inequalities:

$$H_r^1(P) \leqq {}^\alpha H_r(P||U), \qquad r \neq 1, r > 0 \tag{31}$$

for all $P, U \in \Delta_n$, $\alpha = 1$ and 2, where

$$^1H_r(P||U) = (1-r)^{-1} \log\left(\frac{\sum_{i=1}^{n} p_i^r}{\sum_{i=1}^{n} p_i^r u_i^{1-r}}\right), \qquad r \neq 1, r > 0 \tag{32}$$

and

$$^2H_r(P||U) = \frac{r}{1-r} \log\left(\sum_{i=1}^{n} p_i u_i^{r-1/r}\right), \qquad r \neq 1, r > 0. \tag{33}$$

In the limiting case we have

$$\lim_{r \to 1} {}^1H_r(P||U) = \lim_{r \to 1} {}^2H_r(P||U) = H(P||U),$$

where $H(P||U)$ given in Eq. (30) is the well-known inaccuracy measure (Kerridge, 1961). In this case, we know that

$$H(P) \leqq H(P||U) \tag{34}$$

for all $P, U \in \Delta_n$ is the well-known Shannon's or Gibb's inequality. The remaining part of the proof follows from relations (9) and (10), and proposition 2.1(iii) applied to Eqs. (31) and (34).

Inequality 4.a. The following inequality holds:

$$^1H_r(P||U) \leqq {}^2H_r(P||U) + D_r^1(P||U), \qquad r \neq 1, r > 0 \tag{35}$$

for all $P, U \in \Delta_n$, where ${}^1H_r(P||U)$ and ${}^2H_r(P||U)$ are given in Eqs. (32) and (33), respectively, and

$$D_r^1(P||U) = (r-1)^{-1}\log\left(\sum_{i=1}^{n} p_i^r u_i^{1-r}\right), \qquad r \neq 1, r > 0$$

is the directed divergence of order r (Rényi, 1961) given in Section IV.

Proof. We know that [Van der Lubbe (1978)]

$$\sum_{i=1}^{n} p_i^r \begin{cases} \geqq \left[\sum_{i=1}^{n} p_i u_i^{r-1/r}\right]^r, & r > 1, \\ \leqq \left[\sum_{i=1}^{n} p_i u_i^{r-1/r}\right]^r, & 0 < r < 1, \end{cases}$$

i.e.,

$$\frac{\sum_{i=1}^{n} p_i^r}{\sum_{i=1}^{n} p_i^r u_i^{1-r}} \begin{cases} \geqq \dfrac{\left[\sum_{i=1}^{n} p_i u_i^{r-1/r}\right]^r}{\sum_{i=1}^{n} p_i^r u_i^{1-r}}, & r > 1, \\ \leqq \dfrac{\left[\sum_{i=1}^{n} p_i u_i^{r-1/r}\right]^r}{\sum_{i=1}^{n} p_i^r u_i^{1-r}}, & 0 < r < 1. \end{cases} \tag{36}$$

Taking log(·) on both sides of Eq. (36), multiplying by $(1-r)^{-1}(r \neq 1)$, and simplifying we get the required result.

Inequality 5. Bounds on the entropy series. Let $P_\infty = (p_1, p_2, \ldots)$ be a sequence of probability distribution such that $p_n \geqq 0$, $n \geqq 1$, $\Sigma_{n=1}^{\infty} p_n = 1$, with $p_n \geq p_{n+1}, + n$. It is well known (Wyner, (1973) that the entropy series $H(P)$ given by

$$H(P_\infty) = -\sum_{n=1}^{\infty} p_n \log p_n,$$

converges if and only if the series

$$S(P_\infty) = \sum_{n=1}^{\infty} p_n \log n$$

converges. Moreover, the following inequalities (Capocelli, Santis, and Taneja, 1988) hold:

(i) $H(P_\infty) \geqq S(P_\infty) \geqq 0$,

(ii) $H(P_\infty) \leqq L(P_\infty)$,

where

$$L(P_\infty) = S(P_\infty) + \sqrt{S(P_\infty)} + \log[1 + \sqrt{S(P_\infty)}].$$

(iii) $H(P_\infty) \leqq W_k(P_\infty), k = 1, 2, \ldots,$

where

$$W_k(P_\infty) = \sum_{i=0}^{k} \log^i S(P_\infty) + \log^k S(P_\infty) + \beta_k,$$

and

$$\beta_k = \log\left\{\sum_{i=1}^{\infty} \exp_2\left[-\left(\log^{k+1} n + \sum_{i=1}^{k+1} \log^i n\right)\right]\right\} + 0.766k + 8.531$$

is a constant independent of the probability distribution P_∞, and for $x \geqq 0$, $i = 1, 2, \ldots.$

$$\log^0 x = x,$$
$$\log^i x = \begin{cases} \log(\log^{i-1} x), & \text{if } \log^{i-1} x \geqq 1 \\ 0, & \text{otherwise.} \end{cases}$$

(iv) $H(P_\infty) \leqq V(P_\infty)$,

where

$$V(P_\infty) = S(P_\infty) + \log^* S(P_\infty) + 2 + \log 4.26,$$

and

$$\log^* x = \begin{cases} 0, & 0 < x \leqq 1 \\ \log x + \log^*(\log x), & x > 1 \end{cases}$$

i.e.,

$$\log^* x = \log x + \log(\log x) + \log[\log(\log x)] + \cdots$$

with addends all positive.

(v) $\lim_{S(P) \to \infty} [L(P_\infty) - W_1(P_\infty)] = \infty.$

(vi) $\lim_{S(P) \to \infty} [W_k(P_\infty) - W_{k+1}(P_\infty)] = \infty.$

(vii) $\lim_{S(P) \to \infty} [W_k(P_\infty) - V(P_\infty)] = \infty.$

III. Generalized Distance Measures

The quantity $(\sum_{i=1}^{n} p_i^r)^{s-1/r-1}$, $r > 0$ plays an important role in the entropy of order r and degree s. Let us write it in a simplified form, given by

$$G_r^\rho(P) = \left(\sum_{i=1}^{n} p_i^r\right)^\rho, \qquad r > 0, \rho \neq 0, \tag{37}$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$.

The quantity $G_r^\rho(P)$ given in Eq. (37) is either called the *generalized distance measure* (Boekee and Van der Lubbe, 1979; Capocelli et al., 1985) or the *generalized certainty measure* (Van der Lubbe et al., 1984).

Another *generalized distance measure* considered by Capocelli et al. (1985) is given by

$$T_r^\rho(P) = \left(\frac{\sum_{i=1}^{n} p_i^r}{\sum_{i=1}^{n} p_i^\rho}\right)^{1/r-\rho}, \qquad r \neq \rho, r \geqq 0, \rho \geqq 0 \tag{38}$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$.

The quantities (37) and (38), in particular, contain the distance measures considered by Trouborst et al. (1974), Györfi and Nemetz (1975), Devijver (1974), and Vajda (1968).

The generalized distance measures (37) and (38) satisfy some properties (Capocelli et al., 1985) given in the following two propositions:

Proposition 3.1. For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$, we have

(e_1) (i) $G_r^\rho(P)$ is a convex function of P for $r > 1$, $r\rho \geqq 1$ or $0 < r < 1$, $\rho < 0$.

(ii) $G_r^\rho(P)$ is a concave function of P for $0 < r < 1$, $\rho > 0$, $r\rho \leqq 1$.

(e_2) (i) $G_r^\rho(P)$ is a decreasing function of r (ρ fixed and $\rho > 0$).

(ii) $G_r^\rho(P)$ is an increasing function of r (ρ fixed and $\rho < 0$).

(iii) $G_r^\rho(P)$ is a decreasing function of ρ (r fixed and $r > 1$).

(iv) $G_r^\rho(P)$ is an increasing function of ρ (r fixed and $0 < r < 1$).

(e_3) (i) $G_r^\rho(1 - p_{\max}, p_{\max}) \leqq G_r^\rho(P)$

$$\leqq G_r^\rho\left(\frac{1 - p_{\max}}{n-1}, \ldots, \frac{1 - p_{\max}}{n-1}, p_{\max}\right)$$

$0 < r < 1$, $\rho > 0$ (or $r > 1$, $\rho < 0$)

(ii) $G_r^\rho(1-p_{\max}, p_{\max}) \geqq G_r^\rho(P)$

$$\geqq G_r^\rho\left(\frac{1-p_{\max}}{n-1}, \ldots, \frac{1-p_{\max}}{n-1}, p_{\max}\right)$$

$r > 1, \rho > 0$ (or $0 < r < 1, \rho < 0$)

Proposition 3.2. For all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n$, we have

(e_1) (i) $T_r^\rho(P)$ is an increasing function of r (ρ fixed).

(ii) $T_r^\rho(P)$ is an increasing function of ρ (r fixed).

(e_2) (i) $T_r^\rho(P) \leqq p_{\max}$.

(ii) $$T_r^\rho(P) \geqq \left(\frac{p_{\max}^\rho}{\sum_{i=1}^n p_i^\rho}\right)^{1/r-\rho} \cdot p_{\max}, \qquad r > \rho.$$

(iii) $$T_r^\rho(P) \geqq \left(\frac{p_{\max}^r}{\sum_{i=1}^n p_i^r}\right)^{1/r-\rho} \cdot p_{\max}, \qquad \rho > r.$$

Applications of these properties of $G_r^\rho(P)$ and $T_r^\rho(P)$ in obtaining bounds on the probability of error are given in propositions 7.2 and 7.3, respectively.

IV. Generalized Measures of Directed Divergence

Kullback and Leibler (1951) first introduced a measure of information between the two probability distributions as

$$D(P||U) = \sum_{i=1}^n p_i \log \frac{p_i}{u_i}, \tag{39}$$

for all $P, U \in \Delta_n$.

Equation (39) is known in the literature as a function of *discrimination, relative information*, or *directed divergence* between the distributions.

Rényi (1961) first presented a parametric generalization of as

$$D_r^1(P||U) = (r-1)^{-1} \log\left(\sum_{i=1}^n p_i^r u_i^{1-r}\right), \qquad r \neq 1, r > 0 \tag{40}$$

for all $P, U \in \Delta_n$. Another well-known generalization of Eq. (39) is given by

$$D_s^s(P||U) = (1 - 2^{1-s})^{-1}\left[\sum_{i=1}^n p_i^s u_i^{1-s} - 1\right], \qquad s \neq 1, s > 0 \tag{41}$$

for all $P, U \in \Delta_n$. The following limits are easy to check:

$$\lim_{r \to 1} D_r^1(P||U) = \lim_{s \to 1} D_s^s(P||U) = D(P||U).$$

Sharma and Mittal (1977) studied the following two generalizations:

$$D_1^s(P||U) = (1 - 2^{1-s})^{-1}[2^{(s-1)D(P||U)} - 1], \qquad s \neq 1, \tag{42}$$

and

$$D_r^s(P||U) = (1 - 2^{1-s})^{-1}\left[\left(\sum_{i=1}^{n} p_i^r u_i^{1-r}\right)^{s-1/r-1} - 1\right], \qquad s \neq 1, r \neq 1, r > 0 \tag{43}$$

for all $P, U \in \Delta_n$. Again, we can easily verify the following limits:

$$\lim_{s \to 1} D_r^s(P||U) = D_r^1(P||U); \qquad \lim_{s \to 1} D_1^s(P||U) = D(P||U).$$

When $r = s$ in Eq. (43), we have

$$D_r^s(P||U) = D_s^s(P||U).$$

As in the case of generalized entropies, here we can also write

$$D_r^s(P||U) = \eta_s(D_r^1(P||U)), \tag{44}$$

and

$$D_1^s(P||U) = \eta_s(D(P||U)), \tag{45}$$

where $\eta_s\colon [0, \infty) \to \mathscr{R}$ (reals) is given by

$$\eta_s(x) = (1 - 2^{1-s})^{-1}[2^{(s-1)x} - 1], \qquad s \neq 1. \tag{46}$$

The following proposition holds:

Proposition 4.1. The following are true:

(i) $\lim_{s \to 1} \eta_s(x) = x$ for all $x \geqq 0$.

(ii) $\eta_s(x) \geqq 0$ for all $x \geqq 0$ and any s.

(iii) $\eta_s(x)$ is an increasing function of x.

(iv) $\eta_s(x)$ is a convex function of x for $s > 1$.

(v) $\eta_s(x)$ is a concave function of x for $s < 1$.

(vi) $$\eta_s(x)\begin{cases} \geqq x, & s > 1, \\ \leqq x, & s < 1. \end{cases} \tag{47}$$

Proof. Parts (i)–(v) are easy verifications.

(vi) We know that

$$\ln v \leqq v - 1, \qquad v > 0.$$

Substituting $v = 2^{(s-1)x}$, in the above inequality, we get

$$2^{(s-1)x} - 1 \geqq (s-1)x \ln 2. \tag{48}$$

Multiplying both sides of Eq. (48) by $(1 - 2^{1-s})^{-1} (s \neq 1)$ and using Eq. (11) we get the required result.

Let us put the measures given in Eqs. (39), (40), (41), and (42) in a unified way as follows:

$$\mathscr{F}_r^s(P||U) = \begin{cases} D_r^s(P||U), & r \neq 1, s \neq 1, r > 0, \\ D_1^s(P||U), & r = 1, s \neq 1, \\ D_r^1(P||U), & r \neq 1, s = 1, r > 0, \\ D(P||U), & r = 1, s = 1, \end{cases} \tag{49}$$

for all $P, U \in \Delta_n$.

The measure $D_s^s(P||U)$ given in Eq. (41) is not written in a unified expression (49) because it is already contained in $D_r^s(P||U)$ as a particular case when $r = s$.

We call the measure $\mathscr{F}_r^s(P||U)$ a *unified* (r,s)*-directed divergence*.

Remarks. The definitions of $D_1^s(P||U)$ and $D_r^s(P||U)$ given initially by Sharma and Mittal (1979) involve $s > 0$, but here, in our study, we have relaxed this condition. The constant initially considered was $(2^{s-1} - 1)^{-1} \times (s \neq 1)$. Here we have taken $(1 - 2^{1-s})^{-1} (s \neq 1)$ to simplify our study on applications of the measures given in Section VII. To avoid difficulties that arises in the measures when some of the probabilities become zero, we have taken Δ_n instead of Δ_n^0. For a simplified characterization of $D_r^s(P||U)$ refer to Taneja (1984b). More details on the measures $D(P||U)$, $D_r^1(P||U)$, and $D_s^s(P||U)$ can be seen in Mathai and Rathie (1975) and Taneja (1979).

Some important properties of $\mathscr{F}_r^s(P||U)$ are summarized in the following proposition.

Proposition 4.2. For all $P, U \in \Delta_n$, the unified (r,s)-directed divergence, $\mathscr{F}_r^s(P||U)$ satisfies the following:

(i) $\mathscr{F}_r^s(P||U) \geqq 0$ for all $r > 0$ and any s.

(ii) $\mathscr{F}_r^s(P||U)$ is a convex function of the pair $(P, U) \in \Delta_n \times \Delta_n$ for all $s \geqq r > 0$.

(iii) $\mathscr{F}_r^s(P||U)$ is an increasing function of r (s fixed).

(iv) $$\mathscr{F}_r^s\left(\sum_{i=1}^{\sigma} p_i, 1 - \sum_{i=1}^{\sigma} p_i \middle\| \sum_{i=1}^{\sigma} u_i, 1 - \sum_{i=1}^{\sigma} u_i\right) \geqq \mathscr{F}_r^s(P||U)$$

$$\geqq \mathscr{F}_r^s\left(p_1, p_2, \ldots, p_\sigma, \frac{1 - \sum_{i=1}^{\sigma} p_i}{n - \sigma}, \ldots, \frac{1 - \sum_{i=1}^{\sigma} p_i}{n - \sigma} \middle\| \right.$$

$$\left. \times\, u_1, u_2, \ldots, u_\sigma, \frac{1 - \sum_{i=1}^{\sigma} u_i}{n - \sigma}, \ldots, \frac{1 - \sum_{i=1}^{\sigma} u_i}{n - \sigma}\right), 1 \leqq \sigma \leqq n. \tag{50}$$

(v) Let

$$P(c) = \left(\sum_{i=1}^{n} p_i c_{i1}, \quad \sum_{i=1}^{n} p_i c_{i2}, \ldots, \quad \sum_{i=1}^{n} p_i c_{in}\right) \in \Delta_n,$$

and

$$U(c) = \left(\sum_{i=1}^{n} u_i c_{i1}, \quad \sum_{i=1}^{n} u_i c_{i2}, \ldots, \quad \sum_{i=1}^{n} u_i c_{in}\right) \in \Delta_n,$$

where

$$\sum_{i=1}^{n} c_{ik} = \sum_{k=1}^{n} c_{ik} = 1, \qquad c_{ik} \geqq 0, \qquad i, k = 1, 2, \ldots, n. \qquad \text{Then}$$

$$\mathscr{F}_r^s(P(c)||U(c)) \leqq \mathscr{F}_r^s(P||U). \tag{51}$$

(vi) (e_1) $$D_r^s(P||U)\begin{cases} \leqq D_r^1(P||U), & s < 1, \\ \geqq D_r^1(P||U), & s > 1, \end{cases}$$

(e_2) $$D_1^s(P||U)\begin{cases} \leqq D(P||U), & s < 1, \\ \geqq D(P||U), & s > 1, \end{cases}$$

(e_3) $$D_r^s(P||U)\begin{cases} \leqq D_1^s(P||U), & 0 < r < 1, \\ \geqq D_1^s(P||U), & r > 1, \end{cases}$$

(e_4) $$D_r^1(P||U)\begin{cases} \leqq D(P||U), & 0 < r < 1, \\ \geqq D(P||U), & r > 1. \end{cases}$$

Proof. (i) In view of relations (44), (45), and proposition 4.1(iii), it is sufficient to prove the nonnegativity of $D_r^1(P||U)$. However, the nonnegativity of $D_r^1(P||U)$ is already known in the literature (Mathai and Rathie, 1975).

(ii) Proof of this part is based on the following two lemmas (Taneja, 1986b).

Lemma 4.1. The quantity

$$ {}^3M_r(P||U) = \sum_{i=1}^{n} p_i^r u_i^{1-r}, \tag{52} $$

is a convex function of the pair $(P, U) \in \Delta_n \times \Delta_n$ for $r > 1$ or $r < 0$ and is concave for $0 < r < 1$.

Lemma 4.2. The function $\zeta(x) = x^\omega$ is convex for $\omega > 1$ or $\omega < 0$ and is concave for $0 < \omega < 1$.

The proof of Lemma 4.1 can be seen in Ferentinos and Papaioannou (1983) and in Csiszár (1972). Lemma 4.2 is already known in the literature.

Proof of (ii). Let $P_\alpha = (p_{\alpha 1}, p_{\alpha 2}, \ldots, p_{\alpha n}) \in \Delta_n$ and $U_\alpha = (u_{\alpha 1}, u_{\alpha 2}, \ldots, u_{\alpha n}) \in \Delta_n$, $\alpha = 1$ and 2. From Lemma 4.1, we have

$$ \lambda_1 \sum_{i=1}^{n} p_{1i}^r u_{1i}^{1-r} + \lambda_2 \sum_{i=1}^{n} p_{2i}^r u_{2i}^{1-r} \begin{cases} \geqq \sum_{i=1}^{n} (\lambda_1 p_{1i} + \lambda_2 p_{2i})^r (\lambda_1 u_{1i} + \lambda_2 u_{2i})^{1-r}, & r > 1, \\ \leqq \sum_{i=1}^{n} (\lambda_1 p_{1i} + \lambda_2 p_{2i})^r (\lambda_1 u_{1i} + \lambda_2 u_{2i})^{1-r}, & 0 < r < 1, \end{cases} \tag{53} $$

where $\lambda_1 > 0$, $\lambda_2 > 0$, and $\lambda_1 + \lambda_2 = 1$. By the concavity of logarithmic function, we can write

$$ \begin{aligned} &\lambda_1 \log\left(\sum_{i=1}^{n} p_{1i}^r u_{1i}^{1-r} \right) + \lambda_2 \log\left(\sum_{i=1}^{n} p_{2i}^r u_{2i}^{1-r} \right), \\ &\leqq \log\left[\lambda_1 \sum_{i=1}^{n} p_{1i}^r u_{1i}^{1-r} + \lambda_2 \sum_{i=1}^{n} p_{2i}^r u_{2i}^{1-r} \right], \\ &\leqq \log\left[\sum_{i=1}^{n} (\lambda_1 p_{1i} + \lambda_2 p_{2i})^r (\lambda_1 u_{1i} + \lambda_2 u_{2i})^{1-r} \right], \qquad 0 < r < 1. \end{aligned} \tag{54} $$

Inequality (54) is obtained from inequality (53), where we have used the fact that the logarithmic function is increasing. Multiplying by $(r-1)^{-1} (r \neq 1)$ on both sides of (54), we get

$$ \lambda_1 D_r^1(P_1||U_1) + \lambda_2 D_r^1(P_2||U_2) \geqq D_r^1(\lambda_1 P_1 + \lambda_2 P_2||\lambda_1 U_1 + \lambda_2 U_2), \quad 0 < r < 1. $$

This gives the convexity of $D_r^1(P||U)$ for $0 < r < 1$. The convexity of $D_1^s(P||U)$ for $s > 1$ is an immediate consequence of the convexity of $D(P||U)$. The convexity of $D(P||U)$ is well known in the literature (Csiszár and Körner, 1981). By the use of Lemma 4.2 and inequalities (53), we have

$$\lambda_1\left[\sum_{i=1}^n p_{1i}^r u_{1i}^{1-r}\right]^{s-1/r-1} + \lambda_2\left[\sum_{i=1}^n p_{2i}^r u_{2i}^{1-r}\right]^{s-1/r-1}$$

$$\begin{cases} \geqq \left[\sum_{i=1}^n (\lambda_1 p_{1i} + \lambda_2 p_{2i})^r (\lambda_1 u_{1i} + \lambda_2 u_{2i})^{1-r}\right]^{s-1/r-1}, \\ \qquad r > 1, \dfrac{s-1}{r-1} \geqq 1 \left(\text{or } \dfrac{s-1}{r-1} < 0, 0 < r < 1\right), \\ \qquad \text{i.e., } r > 0,\ s > 1,\ s \geqq r \\ \leqq \left[\sum_{i=1}^n (\lambda_1 p_{1i} + \lambda_2 p_{2i})^r (\lambda_1 u_{1i} + \lambda_2 u_{2i})^{1-r}\right]^{s-1/r-1}, \\ \qquad 0 < r < 1, 0 < \dfrac{s-1}{r-1} \leqq 1 \qquad \text{i.e.,} \qquad 0 < r \leqq s < 1. \end{cases} \tag{55}$$

Subtracting 1, multiplying by $(1 - 2^{1-s})^{-1}(s \neq 1)$ in (55), and simplifying we get the convexity of $D_r^s(P||U)$, i.e.,

$$\lambda_1 D_r^s(P_1||U_1) + \lambda_2 D_r^s(P_2||U_2) \geqq D_r^s(\lambda_1 P_1 + \lambda_2 P_2||\lambda_1 U_1 + \lambda_2 U_2),$$

for all $s \geqq r > 0$, $r \neq 1$, $s \neq 1$. Finally, combining all these results we get the convexity of $\mathscr{F}_r^s(P||U)$ in $\Delta_n \times \Delta_n$ for $s \geqq r > 0$.

(iii) In view of proposition 4.1(iii) and the composition relation (44), it is sufficient to prove that $D_r^1(P||U)$ is an increasing function of r. In order to prove this, let us write

$${}^4M_r(P||U) = \left(\sum_{i=1}^n p_i^r u_i^{1-r}\right)^{1/r-1} = \left[\sum_{i=1}^n p_i\left(\frac{p_i}{u_i}\right)^{r-1}\right]^{1/r-1}, \qquad r \neq 1 \tag{56}$$

for all $P, U \in \Delta_n$. Let $p_i/u_i = w_i$, $i = 1, 2, \ldots, n$. Then from Eq. (56) we can write

$${}^3M_r(P||W) = \left[\sum_{i=1}^n p_i w_i^{r-1}\right]^{1/r-1}, \qquad r \neq 1,$$

where $P = (p_1, p_2, \ldots, p_n) \in \Delta_n$ and $W = (w_1, w_2, \ldots, w_n)$, $w_i > 0$, $i = 1, 2, \ldots, n$. We know (Hardy et al., 1934, pp. 15, Theorem 5) that the function given in Eq. (56) is increasing in r. Since log(.) is an increasing function, this proves the required result.

(iv) It can be proved on similar lines as property 3(i).

(v) It follows on the lines of property 11.

(vi) The inequalities given in (e_1) and (e_2) are due to proposition 4.1 (vi) and the composition relations (44) and (45), respectively. The inequalities given in (e_3) and (e_4) are due to part (iii).

V. Generalized Divergence Measures

This section deals with the generalizations of two different kinds of divergence measures. One is known as the *information radius* or the *Jensen difference divergence measure* (Sibson, 1969) and the other is well known as *J-divergence* (Kullback and Leibleŕ, 1951; Jeffreys, 1946).

A. *Information Radius and the J-Divergence*

By using the concavity of Shannon's entropy, we can write

$$\frac{H(P)+H(U)}{2} \leqq H\left(\frac{P+U}{2}\right) \tag{57}$$

for all $P, U \in \Delta_n$.

The difference

$$\begin{aligned} R(P||U) &= H\left(\frac{P+U}{2}\right) - \frac{H(P)+H(U)}{2} \\ &= \sum_{i=1}^{n}\left[\frac{p_i \log p_i + u_i \log u_i}{2} - \left(\frac{p_i+u_i}{2}\right)\log\left(\frac{p_i+u_i}{2}\right)\right] \end{aligned} \tag{58}$$

for all $P, U \in \Delta_n$ is known as the information radius (Sibson, 1969) or Jensen difference divergence measure (Burbea and Rao, 1982). For simplicity, we shall call $R(P||U)$, the *R-divergence.*

Another measure of divergence known in the literature is J-divergence (Kullback and Leibler, 1951; Jeffreys, 1946) and is given by

$$\begin{aligned} J(P||U) &= D(P||U) + D(U||P) \\ &= \sum_{i=1}^{n}(p_i - u_i)\log\frac{p_i}{u_i} \end{aligned} \tag{59}$$

for all $P, U \in \Delta_n$, where $D(P||U)$ is as given in Eq. (39).

By simple calculations, we can write

$$R(P||U) = \frac{1}{2}\left[D\left(P||\frac{P+U}{2}\right) + D\left(U||\frac{P+U}{2}\right)\right] \tag{60}$$

for all $P, U \in \Delta_n$.

B. *Generalizations of R-Divergence*

In this subsection, we shall present three different ways to generalize R-divergence, i.e., the Jensen difference divergence measure given in Eq. (58). These generalizations are as follows.

Taking $\mathscr{E}_r^s(P)$ in place of $H(P)$ in Eq. (58), we have

$$ {}^1\mathscr{V}_r^s(P||U) = \mathscr{E}_r^s\left(\frac{P+U}{2}\right) - \frac{\mathscr{E}_r^s(P) + \mathscr{E}_r^s(U)}{2} $$

for all $P, U \in \Delta_n$, where $\mathscr{E}_r^s(P)$ is the unified (r, s)-entropy given in Eq. (7). More clearly, we have

$$ {}^1\mathscr{V}_r^s(P||U) = \begin{cases} {}^1R_r^s(P||U), & r \neq 1, s \neq 1, r > 0, \\ {}^1R_1^s(P||U), & r = 1, s \neq 1, \\ {}^1R_r^1(P||U), & r \neq 1, s = 1, r > 0, \\ R(P||U), & r = 1, s = 1 \end{cases} \tag{61} $$

for all $P, U \in \Delta_n$, where

$$ {}^1R_r^s(P||U) = (1 - 2^{1-s})^{-1}\left\{\frac{1}{2}\left[\left(\sum_{i=1}^n p_i^r\right)^{s-1/r-1} + \left(\sum_{i=1}^n u_i^r\right)^{s-1/r-1}\right] - \left[\sum_{i=1}^n \left(\frac{p_i + u_i}{2}\right)^r\right]^{s-1/r-1}\right\}, \quad s \neq 1, r \neq 1, r > 0 \tag{62} $$

$$ {}^1R_1^s(P||U) = (1 - 2^{1-s})^{-1}\left[\frac{1}{2}\left(2^{(1-s)H(P)} + 2^{(1-s)H(U)}\right) - 2^{(1-s)H\left(\frac{P+U}{2}\right)}\right], \quad s \neq 1 \tag{63} $$

$$ {}^1R_r^1(P||U) = (r-1)^{-1}\log\left\{\frac{\sum_{i=1}^n \left(\frac{p_i + u_i}{2}\right)^r}{\sqrt{\left(\sum_{i=1}^n p_i^r\right)\left(\sum_{i=1}^n u_i^r\right)}}\right\}, \quad r \neq 1, r > 0. \tag{64} $$

When $r = s$ in Eq. (62), we have

$$ {}^1R_s^s(P||U) = (1 - 2^{1-s})^{-1} \cdot \sum_{i=1}^n \left[\frac{p_i^s + u_i^s}{2} - \left(\frac{p_i + u_i}{2}\right)^s\right], \qquad s \neq 1, s > 0. \tag{65} $$

An alternative way to generalize $R(P||U)$ is to replace $D(P||U)$ by $\mathscr{F}_r^s(P||U)$ in Eq. (60). Then we get

$$ {}^2\mathscr{V}_r^s(P||U) = \frac{1}{2}\left[\mathscr{F}_r^s\left(P\Big|\Big|\frac{P+U}{2}\right) + \mathscr{F}_r^s\left(U\Big|\Big|\frac{P+U}{2}\right)\right] $$

for all $P, U \in \Delta_n$, where $\mathscr{F}_r^s(P||U)$ is given in Eq. (49). More clearly, we have

$$ {}^2\mathscr{V}_r^s(P||U) = \begin{cases} {}^2R_r^s(P||U), & r \neq 1, s \neq 1, r > 0, \\ {}^2R_1^s(P||U), & r = 1, s \neq 1, \\ {}^2R_r^1(P||U), & r \neq 1, s = 1, r > 0, \\ R(P||U), & r = 1, s = 1 \end{cases} \tag{66} $$

for all $P, U \in \Delta_n$, where

$$ {}^2R_r^s(P||U) = \frac{1}{2(1-2^{1-s})}\left\{\left[\sum_{i=1}^n p_i^r\left(\frac{p_i+u_i}{2}\right)^{1-r}\right]^{s-1/r-1} + \left[\sum_{i=1}^n u_i^r\left(\frac{p_i+u_i}{2}\right)^{1-r}\right]^{s-1/r-1} - 2\right\}, $$
$$ r \neq 1, s \neq 1, r > 0 \tag{67} $$

$$ {}^2R_1^s(P||U) = \frac{1}{2(1-2^{1-s})}\{2^{(s-1)D(P||P+U/2)} + 2^{(s-1)D(U||P+U/2)} - 2\}, $$
$$ s \neq 1, \tag{68} $$

$$ {}^2R_r^1(P||U) = \frac{1}{2(r-1)}\left\{\log\left[\sum_{i=1}^n p_i^r\left(\frac{p_i+u_i}{2}\right)^{1-r}\right] + \log\left[\sum_{i=1}^n u_i^r\left(\frac{p_i+u_i}{2}\right)^{1-r}\right]\right\}, \quad r \neq 1, r > 0. \tag{69} $$

When $r = s$ in Eq. (67), we have

$$ {}^2R_s^s(P||U) = (1-2^{1-s})^{-1}\left\{\sum_{i=1}^n \left(\frac{p_i^s+u_i^s}{2}\right)\left(\frac{p_i+u_i}{2}\right)^{1-s} - 1\right\}, \tag{70} $$
$$ s \neq 1, s > 0. $$

There is also a third way to generalize the R-divergence similar to Eqs. (67) and (69) based on an expression given in Eq. (70). These generalizations are as follows:

$$ {}^3R_r^s(P||U) = (1-2^{1-s})^{-1}\left\{\left[\sum_{i=1}^n \left(\frac{p_i^r+u_i^r}{2}\right)\left(\frac{p_i+u_i}{2}\right)^{1-r}\right]^{s-1/r-1} - 1\right\}, \tag{71} $$
$$ r \neq 1, s \neq 1, r > 0, $$

$$ {}^3R_r^1(P||U) = (r-1)^{-1}\log\left\{\sum_{i=1}^n \left(\frac{p_i^r+u_i^r}{2}\right)\left(\frac{p_i+u_i}{2}\right)^{1-r}\right\}, \tag{72} $$
$$ r \neq 1, r > 0. $$

The following limits hold:

$$\lim_{s\to 1} {}^3R_r^s(P\|U) = {}^3R_r^1(P\|U); \qquad \lim_{r\to 1} {}^3R_r^s(P\|U) = {}^3R_1^s(P\|U)$$

When $r = s$ in Eq. (71), we have

$${}^3R_r^s(P\|U) = {}^2R_s^s(P\|U),$$

where ${}^2R_s^s(P\|U)$ is as given in Eq. (70). The last generalizations can be unified as follows:

$${}^3\mathscr{V}_r^s(P\|U) = \begin{cases} {}^3R_r^s(P\|U), & r \neq 1, s \neq 1, r > 0, \\ {}^3R_1^s(P\|U), & r = 1, s \neq 1, \\ {}^3R_r^1(P\|U), & r \neq 1, s = 1, r > 0, \\ R(P\|U), & r = 1, s = 1 \end{cases} \tag{73}$$

for all $P, U \in \Delta_n$.

Remarks. The generalized measures given in Eqs. (61), (66), and (73) are the author's contributions and are presented here for the first time, except Eqs. (64) (Rao, 1982) and (65) (Burbea and Rao, 1982). The measure given in Eq. (70) can be seen in Taneja (1988a).

The following proposition holds:

Proposition 5.1. For all $P, U \in \Delta_n$, we have

(i) ${}^1\mathscr{V}_r^s(P\|U) \geqq 0$ for $(r, s) \in \Gamma_1$,

where Γ_1 is given by Eq. (12).

(ii) ${}^2\mathscr{V}_r^s(P\|U) \geqq 0$ for all $r > 0$ and any s.

(iii) ${}^3\mathscr{V}_r^s(P\|U) \geqq 0$ for all $r > 0$ and any s. (74)

(iv) ${}^2\mathscr{V}_r^s(P\|U) \begin{cases} \leqq {}^3\mathscr{V}_r^s(P\|U), & s \leqq r, \\ \geqq {}^3\mathscr{V}_r^s(P\|U), & s \geqq r. \end{cases}$

Proof. (i) This follows from the concavity of $\mathscr{E}_r^s(P)$ $(P \in \Delta_n)$ for all $(r, s) \in \Gamma_1$ given in property 8.

(ii) This follows by the nonnegativity of $\mathscr{F}_r^s(P\|U)$ given in proposition 4.2(i).

(iii) We can write

$${}^3R_r^s(P\|U) = \eta_s({}^3R_r^1(P\|U)), \qquad r \neq 1, s \neq 1, r > 0 \tag{75}$$

and

$$^3R_1^s(P||U) = \eta_s(R(P||U)), \qquad s \neq 1, \tag{76}$$

where η_s is as given in Eq. (46).

In view of relations (75) and (76), and proposition 4.1(iii), it is sufficient to prove the nonnegativity of $^3R_r^1(P||U)$ $(r \neq 1, r > 0)$ because the nonnegativity of $R(P||U)$ is obvious from Eq. (57). Let us now prove the nonnegativity of $^3R_r^1(P||U)$.

By Lemma 4.2, we can write

$$\frac{p_i^r + u_i^r}{2} \begin{cases} \leqq \left(\dfrac{p_i + u_i}{2}\right)^r, & 0 < r < 1, \\ \geqq \left(\dfrac{p_i + u_i}{2}\right)^r, & r > 1 \end{cases} \tag{77}$$

for all $i = 1,2,\ldots,n$, $P = (p_1,p_2,\ldots,p_n) \in \Delta_n$ and $U = (u_1,u_2,\ldots,u_n) \in \Delta_n$. Multiplying Eq. (77) by $[(p_i + u_i)/2]^{1-r}$ and summing over all $i = 1,2,\ldots,n$, we get

$$\sum_{i=1}^n \left(\frac{p_i^r + u_i^r}{2}\right)\left(\frac{p_i + u_i}{2}\right)^{1-r} \begin{cases} \leqq 1, & 0 < r < 1, \\ \geqq 1, & r > 1. \end{cases} \tag{78}$$

Taking $\log(\cdot)$ on both sides of Eq. (78) and multiplying by $(r-1)^{-1}(r \neq 1)$, we get the required result.

(iv) Again using Lemma 4.2, we can write

$$\left[\sum_{i=1}^n p_i^r\left(\frac{p_i + u_i}{2}\right)^{1-r}\right]^{s-1/r-1} + \left[\sum_{i=1}^n u_i^r\left(\frac{p_i + u_i}{2}\right)^{1-r}\right]^{s-1/r-1}$$
$$\begin{cases} \leqq 2\left[\displaystyle\sum_{i=1}^n \left(\frac{p_i^r + u_i^r}{2}\right)\left(\frac{p_i + u_i}{2}\right)^{1-r}\right]^{s-1/r-1}, \\ \qquad 0 < \dfrac{s-1}{r-1} \leqq 1, \\ \geqq 2\left[\displaystyle\sum_{i=1}^n \left(\frac{p_i^r + u_i^r}{2}\right)\left(\frac{p_i + u_i}{2}\right)^{1-r}\right]^{s-1/r-1}, \\ \qquad \dfrac{s-1}{r-1} \geqq 1 \quad \text{or} \quad \dfrac{s-1}{r-1} < 0. \end{cases} \tag{79}$$

Subtracting 2 from both sides of Eq. (79), multiplying by $(1 - 2^{1-s})^{-1}(s \neq 1)$, and simplifying we get

$$^2R_r^s(P||U) \begin{cases} \leqq {}^3R_r^s(P||U), & s \leqq r, r \neq 1, s \neq 1, \\ \geqq {}^3R_r^s(P||U), & s \geqq r, r \neq 1, s \neq 1 \end{cases} \tag{80}$$

for all $r > 0$. Using the concavity property of the logarithmic function we can write

$$\log\left[\sum_{i=1}^{n} p_i^r\left(\frac{p_i+u_i}{2}\right)^{1-r}\right] + \log\left[\sum_{i=1}^{n} u_i^r\left(\frac{p_i+u_i}{2}\right)^{1-r}\right] \leqq 2\log\left\{\sum_{i=1}^{n}\left(\frac{p_i^r+u_i^r}{2}\right)\left(\frac{p_i+u_i}{2}\right)^{1-r}\right\}, \tag{81}$$

for any $r > 0$. Multiplying Eq. (81) by $(r-1)^{-1}(r \neq 1)$, we get

$$ {}^2R_r^1(P||U)\begin{cases} \leqq {}^3R_r^1(P||U), & r > 1, \\ \geqq {}^3R_r^1(P||U), & 0 < r < 1. \end{cases} \tag{82}$$

In a similar way we can prove that

$$ {}^2R_1^s(P||U)\begin{cases} \leqq {}^3R_1^s(P||U), & 0 < s < 1, \\ \geqq {}^3R_1^s(P||U), & s > 1. \end{cases} \tag{83}$$

Combining Eqs. (80)–(83), we get the required result.

C. Generalizations of J-Divergence

In this subsection we shall present two different ways to generalize the J-divergence given in Eq. (59) involving one and two scalar parameters. The generalizations involving one scalar parameter (Rathie and Sheng, 1981; Bubea and Rao, 1982; Taneja, 1983; Burbea, 1984) are given by

$$J_s^s(P||U) = (1-2^{1-s})^{-1}\left[\sum_{i=1}^{n} p_i^s u_i^{1-s} + \sum_{i=1}^{n} p_i^{1-s}u_i^s - 2\right], \quad s \neq 1, s > 0 \tag{84}$$

$${}^1J_r^1(P||U) = (r-1)^{-1}\left[\log\left(\sum_{i=1}^{n} p_i^r u_i^{1-r}\right) + \log\left(\sum_{i=1}^{n} p_i^r u_i^{1-r}\right)\right], \quad r \neq 1, r > 0 \tag{85}$$

and

$${}^2J_r^1(P||U) = (r-1)^{-1} 2\log\left\{\sum_{i=1}^{n}\left(\frac{p_i^r u_i^{1-r} + p_i^{1-r}u_i^r}{2}\right)\right\}, \quad r \neq 1, r > 0 \tag{86}$$

for all $P, U \in \Delta_n$. We can easily verify that

$$\lim_{s\to 1} J_s^s(P||U) = \lim_{r\to 1} {}^1J_r^1(P||U) = 2\lim_{r\to 1} {}^2J_r^1(P||U) = J(P||U).$$

The generalizations involving two scalar parameters considered by Taneja (1983) are given by

$$ {}^1J_r^s(P\|U) = (1 - 2^{1-s})^{-1}\left[\left(\sum_{i=1}^{n} p_i^r u_i^{1-r}\right)^{s-1/r-1} + \left(\sum_{i=1}^{n} p_i^{1-r} u_i^r\right)^{s-1/r-1} - 2\right], \qquad r \neq 1, s \neq 1, r > 0, \tag{87} $$

and

$$ {}^2J_r^s(P\|U) = (1 - 2^{1-s})^{-1} 2\left\{\left[\sum_{i=1}^{n}\left(\frac{p_i^r u_i^{1-r} + p_i^{1-r} u_i^r}{2}\right)\right]^{s-1/r-1} - 1\right\}, \quad r \neq 1, s \neq 1, r > 0. \tag{88} $$

The following limits are easy to verify:

$$ \lim_{r \to 1} {}^1J_r^s(P\|U) = {}^1J_r^1(P\|U); \lim_{s \to 1} {}^2J_r^s(P\|U) = {}^2J_r^1(P\|U). $$

When $r = s$ in Eqs. (87) and (88) we have

$$ {}^1J_r^s(P\|U) = {}^2J_r^s(P\|U) = J_s^s(P\|U). $$

We can also write

$$ {}^1J_r^s(P\|U) = \eta_s(D_r^1(P\|U)) + \eta_s(D_r^1(U\|P)) \tag{89} $$

and

$$ {}^2J_r^s(P\|U) = 2\eta_s\left[\frac{{}^2J_r^1(P\|U)}{2}\right], \tag{90} $$

where $D_r^1(P\|U)$ and η_s are given by Eqs. (40) and (46), respectively. Also

$$ \begin{aligned} \lim_{r \to 1} {}^1J_r^s(P\|U) &= \eta_s(D(P\|U)) + \eta_s(D(U\|P)) \\ &= {}^1J_1^s(P\|U) \end{aligned} $$

and

$$ \begin{aligned} \lim_{r \to 1} {}^2J_r^s(P\|U) &= \eta_s(J(P\|U)) \\ &= {}^2J_1^s(P\|U). \end{aligned} $$

We can also write

$$ {}^1J_r^1(P\|U) = D_r^1(P\|U) + D_r^1(U\|P). $$

Both the generalizations of J-divergence involving one and two scalar parameters can be unified in the following way:

$$
{}^{\alpha}\mathscr{W}_r^s(P\|U) = \begin{cases} {}^{\alpha}J_r^s(P\|U), & r \neq 1, s \neq 1, r > 0, \\ {}^{\alpha}J_1^s(P\|U), & r = 1, s \neq 1, \\ {}^{\alpha}J_r^1(P\|U), & r \neq 1, s = 1, r > 0, \\ J(P\|U), & r = 1, s = 1 \end{cases} \tag{91}
$$

for all $P, U \in \Delta_n$, and $\alpha = 1$ and 2.

The following proposition holds:

Proposition 5.2. For all $P, U \in \Delta_n$, we have

(i) ${}^{\alpha}\mathscr{W}_r^s(P\|U) \geqq 0$ ($\alpha = 1$ and 2) for all $r > 0$ and any s.

(ii) ${}^{\alpha}\mathscr{W}_r^s(P\|U)$ ($\alpha = 1$ and 2) are convex functions of the pair of distributions $(P, U) \in \Delta_n \times \Delta_n$ for all $s \geqq r > 0$.

(iii)
$$
{}^{1}\mathscr{W}_r^s(P\|U) \begin{cases} \leqq {}^{2}\mathscr{W}_r^s(P\|U), & r \geqq s, \\ \geqq {}^{2}\mathscr{W}_r^s(P\|U), & s \geqq r > 0. \end{cases} \tag{92}
$$

Proof. (i) In view of proposition 4.2(i) and relation (89), the nonnegativity of ${}^{1}J_r^s(P\|U)$ is clear. In view of relation (90), it is sufficient to prove the nonnegativity of ${}^{2}J_r^1(P\|U)$ given in Eq. (86). Its proof is as follows:

By Lemma 4.2, we can write

$$
\sum_{i=1}^n \left(\frac{p_i^r u_i^{1-r} + p_i^{1-r} u_i^r}{2} \right) = \frac{1}{2} \sum_{i=1}^n \left[u_i \left(\frac{p_i}{u_i} \right)^r + p_i \left(\frac{v_i}{p_i} \right)^r \right]
$$
$$
\begin{cases} \leqq \dfrac{1}{2} \left[\sum_{i=1}^n u_i \left(\dfrac{p_i}{u_i} \right) \right]^r + \dfrac{1}{2} \left[\sum_{i=1}^n p_i \left(\dfrac{v_i}{p_i} \right) \right]^r, & 0 < r < 1, \\ \geqq \dfrac{1}{2} \left[\sum_{i=1}^n u_i \left(\dfrac{p_i}{u_i} \right) \right]^r + \dfrac{1}{2} \left[\sum_{i=1}^n p_i \left(\dfrac{v_i}{p_i} \right) \right]^r, & r > 1, \end{cases}
$$

i.e.,

$$
\sum_{i=1}^n \frac{1}{2} [p_i^r u_i^{1-r} + p_i^r u_i^{1-r}] \begin{cases} \leqq 1, & 0 < r < 1, \\ \geqq 1, r > 1. \end{cases} \tag{93}
$$

Taking $\log(\cdot)$ on both sides of Eq. (93), multiplying by $(r-1)^{-1} (r \neq 1)$, and simplifying we get

$$
{}^{2}J_r^1(P\|U) \geqq 0 \text{ for all } r \neq 1, r > 0.
$$

(ii) It can be proved on lines similar to proposition 4.2(ii), where instead of using Lemma 4.1, we use the fact that the function $\Sigma_{i=1}^{n}(p_i^r u_i^{1-r} + p_i^{1-r}u_i^r)$, $r \neq 1$ is convex in the pair $(P, U) \in \Delta_n \times \Delta_n$ for $r > 1$ or $r < 0$ and is concave for $0 < r < 1$.

(iii) Again by the use of Lemma 4.2, we have

$$\left(\sum_{i=1}^{n} p_i^r u_i^{1-r}\right)^{s-1/r-1} + \left(\sum_{i=1}^{n} p_i^{1-r}u_i^r\right)^{s-1/r-1}$$
$$\begin{cases} \leqq 2\left[\sum_{i=1}^{n}\left(\dfrac{p_i^r u_i^{1-r} + p_i^{1-r}u_i^r}{2}\right)\right]^{s-1/r-1}, & 0 < \dfrac{s-1}{r-1} \leqq 1, \\ \geqq 2\left[\sum_{i=1}^{n}\left(\dfrac{p_i^r u_i^{1-r} + p_i^{1-r}u_i^r}{2}\right)\right]^{s-1/r-1}, & \dfrac{s-1}{r-1} \geqq 1 \text{ or } \dfrac{s-1}{r-1} < 0 \end{cases} \tag{94}$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n$ and $U = (u_1, u_2, \ldots, u_n) \in \Delta_n$. Subtracting 2 on both sides of Eq. (94), multiplying by $(1 - 2^{1-s})^{-1}$ $(s \neq 1)$, and simplifying we get

$${}^1J_r^s(P||U)\begin{cases} \leqq {}^2J_r^s(P||U), & s \leqq r, \\ \geqq {}^2J_r^s(P||U), & s \geqq r \end{cases} \tag{95}$$

for all P, $U \in \Delta_n$ and $r > 0$, $r \neq 1$. Using the concavity of the logarithmic function, we can write

$$\log\left(\sum_{i=1}^{n} p_i^r u_i^{1-r}\right) + \log\left(\sum_{i=1}^{n} p_i^{1-r}u_i^r\right)$$
$$\leqq 2\log\left[\sum_{i=1}^{n}\left(\frac{p_i^r u_i^{1-r} + p_i^{1-r}u_i}{2}\right)\right]. \tag{96}$$

Multiplying Eq. (96) by $(r-1)^{-1}$ $(r \neq 1)$, we obtain

$${}^1J_r^1(P||U)\begin{cases} \leqq {}^2J_r^1(P||U), & r > 1, \\ \geqq {}^2J_r^1(P||U), & 0 < r < 1. \end{cases} \tag{97}$$

In a similar way we can show that

$${}^1J_1^s(P||U)\begin{cases} \leqq {}^2J_1^s(P||U), & 0 < s < 1, \\ \geqq {}^2J_1^s(P||U), & s > 1. \end{cases} \tag{98}$$

Combining Eqs. (95)–(98), we get the required result.

For statistical applications of the measures given in Eq. (91) refer to Taneja (1987).

VI. Generalized Entropies for Multivariate Probability Distributions

The idea of entropy measure needs to be developed for multivariate probability distributions, in particular, for bivariate cases, especially in the problems of communication that require analysis of messages sent over a channel and received at the other end. The same is also required in the bounding Bayesian probability of error. In order to develop this idea, let us consider two discrete finite random variables $X = \{1, 2, \ldots, n\}$ and $Y = \{1, 2, \ldots, m\}$ or a joint experiment (X, Y) with joint and individual (marginal) probabilities denoted by

- $a_{ij} = Pr\{X = i, Y = j\}$,
  $$A = (a_{11}, a_{12}, \ldots, a_{1m}, \ldots, a_{n1}, a_{n2}, \ldots, a_{nm}) \in \Delta^0_{nm},$$
- $p_i = Pr\{X = i\}$, $\quad P = (p_1, p_2, \ldots, p_n) \in \Delta^0_n$, $\quad$ and
- $q_j = Pr\{Y = j\}$, $\quad Q = (q_1, q_2, \ldots, q_m) \in \Delta^0_m$

for all $i = 1, 2, \ldots, n; j = 1, 2, \ldots, m$. The conditional probability of $Y = j$ given $X = i$ is denoted by

- $b_{j|i} = Pr\{Y = j \mid X = i\}$, $\quad B_i = (b_{1|i}, b_{2|i}, \ldots, b_{m|i}) \in \Delta^0_m$

for all $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$. Similarly, the conditional probability of $X = i$ given $Y = j$ is denoted by

- $b_{i|j} = Pr\{X = i \mid Y = j\}$, $\quad B_j = (b_{1|j}, b_{2|j}, \ldots, b_{n|j}) \in \Delta^0_n$

for all $i = 1, 2, \ldots, n; j = 1, 2, \ldots, m$.

Let us also denote,

$$P{*}Q = (p_1q_1, p_1q_2, \ldots, p_1q_m, \ldots, p_nq_1, \ldots, p_nq_m) \in \Delta^0_{nm}.$$

The following relations are well known in the literature:

$$a_{ij} = p_i \cdot b_{j|i} = q_j \cdot b_{i|j}, \qquad p_i = \sum_{i=1}^{m} a_{ij}, \qquad \text{and} \qquad q_j = \sum_{i=1}^{n} a_{ij}$$

for all $i = 1, 2, \ldots, n; j = 1, 2, \ldots, m$.

If X and Y are independent random variables, then

$$a_{ij} = p_i q_j, \qquad i = 1, 2, \ldots, n; j = 1, 2, \ldots, m.$$

Based on the above notations, the *joint and individual unified* (r, s)*-entropies* can be written as:

$$\mathscr{E}^s_r(X, Y) = \mathscr{E}^s_r(A),$$

$$\mathscr{E}^s_r(X) = \mathscr{E}^s_r(P),$$

and

$$\mathscr{E}_r^s(Y) = \mathscr{E}_r^s(Q),$$

where $\mathscr{E}_r^s$ is the unified (r, s)-entropy given in Eq. (7). Also, we can easily write $\mathscr{E}_r^s(X, Y, Z)$, etc. Similarly, the *individual conditional unified (r, s)-entropies* are given by

$$\mathscr{E}_r^s(Y \mid X = i) = \mathscr{E}_r^s(B_i), \qquad i = 1, 2, \ldots, n$$

and

$$\mathscr{E}_r^s(X \mid Y = j) = \mathscr{E}_r^s(B_j), \qquad j = 1, 2, \ldots, m.$$

There is no unique way to define the conditional generalized entropies. It has been defined in different ways by different authors. We shall specify here five different ways to define conditional generalized entropies. One is restricted to only entropies of degree s given in Eq. (3), and the other four are for the unified (r, s)-entropy given in Eq. (7). We shall observe that these different approaches in the limiting case reduce to the well-known Shannon's conditional entropy. These five approaches have been divided in two subsections: the first approach is only for the entropy of degree s and the second approach is for the unified entropy.

Henceforth, unless otherwise specified, the letters $X, Y, Z, \ldots, X_1, X_2, \ldots$ etc. will represent the discrete finite random variables.

A. *Entropy of Degree s for Multivariate Probability Distributions*

In this subsection, we shall define a conditional entropy of degree s, which in the limiting case contains Shannon's conditional entropy. This definition was first considered by Daróczy (1970) and satisfies many of the properties of Shannon's case. In order to simplify the results, let us unify these two entropies in the following way:

$$C^s(P) = \begin{cases} (2^{1-s} - 1)^{-1}\left[\sum_{i=1}^{n} p_i^s - 1\right], & s \neq 1, s > 0, \\ -\sum_{i=1}^{n} p_i \log p_i, & s = 1 \end{cases}$$

for all $P = (p_1, p_2, \ldots, p_n) \in \Delta_n^0$.

Define

$$C^s(X \mid Y) = \sum_{j=1}^{m} q_j^s C^s(X \mid Y = j), \qquad s > 0,$$

where

$$C^s(X \mid Y = j) = \begin{cases} (2^{1-s} - 1)^{-1}\left[\sum_{i=1}^{n} b_{i|j}^s - 1\right], & s \neq 1, s > 0, \\ -\sum_{i=1}^{n} b_{i|j} \log b_{i|j}, & s = 1. \end{cases} \tag{99}$$

In a similar way, we can define $C^s(Y \mid X)$.

Define

$$C^s(X, Y \mid Z) = \sum_{\ell=1}^{v} c_\ell^s C^s(X, Y \mid Z = \ell),$$

where

$$C^s(X, Y \mid Z = \ell) = \begin{cases} (2^{1-s} - 1)^{-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{m} b_{ij|\ell}^s - 1\right], & s \neq 1, s > 0, \\ -\sum_{i=1}^{n}\sum_{j=1}^{m} b_{ij|\ell} \log b_{ij|\ell}, & s = 1, \end{cases}$$

and

$$c_\ell = Pr\{Z = \ell\} \text{ for all } \ell = 1, 2, \ldots, v.$$

Also define

$$I^s(X \wedge Y) = C^s(X) - C^s(X \mid Y), \qquad s > 0$$

and

$$I^s(X \wedge Y \mid Z) = C^s(X \mid Z) - C^s(X \mid Y, Z), \qquad s > 0,$$

where

$$C^s(X \mid Y, Z) = \sum_{j=1}^{m}\sum_{\ell=1}^{v} a_{jk}^s C^s(X \mid Y = j, Z = \ell)$$

and

$$C^s(X \mid Y = j, Z = \ell) = \begin{cases} (2^{1-s} - 1)^{-1}\left[\sum_{i=1}^{n} b_{i|j\ell}^s - 1\right], & s \neq 1, s > 0, \\ -\sum_{i=1}^{n} b_{i|j\ell} \log b_{i|j\ell}, & s = 1 \end{cases}$$

for all $j = 1, 2, \ldots, m$ and $\ell = 1, 2, \ldots, v$.

The measure $I^s(X \wedge Y)$ is known in the literature as the *mutual information of degree s*.

Based on the definitions given above, the following propositions hold (Taneja, 1988b).

Proposition 6.1. For all $s > 0$, we have

(i) $C^s(X, Y) = C^s(X) + C^s(Y \mid X) = C^s(Y) + C^s(X \mid Y)$.

(ii) $C^s(X) = C^s(X \mid Y) + I^s(X \wedge Y)$.

(iii) $C^s(X, Y, Z) = C^s(X) + C^s(Y, Z \mid X)$
$= C^s(X, Y) + C^s(Z \mid X, Y)$
$= C^s(X) + C^s(Y \mid X) + C^s(Z \mid X, Y)$.

(iv) $C^s(X_1, X_2, \ldots, X_\delta) = C^s(X_1) + C^s(X_2 \mid X_1) + C^s(X_3 \mid X_1, X_2) + \cdots + C^s(X_\delta \mid X_1, X_2, \ldots, X_{\delta-1})$
$= \sum_{i=1}^{\delta} C^s(X_i \mid X_1, X_2, \ldots, X_{i-1})$.

(v) $C^s(Y, Z \mid X) = C^s(Y \mid X) + C^s(Z \mid X, Y)$.

(vi) $C^s(X \mid Z) = C^s(X \mid Y, Z) + I^s(X \wedge Y \mid Z)$.

(vii) $I^s(X, Y \wedge Z) = I^s(X \wedge Z) + I^s(Y \wedge Z \mid X)$.

(viii) $I^s(X \wedge Y) = C^s(X) + C^s(Y) - C^s(X, Y) = I^s(Y \wedge X)$.

(ix) $C^s(X) + C^s(Y) + C^s(Z) - C^s(X, Y, Z) = I^s(X, Y \wedge Z) + I^s(X \wedge Y)$.

(x) $I^s(X \wedge Z) + I^s(X \wedge Y \mid Z) = I^s(X \wedge Y) + I^s(X \wedge Z \mid Y)$.

(xi) $I^s(X_1, X_2 \wedge X_3 \mid X_4) = I^s(X_1 \wedge X_3 \mid X_4) + I^s(X_2 \wedge X_3 \mid X_1, X_4)$.

The proof of these properties is a simple verification.

Proposition 6.2. For all $s \geqq 1$, we have

(i) $I^s(X \wedge Y) \geqq 0$, i.e., $C^s(X \mid Y) \leqq C^s(X)$.

(ii) $I^s(X \wedge Y \mid Z) \geqq 0$, i.e., $C^s(X \mid Y, Z) \leqq C^s(X \mid Z)$.

The proof of part (i) can be seen in Daróczy (1970), and part (ii) follows form part (i).

Proposition 6.3. We have

(i) $C^s(X, Y) \geqq C^s(X)$ or $C^s(Y)$, $\quad s > 0$.

(ii) $C^s(Y, Z \mid X) \geqq C^s(Y \mid X)$ or $C^s(Z \mid X)$, $\quad s > 0$.

(iii) $C^s(X \mid Y) + C^s(Y \mid Z) \geqq C^s(X \mid Z)$, $\quad s \geqq 1$.

(iv) $C^s(Y|X) + C^s(Z|X) \geqq C^s(Y,Z|X), \qquad s \geqq 1.$

(v) $I^s(X, Y \wedge Z) \geqq I^s(Y \wedge Z|X), \qquad s \geqq 1.$

(vi) If $C^s(X_1, X_2) \neq 0$, then

$$\frac{C^s(X|Y)}{C^s(X,Y)} + \frac{C^s(Y|Z)}{C^s(Y,Z)} \geqq \frac{C^s(X|Z)}{C^s(X,Z)}, \qquad s \geqq 1.$$

Proof. (i) This is obvious from proposition 6.1(i).

(ii) This is obvious from proposition 6.1(v).

(iii) For all $s \geqq 1$, we have

$$\begin{aligned} C^s(X|Y) + C^2(Y|Z) &\geqq C^s(X|Y,Z) + C^s(Y|Z) \quad \text{(proposition 6.2(iii))} \\ &= C^s(X,Y|Z) \geqq C^s(X|Z). \end{aligned}$$

(iv) For all $s \geqq 1$, we have

$$\begin{aligned} C^s(Y|X) + C^s(Z|X) &\geqq C^s(Y|X,Z) + C^s(Z|X), \\ &= C^s(Y,Z|X) \quad \text{(proposition 6.1(v))}. \end{aligned}$$

(v) For all $s \geqq 1$, we have

$$I^s(Y \wedge Z|X) \leqq I^s(X \wedge Z) + I^s(Y \wedge Z|X) = I^s(X, Y \wedge Z).$$

(vi) For all $s \geqq 1$, $C^s(X_1, X_2) \neq 0$, we have

$$\begin{aligned} &\frac{C^s(X|Y)}{C^s(X,Y)} + \frac{C^s(Y|Z)}{C^s(Y,Z)} \\ &= \frac{C^s(X|Y)}{C^s(X|Y)+C^s(Y)} + \frac{C^s(Y|Z)}{C^s(Y|Z)+C^s(Z)}, \\ &\geqq \frac{C^s(X|Y)}{C^s(X|Y)+C^s(Y|Z)+C^s(Z)} + \frac{C^s(Y|Z)}{C^s(X|Y)+C^s(Y|Z)+C^s(Z)}, \\ &= \frac{C^s(X|Y)+C^s(Y|Z)}{C^s(X|Y)+C^s(Y|Z)+C^s(Z)}, \\ &\geqq \frac{C^s(X|Z)}{C^s(X|Z)+C^s(Z)} \geqq \frac{C^s(X|Z)}{C^s(X,Z)}. \end{aligned}$$

Proposition 6.4. Let

$$d_1^s(X,Y) = C^s(X|Y) + C^s(Y|X),$$

$$d_2^s(X,Y) = \frac{d_1^s(X,Y)}{C^s(X,Y)}, \qquad C^s(X,Y) \neq 0,$$

and

$$d_3^s(X, Y) = \begin{cases} \dfrac{C^s(X \mid Y)}{C^s(X)}, & C^s(X) \geqq C^s(Y) > 0, \\ \dfrac{C^s(Y \mid X)}{C^2(Y)}, & C^s(Y) \geqq C^s(X) > 0. \end{cases}$$

Then for all $\alpha = 1, 2$, and 3, we have

(i) $d_\alpha^s(X, Y) \geqq 0, \qquad d_\alpha^s(X, X) = 0, s > 0.$

(ii) $d_\alpha^s(X, Y) = d_\alpha^s(Y, X), \qquad s > 0.$

(iii) $d_\alpha^s(X, Y) + d_\alpha^s(Y, Z) \geqq d_\alpha^s(X, Z), \qquad s \geqq 1.$

This means that for $s \geqq 1$, $d_\alpha^s(X, Y)$ ($\alpha = 1, 2$, and 3) form pseudometric spaces among the random variables.

Proof. For $\alpha = 1$ and 2, the proof follows from proposition 6.3(iii) and (iv), respectively. For $s = 1$, when $\alpha = 1, 2$, and 3 refer to Horibe (1973, 1985). Let us prove the result for $\alpha = 3$. We will prove this in three different cases.

Case 1. When $C^s(X) \geqq C^s(Y) \geqq C^s(Z) > 0$, we have

$$\begin{aligned} d_3^s(X, Z) + d_3^s(Z, Y) &= \frac{C^s(X \mid Z)}{C^s(X)} + \frac{C^s(Y \mid Z)}{C^s(Y)}, \\ &\geqq \frac{C^s(X \mid Z)}{C^s(X)} + \frac{C^s(Y \mid Z)}{C^s(X)}, \\ &\geqq \frac{C^s(X \mid Z)}{C^s(X)} + \frac{C^s(Z \mid Y)}{C^s(X)} \\ &\geqq \frac{C^s(X \mid Y)}{C^s(X)} = d_3^s(X, Y). \end{aligned}$$

Case 2. When $C^s(X) \geqq C^s(Z) \geqq C^s(Y) > 0$, we have

$$\begin{aligned} d_3^s(X, Z) + d_3^s(Z, Y) &= \frac{C^s(X \mid Z)}{C^s(X)} + \frac{C^s(Z \mid Y)}{C^s(Z)}, \\ &\geqq \frac{C^s(X \mid Z) + C^s(Z \mid Y)}{C^s(X)}, \\ &\geqq \frac{C^s(X \mid Y)}{C^s(X)} = d_3^s(X, Y). \end{aligned}$$

Case 3. When $C^s(Z) \geqq C^s(X) \geqq C^s(Y) > 0$, we have

$$\begin{aligned}
d_3^s(X,Z) + d_3^s(Z,Y) &= \frac{C^s(Z\,|\,X)}{C^s(Z)} + \frac{C^s(Z\,|\,Y)}{C^s(Z)}, \\
&= \frac{C^s(X\,|\,Z) + C^s(Z\,|\,Y) + C^s(Z) - C^s(X)}{C^s(Z)}, \\
&\geqq \frac{C^s(X\,|\,Y) + C^s(Z) - C^s(X)}{C^s(Z)}, \\
&= 1 - \frac{C^s(X)}{C^s(Z)}\left[1 - \frac{C^s(X\,|\,Y)}{C^s(X)}\right], \\
&\geqq 1 - \left[1 - \frac{C^s(X\,|\,Y)}{C^s(X)}\right], \\
&= \frac{C^s(X\,|\,Y)}{C^s(X)} = d_3^s(X,Y).
\end{aligned}$$

This completes the proof of the proposition.

Proposition 6.5. The following holds:

(i) $|C^s(X) - C^s(Y)| \leqq d_1^s(X,Y), \qquad s > 0.$

(ii) $|C^s(X\,|\,Z) - C^s(Y\,|\,Z)| \leqq d_1^s(X,Y), \qquad s \geqq 1$ for any Z.

(iii) $|C^s(X_1\,|\,Y_1) - C^s(X_2\,|\,Y_2)| \leqq d_1^s(X_1,Y_1) + d_2^s(Y_1,Y_2), \qquad s \geqq 1.$

(iv) $|I^s(X_1 \wedge Y_1) - I^s(X_2 \wedge Y_2)| \leqq d_1^s(X_1,X_2) + d_2^s(Y_1,Y_2), \qquad s \geqq 1.$

Proof. We have

$$\begin{aligned}
\text{(i) } d_1^s(X,Y) &= C^s(X\,|\,Y) + C^s(Y\,|\,X), \\
&= 2C^s(X,Y) - C^s(X) - C^s(Y), \\
&\geqq \begin{cases} C^s(X) - C^s(Y), \\ C^s(Y) - C^s(X), \end{cases} \\
&= |C^s(X) - C^s(Y)|, \qquad s > 0.
\end{aligned}$$

(ii) For $s \geqq 1$, we know that

$$\begin{aligned}
C^s(X\,|\,Z) &\leqq C^s(X\,|\,Y) + C^s(Y\,|\,Z), \text{ i.e.,} \\
C^s(X\,|\,Z) - C^s(Y\,|\,Z) &\leqq C^s(X\,|\,Y), \\
&\leqq C^s(X\,|\,Y) + C^s(Y\,|\,X), = d_1^s(X,Y). \qquad (100)
\end{aligned}$$

Similarly,

$$C^s(Y \mid Z) - C^s(X \mid Z) \leqq d_1^s(X, Y). \tag{101}$$

Expressions (100) and (101) together give the required result.

(iii) For all $s \geqq 1$, we have

$$\begin{aligned} d_1^s(X_1, X_2) + d_1^s(Y_1, Y_2) &= C^s(X_1 \mid X_2) + C^s(X_2 \mid X_1) + C^s(Y_1 \mid Y_2) + C^s(Y_2 \mid Y_1), \\ &\geq C^s(X_1 \mid X_2) + C^s(X_2 \mid Y_2) + C^s(Y_1 \mid Y_2) \\ &\quad + C^s(Y_2 \mid Y_1) - C^s(X_1 \mid Y_2), \\ &= C^s(X_1 \mid X_2) + C^s(X_2 \mid Y_2) + C^s(Y_1 \mid Y_2) + C^s(Y_2 \mid Y_1) \\ &\quad - C^s(X_1 \mid Y_2) + C^s(X_1 \mid Y_1) - C^s(X_1 \mid Y_1), \\ &\geqq C^s(X_1 \mid X_2) + C^s(X_2 \mid Y_2) + C^s(Y_2 \mid Y_1) - C^s(X_1 \mid Y_1), \end{aligned} \tag{102}$$

where the last inequality is due to the fact that $C^s(X_1 \mid Y_1) + C^s(Y_1 \mid Y_2) - C^s(X_1 \mid Y_2) \geqq 0$. Thus from Eq. (102), we have

$$d_1^s(X_1, X_2) + d_1^s(Y_1, Y_2) \geqq C^s(X_2 \mid Y_2) - C^s(X_1 \mid Y_1). \tag{103}$$

Similarly,

$$d_1^s(X_1, X_2) + d_1^s(Y_1, Y_2) \geqq C^s(X_1 \mid Y_1) - C^s(X_2 \mid Y_2). \tag{104}$$

Expressions (103) and (104) together give the required result.

(iv) We know that

$$C^s(X_1 \mid X_2) + C^s(X_2) = C^s(X_1) + C^s(X_2 \mid X_1).$$

Since $C^s(X_2 \mid X_1) \geqq 0$, this gives

$$C^s(X_1 \mid X_2) \geqq C^s(X_1) - C^s(X_2). \tag{105}$$

From Eqs. (102) and (105), we have

$$\begin{aligned} &d_1^s(X_1, X_2) + d_1^s(Y_1, Y_2) \\ &\quad \geqq C^s(X_2 \mid Y_2) - C^s(X_1 \mid Y_1) + C^s(X_1) - C^s(X_2) + C^s(Y_2 \mid Y_1), \\ &\quad = C^s(Y_1 \mid Y_2) - I^s(X_2 \wedge Y_2) + I^s(X_1 \wedge Y_1) \geqq I^s(X_1 \wedge Y_1) - I^s(X_2 \wedge Y_2). \end{aligned} \tag{106}$$

Similarly,

$$d_1^s(X_1, X_2) + d_1^s(Y_1, Y_2) \geqq I^s(X_2 \wedge Y_2) - I^s(X_1 \wedge Y_1). \tag{107}$$

Expressions (106) and (107) together give the required result.

B. *Unified* (r, s)-*Conditional Entropies*

In the previous subsection, the definition of $C^s(X \mid Y)$ is based on the well-known property of Shannon's entropy, i.e., it is especially defined to satisfy the following property:

$$C^s(X, Y) = C^s(Y) + C^s(X \mid Y), \qquad s > 0. \tag{108}$$

Some authors (Sahoo, 1983; Van der Lubbe et al., 1987) extended Eq. (108) for other entropies, but it didn't give a simplified expression, as in the case of $C^s(X \mid Y)$ given in Eq. (99). In this subsection, we shall use four different ways to define the *unified* (r, s)-*conditional entropies.*

When $s = 1$ in Eq. (99), we have

$$H(X \mid Y) = \sum_{j=1}^{m} q_j H(X \mid Y = j), \tag{109}$$

where

$$H(X \mid Y = j) = -\sum_{i=1}^{n} b_{i|j} \log b_{i|j}, \qquad j = 1, 2, \ldots, m.$$

Let us replace $H(X \mid Y = j)$ given in Eq.(109) by the *unified* (r, s)-*conditional (individual) entropy* $\mathscr{E}_r^s(X \mid Y = j)$ $(j = 1, 2, \ldots, m)$. Then we have

$${}^1\mathscr{E}_r^s(X \mid Y) = \sum_{j=1}^{m} q_j \mathscr{E}_r^s(X \mid Y = j) \tag{110}$$

for all $r > 0$ and any s. More clearly, we have the following individual expressions:

$${}^1H_r^s(X \mid Y) = (2^{1-s} - 1)^{-1}\left[\sum_{j=1}^{m} q_j \left(\sum_{i=1}^{n} b_{i|j}^r\right)^{\frac{s-1}{r-1}} - 1\right],$$

$$s \neq 1, r \neq 1, r > 0, \tag{111}$$

$${}^1H_1^s(X \mid Y) = (2^{1-s} - 1)^{-1}\left[\sum_{j=1}^{n} q_j \exp_2\left((s-1)\sum_{i=1}^{n} b_{i|j} \log b_{i|j}\right) - 1\right], \qquad s \neq 1 \tag{112}$$

$${}^1H_r^1(X \mid Y) = (1 - r)^{-1} \sum_{j=1}^{n} q_j \log\left(\sum_{i=1}^{n} b_{i|j}^r\right), \qquad r \neq 1, r > 0, \tag{113}$$

$${}^1H_s^s(X \mid Y) = (2^{1-s} - 1)^{-1}\left[\sum_{j=1}^{m} q_j \sum_{i=1}^{n} b_{i|j}^s - 1\right], \qquad s \neq 1, s > 0, \tag{114}$$

$${}_t^1H(X \mid Y) = (2^{t-1} - 1)^{-1}\left[\sum_{j=1}^{m} q_j \left(\sum_{i=1}^{n} b_{i|j}^{1/t}\right)^t - 1\right], \qquad t \neq 1, t > 0. \tag{115}$$

We shall now use the expressions given in Eqs. (114) and (115) as the basis for writing an alternative way of defining unified (r, s)-conditional entropies.

Let us define

$$^2H_r^1(X \mid Y) = (1-r)^{-1} \log\left\{\sum_{j=1}^{m} q_j \sum_{i=1}^{n} b_{i|j}^r\right\}, \qquad r \neq 1, r > 0 \tag{116}$$

and

$$^3H_r^1(X \mid Y) = \frac{r}{1-r} \log\left\{\sum_{j=1}^{m} q_j \left(\sum_{i=1}^{n} b_{i|j}^r\right)^{1/r}\right\}, \qquad r \neq 1, r > 0. \tag{117}$$

The definition of $^2H_r^1(X \mid Y)$ is based on expression (114), and the definition of $^3H_r^1(X \mid Y)$ is based on expression (115). In the limiting case we have

$$\lim_{r \to 1} {}^2H_r^1(X \mid Y) = \lim_{r \to 1} {}^3H_r^1(X \mid Y) = H(X \mid Y),$$

where $H(X \mid Y)$ is as given in Eq. (109).

We shall now use expressions (116) and (117) to define the conditional entropies of order r and degree s, using the compositivity relation given in Eq. (9). These definitions are as follows:

$$\begin{aligned} {}^2H_r^s(X \mid Y) &= g_s({}^2H_r^1(X \mid Y)) \\ &= (2^{1-s}-1)^{-1}\left\{\left(\sum_{j=1}^{m}\sum_{i=1}^{n} q_j b_{i|j}^r\right)^{s-1/r-1} - 1\right\}, \qquad s \neq 1, r \neq 1, r > 0, \end{aligned} \tag{118}$$

and

$$\begin{aligned} {}^3H_r^s(X \mid Y) &= g_s({}^3H_r^1(X \mid Y)), \\ &= (2^{1-s}-1)^{-1}\left\{\left[\sum_{j=1}^{m} q_j \left(\sum_{i=1}^{n} b_{i|j}^r\right)^{1/r}\right]^{r(s-1/r-1)} - 1\right\}, \\ &\qquad s \neq 1, r \neq 1, r > 0. \end{aligned} \tag{119}$$

In the limiting case we have

$$\begin{aligned} \lim_{r \to 1} {}^2H_r^s(X \mid Y) &= \lim_{r \to 1} {}^3H_r^s(X \mid Y), \\ &= (2^{1-s}-1)^{-1}\left\{\exp_2\left((s-1)\sum_{j=1}^{m} q_j \sum_{i=1}^{n} b_{i|j} \log b_{i|j}\right) - 1\right\}, \\ &= (2^{1-s}-1)^{-1}[2^{(1-s)H(X|Y)} - 1], \\ &= {}^2H_1^s(X \mid Y) = {}^3H_1^s(X \mid Y), \qquad s \neq 1. \end{aligned}$$

Also we can check that

$$^1H_s^s(X \mid Y) = {}^2H_s^s(X \mid Y) \text{ and } {}_t^1H(X \mid Y) = {}_t^3H(X \mid Y).$$

The exact expressions of ${}_t^2H(X \mid Y)$ and ${}^3H_s^s(X \mid Y)$ are given by

$${}_t^2H(X \mid Y) = (2^{t-1} - 1)^{-1}\left\{\left(\sum_{j=1}^m q_j b_{i|j}^{1/t}\right)^t - 1\right\}, \qquad t \neq 1, t > 0 \tag{120}$$

and

$${}^3H_s^s(X \mid Y) = (2^{1-s} - 1)^{-1}\left\{\left[\sum_{j=1}^m q_j\left(\sum_{i=1}^n b_{i|j}^s\right)^{1/s}\right]^s - 1\right\}, \qquad s \neq 1, s > 0, \tag{121}$$

respectively. Expression (120) is obtained from (118) by taking $r = 1/t$ and $s = 2 - t$. Expression (121) is obtained from (119) by taking $r = s$.

We know that

$$I(X \wedge Y) = H(X) - H(X \mid Y),$$

where $I(X \wedge Y)$ is the well-known mutual information (Ash, 1965) between the random variables X and Y.

Based on the definitions of unified (r, s)-conditional entropies given above, we can generalize $I(X \wedge Y)$ in the following way:

$${}^\alpha\mathscr{N}_r^s(X \wedge Y) = \mathscr{E}_r^s(X) - {}^\alpha\mathscr{E}_r^s(X \mid Y),$$

where $\alpha = 1, 2$, and 3. By simple calculations we can write

$$I(X \wedge Y) = D(A||P^*Q),$$

where

$$D(A||P^*Q) = \sum_{i=1}^n \sum_{j=1}^m a_{ij} \log \frac{a_{ij}}{p_i q_j}$$

is a directed divergence between the distributions A and P^*Q.

We shall now present a fourth way to define the unified (r, s) conditional entropy. This is based on the generalizations of $D(A||P^*Q)$ in terms of the unified (r, s) directed divergence $\mathscr{F}_r^s(A||P^*Q)$ given in Eq. (49). This definition is as follows:

$${}^4\mathscr{E}_r^s(X \mid Y) = \mathscr{E}_r^s(X) - {}^4\mathscr{N}_r^s(X \wedge Y),$$

where

$${}^4\mathscr{N}_r^s(X \wedge Y) = \mathscr{F}_r^s(A||P^*Q).$$

Thus we can write

$${}^\alpha\mathscr{N}_r^s(X \wedge Y) = \mathscr{E}_r^s(X) - {}^\alpha\mathscr{E}_r^s(X \mid Y),$$

where

$$ {}^{\alpha}\mathscr{E}_r^s(X \mid Y) = \begin{cases} {}^{\alpha}H_r^s(X \mid Y), & r \neq 1, s \neq 1, r > 0, \\ {}^{\alpha}H_1^s(X \mid Y), & r = 1, s \neq 1, \\ {}^{\alpha}H_r^1(X \mid Y), & r \neq 1, s = 1, r > 0, \\ H(X \mid Y), & r = 1, s = 1, \end{cases} $$

and

$$ {}^{\alpha}\mathscr{N}_r^s(X \wedge Y) = \begin{cases} {}^{\alpha}I_r^s(X \wedge Y), & r \neq 1, s \neq 1, r > 0 \\ {}^{\alpha}I_1^s(X \wedge Y), & r = 1, s \neq 1, \\ {}^{\alpha}I_r^1(X \wedge Y), & r \neq 1, s = 1, r > 0, \\ I(X \wedge Y), & r = 1, s = 1, \end{cases} $$

for $\alpha = 1, 2, 3$, and 4.

Remarks. ${}^1H_r^1(X \mid Y)$ is defined in a natural way, as is Shannon's entropy. ${}^2H_r^1(X \mid Y)$ can be found in Aczél and Daróczy (1963) and Behara and Nath (1970). ${}^3H_r^1(X \mid Y)$ has been taken by Arimoto (1975) to relate it to Gallager's random coding exponent function. ${}^4H_r^1(X \mid Y)$ has been adopted by Rényi (1960) and is based on the definition of mutual information between random variables.

Based on the above definitions the following proposition holds:

Proposition 6.6. We have

(i) $\mathscr{E}_r^s(X) \geqq 0, \quad {}^{\alpha}\mathscr{E}_r^s(X \mid Y) \geqq 0 \ (\alpha = 1, 2, 3)$

(ii) $\mathscr{E}_r^s(X, Y) \geqq \mathscr{E}_r^s(X)$ or $\mathscr{E}_r^s(Y)$.

(iii) If X and Y are independent random variables, then

$$ \mathscr{E}_r^s(X, Y) = \mathscr{E}_r^s(X) + \mathscr{E}_r^s(Y) + (2^{1-s} - 1) \quad \mathscr{E}_r^s(X)\mathscr{E}_r^s(Y). $$

(iv) ${}^{\alpha}\mathscr{E}_r^s(X \mid Y) \leqq \mathscr{E}_r^s(X)$,

(e_1) for $\alpha = 1$ it is true for $(r, s) \in \Gamma_1$,

(e_2) for $\alpha = 2, 3$, and 4 it is true for all $r > 0$ and any s. (122)

(v) $$ {}^1\mathscr{E}_r^s(X \mid Y) \begin{cases} \geqq {}^2\mathscr{E}_r^s(X \mid Y), & s \leqq r, \\ \leqq {}^2\mathscr{E}_r^s(X \mid Y), & s \geqq r. \end{cases} $$

(vi) ${}^2\mathscr{E}_r^s(X \mid Y) \leqq {}^3\mathscr{E}_r^s(X \mid Y)$. (123)

Parts (i), (ii), (iii), and (vi) are true for all $r > 0$ and any s.

Proof. Parts (i), (ii), and (iii) are easy to verify. Part (iv) (e_1) follows from the concavity of $\mathscr{E}_r^s(X)$ for $(r, s) \in \Gamma_1$ given in property 8. For part (iv) (e_2) when $\alpha = 2$ and 3, it is sufficient to prove the results ${}^2H_r^1(X \mid Y) \leqq H_r^1(X)$ and ${}^3H_r^1(X \mid Y) \leqq H_r^1(X)$. The first follows from Van der Lubbe et al. (1982), and the second follows from Arimoto (1975). For $\alpha = 4$, part (iv) (e_2) holds because of the nonnegativity of $\mathscr{F}_r^s(A \| P^*Q)$ (i.e., $\mathscr{F}_r^s(P \| U)$ for all $r > 0$ and any s given in Section IV. Let us now prove parts (v) and (vi).

(v) From Lemma 4.2, we can write

$$\sum_{j=1}^{m} q_j \left(\sum_{i=1}^{n} b_{i|j}^r \right)^{s-1/r-1} \begin{cases} \geqq \left(\sum_{j=1}^{m} q_j \sum_{i=1}^{n} b_{i|j}^r \right)^{s-1/r-1}, & \dfrac{s-1}{r-1} > 1, \dfrac{s-1}{r-1} < 0 \\ \leqq \left(\sum_{j=1}^{m} q_j \sum_{i=1}^{n} b_{i|j}^r \right)^{s-1/r-1}, & 0 < \dfrac{s-1}{r-1} < 1. \end{cases} \tag{124}$$

Subtracting 1 from both sides of Eq. (124), multiplying by $(2^{1-s} - 1)^{-1}$ $(s \neq 1)$, and simplifying we get

$${}^1H_r^s(X \mid Y) \begin{cases} \geqq {}^2H_r^s(X \mid Y), & s < r, \\ \leqq {}^2H_r^s(X \mid Y), & s > r. \end{cases} \tag{125}$$

When $r = s$ in Eq. (125), we use the equality sign. Using concavity of the logarithmic function we can prove that

$${}^1H_r^1(X \mid Y) \begin{cases} \geqq {}^2H_r^1(X \mid Y), & r > 1, \\ \leqq {}^2H_r^1(X \mid Y), & 0 < r < 1. \end{cases} \tag{126}$$

Similarly, we can prove

$${}^1H_1^s(X \mid Y) \begin{cases} \geqq {}^2H_1^s(X \mid Y), & s < 1, \\ \leqq {}^2H_1^s(X \mid Y), & s > 1. \end{cases} \tag{127}$$

Combining Eqs. (125), (126), and (127) we get the required result.

(vi) Again using Lemma 4.2, we can write

$$\left[\sum_{j=1}^{m} q_j \left(\sum_{i=1}^{n} b_{i|j}^r \right) \right]^{1/r} \begin{cases} \leqq \sum_{j=1}^{m} q_j \left(\sum_{i=1}^{n} b_{i|j}^r \right)^{1/r}, & 0 < r < 1, \\ \geqq \sum_{j=1}^{m} q_j \left(\sum_{i=1}^{n} b_{i|j}^r \right)^{1/r}, & r > 1. \end{cases} \tag{128}$$

For $r = 1$, we use an equality sign in Eq. (128). Taking $\log(\cdot)$ on both sides of Eq. (128), multiplying by $(1 - r)^{-1} (r \neq 1)$, and simplifying, we get

$${}^2H_r^1(X \mid Y) \leqq {}^3H_r^1(X \mid Y), \qquad r \neq 1, r > 0. \tag{129}$$

This gives

$$g_s({}^2H_r^1(X \mid Y)) \leqq g_s({}^3H_r^1(X \mid Y)),$$

i.e.,

$$ {}^2H_r^s(X \mid Y) \leqq {}^3H_r^s(X \mid Y), \qquad r \neq 1, s \neq 1, r > 0. \tag{130}$$

When $r = 1$, we have

$$ {}^2H_1^s(X \mid Y) = {}^3H_1^s(X \mid Y), \qquad s \neq 1. \tag{131}$$

Combining Eqs. (129)–(131), we get the required result.

Proposition 6.7. We have

$$H_r^s(X, Y) \begin{cases} \leqq H_r^s(Y) + {}^1H_r^s(X \mid Y), & s \geqq r, r \cdot \dfrac{s-1}{r-1} \geqq 1, \\ \geqq H_r^s(Y) + {}^1H_r^s(X \mid Y), & r \geqq s, r \cdot \dfrac{s-1}{r-1} \leqq 1. \end{cases} \tag{132}$$

for all $r \neq 1, s \neq 1, r > 0$

Proof. We know (Behara and Chawla, 1974; Rathie and Taneja, 1989) that

$$\left(\sum_{i=1}^{m} \gamma_j\right)^{\rho} - \sum_{j=1}^{m} \gamma_j^{\rho} \begin{cases} \geqq \left(\sum_{j=1}^{m} \delta_j\right)^{\rho} - \sum_{j=1}^{m} \delta_j^{\rho}, & \rho \geqq 1, \rho \leqq 0, \\ \leqq \left(\sum_{j=1}^{m} \delta_j\right)^{\rho} - \sum_{j=1}^{m} \delta_j^{\rho}, & 0 \leqq \rho \leqq 1, \end{cases} \tag{133}$$

where $0 \leqq \delta_j \leqq \gamma_j$, $j = 1, 2, \ldots, m$. We also know (Gallager, 1968, pp. 523) that

$$\sum_{i=1}^{n} a_{ij}^r \begin{cases} \geqq \left(\sum_{i=1}^{n} a_{ij}\right)^r = q_j^r, & 0 < r \leqq 1, \\ \leqq \left(\sum_{i=1}^{n} a_{ij}\right)^r = q_j^r, & r \geqq 1 \end{cases} \tag{134}$$

for all $j = 1, 2, \ldots, m$.

Case 1. $0 < r \leqq 1$. In this case, substituting

$$\delta_j = \sum_{i=1}^{n} a_{ij}^r, \; j = 1, 2, \ldots, m,$$

$$\gamma_j = q_j^r, \; j = 1, 2, \ldots, m,$$

and

$$\rho = \frac{s-1}{r-1}, \qquad r \neq 1, s \neq 1$$

in Eq. (133), we get

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^{r}\right)^{s-1/r-1} - \sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ij}^{r}\right)^{s-1/r-1}$$

$$\begin{cases} \geqq \left(\sum_{j=1}^{m} q_j^r\right)^{s-1/r-1} - \sum_{j=1}^{m} q_j^{r(s-1/r-1)}, \\ \qquad 0 < r < 1\left(\dfrac{s-1}{r-1} \geqq 1 \text{ or } \dfrac{s-1}{r-1} < 0\right), \\ \leqq \left(\sum_{j=1}^{m} q_j^r\right)^{s-1/r-1} - \sum_{j=1}^{m} q_j^{r(s-1/r-1)}, \qquad 0 < r < 1, 0 < \dfrac{s-1}{r-1} \leqq 1, \end{cases}$$

i.e.,

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^{r}\right)^{s-1/r-1} - 1$$

$$\begin{cases} \geqq \sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ij}^{r}\right)^{s-1/r-1} - \sum_{j=1}^{m} q_j^{r(s-1/r-1)} + \left(\sum_{j=1}^{m} q_j^r\right)^{s-1/r-1} - 1, \\ \qquad 0 < r < 1\left(\dfrac{s-1}{r-1} \geqq 1 \text{ or } \dfrac{s-1}{r-1} < 0\right), \\ \leqq \sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ij}^{r}\right)^{s-1/r-1} - \sum_{j=1}^{m} q_j^{r(s-1/r-1)} + \left(\sum_{j=1}^{m} q_j^r\right)^{s-1/r-1} - 1, \\ \qquad 0 < r < 1, 0 < \dfrac{s-1}{r-1} \leqq 1. \end{cases} \tag{135}$$

Multiplying both sides of Eq. (135) by $(2^{1-s} - 1)^{-1} (s \neq 1)$ and simplifying, we get

$$H_r^s(X, Y) \begin{cases} \geqq H_r^s(Y) + V, & s \leqq r < 1, \\ \leqq H_r^s(Y) + V, & r \leqq s < 1, r < 1 < s, \end{cases} \tag{136}$$

where

$$V = (2^{1-s} - 1)^{-1}\left[\sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ij}^{r}\right)^{s-1/r-1} - \sum_{j=1}^{m} q_j^{r(s-1/r-1)}\right],$$
$$s \neq 1, r \neq 1, r > 0. \tag{137}$$

Case 2. $r > 1$. In this case, substituting

$$\delta_j = q_j^r, \qquad j = 1, 2, \ldots, m,$$

$$\gamma_j = \sum_{i=1}^{n} a_{ij}^r, \qquad j = 1, 2, \ldots, m,$$

and

$$\rho = \frac{s-1}{r-1}, \qquad s \neq 1, r \neq 1,$$

into Eq. (133), we get

$$\left(\sum_{j=1}^{m} q_j^r\right)^{s-1/r-1} - \sum_{j=1}^{m} q_j^{r(s-1/r-1)}$$
$$\begin{cases} \geqq \left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^r\right)^{s-1/r-1} - \sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ij}^r\right)^{s-1/r-1}, \\ \qquad r > 1, \left(\frac{s-1}{r-1} \geqq 1 \text{ or } \frac{s-1}{r-1} < 0\right), \\ \leqq \left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^r\right)^{s-1/r-1} - \sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ij}^r\right)^{s-1/r-1}, \qquad r > 1, 0 < \frac{s-1}{r-1} \leqq 1, \end{cases}$$

i.e.,

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^r\right)^{s-1/r-1} - 1$$
$$\begin{cases} \geqq \sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ij}^r\right)^{s-1/r-1} - \sum_{j=1}^{m} q_j^{r(s-1/r-1)} + \left(\sum_{j=1}^{m} q_j^r\right)^{s-1/r-1} - 1, \\ \qquad r > 1, \left(\frac{s-1}{r-1} \geqq 1 \text{ or } \frac{s-1}{r-1} < 0\right), \\ \leqq \sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ij}^r\right)^{s-1/r-1} - \sum_{j=1}^{m} q_j^{r(s-1/r-1)} + \left(\sum_{j=1}^{m} q_j^r\right)^{s-1/r-1} - 1, \\ \qquad r > 1, 0 < \frac{s-1}{r-1} \leqq 1. \end{cases} \tag{138}$$

Multiplying Eq. (138) by $(2^{1-s} - 1)^{-1} (s \neq 1)$ and simplifying, we get

$$H_r^s(X, Y)\begin{cases} \leqq H_r^s(Y) + V, & s \geqq r > 1, \\ \geqq H_r^s(Y) + V, & 1 < s \leqq r, s < 1 < r, \end{cases} \tag{139}$$

where V is as given in Eq. (137). Combining (136) and (139), we get

$$H_r^s(X, Y)\begin{cases} \leqq H_r^s(Y) + V, & s \geqq r, \\ \geqq H_r^s(Y) + V, & r \geqq s \end{cases} \tag{140}$$

for all $r \neq 1, s \neq 1, r > 0$.

We have

$$V = (2^{1-s} - 1)^{-1}\left[\sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ij}^{r}\right)^{s-1/r-1} - \sum_{j=1}^{m} q_j^{r(s-1/r-1)}\right],$$

$$= (2^{1-s} - 1)^{-1}\sum_{j=1}^{m} q_j^{r(s-1/r-1)}\left[\left(\sum_{i=1}^{n} b_{i|j}^{r}\right)^{s-1/r-1} - 1\right],$$

$$= \sum_{j=1}^{m} q_j^{r(s-1/r-1)} H_r^s(X \mid Y = j). \tag{141}$$

When $r = s$ in (6.43), we have

$$V = C^s(X \mid Y),$$

where $C^s(X \mid Y)$ is as given in Section VI.A. In this case we use an equality sign in Eq. (140) i.e., (140) reduces to (108) when $r = s$.

We know that

$$q_j^{r(s-1/r-1)}\begin{cases} \leqq q_j, & r\dfrac{s-1}{r-1} \geqq 1, \\ \geqq q_j, & r\dfrac{s-1}{r-1} \leqq 1 \end{cases} \tag{142}$$

for all $j = 1, 2, \ldots, m$. Expressions (140) and (142) together give the required result.

Proposition 6.8. We have

(i) $H_r^s(X, Y) \leqq H_r^s(X) + H_r^s(Y)$,

(ii) $H_r^s(X_1, X_2, \ldots, X_v) \leqq \sum_{k=1}^{v} {}^1H_r^s(X_k \mid X_1, X_2, \ldots, X_{k-1})$,

(iii) ${}^1H_r^s(Y, Z \mid X) \leqq {}^1H_r^s(Y \mid Z) + {}^1H_r^s(Z \mid X)$,

(iv) ${}^1H_r^s(Y \mid Z) \leqq {}^1H_r^s(Y \mid X) + {}^1H_r^s(X \mid Z)$

for all $s \geqq r$, $r(s - 1/r - 1) \geqq 1$, $r \neq 1$, $s \neq 1$, $r > 0$.

Proof. (i) This follows from propositions 6.6(iv) and 6.7.

(ii) This is an extension of proposition 6.7.

(iii) From Eq. (132), we can write

$${}^1H_r^s(Y, Z \mid X) \leqq {}^1H_r^s(Y \mid X) + {}^1H_r^s(Z \mid X, Y) \tag{143}$$

for all $s \geqq r$, $r(s - 1/r - 1) \geqq 1$, $(r \neq 1, s \neq 1, r > 0)$. Also, from proposition 6.6(iv) we can write

$$ {}^1H_r^s(Z \mid X, Y) \leqq {}^1H_r^s(Z \mid X) \tag{144}$$

for all $(r, s) \in \Gamma_1$. Expressions (143) and (144) together complete the proof of this part.

(iv) We know that

$$\begin{aligned} {}^1H_r^s(Y \mid Z) &\leqq {}^1H_r^s(X, Y \mid Z), \\ &\leqq {}^1H_r^s(X \mid Z) + {}^1H_r^s(Y \mid X, Z), \\ &\leqq {}^1H_r^s(X \mid Z) + {}^1H_r^s(Y \mid X) \end{aligned}$$

for all $s \geqq r$, $r(s - 1/r - 1) \geqq 1$ $(r \neq 1, s \neq 1, r > 0)$, where we have used propositions 6.6(ii) and (iv), and expression (132).

Proposition 6.9. If

$$d_r^s(X, Y) = {}^1H_r^s(X \mid Y) + {}^1H_r^s(Y \mid X),$$

then the following holds:

$$\begin{aligned} &\text{(i)}\ d_r^s(X, Z) \leqq d_r^s(X, Y) + d_r^s(Y, Z), \\ &\text{(ii)}\ |H_r^s(X) - H_r^s(Y)| \leqq d_r^s(X, Y), \\ &\text{(iii)}\ |{}^1H_r^s(X_1 \mid X_2) - {}^1H_r^s(Y_1 \mid Y_2)| \leqq d_r^s(X_1, X_2) + d_r^s(Y_1, Y_2) \end{aligned}$$

for all $s \geqq r$, $r(s - 1/r - 1) \geqq 1$ $(r \neq 1, s \neq 1, r > 0)$.

Proof. (i) This follows from proposition 6.8(iv).

(ii) We have

$$\begin{aligned} d_r^s(X, Y) &= {}^1H_r^s(X \mid Y) + {}^1H_r^s(Y \mid X), \\ &\geqq H_r^s(X, Y) - H_r^s(X) + {}^1H_r^s(X \mid Y) - H_r^s(Y) \quad \text{[proposition 6.8(i)]}, \\ &= 2H_r^s(X, Y) - H_r^s(X) - H_r^s(Y), \\ &\geqq \begin{cases} H_r^s(X) - H_r^s(Y), \\ H_r^s(Y) - H_r^s(X), \end{cases} \\ &= |H_r^s(X) - H_r^s(Y)|, \end{aligned}$$

for all $r \geqq s$, $r(s - 1/r - 1) \geqq 1$ $(r \neq 1, s \neq 1, r > 0)$.

(iii) We have

$$\begin{aligned}
&d_r^s(X_1, X_2) + d_r^s(Y_1, Y_2) \\
&\quad = {}^1H_r^s(X_1 \mid X_2) + {}^1H_r^s(X_2 \mid X_1) + {}^1H_r^s(Y_1 \mid Y_2) + {}^1H_r^s(Y_2 \mid Y_1), \\
&\quad \geqq {}^1H_r^s(X_1 \mid X_2) + {}^1H_r^s(Y_1 \mid Y_2) + {}^1H_r^s(X_2 \mid Y_2) + {}^1H_r^s(Y_2 \mid Y_1) - {}^1H_r^s(Y_2 \mid Y_1), \\
&\quad = {}^1H_r^s(X_1 \mid X_2) + {}^1H_r^s(X_2 \mid Y_2) + {}^1H(Y_1 \mid Y_2) + {}^1H_r^s(Y_2 \mid Y_1) \\
&\qquad - {}^1H_r^s(X_1 \mid Y_2) + {}^1H_r^s(X_1 \mid Y_1) - {}^1H_r^s(X_1 \mid Y_1), \\
&\quad \geqq {}^1H_r^s(X_2 \mid Y_2) - H_r^s(X_1 \mid Y_1) + {}^1H_r^s(X_1 \mid X_2) + {}^1H_r^s(Y_2 \mid Y_1), \\
&\quad \geqq {}^1H_r^s(X_2 \mid Y_2) - {}^1H_r^s(X_1 \mid Y_1) \qquad (145)
\end{aligned}$$

for all $r \geqq s$, $r(s - 1/r - 1) \geqq 1$ $(r \neq 1, s \neq 1, r > 0)$ where we have used proposition 6.8(iv). Similarly,

$$d_r^s(X_1, X_2) + d_r^s(Y_1, Y_2) \geqq {}^1H_r^s(X_1 \mid Y_1) - {}^1H_r^s(X_2 \mid Y_2). \qquad (146)$$

Combining Eqs. (145) and (146), we get the required result.

VII. Applications to Statistical Pattern Recognition

In statistical pattern recognition the key problem is feature selection. Usually the performance of a recognition system is expressed in terms of the probability of error or misclassification. The aim of feature selection is to reduce the number of features without adversely affecting error performance. The feature selection problem can be viewed as the selection of the set of features that minimizes the probability of error, P_e. The computation of P_e, unfortunately, is usually very difficult, involving the determination of the decision regions and the integration of appropriate class-conditional densities over multidimensional spaces. We are, therefore, lead to seek an auxiliary criterion for determining the relative importance of the features and to use it in the selection process. However, unless we know the connection between the auxiliary criterion and the error probability, the use of a feature set chosen according to this criterion, instead of any other, can not be justified. We are thus lead to the choice of an auxiliary criterion that provides a measure of the separability or distance between classes and that has a direct relationship with the probability of error. A variety of such measures have been proposed in the literature and bounds relating P_e to various distance and information measures can be found in Chen (1976) and Kanal (1974). The classification problem is stated as follows:

Suppose we have n pattern classes $X = \{x_1, x_2, \ldots, x_n\}$ with *a priori* probability $p_i = Pr\{X = x_i\}$, $i = 1, 2, \ldots, n$. Let the feature y on Y have a class-conditional probability density function $p(y \mid x_i)$, $i = 1, 2, \ldots, n$. We assume that p_i and $p(y \mid x_i)$ are completely known. Given a feature y on Y, we can calculate the conditional *a posteriori* probability $p(x_i \mid y)$ for each i, by the Bayes rule:

$$p(x_i \mid y) = Pr\{X = x_i \mid Y = y\} = \frac{p_i p(y \mid x_i)}{\sum_{k=1}^{n} p_k p(y \mid x_k)}, \qquad i = 1, 2, \ldots, n.$$

It is well known (Fergunson, 1967) that the decision rule that minimizes the probability of error is the Bayes decision rule, which chooses the hypotheses (pattern classes) with the largest posterior probability. Using this rule, the partial probability of error for a given $Y = y$ is expressed by

$$p(e \mid y) = 1 - \max\{p(x_1 \mid y), p(x_2 \mid y), \ldots, p(x_n \mid y)\}.$$

Prior to observing Y, the probability of error P_e, associated with X is defined as the expected probability of error, i.e.,

$$P_e = E_Y\{p(e \mid y)\} = \int_Y p(e \mid y) p(y)\, dy,$$

where $p(y) = \Sigma_{i=1}^{n} p_i p(y \mid x_i)$ is the unconditional density of Y evaluated at y.

In recent years, researchers have paid attention to the problem of bounding this probability of error for two- or multiple-class problems taking some information, divergence, and distance measures into consideration (Kailath, 1967; Kanal, 1974; Chen, 1976; Boekee and Van der Lubbe, 1979). Our aim here is to give bounds on the probability of error in terms of unified (r, s)-entropy and the distance measures given in Sections II.G and III, respectively. Some particular cases are also considered. Some bounds involving divergence measures given in Sections V.B and V.C are also given.

A. Generalized Entropies, Distance Measures, and Error Bounds

This subsection deals with the upper bounds on the probability of error in terms of the generalized entropies and distance measures given in Sections II.G and III, respectively. Analogues to the Fano-type bounds are also given. Some lower bounds on the probability of error in terms of distance measures are also presented.

Proposition 7.1. We have

$$\text{(i)}\ P_e \leqq \frac{1}{2}\, {}^1\mathscr{E}_r^s(X \mid Y), \qquad (r, s) \in \Gamma_2 \tag{147}$$

and

$$\text{(ii)}\quad {}^1\mathscr{E}_r^s(X \mid Y) \leqq \mathscr{E}_r^s\Big(\underbrace{\frac{P_e}{n-1}, \frac{P_e}{n-1}, \ldots, \frac{P_e}{n-1}}_{n-1 \text{ times}}, 1 - P_e\Big), \qquad (r,s) \in \Gamma_1, \qquad (148)$$

where

$${}^1\mathscr{E}_r^s(X \mid Y) = \int_Y \mathscr{E}_r^s(X \mid Y = y) p(y)\, dy, \qquad r > 0. \qquad (149)$$

Proof. Substituting $P(X \mid Y = y)$ for P in inequalities 3(ii) and 3(iii), we get

$$p(e \mid y) \leqq \frac{1}{2} \mathscr{E}_r^s(X \mid Y = y), \qquad (r,s) \in \Gamma_2 \qquad (150)$$

and

$$\mathscr{E}_r^s(X \mid Y = y) \leqq \mathscr{E}_r^s\left(\frac{p(e \mid y)}{n-1}, \frac{p(e \mid y)}{n-1}, \ldots, \frac{p(e \mid y)}{n-1}, 1 - p(e \mid y)\right), \qquad (151)$$

respectively. Multiplying Eqs. (150) and (151) by $p(y)$ and integrating over Y, we get (147) and (148), respectively, where Eq. (148) follows because of the concavity of $\mathscr{E}_r^s(P)$ for all $(r,s) \in \Gamma_1$.

Equation (147) gives the upper bound on the probability of error in terms of the unified (r,s)-conditional entropy. Equation (148) is a generalization of the well-known Fano inequality or Fano bound in terms of Shannon's entropy.

Using Eqs. (122) and (123) given in Section VI.B we can observe that

$$P_e \leqq \frac{1}{2} {}^1\mathscr{E}_r^s(X \mid Y) \leqq \frac{1}{2} {}^2\mathscr{E}_r^s(X \mid Y) \leqq \frac{1}{2} {}^3\mathscr{E}_r^s(X \mid Y) \qquad (152)$$

for all $s \geqq r > 0$, where ${}^2\mathscr{E}_r^s(X \mid Y)$ and ${}^3\mathscr{E}_r^s(X \mid Y)$ can be written in a similar way as was Eq. (149), using the expressions given in Section VI.B.

Thus from Eq. (152) we can conclude that the bounds obtained in terms of ${}^1\mathscr{E}_r^s(X \mid Y)$ are better for all $s \geqq r > 0$.

Proposition 7.2. We have the following bounds:

$$\text{(i)}\quad G_r^\rho(X \mid Y) \begin{cases} \leqq [(n-1)^{1-r} P_e + (1 - P_e)^r]^\rho, \\ \qquad r > 1, \rho > 0, r\rho \geq 1 \text{ (or } 0 < r < 1, \rho < 0), \\ \geqq [(n-1)^{1-r} P_e^r + (1 - P_e)^r]^\rho, \\ \qquad 0 < r < 1, \rho > 1, r\rho \leqq 1. \end{cases}$$

(ii) (e_1) $P_e \leqq 1 - G_r^\rho(X \mid Y)^{1/(r-1)\rho}$, $\quad r > 0, r \neq 1, \rho \neq 0, r\rho \leqq 1 + \rho$,

(e_2) $P_e \leqq 1 - G_r^\rho(X \mid Y), r > 1$, $\quad \rho > 0$ (or $0 < r < 1, \rho < 0$), $r\rho \geqq 1 + \rho$,

(e_3) $P_e \leqq 1 - G_r^\rho(X \mid Y)$, $\quad r > 1, \rho > 0, r\rho \geqq 1$,

(e_4) $P_e \leqq 1 - G_r^\rho(X \mid Y)$, $\quad r > 1, p > 0, r\rho \leqq 1$,

where

$$G_r^\rho(X \mid Y) = \int_Y \left[\sum_{i=1}^n p(x_i \mid y)^r\right]^\rho p(y)\,dy, \qquad r > 0, \rho \neq 0. \tag{153}$$

Proposition 7.3. We have the following bounds:

(i) $P_e \leqq 1 - T_r^\rho(X \mid Y)$, $\quad r \geqq 0, \rho \geqq 0, r \neq \rho$,

(ii) $P_e \geqq 1 - n^{1/r} T_r^\rho(X \mid Y)$, $\quad r > \rho \geqq 0$,

(iii) $P_e \geqq 1 - n^{1/\rho} T_r^\rho(X \mid Y)$, $\quad \rho > r \geqq 0$.

(iv) $P_e \geqq 1 - T_r^\rho(X \mid Y)^{r-\rho/\rho}$, $\quad r > \rho \geqq 1$,

(v) $P_e \geqq 1 - T_r^\rho(X \mid Y)^{\rho - r/r}$, $\quad \rho > r \geqq 1$,

where

$$T_r^\rho(X \mid Y) = \int_Y \left[\frac{\sum_{i=1}^n p(x_i \mid y)^r}{\sum_{i=1}^n p(x_i \mid y)^\rho}\right]^{1/r-\rho} p(y)\,dy, \qquad r \neq \rho, r \geqq 0, \rho \geqq 0. \tag{154}$$

The proof of propositions 7.2 and 7.3 is based on propositions 3.1 and 3.2, respectively, and it can be seen in Capocelli et al. (1985).

The particular cases of propositions 7.1-7.3 involve known entropies and distance measures and are as follows:

Shannon's entropy. (Chu and Chien, 1966; Hellman and Raviv, 1970). We have

$$P_e \leqq \frac{1}{2} H(X \mid Y) \tag{155}$$

and

$$H(X \mid Y) \leqq -P_e \log P_e - (1 - P_e)\log(1 - P_e) + P_e \log(n - 1), \tag{156}$$

where

$$H(X \mid Y) = -\int_Y \left[\sum_{i=1}^{n} p(x_i \mid y) \log p(x_i \mid y) \right] p(y)\, dy. \tag{157}$$

Chu and Chien (1966) studied the upper bound (155) by using Shannon's inequality. Hellman and Raviv (1970) studied the same bound using the branching property given in Section II.A. The bound (156) is the well-known Fano-type bound. It has also been studied by Chu and Chien (1966) and Kovalveski (1968).

Quadratic entropy. (Vajda, 1968). We have

$$P_e \leqq \Phi_2(X \mid Y), \tag{158}$$

and

$$P_e \geqq \frac{n-1}{n} \left\{ 1 - \sqrt{1 - \frac{n\Phi_2(X \mid Y)}{n-1}} \right\}, \tag{159}$$

where

$$\Phi_2(X \mid Y) = \int_Y \sum_{i=1}^{n} p(x_i \mid y)[1 - p(x_i \mid y)] p(y)\, dy. \tag{160}$$

If we put $r = s = 2$ in Eqs. (147), (148), and (149), we get (158), (159), and (160), respectively.

Cubic entropy. (Chen, 1976). We have

$$P_e \leqq \frac{2}{3} \Phi_3(X \mid Y), \tag{161}$$

and

$$\Phi_3(X \mid Y) \leqq 1 - \left[(1 - P_e)^3 + (n-1)\left(\frac{P_e}{n-1} \right)^3 \right], \tag{162}$$

where

$$\Phi_3(X \mid Y) = 1 - \int_Y \left[\sum_{i=1}^{n} p(x_i \mid y)^3 \right] p(y)\, dy. \tag{163}$$

If we put $r = s = 3$ in Eqs. (147), (148), and (149), we get (161), (162), and (163), respectively.

Entropy of degree s. (Devijver, 1977; Ben-Bassat, 1978; Taneja, 1983).

We have

$$P_e \leqq \frac{1}{2}\,{}^1H_s^s(X \mid Y), \qquad s \neq 1, s > 0, \tag{164}$$

and

$${}^1H_s^s(X \mid Y) \leqq (2^{1-s} - 1)^{-1}\left[(1 - P_e)^s + (n-1)\left(\frac{P_e}{n-1}\right)^s - 1\right],$$
$$s \neq 1, s > 0, \tag{165}$$

where

$${}^1H_s^s(X \mid Y) = (2^{1-s} - 1)^{-1} \int_Y \left[\sum_{i=1}^n p(x_i \mid y)^s - 1\right] p(y)\,dy,$$
$$s \neq 1, s > 0. \tag{166}$$

When $r = s$ with $s \neq 1$, and $s > 0$ in (147), (148), and (149), we get (164), (165), and (166), respectively.

Entropy of order r. (Ben-Bassat and Raviv, 1978). We have

$$P_e \leqq \frac{1}{2}\,{}^1H_r^1(X \mid Y), \qquad 0 < r < 1 \left(\text{or } r > 0, P_e \geqq \frac{1}{2}\right), r \neq 1, \tag{167}$$

and

$${}^1H_r^1(X \mid Y) \leqq (1 - r)^{-1} \log\left[(1 - P_e)^r + (n-1)\left(\frac{P_e}{n-1}\right)^r\right], 0 < r < 1, \tag{168}$$

where

$${}^1H_r^1(X \mid Y) = (1 - r)^{-1} \int_Y \left[\log\left(\sum_{i=1}^n p(x_i \mid y)^r\right)\right] p(y)\,dy, \qquad r \neq 1, r > 0. \tag{169}$$

When $s = 1$ with $r \neq 1, r > 0$ in Eqs. (147), (148), and (149), we get Eqs. (167), (168), and (169), respectively. For the bound (168), also refer to Toussaint (1977) and Taneja (1983). The first condition for the bound given in (167) is $0 < r < 1$, which is because of Eq. (147), but it has been proved independently by Ben-Bassat and Raviv (1978) that it holds for $0 < r \leqq 2$ ($r \neq 1$).

Entropy of kind t. (Boekee and Van der Lubbe, 1980; Taneja, 1982). We have

$$P_e \leqq \frac{1}{2}\,{}_t^1H(X \mid Y), \qquad t \neq 1, t > 0, \tag{170}$$

and

$$ {}^1_tH(X \mid Y) \leqq (2^{t-1}-1)^{-1}\left\{\left[(1-P_e)^{1/t}+(n-1)\left(\frac{P_e}{n-1}\right)^{1/t}\right]^t-1\right\}, \quad t \neq 1, t>0, \tag{171} $$

where

$$ {}^1_tH(X \mid Y) = (2^{t-1}-1)^{-1}\int_Y\left\{\left[\sum_{i=1}^n p(x_i \mid y)^{1/t}\right]^t-1\right\}p(y)\,dy, \qquad t \neq 1, t>0. \tag{172} $$

When $r^{-1}=t=2-s$ in Eqs. (147), (148), and (149), we get Eqs. (170), (171), and (172), respectively.

Entropy of order 1 and degree s. We have

$$ P_e \leqq \frac{1}{2}\,{}^1H_1^s(X \mid Y), \qquad s>1, \tag{173} $$

and

$$ {}^1H_1^s(X \mid Y) \leqq (2^{1-s}-1)^{-1}\{\exp_2((1-s)[-P_e\log P_e-(1-P_e)\log(1-P_e) + P_e\log(n-1)])-1\}, \qquad s>1 \tag{174} $$

where

$$ {}^1H_1^s(X \mid Y) = (2^{1-s}-1)^{-1}\int_Y\left[\exp_2\left((s-1)\sum_{i=1}^n p(x_i \mid y)\log p(x_i \mid y)\right)-1\right]p(y)\,dy, \quad s \neq 1, s>0. \tag{175} $$

When $r=1$ in Eqs. (147), (148), and (149), we get (173), (174), and (175), respectively.

Entropy of order r and degree s. (Taneja, 1985). We have

$$ P_e \leqq \frac{1}{2}\,{}^1H_r^s(X \mid Y), \qquad (r,s) \in \Gamma_2, \tag{176} $$

and

$$ {}^1H_r^s(X \mid Y) \leqq (2^{1-s}-1)^{-1}\left\{\left[(1-P_e)^r+(n-1)\left(\frac{P_e}{n-1}\right)^r\right]^{s-1/r-1}-1\right\}, \quad (r,s) \in \Gamma_1, \tag{177} $$

for all $r \neq 1$, $s \neq 1$, $r > 0$, where

$$ {}^1H_r^s(X \mid Y) = (2^{1-s} - 1)^{-1} \int_Y \left\{ \left[\sum_{i=1}^{n} p(x_i \mid y)^r \right]^{s-1/r-1} - 1 \right\} p(y)\, dy, $$
$$ r \neq 1, s \neq 1, r > 0. \tag{178} $$

The bounds (176) and (177) are the obvious consequences of the bounds (147) and (148), respectively.

Using the inequalities among the entropies given in the Section II.I.1, we can compare some of the upper bounds given above.

Bayesian distance. (Devijver, 1974). We have

$$ 1 - \sqrt{G_2^1(X \mid Y)} \leqq \frac{n-1}{n}\left[1 - \sqrt{\frac{nG_2^1(X \mid Y)}{n-1}}\right] \leqq P_e \leqq 1 - G_2^1(X \mid Y), \tag{179} $$

where

$$ G_2^1(X \mid Y) = 1 - \Phi_2(X \mid Y) = \int_Y \left[\sum_{i=1}^{n} p(x_i \mid y)^2 \right] p(y)\, dy. \tag{180} $$

The inequalities given in Eq. (179) follow from the proposition 7.2 by taking $r = 2$ and $\rho = 1$. Measure $G_2^1(X \mid Y)$ given in Eq. (180) is known as Bayesian distance (Devijver, 1974).

B. Generalized Jensen Difference Divergence Measures and Error Bounds

In Section V.B we gave different generalizations of the Jensen difference divergence measure in the discrete and finite case of the probability distributions. In the same way, we will now write some of the generalizations of the Jensen difference divergence measure between the continuous probability distributions $p(y \mid x_1)$ and $p(y \mid x_2)$. These generalizations are as follows:

$$ \begin{aligned} R = \int_Y \bigg[& \frac{p(y \mid x_1)\log p(y \mid x_2) + p(y \mid x_2)\log p(y \mid x_2)}{2} \\ & - \left(\frac{p(y \mid x_1) + p(y \mid x_2)}{2}\right)\log\left(\frac{p(y \mid x_1) + p(y \mid x_2)}{2}\right)\bigg]\, dy, \end{aligned} $$

$$ \begin{aligned} {}^2R_r^1 = [2(r-1)]^{-1} \bigg\{ & \log\left[\int_Y p(y \mid x_1)^r \left(\frac{p(y \mid x_1) + p(y \mid x_2)}{2}\right)^{1-r} dy\right] \\ & + \log\left[\int_Y p(y \mid x_2)^r \left(\frac{p(y \mid x_1) + p(y \mid x_2)}{2}\right)^{1-r} dy\right\}, \qquad r \neq 1, r > 0, \end{aligned} $$

$$^{3}R_r^1 = (r-1)^{-1}\log\left\{\int_Y \left[\frac{p(y|x_1)^r + p(y|x_2)^r}{2}\right]\left[\frac{p(y|x_1)+p(y|x_2)}{2}\right]^{1-r} dy\right\},$$

$r \neq 1, r > 0,$

$$^{2}R_1^s = [2(1-2^{1-s})]^{-1}\left\{\exp_2\left((s-1)\int_Y p(y|x_1)\log\left(\frac{2p(y|x_1)}{p(y|x_1)+p(y|x_2)}\right)dy\right)\right.$$

$$\left.+\exp_2\left((s-1)\int_Y p(y|x_1)\log\left(\frac{2p(y|x_2)}{p(y|x_1)+p(y|x_2)}\right)dy\right)-2\right\}, \qquad s \neq 1,$$

$$^{3}R_1^s = (1-2^{1-s})^{-1}[2^{(s-1)R}-1], \qquad s \neq 1,$$

$$^{2}R_r^s = [2(1-2^{1-s})]^{-1}\left\{\left[\int_Y p(y|x_1)^r\left(\frac{p(y|x_1)+p(y|x_2)}{2}\right)^{1-r} dy\right]^{s-1/r-1}\right.$$

$$\left.+\left[\int_Y p(y|x_1)^r\left(\frac{p(y|x_1)+p(y|x_2)}{2}\right)^{1-r} dy\right]^{s-1/r-1} - 2\right\},$$

$r \neq 1, s \neq 1, r > 0,$

and

$$^{3}R_r^s = (1-2^{1-s})^{-1}\left\{\left[\int_Y \left(\frac{p(y|x_1)^r + p(y|x_2)^r}{2}\right)\right.\right.$$

$$\left.\left.\times\left(\frac{p(y|x_1)+p(y|x_2)}{2}\right)^{1-r} dy\right]^{s-1/r-1} - 1\right\},$$

$r \neq 1, s \neq 1, r > 0.$

Let us write these generalizations in a unified way:

$$^{\alpha}\mathscr{V}_r^s = \begin{cases} ^{\alpha}R_r^s, & r \neq 1, s \neq 1, r > 0, \\ ^{\alpha}R_1^s, & r = 1, s \neq 1, \\ ^{\alpha}R_r^1, & r \neq 1, s = 1, r > 0, \\ R, & r = 1, s = 1, \end{cases} \tag{181}$$

where $\alpha = 2$ and 3.

We have the following relation [refer to expression (74)]

$$^{2}\mathscr{V}_r^s \begin{cases} \leqq {}^{3}\mathscr{V}_r^s, & s \leqq r, \\ \geqq {}^{3}\mathscr{V}_r^s, & s \geqq r. \end{cases} \tag{182}$$

Let us write the measures relating to ${}^3\mathscr{V}_r^s$ in a more general form for the two-class case as follows:

$$R(p_1, p_2) = \int_Y \left[(p_1 p(y|x_1)) \log(p_2 p(y|x_2)) + (p_2 p(y|x_2)) \log(p_2 p(y|x_2)) \right.$$

$$\left. - (p_1 p(y|x_1) + p_2 p(y|x_2)) \log\left(\frac{p_1 p(y|x_1) + p_2 p(y|x_2)}{2} \right) \right] dy,$$

$${}^3R_r^1(p_1, p_2) = (r-1)^{-1} \log\left\{ \int_Y [(p_1 p(y|x_1))^r + (p_2 p(y|x_2))^r] \right.$$

$$\left. \times \left[\frac{p_1 p(y|x_1) + p_2 p(y|x_2)}{2} \right]^{1-r} dy \right\}, \qquad r \neq 1, r > 0,$$

$${}^3R_1^s(p_1, p_2) = (1 - 2^{1-s})^{-1} [2^{(s-1)R(p_1, p_2)} - 1], \qquad s \neq 1,$$

and

$${}^3R_r^s(p_1, p_2) = (1 - 2^{1-s})^{-1} \left\{ \left[\int_Y ((p_1 p(y|x_1))^r + (p_2 p(y|x_2))^r) \right. \right.$$

$$\left. \left. \times \left(\frac{p_1 p(y|x_1) + p_2 p(y|x_2)}{2} \right) dy \right]^{s-1/r-1} - 1 \right\},$$

$$r \neq 1, s \neq 1, r > 0.$$

Let us write these measures in a unified way:

$${}^3\mathscr{V}_r^s(p_1, p_2) = \begin{cases} {}^3R_r^s(p_1, p_2), & r \neq 1, s \neq 1, r > 0, \\ {}^3R_1^s(p_1, p_2), & r = 1, s \neq 1, \\ {}^3R_r^1(p_1, p_2), & r \neq 1, s = 1, r > 0, \\ R(p_1, p_2), & r = 1, s = 1. \end{cases} \tag{183}$$

When $p_1 = p_2 = 1/2$ in Eq. (183), we have

$${}^3\mathscr{V}_r^s\left(\frac{1}{2}, \frac{1}{2}\right) = {}^3\mathscr{V}_r^s. \tag{184}$$

Let us simplify the measures given in Eq. (183) by using the following relations:

$$p_1 p(y|x_1) = p(y)p(x_1|y); \; p_2 p(y|x_2) = p(y)p(x_2|y);$$

$$p_1 p(y|x_1) + p_2 p(y|x_2) = p(y); \text{ and } p(x_1) + p(x_2) = 1.$$

Thus we have

$$R(p_1, p_2) = \int_Y \Big[[p(y)p(x_1 \mid y)] \log[p(y)p(x_1 \mid y)]$$

$$+ [p(y)p(x_2 \mid y)] \log[p(y)p(x_2 \mid y)] - p(y) \log \frac{p(y)}{2} \Big] dy,$$

$$= 1 - H(X \mid Y), \tag{185}$$

where

$$H(X \mid Y) = \int_Y H(X \mid Y = y) p(y)\, dy,$$

and

$$H(X \mid Y = y) = -p(x_1 \mid y) \log p(x_1 \mid y) - p(x_2 \mid y) \log p(x_2 \mid y).$$

$${}^3R_r^1(p_1, p_2) = (r - 1)^{-1} \log\left\{ \int_Y p(y)^r [p(x_1 \mid y)^r + p(x_2 \mid y)^r] \left(\frac{p(y)}{2}\right)^{1-r} dy \right\},$$

$$= 1 - (1 - r)^{-1} \left\{ \int_Y [p(x_1 \mid y)^r + p(x_2 \mid y)^r] p(y)\, dy \right\}. \tag{186}$$

Using the concavity of the logarithmic function, we can write

$$\log\left\{ \int_Y [p(x_1 \mid y)^r + p(x_2 \mid y)^r] p(y)\, dy \right\},$$

$$\geqq \int_Y \{\log[p(x_1 \mid y)^r + p(x_2 \mid y)^r]\} p(y)\, dy. \tag{187}$$

From Eqs. (186) and (187), we have

$${}^3R_r^1(p_1, p_2) \begin{cases} \leqq 1 - {}^1H_r^1(X \mid Y), & 0 < r < 1, \\ \geqq 1 - {}^1H_r^1(X \mid Y), & r > 1, \end{cases} \tag{188}$$

where

$${}^1H_r^1(X \mid Y) = \int_Y H_r^1(X \mid Y = y) p(y)\, dy$$

and

$${}^1H_r^1(X \mid Y = y) = (1 - r)^{-1} \cdot \log[p(x_1 \mid y)^r + p(x_2 \mid y)^r], \qquad r \neq 1, r > 0.$$

$$\begin{aligned}
{}^3R_1^s(p_1, p_2) &= \eta_s(R(p_1, p_2)), \\
&= \eta_s(1 - H(X \mid Y)), \\
&= (1 - 2^{1-s})^{-1}[2^{(s-1)(1-H(X|Y))} - 1], \\
&= (1 - 2^{1-s})^{-1}\left[\left(\frac{1}{2}\right)^{1-s} 2^{(1-s)H(X|Y)} - 1\right], \\
&= \left(\frac{1}{2}\right)^{1-s}[1 - {}^1H_1^s(X \mid Y)],
\end{aligned} \tag{189}$$

where

$${}^1H_1^s(X \mid Y) = \int_Y H_1^s(X \mid Y = y)p(y)\,dy$$

and

$${}^1H_1^s(X \mid Y = y) = (2^{1-s} - 1)^{-1}\left[\exp_2\left((s-1)\sum_{i=1}^{2} p(x_i \mid y)\log p(x_i \mid y)\right) - 1\right],$$

$s \neq 1$,

$$\begin{aligned}
{}^3R_r^s(p_1, p_2) &= (1 - 2^{1-s})^{-1}\left\{\left[\int_Y p(y)^r(p(x_1 \mid y)^r + p(x_2 \mid y)^r)\right.\right. \\
&\quad \left.\left.\cdot\left(\frac{p(y)}{2}\right)^{1-r} dy\right]^{s-1/r-1} - 1\right\}, \\
&= (1 - 2^{1-s})^{-1}\left\{\left(\frac{1}{2}\right)^{1-s}\left[\int_Y (p(x_1 \mid y)^r + p(x_2 \mid y)^r)\right.\right. \\
&\quad \left.\left.\cdot\, p(y)\,dy\right]^{s-1/r-1} - 1\right\}, \\
&= \left(\frac{1}{2}\right)^{1-s}\left\{1 - (2^{1-s} - 1)^{-1}\left[\left(\int_Y (p(x_1 \mid y)^r + p(x_2 \mid y)^r)\right.\right.\right. \\
&\quad \left.\left.\left.\cdot\, p(y)\,dy\right)^{s-1/r-1} - 1\right]\right\}, \qquad r \neq 1, s \neq 1, r > 0.
\end{aligned} \tag{190}$$

By Lemma 4.2, we can write

$$\left\{\int_Y [p(x_1 \mid y)^r + p(x_2 \mid y)^r]p(y)\,dy\right\}^{s-1/r-1}$$

$$\begin{cases}
\leqq \int_Y [p(x_1 \mid y)^r + p(x_2 \mid y)^r]^{s-1/r-1}p(y)\,dy, & \dfrac{s-1}{r-1} > 1 \text{ or } \dfrac{s-1}{r-1} < 0, \\
\geqq \int_Y [p(x_1 \mid y)^r + p(x_2 \mid y)^r]^{s-1/r-1}p(y)\,dy, & 0 < \dfrac{s-1}{r-1} < 1.
\end{cases} \tag{191}$$

From Eqs. (190) and (191), we obtain

$$ {}^3R_r^s(p_1, p_2) \begin{cases} \leqq \left(\frac{1}{2}\right)^{1-s} [1 - {}^1H_r^s(X \mid Y)], & s \geqq r, r \neq 1, s \neq 1, \\ \geqq \left(\frac{1}{2}\right)^{1-s} [1 - {}^1H_r^s(X \mid Y)], & r \geqq s, r \neq 1, s \neq 1, \end{cases} \tag{192} $$

where

$$ {}^1H_r^s(X \mid Y) = \int_Y H_r^s(X \mid Y = y) p(y)\, dy, $$

and

$$ H_r^s(X \mid Y = y) = (2^{1-s} - 1)^{-1} \{ [p(x_1 \mid y)^r + p(x_2 \mid y)^r]^{s-1/r-1} - 1 \}, $$
$$ r \neq 1, s \neq 1, r > 0. $$

Unifying the results given in Eqs. (185), (188), (189), and (192), we have

$$ {}^3\mathscr{V}_r^s(p_1, p_2) \begin{cases} \leqq \left(\frac{1}{2}\right)^{1-s} [1 - {}^1\mathscr{E}_r^s(X \mid Y)], & s \geqq r, \\ \geqq \left(\frac{1}{2}\right)^{1-s} [1 - {}^1\mathscr{E}_r^s(X \mid Y)], & r \geqq s, \end{cases} \tag{193} $$

where ${}^1\mathscr{E}_r^s(X \mid Y)$ is as given in Eq. (149) with $X = (x_1, x_2)$ and Y as a continuous random variable.

Based on the relations given above we shall now present some error bounds.

1. *Upper Bounds on the Probability of Error in Terms of* ${}^3\mathscr{V}_r^s(p_1, p_2)$ *and* ${}^3\mathscr{V}_r^s$

Proposition 7.4. We have

$$ P_e \leqq \frac{1}{2}[1 - 2^{1-s}\, {}^3\mathscr{V}_r^s(p_1, p_2)], \qquad s \geqq r > 0, \tag{194} $$

and

$$ P_e \leqq \frac{1}{2}[1 - 2^{1-s}\, {}^3\mathscr{V}_r^s], \qquad s \geqq r > 0, \tag{195} $$

where ${}^3\mathscr{V}_r^s(p_1, p_2)$ and ${}^3\mathscr{V}_r^s$ are given by Eqs. (183) and (181), respectively.

Proof. From Eq. (147), we have

$$ P_e \leqq \frac{1}{2}\, {}^1\mathscr{E}_r^s(X \mid Y), \qquad (r, s) \in \Gamma_2. \tag{196} $$

By Eqs. (193) and (196) we get Eq. (194), while Eq. (195) follows from Eq. (184) and (194).

2. *Lower Bounds on* ${}^3\mathscr{V}_r^s(p_1, p_2)$ *and* ${}^3\mathscr{V}_r^s$ *in Terms of the Probability of Error*

When $n = 2$ in Eq. (156), we have

$$H(X \mid Y) \leqq H(P_e), \tag{197}$$

where

$$H(P_e) = -P_e \log P_e - (1 - P_e)\log(1 - P_e).$$

From Eqs. (185) and (197), we have

$$R(p_1, p_2) \leqq 1 - H(P_e). \tag{198}$$

Using Lemma 4.1, we can write

$$\int_Y [p(x_1 \mid y)^r + p(x_2 \mid y)^r] p(y)\, dy$$

$$\begin{cases} \leqq \left[\int_Y p(x_1 \mid y) p(y)\, dy\right]^r + \left[\int_Y p(x_2 \mid y) p(y)\, dy\right]^r, & 0 < r < 1, \\ \geqq \left[\int_Y p(x_1 \mid y) p(y)\, dy\right]^r + \left[\int_Y p(x_2 \mid y) p(y)\, dy\right]^r, & r > 1. \end{cases} \tag{199}$$

Let

$$p(e \mid y) = \min\{p(x_1 \mid y), p(x_2 \mid y)\} = p(x_1 \mid y) \text{ (say)},$$

then from Eq. (199), we get

$$\int_Y [p(x_1 \mid y)^r + p(x_2 \mid y)^r] p(y)\, dy$$

$$\begin{cases} \leqq P_e^r + (1 - P_e)^r, & 0 < r < 1, \\ \geqq P_e^r + (1 - P_e)^r, & r > 1. \end{cases} \tag{200}$$

Taking $\log(\cdot)$ on both sides of Eq. (200), multiplying by $(1 - r)^{-1}(r \neq 1)$, and simplifying, we obtain

$${}^3R_r^1(p_1, p_2) \geqq 1 - H_r^1(P_e), \qquad r \neq 1, r > 0, \tag{201}$$

where

$$H_r^1(P_e) = (1 - r)^{-1} \cdot \log[P_e^r + (1 - P_e)^r], \qquad r \neq 1, r > 0.$$

When $r \to 1$, Eq. (201) reduces to Eq. (198).

We can write

$$^3R_1^s(p_1, p_2) = \eta_s[R(p_1, p_2)]$$

and

$$^3R_r^s(p_1, p_2) = \eta_s[^3R_r^1(p_1, p_2)],$$

where η_s is as given in Eq. (46).

We have

$$\begin{aligned} ^3R_1^s(p_1, p_2) &= \eta_s[R(p_1, p_2)], \\ &\geqq \eta_s[1 - H(P_e)], \\ &= (1 - 2^{1-s})^{-1}\left[\left(\frac{1}{2}\right)^{1-s} 2^{(1-s)H(P_e)} - 1\right], \\ &= \left(\frac{1}{2}\right)^{1-s}[1 - H_1^s(P_e)], \end{aligned} \tag{202}$$

where

$$H_1^s(P_e) = (2^{1-s} - 1)[2^{(1-s)H(P_e)} - 1], \qquad s \neq 1.$$

Also,

$$\begin{aligned} ^3R_r^s(p_1, p_2) &= \eta_s[^3R_r^1(p_1, p_2)], \\ &\geqq \eta_s[1 - H_r^1(P_e)], \\ &= \left(\frac{1}{2}\right)^{1-s}[1 - H_r^s(P_e)], \qquad r \neq 1, s \neq 1, r > 0, \end{aligned} \tag{203}$$

where

$$H_r^s(P_e) = (2^{1-s} - 1)^{-1}\{[P_e^r + (1 - P_e)^r]^{s-1/r-1} - 1\}, \qquad r \neq 1, s \neq 1, r > 0.$$

Combining Eqs. (198), (201), (202), and (203), we have proved the following proposition:

Proposition 7.5. We have

$$^3\mathscr{V}_r^s(p_1, p_2) \geqq \left(\frac{1}{2}\right)^{1-s}[1 - \mathscr{E}_r^s(P_e)], \tag{204}$$

where

$$\mathscr{E}_r^s(P_e) = \begin{cases} H_r^s(P_e), & r \neq 1, s \neq 1, r > 0, \\ H_1^s(P_e), & r = 1, s \neq 1, \\ H_r^1(P_e), & r \neq 1, s = 1, r > 0, \\ H(P_e), & r = 1, s = 1. \end{cases}$$

Particular case of Eq. (204)

When $p_1 = p_2 = 1/2$, then from Eqs. (204) and (184), we have

$$ {}^{3}\mathscr{V}_{r}^{s} \geqq \left(\frac{1}{2}\right)^{1-s} [1 - \mathscr{E}_{r}^{s}(P_e)]. \tag{205} $$

From relation (182) and result (205), we can also obtain

$$ {}^{2}\mathscr{V}_{r}^{s} \geqq \left(\frac{1}{2}\right)^{1-s} [1 - \mathscr{E}_{r}^{s}(P_e)], \qquad s \geqq r > 0, \tag{206} $$

where ${}^{2}\mathscr{V}_{r}^{s}$ is given in Eq. (181) for $\alpha = 2$.

From the inequalities given in Eq. (182), it is quite clear that the result, Eq. (205), is better than Eq. (206).

C. *Generalized Measure of Chernoff, Bhattacharya Distance, and the Probability of Error*

Let

$$ K_r = \int_y \frac{1}{2}[p(y|x_1)^r p(y|x_2)^{1-r} + p(y|x_1)^{1-r} p(y|x_2)^r]\, dy, \qquad r > 0. \tag{207} $$

When $r = 1/2$ in Eq. (207), we have

$$ K_{1/2} = \int_Y \sqrt{p(y|x_1)p(y|x_2)}\, dy = F, \tag{208} $$

where F is the well-known *Bhattacharya distance* or *Matusita's measure of affinity* (Matusita, 1967).

The measure

$$ \int_Y p(y|x_1)^r p(y|x_2)^{1-r}\, dy, \qquad r > 0 \tag{209} $$

is known as the Chernoff measure (Kailath, 1967). Thus, based on Eq. (209), we call K_r given in Eq. (207) the *generalized measure of Chernoff.*

Let us write K_r and F in a more general form involving the prior probabilities p_1 and p_2 given by

$$ K_r(p_1, p_2) = \frac{1}{2}\int_Y [(p_1 p(y|x_1))^r (p_2 p(y|x_2))^{1-r} + (p_2 p(y|x_2))^r (p_1 p(y|x_1))^{1-r}]\, dy, \tag{210} $$

and

$$F(p_1,p_2)=\int_Y \sqrt{(p_1 p(y|x_1))(p_2 p(y|x_2))}\,dy. \tag{211}$$

When $p_1=p_2=1/2$, we have

$$K_r\left(\frac{1}{2},\frac{1}{2}\right)=\frac{1}{2}K_r, \tag{212}$$

and

$$F\left(\frac{1}{2},\frac{1}{2}\right)=\frac{1}{2}F. \tag{213}$$

We can simplify measures (210) and (211) in the following way:

$$K_r(p_1,p_2)=\int_Y K_r(y)p(y)\,dy,$$

and

$$F(p_1,p_2)=\int_Y F(y)p(y)\,dy,$$

where

$$K_r(y)=\frac{1}{2}[p(x_1|y)^r p(x_2|y)^{1-r}+p(x_1|y)^{1-r}p(x_2|y)^r], \qquad r>0$$

and

$$F(y)=\sqrt{p(x_1|y)p(x_2|y)}.$$

Based on the above notations, we have the following proposition:

Proposition 7.6. We have

$$K_r(p_1,p_2)\begin{cases}\geqq \frac{1}{2}F(p_1,p_2)^{2(1-r)}[p_1^{2r-1}+p_2^{2r-1}], & 0<r\leqq\frac{1}{2}\text{ or } r\geqq 1,\\ \leqq \frac{1}{2}F(p_1,p_2)^{2(1-r)}[p_1^{2r-1}+p_2^{2r-1}], & \frac{1}{2}\leqq r\leqq 1.\end{cases} \tag{214}$$

Proof. We have

$$K_r(p_1,p_2)=\frac{1}{2}\int_Y [(p_1p(y|x_1))^r(p_2p(y|x_2))^{1-r} + (p_1p(y|x_1))^{1-r}(p_2p(y|x_1))^r]\,dy,$$

$$K_r(p_1,p_2) = \frac{1}{2}\int_Y \left[p_1 p(y|x_1)\left(\frac{p_2 p(y|x_2)}{p_1 p(y|x_1)}\right)^{1-r} + (p_2 p(y|x_2))\left(\frac{p_1 p(y|x_1)}{p_2 p(y|x_2)}\right)^{1-r}\right] dy,$$

$$= \frac{1}{2} p_1 E_1 \left[\left(\frac{p_2 p(y|x_2)}{p_1 p(y|x_1)}\right)^{1/2}\right]^{2(1-r)} + p_2 E_2 \left[\left(\frac{p_1 p(y|x_1)}{p_2 p(y|x_2)}\right)\right]^{1/2}\Bigg]^{2(1-r)}, \tag{215}$$

where E_1 and E_2 represent the expected values in their respective forms. Using Lemma 4.2 in Eq. (215), we have

$$K_r(p_1,p_2)\begin{cases} \geqq \frac{1}{2} p_1 \left[\int_Y p(y|x_1)\left(\frac{p_2 p(y|x_2)}{p_1 p(y|x_1)}\right)^{1/2} dy\right]^{2(1-r)} \\ \quad + \frac{1}{2} p_2 \left[\int_Y p(y|x_2)\left(\frac{p_1 p(y|x_1)}{p_2 p(y|x_2)}\right)^{1/2} dy\right]^{2(1-r)}, \\ \quad 2(1-r) \geqq 1 \text{ or } (2(1-r) \leqq 0, \\ \leqq \frac{1}{2} p_1 \left[\int_Y p(y|x_1)\left(\frac{p_2 p(y|x_2)}{p_1 p(y|x_1)}\right)^{1/2} dy\right]^{2(1-r)} \\ \quad + p_2 \left[\int_Y p(y|x_2)\left(\frac{p_1 p(y|x_1)}{p_2 p(y|x_2)}\right)^{1/2} dy\right]^{2(1-r)}, \\ \quad 0 \leqq 2(1-r) \leqq 1. \end{cases} \tag{216}$$

Simplifying Eq. (216) we obtain the required result.

Particular cases of Eq. (214)

(i) When $r = 1/2$, we have

$$K_{1/2}(p_1,p_2) = F(p_1,p_2).$$

(ii) When $p_1 = p_2 = 1/2$, we have

$$K_r \begin{cases} \geqq F^{2(1-r)}, & 0 < r \leqq \frac{1}{2}, r \geqq 1, \\ \leqq F^{2(1-r)}, & \frac{1}{2} \leqq r \leqq 1. \end{cases}$$

Proposition 7.7. We have

$$K_r(p_1,p_2) \begin{cases} \geqq K_r(P_e), & r \geqq 1, \\ \leqq K_r(P_e), & 0 < r \leqq 1, \end{cases} \tag{217}$$

where

$$K_r(P_e) = \frac{1}{2}[P_e^r(1 - P_e)^{1-r} + (1 - P_e)^r P_e^{1-r}], \qquad r > 0. \tag{218}$$

Proof. We have

$$K_r(p_1, p_2) = \int_Y K_r(y) p(y)\, dy = E_Y[K_r(y)], \tag{219}$$

where

$$\begin{aligned} K_r(y) &= \frac{1}{2}[p(x_1 \mid y)^r p(x_2 \mid y)^{1-r} + p(x_1 \mid y)^{1-r} p(x_2 \mid y)^r], \\ &= \frac{1}{2}[p(x_1 \mid y)^r (1 - p(x_1 \mid y))^r + p(x_1 \mid y)^{1-r}(1 - p(x_1 \mid y))^r], \qquad r > 0. \end{aligned}$$

Let

$$K_r(p) = \frac{1}{2}[p^r(1 - p)^{1-r} + (1 - p)^r p^{1-r}], \qquad r > 0, 0 < p < 1.$$

It is easy to verify that $K_r(p)$ is a convex function of p for $r > 1$ and a concave function of p for $0 < r < 1$. Therefore, we can write

$$E[K_r(p)] \begin{cases} \geqq A_r[E(p)], & r > 1, \\ \leqq A_r[E(p)], & 0 < r < 1, \end{cases}$$

i.e.,

$$E_Y[K_r(y)] \begin{cases} \geqq K_r[E_Y(y)], & r > 1, \\ \leqq K_r[E_Y(y)], & 0 < r < 1. \end{cases} \tag{220}$$

Also, we can write

$$K_r[E_Y(y)] = \frac{1}{2}[P_e^r(1 - P_e)^{1-r} + (1 - P_e)^r P_e^{1-r}], \qquad r > 0. \tag{221}$$

Expressions (219), (220), and (221) together give the required result.

Particular case of Eq. (217)

(i) When $p_1 = p_2 = 1/2$, then from Eqs. (217) and (212), we have

$$A_r \begin{cases} \geqq P_e^r(1 - P_e)^{1-r} + (1 - P_e)^r P_e^{1-r}, & r \geqq 1, \\ \leqq P_e^r(1 - P_e)^{1-r} + (1 - P_e)^r P_e^{1-r}, & 0 < r \leqq 1. \end{cases} \tag{222}$$

D. Generalizations of J-Divergence and the Probability of Error

In Section V. C, we presented different generalizations of J-divergence in the discrete and finite cases. This section deals with the different generalizations of J-divergence between two continuous distributions $p(y|x_1)$ and $p(y|x_2)$. These generalizations are then related to Bhattacharya distance and the probability of error. We have

$$J = \int_Y [p(y|x_1) - p(y|x_2)] \log\left(\frac{p(y|x_1)}{p(y|x_2)}\right) dy,$$

$${}^1J_r^1 = (r-1)^{-1}\left\{\log\left[\int_Y p(y|x_1)^r p(y|x_2)^{1-r}\, dy\right] + \log\left[\int_Y p(y|x_1)^{1-r} p(y|x_2)^r\, dy\right]\right\}, \qquad r \neq 1, r > 0,$$

$${}^2J_r^1 = (r-1)^{-1} 2 \log\left\{\int_Y \left[\frac{p(y|x_1)^r p(y|x_2)^{1-r} + p(y|x_1)^{1-r} p(y|x_2)^r}{2}\right] dy\right\}, \quad r \neq 1, r > 0,$$

$${}^1J_1^s = (1 - 2^{1-s})^{-1}\left\{\exp_2\left((s-1)\int_Y p(y|x_1) \log p(y|x_1)/p(y|x_2)\, dy\right) + \exp_2\left((s-1)\int_Y p(y|x_2) \log p(y|x_2)/p(y|x_1)\, dy\right) - 2\right\},$$

$${}^2J_1^s = (1 - 2^{1-s})^{-1}[2^{(s-1)J} - 1], \qquad s \neq 1,$$

$${}^1J_r^s = (1 - 2^{1-s})^{-1}\left\{\left[\int_Y p(y|x_1)^r p(y|x_2)^{1-r}\, dy\right]^{s-1/r-1} + \left[\int_Y p(y|x_1)^{1-r} p(y|x_2)^r\, dy\right]^{s-1/r-1} - 2\right\}, \qquad r \neq 1, s \neq 1, r > 0,$$

and

$${}^2J_r^s = (1 - 2^{1-s})^{-1} 2 \times \left\{\left[\int_Y \frac{p(y|x_1)^r p(y|x_2)^{1-r} + p(y|x_1)^{1-r} p(y|x_2)^r}{2}\, dy\right]^{s-1/r-1} - 1\right\}, \quad r \neq 1, s \neq 1, r > 0.$$

The measures given above can be unified in the following way:

$$
{}^{\alpha}\mathscr{W}_r^s = \begin{cases} {}^{\alpha}J_r^s, & r \neq 1, s \neq 1, r > 0, \\ {}^{\alpha}J_1^s, & r = 1, s \neq 1, \\ {}^{\alpha}J_r^1, & r \neq 1, s = 1, r > 0, \\ J, & r = 1, s = 1, \end{cases} \tag{223}
$$

where $\alpha = 1$ and 2.

The following inequalities also hold [refer to expression (92)]:

$$
{}^{1}\mathscr{W}_r^s \begin{cases} \leqq {}^{2}\mathscr{W}_r^s, & r \geqq s, \\ \geqq {}^{2}\mathscr{W}_r^s, & s \geqq r. \end{cases} \tag{224}
$$

Let us write the measures given in Eq. (223) for $\alpha = 2$ in the more general case involving prior probabilities p_1 and p_2 in the following way:

$$
{}^{2}\mathscr{W}_r^s(p_1, p_2) = \begin{cases} {}^{2}J_r^s(p_1, p_2), & r \neq 1, s \neq 1, r > 0, \\ {}^{2}J_1^s(p_1, p_2), & r = 1, s \neq 1, \\ {}^{2}J_r^1(p_1, p_2), & r \neq 1, s = 1, r > 0, \\ J(p_1, p_2), & r = 1, s = 1, \end{cases} \tag{225}
$$

where

$$
\begin{aligned}
J(p_1, p_2) &= \int_Y [p_1 p(y \mid x_1) - p_2 p(y \mid x_2)] \log\left(\frac{p_1 p(x_1 \mid y)}{p_2 p(x_2 \mid y)}\right) dy, \\
&= \int_Y [p(x_1 \mid y) - p(x_2 \mid y)] \left[\log\left(\frac{p(x_1 \mid y)}{p(x_2 \mid y)}\right) \right] p(y)\, dy,
\end{aligned}
$$

$$
\begin{aligned}
{}^{2}J_r^1(p_1, p_2) &= (r-1)^{-1} \log\left\{ \int_Y ([p_1 p(y \mid x_1)]^r [p_2 p(y \mid x_2)]^{1-r} \right. \\
&\quad \left. + [p_1 p(y \mid x_1)]^{1-r} [p_2 p(y \mid x_2)]^r)\, dy \right\}, \\
&= (r-1)^{-1} \log\left\{ \int_Y [p(x_1 \mid y)^r p(x_2 \mid y)^{1-r} \right. \\
&\quad \left. + p(x_2 \mid y)^r p(x_1 \mid y)^{1-r}] p(y)\, dy \right\}, \qquad r \neq 1, r > 0,
\end{aligned}
$$

$$
{}^{2}J_r^s(p_1, p_2) = (1 - 2^{1-s})^{-1} [2^{(s-1)J(p_1, p_2)} - 1], \qquad s \neq 1,
$$

and

$$
\begin{aligned}
{}^2J_r^s(p_1,p_2) &= (1-2^{1-s})^{-1}\Bigg\{\Bigg[\int_Y([p_1p(y\,|\,x_1)]^r[p_2p(y\,|\,x_2)]^{1-r} \\
&\quad + [p_1p(y\,|\,x_1)]^{1-r}[p_2p(y\,|\,x_2)]^r)\,dy\Bigg]^{s-1/r-1} - 1\Bigg\}, \\
&= (1-2^{1-s})^{-1}\Bigg\{\Bigg(\int_Y[p(x_1\,|\,y)^r p(x_2\,|\,y)^{1-r} \\
&\quad + p(x_2\,|\,y)^r p(x_1\,|\,y)^{1-r}]p(y)\,dy\Bigg)^{s-1/r-1} - 1\Bigg\}, \\
&\qquad r\neq 1,\ s\neq 1,\ r>0.
\end{aligned}
$$

When $p_1 = p_2 = 1/2$ in Eq. (225), we have

$$
{}^2\mathscr{W}_r^s\left(\frac{1}{2},\frac{1}{2}\right) = \frac{1}{2}\,{}^2\mathscr{W}_r^s. \tag{226}
$$

If we write

$$
J(y) = [p(x_1\,|\,y) - p(x_2\,|\,y)]\log\left(\frac{p(x_1\,|\,y)}{p(x_2\,|\,y)}\right),
$$

then

$$
J(p_1,p_2) = \int_Y J(y)p(y)\,dy.
$$

Let us consider the auxiliary functions given by

$$
{}^3J_r^1(p_1,p_2) = \int_Y J_r^1(y)p(y)\,dy,
$$

$$
{}^3J_1^s(p_1,p_2) = \int_Y J_1^s(y)p(y)\,dy.
$$

and

$$
{}^3J_r^s(p_1,p_2) = \int_Y J_r^s(y)p(y)\,dy,
$$

where

$$
\begin{aligned}
J_r^1(y) &= (r-1)^{-1}\log[p(x_1\,|\,y)^r p(x_2\,|\,y)^{1-r} + p(x_1\,|\,y)^{1-r}p(x_2\,|\,y)^r], \\
&\qquad r\neq 1,\ r>0, \\
J_1^s(y) &= (1-2^{1-s})^{-1}[2^{(s-1)J(y)} - 1], \qquad s\neq 1,
\end{aligned}
$$

and

$$J_r^s(y) = (1 - 2^{1-s})^{-1}\{[p(x_1 \mid y)^r p(x_2 \mid y)^{1-r} + p(x_1 \mid y)^{1-r} p(x_2 \mid y)^r]^{s-1/r-1} - 1\}, \qquad r \neq 1, s \neq 1, r > 0,$$

respectively.

In a unified way we can write:

$$^3\mathscr{W}_r^s(p_1, p_2) = \int_Y \mathscr{W}_r^s(y) p(y)\, dy, \tag{227}$$

where

$$\mathscr{W}_r^s(y) = \begin{cases} J_r^s(y), & r \neq 1, s \neq 1, r > 0, \\ J_1^s(y), & r = 1, s \neq 1, \\ J_r^1(y), & r \neq 1, s = 1, r > 0, \\ J(y), & r = 1, s = 1. \end{cases} \tag{228}$$

It is easy to check that the following inequalities hold:

$$^3\mathscr{W}_r^s(p_1, p_2) \begin{cases} \leqq {}^2\mathscr{W}_r^s(p_1, p_2), & r \geqq s, \\ \geqq {}^2\mathscr{W}_r^s(p_1, p_2), & s \geqq r. \end{cases} \tag{229}$$

Based on the above considerations we shall now present relations between generalizations of J-divergence, Bhattacharya distance, and the probability of error.

Proposition 7.8. We have

$$^2\mathscr{W}_r^s(p_1, p_2) \geqq \mathscr{W}_r^s(P_e), \tag{230}$$

$$^3\mathscr{W}_r^s(p_1, p_2) \geqq \mathscr{W}_r^s(P_e), \qquad s \geqq r > 0, \tag{231}$$

$$^2\mathscr{W}_r^s \geqq 2\mathscr{W}_r^s(P_e), \tag{232}$$

and

$$^1\mathscr{W}_r^s \geqq 2\mathscr{W}_r^s(P_e), \qquad s \geqq r > 0, \tag{233}$$

where

$$\mathscr{W}_r^s(P_e) = \begin{cases} J_r^s(P_e) = (1 - 2^{1-s})^{-1}[K_r(P_e)^{s-1/r-1} - 1], & r \neq 1, s \neq 1, r > 0, \\ J_1^s(P_e) = (1 - 2^{1-s})^{-1}[2^{(s-1)J(P_e)} - 1], & r = 1, s \neq 1, \\ J_r^s(P_e) = (r-1)^{-1} \log K_r(P_e), & r \neq 1, s = 1, r > 0, \\ J(P_e) = (2P_e - 1) \log\left(\dfrac{P_e}{1 - P_e}\right), & r = 1, s = 1, \end{cases} \tag{234}$$

and $K_r^s(P_e)$ is as given in Eq. (218).

The proof of Eq. (230) follows from relation (217) given in proposition 7.7. Equation (232) follows from (230) and (226) by taking $p_1 = p_2 = 1/2$. Equation (231) follows from relations (229) and (230). Equation (233) follows from relations (232) and (224).

Proposition 7.9. We have

$$
{}^2\mathscr{W}_r^s(p_1, p_2)\begin{cases} \geqq \mathscr{W}_r^s(F; p_1, p_2), & r \geqq \dfrac{1}{2}, \\ \leqq \mathscr{W}_r^s(F; p_1, p_2), & 0 < r \leqq \dfrac{1}{2} \end{cases} \tag{235}
$$

for any s, where

$$
\mathscr{W}_r^s(F; p_1, p_2) = \begin{cases} (1 - 2^{1-s})^{-1}\{[F(p_1, p_2)^{2(1-r)}(p_1^{2r-1} + p_2^{2r-1}]^{s-1/r-1} - 1\}, \\ \qquad r \neq 1, s \neq 1, r > 0, \\ (1 - 2^{1-s})^{-1} \\ \qquad \times [\exp_2((s-1)(p_1 \log p_1 + p_2 \log p_2 - F(p_1, p_2))) - 1], \\ \qquad r = 1, s \neq 1, \\ (r-1)^{-1} \log\{F(p_1, p_2)^{2(1-r)}[p_1^{2r-1} + p_2^{2r-1}]\}, \\ \qquad r \neq 1, s = 1, r > 0, \\ 2[p_1 \log p_1 + p_2 \log p_2 - \log F(p_1, p_2)], \\ \qquad r = 1, s = 1, \end{cases}
$$

and $F(p_1, p_2)$ is as given in Eq. (211).

The proof follows from inequalities (214) given by the proposition 7.6.

Particular case of Eq. (235).

When $p_1 = p_2 = 1/2$, we have

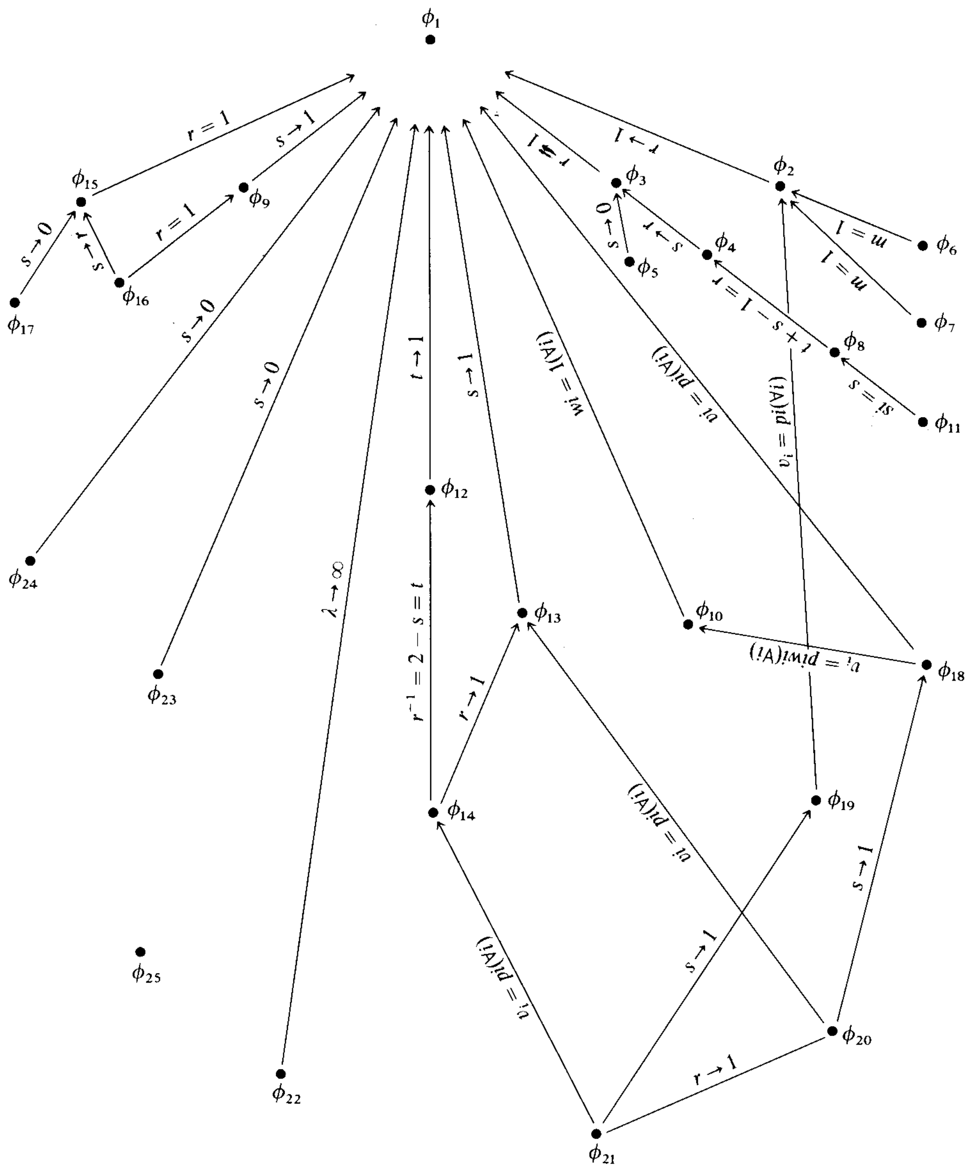
$$
\mathscr{W}_r^s\left(F; \frac{1}{2}, \frac{1}{2}\right) = \begin{cases} (1 - 2^{1-s})^{-1}[F^{2(1-s)} - 1], & s \neq 1, \\ -2 \log F, & s = 1, \end{cases}
$$

where F is as given in Eq. (208).

The particular cases of the propositions 7.8 and 7.9 for $r = 1$ and $s = 1$ can be seen in Toussaint (1974) and in Devijver and Kittler (1981).

Entropy Graph

The following graph indicates how all the entropies given in Section II.F reduce to Shannon's case in the limiting or in the particular case:



REFERENCES

Aczél, J., and Daróczy, Z. (1963). *Publications Mathematicae* **10**, 171–190.

Aczél, J., and Daróczy, Z. (1975). "On Measures of Information and their Characterizations," Academic Press, New York.

Arimoto, S. (1971). *Information and Control* **19**, 181–190.

Arimoto, S. (1975). *Colloq. on Information Theory*, Kesthely, Hungary 41–52.

Arimoto, S. (1976). *IEEE Trans. on Inform. Theory* **IT-20**, 460–473.

Ash, R. (1965). "Information Theory," Interscience New York.

Behara, M., and Chawla, J. M. S. (1974). *In* "Entropy and Ergodic Theory: Selecta Statistica Canadiana." **II**, 15–38.

Behara, M., and Nath, P. (1970), *In* "Probability and Information Theory II" (M. Behara, K. Krickeberg, and J. Wolfowitz, eds.). Springer Verlag, Berlin, pp. 102–137.

Belis, M., and Guiasu S., (1968). *IEEE Trans. on Inform. Theory* **IT-14**, 591–592.

Ben-Bassat, B. (1978). *Information and Control* **39**, 227–242.

Ben-Bassat, B., and Raviv, J. (1978). *IEEE Trans. on Inform. Theory* **IT-24**, 324–331.

Blumer, A. (1982). Ph.D. Thesis, University of Illinois at Urbana-Champaign, Department of Mathematics.

Boekee, D. E., and van der Lubbe, J. C. A. (1979). *Pattern Recognition* **11**, 353–360.

Boekee, D. E., and van der Lubbe, J. C. A. (1980). *Information and Control* **45**, 136–155.

Burbea, J. (1984). *Utilitas Mathematica* **26**, 171–192.

Burbea, J., and Rao, C. R. (1982). *IEEE Trans. on Inform. Theory* **IT-28**, 489–495.

Campbell, L. L. (1965). *Information and Control* **23**, 423–429.

Campbell, L. L, (1985). *Information Sciences* **25**, 199–210.

Campbell, L. L. (1987). Queen's Mathematical Preprint, No. 12.

Capocelli, R. M., and Taneja, I. J. (1984). *Proc. IEEE Intern. Conf. on Systems, Man and Cybernetics*, Oct. 9–12, Halifax, Canada, pp. 43–47.

Capocelli, R. M., and Taneja, I. J. (1985). *Cybernetics and Systems* **16**, 341–376.

Capocelli, R. M., Gargano, L., Vaccaro, U., and Taneja, I. J. (1985). *Proc. IEEE Intern. Conf. on Systems, Man and Cybernetics*, Arizona, U.S.A., November 12–15, pp. 78–82.

Capocelli, R. M., de Santis, A., and Taneja, I. J. (1988). *IEEE Trans. on Inform. Theory* **IT-34**, 134–138.

Chen, C. H., (1976). *Information Sciences* **10**, 159–171.

Chu, J. T. and Chueh, J. E. (1966). *J. Franklin Inst.* **282**, 121–125.

Csiszár, I. (1972). *Periodica Math. Hung.* **2**, 191–213.

Csiszár, I. (1974). *Trans. of the 7th Prague Conf.*, pp. 83–86, Prague; Chechoslovakia.

Csiszár, I. and Kŏrner, J. (1981). "Information Theory: Coding Theorems for Discrete Memoryless System." Academic Press, New York.

Daróczy, Z. (1970). *Information and Control* **16**, 36–51.

Devijver, P. A. (1974). *IEEE Trans. on Comp.* **C-23**, 70–80.

Devijver, P. A. (1977). *Information and Control* **34**, 222–226.

Devijver, P. A., and Kittler, J. V. (1982). "Pattern Recognition: A Statistical Approach." Prentice Hall, London.

Ferentinos, K., and Papaioannou, T. (1983). *J. Comb. Inform. and Syst. Sci.* **8**, 286–294.

Fergunson, T. S. (1967). "Mathematical Statistics." pp. 284–308, Academic Press, New York.

Ferreri, C. (1980). *Statistica* **XL**, 155–168.

Gallager, R. G., (1968). "Information Theory and Reliable Communication. John Wiley and Sons, New York.

Gallager, R. G. (1978). *IEEE Trans. on Inform. Theory* **IT-29**, 668–674.

Guiasu S. (1977). "*Information Theory with Applications.*" McGraw Hill, New York.
Györfi, L., and Nemetz, T. (1975). *Colloq. on Inform. Theory*, Keszthely, Hungary, pp. 309–331.
Hardy, G. H., Littlewood, J. E., and Pólya, G. (1934). "*Inequalities.*" Cambridge University Press, London.
Hartley, R. V. L. (1928). *Bell System Tech. J.* **7**, 535–563.
Hellman, M. E., and Raviv, J. (1970). *IEEE Trans. on Inform. Theory* **IT-16**, 368–372.
Horibe, Y. (1973). *Information and Control* **22**, 403–404.
Horibe, Y. (1985). *IEEE Trans. on Systems, Man, and Cybernetics* **SMC-15**, 641–642.
Jeffreys, H. (1946). *Proc. Royal Soc.*, **A186**, 453–561.
Jelinek, F. (1968a). "*Probabilistic Information Theory.*" McGraw Hill, New York.
Jelinek, F. (1968b). *IEEE Trans. on Inform. Theory* **IT-18**, 765–774.
Jelinek, F., and Schneider, K. (1972). *IEEE Trans. on Inform. Theory* **IT-18**, 765–774.
Kailath, T. (1967). *IEEE Trans. on Commun. Tech.* **COM-15**, 52–60.
Kanal, L. N. (1974). *IEEE Trans. on Inform. Theory* **IT-20**, 687–722.
Kapur, J. N. (1967). *The Math. Seminar* **4**, 78–94.
Kapur, J. N. (1983). *J. Inform. and Optim. Sci.* **4**, 207–232.
Kapur, J. N. (1986). *Indian J. Pure & Appl. Math.* **17**, 429–449.
Kerridge, D. F. (1961). *J. Royal Statist. Soc.* **B23**, 184–194.
Kieffer, J. C. (1979). *Information and Control* **41**, 136–146.
Kovalevski, V. A. (1968). *In* "Character Readers and Pattern Recognition." pp. 3–30, V. A. Kovalevsky, Ed., New York: Spartan.
Kullback, S. and Leibler, R. A. (1951). *Ann. Math. Statist.* **22**, 79–86.
Longo, G. (1980). "Information Theory" (in Italian. Boringhieri, Torino, Italy.
Mangasarian, O. L. (1969). "Nonlinear Programming," Tata McGraw Hill, New Delhi/Bombay.
Marshall, A. W., and Olkin, I. (1979). "Inequalities: Theory of Majorization and Its Application," Academic Press, New York.
Mathai, A. M., and Rathie, P. N. (1975). "Basic Concepts in Information Theory and Statistic." Wiley and Sons, New York.
Matusita, K. (1967). *Ann. Inst. Statist. Math.* **19**, 181–192.
McEliece, R. J. (1977). "The Theory of Information and Coding." Encyclopedia of Mathematics and its Applications, Vol. 3. Addison Wesley, Reading, Massachusetts.
Nath, P. (1975). *Information and Control* **29**, 234–242.
Nyquist, H. (1924). *Bell Syst. Tech. J.* **3**, 324–.
Nyquist, H. (1928). *AIEEE Trans.* **47**, 617–.
Parker, J., D. S. (1979). *SIAM J. Comput.* **9**, 470–489.
Picard, C. F. (1979). *J. Comb. Inform. and Syst. Sci.* **4**, 343–356.
Rao, C. R. (1982). *Theor. Popul. Biology* **21**, 24–43.
Rathie, P. N. (1970). *J. Appl. Probl.* **7**, 124–133.
Rathie, P. N. and Sheng, L. T. (1981). *J. Comb. Inform. and Syst. Sci.* **6**, 197–205.
Rathie, P. N., and Taneja, I. J. (1989). *Information Sciences*, to appear.
Rényi, A. (1960). *MTA III Osztályának Kozl* **10**, 251–282.
Rényi, A. (1961). *Proc. 4th Berk. Symp. Math. Statist. and Probl.*, Vol. 1, pp. 547–461, University of California Press. Berkeley, California.
Sahoo, P. K. (1983). *J. Comb. Inform. and Syst. Sci.* **8**, 263–270.
Sant'anna, A. P. Taneja, I. J. (1985). *Information Sciences* **35**, 145–155.
Shannon, C. E. (1948). *Bell System Tech. J.* **27**, 379–423; 623–656.
Sharma, B. D., and Mittal, D. P. (1975). *J. Math. Sci.* **10**, 28–40.
Sharma, B. D., and Mittal, D. P. (1977). *J. Comb. Inform. and Syst. Sci* **2**, 122–133.
Sharma, B. D., and Taneja, I. J. (1975). *Metrika* **22**, 205–215.
Sharma, B. D., and Taneja, I. J. *Elect. Inform. Kybern.* **13**, 419–433.

Shiva, S. S. G., Ahmed, N. U., and Georganas, N. D. (1973). *J. Appl. Probl.* **10**, 666–670.
Sibson, R. (1969). *Z. Wahrs. und Verw Geb.* **14**, 149–160.
Taneja, I. J. (1975). PhD. Thesis, University of Delhi.
Taneja, I. J. (1979). *J. Comb. Inform. and Syst. Sci.* **4**, 253–274.
Taneja, I. J. (1982). *Proc. IEEE Intern. Conf. on Cybern. and Soc.*, Washington, D.C., October 28–30, pp. 463–466.
Taneja, I. J. (1983a). *IEEE Trans. on Systems, Man, and Cybernetics*, **SMC-13**, 241–242.
Taneja, I. J. (1983b). *J. Comb. Inform. & Syst. Sci.* **8**, 206–212.
Taneja, I. J. (1984a). *Matemática Aplicada e Computacional* **3**, 199–204.
Taneja, I. J. (1984b). *J. Comb. Inform. and Syst. Sci.* **9**, 169–174.
Taneja, I. J. (1985a). *Pattern Recognition Letters* **3**, 361–368.
Taneja, I. J. (1985b). *Proc. Intern. Conf. on Telecommunication and Control*, Rio de Janeiro, Brazil, December 9–12, pp. 48–51.
Taneja, I. J. (1986a). *Information Sciences* **39**, 211–216.
Taneja, I. J. (1986b). *J. Comb. Inform. & Syst. Sci.* **11**, 99–109.
Taneja, I. J. (1987). *Statistical Planning and Inference* **16**, 137–145.
Taneja, I. J. (1988a). *Information Sciences, to appear.*
Taneja, I. J. (1988b). *Tamkang J. Math.* **19**.
Taneja, I. J. (1988c). *Tamkang J. Math.* **19**.
Toussaint, G. T. (1974). *Proc. Second Intern. Joint Conf. on Pattern Recog.*, Copenhagen, Denmark.
Toussaint, G. T. (1977). *IEEE Trans. Systems, Man, and Cybernetics* **SMC-7**, 300–302.
Trouborst, P. M., Backer, E., Boekee, D. E., and Boxma Y. (1974). *Proc. Second Intern. Joint Conf. on Pattern Recog.*, Copenhagen, Denmark.
Vajda, I. (1968). *Inform. Trans. Problems* **4**, 9–19.
van der Lubbe, J. C. A. (1978). *Proc. 8th Prague Conf.*, pp. 253–266. Prague, Chechoslovakia.
van der Lubbe, J. C. A., Boxma, Y. and Boekee, D. E. (1984). *Information Sciences* **32**, 187–215.
van der Lubbe, J. C. A., Boekee, D. E. and Boxma, Y. (1987). *Information Sciences* **41**, 139–169.
van der Pyl, T. (1977). *Colloq. Intern. du C.N.R.S.*, No. 276, Teorie de l'Information, Cahan, France, 4–8 July, pp. 161–171.
Varma, R. S. (1966). *J. Math. Sci.* **1**, 34–48.
Wiener, N. (1948). "Cybernetics." M.I.T. Press, Cambridge.
Wyner, A. D. (1972). *Information and Control* **20**, 176–181.