The Independence Inequality and Its Application to Information Theory

ABU-BAKR EL-SAYED

Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada

The "independence inequality" is introduced, and its motivation and importance in the field of information theory are explained. One way of stating this inequality in words is that the information we get from two experiments will be greatest when the two experiments are independent. Different examples of probability distributions and communication channels for which the independence inequality holds are presented, where we consider different information measures including Shannon entropy, Rényi entropies (entropies of order \( \alpha \)), the generalized entropies (entropies of degree \( \alpha \)), entropies of order \((\alpha, \beta)\), and entropies of degree \((\alpha, \beta)\). It is also shown that Shannon entropy satisfies this inequality for all probability distributions, while any of the other measures satisfies it only for certain distributions. The relation between the independence inequality and some of the known properties (such as the additivity and subadditivity properties) of measures of information is discussed.

INTRODUCTION

In order to introduce the "independence inequality" and give its motivation we need first mention some definitions. We shall use the following notations.

- \( \xi, n \): Finite discrete random variables (of the input and the output respectively)
- \( i, j, \ldots, n \): Positive integers
- \( \{p_i\}, \{q_k\} \): (Marginal) probability distributions (of \( \xi \) and \( n \), respectively)
- \( \{\pi_{ik}\} \): The joint probability distribution (of the two-dimensional random variable \( (\xi, \eta) \))
- \( \{q_{ik}\} \): The conditional probability distribution (of \( \eta \) given \( \xi \)).
- \( H(\{p_i\}) = H(\xi) \): Shannon information measure (Shannon entropy)
- \( \alpha, \beta \): Real numbers
- \( H(\{p_i\}) \): Entropy of order \( \alpha \) (Rényi entropy)
- \( H_\alpha(\{p_i\}) \): Entropy of order \( \alpha, \beta \)
- \( H^\alpha(\{p_i\}) \): Entropy of degree \( \alpha \) (generalized entropy)
Entropy of degree $\alpha, \beta$

$J(\{p_i\})$  
Unspecified entropy, i.e., a function of the probability distribution $\{p_i\}$

$\log$  
logarithm to the base 2

$\Gamma_n$  
Set of complete $n$-ary probability distributions,

\[ \Gamma_n = \{(p_1, p_2, \ldots, p_n) | \sum_{i=1}^{n} p_i = 1, p_i \geq 0, i = 1, 2, \ldots, n; n = 2, 3, \ldots \} \]

**Definition 1.** The following entropies are defined for the set $\Gamma_n$ of complete probability distributions.

A. **Shannon entropy.**

\[
H_n(p_1, \ldots, p_n) = - \sum_{i=1}^{n} p_i \log p_i,
\]

where, for zero probabilities, we use the definition $0 \log 0 := 0$.

B. **Entropy of order $\alpha$, also called Rényi entropy.**

\[
\alpha H_n(p_1, \ldots, p_n) = \frac{1}{1 - \alpha} \log \sum_{i=1}^{n} p_i^\alpha, \alpha \neq 1, 0^\alpha := 0 \text{ for all } \alpha.
\]

C. **Entropy of order $(\alpha, \beta)$.**

\[
\alpha, \beta H_n(p_1, \ldots, p_n) = \frac{1}{\beta - \alpha} \log \frac{\sum_{i=1}^{n} p_i^\alpha}{\sum_{i=1}^{n} p_i^\beta}, \alpha \neq \beta, 0^\gamma := 0 \text{ for all } \gamma.
\]

D. **Entropy of degree $\alpha$:**

\[
H_n^\alpha(p_1, \ldots, p_n) = \frac{1}{2^{1-\alpha} - 1} \left( \sum_{i=1}^{n} p_i^\alpha - 1 \right), \alpha \neq 1, 0^\alpha := 0 \text{ for all } \alpha.
\]

E. **Entropy of degree $(\alpha, \beta)$.**

\[
H_n^{\alpha, \beta}(p_1, \ldots, p_n) = \frac{1}{2^{1-\alpha} - 2^{1-\beta}} \sum_{i=1}^{n} (p_i^\alpha - p_i^\beta), \alpha \neq \beta, 0^\gamma := 0 \text{ for all } \gamma.
\]

(Note: $\alpha, \beta H = \alpha H$ and $H^{\alpha, 1} = H^\alpha$.)

**Definition 2.** Let the discrete constant channel with the space $X = \{x_1, \ldots, x_n\}$ of input symbols and the space $Y = \{y_1, \ldots, y_m\}$ of output symbols be characterized by the $(n, m)$-transition matrix.

\[
Q = [q_{ik}] \quad (i = 1, \ldots, n; k = 1, \ldots, m) \quad \text{with} \quad q_{ik} \geq 0,
\]

\[
\sum_{k=1}^{m} q_{ik} = 1 \quad (i = 1, \ldots, n).
\]
Assume that an arbitrary input probability distribution \((p_1, ..., p_n) \in \Gamma_n\) induces the output distribution \((q_1, ..., q_m) \in \Gamma_m\). The spaces of input and output symbols can be considered as the space of values for discrete random variables \(\xi\) and \(\eta\), respectively, where the distributions of \(\xi\) and \(\eta\) are given by:

\[
P(\xi = x_i) = p_i \quad (i = 1, 2, ..., n),
\]

\[
P(\eta = y_k) = q_k = \sum_{i=1}^{n} p_i q_{ik} \quad (k = 1, 2, ..., m),
\]

\[
P(\eta = y_k \mid \xi = x_i) = q_{ik} \quad (i = 1, 2, ..., n; k = 1, 2, ..., m),
\]

\[
P(\xi = x_i, \eta = y_k) = \pi_{ik} \quad (i = 1, 2, ..., n; k = 1, 2, ..., m).
\]

**Definition 3.** A sequence of functions \(J_n : \Gamma_n \rightarrow R \quad (n = 2, 3, \ldots)\) is said to satisfy the independence inequality if

\[
J_{nm}(p_{11}, p_{12}, \ldots, p_{1m}; p_{21}, p_{22}, \ldots, p_{2m}; \ldots; p_{n1}, p_{n2}, \ldots, p_{nm}) \leq J_{nm}(p_1 q_1, p_1 q_2, \ldots, p_1 q_m; p_2 q_1, p_2 q_2, \ldots, p_2 q_m; \ldots; p_n q_1, p_n q_2, \ldots, p_n q_m)
\]

for all \(n, m\) and for all \((p_{11}, p_{12}, \ldots, p_{1m}; p_{21}, p_{22}, \ldots, p_{2m}; \ldots; p_{n1}, p_{n2}, \ldots, p_{nm}) \in \Gamma_{nm}\).

**Definition 4.** The following two properties of entropies are defined for a sequence of functions \(J_n : \Gamma_n \rightarrow R \quad (n = 2, 3, \ldots)\) for complete probability distributions (cf. Aczél and Daróczy, 1975).

A. **Additivity.**

\[
J_{nm}(p_1 q_1, p_1 q_2, \ldots, p_1 q_m; p_2 q_1, p_2 q_2, \ldots, p_2 q_m; \ldots; p_n q_1, p_n q_2, \ldots, p_n q_m) = J_n(p_1, p_2, \ldots, p_n) + J_m(q_1, q_2, \ldots, q_m)
\]

for all \(n \geq 2, m \geq 2, (p_1, p_2, \ldots, p_n) \in \Gamma_n, (q_1, q_2, \ldots, q_m) \in \Gamma_m\).

In other words, this property says that the information obtained from two independent experiments is the sum of the informations yielded by the individual experiments.

B. **Subadditivity.**

\[
J_{nm}(\pi_{11}, \pi_{12}, \ldots, \pi_{1m}; \pi_{21}, \pi_{22}, \ldots, \pi_{2m}; \ldots; \pi_{n1}, \pi_{n2}, \ldots, \pi_{nm}) \leq J_n \left( \sum_{k=1}^{m} \pi_{1k}, \sum_{k=1}^{m} \pi_{2k}, \ldots, \sum_{k=1}^{m} \pi_{nk} \right) + J_m \left( \sum_{i=1}^{n} \pi_{i1}, \sum_{i=1}^{n} \pi_{i2}, \ldots, \sum_{i=1}^{n} \pi_{im} \right)
\]

for all \(n \geq 2, m \geq 2, ([\pi_{11}, \pi_{12}, \ldots, \pi_{1m}; \pi_{21}, \pi_{22}, \ldots, \pi_{2m}; \ldots; \pi_{n1}, \pi_{n2}, \ldots, \pi_{nm}) \in \Gamma_{nm}\.\)
This property means that the information we get from two (not necessarily independent) experiments is not greater than the sum of the informations we get from the individual experiments.

The additivity property can be written in the abbreviated form

\[ J_{nm}(p_i q_k) = J_n(p_i) + J_m(q_k). \]

Also, since

\[ p_i = \sum_{k=1}^{m} \pi_{ik} \quad \text{and} \quad q_k = \sum_{i=1}^{n} \pi_{ik}, \]

therefore the subadditivity property can be stated in the abbreviated form

\[ J_{nm}(\pi_{ik}) \leq J_n(p_i) + J_m(q_k). \]

Hence, additivity and subadditivity imply

\[ J_{nm}(\pi_{ik}) \leq J_{nm}(p_i q_k), \]

which is the abbreviated form of the independence inequality given in Definition 3. In fact, the independence inequality states that the information we get from two experiments will be greatest when the two experiments are independent.

The following results emphasize the importance of the independence inequality and give different examples of probability distributions and channels for which this inequality is satisfied. We start by formally stating the above discussion.

**Theorem 1.** An additive sequence of functions \( J_n: \Gamma_n \to \mathbb{R} \) \((n = 2, 3, \ldots)\) is subadditive if and only if it satisfies the independence inequality.

**Proof.** Using an abbreviated notation, we have the following.

The additivity of \( J_n \) implies:

\[ J_{nm}(p_i q_k) = J_n(p_i) + J_m(q_k). \tag{1} \]

The subadditivity implies:

\[ J_{nm}(\pi_{ik}) \leq J_n(p_i) + J_m(q_k). \tag{2} \]

And the independence inequality says:

\[ J_{nm}(\pi_{ik}) \leq J_{nm}(p_i q_k). \tag{3} \]

As noticed before, the independence inequality is a consequence of additivity and subadditivity. Also, Eq. (1) together with the inequality (3) imply Eq. (2).

The Shannon entropy, which is both additive and subadditive, (see, for example, Feinstein, 1958), clearly satisfies the independence inequality for any
probability distribution. All the other entropies ($\alpha H$, $H^\alpha$, $\alpha_\beta H$ and $H^{\alpha,\beta}$, at least for some $\alpha, \beta$) will be shown to satisfy this inequality for some, but not all, probability distributions. If an entropy satisfies the independence inequality and is also additive, for some probability distributions, then, according to the above theorem, it will be subadditive as well for these probability distributions. It is also to be emphasized here that an entropy does not have to be additive or subadditive in order to satisfy the independence inequality (e.g., $H^\alpha$ is subadditive (Daróczy 1970), $\alpha H$ is not, but see the following lemma).

**Lemma 1.** Let $\alpha \in \mathbb{R}$, $\alpha \neq 1$. Then, for a given probability distribution, the independence inequality is either satisfied by both $H^\alpha$ and $\alpha H$ or is satisfied by neither. In other words, if $H^\alpha$ satisfies the independence inequality for a certain probability distribution, then for the same distribution $\alpha H$ also does, and vice versa.

**Proof.**

$$H^\alpha(p_1, \ldots, p_n) = \mu \left( \sum_i p_i^\alpha - 1 \right), \mu = (2^{1-\alpha} - 1)^{-1},$$

$$\alpha H(p_1, \ldots, p_n) = \bar{\mu} \cdot \log \sum_i p_i^\alpha, \bar{\mu} = 1/(1 - \alpha).$$

The independence inequality for $H^\alpha$ says:

$$\mu \left( \sum_{i,k} \pi_{ik}^\alpha - 1 \right) \leq \mu \left( \sum_{i,k} (p_i q_k)^\alpha - 1 \right)$$

i.e.,

$$\mu \sum_{i,k} \pi_{ik}^\alpha \leq \mu \sum_{i,k} (p_i q_k)^\alpha$$

And for $\alpha H$, the independence inequality gives:

$$\bar{\mu} \log \sum_{i,k} \pi_{ik}^\alpha \leq \bar{\mu} \log \sum_{i,k} (p_i q_k)^\alpha$$

It is clear that the last two inequalities are equivalent (i.e., each one implies the other) since $x \to \log x$ is an increasing function and both $\mu$ and $\bar{\mu}$ have the same sign at each value of $\alpha$.

In fact, the relation between $H^\alpha(p_1, \ldots, p_n)$ and $\alpha H(p_1, \ldots, p_n)$ can be expressed in the form:

$$\alpha H(p_1, \ldots, p_n) = \bar{\mu} \log \left( 1 + \frac{1}{\mu} H^\alpha(p_1, \ldots, p_n) \right),$$

i.e.,

$$\alpha H(p_1, \ldots, p_n) = F(H^\alpha(p_1, \ldots, p_n)).$$
where

\[ F(x) = \bar{\mu} \log(1 + (x/\mu)) , \]

which is an increasing function of \( x \).

In the following propositions, it will be enough for us to prove (or disprove) the independence inequality for either \( H^\alpha \) or \( \alpha H \), and the proof for the other follows immediately by using the above lemma. (Usually, the proof for \( H^\alpha \) will be given.)

**Proposition 1.** There exist \( \alpha, \beta \) and probability distributions such that the entropies \( H^\alpha, \alpha H, \alpha, \beta H \), and \( H^{\alpha, \beta} \) do not satisfy the independence inequality.

**Proof.** To prove this statement it is enough to give a counter example to this inequality.

(i) \( H^\alpha \): Let

\[ \mu := (2^{1-\alpha} - 1)^{-1}, \]

\[ H^\alpha_{nm}(\pi_{11}, \ldots, \pi_{1m}; \ldots; \pi_{n1}, \ldots, \pi_{nm}) = \mu \left( \sum_{i,k} \pi_{ik}^\alpha - 1 \right), \]

\[ H_{nm}(p_1 q_1, \ldots, p_1 q_m; \ldots; p_n q_1, \ldots, p_n q_m) = \mu \left( \sum_{i,k} (p_i q_k)^\alpha - 1 \right). \]

Let \( \alpha = 2, n = 2 = m, p_1 = 0.1, p_2 = 0.9. \)

\[ (\pi_{ik}) = \begin{pmatrix} 0.02 & 0.08 \\ 0.27 & 0.63 \end{pmatrix}, \]

\[ q_1 = \sum_{i=1}^{2} \pi_{i1} = 0.02 + 0.27 = 0.29, \]

\[ q_2 = \sum_{i=1}^{2} \pi_{i2} = 0.08 + 0.63 = 0.71 (= 1 - q_1), \]

\[ \sum_{i,k} \pi_{ik}^\alpha = 0.02^2 + 0.08^2 + 0.27^2 + 0.63^2 = 0.4766, \]

\[ \sum_{i,k} (p_i q_k)^\alpha = \sum_i p_i^2 \sum_k q_k^\alpha = (0.1^2 + 0.9^2)(0.29^2 + 0.71^2) = 0.4823. \]

Since, for \( \alpha = 2, \mu = -2 < 0 \), therefore

\[ H_{nm}(\pi_{11}, \ldots, \pi_{nm}) > H_{nm}(p_1 q_1, \ldots, p_n q_m). \]

(ii) See Lemma 1.
(iii) $\alpha, \beta H$: The independence inequality, in this case, is equivalent to
\[
\frac{1}{\beta - \alpha} \log \sum_{i,k} \pi_{ik}^\alpha \leq \frac{1}{\beta - \alpha} \log \sum_{i,k} (p_i q_k)^\alpha / \sum_{i,k} (p_i q_k)^\beta,
\]
which implies
\[
\sum_{i,k} \pi_{ik}^\alpha / \sum_{i,k} \pi_{ik}^\beta \geq \sum_{i,k} (p_i q_k)^\alpha / \sum_{i,k} (p_i q_k)^\beta \quad \text{for } \alpha > \beta.
\]

Let $\alpha = 2$, $\beta = 1.5$ and again use the previous probability distribution (of case (i)) to get
\[
\sum_{i,k} \pi_{ik}^\alpha = 0.4766, \sum_{i,k} (p_i q_k)^\alpha = 0.4823 \quad \text{as before},
\]
\[
\sum_{i,k} \pi_{ik}^\beta = 0.021.5 + 0.081.5 + 0.271.5 + 0.631.5 = 0.666,
\]
\[
\sum_{i,k} (p_i q_k)^\beta = \left( \sum_i p_i \right) \left( \sum_k q_k \right) = (0.111.5 + 0.911.5)(0.291.5 + 0.711.5) = 0.670.
\]

LHS of Eq. (4) = $\sum_{i,k} \pi_{ik}^\alpha / \sum_{i,k} \pi_{ik}^\beta = 0.4766/0.666 = 0.7155$.

RHS of Eq. (4) = $\sum_{i,k} (p_i q_k)^\alpha / \sum_{i,k} (p_i q_k)^\beta = 0.4823/0.670 = 0.72 > \text{LHS}$.

(iv) $H^{\alpha, \beta}$: The independence inequality is equivalent to:
\[
\frac{1}{2^{1-\alpha} - 2^{1-\beta}} \sum_{i,k} (\pi_{ik}^\alpha - \pi_{ik}^\beta) \leq \frac{1}{2^{1-\alpha} - 2^{1-\beta}} \sum_{i,k} ((p_i q_k)^\alpha - (p_i q_k)^\beta).
\]

Considering the same values of $\alpha$ and $\beta$ as in (iii), and the same probability distribution (which is the same as in (i)), we get
\[
\frac{1}{2^{1-\alpha} - 2^{1-\beta}} < 0,
\]
\[
\sum_{i,k} (\pi_{ik}^\alpha - \pi_{ik}^\beta) = 0.4766 - 0.666 = -0.1894,
\]
\[
\sum_{i,k} ((p_i q_k)^\alpha - (p_i q_k)^\beta) = 0.4823 - 0.670 = -0.1877,
\]
and again we see that the independence inequality is not satisfied.

*Remark.* In fact, Proposition 1 for the entropies $\alpha H$ (and thus also, by Lemma 1, for $H^\alpha$) with any $\alpha$ follows from Theorem 1 and from Rényi's counter-examples (Rényi, 1970)) to the subadditivity of these entropies.
Let us now give an example of a class of probability distribution for which the entropies $H_\alpha$ and $H_\alpha^a$ ($\alpha \gg 0, \alpha \neq 1$) do satisfy the independence inequality. Later, this example will be generalized to a more general class of probability distributions.

**Example 1.** Consider the doubly uniform probability distribution

$$(q_{ik}) = \begin{pmatrix}
1 - p & \frac{p}{n-1} & \frac{p}{n-1} & \ldots & \frac{p}{n-1} \\
\frac{p}{n-1} & 1 - p & \frac{p}{n-1} & \ldots & \frac{p}{n-1} \\
\frac{p}{n-1} & \frac{p}{n-1} & 1 - p & \frac{p}{n-1} & \ldots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{p}{n-1} & \ldots & \ldots & \ldots & \frac{p}{n-1} \\
\frac{p}{n-1} & \ldots & \ldots & \ldots & \frac{p}{n-1} \\
\frac{p}{n-1} & \ldots & \ldots & \ldots & \frac{p}{n-1} \\
1 - p & \frac{p}{n-1} & \frac{p}{n-1} & \ldots & \frac{p}{n-1}
\end{pmatrix}_{(n,n)}$$

where

$$0 \leq p \leq 1,$$

and let

$$p_i = \frac{1}{n} \quad \text{for all } i.$$

Then,

$$\pi_{ik} = p_i q_{ik} = \frac{1}{n} q_{ik},$$

$$q_k = \sum_{i=1}^{n} \pi_{ik} = \frac{1}{n} \sum_{i=1}^{n} q_{ik} = \frac{1}{n},$$

$$p_i q_k = \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2} \quad \text{for all } i, k.$$

For $H_\alpha$, the independence inequality says:

$$\frac{1}{2^{1-\alpha} - 1} \left( \sum_{i,k} \pi_{ik}^\alpha - 1 \right) \leq \frac{1}{2^{1-\alpha} - 1} \left( \sum_{i,k} (p_i q_k)^\alpha - 1 \right),$$

i.e.,

$$\pi \cdot \sum_{i,k} \pi_{ik}^\alpha \leq \mu \cdot \sum_{i,k} (p_i q_k)^\alpha; \quad (6)$$
where

\[ \mu = (2^{1-\alpha} - 1)^{-1}. \]

\[ \sum_{i,k} (p_i q_{ik})^\alpha = \sum_{i,k} \left( \frac{1}{n^2} \right)^\alpha = n^2 \cdot \left( \frac{1}{n^2} \right)^\alpha = (n^2)^{1-\alpha} = (n^{1-\alpha})^2, \]

\[ \sum_{i,k} q_{ik}^\alpha = \sum_{i,k} \left( \frac{q_{ik}}{n} \right)^\alpha = \frac{1}{n^\alpha} \sum_{i,k} q_{ik}^\alpha \]

\[ = \frac{1}{n^\alpha} \left( n(1 - p)^\alpha + (n^2 - n) \left( \frac{p}{n-1} \right)^\alpha \right) \]

\[ = \frac{1}{n^\alpha} \left( n(1 - p)^\alpha + n(n - 1) \frac{p^\alpha}{(n-1)^\alpha} \right) \]

\[ = n^{1-\alpha}((1 - p)^\alpha + p^\alpha(n - 1)^{1-\alpha}). \]

Let

\[ f(p) := (1 - p)^\alpha + p^\alpha(n - 1)^{1-\alpha}. \]

Then

\[ f'(p) = \alpha(1 - p)^{\alpha-1} - 1 + \alpha p^{\alpha-1}(n - 1)^{1-\alpha}, \]

\[ f''(p) = \alpha(\alpha - 1)[(1 - p)^{\alpha-2} + (n - 1)^{1-\alpha} \cdot p^{\alpha-2}], \]

\[ f'(p) = 0 \text{ if and only if } (1 - p)^{\alpha-1} = p^{\alpha-1}(n - 1)^{1-\alpha} \]

\[ \iff \left( \frac{p}{1-p} \right)^{\alpha-1} = (n - 1)^{\alpha-1} \iff \frac{p}{1-p} = n - 1 \]

\[ \iff p = \frac{n-1}{n}, \]

i.e., \( f(p) \) has a critical point at \( p_c = (n - 1)/n \) and \( f(p_c) = n^{1-\alpha} \).

(i) If \( 0 \leq \alpha < 1 \), then \( \mu > 0 \) and \( p_c \) is a maximum point \( (f''(p_c) < 0) \). Thus, the inequality (6), which is equivalent to

\[ \mu \cdot n^{1-\alpha} f(p) \leq \mu \cdot (n^{1-\alpha})^3 \]

is satisfied.

(ii) If \( \alpha > 1 \), then \( \mu < 0 \) and \( p_c \) is a minimum point, and again Eq. (6) is satisfied.

Hence, Eq. (6) is always true for all \( \alpha \geq 0, \alpha \neq 1 \), which was to be shown.
Proposition 2. The independence inequality is satisfied by the following set of entropies:

\[
\begin{align*}
E_1 & \left( i \right) - H^a, \quad \alpha \geq 0, \quad \alpha \neq 1 \\
E_2 & \left( ii \right) - \sum H_i, \quad \alpha \geq 0, \quad \alpha \neq 1 \\
E_3 & \left( iii \right) - a \beta H, \quad 0 \leq \alpha < 1 < \beta \quad \text{(or } 0 \leq \beta < 1 < \alpha) \\
E_4 & \left( iv \right) - H^\alpha \beta, \quad 0 \leq \alpha < 1 < \beta \quad \text{(or } 0 \leq \beta < 1 < \alpha)
\end{align*}
\]

for all probabilities that give rise to an equiprobable "output" distribution \((q_k = 1/m, k = 1, 2, \ldots, m)\).

Proof.

(i) \(H^a\):

(a) Let \(0 \leq \alpha < 1\). Thus, \(\mu = 1/(2^{1-a} - 1) > 0\).

\[
\sum \frac{1}{m} q_{ik}^\alpha \leq \left( \sum \frac{1}{m} q_{ik} \right)^\alpha \quad \text{for all } i = 1, 2, \ldots, n,
\]

(by the concavity of the function \(x \mapsto x^\alpha, \quad 0 \leq \alpha < 1\)). Since \(\sum q_{ik} = 1, \quad i = 1, 2, \ldots, n\), therefore

\[
\sum q_{ik}^\alpha \leq m \left( \frac{1}{m} \right)^\alpha = \sum q_k^\alpha \quad \text{for all } i.
\]

Multiplying by \(p_i\) and summing over \(i\), we get

\[
\sum p_i q_{ik}^\alpha \leq \sum p_i \sum q_k^\alpha,
\]

\[
\sum (p_i q_{ik})^\alpha \leq \sum (p_i q_k)^\alpha,
\]

\[
\mu \left( \sum (\pi_{ik}^\alpha - 1) \right) \leq \mu \left( \sum (p_i q_k)^\alpha - 1 \right).
\]

which is the independence inequality for \(H^a\).

(b) Let \(\alpha > 1\). Thus, \(\mu < 0\), and now we get

\[
\sum q_{ik}^\alpha \geq \sum q_k^\alpha \quad \forall i,
\]

\[
\sum (p_i q_{ik})^\alpha \geq \sum (p_i q_k)^\alpha,
\]

\[
\mu \left( \sum (\pi_{ik}^\alpha - 1) \right) \leq \mu \left( \sum (p_i q_k)^\alpha - 1 \right).
\]
(ii) See Lemma 1.

(iii) \( a, b H \):

\[
0 \leq \alpha < 1 \Rightarrow \sum_{i,k} \pi_{ik}^\alpha \leq \sum_{i,k} (p_i q_k)^\alpha,
\]
\[
\beta > 1 \Rightarrow \sum_{i,k} \pi_{ik}^\beta \geq \sum_{i,k} (p_i q_k)^\beta,
\]
\[
\frac{\sum_{i,k} \pi_{ik}^\alpha}{\sum_{i,k} \pi_{ik}^\beta} \leq \frac{\sum_{i,k} (p_i q_k)^\alpha}{\sum_{i,k} (p_i q_k)^\beta},
\]
\[
\frac{1}{\beta - \alpha} \log \frac{\sum_{i,k} \pi_{ik}^\alpha}{\sum_{i,k} \pi_{ik}^\beta} \leq \frac{1}{\beta - \alpha} \log \frac{\sum_{i,k} (p_i q_k)^\alpha}{\sum_{i,k} (p_i q_k)^\beta},
\]

(iv) \( H^{a, b} \):

\[
0 \leq \alpha < 1 \Rightarrow \sum_{i,k} \pi_{ik}^\alpha \leq \sum_{i,k} (p_i q_k)^\alpha,
\]
\[
\beta > 1 \Rightarrow \sum_{i,k} \pi_{ik}^\beta \geq \sum_{i,k} (p_i q_k)^\beta,
\]
\[
\sum_{i,k} (\pi_{ik}^\alpha - \pi_{ik}^\beta) \leq \sum_{i,k} ((p_i q_k)^\alpha - (p_i q_k)^\beta),
\]
\[
\delta \sum_{i,k} (\pi_{ik}^\alpha - \pi_{ik}^\beta) \leq \delta \sum_{i,k} ((p_i q_k)^\alpha - (p_i q_k)^\beta)
\]

where \( \delta = (2^{1-\alpha} - 2^{1-\beta})^{-1} \).

Remark 1. The class of probabilities discussed in Example 1 is a subclass of the probabilities considered in the above proposition, in which no condition on the “input” probabilities \( (p_i) \) or the conditional probabilities \( (q_{ik}) \) is assumed. In fact, this proposition involves, among others, all channels which are uniform from the output (i.e., the columns of the transition matrix are permutations of the same set of \( n \) numbers). In such channels, a uniform input probability distribution \( (p_i = 1/n; i = 1, 2, \ldots, n) \) results in a uniform output probability distribution \( (q_k = 1/m; k = 1, 2, \ldots, m) \), because \( q_k = \sum_i p_i q_{ik} = (1/n) \sum_i q_{ik} = (1/n) \cdot (n/m) = (1/m) \). The previously mentioned example represented a channel which is uniform both from the output and the input (doubly uniform). The next proposition deals with channels that are uniform from the input (i.e., the rows of the transition matrix are permutations of the same set of \( m \) numbers).

Remark 2. Neither of the following entropies satisfies the independence inequality for all probabilities yielding a uniform “output” distribution \( (q_k = 1/m, k = 1, 2, \ldots, m) \):
(i) \( H^{\alpha} \), \( \alpha < 0 \);
(ii) \( \alpha H \), \( \alpha < 0 \);
(iii) \( \alpha, \beta H \), \( \alpha < 0 \leq \beta < 1 \) (or \( \beta < 0 \leq \alpha < 1 \))
or \( \alpha \leq 0 < \beta < 1 \) (or \( \beta \leq 0 < \alpha < 1 \));
(iv) \( H_{\alpha, \beta} \), \( \alpha < 0 \leq \beta < 1 \) (or \( \beta < 0 \leq \alpha < 1 \))
or \( \alpha \leq 0 < \beta < 1 \) (or \( \beta \leq 0 < \alpha < 1 \)).

The proof goes along lines similar to those of the proof of the proposition, and we only remark that for \( \alpha < 0 \), the function \( x \rightarrow x^{\alpha} \) is convex \( ((x^{\alpha})'' = \alpha(\alpha - 1)x^{\alpha-2} > 0) \) and we have (with \( q_k = 1/m \) for all \( k \)):

\[
\sum_k q_{ik}^{\alpha} \geq \sum_k q_k^{\alpha} \quad \forall i = 1, 2, \ldots, n.
\]

Also, since the \((\alpha, \beta)\) entropies \((\alpha, \alpha H)\) and \(H_{\alpha, \beta}\) are defined for \( \alpha \neq \beta \), we should note that, if \( \alpha = 0 \) is included, then \( \beta = 0 \) should be excluded and vice versa.

**Proposition 3.** Let the rows of the probability matrix \((q_{ik})_{n \times m}\) be permutations of the same set of \( m \) numbers. Then the following entropies satisfy the independence inequality.

(i) \( H^{\alpha} \), \( \alpha \geq 0 \), \( \alpha \neq 1 \);
(ii) \( \alpha H \), \( \alpha \geq 0 \), \( \alpha \neq 1 \);
(iii) \( \alpha, \beta H \), \( 0 \leq \alpha < 1 \leq \beta \) (or \( 0 \leq \beta < 1 \leq \alpha \));
(iv) \( H_{\alpha, \beta} \), \( 0 \leq \alpha < 1 \leq \beta \) (or \( 0 \leq \beta < 1 \leq \alpha \)).

**Proof.** Let such a set of \( m \) numbers be \((a_1, a_2, \ldots, a_m)\), and let \( \sum_{i=1}^{m} a_i^{\alpha} = A \).

\[
\sum_k q_{ik}^{\alpha} = \sum_{i=1}^{n} a_i^{\alpha} = A, \quad i = 1, 2, \ldots, n;
\]

\[
\sum_k \sum_j p_jq_{jk}^{\alpha} = \sum_k \sum_j p_j a_j^{\alpha} = \sum_j p_j \cdot A = A;
\]

\[
\sum_k q_{ik}^{\alpha} = \sum_k \sum_j p_j a_j^{\alpha}, \quad i = 1, 2, \ldots, n;
\]

For \( 0 \leq \alpha < 1 \),

\[
\sum_k \sum_j p_j a_j^{\alpha} \leq \sum_k \left( \sum_j p_j a_j^{\alpha} \right) \sum_k q_k^{\alpha};
\]

\[
\sum_k q_{ik}^{\alpha} \leq \sum_k q_k^{\alpha} \quad 0 \leq \alpha < 1.
\]
Then the proofs of both (i) and (ii) can be completed as in the previous proposition.

Also, for \( \beta > 1 \), we conclude: \( \sum_k q_{ik}^\beta \geq \sum_k q_{ik}^\beta \), which leads to: \( \sum_{i,k} \pi_{ik}^\beta \geq \sum_{i,k} (p_{ik} q_{ik})^\beta \), and again the proofs of both (iii) and (iv) can be completed as was done in the previous proposition.

We can easily prove the following remark.

Remark 1. Neither of the following entropies satisfies the independence inequality, when the rows of the probability matrix \((q_{ik})_{n,m}\) are permutations of the same set of \( m \) numbers:

(i) \( H^\alpha \), \( \alpha < 0 \),
(ii) \( _\alpha H \), \( \alpha < 0 \),
(iii) \( _{\alpha,\beta} H \), \( \alpha < 0 \leq \beta < 1 \),
\quad or \( \alpha \leq 0 < \beta < 1 \),
(iv) \( H^{\alpha,\beta} \), \( \alpha < 0 \leq \beta < 1 \),
\quad or \( \alpha \leq 0 < \beta < 1 \).

Remark 2. We notice that, in the proof of the above proposition, the row-permutation property of the matrix \((q_{ik})\) (i.e., the rows are permutations of the same set of numbers) was only useful in getting the equation

\[ \sum_k q_{ik}^\alpha = A \text{ "Const." for all } i = 1, 2, \ldots, n \]

(and later, \( \sum_k q_{ik}^\alpha = \text{Const.}; i = 1, 2, \ldots, n \)).

In fact, all of the proof (of (i) and (ii) in particular) remains true for any conditional probability distribution \((q_{ik})\) which satisfies the above equation, but not necessarily the row-permutation property. Later we will elaborate on such distributions and their existence. In other words, the above proposition can be generalized for parts (i) and (ii) to the following proposition.

**Proposition 4.** Let \( \alpha \geq 0, \alpha \neq 1 \). If the elements of the conditional probability matrix \((q_{ik})_{n,m}\) satisfy the equation \( \sum_{k=1}^{m-1} q_{ik}^\alpha = A \) for all \( i = 1, 2, \ldots, n \), where \( A \) is some constant, then the entropies \( _\alpha H \) and \( H^\alpha \) satisfy the independence inequality.

For every given \( \alpha (\alpha \geq 0, \alpha \neq 1) \) there exists an infinite number of conditional probability distributions, each of which satisfies the above equation, and which can be found as follows. For \( y_k \geq 0, k = 1, \ldots, m \), \( m \equiv \sum_{k=1}^{m-1} y_k = 1 \), the function \( f(y_1, y_2, \ldots, y_{m-1}) := f(y_1 \ldots y_{m-1}, y_m) := \sum_{k=1}^{m-1} y_k^\alpha = \sum_{k=1}^{m-1} y_k^\alpha + y_m^\alpha = \sum_{k=1}^{m-1} y_k^\alpha + (1 - \sum_{k=1}^{m-1} y_k)^\alpha \) represents a continuous surface "S", in the \( m \)-dimensional space.

The function

\[ f(y_1 \ldots y_m) = \sum_{k=1}^{m} y_k^\alpha, \]
under the condition $\sum_k y_k = 1$, has a maximum and a minimum which are,
respectively, given by

$$f(y_1, \ldots, y_m) = \max(1, m^{1-\alpha})$$

and

$$\bar{f} = \min(y_1, \ldots, y_m) = \min(1, m^{1-\alpha}).$$

(For example, if $0 \leq \alpha < 1$, then $\bar{f} = m^{1-\alpha}$ (corresponding to the uniform
distribution $y_k = 1/m, k = 1, \ldots, m$) and $\bar{f} = 1$). This can be easily shown using,
for example, the Lagrangian method, or by just partially differentiating the function

$$f(y_1, \ldots, y_m) = \sum_{k=1}^{m-1} y_k^\alpha + \left(1 - \sum_{k=1}^{m-1} y_k\right)^\alpha$$

w.r.t. the individual elements $y_k$; also, this is clear from the maximality and the nonnegativity of R\'enyi entropies. Hence, the function $g$ has a max $\bar{g} = \bar{f}$ and
a min $\bar{g} = \bar{f}$. The intersection of the surface $S$ with the plane $g(y_1, \ldots, y_{m-1}) = A$, where $\bar{g} < A < \bar{g}$ gives a curve (whose projection on the plane $g = 0$ is a curve)
with an infinite number of points, each having $(m - 1)$ coordinates $(y_1, \ldots, y_{m-1})$.
Choose any $n$ of these points (possibly with repetitions) and let the coordinates
of the $j$th point (where $j = 1, 2, \ldots, n$) be called $(q_{j,1}, \ldots, q_{j,m-1})$, i.e., the $n$ points
are $(q_{11}, \ldots, q_{1,m-1}), (q_{21}, \ldots, q_{2,m-1}), \ldots, (q_{n1}, \ldots, q_{n,m-1})$.

(Note: In the previous proposition these points are chosen so that the $n$-tuples
$(q_{i1}, \ldots, q_{im}), i = 1, 2, \ldots, n$ are symmetric, i.e., the coordinates are permutations
of the same set of numbers).

Now, the elements of the $(n, m)$-matrix $(q_{jk})_{n,m}$ where

$$q_{i,m} = 1 - \sum_{k=1}^{m-1} q_{i,k}, \quad j = 1, \ldots, n$$

satisfy the condition $\sum_{k=1}^{m-1} q_{jk}^\alpha = A, \ j = 1, 2, \ldots, n$, and for all probability
matrices $(q_{jk})_{n,m}$ chosen in such a way, the independence inequality is satisfied
by the entropies $sH$ and $H^\alpha$.

In Proposition 2 we considered the case where the “output” distribution
$(q_k, k = 1, 2, \ldots, m)$ is uniform. Now, we will consider the case of a uniform
“input” distribution $(p_i = 1/n, i = 1, 2, \ldots, n)$ and will see that the same set
of entropies will also satisfy the independence inequality.

**Proposition 5.** The independence inequality is satisfied by the set $E$ of entropies
(defined in Proposition 2) if the (input) probability distribution $(p_i, i = 1, 2, \ldots, n)$
is uniform.

**Proof.** Let

$$L := \sum_{i,k} (p_i q_{ik})^\alpha = \left(\frac{1}{n}\right)^\alpha \sum_{i,k} q_{ik}^\alpha,$$
\[ R := \sum_{i,k} (p_i q_{ik})^\alpha = \left(\frac{1}{n}\right)^\alpha \sum_{i,k} q_{ik}^\alpha = \left(\frac{1}{n}\right)^\alpha \cdot n \sum_k q_k^\alpha = \left(\frac{1}{n}\right)^\alpha \cdot n^1 \sum_k \left(\sum_i q_{ik}\right)^\alpha. \]

(a) For \(0 \leq \alpha < 1\):

\[ \left(\sum_i \frac{1}{n} q_{ik}\right)^\alpha \geq \sum_i \frac{1}{n} q_{ik}^\alpha, \]

\[ \sum_i q_{ik}^\alpha \leq n^{1-\alpha} \left(\sum q_{ik}\right)^\alpha, \]

\[ \sum_{i,k} q_{ik}^\alpha \leq n^{1-\alpha} \sum_k \left(\sum q_{ik}\right)^\alpha. \]

Hence, \( L \leq R \) and, since \((0 \leq \alpha < 1) 1 - \alpha > 0 \) and \(\mu > 0 \) (where \(\mu = (2^{1-\alpha} - 1)^{-1}\)), therefore the independence inequality is satisfied by \(\pi H\) and \(H^2\).

(b) For \(\alpha > 1\): \(1 - \alpha < 0\), \(\mu < 0\), and \(L > R\); and again the inequality in question is satisfied. Thus, the proof for the set \(E_1\) (defined in Proposition 2) is completed. The proof for the set \(E_2\) is similar to that in Proposition 2.

**Corollary.** The independence inequality is satisfied by the set \(E\) of entropies if the joint probability distribution \((\pi_{ik}, i = 1, 2, \ldots, n; k = 1, 2, \ldots, m)\) is uniform.

**Proof.** A uniform joint probability distribution \((\pi_{ik} = (1/\text{n}m) \forall i, k)\) results in an equiprobable \((p_i)\) distribution.

\[ p_i = \sum_k \pi_{ik} = \sum_{k=1}^m \frac{1}{\text{n}m} = m \cdot \frac{1}{\text{n}m} = \frac{1}{n} \forall i. \]

The following statement can be considered as a corollary to any of the Propositions 2 and 3.

**Corollary.** If the conditional probability distribution \((q_{ik}, i = 1, 2, \ldots, n; k = 1, 2, \ldots, m)\) is uniform, then the set \(E\) of entropies satisfies the independence inequality.

**Proof.** It is clear that the row-permutation property of the matrix \((q_{ik})_{n,m}\) is satisfied. Also,

\[ q_k = \sum_i \pi_{ik} = \sum_i p_i q_{ik} = \sum_i p_i \cdot \frac{1}{m} = \frac{1}{m} \sum_i p_i = \frac{1}{m}. \]
We can summarize the cases, we have discussed, in which the independence inequality is satisfied by the entropies $H^\alpha$, $\alpha H$, $\alpha J H$ and $H^{\alpha, \beta}$ in the following theorem.

**Theorem 2.** The independence inequality is satisfied by the following set of entropies:

(i) $H^\alpha$, $\alpha \geq 0$, $\alpha \neq 1$;
(ii) $\alpha H$, $\alpha \geq 0$, $\alpha \neq 1$;
(iii) $\alpha, J H$, $\beta > 1 > \alpha \geq 0$ (or $\alpha > 1 > \beta \geq 0$);
(iv) $H^{\alpha, \beta}$, $\beta > 1 > \alpha \geq 0$ (or $\alpha > 1 > \beta \geq 0$);

if any of the following conditions is satisfied:

1. The (input) probability distribution $(p_i, i = 1, 2, ..., n)$ is uniform (in particular, if the joint probability distribution $(\tau_{ik}, i = 1, 2, ..., n; k = 1, 2, ..., m)$ is uniform).
2. The (output) probability distribution $(q_k, k = 1, 2, ..., m)$ is uniform (in particular, if the conditional probability distribution $(q_{ik}, i = 1, 2, ..., n; k = 1, 2, ..., m)$ is uniform).
3. The conditional probability matrix $(q_{ik})_{n \times m}$ has the row-permutation property.

Furthermore, if there exists a constant $A$ such that

$$\sum_{k=1}^{m} q_{ik}^\alpha = A \quad \text{for all} \quad i = 1, 2, ..., n,$$

where $\alpha \geq 0$, $\alpha \neq 1$, then the entropies (i) and (ii) also satisfy the independence inequality.

**Corollary.** The additive entropies

(i) $\alpha H$, $\alpha \geq 0$, $\alpha \neq 1$

and

(ii) $\alpha, J H$, $0 \leq \alpha < 1 < \beta$

are subadditive for any of the probability distributions mentioned in the above theorem.

**Proof.** It is known that the entropies $\alpha H$ and $\alpha, J H$ are additive (see, for example, Aczél and Daróczy, 1975). Thus for any of the probability distributions of the above theorem, the entropies (i) and (ii) are both additive and satisfy the independence inequality, and hence, by Theorem 1, they are also subadditive.

Received: July 9, 1976; revised: October 29, 1976
REFERENCES


