

**CONTRIBUTION OF CHANNEL EQUIVOCATION
FOR THE DEVELOPMENT OF SOURCE CODING THEOREMS**

Om Parkash¹, Priyanka Kakkar^{2 §}

^{1,2}Department of Mathematics
Guru Nanak Dev University
Amritsar, 143005, INDIA

Abstract: The present communication deals with the development of new coding theorems in terms of channel equivocation, that is, coding is done for a source which selects a new set of source statistics after each output symbol is received from the channel. New proof for Fano's bound on Shannon's equivocation is provided by using log sum inequality. Moreover, bounds on various generalizations of Shannon's equivocation have been provided.

AMS Subject Classification: 94A24, 94A15, 94A29

Key Words: entropy, source coding, uniquely decipherable code, instantaneous code, Kraft's Inequality, equivocation, Jensen's inequality, Log Sum inequality

1. Introduction

The main purpose in source coding is to encode the source that produces symbols $X = (x_1, x_2, \dots, x_n)$ with probabilities $p = (p_1, p_2, \dots, p_n)$ to a codeword c_i of length l_i expressed using the D letters of the alphabet. The first source coding theorem investigated by Shannon [4] proved that entropy of the source provides

Received: March 5, 2014

© 2014 Academic Publications, Ltd.
url: www.acadpubl.eu

§Correspondence author

a lower bound to the average number of code symbols needed to encode each source symbol. By making use of Kraft's [8] inequality, which is necessary and sufficient condition for the code to be uniquely decipherable given by

$$\sum_{i=1}^n D^{-l_i} \leq 1, \quad (1.1)$$

Shannon [4] proved the following result:

$$H(X) \leq L < H(X) + 1 \quad (1.2)$$

where

$$L = \sum_{i=1}^n p_i l_i \quad (1.3)$$

is the mean codeword length and

$$H(X) = \sum_{i=1}^n p_i \log_D \frac{1}{p_i} \quad (1.4)$$

is the Shannon's [4] entropy.

Later on, Campbell [9] and Kapur [7] proved the source coding theorems for their own exponentiated mean codeword lengths in the form of following inequalities

$$R_\alpha(p) \leq L_\alpha < R_\alpha(p) + 1 \quad (1.5)$$

and

$$R_\alpha(p) \leq L^\alpha < R_\alpha(p) + 1 \quad (1.6)$$

respectively, where

$$L_\alpha = \frac{\alpha}{1-\alpha} \log_D \left(\sum_{i=1}^n p_i D^{\frac{l_i(1-\alpha)}{\alpha}} \right)$$

is Campbell's [9] mean codeword length,

$$L^\alpha = \frac{1}{\alpha-1} \log_D \left(\frac{\sum_{i=1}^n p_i^\alpha D^{\frac{l_i(1-\alpha)}{\alpha}}}{\sum_{i=1}^n p_i^\alpha} \right), \alpha \neq 1, \alpha > 0$$

is Kapur's [7] mean codeword length and

$$R_\alpha(p) = \frac{1}{1-\alpha} \log_D \sum_{i=1}^n p_i^\alpha$$

is Renyi's [1] measure of entropy.

The above mentioned Kraft's inequality plays an important role in coding theory and many source coding theorems have been proved with the help of this inequality. Some contribution of this inequality to coding theory has recently been provided by Parkash and Priyanka [11] whereas Parkash and Kakkar [13] have obtained optimum probability distribution for minimum redundancy of source coding. Recently, Parkash and Kakkar [12] introduced two mean code-word lengths and proved the following source coding theorems:

Theorem 1. For all uniquely decipherable codes (or instantaneous codes), the lower bound of new mean codeword length $L(\beta)$ lies between $K^\beta(p)$ and $K^\beta(p) + 1$, where

$$L(\beta) = \frac{\sum_{i=1}^n p_i^\beta l_i}{\sum_{i=1}^n p_i^\beta} \quad ; \quad \beta > 0$$

and

$$K^\beta(p) = - \frac{\sum_{i=1}^n p_i^\beta \log_D \left(\frac{p_i^\beta}{\sum_{i=1}^n p_i^\beta} \right)}{\sum_{i=1}^n p_i^\beta} \quad ; \quad \beta > 0$$

is a new measure of entropy introduced by Parkash and Kakkar [12].

Theorem 2. For all uniquely decipherable codes (or instantaneous codes), the exponentiated mean codeword length $L(\alpha, \beta)$ satisfies the following relation:

$$E_\beta^\alpha(p) \leq L(\alpha, \beta) < E_\beta^\alpha(p) + 1$$

where

$$L(\alpha, \beta) = \frac{\alpha}{1 - \alpha} \log_D \left[\frac{\sum_{i=1}^n p_i^{\beta/\alpha} D^{l_i(1-\alpha)/\alpha}}{\sum_{i=1}^n p_i^{\beta/\alpha}} \right] \quad ; \quad \alpha > 0, \beta > 0, \alpha \neq 1$$

and

$$E_\beta^\alpha(p) = \frac{1}{\alpha - 1} \log_D \left[\frac{\left(\sum_{i=1}^n p_i^{\beta/\alpha} \right)^\alpha}{\sum_{i=1}^n p_i^\beta} \right]$$

is Kapur's [7] two parameter additive measure of entropy.

Basically, the above results take into account only the source that produces the symbols to be encoded and not the channel from which the input alphabet

is transmitted. Thus, it becomes interesting to take into consideration the accountability of the channel for its participation in proving the new source coding theorems.

Also, one of the fundamental concepts in information theory is the relation between entropy and error probability, and one of the pioneer related with such studies is Fano's [15], who developed lower bound on the error probability of decoders which plays an important role in deriving other theorems and criteria in information theory. Dong and Fan [17] extended Fano's inequality and introduced the lower bounds on the mutual information between two random variables. Ho and Verdu [16] found a different upper bound on the conditional entropy (equivocation) in terms of the error probability and the marginal distribution of the random variable. Moreover, the authors found the new lower bound on the conditional entropy for countably infinite alphabets. Hu and Xing [2] derived analytical upper and lower bounds between entropy and error probability based on the closed-form solutions of conditional entropy without involving any approximation. The objective of the present paper is to highlight the role of channel in source coding theorems.

In section 2, coding theorems are provided for a source not with a fixed set of source statistics but for a source which selects a new set of source statistics after each output symbol is received from the channel. Section 3 provides the new proof for Fano's [5] bound on Shannon's equivocation using log sum inequality as well as provides new bounds on various generalizations of Shannon's equivocation.

2. Coding Theorems in Terms of Channel Equivocation

An information channel is described by giving an input alphabet $A = \{a_i, i = 1, 2, \dots, r\}$, an output alphabet $B = \{b_j, j = 1, 2, \dots, s\}$ and a set of conditional probabilities $p(b_j/a_i)$ for all i and j where $p(b_j/a_i)$ is the probability that the output symbol b_j will be received if the input symbol a_i is sent. Let us suppose that the input symbols $A = \{a_i, i = 1, 2, \dots, r\}$ occur according to the probabilities $p(a_i), i = 1, 2, \dots, r$ for transmission through the channel and the output symbols $B = \{b_j, j = 1, 2, \dots, s\}$ occur according to the probabilities $p(b_j), j = 1, 2, \dots, s$. So, a priori entropy of A is given by

$$H(A) = \sum_A p(a_i) \log_2 \frac{1}{p(a_i)} \quad (2.1)$$

and the posterior entropy of A , when b_j is received is given by

$$H(A/b_j) = \sum_A p(a_i/b_j) \log_2 \frac{1}{p(a_i/b_j)} \quad (2.2)$$

So, the average number of binary digits necessary to represent an input alphabet a_i , if we are given an output symbol, is the average a posterior entropy

$$\sum_B p(b)H(A/b).$$

It is to be mentioned that we restrict ourselves to the binary case so the logarithms are taken to the base 2 in further discussion.

It has been shown in [10] that

$$H(A/B) \leq \bar{L} < H(A/B) + 1 \quad (2.3)$$

where $\bar{L} = \sum_{A,B} p(a_i, b_j)l_{ij}$ and $H(A/B) = \sum_{A,B} p(a, b) \log_2 \frac{1}{p(a/b)}$, is the channel equivocation.

One major difference between Shannon's noiseless coding theorem and inequality (2.3) is that the former applies to all uniquely decipherable codes, instantaneous or not, whereas the latter applies only to instantaneous codes.

Next, we extend the source coding theorems given by Parkash and Kakkar [12] in the context of channel equivocation.

Extension of Theorem 1. For this purpose, we make use of the concept of escort distribution as follows:

If $p(a_1), p(a_2), \dots, p(a_r)$ is the original distribution of the input alphabet, then its escort distribution is given by $P(a_1), P(a_2), \dots, P(a_r)$ where $P(a_i) = \frac{p^\beta(a_i)}{\sum_{i=1}^n p^\beta(a_i)}$ for some parameter $\beta > 0$. Let us associate the probabilities $P(a_i), i = 1, 2, \dots, r$ with the input alphabet A of the channel, called a priori probabilities and let us suppose that output symbols occur with probabilities $P(b_j), j = 1, 2, \dots, s$. In this case, the average number of binary digits necessary to represent an input alphabet a_i is the average a posterior entropy, $\sum_B p(b_j)H(A/b_j)$ for the given output symbol.

Next, we construct s binary codes one for each of the possible received symbols b_j and we use the j^{th} binary code to encode the transmitted symbol a_i when the output of the channel is b_j . Let the word lengths of s codes be as shown in Table 1. If we require each code to be instantaneous, we may apply the first part of theorem 1 to each code separately which gives

$$K^\beta(A/b_j) \leq \sum_{i=1}^n P(a_i/b_j)l_{ij} = L_{ij}(\beta) \quad (2.4)$$

Table 1: Word lengths for s codes

Input Symbol	Code 1	Code2	...	code s
a_1	l_{11}	l_{12}	...	l_{1s}
a_2	l_{21}	l_{22}	...	l_{2s}
...
a_r	l_{r1}	l_{r2}	...	l_{rs}

where $L_{ij}(\beta)$ is the average length of the j^{th} code. Here, the conditional probabilities are used instead of marginal probabilities because j^{th} code is used only when b_j is the received symbol. So, averaging with respect to the received symbols b_j will give us the average number of binary digits required for each member of the input alphabet.

Multiplying both sides of (2.4) by $P(b_j)$ and summing over all B gives

$$\sum_B K^\beta(A/b_j)P(b_j) \leq \sum_{i=1}^n P(a_i, b_j)l_{ij} = \bar{L}(\beta) \quad (2.5)$$

where $\bar{L}(\beta)$ is the average number of binary digits for each member of the input alphabet averaged with respect to both input and output symbols.

Next, we illustrate a coding procedure in order to show that the bound (2.5) can be achieved.

We select l_{ij} , the codeword length corresponding to the input a_i , when b_j is the output of the channel as the unique integer satisfying

$$\log_2 \frac{1}{P(a_i/b_j)} \leq l_{ij} < \log_2 \frac{1}{P(a_i/b_j)} + 1 \quad (2.6)$$

First, we need to check that the codeword lengths chosen in this manner satisfy the Kraft's inequality and therefore acceptable as the word lengths of an instantaneous code.

Now, from L.H.S of inequality (2.6), we have

$$\sum_i 2^{-l_{ij}} \leq 1 \quad \text{for each } j. \quad (2.7)$$

So, equation (2.6) defines an acceptable set of codeword lengths l_{ij} , for an instantaneous code for each j .

Now, multiplying equation (2.6) by $P(a_i, b_j) = P(a_i/b_j)P(b_j)$ and summing over A and B , we get

$$\sum_B P(b_j)H(A/b_j) \leq \sum_{A,B} P(a_i, b_j)l_{ij} < \sum_B P(b_j)H(A/b_j) + 1 \quad (2.8)$$

Now,

$$\sum_B P(b_j)H(A/b_j) = \sum_B \sum_A P(a_i, b_j) \log \frac{1}{P(a_i/b_j)} = K^\beta(A/B)$$

So, equation (2.8) can be written as

$$K^\beta(A/B) \leq \sum_{A,B} P(a_i, b_j)l_{ij} < K^\beta(A/B) + 1 \tag{2.9}$$

So, result (2.9) is an extension of theorem 1 in terms of channel equivocation.

Note. One major difference between noiseless coding theorem 1 and inequality (2.9) is that the former applies to all uniquely decipherable codes, instantaneous or not, whereas the latter applies only to instantaneous codes. This fact is because of the reason that in the latter case, though each of the code used is uniquely decipherable but it is also not generally true that a sequence of codewords from a known sequence of uniquely decipherable codes is uniquely decipherable. Hence, the codes must all be instantaneous.

Remark. For $\beta = 1, K^\beta(A/B)$ becomes $H(A/B)$. So, it is an extension of Shannon's equivocation.

Extension of Theorem 2. Exponentiated mean codeword length $L(\alpha, \beta)$ and Kapur's [7] entropy $E_\beta^\alpha(p)$ respectably can also be written as

$$L(\alpha, \beta) = \frac{\alpha}{1 - \alpha} \log_D \left(\sum_{i=1}^n Q_i D^{\frac{l_i(1-\alpha)}{\alpha}} \right)$$

and

$$E_\beta^\alpha(p) = \frac{1}{1 - \alpha} \log_D \sum_{i=1}^n Q_i^\alpha,$$

where $Q_i = \frac{p_i^{\beta/\alpha}}{\sum_{i=1}^n p_i^{\beta/\alpha}}$.

Let us associate the probabilities $Q(a_i) = \frac{p_i^{\beta/\alpha}}{\sum_{i=1}^n p_i^{\beta/\alpha}}, \alpha > 0, \beta > 0$, with the input alphabet A of the channel and let us suppose that output symbols occur with probabilities $Q(b_j)$. So, $\sum_B Q(b_j)E_\beta^\alpha(A/b_j)$ represents the average number of binary digits necessary to represent an input alphabet a_i , if we are given an output symbol. On similar lines as in extension of theorem 1, s binary codes are constructed one for each of the possible received symbols b_j and j^{th} binary code is used to encode the transmitted symbol a_i when the output of

the channel is b_j . Then, we may apply the first part of theorem 2 to each code separately which gives

$$E_{\beta}^{\alpha}(A/b_j) \leq \frac{\alpha}{1-\alpha} \log_2 \sum_A Q(a_i/b_j) D^{l_{ij}(\frac{1-\alpha}{\alpha})} = L_{ij}(\alpha, \beta) \quad (2.10)$$

where $L_{ij}(\alpha, \beta)$ is the average length of the j^{th} code. So, averaging with respect to the received symbols b_j will give us the average number of binary digits required for each member of the input alphabet.

Multiplying both sides of (2.10) by $Q(b_j)$ and summing over all B gives

$$\sum_B E_{\beta}^{\alpha}(A/b_j) Q(b_j) \leq \frac{\alpha}{1-\alpha} \sum_B Q(b_j) \log_2 \sum_A Q(a_i/b_j) D^{l_{ij}(\frac{1-\alpha}{\alpha})} = \bar{L}(\alpha, \beta) \quad (2.11)$$

where $\bar{L}(\alpha, \beta)$ is the average number of binary digits for each member of the input alphabet averaged with respect to both input and output symbols.

Next, we illustrate a coding procedure in order to show that the bound (2.11) can be achieved.

We select l_{ij} , the codeword length corresponding to the input a_i , when b_j is the output of the channel as the unique integer satisfying

$$-\log_2 \frac{Q^{\alpha}(a_i/b_j)}{\sum_A Q^{\alpha}(a_i/b_j)} \leq l_{ij} < -\log_2 \frac{Q^{\alpha}(a_i/b_j)}{\sum_A Q^{\alpha}(a_i/b_j)} + 1 \quad (2.12)$$

Now, from inequality (2.12), we get

$$\sum_i 2^{-l_{ij}} \leq 1 \quad \text{for each } j. \quad (2.13)$$

So, equation (2.12) defines an acceptable set of codeword lengths l_{ij} , for an instantaneous code for each j . Now, we consider the following cases:

Case I: When $0 < \alpha < 1$, $\frac{1-\alpha}{\alpha} > 0$. Now, multiplying (2.12) by $\frac{1-\alpha}{\alpha}$ and raising to the power 2, we get the following expression upon multiplying the resulting equation by $Q(a_i/b_j)$ and summing over all members of A :

$$\left(\sum_A Q^{\alpha}(a_i/b_j) \right)^{1/\alpha} \leq \sum_A Q(a_i/b_j) 2^{l_{ij}(\frac{1-\alpha}{\alpha})} < \left(\sum_A Q^{\alpha}(a_i/b_j) \right)^{1/\alpha} \cdot 2^{\frac{(1-\alpha)}{\alpha}} \quad (2.14)$$

Equation (2.13) further gives

$$\frac{1}{1-\alpha} \sum_B Q(b_j) \log_2 \sum_A Q^{\alpha}(a_i/b_j) \leq \bar{L}(\alpha, \beta)$$

$$< \frac{1}{1-\alpha} \sum_B Q(b_j) \log_2 \sum_A Q^\alpha(a_i/b_j) + 1 \tag{2.15}$$

Now,

$$E_\beta^\alpha(A/B) = \frac{1}{1-\alpha} \sum_B Q(b_j) \log_2 \sum_A Q^\alpha(a_i/b_j) \text{ as given in [3]}$$

So, equation (2.14) can be written as

$$E_\beta^\alpha(A/B) \leq \bar{L}(\alpha, \beta) < E_\beta^\alpha(A/B) + 1 \tag{2.16}$$

Case II: When $\alpha > 1$, $\frac{1-\alpha}{\alpha} < 0$, we have

$$E_\beta^\alpha(A/B) \leq \bar{L}(\alpha, \beta) < E_\beta^\alpha(A/B) + 1$$

Again, this result applies only to instantaneous codes.

Remark. $\lim_{\alpha \rightarrow 1} E_\beta^\alpha(A/B) = K^\beta(A/B)$ and

$$\lim_{\alpha \rightarrow 1, \beta \rightarrow 1} E_\beta^\alpha(A/B) = H(A/B).$$

So, $E_\beta^\alpha(A/B)$ is a generalization of equivocation $K^\beta(A/B)$ and Shannon's [4] equivocation.

Similar results can also be proved for Campbell's [9] and Kapur's [7] mean codeword lengths.

3. Bounds on Equivocation in Terms of Probability of Error

The probability of error for a channel with an input alphabet $A = \{a_i, i = 1, 2, \dots, r\}$, an output alphabet $B = \{b_j, j = 1, 2, \dots, s\}$ is given by

$$p_E = 1 - \sum_B p[d(b)/b]p(b) \tag{3.1}$$

where $d(b_j)$ is any function specifying a unique input symbol for each output symbol and is called decision rule.

Now, the conditional maximum likelihood decision rule says that in order to minimize channel error probability, we should use that decision rule which

chooses for each output symbol the input symbol with the highest probability. So, (3.1) is minimized if we choose

$$d(b_j) = a^* \quad \text{for each } j$$

where a^* is defined by $p(a^*/b_j) \geq p(a_i/b_j)$ for all i . So, (3.1) can be written as

$$\begin{aligned} p_E &= 1 - \sum_B p[d(b), b] \\ &= 1 - \sum_B p(a^*, b) \end{aligned} \quad (3.2)$$

The probability $p[d(b), b]$ is the joint probability such that $d(b_j) = a^*$ is transmitted and b_j is received.

Next, we define $\bar{p}_E = 1 - p_E$, so that (3.2) can be written as

$$\bar{p}_E = \sum_B p(a^*, b) \quad (3.3)$$

Since $\sum_{A,B} p(a, b) = 1$, (3.2) can also be written as

$$p_E = \sum_{B, A-a^*} p(a, b) \quad (3.4)$$

New Proof of Fano's Bound using Log Sum Inequality. To prove Fano's [15] bound, we make use of log sum inequality given below:

For non-negative numbers x_1, x_2, \dots, x_n and y_1, y_2, \dots, y_n , the log-sum inequality is expressed as

$$\sum_{i=1}^n x_i \log \frac{x_i}{y_i} \geq \sum_{i=1}^n x_i \log \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i} \quad (3.5)$$

with equality if and only if $\frac{x_i}{y_i}$ is constant for all i .

Now,

$$\begin{aligned} H(A/b_j) &= - \sum_{A-a^*} p(a/b_j) \log_2 p(a/b_j) - p(a^*/b_j) \log_2 p(a^*/b_j) \\ &\leq - \sum_{A-a^*} p(a/b_j) \log_2 \sum_{A-a^*} p(a/b_j) - p(a^*/b_j) \log_2 p(a^*/b_j) \\ &\quad + \sum_{A-a^*} p(a/b_j) \log_2 (r-1) \quad \text{using (3.5)} \end{aligned} \quad (3.6)$$

Let $V = \sum_{A-a^*} p(a/b_j)$, then $p(a^*/b_j) = 1 - V$. So, (3.6) can be written as

$$H(A/b_j) \leq f(V) + \sum_{A-a^*} p(a/b_j) \log_2(r - 1) \tag{3.7}$$

where

$$f(V) = -V \log_2 V - (1 - V) \log_2(1 - V)$$

Multiplying both sides of (3.7) with $p(b_j)$ and summing over B , we get

$$\sum_B p(b)H(A/b) \leq \sum_B p(b)f(V) + \sum_B p(b) \sum_{A-a^*} p(a/b) \log_2(r - 1) \tag{3.8}$$

Since $f(V)$ is concave function of V , applying Jensen's [6] inequality to the first part of R.H.S. of (3.8), we get

$$\begin{aligned} H(A/B) &\leq - \left(\sum_{B,A-a^*} p(a,b) \right) \log_2 \left(\sum_{B,A-a^*} p(a,b) \right) \\ &\quad - \left(1 - \left(\sum_{B,A-a^*} p(a,b) \right) \right) \log_2 \left(1 - \left(\sum_{B,A-a^*} p(a,b) \right) \right) \\ &\quad + \log_2(r - 1) \sum_{B,A-a^*} p(a,b) \\ &= -p_E \log_2 p_E - (1 - p_E) \log_2(1 - p_E) + \log_2(r - 1)p_E \end{aligned}$$

that is,

$$H(A/B) \leq H(p_E) + p_E \log_2(r - 1) \tag{3.9}$$

which is the Fano's [15] bound on equivocation.

The equality is achieved when $p(a/b) = \frac{p_E}{r-1}$ for all b , $a \neq a^*$ and $p(a^*/b) = \bar{p}_E$ for all b .

Fano's bound on Renyi's Equivocation for $\alpha > 1$. The Renyi's [1] equivocation (conditional Renyi's entropy) is given by

$$H_\alpha(A/B) = \frac{1}{1 - \alpha} \sum_B p(b) \log_2 \sum_A p^\alpha(a/b), \quad \alpha > 0 \quad \alpha \neq 1$$

and it approaches to Shannon's[4] equivocation when $\alpha \rightarrow 1$.It has been shown in [5] that

$$H_\alpha(A/B) \leq \frac{1}{1 - \alpha} \log_2 \left((r - 1)^{1-\alpha} p_E^\alpha + (1 - p_E)^\alpha \right) \quad \text{for } 0 < \alpha < 1$$

We will show that for $\alpha > 1$,

$$H_\alpha(A/B) \leq H(p_E) + p_E \log_2(r - 1) \tag{3.10}$$

For this, we need to show that $H_\alpha(A/B) \leq H(A/B), \alpha > 1$.

Consider

$$\sum_A p(a/b) \log_2 \frac{p(a/b)}{p^\alpha(a/b)} \geq \sum_A p(a/b) \log_2 \frac{\sum_A p(a/b)}{\sum_A p^\alpha(a/b)} \text{ for all } b. \text{ (using(3.5))}$$

or

$$\sum_A p(a/b) \log_2 p^{\alpha-1}(a/b) \leq \sum_A p(a/b) \log_2 \frac{\sum_A p^\alpha(a/b)}{\sum_A p(a/b)} \tag{3.11}$$

Since $\sum_A p(a/b) = 1$ for all b , equation (3.11) becomes

$$(\alpha - 1) \sum_A p(a/b) \log_2 p(a/b) \leq \log_2 \sum_A p^\alpha(a/b) \tag{3.12}$$

Multiplying both sides of (3.12) by $p(b)$ and summing over B , we get

$$(\alpha - 1) \sum_B p(b) \sum_A p(a/b) \log_2 p(a/b) \leq \sum_B p(b) \log_2 \sum_A p^\alpha(a/b) \tag{3.13}$$

Now, for $\alpha > 1$, $(1 - \alpha) < 0$. So, (3.13) becomes

$$-\sum_{A,B} p(a,b) \log_2 p(a/b) \geq \frac{1}{1 - \alpha} \sum_B p(b) \log_2 \sum_A p^\alpha(a/b)$$

that is,

$$H_\alpha(A/B) \leq H(A/B) \tag{3.14}$$

So, using (3.14), (3.10) follows from (3.9).

Equality is achieved in (3.10) when $p(a/b) = \frac{p_E}{r-1}$ for all $b, a \neq a^*$ and $p(a^*/b) = \bar{p}_E$ for all b and $\alpha \rightarrow 1$.

Remark. For $0 < \alpha < 1$, $H(A/B) \leq H_\alpha(A/B)$ and so

$$H(A/B) \leq \frac{1}{1 - \alpha} \log_2 ((r - 1)^{1-\alpha} p_E^\alpha + (1 - p_E)^\alpha) \text{ for } 0 < \alpha < 1$$

with equality when $p(a/b) = \frac{p_E}{r-1}$ for all $b, a \neq a^*$ and $p(a^*/b) = \bar{p}_E$ for all b and $\alpha \rightarrow 1$.

Bound on generalized equivocation $K^\beta(A/B)$. Generalized equivocation $K^\beta(A/B)$ can be written as

$$K^\beta(A/B) = \sum_{B,A-a^*} P(a,b) \log_2 \frac{1}{P(a,b)} + \sum_B P(a^*,b) \log_2 \frac{1}{P(a^*,b)} \quad (3.15)$$

where $P(a_i) = \frac{p^\beta(a_i)}{\sum_{i=1}^n p^\beta(a_i)}$, $\beta > 0$ is the associated escort distribution. Using log sum inequality, it can be easily proved that

$$K^\beta(A/B) \leq K^\beta(P_E) + P_E \log_2(r-1) \quad (3.16)$$

Again, for $\beta = 1$, it becomes Fano's inequality given by (3.9).

Remark. Note the difference between $p_E = \sum_{B,A-a^*} p(a,b)$ and $P_E = \sum_{B,A-a^*} P(a,b)$. The former involves general probability distribution whereas latter involves its associated escort probability distribution.

Bound on generalized equivocation $E_\beta^\alpha(A/B)$ for $0 < \alpha < 1$, $\beta > 0$. Generalized equivocation $E_\beta^\alpha(A/B)$ is defined as

$$E_\beta^\alpha(A/B) = \frac{1}{1-\alpha} \sum_B Q(b_j) \log_2 \sum_A Q^\alpha(a_i/b_j), \quad \alpha > 0 \quad \alpha \neq 1$$

where $Q(a_i) = \frac{p^{\beta/\alpha}(a_i)}{\sum_{i=1}^n p^{\beta/\alpha}(a_i)}$, $\alpha, \beta > 0$.

In order to find bound on equivocation, we will make use of the following inequality: If x, y are positive r tuples and $0 < t < 1$, then

$$\sum_{i=1}^r x_i^t y_i^{1-t} \leq \left(\sum_{i=1}^r x_i \right)^t \left(\sum_{i=1}^r y_i \right)^{1-t} \quad (\text{Refer [14], Pg. 129}) \quad (3.17)$$

Now, for $0 < \alpha < 1, \beta > 0$

$$\begin{aligned} E_\beta^\alpha(A/B = b_j) &= \frac{1}{1-\alpha} \log_2 \sum_A Q^\alpha(a/b_j) \\ &= \frac{1}{1-\alpha} \log_2 \left(\sum_A Q^\alpha(a/b_j) + Q^\alpha(a^*/b_j) \right) \end{aligned} \quad (3.18)$$

Letting $y_i = 1/r$ in (3.17), it follows that for $0 < t < 1$, we have

$$\sum_{i=1}^r x_i^t \leq \left(\sum_{i=1}^r x_i \right)^t \quad (3.19)$$

Using (3.19) in (3.18), we get

$$\begin{aligned} E_{\beta}^{\alpha}(A/B = b_j) &= \frac{1}{1-\alpha} \log_2 \left((r-1)^{1-\alpha} \left(\sum_{A-a^*} Q(a/b_j) \right)^{\alpha} + Q^{\alpha}(a^*/b_j) \right) \\ &\leq \frac{1}{1-\alpha} \log_2 \left((r-1)^{1-\alpha} (1 - Q(a^*/b_j))^{\alpha} + Q^{\alpha}(a^*/b_j) \right) \end{aligned} \quad (3.20)$$

Put $Q(a^*/b_j) = z$. So, (3.20) can be written as

$$E_{\beta}^{\alpha}(A/B = b_j) \leq \frac{1}{1-\alpha} \log_2 \left((r-1)^{1-\alpha} (1-z)^{\alpha} + z^{\alpha} \right) \quad (3.21)$$

Let

$$g(z) = \frac{1}{1-\alpha} \log_2 \left((r-1)^{1-\alpha} (1-z)^{\alpha} + z^{\alpha} \right)$$

then $g''(z) < 0$ which shows that $g(z)$ is a concave function of z for $0 < \alpha < 1$. So, multiplying both sides of (3.21) with $Q(b)$ and summing over B , we get

$$E_{\beta}^{\alpha}(A/B) \leq \frac{1}{1-\alpha} \sum_B Q(b) \log_2 \left((r-1)^{1-\alpha} (1-z)^{\alpha} + z^{\alpha} \right) \quad (3.22)$$

Applying Jensen's inequality on the R.H.S of (3.22), we get

$$E_{\beta}^{\alpha}(A/B) \leq \frac{1}{1-\alpha} \log_2 \left((r-1)^{1-\alpha} Q_E^{\alpha} + (1 - Q_E)^{\alpha} \right) \quad (3.23)$$

where $Q_E = \sum_{B, A-a^*} Q(a, b)$ is the probability of error involving escort distribution and $\bar{Q}_E = 1 - Q_E$.

So, (3.23) provides the bound on generalized equivocation $E_{\beta}^{\alpha}(A/B)$ for $0 < \alpha < 1$.

Acknowledgments

The authors are thankful to Council of Scientific and Industrial Research, New Delhi for providing the financial assistance for the preparation of the manuscript.

References

- [1] A. Renyi, On measures of entropy and information, Proc. Fourth Berkeley Symp. Math. Statist. Prob., 1(1961), 547-561.
- [2] B. G. Hu, H. J. Xing, A new approach of deriving bounds between entropy and error from joint distribution: case study for binary classifications, doi:<http://arxiv.org/abs/1303.0943v1> (2013).
- [3] C. Cachin, Entropy Measures and Unconditional Security in Cryptograph, Ph.D. diss., Dept. Comput. Inf. Sci., Swiss Federal Inst. Technol., Zurich, Switzerland (1997).
- [4] C. E. Shannon, A mathematical theory of communication, Bell System Tech. J., 27 (1948), 379-423.
- [5] G.T. Toussaint, A generalization of Shannon's equivocation and the Fano bound, IEEE Transactions on Systems, Man, and Cybernetics, 7 (4) (1977), 300-302.
- [6] J.L.W.V. Jensen, Sur les fonctions convexes et les inegalites entre les valeurs moyennes, Acta Math, 30 (1906), 175-193.
- [7] J. N. Kapur, Entropy and Coding, Mathematical Sciences Trust Society, New Delhi (1998).
- [8] L. G. Kraft, A Device for Quantizing Grouping and Coding Amplitude Modulated Pulses, M.S. Thesis, Electrical Engineering Department, Massachusetts Institute of Technology (1949).
- [9] L. L. Campbell, A coding theorem and Renyi's entropy, Information and Control, 8 (1965), 423-429.
- [10] N. Abramson, Information Theory and Coding, McGraw-Hill, New York (1963).
- [11] O. Parkash, Priyanka, Contribution of Kraft's inequality to coding theory, Pacific-Asian Journal of Mathematics, 5(1) (2011), 45-53.
- [12] O. Parkash, P. Kakkar, Development of two new mean codeword lengths, Information Sciences, 207 (2012), 90-97.
- [13] O. Parkash, P. Kakkar, Optimum probability distribution for minimum redundancy of source coding, Applied Mathematics, 5 (2014), 96-105.

- [14] P.S. Bullen, *A Dictionary of Inequalities: Pitman Monographs and Surveys in Pure and Applied Mathematics 97*, Addison Wesley Longman Limited (1998).
- [15] R. M. Fano, *Transmission of Information: A Statistical Theory of Communication*, Massachusetts Institute of Technology, New York (1961).
- [16] S. W. Ho, S. Verdu, On the interplay between conditional entropy and error probability, *IEEE Transactions on Information Theory*, 56(12) (2010), 5930-5942.
- [17] Y. Dong, P. Fan 2013. An extended Fano's inequality for the finite block-length coding, *IEEE International Conference on Communications, China*, (2013), 6-10.