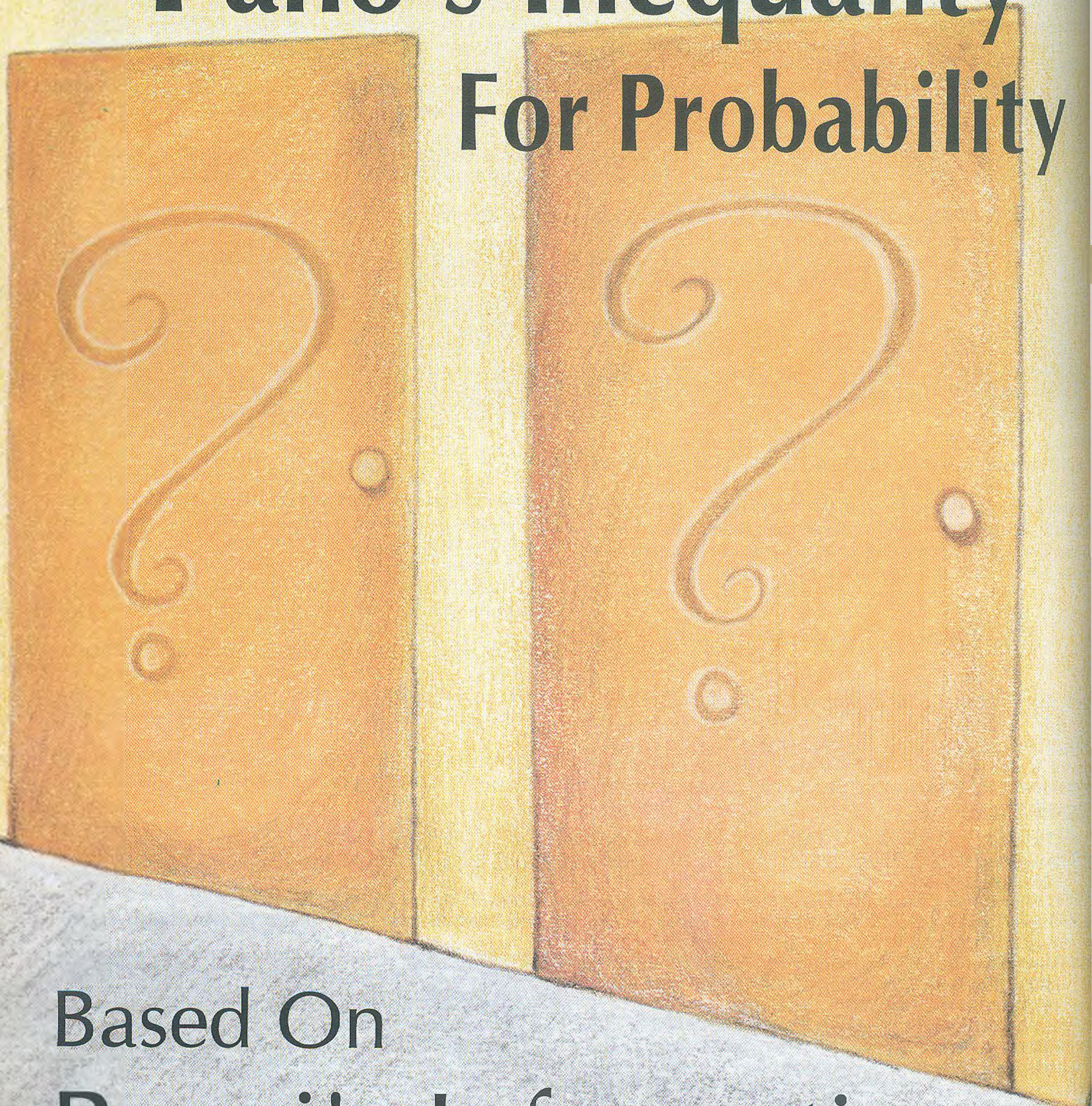
Fano's Inequality For Probability Based On Renyi's Information



Dr. Priti Gupta & Mr. Hari Arora

ABSTRACT

An upper bound to error probability has been presented in terms of Shannon entropy [6]. In this paper, we obtain Fano's bound for probability based on Renyi's entropy [5]. Further, lower bound for average probability of error is calculated in terms of channel capacity.

INTRODUCTION

Let us consider the decision theory problem of classifying observation X as coming from one of the m possible classes (hypothesis) $\theta = \{\theta 1, \theta_2, ..., \theta_n\}$. Let $P_i = Pr\{\theta = \theta_i\}$, i = 1,2,......., n denote the prior probability of the classes and let $f_1(x)$, $f_2(x)$, ..., $f_n(x)$ denote the conditional density functions given the true class i.e. $f_i(x) = Pr\{X = x / \theta = \theta_i\}$, i = 1,2, ...n. We assume that $f_i(x)$ and p_i, i = 1,2,........n are completely known. Given that the observation X = x, we can conclude that the conditional probability of θ by the Bayes rule:

$$P(\theta_i / x) = Pr\{\theta = \theta_i / X = x\}$$

$$= \frac{p_i f_i(x)}{\sum_{j=1}^{n} p_j f_j(x)} \qquad i = 1, 2, ...n. \quad (1.1)$$

It is well known that the decision rule, which minimizes the probability of error, is the Baye's decision rule, which chooses the hypothesis with the largest posterior probability. Using the rule, the probability of error for given X = x is expressed by

$$P(e / x) = 1 - \max [P(\theta_1 / x), P(\theta_2 / x), ... P(\theta_n / x)].$$

Prior to observing X, the probability of error P (e) associated with X is defined as the expected probability after observing it. i.e.,

$$P(e) = E_x[1 - \max\{P(\theta_1/x), P(\theta_2/x), \ldots\ldots P(\theta_n/x)\}]$$

$$= 1 - E_x[\max\{P(\theta_1/x), P(\theta_2/x), \ldots\ldots P(\theta_n/x)\}]$$

Given an arbitrary code (s, n) consisting of words $x^{(1)}, x^{(2)}, \ldots\ldots x^{(s)}$. Let $\mathbf{X} = (X_1, X_2, X_3, \ldots\ldots X_n)$ be a random vector that equals $\mathbf{x}^{(i)}$ with probability $p(\mathbf{x}^{(i)})$, i = 1,2,3,......s, where $\sum p(\mathbf{x}^{(i)}) = 1$. [In other words, we are choosing a code word at random according to the distribution $p(\mathbf{x}^{(i)})$]. Let $\mathbf{Y} = (Y_1, Y_2, \ldots\ldots Y_n)$ be the corresponding output sequence. If P(e) is the probability of error of the code, computed for the given input distribution, then

$$H(X/Y) \leq H\{p(e), 1 - p(e)\} + p(e)\log(s-1). \tag{1.2}$$

In the development of the above-mentioned bound, we utilize several theoretic quantities as defined by Shannon. These are the joint entropy, Conditional entropy, and mutual information.

For a discrete random variable X, Shannon's entropy [6] is given by

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log p(x_i). \tag{1.3}$$

Based on this definition, the joint entropy, mutual information and conditional entropy are defined as

$$H(X, Y) = -\sum_{i=1}^{n}\sum_{j=1}^{n} p(x_i, y_j)\log p(x_i, y_j)$$

$$I(X, Y) = -\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} p(x_i, y_j)\log p(x_i, y_j)}{p(x_i)\,p(y_j)}$$

$$H(X/Y) = \sum_{j=1}^{n} H(X/y_j)\,p(y_j), \tag{1.4}$$

where $$H(X/y_j) = -\sum_{i=1}^{n} p(x_i/y_j)\log p(x_i/y_j) \tag{1.5}$$

and $p(x_i, y_j)$ and $p(x_i / y_j)$ are respectively the joint and the conditional probabilities of X and Y.

Renyi's entropy [5] for X is given by

$$H_\alpha(X) = \frac{1}{1-\alpha}\log\sum_{i=1}^{n} p^\alpha(x_i), \tag{1.6}$$

Where α is a real positive constant different from 1. The (average) mutual information and (average) conditional entropy are consequently

$$H_\alpha(X, Y) = \frac{1}{1-\alpha}\log\sum_{i=1}^{n}\sum_{j=1}^{n} p^\alpha(x_i, y_j) \tag{1.7}$$

$$I_\alpha(X, Y) = \frac{1}{1-\alpha}\log\sum_{i=1}^{n}\sum_{j=1}^{n}\{p^\alpha(x_i, y_j)\}/\{p^{\alpha-1}(x_i)\,p^{\alpha-1}(y_j)\} \tag{1.8}$$

$$H_\alpha(Y/X) = \sum_{i=1}^{n} p(x_i)\,H_\alpha(Y/x_i), \tag{1.9}$$

where

$$H_\alpha(Y/x_i) = \frac{1}{1-\alpha}\log\sum_{j=1}^{n} p^\alpha(y_j/x_i)$$

A large amount of work on probability of error has been done by M.E. Hellman and J. Raviv [3], D.G. Lainiotis [4]. In this paper, we extend our idea of Fano's bound on the probability of error to a family of lower bounds based on Renyi's definition of entropy and mutual information. We relate the probability of error of a code to Renyi's entropy, a generalization of Shannon's entropy. In section I, A systematic method of computing Fano's bound for probability based on Renyi's information is presented and in section II, the lower bound for the average probability of error is calculated in terms of channel capacity by using Renyi's entropy. Shannon measure does not depend upon extraneous factors. But in practical situations extraneous factors plays an important role. In this paper, Bounds derived for probability of error depends upon parameter α, which represents these extraneous factors such as environmental factors, cost factors etc. As a particular case when $\alpha \to 1$, our result reduces to that one corresponding to Shannon's entropy [6].

FANO'S INEQUALITY USING RENYI'S ENTROPY

In order to find the Fano's bound for probability based on Renyi's information we use Jensen's inequality, which is as follows:

Assume g (x) is convex (if concave reverse inequality), $x \in [a, b]$ then for

$$\sum_{i=1}^{n} w_i = 1, \quad w_i > 0, \quad \text{we have}$$

$$g\left[\sum_{i=1}^{n} w_i x_i\right] \leq \sum_{i=1}^{n} w_i\, g(x_i). \qquad (2.1)$$

We also write the conditional probability of error given a specific class as

$$P(e/x_i) = \sum_{i\neq j} p(y_j/x_i) \qquad (2.2)$$

$$1 - P(e/x_i) = p(y_i/x_i) \qquad (2.3)$$

Theorem: 2.1 Given an arbitrary code (s, n) consisting of words $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots\ldots\ldots\ldots \mathbf{x}^{(s)}$. Let $\mathbf{X} = (X_1, X_2, X_3, \ldots\ldots X_n)$ be a random vector that equals $\mathbf{x}^{(i)}$ with probability $p(\mathbf{x}^{(i)})$, $i = 1,2,3,\ldots\ldots s$, where $\sum p(\mathbf{x}^{(i)}) = 1$. [In other words, we are choosing a code word at random according to the distribution $p(\mathbf{x}^{(i)})$]. Let $\mathbf{Y} = (Y_1, Y_2, \ldots\ldots\ldots\ldots Y_n)$ be the corresponding output sequence. If P(e) is the probability of error of the code, computed for the given input distribution, then

$$H_\alpha(X/Y) \leq H_\alpha\{p(e), 1 - p(e)\} + p(e)\log(s-1). \qquad (2.4)$$

Proof: Consider Renyi's conditional entropy [5] of Y given

$$H_\alpha(Y/x_i) = \frac{1}{1-\alpha} \log \sum_{i=1}^{n} p^\alpha(y_j/x_i)$$

$$= \frac{1}{1-\alpha} \log\left[\sum_{i\neq j} p^\alpha(y_j/x_i) + p^\alpha(y_i/x_i)\right] \qquad (2.5)$$

$$= \frac{1}{1-\alpha} \log\left[p^\alpha(e/x_i) \sum_{i\neq j} \{p(y_j/x_i)/p(e/x_i)\}^\alpha + \{1 - p(e/x_i)\}^\alpha\right].$$

Using Jensen's inequality, (2.2) and (2.3), we obtain two inequality for $\alpha > 1$ and $\alpha < 1$ cases

$\alpha > 1$

$$H_\alpha(Y/x_i) \leq p(e/x_i)\frac{1}{1-\alpha} \log p^{\alpha-1}(e/x_i) \sum_{i\neq j} \{p(y_j/x_i)/p(e/x_i)\}^\alpha$$

$$+ \{1 - p(e/x_i)\}\frac{1}{1-\alpha} \log\{1 - p(e/x_i)\}^{\alpha-1}. \qquad (2.6)$$

$\alpha < 1$

or $\quad H_{\alpha}(Y / x_i) \geq \quad p(e / x_i) \frac{1}{1-\alpha} \log p^{\alpha-1}(e / x_i) \sum_{i \neq j} \{ p(y_j / x_i) / p(e / x_i) \}^{\alpha}$

$$+ \{ 1 - p(e / x_i) \} \frac{1}{1-\alpha} \log \{ 1 - p(e / x_i) \}^{\alpha-1} \quad (2.7)$$

$$= H(e / x_i) + p(e / x_i) \frac{1}{1-\alpha} \log \sum_{i \neq j} \{ p(y_j / x_i) / p(e / x_i) \}^{\alpha}.$$

Recall that for (s – 1) point entropy, we have

$$\frac{1}{1-\alpha} \log \sum_{i \neq j} \{ p(y_j / x_i) / p(e / x_i) \}^{\alpha} \leq \log(s-1). \quad (2.8)$$

equality being achieved for a uniform distribution. Hence, for $\alpha > 1$ from (2.6) and (2.8) we obtain

$$H_{\alpha}(Y / x_i) \leq H_{\alpha}(e / x_i) + p(e / x_i) \log(s-1).$$

Finally, using Baye's rule on the conditional distributions and entropies we get the lower bound for P (e).

$$H_{\alpha}(Y / X) \leq H_{\alpha}(e) + p(e) \log(s-1)$$

or

$$H_{\alpha}(X / Y) \leq H_{\alpha}\{ p(e), 1 - p(e) \} + p(e) \log(s-1).$$

Theorem: 2.2 The average probability of error $\overline{p(e)}$ of any code (s, n) satisfy

$$\overline{p(e)} \geq 1 - (n C_{\alpha} + \log 2) / (\log s) \quad (2.9)$$

where C_{α} is the channel capacity. Consequently if $s \geq 2^{n(C+\delta)}$ where $\delta > 0$, then

$$n(C_{\alpha} + \delta) \leq nC_{\alpha} + 1 \text{ or } \overline{p(e)} \geq 1 - (C_{\alpha} + 1/n) / (C_{\alpha} + \delta) \rightarrow 1 - [C_{\alpha} / (C_{\alpha} + \delta)]$$

Thus if $R > C_{\alpha}$, no sequence of codes ($[2^{nR}]$, n) can have an average probability of error which $\rightarrow 0$ as $n \rightarrow \infty$, hence no sequence of codes ($[2^{nR}]$, n, λ_n) can exist with $\lim_{n \to \infty} \lambda_n = 0$

Proof: Choose a code word at random with all words equally likely, that is let X and Y be as in the Fano's inequality with $p(x^{(i)}) = 1/s$, i = 1,2, ..., s. Then H (X) = log s so that

$$I_{\alpha}(X / Y) = \log s - H_{\alpha}(X / Y) \quad (2.10)$$

Let $X_1, X_2, \ldots X_n$ be a sequence of inputs to a discrete memoryless channel, and $Y_1, Y_2, \ldots Y_n$ the corresponding outputs. Then

$I_{\alpha}(X_1, X_2, \ldots, X_n / Y_1, Y_2, \ldots, Y_n) \leq \sum_{i=1}^{n} I_{\alpha}(X_i / Y_i)$ with equality if and only if $Y_1, Y_2, \ldots Y_n$ are independent.

Using above, we have

$$I_{\alpha}(X / Y) \leq \sum_{i=1}^{n} I_{\alpha}(X_i / Y_i) \quad (2.11)$$

Since $I_\alpha(X_i / Y_i) \leq C_\alpha$ (by definition of capacity), (2.10) and (2.11) yield

$$\log s - H_\alpha(X/Y) \leq n C_\alpha \quad (2.12)$$

By Theorem (2.1),

$$H_\alpha(X/Y) \leq H_\alpha\{\overline{p(e)}, 1 - \overline{p(e)}\} + \overline{p(e)} \log(s-1)$$

Hence

$$H_\alpha(X/Y) \leq \log 2 + \overline{p(e)} \log(s) \quad (2.13)$$

The result now follows from equation (2.12) and (2.13).

i.e

$$\log s \leq (n C_\alpha + \log 2) / (1 - \overline{p(e)})$$

or

$$\overline{p(e)} \geq 1 - (n C_\alpha + \log 2) / (\log s)$$

PARTICULAR CASES

(i) When $\alpha \rightarrow 1$ equation (2.4) reduces to (1.1) refer Ash R. [1]
(ii) When $\alpha \rightarrow 1$ equation (2.11) and (2.12) reduces to (2.8) refer Ash R. [1].

CONCLUSIONS

Fano's inequality is an important outcome in Shannon's information theory. This bound is widely appreciated and has acquired wide application in the different fields of communication theory. Fano's lower bound has considerably significant effect as it provides the analyst to find limit of attainable performance in communication channel, whereas, the upper bound, on the other hand, assures that the worst-case performance of the final product is improved with in the known bounds. However, Fano's bound for probability based on Renyi's entropy and the expression for average probability of error is discussed in the present paper. It has amply been demonstrated under numerical dimension the application of proposed bounds to realistic situations (problems) in communication theory. However, either one of these bounds can be utilized in existing practice interchangeably. Finally a candid view has been derived from the study is that these kinds of information which is generally theoretic bounds always require an information which is generally sufficient to get an estimate of the probability of error itself. As such these bounds could be favourably helpful in determining the confidence interval for this probability.

Concludingly, it can favourably asserted herewith that with the help of Fano's inequality we can also propose to derive the relationship among entropies.

REFERENCES

1. *Ash R., Information Theory, Interscience publishers, New York 1968.*
2. *Erdogmus D. and Principe J.C., "Information Transfer Through Classifiers and its Relation to Probability of Error", Intl. Joint Conf. On Neural Networks, pp. 50-54, July 2001.*
3. *Hellman M.E. and Raviv J., "Probability of Error, Equivocation and the Chernoff Bound", IEEE Trans. Inform. Theory, vol. IT 16, pp. 368-372, 1970.*
4. *Lainiotis D.G, "A Class of Upper Bounds on Probability of Error for Multihypothesis Pattern Recognition", IEEE Trans. Inform. Theory, vol. IT 15, pp. 730 731, 1969.*
5. *Renyi A., "On Measures of Entropy and Information", in Proc. 4th Berkeley Symp. Math. And Probability, vol. 1, pp. 547 561, 1960.*
6. *Shannon C.E., "A Mathematical Theory of Communication", Bell system Tech. Journal, vol.27, 379 423, 1948.*