

REFERENCES

- [1] A. Rényi, "On some basic problems of statistics from the point of view of information theory," in *Proc. 5th Berkeley Symp. on Mathematical Statistics and Probability*, vol. 1, 1967, pp. 531–543.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [3] A. Rényi, "On the amount of information concerning an unknown parameter in a sequence of observations," *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, vol. 9, pp. 617–625, 1965.
- [4] M. Hellman and J. Raviv, "Probability of error, equivocation, and the Chernoff bound," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 368–372, July 1970.
- [5] M. Feder and N. Merhav, "Relations between entropy and error probability," *IEEE Trans. Inform. Theory*, vol. 40, pp. 259–266, Jan. 1994.
- [6] S. Arimoto, *Probability, Information and Entropy* (in Japanese). Tokyo: Morikita Pub. Co., 1980.

A Lower Bound on the Probability of Error in Multihypothesis Testing

H. Vincent Poor, *Fellow, IEEE*, and Sergio Verdú, *Fellow, IEEE*

Abstract—Consider two random variables X and Y , where X is finitely (or countably-infinitely) valued, and where Y is arbitrary. Let ϵ denote the minimum probability of error incurred in estimating X from Y . It is shown that

$$\epsilon \geq \sup_{0 \leq \alpha \leq 1} (1 - \alpha)P(\pi(X|Y) \leq \alpha)$$

where $\pi(X|Y)$ denotes the posterior probability of X given Y . This bound finds information-theoretic applications in the proof of converse channel coding theorems. It generalizes and strengthens previous lower bounds due to Shannon, and to Verdú and Han.

Index Terms—Hypothesis testing, probability of error, Shannon theory, Converse Channel Coding Theorem.

I. INTRODUCTION

Consider two random variables X and Y , where Y is arbitrary and where X takes values in a finite or countably infinite set \mathcal{X} . The minimum-error-probability estimate of X conditioned on the observation of Y is given by

$$\hat{X} = \arg \left\{ \max_{k \in \mathcal{X}} \pi(k|Y) \right\} \quad (1)$$

where

$$\pi(k|Y) \triangleq P(X = k|Y). \quad (2)$$

The minimum error probability incurred in testing among the values of X is thus given by

$$\epsilon = P(\hat{X} \neq X) \equiv 1 - E \left\{ \max_{k \in \mathcal{X}} \pi(k|Y) \right\}. \quad (3)$$

Manuscript received September 8, 1994; revised May 16, 1995. This work was supported by the U. S. Army Research Office under Grant DAAH04-93-G-0219.

The authors are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA.
IEEE Log Number 9414768.

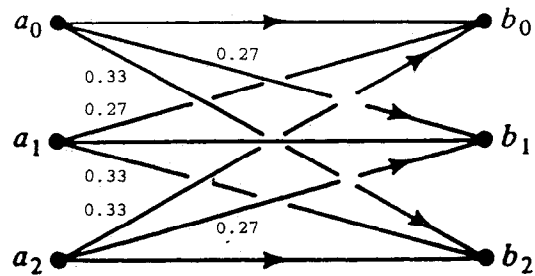


Fig. 1. A ternary hypothesis testing problem with $X \in \{a_0, a_1, a_2\}$ and with a ternary observation $Y \in \{b_0, b_1, b_2\}$.

The maximum occurring in the argument of the expectation of (3) often makes the minimum error probability ϵ difficult to deal with directly. For this reason, bounds on ϵ are of interest in areas for which multihypothesis testing is of central importance, such as digital communications, information theory, and pattern recognition. Such bounds basically have two uses—computational and analytical. In particular, some error-probability bounds are useful because they are simpler to compute than is the actual error probability, and thereby provide a means for performance prediction in multihypothesis testing; whereas other such bounds are of use because they provide analytically tractable means of assessing the behavior of the error probability as various asymptotes are approached. Bounds of this latter type play an important role in the development and proof of coding theorems, and this correspondence presents a new lower bound of this type.

Two classical lower bounds on the multihypothesis error probability that have found use in proving coding theorems are the Fano and Shannon inequalities, which place lower bounds on ϵ in testing among $|\mathcal{X}| = M < \infty$ equiprobable hypotheses. In particular, the Fano inequality (e.g., [1]) is given by

$$\epsilon \geq 1 - \frac{I(X; Y) + \log 2}{\log M} \quad (4)$$

where $I(X; Y)$ denotes the mutual information between X and Y , defined as the expected value (over the joint distribution of X and Y) of the information density

$$i_{XY}(X; Y) = \log \frac{\pi(X|Y)}{P_X(X)} \quad (5)$$

where P_X denotes the probability mass function of X .

The Shannon bound [2] is expressed in terms of the cumulative probability distribution function (cdf) of the information density instead of its average; namely

$$\epsilon \geq \frac{1}{2} P \left(i_{XY}(X; Y) \leq \log \frac{M}{2} \right) \quad (6)$$

$$= \frac{1}{2} P \left(\pi(X|Y) \leq \frac{1}{2} \right) \quad (7)$$

where (7) readily follows from (5) and the assumption that the hypotheses are equiprobable, i.e.

$$P_X(k) = \frac{1}{M}, \quad k \in \mathcal{X}.$$

In the case of nonequiprobable and possibly countably infinitely valued X , the Fano inequality (4) has recently been generalized by Han and Verdú [3]; viz.

$$\epsilon \geq 1 + \frac{I(X; Y) + \log 2}{\log (\max_{k \in \mathcal{X}} P_X(k))}. \quad (8)$$

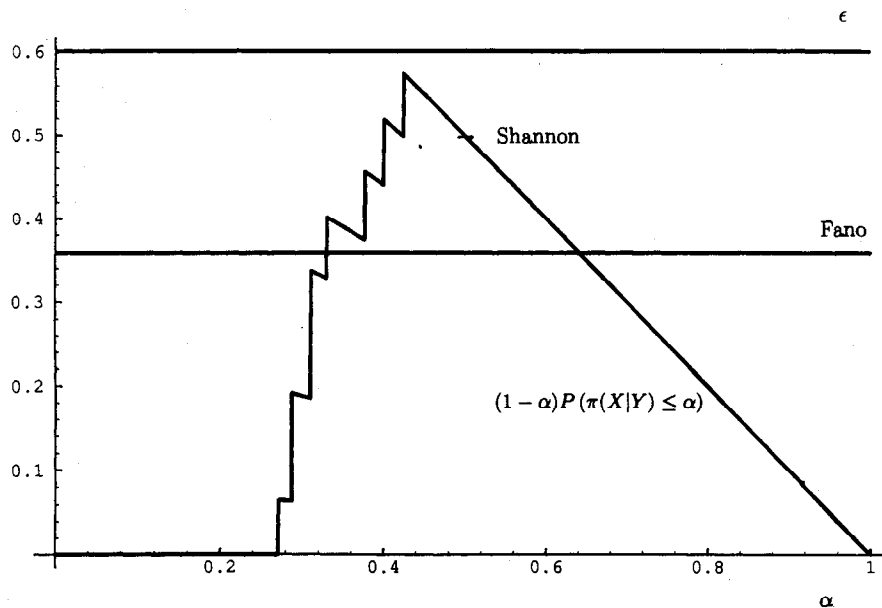


Fig. 2. Lower bounds on the minimum error probability ϵ for the example in Fig. 1 with equiprobable hypotheses.

Another recent lower bound that holds for equiprobable X was obtained by Verdú and Han [4]; namely

$$\epsilon \geq P(\pi(X|Y) \leq \alpha) - \alpha, \quad 0 \leq \alpha \leq 1. \quad (9)$$

This latter bound was used in [4] to provide a converse to the channel coding theorem, which leads to a general expression for channel capacity. As noted in [4], the Fano bound (4) (or its generalization (8)) is not powerful enough to provide such a general result.

Note that each of the above bounds replaces the maximum occurring in (3) with a simpler quantity—either the cdf or the expectation of the information density—that can be analyzed asymptotically via ergodic or large deviations properties. This is the essence of the usefulness of these bounds.

In this correspondence we present a new lower bound on ϵ that is always tighter than both the Shannon and Verdú–Han bounds of (7) and (9), and that shares their advantage over the Fano inequality (4), (8) of providing general converses in channel coding. Moreover, this bound has a simple proof that holds not only for a finite number of equiprobable hypotheses, but also for countably many hypotheses with an arbitrary prior distribution.

II. NEW BOUND

The new bound is summarized in the following result, a proof of which is found in Section III.

Theorem 1: Suppose X and Y are random variables, with X taking on a finite (or countably infinite) number of values. The minimum probability of error ϵ in estimating X from Y satisfies the inequality

$$\epsilon \geq (1 - \alpha)P(\pi(X|Y) \leq \alpha), \quad (10)$$

for each $\alpha \in [0, 1]$, where $\pi(X|Y)$ is defined as in (2).

To gain an intuitive understanding of this bound, note that the cdf of the posterior probabilities $\pi(X|Y)$ will be close to 1 only for values of α close to 1 if the hypothesis-testing problem is one of “low noise.” Conversely, this cdf will be close to 1 for all but small values of α in “high-noise” situations. Once the cdf is multiplied by $(1 - \alpha)$ the function attains a low/high maximum, respectively.

It is easy to find examples where the bound given in Theorem 1 is tighter than Fano’s bound (4) or its generalizations (8). (One such

example is shown in Figs. 1 and 2.) In addition, the lower bound in Theorem 1 is stronger and more general than the bounds by Shannon (7) and Verdú and Han (9). The Shannon bound (7) corresponds to the special case where $\alpha = 1/2$ and X is equally likely to take any of M values. The Shannon bound can be used to prove a general channel capacity converse (such as the one in [4]), and an upper bound on the reliability function (optimum error exponent). However, in contrast to the Verdú–Han bound (9), the Shannon bound does not lead to a tight result for ϵ -capacity. The Verdú–Han bound (9) also corresponds to the case of equiprobable hypotheses, and it is weaker than the bound (10) by the amount

$$\alpha P(\pi(X|Y) > \alpha). \quad (11)$$

The bound (9) can be used to prove a general channel capacity converse and a formula for ϵ -capacity [4]. However, (9) does not lead to a tight result on the reliability function. Therefore, Theorem 1 gives the first known bound to result in both general formulas for channel capacity and ϵ -capacity plus the following converse result for the channel reliability function:

Proposition: For any $R > 0$, define the channel reliability function $E(R)$ as the largest scalar $\beta > 0$ such that there exists a sequence of (n, M, ϵ_n) codes with

$$\beta \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\epsilon_n} \quad (12)$$

and

$$R < \liminf_{n \rightarrow \infty} \frac{\log M}{n}. \quad (13)$$

Then the following general upper bound on $E(R)$ holds:

$$E(R) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \sup_{X^n} \log \left[1/P \left(\frac{1}{n} i_{X^n Y^n}(X^n, Y^n) \leq R \right) \right]. \quad (14)$$

Proof of the Proposition: Let us use the bound of Theorem 1 with arbitrary fixed $\alpha \in (0, 1)$ for any (n, M, ϵ_n) code. In this case X of Theorem 1 takes M equiprobable values on the codebook, which is a subset of the set of input n -strings. Using the same reasoning as

that leading to (7), Theorem 1 implies

$$\begin{aligned} \frac{1}{n} \log \frac{1}{\epsilon_n} &\leq \frac{1}{n} \log \frac{1}{1-\alpha} \\ &+ \frac{1}{n} \sup_{X^n} \log \left[1/P \left(\frac{1}{n} i_{X^n Y^n}(X^n, Y^n) \right. \right. \\ &\quad \left. \left. \leq \frac{\log M}{n} + \frac{\log \alpha}{n} \right) \right] \\ &\leq \frac{1}{n} \log \frac{1}{1-\alpha} \\ &+ \frac{1}{n} \sup_{X^n} \log \left[1/P \left(\frac{1}{n} i_{X^n Y^n}(X^n, Y^n) \leq R \right) \right] \end{aligned}$$

which holds for all sufficiently large n because of (13). Then, (14) follows from (12) and the above inequality. \square

We conjecture that the bound (14) is in fact tight; however, the known approaches to the constructive part of the coding theorem are not sufficient to prove this conjecture even for the simplest channels (for which the reliability function is not yet known for all rates). For example, in the case of a binary-symmetric channel, the evaluation of the right-hand side of (14) is an interesting unsolved large-deviations/optimization problem.

III. PROOF OF THE BOUND

Theorem 1 admits a very simple proof that is quite different from the proofs of the special cases in [2] and [4].

Proof: Without loss of generality, we list the elements of \mathcal{X} as the positive integers $1, 2, \dots$. Let Z_1, Z_2, \dots , denote the random variables $\pi(1|Y), \pi(2|Y), \dots$, placed in decreasing order, pointwise in the sample space¹ (it is immaterial how ties are resolved). First note from (3) that

$$\epsilon = 1 - E\{Z_1\}. \quad (15)$$

For any $\alpha \in [0, 1]$, we can write

$$P(\pi(X|Y) > \alpha) = E \left\{ \sum_{k \in \mathcal{X}} \pi(k|Y) 1\{\pi(k|Y) > \alpha\} \right\} \quad (16)$$

where the expectation is with respect to the unconditional distribution of Y . The argument of the expected value in (16) can be written as

$$\sum_{k \in \mathcal{X}} \pi(k|Y) 1\{\pi(k|Y) > \alpha\} = \sum_{k \in \mathcal{X}} Z_k 1\{Z_k > \alpha\}. \quad (17)$$

Dropping all but the first term

$$P(\pi(X|Y) > \alpha) \geq E\{Z_1 1\{Z_1 > \alpha\}\}. \quad (18)$$

In view of (15) and (18), all we need to do is to relate $E\{Z_1\}$ to $E\{Z_1 1\{Z_1 > \alpha\}\}$ using the fact that $0 \leq Z_1 \leq 1$. Since $Z_1 \leq 1$ note that, for any $\alpha \in [0, 1]$ we have

$$Z_1 = \alpha Z_1 + (1-\alpha)Z_1 \leq \alpha + (1-\alpha)Z_1 1\{Z_1 > \alpha\} \quad (19)$$

which is tantamount to upperbounding Z_1 by $\alpha + (1-\alpha)Z_1$ when $\alpha \leq Z_1 \leq 1$, and by α , otherwise.

Thus on combining (18) and (19), we have

$$\begin{aligned} E\{Z_1\} &\leq \alpha + (1-\alpha)E\{Z_1 1\{Z_1 > \alpha\}\} \\ &\leq \alpha + (1-\alpha)P(\pi(X|Y) > \alpha) \end{aligned} \quad (20)$$

which, together with (15), implies the bound. \square

¹That is, for each point ω in the underlying sample space, $Z_1(\omega), Z_2(\omega), \dots$, denotes the ordered sequence $\pi(1|Y(\omega)), \pi(2|Y(\omega)), \dots$.

REFERENCES

- [1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [2] C. E. Shannon, "Certain results in coding theory for noisy channels," *Inform. Contr.*, vol. 1, pp. 6-25, Sept. 1957.
- [3] T. S. Han and S. Verdú, "Generalizing the Fano inequality," *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1247-1251, July 1994.
- [4] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1147-1157, July 1994.

Asymptotic Efficiency of A Sequential Multihypothesis Test

Venugopal V. Veeravalli, *Member, IEEE*,
and Carl W. Baum, *Member, IEEE*

Abstract—A sequential multihypothesis test known as the MSPRT is generalized to account for nonuniform decision costs. Bounds on error probabilities and asymptotic expressions for the stopping time and error probabilities are given. A key result of this correspondence is a proof that the generalized MSPRT is asymptotically efficient.

Index Terms—Sequential analysis, hypothesis testing, informational divergence.

I. INTRODUCTION

The sequential testing of more than two hypotheses has important applications in direct-sequence signal acquisition [1], [2], multiple-resolution-element radar [3], and other areas. Published work on sequential multihypothesis testing has generally taken two approaches. One approach has aimed at determining a Bayes optimal test, where optimality has been defined in terms of the minimization of a linear combination of two quantities: the expected decision cost and the expected number of observations taken by the test. A recursive solution to the Bayesian optimization problem has in fact been obtained [4]–[6], but unfortunately, this solution is very complex and impractical except in a few special cases.

A second approach has focused on extending and generalizing the sequential probability ratio test (SPRT), a binary test, to incorporate more than two hypotheses. A survey of many of these tests is found in [7]. Although these tests are of low complexity, they have been developed without much consideration to optimality.

In [8], a test is given that incorporates both approaches. The test, called the M -ary Sequential Probability Ratio Test (MSPRT), is a generalization of the SPRT. The MSPRT has a simple structure that facilitates implementation, and it is also based on the solution to the Bayesian optimization problem. It is shown in [8] that the MSPRT approximates the Bayes optimal test, and an example demonstrates that, in at least some cases, the MSPRT is asymptotically optimal as the cost per observation decreases to zero. The MSPRT test structure

Manuscript received October 16, 1994; revised April 19, 1995. The material in this correspondence has appeared in part in the *Proceedings of the 1994 IEEE International Symposium on Information Theory*.

V. V. Veeravalli is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251-1892 USA.

C. W. Baum is with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634-0915 USA.
IEEE Log Number 9414775.