The Error Exponent for Finite-Hypothesis Channel Identification

Patrick Mitran and Aleksander Kavčić
Harvard University
33 Oxford St. Cambridge, MA 02138
{mitran, kavcic}@deas.harvard.edu

Abstract — **We consider the issue of signal selection in hypothesis testing. In particular, we model each hypothesis as a discrete memoryless channel. We first derive the Bayesian error exponent as a function of the limiting empirical distribution of the input sequence. We show that in the case of discriminating between two hypotheses, the asymptotically optimal input sequence consists of always repeating the same input. Finally, we derive an efficient method to evaluate the error exponent as a function of the limiting distribution.**

I. Introduction

In traditional hypothesis testing, we are given a set of hypotheses $\mathcal{H}$. For each hypothesis $h \in \mathcal{H}$, we know the probability law for an observable variable Y, i.e., we know $P_Y^h(y)$. We make n observations $y_1^n = [y_1, y_2, \ldots, y_n]$ and based on these observations, we need to infer the hypothesis $h \in \mathcal{H}$. This is a well known problem with well known solutions in the context of Bayesian and Neyman-Pearson decision making [1]. Furthermore, the type 2 error exponent (for Newman-Pearson) or average error exponent (for Bayesian) detection is well known and may be derived by the method of types [2, 3].

However, let us suppose that the observable variable Y is obtained as the response to an input variable X, which we control. In particular, each hypothesis h may be viewed as a memoryless channel $P_{Y|X}^h(y|x)$. We refer to this type of problem as a *finite-hypothesis channel identification* or *channel detection* problem. In finite-hypothesis channel identification, we are given a finite set of hypotheses H. The objective is to select a set of input signals $x_1^n = [x_1, x_2, ..., x_n]$ according to some policy. We transmit these n signals and only then, observe the n responses $y_1^n = [y_1, ..., y_n]$. We make a decision on $h \in \mathcal{H}$ after we observe all outputs y_1^n.

II. Preliminaries

If a sequence x_1^n is known to generate y_1^n, then a maximum a posteriori detector (which maximizes the probability of correct detection) chooses the hypothesis $\hat{h} = \arg\max_h P[h|x_1^n, y_1^n]$. This may be evaluated using Bayes' rule. The probability of correct detection is then $P[\hat{h}|x_1^n, y_1^n]$.

If the input sequence x_1^n is not randomized (i.e., it is fixed in advance), then, it is clear that the average probability of correct detection is $P_n^c = E_{Y_1^n} P[\hat{H}|x_1^n, Y_1^n]$.

Furthermore, by the memoryless assumption, P_n^c depends only on the *type* of the sequence x_1^n. If we randomize the sequence X_1^n, then we average the performance over several types. Clearly, by the memoryless and time-invariant assumptions, it is best to choose the input sequence x_1^n from the best sequence type.

We let $\mathcal{E}_X$ be the set of all distributions on X and $\mathcal{E}_{Y|X}$ denote the set of all conditional distributions of Y given X. Furthermore, we let $Q_{x_1^n}$ denote the type of x_1^n and we let $\mathcal{Q}_X^n$ denote the set of all types $Q_{x_1^n}$.

Definition 1. *Let $Q_n \in \mathcal{Q}_X^n$, $n = 1, \ldots, \infty$, be a sequence of types for X with $Q_n \to Q_X \in \mathcal{E}_X$ as $n \to \infty$.*

1. *The type error exponent of Q_X is defined to be $E_t(Q_X) = \lim_{n\to\infty} -\frac{1}{n} \log\left(1 - P_n^c\right)\big|_{x_1^n \in T(Q_n)}$.*
2. *The optimal error exponent is $E_t = \max_{Q_X \in \mathcal{E}_X} E_t(Q_X)$.*

III. Results

Theorem 1. *The type error exponent is given by*

$$E_t(Q_X) = \min_{\substack{h,h' \in \mathcal{H} \\ h \neq h'}} \min_{Q_{Y|X} \in \mathcal{E}_{Y|X}} \max\{D(Q_{Y|X}Q_X || P_{Y|X}^h), D(Q_{Y|X}Q_X || P_{Y|X}^{h'})\},$$

where $P_{Y|X}^h(y|x)$ is the channel under hypothesis $h \in \mathcal{H}$. Futhermore, the minimizing $Q_{Y|X}$ results in equality of the two terms inside the max.

Corollary 1. *In the case of a binary hypotheses scenario, the optimal error exponent $E_t = \max_{Q_X \in \mathcal{E}_X} E_t(Q_X)$ is achieved by a vertex of the simplex $\mathcal{E}_X$, (i.e., the optimal input sequence is constant).*

We now propose a method to efficiently evaluate $E_t(Q_X)$ numerically. The maximizing Q_X may then be found using standard numerical techniques. Given two distinct hypotheses $h_1, h_2 \in \mathcal{H}$, it will suffice to be able to evaluate D_{h_1,h_2}, defined as follows

$$D_{h_1,h_2}(Q_X) \triangleq \min_{Q_{Y|X} \in \mathcal{E}_{Y|X}} \max\{D(Q_{Y|X}Q_X || P_{Y|X}^{h_1}), D(Q_{Y|X}Q_X || P_{Y|X}^{h_2})\} \quad (1)$$

Theorem 2. *If $D(P_{Y|X}^{h_2}Q_X || P_{Y|X}^{h_1}) = 0$ then $Q_{Y|X} = P_{h_2}$ is a global minimizer of* (1). *Otherwise, the global minimizing $Q_{Y|X}$ in* (1) *has the form*

$$Q_{Y|X}^{(\lambda)}(y|x) = \frac{P_{Y|X}^{h_1}(y|x)^\lambda P_{Y|X}^{h_2}(y|x)^{1-\lambda}}{\sum_{y'} P_{Y|X}^{h_1}(y'|x)^\lambda P_{Y|X}^{h_2}(y'|x)^{1-\lambda}}, \quad (2)$$

for some $\lambda = \lambda^*$. Furthermore, this λ^* is the unique zero of $L(\lambda) = D(Q_{Y|X}^{(\lambda)}Q_X || P_{Y|X}^{h_1}) - D(Q_{Y|X}^{(\lambda)}Q_X || P_{Y|X}^{h_2})$ in the interval $[0.0, 1.0]$. *Hence, λ^* may be computed efficiently using the bisection method.*

References

[1] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part 1,* New York: Wiley, 1968.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems,* New York: Academic, 1981.

[3] T. M. Cover and J. A Thomas, *Elements of Information Theory,* New York: Wiley, 1991.