ALBANIAN JOURNAL
OF MATHEMATICS
Volume 2, Number 3, Pages 179–184
ISSN 1930-1235: (2008)

ON IDENTIFICATION FOR SOURCES EXTENDED TO MODEL WITH LIES

Zlatko Varbanov

Department of Mathematics and Informatics
Veliko Tarnovo University, 5000 Veliko Tarnovo
e-mail:vtgold@yahoo.com

1. Introduction

The classical transmission problem deals with the question how many possible messages can we transmit over a noisy channel? Transmission means there is an answer to the question "What is the actual message?"

In the identification problem we deal with the question how many possible messages the receiver of a noisy channel can identify? Identification means there is an answer to the question "Is the actual message u?". Here u can be any member of the set of possible messages.

Allowing randomized encoding the optimal code size grows double exponentially in the block length and somewhat surprisingly the second order capacity equals Shannon's first order transmission capacity (see [5]).

Thus, Shannon's Channel Coding Theorem for Transmission is paralleled by a Channel Coding Theorem for Identification. It seems natural to look for such a parallel for sources, in particular for noiseless coding. This was suggested by Ahlswede in [1].

Let $(\mathcal{U}, P)$ be a source, where $\mathcal{U} = \{1, 2, \ldots, N\}, P = \{P_1, P_2, \ldots, P_N\}$, and let $\mathcal{C} = \{c_1, c_2, \ldots, c_N\}$ be a binary prefix code (PC) for this source with $||c_u||$ as length of c_u. Introduce the random variable U with $\text{Prob}(U = u) = p_u$ for $u = 1, 2, \ldots, N$ and the random variable C with $C = c_u = (c_1, c_2, \ldots, c_{u||c_u||})$ if $U = u$.

We use the PC for noiseless identification, that is user u wants to know whether the source output equals u, that is, whether C equals c_u or not. The user iteratively checks whether C coincides with c_u in the first, second, etc. letter and stops when the first different letter occurs or when $C = c_u$. The problem is: **What is the expected number $L_{\mathcal{C}}(P, u)$ of checkings?**

In order to calculate this quantity we introduce for the binary tree $T_{\mathcal{C}}$, whose leaves are the codewords $c_1, c_2, \ldots, c_N$, the sets of leaves $\mathcal{C}_{ik}(1 \le i \le N; 1 \le k)$, where $\mathcal{C}_{ik} = \{c \in \mathcal{C} : c$ coincides with c_i exactly until the k'th letter of $c_i\}$. If C takes a value in $\mathcal{C}_{uk}, 0 \le k \le ||c_u|| - 1$, the answers are k times "Yes" and 1 time "No". For $C = c_u$ we have

$$L_{\mathcal{C}}(P, u) = \sum_{k=0}^{||c_u||-1} P(C \in \mathcal{C}_{uk})(k + 1) + ||c_u|| P_u.$$

Partially supported by RD491-09/2008 project, Veliko Tarnovo University.