On Logarithmically Asymptotically Optimal Hypothesis Testing for Arbitrarily Varying Sources with Side Information

R. Ahlswede, Ella Aloyan, and E. Haroutunian*

Abstract. The asymptotic interdependence of the error probabilities exponents (reliabilities) in optimal hypotheses testing is studied for arbitrarily varying sources with state sequence known to the statistician. The case when states are not known to the decision maker was studied by Fu and Shen.

1 Introduction

On the open problems session of the Conference in Bielefeld (August 2003) Ahlswede formulated among others the problem of investigation of "Statistics for not completely specified distributions" in the spirit of his paper [1]. In this paper, in particular, coding problems are solved for arbitrarily varying sources with side information at the decoder. Ahlswede proposed to consider the problems of inference for similar statistical models. It turned out that the problem of "Hypothesis testing for arbitrarily varying sources with exponential constraint" was already solved by Fu and Shen [2]. This situation corresponds to the case, when side information at the decoder (in statistics this is the statistician, the decision maker) is absent.

The present paper is devoted to the same problem when the statistician has the possibility to make decisions after receiving the complete sequence of states of the source. This, still simple, problem may be considered as a beginning of the realization of the program proposed by Ahlswede.

This investigation is a development of results from [3]-[9] and may be continued in various directions: the cases of many hypotheses, non complete side information, sources of more general classes (Markov chains, general distributions), identification of hypotheses in the sense of [10].

2 Formulation of Results

An arbitrarily varying source is a generalized model of a discrete memoryless source, distribution of which varies independently at any time instant within a certain set. Let $\mathcal{X}$ and $\mathcal{S}$ be finite sets, $\mathcal{X}$ the source alphabet, $\mathcal{S}$ the state alphabet. $\mathcal{P}(\mathcal{S})$ is a set of all possible probability distributions P on $\mathcal{S}$. Suppose a statistician makes decisions between two conditional probability distributions

* Work was partially supported by INTAS grant 00738.

R. Ahlswede et al. (Eds.): Information Transfer and Combinatorics, LNCS 4123, pp. 547–552, 2006.

of the source: $G_1 = \{G_1(x|s), x \in \mathcal{X}, s \in \mathcal{S}\}$, $G_2 = \{G_2(x|s), x \in \mathcal{X}, s \in \mathcal{S}\}$, thus there are two alternative hypotheses $H_1 : G = G_1$, $H_2 : G = G_2$. A sequence $\mathbf{x} = (x_1, \dots, x_N)$, $\mathbf{x} \in \mathcal{X}^N$, $N = 1, 2, ...$, is emitted from the source, and sequence $\mathbf{s} = (s_1, \dots, s_N)$ is created by the source of states. We consider the situation when the source of states is connected with the statistician who must decide which hypothesis is correct on the base of the data $\mathbf{x}$ and the state sequence $\mathbf{s}$. Every test $\varphi^{(N)}$ is a partition of the set $\mathcal{X}^N$ into two disjoint subsets $\mathcal{X}^N = \mathcal{A}_{\mathbf{s}}^{(N)} \bigcup \overline{\mathcal{A}}_{\mathbf{s}}^{(N)}$, where the set $\mathcal{A}_{\mathbf{s}}^{(N)}$ consists of all vectors $\mathbf{x}$ for which the first hypothesis is adopted using the state sequence $\mathbf{s}$.

Making decisions about these hypotheses one can commit the following errors: the hypothesis H_1 is rejected, but it is correct, the corresponding error probability is

$$\alpha_{1|2}^{(N)}(\varphi^{(N)}) = \max_{\mathbf{s} \in \mathcal{S}^N} G_1^N(\overline{\mathcal{A}}_{\mathbf{s}}^{(N)} \mid \mathbf{s}),$$

if the hypothesis H_1 is adopted while H_2 is correct, we make an error with the probability

$$\alpha_{2|1}^{(N)}(\varphi^{(N)}) = \max_{\mathbf{s} \in \mathcal{S}^N} G_2^N(\mathcal{A}_{\mathbf{s}}^{(N)} \mid \mathbf{s}).$$

Let us introduce the following error probability exponents or "reliabilities" $E_{1|2}$ and $E_{2|1}$, using logarithmical and exponential functions at the base e:

$$\overline{\lim_{N\to\infty}} - N^{-1} \ln \alpha_{1|2}^{(N)}(\varphi^{(N)}) = E_{1|2}, \tag{1}$$

$$\overline{\lim_{N\to\infty}} - N^{-1} \ln {\alpha_{2|1}}^{(N)}(\varphi^{(N)}) = E_{2|1}. \tag{2}$$

The test is called *logarithmically asymptotically optimal* (LAO) if for given $E_{1|2}$ it provides the largest value to $E_{2|1}$. The problem is to state the existence of such tests and to determine optimal dependence of the value of $E_{2|1}$ from $E_{1|2}$.

Now we collect necessary basic concepts and definitions. For $\mathbf{s}{=}(s_1, \dots, s_N)$, $\mathbf{s} \in \mathcal{S}^N$, let $N(s \mid \mathbf{s})$ be the number of occurrences of $s \in \mathcal{S}$ in the vector $\mathbf{s}$. The type of $\mathbf{s}$ is the distribution

$$P_{\mathbf{s}} = \{P_{\mathbf{s}}(s), s \in \mathcal{S}\}$$

defined by

$$P_{\mathbf{s}}(s) = \frac{1}{N} N(s \mid \mathbf{s}), \qquad s \in \mathcal{S}.$$

For a pair of sequences $\mathbf{x} \in \mathcal{X}^N$ and $\mathbf{s} \in \mathcal{S}^N$, let $N(x, s|\mathbf{x}, \mathbf{s})$ be the number of occurrences of the pair $(x, s) \in \mathcal{X} \times \mathcal{S}$ in the pair of vectors $(\mathbf{x}, \mathbf{s})$. The joint type of the pair $(\mathbf{x}, \mathbf{s})$ is the distribution

$$Q_{\mathbf{x},\mathbf{s}} = \{Q_{\mathbf{x},\mathbf{s}}(x, s),\, x \in \mathcal{X},\, s \in \mathcal{S}\}$$

defined by

$$Q_{\mathbf{x},\mathbf{s}}(x, s) = \frac{1}{N} N(x, s \mid \mathbf{x}, \mathbf{s}), \qquad x \in \mathcal{X}, s \in \mathcal{S}.$$

The conditional type of $\mathbf{x}$ for given $\mathbf{s}$ is the conditional distribution

$$Q_{\mathbf{x}|\mathbf{s}} = \{Q_{\mathbf{x}|\mathbf{s}}(x|s),\ x \in \mathcal{X},\ s \in \mathcal{S}\}$$

defined by

$$Q_{\mathbf{x}|\mathbf{s}}(x|s) = \frac{Q_{\mathbf{x},\mathbf{s}}(x,s)}{P_{\mathbf{s}}(s)}, \qquad x \in \mathcal{X}, s \in \mathcal{S}.$$

Let X and S are random variables with probability distributions $P = \{P(s), s \in \mathcal{S}\}$ and $Q = \{Q(x|s), x \in \mathcal{X}, s \in \mathcal{S}\}$. The conditional entropy of X with respect to S is:

$$H_{P,Q}(X \mid S) = -\sum_{x,s} P(s)Q(x|s) \ln Q(x|s).$$

The conditional divergence of the distribution $P \circ Q = \{P(s)Q(x|s), x \in \mathcal{X}, s \in \mathcal{S}\}$ with respect to $P \circ G_m = \{P(s)G_m(x|s),\ x \in \mathcal{X}, s \in \mathcal{S}\}$ is defined by

$$D(P \circ Q || P \circ G_m) = D(Q \parallel G_m|P) = \sum_{x,s} P(s)Q(x|s) \ln \frac{Q(x|s)}{G_m(x|s)}, \qquad m = 1, 2.$$

The conditional divergence of the distribution $P \circ G_2 = \{P(s)G_2(x|s), x \in \mathcal{X}, s \in \mathcal{S}\}$ with respect to $P \circ G_1 = \{P(s)G_1(x|s),\ x \in \mathcal{X}, s \in \mathcal{S}\}$ is defined by

$$D(P \circ G_2 || P \circ G_1) = D(G_2 \parallel G_1|P) = \sum_{x,s} P(s)G_2(x|s) \ln \frac{G_2(x|s)}{G_1(x|s)}.$$

Similarly we define $D(G_1 \parallel G_2|P)$.

Denote by $\mathcal{P}^N(\mathcal{S})$ the space of all types on $\mathcal{S}$ for given N, and $\mathcal{Q}^N(\mathcal{X}, \mathbf{s})$ the set of all possible conditional types on $\mathcal{X}$ for given $\mathbf{s}$. Let $\mathcal{T}^{(N)}_{P_{\mathbf{s}},Q}(X \mid \mathbf{s})$ be the set of vectors $\mathbf{x}$ of conditional type Q for given $\mathbf{s}$ of type $P_{\mathbf{s}}$.

It is known [3] that

$$\mid \mathcal{Q}^N(\mathcal{X}, \mathbf{s}) \mid \leq (N+1)^{|\mathcal{X}||\mathcal{S}|}, \tag{3}$$

$$(N+1)^{-|\mathcal{X}||\mathcal{S}|} \exp\{NH_{P_{\mathbf{s}},Q}(X|S)\} \leq \mid \mathcal{T}^{(N)}_{P_{\mathbf{s}},Q}(X \mid \mathbf{s}) \mid \leq \exp\{NH_{P_{\mathbf{s}},Q}(X|S)\}. \tag{4}$$

Theorem. For every given $E_{1|2}$ from $(0, \min\limits_{P \in \mathcal{P}(\mathcal{S})} D(G_2 \parallel G_1 \mid P))$

$$E_{2|1}(E_{1|2}) = \min_{P \in \mathcal{P}(\mathcal{S})} \min_{Q: D(Q||G_1|P) \leq E_{1|2}} D(Q \parallel G_2 \mid P). \tag{5}$$

We can easily infer the following

Corollary. (Generalized lemma of Stein): When $\alpha^{(N)}_{1|2}(\varphi^{(N)}) = \varepsilon > 0$, for N large enough

$$\alpha^{(N)}_{2|1}(\alpha^{(N)}_{1|2}(\varphi^{(N)}) = \varepsilon) = \min_{P \in \mathcal{P}(\mathcal{S})} \exp\{-ND(G_1 \parallel G_2|P)\}.$$

3 Proof of the Theorem

The proof consists of two parts. We begin with demonstration of the inequality

$$E_{2|1}(E_{1|2}) \geq \min_{P\in\mathcal{P}(\mathcal{S})} \min_{Q:D(Q||G_1|P)\leq E_{1|2}} D(Q||G_2|P). \tag{6}$$

For $\mathbf{x} \in \mathcal{T}^{(N)}_{P_\mathbf{s},Q}(X|\mathbf{s})$, $\mathbf{s} \in \mathcal{T}^{(N)}_{P_\mathbf{s}}(\mathcal{S})$, $m = 1, 2$ we have,

$$G^N_m(\mathbf{x} \mid \mathbf{s}) = \prod_{n=1}^{N} G_m(x_n|s_n) = \prod_{x,s} G_m(x|s)^{N(x,s|\mathbf{x},\mathbf{s})} = \prod_{x,s} G_m(x|s)^{NP_\mathbf{s}(s)Q_{\mathbf{x}|\mathbf{s}}(x|s)} =$$

$$= \prod_{x,s} \exp\{NP_\mathbf{s}(s)(s)Q_{\mathbf{x}|\mathbf{s}}(x|s)\ln G_m(x|s)\} = \prod_{x,s} \exp\{N[P_\mathbf{s}(s)(s)Q_{\mathbf{x}|\mathbf{s}}(x|s)\ln G_m(x|s) -$$

$$-P_\mathbf{s}(s)(s)Q_{\mathbf{x}|\mathbf{s}}(x|s)\ln Q_{\mathbf{x}|\mathbf{s}}(x|s) + P_\mathbf{s}(s)(s)Q_{\mathbf{x}|\mathbf{s}}(x|s)\ln Q_{\mathbf{x}|\mathbf{s}}(x|s)]\} =$$

$$= \exp\{N\sum_{x,s}(-P_\mathbf{s}(s)(s)Q_{\mathbf{x}|\mathbf{s}}(x|s)\ln \frac{Q_{\mathbf{x}|\mathbf{s}}(x|s)}{G_m(x|s)} + P_\mathbf{s}(s)(s)Q_{\mathbf{x}|\mathbf{s}}(x|s)\ln Q_{\mathbf{x}|\mathbf{s}}(x|s))\} =$$

$$= \exp\{-N[D(Q \parallel G_m|P) + H_{P_\mathbf{s},Q}(X \mid S)]\}. \tag{7}$$

Let us show that the optimal sequence of tests $\varphi^{(N)}$ for every $\mathbf{s}$ is given by the following sets

$$\mathcal{A}^{(N)}_\mathbf{s} = \bigcup_{Q:D(Q\|G_1|P_\mathbf{s})\leq E_{1|2}} \mathcal{T}^{(N)}_{P_\mathbf{s},Q}(X \mid \mathbf{s}). \tag{8}$$

Using (4) and (7) we see that

$$G^N_m(\mathcal{T}^{(N)}_{P_\mathbf{s},Q}(X|\mathbf{s}) \mid \mathbf{s}) = \mid \mathcal{T}^{(N)}_{P_\mathbf{s},Q}(X \mid \mathbf{s}) \mid G^N_m(\mathbf{x}|\mathbf{s}) \leq \exp\{-ND(Q \parallel G_m|P_\mathbf{s})\}.$$

We can estimate both error probabilities

$$\alpha^{(N)}_{1|2}(\varphi^{(N)}) = \max_{\mathbf{s}\in\mathcal{S}^N} G^N_1(\overline{\mathcal{A}}^{(N)}_\mathbf{s} \mid \mathbf{s}) = \max_{\mathbf{s}\in\mathcal{S}^N} G^N_1(\bigcup_{Q:D(Q||G_1|P_\mathbf{s})>E_{1|2}} \mathcal{T}^{(N)}_{P_\mathbf{s},Q}(X \mid \mathbf{s}) \mid \mathbf{s}) \leq$$

$$\leq \max_{\mathbf{s}\in\mathcal{S}^N} (N+1)^{|\mathcal{X}|\mathcal{S}|} \max_{Q:D(Q||G_1|P_\mathbf{s})>E_{1|2}} G^N_1(\mathcal{T}^{(N)}_{P_\mathbf{s},Q}(X \mid \mathbf{s}) \mid \mathbf{s}) \leq$$

$$\leq (N+1)^{|\mathcal{X}||\mathcal{S}|} \max_{P_\mathbf{s}\in\mathcal{P}^N(\mathcal{S})} \max_{Q:D(Q||G_1|P_\mathbf{s})>E_{1|2}} \exp\{-ND(Q||G_1|P_\mathbf{s})\} =$$

$$= \max_{P_\mathbf{s}\in\mathcal{P}^N(\mathcal{S})} \max_{Q:D(Q||G_1|P_\mathbf{s})>E_{1|2}} \exp\{|\mathcal{X}||\mathcal{S}|\ln(N+1) - ND(Q||G_1|P)\} =$$

$$= \max_{P_\mathbf{s}\in\mathcal{P}^N(\mathcal{S})} \max_{Q:D(Q||G_1|P_\mathbf{s})>E_{1|2}} \exp\{-N[D(Q||G_1|P_\mathbf{s}) - o(1)]\} \leq$$

$$\leq \exp\{-N(E_{1|2} - o(1))\},$$

where $o(1) = N^{-1}|\mathcal{X}||\mathcal{S}|\ln(N+1) \to 0$, when $N \to \infty$. And then

$$\alpha_{2|1}^{(N)}(\varphi^{(N)}) = \max_{\mathbf{s}\in\mathcal{S}^N} G_2^N(\mathcal{A}_{\mathbf{s}}^{(N)}|\mathbf{s}) = \max_{\mathbf{s}\in\mathcal{S}^N} G_2^N(\bigcup_{Q:D(Q||G_1|P_{\mathbf{s}})\leq E_{1|2}} \mathcal{T}_{P_{\mathbf{s}},Q}^{(N)}(X|\mathbf{s})|\mathbf{s}) =$$

$$= \max_{\mathbf{s}\in\mathcal{S}^N} \sum_{Q:D(Q||G_1|P_{\mathbf{s}})\leq E_{1|2}} G_2^N(\mathcal{T}_{P_{\mathbf{s}},Q}^{(N)}(X|\mathbf{s})|\mathbf{s}) \leq$$

$$\leq (N+1)^{|\mathcal{X}||\mathcal{S}|} \max_{P_{\mathbf{s}}\in\mathcal{P}^N(\mathcal{S})} \max_{Q:D(Q||G_1|P_{\mathbf{s}})\leq E_{1|2}} \exp\{-ND(Q||G_2|P_{\mathbf{s}})\} =$$

$$= \max_{P_{\mathbf{s}}\in\mathcal{P}^N(\mathcal{S})} \max_{Q:D(Q||G_1|P_{\mathbf{s}})\leq E_{1|2}} \exp\{-N[D(Q||G_2|P_{\mathbf{s}}) - o(1)]\} =$$

$$= \exp\{-N(\min_{P_{\mathbf{s}}\in\mathcal{P}^N(\mathcal{S})} \min_{Q:D(Q||G_1|P_{\mathbf{s}})\leq E_{1|2}} D(Q||G_2|P_{\mathbf{s}}) - o(1))\}.$$

So with $N \to \infty$ we get (6)

Now we pass to the proof of the second part of the theorem. We shall prove the inequality inverse to (6). First we can show that this inverse inequality is valid for test $\varphi^{(N)}$ defined by (8). Using (4) and (7) we obtain

$$\alpha_{2|1}^{(N)}(\varphi^{(N)}) = \max_{\mathbf{s}\in\mathcal{S}^N} G_2^N(\mathcal{A}_{\mathbf{s}}^{(N)}|\mathbf{s}) = \max_{\mathbf{s}\in\mathcal{S}^N} G_2^N(\bigcup_{Q:D(Q||G_1|P_{\mathbf{s}})\leq E_{1|2}} \mathcal{T}_{P_{\mathbf{s}},Q}^{(N)}(X|\mathbf{s})|\mathbf{s})) \geq$$

$$\geq \max_{\mathbf{s}\in\mathcal{S}^N} \max_{Q:D(Q||G_1|P_{\mathbf{s}})\leq E_{1|2}} G_2^N(\mathcal{T}_{P_{\mathbf{s}},Q}^{(N)}(X|\mathbf{s})|\mathbf{s}) \geq$$

$$\geq \max_{P_{\mathbf{s}}\in\mathcal{P}^N(\mathcal{S})} (N+1)^{-|\mathcal{X}||\mathcal{S}|} \max_{Q:D(Q||G_1|P_{\mathbf{s}})\leq E_{1|2}} \exp\{-ND(Q||G_2|P_{\mathbf{s}})\} =$$

$$= \exp\{-N(\min_{P_{\mathbf{s}}\in\mathcal{P}^N(\mathcal{S})} \min_{Q:D(Q||G_1|P_{\mathbf{s}})\leq E_{1|2}} D(Q||G_1|P_{\mathbf{s}}) + o(1))\}.$$

So with $N \to \infty$ we get that for this test

$$E_{2|1}(E_{1|2}) \leq \min_{P\in\mathcal{P}(\mathcal{S})} \min_{Q:D(Q||G_1|P)\leq E_{1|2}} D(Q||G_2|P). \tag{9}$$

Then we have to be convinced that any other sequence $\tilde{\varphi}^N$ of tests defined for every $\mathbf{s} \in \mathcal{S}^N$ by the sets $\tilde{\mathcal{A}}_{\mathbf{s}}^N$ such, that

$$\alpha_{1|2}{}^{(N)}(\tilde{\varphi}^{(N)}) \leq \exp\{-NE_{1|2}\}, \tag{10}$$

and

$$\alpha_{2|1}^{(N)}(\tilde{\varphi}^{(N)}) \leq \alpha_{2|1}^{(N)}(\varphi^{(N)}),$$

in fact coincide with $\varphi^{(N)}$ defined in (8). Let us consider the sets $\tilde{\mathcal{A}}_{\mathbf{s}}^N \bigcap \mathcal{A}_{\mathbf{s}}^N$, $\mathbf{s} \in \mathcal{S}^N$. This intersection cannot be void, because in that case $\overline{\tilde{\mathcal{A}}_{\mathbf{s}}^N \bigcap \mathcal{A}_{\mathbf{s}}^N}$ will be equal to $\mathcal{X}^N = \overline{\mathcal{A}_{\mathbf{s}}^N} \bigcup \overline{\tilde{\mathcal{A}}^N}_{\mathbf{s}}$ and the probabilities $G_1^N(\overline{\mathcal{A}_{\mathbf{s}}^N}|\mathbf{s})$ and $G_1^N(\overline{\tilde{\mathcal{A}}_{\mathbf{s}}^N}|\mathbf{s})$ cannot be small simultaneously.

Now because

$$G_1^N(\overline{\mathcal{A}_{\mathbf{s}}^N \bigcap \tilde{\mathcal{A}}_{\mathbf{s}}^N}|\mathbf{s}) \leq$$

$$G_1^N(\overline{\mathcal{A}_{\mathbf{s}}^N})|\mathbf{s}) + G_1^N(\overline{\tilde{\mathcal{A}}_{\mathbf{s}}^N})|\mathbf{s}) \leq 2 \cdot \exp\{-NE_{1|2}\} = \exp\{-N(E_{1|2} + o(1)\},$$

from (6) we obtain

$$G_2^N(\mathcal{A}_{\mathbf{s}}^N \bigcap \tilde{\mathcal{A}}_{\mathbf{s}}^N|\mathbf{s}) \leq G_2^N(\mathcal{A}_{\mathbf{s}}^N|\mathbf{s}) \leq \exp\{-N(\min_{P\in\mathcal{P}(\mathcal{S})} \min_{Q:D(Q||G_1|P)\leq E_{1|2}} D(Q||G_2|P))\}$$

and so we conclude that if we exclude from $\tilde{\mathcal{A}}_{\mathbf{s}}^N$ the vectors $\mathbf{x}$ of the types $\mathcal{T}_{P,Q}^N(X|\mathbf{s})$ with $D(Q||G_1|P) > E_{1|2}$ we do not make reliabilities of the test $\tilde{\varphi}^{(N)}$ worse. It is left to remark that when we add to $\tilde{\mathcal{A}}_{\mathbf{s}}^N$ all types $\mathcal{T}_{P,Q}^N(X|\mathbf{s})$ with $D(Q||G_1|P) \leq E_{1|2}$, that is we take $\tilde{\mathcal{A}}_{\mathbf{s}}^N = \mathcal{A}_{\mathbf{s}}^N$, we obtain that (6) and (9) are valid, that is the test $\varphi^{(N)}$ is optimal.

References

1. R. Ahlswede, Coloring hypergraphs: a new approach to multi-user source coding, I and II, J. Combin., Inform. System Sciences, Vol. 4, No. 1, 75-115, 1979, Vol. 5, No. 2, 220-269, 1980.
2. F.-W. Fu and S.-Y. Shen, Hypothesis testing for arbitrarily varying sources with exponential-type constraint, IEEE Trans. Inform. Theory, Vol. 44, No. 2, 892-895, 1998.
3. I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, Academic press, New York, 1981.
4. W. Hoeffding, Asymptotically optimal tests for multinomial distributions, Ann. Math. Stat., Vol. 36, 369-401, 1965.
5. I. Csiszár and G. Longo, On the error exponent for source coding and for testing simple statistical hypotheses, Studia Sc. Math. Hungarica, Vol. 6, 181-191, 1971.
6. G. Tusnady, On asymptotically optimal tests, Ann. Statist., Vol. 5, No. 2, 385-393, 1977.
7. G. Longo and A. Sgarro, The error exponent for the testing of simple statistical hypotheses, a combinatorial approach, J. Combin., Inform. System Sciences, Vol. 5, No 1, 58-67, 1980.
8. L. Birgé, Vitesse maximales de décroissance des erreurs et tests optimaux associés, Z. Wahrsch. verw. Gebiete, Vol. 55, 261–273, 1981.
9. E. A. Haroutunian, Logarithmically asymptotically optimal testing of multiple statistical hypotheses, Probl. Control and Inform. Theory, Vol. 19, No. 5–6, 413–421, 1990.
10. R. Ahlswede and E. Haroutunian, On logarithmically asymptotically optimal testing of hypotheses and identification: research program, examples and partial results, preprint.