On some testing of hypothesis problems with information constraints *

Burnashev M. V.[†], Amari S.[‡], Han T. S.[§]

Abstract

A problem of hypothesis testing is considered where some part of data can not be directly observed. Our helper observes those data and can send us some limited amount of information about them. What kind of that information allows us to make the best statistical inferences ? In particular, what is the minimal information sufficient in order to get the same results as if we could observe directly all data ? Some upper bounds for that minimal amount of information and some related results are obtained.

Key words and phrases: hypothesis testing, error probability, rate of transmission, critical rate.

I. Introduction. General description of the problem

In this paper we would like to draw specialist's attention to one rather new and essentially open class of problems and to present some particular results on that matter. Although those problems have been started to investigate in information theory [1 - 5], they have statistical essence and knowledge of information theory is not needed for their understanding.

---

*The research described in this publication was made possible in part by Grant N 98-01-04108 from the Russian Fund for Fundamental Research.

[†]Institute for Problems of Information Transmission, RAS, Bolshoi Karetnii 19, 101447, Moscow, Russia.

[‡]RIKEN Brain Science Institute, Wako-shi, Hirosawa 2-1, Saitama 351-0198, Japan.

[§]Graduate School of Information Systems, University of Electro–Communications, Chfugaoka 1 - 5 - 1, Chofu, Tokyo 182, Japan.

For that purpose we consider some particularly interesting cases of the following general problem.

Let a "statistician" should make a certain statistical inferences concerning the system state (for example, to estimate some unknown parameter, to test some hypotheses, etc.). There are two sets of data (observations): set $\mathcal{A}$ ("available") and set $\mathcal{R}$ ("remote").

Statistician directly observes all data from the set $\mathcal{A}$.

He can not observe directly data from the remote set $\mathcal{R}$, but his "helper" can observe them. Moreover the helper is allowed to send statistician some limited amount of information about those data.

The problem is: what kind of information (limited) about those remote data should send the helper that will allow statistician to make the best possible statistical inferences (for example, to get the minimum mean–square error for parameter estimate, etc.) ?

There are many practical situations when we meet this kind of problem. For example, in some applications the set $\mathcal{R}$ can be regarded as some "nuisance noise" that has "contaminated" already the data from the set $\mathcal{A}$, and therefore we would like to "remove" (as much as possible) that "contamination" in order to improve our statistical inferences.

We will deal below with discrete–time models and moreover, by "limited amount of information" we will mean that the helper can send information with communication rate not exceeding some prescribed value $R > 0$ (see Example 1 below). Of course, if R is such large that the helper can simply resend to statistician all data from $\mathcal{R}$ then we come back to traditional statistical problem (that is not of our interest here). For that reason it is natural to assume that R is small enough in order to avoid such primitive resending.

Nevertheless, even with such assumption there are some cases (sometimes, probably, natural) when the optimal solution can be obtained quite easily. Certainly, it will always be the case when data from sets $\mathcal{A}$ and $\mathcal{R}$ both represent independent observations of the same phenomenon. The following particular example illustrates such case.

Example 1. Let all data $\{x_1, \ldots, x_n\}$ from the set $\mathcal{A}$ and all data $\{x_{n+1}, \ldots, x_{n+m}\}$ from the set $\mathcal{R}$ have the form

$$x_i = \theta + \xi_i \, ; \qquad i = 1, \ldots, n+m \, ; \quad -\infty < a < \theta < b < \infty \, ,$$