Selected Topics of Information Theory and Mathematical Statistics

Marat V. Burnashev (Moscow)

1 Testing of hypotheses and Identification

Identification became a quite popular topic in Information Theory. It is a very good example of interaction between Mathematical Statistics (Testing of Hypotheses) and Information Theory. There are still many interesting open problems in it. I am going to devote a good part of these lectures to that topic. In order to come smoothly to identification problems we will need first some knowledge on testing of hypotheses and related notions.

1.1 Preliminaries. Testing of hypotheses and L_1 - distance.

Since later in identification we will deal only with finite alphabets all necessary preliminaries will be given here also only for finite alphabets.

L_1 - distance. Let $Y = (y_1, y_2, \ldots, y_K)$ be some finite set (alphabet). *Probability measure* P on Y is defined by any collection of *nonnegative values* $\{P(y_1), P(y_2), \ldots, P(y_K)\}$ satisfying the condition:

$$\sum_{i=1}^{K} P(y_i) = 1$$

L_1 - *distance* between a pair of probability measures P and Q on Y is defined as

$$\|P - Q\| = \sum_{i=1}^{K} |P(y_i) - Q(y_i)|$$

(sometimes it is called a *variational distance* and sometimes the factor 1/2 in front of the sum is used).

Support of a measure P is defined as

$$\text{supp } P = \{y \in Y : P(y) > 0\}$$

(i.e. it consists of all points y of the set Y where the measure P is strongly positive).

Examples. If $P = Q$ (i.e. $P(y) = Q(y)$ for any $y \in Y$) then obviously $\|P - Q\| = 0$. If supports of measures P and Q do not intersect (i.e. measures P and Q are *orthogonal* to each other) then $\|P - Q\| = 2$.

L_1 - distance between a probability measure P and some collection of probability measures $\{Q_\alpha, \alpha \in \mathcal{A}\}$ (where $\mathcal{A}$ is an arbitrary index set) is defined as

$$\|P - \{Q_\alpha, \alpha \in \mathcal{A}\}\| = \inf_{\alpha \in \mathcal{A}} \|P - Q_\alpha\|$$

We will need one more operation with probability measures.

Let $\{Q_\alpha, \alpha \in \mathcal{A}\}$ (where $\mathcal{A}$ is an arbitrary index set) be some collection of probability measures. *Convex hull* of the family $\{Q_\alpha, \alpha \in \mathcal{A}\}$ (denoted by conv $\{Q_\alpha, \alpha \in \mathcal{A}\}$ or simply conv $\{Q_\alpha\}$) is the set of all possible finite convex linear combinations of measures from $\{Q_\alpha\}$. In other words, conv $\{Q_\alpha\}$ consists of all measures F that can be represented in the form

$$F = \sum_{i \in \mathcal{A}} c_i Q_i$$

with

$$c_i \geq 0 \text{ for all } i \ , \ \text{ and } \ \sum_{i \in \mathcal{A}} c_i = 1 \ ,$$

and only finite number of c_i's are not equal to zero.

We denote by $\mathcal{P}(Y)$ the set of *all* probability measures on Y. The set $\mathcal{P}(Y)$ is a *complete* set with respect to L_1 - distance (metric). In other words, if $\{Q_n, \ n = 1, 2, \ldots, \}$ is a *fundamental* sequence (i.e. $\|Q_k - Q_m\| \leq \epsilon$ if $k, m \geq n(\epsilon)$ for any $\epsilon > 0$) then it has a *limit* $Q_0 \in \mathcal{P}(Y)$ (i.e. it is also a *probability* measure).

Example. Let on the set $Y = (y_1, y_2, \ldots, y_K)$ be given the collection of probability measures $\{Q_i \ , \ i = 1, \ldots, K\}$ such that $Q_i(y_i) = 1$ for any $i = 1, \ldots, K$ (it is clear that then $Q_i(y_j) = 0$ for any $i \neq j$). Then $\text{conv}\{Q_i\} = \mathcal{P}(Y)$.

Testing of simple hypotheses. Let on a finite space Y there are given two probability measures: P and Q. We get *one* observation $y \in Y$ according to some distribution F. We know that an unknown distribution F satisfies one of the following two possibilities (hypotheses):

$$H_0 : \ F = P \ ,$$

or

$$H_1 : \ F = Q$$