

## Data compressors beyond data compression.

**Boris Ryabko, Nadezhda Savina**

*Institute of Computational Technologies of Siberian Branch of RAS,  
Siberian University of Consumer Cooperatives, Novosibirsk, Russia*

In the last thirty years, it was recognized that data compressors can be used for many purposes which are far from file compaction. In particular, it was shown that methods of data compression can be used for prediction and hypothesis testing for time series, mathematical statistics, randomness testing and cryptography, see *B. Ryabko, J. Astola, M. Malyutov. Compression-Based Methods of Statistical Analysis and Prediction of Time Series Springer, 2016.*

The main goal of the report is to give a compression-based solution for the following problems:

- i) Homogeneity test, where there are several sequences  $x_1^1 x_2^1 \dots x_{n_1}^1, \dots, x_1^k x_2^k \dots x_{n_k}^k, y_1^1 y_2^1 \dots y_{m_1}^1, \dots, y_1^s y_2^s \dots y_{m_s}^s$ , generated either by a single source or by two different ones, and two corresponding hypotheses.
- ii) Classification problems, where there are samples  $x_1^1 x_2^1 \dots x_{n_1}^1, \dots, x_1^k x_2^k \dots x_{n_k}^k, y_1^1 y_2^1 \dots y_{m_1}^1, \dots, y_1^s y_2^s \dots y_{m_s}^s$  generated two different (but unknown) sources and  $z_1 \dots z_l$  is generated by one of the two. The goal is to determine which of them generated  $z_1 \dots z_l$ .
- iii) Estimation of a so-called measurement of interdependence, or the association.

**A distinction of the suggested method from other approaches is that it belongs to the framework of mathematical statistics.**