Kozhevin A. A. (Moscow, Russia). **An information theory-based approach to feature selection**

The talk is devoted to the estimators of conditional entropy [1] and mutual information in the mixed model [2] and to the new result for feature selection procedure based on information theory approach.

Mixed model is described in the paper [1]. Let $\zeta_n = \{(X^i, Y^i)\}_{i=1}^n$ be a sample of i.i.d. observations, $(X^1, Y^1) \sim (X, Y)$ where $X = (X_1, \ldots, X_d)$ is absolutely continuous random vector in $\mathbb{R}^d$, Y is a random variable taking values in a finite set M. The set of indices $S = \{s_1, \ldots s_m\} \subset \{1, \ldots, d\}$ ($s_i \neq s_j$ for $i \neq j$) and the set of factors X_S, where $u_L = (u_{l_1}, \ldots, u_{l_m})$ for $u = (u_1, \ldots, u_d)$ and $L = \{l_1, \ldots, l_m\}$, are called relevant if for each $y \in M$ relation $f_{Y|X}(y|X) = f_{Y|X_S}(y|X_S)$ is valid a.s. Let $Q_m = \{\{l_1, \ldots, l_m\} \subset \{1, \ldots, d\} \colon l_i \neq l_j, i \neq j\}$. For each $L \in Q_m$ define $\zeta_{n,L} = \{(X_L^i, Y^i)\}_{i=1}^n$ and estimate mutual information $I(X_L; Y)$ for each sample $\zeta_{n,L}$ by the method proposed in [2]. The resulting estimates we denote as $\hat{I}_{n,k,L}$ where $k \in \{1, \ldots, n-1\}$ is a parameter of the method.

Define $\hat{S}_{n,k} = argmax_{L \in Q_m} \hat{I}_{n,k,L}$. In case the maximum $\hat{I}_{n,k,L}$ is reached at several sets Q_m, $\hat{S}_{n,k}$ can be defined as the first of such sets in the sense of lexicographical order. The following new result is valid.

Theorem. Let m be known, relevant set of factors of length m be unique. Density f_X is strictly positive, for each $L \subset \{1, \ldots, d\}$ and $y \in M$ density $f_{X_L,Y}(\cdot, y)$ is C_0-constricted ($C_0 > 0$) and for some $\varepsilon > 0$ relation $\mathsf{E}|\log f_{X_L}(X_L)|^{2+\varepsilon} < \infty$ is valid. Then $\mathsf{P}(\hat{S}_{n,k} = S) \to 1$ when $n \to \infty$ for each $\alpha \in (0, 1)$ and $k \propto n^\alpha$.

REFERENCES

1. *Bulinski, A., Kozhevin, A.* Statistical Estimation of Conditional Shannon Entropy. ESAIM: Probability and Statistics (Accepted for publication November 28, 2018), p. 1-37, DOI: https://doi.org/10.1051/ps/2018026.

2. *Bulinski, A., Kozhevin, A.* Statistical Estimation of Mutual Information for Mixed Model. (to appear)

This work is supported by the Lomonosov Moscow State University under grant “Modern Problems of the Fundamental Mathematics and Mechanics”.