Bulinski A.V. (Moscow, Russia) — **Asymptotic behavior of entropy estimates**. The concept of entropy is fundamental in physics and mathematics. Significant contributions to the development of this notion were made by L.Boltzmann, J.Gibbs, M.Plank, C.Shannon, A.N.Kolmogorov, Ya.G.Sinai, A.Renyi, C.Tsallis, A.S.Holevo. In fact, there are different definitions of entropy. We discuss important problems where one employs statistical estimates of the due entropy. For example such estimates are used in the detection of material inhomogeneities (see, e.g., [1]). They are applied also in the feature selection theory (see, e.g., [2]) having applications in medical and biological studies. Here one employs the estimates of the mutual information involving entropy estimates. One can indicate a number of other applications mentioned, e.g., in [7]. There are various complementary approaches to the entropy estimation, we refer, e.g., to the works by L.F.Kozachenko and N.N.Leonenko (1987), P.Hall and S.C.Morton (1993), A.B.Tsybakov and E.C.Van der Meulen (1996), E.G.Miller (2003), L.Paninski (2003), D.Stowell and M.D.Plumbley (2009), K.Sricharan et al. (2013), E.Archer et al. (2016), A.Charzynska and A.Gambin (2016), S.Delattre and N.Fournier (2017).

Along with a survey we concentrate on our quite new results. The behavior of the Kozachenko - Leonenko estimates for the (differential) Shannon entropy is considered when the number of i.i.d. vector-valued observations tends to infinity. In [3] the asymptotic unbiasedness and L^2-consistency of the estimates are established under wide conditions. The assumptions employed involve analogues of the Hardy - Littlewood maximal function. It is shown that the results are valid in particular for the entropy estimation of any nondegenerate Gaussian vector.

Then we turn to the new statistical estimates of conditional Shannon entropy proposed in [4] in the framework of the model describing a discrete response variable depending on a vector of d factors having a density w.r.t. the Lebesgue measure in $\mathbb{R}^d$. Namely, the mixed-pair model (X, Y) is considered where X and Y take values in $\mathbb{R}^d$ and an arbitrary finite set M, respectively. Such models include, for instance, the famous logistic regression (see, e.g., [6]). In contrast to the well-known Kozachenko – Leonenko estimates of unconditional entropy the proposed estimates are constructed by means of the certain k-nearest neighbor statistics (where $k = k_n$ depends on amount of observations n) and a random number of i.i.d. observations contained in the balls of specified random centers and random radii. The asymptotic unbiasedness and L^2-consistency of the new estimates are established under simple conditions as well. It is worth to emphasize that our estimates construction (cf. [5]) does not suppose the existence of a topological structure on a set M.

REFERENCES

1. *Alonso-Ruiz, P., Spodarev, E.* Entropy-based inhomogeneity detection in porous media, arXiv:1611.02241v1, 2016, pp. 1–18.
2. *Bennasar M., Hicks Y., Setchi R.* Feature selection using joint mutual information maximisation. Expert Systems with Applications, 2015, 42, pp. 8520–8532.
3. *Bulinski A., Dimitrov D.* Statistical estimation of the Shannon entropy, arXiv:1801.02050v1, 2018, pp. 1–28.
4. *Bulinski A., Kozhevin A.* Statistical Estimation of Conditional Shannon Entropy. arXiv:1804.08741v1, 2018, pp. 1-33.
5. *Gao W., Kannan S., Oh S., Viswanath P.* Estimating mutual information for discrete-continuos mixtures. arXiv:1709.06212v2, 2018, pp. 1-25.
6. *Massaron L., Boschetti A.* Regression Analysis with Python. Packt Publishing Ltd., Birmingham, 2016.
7. *Pál, D., Póczos, B., Szepesvári C.* Estimation of Rényi entropy and mutual information based on generalized nearest-neighbor graphs. arXiv:1003.1954v2, 2010, pp. 1–24.

The work is supported by the Russian Science Foundation under grant 14-21-00162 and performed at the Steklov Mathematical Institute of Russian Academy of Sciences.