Булинский А.В. (Москва, Россия) — **Асимптотическое поведение оценок энтропии**.

Понятие энтропии является фундаментальным в физике и математике. Значительный вклад в развитие этого понятия внесли L.Boltzmann, J.Gibbs, M.Plank, C.Shannon, А.Н.Колмогоров, Я.Г.Синай, A.Renyi, C.Tsallis, А.С.Холево. В действительности, имеются различные определения энтропии. Мы рассматриваем важные задачи, в которых используются статистические оценки надлежащей энтропии. Например, такие оценки применяются для выявления неоднородностей материалов ([1]). Они также полезны в теории выбора значимых факторов ([2]), имеющей применения в медицинских и биологических исследованиях. Для этого употребляются оценки взаимной информации, использующие оценки энтропии. Можно указать множество других приложений, упомянутых, например, в [7]. Есть дополняющие друг друга подходы к оцениванию энтропии. В этой связи укажем на работы Л.Ф.Козаченко и Н.Н.Леоненко (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).

Наряду с обзором мы остановимся на наших недавних работах. Асимптотическое поведение оценок Козаченко - Леоненко для дифференицальной энтропии Шеннона рассматривается, когда число независимых одинаково распределенных векторно-значных наблюдений стремится к бесконечности. В [3] асимптотическая несмещенность и L^2-состоятельность оценок устанавливаются при весьма широких условиях, которые используют аналоги максимальной функции Харди - Литтлвуда. В частности, результаты верны для оценки энтропии произвольного невырожденного гауссовского вектора.

Далее мы обращаемся к новым оценкам условной энтропии Шеннона, введенным в [4]. Рассматривается модель, описывающая дискретный отклик, зависящий от d факторов, имеющих плотность распределения по мере Лебега в $\mathbb{R}^d$. А именно, изучается модель смешанной пары (X, Y), когда X и Y принимают значения соответственно в $\mathbb{R}^d$ и произвольном конечном множестве M. Такие модели включают в себя знаменитую логистическую регрессию (см., напр., [6]). В отличие от хорошо известных оценок Козаченко - Леоненко для безусловной энтропии, предложенные оценки строятся при помощи некоторых статистик k-ближайших соседей (где $k = k_n$ зависит от общего числа наблюдений n) и случайного числа независимых одинаково распределенных наблюдений, содержащихся в определенных шарах со случайными центрами и случайными радиусами. Асимптотическая несмещенность и L^2-состоятельность новых оценок доказываются также при простых условиях. Следует подчеркнуть, что наша конструкция оценок (ср. с [5]) не предполагает наличия топологической структуры на 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.

Работа выполнена в Математическом институте имени В.А.Стеклова РАН при поддержке Российского научного фонда, грант 14-21-00162.