MASCOT-NUM 2019

JOURDAN Astrid
EISTI

Abstract

Space-filling designs based on Rényi entropy

Space-filling designs are commonly used for selecting the input values of time-consuming computer codes. Since the true relation between the computer response and inputs is not known, the designs should allow one to fit a variety of models and should provide information about all portions of the experimental region. One strategy for selecting the values of the inputs which to observe the response is to choose these values so they are spread evenly throughout the experimental region, according to a "space-filling criterion". Many space-filling criteria have been investigated in the literature. Some of them quantify how the points fill up the space using the distance between points, such as the maximin distance [5] or the Audze-Eglais criterion [1]. Others measure the difference between the empirical distribution of the design points and the uniform distribution, such as the discrepancy ([9], [3]) or Kullback-Leibler criterion [6].In this paper, we use results discussed in Pronzato's work ([10], [8]) to build space-filling designs based on Renyi's entropy. Suppose that the points $x_1,...,x_n$ of the design D, are *n* independent observations of the random vector $X=(X_1,...,X_d)$ with absolutely continuous density function *f* concentrated on the unit cube $[0,1]^d$ (we reduce the design space to the unit cube). Rényi entropy,

$$H_q(D) = \frac{1}{1-q} \ln \int_E f(x)^q dx \quad \text{, with } q \in]0,1[$$

measures the difference between f and the uniform density function in so far as, one always has Hq(D) (D), zero being uniquely attained by the uniform density. This latter property confirms that maximizing Rényi entropy makes f converge toward the uniform density. We investigate three ways for estimating the entropy:

- a Monte Carlo method [2] where the unknown density function f is replaced by its kernel density estimate [11],• an estimation based on the nearest neighbor distance [7],• a method based on the minimum spanning tree built from the design points [4].

References

[1] Audze, P. and V. Eglais. New approach for planning out of experiments, Problems of Dynamics and Strengths, 35:104-107, 1977.[2] Beirlant J., Dudewicz E.J., Györfi L., Van Der Meulen E.C. Nonparametric entropy estimation : an overview. Int. J. Math. Stat. Sci., 6(1):17-39, 1997.[3] Fang K.T., Li R., Sudjianto A. Design and modeling for computer experiments. Chapman&Hall, London, 2006.[4] Hero A., Bing Ma, Michel O., Gorman J. Applications of entropic spanning graphs. Signal Processing Magazine, IEEE, 19:85 – 95, 2002.

[5] Johnson M.E., Moore L.M., Ylvisaker D. Minimax and maximin distance design. J. Statist. Plann. Inf., 26:131-148, 1990.[6] Jourdan A. et Franco J. Optimal Latin hypercube designs for the Kullback-Leibler criterion. AStA Advances in Statistical Analysis, 94(4):341-351, 2010.[7] Kosachenko L.F., Leonenko N.N.. Sample estimate of entropy of a random vector. Problem of Information Transmission, 23:95-101, 1987.[8] Leonenko N., Pronzato L.,Savani V., A class of Rényi information estimators for multidimensional densities, Ann. Statist. 36:2153 - 2182; correction by Leonenko and Pronzato 2010, Ann. Statist., 38:3837 – 3838, 2008.[9]Niederreiter H. Point sets and sequences with small discrepancy. Monasth. Math., 104:273-337, 1987.[10] Pronzato L. Minimax and maximin space-filling designs: some properties and methods for construction. Journal de la Société Française de Statistique, 158(1), 2017.[11]Silverman B.W. Density estimation for statistics and data analysis. Chapman & Hall, London, 1986.