

Optimal design of sampling survey for efficient parameter estimation

Wei Zheng ¹

For many tasks of data analysis, we may only have the information of the explanatory variable and the evaluation of the response values are quite expensive. While it is impractical or too costly to obtain the responses of all units, a natural remedy is to judiciously select a good sample of units, for which the responses are to be evaluated. In this paper, we adopt the classical criteria in design of experiments to quantify the information of a given sample regarding parameter estimation. Then, we provide a theoretical justification for approximating the optimal sample problem by a continuous problem, for which fast algorithms can be further developed with the guarantee of global convergence. Our results have the following novelties: (i) The statistical efficiency of any candidate sample can be evaluated without knowing the exact optimal sample; (ii) It can be applied to a very wide class of statistical models; (iii) It can be integrated with a broad class of information criteria; (iv) It is much faster than existing algorithms. (v) A geometric interpretation is adopted to theoretically justify the relaxation of the original combinatorial problem to continuous optimization problem.

¹Department of Business Analytics and Statistics, University of Tennessee, Knoxville, TN 37966, USA, E-mail: wzheng9@utk.edu