Optimal Transport Methods in Survey Sampling

Yves Tillé

Université de Neuchâtel

Optimal transport methods dealt with the origin of material transfers and the optimal allocation of resources for their transport. These methods can also be used in survey sampling theory. A first application is statistical matching, which aims to integrate two statistical sources. These sources can be two samples or one sample and the whole population. When two samples have been selected from the same population and information has been collected on different variables of interest, it is interesting to match the two surveys in order to analyze, for example, contingency tables or covariances. In this paper, we propose an efficient method for matching two samples, each of which may contain a weighting system. Several variants are proposed to create a directly usable file that integrates data from both information sources. A second application is the estimation of complex statistics when specific weighting schemes are used for each variable of interest. Indeed, many recently proposed survey weighting methods provide weights that depend on the variable of interest, thus deviating from the universal nature of the weights proposed by survey data processing methods. Simulations show that the optimal transport method for weight harmonization can provide accurate parameter estimates in different scenarios, especially in situations where the differences between the sample and population distributions are large. It therefore appears to be a versatile solution.