Credible Distributions for Ranking Entities

Gauri Sankar Datta

Department of Statistics, University of Georgia, Athens, GA and Center for Statistical Research and Methodology, U.S. Census Bureau, Suitland, MD

Abstract:
Inference on overall ranking of a set of entities, which may be individuals, small areas, hospitals, herds of animals, etc. is a problem of significant interest. Effective implementation of welfare and well-being programs overseen by government or private organizations, identifying the better yielding crop varieties or top-breeding herds, requires reliable estimation of the rank vector. Implicit estimation of the component ranks based on point estimates of entity means does not account for uncertainty in these estimates. Even then users often routinely consider ranks based on published estimates of subpopulation or small area means as true ranks to make decisions and develop policy goals. Treating estimated ranks without any regard to them being subject to estimation uncertainty is problematic. Recently, several researchers recognized the problem and proposed in a pioneering paper an interesting and useful solution to this problem using the frequentist approach to statistical inference. Recognizing the importance of the problem, in the current manuscript we proposed an alternative solution based on the Bayesian approach to inference. The Bayesian solution to this problem developed here is competitive, more effective and informative, and is as easy to implement as it is to compute the posterior means and variances of the entity means. Using established hierarchical Bayesian modeling and widely accepted noninformative prior distributions, we developed various empirical credible sets of the entity means to be ranked. Using the credible sets we created, novel credible distributions for the rank vector of the entities. Applications of the proposed method to ranking major league baseball players, states in terms of commuting times to workplace, and states in terms of median incomes show the method's widespread effectiveness. Evaluation of the proposed procedure in terms of accuracy and stability in two of our applications and a simulation study mimicking the baseball application show that the proposed procedure is substantially better than its frequentist counterpart. Finally, while covariates cannot be incorporated in the frequentist procedure, the proposed solution based on hierarchical Bayesian model has no difficulty in incorporating such information.

Keywords and phrases: Bayes; Credible sets; Fay-Herriot model; hierarchical Bayes; highest posterior density; small area estimation; unstructured Bayes.

This is a joint work with Ms. Yiren Hou, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA, and Professor Abhyuday Mandal, Department of Statistics, University of Georgia, Athens, GA, 30606, USA.