Non-Bayesian Parametric Missing-Mass Estimation

Tirza Routtenberg, Shir Cohen, and Lang Tong

Abstract:
We consider the classical problem of missing-mass estimation, which deals with estimating the total probability of unseen elements in a sample. The missing-mass estimation problem has various applications in machine learning, statistics, language processing, ecology, sensor networks, and others. In this work, we introduce a frequentist, non-Bayesian parametric model of the problem of missing-mass estimation. We introduce the concept of missing-mass unbiasedness by using the Lehmann unbiasedness definition. We derive a non-Bayesian CCRB-type lower bound on the missing-mass MSE (mmMSE) based on the missing-mass unbiasedness. Based on the new bound, we propose a new method to improve existing estimators by an iterative missing-mass Fisher-scoring method. We demonstrate via numerical simulations that the proposed bound is a valid and informative lower bound on the mmMSE of state-of- the-art estimators. We also show that the performance of the Laplace estimator is improved by using the new missing-mass Fisher-scoring method.