Goodness-of-fit Tests under Permutations

Chen Chen

Preprint no. 2019-12

Abstract

Several new goodness-of-fit tests are proposed on countable alphabets, where certain fundamental statistical concepts associated with random variables, such as cumulative distribution functions, characteristic functions and moments, may not exist. An entropic perspective by ways of the entropic basis, derived from the well-known Turing's formula, is introduced. A new characterization theory of probability distributions on alphabets is established by means of the entropic basis. Based on this logic framework several goodness-of-fit tests are developed.

Toward developing the new goodness-of-fit tests, a one-to-one correspondence between a given probability distribution and its entropic basis is first established. In case the cardinality of underlying distribution is finite, say K, the first K entropic moments uniquely determine the underlying probability distribution up to a permutation on the index set. For each of the entropic moments, an uniformly minimum variance unbiased estimator (UMVUE) is introduced. Based on the sampling distribution of the UMVUEs of the entropic moments and the multivariate delta method, two new Chi-squared goodness-of-fit tests are constructed and their asymptotic distributional properties are established in theory. However it is also observed that these new tests are difficult to implement numerically. To alleviate the computational difficulty in implementation, a heuristic exact test for goodness-of-fit is proposed. The performance of the proposed tests is evaluated by simulation studies under a range of distributions. The new tests are also illustrated in several real life applications.

Department of Mathematics and Statistics, UNC Charlotte, Charlotte, NC 28223