

1079-62-419

Brian D. White* (bdwhite@mail.usf.edu). *Applications of the Brownian Bridge and the Kolmogorov-Smirnov Test to Clustering.*

The two-sample Kolmogorov-Smirnov test is a nonparametric test for the equality of continuous, one-dimensional probability distributions that can be used to compare two independent samples. The K-S statistic is the supremum of distances that quantifies a distance between the empirical distribution functions of the two samples. The Kolmogorov distribution is the distribution of the supremum of the Brownian bridge. Under the null hypothesis, the K-S statistic converges to the Kolmogorov distribution. The test does not require specification of the hypothesized distribution yet is sensitive to differences in location and shape of the empirical distribution functions of the two samples. In insurance, the cause and nature of an injury as well as body part involved play an important role in estimating the ultimate claim size. Given the large number of possible levels, categorical variables such as these often pose a challenge to an analyst. To reduce the levels to a manageable number, clustering methods are commonly used. We present a novel approach to the use of the K-S statistic to clustering based on both location and shape. We show that this approach is often more robust than methods that focus only on location, particularly for long tail distributions such as claim sizes. (Received January 18, 2012)