Pearson’s Chi-squared statistics: approximation theory and beyond

Abstract:

Pearson’s Chi-squared test is very widely used in practice and has a long history. The validity with a large number of cells or small expected frequencies has been open for a long time. We provide a solution to this open problem by rigorously establishing a distributional approximation of Pearson’s Chi-squared test statistic by using a high-dimensional central limit theorem for quadratic forms of random vectors. We also propose a modified chi-squared statistic with a faster convergence rate and propose the concept of adjusted degrees of freedom. Our procedure is applied to goodness-of-fit test for the social life feeling data and the Rochdale data.