Goodness-of-fit for sparse distributions in high energy physics

B.D. Yabsley

School of Physics, University of Sydney, NSW 2006 Australia

Abstract

We examine the behaviour of Pearson’s chi-square $\chi^2_P$, the “likelihood chi-square” $\chi^2_L$, and Zelterman’s $D^2$ statistic for some model high energy physics problems where the data are sparse. Maximum likelihood fits to sparse data are becoming common in the field, leading to problems in assessing the goodness-of-fit because typical bin populations are low (or because the fit is unbinned). We assess the performance of $\chi^2_P$, $\chi^2_L$, and $D^2$ in this regime (where they follow non-$\chi^2$ distributions) as measures of goodness-of-fit for sparse binned maximum likelihood fits. Power against likely alternative hypotheses, and the usefulness of asymptotic distributions (rather than toy MC study), are assessed.