Computer-enabled metrics of statistical significance

Mark Tygert

Facebook Artificial Intelligence Research, USA

Email: tygert@aya.yale.edu

Goodness-of-fit tests based on the Euclidean distance often outperform chi-square and other classical tests (including the standard exact tests) by at least an order of magnitude when the model being tested for goodness-of-fit is a discrete probability distribution that is not close to uniform. Goodness-of-fit tests based on the Euclidean distance are now practical and convenient: although the actual values taken by the Euclidean distance and similar goodness-of-fit statistics are seldom interpretable without the aid of a computer, black-box software can rapidly calculate their precise significance. This is joint with Rachel Ward of the University of Texas at Austin and Will Perkins of the University of Birmingham, England.