

A saga of the CMEs

Jeff Wu

Coca-Cola Chair in Engineering Statistics, Georgia Tech

Abstract

A two-factor interaction is the difference between two conditional main effects (CMEs). In many investigations, CMEs give a better description of how factor A “interacts” with factor B in the sense that the presence or absence of gene A (or compound A) affects the effect of gene B (or compound B) on a particular disease. In this talk I will share the story of my association with the use of CMES in developing new analysis methods over the last three decades. In 1988 when I started my GM/NSERC Chair in Waterloo, I realized from analyzing a designed experiment data at GM that the use of CMEs as a reparametrization of the factorial effects space can “de-alias” aliased interactions in resolution IV or III designs. Such feat was deemed impossible using the existing paradigms. But, alas, the discovery was ahead of its time. I did not further pursue it because a paper would likely be rejected and the possibility it raised be ridiculed by traditionalists. Fast forward to the winter of 2011 when I was informed that I would give the 2011 Fisher Lecture at the JSM. I knew my time had finally come. That original idea, properly expanded, would be presented in the lecture and the paper (JASA, 2015) would not be rejected☺. A fully developed paper by Su-Wu appeared in JQT in 2016 (and won the Brumbaugh award). It proposes some principles governing the relationships between CMEs and traditional factorial effects and a simple and effective analysis procedure. It works well on real data from resolution IV designs: achieving model parsimony and, more importantly, the CME terms in the model are more interpretable in the engineering context. By then my team was on full speed to pursue the idea further. Another talented student Simon Mak joined me to use the CMEs as the basis functions in variable selection for general binary input factors with no restrictions on the relationships between factors (while fractional factorial designs employ orthogonal factors). Without factorial orthogonality, new groupings of CMEs are proposed. A new variable selection procedure, called cmenet, employs two principles (CME coupling and CME reduction) to effectively navigate the selection algorithm. Applied to a gene association study, cmenet not only provides more parsimonious models and improved predictive performance over existing methods, but also reveals important insights on gene activation behavior which can guide further experiments. The talk concludes with a recent work on how CMEs can be used to improve modeling and optimization in online e-commerce.