Abstract:

The agreement problem has a long history starting with the correlation coefficient for agreement over 100 years ago and it covers a broad range of data with applications arising from many different fields. In medical and other related sciences, clinical or experimental measurements usually serve as a basis for diagnostic, prognostic, therapeutic, and performance evaluations. Examples can be assessing the reliability of multiple raters (or measurement methods), assessing the suitability for tumor evaluation of using a local laboratory or a central laboratory in a randomized clinical trial (RCT), validating surrogate endpoints in a study, determining that the important outcome measurements are interchangeable among the evaluators in an RCT. Any elegant study design cannot overcome the damage by unreliable measurement. Many methods have been developed to assess the agreement of two measurement methods. However, there is little attention to quantify how good the agreement of two measurement methods is. In this talk, similar to the type I error and the power in describing a hypothesis testing, we propose quantifying an agreement assessment using two rates: the discordance rate and the tolerance probability. This approach is demonstrated through examples.