Evaluating the reproducibility of two studies of a large number of objects: Modified Kendall Rank-Order Association Test

Tian Zheng[*] **and Shaw-Hwa Lo**

Department of Statistics, Columbia University

E-Mail: tzheng@stat.columbia.edu

Abstract: Assessing the reproducibility of research studies can be difficult, especially when the number of objects involved is large. In such situations, there is only a small set of those objects that are truly relevant to the scientific questions. For example, in microarray analysis, despite data sets containing expression levels for tens of thousands of genes, it is expected that only a small fraction of these genes are regulated by the treatment in a single experiment. In such cases, it is acknowledged that reproducibility of two studies is high only for objects with real signals. One way to assess reproducibility is to measure the associations between the two sets of data. The traditional association methods suffered from the lack of adequate power to detect the real signals, however. We present in this talk the use of a modified Kendall rank-order test of association, based on truncated ranks. Simulation results show that the proposed procedure increases the capacity to detect the real signals considerably. Applications to gene expression analysis and genetic epidemiology will be discussed.

1