Probabilistic Inference: Test and Multiple Tests

Jianchun Zhang, Jun Xie, and Chuanhai Liu*

September 30, 2009

Abstract

We view that real-world scientific inference about an assertion on unknown quantities is to produce a probability triplet (p, q, r), conditioned on available data. The probabilities p and q are *for* and *against* the truth of the assertion, whereas $r = 1 - p - q$ is the remaining probability called the probability of "don't know". Such a (p, q, r)-formulation provides a promising way of resolving long-lasting fundamental issues with hypothesis testing. With a brief review of the concept of inferential models proposed in the recent literature for producing (p, q, r) triplets for assertions, we focus on a particular inferential model for inference about an unobserved sorted uniform sample. We show how this inferential model can be used for (i) single tests, (ii) robust estimation of the empirical null distribution in the context of the local FDR method of B. Efron, and (iii) large-scale simultaneous hypothesis problems, including the many-normal-means problem and the problem of identifying significantly expressed genes in the context of microarray data analysis. These examples indicate that hypothesis testing problems can be formulated and solved in a new way of probabilistic inference.

KEY WORDS: Discrete optimization; Goodness-of-fit; Inferential model; Predictive random sets; Many-normal-means; Robust estimation; Sampling model.

*Department of Statistics, Purdue University, West Lafayette, IN 47906. E-mail: zhang10, junxie, chuanhai@purdue.edu.