Power and Sample Size Estimation in Microarray Studies

Wei-Jiun Lin
National Center for Toxicological Research,
U.S. Food and Drug Administration

Abstract

Before conducting a microarray experiment, one important issue that needs to be determined is the number of arrays required in order to have adequate power to identify differentially expressed genes. Common methods for sample size estimation are formulated as the minimum sample size necessary to achieve a specified sensitivity (proportion of detected truly differentially expressed genes) *on average* at a specified significance level. Unfortunately, the probability of detecting the specified sensitivity in such a formulation can be low. We formulate the sample size problem as the number of arrays needed to achieve a specified sensitivity with *95% probability*. A permutation method using a small pilot dataset to estimate sample size is proposed. This method accounts for correlation and effect size heterogeneity among genes.