SEQUENTIAL DETERMINATION OF SAMPLE SIZE FOR ROBUST LINEAR REGRESSION: APPLICATION TO MICROARRAY EXPERIMENTAL DESIGN

G. Durrieu[1], L. Briollais[2]

[1]*University of Bordeaux 1 and UMR CNRS, Bordeaux, France*
[2]*Samuel Lunenfeld Research Institute, Toronto, Canada*

Email: *laurent@mshri.on.ca*

Microarray experiments has become an important tool for simultaneously screening thousands of genes for changes in their patterns of expression in a single experiment. Because microarray are expensive experiments and the amount of DNA samples available is often limited, especially in human tumor samples, planning efficient designs is a fundamental aspect of the experiment. In this work, we show the interest of a sequential design that could include one or more interim analyses. The sample size necessary to detect differentially expressed genes with accurate precision is updated at each phase. The experiment stops when enough samples have been collected. Our method is based on the construction of confidence intervals for gene expression using the first component of a non parametric linear regression model. A robust approach is used to estimate the parameters of this model. With this method, we can establish with a minimal amount of data, an interval to which the gene expression value belongs with a given level of confidence. We describe an application of the method to a dose-response microarray experiment, that searches to identify genes differentially expressed in breast cancer cell-lines after stimulation with various concentrations of estrogen. In this real application, we always estimated the stopping variable for the sample size determination to be smaller than the actual sample size available to conduct the experiment. It means that we can obtain a very good accuracy for gene expression measurement without compromising the cost and size of the study. This is a fundamental result since more efficient designs for microarray experiments could be allowed by the use of a sequential approach.