

# Evaluating normalization techniques for cDNA microarray data

Júlia Maria Pavan Soler  
Fernando Henrique Ferraz Pereira da Rosa  
Department of Statistics, University of São Paulo, SP

## Abstract

Studies in genetics involving microarray experiments allow simultaneous comparison and quantification of gene expression on a large scale. There are many sources of variation which play an important role on the analysis of data coming from such experiments.

The source of variation one is usually interested in is the biological one, but in order to infer appropriate conclusions from it, the other sources of variation must first be controlled. This is particularly important when dealing with microarray experiments because the source of variation due to the technical procedures is considerably large [2].

In order to address this issue techniques which aim to control the technical variation were developed [1]. In this work, we evaluate these procedures, known as 'normalization techniques', analyzing their properties and comparing their performances.

As a result of this study we propose a new normalization technique, based on scatter plot smoothing via smoothing splines [3], which up to the moment has shown good results in controlling the technical variation. More specifically, this new method appears to be more successful in this control than the other methods currently available [5].

We also propose and apply two different methods to compare different normalization techniques and to assess which is better, based on a union/intersection metric [4].

## References

- [1] Yidong Chen, Edward R. Dougherty, and Michael L. Bittner. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Journal of Biomedical Optics*, 2(4):364-374, 1997.
- [2] G. A. Churchill. Fundamentals of experimental design for cDNA microarrays. *Nat. Gen. Sup.*, 32:490-495, 2002.
- [3] Grace Wahba and P. Craven. Smoothing noisy data with spline functions. *Numerische Mathematik*, (31):337-403, 1979.
- [4] Dennis M. Wilkinson and Bernardo A. Huberman. A method for finding communities of related genes.
- [5] Yee Hwa Yang, Sandrine Dudoit, Percy Luu, and Terence P. Speed. Normalization for cDNA microarray data. Technical Report 589, January 2001.