

**MEASUREMENT SCALE CATEGORIZATION ACCORDING TO THE MEASUREMENT ERROR
A GENE REDUCTION APPROACH TO THE MICROARRAY DATA ANALYSIS**

Yoichi M. Ito, Yasuo Ohashi

Department of Biostatistics, School of Public Health, The University of Tokyo, JAPAN

In construction of the prediction model based on microarray data, selection of the important genes for inclusion in the model must be needed because the sample size is usually limited in contrast to the huge number of genes. This is called feature selection. It is well known that the measurement error of the microarray data is prone to be large, therefore the subtle difference of gene expression cannot be detected. In this study, we propose the measurement scale categorization according to both the measurement error and the categorization error. This approach is to break continuous measurement scale into some prespecified categories and by which the expression for a gene can be classified to one of a finite expression patterns. In the feature selection, we consider these gene expression patterns instead of the original gene expressions. The number of expression patterns is small when the sample size is small, therefore this approach can reduce the total number of genes for feature selection drastically. The simulation study revealed that the power loss associated with this approach was not so large when the effect size is 0.5. We applied this approach to a clinical data for primary breast cancer.