

Zhang, Cun-Hui (Rutgers University, USA) and Jian Huang

Two-Way Semi-Linear Regression with Applications to Microarray Data

Abstract: We consider a two-way semi-linear regression model

$$y_{ij} = \beta_j' z_i + f_i(x_{ij}) + \epsilon_{ij}$$

which extends and combines the standard ANOVA and semi-linear regression models. A Gauss-Seidel algorithm provides joint estimates of β_j and f_i based on the least squares estimates of β_j and linear smoothing estimates of f_i . Methods are developed for the inference about linear functionals of $\beta \equiv (\beta_1, \dots, \beta_J)$. The performance of our methods, which hinges on the invertibility of a high-dimensional conditional information operator for the estimation of β , is investigated under mild conditions on the distribution and the dimensionality of $\{x_{ij}\}$ and $\{\epsilon_{ij}\}$.

Our results provide tools for proper normalization of cDNA microarray data to ensure meaningful down-stream analyses. The methodologies have two important features. First, it does not require pre-selection of constantly expressed genes or the assumption that the percentage of differentially expressed genes is small. Second, when used for detection of differentially expressed genes, it incorporates variations due to normalization in the assessment of uncertainty in the estimated differences in gene expressions. This reflects a basic feature of the microarray data in which self calibration in the data is required when making statistical inference.