

Variance Estimation in the Analysis of Microarray Data

Yuedong Wang

University of California, Santa Barbara, USA yuedong@pstat.ucsb.edu

Yanyuan Ma*

Texas A&M University, College Station, USA ma@stat.tamu.edu

Raymond J. Carroll

Texas A&M University, College Station, USA carroll@stat.tamu.edu

May 7, 2013

In microarrays analysis, one main problem is that conventional estimates of the variances are unreliable due to the small number of replications. It is commonly observed that the variance increases proportionally with the intensity level, leading to modeling variance as a function of mean through the constant coefficient of variation model and the quadratic variance-mean model. Because the means are unknown and estimated with few degrees of freedom, naive methods that use the sample mean in place of the true mean are biased. We propose three methods for overcoming this bias. The first two are variations of a heteroscedastic-simulation-extrapolation estimator, the third is based on semiparametric information calculations. Theory and Simulations show the power of our methods and their lack of bias compared to the naive method. The methodology is illustrated using microarray data from leukemia patients.

Key Words: Heteroscedasticity; measurement error; microarray; semiparametric methods; simulation-extrapolation; variance function estimation.