Read Paper


In microarray experiments, scientists and biologists are often interested in identifying differentially expressed genes, genes whose mean expression intensities are different across patient and control. Two sample $t$-test is a commonly practiced routine to carry out the statistical inference. However, this simple method has been found to be a less powerful procedure (Callow et al. 2000; Cui et al. 2005), primarily resulted from the conventional pooled variance estimator is unstable due to the limited number of arrays. Moreover, it is commonly observed that the variance of gene expression is not constant but inflates proportionally with the intensity. Assuming the variance being a function of mean, various parametric and nonparametric methods have been proposed in the literature (Rocke and Durbin 2001; Kamb and Ramaswami 2001).

In this paper, the authors assumed a modified classical measurement error model. The key difference lies on the fact that the modified model, taking the biological phenomenon into consideration, assumed the variance is not constant but depends on the true value.
More specifically, the measurement model has the form

$$Y_{ij} = X_i + g^{1/2}(X_i; \theta) \varepsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m$$  \hspace{1cm} (1)

where $Y_{ij}$ is the observed gene expression from the $i$th gene in the $j$th array, $X_i$ is the expected expression value of the $i$th gene and $\varepsilon_{ij}$ is independent random error with standard normal distribution and $\theta$ is a $d$-dimensional unknown parameter vector.

The authors primarily focused on estimation of the unknown parameter $\theta$ in parametric forms. Two popular parametric forms of $g(\cdot)$ were considered in this paper. The first one is called “constant coefficient of variation model”, which assumes

$$g(x; \theta) = \theta x^2, \quad \theta \geq 0$$  \hspace{1cm} (2)

and the other one is called “quadratic mean-variance model”, which assumes

$$g(x; \theta) = \alpha + \beta x^2, \quad \alpha, \beta \geq 0$$  \hspace{1cm} (3)

Two estimating procedures, least square and moment estimators are used in these models. For instance, the moment estimator of $\theta$ in (2) is given by

$$\hat{\theta}_M = \frac{n^{-1} \sum_{i=1}^{n} S_i}{n^{-1} \sum_{i=1}^{n} X_i^2}$$  \hspace{1cm} (4)

where $S_i$ as the sample variance of $Y_{i1}, \ldots, Y_{im}$. These estimators, however, require the knowledge of $X_i$. A naive approach is to replace $X_i$ by its unbiased estimator $\bar{Y}_{i+}$, the sample mean of the $i$th gene over $m$ arrays. The authors showed the naive estimators were not consistent estimators. More specifically, the parameter estimates $\theta$ in (2) and $\beta$ in (3) were underestimated, which mirrors the classical attenuation problem in measurement error model.

To overcome the limitation of naive estimators, the authors proposed three different types of estimators, simulation-extrapolation moment estimator, permutation simulation-extrapolation estimator and semiparametric estimator. Taking advantage of the replicates
of $Y_{ij}$’s, the first one is a SIMEX-type estimator (Devanarayan and Stefanski 2002). However, the first method does not satisfy the usual requirement in most measurement error model, that the measurement error is independent of the response. Forcing the independence between response and measurement error, the second method was a modification of the first one. The third method adopted a different perspective from the previous ones. By treating the unobservable variables as latent, the problem of estimating $\theta$ can be formulated in a semiparametric framework, where a set of influence functions and estimating equations were derived.

Asymptotic consistency are established for all three methods. However, the first method has its limitation that can only be useful for moment estimators, whereas the remaining two methods lead to consistent moment and least square estimators. In their simulation study, they demonstrated that the permutation SIMEX and semiparametric methods performed well in terms of correcting the bias arisen from the naive approach and providing more efficient estimators. However, the effectiveness of classical SIMEX was limited in moment-based estimators in a constant coefficient of variation model. All these findings were coherent with the theoretical results. Moreover, when data are generated from (2), moment-based estimators consistently outperformed those least square estimators.

One potential future research trajectory can work on comparing the proposed methods with alternative methods like VarMixt (Delmar et al. 2005) and Haar-Fisz (Motakis et al. 2006) methods.
REFERENCES


