Neural, Parallel, and Scientific Computations xx (200x) xx-xx

BEHRENS-FISHER'S DISTRIBUTION FOR SELECTING DIFFERENTIALLY EXPRESSED GENES

NABIN K. MANANDHAR SHRESTHA AND KANDETHODY M. RAMACHANDRAN

Department of Mathematics and Statistics
University of South Florida
Tampa, FL 33620, USA.

This work is dedicated to the distinguished Professor of Statistics C. P. Tsokos.

ABSTRACT. Microarray expression experiments allow the recording of expression levels of thousands of genes simultaneously. These experiments primarily consists of either monitoring a gene multiple times under many conditions or evaluating each gene in a single environment but in different tissue types. In the two-sample microarray expression, the variance of the expression levels in control and treatment conditions are generally different due to the nature and response of the mRNA at the different conditions. Some of the genes might have the same variance but most of the genes have both mean and variance altered. In this paper we have proposed the Bayesian version of t-test that assumes the different variances in control and treatment groups. The efficiency of new method is demonstrated by simulation results. This method is easy to implement and give the better power compared to its equal variance counterpart and the SAM.

Key Words Behrens-Fisher Distribution, Differentially expressed genes, Microarrays, SAM, Bootstrap.

1. Introduction

There is a large volume of literature behind the two sample t-test [3], [5],[10]. There are different versions and variants of these test- non-parametric and Bayesian version. In such tests, one is primarily interested on the identification of differentially expressed genes under two different conditions, so that those particular genes of interest are further studied. Most prevalent methods used in the literature for identification of differentially expressed genes are fold change [3], and the simple two-sample t- tests [1], [2], [3], [10].

The fold-change approach considers genes as differentially expressed if its average expression level changes by more than a constant factor, generally 2-fold change. But this rule does not take account of the change in the variances. The another approach for finding genes as differentially expressed is the use of t-test. The idea is simple

Received xxxx, 200x

but intuitive for the gene expression data. The reason behind this - it takes the variances of each of the genes in account together with the means. In this test, a gene is said to be expressed if $|t|$ exceeds a certain threshold depending on the confidence level selected. Since the distance between the sample means are standardized by the variances, this approach is better than the fold-change method. The gene expression data shows that there is an inherent limitation behind using the simple two-sample t-test. The variances of the genes depends on the expression level [1], [3], [10]. To identify the genes that are actually expressed, one should consider this fact. This fact was considered by previous researchers [2], [10].

There are two inherent problems in microarray experiments: First, the number of replications is very small, and second the number of genes to be tested is usually large and the test is to be repeated thousands of times. Since the small population size is very common in microarray studies, the sample estimates of variances are not appropriate for the testing. Therefore, we have to consider the Bayesian approach to approximate the variances in two conditions. To remove the second defect arising from multiple comparison, we use the False Discovery Rate (FDR) correction [6]. Applying the hypothesis test repeatedly gene by gene for several thousands of genes, there is a great chance of selecting false genes as differentially expressed, even though the significance level is set very small. For the test to be reliable, the probability of selecting true positive should be high. To control the false positive rate, we have applied the FDR correction, in which the p -values for each of the gene is compared with its corresponding threshold. A gene is, then, said to be differentially expressed if the p-value is less than the threshold.

The Behrens - Fisher problem arises when one seeks to make inferences about the means of two normal populations without assuming the variances are equal. But, in many practical situations the populations do not have same variances. Here, we briefly review the two sample t-test.

Let $\mathbf{x} = (x_1, x_2, ..., x_m)$ and $\mathbf{y} = (y_1, y_2, ..., y_n)$ be two independent samples from two normal populations with means μ_x and μ_y and equal variances $\sigma_x^2 = \sigma_y^2 = \sigma^2$ respectively. The sample mean and sample variance for these samples $\mathbf{x}$ and $\mathbf{y}$ are

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i \ , \quad s_x^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2$$

and

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j \ , \quad s_y^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \bar{y})^2$$

respectively. The population variance is estimated by the pooled sample variance

$$s^2 = \frac{(m-1)s_x^2 + (n-1)s_y^2}{m+n-2}$$

Then the sufficient statistics for μ_x, μ_y and σ^2 are $\bar{x}$, $\bar{y}$ and s^2 respectively. Furthermore, $\bar{y} - \bar{x}$ has a normal distribution with mean $\delta = \mu_y - \mu_x$ and variance $(1/m + 1/n)\sigma^2$. Then,

$$t = \frac{\delta - (\bar{y} - \bar{x})}{\sqrt{(1/m + 1/n)s^2}}$$

is distributed as student's t-statistic with $(m + n - 2)$ degrees of freedom.
Now, let the samples $\mathbf{x}$ and $\mathbf{y}$ are from normal distributions with unequal variances σ_x^2 and σ_y^2 respectively. In this case neither a pivotal statistic nor an exact confidence interval procedure exist [11]. We can take a statistic

$$t^* = \frac{\delta - (\bar{y} - \bar{x})}{\sqrt{s_x^2/m + s_y^2/n}} \sim t_{[\min(\nu_1, \nu_2)]}$$

where, $\nu_1 = m - 1, \quad \nu_2 = n - 1$

If the sample sizes m and n are large, then the both t and t^* statistics give almost the same result. In the microarray experiments the sample sizes are relatively small, thus motivating us to look for an alternative.

2. **Sampling Distribution for Non-homogeneous Variance**

2.1. **Bayesian Approach.** For the samples $\mathbf{x} = (x_i) \overset{iid}{\sim} \quad N(\mu, \sigma^2)$ and $\mathbf{y} = (y_j) \overset{iid}{\sim}$ $N(\mu + \Delta\mu, \tau^2)$. where, $i = 1, 2, ..., m$; and $j = 1, 2, ..., n$.
The density for $\mathbf{x} = (x_1, x_2, ., ..., x_m)$ can be written as ,

$$f(\mathbf{x}) = \frac{1}{\sigma^m (2\pi)^{\frac{m}{2}}} \exp\Big[\frac{-1}{2\sigma^2}\{(m-1)s_x^2 + m(\bar{x} - \mu)^2\}\Big] \tag{2.1}$$

Similarly, the density of $\mathbf{y}$ is,

$$f(\mathbf{y}) = \frac{1}{\tau^n (2\pi)^{\frac{n}{2}}} \exp\Big[\frac{-1}{2\tau^2}\{(n-1)s_y^2 + n(\bar{y} - \mu - \Delta\mu)^2\}\Big] \tag{2.2}$$

Assuming the independency of the location parameter μ and scale parameter σ^2, the joint prior for μ and σ^2 is the product

$$p(\mu, \sigma^2) = p(\mu)p(\sigma^2) \tag{2.3}$$

Similarly, the joint prior for $\mu + \Delta\mu$ and τ^2 is

$$p(\mu + \Delta\mu, \tau^2) = p(\mu + \Delta\mu)p(\tau^2) \tag{2.4}$$

Finally, the joint prior for $\mu, \sigma^2, \mu + \Delta\mu, \tau^2$ is

$$\begin{aligned} \text{prior} &= p(\mu, \sigma^2, \mu + \Delta\mu, \tau^2) \\ &= p(\mu)p(\sigma^2)p(\mu + \Delta\mu)p(\tau^2) \end{aligned} \tag{2.5}$$

Since $(\bar{x}, s_x^2)$ and $(\bar{y}, s_y^2)$ are sufficient statistics for (μ, σ^2) and $(\mu + \Delta\mu, \tau^2)$ respectively, we have the joint posterior distribution is given by the *Bayes' Rule* [8],

$$p(\mu, \Delta\mu, \sigma^2, \tau^2|\mathbf{x}, \mathbf{y}) \quad \propto \quad (\text{prior})\frac{1}{\sigma^m \tau^n} \quad \exp\Big(\frac{-C_\mu}{2\sigma^2}\Big)\exp\Big(\frac{-D_{\mu+\Delta\mu}}{2\tau^2}\Big) \tag{2.6}$$

where,

$C_\mu = (m-1)s_x^2 + m(\bar{x} - \mu)^2;$

$D_{\mu+\Delta\mu} = (n-1)s_y^2 + n(\bar{y} - (\mu + \Delta\mu))^2$

We obtain the marginal posterior density of $(\Delta\mu, \sigma^2, \tau^2)$ by integrating (2.6) with respect to μ. The marginal posterior is

$$p(\Delta\mu, \sigma^2, \tau^2|\mathbf{x}, \mathbf{y}) \propto \int_{-\infty}^{\infty} (\text{prior})\Big[\frac{1}{\sigma^m \tau^n}\exp\Big(\frac{-C_\mu}{2\sigma^2} - \frac{D_{\mu+\Delta\mu}}{2\tau^2}\Big)\Big]d\mu \tag{2.7}$$

Let us assume that the priors for μ and $\mu + \Delta\mu$ are flat priors, *i.e.* $p(\mu) = 1$, $p(\mu + \Delta\mu) = 1$ and the priors for σ^2 and τ^2 are scaled inverse-χ^2 distributions. With these priors the posterior is of the same form [3]. So, they are the conjugate priors for the normal likelihoods. *i.e.* $p(\sigma^2) = I(\sigma^2; \nu_0, \sigma_0^2)$ and $p(\tau^2) = I(\tau^2; \eta_0, \tau_0^2)$ where,

$$p(\sigma^2) \propto \sigma^{(-\nu_0+2)} exp(-\frac{1}{2\sigma^2}\nu_0\sigma_0^2) \tag{2.8}$$

$$p(\tau^2) \propto \tau^{(-\eta_0+2)} exp(-\frac{1}{2\tau^2}\eta_0\tau_0^2) \tag{2.9}$$

where, $\alpha = (\nu_0, \eta_0, \sigma_0^2, \tau_0^2)$ is the hyper-parameters that should be estimated from the data.

Hence from equations (2.7), (2.8) and (2.9), the marginal posterior density of $(\Delta_\mu, \sigma^2, \tau^2)$ is

$$\begin{aligned} p(\Delta\mu, \sigma^2, \tau^2|\mathbf{x}, \mathbf{y}) &\propto \frac{1}{\sigma^{m+\nu_0+2}} \cdot \frac{1}{\tau^{n+\eta_0+2}} \int_{-\infty}^{\infty} \exp\Big(-\frac{C_\mu + \nu_0\sigma_0^2}{2\sigma^2}\Big)\exp\Big(-\frac{D_{\mu+\Delta\mu} + \eta_0\tau_0^2}{2\tau^2}\Big)d\mu \\ &= \int_{-\infty}^{\infty} \Big[\frac{1}{\sigma^{m+\nu_0+2}}\exp\Big(-\frac{C_\mu + \nu_0\sigma_0^2}{2\sigma^2}\Big)\Big]\Big[\frac{1}{\tau^{n+\eta_0+2}}\exp\Big(-\frac{D_{\mu+\Delta\mu} + \eta_0\tau_0^2}{2\tau^2}\Big)\Big]d\mu \end{aligned}$$

The marginal posterior of Δ_μ is obtained by

$$p(\Delta\mu|\mathbf{x}, \mathbf{y}) \propto \int_{\mu=-\infty}^{\infty} \Big[\int_0^{\infty} \frac{1}{\sigma^{m+\nu_0+2}}\exp\Big(-\frac{C_\mu + \nu_0\sigma_0^2}{2\sigma^2}\Big)d\sigma^2\Big].$$

$$\Big[\int_0^\infty \frac{1}{\tau^{n+\eta_0+2}}\exp\Big(-\frac{D_{\mu+\Delta\mu}+\eta_0\tau_0^2}{2\tau^2}\Big)d\tau^2\Big]d\mu$$

$$= \int_{\mu=-\infty}^{\infty} I1.\ I2\ d\mu \tag{2.10}$$

Now,

$$I1 = \int_0^\infty \frac{1}{\sigma^{m+\nu_0+2}}\exp\Big(-\frac{A}{2\sigma^2}\Big)d\sigma^2,$$

where,

$A = C_\mu + \nu_0\sigma_0^2$

Changing the variable $u = \frac{A}{2\sigma^2}$ and after some computation, we get

$I1 = \int_\infty^0 (\frac{A}{2u})^{-(\frac{m+\nu_0+2}{2})} e^{-u}(\frac{-A}{2u^2})du$

$= A^{-\frac{(m+\nu_0)}{2}} \int_0^\infty e^{-u}u^{(\frac{m+\nu_0+2}{2})-2}du$

$= A^{-\frac{(m+\nu_0)}{2}} \int_0^\infty e^{-u}u^{\frac{(m+\nu_0)}{2}-1}du$

This being non-normalized gamma integral, the above integral is,

$I1 \propto A^{-\frac{(m+\nu_0)}{2}}$

$= (C_\mu + \nu_0\sigma_0^2)^{-\frac{(m+\nu_0)}{2}}$

$= [(m-1)s_x^2 + m(\bar{x}-\mu)^2 + \nu_0\sigma_0^2]^{-\frac{(m+\nu_0)}{2}}$

$= [1 + \frac{m(\bar{x}-\mu)^2}{(m-1)s_x^2+\nu_0\sigma_0^2}]^{-\frac{(m+\nu_0)}{2}}$

i.e.

$$I1 \propto [1 + \frac{m(\bar{x}-\mu)^2}{v_m\sigma_m^2}]^{-\frac{(m+\nu_0)}{2}} \tag{2.11}$$

where, $v_m = m + \nu_0 - 1, \qquad v_m\sigma_m^2 = (m-1)s_x^2 + \nu_0\sigma_0^2.$

Similarly, we get the another factor in (2.10) as

$$I2 \propto [1 + \frac{n(\bar{y}-\mu-\Delta_\mu)^2}{w_n\tau_n^2}]^{-\frac{(n+\eta_0)}{2}} \tag{2.12}$$

where, $w_n = n + \eta_0 - 1, \qquad w_n\tau_n^2 = (n-1)s_y^2 + \eta_0\tau_0^2.$

Substituting (2.11) and (2.12) in (2.10), we get

$$
\begin{aligned}
p(\Delta\mu|\mathbf{x},\mathbf{y}) &= k\int_{\mu=-\infty}^{\infty}\Big[1+\frac{m(\bar{x}-\mu)^2}{v_m\sigma_m^2}\Big]^{-\frac{1}{2}(m+\nu_0)}\Big[1+\frac{n(\bar{y}-(\Delta\mu+\mu))^2}{w_n\tau_n^2}\Big]^{-\frac{1}{2}(n+\eta_0)}d\mu \\
&= k\int_{\mu=-\infty}^{\infty}\Big[1+\frac{m(\bar{x}-\mu)^2}{v_m\sigma_m^2}\Big]^{-\frac{1}{2}(v_m+1)}\Big[1+\frac{n(\bar{y}-(\Delta\mu+\mu))^2}{w_n\tau_n^2}\Big]^{-\frac{1}{2}(w_n+1)}d\mu \qquad (2.13)
\end{aligned}
$$

This is the pdf of the Behrens-Fisher distribution, where k is given by,

$$
k=\Big[Beta\Big(\frac{v_m}{2},\frac{1}{2}\Big)Beta\Big(\frac{w_n}{2},\frac{1}{2}\Big)\sqrt{v_m w_n}\Big]^{-1}
$$

Now, we can apply the Behrens-Fisher distribution for testing the hypothesis regarding the two population means, using two samples drawn from the population with different means and different variances.

2.2. **Test Statistic.** Let us define the statistic, called the BF-statistic

$$
\begin{aligned}
B &= \frac{\Delta\mu-(\bar{y}-\bar{x})}{(\frac{\sigma_m^2}{m}+\frac{\tau_n^2}{n})^{\frac{1}{2}}} \\
&= \frac{(\mu+\Delta\mu)-\bar{y}}{\tau_n/\sqrt{n}}cos\theta-\frac{(\mu-\bar{x})}{\sigma_m/\sqrt{m}}sin\theta \\
&= B_y cos\theta - B_x sin\theta
\end{aligned}
$$

where,

$$
tan\theta=\frac{\sigma_m/\sqrt{m}}{\tau_n/\sqrt{n}},\qquad 0\le\theta\le\frac{\pi}{2}
$$

$$
B_x=\frac{(\mu-\bar{x})}{\sigma_m/\sqrt{m}},\qquad B_y=\frac{(\mu+\Delta\mu)-\bar{y}}{\tau_n/\sqrt{n}}
$$

and, B_x and B_y are independently distributed as t- statistics, $t_{(v_m)}$ and $t_{(w_n)}$ respectively. Hence, under the sampling distribution, $p(\mathbf{x},\mathbf{y}|\mu,\sigma^2,\tau^2)$, the statistic B is distributed as the Behrens- Fisher distribution with v_m and w_n degrees of freedom. That is,

$$
B\sim BF(v_m,w_n,\theta)
$$

with pdf

(2.14)

$$
f(\beta|\mu,\sigma^2,\tau^2)=k\int_{-\infty}^{\infty}\Big[1+\frac{(\alpha cos\theta-\beta sin\theta)^2}{v_m}\Big]^{-\frac{v_m+1}{2}}\Big[1+\frac{(\alpha sin\theta+\beta cos\theta)^2}{w_n}\Big]^{-\frac{w_n+1}{2}}d\alpha,
$$

where,

$$\alpha = B_y sin\theta + B_x cos\theta, \quad \beta = B_y cos\theta - B_x sin\theta$$

which is same as (2.13).
Hence, we have proved the following :

Theorem 2.1. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be two independent samples with sample sizes* m *and* n *respectively from normal distributions with different means* μ *and* $(\mu + \Delta_\mu)$ *and variances* σ^2 *and* τ^2*. If the priors* ν_0 *and* η_0 *for means are flat priors and priors for variances* σ_0^2 *and* τ_0^2 *are scaled inverse* χ^2*- distributions, the posterior distribution of* Δ_μ *is the Behrens-Fisher distribution with* $v_m = m + \nu_0 - 1$*,* $w_n = n + \eta_0 - 1$ *degrees of freedom.*

Due to the complexity of the pdf of the BF-distribution as given in (2.14), it is very hard to compute the corresponding probabilities, especially due to the possibility of the fractional degrees of freedom. In addition, there are no uniformly most powerful unbiased tests for all sample sizes for the BF- problem [13]. Because of this, there are various types of approximations available in the literature [9],[13]. Due to the simplicity of application as well as availability of R-code (*www.r-project.org*) for computing t- values even for fractional degrees of freedom, we use Patil's approximation [9] in this work as follows.

Let

$f_1 = \left(\frac{w_n}{w_n-2}\right) cos^2\theta + \left(\frac{v_m}{v_m-2}\right) sin^2\theta$

$f_2 = \frac{w_n^2}{(w_n-2)^2(w_n-4)} cos^4\theta + \frac{v_m^2}{(v_m-2)^2(v_m-4)} sin^4\theta$

$a^2 = \frac{(b-2)}{b} f_1$

$b = 4 + \frac{f_1^2}{f_2}$

$cos^2\theta = \frac{\frac{\tau_n^2}{n}}{\left(\frac{\tau_n^2}{n} + \frac{\sigma_m^2}{m}\right)}, \qquad sin^2\theta = 1 - cos^2\theta.$

Then, the statistic

$$\frac{B}{a} \sim t_{(b)}. \tag{2.15}$$

That is, B has approximately t-distribution with b degrees of freedom ($b \geq 1$, b may not be an integer [7]) and scale parameter a. This statistic B can also be denoted as $B \sim t(0, a^2, b)$. It was noted [9] that the formula (2.15) is valid only for $v_m, w_n \geq 5$ and works quite well for $v_m, w_n \geq 7$.

The corresponding Bayesian test statistic [5] for the equal variance case is

$$E = \frac{\Big(\Delta_\mu - (\bar{y} - \bar{x})\Big)}{\upsilon\sqrt{\frac{1}{m} + \frac{1}{n}}} \tag{2.16}$$

which is distributed as t distribution with δ degrees of freedom, where,

$$\begin{cases} \delta = m + n + \nu_0 - 2 \\ \delta\upsilon^2 = \nu_0\sigma_0^2 + (m-1)s_x^2 + (n-1)s_y^2 \end{cases} \tag{2.17}$$

3. Calculation of Prior d.f. and Prior Variance

There are many possible ways one can choose p and q (and so prior variances and prior means) in the above expressions. In each of the methods, we use the sample variances of only particular genes in both control and treatment conditions, and apply equations (3.1),(3.2),(3.3) and (3.4). In this work we choose p and q in the following manners and compare the effect of each of these in terms of FDR.

Method 1: Window Method

In this method, we calculate the prior degrees of freedom and prior variances by taking those genes that are similar in variances to that of the gene of interest both in control and treatment condition within pre-chosen window size. To calculate the prior variance of a gene g in control condition, we calculate its variance in control condition. Then we calculate the variance of all other genes and take only those p genes whose variances are close to that of gene g. Now the prior variance for gene g is the mean of variances of these p genes in control condition. Similarly the prior variance for gene g in treatment condition is calculated by taking q genes with similar variances to that of gene g.

For each gene g, the prior variances and prior means will be different. We have dropped the subscript g from them. For p genes with similar variances and each having m replicates in control condition, the prior degree of freedom for the variance can be calculated as,

$$\hat{\nu_0} = p(m-1) \tag{3.1}$$

Similarly, for q genes with similar variances and each having n replicates in treatment condition, the prior degrees of freedom for the variance is given by

$$\hat{\eta_0} = q(n-1) \tag{3.2}$$

The prior variances for control and treatment conditions are calculated as sample variance of similar genes, which are the means of variances of similar genes in these two conditions respectively.

$$\hat{\sigma_0}^2 = \frac{1}{\nu_0}\sum_{k=1}^{p}\sum_{i=1}^{m}(x_{k,i} - \bar{x}_k)^2 \tag{3.3}$$

$$\hat{\tau_0}^2 = \frac{1}{\eta_0}\sum_{k=1}^{q}\sum_{j=1}^{n}(y_{k,j} - \bar{y}_k)^2 \tag{3.4}$$

where,
$\bar{x}_k$ is the mean response of gene k in the control condition
$\bar{y}_k$ is the mean response of gene k in the treatment condition
$x_{k,i}$ is the response i of gene k in the control condition
$y_{k,i}$ is the response i of gene k in the treatment condition

Method 2: Similar Variance Method

Another way could be to choose the absolute difference of variances in both control and treatment group using some cut-off values. For a gene g, let s^2_{cg} and s^2_{tg} be the sample variances in control and treatment conditions respectively. Then choose p_g= number of genes j such that $|s^2_{cg} - s^2_{cj}| \leq k$, where $k = |min\{s^2_{cg} -$ 95% Lower χ^2 CL of $s^2_{cg}, s^2_{cg} -$ 95% Upper χ^2 CL of $s^2_{cg}\}|$ in the control condition. Similarly, choose q_g = number of genes j such that $|s^2_{tg} - s^2_{tj}| \leq k$, where k $= |min\{s^2_{ct} -$ 95% Lower χ^2 CL of $s^2_{ct}, s^2_{ct} -$ 95% Upper χ^2 CL of $s^2_{ct}\}|$ in the treatment condition. So, in this method the values of p and q varies according to each gene. For simplicity, we omit the subscript g in the notation.

Method 3: Bootstrapping Method

Instead of taking those genes which have similar variances within a window or as in similar variance method, it is more practical to take bootstrap samples to calculate prior variances for the actual data. Because, in the window method the researcher should consider all possible window size and the determination of it may be time consuming. On the other hand, the similar variance method selects huge number of genes and so the computation time is exponentially increased. In bootstrapping method we take the $p = q = 1\%$ bootstrap samples from the pool of all genes together with the gene of interest from the control and treatment conditions respectively to determine the prior d.f. and variance for a gene. Then for p genes each having m replicates in the control condition, the prior degrees of freedom is given by $\nu_0 = p(m-1)$ and the

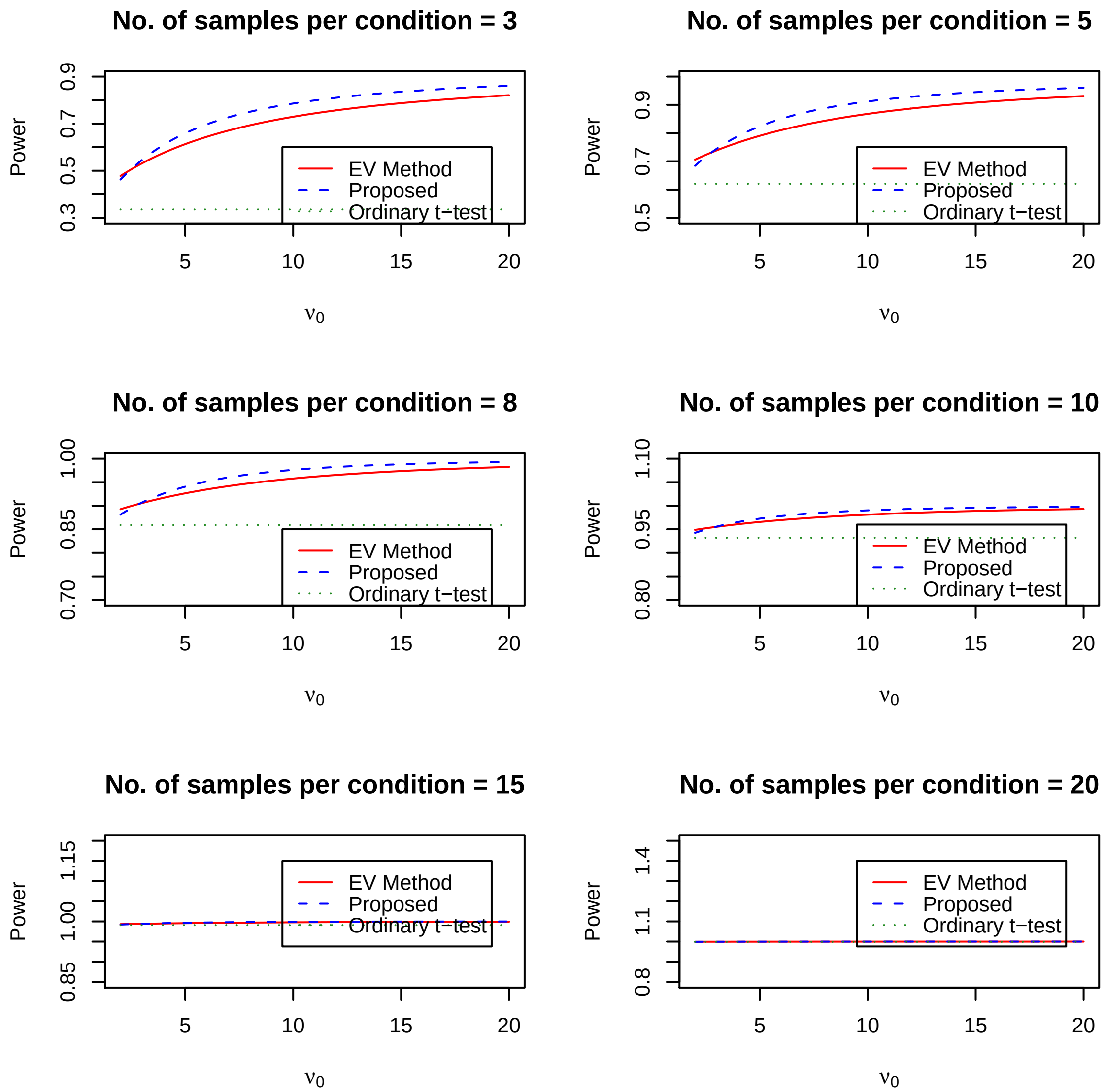

FIGURE 1. Power comparison for different sample size and priors using Method 1

prior variance σ_0^2 is estimated by the variance of the p bootstrap samples in the control condition. The prior degrees of freedom in the treatment condition $\eta_0 = q(n-1)$ and variance τ_0^2 are calculated similarly from the treatment condition.

4. **Simulation and Result**

To implement the theory, we have simulated 10,000 genes from normal distributions having mean $\mu_1 = 0$ variance $\sigma_1^2 = 1.5$ in the normal conditions and mean $\mu_2 =$ 0.2 and variance $\sigma_2^2 = 1.2$ in the treatment conditions. Each gene was replicated 10 times in each of the control and treatment conditions. Without loss of generality, we have set 2 percent of the genes as differentially expressed, beginning from the first quarter onwards. The differentially expressed genes in the treatment condition was generated by the normal distribution with mean of 5 and variance 2.

In the simplest case, when we choose $\nu_0 = \eta_0 = 0$, then the BF statistic reduces to the two sample t-test. The parameters ν_0 and η_0 represents the degree of confidence in the background variances σ_0^2 and τ_0^2 versus the empirical variances of control and treatment respectively. We have chosen the values of these two variances in the wide range beginning from 0. The values of ν_0 and η_0 are increased. In each of the calculation of variances of similar genes in the two conditions, we have taken the window size equal to the the replicates in each condition, *i.e.*, taking $p/2$ genes in control group immediately above and below the gene under considerations. The mean is taken of those ordered variances corresponding to the gene of interest. Since the Behrens-Fisher distribution does not give the good result unless v_m and w_n is greater than 5 and gives very good result when it exceeds 7 [9], we have chosen the values of the prior degrees of freedom such that v_m and w_n exceeds 7. We have chosen those genes as differentially expressed whose p-values are smaller than ranked values of the statistic given by FDR criterion. Furthermore we have chosen the level of significance $\alpha = 0.05$. The number of genes that are found differentially expressed in our method and in the equal variance cases are compared in the table. We have run the data 5 times and averaged the number of genes selected as differentially expressed in each time.

In analyzing the fold change method with the BF method, we have found that all the genes associated with the large - fold change are not necessarily statistically significant in the Bayesian BF method.

5. **Power Comparisons**

In the microarray data, the number of microarrays are few generally 4 or 5 because of the high cost of production. So, we have to find the method which describes the high power even for the small number of samples. Here, we have compared the power obtained by our method with that of equal variance test of Fox-Dimmic [5]. The application of treatment not only affects the averages but it affects the expression measurements, so it affects variances as well. It is a common experience in statistical analysis that as the number of samples increases, the power of the test also increases. But in our case, the power depends on the priors as well. Even for the small values of priors, the power of the proposed method seems optimum than the equal variance method.

We see from the graph of power (Figure 1) and power comparison table (Table 1) that, at small values of ν_0 the power obtained by the Equal variance test is preferred

Table 1. Table of Power Comparison

$m = n$	$\nu_0 = \eta_0$	power BF	Power EV
2	16	0.5617153	0.4532222
3	16	0.7179875	0.6035196
4	16	0.820609	0.7177347
5	16	0.8865956	0.8020212
2	8	0.4969753	0.4067113
3	8	0.6431226	0.5496432
4	8	0.7496756	0.6657278
5	8	0.8262618	0.7566512
2	4	0.3985896	0.3444951
3	4	0.5350779	0.4867308
4	4	0.6495701	0.6098946
5	4	0.8395813	0.7105832
2	2	0.3057716	0.2803869
3	2	0.4221825	0.429857
4	2	0.545814	0.5631983
5	2	0.6547838	0.6739527

than our method. But eventually, our proposed method seems to have better result. In small values of m and ν_0, the corresponding values for $v_m = m + \nu_0 - 1$ are smaller and it does not give appropriate values unless it exceeds 7. Similarly, for the small values of w_n. So, we have to choose the values of ν_0 or η_0 and m or n such that the values of v_m and w_n is at least 5. Holding $\nu_0 = \eta_0 = 4$ we see the difference of power in EV method and the new method increases significantly as the sample size increase from 2 to 5. The difference in the power between this proposed method and the EV method is less pronounced for the small sample sizes and small values of the prior degrees of freedom. But it is more distinct as the prior degree of freedom increases. We have chosen the prior variances σ_0^2=0.7 and τ_0^2=0.3.

In this simulated study, we have used the proportion of genes that are actually differentially expressed and detected by our method as the criterion for comparing the proposed method. As this method is not suitable if there is small values of m or n and p or q that makes v_m and w_n small, which is seen clearly from the table (Table 2) that for $m = n = 2$ and choosing any values of the priors ν_0 and τ_0 only selects 10% of the genes that are actually expressed, although it reports large amount of genes as differentially expressed. We have found that the genes marked as differentially expressed by the fold-change (fold change = 2) method is almost 32% of the total number of genes included in the simulated study. This result is highly unacceptable as we have assumed only 2% of the genes as differentially expressed. On the another

Replications		DE genes				Common in actual		Proportion	
$m = n$	$p = q$	FC	t-test	EV	BF	BF	EV	BF	EV
2	2	3421	1	157.5	239	21.5	23	0.09	0.15
2	4	3430.5	0.5	544.5	263	35.5	61	0.13	0.11
2	8	3439.67	0	1145.67	733.33	85.33	119	0.12	0.1
2	20	3459	0	1693	1051	110	138	0.1	0.08
2	100	3374.5	0	2126.5	1187	121	154	0.1	0.07
3	2	3347	0.33	134.33	45	20	57.67	0.44	0.43
3	4	3392	0.33	344	226	79	111	0.35	0.32
3	8	3314.33	0	569.33	357.33	107.67	133.67	0.3	0.23
3	20	3371	0	832	479.67	133	158	0.28	0.19
3	100	3320.5	0.5	1032	568	135	165	0.24	0.16
4	2	3293.33	0.5	165	109.67	74.67	104	0.68	0.63
4	4	3317.4	0.33	279.4	204.4	114.8	141.4	0.56	0.51
4	8	3353.33	1.2	409	282	131	156.67	0.46	0.38
4	20	3378.5	0.67	568	364	151.5	175.5	0.42	0.31
4	100	3311	1	648.5	389	165	184	0.42	0.28
5	2	3332.33	1	208	167.33	131.33	158	0.78	0.76
5	4	3273	36.33	260	227.33	153.67	166.67	0.68	0.64
5	8	3267	23	320.5	240.5	155	176	0.64	0.55
5	20	3295	36.5	424.33	307	168.67	186.67	0.55	0.44
5	100	3251.67	51.67	453.33	326.67	174	184.33	0.53	0.41
10	2	3169	28.67	226.5	225.5	200	200	0.89	0.88
10	4	3133.67	215	245.33	242	198	199.33	0.82	0.81
10	8	3145.33	212.33	264.33	263.67	199.67	199.67	0.76	0.76
10	20	3168.33	217	269.67	264.67	200	200	0.76	0.74
10	100	3094	211.5	266.5	268	199	199.5	0.74	0.75

TABLE 2. *Comparison of BF test with other tests based on the proportion of actually DE genes selected in Window Method .*

extreme, we ran two-sample t-test and found that it selected few genes as differentially expressed as the sample sizes was increased to 4 and the window size was 20. After that it selected many genes and the result was getting better when we took sample size of 10. In this case, it selected almost the same genes as our method. So, this method failed for the small sample size, because it could not select the genes that are actually differentially expressed.

Replications		DE genes				Common in actual				Proportion			
m	n	BF	EV	SAM	t-test	BF	EV	SAM	t-test	BF	EV	SAM	t-test
3	3	22	24	11	15	15	15	11	7	0.68	0.63	1	0.47
4	4	24	23	20	26	19	19	19	3	0.79	0.83	0.95	0.12
5	5	22	22	22	29	20	20	20	19	0.91	0.91	0.91	0.66
6	6	23	21	20	33	20	19	19	19	0.87	0.9	0.95	0.58
10	10	25	23	24	36	20	19	20	20	0.8	0.83	0.83	0.56
12	12	23	26	23	29	20	20	20	20	0.87	0.77	0.87	0.69
15	15	20	20	51	34	20	20	20	20	1	1	0.39	0.59
3	4	29	28	16	30	18	18	16	12	0.62	0.64	1	0.4
4	6	24	25	24	27	20	20	20	17	0.83	0.8	0.83	0.63
10	15	26	26	34	37	20	20	20	20	0.77	0.77	0.59	0.54

TABLE 3. *Comparision of BF test with other tests based on the proportion of actually DE genes selected in Similar Variance Method .*

6. **Comparison of Window Method with Other Methods**

All three tests, except fold-change method, gave almost the same conclusion as our proposed *similar variance* method. Initially, when the sample size and window size both are small, $m = 2$ and $p = 2$, then the method of equal variance test seemed better in terms of proportion of actual genes selected. This is natural because our method does not give the better result for small values of v_m or w_n. However, as we have greater values of sample size and window size, our method excels the equal variance Bayesian test counterpart. We have seen from the result that almost all of the actually DE genes were selected by this proposed method when $m = n = 10$ and $p = q = 8$. But, the proportion of genes selected compared to the actual set of DE genes is just 75.7%.

The maximum proportion of genes selected is 88.7% when $m = n = 10$ and $p = q = 2$. In this case all of the genes actually DE are selected. Our method and equal variance test selected almost the same number of genes on average. Unless $p = q = 8$ and for $m = n = 2$, both the equal variance and unequal variance method selected enormous number of genes, most of them bogus. The proportion of true genes is very low, about 11%. When $m = n = 3$ and $p = q = 2$, our method selected 45 genes of them 20 are actually DE, but equal variance method selected about 134 genes, of them about 58 are actually DE. Hence the proportion of true genes are 0.44 and 0.42 respectively.
Taking the sample sizes constant, we found that,the number of genes selected by all three tests by Method 1 is proportional to the window size. But, most of them are

Replicates	Common in Actual DE				Genes Selected as DE				Proportion of Genes Selected			
m, n	BF	EV	SAM	t-test	BF	EV	SAM	t-test	BF	EV	SAM	t-test
2, 2	73	63	29	15	75	63	29	115	0.973	1	1	0.13
3, 3	144	139	136	79	147	142	141	183	0.98	0.979	0.965	0.432
4, 4	179	184	175	128	182	190	179	254	0.984	0.968	0.978	0.504
5, 5	195	193	191	159	200	197	206	278	0.975	0.98	0.927	0.572
6, 6	197	198	198	188	203	203	231	325	0.97	0.975	0.857	0.578
8, 8	199	199	199	198	212	209	270	345	0.939	0.952	0.737	0.574
10, 10	200	200	200	200	209	206	332	342	0.957	0.971	0.602	0.585
12, 12	200	200	200	200	207	207	246	337	0.966	0.966	0.813	0.593
4, 3	168	170	151	94	173	173	153	230	0.971	0.983	0.987	0.409
4, 5	186	186	183	144	189	192	190	285	0.984	0.969	0.963	0.505
5, 6	198	198	198	162	201	201	215	323	0.985	0.985	0.921	0.502
6, 8	198	199	198	196	207	207	222	325	0.957	0.961	0.892	0.603

TABLE 4. *Comparision of BF test with other tests based on the proportion of actually DE genes selected in Bootstrap Method .*

bogus, *i.e*, genes seem to be differentially expressed but are not really. Even though the window size is small, the proportion of genes selected by these methods increased as window size decreased. This means that for a fixed sample size, small values of the priors ν_0 and τ_0 are preferred.

To see the performance of method 2, we have simulated 1000 genes as in the other two methods. Because it selects huge number of genes with similar variance according to our cut-of criterion, the computation is time consuming (which may not be good for large number of genes and low memory computer). This method compares with the equal variance and SAM as seen from the proportion and the number of truly differentially expressed genes selected. In this case we have introduced only 20 genes as differentially expressed. All three tests- BF, EV and the SAM were quite competitive in this method. But, EV and BF are relatively more competitive. We see that SAM selects very few genes when the sample sizes are small. But as the sample size was increased to 15, it selected more genes. On the other hand, both BF and the EV method selected almost the similar number of genes even though the sample size was increased. The result is shown in Table 4.

Table 4. shows the number of gene selected by the Method 3. We have compared the actual number of genes selected by different methods including SAM. We see that the genes selected by our method is comparable to EV test and SAM. We notice that SAM selects small number of genes relatively in the small sample sizes. Although our BF method chooses more genes, it selects genes that are actually expressed. The proportion of actually DE genes selected by BF method is comparable with the EV

	Proportion of DE genes					FDR					FNR				
m=n	3	4	5	6	10	3	4	5	6	10	3	4	5	6	10
I.	0.35	0.56	0.67	0.74	0.82	0.65	0.44	0.33	0.26	0.18	0.61	0.43	0.24	0.13	0.01
II.	0.68	0.79	0.91	0.87	0.8	0.32	0.21	0.09	0.13	0.2	0.25	0.05	0	0	0
III.	0.98	0.98	0.98	0.97	0.96	0.02	0.02	0.03	0.03	0.04	0.28	0.11	0.03	0.02	0

TABLE 5. *Comparison of three methods (I = Window Method, II = Similar Variance Method, III = Bootstrap Method) according to Proportion of truly De genes selected, FDR and FNR .*

test, and performs better than the SAM method. The number of genes selected by different methods increases as the sample size increases. But the EV test and BF test select more actually expressed genes than the SAM and t-test. This means the the false positive rate of these two tests are smaller than that of SAM and t-test. The false positive rate is defined as probability of false positives among the positive findings. False Positive Rates for the samples having $m = n = 5$ are: (BF, 0.025), (EV, 0.020), (SAM, 0.072) and (t-test, 0.42). This means that 2.5% genes selected by BF method are bogus whereas 7.2% genes selected by the SAM method are bogus. The false discovery rate is defined as the expected proportion of false positives among the positive findings. So, our method (Method 2.) seems better than SAM and t-test while selecting the truly differentially expressed genes.

7. **Comparison of Proposed THREE Methods**

The above table (Table 5.) shows the comparison of aforementioned three methods. Here we have calculated the false discovery rate (FDR), which is the expected proportion of false positives among the positive findings; Proportion of DE genes; and False Negative rate (FNR), which is the expected proportion of true negatives among the truly DE genes for each sample and compared them. It has been found that the proportion of truly DE genes in the Method III is higher than in the other two methods, taking the same sample sizes in all three methods. This means that bootstrap method selects more truly DE genes than other two method. Similarly, the FDR is also small in bootstrap method than other two methods. And the FNR is smaller in method III than in method I and comparable to that of Method II. Hence it seems that the bootstrap method is the best method in selecting the truly DE genes than window and similar variance method in our simulated data.

8. **Conclusions**

We have proposed three new methods to get the differentially expressed genes if the variances are different in samples. In all of our methods, our test performed better than the equal variance (EV) test (see Table 2) and SAM (see Table 3 and Table 4). Furthermore, we have compared the three methods based on the proportion of truly differentially expressed genes selected, FDR , and FNR criterion (see Table 5) and found that the bootstrap method gave the best result among three proposed methods. We have implemented our theory in R statistical software, *http://www.r-project.org.*

REFERENCES

[1] Sartor *et. al.* : Intensity-based hierarchical Bayes method improves testing for differentially expressed genes in microarray experiments ; *BMC Bioinformatics*, Vol. **7**, 2006.

[2] Gottardo *et. al.*: Statistical analysis of Microarray Data: A Bayesian approach; *Biostatistics*, Vol. **1**, 2002.

[3] Baldi and Long : A Bayesian Framework for the Analysis of Microarray Expression Data: Regularized t-test and Statistical Inferences of Gene Changes; *Bioinformatics*, Vol. **17**, 2001.

[4] Pawitan *et. al.*: False Discovery Rate, Sensitivity and Sample Size for Microarray Studies; *Bioinformatics*, Vol. **21**, 2005.

[5] Fox and Dimmic: A two-sample t-test for Microarray data; BMC *Bioinformatics*, Vol. **7**, 2006.

[6] Aidong Zhang: Advanced Analysis of Gene Expression Microarray Data, *World Scientific*, 2006.

[7] Box and Tiao : Bayesian Inference in Statistical Analysis ; *Addison-Wesley*, 1973.

[8] Gelman, Carlin, Stern and Rubin: Bayesian Data Analysis; 2nd edition, *Chapman & Hall*, 2004.

[9] Patil, V. H. : Approximation to the Behrens-Fisher Distributions; *Biometrika*, Vol. **1**, 1965.

[10] Tusher, Tibshirani and Chu: Significance Analysis of Microarrays applied to the Ionizing Radiation Response ; *PNAS*, Vol. **98**, 2001.

[11] Kim and Cohen: On the Behrens-Fisher Problem : A Review ; *Journal of Educational and Behavioral Statistics*, Vol. **23**, 1998.

[12] Schwender, H.: Identifying Differentially Expressed Genes with siggenes; *A Bioconductor Package* - Version **2.1**.

[13] Best and Rayner: Welch's Approximate Solution for the Behrens Fisher Problem ; *Technometrics*, Vol. **29 (II)**, 1987.