Testing the Mean Vector of a Population in High Dimension

Edgard M. Maboudou-Tchao* Ivair Silva † Jeremiah Perez‡

Abstract

Traditional multivariate tests, Hotelling's test or Wilk test, are designed for a test of the mean vector under the assumption that the number of observations is larger than the number of variables. For high dimensional data, where the number of features is nearly as large as or larger than the number of observations, existing tests do not provide a satisfactory solution because the estimated covariance matrix is singular. In this paper, we consider a test for the mean vector of independent and identically distributed multivariate normal random vectors where the dimension is larger than or equal to the number of observations. To solve this problem, we propose a modified Hotelling statistic. Simulation results show that the proposed test is superior to other tests available in the literature. However, since we do not know the theoretical distribution of this modified statistic, Monte Carlo methods were used to reach a decision. Instead of using the conventional Monte Carlo methods, which perform a fixed-number of simulations, we suggest using the sequential Monte Carlo test in order to decrease the number of simulations needed to reach a decision. Simulation results show that the sequential Monte Carlo test is always preferable to a fixed-sample test, especially when using computationally intensive statistical methods.

Key Words: Covariance matrix; slicing; flip-flop algorithm; sequential Monte Carlo

1. Introduction

The area of high-dimensional statistics deals with estimation in the "large P, small N" setting, where P is the number of variables and N the number of observations. It is rare to find a dataset large enough to compute a non-singular covariance matrix. Advances in computing have made high-dimensional data analysis possible in a number of important applications, including gene arrays, climate studies, spectroscopy, functional magnetic resonance imaging, and data mining. For this setting, traditional statistical techniques based on small or medium sample sizes may not be applicable because of the 'curse of dimensionality'. It is well known that the empirical covariance matrix for samples of size N from a P-variate Gaussian distribution is not a good estimator of the population covariance matrix if P is larger than N. Such high-dimensional scaling can lead to dramatic breakdowns in many classical procedures. In the absence of additional model assumptions, it is often impossible to obtain consistent procedures when $N < P$. The methods that deal with high dimensional data sets are usually based on a set of regularizing assumptions to reduce the complexity of the problem.

Assume that we observe a set of N vectors $\mathbf{x}_i, i = 1, 2, \ldots, N$, where each vector, $\mathbf{x}_i$ is P dimensional. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ be independent and identically distributed random vectors with mean vector $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. Suppose we would like to determine whether the population mean $\boldsymbol{\mu}$ is significantly different from a hypothesized mean vector $\boldsymbol{\mu}_0$. Without loss of generalities, we will assume that $\boldsymbol{\mu}_0 = \mathbf{0}$. This problem can be formulated as the following hypothesis test:

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \quad vs \quad H_a : \boldsymbol{\mu} \neq \mathbf{0}$$

*Department of Statistics, University of Central Florida, Orlando, FL

†Departamento de Estatistica, Universidade Federal de Minais Gerais, Belo Horizonte, Brazil

‡Department of Statistics, University of Central Florida, Orlando, FL

2. Covariance estimation for data with fewer observations than the dimension

In the general case, assume that we observe a set of N vectors $\mathbf{x}_i$, $i = 1, 2, \ldots, N$, where each vector, $\mathbf{x}_i$ is P dimensional. Without loss of generality, assume that $\mathbf{x}_i$ has zero mean. If the vectors $\mathbf{x}_i$ are identically distributed, then the sample covariance matrix is given by

$$\mathbf{S} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i'$$

and $\mathbf{S}$ is an unbiased estimator of the true covariance matrix $\mathbf{\Sigma} = E(\mathbf{S})$. While $\mathbf{S}$ is an unbiased estimate of the true covariance matrix, it is also singular when $N < P$. In practice, N may be much smaller than P and so most of the eigenvalues of $\mathbf{\Sigma}$ are incorrectly estimated as zero. Several methods have been proposed to regularize the estimate of $\mathbf{\Sigma}$ so that it is not singular. Shrinkage estimators are a class of estimators which regularize the covariance matrix by shrinking it toward some target structure. They generally have the form $\hat{\mathbf{\Sigma}} = \alpha\mathbf{D} + (1-\alpha)\mathbf{S}$ where $\mathbf{D}$ is some positive definite matrix. Some popular choices for $\mathbf{D}$ are the identity matrix or its scaled version. The shrinkage estimate α is estimated by cross-validation or bootstrap methods. The concept of robust estimation of an inverse covariance matrix was first introduced by Dempster, 1972 who suggested that the number of parameters to be estimated be reduced by setting some elements of the precision matrix or inverse covariance matrix $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$ to zero. A number of methods have been proposed for regularizing the estimate by making either the covariance or its inverse sparse. In the absence of model assumptions when $P > N$, an active line of statistical research is based on imposing various restrictions on the model – for instance, sparsity. In this article, we will use a technique called 'slicing' that is suitable for obtaining nonsingular estimates of the covariance matrix of high dimensional data in the "large P, small N" setting. Slicing was first introduced in Akdemir, 2011.

2.1 Matrix-Variate Normal Distribution

The matrix variate random variable with Kronecker delta covariance structure has been studied intensively in Gupta and Nagar (2000) and by many others. Several authors have used this kind of model for analyzing matrix variate data. Naik and Rao (2001) used the structure $\mathbf{\Sigma} = \mathbf{\Omega} \otimes \mathbf{\Psi}$ for the analysis of the data using a MANOVA model, Roy and Khattree (2003) used the same structure in discriminant analysis of repeated measures data. Lu and Zimmerman (2005) have also considered this structure in their work. Krzysko and Skorzybut (2009) have also used this structure to establish discriminant analysis of multivariate repeated measures data. In this section, we will review some useful properties of this distribution. Here, the random matrix is the fundamental element instead of the random vector. Our discussion of this distribution follows from Gupta and Nagar (2000).

Definition 2.1. *The "vec" operator transforms a $p \times q$ matrix into a vector of length pq.*

Definition 2.2. *(unstructured) The random matrix $\mathbf{X}(p \times q)$ is said to have a matrix variate normal distribution with mean matrix $\mathbf{M}(p \times q)$ and covariance matrix $\mathbf{\Sigma}$ where $\mathbf{\Sigma}$ is a $pq \times pq$ positive definite matrix, if $vec(\mathbf{X}') \sim N_{p\times q}(vec(\mathbf{M}'), \mathbf{\Sigma})$*

Definition 2.3. *(structured) The random matrix $\mathbf{X}(p\times q)$ is said to have a matrix variate normal distribution with mean matrix $\mathbf{M}(p \times q)$ and Kronecker delta structured covariance matrix $\mathbf{\Omega} \otimes \mathbf{\Psi}$ where $\mathbf{\Omega}$ is a $p \times p$ positive definite matrix, $\mathbf{\Psi}$ is a $q \times q$ positive definite matrix, if $vec(\mathbf{X}') \sim N_{p\times q}(vec(\mathbf{M}'), \mathbf{\Omega} \otimes \mathbf{\Psi})$*

The matrix variate normal distribution with Kronecker delta covariance structure is denoted by $\mathbf{X} \sim N_{p\times q}(\mathbf{M}, \mathbf{\Omega}, \mathbf{\Psi})$. $\mathbf{M}$ is the mean matrix, $\mathbf{\Omega}$ is sometime called the row covariance matrix, and $\mathbf{\Psi}$ is known as the column covariance matrix. The density, $f(\mathbf{M}, \mathbf{\Omega}, \mathbf{\Psi}))$, of this distribution is

$$f(\mathbf{M}, \mathbf{\Omega}, \mathbf{\Psi}) = (2\pi)^{-\frac{pq}{2}} |\mathbf{\Omega}|^{-\frac{q}{2}} |\mathbf{\Psi}|^{-\frac{p}{2}} etr\left(-\frac{1}{2}\mathbf{\Omega}^{-1}(\mathbf{X}-\mathbf{M})\mathbf{\Psi}^{-1}(\mathbf{X}-\mathbf{M})'\right), \tag{1}$$

where etr is the exponential of the trace function.
Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$ be iid $N_{p\times q}(\mathbf{M}, \mathbf{\Omega}, \mathbf{\Psi})$, $N > max(p, q)$, then the likelihood of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$ is

$$L(\mathbf{X}, \mathbf{M}, \mathbf{\Omega}, \mathbf{\Psi}) = (2\pi)^{-\frac{pqN}{2}} |\mathbf{\Omega}|^{-\frac{qN}{2}} |\mathbf{\Psi}|^{-\frac{pN}{2}} etr\left(-\frac{1}{2}\sum_{i=1}^{N}\mathbf{\Omega}^{-1}(\mathbf{X}_i-\mathbf{M})\mathbf{\Psi}^{-1}(\mathbf{X}_i-\mathbf{M})'\right) \tag{2}$$

The Maximum likelihood estimator (MLE) of $\mathbf{M}$ is $\bar{\mathbf{X}}$ where $\bar{\mathbf{X}} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{X}_j$. Assume that ψ_{qq} = 1 and let $\mathbf{X}_{ic} = \mathbf{X}_i - \bar{\mathbf{X}}$, the MLE of $\mathbf{\Psi}$ is

$$\hat{\mathbf{\Psi}} = \frac{1}{Np}\sum_{i=1}^{N}\mathbf{X}_{ic}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X}_{ic} \tag{3}$$

Similarly, the MLE of $\mathbf{\Omega}$ is

$$\hat{\mathbf{\Omega}} = \frac{1}{Nq}\sum_{i=1}^{N}\mathbf{X}_{ic}\hat{\mathbf{\Psi}}^{-1}\mathbf{X}_{ic}' \tag{4}$$

With the following condition

$$\sum_{i=1}^{N}\mathbf{X}_{icq}'\mathbf{\Omega}^{-1}\mathbf{X}_{icq} = Np \tag{5}$$

where $\mathbf{X}_{ic} = (\mathbf{X}_{ic1} : \mathbf{X}_{icq})$ and $\mathbf{X}_{icq}$ is $p \times 1$, The maximum likelihood estimates of $\mathbf{\Omega}$ and $\mathbf{\Psi}$ are obtained by solving simultaneously and alternatively the equations (2.3) and (2.4) subject to the condition (2.5). This is the so called "flip-flop" algorithm. We can summarize the results above in the following theorem.

Theorem 2.1. *(Srivastava et al. (2008)) Assume that $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$ are iid $N_{p\times q}(\mathbf{M}, \mathbf{\Omega}, \mathbf{\Psi})$ with ψ_{qq} = 1. If $N > max(p, q)$, then the maximum likelihood estimation equations given by equations (2.3) and (2.4) subject to the condition (2.5) will always converge to the unique maximum.*

When a structured covariance matrix (Definition 2.3) is available, the covariance matrix is estimated using the results of Srivastava et al. (2008). On the other hand, when the covariance matrix is unstructured, our simulations have shown that the estimates suggested by Srivastava et al. (2008) can also be used. Also, the main advantage in using a Kronecker structure is the decrease in the number of parameters.

2.2 Slicing and covariance estimation

Suppose that we have P component vectors $\mathbf{x}_i, i = 1, 2, \ldots, N$ that are independent multivariate normal random vectors with mean vector $\mathbf{\mu}_0$ and covariance matrix $\mathbf{\Sigma}$.

Definition 2.4. *A P vector $\mathbf{x}$ is said to be sliced into a p and q matrix $\mathbf{X}$ when $\mathbf{x}$ is written as a matrix $\mathbf{X}$ with p rows and q columns (we assume that there exist two integers p and q with $P = p \times q$). The p dimensional columns of $\mathbf{X}$ are obtained by slicing the vector $\mathbf{x}$ into q vectors. Then, by stacking these column vectors, the p and q matrix $\mathbf{X}$ is obtained.*

According to the definition above, the "slicing" operator transforms a vector of length pq into a $p \times q$ matrix. The vec operator applied to any slicing of $\mathbf{x}$ gives $\mathbf{x}$.
Given $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ following a multivariate normal distribution $N_P(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, by slicing the $\mathbf{x}$ into a matrix and by assuming that the model in Definition 2.3 holds this random matrix, a nonsingular covariance matrix of $\mathbf{x}$ can be obtained even when $N < P$. Also, we suggest that the condition for slicing to produce nonsingular estimates of the covariance matrix is $Np > q$. We additionally assume that $p < q$.

3. One-sample testing for high-dimensional data

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ be independent and identically distributed random vectors with mean vector $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. We are interested in the hypothesis testing problem of the form

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \quad vs \quad H_a : \boldsymbol{\mu} \neq \mathbf{0}$$

when the sample size N is smaller than or equal to the dimension P, that is $N \leq P$.
The traditional Hotelling T^2 test used the test statistics

$$T^2 = N\bar{\mathbf{x}}'\mathbf{S}^{-1}\bar{\mathbf{x}}$$

where the sample mean vector $\bar{\mathbf{x}}$ and the sample covariance matrix are defined, respectively by

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i, \quad and \quad \mathbf{S} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'.$$

For these settings, the covariance matrix $\mathbf{S}$ is singular and $\mathbf{S}^{-1}$ does not exist. So, the Hotelling test cannot be used. We therefore look for another estimate of the covariance matrix that is nonsingular and use that estimate instead of the empirical covariance matrix. One such estimate is the one proposed by Akdemir, 2011 and discussed above. So, a nonsingular estimate of $\boldsymbol{\Sigma}$ is $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Psi}} \otimes \hat{\boldsymbol{\Omega}}$ and the modified Hotelling statistic is defined as

$$T_N^2 = N\bar{\mathbf{x}}'(\hat{\boldsymbol{\Psi}}^{-1} \otimes \hat{\boldsymbol{\Omega}}^{-1})\bar{\mathbf{x}} \tag{6}$$

where $\hat{\boldsymbol{\Omega}}$ is $p \times p$ and $\hat{\boldsymbol{\Psi}}$ is $q \times q$.
To be able to use the proposed modified statistic in practice, one needs to know the distribution of this statistic under the null hypothesis before computing either a p-value or a critical value. Obtaining the exact or asymptotic null distribution is not an easy task, so the null distribution is obtained by simulation. One technique that can be easily used is the Monte Carlo method. Recall that the key to use Monte Carlo is to be able to simulate the desired statistic under the null hypothesis. In our case, the modified Hotelling statistic can be easily simulated under H_0 as follows. When H_0 is true, then $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ following a multivariate normal distribution $N_P(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$, where μ_0 is a P-dimensional vector with entries equal to 0. A simulated T_N^2 under H_0 is the result of applying a given simulated observation from $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ under the null to the expression (6). Let T_0 be the observed value of the test statistic T_N^2 based on the available sample of N observations and P variables. The Monte Carlo test is based on compare T_0 with an empirical distribution constructed with $(m-1)$ simulated statistics under H_0. A more detailed description about the Monte Carlo p-value and its properties will be seen in Section 4.

3.1 Other tests

Several tests have been proposed to solve the problem

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \quad vs \quad H_a : \boldsymbol{\mu} \neq \mathbf{0}$$

when $N \leq P$. Some authors developed tests which do not require the nonsingularity of the sample covariance matrix $\mathbf{S}$.
Assume that $n = N - 1$, Dempster, 1958 proposed a test based on the statistic

$$T_D = \frac{n\bar{\mathbf{x}}'\bar{\mathbf{x}}}{tr\mathbf{S}} \tag{7}$$

Bai and Saranadasa (1996) proposed another test statistic for testing the hypothesis (1.1), which is given by

$$T_{BS} = \frac{n\bar{\mathbf{x}}'\bar{\mathbf{x}} - tr\mathbf{S}}{\left[\frac{2n(n+1)}{(n-1)(n+2)}\left(tr\mathbf{S}^2 - \frac{(tr\mathbf{S})^2}{n}\right)\right]^{\frac{1}{2}}} \tag{8}$$

Srivastava and Du (2008) proposed another test statistic for testing the hypothesis (1.1). The test is based on the test statistic

$$T_{SD} = \frac{N\bar{\mathbf{x}}'\mathbf{D}_S^{-1}\bar{\mathbf{x}} - \frac{nP}{n-1}}{\sqrt{2\left(tr\mathbf{R}^2 - \frac{P^2}{n}\right)c_{P,n}}} \tag{9}$$

where the diagonal matrix of sample variances is defined by $\mathbf{D}_S = diag(s_{11}, \ldots, s_{PP})$, where $s_{11}, \ldots, s_{PP}$ are the diagonal elements of $\mathbf{S}$. The sample correlation matrix $\mathbf{R}$ is defined by $\mathbf{R} = \mathbf{D}_S^{-\frac{1}{2}}\mathbf{S}\mathbf{D}_S^{-\frac{1}{2}} = (r_{ij})$, where r_{ij} is the sample correlation between the i^{th} and j^{th} components of the random vector based on N observations and $r_{ij} = 1, i = 1, \ldots, P$. The adjustment coefficient $c_{P,n} \to 1$ in probability as $(n, P) \to \infty$. The one particular choice of $c_{P,n}$ that we use in this paper is the choice that is provided by Srivastava and Du (2008) and is given by $c_{P,n} = 1 + \frac{tr\mathbf{R}^2}{P^{\frac{3}{2}}}$. Srivastava and Du (2008) showed that their proposed test statistic has better powers than Dempster's T and Bai-Saranadasa's T.

4. Monte Carlo tests

To decide whether H_0 should be rejected or not, we must know the distribution of our test statistic under H_0 to obtain the exact p-value, which is used to perform the exact hypothesis test. Since we are using the estimate of the covariance matrix proposed by Akdemir and Gupta (2011), the modified Hotelling's statistic we propose here does not have the same distribution as Hotelling's T^2, and it is too cumbersome to obtain the analytical distribution of T_N^2. However, it is simple to simulate T_N^2 under the null hypothesis, hence, the Monte Carlo (MC) test can be used as an alternative way to perform the test.

4.1 Conventional Monte Carlo tests

Let U be a test statistic that can be simulated under H_0. The fixed-size or conventional Monte Carlo test first calculates the observed value of U, u_0, using the available sample. Then, a sample of $(m-1)$ test statistics is generated from U under the null hypothesis. Denote each simulated value by $u_i, i = 1, \ldots, (m-1)$. The conventional Monte Carlo p-value is

$$P_{mc} = \frac{1 + \sum_{i=1}^{m-1} I(u_i \geq u_0)}{m} \tag{10}$$

The test criterion is based on rejecting the null hypothesis if $P_{mc} \leq \alpha$, where α is the significance level. Denote the conventional Monte Carlo test procedure by MC_m. An important property is that MC_m has significance level equal to α. When U has a continuous distribution, and m is a multiple of $\lfloor 1/\alpha \rfloor$, the probability of type I error is equal to α, and if the distribution of U is discrete, then the size of the MC test is at most α (Silva et al., 2009). Additionally, Jockel (1984) has proved for a large class of test statistics that, if we choose m large enough, as $m = 10000$, for example, the power loss of the MC test, with respect to the exact test, is smaller than 2%.

4.2 Sequential Monte Carlo tests

A weakness of the conventional Monte Carlo test is that the number of simulations $(m-1)$ is fixed. Typically, we set $m = 10000$ or $m = 1000$. Recall that the method proposed in this paper uses two levels of simulation. This means that there is a first Monte Carlo simulation, which is the estimation of the covariance matrix, nested inside a second Monte Carlo simulation, which is the estimation of the distribution of the test statistic under H_0. So, the simulation of each independent copy of the test statistic can take a long time when dealing with high dimensional data and computationally intensive statistical methods needed to estimate the covariance matrix.

Therefore, it would be beneficial if we could shorten the number of simulations required to make a decision concerning the null hypothesis, while having the same power as a test that run all $(m-1)$ simulations.

Besag & Clifford (1991) developed a sequential Monte Carlo test to obtain p-values without fixing the number of simulations. The sequential Monte Carlo test is based on the idea that if there is little evidence against the null hypothesis early in the Monte Carlo procedure, then it is wasteful to run all $(m-1)$ simulations. It keeps simulating by Monte Carlo from the null hypothesis distribution until h of the simulated values are larger than the observed value u_0. There is also an upper limit $(w-1)$ for the total number of simulations. The p-value is based on the proportion of simulated values larger than or equal to u_0 at the stopping time. In other words, simulate independently and sequentially the random values $U_1, U_2, \ldots, U_L$ from the same distribution as U under the null hypothesis. The random variable L has possible values $h, h+1, \ldots, w-1$ and its value is determined in the following way: L is the first time when there are h simulated values larger than u_0. If this has not occurred at step $w-1$, then let $L = w-1$.

Let g be the number of simulated test statistics, u_i, larger than the observed value of the test statistic, u_0 at termination. Let l be the number of Monte Carlo simulations performed and $w-1$ be the upper limit for the total number of Monte Carlo simulations to be performed during the entire procedure. The sequential Monte Carlo test first calculates the observed value of the test statistic, u_0, using the available sample of N observations and P dimensions. Then, it keeps simulating by Monte Carlo from the null hypothesis distribution until

$$l = w - 1 \;\; and \;\; g < h$$

The sequential Monte Carlo p-value is

$$p_s = \begin{cases} h/l & \text{if } g = l; \\ (g+1)/w & \text{if } g < l\,. \end{cases} \tag{11}$$

The null hypothesis is rejected with α level if $p_s \leq \alpha$. Denote this sequential procedure by MC_h. Silva et al. (2009) demonstrated that MC_m and MC_h has same power if we take $h = \lfloor \alpha m \rfloor$, where $\lfloor x \rfloor$ is the floor of x, the largest integer smaller than x. Also, Silva et al. (2009) proved that, for $w \geq h/\alpha + 1$, the power of MC_h is constant. Then, in order to save execution time, it is convenient to use $w = \lfloor h/\alpha \rfloor + 1$. By combining these two rules, we need to have $w = m+1$ in order to have the same power between MC_m and

MC_h and therefore, minimizing the choice of w. In other words, they showed that, given a conventional Monte Carlo test based on $(m-1)$ simulations, there is always a sequential Monte Carlo test with the same power but typically requiring a smaller number of simulations.

5. The power of T_N^2

In this section, we compare the power of our proposed test statistic with the powers from the main important statistics in the literature described in this paper. To compare the four tests based on the statistics $(T_N^2, T_{BS}, T_D, T_{SD})$, we define the attained significance levels and empirical powers similar to how Srivastava and Du (2008) defined them and let $\boldsymbol{\mu}_0 = \mathbf{0}$. We compare the four tests on the same scale, i.e., we use the conventional MC test instead of the asymptotic distributions, as described by Srivastava and Du (2008), to compare them.

5.1 Simulation Design

To calculate the attained significance levels and powers we first simulate m replications of the data set under the null hypothesis. Then, we select the $(m\alpha)^{th}$ largest value of the test statistic as the empirical critical point, denoted $\hat{t}_{1-\alpha}$. With K replications of the data set under the null hypothesis, we compute the attained significance level as

$$\hat{\alpha} = \frac{1}{K}\sum_{i=1}^{K} I(t_i^0 \geq \hat{t}_{1-\alpha}) \tag{12}$$

where t_i^0 is the value of the test statistic based on the data sets simulated from the null hypothesis. We have chosen K as 5000 and α equal to 0.05. Note that the comparison $t_i^0 \geq \hat{t}_{1-\alpha}$ is essentialy the conventional Monte Carlo test. Therefore, as the test statistics treated here have continuous distributions, the probability of type I error is proved to be equal to α, and so it is not really necessary to verify if $\hat{\alpha}$ is close to α, because this is a theoretical result valid for any test statistic.
Next, we have simulated another K replications of the data set under the alternative hypothesis to calculate the empirical power by:

$$\hat{\beta} = \frac{1}{K}\sum_{i=1}^{K} I(t_i^A \geq \hat{t}_{1-\alpha}) \tag{13}$$

where t_i^A is the value of the test statistic based on the data sets simulated from the alternative hypothesis. Again, K has been chosen as 5000 in our simulations. Also, the parameter selection is done as described by Srivastava and Du (2008). Recall that the covariance can be defined by $\boldsymbol{\Sigma} = \mathbf{D}^{-\frac{1}{2}}\mathbf{R}\mathbf{D}^{-\frac{1}{2}}$. We will consider six different covariance structures in our simulation. Covariance structures 1, 2, and 3 are constructed from an independent correlation structure

$$\mathbf{R} = \mathbf{I}_p = diag(1, 1, \ldots, 1),$$

while covariance structures 4, 5, and 6 are obtained from an equal correlation structure

$$\mathbf{R} = \mathbf{R}_1 = (\rho_{ij}) : \rho_{ij} = 0.25, \qquad i \neq j.$$

Covariance structures 1 and 4 will also be constructed from a diagonal matrix of variances

$$\mathbf{D} = \mathbf{I}_p.$$

Covariance structures 2 and 5 will be constructed by letting

$$\mathbf{D} = \mathbf{D}_1 = diag(\sigma_{11}, \ldots, \sigma_{PP})$$

where $\sigma_{11}^{\frac{1}{2}}, \ldots, \sigma_{PP}^{\frac{1}{2}} \sim Unif(2, 3)$.
Covariance structures 3 and 6 will be constructed by letting

$$\mathbf{D} = \mathbf{D}_2 = diag(\sigma_{11}, \ldots, \sigma_{PP})$$

where $\sigma_{11}, \ldots, \sigma_{PP} \sim \chi_3^2$. For the alternative hypothesis, we choose $\boldsymbol{\mu} = \boldsymbol{\nu} = (\nu_1, \ldots, \nu_P)' : \nu_{2k-1} = 0$ and $\nu_{2k} \sim Unif(-1/2, 1/2), k = 1, \ldots, p/2$
To construct the ROC Curves, we used the setting corresponding for $N = 40$, $P = 100$ and computed the empirical powers at different significance levels α.

5.2 Simulation Results

The attained significance levels of the tests T_N^2, T_{BS}, T_D, and T_{SD} approximate the nominal level $\alpha =$ 0.05 reasonably well in all cases. When $\mathbf{R} = \mathbf{I}_p$, the powers of all the tests are very close to each other, as shown in Figures 1, 2, and 3 (tables of empirical powers, left, and ROC curves, right). When $\mathbf{R} = \mathbf{R}_1$ and $\mathbf{D} = \mathbf{D}_2$, the powers of T_N^2 and T_{SD} are close to each other but better than T_{BS} and T_D, as shown in Figure 6 (table of empirical powers, left, and ROC curve, right). However, when $\mathbf{R} = \mathbf{R}_1$ and $\mathbf{D} = \mathbf{D}_1$, the powers of T_N^2 are substantially better than those of the other tests, as shown in Figure 5 (table of empirical powers, left, and ROC curve, right).

5.3 Comparing Monte Carlo Tests

In our simulations we would like to compare the p-values for the conventional and the sequential Monte Carlo tests under different simulated samples. As mentioned before, the literature has proved analitically that MC_m and MC_h have same power by choosing $h = \lfloor \alpha m \rfloor$. Then, we can use our simulation results to check if the assumptions, calculations and computational model have accuracy with the results expected theoretically. We basically would like to verify if the decision made by the conventional Monte Carlo test is the same as the decision made by the sequential Monte Carlo test for our simulations. The simulated samples were generated from a multivariate normal distribution under different covariance structures and mean vectors. We set α to 0.01 and $m = 1000$. This yields $h = 10$.
As shown in Table 1, for every example, the conventional and sequential Monte Carlo tests reached the same decision as expected theoretically. The p-values of the two Monte Carlo tests are very close to each other.
We emphasize that, if there is little evidence against the null hypothesis early in the Monte Carlo procedure, the sequential Monte Carlo test requires a small number of simulations to reach the same decision than MC_m. But this execution time reduction is not expressive when the null hypothesis is false.

6. Example

In this section, we apply the conventional and sequential Monte Carlo tests using modified Hotelling's T^2 to a DNA microarray data set. We will use Alon's Colon Cancer Dataset, Alon et al. (1999) (http://genomics-pubs.princeton.edu/oncology/). The dataset consists of 2000 genes measured on 62 patients: 40 diagnosed

with colon cancer and 22 healthy patients. We would like to see if the population mean vector $\boldsymbol{\mu}$ for only the colon cancer patients is significantly different from some $\boldsymbol{\mu}_0$. We let $m = 1000$ and $\alpha = 0.01$.
In a first scenario, the four tests are giving a p-value of 0 giving evidence that the mean vector is not $\boldsymbol{\mu}_0$. In a second scenario, we compare the conventional Monte Carlo to the sequential Monte Carlo. The results are shown in Table 2. The p-value of the conventional Monte Carlo test is 0.854 and the p-value of the sequential Monte Carlo test is 0.821. Therefore, the estimated p-values are fairly close to each other. Since the p-values for both tests are greater than = 0.05, we cannot reject the null hypothesis. Therefore, both tests conclude that $\boldsymbol{\mu}$ is not significantly different from $\boldsymbol{\mu}_0$. Thus, whether using conventional or sequential Monte Carlo tests, we made the same decision. However, it took less simulations to reach the decision when using sequential Monte Carlo. It took 28 simulations to reach the decision using sequential Monte Carlo, while it took 1000 simulations to reach the same decision using conventional Monte Carlo. Therefore, in order to save time, especially when dealing with high dimensional data or computationally intensive statistical methods, sequential Monte Carlo should be adopted.

7. Conclusion

In this paper, we have proposed a new approach to test the mean vector of a population when the number of variables is larger than the number of observations. The statistic proposed in this paper is based on a modification of the well known Hotelling's T^2 by replacing the singular empirical covariance matrix by a nonsingular estimate of the covariance matrix. This estimator is based on the work of Akdemir and Gupta (2011) and is used here because of certain desirable properties of that estimator such as being positive definite and full rank even when the empirical covariance is singular. Even though this nonsingular estimate used the assumption that the covariance matrix has a Kronecker structure, the statistic proposed seems to be performing well. The performance of this test was discussed and compared to some existing tests, for high dimensional data, like Srivastava test, Dempster test, and Bai test. The simulations show that the proposed test has good performances, not always the best but very close to the best, compared to the existing tests considered in this current paper. The results obtained indicate that the overall performance of this statistic makes it a new appealing tool for testing the mean vector of a population in high dimensional problems.

REFERENCES

Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., and Levine, A. J. (1999) "Broad Patterns of Gene Expression Revealed by Clustering Analysis of Tumor and Normal Colon Tissues Probed by Oligonucleotide Arrays", *Proceedings of the National Academy of Sciences of the United States of America*, 96, 12, 6745

Akdemir, D, and Gupta, A.K. (2011), "Array Variate Random Variables with Multiway Kronecker Delta Covariance Matrix Structure," *Journal of Algebraic Statistics*, 2, 1, 98–113

Bai, Z., and Saranadasa, H. (1996), "Effect of high dimension: an example of a two sample problem", *Statist. Sinica*, 6, 311–329.

Besag, J., and Clifford, P. (1991), "Sequential Monte Carlo p-values", *Biometrika*, 78, 2, 301–304

Dempster, A. P. (1972), "Covariance selection," *Biometrics*, 28, 157–175.

Dempster, A. P. (1958), "A high dimensional two sample significance test," *Ann. Math. Statist.*, 29, 995–1010.

Gupta, A. K., and Nagar, D. K. (2000), *Matrix Variate Distributions*, Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics

Krzyśko, M., and Skorzybut, M. (2009), "Discriminant Analysis of Multivariate Repeated Measures Data with a Kronecker Product Structured Covariance Matrices," *Statistical Papers*, 50, 4, 817–835

Lu, N., and Zimmerman, D. L. (2005), "The Likelihood Ratio Test for a Separable Covariance Matrix," *Statistics & Probability Letters*, 73, 4, 449–457

Naik, D. N., and Rao, S. S. (2001), "Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix," *Journal of Applied Statistics*, 28, 1, 91–105.

Roy, A., and Khattree, R. (2003), "Tests for Mean and Covariance Structures Relevant in Repeated Measures Based Discriminant Analysis", *Journal of Applied Statistical Science*, 12, 2, 91–104.

Silva, I., Assuncao, R., and Costa, M. (2009), "Power of the sequential Monte Carlo Test," *Sequential Analysis*, 28, 2, 163–174

Srivastava, M. S., and Du, M. (2008), "A Test for the Mean Vector with Fewer Observations than the Dimension." *Journal of Multivariate Analysis*, 99, 386–402

Srivastava, M. S., von Rosen, T., and Von Rosen, D. (2008), "Models with a Kronecker Product Covariance Structure: Estimation and Testing," *Mathematical Methods of Statistics*, 17, 4, 357–370.

Appendix

	$N = 20$, $P = 500$	$N = 30$, $P = 1000$	$N = 40$, $P = 2000$
p-value using Conventional MC	0.015	0.239	0.12
p-value using Sequential MC	0.01592	0.2173	0.1086
Number of Sequential MC	314	23	46

Table 1: Comparing the p-values of the Conventional and Sequential Monte Carlo tests using the modified Hotelling's T^2.

	T_N^2 using Conventional MC	T_N^2 using Sequential MC
p-value	0.854	0.821
Number of Simulations	1000	28

Table 2: Observed p-values for colon cancer data.

P	N	T_N^2	T_{SD}	T_D	T_{BD}
60	30	0.975	0.978	0.987	0.988
100	40	1.0	1.0	1.0	1.0
	60	1.0	1.0	1.0	1.0
	80	1.0	1.0	1.0	1.0
150	40	1.0	1.0	1.0	1.0
	60	1.0	1.0	1.0	1.0
	80	1.0	1.0	1.0	1.0
200	40	1.0	1.0	1.0	1.0
	60	1.0	1.0	1.0	1.0
	80	1.0	1.0	1.0	1.0
400	40	1.0	1.0	1.0	1.0
	60	1.0	1.0	1.0	1.0
	80	1.0	1.0	1.0	1.0

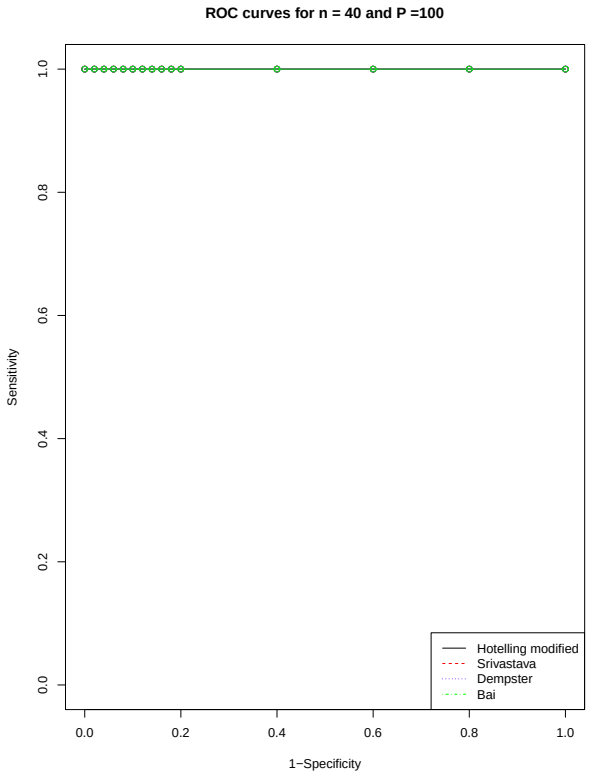

Figure 1: Empirical powers under the alternative hypothesis, when the diagonal matrix of variances is the identity matrix and the population correlation matrix is the identity matrix

P	N	T_N^2	T_{SD}	T_D	T_{BD}
60	30	0.211	0.261	0.251	0.245
100	40	0.432	0.542	0.459	0.461
	60	0.742	0.8	0.760	0.762
	80	0.938	0.942	0.937	0.936
150	40	0.608	0.659	0.668	0.662
	60	0.858	0.886	0.848	0.842
	80	0.996	0.999	0.995	0.995
200	40	0.7	0.781	0.735	0.734
	60	0.969	0.980	0.969	0.968
	80	0.996	0.999	0.994	0.994
400	40	0.912	0.948	0.940	0.939
	60	0.997	0.998	0.997	0.997
	80	1.0	1.0	1.0	1.0

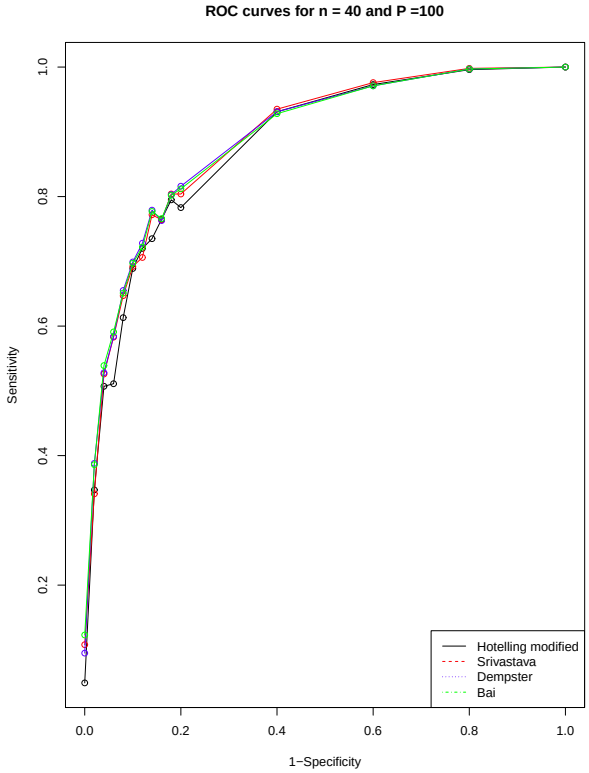

Figure 2: Empirical powers under the alternative hypothesis, when the correlation matrix has independent correlation structure and the square root of the variances are independently and identically Unif(2,3) distributed

P	N	T_N^2	T_{SD}	T_D	T_{BD}
60	30	0.436	0.813	0.366	0.373
100	40	0.973	1.0	0.877	0.877
	60	1.0	1.0	1.0	1.0
	80	0.999	1.0	0.995	0.995
150	40	0.995	1.0	0.979	0.974
	60	1.0	1.0	1.0	1.0
	80	1.0	1.0	1.0	1.0
200	40	0.996	1.0	0.980	0.978
	60	1.0	1.0	1.0	1.0
	80	1.0	1.0	1.0	1.0
400	40	1.0	1.0	1.0	1.0
	60	1.0	1.0	1.0	1.0
	80	1.0	1.0	1.0	1.0

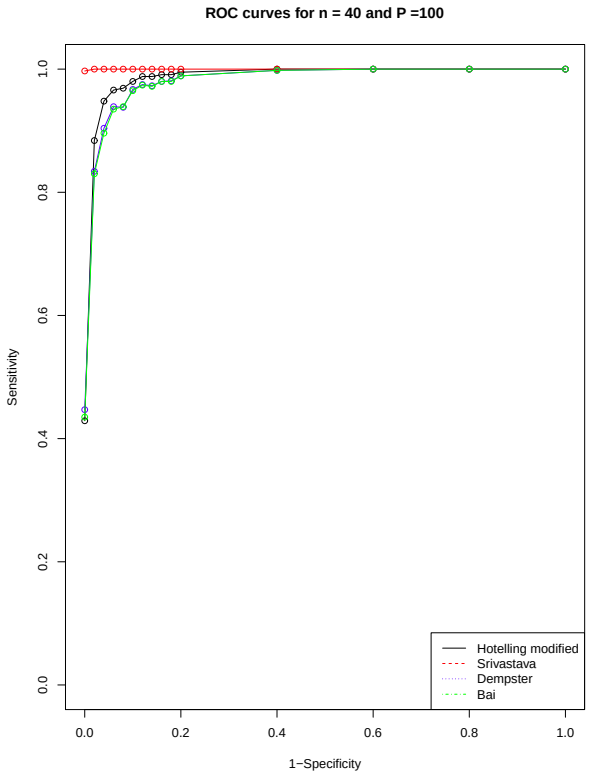

Figure 3: Empirical powers under the alternative hypothesis, when the correlation matrix has independent correlation structure and the variances are independently and identically χ_3^2 distributed.

P	N	T_N^2	T_{SD}	T_D	T_{BD}
60	30	0.999	0.660	0.736	0.658
100	40	1.0	0.881	0.961	0.915
	60	1.0	1.0	1.0	1.0
	80	0.999	1.0	0.995	0.995
150	40	1.0	0.965	1.0	0.990
	60	1.0	1.0	1.0	1.0
	80	1.0	1.0	1.0	1.0
200	40	1.0	0.943	1.0	0.971
	60	1.0	1.0	1.0	1.0
	80	1.0	1.0	1.0	1.0
400	40	1.0	0.954	1.0	0.994
	60	1.0	1.0	1.0	1.0
	80	1.0	1.0	1.0	1.0

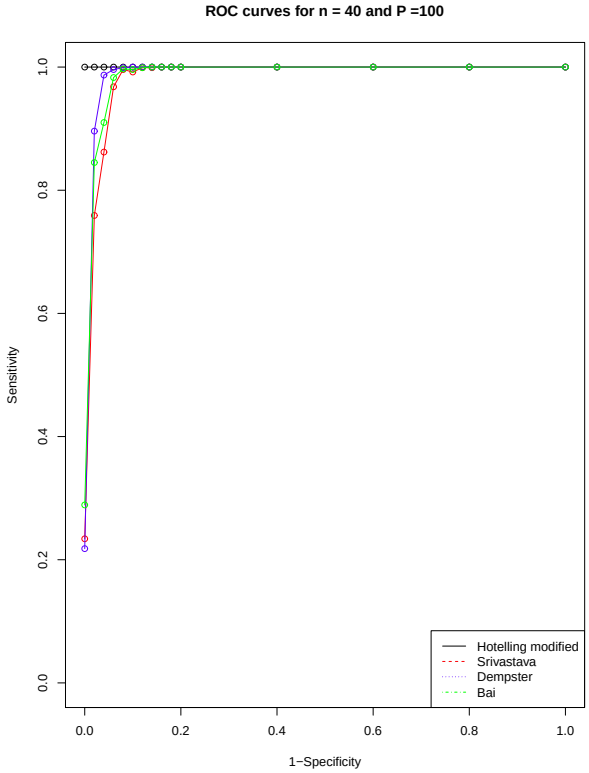

Figure 4: Empirical powers under the alternative hypothesis, when the diagonal matrix of variances is the identity matrix and the population correlation matrix has an equal correlation structure such that $\rho_{ij} = 0.25$, when $i \neq j$ and $\rho_{ij} = 1$, otherwise.

P	N	T_N^2	T_{SD}	T_D	T_{BD}
60	30	0.260	0.082	0.094	0.087
100	40	0.571	0.081	0.071	0.077
	60	0.814	0.139	0.132	0.131
	80	0.966	0.236	0.194	0.219
150	40	0.663	0.096	0.101	0.102
	60	0.949	0.161	0.144	0.144
	80	0.995	0.232	0.193	0.198
200	40	0.793	0.107	0.084	0.095
	60	0.971	0.165	0.144	0.132
	80	0.996	0.195	0.176	0.205
400	40	0.968	0.087	0.064	0.080
	60	1.0	0.120	0.135	0.117
	80	1.0	0.304	0.276	0.263

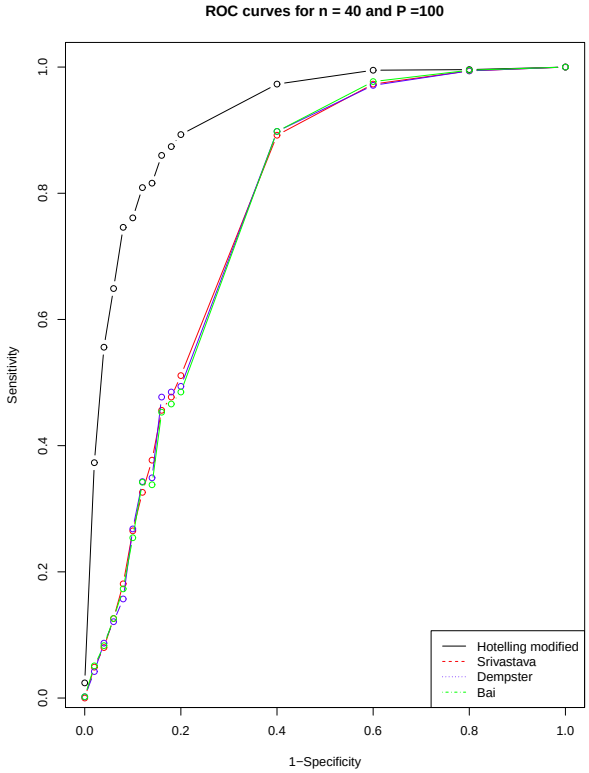

Figure 5: Empirical powers under the alternative hypothesis, when the square root of the variances are identically and independently Unif(2,3) distributed and the population correlation matrix has an equal correlation structure such that $\rho_{ij} = 0.25$, when $i \neq j$ and $\rho_{ij} = 1$, otherwise.

P	N	T_N^2	T_{SD}	T_D	T_{BD}
60	30	0.926	0.772	0.267	0.287
100	40	0.825	0.429	0.112	0.121
	60	0.993	0.983	0.430	0.388
	80	0.995	0.885	0.532	0.545
150	40	0.996	0.664	0.275	0.297
	60	1.0	1.0	0.668	0.624
	80	1.0	0.996	0.665	0.632
200	40	1.0	0.991	0.235	0.246
	60	1.0	1.0	0.305	0.282
	80	1.0	1.0	0.578	0.593
400	40	0.999	0.923	0.247	0.272
	60	1.0	1.0	0.354	0.364
	80	1.0	1.0	0.828	0.724

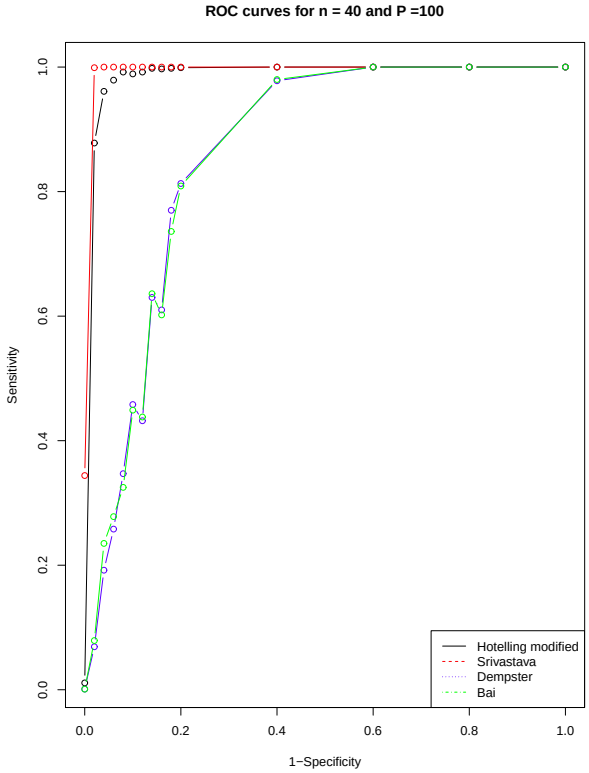

Figure 6: Empirical powers under the alternative hypothesis, when the variances are identically and independently χ_3^2 distributed and the population correlation matrix has an equal correlation structure such that $\rho_{ij} = 0.25$, when $i \neq j$ and $\rho_{ij} = 1$, otherwise.