A Review of Some Developments of Multivariate Isotonic Regression

D.D.S. KULATUNGA

Department of Mathematics, Faculty of Science
University of Kelaniya, Kelaniya

ABSTRACT

The theory of Multivariate Isotonic Regression plays a key role in the field of statistical inference under order restrictions for vector valued parameters. This paper which reviews some developments of multivariate isotonic regression includes an algorithm for computing multivariate isotonic regression along with its convergence, estimation and tests in multivariate normal distribution and some applications.

Key words: Likelihood ratio test, Order restriction Multivariate isotonic regression, Multivariate normal distributions.

INTRODUCTION

Many situations occur in statistical analysis where the statistician has prior knowledge in the form of order restrictions on the parameters under investigation. In the most common case where data are arranged in ordered groups, the mean value of a random variable is assumed to change monotonically with the ordering of the groups. For example if we want to know whether the average height of the children of an area increase year by year, we expect it to be possible to make better statistical inference both in estimation and hypothesis testing on the group means when this prior information is fully utilized than when it is ignored. Thus statistical inference in the presence of order restrictions is an important area in statistical analysis. Isotonic regression theory plays a key role in this field. Most of the theory related to this has been reviewed in Barlow *et al.* (1972) and Robertson *et al.*, (1988).

The method of maximum likelihood is one of the general approaches to the inference under order restrictions. Bartholomew (1959a; 1959b; 1961) derived the likelihood ratio test for the problem of testing the equality of the means of several normal populations against an order restriction on means. Suppose that we have a sample of independent observations $x_1, x_2, \ldots, x_k$, where x_i is drawn from a normal population with unknown mean θ_i and known variance $\sigma_i^2, i = 1,\ldots,k$. We are concerned with the testing problem H_0: $\theta_1 = \theta_2 = \cdots = \theta_k$ versus H_1: $(\theta_1, \theta_2, \ldots, \theta_k)$ satisfies some order restriction. A particular form of H_1 to which a considerable attention has been devoted is $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$, not all θ_i's equal, referred to as the case of simple order. The maximum likelihood estimates under H_1 lead to the following quadratic programming problem:

$$\text{minimize} \sum_{i=1}^{k} (x_i - \theta_i)^2 w_i \tag{1}$$

under the restriction that $(\theta_1, \theta_2, \ldots, \theta_k)$ is constrained by some partial order, where $w_i = 1/\sigma^2$. The solution to this is furnished by the isotonic regression of $x_1, x_2, \ldots, x_k$ with weights $w_2, \ldots, w_k$.

A number of algorithms have been developed in the literature for calculating the isotonic regression and they are discussed in detail in Section 2.3 of Barlow *et al.* (1972) and Section 1.4 of Robertson *et al.* (1988). In the case of simple order a simple iterative âlgorithm called Pooled Adjacent Violators Algorithm (PAVA) is available. Cran (1980) gives a subroutine for computing the isotonic regression using up and down blocks algorithm in the case of simple ordering. Some results and algorithms related to this can also be found in Geng and Shi (1990), Hansohm (2007) and Hansohm and Hu (2008).

A multivariate generalization of the isotonic regression including the multivariate extensions of well-known Batholomew's $\bar{\chi}^2$ and $\bar{E}^2$ statistics was first considered by Sasabuchi *et al.* (1983). This theory enables us to study statistical inference for ordered vector valued parameters or sets of ordered parameters. Let $\underline{X_i}$ be independent p-variate normal vectors with mean vectors $\underline{\theta_i}$ and covariance matrices Λ_i, $i = 1,\ldots,k$. We are concerned with the problem of testing H_0: $\underline{\theta_1} = \underline{\theta_2} = \ldots = \underline{\theta_k}$ against H_1: $(\underline{\theta_1}, \underline{\theta_2}, \ldots, \underline{\theta_k})$ satisfies some order restriction. A particular form of H_1 to which much attention is paid is $\underline{\theta_1} \leq \underline{\theta_2} \leq \cdots \leq \underline{\theta_k}$, where $\underline{\theta_i} \leq \underline{\theta_j}$ means that all the elements of $\underline{\theta_j} - \underline{\theta_i}$ are non negative. The standard method of finding the maximum likelihood estimator under H_1 leads to the solution of the multivariate generalization of the quadratic programming problem given by (1), which is

$$\text{minimize} \sum_{i=1}^{k} (\underline{X_i} - \underline{\theta_i})' \Lambda_i^{-1} (\underline{X_i} - \underline{\theta_i}) \tag{2}$$

under the restriction that $\underline{\theta}=(\theta_1, \theta_2, \ldots, \theta_k)$ is isotonic with respect to some partial order. The solution to (2) is the multivariate isotonic regression (MIR).

Sasabuchi *et al.* (1983) proposed an iterative algorithm for the computation of MIR in the bivariate case and studied its convergence. Again Sasabuchi *et al.* (1992) extended this algorithm to the general multivariate case and obtained conditions for the convergence of the algorithm. Sasabuchi *et al.* (2003) examined the convergence of the algorithm when the dimension is three and four through simulations and observed that the findings are exactly similar to those in the bivariate case. A Fortran program for the computation of the general MIR along with two examples of different types to illustrate the applications of MIR is given in Fernando and Kulatunga (2007). The convergence of the algorithm, when the dimension is greater than or equal to five, has also been investigated there. Although the convergence of the algorithm is not proved except under certain conditions, the simulation studies show that the algorithm converges in situations where the conditions are not satisfied.

Kulatunga and Sasabuchi (1984) derived the likelihood ratio test for testing H_0 versus H_1 in the multivariate case, studied its null distribution and observed that the null distribution of the likelihood ratio test cannot be computed except for some special situations. Nomakuchi and Shi (1988) proposed another test by combining best linear tests considered by Abelson and Tukey (1963), Schaafsma and Smid (1966) and Kudõ's (1963) multivariate analogue of the one-sided test. They have studied the power of the test and their results indicate that their test is not uniformly more powerful than the likelihood ratio test. It has been observed that the likelihood ratio test would be desirable but the covariance structure of the covariance matrices other than the independence with equal covariance matrices causes grave problems in distributional theory. This leads Sasabuchi *et al.* (1998) to propose three test procedures and study them through extensive simulations in the bivariate case. The same test procedures were again studied by Sasabuchi *et al.* (2003) when the dimension is larger than two and obtained a few different results. When covariance matrices are common but unknown, the testing problem was recently considered by Sasabuchi *et al.,* (2003). They proposed a test statistic, studied its upper tail probability under the null hypothesis and estimated its critical points.

DEFINITIONS

Let $K = \{1, \ldots, k\}$ be a finite set on which a partial order $\ll$ is defined. The partial order on K may or may not be the natural order among positive integers $1 \ll 2 \ll \ldots \ll k$, which is called the simple order.

Definition 2.1. A real vector $(\theta_1, \ldots, \theta_k)$ is said to be *isotonic* with respect to the partial order $\ll$, if $\mu, \nu \in K$ and $\mu \ll \nu$ imply $\theta_\mu \leq \theta_\nu$

Definition 2.2. Given real numbers $x_1, \ldots, x_k$ and positive numbers $w_1, \ldots, w_k$, a vector $(\hat{\theta}_1, \ldots, \hat{\theta}_k)$ is said to be the *univariate isotonic regression* of $x_1, \ldots, x_k$ with weights $w_1, \ldots, w_k$ if it is isotonic and minimizes

$\sum_{v=1}^{k}(x_v - \theta_v)^2 w_v$ under the restriction that $(\theta_1, \ldots, \theta_k)$ is isotonic.

Definition 2.3. A $p \times k$ real matrix $\theta = (\theta_1, \ldots, \theta_k)$ is said to be isotonic with respect to the partial order $\ll$, if $\mu, \nu \in K$ and $\mu \ll \nu$ imply $\theta_\mu \leq \theta_\nu$, where $\theta_\mu \leq \theta_\nu$ means all the elements of $\theta_\nu - \theta_\mu$ are non negative.

Throughout this paper $\min^*_\theta(.)$ denotes the minimum for all θ isotonic with respect to the partial order $\ll$.

Definition 2.4. Given p dimensional real vectors $x_1, \ldots, x_k$ and $p \times p$ positive definite matrices $\Lambda_1, \ldots, \Lambda_k$, a $p \times k$ matrix $(\hat{\theta}_1, \ldots, \hat{\theta}_k)$ is said to be the *multivariate isotonic regression,* in fact p-variate isotonic regression, of $x_1, \ldots, x_k$ with weights $\Lambda_1^{-1}, \ldots, \Lambda_k^{-1}$ if it is isotonic and satisfies

$$\min{}^*_\theta \textstyle\sum_v (x_v - \theta_v)' \Lambda_v^{-1} (x_v - \theta_v) = \sum_v (x_v - \hat{\theta}_v)' \Lambda_v^{-1} (x_v - \hat{\theta}_v).$$

To give a geometric interpretation of the MIR, for given $p \times k$ real matrices, $y = (y_1, y_2, \ldots, y_k)$ and $z = (z_1, z_2, \ldots, z_k)$, we define $\langle y, z\rangle = \sum_v y_v' \Lambda_v^{-1} z_v$, and $d(y, z) = \sqrt{\langle y - z, y - z\rangle}$.

It is clear that $\langle \cdot,\cdot \rangle$ and d are inner product and metric in pk-dimensional Euclidean space. Further define Θ to be the set of all $p \times k$ isotonic matrices. It follows easily that Θ is a closed convex cone and the multivariate isotonic regression of $x_1, \ldots, x_k$ with weights $\Lambda_1^{-1}, \ldots, \Lambda_k^{-1}$ is the closest point of $x = (x_1, \ldots, x_k)$ to Θ with respect to the metric d. Then from the general theory of convex analysis it follows that, given any partial order and weights, the MIR exists uniquely for any p dimensional real vectors $x_1, \ldots, x_k$.

AN ITERATIVE ALGORITHM FOR COMPUTING MULTIVARIATE ISOTONIC REGRESSION

In this section the algorithm for computing multivariate isotonic regression proposed in Sasabuchi *et al.* (1992) is given.

According to Definition 2.4, the problem is to find a $p \times k$ real matrix

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k) = \begin{pmatrix} \theta_{11} & \cdots & \theta_{1k} \\ \vdots & \vdots & \vdots \\ \theta_{p1} & \cdots & \theta_{pk} \end{pmatrix} = \begin{pmatrix} \theta^1 \\ \theta^2 \\ \vdots \\ \theta^p \end{pmatrix}$$

for given p-dimensional real vectors $x_1, \ldots, x_k$ and $p \times p$ positive definite matrices $\Lambda_1, \ldots, \Lambda_k$ which minimizes $L(\theta^1, \ldots, \theta^p) = \sum_{\nu=1}^{k} (x_\nu - \theta_\nu)' \Lambda_\nu^{-1} (x_\nu - \theta_\nu)$ under the condition that $\boldsymbol{\theta}$ is isotonic. As seen above, the right-hand side is expressed as a function of $\theta_1, \ldots, \theta_k$. But L is defined as a function of $\theta^1, \ldots, \theta^p$ because of the convenience in describing this algorithm. When the weight matrices are diagonal, as in the case of bivariate isotonic regression (cf. Sasabuchi *et al.*, 1983), $L(\theta^1, \ldots, \theta^p)$ can be written as follows:

$$L(\theta^1, \ldots, \theta^p) = \sum_{i=1}^{p} \sum_{\nu=1}^{k} (x_{i\nu} - \theta_{i\nu})^2 \lambda_{\nu ii}^{-1},$$

where $x_\nu = (x_{1\nu}, \ldots, x_{p\nu})$ and $\lambda_{\nu ii}$ is the (i,i)th element of Λ_ν. Thus the multivariate isotonic regression can be obtained easily by applying the methods of computing univariate isotonic regression to each term of the sum separately.

Now suppose that at least one weight matrix is not diagonal and in this case L can be rewritten in the following p forms:

$$L(\theta^1, \ldots, \theta^p) = f_1(\theta^2, \ldots, \theta^p) + g_1(\theta^1, \ldots, \theta^p)$$
$$= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$= f_i(\theta^1, \ldots, \theta^{i-1}, \theta^{i+1}, \ldots, \theta^p) + g_i(\theta^1, \ldots, \theta^p)$$
$$= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$= f_p(\theta^1, \ldots, \theta^{p-1}) + g_p(\theta^1, \ldots, \theta^p), \text{ where}$$

$$f_i(\theta^1, \ldots, \theta^{i-1}, \theta^{i+1}, \ldots, \theta^p) = \sum_{\nu=1}^{k} (x_{\nu(i)} - \theta_{\nu(i)})' \Lambda_{\nu(i)}^{-1} (x_{\nu(i)} - \theta_{\nu(i)}),$$

$$g_i(\theta^1, \ldots, \theta^p)$$
$$= \sum_{\nu=1}^{k} w_{\nu:(i)}^{-1} \{(x_{i\nu} - \theta_{i\nu}) - \lambda'_{\nu(i)} \Lambda_{\nu(i)}^{-1} (x_{\nu(i)} - \theta_{\nu(i)})\}^2,$$

$$w_{\nu:(i)} = \lambda_{\nu ii} - \lambda'_{\nu(i)} \Lambda_{\nu(i)}^{-1} \lambda_{\nu(i)}, \quad i = 1, \ldots, p; \quad \nu = 1, \ldots, k,$$

$x_{\nu(i)}$ and $\theta_{\nu(i)}$ are the column vectors obtained after deleting the ith element of x_ν and θ_ν respectively. $\Lambda_{\nu(i)}$ is the $(p-1) \times (p-1)$ submatrix obtained after deleting the ith row and column of Λ_ν and $\lambda_{\nu(i)}$ is the vector obtained after deleting the ith element of the ith column of Λ_ν.

Sasabuchi *et al.* (1992) described an algorithm of stepwise approximation to the MIR, where the nth approximation of the ith row vector θ^i computed at (n, i) step $(n = 1, 2, \ldots, i = 1, \ldots, p)$ is denoted by $\theta^{i(n)} = (\theta_{i1}^{(n)}, \ldots, \theta_{ik}^{(n)})$, $i = 1, \ldots, p$.

From the (1,1) step to $(1, p-1)$step of our algorithm, the first approximation $\theta^{1(1)}, \ldots\ldots, \theta^{p-1(1)}$ of $\theta^1, \ldots, \theta^{p-1}$ is taken arbitrarily. For instance, $\theta^{i(1)} = (x_{i1}, \ldots, x_{ik})$, $i = 1, \ldots, p-1$.

In the $(1, p)$ step, find the first approximation $\theta^{p(1)}$ of θ^p. In fact find isotonic $\theta^{p(1)}$ such that

$$\min_{\theta^p}{}^* L(\theta^{1(1)}, \ldots, \theta^{p-1(1)}, \theta^p) = L(\theta^{1(1)}, \ldots, \theta^{p-1(1)}, \theta^{p(1)}) \text{ or equivalently}$$
$$\min_{\theta^p}{}^* g_p(\theta^{1(1)}, \ldots, \theta^{p-1(1)}, \theta^p) = g_p(\theta^{1(1)}, \ldots, \theta^{p-1(1)}, \theta^{p(1)}).$$

Then $\theta_{p\nu}^{(1)}$ is just the univariate isotonic regression of $\{x_{p\nu} - \lambda'_{\nu(p)} \Lambda_{\nu(p)}^{-1} (x_{\nu(p)} - \theta_{\nu(p)}^{(1)})\}$ with weights

$$w_{v:(p)}^{-1} = \left(\lambda_{vpp} - \lambda_{v(p)}\Lambda_{v(p)}^{-1}\lambda_{v(p)}\right)^{-1}.$$

After computing $\theta^{1(n)}, \ldots, \theta^{p(n)}$ at $(n,1), \ldots, (n,p)$ steps, respectively, proceed to $(n+1,1)$step, where $\theta^{1(n+1)}$ is determined by the following relation with the condition that $\theta^{1(n+1)}$ is isotonic:

$\min_{\theta^1}{}^{*} L(\theta^1, \theta^{2(n)}, \ldots, \theta^{p(n)}) = L(\theta^{1(n+1)}, \theta^{2(n)}, \ldots, \theta^{p(n)})$ or equivalently

$$\min_{\theta^1}{}^{*} g_1(\theta^1, \theta^{2(n)}, \ldots, \theta^{p(n)}) = g_1(\theta^{1(n+1)}, \theta^{2(n)}, \ldots, \theta^{p(n)}).$$

In general, at $(n+1, i)$ step, find an isotonic $\theta^{i(n+1)}$ such that

$$\min_{\theta^i}{}^{*} L(\theta^{1(n+1)}, \ldots, \theta^{i-1(n+1)}, \theta^i, \theta^{i+1(n)}, \ldots, \theta^{p(n)}) = (\theta^{1(n+1)}, \ldots, \theta^{i-1(n+1)}, \theta^{i(n+1)}, \theta^{i+1(n)}, \ldots, \theta^{p(n)})$$

or equivalently

$$\min_{\theta^i}{}^{*} g_i(\theta^{1(n+1)}, \ldots, \theta^{i-1(n+1)}, \theta^i, \theta^{i+1(n)}, \ldots, \theta^{p(n)}) = g_i(\theta^{1(n+1)}, \ldots, \theta^{i-1(n+1)}, \theta^{i(n+1)}, \theta^{i+1(n)}, \ldots, \theta^{p(n)})$$

Continuing in this manner, at $(n+1, p)$ step find an isotonic $\theta^{p(n+1)}$ such that

$\min_{\theta^p}{}^{*} L(\theta^{1(n+1)}, \ldots, \theta^{p-1(n+1)}, \theta^p) = L(\theta^{1(n+1)}, \ldots, \theta^{p-1(n+1)}, \theta^{p(n+1)})$ or equivalently

$$\min_{\theta^p}{}^{*} g_p(\theta^{1(n+1)}, \ldots, \theta^{p-1(n+1)}, \theta^p) = g_p(\theta^{1(n+1)}, \ldots, \theta^{p-1(n+1)}, \theta^{p(n+1)}).$$

Here an iterative algorithm has been demonstrated in which each cycle of recurrent computations consists of p subcycles; each revises one row vector while the other p -1 row vectors are tentatively fixed. In fact this algorithm involves iterative computation of univariate isotonic regressions. Thus at each step $\theta^{1(n)}, \ldots, \theta^{p(n)}$ are determined uniquely.

CONVERGENCE OF THE ALGORITHM

The utility of the algorithm described in §3 lies in the following theorem which states that if the sequences $\{\theta^{i(n)}\}$, $i = 1, \ldots, p$ converge, then the resulting values are the MIR.

THEOREM 1 (Sasabuchi, Inutsuka and Kulatunga,1992)

If $\lim_{n\to\infty} \theta^{i(n)} = \theta^{i(\infty)}, \quad i = 1, \ldots, p$

exists, then the $p \times k$ matrix $\theta^{(\infty)} = (\theta^{1(\infty)'}, \ldots, \theta^{p(\infty)})$ is the MIR of $x_1, \ldots, x_k$ with weights $\Lambda_1^{-1}, \ldots, \Lambda_k^{-1}$.

The convergence of the algorithm has not been proved yet in general. However, Sasabuchi *et al.* (1983, 1992) obtained some results on the convergence of the algorithm and they are stated below:

COROLLARY 1 (Sasabuchi, Inutsuka and Kulatunga (1992))

When the weight matrices Λ_ν's take the form, for $\nu = 1, \ldots, k; \ i = 1, \ldots, p,$

$\lambda_{\nu ij} = \sigma_\nu^2 \rho_\nu, \ i \neq j, \qquad \lambda_{\nu ij} = \sigma_\nu^2,$ where $\lambda_{\nu ij}$ is the (i,j)-*th* element of Λ_ν, the sequences $\{\phi^{(n)}\}$ converge if $-(2p-3)^{-1} < \rho_\nu < 1, \quad \nu = 1, \ldots, k.$

Recently Fernando and Kulatunga (2007) studied the convergence of the algorithm through extensive simulations when $\Lambda_1, \ldots, \Lambda_k$ are equally correlated matrices whose diagonal elements are 1 and off-diagonal elements are ρ. These matrices are positive definite if and only if $-(p-1)^{-1} < \rho < 1$. From the above corollary, the algorithm converges to the MIR if $-(2p-3)^{-1} < \rho < 1$. This study was conducted by generating 1000 sets of k, p-variate random numbers with mean vectors $(0,0, \ldots, 0)'$, unit variance and correlation coefficient ρ, where k = 2 (2) 10 and ρ = -0.2 (0.1) 0.9 for p = 5 and ρ = -0.1 (0.1) (0.9) for p = 6 in order to ensure that the covariance matrices are positive definite. The iterative computation is terminated when the maximum norm of the difference between *(n-1)th* approximation and the *nth* one is smaller than 10^{-5}. It is revealed in the simulation study that the condition of Corollary 4.1 is not necessary for the convergence of the algorithm. For example when p = 5 and ρ = -0.2, the weight matrices are positive definite but the condition for the convergence is not satisfied. This fact has also been observed in the previous studies conducted for different dimensions (see Sasabuchi *et al.* (2003).

ESTIMATION AND TEST IN MULTIVARIATE NORMAL DISTRIBUTION

Let X_v be independent p- variate normal random vector with mean vectors θ_v and covariance matrices Λ_v ($v = 1 \ldots k$). Then it follows from the Definition 2.4 that, when Λ_v's are known, $\hat{\theta}_v$, the MIR of X_v with weights Λ_v^{-1} furnishes the maximum likelihood estimate of θ_v under the restriction that $(\theta_1, \ldots, \theta_k)$ is isotonic. When $\Lambda_v = \sigma^2\Gamma_v$, where σ^2 is unknown scalar factor and Γ_v is known, the maximum likelihood estimates of θ_v and σ^2 are θ_v^* and $\Sigma_v(X_v - \theta_v^*)' \Gamma_v^{-1}(X_v - \theta_v^*)/pk$ respectively, where θ_v^* is the MIR of X_v with weights Γ_v^{-1} (see Sasabuchi *et al.*, 1983).

Now we consider the testing problem having null and alternative hypotheses as H_o: $\theta_1 = \cdots = \theta_k$ and H_o: $(\theta_1, \ldots, \theta_k)$ isotonic. The following theorem gives the likelihood ratio test statistic in the multivariate case.

THEOREM 2. (Sasabuchi, Inutsuka and Kulatunga, 1983)

When Λ_v is known, the likelihood ratio test rejects H_0 for large values of

$$\bar{\chi}^2_{k,p} = \Sigma_v \left(\hat{\theta}_v - \bar{X}\right)' \Lambda_v^{-1} \left(\hat{\theta}_v - \bar{X}\right),$$

where $\bar{X} = \left(\Sigma_v \Lambda_v^{-1}\right)^{-1} \Sigma_v \Lambda_v^{-1} X_v$ and $\hat{\theta}_v$ is the MIR of x_v with weights Λ_v^{-1}. If $\Lambda_v = \sigma^2\Gamma_v$, where σ^2 is unknown and Γ_v is known and an independent estimate S of σ^2 based on a χ^2 distribution with m degrees of freedom is available, the likelihood ratio test rejects H_0 for large values of

$$\bar{E}^2_{k,p} = \sum_{v=1}^{k}\left(\theta_v^* - \bar{X}\right)' \Gamma_v^{-1}\left(\theta_v^* - \bar{X}\right) \Big/ \left\{\sum_{v=1}^{k}\left(X_v - \bar{X}\right)' \Gamma_v^{-1}\left(X_v - \bar{X}\right) + mS\right\},$$

where θ_v^* is the MIR of x_v with weights Γ_v^{-1}. If such an S is not available, i.e. only $X_1, \ldots, X_k$ are available, the likelihood ratio test statistic based on the joint distribution of $X_1, \ldots, X_k$ is obtained by putting $mS=0$ in $\bar{E}^2_{k,p}$.

Null Distribution of $\bar{\chi}^2_{k,p}$ when covariance matrices are known

The testing problem can be converted to the multivariate analogue of the one-sided test of Kudo (1963) or to its generalization of Kudo and Choi (1975), but the null distribution of the test statistic cannot be computed exactly except for some special situations. According to the theory of multivariate analogue of one sided test, the null distribution of $\overline{\chi}^2_{k,p}$ is a weighted sum of χ^2 distribution functions in the case of simple order alternative. The main problem arises here in the computation of weights. But according to Sun (1988) the upper tail probabilities of $\overline{\chi}^2_{k,p}$ can be computed only for the values of k and p *such* that $(k-1)p \leq 9$. However, Kulatunga and Sasabuchi (1984) showed that, when Λ_i's are diagonal each of the weights is expressed as a convolution of the weights of the null distribution of Bartholomew's (1959) $\overline{\chi}^2_k$ statistic. In addition if Λ_i's are all equal, the weights are computed by a recurrence formula given and tabulated in Kulatunga (1984). Hence when Λ_i's are all equal and diagonal, the null distribution of $\overline{\chi}^2_{k,p}$ and thus it's upper tail percentage points can be computed exactly. The above results are summarized as follows:

When Λ_v's are diagonal with diagonal elements $d_{v1}, \ldots, d_{vp}$, $v = 1, \ldots, k$, the likelihood ratio statistic is of the form

$$\overline{\chi}^2_{k,p} = \Sigma_{i=1}^{p} \Sigma_{v=1}^{k} d_{vi}^{-1} \left(\hat{\theta}_{vi} - \overline{X}_i\right)^2, \tag{3}$$

where $\overline{X}_i = \left(\Sigma_{v=1}^{k} d_{vi}^{-1}\right)^{-1}\left(\Sigma_{v=1}^{k} d_{vi}^{-1} X_{vi}\right)$ and $(\hat{\theta}_{1i}, \ldots, \hat{\theta}_{ki})$ is the univariate isotonic regression of $(X_{1i}, \ldots, X_{ki})$ with weights $d_{1i}^{-1}, \ldots, d_{ki}^{-1}$, $i = 1, \ldots, p$.

THEOREM 3. Let Λ_v's be common and diagonal and the alternative be $H_1': \theta_1 \leq \ldots \leq \theta_k$.

If H_0 is true, then for $c > 0$,

$$P(\overline{X}^2_{k,p} \geq c = \sum_{l=p+1}^{kp} Q(l,k,p) P(X^2_{l-p} \geq c) \tag{4}$$

and $\qquad P\left(\overline{\chi}^2_{k,p} = 0\right) = k^{-p},$

where χ^2_{l-p} denotes the chi square random variable with $l-p$ degrees of freedom and $Q(l,k,p)$ is the convolution of probabilities $P(l,k)$ used in order restricted inference. (cf. Kulatunga, 1984)

A computer program for computing $Q(l,k,p)$ is given by Shi *et al.* (1988).

As stated above, it is observed that the covariance structure other than the independence with equal covariance matrices causes gave problems in computing the null distribution of $\bar{\chi}^2_{k,p}$. Thus Sasabuchi *et al.* (1998) proposed the following three test procedures and studied them by extensive simulations when the dimension is 2 and the alternative hypothesis is the simple order.

Test 1 Ignore all the off- diagonal elements of Λ_υ ($\upsilon = 1, \ldots, k$) and replace them with zero. In other words we construct the test as if independence was assumed in the multivariate variables. Compute the statistic (3) and use the right hand side of (5.2) as an approximate evaluation of significance probability.

Test 2 Compute the statistic $\bar{\chi}^2_{k,p}$, but the right hand side of (4) is used as an approximate evaluation of significance probability. That is we use the original likelihood ratio test statistic, but choose critical points as if independence in the multivariate variables was assumed.

Test 3 Use the true likelihood ratio test statistic $\bar{\chi}^2_{k,p}$ and its critical points estimated by simulation.

One reason for proposing T1 and T2 is that the convenience in computation. But an important theoretical reason can also be given. In the univariate case of the problem considered here, Robertson and Wright (1983) have shown that the equal weight approximation is satisfactory for the null distribution of $\bar{\chi}^2_k$ statistic except in extreme cases. The reason for this is that the level probabilities $P(l,k)$ are fairly robust with respect to the weights. The null distribution of $\bar{\chi}^2_{k,p}$ statistic is also expressed as a weighted sum of χ^2 distributions and the weights correspond to the $P(l,k)$'s in the univariate case. On the analogy of the univariate case, we can expect that these weights are also rather robust with respect to the covariance matrices. This consideration motivates the authors to set covariances equal to zero and to take common (equal) diagonal matrices case as an approximate distribution.

In the simulation study conducted by generating 100,000 sets of k bivariate normal random numbers with mean vectors (0,0), unit variance and correlation coefficients ρ, Sasabuchi *et al.*, (1998) estimated the effective sizes of T1 and T2, and critical points of T3 and conducted a power comparison of the 3 test procedures . They have observed that the effective sizes of T1 and T2 are larger than the given nominal significance level when $\rho > 0$, and smaller than the nominal significance level when $\rho < 0$. Thus, when $\rho < 0$, T1 and T2 are conservative and we can compare the powers of these tests only when $\rho < 0$.

Power companion was done by selecting three directions to fully represent this alternative region. The alternative region in this problem is a convex polyhedral cone in the 2k-diemensional Euclidian space. The first direction is near the center of the cone, the third direction is near its edge, and the second is between these two. In the univariate case, Robertson *et al.* (1998) suggested that maximum and minimum powers over H_1 of $\bar{\chi}^2_k$ test occur at the above first and third directions respectively. Similar observations can be expected in the multivariate case too. The results of the simulation showed that

(i) T3 has the highest power near the center of the alternative and the lowest power near its edge. Also the power of T3 is rather high and stable over H_1' and uniformly much higher than that of the usual χ^2 statistic used in unrestricted alternative hypothesis. (ii) T3 is uniformly most powerful of the these test procedures and (iii) The power of T1 is uniformly lower than that of T2. When ρ is small, the powers of T1 are even lower than that of usual χ^2 test (iv) The powers of T2 when $\rho < 0$, and T3 are considerably higher than those of χ^2 test. Also there is no significant difference in powers of T2 and T3.

Thus T3 performs best, but it is based on the critical points of $\bar{\chi}^2_{k,p}$ estimated by simulation. It is rather unpalatable to expect researchers to set up the machinery to perform simulations in general to find critical points. Based on the results of the simulation study the following recommendations were made:

(a) T1 cannot be recommended.
(b) When $\rho < 0$ they have recommended T2 for practical use because it is simpler than the likelihood ratio test while the overall power is satisfactory.
(c) When the correlation coefficient is positive, T3 is recommended.

This idea is quite similar to those of Robertson and Wright (1982, 1983) where in the univariate problem they consider equal weight approximation or probability bounds of $\bar{\chi}^2_k$ statistic for unequal weights case.

The simulation study considered in Sasabuchi *et al.* (1998) is modest in the sense that only the bivariate case has been examined there. In general compared to the bivariate case, higher dimensional cases often have more difficulties and complexities, and sometimes show different phenomenon. The same test procedures have again been studied by extensive simulation by Sasabuchi *et al.* (2003) and they obtained a few findings which are different from those of the bivariate case. In fact the above findings mentioned in (i), (ii) and (iii) for the bivariate case are quite similar in the general case. But in the multivariate case the power of T2 is uniformly higher than that of the χ^2 test, but much lower than that of T3. Also it has been noted that the difference between the power of T2 and that of T3 becomes larger as the dimension p increases. Because of this reason Sasabuchi *et al.* (2003) suggested that we should not use T2 as an alternative to the true likelihood ratio test T3 when the dimension is greater than or equal to 3. Although T3 needs to estimate its critical points by simulation, this might not be a serious problem for modern day computers with high speed CPU.

Testing Problem when Covariance Matrices are Common and Unknown

When the covariance matrices are known the problem of testing the homogeneity of several normal mean vectors against some order alternatives has been studied to some extent in the literature. Recently Sasabuchi et al. (2003) considered the same problem where the covariance matrices are assumed to be common but unknown. In the univariate case, this problem has been studied by Bartholomew (1961) and many others, and the likelihood ratio test statistic is well known as $\overline{E}_k^2$. Sasabuchi and Kulatunga (1985) showed that moment approximation of the null distribution of $\overline{E}_k^2$ statistic can be used in most practical applications.

The problem and the proposed test statistics

Suppose $X_{i1}, \ldots, X_{iN_i}$ are random samples from a p-variate normal distribution $N_p(\mu_i, \Sigma), i = 1, \ldots, k,$ where Σ is assumed to be unknown and $N_1 + \ldots + N_k > p + k$.

When Σ is known, by Theorem 1, the critical region of the likelihood ratio test for H_o: $\theta_1 = \cdots = \theta_k$ verses H_1': $\theta_1 \leq \ldots \leq \theta_k$ is given by

$$\overline{\chi}^2_{k,p} = \sum_{I=1}^{k} N_i \left(\hat{\theta}_i - \overline{X}\right) \Sigma^{-1} \left(\hat{\theta}_i - \overline{X}\right) \geq c_1,$$

where $(\hat{\theta}_1, \ldots, \hat{\theta}_k)$ is the MIR of $\overline{X}_1, \ldots, \overline{X}_k$ with weights $N_1\Sigma^{-1}, \ldots, N_k\Sigma^{-1}$, c_1 is a positive constant depending on the significance level, $\overline{X}_i = N_i^{-1} \sum_{j=1}^{N_i} X_{ij}$, $i = 1, \ldots, k$, and $\overline{X} = \left(\sum_{I=1}^{k} N_i\right)^{-1} \sum_{i=1}^{k} N_i \overline{X}_i$

In the present problem Σ is unknown and the likelihood ratio test for H_0 versus H_1' has not been obtained yet. Replacing the unknown covariance matrix in the $\bar{\chi}^2_{k.p}$ statistic by its estimator

$S = \sum_{i=1}^{k}\sum_{j=1}^{N_i}(X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'$. Sasabuchi *at el.* (2003) proposed the following test with a mathematical justification of the plausibility of this statistic:

$$\bar{T}^2 = \sum_{i=1}^{k} N_i (\hat{\theta}_i - \bar{X})' S^{-1}(\hat{\theta}_i - \bar{X}) \geq c_2 \quad \Longrightarrow \quad \text{reject } H_0,$$

where $(\hat{\theta}_1, \ldots, \hat{\theta}_k)$ is the MIR of $\bar{X}_1, \ldots, \bar{X}_k$ with weights $N_1 S^{-1}, \ldots, N_k S^{-1}$, and c_2 is a positive constant depending on the significance level.

To conduct the test using $\bar{T}^2$ it is required to compute the supremum of its upper tail probability under H_0, that is,

$$\sup_{\Sigma} \sup_{H_0} P_{\mu,\Sigma}(\bar{T}^2 \geq c)$$

arbitrary constant c. Here $P_{\mu,\Sigma}$ denotes the probability measure corresponding to the parameters $\theta = (\theta_1, \ldots, \theta_k)$ and Σ, sup_{H_0} denotes the supremum for $\theta_1, \ldots, \theta_k$ with $\theta_1 = \cdots = \theta_k$, and sup_Σ denotes the supremum for all $p \times p$ positive definite real matrices.

They have proved that for any real number c,

$$\sup_{\Sigma} \sup_{H_0} P_{\mu,\Sigma}(\bar{T}^2 \geq c) = P_{0,I_p}(T^* \geq c),$$

where

$$T^* = \sum_{i=1}^{k} N_i(\bar{X}_i - \bar{X})' S^{-1}(\bar{X}_i - \bar{X}) - \frac{1}{s_{11}} \sum_{i=1}^{k} N_i (\bar{X}_{i1} - \hat{\theta}_{i1})^2,$$

$\bar{X}_{i1}$ is the first component of $\bar{X}_i$, $i = 1, 2, \ldots, k$, s_{11} is the $(1,1)$th element of S, and $(\hat{\theta}_{11}, \ldots, \hat{\theta}_{k1})$ is the MIR of $\bar{X}_{11}, \ldots, \bar{X}_{k1}$ with weights $N_1, \ldots, N_k$.

Though the exact distribution of T^* is not known, the approximate value of $P_{0,I_p}(T^* \geq c)$ can be easily obtained using Monto Carlo simulation generating standard normal random numbers. Upper α points estimated by simulation for various values of p and k when $\alpha = 0.01, 0.05$ are given in Table 1 of Sasabuchi *et al.* (2003).

SOME APPLICATIONS OF MULTIVARIATE ISOTONIC REGRESSION.

Two types of numerical examples are given to illustrate the applications of MIR in Fernando and Kulatunga (2007).

Example 1 (This example is taken from *p*.257 of Johnson and Wichern, 1998)

The Wisconsin Department of Health and Social Services reimburses nursing homes in the state for the services provided. Nursing homes in the state are classified into 3 groups: private, non-profit and government. A study was conducted to investigate the effects of 3 groups on costs. Four costs, computed on a per-patient-day basis and measured in hours per-patient-day were selected for analysis: cost of nursing labour ($i=1$), cost of dietary labour ($i=2$), cost of plant operation and maintenance labour ($i=3$), cost of housekeeping and laundry labour ($i=4$).

The simultaneous homogeneity of mean vectors against an increasing trend in three independent groups of data is considered here.

Summary statistics for each of the 3 groups (types) are given below.

Group	Number of observations	Sample mean vectors	
ν = Year 1	n_1=271	$\underline{x}_1= \begin{bmatrix} 2.066 \\ 0.480 \\ 0.082 \\ 0.360 \end{bmatrix}$	$\underline{x}_2= \begin{bmatrix} 2.167 \\ 0.596 \\ 0.124 \\ 0.418 \end{bmatrix}$
ν = Year 2	n_2=138	$\underline{x}_3= \begin{bmatrix} 2.273 \\ 0.521 \\ 0.125 \\ 0.383 \end{bmatrix}$	
ν = Year 3	n_3=107		

Sample covariance matrices

$$S_1 = \begin{bmatrix} 183.678 & 116.951 & 79.750 \\ 116.951 & 249.093 & 73.872 \\ 79.750 & 73.872 & 93.414 \end{bmatrix}, S_2 = \begin{bmatrix} 207.228 & 117.880 & -21.613 \\ 117.880 & 227.975 & -13.454 \\ -21.613 & -13.454 & 224.221 \end{bmatrix},$$

$$S_3 = \begin{bmatrix} 304.630 & 116.330 & 78.456 \\ 116.330 & 278.237 & -16.237 \\ 78.456 & -16.237 & 152.829 \end{bmatrix}$$

By examining the results one may suspect an increasing trend in average costs of the four variables in the three groups. Thus one we would be interested in testing $H_0: \underline{\theta}_1 = \underline{\theta}_2 = \underline{\theta}_3$ versus $H_1: \underline{\theta}_1 \leq \underline{\theta}_2 \leq \underline{\theta}_3$. That is, to see whether there are no group effects or equivalently no difference in average costs of four variables among the 3 groups against an increasing trend. Maximum likelihood estimates of $\underline{\theta}_1, \underline{\theta}_2, \underline{\theta}_3$ under H_1 are the MIR of $\underline{x}_1, \underline{x}_2, \underline{x}_3$ with weights s_1, s_2, s_3. Steps of the algorithm for MIR are given in Table 1. This explains the steps of the algorithm in the 4-diemensional case. Here 7 cycles of approximations are required for the convergence. Each cycle of recurrent computations of the algorithm consist s of 4 substeps; one revises one row vector while the other there are tentatively fixed. In each substep, univariate isotonic regression with weights computed using the entries of s_1, s_2, s_3 is applied on the row vectors of the data $\underline{x}_1, \underline{x}_2, \underline{x}_3$ recursively.

Example 2

It is expected in general that the students in the universities would show a gradual improvement in performance in each subject they follow year by year. To test this we have a taken random sample of students at the University of Kelaniya, who followed continuous course units in Pure Mathematics ($i=1$), Physics $i = 2$), Statistics ($i=3$), during the last three academic years.

In this example the simultaneous homogeneity of mean vectors against an increasing trend of a single group at 3 different time points is considered. Summary results are as follows.

Group	Sample mean vectors
v = Year 1 v = Year 2 v = Year 3	$\underline{x}_1 = \begin{bmatrix} 47.711 \\ 44.750 \\ 49.263 \end{bmatrix}$ $\underline{x}_2 = \begin{bmatrix} 45.431 \\ 43.650 \\ 45.530 \end{bmatrix}$. $\underline{x}_3 = \begin{bmatrix} 55.770 \\ 49.09 \\ 55.03 \end{bmatrix}$

Sample covariance matrices

$$S_1 = \begin{bmatrix} 183.678 & 116.951 & 79.750 \\ 116.951 & 249.093 & 73.872 \\ 79.750 & 73.872 & 93.414 \end{bmatrix}, S_2 = \begin{bmatrix} 207.228 & 117.880 & -21.613 \\ 117.880 & 227.975 & -13.454 \\ -21.613 & -13.454 & 224.221 \end{bmatrix}, S_3 = \begin{bmatrix} 304.630 & 116.330 & 78.45 \\ 116.330 & 278.237 & -16.2 \\ 78.456 & -16.237 & 152.82 \end{bmatrix}$$

Maximum likelihood estimates under the restriction that the marks follow an upward trend simultaneously in each subject year by year are obtained by the algorithm. After 13 iterations, the algorithm converges to the MIR and the values are given below.

$\theta^{1(13)}$	46.065	46.075	55.770
$\theta^{2(13)}$	43.449	43.998	49.090
$\theta^{3(13)}$	47.903	47.903	55.030

The steps are illustrated in the Figure 1.

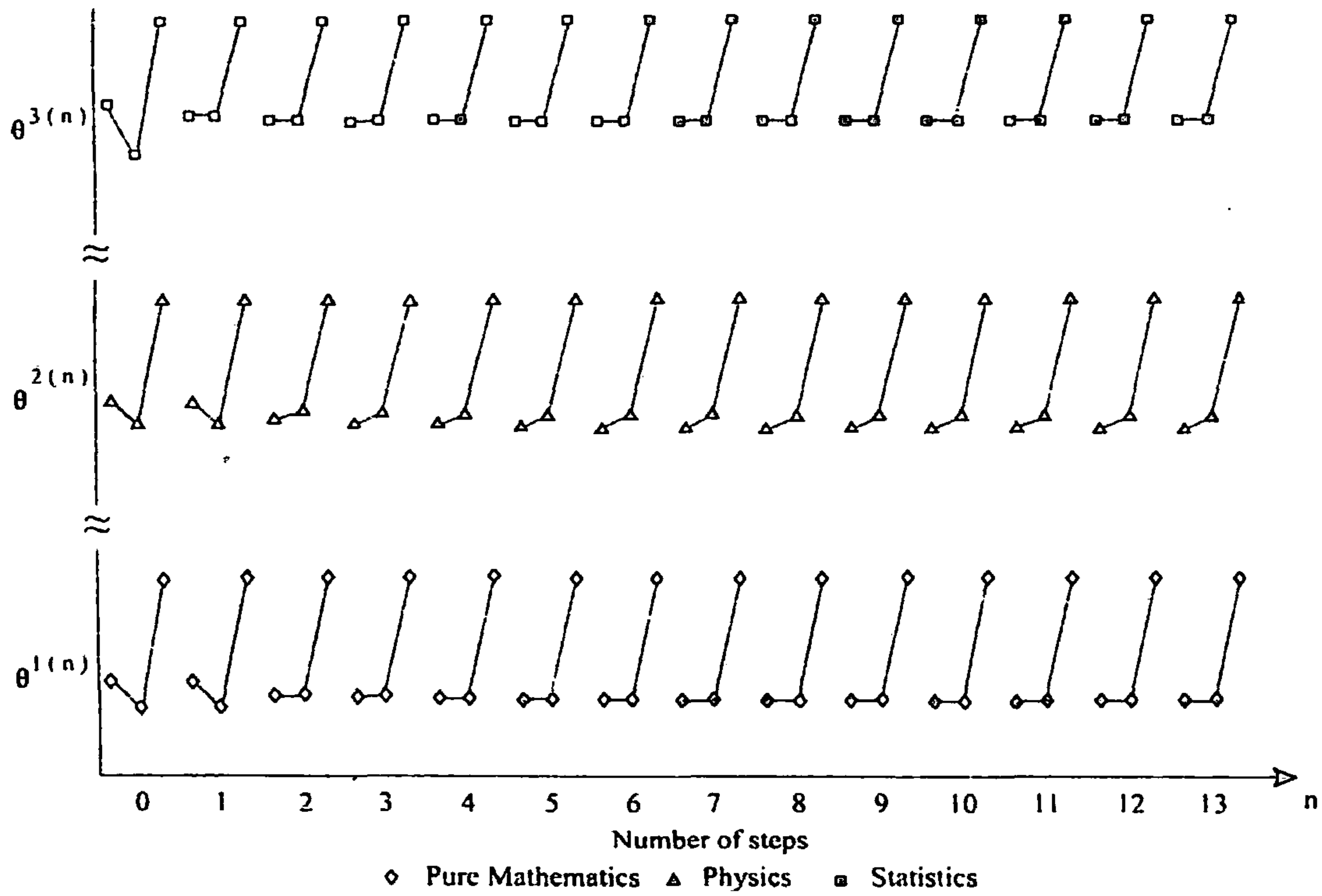

Figure 1. Graphical illustration of the steps

Table 1: Illustration of the steps of the algorithm

$\theta^{1(1)}$	2.066	2.167	2.273
$\theta^{2(1)}$	0.480	0.596	0.521
$\theta^{3(1)}$	0.082	0.124	0.125
$\theta^{4(1)}$	0.360	0.397	0.397
$\theta^{1(2)}$	2.066	2.126	2.282
$\theta^{2(2)}$	0.480	0.551	0.551
$\theta^{3(2)}$	0.082	0.116	0.125
$\theta^{4(2)}$	0.360	0.397	0.397
$\theta^{1(3)}$	2.066	2.126	2.282
$\theta^{2(3)}$	0.480	0.548	0.548
$\theta^{3(3)}$	0.548	0.116	0.125
$\theta^{4(3)}$	0.360	0.396	0.396
$\theta^{1(4)}$	2.066	2.130	2.327
$\theta^{2(4)}$	0.480	0.551	0.551
$\theta^{3(4)}$	0.082	0.116	0.125
$\theta^{4(4)}$	0.360	0.398	0.398
$\theta^{1(5)}$	2.066	2.129	2.324
$\theta^{2(5)}$	0.480	0.551	0.551
$\theta^{3(5)}$	0.082	0.116	0.125
$\theta^{4(5)}$	0.360	0.398	0.398
$\theta^{1(6)}$	2.066	2.133	2.328
$\theta^{2(6)}$	0.480	0.551	0.551
$\theta^{3(6)}$	0.082	0.116	0.125
$\theta^{4(6)}$	0.360	0.398	0.398
$\theta^{1(7)}$	2.066	2.132	2.329
$\theta^{2(7)}$	0.480	0.552	0.552
$\theta^{3(7)}$	0.082	0.117	0.125
$\theta^{4(7)}$	0.360	0.398	0.398

DEDICATION

This paper is dedicated to the memory of late Professor Akio Kudo who passed away in February 2003.

REFERENCES

Abelson, R.P. and J.W. Tukey (1963). Efficient utilization of non-numerical information in quantitative analysis: General theory and the case of simple order. *Annals of Mathematical Statistics*, 34: 1347- 1369.

Barlow, R.E., D.J. Bartholomew, J.M. Bremner and H.D. Brunk (1972). *Statistical inference under Order Restrictions. The Theory and Application of Isotonic Regression.* Wiley, New York.

Bartholomew, D.J. (1959a). A test of homogeneity for ordered alternatives. *Biometrika.* 46: 36-48.

Bartholomew, D.J. (1959b). A test of homogeneity for ordered alternatives II. *Biometrika.* 46: 328-335.

Bartholomew, D.J. (1961). A test of homogeneity of means under restricted alternatives (with discussion). *Journal of Royal Statistical Society,* B23: 239-281.

Cran, G.W. (1980). Algorithm AS 149, Amalgamation of means in case of simple ordering. *Applied Statistics,* 29: 209-211.

Fernando W.T.P.S. and D.D.S. Kulatunga (2007). On the computation and some applications of multivariate isotonic regression. *Computational Statistics and Data Analysis,* 52: 702-712.

Geng, Z. and N.Z. Shi (1990). Isotonic regression for umbrella ordering. *Applied Statistics,* 39: 397-402.

Hansohm, J. (2007) Algorithms and error estimations for monotone regression on partially preordered sets. *Journal of multivariate Analysis,* 98: 1043-1050.

Hansohm, J. and .X.M. Hu. (2008) Merge and chop in the computation for isotonic regression, *Journal of Statistical Planning and Inference,* 165-170.

Johnson, R. A. and D.W. Wichern (1998). *Applied Multivariate Statistical Analysis.* Prentice Hall Inc. N.J.

Kulatunga, D.D.S. (1984). Convolutions of the probabilities *P(l,k)* used in order restricted inferences. *Mem. Fac. Sci. Kyushu Unive. Ser. A. (Math),* 38: 9-15.

Kulatunga, D.D.S. and S. Sasabuchi (1984). A test of homogeneity of mean vectors against multivariate isotonic alternatives. *Mem. Fac. Sci. Kyushu Unive. Ser. A. (Math)*, 38: 151-161.

Kudô, A. (1963).A multivariate analogue of the one-sided test. *Biometrika*, 50: 403-417.

Kudô, A. and J.R. Choi (1975) A Generalized Multivariate Analogue of the One-Sided Test. *Memoirs of the Faculty of Science, Kyushu Univ. Ser. A.(Math.)* 29: 303-28.

Nomakuchi. K. and N.Z. Shi (1988). A test for a multiple isotonic regression problem. *Biometrika*, 75: 181-184.

Roberston, T., F.T. Wright and R.L. Dykstra (1988). *Order Restricted Statistical Inference.* Wiley, New York.

Sasabuchi, S., M. Inutsuka and D.D.S. Kulatunga (1983). A multivariate version of isotonic regression, *Biometrika*, 70: 465-472.

Sasabuchi, S. and D.D.S. Kulatunga (1985). Some approximations for the null distribution of the $\bar{E}^2$ statistic used in order restricted inference, *Biometrika*, 72: 476-480.

Sasabuchi, S., M. Inutsuka and D.D.S. Kulatunga (1992). An algorithm for computing multivariate isotonic regression, *Hiroshima Mathematical Journal*, 22: 551-560.

Sasabuchi, S., D.D.S. Kulatunga and H. Saito (1998). Comparison of powers of some tests for testing homogeneity under order restrictions in multivariate normal means. *American Journal of Mathematical and Management Sciences*, 18: 131-158.

Sasabuchi, S., T. Miura and H. Oda (2003). Estimation and test of several multivariate normal means under an order restriction when the dimension is larger than two. *Journal of Statistical Computation and Simulation.*, 73: 619-641.

Sasabuchi, S., K. Tanaka and T. Tsukamota (2003). Testing homogeneity of multivariate normal mean vectors under an order restriction when the covariance matrices are common but unknown. *Annals of Statistics*, 31: 1517-1536.

Schaafsma, W. L.J. and Smid (1966). Most stringent somewhere most powerful tests against alternatives restricted by a number of linear inequalities. *Annals of Mathematical Statistics*, 37: 1161-1172

Shi, N.-Z., A. Kudô, M. Fukugawa and T. Sakata (1988). Testing equality of proportions against trend and applications in study of aging effects. *Bulletin of the Biometric Society of Japan*, 9: 61-72.

Sun, H.J. (1988). A Fortran subroutine for computing normal orthant probabilities of dimensions up to nine. *Communications in Statistics. Simulation and Computation*, 17: 1097-1111.