

A New Test for Simple Tree Alternative in a $2 \times k$ Table

Parthasarathi Chakrabarti

*Department of Statistics, R.K.M. Residential College,
Narendrapur, Kolkata 700 103, India
partho.stat@gmail.com*

Uttam Bandyopadhyay*

*Department of Statistics, University of Calcutta
35 Ballygunge Circular Road, Kolkata 700 019, India
ubandyopadhyay08@gmail.com*

"Tgegkxgf "42"QevqdgT"4238

""Ceeegr vgf "53"O c{ "4239

This paper considers simple tree order restriction in $2 \times k$ cohort study and provides a consistent test in which the usual multiple comparison test statistics are modified by using the characteristic roots of a consistent estimator of the associated correlation matrix. The relevant performance measures of the proposed test are obtained and are compared numerically with existing competitors via simulation. It is shown that the proposed test is comparable to or better than the competitors in terms of type I error rate and power. Finally, data study illustrates the use of such a test.

Keywords: order restriction; simple tree; empirical size; empirical power; bootstrap.


2000 Mathematics Subject Classification: 62F30, 62G10

1. Introduction

Testing the equality of multiple mortality rates from different exposure categories against an ordered alternative occurs frequently in epidemiological studies. For example, consider the cohort study by Gupta and Mehta (2000) in which the age adjusted mortality rates among women in Mumbai, India using mishri (roasted, powdered form of tobacco used to clean teeth) and betel nut are, respectively, 12.3 and 12.6 per 1000 per annum, whereas such rate for control group is 9.9. Hence, it would be reasonable to assume the simple tree restriction $\pi_1 \leq \pi_2, \pi_3$, where π_1 , π_2 and π_3 represent, respectively, the risks of dying among women for the control group, for those who use mishri and for those who chew betel nuts. In general, if $H : \pi_1 = \pi_2 = \dots = \pi_k$ represents no restriction on mortality rates for k exposure categories, H can be tested against the patterned alternative $H_{st} - H$, where $H_{st} : \pi_1 \leq \pi_2, \pi_3, \dots, \pi_k$.

*corresponding author.

Several tests are available in the literature for testing H against $H_{st} - H$. These are, for example, based on restricted maximum likelihood estimator (RMLE), multiple comparison procedures and non parametric kernels (see, for example, Fligner and Wolfe, 1982; Magel, 1988; Desu et al., 1996). While detecting order restrictions on binomial probabilities based on a $2 \times k$ cohort study, multinomial allocation probabilities corresponding to the exposure levels play an important role. The existing tests to detect simple tree order restriction in a $2 \times k$ table, where allocation probabilities are unbalanced, occasionally fail to attain the nominal level for small values of π_1 . Our aim is to propose a multiple comparison consistent test using the characteristic roots of a consistent estimator of the associated correlation matrix based on the multinomial allocation probabilities, in which this short fall has been overcome.

Among the RMLE based approaches, the work on confidence interval estimation subject to order restriction (Hwang and Peddada, 1994) is based on modified generalised isotonic regression estimator (MGIRE). A number of testing procedures are obtained following MGIRE (see, for example, Peddada et al., 2001; Peddada and Haseman, 2006; Teoh et al., 2008). In this paper we choose an MGIRE based test as competitor and is referred to as the MGIRE test. Other RMLE based procedures to detect simple tree alternative are, for example, due to Wright and Tran (1985), Conaway et al. (1991), Singh et al. (1993), Futschik and Pflug (1998), Tsai (2004). Multiple comparison procedure (Bretz et al., 2001, 2003; Genz, 2004; Schaarschmidt et al., 2008; Hothorn et al., 2009), based on normal and binary responses, is proposed as a method in which the cut off points of the related tests are obtained from the distribution functions of multivariate normal and multivariate t distributions and are provided numerically through the R-packages *mnormt* and *mvtnorm*. In our setting we also choose one of such tests under binary response as another competitor and call the corresponding test as the GBH (Genz-Bretz-Hothorn) test. Besides these multiple comparison tests some single contrast tests are available to detect order restriction among binomial probabilities  see, for example, Leuraud and Benichou, 2001, 2004; Bretz and Hothorn, 2003; Bandyopadhyay and Chakrabarti, 2013 and the references there in). Our numerical computation shows that for small sample size the MGIRE and GBH tests often fail to attain the nominal level under unbalanced allocation as compared to that under balanced allocation. The proposed test overcomes such shortfall and increases its power locally.

The outline of the paper is as follows. Section 2 provides the data layout and notations. Section 3 contains some asymptotics and formulation of the proposed test. Section 4 describes competitors of the proposed test. Simulation results on size and power of the tests are given in Section 5. Section 6 contains data study. The paper concludes with some discussions in Section 7, followed by some technical details in Appendices A and B.

2. Data layout and notations

Consider a cohort study on n individuals, where the dichotomous response variable Y , indicating survival status, is recorded for the exposure X consisting of the levels x_1, x_2, \dots, x_k , measured in a nominal scale, satisfying $x_1 \preceq x_2, x_3, \dots, x_k$. Let $p_j = P(X = x_j) > 0$, the chance of occurrence of the exposure level x_j , $j = 1, 2, \dots, k$ with $\sum_{j=1}^k p_j = 1$, and $\pi_j = P(Y = 1|X = x_j) = 1 - P(Y = 0|X = x_j)$, the mortality rate at x_j , $j = 1, 2, \dots, k$. Define $n_j = \#(X = x_j)$ as the number of individuals observed at x_j and $s_j = \#(Y = 1|X = x_j)$ as the disease count at x_j , $j = 1, 2, \dots, k$, where $n = \sum_{j=1}^k n_j$.

Let us write $\mathbf{n}^T = (n_1, n_2, \dots, n_k)$, $\mathbf{p}^T = (p_1, p_2, \dots, p_k)$ and $\boldsymbol{\pi}^T = (\pi_1, \pi_2, \dots, \pi_k)$. Evidently, the distribution of \mathbf{n} is multinomial on k categories with index n and parameter \mathbf{p} . Further (s_1, s_2, \dots, s_k) ,

conditioning on \mathbf{n} , constitutes k –independent binomial random variables, where s_j follows binomial distribution with index n_j and parameter π_j , $j = 1, 2, \dots, k$. In order to understand the simple tree order of the mortality rates at different exposure levels, H is tested against $H_{st} - H$.

In the subsequent discussions, \hat{p}_j and $\hat{\pi}_j$ are used to denote, respectively, the observed proportions of individuals and successes at x_j , where $\hat{p}_j = n_j/n$ and $\hat{\pi}_j = s_j/n_j$. Then the overall proportion of success is obtained by $\hat{\pi} = \frac{1}{n} \sum_{j=1}^k n_j \hat{\pi}_j = \hat{\mathbf{p}}^T \hat{\boldsymbol{\pi}}$, where $\hat{\mathbf{p}}^T = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k)$ and $\hat{\boldsymbol{\pi}}^T = (\hat{\pi}_1, \hat{\pi}_2, \dots, \hat{\pi}_k)$. If n_j vanishes for some $j = 1, 2, \dots, k$, dirichlet prior is used to choose $\hat{p}_j = \frac{n_j+1/k}{n+1}$, $j = 1, 2, \dots, k$. Similarly, if $\hat{\pi}$ is found to be 0 or 1 for a specific sample, we choose $\hat{\pi} = \frac{\sum_{j=1}^k n_j \hat{\pi}_j + 1/2}{n+1}$ by use of beta prior.

3. Proposed test and related asymptotic results

A naive test, analogous to Dunnett's procedure (1955), can be constructed through Bonferroni's correction in which H is rejected at level α against $H_{st} - H$ if and only if

$$T = \max \left\{ t_j = \frac{\sqrt{n}(\hat{\pi}_j - \hat{\pi}_1)}{\sqrt{\hat{\pi}(1-\hat{\pi})(\frac{1}{\hat{p}_1} + \frac{1}{\hat{p}_j})}}, j = 2, 3, \dots, k \right\}$$

exceeds $\tau_{\alpha/(k-1)}$, where τ_{α} is the $(1 - \alpha)^{\text{th}}$ quantile of standard normal distribution, $0 < \alpha < 1$. Such a test is referred to as the T -test. In this paper a modification of the T -test is proposed by standardizing $\mathbf{t} = (t_2, t_3, \dots, t_k)^T$ through the estimators of the characteristic roots of the correlation matrix of \mathbf{t} . Towards such modification, H is expressed in terms of multiple contrasts of $\boldsymbol{\pi}$ by

$$H : C\boldsymbol{\pi} = \mathbf{0},$$

where $C^{(k-1) \times k} = (-\mathbf{1}_{k-1} \mathbf{e}_1 \mathbf{e}_2 \dots \mathbf{e}_{k-1})$ with \mathbf{e}_j , $j = 1, 2, \dots, k-1$ as $(k-1)$ component independent unit vectors and $\mathbf{1}_{k-1} = \sum_{j=1}^{k-1} \mathbf{e}_j$. Then, H is rejected against $H_{st} - H$ if and only if H_j is rejected against H_j^a for at least one j , where $H_j : \pi_1 = \pi_j$ and $H_j^a : \pi_1 < \pi_j$, $j = 2, 3, \dots, k$. Furthermore, an upper tail test based on t_j is appropriate for the testing problem (H_j, H_j^a) , $j = 2, 3, \dots, k$. Hence, combining all such component tests, the resulting test becomes the T -test. Now, we consider the proposed modification.

Modifying T :

It is not difficult to see that, for $0 < p_j < 1$, $j = 1, 2, \dots, k$, as $n \rightarrow \infty$,

$$\left\{ \frac{\sqrt{n}(\hat{\pi}_j - \pi_j)}{\sqrt{\frac{\hat{\pi}_j(1-\hat{\pi}_j)}{\hat{p}_j}}}, j = 1, 2, \dots, k \right\}$$

converges in distribution to $N_k(\mathbf{0}, I)$, the k -variate normal distribution with mean vector $\mathbf{0}$ and unit dispersion matrix I . That means, for large n ,

$$\sqrt{n}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \overset{a}{\sim} N_k(\mathbf{0}, \Sigma(\hat{\mathbf{p}})),$$

where

$$\Sigma(\hat{\mathbf{p}}) = \text{Diag} \left(\frac{\pi_1(1-\pi_1)}{\hat{p}_1}, \frac{\pi_2(1-\pi_2)}{\hat{p}_2}, \dots, \frac{\pi_k(1-\pi_k)}{\hat{p}_k} \right),$$

which under H reduces to $\pi(1 - \pi)\text{Diag}(\hat{p}_1^{-1}, \hat{p}_2^{-1}, \dots, \hat{p}_k^{-1})$, and the notation “ $u_n \stackrel{a}{\sim} v_n$ ” is used to mean that asymptotic distributions of the random variables u_n and v_n are same. Therefore, the asymptotic distribution of \mathbf{t} , shown in Appendix A, is $N_{k-1}(\mathbf{0}, R(\mathbf{p}))$, under H , where $R(\mathbf{p})$ is the correlation matrix with elements

$$r_{ij}(\mathbf{p}) = 1 \quad \text{if } 1 \leq i = j \leq k-1,$$

$$= \sqrt{\frac{p_{i+1}p_{j+1}}{(p_1 + p_{i+1})(p_1 + p_{j+1})}} \quad \text{if } 1 \leq i \neq j \leq k-1.$$

Let $\lambda_j = \lambda_j(\mathbf{p}) > 0$, $j = 1, 2, \dots, k-1$ be the characteristic roots of $R(\mathbf{p})$ and \mathbf{w}_j be the unit norm characteristic vector corresponding to λ_j , $j = 1, 2, \dots, k-1$. Then, setting $W = (\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_{k-1})$, it follows that

$$W^T R(\mathbf{p}) W = \Lambda(\mathbf{p}) = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_{k-1}).$$

Hence, there exists a positive definite matrix $R^{1/2}(\mathbf{p})$ for which, as $n \rightarrow \infty$,

$$R^{-1/2}(\mathbf{p})\mathbf{t} \rightarrow N_{k-1}(\mathbf{0}, I)$$

in distribution under H , where

$$R^{-1/2}(\mathbf{p}) = W \Lambda^{-1/2}(\mathbf{p}) W^T, \quad \Lambda^{-1/2}(\mathbf{p}) = \text{Diag} \left(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \dots, \frac{1}{\sqrt{\lambda_{k-1}}} \right).$$

Since, $R^{-1/2}(\hat{\mathbf{p}}) \rightarrow R^{-1/2}(\mathbf{p})$ in probability, we get, as $n \rightarrow \infty$,

$$\mathbf{t}_m = (t_2^m, t_3^m, \dots, t_k^m)^T = R^{-1/2}(\hat{\mathbf{p}})\mathbf{t} \rightarrow N_{k-1}(\mathbf{0}, I) \quad (1)$$

in distribution under H , and hence T can be modified by

$$T_m = \max\{t_2^m, t_3^m, \dots, t_k^m\}. \quad (2)$$

As usual, an upper tail test based on T_m would be appropriate. Such a test can be described by the critical region

$$w : T_m > T_{m\alpha}, \quad (3)$$

where, for given $\alpha : 0 < \alpha < 1$, $T_{m\alpha}$ is obtained approximately from the relation

$$\lim_{n \rightarrow \infty} P_H\{T_m \leq T_{m\alpha}\} = 1 - \alpha,$$

which, by use of (1) and (2), yields the approximate relation

$$\{\Phi(T_{m\alpha})\}^{k-1} = 1 - \alpha \quad (4)$$

with $\Phi(\cdot)$ as the distribution function of standard normal variable. Thus the test (referred to as the T_m -test), given by (3) and (4), is asymptotically level α test for the testing problem $(H, H_{st} - H)$ and is a modification of the T -test. It is shown (see Appendix B) that the test is consistent.

4. Competitors

MGIRE test

Here components of $\boldsymbol{\pi}$ are estimated (subject to a general order restriction) by

$$\tilde{\pi}_1 = \min \left\{ \frac{\sum_{j=1}^l \hat{p}_j \hat{\pi}_j}{\sum_{j=1}^l \hat{p}_j}, l = 1, 2, \dots, k \right\}$$

and

$$\tilde{\pi}_j = \max\{\tilde{\pi}_1, \hat{\pi}_j\}, j = 2, 3, \dots, k.$$

Then, incorporating Bonferroni's corrections, the test, described by the critical region

$$T_{MGIRE} = \max \left\{ \frac{\sqrt{n}(\tilde{\pi}_j - \tilde{\pi}_1)}{\sqrt{\hat{\pi}(1 - \hat{\pi})(\frac{1}{\hat{p}_j} + \frac{1}{\hat{p}_1})}}, j = 2, 3, \dots, k \right\} > \tau_{\alpha/(k-1)},$$

is asymptotically level α test for the testing problem $(H, H_{st} - H)$ and is used as a competitor of the proposed tests. From Appendix B, it is not difficult to see that under any $\boldsymbol{\pi}$,

$$\frac{1}{\sqrt{n}} T_{MGIRE} \rightarrow \max\{\theta_j, j = 1, 2, \dots, k-1\}$$

in probability, which is 'zero' or positive according as $\boldsymbol{\pi} \in H$ or $\boldsymbol{\pi} \in H_{st} - H$. This implies that the *MGIRE* test is consistent for testing H against $\boldsymbol{\pi} \in H_{st} - H$.

GBH test

Here H is rejected at level α against $H_{st} - H$ if and only if

$$T_{GBH} = \max \left\{ t_j = \frac{\sqrt{n}(\hat{\pi}_j - \hat{\pi}_1)}{\sqrt{\hat{\pi}(1 - \hat{\pi})(\frac{1}{\hat{p}_1} + \frac{1}{\hat{p}_j})}}, j = 2, 3, \dots, k \right\}$$

exceeds the $100(1 - \alpha)$ equi-percentage point, $c_{k-1, R(\hat{\mathbf{p}}), 1-\alpha}$, of $N_{k-1}(\mathbf{0}, R(\hat{\mathbf{p}}))$, the approximated null distribution of \mathbf{t} . The consistency of this test for testing H against $\boldsymbol{\pi} \in H_{st} - H$ can be established by the same technique as in the previous test.

5. Simulation study

We perform a simulation study with hundred thousand replications taking $k = 3$ and, for the purpose of illustration, the nominal level (α) is chosen at 0.05. The proposed test and the competitors are compared with respect to both empirical type I error rate and empirical power. Empirical type I error rate (power) of a test is computed by that proportion of hundred thousand replications of the experiment under H ($H - H_{st}$), in which the test statistic exceeds the 0.95th quantile of its asymptotic null distribution.

For a $2 \times k$ cohort data, setting $\mathbf{p} = \hat{\mathbf{p}}$ and the common success probability under H at $\pi = \hat{\pi}$, 100,000 tables, similar to the data, are generated. If the bootstrap percentile points of the simulated null distributions of the statistics agree with the percentile points of the asymptotic null distributions of the respective statistics, P-values of the tests are obtained using the approximate null distributions,

otherwise P-values are determined by bootstrapping (See, for example, Efron and Tibshirani (1993), Noreen (1989) and Romano (1988, 1989) for details) in which the proportion of cases the test statistics, evaluated from all such 100,000 tables, exceed the respective observed values obtained from the data set.

Similarly, if the empirical type I error rates do not agree with the nominal level, the powers of the corresponding test are evaluated using empirical cut-off point (0.95^{th} quantile of the simulated null distribution of the test statistic) instead of the approximate cut-off point.

The simulation study is performed for different choices of n and \mathbf{p} . For illustration, we choose $n = 100, 200, 300, 400$ and 500 for both balanced ($p_1 = p_2 = p_3$) and unbalanced situations. As most of the cohort studies indicate highly unbalanced situations, we take $\mathbf{p} = (0.9, 0.05, 0.05)$ (more allocation towards control) and $\mathbf{p} = (0.1, 0.45, 0.45)$ (less allocation towards control) for the present computation. For balanced allocation $\rho = r_{12}(\mathbf{p})$ is equal to 0.5 and for $\mathbf{p} = (0.9, 0.05, 0.05)$, $(0.1, 0.45, 0.45)$ ρ is, respectively, equal to 0.053 and 0.818 . $\boldsymbol{\pi}$ is chosen from $\{0.1, 0.3, 0.5\}$ in order to ensure the conformity of the type I error rates to the nominal level. The empirical powers of the tests are obtained under the following cases of the parametric configurations:

Case A: $\boldsymbol{\pi}$ lying in the boundary of the alternative region, such as: $\pi_1 = \pi_3 < \pi_2$.

Case B: $\boldsymbol{\pi}$ is well within the alternative region, such as: (B1) $\pi_1 < \pi_2 = \pi_3$, (B2) $\pi_1 < \pi_3 < \pi_2$.

For revealing the behaviour of the tests under *Case A*, we choose $\boldsymbol{\pi} = (0.1, 0.2, 0.1)$, and that under *Case B*, we choose $\boldsymbol{\pi} = (0.1, 0.2, 0.2)$ and $(0.1, 0.3, 0.2)$ for (B1) and (B2), respectively.

Simplification: $k = 3$.

Setting $\hat{\rho} = \sqrt{\frac{\hat{p}_2 \hat{p}_3}{(\hat{p}_1 + \hat{p}_2)(\hat{p}_1 + \hat{p}_3)}}$, we get $R(\hat{\mathbf{p}}) = \begin{pmatrix} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{pmatrix}$, which gives

$$\Lambda(\hat{\mathbf{p}}) = \begin{pmatrix} 1 - \hat{\rho} & 0 \\ 0 & 1 + \hat{\rho} \end{pmatrix} \text{ and } W = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix},$$

and hence

$$R^{-1/2}(\hat{\mathbf{p}}) = \frac{1}{2} \begin{pmatrix} \frac{1}{\sqrt{1+\hat{\rho}}} + \frac{1}{\sqrt{1-\hat{\rho}}} & \frac{1}{\sqrt{1+\hat{\rho}}} - \frac{1}{\sqrt{1-\hat{\rho}}} \\ \frac{1}{\sqrt{1+\hat{\rho}}} - \frac{1}{\sqrt{1-\hat{\rho}}} & \frac{1}{\sqrt{1+\hat{\rho}}} + \frac{1}{\sqrt{1-\hat{\rho}}} \end{pmatrix}.$$

Consequently, T_m becomes

$$T_m = \max \{a(\hat{\rho})t_2 + b(\hat{\rho})t_3, b(\hat{\rho})t_2 + a(\hat{\rho})t_3\},$$

where

$$a(\hat{\rho}) = \frac{1}{2} \left(\frac{1}{\sqrt{1+\hat{\rho}}} + \frac{1}{\sqrt{1-\hat{\rho}}} \right) \text{ and } b(\hat{\rho}) = \frac{1}{2} \left(\frac{1}{\sqrt{1+\hat{\rho}}} - \frac{1}{\sqrt{1-\hat{\rho}}} \right).$$

Result:

Computation of Type I error rate

In Table 1, the entries, showing maximum departure of the type I error rates from the nominal level (more than 10 %departure from the nominal level) for different choices of $\boldsymbol{\pi}$ and n , are marked in bold faces. The table shows that under balanced allocation and unbalanced allocation probabilities

(0.9, 0.05, 0.05) the T_m test and its competitors, except one exception, have similar behaviour. Again, unlike the T_m test, type I error rates of the MGIRE test do not agree with the nominal level under the allocation probabilities (0.1, 0.45, 0.45). However, in this situation, the GBH test maintains the nominal level except for small values of π . The more the increase in ρ , more is the deviation of the type I error rates for the MGIRE and GBH tests from the nominal level.

Table 1. Empirical type I error rate: T_m , MGIRE and GBH tests ($\alpha = 0.05$).

π	n	T_m $\mathbf{p} = (1/3, 1/3, 1/3)$	MGIRE	GBH	T_m $\mathbf{p} = (0.09, 0.05, 0.05)$	MGIRE	GBH	T_m $\mathbf{p} = (0.1, 0.45, 0.45)$	MGIRE	GBH
0.1	100	0.058	0.052	0.053	0.099	0.099	0.098	0.039	0.004	0.009
	200	0.055	0.050	0.052	0.083	0.083	0.083	0.404	0.014	0.026
	300	0.055	0.052	0.053	0.082	0.082	0.081	0.045	0.023	0.036
	400	0.053	0.050	0.050	0.077	0.077	0.076	0.045	0.029	0.040
	500	0.054	0.051	0.051	0.073	0.074	0.073	0.048	0.028	0.041
0.3	100	0.053	0.049	0.051	0.070	0.069	0.070	0.050	0.029	0.043
	200	0.050	0.049	0.051	0.063	0.064	0.063	0.047	0.033	0.043
	300	0.052	0.052	0.053	0.060	0.061	0.060	0.047	0.033	0.044
	400	0.049	0.049	0.050	0.061	0.061	0.061	0.050	0.034	0.046
	500	0.049	0.047	0.048	0.059	0.060	0.060	0.048	0.033	0.046
0.5	100	0.052	0.049	0.050	0.042	0.042	0.043	0.053	0.040	0.053
	200	0.049	0.048	0.049	0.048	0.048	0.048	0.050	0.039	0.051
	300	0.050	0.050	0.051	0.048	0.049	0.048	0.049	0.038	0.050
	400	0.049	0.046	0.047	0.052	0.052	0.052	0.052	0.040	0.051
	500	0.051	0.048	0.050	0.048	0.049	0.048	0.051	0.037	0.049

Computation of empirical power

Table 2 and Table 3 show, respectively, the empirical powers of the tests under *Case A* and *Case B*. For each of the given choices of \mathbf{p} , π and n maximum powers are marked in bold faces.

Case A:

Table 2. Empirical power: T_m , MGIRE and GBH tests ($\alpha = 0.05$, Case A).

π	n	T_m $\mathbf{p} = (1/3, 1/3, 1/3)$	MGIRE	GBH	T_m $\mathbf{p} = (0.09, 0.05, 0.05)$	MGIRE	GBH	T_m $\mathbf{p} = (0.1, 0.45, 0.45)$	MGIRE	GBH
(0.1, 0.2, 0.1)	100	0.261	0.230	0.240	0.116	0.115	0.116	0.280	0.240	0.243
	200	0.461	0.397	0.407	0.166	0.162	0.164	0.463	0.289	0.290
	300	0.623	0.549	0.561	0.292	0.290	0.292	0.620	0.384	0.384
	400	0.745	0.668	0.679	0.267	0.264	0.267	0.735	0.463	0.463
	500	0.832	0.764	0.772	0.307	0.306	0.307	0.813	0.535	0.537

Table 2 shows the empirical powers of all the tests for the given choices of π lying in a boundary of parametric space under $H_{st} - H$. For the given choices of n , the T_m test is found to be more powerful than the MGIRE and GBH tests under both balanced and unbalanced allocation probabilities. Based on this empirical power comparison, an approximate ordering of the tests is T_m , GBH , $MGIRE$, in which the T_m -test is the best in terms of having maximum empirical power.

Case B: Table 3 shows numerical computations of empirical power under both *Case B1* and *Case B2*. Here, under *Case B1*, the GBH test is found to be more powerful than the MGIRE and T_m

Table 3. Empirical power: T_m , MGIRE and GBH tests ($\alpha = 0.05$, Case B).

π	n	T_m	MGIRE	GBH	T_m	MGIRE	GBH	T_m	MGIRE	GBH
		$p = (1/3, 1/3, 1/3)$			$p = (0.09, 0.05, 0.05)$			$p = (0.1, 0.45, 0.45)$		
(0.1, 0.2, 0.2)	100	0.245	0.302	0.312	0.175	0.178	0.178	0.137	0.259	0.260
	200	0.422	0.507	0.521	0.265	0.268	0.270	0.206	0.311	0.312
	300	0.572	0.667	0.681	0.346	0.351	0.352	0.282	0.397	0.398
	400	0.683	0.775	0.788	0.404	0.409	0.411	0.347	0.492	0.493
	500	0.792	0.866	0.874	0.486	0.490	0.494	0.392	0.563	0.564
(0.1, 0.3, 0.2)	100	0.519	0.566	0.581	0.285	0.285	0.287	0.374	0.468	0.469
	200	0.810	0.847	0.854	0.444	0.448	0.450	0.629	0.634	0.635
	300	0.939	0.957	0.960	0.577	0.580	0.582	0.803	0.777	0.779
	400	0.980	0.989	0.990	0.683	0.688	0.690	0.903	0.879	0.880
	500	0.995	0.997	0.998	0.759	0.763	0.765	0.951	0.905	0.926

tests under both balanced and unbalanced allocation probabilities. Based on this empirical power comparison, like *Case A*, an approximate ordering of the tests is *GBH*, *MGIRE*, T_m . Under *Case B2* for balanced allocation probabilities and allocation probabilities (0.9, 0.05, 0.05) ordering of the tests with respect to empirical powers remains unaltered with an insignificant variation among the empirical powers. Under allocation probabilities (0.1, 0.45, 0.45) corresponding to *Case B2*, for $n \geq 300$, T_m test is found to be more powerful, whereas ordering of the tests remains same as in *Case B1* for $n = 100$ and 200.

6. Data study

Example 1:


The data, given in Table 4, are extracted from the cohort study (Gupta and Mehta, 2000) on the risk of mortality among tobacco users in Mumbai, India, 

Table 4. Mortality risk by use of mishri and betel nut among women.

category	frequency	\hat{p}	mortality risk ($\hat{\pi}$)
Control	64414	0.5225	0.0099
Mishri	56515	0.4585	0.0123
Betel nut	2343	0.0190	0.0126
total	123272	1	-

where users are classified gender-wise into smoking groups (smoking cigarette and bidi (tobacco hand rolled in temburni leaf and flaked)) and consuming smokeless tobacco (mishri, betel quid, betel nut, etc). Table 4 shows the risk of mortality among women in Mumbai from the use of mishri and betel nut. Here, the P-values of all the tests, proposed and competitors, are obtained by bootstrapping. In addition the bootstrap 0.95th percentile points of the simulated null distributions of such test statistics are obtained at various sample sizes. Furthermore, setting $p = \hat{p}$ and $\pi = \hat{\pi}$, another 100,000 tables are generated for those sample sizes. From each such tables the test statistics are computed. Finally, the powers of the tests are obtained as the proportions of cases in which such test statistics exceed the respective bootstrap percentile points. Estimated P-values and powers of the tests corresponding to Example 1 are given in Table 5.

Table 5. P-values and powers of the tests obtained by bootstrapping .

P-value/Power	T_m	$MGIRE$	GBH
P-value	0.00020	0.00025	0.00025
Power			
$n = 123272$	0.979	0.980	0.981
$n = 50000$	0.722	0.728	0.735
$n = 25000$	0.434	0.442	0.451
$n = 10000$	0.208	0.209	0.216
$n = 5000$	0.139	0.134	0.139
$n = 1000$	0.075	0.073	0.074

It is observed (Table 5) that all the tests, proposed and competitors, strongly reject the null hypothesis of no difference among the risks of mortality, where the T_m test has the least P-value. For different choices of n in Table 5 we see that empirical powers of the T_m , GBH and $MGIRE$ tests are approximately equal. For $n > 5,000$, an approximate ordering of the tests with respect to empirical power is GBH , $MGIRE$, T_m , in which the GBH -test is the best in terms of having maximum power. However, for $n \leq 5,000$, the ordering becomes T_m , GBH , $MGIRE$.

Example 2:

All the tests are applied to another data set (Graubard and Korn, 1987) relating to the effect of maternal alcoholism on congenital sex organ malformation among infants. The information on alcohol consumption is collected from would-be mothers after the first trimester and the malformations among infants are recorded following childbirth. Alcohol consumption categories are classified as average number of drinks per day. The data set is summarized in Table 6.

Table 6. Risk of infant's sex organ malformation for maternal alcoholism.

average number of drinks/day	frequency of mothers	\hat{p}	risk of malformation
< 1	31616	0.9706	0.0027
$1 - 2$	793	0.0243	0.0063
> 2	165	0.0051	0.0121
Total	32574	1	–

Adopting the similar technique, as used in Example 1, P-values and powers of the tests are determined and exhibited in Table 7.

Table 7. P-values and powers of the tests obtained by resampling.

P-value/Power	T_m	$MGIRE$	GBH
P-value	0.062	0.062	0.062
Power			
$n = 32574$	0.629	0.631	0.633
$n = 10000$	0.323	0.323	0.323
$n = 5000$	0.264	0.228	0.228
$n = 1000$	0.130	0.130	0.130

Table 6 shows that \hat{p}_1 is almost unity and \hat{p}_2 is significantly larger than \hat{p}_3 . Thus, the sample corresponds to an extremely unbalanced situation. Here, as the P-values suggest, all the T_m , GBH , $MGIRE$ tests strongly reject the null hypothesis. The T_m test is found to be more powerful than its competitors for $n \leq 10,000$.

7. Discussion

The failure of the type I error rate to attain the nominal level occurs more frequently in the $MGIRE$ and GBH tests than in the T_m test under unbalanced allocation probabilities. On the boundary of the parameter space under $H_{st} - H$, that is, under *Case A*, the T_m test is found to be locally more powerful than its competitors. Power of the T_m test in this case becomes significantly more as compared to that of its competitors with the increase in the value of ρ . Thus, for unbalanced allocation probabilities yielding high values of $r_{ij}(\mathbf{p})$'s, the T_m test can be preferred for its agreement of type I error rate with the nominal level.

References

- [1] A. Futschik and G. C. Pflug, The likelihood ratio test for simple tree order: A useful asymptotic expansion, *J. Stat. Plan. Inference* **70** (1998) 57–68.
- [2] A. Genz, Numerical computation of rectangular bivariate and trivariate normal and t-probabilities, *Stat. Comput.* **14** (2004) 251–260.
- [3] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, (Chapman & Hall, New York, 1993).
- [4] B. I. Graubard and E. L. Korn, Choice of column scores for testing independence in ordered 2 x K contingency tables, *Biometrics* **43** (1987) 471–476.
- [5] B. Singh, M. J. Schell and F. T. Wright, The power functions of the likelihood ratio tests for a simple tree ordering in normal means: Unequal weights, *Comm. Statist. Theory Methods A* **22** (1993) 425–449.
- [6] C. W. Dunnett, A multiple comparison procedure for comparing several treatments with a control, *Canad. J. Statist.* **50** (1955) 1096–1121.
- [7] E. Noreen, *Computer Intensive Methods for Testing Hypotheses: An Introduction*, (Wiley, New York, 1989).
- [8] E. Teoh, A. Nyska, U. Wormser and S. D. Peddada, Statistical inference under order restrictions on both rows and columns of a matrix, with an application in toxicology, *IMS Collections Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen* **1** (2008) 62–77.
- [9] F. Bretz, A. Genz and L. A. Hothorn, On the numerical availability of multiple comparison procedures, *Biom. J.* **43**(5) (2001) 645–656.
- [10] F. Bretz and L. A. Hothorn, Statistical analysis of monotone or non-monotone dose–response data from in vitro toxicological assays, *ATLA-ALTERN LAB ANIM* **31**(1) (2003) 81–96.
- [11] F. Schaarschmidt, M. Sill and L. A. Hothorn, Approximate simultaneous confidence intervals for multiple contrasts of binomial proportions, *Biom. J.* **50**(5) (2008) 782–792.
- [12] F. T. Wright and T. Tran, Approximating the level probabilities in order restricted inference: The simple tree ordering, *Biometrika* **71**(2) (1985) 429–439.
- [13] J. Hwang and S. Peddada, Confidence interval estimation subject to order restrictions, *Ann. Stat.* **22** (1994) 67–93.
- [14] J. Romano, A bootstrap revival of some nonparametric distance tests, *J. Amer. Statist. Assoc.* **83** (1988) 698–708.
- [15] J. Romano, Bootstrap and randomization tests of some nonparametric hypotheses, *Ann. Stat.* **17** (1989) 141–159.
- [16] K. Leuraud and J. Benichou, A comparison of several methods to test for the existence of a monotonic dose-response relationship in clinical and epidemiological studies, *Stat. Med.* **20** (2001) 3335–3351.

- [17] K. Leuraud and J. Benichou, Tests for monotonic trend from case-control data: Cochran-Armitage-Mantel trend test, isotonic regression and single and multiple contrast tests, *Biom. J.* **46** (2004) 731–749.
- [18] L. A. Hothorn, M. Vaeth and T. Hothorn, Trend tests for the evaluation of exposure-response relationships in epidemiological exposure studies, *Epidemiol Perspect Innov.* **6** (2009) 1.
- [19] M. A. Fligner and D. A. Wolfe, Distribution-free tests for comparing several treatments with a control, *Stat. Neerl.* **36** (1982) 119–127.
- [20] M. Conaway, C. Pillers, T. Robertson and J. Sconing, A circular-cone test for testing homogeneity against a simple tree order. *Canad. J. Statist.* **19** (1991) 283–296.
- [21] M. M. Desu, S. Park and S. Chakraborti, Linear rank statistics for the simple tree alternatives, *Biom. J.* **38**(3) (1996) 359–373.
- [22] M. T. Tsai, Maximum likelihood estimation of covariance matrices under simple tree ordering, *J. Multivar. Anal.* **89** (2004) 292–303.
- [23] P. C. Gupta and H. C. Mehta, Cohort study of all-cause mortality among tobacco users in Mumbai, India, *Bull. World Health Organ.* **78**(7) (2000) 877–883.
- [24] R. Magel, A test for the equality of k medians against the simple tree alternative under right censorship, *Comm. Statist. Simul. Comput. B* **17** (1988) 917–925.
- [25] S. D. Peddada and J. K. Haseman, Tests for a simple tree order restriction with application to dose-response studies, *Appl. Stat.* **55**(4) (2006) 493–506.
- [26] S. Peddada, K. Prescott and M. Conaway, Tests for order restrictions in binary response, *Biometrics* **57** (2001) 1219–1227.
- [27] U. Bandyopadhyay and P. Chakrabarti, Single contrast tests for detecting trends in binomial proportions, *J. Korean Statist. Soc.* **42** (2013) 235–246.

Appendix A

Asymptotic Distribution under H Setting

$$M(\hat{\mathbf{p}}) = \text{Diag} \left(\sqrt{\frac{\hat{p}_1 \hat{p}_2}{\hat{p}_1 + \hat{p}_2}}, \sqrt{\frac{\hat{p}_1 \hat{p}_3}{\hat{p}_1 + \hat{p}_3}}, \dots, \sqrt{\frac{\hat{p}_1 \hat{p}_k}{\hat{p}_1 + \hat{p}_k}} \right),$$

it follows that, under H , for large n ,

$$\sqrt{n}M(\hat{\mathbf{p}})C\hat{\boldsymbol{\pi}} \overset{a}{\sim} N_{k-1}(\mathbf{0}, \Sigma_H(\hat{\mathbf{p}})),$$

where

$$\Sigma_H(\hat{\mathbf{p}}) = M(\hat{\mathbf{p}})C\Sigma(\hat{\mathbf{p}})C^T M^T(\hat{\mathbf{p}}) = \pi(1 - \pi)R(\hat{\mathbf{p}})$$

with $R(\hat{\mathbf{p}})$ given by Section 3 when \mathbf{p} is estimated by $\hat{\mathbf{p}}$. Hence the statistic \mathbf{t} is identified as

$$\mathbf{t} = \sqrt{\frac{n}{\hat{\pi}(1 - \hat{\pi})}} M(\hat{\mathbf{p}})C\hat{\boldsymbol{\pi}}.$$

Now, using the fact that $\hat{\mathbf{p}} \rightarrow \mathbf{p}$ in probability, it follows that, under H ,

$$\mathbf{t} \rightarrow N_{k-1}(\mathbf{0}, R(\mathbf{p}))$$

in distribution as $n \rightarrow \infty$.

Appendix B

Consistency

First, we prove the following result.

Result B.1: Let $A = (\mathbf{a}_1 \mathbf{a}_2 \cdots \mathbf{a}_d)$ be a positive definite symmetric matrix, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_d)^T$ be a vector of non-negative elements with $\boldsymbol{\alpha} \neq \mathbf{0}$. Then

$$\max\{\mathbf{a}_j^T \boldsymbol{\alpha}, j = 1, 2, \dots, d\} > 0.$$

Proof: Assume that the assertion is false. Then, by the given conditions, we have $\boldsymbol{\alpha}^T A \boldsymbol{\alpha} \leq 0$. But this is a contradiction as A is positive definite. Hence the result follows.

Next, writing $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_{k-1})^T$ with

$$\theta_j = \frac{(\pi_{j+1} - \pi_j)}{\sqrt{\pi(1-\pi)}} \sqrt{\frac{p_1 p_{j+1}}{p_1 + p_{j+1}}}, j = 1, 2, \dots, k-1,$$

we can find a vector valued function $\mathbf{g} = (g_1, g_2, \dots, g_{k-1})^T$ of $\boldsymbol{\theta}$ such that as $n \rightarrow \infty$,

$$\frac{1}{\sqrt{n}} \mathbf{t}_m = \frac{1}{\sqrt{n}} R^{-\frac{1}{2}}(\hat{\mathbf{p}}) \mathbf{t} = \frac{1}{\sqrt{\hat{\pi}(1-\hat{\pi})}} R^{-\frac{1}{2}}(\hat{\mathbf{p}}) M(\hat{\mathbf{p}}) C \hat{\boldsymbol{\pi}}$$

converges to

$$\mathbf{g}(\boldsymbol{\theta}) = R^{-\frac{1}{2}}(\mathbf{p}) \boldsymbol{\theta}$$

in probability for any $\boldsymbol{\pi}$, where $M(\hat{\mathbf{p}})$ is defined in Appendix A. It is obvious that $\boldsymbol{\theta} = \mathbf{0}$ when $\boldsymbol{\pi} \in H$ and $\theta_j \geq 0, j = 1, 2, \dots, k-1$ with $\boldsymbol{\theta} \neq \mathbf{0}$ when $\boldsymbol{\pi} \in H_{st} - H$. Furthermore, as $n \rightarrow \infty$,

$$\frac{1}{\sqrt{n}} T_m \rightarrow \mu(\boldsymbol{\theta}) = \max\{g_j(\boldsymbol{\theta}), j = 1, 2, \dots, k-1\}$$

in probability, where $g_j(\boldsymbol{\theta}) = \mathbf{a}_j^T \boldsymbol{\theta}$ with \mathbf{a}_j as the j^{th} column of the symmetric positive definite matrix $R^{-\frac{1}{2}}(\mathbf{p})$. Hence, by use of Result B.1, we get

$$\begin{aligned} \mu(\boldsymbol{\theta}) &= 0 \text{ if } \boldsymbol{\pi} \in H \\ &> 0 \text{ if } \boldsymbol{\pi} \in H_{st} - H. \end{aligned}$$

This implies that the proposed test, described by (3), is consistent for testing H against any $\boldsymbol{\pi}$ under $H_{st} - H$.