Annals of the Institute of Statistical Mathematics (2024) 76:649–678
https://doi.org/10.1007/s10463-024-00897-7

Testing against ordered alternatives in one-way ANOVA model with exponential errors

Anjana Mondal[1] · Markus Pauly[2,3] · Somesh Kumar[1]

Received: 16 May 2023 / Revised: 28 December 2023 / Accepted: 10 January 2024 /
Published online: 12 March 2024

Abstract

In this paper, a one-way heteroscedastic ANOVA model is considered with exponentially distributed errors. The likelihood ratio test (LRT) and two multiple comparison tests are developed for testing against ordered alternatives. A parametric bootstrap (PB) approach is proposed for implementation of tests and its asymptotic accuracy is proved. An extensive simulation study shows that all the proposed tests are accurate in terms of achieving the nominal size value, even for small samples. The proposed simultaneous confidence intervals are also seen to maintain the preassigned coverage probability. The powers of these tests are compared with a recently proposed test, which is quite conservative. Finally, the proposed tests are illustrated with the help of three data sets related to medical studies. We have developed an 'R' package for implementing our test procedures and shared it on the open platform 'GitHub.'

Keywords Asymptotic accuracy · Likelihood ratio test · Multiple comparison tests · Parametric bootstrap

✉ Somesh Kumar
smsh@maths.iitkgp.ac.in

Anjana Mondal
mondal.anjana25@gmail.com

Markus Pauly
pauly@statistik.tu-dortmund.de

[1] Department of Mathematics, Indian Institute of Technology Kharagpur, Kharagpur 721032, India

[2] Department of Statistics, TU Dortmund University, 44227 Dortmund, Germany

[3] Research Center Trustworthy Data Science and Security, UA Ruhr, 44227 Dortmund, Germany

1 Introduction

One-way ANOVA is one of the most commonly used statistical tools in experimental studies, e.g., drug development, agricultural and genetic trials, product design in engineering, etc. The classical F test has been developed under the assumption of normal errors and homogeneous error variances. However, it has been observed that this test is not robust when errors are non-normal and/or group variances are heterogeneous. Donaldson (1968) studied in detail the size and power performance of this test where error distributions are exponential or lognormal. He carried out detailed numerical computations and showed the non-robustness for various cases. Testing equality of treatment effects in a two-way additive model under inverse Gaussian distributions has been studied by Fries and Bhattacharyya (1983). Şenoğlu and Tiku (2001) have considered a two-way ANOVA model with interactions when errors have non-normal distributions in a location-scale family. For testing the main and the interaction effects, they developed F-statistics using the method of modified maximum likelihood estimators. Among skewed distributions, the exponential distribution is one of the most widely used models with applications in reliability, queuing theory, survival analysis, etc. In a Poisson process, the time elapsed between two successive epochs, waiting time for an occurrence, etc. follow exponential distributions. In a dose-response experiment, the location parameter of an exponential distribution is used to represent the guaranteed survival time and scale parameter as the mean survival time in addition to the guaranteed survival time. Our aim in the present study is to develop tests in a one-way ANOVA with exponential error distributions, heteroscedasticity and unbalanced sample sizes against ordered alternatives.

Consider a completely randomized design (CRD) with k treatments and a total of $N = \sum_{i=1}^{k} n_i$ experimental units. Let $X_{i1}, X_{i2}, \ldots, X_{in_i}$ denote the observations on the ith treatment, $i = 1, 2, \ldots, k$, satisfying the following relation

$$X_{ij} = \mu_i + \epsilon_{ij}, \quad j = 1, \ldots, n_i \,; i = 1, \ldots, k. \tag{1}$$

In this model, μ_i denotes the effect of the ith treatment and ϵ_{ij}s are independently distributed errors. We take the error distribution to be exponential with location and scale parameters 0 and σ_i , respectively. That is, $\epsilon_{ij} \sim \exp(0, \sigma_i)$ for $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$, where σ_is are taken to be completely unknown. Then, the problem of testing the homogeneity of the treatment effects in the one-way ANOVA model (1) is equivalent to the problem of testing the equality of location parameters of k exponential populations when scale parameters are unknown.

In the traditional ANOVA, the alternative hypothesis is taken to be 'at least one inequality among treatment effects.' However, in some cases, it is reasonable to assume that there is a simple ordering among the treatment effects μ_i for $i = 1, \ldots, k$. For example, in dose-response experiments, an increased ingredient dose either increases survival probability and guaranteed effective duration or may have an insignificant increase (see Chuang-Stein and Agresti 1997). In quality control, new technologies are introduced so that the guaranteed lives of components increase and so the quality of products improves. In such situations, testing the hypothesis

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ against the ordered alternative $H_1 : \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$ with at least one strict inequality is appropriate and informative.

For the model described in (1), the errors are assumed to be exponentially distributed. Then, μ_is represent minimum guarantee times for the response variable in the ith group, $i = 1, 2, \ldots, k$. There are several significant applications, especially in clinical studies. Gill et al. (2017) have considered the data on the number of days of survival for the inoperable lung tumor patients classified into four groups such as Squamous, Small, Adeno, and Large—call these Type 1, 2, 3, and 4, respectively. Based on medical experience, it is expected that the minimum survival times for patients with these four types will be in increasing order. These data sets are shown to be exponentially distributed. Another interesting application is comparing the effectiveness of four different drugs for the treatment of leukemia patients. The data from the clinical trial reported in Wu and Wu (2005) show that the remission times are exponentially distributed. It is useful to find the relative efficiency of drugs by ranking their minimum remission times. In the investigation of the number of positive cases of COVID-19 patients, it is desirable that the minimum number of positive cases be in increasing order according to the population density of states. Here, also, the number of COVID-19 patients in specified states for a certain period is seen to follow exponential distributions. In Sect. 7, we have implemented the test procedures developed in this paper to all these data sets.

Testing against ordered alternatives for the normal error distribution in one-way ANOVA has been studied by many statisticians (see e.g., Barolow et al. 1972; Robertson et al. 1988; Hayter 1990; Marcus 1976; Liu et al. 2000 and the references therein). The cases of homogeneous and unknown or heterogeneous but known error variances have been investigated in detail. For normal models, the test procedures proposed by Williams (1971) and Williams (1972) are Max-T test and suggested by the National Toxicological Program. The procedures have been implemented as a multiple contrast test by Bretz (2006) and generalized to variance heteroscedasticity by Hasler and Hothorn (2008). For skewed distributions, Shirley (1997) proposed nonparametric tests equivalent to Williams (1971) and Williams (1972) and generalized to heteroscedasticity by Konietschke and Hothorn (2012). The general case of heterogeneous (unknown) error variances and unbalanced samples has been recently considered by Mondal et al. (2023). They proposed PB tests for testing H_0 against H_1 when error distributions are normal. However, these tests become either too conservative or too liberal when error distributions are taken to be exponential with heteroscedastic variances and group sample sizes are not large. Therefore, there is a need of considering the problem of testing against ordered alternatives with exponentially distributed errors. In this paper, we develop tests for this problem which are seen to achieve nominal size and attain high empirical power values under departure from the null hypothesis.

The problem of testing the equality of location parameters of exponential distributions against natural alternatives has been studied by several researchers (see, e.g., Chen 1982; Singh and Narayan 1983; Singh 1985 and Hsieh 1986). Of late, there has been an interest in this problem when ordered alternatives are considered. Dhawan and Gill (1997) have proposed simultaneous test procedures when scale parameters are common and known or common and unknown. Singh et al. (2006) have

proposed simultaneous test procedures for testing the equality of location parameters against natural as well as ordered alternatives. Maurya et al. (2011) proposed a test based on the restricted maximum likelihood estimators of parameters for testing against simple ordered as well as tree ordered alternatives. However, they considered unknown but equal scale parameters and balanced group sizes. Singh and Singh (2013) have developed simultaneous one-sided and two-sided confidence intervals for successive pairwise differences of location parameters using two-stage procedures. Gill et al. (2017) proposed a test procedure for the testing problem H_0 against H_1 when scaling parameters σ_is are unknown and unequal and data is unbalanced. They showed that their test is superior to one-stage and two-stage tests of Singh and Singh (2013). So, except the study of Gill et al. (2017), the existing literature on testing against ordered restricted location parameters of two-parameter exponential distributions deals with homogeneous scale parameters (known or unknown). Hence, in this article, we consider this problem with heterogeneous scale parameters and compare our proposed tests with the existing test of Gill et al. (2017).

This article is organized as follows. In Sect. 2, we develop the likelihood ratio test for testing H_0 vs. H_1. Subsection 2.1 presents the parametric bootstrap algorithm to find the critical value, power and size value of LRT. In this section, we also show the asymptotic accuracy of the parametric bootstrap for the LRT. Simultaneous tests Max-T and Min-T and the corresponding PB are presented in Sect. 3. Details of the parametric bootstrap algorithm for finding critical value, power and size value are given in Sect. 3.1. The proof of the asymptotic accuracy of PB used for Max-T and Min-T is also developed in this section. The Max-T test evaluates simultaneous confidence intervals for successive pairwise differences of location parameters in Sect. 4. A short detail of existing Gill's test (Gill et al. 2017) is provided in Sect. 5. Simulated size and power values of the PB tests and existing Gill's test are tabulated and analyzed in Sect. 6. Details of an 'R' package are given in Sect. 6. In Sect. 7, our tests are applied to three real data sets.

2 Likelihood ratio test

In this section, we develop the LRT for testing H_0 against H_1 in the one-way ANOVA model defined in Equation (1). Due to the fact that $\epsilon_{ij} \sim \exp(0, \sigma_i)$, the observations X_{ij}, satisfying the model (1) follow an $\exp(\mu_i, \sigma_i)$ distribution for $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, n_i$. We denote the parameter space under the hypothesis H_0 and the full parameter space by Ω_0 and Ω, respectively. These are defined as follows:

$$\Omega_0 = \{(\mu_1, \ldots, \mu_k, \sigma_1, \ldots, \sigma_k) : \mu_1 = \cdots = \mu_k = \mu(\text{ say}),\ \sigma_i > 0, i = 1, 2, \ldots, k\}$$

and

$$\Omega = \{(\mu_1, \ldots, \mu_k, \sigma_1, \ldots, \sigma_k) : \mu_1 \leq \cdots \leq \mu_k,\ \sigma_i > 0, i = 1, 2, \ldots, k\}.$$

Let us define $X_{i(1)} = \min_{1\le j\le n_i} X_{ij}$, $\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$, $i = 1, 2, \ldots, k$ and $X_{(1)} = \min_{1\le i\le k} X_{i(1)}$. Let $\hat{\mu}_{\Omega_0}$ and $\hat{\sigma}_{i\Omega_0}$ be the maximum likelihood estimators (MLEs) of μ and σ_i for $i = 1, 2, \ldots, k$ under Ω_0. Also let $\hat{\mu}_{i\Omega}$ and $\hat{\sigma}_{i\Omega}$ denote the MLEs of μ_i and σ_i for $i = 1, 2, \ldots, k$ under Ω. These are obtained as $\hat{\mu}_{\Omega_0} = X_{(1)}$, $\hat{\sigma}_{i\Omega_0} = \bar{X}_i - X_{(1)}$, $\hat{\mu}_{i\Omega} = \min(X_{i(1)}, \cdots, X_{k(1)})$, $\hat{\sigma}_{i\Omega} = \bar{X}_i - \hat{\mu}_{i\Omega}, i = 1, 2, \ldots, k$ (see Vijayasree et al. 1995). The likelihood ratio is then evaluated as

$$\Lambda = \prod_{i=1}^{k} \left(\frac{\hat{\sigma}_{i\Omega}}{\hat{\sigma}_{i\Omega_0}} \right)^{n_i}. \tag{2}$$

We reject H_0 at level α if $\Lambda < \lambda$, where λ is the α-quantile of Λ under H_0.

For $k \ge 3$, the exact null distribution of Λ cannot be obtained in a closed analytical form. Therefore, for determining critical values, we use the parametric bootstrap approach. In the following section, we provide an algorithm for finding the critical value, size and power values of the LRT using the classical PB method.

Remark 1 Note that when $k = 2$, the testing problem is a classical two sample problem with one-sided alternative. A reduced LRT was proposed by Epstein and Tsao (1953). However, this approach cannot be extended to more than two populations.

2.1 Parametric bootstrap scheme for LRT

Here, we use the classical PB, where we first choose an estimator of parameters and then samples are generated from the distribution, replacing parameters by their estimators. Then, the test statistic value is calculated from the bootstrap sample.

For convenience of notation, we use an unbiased estimator S_i of σ_i given by $S_i = \frac{n_i}{(n_i-1)}(\bar{X}_i - X_{i(1)}), i = 1, 2, \ldots, k$. It can be easily seen that under H_0, the distribution of the test statistic Λ is independent of location parameter. Hence, without loss of generality, we choose $\mu = 0$. Given the observations X_{ij}, we generate parametric bootstrap samples X^*_{ij} from $\exp(0, S_i)$, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$.

Define $\hat{\sigma}^*_{i\Omega_0} = \bar{X}^*_i - X^*_{(1)}$, $\hat{\sigma}^*_{i\Omega} = \bar{X}^*_i - \hat{\mu}^*_{i\Omega}$, where $X^*_{i(1)} = \min_{1\le j\le n_i} X^*_{ij}$, $\bar{X}^*_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X^*_{ij}$, $\hat{\mu}^*_{i\Omega} = \min(X^*_{i(1)}, \ldots, X^*_{k(1)})$, $i = 1, 2, \ldots, k$ and $X^*_{(1)} = \min_{1\le i\le k} X^*_{i(1)}$. Then, the parametric bootstrap likelihood ratio statistic is obtained as:

$$\Lambda^* = \prod_{i=1}^{k} \left(\frac{\hat{\sigma}^*_{i\Omega}}{\hat{\sigma}^*_{i\Omega_0}} \right)^{n_i}. \tag{3}$$

The α-quantiles of the distribution of Λ^* are then used as critical values for the LRT. They are usually derived by the Monte Carlo method. As we will later investigate size and power values in a simulation study, we already give a general algorithm for this, which contains the Monte Carlo procedure (Steps 2–4) for finding critical points:

Algorithm 1

Step 1. At the rth iteration of simulation, sample observations X_{ijr}, $j = 1, 2, \ldots, n_i$; $i = 1, 2, \ldots, k$ are generated from $\exp(\mu, \sigma_i)$, and the value of the likelihood ratio statistic Λ, as defined in (2), and call it Λ_r, is computed. Also calculate $S_{ir} = \frac{n_i}{(n_i-1)}(\bar{X}_{ir} - X_{ir(1)})$ from this sample.

Step 2. Now, generate parametric bootstrap samples X^*_{ijhr} from $\exp(0, S_{ir})$, $h = 1, \ldots, H; j = 1, \ldots, n_i; i = 1, \ldots, k$. Here, H denotes the number of simulations for the rth iteration in this inner bootstrap simulation loop.

Step 3. For each of these bootstrap iterations, say $h = s$, for the bootstrap sample $X^*_{ijsr}, j = 1, \ldots, n_i; i = 1, \ldots, k$, calculate the bootstrap likelihood ratio Λ^*_{rs}, defined in (3).

Step 4. Now, for each fixed r, after H runs of simulation, all values of Λ^*_{rs} are sorted as $\Lambda^*_{r(1)} \leq \cdots \leq \Lambda^*_{r(H)}$ and choose $\lambda^*_r = \Lambda^*_{r(\lfloor \alpha H \rfloor)}$ as the bootstrap critical value for the rth simulation run of the outer loop. This is also called the empirical α -quantile among the different values of $\Lambda^*_{r1}, \ldots, \Lambda^*_{rH}$.

Step 5. Repeat Steps 1–4 B number of times and then you will get B number of values of the test statistic, say, $\Lambda_1, \Lambda_2, \ldots, \Lambda_B$ and B number of critical values, say, $\lambda^*_1, \lambda^*_2, \ldots, \lambda^*_B$.

Step 6. Now the estimated size of the test is given by

$$\alpha_{LR} = \frac{\text{number of times } \Lambda_r \text{ less than } \lambda^*_r}{B} = \frac{1}{B}\sum_{r=1}^{B} \mathbf{1}\{\Lambda_r < \lambda^*_r\}.$$

Remark 2 The steps 2–4 for finding critical points are included in the above mentioned algorithm. For a given data set $X_{ij}, j = 1, \ldots, n_i; i = 1, \ldots, k$, one needs to find the values of S_i for $i = 1, \ldots, k$. Then generate bootstrap samples X^*_{ij} from $\exp(0, S_i)$ for $j = 1, \ldots, n_i; i = 1, \ldots, k$. After that, evaluate the bootstrap test statistic value Λ^*. For finding the critical point, all these steps are to be generated H (say, $H = 5000$) number of times. Then, evaluate the bootstrap critical point as the α-quantile of Λ^*, i.e., $\lambda^* = \Lambda_{\lfloor \alpha H \rfloor}$.

Remark 3 In Step 6, α_{LR} gives the size value of the LRT. If in Step 1, if X_{ij}'s are generated from $\exp(\mu_i, \sigma_i)$ such that μ_is are taken in the order $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$ with at least one inequality, then Step 6 provides the power value of LRT. From a given sample, the critical value can be calculated by following Steps 2-4.

To get a valid bootstrap test, we need to show that the conditional distribution of the bootstrap statistic Λ^* given the samples always converges to the null distribution of the likelihood ratio statistic Λ (see Pauly et al. 2015, 2016; Konietschke et al. 2015; Mondal et al. 2023, 2023). To establish this, we use the concept of Kullback–Leibler (KL) divergence between two distributions. The following lemma will be helpful in this regard.

Lemma 1 *Let* $\boldsymbol{U} = (U_1, \ldots, U_m)$ *and* $\boldsymbol{U}^* = (U_1^*, \ldots, U_m^*)$, $m \geq 1$, *be continuous random vectors with joint cumulative distribution functions* G *and* G^*, *respectively. For a positive integer* n, *let* $V_i = U_i^n$, $V_i^* = U_i^{*n}$, $i = 1, 2, \ldots, m$; $\boldsymbol{V} = (V_1, \ldots, V_m)$ *and* $\boldsymbol{V}^* = (V_1^*, \ldots, V_m^*)$ *and let* H *and* H^* *be the joint cumulative distribution functions of* $\boldsymbol{V}$ *and* $\boldsymbol{V}^*$, *respectively. Let* $D_{KL}(G \,||\, G^*)$ *denote the KL divergence between* G *and* G^*. *If* U_i*'s and* U_i^**'s are positive valued, then*

$$D_{KL}(H \,||\, H^*) = D_{KL}(G \,||\, G^*).$$

Proof Since random variables are positive valued, the function $v = u^n$ is bijective with only one real inverse image, say $u = v^{1/n}$. Denote the joint density functions of $\boldsymbol{U}$ and $\boldsymbol{U}^*$ by g and g^*, respectively. Then the joint density functions of $\boldsymbol{V}$ and $\boldsymbol{V}^*$, say h and h^*, are given by

$$h(\boldsymbol{v}) = g\left(v_1^{1/n}, \ldots, v_m^{1/n}\right) \frac{1}{n^m} \left(\prod_{i=1}^{m} v_i\right)^{\frac{1}{n}-1}$$

and

$$h^*(\boldsymbol{v}) = g^*\left(v_1^{1/n}, \ldots, v_m^{1/n}\right) \frac{1}{n^m} \left(\prod_{i=1}^{m} v_i\right)^{\frac{1}{n}-1},$$

where $\boldsymbol{v} = (v_1, \ldots, v_m),\ v_i > 0;\ i = 1, \ldots, m$.

Now,

$$\begin{aligned} D_{KL}(H \,||\, H^*) &= \int g(v_1^{1/n}, \ldots, v_m^{1/n}) \frac{1}{n^m} \left(\prod_{i=1}^{m} v_i\right)^{\frac{1}{n}-1} \ln\left(\frac{g(v_1^{1/n}, \ldots, v_m^{1/n})}{g^*(v_1^{1/n}, \ldots, v_m^{1/n})}\right) d\boldsymbol{v} \\ &= \int g(u_1, \ldots, u_m) \ln\left(\frac{g(u_1, \ldots, u_m)}{g^*(u_1, \ldots, u_m)}\right) d\boldsymbol{u},\ \ \boldsymbol{u} = (u_1, \ldots, u_m) \\ &= D_{KL}(G \,||\, G^*). \end{aligned}$$

□

The following theorem shows that the Kolmogorov distance between the conditional distribution function of Λ^* and the null distribution function of Λ converges to zero as $\min_{1 \leq i \leq k} n_i \to \infty$.

Theorem 1 *Let* $F_{\Lambda|H_0}(x)$ *denote the null distribution function of* Λ *and* $F_{\Lambda^*|\boldsymbol{X}}(x)$ *denote the conditional distribution function of* Λ^* *given the observation* $\boldsymbol{X} = (X_{11}, X_{12}, \cdots, X_{kn_k})$, *where* $X_{ij} \sim \exp(\mu_i, \sigma_i)$ *for* $j = 1, \ldots, n_i; i = 1, \ldots, k$. *Then,*

$$\sup_x |F_{\Lambda^*|\boldsymbol{X}}(x) - F_{\Lambda|H_0}(x)| \xrightarrow{P} 0 \ \text{ as } \ \min_{1 \leq i \leq k} n_i \to \infty.$$

Proof Under H_0, the distribution of the likelihood ratio test statistic Λ does not depend on the common location parameter μ. So without loss of generality, we take $\mu = 0$. Hence under H_0, $X_{ij} \sim \exp(0, \sigma_i), j = 1, 2, \ldots, n_i;\ i = 1, 2, \ldots, k$, and given $\boldsymbol{X}$, $X^*_{ij} \sim \exp(0, S_i),\ j = 1, 2, \ldots, n_i;\ i = 1, 2, \ldots, k$.

Let $\mathcal{L}_i$ denote the null distribution function of $X_{i(1)}$ and $\mathcal{L}^*_i$ denote the conditional distribution function of $X^*_{i(1)}$ given X_{ij}. It is known that as $n_i \to \infty$, $S_i \xrightarrow{a.s.} \sigma_i$ for $i = 1, 2, \ldots, k$. Hence, Kullback–Leibler (KL) divergence between $\mathcal{L}_i$ and $\mathcal{L}^*_i$ converges to zero almost surely, that is

$$D_{KL}(\mathcal{L}_i \,||\, \mathcal{L}^*_i) = \ln\left(\frac{\sigma_i}{S_i}\right) + \frac{S_i}{\sigma_i} - 1 \xrightarrow{a.s.} 0 \text{ as } n_i \to \infty,\ i = 1, 2, \ldots, k. \tag{4}$$

Again $S_i \sim \text{Gamma}\left(n_i - 1, \frac{n_i - 1}{\sigma_i}\right)$, say $\mathcal{G}_i$, and given the observations $\boldsymbol{X}$, conditional distribution of S^*_i is $\text{Gamma}\left(n_i - 1, \frac{n_i - 1}{S_i}\right)$, say $\mathcal{G}^*_i$, where $\text{Gamma}(\eta, \xi)$ denotes a gamma distribution with shape parameter η and scale parameter $1/\xi$. We obtain the KL divergence of $\mathcal{G}_i$ from $\mathcal{G}^*_i$ as

$$D_{KL}(\mathcal{G}_i \,||\, \mathcal{G}^*_i) = (n_i - 1)\left[\ln\left(\frac{S_i}{\sigma_i}\right) + \frac{\sigma_i}{S_i} - 1\right],\ i = 1, 2, \ldots, k.$$

As $\frac{\sigma_i}{S_i} \xrightarrow{a.s.} 1$, the Taylor's expansion of $\ln\left(1 + \frac{\sigma_i}{S_i} - 1\right)$ reduces the above identity to

$$D_{KL}(\mathcal{G}_i \,||\, \mathcal{G}^*_i) = \frac{(n_i - 1)}{2}\left(\frac{\sigma_i}{S_i} - 1\right)^2\left[1 - \frac{2}{3}\left(\frac{\sigma_i}{S_i} - 1\right) + \frac{2}{4}\left(\frac{\sigma_i}{S_i} - 1\right)^2 - \cdots\right]. \tag{5}$$

Clearly, the series inside the square bracket converges to 1 almost surely. Using Markov inequality, we get

$$P\left(\left|\frac{\sigma_i}{S_i} - 1\right| > \epsilon\right) \le \frac{1}{\epsilon^4} E\left(\frac{\sigma_i}{S_i} - 1\right)^4 = \frac{3(n_i - 1)^4(n_i + 4)}{(n_i - 2)^4(n_i - 3)(n_i - 4)(n_i - 5)\epsilon^4}.$$

Hence,

$$D_{KL}(\mathcal{G}_i \,||\, \mathcal{G}^*_i) \xrightarrow{P} 0 \text{ as } n_i \to \infty,\ i = 1, 2, \ldots, k. \tag{6}$$

Since $X_{i(1)}$ and S_i; and $X^*_{i(1)}$ and S^*_i are independent for $i = 1, 2, \ldots, k$,

$$D_{KL}\big((\mathcal{L}_i, \mathcal{G}_i) \,||\, (\mathcal{L}^*_i, \mathcal{G}^*_i)\big) = D_{KL}\big(\mathcal{L}_i \,||\, \mathcal{L}^*_i\big) + D_{KL}(\mathcal{G}_i \,||\, \mathcal{G}^*_i).$$

By (4) and (6),

$$D_{KL}\big((\mathcal{L}_i, \mathcal{G}_i) \,||\, (\mathcal{L}^*_i, \mathcal{G}^*_i)\big) \xrightarrow{P} 0 \text{ for } i = 1, 2, \ldots, k. \tag{7}$$

Let $\mathcal{L} = (\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_k)$, $\mathcal{G} = (\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_k)$, $\mathcal{L}^* = (\mathcal{L}_1^*, \mathcal{L}_2^*, \ldots, \mathcal{L}_k^*)$ and $\mathcal{G}^* = (\mathcal{G}_1^*, \mathcal{G}_2^*, \ldots, \mathcal{G}_k^*)$. By additive property of KL divergence and (7)

$$D_{KL}((\mathcal{L}, \mathcal{G}) \,||\, (\mathcal{L}^*, \mathcal{G}^*)) \xrightarrow{P} 0 \text{ as } \min_{1 \le i \le k} n_i \to \infty. \tag{8}$$

The mapping from $(X_{1(1)}, \ldots, X_{k(1)}, S_1, \ldots, S_k)$ to $\boldsymbol{U} = (\hat{\sigma}_{1\Omega_0}, \cdots, \hat{\sigma}_{k\Omega_0}, \hat{\sigma}_{1\Omega}, \cdots, \hat{\sigma}_{k\Omega})$ is a many-one onto function. Similarly, the mapping from $(X^*_{1(1)}, \ldots, X^*_{k(1)}, S^*_1, \ldots, S^*_k)$ to $\boldsymbol{U}^* = \left(\hat{\sigma}^*_{1\Omega_0}, \cdots, \hat{\sigma}^*_{k\Omega_0}, \hat{\sigma}^*_{1\Omega}, \ldots, \hat{\sigma}^*_{k\Omega}\right)$ is also a many-one onto function. But these use only permutations and linear transformations. So, if we consider $\boldsymbol{\tau}$ as the null distribution of $\boldsymbol{U}$ and $\boldsymbol{\tau}^*$ as the conditional distribution of $\boldsymbol{U}^*$ given $\boldsymbol{X}$, then it can be shown that

$$D_{KL}(\boldsymbol{\tau} \,||\, \boldsymbol{\tau}^*) = D_{KL}((\mathcal{L}, \mathcal{G}) \,||\, (\mathcal{L}^*, \mathcal{G}^*)).$$

By (8),

$$D_{KL}(\boldsymbol{\tau} \,||\, \boldsymbol{\tau}^*) \xrightarrow{P} 0 \text{ as } \min_{1 \le i \le k} n_i \to \infty. \tag{9}$$

Further, note that components of $\boldsymbol{U}$ and $\boldsymbol{U}^*$ are positive valued. Let $\boldsymbol{\tau}^{(\boldsymbol{n})}$ denote the null distribution function of

$$\boldsymbol{U}^{(\boldsymbol{n})} = \left(\hat{\sigma}^{n_1}_{1\Omega_0}, \ldots, \hat{\sigma}^{n_k}_{k\Omega_0}, \hat{\sigma}^{n_1}_{1\Omega}, \ldots, \hat{\sigma}^{n_k}_{k\Omega}\right)$$

and let $\boldsymbol{\tau}^{*(\boldsymbol{n})}$ be the conditional distribution function of

$$\boldsymbol{U}^{*(\boldsymbol{n})} = \left(\hat{\sigma}^{*n_1}_{1\Omega_0}, \ldots, \hat{\sigma}^{*n_k}_{k\Omega_0}, \hat{\sigma}^{*n_1}_{1\Omega}, \ldots, \hat{\sigma}^{*n_k}_{k\Omega}\right)$$

given $\boldsymbol{X}$. Using Lemma 1, we have

$$D_{KL}(\boldsymbol{\tau}^{(\boldsymbol{n})} \,||\, \boldsymbol{\tau}^{*(\boldsymbol{n})}) = D_{KL}(\boldsymbol{\tau} \,||\, \boldsymbol{\tau}^*). \tag{10}$$

Hence, by (9) and (10)

$$D_{KL}(\boldsymbol{\tau}^{(\boldsymbol{n})} \,||\, \boldsymbol{\tau}^{*(\boldsymbol{n})}) \xrightarrow{P} 0 \text{ as } \min_{1 \le i \le k} n_i \to \infty, \tag{11}$$

Now, Λ and Λ^* being continuous functions of $\boldsymbol{U}^{(\boldsymbol{n})}$ and $\boldsymbol{U}^{*(\boldsymbol{n})}$, respectively, by the continuous mapping theorem and (11) the required result holds. □

This theorem shows that asymptotically the conditional distribution of Λ^* always approximates the null distribution of Λ. Hence, it ensures that the bootstrap critical value always converges in probability to the critical value that is obtained using the null distribution of Λ. Precisely, it says, if λ^* is the α-quantile of Λ^*, given the observations $\mathbf{X}$ and λ is the α-quantile of Λ under H_0, then $\lambda^* \xrightarrow{P} \lambda$ as $\min_{1 \le i \le k} n_i \to \infty$.

3 Min-T and max-T tests

In this section, we develop two more tests for H_0 against H_1 based on simultaneous comparisons. Since ordered alternative is used, we need to take pairwise differences of successive sample minimums. The idea is similar to the one used for normal populations where sample means are used (see Mondal et al. 2023).

Consider the pairwise hypotheses testing problems $H_{0i} : \mu_i = \mu_{i+1}$ against $H_{1i} : \mu_i < \mu_{i+1}$ for $i = 1, 2, \ldots, k-1$. Then H_0 and H_1 can be written as $H_0 = \cap_{i=1}^{k-1} H_{0i}$ and $H_1 = \cup_{i=1}^{k-1} H_{1i}$. If we consider a particular form of H_1, say, $H_1^* : \mu_1 < \cdots < \mu_k$, then it can be expressed as $H_1^* : \cap_{i=1}^{k-1} H_{1i}$. So, we propose two union-intersection type tests for H_0 against H_1 and H_0 against H_1^*, respectively.

A natural Behrens–Fisher-type test statistic for H_{0i} vs. H_{1i} is

$$W_i = \frac{X_{i+1\,(1)} - X_{i\,(1)}}{\sqrt{\frac{(n_i-1)}{n_i^3} S_i^2 + \frac{(n_{i+1}-1)}{n_{i+1}^3} S_{i+1}^2}}, \; i = 1, 2, \ldots, k-1,$$

where the term within square roots in the denominator is the UMVUE of variance of $X_{i+1\,(1)} - X_{i(1)}$.

Based on W_is, we now define the following test statistics for testing H_0 vs. H_1:

$$T_1 = \min_{1 \le i \le k-1} W_i \tag{12}$$

and

$$T_2 = \max_{1 \le i \le k-1} W_i. \tag{13}$$

We call the above proposed tests Min-T and Max-T, respectively. Note that LRT takes information from the whole data set under both parameter spaces. Hence, it would be more powerful than the Min-T and Max-T tests. However, when the null hypothesis is rejected, LRT is not helpful for post-hoc analysis. For this reason, we propose these procedures. The test statistics T_2 can be used to find the simultaneous confidence intervals for successive pairwise differences between treatment effects (μ_is). It is helpful to investigate the exact differences among μ_is which lead to rejection of H_0. Again, T_1 is helpful in detecting the strict ordering among μ_is.

Note that there is some difference in the rejection criteria used by T_1 and T_2. The test statistic T_1 rejects the null hypothesis H_0 if and only if H_{0i} is rejected for all $i = 1, 2, \ldots, k$. Meanwhile, H_0 is rejected by T_2 if and only if at least one of the null hypotheses H_{0i} is rejected. At level α, the Min-T test rejects H_0 if $T_1 > d$, where d is the $(1-\alpha)$-quantile of T_1 under the null hypothesis H_0. The null hypothesis H_0 is rejected by Max-T test if $T_2 > e$, where e is the $(1-\alpha)$-quantile of T_2 under H_0.

The distributions of the test statistics T_i, $i = 1, 2$ under H_0 cannot be obtained in a closed form. In order to determine the critical values using the null distributions of T_i, $i = 1, 2$, one needs to solve some multiple integrals numerically. Instead of doing this, we use the parametric bootstrap for its simplicity. This involves doing simulations a large number of times. However, we have provided an 'R' package to find

test statistics and critical values, which makes this process worthwhile. Also, in simulation studies, it is seen that the tests have very good size and power performance.

3.1 Parametric bootstrap Min-T and Max-T tests

Given samples X_{ij} from $\exp(\mu_i, \sigma_i)$, we generate parametric bootstrap samples X_{ij}^* from $\exp(0, S_i)$, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$. Define $S_i^* = \frac{n_i}{n_i - 1}(\bar{X}_i^* - X_{i(1)}^*)$ for $i = 1, 2, \ldots, k$. The PB statistic corresponding to W_i is

$$W_i^* = \frac{X_{i+1\,(1)}^* - X_{i(1)}^*}{\sqrt{\frac{(n_i-1)}{n_i^3} S_i^{*2} + \frac{(n_{i+1}-1)}{n_{i+1}^3} S_{i+1}^{*2}}}, \; i = 1, 2, \ldots, k-1.$$

Now, define the parametric bootstrap Min-T and Max-T statistics by

$$T_1^* = \min_{1 \le i \le k-1} W_i^* \tag{14}$$

and

$$T_2^* = \max_{1 \le i \le k-1} W_i^*, \tag{15}$$

respectively.

The procedure for finding the critical points, size and power values using the bootstrap statistics is similar to that in Algorithm 1 with a few changes. In Step 1, calculate the value of T_1 and T_2, defined in (12) and (13) for Min-T and Max-T tests, respectively. Using bootstrap samples X_{ij}^* calculate the bootstrap statistics T_1^* and T_2^*, defined in (14) and (15), respectively. In Step 4, after arranging the B number of values of statistics T_1^* and T_2^* in respective increasing orders, the critical values of Min-T and Max-T, respectively, are defined as $d_r^* = T_{1[(1-\alpha)H]}^*$ and $e_r^* = T_{2[(1-\alpha)H]}^*$. Now in Step 6, we define the power values as

$$\alpha_{Min} = \frac{\text{number of times } T_{1r} \text{ more than } d_r^*}{B} = \frac{1}{B} \sum_{r=1}^{B} 1\{T_{1r} > d_r^*\}$$

and

$$\alpha_{Max} = \frac{\text{number of times } T_{2r} \text{ more than } e_r^*}{B} = \frac{1}{B} \sum_{r=1}^{B} 1\{T_{2r} > e_r^*\}.$$

In the following theorems, we establish the asymptotic accuracy of PB tests T_1^* and T_2^* as we have done earlier for LRT. First in Theorem 2, we prove that given the sample $\boldsymbol{X}$, the bootstrap statistic S_i^* converges in probability to σ_i as $n_i \to \infty$, for $i = 1, 2, \ldots, k$ almost surely (a.s.). Bickel and Freedman (1981) have shown that given the sample $\boldsymbol{X}$, the sample variance converges in probability to population variance. However, their result is not directly applicable here.

Theorem 2 *Suppose $X_1, X_2, \ldots$ are independent and identically distributed* $\exp(\mu, \sigma)$ *random variables. Given the observations X_1, X_2, …, X_n, bootstrap samples X_1^*, X_2^*, …, X_n^* are generated from* $\exp(0, S)$, *where* $S = \frac{n}{n-1}\left(\bar{X} - \tilde{X}_{(1)}\right)$, $\tilde{X}_{(1)} = \min_{1 \le j \le n} X_j$. *Then along almost all sample sequences $X_1, X_2, \ldots$, given $(X_1, X_2, \ldots, X_n)$, as $n \to \infty$, $S^* \to \sigma$ in conditional probability, that is, for $\epsilon > 0$,*

$$P\left(\mid S^* - \sigma \mid > \epsilon \mid (X_1, X_2, \ldots, X_n)\right) \to 0 \; a.s.,$$

where $S^* = \frac{n}{n-1}(\bar{X}^* - \tilde{X}^*_{(1)})$, $\bar{X}^* = \frac{1}{n}\sum_{j=1}^{n} X_j^*$, $\tilde{X}^*_{(1)} = \min_{1 \le j \le n} X_j^*$.

Proof It is known that the distribution of S is Gamma $\left(n-1, \frac{n-1}{\sigma}\right)$ and the conditional distribution of S^* given $\boldsymbol{X}$ is Gamma $\left(n-1, \frac{n-1}{S}\right)$.

We get,

$$E((S - S^*)^2) = E_{\boldsymbol{X}}(S^2) + E_{\boldsymbol{X}}\left[E_{S^*|\boldsymbol{X}}(S^{*2})\right] - 2E_{\boldsymbol{X}}\left[SE_{S^*|\boldsymbol{X}}(S^*)\right] = \frac{n\sigma^2}{(n-1)^2}.$$

Therefore, in conditional probability, $S^* - S \to 0$ as $n \to \infty$. Now, the proof follows from the fact that $S \xrightarrow{a.s.} \sigma$ as $n \to \infty$. □

Theorem 3 *Let $F_{\mathbf{W}^*|\mathbf{X}}(\boldsymbol{w})$ denote the conditional distribution function of $\mathbf{W}^*$ given the observations $\mathbf{X}$ and $F_{\mathbf{W}|H_0}(\boldsymbol{w})$ denote the null distribution function of $\mathbf{W}$, where $\mathbf{W} = (W_1, W_2, \ldots, W_{k-1})$, $\mathbf{W}^* = (W_1^*, W_2^*, \ldots, W_{k-1}^*)$. Then,*

$$\sup_{\boldsymbol{w} \in \mathbb{R}^{k-1}} |F_{\mathbf{W}^*|\boldsymbol{X}}(\boldsymbol{w}) - F_{\mathbf{W}|H_0}(\boldsymbol{w})| \xrightarrow{P} 0 \;\text{ as }\; \min_{1 \le i \le k} n_i, \; N \to \infty.$$

Proof We rewrite W_i and W_i^* as

$$W_i = \frac{N(X_{i+1\,(1)} - X_{i(1)})}{\sqrt{\frac{N^2(n_i-1)}{n_i^3}S_i^2 + \frac{N^2(n_{i+1}-1)}{n_{i+1}^3}S_{i+1}^2}}, \; i = 1, 2, \ldots, k-1 \tag{16}$$

$$W_i^* = \frac{N(X^*_{i+1\,(1)} - X^*_{i(1)})}{\sqrt{\frac{N^2(n_i-1)}{n_i^3}S_i^{*2} + \frac{N^2(n_{i+1}-1)}{n_{i+1}^3}S_{i+1}^{*2}}}, \; i = 1, 2, \ldots, k-1. \tag{17}$$

Under H_0, W_i's are independent of the location parameter, so we can choose $\mu = 0$. Hence, under H_0, $NX_{i(1)} \sim \exp\left(0, \frac{N\sigma_i}{n_i}\right)$ for $i = 1, 2, \ldots, k$. Given X_{ij}, $j = 1, 2, \ldots, n_i$; $i = 1, 2, \ldots, k$, $NX^*_{i(1)} \sim \exp\left(0, \frac{NS_i}{n_i}\right)$, $i = 1, 2, \ldots, k$. Let $\mathcal{V}_i$ be the null distribution function of $NX_{i(1)}$ and $\mathcal{V}_i^*$ be the conditional distribution function of $NX^*_{i(1)}$ given $\boldsymbol{X}$. The Kullback–Leibler divergence of $\mathcal{V}_i$ from $\mathcal{V}_i^*$ is obtained as

$$D_{KL}\left(\mathcal{V}_i||\mathcal{V}_i^*\right) = \log\left(\frac{\sigma_i}{S_i}\right) + \frac{S_i}{\sigma_i} - 1.$$

Since $S_i \xrightarrow{a.s.} \sigma_i$ as $n_i \to \infty$,

$$D_{KL}\left(\mathcal{V}_i||\mathcal{V}_i^*\right) \xrightarrow{a.s.} 0 \text{ as } n_i \to \infty. \tag{18}$$

Let $\mathcal{V}$ denote the null distribution function of $\mathbf{Y}$=$N(X_{1(1)}, X_{2(1)}, \cdots, X_{k(1)})$ and $\mathcal{V}^*$ denote the conditional distribution function of $\mathbf{Y}^*$=$N(X^*_{1(1)}, X^*_{2(1)}, \cdots, X^*_{k(1)})$ given $\boldsymbol{X}$. By (18), $D_{KL}\left(\mathcal{V} \,||\, \mathcal{V}^*\right) \xrightarrow{a.s.} 0$ as $\min_{1\le i\le k} n_i \to \infty$. This yields

$$\sup_{\boldsymbol{y}} \mid F_{\mathbf{Y}|H_0}(\boldsymbol{y}) - F_{\mathbf{Y}^*|\mathbf{X}}(\boldsymbol{y}) \mid \xrightarrow{a.s.} 0 \text{ as } \min_{1\le i\le k} n_i, N \to \infty. \tag{19}$$

Let $\mathbf{Z} = A\mathbf{Y}$, where $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_{k-1})$, $Z_i = N(X_{i+1\ (1)} - X_{i(1)})$, $i = 1, 2, \ldots, k-1$ and

$$A = \begin{bmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{bmatrix}.$$

By Equation (19) and the continuous mapping theorem, we get,

$$\sup_{z} \mid F_{\mathbf{Z}|\mathbf{X}}(z) - F_{\mathbf{Z}^*|H_0}(z) \mid \xrightarrow{a.s.} 0. \tag{20}$$

Let $\hat{\boldsymbol{D}} = \text{diag}(\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_{k-1})$, $\hat{\boldsymbol{D}}^* = \text{diag}(\hat{d}^*_1, \hat{d}^*_2, \ldots, \hat{d}^*_{k-1})$ and $\boldsymbol{D}$=diag(d_1, d_2, $\cdots$, d_{k-1}), where $\hat{d}_i = \frac{N^2(n_i-1)}{n_i^3} S_i^2 + \frac{N(n_{i+1}-1)}{n_{i+1}^3} S_{i+1}^2$, $\hat{d}^{*2}_i = \frac{N^2(n_i-1)}{n_i^3} S_i^{*2} + \frac{N(n_{i+1}-1)}{n_{i+1}^3} S_{i+1}^{*2}$ and $d_i = \left(\frac{\sigma_i}{\nu_i}\right)^2 + \left(\frac{\sigma_{i+1}}{\nu_{i+1}}\right)^2$.

Consider the asymptotic setup

$$\frac{n_i}{N} \to \nu_i \text{ as } N, \min_{1\le i\le k} n_i \to \infty,\ i = 1, 2, \ldots, k. \tag{21}$$

By Theorem 2, under the asymptotic set up (21),

$$\hat{d}_i \xrightarrow{a.s.} d_i \text{ and } \hat{d}^*_i \xrightarrow{P} d_i,\ i = 1, 2, \ldots, k-1. \tag{22}$$

Hence, as $\min_{1\le i\le k} n_i$, $N \to \infty$, $\hat{\boldsymbol{D}} \xrightarrow{a.s.} \boldsymbol{D}$ and $\hat{\boldsymbol{D}}^* \xrightarrow{P} \boldsymbol{D}$. We have $\mathbf{W} = \mathbf{D}^{-1}\mathbf{Z}$ and $\mathbf{W}^* = \mathbf{D}^{*-1}\mathbf{Z}$. Therefore, the result follows by the continuous mapping theorem and (20). □

4 Simultaneous confidence intervals

If the null hypothesis is rejected, it is of interest to find the pairs for which the rejection occurs. In this section, we develop simultaneous confidence intervals (SCI) for the successive pairwise differences of the location parameters; $\mu_{i+1} - \mu_i,\ i = 1, 2, \ldots, k-1$ using the Max-T test. Inverting the statistics W_is, we get $(1-\alpha)$-level one-sided simultaneous confidence intervals for $\mu_{i+1} - \mu_i,\ i = 1, 2, \ldots, k-1$ as

$$\left[X_{i+1\ (1)} - X_{i(1)} - e\sqrt{\frac{(n_i-1)}{n_i^3}S_i^2 + \frac{(n_{i+1}-1)}{n_{i+1}^3}S_{i+1}^2},\ \infty\right),$$

where e is the critical value of the Max-T test, defined in Sect. 3.

As we have already discussed, we use bootstrap critical value for its simplicity; here also we use bootstrap critical point. We call the corresponding confidence intervals as the bootstrap confidence intervals. This is simply

$$\left[X_{i+1\ (1)} - X_{i(1)} - e^*\sqrt{\frac{(n_i-1)}{n_i^3}S_i^2 + \frac{(n_{i+1}-1)}{n_{i+1}^3}S_{i+1}^2},\ \infty\right).$$

5 Gill's test

As we want to compare the powers of LRT, Min-T and Max-T with Gill's test (Gill et al. 2017), here we describe Gill's test. For the hypothesis testing problem of equality of exponential location parameters H_0 against ordered alternatives H_1 with unknown and unequal scale parameters, Gill proposed the test statistic

$$G_j = \max_{1\le i\le j} \frac{(X_{j(1)} - X_{i(1)})}{(S_j/n_j)},\ j = 2, \ldots, k.$$

The test rejects H_0 at a level α, if at least one of the observed values of G_j is greater than g_j, where $g_j = F^{-1}_{2,2n_j-2}((1-\alpha)^{1/(k-1)}),\ j = 2, \ldots, k$.

6 Size and power analysis

We have proved theoretically that the parametric bootstrap test statistics always approximates the null distributions of the main test statistics for large sample sizes. In this section, we investigate the performance of the tests for small, medium and highly unbalanced sample sizes. We have also considered large samples to exhibit the validation of theoretical results. Carrying out extensive simulations, we tabulate here the simulated size and power values of the proposed PB based LRT, Min-T and Max-T tests. We also tabulate size and power values of Gill's test for the same

configurations as taken for the other tests, so that we can compare Gill's test with our test procedures. Moreover, the power values of LRT, Min-T, and Max-T tests are also tabulated for the combinations of parameters as given in Gill et al. (2017).

The steps of Algorithm 1 are followed to find the size values of LRT, Min-T and Max-T, given in Sects. 2.1 and 3.1. For power calculations, μ_is in Step 1 are taken to be in a monotonic increasing order and other steps are retained the same as in Algorithm 1. For a given number of groups k, common location parameter μ, scale parameters $\sigma_1, \sigma_2, \ldots \sigma_k$ and sample sizes $n_1, n_2, \ldots, n_k$, exponential populations X_{ij} are generated from $\exp(\mu_i, \sigma_i)$, $j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, k$. Then, for each sample, the bootstrap critical value is calculated by generating the bootstrap samples 5000 times and it is counted if the null hypothesis is rejected. This process is repeated 10,000 times and the proportion of times the null hypothesis is rejected is calculated as a simulated size value. Note that Gill et al. (2017) have not tabulated the size values of their test. For the purpose of comparison, we have additionally estimated the size values of their test. For finding the size values of Gill's test, samples X_{ij} are generated from ith exponential population with location μ and scale σ_i for $j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, k$. Then, the observed value of the test statistic and the corresponding critical values are calculated from the data. The proportion of times H_0 is rejected is considered as the size value. We have also compared proposed simultaneous confidence intervals and existing Gill's confidence intervals on the basis of the coverage probability. For this, we have generated samples X_{ij} from $\exp(\mu_i, \sigma_i)$ for $j = 1, 2, \ldots, n_i, i = 1, 2, \ldots, k$. As in Gill et al. (2017), we have taken $\mu_i = 0,\ i = 1, 2, \cdots, k$. From the samples, simultaneous intervals for successive pairwise differences $\mu_{i+1} - \mu_i,\ i = 1, 2, \ldots, k-1$ are evaluated and stored as value one if all these confidence intervals contain zero. The process is repeated 10,000 times. The proportion of times all the confidence intervals contain zero is evaluated as the coverage probability. All the simulations are conducted in freely available 'R' software.

The simulated size values of the parametric bootstrap based tests and the existing Gill's test are tabulated in Tables 1, 2 and 3. We consider $k = 2$ (number of groups) in Table 1, $k = 3$ in Table 2 and $k = 9$ in Table 3. Very small, moderate, large and heavily unbalanced sample sizes and several types of σ_i's are considered in both the tables. For $k = 2$, Min-T and Max-T tests reduce to single paired comparison tests. We call this T test. In Table 1, we have presented size values of tests for $k = 2$ by taking sample sizes $\boldsymbol{N_1} = (7, 8)$, $\boldsymbol{N_2} = (12, 14)$, $\boldsymbol{N_3} = (18, 8)$, $\boldsymbol{N_4} = (20, 20)$, $\boldsymbol{N_5} = (10, 30)$, $\boldsymbol{N_6} = (7, 80)$, $\boldsymbol{N_7} = (45, 50)$ and scale parameters (1, 1), (1, 2), (0.4, 0.1), (4, 8). In Table 2, the sample sizes considered are $\boldsymbol{N_8} = (7, 8, 9)$, $\boldsymbol{N_9} = (10, 8, 9)$, $\boldsymbol{N_{10}} = (10, 30, 15)$, $\boldsymbol{N_{11}} = (20, 30, 25)$, $\boldsymbol{N_{12}} = (30, 30, 30)$, $\boldsymbol{N_{13}} = (70, 80, 60)$, $\boldsymbol{N_{14}} = (10, 80, 10)$ and $\boldsymbol{N_{15}} = (50, 40, 7)$. Note that $\boldsymbol{N_p} = (n_1, n_2), p = 1, 2, \ldots, 7$ and $\boldsymbol{N_p} = (n_1, n_2, n_3)$ for $p = 8, \ldots, 15$. Also, scale parameters are taken to be equal and/or small, unequal and/or large, namely $(\sigma_1, \sigma_2, \sigma_3) \in \{$ (1, 1, 1), (0.5, 0.2, 0.8), (1, 1.5, 2), (1, 1, 2), (1, 2, 3), (2.5, 3, 2) $\}$. In Table 3, the sample sizes for nine groups are taken as $\boldsymbol{N_{16}} = (7, 7, 8, 7, 7, 8, 7, 7, 8)$, $\boldsymbol{N_{17}} = (10, 8, 9, 10, 8, 9, 10, 8, 9)$, $\boldsymbol{N_{18}} = (20, 30, 25, 20, 30, 25, 20, 30, 25)$, $\boldsymbol{N_{19}} = (30, 30, 30, 30, 30, 30, 30, 30, 30)$, $\boldsymbol{N_{20}} = (10, 30, 10, 10, 30, 10, 10, 30, 10)$, $\boldsymbol{N_{21}} = (70, 80, 60, 70, 80, 60, 70, 80, 60)$, $\boldsymbol{N_{22}} = (70, 60, 8, 70, 60, 8, 70, 60, 8)$, where $\boldsymbol{N_q} = (n_1, n_2, \ldots, n_9)$,

Table 1 Size values of LRT, Min-T/Max-T and Gill's test for $k = 2$ when $\boldsymbol{N_1} = (7, 8)$, $\boldsymbol{N_2} = (12, 14)$, $\boldsymbol{N_3} = (18, 8)$, $\boldsymbol{N_4} = (20, 20)$, $\boldsymbol{N_5} = (10, 30)$, $\boldsymbol{N_6} = (7, 80)$, $\boldsymbol{N_7} = (45, 50)$

(σ_1, σ_2)	Tests	$\boldsymbol{N_1}$	$\boldsymbol{N_2}$	$\boldsymbol{N_3}$	$\boldsymbol{N_4}$	$\boldsymbol{N_5}$	$\boldsymbol{N_6}$	$\boldsymbol{N_7}$
(1,1)	LRT	0.0556	0.0522	0.0590	0.0506	0.0530	0.0494	0.0464
	Min-T/Max-T	0.0698	0.0582	0.0734	0.0548	0.0482	0.0532	0.0518
	Gill's test	0.0221	0.0229	0.0345	0.0260	0.0132	0.0037	0.0242
(1,2)	LRT	0.0592	0.0484	0.0548	0.0562	0.0530	0.0480	0.0502
	Min-T/Max-T	0.0662	0.0652	0.0636	0.0612	0.0536	0.0526	0.0618
	Gill's test	0.0329	0.0312	0.0410	0.0335	0.0202	0.0073	0.0315
(0.4,0.1)	LRT	0.0534	0.0614	0.0550	0.0550	0.0492	0.0210	0.0562
	Min-T/Max-T	0.0550	0.0536	0.0674	0.0512	0.0494	0.0490	0.0514
	Gill's test	0.0088	0.0085	0.0182	0.0103	0.0041	0.0011	0.0083
(4,8)	LRT	0.0582	0.0492	0.0558	0.0502	0.0556	0.0498	0.0514
	Min-T/Max-T	0.0616	0.0636	0.0616	0.0508	0.0534	0.0480	0.0514
	Gill's test	0.0320	0.0312	0.0408	0.0340	0.0196	0.0079	0.0318

$q = 16, 17, \ldots, 22$. Here, several combinations of scale parameters, say, $(\sigma_1, \sigma_2, \cdots, \sigma_9) \in \{(1, 1, \cdots, 1), (1, 1.5, 2, 1, 1.5, 2, 1, 1.5, 2, 1, 1.5, 2), (1, 2, 3, 1, 2, 3, 1, 2, 3), (0.5, 0.2, 0.8, 0.5, 0.2, 0.8, 0.5, 0.2, 0.8), (2.5, 3, 2, 2.5, 3, 2, 2.5, 3, 2), (1, 2, 2, 1, 2, 2, 1, 2, 2), (1, 1, 2, 1, 1, 2, 1, 1, 2)\}$ are considered.

Power plots of LRT, Min-T/Max-T and Gill's test for $k = 2$ are given in Fig. 1. Here, we have taken treatment effects $(\mu_1, \mu_2) = c(1.2, 1.4)$, sample sizes (12, 14), scale parameters (1, 2) and (3, 2). The value of c ranges from 0 (0 gives the sizes) to 5, with an increment of 0.2. Here the configurations for μ_is are chosen so that by increasing the value of c, the difference between μ_is increases and consequently, we can see an increase in the power values. In Figs. 2, 3, 4, 5, 6, 7 and 8, power curves of PB based LRT, Max-T, Min-T and Gill's tests are shown for $k = 3$. In Figs. 2 and 3 strictly increasing μ_is, say, $(\mu_1, \mu_2, \mu_3) = c(1, 1.01, 1.02)$ is taken. In Fig. 4, the choices of (μ_1, μ_2, μ_3) and $(\sigma_1, \sigma_2, \sigma_3)$ are $c(2, 2.05, 3)$ and (2, 4, 6), respectively. In Figs. 5 and 6, increasing μ_is with one equality, say $(\mu_1, \mu_2, \mu_3) = c(1, 1.02, 1.02)$ is chosen. In Figs. 2, 3, 5, and 6, samples are chosen $(n_1, n_2, n_3) \in \{(20, 30, 25), (40, 60, 50), (70, 80, 100)\}$. In Fig. 4, sample sizes (8, 10, 12) and (15, 18, 20) are considered. In Figs. 2, 5, equal or small scale parameters $(\sigma_1, \sigma_2, \sigma_3) = (1, 1, 1)$ and in Figs. 3, 6 unequal scales $(\sigma_1, \sigma_2, \sigma_3) = (1, 2, 3)$ are considered. Figures 7 and 8 provide the power curves of the LRT, Max-T and Min-T tests for $k = 4$. Here also strictly increasing μ_is, that is, $\boldsymbol{\mu} = c(1, 1.01, 1.02, 1.03)$ and monotonically increasing μ_is with one equality, say, $\boldsymbol{\mu} = c(1, 1, 1.04, 1.05)$ are considered. For both combinations of μ_is, unequal scale parameters (1, 1.2, 1.4, 2) are taken. In both figures, the choices of sample sizes are taken as moderate (20, 30, 25, 20) and large (40, 60, 50, 40)). The power values of the tests are tabulated for the configurations of parameters as considered by Gill et al. (2017) in Tables 4, 5, 6 and 7. In Table 8, we have presented the estimated coverage probabilities. For the purpose of comparison, we have chosen the same configurations of

Table 2 Size values of LRT, Min-T, Max-T and Gill's test for $k = 3$ when $\boldsymbol{N_8} = (7, 8, 9)$, $\boldsymbol{N_9} = (10, 8, 9)$, $\boldsymbol{N_{10}} = (10, 30, 15)$, $\boldsymbol{N_{11}} = (20, 30, 25)$, $\boldsymbol{N_{12}} = (30, 30, 30)$, $\boldsymbol{N_{13}} = (70, 80, 60)$, $\boldsymbol{N_{14}} = (10, 80, 10)$, $\boldsymbol{N_{15}} = (50, 40, 7)$

$(\sigma_1, \sigma_2, \sigma_3)$	Tests	$\boldsymbol{N_8}$	$\boldsymbol{N_9}$	$\boldsymbol{N_{10}}$	$\boldsymbol{N_{11}}$	$\boldsymbol{N_{12}}$	$\boldsymbol{N_{13}}$	$\boldsymbol{N_{14}}$	$\boldsymbol{N_{15}}$
(1, 1, 1)	LRT	0.0565	0.0585	0.0507	0.0500	0.0542	0.0521	0.0494	0.0568
	Max-T	0.0415	0.0575	0.0492	0.0512	0.0482	0.0510	0.0542	0.0580
	Min-T	0.0615	0.0630	0.0578	0.0572	0.0568	0.0470	0.0540	0.0550
	Gill's	0.0289	0.0307	0.0248	0.0257	0.0294	0.0294	0.0254	0.0363
(1, 1.5, 2)	LRT	0.0561	0.0559	0.0479	0.0531	0.0478	0.0511	0.0471	0.0530
	Max-T	0.0450	0.0553	0.0541	0.0547	0.0494	0.0539	0.0538	0.0520
	Min-T	0.0663	0.0716	0.0520	0.0578	0.0528	0.0564	0.0480	0.0600
	Gill's	0.0328	0.0361	0.0287	0.0314	0.0344	0.0342	0.0278	0.0415
(1, 2, 3)	LRT	0.0549	0.0594	0.0547	0.0508	0.0526	0.0488	0.0554	0.0548
	Max-T	0.0502	0.0552	0.0549	0.0504	0.0504	0.0518	0.0540	0.0560
	Min-T	0.0632	0.0630	0.0548	0.0566	0.0508	0.0514	0.0510	0.0510
	Gill's	0.0355	0.0393	0.0319	0.0342	0.0371	0.0367	0.0292	0.0430
(0.5, 0.2, 0.8)	LRT	0.0538	0.0576	0.0478	0.0521	0.0513	0.0487	0.0471	0.0510
	Max-T	0.0600	0.0525	0.0579	0.0530	0.0549	0.0520	0.0486	0.0522
	Min-T	0.0632	0.0636	0.0524	0.0520	0.0556	0.0526	0.0500	0.0535
	Gill's	0.0267	0.0307	0.0262	0.0281	0.0287	0.0279	0.0260	0.0330
(2.5, 3, 2)	LRT	0.0566	0.0589	0.0551	0.0521	0.0540	0.0510	0.0510	0.0550
	Max-T	0.0590	0.0572	0.0523	0.0490	0.0498	0.0519	0.0571	0.0600
	Min-T	0.0582	0.0602	0.0512	0.0546	0.0578	0.0450	0.0512	0.0500
	Gill's	0.0269	0.0298	0.0242	0.0258	0.0292	0.0287	0.0251	0.0371
(1, 2, 2)	LRT	0.0567	0.0551	0.0541	0.0512	0.0532	0.0518	0.0507	0.0527
	Max-T	0.0590	0.0533	0.0543	0.0550	0.0499	0.0507	0.0584	0.0550
	Min-T	0.0630	0.0608	0.0572	0.0506	0.0598	0.0542	0.0558	0.0620
	Gill's	0.0328	0.0373	0.0286	0.0331	0.0353	0.0359	0.0274	0.04207
(1, 1, 2)	LRT	0.0531	0.0541	0.0517	0.0495	0.0508	0.0505	0.0481	0.0529
	Max-T	0.0547	0.0536	0.0532	0.0523	0.0526	0.0524	0.0546	0.0600
	Min-T	0.0670	0.0678	0.0560	0.0548	0.0614	0.0520	0.0508	0.0575
	Gill's	0.0313	0.0348	0.0273	0.0303	0.0336	0.0327	0.0258	0.0383

sample sizes and scale parameters as taken in Gill's paper. In all tabulated results, nominal value of α is taken as 0.05.

In the following remarks, we analyze the results of our extensive simulation study.

Remark 4 For $k = 2$, size values of LRT and the T test achieve pre-defined level 0.05 with a few exceptions. For $k = 3$, we observe from Table 2 that simulated size values of the parametric bootstrap based LRT, Min-T, and Max-T maintain the preassigned 5% level, for small, moderate, large, and highly unbalanced sample sizes. However, sometimes for very small sample sizes, e.g., $((n_1, n_2, n_3) = (7, 8, 9))$ all three tests become slightly liberal. However, for all the configurations of sample sizes and scale

Table 3 Size of LRT, Min-T, Max-T and Gill's test for $k = 9$ when $\boldsymbol{N_{16}} = (7, 7, 8, 7, 7, 8, 7, 7, 8)$, $\boldsymbol{N_{17}} = (10, 8, 9, 10, 8, 9, 10, 8, 9)$, $\boldsymbol{N_{18}} = (20, 30, 25, 20, 30, 25, 20, 30, 25)$, $\boldsymbol{N_{19}} = (30, 30, 30, 30, 30, 30, 30, 30, 30)$, $\boldsymbol{N_{20}} = (10, 30, 10, 10, 30, 10, 10, 30, 10)$, $\boldsymbol{N_{21}} = (70, 80, 60, 70, 80, 60, 70, 80, 60)$ and $\boldsymbol{N_{22}} = (70, 60, 8, 70, 60, 8, 70, 60, 8)$

$\boldsymbol{\sigma}$	Tests	$\boldsymbol{N_{16}}$	$\boldsymbol{N_{17}}$	$\boldsymbol{N_{18}}$	$\boldsymbol{N_{19}}$	$\boldsymbol{N_{20}}$	$\boldsymbol{N_{21}}$	$\boldsymbol{N_{22}}$
(1, 1, 1,	LRT	0.0568	0.0592	0.0546	0.0552	0.0554	0.0496	0.0524
1, 1, 1,	Max-T	0.0354	0.0374	0.0474	0.0472	0.0508	0.0476	0.0580
1, 1, 1)	Min-T	0.0542	0.0504	0.0492	0.0492	0.0492	0.0540	0.0536
	Gill's	0.0384	0.0401	0.0386	0.0391	0.0356	0.0381	0.0417
(1, 1.5, 2,	LRT	0.0602	0.0478	0.0518	0.0518	0.0568	0.0528	0.0520
1, 1.5, 2,	Max-T	0.0408	0.0454	0.0521	0.0518	0.0504	0.0486	0.0570
1, 1.5, 2)	Min-T	0.0520	0.0514	0.0536	0.0556	0.0530	0.0570	0.0562
	Gill's	0.0406	0.0387	0.0384	0.0405	0.0396	0.0391	0.0424
(1, 2, 3,	LRT	0.0558	0.0540	0.0534	0.0552	0.0564	0.0482	0.0542
1,2,3	Max-T	0.0464	0.0511	0.0489	0.0449	0.0622	0.0516	0.0526
1,2,3)	Min-T	0.0514	0.0530	0.0534	0.0534	0.0506	0.0574	0.0498
	Gill's	0.0397	0.0400	0.0384	0.0413	0.0394	0.0423	0.0427
(0.5, 0.2, 0.8,	LRT	0.0520	0.0606	0.0534	0.0506	0.0498	0.0548	0.0448
0.5, 0.2, 0.8,	Max-T	0.0518	0.0527	0.0515	0.0522	0.0494	0.0520	0.0508
0.5, 0.2, 0.8)	Min-T	0.0578	0.0516	0.0550	0.0550	0.0486	0.0550	0.0528
	Gill's	0.0379	0.0394	0.0406	0.0391	0.0393	0.0384	0.0409
(2.5, 3, 2,	LRT	0.0530	0.0616	0.0520	0.0554	0.0534	0.0518	0.0552
2.5, 3, 2,	Max-T	0.0400	0.0403	0.0472	0.0484	0.0540	0.0456	0.0590
2.5, 3, 2)	Min-T	0.0554	0.0502	0.0488	0.0488	0.0542	0.0476	0.0554
	Gill's	0.0392	0.0398	0.0372	0.0391	0.0390	0.0401	0.0419
(1, 2, 2,	LRT	0.0532	0.0532	0.0484	0.0486	0.0536	0.0510	0.0548
1, 2, 2,	Max-T	0.0482	0.0496	0.0446	0.0464	0.0566	0.0512	0.0518
1, 2, 2)	Min-T	0.0552	0.0536	0.0552	0.0552	0.0538	0.0500	0.0620
	Gill's	0.0403	0.0420	0.0402	0.0413	0.0386	0.0407	0.0441
(1, 1, 2,	LRT	0.0558	0.0514	0.0516	0.0506	0.0482	0.0484	0.0526
1, 1, 2,	Max-T	0.0402	0.0458	0.0520	0.0494	0.0600	0.0552	0.0548
1, 1, 2)	Min-T	0.0544	0.0538	0.0520	0.0520	0.0516	0.0484	0.0524
	Gill's	0.0393	0.0392	0.0386	0.0396	0.0394	0.0396	0.0416

parameters, Gill's test is very conservative and has substantially deflated type I error rate. In Table 3, number of groups is large, that is, $k = 9$. It can be noticed that all the proposed tests attain the nominal level 0.05 for all the choices of sample sizes and scale parameters. Here also Gill's test is seen to be quite conservative. Therefore, we conclude that for any number of groups, parametric bootstrap based LRT, Min-T and Max-T behave as exact tests for any combinations of sample sizes and parametric configurations. However, Gill's test is quite conservative.

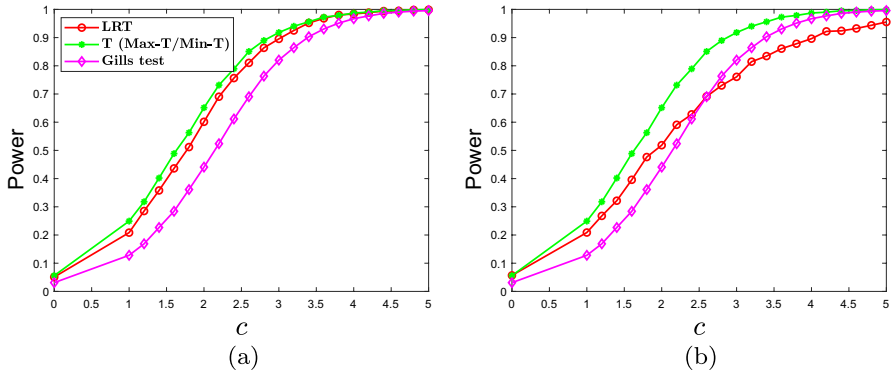

Fig. 1 Power curves of LRT, T (Min-T/Max-T) and Gill's test for treatment effects $(\mu_1, \mu_2) = c(1.2, 1.4)$ and sample sizes $(n_1, n_2) = (12, 14)$ when: **(a)** scale $(\sigma_1, \sigma_2) = (1, 2)$ **(b)** scale $(\sigma_1, \sigma_2) = (3, 2)$

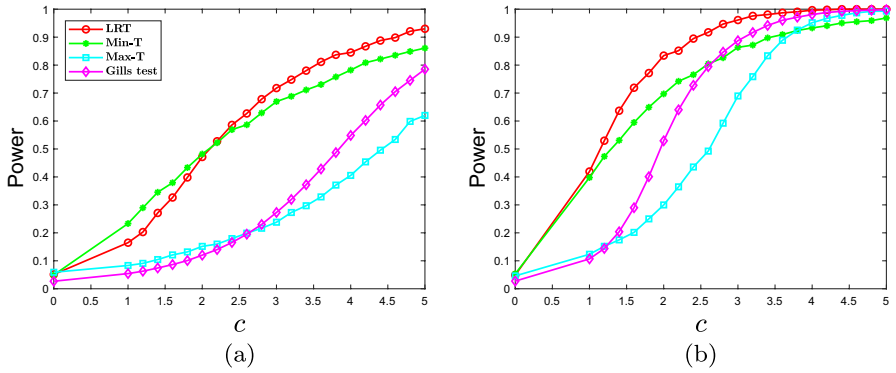

Fig. 2 Power curves of LRT, Min-T, Max-T, and Gill's test for homogeneous scale parameters (1,1,1) and treatment effects $(\mu_1, \mu_2, \mu_3) = c(1, 1.01, 1.02)$ when: **(a)** moderate sample sizes (40, 60, 50) **(b)** large sample sizes (70, 80, 100)

Remark 5 Figures 1, 2, 3, 4, 5, 6, and 7 show that the powers of all the tests increase as the value of c increases that is, when differences between μ_is increase. This shows good discriminatory power of the tests. Further, power values decrease if any σ_i increases. Power values increase as sample sizes increase. Therefore, all the proposed tests give expected power behavior.

Remark 6 From Figures 1, 2, 3, and 4, irrespective of sample sizes and parameter configurations, LRT is seen to perform better than all other tests. Among Max-T and Min-T, Min-T has higher power when μ_is are in strictly increasing order. An exception to this is seen in Fig. 4. Here Max-T is observed to perform better than the Min-T. This is because of the fact that the minimum difference 0.05 is very much less than the maximum difference of 1. It is seen that Gill's test has less power value than all the PB tests. However, for large sample sizes (70, 80, 100), Gill's test is

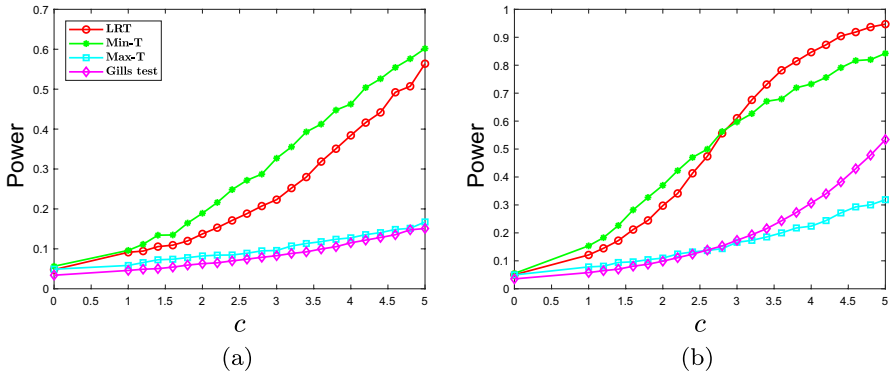

Fig. 3 Power curves of LRT, Min-T, Max-T, and Gill's test for unequal scale parameters are (1,2,3) and treatment effects $(\mu_1, \mu_2, \mu_3) = c(1, 1.01, 1.02)$ when: **(a)** moderate sample sizes (40, 60, 50) **(b)** large sample sizes (70, 80, 100)

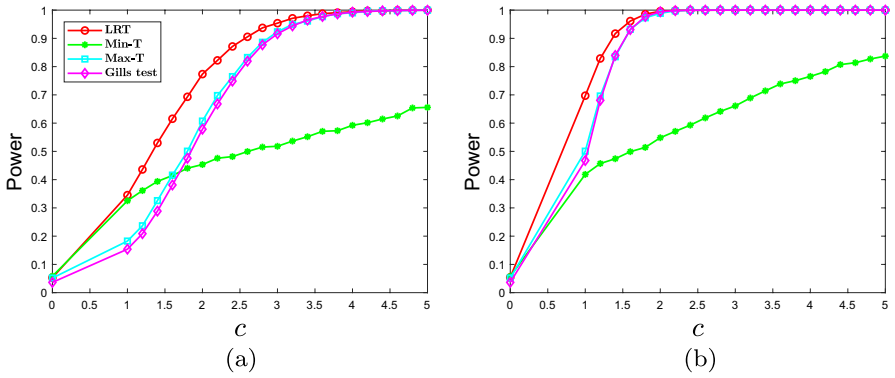

Fig. 4 Power curves of LRT, Min-T, Max-T, and Gill's test for unequal scale parameters (2,4,6) and treatment effects $(\mu_1, \mu_2, \mu_3) = c(2, 2.05, 3)$ when: **(a)** sample sizes (8, 10, 12) **(b)** moderate sample sizes (15, 18, 20)

observed to give better power values than the Max-T test but less than those of LRT. Similarly, in Fig. 4, it has higher power than that of the Min-T test, but still less than that of LRT.

Remark 7 In Figs 5 and 6, where two μ_is are taken to be equal (that is, $(\mu_1, \mu_2, \mu_3) = c(1, 1.02, 1.02)$), it is observed that the Min-T test does not perform well. The power of Min-T increases up to a certain value (not 1) and gets stabilized there both for equal (small) and unequal (large) scale parameters. When sample sizes are (40, 60, 50) and (20, 30, 25), initially, power values of Min-T are observed to increase faster than LRT and Max-T, but at a certain value of c, it gets stabilized. However, the power of the LRT and Max-T increases, and they become superior to the Min-T test. The same thing happens for the other two combinations of sample

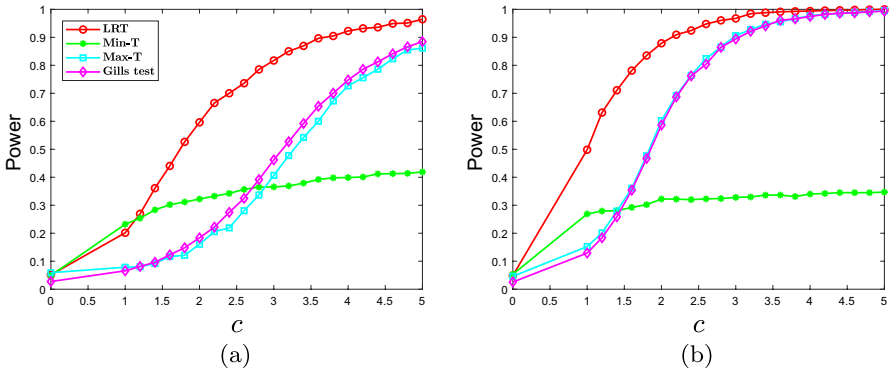

Fig. 5 Power curves of LRT, Min-T, Max-T, and Gill's test for homogeneous unequal scale parameter (1,1,1) and treatment effects $(\mu_1, \mu_2, \mu_3) = c(1, 1.02, 1.02)$ when: **(a)** moderate sample sizes (40, 60, 50) **(b)** large sample sizes (70, 80, 100)

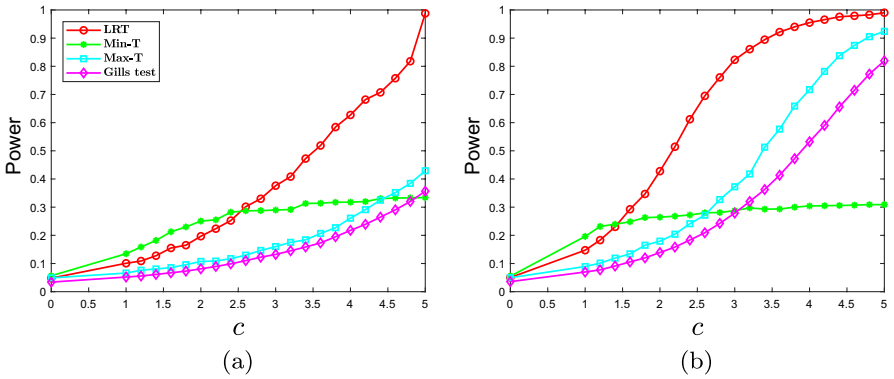

Fig. 6 Power curves of LRT, Min-T, Max-T, and Gill's test for homogeneous unequal scale parameter (1,2,3) and treatment effects $(\mu_1, \mu_2, \mu_3) = c(1, 1.02, 1.02)$ when: **(a)** moderate sample sizes (40, 60, 50) **(b)** large sample sizes (70, 80, 100)

sizes. However, in these cases, almost from the initial value of c power of Min-T is less than LRT and Max-T. The power of Gill's test is observed to be less than the other tests for all the combinations of sample sizes and scales. Although the power of the Min-T test gets stabilized for some value but the power of Gill's test increases. Hence, for a large sample or a large value of c, Gill's test becomes superior to the Min-T test in case of at least one equality in the combination $c(\mu_1, \mu_2, \cdots, \mu_k)$. However, LRT remains superior in these cases too.

Remark 8 For $k = 4$, it is observed from Fig. 7 (μ_is are taken in strictly increasing order), that LRT has more power than all others when scale parameters are equal. However, when scale parameters are distinct, Min-T has more power than LRT in

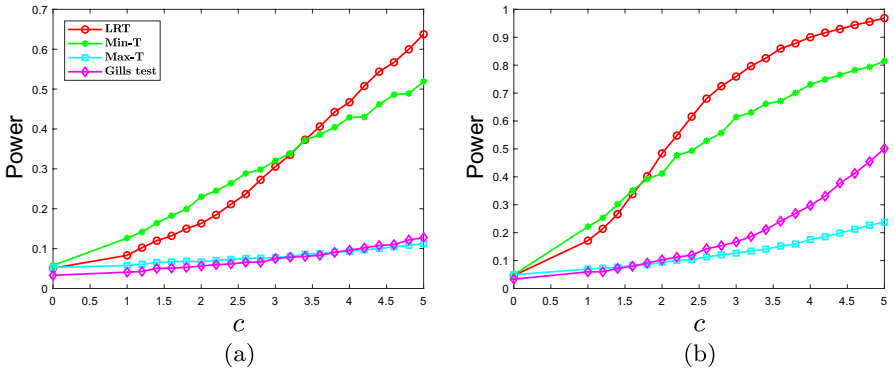

Fig. 7 Power curves of LRT, Min-T, Max-T, and Gill's test for homogeneous unequal scale parameter (1,1.2,1.4,2), when $(\mu_1, \mu_2, \mu_3) = c(1, 1.01, 1.02, 1.03)$: **(a)** for moderate sample sizes (20, 30, 25, 20) **(b)** for large sample sizes (40, 60, 50, 40)

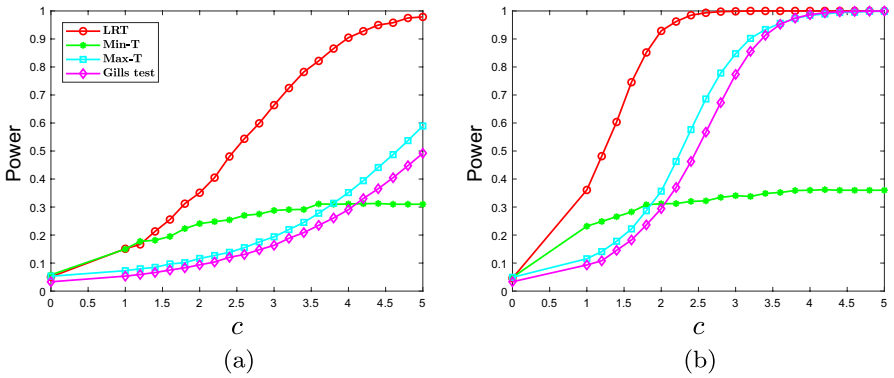

Fig. 8 Power curves of LRT, Min-T, Max-T, and Gill's test for homogeneous unequal scale parameter (1,1.2,1.4,2) and treatment effects $(\mu_1, \mu_2, \mu_3) = c(1, 1, 1.04, 1.05)$ when: **(a)** moderate sample sizes (20, 30, 25, 20) **(b)** large sample sizes (40, 60, 50, 40)

some cases for small and moderate samples. It is seen that among Min-T and Max-T, Min-T is powerful. For moderate and large samples, the power values of Gill's test are higher than that of the Max-T in some cases but less than those of LRT and Min-T.

Remark 9 In Fig. 8, two μ_is are taken to be equal, that is, $\mu_1 = \mu_2$. Again here, LRT has more power than all other tests. For moderate and large samples, Gill's test and Max-T are more powerful than Min-T. This is due to the fact that they discriminate better when at least one inequality is there, whereas Min-T is better when all effects are distinct. Similar behavior of all tests is observed from Tables 4, 5, 6 and 7 also. These configurations have been considered in Gill et al. (2017). From Table 8, it can be seen that in attaining the preassigned coverage probability 0.95, the proposed

Table 4 Power values of LRT, Max-T, Min-T and Gill's test when $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1, 1.1, 1.2, 1.3)$

$(\mu_1, \mu_2, \mu_3, \mu_4)$	m	LRT	Min-T	Max-T	Gill's test
(0, 0, 0, 0.3)	10	0.3870	0.1800	0.1896	0.1427
	20	0.8452	0.1934	0.7916	0.7115
	30	0.9990	0.1862	0.9989	0.9975
(0, 0.1, 0.2, 0.3)	10	0.7062	0.4762	0.1028	0.1574
	20	0.9852	0.7750	0.2588	0.6819
	30	0.9996	0.9056	0.5414	0.9819
(0, 0, 0, 0.4)	10	0.5226	0.1864	0.3238	0.2515
	20	0.9706	0.1936	0.9560	0.9539
	30	1	0.1882	1	1
(0, 0.2, 0.3, 0.4)	10	0.8896	0.5700	0.1536	0.3090
	20	0.9978	0.8112	0.6070	0.9417
	30	1	0.9362	0.4562	0.9990
(0, 0, 0, 0.5)	10	0.6330	0.1886	0.4952	0.4121
	20	0.9990	0.1890	0.9982	0.9974
	30	1	0.1930	1	1
(0, 0.3, 0.4, 0.5)	10	0.9576	0.5876	0.2880	0.4980
	20	0.9996	0.8354	0.9354	0.9901
	30	1	0.9288	0.9968	1

Table 5 Power values of LRT, Max-T, Min-T and Gill's test when $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1, 1.1, 1.2, 1.3)$

$(\mu_1, \mu_2, \mu_3, \mu_4)$	(n_1, n_2, n_3, n_4)	LRT	Min-T	Max-T	Gill's test
(0, 0, 0, 0.3)	(21, 23, 24, 26)	0.9792	0.1916	0.9492	0.7808
	(21,22,23,24)	0.9632	0.1990	0.9108	0.7898
	(31,31,31,31)	0.9992	0.2010	0.9944	0.9922
(0, 0.1, 0.2, 0.3)	(21,23,24,26)	0.9948	0.819	0.3538	0.7609
	(21,22,23,24)	0.9942	0.8202	0.3440	0.7350
	(31,31,31,31)	0.9998	0.9208	0.5708	0.9711
(0, 0, 0, 0.4)	(16,18,19,20)	0.9768	0.1894	0.9634	0.7714
	(21,21,21,21)	0.9824	0.1956	0.9752	0.9269
	(30,30,30,31)	1	0.2008	1	1
(0, 0.2, 0.3, 0.4)	(16,17,19,20)	0.9942	0.7832	0.4542	0.7840
	(21,21,21,21)	0.9984	0.8464	0.6568	0.9121
	(30,30,30,30)	1	0.9272	0.9564	0.9983
(0, 0, 0, 0.5)	(14, 15, 16, 17)	0.9772	0.1844	0.9640	0.7691
	(21,21,21,21)	0.9988	0.1822	0.9980	0.9929
	(30,30,30,30)	1	0.2026	1	1
(0, 0.3, 0.4, 0.5)	(14, 15, 16, 17)	0.9956	0.1900	0.6806	0.8228
	(21,21,21,21)	1	0.7368	0.9522	0.9873
	(30,30,30,30)	1	0.8664	0.9978	0.9999

Table 6 Power values of LRT, Max-T, Min-T and Gill's test when $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1, 1, 1, 1)$

$(\mu_1, \mu_2, \mu_3, \mu_4)$	m	LRT	Min-T	Max-T	Gill's test
(0, 0, 0, 0.3)	10	0.5198	0.1744	0.3200	0.2351
	30	1	0.1876	0.9996	1
(0, 0.1, 0.2, 0.3)	10	0.7946	0.5272	0.1180	0.2310
	30	1	0.9274	0.7462	0.9974
(0, 0, 0, 0.4)	10	0.6880	0.1922	0.5338	0.4293
	20	0.999	0.1802	0.9964	0.9976
	30	1	0.1870	1	1
(0, 0.2, 0.3, 0.4)	10	0.9252	0.6130	0.1780	0.4436
	20	0.9990	0.8602	0.7284	0.9832
	30	1	0.9492	0.9842	1
(0, 0, 0, 0.5)	10	0.8016	0.1820	0.7496	0.6238
	20	0.9998	0.1866	0.9998	1
	30	1	0.1900	1	1
(0, 0.3, 0.4, 0.5)	10	0.9722	0.6322	0.3302	0.6695
	20	1	0.8778	0.9548	0.9970
	30	1	0.9536	0.9988	1

Table 7 Power values of LRT, Max-T, Min-T and Gill's test when $(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1, 1, 1, 1)$

$(\mu_1, \mu_2, \mu_3, \mu_4)$	(n_1, n_2, n_3, n_4)	LRT	Min-T	Max-T	Gill's test
(0, 0, 0, 0.3)	(21,21,21,21)	0.9852	0.1856	0.9642	0.7804
	(31,31,30,31)	1	0.2048	0.9998	1
(0, 0.1, 0.2, 0.3)	(21,21,21,21)	0.9972	0.8366	0.3622	0.7627
	(30,30,30,30)	0.9998	0.9288	0.7330	0.9961
(0, 0, 0, 0.4)	(16,16,16,16)	0.9752	0.1822	0.9562	0.7747
	(21,21,21,21)	0.9994	0.1820	0.9984	0.9925
	(30,30,30,30)	1	0.1870	1	1
(0, 0.2, 0.3, 0.4)	(16,16,16,16)	0.9930	0.7950	0.4544	0.7976
	(21,21,21,21)	0.9996	0.8694	0.7808	0.9718
	(30,30,30,30)	1	0.9514	0.9810	0.9997
(0, 0, 0, 0.5)	(14,14,14,14)	0.9754	0.1896	0.9712	0.7897
	(21,21,21,21)	1	0.1928	0.9998	1
	(30,30,30,30)	1	0.1894	1	1
(0, 0.3, 0.4, 0.5)	(14,14,14,14)	0.9972	0.7694	0.6816	0.8421
	(21,21,21,21)	1	0.8784	0.9714	0.9964
	(30,30,30,30)	1	0.9514	0.9994	1

simultaneous confidence interval using the Max-T performs better than that of Gill et al. (2017).

Table 8 The coverage probability

Equal sample	$(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1, 1.2, 1.3, 1.4)$		$(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = (1, 1, 1, 1)$	
	Max-T	Gills	Max-T	Gills
10	0.9521	0.9628	0.9508	0.9626
11	0.9536	0.9618	0.9538	0.9560
12	0.9560	0.9596	0.9536	0.9583
13	0.9540	0.9642	0.9536	0.9617
14	0.9544	0.9631	0.9534	0.9662
15	0.9470	0.9604	0.9522	0.9609
20	0.9488	0.9590	0.9562	0.9612
25	0.9488	0.9622	0.9512	0.9637
30	0.9572	0.9630	0.9538	0.9641
35	0.9528	0.9635	0.9492	0.9608

Remark 10 Therefore, as a conclusion, we say that when μ_i's are in strict order LRT and Min-T performs well for small to large sample sizes and even for very small difference among μ_i's. For large sample sizes and small scales, Max-T also performs well. When there is any equality among μ_i's LRT and Max-T are desirable to use. Although sometimes Gill's test gives more power than Min-T or Max-T, it is not desirable as it does not control the size and behaves quite conservatively.

We have considered small differences in values of μ_is in tables to show the actual sensitivity of different test procedures to departure from the null hypothesis. For large differences all tests are good.

Remark 11 In this paper, the tests have been developed for ordered alternatives. Therefore, in the case of alternatives not obeying monotonic increasing order, these will not perform so well. For example, in the table below, we have reported the power values of the proposed tests along with the existing Gill's test for ordered and non-ordered parameters. Here, it can be seen that for the reverse order value of (μ_1, μ_2, μ_3), say, (2, 1.5, 1), all the tests give zero power value. For (1.5,2,1), the Max-T gives a power value of 0.4314, although it is less than the case of (1,1.5,2). The Max-T test rejects null hypothesis when there is at least one order among μ_i's. Hence, in this case, it does not give zero power (Table 9).

Table 9 Power values of tests for ordered and non-ordered treatment effects

(μ_1, μ_2, μ_3)	LRT	Min-T	Max-T	Gill's test
(1,1.5,2)	0.994	0.847	0.914	0.964
(2,1.5,1)	0	0	0	0
(1.5,2,1)	0	0	0.4314	0.1153

Remark 12 We have published an 'R' Package, namely 'Exptest' in the open source GitHub account 'AnjanaStat.' Inside this package, the functions LRT, MinT and MaxT, expSCI are included. The functions LRT, MinT, and MaxT are developed for finding test statistic value and critical value of the tests LRT, Min-T, and Max-T, respectively. The link to access the codes for this package is: https://github.com/AnjanaStat/Exptest. Data sets are to be imported in a systematic way. The groups with increasing order of location parameters are to be placed in increasing columns of the 'excel sheet.' This is how the data set is to be imported in 'R.' The output of each of these functions is a vector, whose first component is the test statistic value and the second is the corresponding critical value. The function for finding simultaneous confidence intervals is 'expSCI.' It gives the lower bounds for the successive pairwise differences of location parameters.

7 Numerical examples

Example 1 We consider the Tumor data given in Table 3 of Gill et al. (2017). We take the survival times of patients with tumor types Squamous, Small, Adeno, and Large as group 1, 2, 3, and 4, respectively. We obtain the test statistic value (critical value) of LRT, Max-T, and Min-T, respectively, as 0.00153 (0.0119), 9.1127 (4.8099), and −1.5237 (0.0524). As the test statistic value of LRT is less than the critical value, LRT rejects H_0. Max-T also rejects H_0, since the test statistic value is greater than the critical value. However, Min-T could not reject the null hypothesis here. This is due to the fact that there is some equality among μ_is. The simultaneous confidence intervals obtained by using Max-T test are as follows: $\mu_2 - \mu_1 \in [-15.2688, \infty)$, $\mu_3 - \mu_2 \in [-42.5006, \infty)$, $\mu_4 - \mu_3 \in [45.6562, \infty)$. From this, it is clear that $\mu_1 = \mu_2 = \mu_3 < \mu_4$.

Note that Gill's test also rejects the null hypothesis. Hence, the minimum survival times of patients for the four tumor types are in increasing order.

Example 2 We consider the data set consisting of the duration of remission by four drugs from Wu and Wu (2005). The observations corresponding to each drug are two-parameter exponential distributed. It is noted that the scale parameters are different. Let μ_i denote the minimum duration of remission time of Drug i, $i = 1, 2, 3, 4$. We would like to test $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ against $H_1 : \mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4$ with

Table 10 Test statistic and critical values

Tests	Test statistic value	Critical value	Decision
LRT	$5.4897e^{-16}$	$1.0590e^{-02}$	Rejected
Min-T	5.4468	0.1829	Rejected
Max-T	13.8729	3.6098	Rejected
Gill's test	$W_2 = 15.6955$	4.5479	Rejected
	$W_3 = 12.7293$		
	$W_4 = 17.1034$		

Table 11 Simultaneous confidence intervals

Max-T		Gill's test	
Differences	SCI	Differences	SCI
$\mu_2 - \mu_1$	$[0.8986, \infty)$	$\mu_2 - \mu_1$	$[0.8530, \infty)$
$\mu_3 - \mu_2$	$[0.3071, \infty)$	$\mu_3 - \mu_2$	$[0.1217, \infty)$
		$\mu_3 - \mu_1$	$[0.3227, \infty)$
$\mu_4 - \mu_3$	$[0.6277, \infty)$	$\mu_4 - \mu_3$	$[0.5003, \infty)$
		$\mu_4 - \mu_2$	$[1.3573, \infty)$
		$\mu_4 - \mu_1$	$[2.5583, \infty)$

at least one strict inequality. The test statistic and critical values of the tests and the decision are presented in Table 10. The simultaneous confidence intervals for the successive differences of location parameters are given in Table 11.

From Tables 10 and 11, it is clear that the minimum duration of remission for drugs 1, 2, 3, and 4 is strictly increasing.

Table 12 Number of positive cases of Covid-19 patients

Odisha			Karnataka			Delhi		
20	737	2994	128	1056	3796	576	6923	22424
42	828	3140	175	1092	4063	720	7353	23645
42	876	3250	181	1135	4123	903	7639	26334
48	978	3386	197	1147	4212	1069	7658	27654
50	1020	3477	232	1152	4320	1154	7998	28936
51	1030	3723	247	1234	4688	1561	8470	29943
54	1052	3909	260	1246	5921	1578	8895	31309
54	1103	4055	279	1328	6245	1640	9333	34687
60	1189	4163	315	1462	6516	1707	9755	36824
68	1269	4338	384	1605	6824	2003	10054	41182
94	1517	4512	390	1743	7000	2081	10231	42829
103	1593	4677	408	1959	7530	2156	10554	44688
111	1660	4856	427	2089	7734	2248	11659	49979
122	1723	5160	445	2182	7944	2375	12141	56746
157	1819	5303	454	2283	8281	2514	14053	59746
169	1948	5470	474	2343	8697	2625	14465	62655
209	2158	5752	503	2418	9150	3314	15257	66602
294	2213	5962	512	2567	9721	3439	16281	70390
538	2245	6180	535	2781	10560	4122	16758	77240
611	2312	6350	565	2922	11005	4549	17386	83077
640	2388	6614	601	3221	13190	5104	18742	85161
670	2478	7065	614	3243	14295	5532	19844	87360
672	2564		705	3408		5980	19873	
682	2608		925	3421		6542	22132	

Table 13 Test statistic and critical values

Tests	Test statistic value	Critical value	Decision
LRT	0.0135	0.0164	Rejected
Min-T	1.5638	0.9620	Rejected
Max-T	2.0670	3.2102	Not rejected
Gill's test	$W_2 = 2.4167$	4.2002	Not rejected
	$W_3 = 1.8938$		

Example 3 We choose random observations on positive Covid-19 cases of patients for March, April and June 2020. We have considered the states of Odisha, Karnataka, and Delhi. The raw data set contains several others states. It is available in https://www.kaggle.com/datasets/sudalairajkumar/covid19-in-india. The sampled data set is given in Table 12. Let μ_1, μ_2 and μ_3 denote the minimum number of positive cases in Odisha, Karnataka and Delhi, respectively. The q-q plots yield that observations in the three cities follow two-parameter exponential distributions. The quadratic test of Nelson (1982) and the modified marginal likelihood ratio test of Thiagarajah and Paul (1990) show that the scale parameters are unequal. Our aim is to test that $H_0 : \mu_1 = \mu_2 = \mu_3$ against $H_1 : \mu_1 \leq \mu_2 \leq \mu_3$ with at least one strict inequality.

The test statistic and critical values are written in Table 13. Here, we observe that LRT and Min-T reject the null hypothesis. However, the null hypothesis is not rejected by Max-T and Gill's test. It is seen in Section 6 that the LRT and Min-T tests have higher power values than those of the Max-T and Gill's tests, when treatment effects are in a strictly increasing order. Due to this, here the Max-T and Gill's tests do not reject the null hypothesis. Therefore, the decisions of LRT and Min-T are preferred in this case. A rejection by Min-T indicates a strictly increasing order among μ_is (Table 1).

8 Concluding remark

In this paper, we have considered the testing problem of equality of location parameters of several exponential populations against ordered alternatives. We have developed likelihood ratio test and two heuristic tests based on multiple comparisons of successive differences between treatment effects, namely Max-T and Min-T. Moreover, simultaneous confidence intervals for successive pairwise differences of ordered treatment effects are evaluated. Parametric bootstrap is used for implementation of the tests and accordingly, an 'R' package has been published in GitHub. Asymptotic accuracy of parametric bootstrap is established for all three test procedures. Extensive simulation study shows that the tests LRT, Max-T and Min-T achieve nominal size values for small, medium, highly unbalanced sample sizes. However, a few exceptions occur for very small sample sizes. We have also compared our test

procedures with an existing test. The existing test is observed to be conservative all the time. Power comparison also shows that existing test has lower power values than those of proposed test procedures. For some particular configurations of parameters, it gives higher power values than Max-T or Min-T. All the tests are implemented using three real data sets. In one data set, it is seen that although LRT and Min-T tests reject the null hypothesis, the existing test cannot reject it because of lower power.

Acknowledgements The authors are sincerely thankful to editor-in-chief, associate editor and two reviewers for their significant suggestions which have helped in improving the manuscript.

References

Barolow, R. E., Bartholomew, D. J., Bremner, J. M., Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*. New York: Wiley.

Bickel, P. J., Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics, 9*(6), 1196–1217.

Bretz, F. (2006). An extension of the Williams trend test to general unbalanced linear models. *Computational Statistics and Data Analysis, 50*(7), 1735–1748.

Chen, H. J. (1982). A new range statistic for comparisons of several exponential location parameters. *Biometrika, 69*, 257–260.

Chuang-Stein, C., Agresti, A. (1997). Tutorial in biostatistics: A review of tests for detecting a monotone dose-response relationship with ordinal response data. *Statistics in Medicine, 16*(22), 2599–2618.

Dhawan, A. K., Gill, A. N. (1997). Simultaneous one-sided confidence intervals for the ordered pairwise differences of exponential location parameters. *Communications in Statistics-Theory and Methods, 26*, 247–262.

Donaldson, T. S. (1968). Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F ratio. *Journal of the American Statistical Association, 63*, 660–667.

Epstein, B., Tsao, C. K. (1953). Some tests based on ordered observations from two exponential populations. *The Annals of Mathematical Statistics, 24*, 458–466.

Fries, A., Bhattacharyya, G. K. (1983). Analysis of two-factor experiments under an inverse Gaussian model. *Journal of the American Statistical Association, 78*, 820–826.

Gill, A. N., Goyal, A., Maurya, V. (2017). Simultaneous inferences for ordered exponential location parameters under unbalanced data and heteroscedasticity of scale parameters. *Communications in Statistics - Simulation and Computation, 46*, 3129–3139.

Hasler, M., Hothorn, L. A. (2008). Multiple contrast tests in the presence of heteroscedasticity. *Biometrical Journal Journal of Mathematical Methods in Biosciences, 50*(5), 793–800.

Hayter, A. J. (1990). A one-sided studentized range test for testing against a simple ordered alternative. *Journal of the American Statistical Association, 85*(411), 778–785.

Hsieh, H. K. (1986). An exact test for comparing location parameters of k exponential distributions with unequal scales based on type II censored data. *Technometrics, 28*, 157–164.

Konietschke, F., Hothorn, L. A. (2012). Evaluation of toxicological studies using a nonparametric Shirley-type trend test for comparing several dose levels with a control group. *Statistics in Biopharmaceutical Research, 4*(1), 14–27.

Konietschke, F., Bathke, A. C., Harrar, S. W., Pauly, M. (2015). Parametric and non parametric bootstrap methods for general MANOVA. *Journal of Multivariate Analysis, 140*, 291–301.

Liu, W., Miwa, T., Hayter, A. J. (2000). Simultaneous confidence interval estimation for successive comparisons of ordered treatment effects. *Journal of Statistical Planning and Inference, 88*, 75–86.

Marcus, R. (1976). The powers of some tests of the equality of normal means against an ordered alternative. *Biometrika, 5*(1), 37–42.

Maurya, V., Goyal, A., Gill, A. N. (2011). Simultaneous testing for the successive differences of exponential location parameters under heteroscedasticity. *Statistics and Probability Letters, 81*, 1507–1517.

Mondal, A., Pauly, M., Kumar, S. (2023). Testing for ordered alternatives in heteroscedastic ANOVA under normality. *Statistical Papers, 64*(6), 1913–1941.

Mondal, A., Sattler, P., Kumar, S. (2023). Testing against ordered alternatives in a two-way model without interaction under heteroscedasticity. *Journal of Multivariate Analysis., 196*, 105177. https://doi.org/10.1016/j.jmva.2023.105177

Nelson, W. (1982). *Applied Life Data Analysis*. New York: Wiley.

Pauly, M., Brunner, E., Konietschke, F. (2015). Asymptotic permutation tests in general factorial designs. *Journal of the Royal Statistical Society Series B Statistical Methodology, 77*(2), 461–473.

Pauly, M., Asendorf, T., Konietschke, F. (2016). Permutation-based inference for the AUC: A unified approach for continuous and discontinuous data. *Biometrical Journal, 58*(6), 1319–1337.

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

Şenoğlu, B., Tiku, M. L. (2001). Analysis of variance in experimental design with nonnormal error distributions. *Communications in Statistics-Theory and Methods, 30*, 1335–1352.

Shirley, E. (1997). A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. *Biometrics, 33*, 386–389.

Singh, N. (1985). A simple and asymptotically optimal test for the equality of $k(\geq 2)$ exponential distributions based on type II censored samples. *Communications in Statistics-Theory and Methods, 14*, 1615–1625.

Singh, N., Narayan, P. (1983). The Likelihood ratio test for the equality of two parameter exponential distributions based on type II censored samples. *Journal of Statistical Computation and Simulation, 18*(4), 287–297.

Singh, P., Singh, N. (2013). Simultaneous confidence intervals for ordered pairwise differences of exponential location parameters under heteroscedasticity. *Statistics and Probability Letters, 83*, 2673–2678.

Singh, P., Abebe, A., Mishra, S. N. (2006). Simultaneous testing for the successive differences of exponential location parameters. *Communications in Statistics - Simulation and Computation, 35*, 547–561.

Thiagarajah, K., Paul, S. R. (1990). Testing for the equality of scale parameters of $k(\geq 2)$ exponential populations based on complete and type II censored samples. *Communications in Statistics - Simulation and Computation, 19*(3), 891–902.

Vijayasree, G., Misra, N., Singh, H. (1995). Componentwise estimation of ordered parameters of $k(\geq 2)$ exponential populations. *Annals of the Institute of Statistical Mathematics, 47*, 287–307.

Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics, 27*(1), 103–117.

Williams, D. A. (1972). The comparison of several dose levels with a zero dose control. *Biometrics, 28*, 519–531.

Wu, S. F., Wu, C. C. (2005). Two stage multiple comparisons with the average for exponential location parameters under heteroscedasticity. *Journal of Statistical Planning and Inference, 134*, 392–408.