

Article

Empirical Likelihood Ratio Tests for Homogeneity of Multiple Populations in the Presence of Auxiliary Information

Ronghuo Wu^{1,2} and Yongsong Qin^{3,*}¹ Department of Mathematics and Statistics, Yulin Normal University, Yulin 537000, China; rhw@ylu.edu.cn² Key Laboratory of Complex System Optimization and Big Data Processing in Department of Guangxi Education, Yulin Normal University, Yulin 537000, China³ Department of Statistics, Guangxi Normal University, Guilin 541004, China

* Correspondence: ysqin@gxnu.edu.cn; Tel.: +86-135-9732-2048

Abstract: The empirical likelihood ratio test (ELRT) statistic is constructed for testing the homogeneity of several nonparametric populations in the presence of some auxiliary information. It is shown—under some regularity conditions and under the null hypothesis that all distribution functions of the populations are equal—that the asymptotic distribution of the ELRT is a chi-squared distribution. The proposed ELRT could be more powerful than the Kruskal–Wallis test, as extra information can be efficiently employed by ELRT. The advantage of ELRT over T&P (2006) is that researchers do not need to select approximately normal statistics for inter-group comparisons, and ELRT is more suitable for the multi-population consistency test with a small sample size.

Keywords: test of homogeneity; estimating equation; empirical likelihood ratio test; auxiliary information; Kruskal–Wallis test

MSC: 62E20



Citation: Wu, R.; Qin, Y. Empirical Likelihood Ratio Tests for Homogeneity of Multiple Populations in the Presence of Auxiliary Information. *Mathematics* **2022**, *10*, 2341. <https://doi.org/10.3390/math10132341>

Academic Editor: Alicia Nieto-Reyes

Received: 24 May 2022

Accepted: 2 July 2022

Published: 4 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Suppose that there are k ($k \geq 2$) populations, and the distribution function of the i -th population is $F(x; \theta_i)$ ($1 \leq i \leq k$), where $\theta_i \in A \subseteq R^p$, θ_i are parameter vectors ($1 \leq i \leq k$), and A is the parameter space. In other words, the k populations share the same type of distribution but may have different structures as θ_i ($1 \leq i \leq k$) varies. Consider the hypothesis

$$H_0 : \theta_1 = \theta_2 = \dots = \theta_k. \quad (1)$$

This test for homogeneity arises, for example, in the comparison of a number of different treatments, processes, varieties, or locations, when one wishes to test whether these differences have any effect on an outcome X , where X can be a scalar or a vector.

If F is the normal distribution function, the standard analysis of variance (ANOVA) for testing the above hypothesis has been widely used by a number of investigators. For example, Dou [1] employed this method in a parametric study of a developed statistical model. However, the standard ANOVA is not suitable for other distributions.

Due to the complexity of the real world, the form of F may not be known in many applications. In this nonparametric setting, the Kruskal–Wallis test (KWT) provides tests of the null hypothesis that independent samples from two or more groups come from identical populations. Refer to Lehmann [2] for the theory and applications of KWT.

Here, we provide a brief definition for KWT and its limiting distributions. First, the data of all samples in a single series are arranged in an ascending order, and a rank is assigned to each data in the ascending order too. In the case of a repeated value, or a tie, assign ranks to them by averaging their rank position. For example, if the sample number is even, the rank of the median is the average rank of the two numbers before and after it. The KWT statistics for the k independent samples, each of size n_i , is

$$T = \frac{1}{s^2} \sum_{i=1}^k \left\{ \frac{R_i^2}{n_i} - \frac{n(n+1)^2}{4} \right\},$$

where $n = \sum_{i=1}^k n_i$, and R_i is the sum of the ranks (from all samples pooled). For the i -th sample, we have

$$s^2 = \frac{1}{n-1} \sum_{i=1}^k \sum_{j=1}^{n_i} \left\{ R_{ij}^2 - \frac{n(n+1)^2}{4} \right\},$$

where R_{ij} is the rank (from all samples pooled) of the j -th observation in the i -th sample. The null hypothesis of this test is that all k distribution functions are equal. It is shown, under the null hypothesis and some regularity conditions, that

$$T \xrightarrow{d} \chi_{k-1}^2.$$

As mentioned above, when the form of the F may not be known in many applications. KWT can be used to perform consistency tests for multiple populations. As KWT constructs statistics based on sample rank, its test efficacy is good when the sample size is large. However, when the sample size is small, the statistics constructed based on sample rank carry much less sample information. In other words, KWT is obviously going to be a lot worse. So we introduce the empirical likelihood method; here, we provide a brief definition for it.

The empirical likelihood method as a nonparametric technique for statistical inference in the nonparametric setting was introduced by Owen [3,4] and has many advantages over other nonparametric test methods such as the normal-approximation-based method and the bootstrap method, as put forward by Hall and La Scala [5] and Hall [6]. The Wilks' theorem, Bartlett correction and the ability of using auxiliary information are three striking properties of the empirical likelihood methods. Chen and Qin [7] proved that the empirical likelihood method can be seamlessly applied to finite population estimation problems, and more accurate statistical inference can be obtained through the effective use of auxiliary information. Zhang [8] developed a new class of M function estimators and quantile estimators with some auxiliary information, using the empirical likelihood technique. A natural question is whether and how an empirical likelihood method can efficiently use the auxiliary information to decide whether several samples should be regarded as those that come from the same population. In this paper, an empirical likelihood ratio test (ELRT) statistics is constructed for testing the homogeneity of several nonparametric populations in the presence of some auxiliary information. Since the auxiliary information is not employed in the KWT method, KWT may be less powerful than ELRT in the field of population distribution consistency. A comprehensive comparison between ELRT and KWT was conducted and is presented in Section 3.

We note that there exist a few other approaches which allow to incorporate auxiliary information in statistics testing. For example, the method based on the auxiliary information in a form of vectors of unbiased estimates in Tarima and Pavlov [9] (T&P (2006)) may be used in the context of this article. The asymptotic properties of the above work are analyzed by Albertus [10]. Tarima and Pavlov used additional information to construct parameter estimation statistics, and completed parameter estimation by adding data sources. Therefore, we will compare ELRT and T&P (2006) separately in the numerical simulation part.

The form of F is unknown in the present study. However, it is assumed that some auxiliary information about the distribution function $F(x; \theta)$ is available in the sense that there exist r ($r > p$) known functions $g_1(x; \theta), g_2(x; \theta), \dots, g_r(x; \theta)$ such that

$$E\{g(X; \theta)\} = 0, \tag{2}$$

where $X \sim F(x; \theta)$ and $g(x; \theta) = (g_1(x; \theta), g_2(x; \theta), \dots, g_r(x; \theta))^T$ is an r -dimensional vector.

Equation (2) defines a group of estimating equations. Those equations are widely applicable and particularly powerful when the data model is not specified by a full parametric likelihood function, as elaborated by Hansen [11] and Godambe and Heyde [12] among many others. Qin and Lawless [13] showed that an empirical likelihood approach produces a semiparametric efficient parameter estimate. In this study, $r > p$ is required. Excellent explanations related to this requirement are given by Qin and Lawless [13] and Zhang [8]. More related results of statistical inference using the estimating equations can be found in Wang and Chen [14] and Zhou et al. [15], among others.

This assumption (2) is natural in practice; as in most commonly used distribution families, the distribution is usually determined by some of its moments, such as the mean, variance, skewness, kurtosis and so on. For example, if X is the amount of a type of grains and one suspects that there could be some differences in the amount of the grains, among several populations, caused by the amount of fertilizer, we may set $\theta_{(1)} = EX$ and $g_1(x; \theta_{(1)}) = x - \theta_{(1)}$. On the other hand, if one suspects that the use of the fertilizer may not only cause the change of amount of grains but also the change of the variance of X , then we may set $\theta_{(1)} = EX, \theta_{(2)} = EX^2, g_1(x; \theta) = x - \theta_{(1)}$ and $g_2(x; \theta) = X^2 - \theta_{(2)}$, where $\theta = (\theta_{(1)}, \theta_{(2)})'$. These could be initially assessed by comparing the histograms of the data sets of populations which are under consideration. In addition to the above (partial) information, we may know some extra information. For example, we may know some moments of X or may know that the distribution of X is symmetric about some points.

Based on (2), we will construct an empirical likelihood ratio test (ELRT) to test H_0 . It is shown that the limiting distribution of the ELRT under H_0 is $\chi^2_{(r-p)(k-1)}$, and thus the testing method for H_0 is ready to use, where $\chi^2_{(r-p)(k-1)}$ is the chi-squared random variable with $(r - p)(k - 1)$ degrees of freedom.

The rest of the paper is organized as follows. The main results of this study are presented in Section 2. Results of a simulation study on the finite sample performance of the ELRT are reported in Section 3. We conclude and give some remarks on our future work in Section 4. Finally, the proof of the main results is presented in Section 5.

2. Main Results

For $1 \leq i \leq k$, suppose that data $X_{ij}(j = 1, 2, \dots, n_i)$ are independently distributed as $F(x; \theta_i)$ (unknown) and that all $X_{ij}(j = 1, 2, \dots, n_i; i = 1, 2, \dots, k)$ are independent. Let $n = \sum_{i=1}^k n_i, \theta_i \in \Theta \subset R^p$ and Θ is the parameter space of θ and an open set of R^p . Let

$$B = \{p_{ij}, j = 1, 2, \dots, n_i; i = 1, 2, \dots, k | p_{ij} \geq 0, \sum_{i,j} p_{ij} = 1, \sum_{i,j} p_{ij} g(X_{ij}; \theta) = 0\},$$

$$B_i = \{q_{ij}, j = 1, 2, \dots, n_i | q_{ij} \geq 0, \sum_j q_{ij} = 1, \sum_j q_{ij} g(X_{ij}; \theta) = 0\}, 1 \leq i \leq k.$$

Here, p_{ij} is the probability mass, which represents the probability that the random variable $g(X, \theta)$ values $g(X_{ij}, \theta)$, both of which are non-negative, and the sum is 1. Similarly, q_{ij} represents the probability that the random variable $g(X, \theta)$ values $g(X_{ij}, \theta)$ when $1 \leq i \leq k$.

Applying the method proposed by Qin and Lawless [13], the ELRT for testing H_0 can be defined as

$$\lambda_n = \frac{\sup_{\theta \in \Theta} \sup_{\{p_{ij}\} \in B} \prod_{i=1}^k \prod_{j=1}^{n_i} (n p_{ij})}{\prod_i \left\{ \sup_{\theta \in \Theta} \sup_{\{q_{ij}\} \in B_i} \prod_{j=1}^{n_i} (n_i q_{ij}) \right\}}. \tag{3}$$

The ELRT rejects H_0 for large values of $-2 \log \lambda_n$.

For every $1 \leq i \leq k$, assume that 0 is in the convex hull of $\{g(X_{ij}; \theta), 1 \leq j \leq n_i\}$. Then, according to the Lagrange multiplier method, one can obtain

$$\sup_{\{p_{ij}\} \in B} \prod_{i=1}^k \prod_{j=1}^{n_i} (np_{ij}) = \prod_{i,j} \{1 + t'(\theta)g(X_{ij}; \theta)\}^{-1},$$

where $t(\theta)$ is the solution of the equation

$$\frac{1}{n} \sum_{i,j} \frac{g(X_{ij}; \theta)}{1 + t'(\theta)g(X_{ij}; \theta)} = 0. \tag{4}$$

Similarly, for $1 \leq i \leq k$,

$$\sup_{\{q_{ij}\} \in B_i} \prod_{j=1}^{n_i} (n_i q_{ij}) = \prod_j \{1 + t'_i(\theta)g(X_{ij}; \theta)\}^{-1},$$

where $t_i(\theta)$ is the solution of the equation

$$\frac{1}{n_i} \sum_j \frac{g(X_{ij}; \theta)}{1 + t'_i(\theta)g(X_{ij}; \theta)} = 0. \tag{5}$$

Hence,

$$\lambda_n = \frac{\sup_{\theta \in \Theta} \prod_{i,j} \{1 + t'(\theta)g(X_{ij}; \theta)\}^{-1}}{\prod_i \left\{ \sup_{\theta \in \Theta} \prod_j \{1 + t'_i(\theta)g(X_{ij}; \theta)\}^{-1} \right\}}. \tag{6}$$

The log-empirical likelihood function for data $X_{ij}(j = 1, 2, \dots, n_i)$ is therefore defined by

$$\ell_n(\theta) = \sum_{i=1}^k \sum_{j=1}^{n_i} \log \{1 + t'(\theta)g(X_{ij}; \theta)\}^{-1}. \tag{7}$$

Suppose that there is a $\hat{\theta}_n$ that maximizes $\ell_n(\theta)$, then the $\hat{\theta}_n$ is called the maximum empirical likelihood estimator (MELE) of θ . Suppose that, in addition, $\ell_n(\theta)$ is differentiable in θ , then $\hat{\theta}_n$ will be a solution of the empirical likelihood equation

$$\frac{\partial}{\partial \theta} \ell_n(\theta) = \left(\frac{\partial}{\partial \theta} \ell_n(\theta) \right)_{p \times 1} = \sum_{i,j} \frac{1}{1 + t'(\theta)g(X_{ij}; \theta)} \left(\frac{\partial g_u(X_{ij}; \theta)}{\partial \theta_v} \right)' t(\theta) = 0, \tag{8}$$

where $\frac{\partial g_u(X_{ij}; \theta)}{\partial \theta_v}$ is the (u, v) -element of the $r \times p$ matrix $\left(\frac{\partial g_u(X_{ij}; \theta)}{\partial \theta_v} \right)$.

Similarly, for $1 \leq i \leq k$, the log-empirical likelihood function for $X_{ij}(j = 1, 2, \dots, n_i)$ is defined by

$$\ell_{ni}(\theta) = \sum_{j=1}^{n_i} \log \{1 + t'_i(\theta)g(X_{ij}; \theta)\}^{-1}. \tag{9}$$

The MELE of θ under the i -th sample is denoted as $\hat{\theta}_{ni}$, which is a solution of the empirical likelihood equations

$$\frac{\partial}{\partial \theta} \ell_{ni}(\theta) = \sum_j \frac{1}{1 + t'_i(\theta)g(X_{ij}; \theta)} \left(\frac{\partial g_u(X_{ij}; \theta)}{\partial \theta_v} \right)' t_i(\theta) = 0. \tag{10}$$

We assume that all $\hat{\theta}_n$ and $\hat{\theta}_{ni}$ are consistent estimators of θ as $\min_{1 \leq i \leq k} n_i \rightarrow \infty$. Then the λ_n can be rewritten as

$$\lambda_n = \frac{\prod_{i,j} \{1 + t'(\hat{\theta}_n)g(X_{ij}; \hat{\theta}_n)\}^{-1}}{\prod_i \left\{ \prod_j \{1 + t'_i(\hat{\theta}_{ni})g(X_{ij}; \hat{\theta}_{ni})\}^{-1} \right\}} \tag{11}$$

Let X be a population with a distribution $F(x; \theta)$, θ_0 be the true value of θ , and $\|M\|$ be the L_2 -norm of a matrix M . To obtain the asymptotic distribution of λ_n , we need some regularity conditions as follows Qin and Lawless [13] (pp. 305–306):

(A) $E\{g(X; \theta_0)g'(X; \theta_0)\}$ is positive definite, $\partial g(X; \theta)/\partial \theta$ is continuous in a neighborhood of θ_0 , $\|\partial g(x; \theta)/\partial \theta\|$ and $\|g(x; \theta)\|^3$ are bounded by a function $G(x)$ in this neighborhood, and the rank of $E\{g(X; \theta_0)g'(X; \theta_0)\}$ is p , where $E\{G(X)\} < \infty$.

(B) $\partial^2 g(x; \theta)/\partial \theta \partial \theta'$ is continuous in θ in a neighborhood of θ_0 and $\|\partial^2 g(x; \theta)/\partial \theta \partial \theta'\|$ is bounded by a function $G(x)$ in this neighborhood, where $E\{G(X)\} < \infty$.

The main results of this study are presented as follows.

Theorem 1. Suppose assumptions (A) and (B) are satisfied, then, under H_0 , as $\min_{1 \leq i \leq k} n_i \rightarrow \infty$, we have

$$-2 \log \lambda_n \xrightarrow{d} \chi^2_{(r-p)(k-1)} \tag{12}$$

where $\chi^2_{(r-p)(k-1)}$ is the chi-squared random variable with $(r - p)(k - 1)$ degrees of freedom.

Remark 1. If we use λ_n in Equation (11) in stead of Equation (6) as the original definition, where $\hat{\theta}_n$ and $\hat{\theta}_{ni}$ are the roots of related likelihood equations, then Theorem 1 still holds true. This can be seen from the proof of Theorem 1. In other words, $\hat{\theta}_n$ and $\hat{\theta}_{ni}$ do not need to be the MELEs to have the results of Theorem 1.

To sum up, we constructed an ELRT statistic for testing the homogeneity of several nonparametric populations in the presence of some auxiliary information when the population distribution is unknown, and proved the asymptotic distribution of ELRT as a chi-square distribution under some regularity conditions when the null hypothesis is true. Next we will begin the numerical simulation. In this section, we will calculate the rejection rates of ELRT and compared with those of the Kruskal–Wallis test under several alternatives and compare the powers of them.

3. Simulation Results

Several commonly used distribution families were used in our simulations. The collective distribution and related parameter information are shown in Table 1.

Table 1. Distribution families investigated in simulations.

Density Function	Distribution Name	Notation	Parameter Space	Value
$p^x(1 - p)^{1-x}$	Bernoulli(p)	$b(p)$	$(0, 1)$	$x = 0, 1$
$\frac{\lambda^x}{x!} e^{-\lambda}$	Poisson(λ)	$P(\lambda)$	$(0, \infty)$	$x = 0, 1, \dots$
$\frac{1}{\theta} e^{-x/\theta}$	Exponential(θ)	$E(\theta)$	$(0, \infty)$	$(0, \infty)$

In this study, only three populations were compared. In the simulations, it was supposed that we only know the means of the populations. On the one hand, under the combination of sample size, we took the true value of the distribution under the null hypothesis to generate three distribution populations, and calculated the value of $-2 \log \lambda_n$. The simulation was repeated 5000 times to obtain 5000 corresponding $-2 \log \lambda_n$ values. Then, the quantiles of $-2 \log \lambda_n$ samples obtained were compared with the quantiles of the Chi-square distribution in Theorem 1. Finally, the Q-Q diagram of ELRT was made as well

as the Q-Q diagram of KWT under the same conditions (Figures 1–6). Here the abscissa is the theoretical quantile value, and the ordinate is the quantile value of the distribution population. It can be seen from Figures 1–6 that when the null hypothesis is true, the Q-Q diagrams of ERLT and KWT can prove that the asymptotic distribution of the test statistics given in this chapter obeys the Chi-square distribution when the null hypothesis is true.

At the same time, the simulated rejection rates of ELRT and KWT under several alternatives were compared using 5000 Monte Carlo trials with various sample sizes. It should be noted here that the rejection rate is calculated as follows:

$$reject.rate.KWT = \text{sum}(reslt.KWT > quantl.KWT) / m$$

where the *reslt. KWT* is calculated from the samples by KWT statistics, and *quantl. KWT* is calculated from the samples under the alternatives by Chi-square distribution, and *m* is the number of samples. The significant level was always set as 0.05 in the simulations. Results of these comparisons were reported in Table 2. In addition, we simulated the rejection rates of KWT and ELRT under the original hypothesis, and the results are shown in Table 3. From these results, it can be seen that the simulated powers are quite good for both tests, even for moderate sample sizes with better performance, as sample sizes increase and ELRT performs better than KWT.

On the other hand, we consider that T&P (2006) also performs parameter estimation research based on additional information, so we will separately compare the ELRT proposed by T&P (2006) in this paper. The results are shown in Table 4. We can see some interesting results from the comparison results. For example, T&P (2006) is more dependent on the normal sample, that is to say, when the comparison sample is biased to the normal sample, the test efficacy of T&P (2006) is very effective; when deviating from the normal condition, T&P (2006) showed poor test efficacy compared with ELRT. In other words, the advantage of ELRT over T&P (2006) is that researchers do not need to select approximately normal statistics for inter-group comparisons. At the same time, from the perspective of the sample size, ELRT is more suitable for the multi-population consistency test with a small sample size.

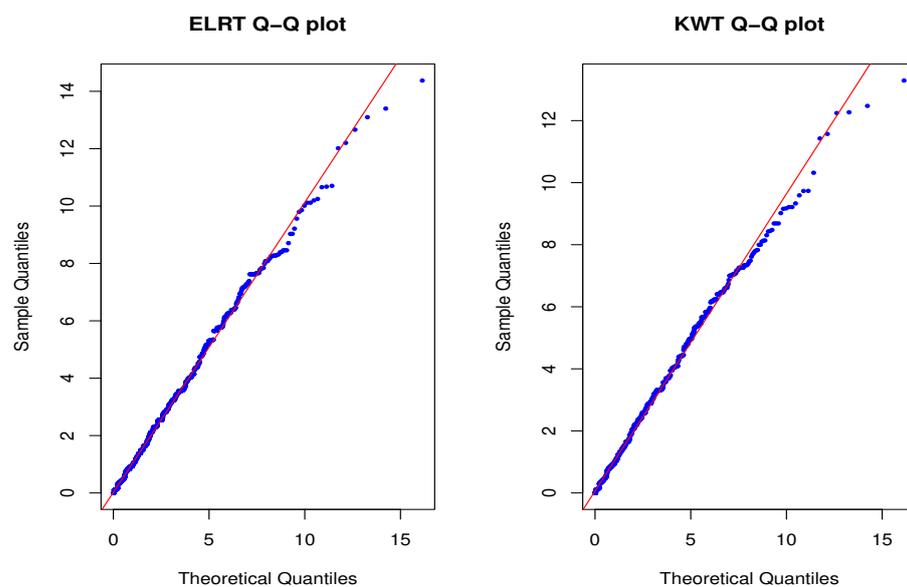


Figure 1. $p_0 = 0.25$, the sample size is (30,30,40), the Q-Q plots of ELRT and KWT for the overall comparison of the 3 Bernoulli distributions.

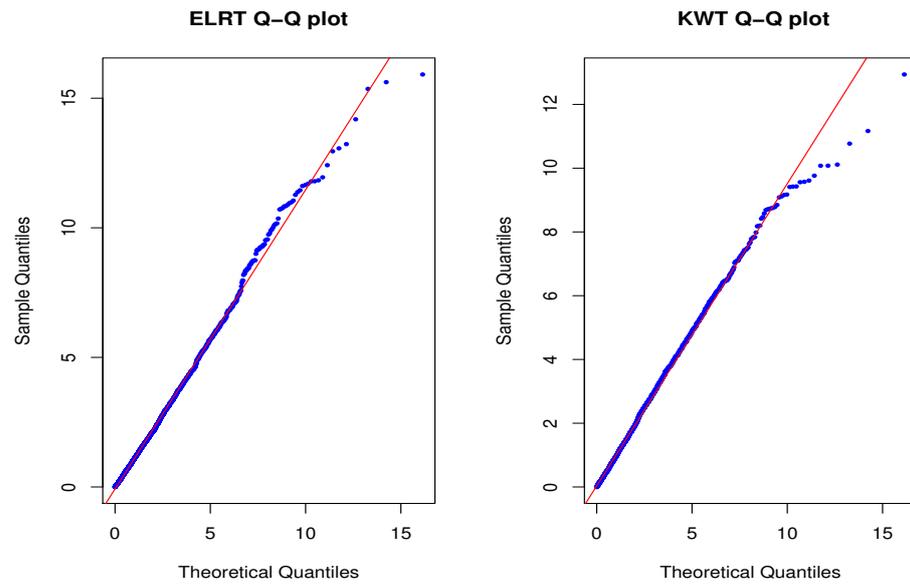


Figure 2. $\lambda_0 = 1$, the sample size is (30,30,40), the Q-Q plots of ELRT and KWT for the overall comparison of the 3 Poisson distributions.

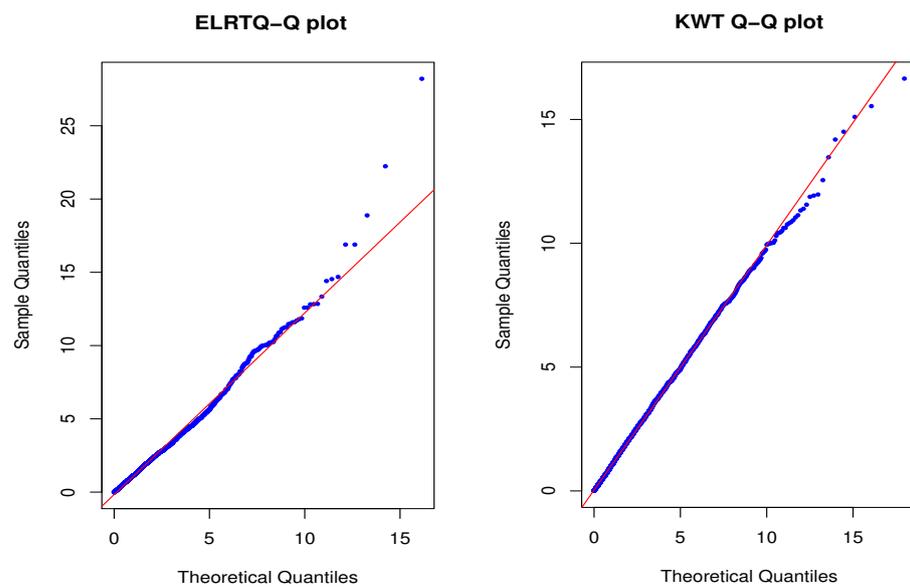


Figure 3. $\theta_0 = 1$, the sample size is (30,30,40), the Q-Q plots of ELRT and KWT for the overall comparison of the 3 exponential distributions.

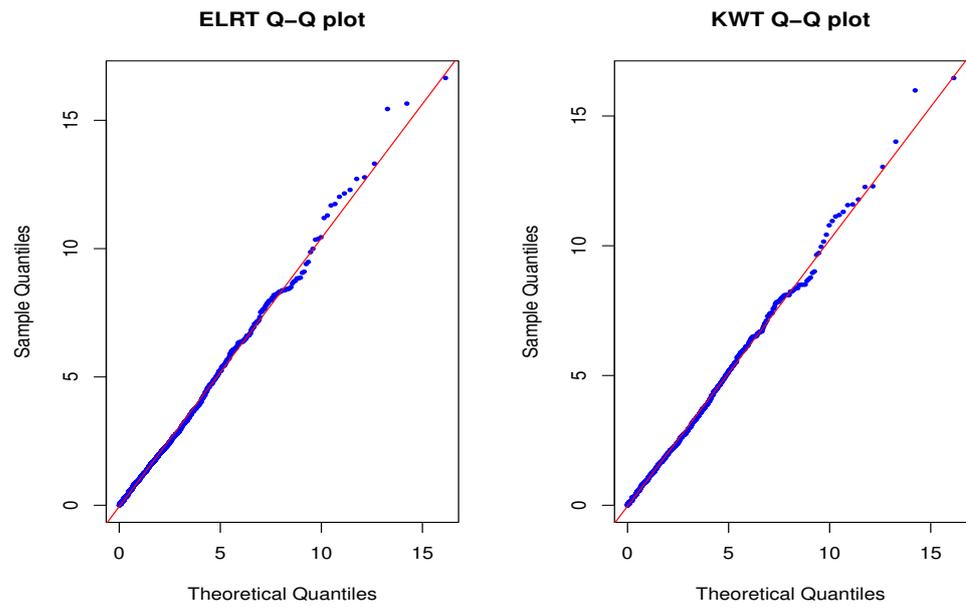


Figure 4. $p_0 = 0.25$, the sample size is (110,110,120), the Q-Q plots of ELRT and KWT for the overall comparison of the 3 Bernoulli distributions.

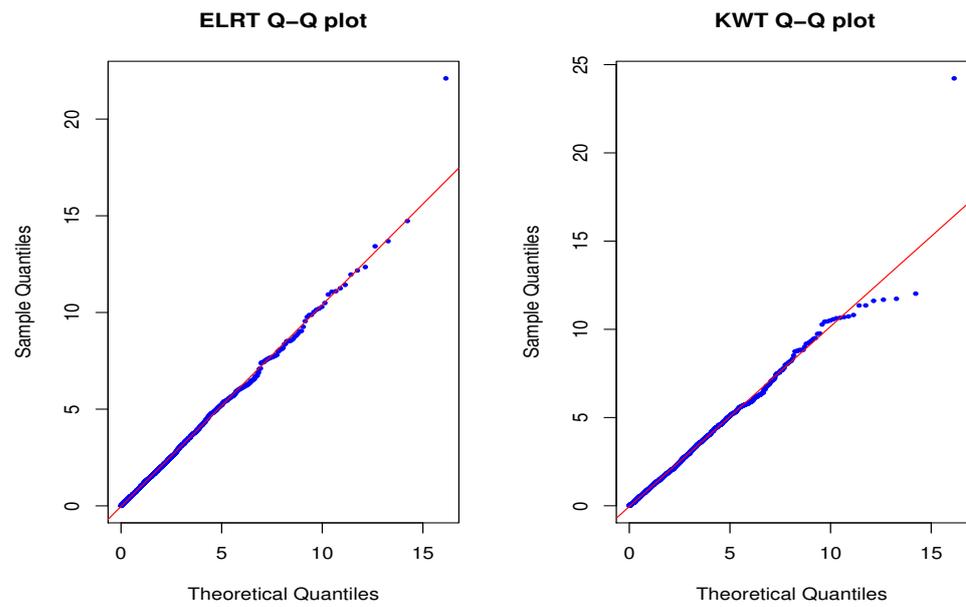


Figure 5. $\lambda_0 = 1$, the sample size is (110,110,120), the Q-Q plots of ELRT and KWT for the overall comparison of the 3 Poisson distributions.

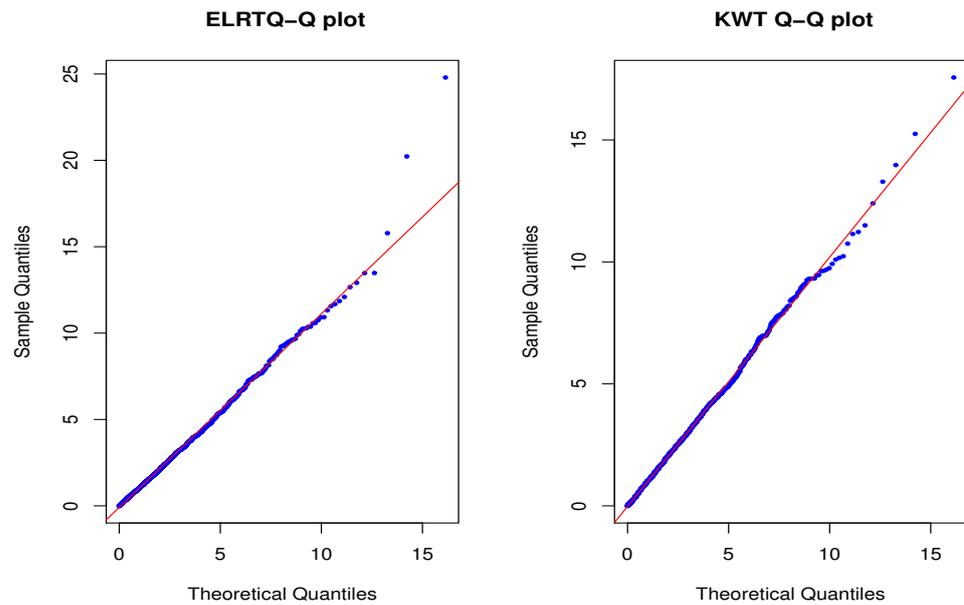


Figure 6. $\theta_0 = 1$, the sample size is (110,110,120), the Q-Q plots of ELRT and KWT for the overall comparison of the 3 exponential distributions.

Table 2. Rejection rates of ELRT and KWT under sample sizes (30,30,40) and (110,110,120) and different alternatives indicated in terms of parameters.

Distribution	Alternative Hypothesis Combination	(30, 30, 40)		(110, 110, 120)	
		ELRT	KWT	ELRT	KWT
$b(p)$	(0.1, 0.3, 0.5)	0.8303	0.8154	0.9376	0.8952
	(0.3, 0.3, 0.5)	0.8210	0.8103	0.9223	0.9109
$P(\lambda)$	(0.1, 0.3, 0.4)	0.7756	0.7301	0.9012	0.8950
	(0.2, 0.2, 0.4)	0.8020	0.7850	0.9051	0.8700
$E(\theta)$	(0.2, 0.3, 0.5)	0.8425	0.7866	0.9133	0.8806
	(0.2, 0.2, 0.3)	0.8644	0.8157	0.9256	0.9102

Table 3. Rejection rates ELRT and KWT under sample sizes (30,30,40) and (110,110,120) and corresponding null hypothesis.

Distribution	Alternative Hypothesis Combination	(30, 30, 40)		(110, 110, 120)	
		ELRT	KWT	ELRT	KWT
$b(p), p_0 = 0.25$	(0.25, 0.25, 0.25)	0.0490	0.0511	0.0510	0.0505
$P(\lambda), \lambda_0 = 1$	(1, 1, 1)	0.0424	0.0455	0.0490	0.0510
$E(\theta), \theta_0 = 1$	(1, 1, 1)	0.0415	0.0422	0.0480	0.0498

Table 4. Rejection rates ELRT and T&P (2006) under sample sizes (30, 30, 40) and (110, 110, 120) and different alternatives indicated in terms of parameters.

Distribution	Alternative Hypothesis Combination	(30, 30, 40)		(110, 110, 120)	
		ELRT	T&P (2006)	ELRT	T&P (2006)
$P(\lambda)$	(0.1, 0.3, 0.4)	0.7756	0.0.640	0.9012	0.9930
	(0.2, 0.2, 0.4)	0.8020	0.3750	0.9051	0.7800
	(2, 3, 4)	0.9260	0.3633	0.9722	0.9933
$E(\theta)$	(0.2, 0.3, 0.5)	0.8425	0.5230	0.9133	0.9820
	(0.2, 0.2, 0.3)	0.8644	0.1730	0.9256	0.6120
	(2, 3, 4)	0.6810	0.1100	0.8550	0.8610

4. Conclusions

In this study, we discussed the consistency test of the population when the population distribution is unknown, and constructed an ELRT statistic for testing the homogeneity of several nonparametric populations in the presence of some auxiliary information. Meanwhile, we proved the distribution of ELRT both theoretically and numerically and calculated the rejection rates of ELRT and compared with those of the Kruskal–Wallis test under several alternatives. In addition, the efficacy of ELRT and T&P (2006) were compared separately.

The results show that, firstly, the asymptotic distribution of ELRT as a chi-square distribution under some regularity conditions when the null hypothesis is true. Secondly, the rejection rates of ELRT are bigger than those of KWT, as the sample sizes increase when the sample is small. In other words, the proposed ELRT could be more powerful than the Kruskal–Wallis test, as extra information can be more efficiently employed by ELRT. Thirdly, the advantage of ELRT over T&P (2006) is that researchers do not need to select approximately normal statistics for inter-group comparisons. At the same time, compared with T&P (2006), ELRT is more suitable for multi-population consistency test with small sample size.

This discussion will be applied to the field of biological information. For example, when two samples are from the data of an experimental group and a control group, the statistics we constructed will be able to test whether the experimental processing is effective. If the overall distributions of the two data are equal, it means that the experimental processing is ineffective, otherwise it means that the experimental processing is effective. Although some good main conclusions and simulation results were obtained in this paper, there are still many problems to be further discussed in the future. On one hand, the study presented in this paper is based on simple random samples, so more complex cases (such as mixed cases or dependent samples) should be considered. On the other hand, the simulations of one-parameter distributions were completed in this paper, while the simulations of multi-parameter distributions still need to be completed. Therefore, in the future, we will continue to complete the simulation of population consistency for multi-parameter distributions by ELRT and construct a new ELRT statistics above multi-population consistency under complex samples.

5. Proofs

We first state a lemma which will be used in the proof of Theorem 1.

Lemma 1. Let $A_k = (a_{ij})$ be a $k \times k$ ($k \geq 2$) symmetric matrix, $r_i > 0$ for all $1 \leq i \leq k$ and $\sum_{i=1}^k r_i = 1$, where $a_{ii} = r_i^{-1} - 1$ for $1 \leq i \leq k$ and $a_{ij} = -1$ for $i \neq j, 1 \leq i, j \leq k$. Let $B_k = (b_{ij})$ be a $k \times k$ diagonal matrix and $C_k = B_k A_k B_k$ with $b_{ii} = r_i^{1/2}$ for $1 \leq i \leq k$, then $C_k = C_k'$ and C_k is an idempotent matrix with $tr(C_k) = k - 1$.

Proof of Lemma 1. Let $R_k = B_k^2$ and $\mathbf{1}_k = (1, 1, \dots, 1)'$. Then $A_k = R_k^{-1} - \mathbf{1}_k \mathbf{1}_k'$. It can be shown that $R_k^{1/2} \mathbf{1}_k \mathbf{1}_k' R_k^{1/2} = ((r_i r_j)^{1/2})_{k \times k}$, where $(r_i r_j)^{1/2}$ is the (i, j) element of the matrix. Combining with $\sum_{i=1}^k r_i = 1$, one can show that $((r_i r_j)^{1/2})_{k \times k}$ is an idempotent matrix. Notice that $C_k = I_k - R_k^{1/2} \mathbf{1}_k \mathbf{1}_k' R_k^{1/2}$. It follows that C_k is an idempotent matrix and $tr(C_k) = \sum_{i=1}^k (1 - r_i) = k - 1$, it is clear that $C_k = C_k'$. The proof of Lemma 1 is thus complete. \square

Proof of Theorem 1. Let

$$S_{11} = -E\{g(X; \theta_0)g'(X; \theta_0)\}, S_{12} = E\{\partial g(X; \theta) / \partial \theta |_{\theta = \theta_0}\},$$

$$S_{21} = S_{12}', S_{22.1} = S_{21}(-S_{11})^{-1}S_{12}.$$

Throughout the proof, we assume that H_0 holds true and the true value of θ is θ_0 . Rewrite λ_n as

$$\lambda_n = \lambda_{n1} / \left(\prod_i \lambda_{n2.i} \right),$$

where

$$\lambda_{n1} = \prod_{i,j} \{1 + t'(\hat{\theta}_n)g(X_{ij}; \hat{\theta}_n)\}^{-1}$$

and

$$\lambda_{n2.i} = \prod_j \{1 + t'_i(\hat{\theta}_{ni})g(X_{ij}; \hat{\theta}_{ni})\}^{-1}.$$

Employing the result in the proof of Theorem 2 in Qin and Lawless [13], we have

$$-2 \log \lambda_{n1} = \left(n^{-1/2} \sum_{i=1}^k Y_i \right)' A \left(n^{-1/2} \sum_{i=1}^k Y_i \right) + o_p(1), \tag{13}$$

where

$$A = -S_{11}^{-1} \{I + S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1}\}, Y_i = \sum_{j=1}^{n_i} g(X_{ij}; \theta_0), 1 \leq i \leq k,$$

with A being an identity matrix.

Similarly, for $1 \leq i \leq k$,

$$-2 \log \lambda_{n2.i} = (n_i^{-1/2} Y_i)' A (n_i^{-1/2} Y_i) + o_p(1).$$

It follows that

$$-2 \log \lambda_n = n^{-1/2} (Y_1', \dots, Y_k') \{M_k \otimes A\} \cdot n^{-1/2} (Y_1', \dots, Y_k')' + o_p(1), \tag{14}$$

where \otimes is the Kronecker product and $M_k = (m_{ij})$ be a $k \times k$ symmetric matrix with $m_{ii} = \frac{n}{n_i} - 1$ for $1 \leq i \leq k$ and $a_{ij} = -1$ for $i \neq j, 1 \leq i, j \leq k$.

Let $N_k = \text{diag}((n/n_1)^{1/2}, \dots, (n/n_k)^{1/2})_{k \times k}$ and

$$Z_n = \{N_k \otimes (-S_{11})^{-1/2}\} (n^{-1/2} Y_1', \dots, n^{-1/2} Y_k')'.$$

According to the central limiting theorem,

$$Z_n \xrightarrow{d} N(0, I_{pk}). \tag{15}$$

From (14), we have

$$-2 \log \lambda_n = Z_n' \left\{ N_k \otimes (-S_{11})^{-1/2} \right\}^{-1} \left\{ M_k \otimes A \right\} \left\{ N_k \otimes (-S_{11})^{-1/2} \right\}^{-1} Z_n + o_p(1). \quad (16)$$

It can be shown, by the properties of the Kronecker product, that

$$\begin{aligned} & \left\{ N_k \otimes (-S_{11})^{-1/2} \right\}^{-1} \left\{ M_k \otimes A \right\} \left\{ N_k \otimes (-S_{11})^{-1/2} \right\}^{-1} \\ &= \left\{ N_k^{-1} \otimes (-S_{11})^{1/2} \right\} \left\{ M_k \otimes A \right\} \left\{ N_k^{-1} \otimes (-S_{11})^{1/2} \right\} \\ &= S, \end{aligned} \quad (17)$$

where $S = \left(N_k^{-1} M_k N_k^{-1} \right) \otimes \left((-S_{11})^{1/2} A (-S_{11})^{1/2} \right)$. It is clear that $(-S_{11})^{1/2} \cdot A \cdot (-S_{11})^{1/2}$ is symmetric and idempotent with a trace equal to $r - p$. On the other hand, using Lemma 1, we can see that $N_k^{-1} M_k N_k^{-1}$ is symmetric and idempotent with a trace equal to $k - 1$. It follows that S must be symmetric and idempotent with a trace equal to $(r - p)(k - 1)$.

Theorem 1 is therefore proved by following Equations (14)–(17). \square

Author Contributions: Conceptualization, R.W. and Y.Q.; methodology, R.W. and Y.Q.; software, R.W.; validation, Y.Q.; formal analysis, Y.Q.; investigation, R.W. and Y.Q.; resources, R.W. and Y.Q.; data curation, R.W.; writing—original draft preparation, R.W.; writing—review and editing, R.W. and Y.Q.; visualization, R.W.; supervision, Y.Q.; project administration, Y.Q.; funding acquisition, R.W. and Y.Q. All authors have read and agreed to the published version of the manuscript.

Funding: The project was partially supported by the National Natural Science Foundation of China (12061017), the Natural Science Foundation of Guangxi (AD19245102, 2020GXNSFAA159155), Improving the Basic Scientific Research Ability of Young and Middle-aged Teachers in Guangxi Universities under Grant 2022KY0575, Yulin Normal University Research Grant (2018YJKY29).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable to this article, as no datasets were generated or analyzed during the current study. All data generated or analyzed in this study are generated by the corresponding probability distribution, and its parameters are presented in Table 2 of the numerical simulation.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Dou, Y.Q.; Ricks, T.M.; DuBien, J.L.; Lacy, T.E., Jr.; Liu, Y.-C. Response surface modeling to facilitate the parametric study of transversely impacted pressurized pipelines. *Thin-Walled Struct.* **2017**, *119*, 646–652. [[CrossRef](#)]
- Lehmann, E.L. *Nonparametrics: Statistical Methods Based on Ranks*; McGraw-Hill International Book Company: New York, NY, USA, 1975.
- Owen, A.B. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **1988**, *75*, 237–249. [[CrossRef](#)]
- Owen, A.B. Empirical likelihood ratio confidence regions. *Ann. Stat.* **1990**, *18*, 90–120. [[CrossRef](#)]
- Hall, P. *The Bootstrap and Edgeworth Expansion*; Springer: New York, NY, USA, 1992.
- Hall, P.; La Scala, B. Methodology and algorithms of empirical likelihood. *Int. Stat. Rev.* **1990**, *55*, 109–127. [[CrossRef](#)]
- Chen, J.; Qin, J. Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika* **1993**, *80*, 107–116. [[CrossRef](#)]
- Zhang, B. M -estimation and quantile estimation in the presence of auxiliary information. *J. Stat. Plan. Inference* **1995**, *44*, 77–94. [[CrossRef](#)]
- Sergey, T.; Dmitri, P. Using auxiliary information in statistical function estimation. *Esaim Probab. Stat.* **2005**, *10*, 11–23.
- Albertus, M. Asymptotic Z and chi-squared tests with auxiliary information. *Metrika* **2022**. [[CrossRef](#)]
- Hansen, L.P. Large sample properties of generalized method of moments estimators. *Econometrica* **1982**, *50*, 1029–1054. [[CrossRef](#)]
- Godambe, V.P.; Heyde, C.C. Quasi-likelihood and optimal estimation. *Int. Stat. Rev.* **1987**, *55*, 231–244. [[CrossRef](#)]
- Qin, J.; Lawless, J. Empirical likelihood and general estimating equations. *Ann. Stat.* **1994**, *22*, 300–325. [[CrossRef](#)]
- Wang, D.; Chen, S.X. Empirical likelihood for estimating equations with missing values. *Ann. Stat.* **2009**, *37*, 490–517. [[CrossRef](#)]
- Zhou, Y.; Wan, A.T.K.; Wang, X. Estimating equations inference with missing data. *J. Am. Stat. Assoc.* **2008**, *44*, 1187–1199. [[CrossRef](#)]