Empirical likelihood-based tests for stochastic ordering

HAMMOU EL BARMI[1] and **IAN W. MCKEAGUE**[2]

[1]Department of Statistics and Computer Information Systems, Baruch College, The City University of New York, One Baruch Way, New York, NY 10010, USA

[2]Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 West 168th Street, 6th Floor, New York, NY 10032, USA

Abstract

This paper develops an empirical likelihood approach to testing for the presence of stochastic ordering among univariate distributions based on independent random samples from each distribution. The proposed test statistic is formed by integrating a localized empirical likelihood statistic with respect to the empirical distribution of the pooled sample. The asymptotic null distribution of this test statistic is found to have a simple distribution-free representation in terms of standard Brownian bridge processes. The approach is used to compare the lengths of rule of Roman Emperors over various historical periods, including the "decline and fall" phase of the empire. In a simulation study, the power of the proposed test is found to improve substantially upon that of a competing test due to El Barmi and Mukerjee.

Keywords

distribution-free; nonparametric likelihood ratio testing; order restricted inference

1. Introduction

Comparing random variables in terms of their distributions can provide an understanding of underlying causal mechanisms and risks. In addition, knowledge of an ordering of distributions can be useful for increasing the efficiency of estimation procedures, as is well documented in the literature on order restricted inference; see, for example, the comprehensive monograph of Silvapulle and Sen [20]. There are many types of ordering for the comparison of univariate distributions. These include, with increasing generality, likelihood ratio ordering, uniform stochastic ordering (equivalent to hazard rate ordering), stochastic ordering, and increasing convex ordering (of interest in economics and actuarial science); see Shaked and Shanthikumar [19] for an overview.

The aim of this paper is to develop an empirical likelihood approach to testing for the presence of the classical type of stochastic ordering. Such ordering often arises in the biomedical sciences and reliability engineering, for example, with lifetime distributions of human populations exposed to higher risk, or of engineering systems under greater stress. The notion of stochastic ordering is due to Lehmann [11] who defined a random variable X_1 to be *stochastically larger* than a random variable X_2 if $F_1(x) \leq F_2(x)$ for all x (with strict inequality for some x), where F_1 and F_2 are the corresponding cdfs; we write this as $F_1 \succ F_2$.

For a stochastic ordering of k distributions, we write $F_1 \succ F_2 \succ \cdots \succ F_k$ if $F_j(x) \geq F_{j+1}(x)$ for all x and $j = 1, \ldots, k-1$, with strict inequality for some x and some j.

There is an extensive literature on the problem of testing for equality of two distributions against the alternative that they are stochastically ordered. Lee and Wolf [10] proposed a Mann–Whitney–Wilcoxon-type test. Robertson and Wright [17] studied the corresponding likelihood test (LRT) in the one- and two-sample cases when the distributions are discrete. They showed that the limiting distributions are chi-bar square. Their results indicate that, in the two-sample case, the LRT is not asymptotically distribution free. They also obtained the least favorable distribution in this case. Other tests are discussed in Dykstra, Madsen and Fairbanks [3], Franck [7] and Mau [12]. For more than two populations, Wang [21] discussed the LRT in the multinomial case; El Barmi and Johnson [5] showed that the limiting distribution of his test statistic is of chi-bar square type and gave the expression of the weighting values. Also in the k-sample case ($k \geq 2$), El Barmi and Mukerjee [6] provided an asymptotically distribution-free test based on the sequential testing procedure originally introduced by Hogg [8]. This test is applicable in both the multinomial and the continuous cases, with or without censoring. Recently, Baringhaus and Grübel [1] introduced a nonparametric two-sample test for the more general hypothesis of increasing convex ordering; their test is not asymptotically distribution-free, however, and requires the critical values to be obtained via a bootstrap procedure.

The contribution of the present paper is to provide empirical likelihood based k-sample tests for alternatives that are stochastically ordered. The empirical likelihood (EL) method was originally introduced by Owen [15,16] for the purpose of finding confidence regions for parameters defined by general classes of estimating equations. It combines the flexibility of nonparametric methods with the efficiency of likelihood-ratio-based inference. Inference based on EL has many attractive properties: estimation of variance is typically not required, the range of the parameter space is automatically respected and confidence regions have greater accuracy than those based on the Wald approach. Einmahl and McKeague [4] developed a localized version of EL, to allow nonparametric hypothesis testing, and showed via simulation studies that it outperforms (in terms of power) the corresponding Cramér–von Mises statistics for a variety of classical testing problems. Their approach is restricted to omnibus alternatives, whereas ordered alternatives are often more useful because they can provide a more direct interpretation of the result of the test.

The development of the proposed test statistic and results on its asymptotic null distribution are given in Section 2. First we consider the special case of testing whether a distribution function is stochastically larger than a specified distribution function, based on a single sample. Once the theory has been developed in this one-sample case, it is relatively straightforward to extend the approach to the general k-sample setting in which all the distribution functions are unknown. Section 3 presents the results of a simulation study in which we find that the proposed test has superior power to the test of El Barmi and Mukerjee [6], which is the only previous test to have been developed for ordered alternatives in this setting. Section 3 also contains an application of the proposed test to a comparison of the lengths of rule of Roman Emperors over various historical periods. Some concluding remarks are given in Section 4. Proofs of the main results are collected in Section 5.

2. Empirical likelihood approach

2.1. Stochastic ordering relative to a specified distribution

Suppose we are given a random sample $X_1, X_2, \ldots, X_n$ from the cdf F, and we want to test the null hypothesis $H_0: F = F_0$ versus $H_1: F \succ F_0$, where F_0 is a specified cdf.

Adapting the approach of Einmahl and McKeague [4] to the present setting, we first need to consider testing the "local" null hypothesis H_0^x: $F(x) = F_0(x)$ versus the alternative H_1^x: $F(x) < F_0(x)$, where x is fixed. The empirical likelihood procedure in this case rejects H_0^x for small values of

$$R(x) = \frac{\sup\{L(F): F(x) = F_0(x)\}}{\sup\{L(F): F(x) \leq F_0(x)\}}, \quad (1)$$

where the suprema are over cdfs F that are supported by the data points, $L(F)$ is the nonparametric likelihood function and, by convention, $\sup \emptyset = 0$ and $0/0 = 1$. For F having point mass p_i at X_i, define the new parameters $\theta_i = p_i/\phi$ and $\psi_i = p_i/(1 - \phi)$, where $0 < \phi = F(x) < 1$. In terms of this new parameterization, with $\hat{F}$ denoting the empirical cdf, we need to maximize

$$L(F) = \prod_{i=1}^{n} p_i = \left\{ \prod_{i: X_i \leq x} \theta_i \right\} \left\{ \prod_{i: X_i > x} \psi_i \right\} \phi^{n\hat{F}(x)} [1 - \phi]^{n(1 - \hat{F}(x))}, \quad (2)$$

subject to the constraint

$$\sum_{i: X_i \leq x} \theta_i = \sum_{i: X_i > x} \psi_i = 1,$$

with either $\phi = F_0(x)$ under H_0^x, or $\phi < F_0(x)$ under H_1^x. Note that the three terms in the right-hand side of (2) can be maximized separately. As the constraints for the first two terms of (2) are the same for both the numerator and the denominator of (1), these terms cancel and make no contribution to $R(x)$. The third term of (2) is maximized by $\phi = F_0(x)$ under H_0^x, or $\phi = F_0(x) \wedge \hat{F}(x)$ under H_1^x. Consequently,

$$R(x) = \begin{cases} 1, & \text{if } \hat{F}(x) > F_0(x), \\ \left[\frac{F_0(x)}{\hat{F}(x)} \right]^{n\hat{F}(x)} \left[\frac{1 - F_0(x)}{1 - \hat{F}(x)} \right]^{n(1 - \hat{F}(x))}, & \text{if } \hat{F}(x) \leq F_0(x), \end{cases}$$

with the convention that any term raised to a zero power is set to 1. Using a second-order Taylor expansion of $\log(1 + y)$ about $y = 0$, it can be shown (see the proof of the theorem below) that, for a given x, such that $0 < F_0(x) < 1$, under H_0^x,

$$\begin{aligned} -2 \log R(x) &= n\left(\hat{F}(x) - F_0(x)\right)^2 \left[\frac{1}{\hat{F}(x)} + \frac{1}{1 - \hat{F}(x)} \right] I[0 < \hat{F}(x) \leq F_0(x)] + o_p(1) \\ &\xrightarrow{d} Z^2 I(Z \geq 0), \end{aligned}$$

using the CLT and the continuous mapping theorem, where $Z \sim N(0, 1)$. That is, the asymptotic null distribution of $-2 \log R(x)$ is chi-bar square.

To test H_0 against H_1, we introduce the integral-type test statistic

$$T_n = -2 \int_{-\infty}^{\infty} \log(R(x)) \, dF_0(x).$$

Here the range of integration is actually restricted to the interval $[X_{(1)}, X_{(n)}]$, where $X_{(1)}$ and $X_{(n)}$ are the smallest and largest order statistics in the sample, because the integrand vanishes outside this interval. The following result gives the asymptotic null distribution of T_n.

Theorem 1—*If F_0 is continuous, then under H_0,*

$$T_n \xrightarrow{d} \int_0^1 \frac{B^2(t)}{t(1-t)} I(B(t) \geq 0)\, \mathrm{d}t,$$

where B is a standard Brownian bridge.

<u>**Remark 1:**</u> An alternative test statistic is obtained by integrating with respect to the empirical cdf (instead of F_0),

$$T_n^* = -2 \int_{-\infty}^{\infty} \log(R(x))\, \mathrm{d}\widehat{F}(x).$$

It can be shown using a martingale argument (see Section 5), that T_n^* has the same asymptotic null distribution as T_n.

2.2. Stochastic ordering among *k* distributions

Suppose now that we are given a random sample of size n_j from the cdf F_j, for $j = 1, \ldots, k$, the k samples are independent and we want to test the null hypothesis $H_0: F_1 = \cdots = F_k$ versus $H_1: F_1 \succ \cdots \succ F_k$. We assume that the proportion $w_j = n_j/n$ of observations in the jth sample remains fixed as the total sample size $n \to \infty$, with $0 < w_j < 1$ for all $j = 1, \ldots, k$.

Adapting the approach of Section 2.1, we now consider the localized empirical likelihood function

$$R(x) = \frac{\sup\{\prod_{j=1}^k L(F_j): F_j(x) = F_{j+1}(x),\ j = 1, \ldots, k-1\}}{\sup\{\prod_{j=1}^k L(F_j): F_j(x) \leq F_{j+1}(x),\ j = 1, \ldots, k-1\}}, \quad (3)$$

where, in each supremum, F_j is supported by the observations in the jth sample. Applying the same parameterization used in (2), separately for each F_j, and making the same cancelation in the numerator and denominator, it suffices to maximize

$$\prod_{j=1}^k \phi_j^{n_j \widehat{F}_j(x)} [1 - \phi_j]^{n_j (1 - \widehat{F}_j(x))} \quad (4)$$

subject to the constraint $0 < \phi_1 = \cdots = \phi_k < 1$, or $0 < \phi_1 \cdots \phi_k < 1$, depending on whether it is the numerator or the denominator of (3). Here $\hat{F}_j$ is the empirical cdf based on the jth sample. Under the first of these constraints, (4) is maximized by $\phi_j = \hat{F}(x)$, where $\hat{F}$ is the empirical cdf of the pooled sample. Under the second constraint, this is the classical bioassay problem, as discussed in Robertson *et al.* [18], page 32, and it follows that (4) is maximized by

$$\phi_j = E_{\mathbf{w}}(\widehat{\boldsymbol{\phi}}|I)_j \equiv \tilde{F}_j(x),$$

where $E_{\mathbf{w}}(\hat{\boldsymbol{\phi}}|I)$ is the weighted least squares projection of $\hat{\boldsymbol{\phi}} = (\hat{F}_1(x), \ldots, \hat{F}_k(x))^T$ onto $I = \{\mathbf{z} \in \mathbb{R}^k : z_1 \quad z_2 \quad \cdots \quad z_k\}$, with weights w_j. In passing, we mention that several algorithms have been developed for computing this projection, including the pool-adjacent-violators algorithm, see Robertson *et al.* [18]. We now have

$$R(x) = \prod_{j=1}^{k} \left[\frac{\widehat{F}(x)}{\tilde{F}_j(x)} \right]^{n_j \widehat{F}_j(x)} \left[\frac{1 - \widehat{F}(x)}{1 - \tilde{F}_j(x)} \right]^{n_j (1 - \widehat{F}_j(x))} \quad (5)$$

under the convention that any term raised to a zero power is set to 1.

To test H_0 against H_1, we propose the test statistic

$$T_n = -2 \int_{-\infty}^{\infty} \log R(x) \, \mathrm{d}\widehat{F}(x). \quad (6)$$

The following theorem gives the asymptotic null distribution of T_n.

Theorem 2—*Under H_0 and assuming that the common distribution function F is continuous,*

$$T_n \xrightarrow{d} \sum_{j=1}^{k} w_j \int_0^1 \frac{\left(E_{\mathbf{w}}\left[\mathbf{B}(t) | I \right]_j - \overline{B}(t) \right)^2}{t(1-t)} \, \mathrm{d}t, \quad (7)$$

where $\mathbf{B} = (B_1/\sqrt{w_1}, B_2/\sqrt{w_2}, \ldots, B_k/\sqrt{w_k})^T$, the processes $B_1, B_2, \ldots, B_k$ are independent standard Brownian bridges, and $\overline{B} = \sum_{j=1}^{k} \sqrt{w_j} B_j$.

Remark 2: For the two-sample case, it can be shown that the limiting distribution in the above result coincides with that in the one-sample case (Theorem 1); the equivalence arises from the fact that $B = \sqrt{w_2} B_1 - \sqrt{w_1} B_2$ is a standard Brownian bridge. Moreover, when testing against the unrestricted alternative $F_1 \quad F_2$, the limiting distribution of the corresponding test statistic (see Einmahl and McKeague [4], Theorem 2a) is the same apart from the presence of the indicator $I(B(t) \quad 0)$ in the integrand.

3. Numerical examples

In this section we discuss some numerical examples illustrating the proposed test for a comparison of two or more distributions developed in Section 2.2.

To implement the proposed test we first need to obtain critical values for T_n. The null distribution of T_n is not tractable, even asymptotically, but it is asymptotically distribution free. We use simulation to approximate selected critical values as provided in Table 1. These critical values are based on 100 000 data sets distributed as $N(0, 1)$, with sample sizes of $n_i = 100$, $i = 1, \ldots, k$, in each case. The (Fortran) program used to compute the critical values in Table 1 is available online in the supplemental files.

3.1. Simulation study

Here we present the results of a simulation study designed to compare the performance of T_n with the test statistic S_n of El Barmi and Mukerjee [6], which is defined as the maximum of a sequence of (one-sided) two-sample Kolmogorov–Smirnov test statistics. As far as we know, S_n is the only previously developed test statistic when $k \quad 3$.

Tables 2 and 3 give the results for a variety of distributions and sample sizes, for $k = 2$ and $k = 3$, respectively. In each case, 10 000 data sets were used to approximate the power at a nominal level of $\alpha = 0.05$, with critical values for T_n taken from Table 1; critical values for S_n are obtained from its asymptotic distribution, which is available in a closed form. In all cases, T_n has greater power than S_n and has better agreement with the nominal level of the test.

3.2. Lengths of rule of Roman Emperors

A recent article of Khmaladze, Brownrigg and Haywood [9] reached the interesting conclusion that the lengths of rule of Roman Emperors were exponentially distributed, implying that their reigns ceased unexpectedly ("brittle power"). It is also of interest to examine whether there were *changes* in the distribution of rule lengths, especially during the "decline and fall" phase of the empire. We use the list of $n = 70$ Roman Emperors from Augustus to Theodossius, covering 27 BC to 395 AD. Our analysis is based on the chronology of Parkin (see Khmaladze *et al.* [9] for further details). The (Fortran) programs used for the two analyzes are available online in the supplemental files.

First we consider whether there is an effect on duration of rule due to the Crisis of the Third Century (235–284 AD), when the Roman Empire nearly collapsed under the pressure of civil war (among other things!). Figure 1 shows the empirical survival function of durations of rule for the Principate (27 BC–235 AD), which was the relatively stable period preceding the Crisis, compared with the period after 235 AD; the sample sizes are $n_1 = 29$ and $n_2 = 41$, respectively. The two distributions appear to be exponential, and the likelihood ratio test of stochastic ordering under this assumption has p-value 0.195; the corresponding unrestricted likelihood-ratio test has p-value 0.390. Applying our proposed test (with $k = 2$) to assess whether the duration of rule is stochastically shorter after the Principate, we obtain $T_n = 0.3161$ with a p-value of 0.424. This compares with a p-value of 0.575 based on S_n.

The period 285–395 AD forms part of what is known as the Dominate, the despotic later phase of the empire. Inspection of Figure 2 suggests that the exponential hypothesis is not tenable for each separate period, so our nonparametric approach is more reasonable. The plot also suggests that the rule lengths are stochastically ordered as Dominate ≻ Principate ≻ Crisis. Applying our approach to formally test this hypothesis, we find that T_n has a p-value of 0.0002, compared with a p-value of 0.0017 for S_n. Under the assumption of exponential distributions, the likelihood ratio test has p-value less than 0.0007.

4. Discussion

In this paper we have developed a novel empirical likelihood approach to the important problem of nonparametrically testing for the presence of stochastic ordering based on k independent samples. The proposed tests are computationally efficient to implement, and could be used with massive data sets because they do not rely on the bootstrap or any other simulation technique, and they reduce to a local test for an ordering of binomial probabilities, which only requires a single sweep through the pooled data in the k groups.

Various extensions of the proposed tests are possible. In change-point problems, for example, it is of interest to test whether there is a sudden change in the distribution of a sequence of independent random variables $X_1, \ldots, X_n$. Einmahl and McKeague [4] developed an EL-based change-point test for the presence of an (unknown) change-point $\tau \in \{2, \ldots, n\}$ such that

$$X_1, \ldots, X_{\tau-1} \sim F_1 \quad \text{and} \quad X_\tau, \ldots, X_n \sim F_2.$$

They only considered the unrestricted alternative $F_1 \neq F_2$, but it is also of interest to consider the ordered alternative $F_1 \succ F_2$. This can be done by extending the two-sample case to allow the sample sizes to depend on an additional local parameter, namely $t \in [1/n, 1)$, with $n_1 = \lfloor nt \rfloor$ and $n_2 = n - \lfloor nt \rfloor$. The resulting test statistic has a limiting distribution of the same form as in Theorem 2 of Einmahl and McKeague [4], involving the integral of a four-sided tied-down Wiener process $W_0(t, y)$, except that the integrand now includes the indicator $I(W_0(t, y) \geq 0)$.

Our approach also naturally extends to non-monotonic alternatives, namely to testing whether $F_1, F_2, \ldots, F_k$ are isotonic with respect to a quasi-order on $\{1, 2, \ldots, k\}$. A relation $\lesssim$ on $\{1, 2, \ldots, k\}$ is a *quasi-order* if it is reflexive and transitive (and a partial order if, in addition, it is antisymmetric). We say that $F_1, F_2, \ldots, F_k$ are *isotonic* with respect to $\lesssim$ if $F_i \succ F_j$ whenever $i \lesssim j$. Examples of such ordered alternatives include $F_1 \succ F_i$, $i = 2, \ldots, k$ (tree ordering) and $F_1 \succ F_2 \succ \cdots \succ F_{i_0} \prec F_{i_0+1} \prec \cdots \prec F_k$, where i_0 is known (umbrella ordering). The localized empirical likelihood (3) extends naturally to such ordered alternatives, the only difference being that in $\phi_j = E_{\mathbf{w}}(\hat{\phi}|I)_j$ the set I is now the isotonic cone corresponding to $\lesssim$. For examples, in the case of tree ordering, the cone becomes $I = \{\mathbf{z} \in \mathbb{R}^k : z_1 \geq z_i, i = 2, \ldots, k\}$. The ϕ_j can be computed using quadratic programming or algorithms described in Robertson, Wright and Dykstra [18], one of the most general being the lower-sets algorithm. The limiting distribution of the resulting test statistic is obtained by taking I in (7) as the isotonic cone corresponding to $\lesssim$.

An important and challenging problem for future research in this area would be to develop EL-based tests for stochastic ordering based on *censored* data. EL methods are well developed for the comparison of survival functions from right-censored data, see McKeague and Zhao [13, 14], but these methods only apply to omnibus alternatives. The complication in extending the present tests to right-censored data arises because the EL ratio would then no longer have such an explicit form as in (5), and Lagrange multipliers would be involved. This extension is beyond the scope of the present paper.

5. Proofs

Proof of Theorem 1

For $0 < \varepsilon < 1$, let $x_\varepsilon, y_\varepsilon$ be real numbers such that $F_0(x_\varepsilon) = 1 - F_0(y_\varepsilon) = \varepsilon/2$. Then decompose the test statistic as $T_n = T_{1n} + T_{2n}$, where

$$T_{1n} = -2 \int_{x_\varepsilon}^{y_\varepsilon} \log(R(x)) \, dF_0(x)$$

and

$$T_{2n} = -2 \int_{[x_\varepsilon, y_\varepsilon]^c} \log(R(x)) \, dF_0(x).$$

By appealing to Theorem 4.2 of Billingsley [2], note that, to complete the proof of the theorem, it suffices to show that for fixed ε,

$$T_{1n} \xrightarrow{d} \int_{\varepsilon/2}^{1-\varepsilon/2} \frac{B^2(t)}{t(1-t)} I(B(t) \geq 0) \, dt \quad (8)$$

as $n \to \infty$, and, for each $\delta > 0$, that $\limsup_{n \to \infty} P(|T_{2n}| \geq \delta) \to 0$ as $\varepsilon \to 0$.

First consider T_{1n}. Using the inequality $|\log(1 + y) - y + y^2/2| \leq |y|^3/3$ when $|y| \leq 1/2$, the Glivenko–Cantelli theorem and Donsker's theorem, we have

$$\limsup_{n\to\infty} \sup_{x\in[x_\varepsilon, y_\varepsilon]} \left| \log\left(R(x)\right) + \tfrac{n}{2}\left(\widehat{F}(x) - F_0(x)\right)^2 \left[\tfrac{1}{\widehat{F}(x)} + \tfrac{1}{1-\widehat{F}(x)} \right] I[\widehat{F}(x) \leq F_0(x)] \right|$$

$$\leq \limsup_{n\to\infty} \sup_{x\in[x_\varepsilon, y_\varepsilon]} \tfrac{n}{3}|\widehat{F}(x) - F_0(x)|^3 \left[\tfrac{1}{\widehat{F}(x)} + \tfrac{1}{1-\widehat{F}(x)} \right] = 0,$$

almost surely. Then, noting that $\hat{F}(x) = \hat{\Gamma}(F_0(x))$, where $\hat{\Gamma}$ is the empirical cdf of $V_i = F_0(X_i) \sim U(0,1)$, $i = 1, \ldots, n$, and changing variables in the integration to $t = F_0(x)$, it follows that

$$\begin{aligned} T_{1n} &= \int_{\varepsilon/2}^{1-\varepsilon/2} n(\widehat{\Gamma}(t) - t)^2 \left[\tfrac{1}{\widehat{\Gamma}(t)} + \tfrac{1}{1-\widehat{\Gamma}(t)} \right] I[\sqrt{n}(\widehat{\Gamma}(t) - t) \leq 0]\, dt + o_p(1) \\ &= \int_{\varepsilon/2}^{1-\varepsilon/2} \tfrac{\widehat{U}(t)^2}{t(1-t)} I[\widehat{U}(t) \leq 0]\, dt + o_p(1), \end{aligned} \qquad (9)$$

where $\widehat{U}(t) = \sqrt{n}(\widehat{\Gamma}(t) - t)$ is the uniform empirical process. Note that (for any fixed $0 < \varepsilon < 1$) the functional

$$f \mapsto \int_{\varepsilon/2}^{1-\varepsilon/2} \tfrac{f(t)^2}{t(1-t)} I(f(t) \leq 0)\, dt, \quad f \in D[0,1],$$

is continuous when the Skorohod space $D[0, 1]$ is equipped with the uniform norm. By Donsker's theorem, $\hat{U}$ converges weakly to B in $D[0, 1]$, so applying the continuous mapping theorem to the leading term in (9) establishes (8).

Finally we need to verify the claim concerning T_{2n}. This follows immediately from a corresponding result in Einmahl and McKeague [4], who considered the test of the null hypothesis $F = F_0$ versus the (omnibus) alternative $F \neq F_0$, with the same integral-type test statistic as T_n except that the integrand does not vanish when $\hat{F}(x) > F_0(x)$. This completes the proof.

Proof for Remark 1

The asymptotic distribution of T_n^* can be obtained following the same steps as the proof of Theorem 1, except that the leading term in T_{1n} now becomes

$$\int_{\varepsilon/2}^{1-\varepsilon/2} \tfrac{\widehat{U}(t)^2}{t(1-t)} I[\widehat{U}(t) \leq 0]\, d\widehat{\Gamma}(t) = \int_{\varepsilon/2}^{1-\varepsilon/2} V(t-)\, d\widehat{\Gamma}(t) + o_p(1),$$

where

$$V(t) = \tfrac{\widehat{U}(t)^2}{t(1-t)} I[\widehat{U}(t) \leq 0, \; \varepsilon/2 < t \leq 1 - \varepsilon/2].$$

Note that

$$M(t) = \widehat{\Gamma}(t) - \int_0^t [1 - \widehat{\Gamma}(s-)](1-s)^{-1} ds$$

is a martingale wrt to the natural filtration defined by $\hat{\Gamma}$, and its predictable quadratic variation process is $\langle M \rangle (t) = n^{-1} \int_0^t [1 - \widehat{\Gamma}(s-)](1-s)^{-1} \, ds$. Also note that $V(t-)$ is a predictable process because it is adapted and left-continuous. Write

$$\int_{\varepsilon/2}^{1-\varepsilon/2} V(t-) \, d\widehat{\Gamma}(t)$$
$$= \int_{\varepsilon/2}^{1-\varepsilon/2} V(t-) \, dM(t) + \int_{\varepsilon/2}^{1-\varepsilon/2} V(t-)[1-\widehat{\Gamma}(t-)](1-t)^{-1} dt.$$

Using a basic property of martingale integrals, the second moment of the first term above is

$$E \int_{\varepsilon/2}^{1-\varepsilon/2} V(t-)^2 \, d \langle M \rangle (t) = O(1/n),$$

so this term tends in probability to zero. The second term in the above display can be handled in the same way as the main term T_{1n} in the proof of Theorem 1, and has the same limit distribution.

Proof of Theorem 2

The proof is similar to the proof of Theorem 1, so we only indicate the main steps. Using the Taylor expansion of $\log(1+y)$, as before, and the (uniform) consistency of $\tilde{F}_j$ as an estimator of $F_j = F$ (see, e.g., El Barmi and Mukerjee [6], page 253), for each fixed x, such that $0 < t = F(x) < 1$, we have

$$
\begin{aligned}
-2 \log R(x) &= \sum_{j=1}^k n_j \left(\tilde{F}_j(x) - \widehat{F}(x) \right)^2 \left[\frac{1}{\tilde{F}_j(x)} + \frac{1}{1-\tilde{F}_j(x)} \right] + o_p(1) \\
&= \sum_{j=1}^k w_j \frac{\left[\sqrt{n}(\tilde{F}_j(x)-F(x)) - \sqrt{n}(\widehat{F}(x)-F(x)) \right]^2}{F(x)(1-F(x))} + o_p(1) \\
&= \sum_{j=1}^k w_j \frac{(E_{\mathbf{w}}[\widehat{\mathbf{U}}(t)|I]_j - \overline{U}(t))^2}{t(1-t)} + o_p(1) \\
&\xrightarrow{d} \sum_{j=1}^k w_j \frac{(E_{\mathbf{w}}[\mathbf{B}(t)|I]_j - \overline{B}(t))^2}{t(1-t)},
\end{aligned}
$$

where $\widehat{\mathbf{U}} = (\widehat{U}_1/\sqrt{w_1}, \ \widehat{U}_2/\sqrt{w_2}, \dots, \widehat{U}_k/\sqrt{w_k})^T$, $\widehat{U}_j(t) = \sqrt{n_j}(\widehat{F}_j(x) - F(x))$ are independent uniform empirical processes, and $\overline{U} = \sum_{j=1}^k \sqrt{w_j}\widehat{U}_j$. Donsker's theorem and the continuous mapping theorem have been used as before, but we have also used the fact that $E_{\mathbf{w}}(\cdot|I)$ is a continuous function on $\mathbb{R}^k$.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Baringhaus L, Grübel R. Nonparametric two-sample tests for increasing convex order. Bernoulli. 2009; 15:99–123. MR2546800.

2. Billingsley, P. Convergence of Probability Measures. New York: Wiley; 1968. MR0233396

3. Dykstra RL, Madsen RW, Fairbanks K. A nonparametric likelihood ratio test. J Statist Comput Simulation. 1983; 18:247–264. MR074655.

4. Einmahl JHJ, McKeague IW. Empirical likelihood based hypothesis testing. Bernoulli. 2003; 9:267–290. MR1997030.

5. El Barmi H, Johnson M. A unified approach to testing for and against a set of linear inequality constraints in the product multinomial setting. J Multivariate Anal. 2006; 97:1894–1912. MR2298895.

6. El Barmi H, Mukerjee H. Inferences under a stochastic ordering constraint: The *k*-sample case. J Amer Statist Assoc. 2005; 100:252–261. MR2156835.

7. Franck WE. A likelihood ratio test for stochastic ordering. J Amer Statist Assoc. 1984; 79:686–691. MR0763587.

8. Hogg RV. Iterated tests of the equality of several distributions. J Amer Statist Assoc. 1962; 57:579–585. MR0159373.

9. Khmaladze E, Brownrigg R, Haywood J. Brittle power: On Roman Emperors and exponential lengths of rule. Statist Probab Lett. 2007; 77:1248–1257. MR2392795.

10. Lee YJ, Wolfe DA. A distribution-free test for stochastic ordering. J Amer Statist Assoc. 1976; 71:722–727. MR0443208.

11. Lehmann EL. Ordered families of distributions. Ann Math Statist. 1955; 26:399–419. MR0071684.

12. Mau J. A generalization of a nonparametric test for stochastically ordered distributions to censored survival data. J Roy Statist Soc Ser B. 1988; 50:403–412. MR0970976.

13. McKeague IW, Zhao Y. Simultaneous confidence bands for ratios of survival functions via empirical likelihood. Statist Probab Lett. 2002; 60:405–415. MR1947180.

14. McKeague IW, Zhao Y. Comparing distribution functions via empirical likelihood. Int J Biostat. 2005; 1:20. Art. 5, (electronic). MR2232230.

15. Owen A. Empirical likelihood ratio confidence regions. Ann Statist. 1990; 18:90–120. MR1041387.

16. Owen AB. Empirical likelihood ratio confidence intervals for a single functional. Biometrika. 1988; 75:237–249. MR0946049.

17. Robertson T, Wright FT. Likelihood ratio tests for and against a stochastic ordering between multinomial populations. Ann Statist. 1981; 9:1248–1257. MR0630107.

18. Robertson, T.; Wright, FT.; Dykstra, RL. Order Restricted Statistical Inference. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Chichester: Wiley; 1988. MR0961262

19. Shaked, M.; Shanthikumar, GJ. Stochastic Orders. New York: Springer; 2006.

20. Silvapulle, MJ.; Sen, PK. Constrained Statistical Inference: Inequality, Order, and Shape Restrictions. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley-Interscience; 2005. MR2099529

21. Wang Y. A likelihood ratio test against stochastic ordering in several populations. J Amer Statist Assoc. 1996; 91:1676–1683. MR1439109.
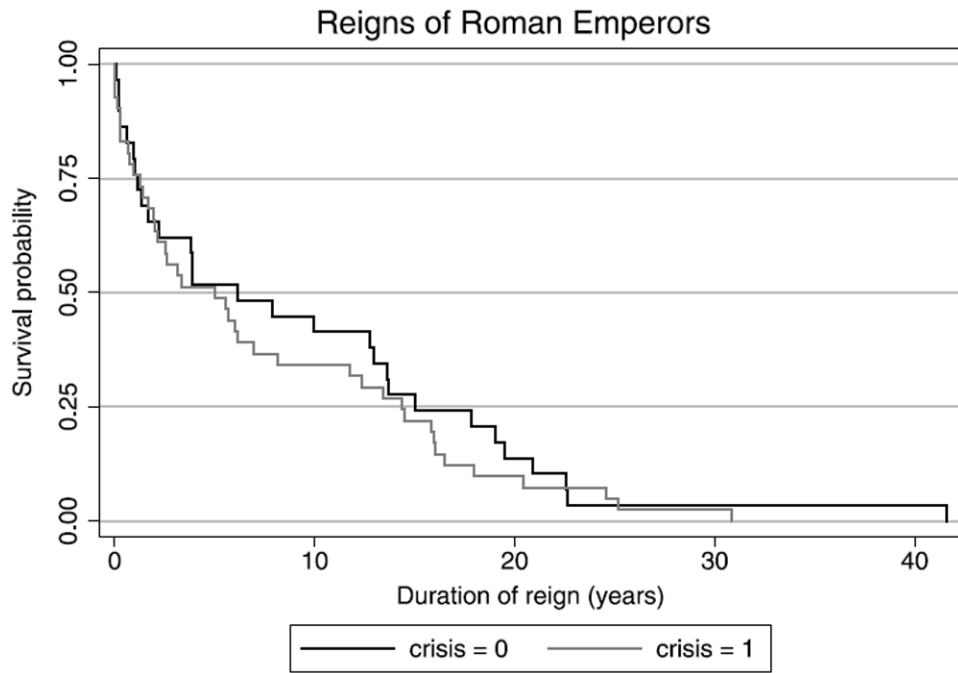
Figure 1.
Empirical survival functions of durations of rule of the first 70 Roman Emperors before 235
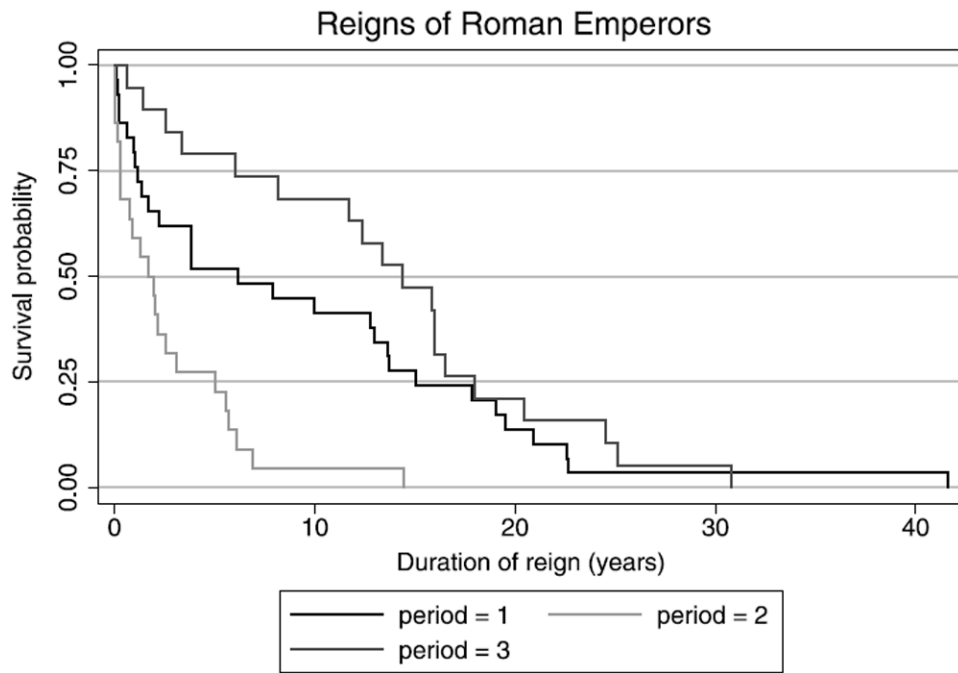AD (crisis = 0), and after 235 AD (crisis = 1).

Figure 2.
Empirical survival functions of durations of rule during the Principate, 27 BC–235 AD
(period = 1), the Crisis, 235–284 AD (period = 2) and the Dominate, 284–395 AD (period =
3).

Table 1

Selected critical points of T_n

	Significance level α		
k	0.01	0.05	0.10
2	3.185	1.821	1.288
3	4.128	2.613	1.943
4	4.663	3.107	2.404
5	5.144	3.470	2.701

Table 2

Power comparison of tests for stochastic ordering of $k = 2$ distributions at level $\alpha = 0.05$

Distributions		$n_1 = 50, n_2 = 30$		$n_1 = 30, n_2 = 50$		$n_1 = 50, n_2 = 50$	
F_1	F_2	T_n	S_n	T_n	S_n	T_n	S_n
Uni(0, 1)	Uni(0, 1)	0.064	0.038	0.051	0.045	0.051	0.036
Uni(0, 1.1)	Uni(0, 1)	0.143	0.104	0.162	0.111	0.199	0.125
Uni(0, 2)	Uni(0, 1)	0.911	0.816	0.912	0.818	0.908	0.815
Uni(0.1, 1.1)	Uni(0, 1)	0.377	0.244	0.357	0.246	0.468	0.287
Exp(1)	Exp(1)	0.063	0.037	0.048	0.041	0.047	0.036
Exp(1)	Exp(1.1)	0.123	0.076	0.091	0.068	0.108	0.076
Exp(1)	Exp(2)	0.782	0.716	0.813	0.718	0.909	0.815
0.1 + Exp(1)	Exp(1)	0.207	0.118	0.137	0.105	0.195	0.127
N(0, 1)	N(0, 1)	0.063	0.037	0.049	0.040	0.051	0.036
N(0.1, 1)	N(0, 1)	0.132	0.081	0.100	0.079	0.122	0.079
N(0.5, 1)	N(0, 1)	0.646	0.530	0.690	0.540	0.771	0.628
N(1, 1)	N(0, 1)	0.992	0.975	0.991	0.975	0.993	0.976

Table 3

Power comparison of tests for stochastic ordering of $k = 3$ distributions at level $\alpha = 0.05$

Distributions			$n_1 = n_2 = n_3 = 30$		$n_1 = n_2 = n_3 = 50$	
F_1	F_2	F_3	T_n	S_n	T_n	S_n
Uni(0, 1)	Uni(0, 1)	Uni(0, 1)	0.038	0.033	0.045	0.039
Uni(0, 1.1)	Uni(0, 1)	Uni(0, 1)	0.455	0.370	0.740	0.647
Uni(0, 1.1)	Uni(0, 1.1)	Uni(0, 1)	0.389	0.319	0.651	0.633
Uni(0.1, 1.1)	Uni(0, 1)	Uni(0, 1)	0.948	0.884	0.999	0.885
Exp(1)	Exp(1)	Exp(1)	0.041	0.019	0.049	0.045
Exp(1)	Exp(1)	Exp(1.1)	0.076	0.033	0.098	0.067
Exp(1)	Exp(1.1)	Exp(1.1)	0.067	0.029	0.098	0.073
Exp(1)	Exp(1.1)	Exp(1.2)	0.116	0.046	0.171	0.109
Exp(1)	Exp(1.25)	Exp(1.5)	0.313	0.121	0.507	0.321
N(0, 1)	N(0, 1)	N(0, 1)	0.042	0.035	0.049	0.035
N(0.1, 1)	N(0, 1)	N(0, 1)	0.272	0.183	0.423	0.292
N(0.1, 1)	N(0.1, 1)	N(0, 1)	0.246	0.151	0.393	0.249
N(0.5, 1)	N(0.25, 1)	N(0, 1)	1.000	0.993	1.000	1.000