Bernoulli **27**(4), 2021, 2556–2563
https://doi.org/10.3150/21-BEJ1322

Testing against uniform stochastic ordering with paired observations

DEWEI WANG[1] and CHUAN-FA TANG[2]

[1]*Department of Statistics, University of South Carolina, 1523 Greene Street, Columbia, SC 29208, USA. E-mail: deweiwang@stat.sc.edu*
[2]*Department of Mathematical Sciences, University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080, USA. E-mail: chuan-fa.tang@utdallas.edu*

This article develops a two-sample nonparametric goodness-of-fit (GOF) test for uniform stochastic ordering (USO) when observations are taken in pairs. We propose a data-driven critical value that controls the type I error and yields a consistent test. A simulation study illustrates the finite-sample performance of our test. All the proofs are included in the supplemental file.

Keywords: Bootstrap; copula; hazard rate ordering; least favorable configuration; order-restricted inference; ordinal dominance curve

1. Introduction

A random variable X is smaller than Y in uniform stochastic ordering, denoted by $X \preceq Y$ throughout this article, if $\mathrm{pr}(X > t|X > t_0) \leq \mathrm{pr}(Y > t|Y > t_0)$ for all $t > t_0$. Let F and G be the distribution functions of X and Y, respectively. When F and G are absolutely continuous, USO is also known as hazard rate ordering, an important characterization in reliability, econometrics, actuarial sciences, and biomedical studies (see, e.g., Dykstra [6], Navarro and Shaked [10], Da and Ding [3], Balakrishnan et al. [2]).

Acknowledging USO when it exists improves inference. For example, nonparametric estimators of F and G subject to a USO constraint are more efficient than unrestricted estimators when the constraint is true (Rojo and Samaniego [12,13], Mukerjee [9], Arcones and Samaniego [1]). On the other hand, wrongly acknowledging USO could compromise inference by inducing a large bias.

Nonparametric GOF tests for USO started from discrete settings (Dardanoni and Forcina [4], Park et al. [11]). Arcones and Samaniego [1] offered a one-sample test for continuous data. Tang et al. [14] proposed a two-sample GOF test that targets specifically at the situation where X and Y are independent.

In this article, we extend the two-sample test in Tang et al. [14] to account for paired observations where X and Y are likely not independent. Though our test statistic remains the same, we use a different way to find a data-driven critical value. We show that our critical value controls the type I error asymptotically and yields a consistent test regardless of the correlation between X and Y. Moreover, when X and Y are independent, our simulation shows that the new test is more powerful than the one in Tang et al. [14].

1.1. Hypotheses and test statistic

The hypotheses of our GOF test are $\mathcal{H}_0 : X \preceq Y$ versus $\mathcal{H}_1$: not $\mathcal{H}_0$. As introduced in Tang et al. [14], $X \preceq Y$ holds if and only if the ordinal dominance curve (ODC) $R = FG^{-1}$ is *star-shaped*, that is,

1350-7265

the slope function, $r(u) = \{1 - R(u)\}/(1 - u)$, of the secant line from the point $(1, 1)$ to $(u, R(u))$ is non-increasing in $u \in [0, 1)$ (see Figure 1 in Tang et al. [14] for an illustration). Following their work, we assume that the joint density of (X, Y) is continuous and that the first derivative of R is bounded over $[0, 1]$. Denote the parameter space of R by Θ, the collection of non-decreasing and continuously differentiable functions from $[0, 1]$ to $[0, 1]$. Then, the hypotheses can be expressed equivalently as

$$\mathcal{H}_0 : R \in \Theta_0 = \{\theta \in \Theta : \theta \text{ is star-shaped}\} \quad \text{versus} \quad \mathcal{H}_1 : R \in \Theta \setminus \Theta_0.$$

Let the data be $\{(X_i, Y_i)^{\mathrm{T}}\}_{i=1}^n$, n independent and identically distributed copies of $(X, Y)^{\mathrm{T}}$. We use the same test statistic as in Tang et al. [14], written by

$$M_{n,p} = n^{1/2} \|\mathcal{M}R_n - R_n\|_p,$$

for $1 \leq p \leq \infty$, where $n^{1/2}$ is a normalizing constant, $\|\cdot\|_p$ denotes the L^p norm, R_n is an unrestricted estimator of R, and $\mathcal{M}R_n$ is a restricted estimator with respect to $\mathcal{H}_0$. Specifically, $R_n = F_n G_n^{-1}$ is the empirical ODC, where $F_n(x)$ is the empirical distribution function of X, and $G_n^{-1}(u)$ is the empirical quantile function of Y; $\mathcal{M}R_n(u)$ is the smallest star-shaped function that is at least as large as R_n calculated by $\mathcal{M}R_n = 1 - (1 - u) \inf_{0 \leq v \leq u} [\{1 - R_n(v)\}/(1 - v)]$ for $u \in [0, 1)$ and $\mathcal{M}R_n(1) = 1$. The operator $\mathcal{M}$ is called the *least star-shaped majorant operator*. Large values of $M_{n,p}$ are evidence against $\mathcal{H}_0$.

1.2. Non-differentiability

Let $\mathcal{D} = \mathcal{M} - \mathcal{I}$, where $\mathcal{I}$ is the identity operator; i.e., $\mathcal{I}R = R$. Lemma 5 in Tang et al. [14] showed that neither $\mathcal{M}$ nor $\mathcal{D}$ is Hadamard directionally differentiable at $R \in \Theta_0$ unless R is of a special shape. The non-differentiability brings two unique difficulties to locate a valid critical value.

First, the non-differentiability does not permit the use of the functional delta method to derive the asymptotic distribution of $M_{n,p}$ under $\mathcal{H}_0$. Lemma 1.1 implies that $n^{1/2}(R_n - R)$ converges weakly to a stochastic process $\mathcal{T}_{R,C}$ as $n \to \infty$. The $\mathcal{T}_{R,C}$ satisfies

$$\mathcal{T}_{R,C}(u) = \mathcal{B}_C\{R(u), 1\} - R'(u)\mathcal{B}_C(1, u)$$

for $u \in [0, 1]$ where $\{\mathcal{B}_C(u, v) : 0 \leq u, v \leq 1\}$ is a mean-zero Gaussian process satisfying (1.1) and R' is the derivative of R. This lemma is similar to Theorem 2.2 in Hsieh and Turnbull [8] with X and Y now being dependent. For any $R \in \Theta_0$, $\mathcal{D}R = 0$. The term $\mathcal{M}R_n - R_n$ in $M_{n,p}$ equals $\mathcal{D}R_n - \mathcal{D}R$. If $\mathcal{D}$ is Hadamard directional differentiable at R with a derivative $d\mathcal{D}_R$, then Lemma 1.1 easily implies that $n^{1/2}(\mathcal{M}R_n - R_n)$ converges weakly to $d\mathcal{D}_R\mathcal{T}_{R,C}$. Unfortunately, the differentiability does not hold for all $R \in \Theta_0$.

Lemma 1.1. *Let C be the bivariate copula induced by the joint distribution of X and Y. There exists a sequence of mean-zero Gaussian processes, $\{\mathcal{B}_C^{(n)}(u, v) : 0 \leq u, v, \leq 1\}_{n=1}^{\infty}$, with covariance structure*

$$cov\big\{\mathcal{B}_C^{(n)}(u_1, v_1), \mathcal{B}_C^{(n)}(u_2, v_2)\big\} = C(u_1 \wedge u_2, v_1 \wedge v_2) - C(u_1, v_1)C(u_2, v_2), \tag{1.1}$$

where $u \wedge v = \min(u, v)$, such that the following holds almost surely,

$$\lim_{n\to\infty} \sup_{0 \leq u \leq 1} \Big| n^{1/2}\big\{R_n(u) - R(u)\big\} - \big[\mathcal{B}_C^{(n)}\big\{R(u), 1\big\} - R'(u)\mathcal{B}_C^{(n)}(1, u)\big]\Big| = 0.$$

Second, the non-differentiability prevents the standard bootstrap from estimating the sampling distribution of $M_{n,p}$ and thus brings difficulty to find a valid critical value. As discussed by Dümbgen [5] and further investigated by Fan and Santos [7], standard bootstrap inference does not work with operators that are not Hadamard directionally differentiable. Therefore, it is not surprising to see that by sampling the observed pairs with replacement, bootstrap versions of $M_{n,p}$ failed to estimate the sampling distribution of $M_{n,p}$.

These two challenges also presented when X and Y are independent. Tang et al. [14] pointed out that weak convergence of $n^{1/2}(\mathcal{M}R_n - R_n)$ is not a necessary prerequisite to derive the limiting distribution of $M_{n,p}$ and obtained the distribution without using the functional data method. Furthermore, they identified the *least favorable configuration* to compute a valid critical value. Though using the least favorable configuration controls the probability of type I error, it yields a quite conservative test and weakens the power of the test. Moreover, when X and Y are dependent, it is not clear whether the least favorable configuration still exists or not. In this article, we follow Tang et al. [14] to address the first challenge for paired observations, but provide a different critical value which not only works for dependent cases but also improves the power of the test when X and Y are truly independent.

2. Critical value

Now we explain the construction of our new critical value. We start with the limiting distribution of $M_{n,p}$ under the null hypothesis. This helps us quantify the probability of type I error asymptotically. For every $R \in \Theta_0$, we follow Tang et al. [14] to partition [0, 1] into a non-strictly star-shaped region S_1 and a strictly star-shaped region $[0, 1] \setminus S_1$. If S_1 is not empty, we can write S_1 as a union of pairwise disjoint intervals of the form $[a_k, b_k]$ for $0 \le a_k < b_k \le 1$.

Theorem 2.1. *For $R \in \Theta_0$, if $S_1 = \varnothing$, define $\mathcal{D}_{S_1}\mathcal{T}_{R,C}(u) = 0$ for all $u \in [0, 1]$; if $S_1 = \cup_k [a_k, b_k]$, define*

$$\mathcal{D}_{S_1}\mathcal{T}_{R,C}(u) = \sum_k \left[(1-u) \sup_{a_k \le v \le u} \left\{\frac{\mathcal{T}_{R,C}(v)}{1-v}\right\} - \mathcal{T}_{R,C}(u)\right] I(a_k \le u \le b_k)$$

for $u \in [0, 1)$ and $\mathcal{D}_{S_1}\mathcal{T}_{R,C}(1) = 0$. The $M_{n,p}$ converges in distribution to $\|\mathcal{D}_{S_1}\mathcal{T}_{R,C}\|_p$ for $1 \le p \le \infty$ as $n \to \infty$.

When C is the independence copula, or equivalently, X and Y are independent, Theorem 2.1 simplifies to Theorem 1 in Tang et al. [14]. The simplified version facilitates the finding of Theorem 2 in Tang et al. [14] that the unique least favorable configuration is R_0, where R_0 satisfies $R_0(u) = u$ for all $u \in [0, 1]$ and is known as the *equal distribution line* (i.e., R_0 holds when $F = G$). It means that, when X and Y are independent,

$$\lim_{n\to\infty} \mathrm{pr}_{R\in\Theta_0}(M_{n,p} > t) \le \lim_{n\to\infty} \mathrm{pr}_{R=R_0}(M_{n,p} > t);$$

that is, controlling the type I error at R_0 controls the type I error at all $R \in \Theta_0$. When C is not the independence copula, the dependence between X and Y makes the identification of the least favorable configuration very challenging.

The following theorem suggests an alternative way to control the type I error.

Theorem 2.2. *Let $\mathcal{L}_{R,C}(u) = \mathcal{B}_C\{R(u), 1\} - r(u)\mathcal{B}_C(1, u)$ for $u \in [0, 1]$. If R is star-shaped,*

$$\lim_{n\to\infty} \mathrm{pr}(M_{n,p} > t) \le \mathrm{pr}\big(\|\mathcal{D}_{[0,1]}\mathcal{L}_{R,C}\|_p > t\big) \tag{2.1}$$

for all $p \in [1, \infty]$ and all $t \in \mathbb{R}$, where

$$\mathcal{D}_{[0,1]}\mathcal{L}_{R,C}(u) = \left[(1-u) \sup_{0\le v\le u}\left\{\frac{\mathcal{L}_{R,C}(v)}{1-v}\right\} - \mathcal{L}_{R,C}(u)\right] I \quad (0 \le u < 1),$$

where the inequality in (2.1) becomes equality only when $R = R_0$.

When $R \in \Theta_0$, letting the critical value be c, Theorem 2.2 reveals that the probability of type I error $\mathrm{pr}(M_{n,p} > c)$ is asymptotically bounded by $\mathrm{pr}(\|\mathcal{D}_{[0,1]}\mathcal{L}_{R,C}\|_p > c)$. Therefore, there is no need to find the least favorable configuration to bounded the type I error. If we can estimate the distribution of $\|\mathcal{D}_{[0,1]}\mathcal{L}_{R,C}\|_p$, taking the $1-\alpha$ quantile of the estimator suffices to control the size of the test to be no larger than α.

The $\mathcal{L}_{R,C}$ involves three terms: $\mathcal{B}_C$, R, and r. We first use bootstrap to estimate $\mathcal{B}_C$. By sampling $\{(X_i^*, Y_i^*)^{\mathrm{T}}\}_{i=1}^n$ with replacement from $\{(X_i, Y_i)^{\mathrm{T}}\}_{i=1}^n$, the bootstrap estimator is

$$\mathcal{B}_C^*(u, v) = n^{-1/2}\sum_{i=1}^n \big[I\big\{X_i^* \le F_n^{-1}(u), Y_i^* \le G_n^{-1}(v)\big\} - I\big\{X_i \le F_n^{-1}(u), Y_i \le G_n^{-1}(v)\big\}\big]$$

for $0 \le u, v \le 1$, where F_n^{-1} and G_n^{-1} are the empirical quantile function associated with F and G, respectively. Furthermore, we note that (2.1) only holds for a star-shaped R. We estimate R and r by their restricted estimators (subject to $\mathcal{H}_0$) $\mathcal{M}R_n$ and $\widehat{r}_n$, respectively, where $\widehat{r}_n(u) = \{1 - \mathcal{M}R_n(u)\}/(1-u)$ for $u \in [0, 1)$ and $\widehat{r}_n(1) = \lim_{u\uparrow 1}\{1 - \mathcal{M}R_n(u)\}/(1-u)$. Finally, we estimate $\mathcal{L}_{R,C}(u)$ by

$$\mathcal{L}_{R,C}^*(u) = \mathcal{B}_C^*\big\{\mathcal{M}R_n(u), 1\big\} - \widehat{r}_n(u)\mathcal{B}_C^*(1, u) \quad \text{for } u \in [0, 1].$$

The following lemma demonstrates that the distribution of $\|\mathcal{D}_{[0,1]}\mathcal{L}_{R,C}^*\|_p$ conditional on the data well approximates the distribution of $\|\mathcal{D}_{[0,1]}\mathcal{L}_{R,C}\|_p$ for all $p \in [1, \infty]$.

Lemma 2.1. *If R is star-shaped,*

$$\sup_{t\in\mathbb{R}}\big|\mathrm{pr}_*\big(\big\|\mathcal{D}_{[0,1]}\mathcal{L}_{R,C}^*\big\|_p > t\big) - \mathrm{pr}\big(\|\mathcal{D}_{[0,1]}\mathcal{L}_{R,C}\|_p > t\big)\big|$$

converges to zero in probability, for all $p \in [1, \infty]$, as $n \to \infty$, where pr_* denotes the probability conditional on the observed data.

Lemma 2.1 suggests that, at significance level α, the $1-\alpha$ quantile of the distribution of $\|\mathcal{D}_{[0,1]}\mathcal{L}_{R,C}^*\|_p$ conditional on the observe data asymptotically controls the size of the test to be no larger than α. We denote that quantile by $\widehat{c}_{\alpha,p}$ and reject $\mathcal{H}_0$ if $M_{n,p} > \widehat{c}_{\alpha,p}$. In practice, we can repeat the bootstrap K times and take $\widehat{c}_{\alpha,p}$ to be the $1-\alpha$ sample quantile of the K bootstrap versions of $\|\mathcal{D}_{[0,1]}\mathcal{L}_{R,C}^*\|_p$.

Theorem 2.3. *For $p \in [1, \infty]$, if $R \in \Theta_0$*

$$\lim_{n\to\infty} \mathrm{pr}(M_{n,p} > \widehat{c}_{\alpha,p}) \le \alpha \tag{2.2}$$

where the equality holds when $R = R_0$; *if* $R \in \Theta \setminus \Theta_0$,

$$\lim_{n\to\infty} \text{pr}(M_{n,p} > \widehat{c}_{\alpha,p}) = 1.$$

Finally, Theorem 2.3 concludes that the $\widehat{c}_{\alpha,p}$ yields an asymptotic size α test, and the test is consistent. It is important to note that using this critical value, the test could still be conservative because $M_{n,p}$ converges in distribution to a degenerated term 0 when R is strictly star-shaped. But our simulation shows the new critical value yields a larger power than the one used in Tang et al. [14] when X and Y are truly independent.

3. Simulation

We use simulation to assess the finite-sample performance of our test. To assess size properties, we consider the equal distribution line R_0 and four star-shaped ODCs R_1, R_2, R_3 and R_4 shown in Figure 3 of Tang et al. [14]. All of our simulation results are based on 1000 Monte Carlo data sets. To generate these data sets, we let $F(u) = R_k(u)$ and $G(u) = u$ for $u \in [0, 1]$, where $k = 0, 1, \ldots, 4$, and take C to be the Gaussian copula with a correlation coefficient ρ. Then, we sample $\{(U_i, V_i)^{\mathrm{T}}\}_{i=1}^n$ from C and compute $X_i = F^{-1}(U_i)$ and $Y_i = G^{-1}(V_i)$. As a result, the pairs are from a joint distribution with its ODC being R_k. Throughout this section, we use $K = 1000$ to compute $\widehat{c}_{\alpha,p}$, because it is sufficient to approximate the distribution of $M_{n,p}$ at $R = R_0$ for $n \in \{200, 400\}$ and $\rho \in [-0.9, 0.9]$. In practice, one could first estimate C by $\widehat{C}_n$, then choose a K such that the distribution of $M_{n,p}$ at $R = R_0$ and $C = \widehat{C}_n$ can be well approximated.

Table 1 summarizes Monte Carlo estimates of the probability of rejecting $\mathcal{H}_0$ when $R = R_k$, for $k \in \{0, \ldots, 4\}$, $n \in \{200, 400\}$, $p \in \{1, 2, \infty\}$, and $\rho \in \{-0.8, 0, 0.8\}$, at $\alpha = 0.05$. Because the equality in (2.2) holds when $R = R_0$, we first check the finite-sample performance when $R = R_0$. Using 1000 data sets has 0.018 as the margin of error in estimating the probability of type I error. We see that our test confers the nominal size for all considered ρ. Furthermore, the estimated rejection rates for R_1, R_2, R_3 and R_4 are all bounded by $\alpha = 0.05$, which reinforces the inequality (2.2) in Theorem 2.3. In addition, we see that $p = \infty$ often yields the least conservative results than $p \in \{1, 2\}$.

Table 1. Size study. Estimated probability ($\times 10^3$) of rejecting $\mathcal{H}_0$ at $\alpha = 0.05$ for $n \in \{200, 400\}$, $\rho \in \{-0.8, 0, 0.8\}$, and $p \in \{1, 2, \infty\}$. The considered star-shaped ODCs are $R_0, R_1, \ldots, R_4$

		$p=1$			$p=2$			$p=\infty$		
	ρ	−0.8	0	0.8	−0.8	0	0.8	−0.8	0	0.8
R_0	$n=200$	60	62	61	58	56	51	51	52	52
	$n=400$	65	56	65	61	55	54	65	57	61
R_1	$n=200$	20	21	15	32	27	29	49	42	44
	$n=400$	17	11	8	25	22	23	40	36	32
R_2	$n=200$	0	0	0	0	0	0	5	2	2
	$n=400$	0	0	0	0	0	0	4	2	0
R_3	$n=200$	34	25	11	38	28	26	47	39	43
	$n=400$	32	22	11	38	37	29	48	43	42
R_4	$n=200$	0	0	0	0	0	0	2	2	0
	$n=400$	0	0	0	0	1	0	3	2	0

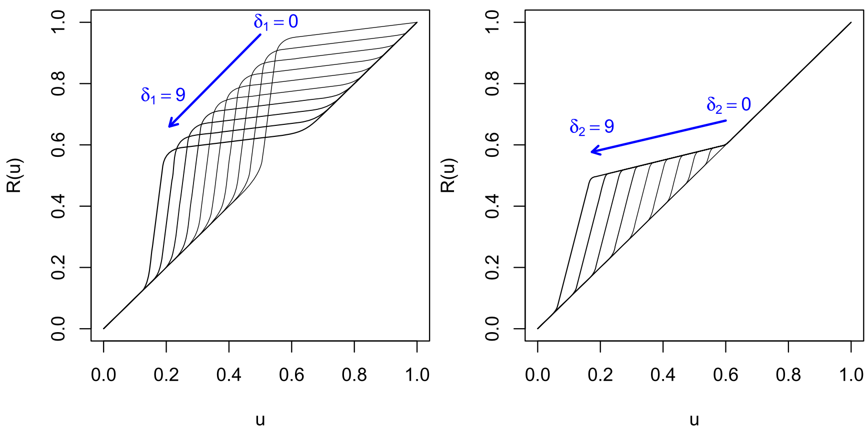

Figure 1. ODC sequences. Each sequence is indexed by a parameter δ_1 (left) and δ_2 (right), where $\delta_1, \delta_2 \in \{0, \ldots, 9\}$ and the $\delta_2 = 0$ member is R_0.

To assess power properties, we consider two sequences of ODCs displayed in Figure 1. Each sequence is indexed by a parameter δ whose $\delta = 0$ member is star-shaped (i.e., $X \preceq Y$ holds) and other members, $\delta \in \{1, \ldots, 9\}$, gradually move away from $\mathcal{H}_0$ as δ increases. The closed-form expression for each sequence is included in the supplementary file [15]. We repeat the aforementioned Monte Carlo procedure to estimate the rejection rate for each member of the two sequences. The first row in Figure 2 presents the results for $n = 200$, $p = \infty$, $\rho \in \{-0.8, 0, 0.8\}$, and $\alpha = 0.05$. When $\rho = 0$, the Gaussian copula reduces to the independence copula. We then compared with Tang et al. [14]. Results of this comparison are presented in the second row in Figure 2. The figures for $p \in \{1, 2\}$ and for $n = 400$ are of a similar pattern and are included in the supplemental file [15].

Subfigures in the first row in Figure 2 present estimated power curves. We see that the correlation ρ affects the power of the test; that is, as the correlation increases, the power increases. However, regardless of the value of ρ, the overall trend of the estimated power curves is the same. Because each ODC sequence starts with a star-shaped member, the power curves all start with a value less or around $\alpha = 0.05$. As the index parameter δ increases, all curves approach one. This trend reinforces Theorem 2.3 as well.

The second row in Figure 2 presents the comparison between using the new critical value and the one suggested in Tang et al. [14] when X and Y are independent ($\rho = 0$). The power gains of using our data-driven critical value are promising, even Tang et al.'s [14] test is provided that X and Y are independent in advance.

Acknowledgements

We would like to thank the Editor, the Associate Editor, and two referees for their constructive comments which have greatly improved the presentation of this article. This research was supported in part by R03 AI135614 grant from the National Institutes of Health and an ASPIRE grant from the Office of the Vice President for Research at the University of South Carolina.

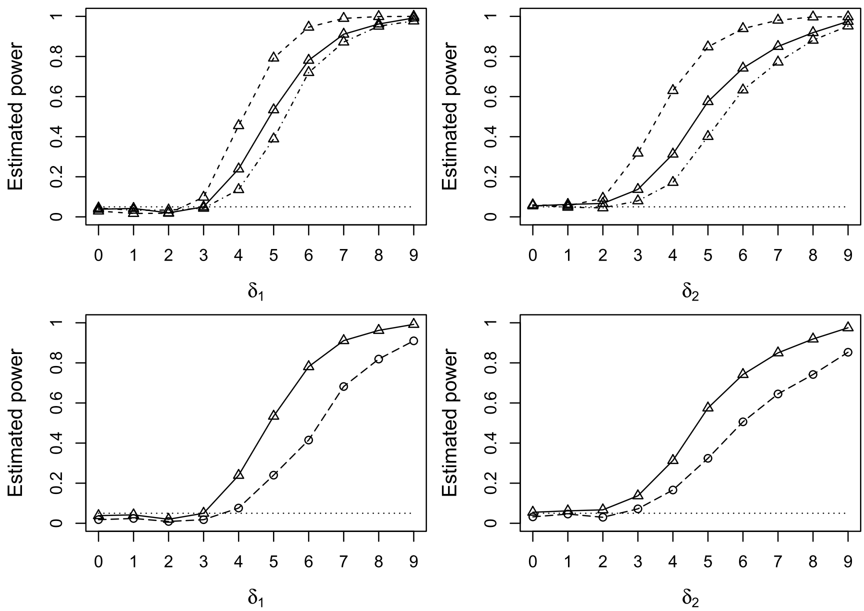

Figure 2. Estimated power curves using $p = \infty$ for sample sizes $n = 200$, $\rho \in \{-0.8, 0, 0.8\}$, and $\alpha = 0.05$. Curves using the new critical value are shown in triangles and connected by dot-dashes ($\rho = -0.8$), lines ($\rho = 0$), or dashes ($\rho = 0.8$). The second row focuses on $\rho = 0$ to compare with Tang et al.'s [14] critical value. Curves using Tang et al. [14] are shown in circles. The horizontal dotted line marks $\alpha = 0.05$. The left and right columns correspond to the δ_1 and δ_2 ODC sequences, respectively.

Supplementary Material

Supplement to "Testing against uniform stochastic ordering with paired observations" (DOI: 10.3150/21-BEJ1322SUPP; .pdf). The supplementary material includes proofs of Lemmas 1.1–2.1 and Theorems 2.1–2.3, closed-form expression of the ODCs plotted in Figure 1, and additional simulation results.

References

[1] Arcones, M.A. and Samaniego, F.J. (2000). On the asymptotic distribution theory of a class of consistent estimators of a distribution satisfying a uniform stochastic ordering constraint. *Ann. Statist.* **28** 116–150. MR1762906 https://doi.org/10.1214/aos/1016120367

[2] Balakrishnan, N., Zhang, Y. and Zhao, P. (2018). Ordering the largest claim amounts and ranges from two sets of heterogeneous portfolios. *Scand. Actuar. J.* **1** 23–41. MR3765136 https://doi.org/10.1080/03461238.2017.1278717

[3] Da, G. and Ding, W. (2016). Component level versus system level k-out-of-n assembly systems. *IEEE Trans. Reliab.* **65** 425–433.

[4] Dardanoni, V. and Forcina, A. (1998). A unified approach to likelihood inference on stochastic orderings in a nonparametric context. *J. Amer. Statist. Assoc.* **93** 1112–1123. MR1649205 https://doi.org/10.2307/2669854

[5] Dümbgen, L. (1993). On nondifferentiable functions and the bootstrap. *Probab. Theory Related Fields* **95** 125–140. MR1207311 https://doi.org/10.1007/BF01197342

[6] Dykstra, R., Kochar, S. and Rojo, J. (1997). Stochastic comparisons of parallel systems of heterogeneous exponential components. *J. Statist. Plann. Inference* **65** 203–211. MR1622774 https://doi.org/10.1016/S0378-3758(97)00058-X

[7] Fang, Z. and Santos, A. (2019). Inference on directionally differentiable functions. *Rev. Econ. Stud.* **86** 377–412. MR3936869 https://doi.org/10.1093/restud/rdy049

[8] Hsieh, F. and Turnbull, B.W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann. Statist.* **24** 25–40. MR1389878 https://doi.org/10.1214/aos/1033066197

[9] Mukerjee, H. (1996). Estimation of survival functions under uniform stochastic ordering. *J. Amer. Statist. Assoc.* **91** 1684–1689. MR1439110 https://doi.org/10.2307/2291596

[10] Navarro, J. and Shaked, M. (2006). Hazard rate ordering of order statistics and systems. *J. Appl. Probab.* **43** 391–408. MR2248572 https://doi.org/10.1239/jap/1152413730

[11] Park, C.G., Lee, C.-I.C. and Robertson, T. (1998). Goodness-of-fit test for uniform stochastic ordering among several distributions. *Canad. J. Statist.* **26** 69–81. MR1624377 https://doi.org/10.2307/3315674

[12] Rojo, J. and Samaniego, F.J. (1991). On nonparametric maximum likelihood estimation of a distribution uniformly stochastically smaller than a standard. *Statist. Probab. Lett.* **11** 267–271. MR1097984 https://doi.org/10.1016/0167-7152(91)90154-J

[13] Rojo, J. and Samaniego, F.J. (1993). On estimating a survival curve subject to a uniform stochastic ordering constraint. *J. Amer. Statist. Assoc.* **88** 566–572. MR1224382

[14] Tang, C.-F., Wang, D. and Tebbs, J.M. (2017). Nonparametric goodness-of-fit tests for uniform stochastic ordering. *Ann. Statist.* **45** 2565–2589. MR3737902 https://doi.org/10.1214/16-AOS1535

[15] Wang, D. and Tang, C.-F. (2021). Supplement to "Testing against uniform stochastic ordering with paired observations." https://doi.org/10.3150/21-BEJ1322SUPP

Received March 2020