Journal of Statistical Planning and Inference 141 (2011) 3244–3255

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

A class of exceedance-type statistics for the two-sample problem

Eugenia Stoimenova [a,*], N. Balakrishnan [b]

[a] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G.Bontchev str., block 8, 1113 Sofia, Bulgaria
[b] Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada L8S 4K1

ARTICLE INFO

Article history:
Received 28 November 2010
Received in revised form
4 April 2011
Accepted 6 April 2011
Available online 14 April 2011

Keywords:
Rank tests
Two-sample problem
Lehmann alternative
Exceedance statistics

ABSTRACT

We consider here a class of test statistics based on exceeding observations and develop exceedance-type tests for the two-sample hypothesis testing problem. The exact distribution of the statistics are derived under the null hypothesis as well as under the Lehmann alternative, and then a comparative power study is carried out.

1. Introduction

Let $X_1,\dots,X_m$ and $Y_1,\dots,Y_n$ be random samples from continuous distribution functions F and G, respectively. A standard problem is that of testing the hypothesis H_0 that F and G are identical against the one-sided alternative H_A that Y's are stochastically larger than X's with strict inequality for some x:

$$H_A: \; F(x) \geq G(x). \tag{1}$$

For this hypothesis testing problem, a test can be based on the number of observations from one sample which exceed (precede) some threshold specified by the other sample. The simplest such test is due to Rosenbaum (1954) which is based on the number of observations in the X-sample before smallest observation in the Y-sample.

Use of the r-th order statistic from the Y-sample, i.e., $Y_{(r)}$, instead of the smallest order statistic $Y_{(1)}$, results in the well-known precedence test; see, for example, van der Laan and Chakraborti (2001) and Balakrishnan and Ng (2006). In this case, the test statistic is simply the number of observations in the X-sample that are smaller than $Y_{(r)}$. Large values of this statistic lead to the rejection of the null hypothesis of equality of the two distributions. Precedence tests are useful in life-testing experiments wherein the data become available naturally in order of size. The experiment is terminated after a certain number of failures. Many generalizations of precedence tests are possible, including using record statistics instead of order statistics (Balakrishnan et al., 2008a), and for handling censored data Balakrishnan et al. (2008b), and Bairamov (2006). Interested readers may refer to Balakrishnan and Ng (2006) for a comprehensive treatment of precedence tests and their generalizations.

In the general situation of testing equality of two distributions, the precedences and exceedances are both necessary with respect to thresholds from both samples. The number of exceedances in the Y-sample with respect to a threshold

* Corresponding author. Tel.: +359 2 979 3855; fax: +359 2 971 3649.
E-mail addresses: jeni@math.bas.bg, jenistoimenova@gmail.com (E. Stoimenova), bala@mcmaster.ca (N. Balakrishnan).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2011.04.010

from the X-sample could be used along with the number of precedences in the X-sample with respect to a threshold from the Y-sample. Such a test has been studied by Stoimenova (to appear) using extremal order statistics as thresholds.

In this paper, we propose a family of rank statistics for the two-sample problem in which test statistics are based on thresholds from both samples. For equal sample sizes, the exceedance statistics are defined as

$$A_r = \text{the number of } Y\text{'s larger than } X_{(n-r)},$$

$$B_r = \text{the number of } X\text{'s smaller than } Y_{(1+r)}, \tag{2}$$

where $0 \le r < n$.

To test H_0 versus H_A in (1), we propose a test based on the statistic

$$M_r = \max\{n-A_r, n-B_r\}. \tag{3}$$

Evidently, small values of M_r lead to the rejection of H_0 in favor of the stochastically ordered alternative H_A.

The following example is useful for the numerical illustration of the proposed M_r-test statistic.

Example 1. The data set is taken from Nelson (1982, Table 2.1, p. 510). Consider X- and Y-samples to be the data from appliance cord life under Tests 2 and 1, respectively. These lifetimes are presented in Table 1.

In this case, we have $n=9$ and for $r=0,1,2,3$ we have presented in Table 2 the values of the exceedance statistics A_r and B_r and the corresponding M_r-test statistics.

In this paper, we study the distribution of the test statistic M_r defined in (3) for the case of equal sample sizes under the null hypothesis as well as under alternatives of the form $H_{LE}: \; G(x) = 1-\{1-F(x)\}^{1/\eta}, \; \eta > 1$. This subclass of H_A was introduced first by Lehmann (1953), and clearly, it can be considered as a rough approximation for a shift of G to the right with respect to F. The form of Lehmann alternatives facilitates the derivation of the distributions of some rank statistics.

The family of test statistics can be extended for unbalanced sample size case as well. Suppose the X-sample is of size m and the Y-sample is of size n. Then, we need to change the form of M_r in (3) suitably. Instead of taking r, we need to take a proportion, say p, and take $r_1 = [(m+1)(1-p)]$ and $r_2 = [(n+1)p]$, where $[\cdot]$ denotes the integer part. Clearly, these r_1 and r_2 will correspond to $(1-p)$-th and p-th sample quantiles from the two samples. Then, we modify the test statistic in (3) for this case to be

$$M_r = \max\{n-A_{r_1}, m-B_{r_2}\}.$$

The rest of this paper is organized as follows. First, we give a useful representation for the distribution of the M_r-statistic in Section 2 and derive the joint distribution of A_r and B_r in Section 3. The null distribution of the M_r-statistic is derived in Section 4. Critical values for $r=0,\ldots,10$ for some selected choices of sample sizes are presented. Exact distribution of the M_r-statistic under Lehmann alternatives is derived in Section 5. Next, we study the exact power properties of the tests under the Lehmann alternative, and for this purpose we derive the exact power function of these tests under the Lehmann alternatives in Section 6. In Section 7, a comparison is made with known precedence tests and here some examples are also presented to illustrate all the tests. Finally, we present in the Appendix near 5% critical values and exact levels of significance of M_r for $r=0,1, \ldots,10$ and $n=6,\ldots,25$.

2. Distribution of the test statistic

Let $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_n$ be random samples from continuous distribution functions F and G, respectively. Let A_r and B_r be as defined in (2).

Proposition 1. *The distribution of the M_r-statistic, $M_r = \max\{n-A_r, n-B_r\}$, is represented by*

$$P(M_r = n) = P(B_r = 0) + P(B_r > 0, A_r = 0), \tag{4}$$

Table 1
Appliance cord life times.

X	57.5	77.8	88.0	98.4	102.1	105.3	139.3	143.9	148.0
Y	96.9	100.3	100.8	103.3	103.4	105.4	122.6	151.3	162

Table 2
Exceedance statistics and the M_r-test statistic.

r	A_r	B_r	M_r
0	2	3	7
1	2	4	7
2	2	4	7
3	4	5	5