ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 156, 2002

*Agnieszka Rossa**

SIMULATION STUDY OF TWO-SAMPLE KOLMOGOROV–SMIRNOV TEST IN RANDOMLY CENSORED DATA

Abstract. The paper deals with a problem of testing the non-parametric hypothesis that two populations are equally distributed in the situation when the observations are subject to random censoring. A general metric for measuring the distance between two distributions is the Kolmogorov metric and the corresponding test is the Two-Sample Kolmogorov–Smirnov test. In the report below we present results of a simulation study performed for three versions of the Two-Sample Kolmogorov–Smirnov test for censored data. These three versions are generated by three methods of treating censored observations. Basic statistical properties of these tests are inspected by means of Monte Carlo simulations.

Key words: censored data, Kolmogorov–Smirnov test, Monte Carlo simulations.

INTRODUCTION

Censored data are fundamentally different from other types of data in the sense that the response of interest (the time until a specified event) is not always fully observed because some causes can interrupt the observation before the event occurs. Randomly censored data occur frequently in many fields of applied statistics: e.g. industrial applications and technology (reliability theory and life-testing), medical and biological studies (survival time), economic studies (e.g. when one is preparing a report on the duration of a phenomena and some of them are still in run) etc. References, especially in medical applications, are in abundance (e.g. Altman (1991), Marubini and Valsecchi (1996)).

We are interested in testing the non-parametric hypothesis that two populations are equally distributed in the situation when the observations are subject to random censoring. Typically the Mann–Whitney–Wilcoxon

* Dr., Department of Statistical Methods, University of Łódź.

test (e.g. Gehan (1965)), the log-rank test (e.g. Marubini and Valescchi (1996)), or some other combinatorial tests are used. However applying such tests we are not able to control their critical regions according to our knowledge concerning the alternative hypothesis and the tests may not reflect the real-life situation under consideration. For instance the Mann–Whitney–Wilcoxon test is constructed for alternatives formed by shifting the distribution specified in the null hypothesis.

A general metric for measuring the distance between two distributions is the Kolmogorov metric and the corresponding test is the Two-Sample Kolomogorov–Smirnov test. It measures the distance between two empirical distribution functions in terms of Kolmogorov metric. In the report below we present results of a simulation study performed for three versions of the Two-Sample Kolmogorov–Smirnov test under random censoring. These three versions are generated by three methods of treating censored observations. As a result we are able to assess how much we lose in the effect of censoring.

II. STATEMENT OF THE PROBLEM

Let X and Y be positive random variables representing failure time in two populations of individuals. Let F and G denote unknown continuous cumulative distribution functions of X and Y, respectively. The problem is to test the null hypothesis

$$\mathrm{H}_0 : F = G$$

against a general alternative

$$\mathrm{H}_1 : F \neq G$$

Let Z be a positive random variable (censoring variable) independent on X and Y, distributed according to a cumulative distribution function H. In consequence, what we observe are two censored samples

$$(X'_1, \delta_1^{(1)}), \quad (X'_2, \delta_2^{(1)}), \ \ldots, \ (X'_m, \delta_m^{(1)}) \tag{1}$$

and

$$(Y'_1, \delta_1^{(2)}), \quad (Y'_2, \delta_2^{(2)}), \ \ldots, \ (Y'_n, \delta_n^{(2)})$$

where

$$X'_k = \min(X_k, Z_k) \quad \text{and} \quad \delta_k^{(1)} = I(X_k \leqslant Z_k), \quad \text{for } k = 1, 2, \ldots, m,$$

$$Y_l' = \min(Y_l, Z_l) \quad \text{and} \quad \delta_l^{(1)} = I(X_l \leqslant Z_l), \quad \text{for } l = 1, 2, \ldots, n,$$

where I denotes the indicator function. $X_1, X_2, \ldots, X_m$ are independent and identically distributed random variables with cumulative distribution function F, and $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed random variables with cumulative distribution function G.

To assess the behaviour of the Two-Sample Kolmogorov–Smirnov test under random censoring we consider four statistics D, D_1, D_2 and D_{KM} defined below.

First, let us consider a standard case, when the data are not subject to random censoring. Let F_m and G_n be empirical distribution functions from two uncensored samples $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$

$$F_m(x) = \frac{1}{m}\sum_{k=1}^{m} I(X_k \leqslant x) \quad \text{and} \quad G_n(x) = \frac{1}{n}\sum_{l=1}^{n} I(Y_l \leqslant x)$$

thus the standard Kolmogorov metric takes the form

$$D = \sup_x |F_m(x) - G_n(x)| \tag{2}$$

Now, let us consider two censored samples given in (1). Let F_m' and G_n' be empirical distribution functions calculated from these two samples

$$F_m'(x) = \frac{1}{m}\sum_{k=1}^{m} I(X_k' \leqslant x) \quad \text{and} \quad G_n'(x) = \frac{1}{n}\sum_{l=1}^{n} I(Y_l' \leqslant x)$$

then the Kolmogorov metric for censored samples will be defined as follows

$$D_1 = \sup_x |F_m'(x) - G_n'(x)| \tag{3}$$

Let m' and n' denote numbers of uncensored observations in both samples (1). It is clear that

$$m' = \sum_{k=1}^{m} \delta_k^{(1)} \quad \text{and} \quad n' = \sum_{l=1}^{n} \delta_l^{(2)}$$

Denote by $\tilde{F}_m$ and $\tilde{G}_n$ two empirical distribution functions calculated from these uncensored observations in both samples. Thus

$$\tilde{F}(x) = \frac{1}{m'}\sum_{\{k:\delta_k^{(1)}=1\}} I(X_k \leqslant x) \quad \text{and} \quad \tilde{G}_n(x) = \frac{1}{n'}\sum_{\{l:\delta_l^{(2)}=1\}} I(Y_l \leqslant x)$$

The Kolmogorov metric based on reduced samples takes the form

$$D_2 = \sup_x |\tilde{F}_m(x) - \tilde{G}_n(x)| \tag{4}$$

Let F_m^{KM}, G_n^{KM} be the non-parametric Kaplan–Meier estimators (see Kaplan and Meier (1958)) of the respective cumulative distribution functions F and G

$$F_m^{KM}(x) = 1 - \prod_{x_{(k)} \leqslant x} \left(\frac{m-k}{m-k+1} \right)^{\delta_k^{(1)}}, \quad \text{where} \quad x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(m)}$$

$$G_n^{KM} = 1 - \prod_{y_{(l)} \leqslant x} \left(\frac{n-l}{n-l+1} \right)^{\delta_l^{(2)}}, \quad \text{where} \quad y_{(1)} \leqslant y_{(2)} \leqslant \ldots \leqslant y_{(n)}$$

The Kolmogorov metric based on the Kaplan-Meier estimators will be defined as follows

$$D_{KM} = \sup_x |F_m^{KM}(x) - G_n^{KM}(x)| \tag{5}$$

III. SIMULATIONS

The aim was to study power performance of the presented test procedures based on statistics D, D_1, D_2 and D_{KM} given in (2), (3), (4) and (5), respectively. In order to determine achieved significance levels and powers as the distribution G and H were varied, we performed a Monte Carlo study for test statistics: D, D_1 and D_2. Power performance of D_{KM} has to be considered separately. To control the size of tests D, D_1 and D_2 the randomised tests were employed.

For variables X and Y exponential distributions F and G were considered, i.e. $F \sim \mathrm{E}(0, 1)$, $G \sim \mathrm{E}(0, c)$, where pdf $\mathrm{E}(0, c) \propto e^{-x/c}$. For censoring variable Z a gamma distribution H was assumed i.e. $H \sim \mathrm{Gamma}\ (a, 1)$, where pdf $\mathrm{Gamma}\ (a, b) \propto x^{a-1} e^{-x/b}$.

To measure the degree of censoring the following probabilities p_1 and p_2 were evaluated

$$p_1 = P(X > Z | X \sim F,\ Z \sim H) = \frac{1}{2^a} \tag{6}$$

$$p_2 = P(Y > Z \mid Y \sim G,\ Z \sim H) = \frac{1}{\left(1 + \frac{1}{c}\right)^a} \tag{7}$$

For some combinations of parameters a and c two censored samples of size $m = n = 10$ were generated. Based on these data it was determined whether D, D_1 and D_2 tests reject H_0 at significance level 0.1. The percentages of rejections out of the 10 000 replicates (i.e. their simulated powers) were computed. Table 1 summarises the simulated powers. The first block of Table 1 comprises results obtained for various values of parameter c and for $a = 1$, while the second one comprises results obtained for various values of parameter c and for $a = 3$. Notice, that when $c = 1$ the distribution functions F and G are equal. Thus in this case the simulated rate of rejections of H_0 reflects a simulated significance level which agrees with the nominal one equal to 0.1.

Table 1

Simulated powers of the 10%-level tests D, D_1 and D_2, sample sizes $m = n = 10$
$F \sim E(0, 1)$, $G \sim E(0, c)$, $H \sim \text{Gamma}(a, 1)$
10 000 replications for each combination of parameters a and c

Values of scale parameter c	Power values of tests			Simulated and exact fractions of censoring			
	D	D_1	D_2	$\hat{p}_1$	p_1	$\hat{p}_2$	p_2
		$a = 1$					
1.0000	0.1002	0.1002	0.1040	0.499	0.500	0.502	0.500
1.5000	0.4713	0.1154	0.1063	0.499	0.500	0.601	0.600
2.0000	0.6205	0.1374	0.1120	0.499	0.500	0.669	0.667
2.5000	0.7311	0.1586	0.1181	0.499	0.500	0.716	0.714
3.0000	0.8121	0.1730	0.1215	0.499	0.500	0.751	0.750
3.5000	0.8645	0.1876	0.1246	0.499	0.500	0.779	0.778
4.0000	0.9005	0.1999	0.1228	0.499	0.500	0.801	0.800
4.5000	0.9283	0.2076	0.1202	0.499	0.500	0.819	0.818
5.0000	0.9453	0.2148	0.1169	0.499	0.500	0.834	0.833
5.5000	0.9580	0.2216	0.1123	0.499	0.500	0.847	0.846
6.0000	0.9686	0.2284	0.1097	0.499	0.500	0.857	0.857
7.0000	0.9782	0.2349	0.1012	0.499	0.500	0.876	0.875
8.0000	0.9851	0.2422	0.0942	0.499	0.500	0.889	0.889
		$a = 3$					
1.0000	0.1016	0.1017	0.1015	0.125	0.125	0.126	0.125
1.5000	0.1898	0.1604	0.1310	0.125	0.125	0.217	0.216
2.0000	0.3307	0.2639	0.1704	0.125	0.125	0.298	0.296

Table 1 (contd.)

Values of scale parameter c	Power values of tests			Simulated and exact fractions of censoring			
	D	D_1	D_2	$\hat{p}_1$	p_1	$\hat{p}_2$	p_2
		$a = 3$					
2.5000	0.4783	0.3643	0.2057	0.125	0.125	0.367	0.364
3.0000	0.5918	0.4438	0.2299	0.125	0.125	0.424	0.422
3.5000	0.6870	0.5115	0.2479	0.125	0.125	0.472	0.471
4.0000	0.7616	0.5681	0.2589	0.125	0.125	0.514	0.512
4.5000	0.8154	0.6139	0.2636	0.125	0.125	0.551	0.548
5.0000	0.8575	0.6494	0.2649	0.125	0.125	0.581	0.579
5.5000	0.8865	0.6788	0.2697	0.125	0.125	0.608	0.606
6.0000	0.9103	0.7040	0.2719	0.125	0.125	0.633	0.630
50.0000	1.0000	0.9210	0.1150	0.125	0.125	0.940	0.942
80.0000	1.0000	0.9320	0.0740	0.125	0.125	0.967	0.963

IV. RESULTS AND CONCLUSIONS

The standard Two-Sample Kolmogorov–Smirnov test based on statistic D was considered as a benchmark for our study of D_1 and D_2 for censored data. It is obvious that the loss of power for statistics D_1 and D_2 is due to censoring. However, the influence of censoring on D_1 and D_2 differs markedly. The loss of power for D_1 is caused by the fact that the Kolmogorov distance between F' and G' is smaller than the distance between F and G. On the other hand, it can be seen that the power performance for the second statistic D_2 is very sensitive to the sample sizes. We can notice, that the statistic D_2 was calculated for reduced samples, obtained by elimination of censored observations. Thus the sample sizes were random in this case. What is more, the power values obtained for D_1 are greater than these ones obtained for D_2. This seems to be true for different alternative hypotheses. Thus statistic D_2 is markedly less powerful than statistic D_1.

It is also worth noting, that D_2 exhibits non-monotone change in its power. We can observe that the power values of D_2 decrease when the censoring fraction increases. For heavy censoring the power of D_2 drops even beneath the assumed significance level, so the test appears to be biased in such cases. Due to a serious loss of power for D_2 the Kolmogorov–Smirnov test for random reduced samples cannot be recommended when censoring is present. The Monte Carlo study has clearly shown that D_1 test is much better than D_2.

V. REMARKS ON DISTRIBUTION OF STATISTIC D_{KM}

The main problem in applying the statistic D_{KM} lies in finding its exact or approximate distribution under the null hypothesis. Unfortunately, the distribution of D_{KM} depends in a very complicated way on the censoring distribution H. As yet, we are able to derive the distribution of statistic D_{KM} under rather strong hypothesis: $F = G = H$. For instance for $m = n = 10$ and for the assumed pattern of censoring in both ordered samples given by the assumed order of ones and zeros in the sequence of δ-values i.e. for $\delta^{(1)} = (1, 0, 1, 0, 1, 0, 1, 0, 0, 1)$ and $\delta^{(2)} = (1, 0, 0, 1, 0, 1, 0, 1, 0, 1)$ we obtained the distribution of statistic D_{KM}, given in Table 2.

Table 2

The exact and simulated distribution of statistic D_{KM}
for $\delta^{(1)} = (1, 0, 1, 0, 1, 0, 1, 0, 0, 1)$ and $\delta^{(2)} = (1, 0, 0, 1, 0, 1, 0, 1, 0, 1)$
$m = n = 10$

d_i	0.411	0.489	0.492	0.508	0.617	0.656	0.771	0.788	0.900
Exact probabilities $p_i = P(D_{KM} = d_i)$	0.392	0.002	0.456	0.001	0.089	0.038	0.015	0.005	0.002
Simulated probabilities $\hat{p}_i$	0.389	0.002	0.458	0.002	0.092	0.036	0.014	0.005	0.002

This distribution differs from the distribution of statistic D given in Table 3.

Table 3

The exact distribution of statistic D
$m = n = 10$

d_i	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Probabilities $P(D = d_i)$	0.006	0.207	0.369	0.250	0.115	0.040	0.011	0.002	0.000	0.000

What is more the distribution of statistic D_{KM} changes when the pattern of censoring in one or both samples is changed. Thus D_{KM} can hardly be applied for testing the equality $F = G$ without referring to the pattern of censoring.

VI. ACKNOWLEDGEMENTS

I am grateful to Professor Ryszard Zieliński for numerous discussions, suggestions and help with the numerical computations.

REFERENCES

Altman D. G. (1991), *Practical Statistics for Medical Research*, Chapman & Hall.

Gehan E. (1965), *A Generalized Wicoxon Test for Comparing Arbitrarily Singly Censored Samples*, Biometrika, **52**, 203–223.

Kaplan E. L., Meier, P. (1958), *Nonparametric Estimation From Incomplete Observations*, JASA, **53**, 457–481.

Marubini E., Valsecchi, M. G. (1996), *Analysing Survival Data from Clinical Trials and Observation Studies*, Wiley.

Agnieszka Rossa

TEST ZGODNOŚCI KOŁMOGOROWA-SMIRNOWA DLA DANYCH LOSOWO CENZUROWANYCH - ANALIZA SYMULACYJNA

(Streszczenie)

W artykule przedstawione są trzy wersje testu zgodności Kołmogorowa–Smirnowa dla danych prawostronnie cenzurowanych. Poszczególne testy różnią się sposobem podejścia do obserwacji cenzurowanych. Moc testów została zbadana i porównana za pomocą symulacji Monte Carlo.