

# THE OPTIMAL TWO-SAMPLE TEST WITH LIFETIME DATA UNDER THE WALD MAXIMIN MODEL OF DECISION MAKING UNDER RISK AND UNCERTAINTY

P. A. Philonenko, S. N. Postovalov

*Novosibirsk State Technical University, 630073, Novosibirsk, Russia*

UDC 519.234.3

In this paper, the two-sample problem with right-censored data is considered. Using the maximin Wald model of the game theory and using the Monte-Carlo method for the simulation of statistical test power, we research the two-sample tests stability on various types of an alternative hypothesis. The represented results are recommendations on the two-sample tests application on a certain type of an alternative hypothesis. Moreover, we propose a new two-sample *MIN3* test for right-censored data that the test power is close to the corresponding Wald test statistic.

**Keywords:** survival analysis, right-censored data, two-sample problem, Monte-Carlo method, test power, proposed *MIN3* test.

## Introduction

In hypothesis testing, there are always such two alternative hypotheses A and B where one two-sample test is more preferable in the alternative A and less preferable in the alternative B than another two-sample test [1]. Hence, the most powerful two-sample test does not exist in the general case. However, the analysis of two-sample tests power by the Wald test [1] for decision making under risk and uncertainty can determine what the test is more preferable in a certain type of alternative hypotheses. We have constructed the types of alternative hypotheses and determined three two-sample tests complementing each other in terms of the test power [1]. They are Multiple Crossing [2], Single Corssing [3] and weighted Kaplan-Meier [4] two-sample tests. Using test statistics of these tests, we have proposed a new two-sample test MIN3 that has the test power close to the maximum of these tests power. The proposed test is a Wald optimal test and its test power do not have the difference more than 5% from the corresponding Wald test statistic.

In section 1, we consider the problem statement. In section 2, we represented applied two-sample tests for right-censored data. In section 3, we consider results of the simulations.

## 1 Problem Statement

Suppose that we have two samples of continues variables  $\xi_1$  and  $\xi_2$  respectively,  $X_1 = \{t_{11}, t_{12}, \dots, t_{1n_1}\}$  and  $X_2 = \{t_{21}, t_{22}, \dots, t_{2n_2}\}$  of two survival distributions  $S_1(t)$  and  $S_2(t)$ . The observation  $t_{ij} = \min(T_{ij}, C_{ij})$ , where  $T_{ij}$  and  $C_{ij}$  are the failure and censoring times for the  $j$ -th object of the  $i$ -th group.  $T_{ij}$  and  $C_{ij}$  are i.i.d. with CDF  $F_i(t)$  and  $F_i^C(t)$  respectively. Survival curve means the probability of survival in the time interval  $(0, t)$

$$S_i(t) = P\{\xi_i > t\} = 1 - F_i(t).$$

Then the null hypothesis is

$$H_0 : S_1(t) = S_2(t).$$

---

This research has been supported by Russian Ministry of Education and Science as a part of the state task (project 1.1009.2017/4.6) and by Novosibirsk State Technical University as a young scientist project (project C-15, 2017).

Further, we will suppose that the elements of samples are ordered:  $t_{11} < \dots < t_{1n_1}$  and  $t_{21} < \dots < t_{2n_2}$ . Also, we pool these samples and sort their elements by ascending order  $T = X_1 \cup X_2 = \{t_1, t_2, \dots, t_n\}$ , where  $t_1 < \dots < t_n$  and  $n = n_1 + n_2$ .

Denote the sample indicator  $v_i$  and the censoring indicators  $c_{ij}$  and  $c_i$  as

$$v_i = \begin{cases} 0, & t_i \in X_1, \\ 1, & t_i \in X_2 \end{cases} \quad c_{ij} = \begin{cases} 0, & t_{ij} \text{ is failure,} \\ 1, & t_{ij} \text{ is censored} \end{cases} \quad c_i = \begin{cases} 0, & t_i \text{ is failure,} \\ 1, & t_i \text{ is censored.} \end{cases}$$

## 2 Applied Methods

### 2.1 Two-Sample Tests

In the paper, we apply various two-sample tests for right-censored data. They are the Gehan's generalized Wilcoxon test (denoted by  $G$ ) [5], Peto and Peto's generalized Wilcoxon test (denoted by  $P$ ) [6], log-rank test (denoted by  $LG$ ) [7], Cox-Mantel test (denoted by  $CM$ ) [8],  $Q$ -test (denoted by  $Q$ ) [9], maximum value test (denoted by  $MAX$ ) [10], Bagdonavicius-Nikuln tests for generalized Cox [11], multiple [2] and single [3] crossing models (denoted by  $BN1, BN2, BN3$  respectively), weighted Kaplan-Meier test [4] (denoted by  $WKM$ ) and proposed  $MIN3$  test (denoted by  $MIN3$ ).

The  $MIN3$  test statistic is

$$MIN3 = \min(pv_{BN2}, pv_{BN3}, pv_{WKM}),$$

where  $pv_S$  is a calculated  $p$ -value of two-sample test with test statistic  $S$ .

### 2.2 Maximin Wald Test for Decision-Making Under Risk and Uncertainty

Let  $r$  be a strategy of the statistical test application with the test statistic  $S$  on the alternative hypothesis  $H$ . Then, for every strategy  $r$  can be set the utility using the utility function  $U(r|H)$  [12] (or abbreviated as  $U(r)$ ) [1].

The Wald test is one of tests for decision-making under risk and uncertainty [12]. It is the test of a "careful observer" and the Wald test optimizes the utility (the test power) under an assumption that the environment is unfavorable for the observer, i.e. power values of all tests on the alternative hypothesis  $H$  are minimal. The decision rule is as follows:

$$W = \max_{i=1, k} \min_{j=1, m} U(r_i|H_j),$$

where  $k$  is a number of tests,  $m$  is a number of alternative hypotheses. In fact, the best test in according to the Wald test is a statistical test whose minimal test power among all alternative hypotheses is maximal among all statistical tests.

## 3 Simulations

In this section, we consider values of the maximin Wald test statistic for various types of alternative hypotheses. The types of alternative hypotheses are represented in Table 1. There are types with 0, 1 and 2 points of intersections between survival functions  $S_1(t)$  and  $S_2(t)$ . Every type contains three alternative hypotheses with various distribution of failure time  $F(t)$ . Using the Monte-Carlo method, we simulate test power and then compute the corresponding the maximin Wald test statistic. The greater test power, the higher accuracy of statistical conclusions. The results of simulation are represented in Tables 2–10. The size of simulation is 150 000 Monte-Carlo replications. The test size  $\alpha$  is  $\alpha = 0.05$ . The sample sizes  $n_1, n_2$  are  $n_1 = n_2 = 200$  observations. The distributions of censored times  $F^C(t)$  are Weibull, Gamma and Exponential and censored rate in the range from 10% to 50%.

One can see that test power of the  $MIN3$  test is close to the corresponding the maximin Wald test statistic. It means that the  $MIN3$  test power is close to the maximum on the alternative hypothesis when values of all test power are minimal. The maximal difference between  $MIN3$  test power and corresponding  $W$  statistic is less than 0.05. It makes possible to conclude the  $MIN3$  two-sample test is a stable two-sample test for various types of an alternative hypothesis.

Table 1: Types of alternative hypotheses

$H_i$	Type of alternative hypothesis	$S_1 - S_2$	Intersections	$L^1$	$ \mu_1 - \mu_2 $	$ \sigma_1 - \sigma_2 $
$H_{01}$	Difference in early time without an intersection	$Exp - Exp$	-	0.099	0.10	0.10
$H_{02}$		$We - LgN$	-	0.096	0.10	0.66
$H_{03}$		$LgN - Exp$	-	0.109	0.18	0.00
$H_{04}$	Difference in middle time without an intersection	$\Gamma - Exp$	-	0.075	0.07	0.00
$H_{05}$		$Exp - We$	-	0.100	0.10	0.08
$H_{06}$		$Exp - We$	-	0.162	0.16	0.16
$H_{07}$	Difference in late time without an intersection	$Exp - \Gamma$	-	0.089	0.09	0.03
$H_{08}$		$We - Exp$	-	0.116	0.13	0.57
$H_{09}$		$We - LgN$	-	0.107	0.11	0.41
$H_{11}$	One point of intersection in early time	$Exp - Exp$	0.363	0.100	0.09	0.05
$H_{12}$		$\Gamma - \Gamma$	0.763	0.078	0.06	0.35
$H_{13}$		$We - Exp$	0.571	0.125	0.12	0.00
$H_{14}$	One point of intersection in middle time	$We - \Gamma$	0.611	0.081	0.08	0.01
$H_{15}$		$Exp - We$	0.843	0.097	0.04	0.12
$H_{16}$		$\Gamma - Exp$	1.040	0.068	0.04	0.00
$H_{17}$	One point of intersection in late time	$We - Exp$	1.346	0.099	0.01	0.00
$H_{18}$		$\Gamma - \Gamma$	1.878	0.132	0.03	0.15
$H_{19}$		$We - Exp$	3.626	0.097	0.09	0.47
$H_{21}$	Two points of intersections in early and middle	$LgN - We$	0.243, 0.655	0.071	0.01	0.04
$H_{22}$		$Exp - LgN$	0.814, 1.038	0.079	0.08	0.48
$H_{23}$		$We - Exp$	0.577, 1.327	0.105	0.09	0.53
$H_{24}$	Two points of intersections in early and late	$LgN - \Gamma$	0.683, 3.074	0.095	0.06	0.13
$H_{25}$		$LgN - Exp$	0.232, 0.831	0.068	0.01	0.19
$H_{26}$		$We - LgN$	1.018, 2.381	0.131	0.06	0.60
$H_{27}$	Two points of intersections in middle and late	$We - LgN$	1.254, 3.265	0.099	0.11	0.09
$H_{28}$		$\Gamma - LgN$	0.793, 2.994	0.088	0.07	0.11
$H_{29}$		$Exp - LgN$	1.321, 3.463	0.104	0.11	0.08

## Conclusions

Thus, in the paper the types of alternative hypotheses have been represented for 0, 1 and 2 points of intersections between underlying survival functions. Every type contains three alternative hypotheses with various distributions of failure time. The results of simulation have been represented that the proposed two-sample *MIN3* test power is close to the corresponding to the maximin Wald test statistic for decision-making under risk and uncertainty. Hence, the *MIN3* test is a stable two-sample test for the uncertainty of a alternative hypothesis. It makes possible to recommend the *MIN3* for applying in two-sample problem solving under right-censored data.

## References

- [1] P. Philonenko, S. N. Postovalov, Test power in two-sample problem testing as the utility function in the theory of decision making under risk and uncertainty // Actual problems of electronic instrument engineering (APEIE-2016) : XIII International Scientific-Technical Conference, Novosibirsk, Russia, October 03-06, 2016 : v.1, p.2. — pp.369–373, 2016 ISBN 978-5-7782-2991-4.
- [2] Bagdonavicius V.B., Nikulin M., On goodness-of-fit tests for homogeneity and proportional hazards, Applied Stochastic Models in Business and Industry, vol. 22, no. 1, pp. 607–619, 2006.
- [3] Bagdonavicius V.B., J. Kruopis, M. Nikulin, Nonparametric tests for censored data, John Wiley & Sons, Inc., New York, 2010, 233 pp.
- [4] M.S. Pepe and T.R. Fleming, Weighted Kaplan-Meier statistics: A class of distance tests for censored survival data, Biometrics 45 (1989), pp. 497–507.

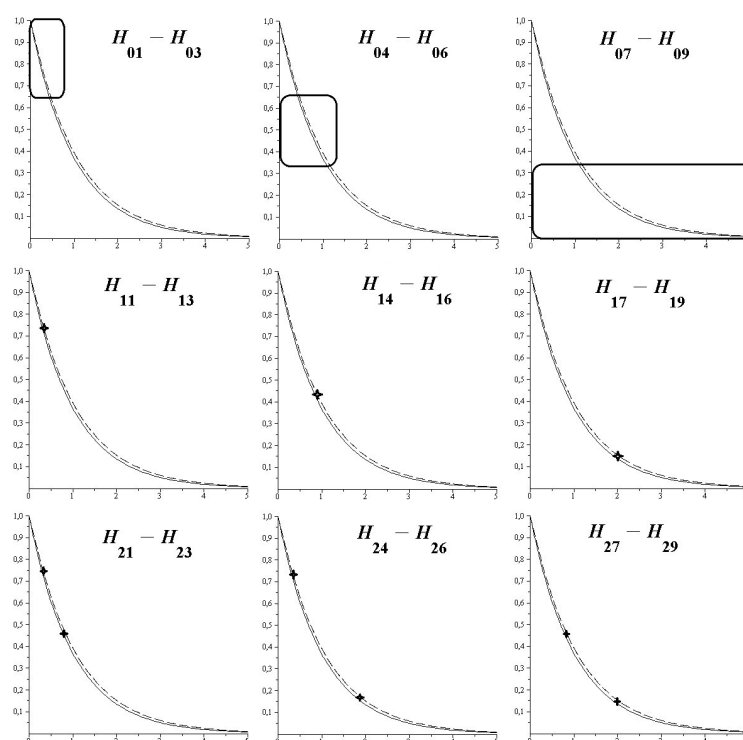


Figure 1: Types of alternative hypotheses (the star means the area of intersection)

- [5] Gehan E. A. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples // Biometrika. V. 52, N 1/2. P. 203–223, (1965).
- [6] Peto R., Peto J. Asymptotically efficient rank invariant test procedures // J. Royal Statist. Soc. Ser. A (General). V.135, N 2. P.185–207, (1972).
- [7] Lee E. T., Wang J. W. Statistical Methods for Survival Data Analysis. N. Y: Wiley, 2003. (Wiley Series in Probability and Statistics).
- [8] Mantel, N., Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration. Cancer Chemotherapy Reports, 50, 163–170, (1966).
- [9] Ruvie Lou Maria C. Martinez A pretest for choosing between logrank and wilcoxon tests in the two-sample problem / Ruvie Lou Maria C. Martinez, Joshua D. Naranjo // International Journal of Statistics, vol.LXVIII, n.2. 111–125 pp, (2010).
- [10] P. Philonenko, S. Postovalov, A new two-sample test for choosing between log-rank and Wilcoxon tests with right-censored data, Journal of Statistical Computation and Simulation, 85:14, 2761–2770, (2015).
- [11] Bagdonavicius V.B., Levulienė R.J., Nikulin M.S., et al., Tests for equality of survival distributions against non-location alternatives, Lifetime Data Analysis, vol. 10, no. 4, pp. 445–460, 2004.
- [12] Wald, A., Statistical decision functions which minimize the maximum risk. The Annals of Mathematics, 46(2), 265–280, (1945).

*Petr Aleksandrovich Philonenko — Postgraduate student, Junior Researcher of Theoretical and Applied Informatics department Novosibirsk State Technical University; e-mail: petr-filonenko@mail.ru;*

*Sergey Nikolaevich Postovalov — Doctor of Science, Associated Professor of Theoretical and Applied Informatics department Novosibirsk State Technical University; e-mail: postovalov@ngs.ru.*

*Received date — May 31, 2017.*

Table 2: The minimal values of tests power on the alternative hypotheses  $H_{01} - H_{03}$ 

Test	Censored rate, %				
	10	20	30	40	50
$G$	0.315	0.345	0.402	0.472	0.531
$P$	0.315	0.318	0.346	0.376	0.427
$LG$	0.142	0.127	0.139	0.160	0.209
$CM$	0.141	0.129	0.137	0.163	0.212
$Q$	0.259	0.262	0.278	0.291	0.303
$MAX$	0.272	0.296	0.346	0.413	0.478
$BN1$	0.190	0.246	0.320	0.401	0.434
$BN2$	0.304	0.336	0.370	0.429	0.490
$BN3$	0.291	0.350	0.421	0.452	0.489
$WKM$	0.323	0.346	0.391	0.468	0.535
$MIN3$	0.389	0.423	0.468	0.528	0.597
$W$	<b>0.389</b>	<b>0.423</b>	<b>0.468</b>	<b>0.528</b>	<b>0.597</b>

Table 3: The minimal values of tests power on the alternative hypotheses  $H_{04} - H_{06}$ 

Test	Censored rate, %				
	10	20	30	40	50
$G$	0.178	0.188	0.190	0.203	0.207
$P$	0.175	0.178	0.181	0.180	0.180
$LG$	0.101	0.109	0.115	0.129	0.132
$CM$	0.101	0.107	0.118	0.124	0.133
$Q$	0.144	0.149	0.147	0.154	0.156
$MAX$	0.157	0.162	0.170	0.177	0.180
$BN1$	0.146	0.148	0.148	0.150	0.148
$BN2$	0.128	0.127	0.127	0.131	0.134
$BN3$	0.158	0.160	0.160	0.157	0.156
$WKM$	0.182	0.188	0.194	0.200	0.206
$MIN3$	0.157	0.161	0.168	0.170	0.180
$W$	<b>0.182</b>	<b>0.188</b>	<b>0.194</b>	<b>0.203</b>	<b>0.207</b>

Table 4: The minimal values of tests power on the alternative hypotheses  $H_{07} - H_{09}$ 

Test	Censored rate, %				
	10	20	30	40	50
$G$	0.089	0.083	0.077	0.071	0.066
$P$	0.093	0.086	0.082	0.075	0.069
$LG$	0.191	0.170	0.125	0.095	0.077
$CM$	0.192	0.170	0.126	0.095	0.077
$Q$	0.178	0.138	0.106	0.086	0.073
$MAX$	0.182	0.142	0.106	0.087	0.075
$BN1$	0.160	0.155	0.145	0.092	0.069
$BN2$	0.132	0.130	0.123	0.112	0.093
$BN3$	0.162	0.155	0.131	0.086	0.066
$WKM$	0.091	0.082	0.075	0.070	0.066
$MIN3$	0.165	0.164	0.127	0.100	0.083
$W$	<b>0.192</b>	<b>0.170</b>	<b>0.145</b>	<b>0.112</b>	<b>0.093</b>

Table 5: The minimal values of tests power on the alternative hypotheses  $H_{11} - H_{13}$ 

Test	Censored rate, %				
	10	20	30	40	50
$G$	0.054	0.051	0.049	0.048	0.052
$P$	0.057	0.052	0.051	0.050	0.049
$LG$	0.072	0.067	0.063	0.058	0.051
$CM$	0.071	0.069	0.066	0.058	0.052
$Q$	0.063	0.061	0.057	0.053	0.051
$MAX$	0.064	0.062	0.060	0.058	0.057
$BN1$	0.070	0.069	0.069	0.071	0.072
$BN2$	0.067	0.068	0.069	0.072	0.079
$BN3$	0.074	0.073	0.074	0.075	0.078
$WKM$	0.054	0.050	0.050	0.049	0.051
$MIN3$	0.068	0.067	0.067	0.070	0.074
$W$	<b>0.074</b>	<b>0.073</b>	<b>0.074</b>	<b>0.075</b>	<b>0.079</b>

Table 6: The minimal values of tests power on the alternative hypotheses  $H_{14} - H_{16}$ 

Test	Censored rate, %				
	10	20	30	40	50
$G$	0.049	0.050	0.050	0.053	0.056
$P$	0.051	0.049	0.050	0.049	0.052
$LG$	0.055	0.051	0.049	0.049	0.050
$CM$	0.055	0.050	0.049	0.050	0.049
$Q$	0.058	0.058	0.058	0.057	0.051
$MAX$	0.064	0.064	0.058	0.055	0.056
$BN1$	0.128	0.117	0.098	0.086	0.076
$BN2$	0.104	0.101	0.089	0.080	0.073
$BN3$	0.139	0.119	0.100	0.086	0.078
$WKM$	0.050	0.050	0.051	0.053	0.057
$MIN3$	0.112	0.097	0.085	0.075	0.071
$W$	<b>0.139</b>	<b>0.119</b>	<b>0.100</b>	<b>0.086</b>	<b>0.078</b>

Table 7: The minimal values of tests power on the alternative hypotheses  $H_{17} - H_{19}$ 

Test	Censored rate, %				
	10	20	30	40	50
$G$	0.163	0.180	0.200	0.219	0.243
$P$	0.155	0.165	0.171	0.177	0.189
$LG$	0.052	0.064	0.083	0.100	0.115
$CM$	0.053	0.066	0.083	0.099	0.114
$Q$	0.146	0.150	0.160	0.168	0.169
$MAX$	0.135	0.153	0.170	0.186	0.205
$BN1$	0.207	0.202	0.201	0.195	0.196
$BN2$	0.171	0.169	0.171	0.174	0.173
$BN3$	0.227	0.223	0.218	0.211	0.209
$WKM$	0.164	0.180	0.199	0.223	0.242
$MIN3$	0.199	0.202	0.207	0.215	0.227
$W$	<b>0.227</b>	<b>0.223</b>	<b>0.218</b>	<b>0.223</b>	<b>0.243</b>

Table 8: The minimal values of tests power on the alternative hypotheses  $H_{21} - H_{23}$ 

Test	Censored rate, %				
	10	20	30	40	50
$G$	0.050	0.049	0.051	0.052	0.051
$P$	0.050	0.049	0.049	0.050	0.051
$LG$	0.058	0.065	0.066	0.056	0.050
$CM$	0.060	0.065	0.067	0.055	0.051
$Q$	0.060	0.062	0.062	0.055	0.052
$MAX$	0.063	0.065	0.063	0.063	0.060
$BN1$	0.066	0.063	0.062	0.064	0.064
$BN2$	0.087	0.083	0.085	0.076	0.070
$BN3$	0.060	0.061	0.062	0.064	0.065
$WKM$	0.050	0.050	0.052	0.053	0.051
$MIN3$	0.093	0.081	0.074	0.068	0.063
$W$	<b>0.093</b>	<b>0.083</b>	<b>0.085</b>	<b>0.076</b>	<b>0.070</b>

Table 9: The minimal values of tests power on the alternative hypotheses  $H_{24} - H_{26}$ 

Test	Censored rate, %				
	10	20	30	40	50
$G$	0.050	0.051	0.051	0.052	0.054
$P$	0.050	0.051	0.050	0.050	0.050
$LG$	0.050	0.050	0.050	0.051	0.049
$CM$	0.049	0.050	0.050	0.051	0.050
$Q$	0.049	0.049	0.048	0.051	0.052
$MAX$	0.050	0.052	0.054	0.058	0.059
$BN1$	0.054	0.050	0.051	0.049	0.053
$BN2$	0.124	0.115	0.109	0.100	0.097
$BN3$	0.049	0.049	0.052	0.051	0.063
$WKM$	0.050	0.049	0.048	0.051	0.053
$MIN3$	0.103	0.094	0.088	0.081	0.077
$W$	<b>0.124</b>	<b>0.115</b>	<b>0.109</b>	<b>0.100</b>	<b>0.097</b>

Table 10: The minimal values of tests power on the alternative hypotheses  $H_{27} - H_{29}$ 

Test	Censored rate, %				
	10	20	30	40	50
$G$	0.059	0.064	0.075	0.094	0.131
$P$	0.060	0.060	0.062	0.070	0.083
$LG$	0.050	0.050	0.050	0.050	0.051
$CM$	0.050	0.050	0.050	0.050	0.052
$Q$	0.058	0.063	0.065	0.072	0.070
$MAX$	0.056	0.063	0.073	0.089	0.117
$BN1$	0.050	0.054	0.070	0.100	0.147
$BN2$	0.221	0.219	0.221	0.237	0.275
$BN3$	0.059	0.075	0.103	0.136	0.180
$WKM$	0.059	0.065	0.078	0.095	0.131
$MIN3$	0.178	0.184	0.188	0.215	0.256
$W$	<b>0.221</b>	<b>0.219</b>	<b>0.221</b>	<b>0.237</b>	<b>0.275</b>