Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 725–730
DOI: 10.15439/2024F223
ISSN 2300-5963 ACSIS, Vol. 39

Non-parametric comparison of survival functions with censored data: A computational analysis of exhaustive and Monte Carlo approaches

Lubomír Štěpánek[1, 2, 3], Filip Habarta[1], Ivana Malá[1], Luboš Marek[1]
[1]Department of Statistics and Probability
[2]Department of Mathematics
Faculty of Informatics and Statistics
Prague University of Economics and Business
W. Churchill's square 4, 130 67 Prague, Czech Republic
{lubomir.stepanek, filip.habarta, malai, marek}@vse.cz
&
[3]Institute of Biophysics and Informatics
First Faculty of Medicine
Charles University
Salmovská 1, Prague, Czech Republic
lubomir.stepanek@lf1.cuni.cz

***Abstract*—Comparison of two survival functions, which describe the probability of not experiencing an event of interest by a given time point in two different groups, is a typical task in survival analysis. There are several well-established methods for comparing survival functions, such as the log-rank test and its variants. However, these methods often come with rigid statistical assumptions. In this work, we introduce a non-parametric alternative for comparing survival functions that is nearly free of assumptions. Unlike the log-rank test, which requires the estimation of hazard functions derived from (or facilitating the derivation of) survival functions and assumes a minimum number of observations to ensure asymptotic properties, our method models all possible scenarios based on observed data. These scenarios include those in which the compared survival functions differ in the same way or even more significantly, thus allowing us to calculate the p-value directly. Individuals in these groups may experience an event of interest at specific time points or may be censored, i.e., they might experience the event outside the observed time points. Focusing on all scenarios where survival probabilities differ at least as much as observed usually requires computationally intensive calculations. Censoring is treated as a form of noise, increasing the range of scenarios that need to be calculated and evaluated. Therefore, to estimate the p-value, we compare a computationally exhaustive approach that computes all possible scenarios in which groups' survival functions differ as observed or more, with a Monte Carlo simulation of these scenarios, alongside a traditional approach based on the log-rank test. Our proposed method reduces the first type error rate, enhancing its utility in studies where robustness against false positives is critical. We also analyze the asymptotic time complexity of both proposed approaches.**

I. Introduction

SURVIVAL analysis encompasses the study of various time-to-event data, not limited to historical focuses such as death or disease occurrence what could be primarily indicated by its name. This statistical field characterizes when an event of interest happens or, alternatively, when it does not occur which leads to censoring. The dual nature of the data – combining both the timing of events and their occurrence (or non-occurrence) – distinguishes survival analysis from other statistical methods. Typically, we describe time to an event of interest occurrence for a given individual or a group of individuals using a survival function, which is a function that returns a probability of non-experiencing the event of interest before a given time point.

When comparing two distinct time-to-event survival curves from separate groups, the log-rank test is commonly employed to evaluate the differences, as suggeted by Mantel (1966) [1]. The use of the log-rank test is particularly challenged when comparing non-crossing time-event survival curves with uneven censoring between the groups. To enhance the test's efficiency and its robustness against assumptions, various modifications have been proposed. Kong (1997) optimized the log-rank test by adjusting the hazard functions [2], while Song et al. (2007) explored covariate matrix decomposition to establish minimal sample sizes that validate the use of the log-rank test [3]. Additionally, weighted observations have been utilized to correct the test's accuracy, with weights typically greater for earlier events which have more numerous observations, as suggested Peto and Peto (1972) [4], Yang and Prentice (2009) [5], and Li (2018) [6].

Our proposed non-parametric method challenges these limitations by avoiding the estimation of hazard functions and the use of weighting schemes. It explores every conceivable scenario, including those affected by censoring, thus offering a comprehensive approach that traditional methods often can-

not handle due to their computational demands and reliance on strict assumptions. By setting potential event time points for individuals observed as censored, as Štěpánek initially suggested in [7] and [8], we remove data noise and improve the estimation of the survival function for each time point, making our method more adaptable to real-world data.

This approach rigorously evaluates whether the observed differences in survival functions are statistically significant, thereby reducing false positive rates and enhancing methodological robustness. We have developed and applied two computational strategies to estimate p-values: a detailed, computationally intensive approach, and a Monte Carlo simulation, both designed to handle the complexity of evaluating multiple scenarios. For both approaches, we analyzed their asymptotic time complexities.

The manuscript firstly revisits the principles of the log-rank test and its limitations, then introduces our method which involves intensive computational modeling of all potential event scenarios, including censored cases. We provide a relatively detailed analysis of the proposed method's asymptotic time complexity, particularly its p-value calculation using the computationally exhaustive approach and Monte Carlo estimation. The effectiveness of our method is demonstrated through simulation studies comparing p-values calculated using both the log-rank test and our proposed approaches, focusing on their ability to maintain low first type error rates.

II. Traditional methods for comparison of survival functions

Firstly, we revisit the log-rank test, examining its principles, assumptions, and limitations.

A. Foundations, assumptions, and limitations of the log-rank test

Foundations of the log-rank test. Assume there are k distinct time points where an event of interest could occur, denoted as t_j for $j \in \{1, 2, 3, \ldots, k\}$, and arranged in an ordered tuple $(t_1, t_2, \ldots, t_k)^T$. Consider two groups of subjects, labeled as group 1 and 2, $g \in \{1, 2\}$. At each time point t_j, there are $r_{1,j}$ and $r_{2,j}$ individuals at risk (who have not yet experienced the event or have been censored) in groups 1 and 2, respectively, and $d_{1,j}$ and $d_{2,j}$ individuals in each group who have experienced the event. This setup leads to the construction of the contingency table in Table I.

TABLE I
NUMBERS OF INDIVIDUALS EXPERIENCING THE EVENT OF INTEREST IN BOTH GROUPS (1 AND 2) AT TIME POINT t_j.

	event of interest at time t_j		
group	yes	no	total
1	$d_{1,j}$	$r_{1,j} - d_{1,j}$	$r_{1,j}$
2	$d_{2,j}$	$r_{2,j} - d_{2,j}$	$r_{2,j}$
total	d_j	$r_j - d_j$	r_j

The log-rank test evaluates the null hypothesis H_0 that both groups experience identical event rates over time, conditional on fixed rates in the past being the same. Under H_0, the observed numbers of events, $D_{1,j}$ and $D_{2,j}$, are modeled as random variables following a hypergeometric distribution with parameters $(r_j, r_{g,j}, d_j)$ for both $g \in \{1, 2\}$. The expected value of $D_{g,j}$ is $\mathbb{E}(D_{g,j}) = r_{g,j}\frac{d_j}{r_j}$, and the variance is $\mathrm{var}(D_{g,j}) = \frac{r_{1,j}r_{2,j}d_j}{r_j^2}\left(\frac{r_j - d_j}{r_j - 1}\right)$. We then compare the observed numbers of events, $d_{g,j}$, for all j, to their expected values. The test statistic for both groups is computed as follows,

$$\chi^2_{\text{log-rank}} = \frac{\left(\sum_{j=1}^{k}(d_{g,j} - \mathbb{E}(D_{g,j}))\right)^2}{\sum_{j=1}^{k}\mathrm{var}(D_{g,j})} \sim \chi^2(1),$$

where $g \in \{1, 2\}$. Under H_0, the test statistic follows a χ^2 distribution with one degree of freedom. For sufficiently large r_j (at least 30), $\sqrt{\chi^2_{\text{log-rank}}}$ approximates a standard normal distribution. Since $\chi^2_{\text{log-rank}} \sim \chi^2(1)$, the test statistic can be uniquely transformed into a p-value, representing the conditional probability of observing a test statistic as extreme as or more extreme than the one observed, assuming H_0 is true.

Assumptions and limitations of the log-rank test. It is crucial that right censoring does not differentially affect the event occurrences in both groups. The proportions of censored observations should be nearly equal in both groups; otherwise, the test statistic $\chi^2_{\text{log-rank}}$ could be biased for either group.

Moreover, both the initial total number of individuals at risk and the initial number not experiencing the event should be large enough to meet the Cochrane criteria for minimal sample size for χ^2 tests. If these conditions are not met, the $\chi^2_{\text{log-rank}}$ statistic may not fulfill its asymptotic properties, making the estimate numerically unstable.

Additionally, the robustness and statistical power of the log-rank test can be compromised if the proportions $\frac{r_{1,j}}{r_j}$ and $\frac{r_{2,j}}{r_j}$ are not constant across all time points. Significant changes in the survival curves' trends, mutual distances, or crossings can decrease the test's power, making it less likely to reject H_0 when the survival curves actually differ.

III. The proposed method for comparison of survival functions

This section explores our non-parametric approach, which examines all conceivable scenarios where an event could occur at different feasible times for each individual, including those unobserved due to censoring. By adopting all time points for events that come from observed data and calculating a number of all possible scenarios how the the events could be registered by individuals in time, we estimate the survival function based on the proportion of individuals who have not experienced the event. We then assess whether the survival functions of the compared groups statistically differ by evaluating the sum of group-based mutual differences across all time points. Theoretically, the proportion of scenarios where the differences are as large or larger than observed corresponds to the p-value, indicating the probability of these findings under the null hypothesis that the survival functions are equivalent. Through this approach, we calculate a range of p-values and

determine whether to reject the null hypothesis based on a predefined confidence level, thus assessing if the survival functions statistically differ.

A. Foundations of the proposed method for comparison of survival functions

We assume two groups of individuals we want to compare so that for each of them we know the same amount of information, coming from data on input. We assume $k \in \mathbb{N}$ distinct time points, denoted $t_1, t_2, \ldots, t_k$. For each individual of both groups, we know two pieces of information $(\Upsilon_i, \mathcal{E}_i)$, where $\Upsilon_i \in \{t_1, t_2, \ldots, t_k\}$ and $\mathcal{E}_i \in \{\text{event occured}, \text{event did not occur}\}$. While Υ_i indicates in which time point an event of interest happened to i-th individual, $\mathcal{E}_i$ describes if it was in fact the event of interest, or rather censoring. Having such information for each individual in a group of $n \in \mathbb{N}$ individuals at all, we can transform pairs $(\Upsilon_i, \mathcal{E}_i)$ for $\forall i \in \{1, 2, \ldots, n\}$ into grid as in Fig. 1. In lines of the grid in Fig. 1, there are individuals ordered from the one who experienced the event of interest or censoring as first. Thus, if $\boldsymbol{\Upsilon} = (\Upsilon_1, \Upsilon_2, \ldots, \Upsilon_n)^T = (\tau_1, \tau_2, \ldots, \tau_n)^T$ where $\forall \tau_i \in \{t_1, t_2, \ldots, t_k\}$, it is $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_n$. While the black dots stand for time points where individuals have not register the event yet including the last point in a row, when the event is experienced, gray dots indicate that an individual was censored at some time point, thus, in theory, the gray dots could be in changed in black if we would know when the event happened (when individual is not censored). White squares stand for time points where an individual experienced neither the event nor censoring.

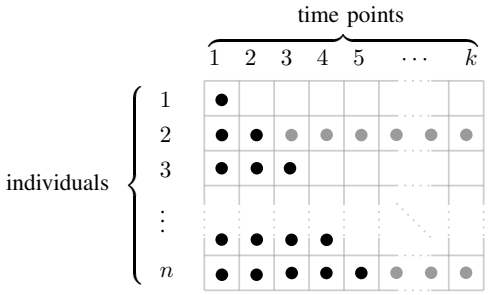

Fig. 1. The grid of k time points displays n individuals ordered by the timing of their event of interest or censoring. Black dots represent time points where the event has not yet occurred up to the last black point in a row standing for the event, while gray dots indicate censoring, suggesting that these could become black if the event timing were known. White squares denote time points where neither the event nor censoring has occurred.

We want to test the following null hypothesis H_0 against the alternative H_1 as follows,

H_0 : survival functions do not differ between the groups,
H_1 : survival functions differ between the groups.

For purposes of statistical inference, we need to calculate p-value as a probability of observing data the same way or even more contrary the hypothesis H_0 that the survival functions do not differ between the groups. Addressing the p-value calculation in a non-parametric fashion, we have to calculate a number of all scenarios that favor the p-value's definition meaning. The number of scenarios in contradiction to H_0 depends on when censoring happen to censored individuals.

Let $\mathcal{C} \subseteq \{1, 2, \ldots, n\}$ be a subset of individuals' indices that have been censored. Considering the censoring arrangement, for each such arrangement of values $\tau_i \in \{t_1, t_2, \ldots, t_k\}$ where $i \in \mathcal{C}$, we can calculate a unique p-value, since the entry grids as in Fig. 1 differ as τ_i vary. To do this, we need updated times of events with respect to censoring for all individuals, denoted as $\boldsymbol{\Upsilon}' = (\Upsilon'_1, \Upsilon'_2, \ldots, \Upsilon'_n)^T = (\tau'_1, \tau'_2, \ldots, \tau'_n)^T$ where

$$\begin{cases} \tau'_i \geq \tau_i, & \forall i \in \mathcal{C}, \\ \tau'_i = \tau_i, & \forall i \in \{1, 2, \ldots, n\} \setminus \mathcal{C}. \end{cases} \tag{1}$$

As an illustration, we can compare Fig. 2 and Fig. 3. In both figures, there is $n = 6$ and time points $t \in \{1, 2, \ldots, 8\}$. Using previous notations, obviously it is $(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6)^T = (1, 1, 2, 2, 3, 4)^T$ and $\mathcal{C} = (2, 4)^T$ in Fig. 2. In Fig. 3, we set time points of possible event registrations for censored individuals as follows, $\tau_2 = 2$ and $\tau_4 = 6$, thus, $(\tau'_1, \tau'_2, \tau'_3, \tau'_4, \tau'_5, \tau'_6)^T = (1, 2, 2, 6, 3, 4)^T$.

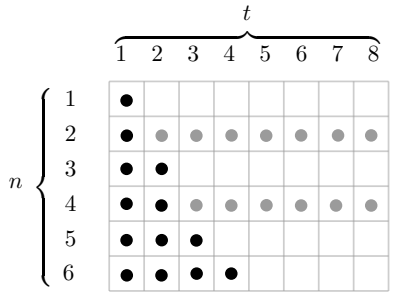

Fig. 2. A grid for an initial dataset with $n = 6$ individuals across $t \in \{1, 2, \ldots, 8\}$ time points, where $(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6)^T = (1, 1, 2, 2, 3, 4)^T$ and censored times $\mathcal{C} = (2, 4)^T$.

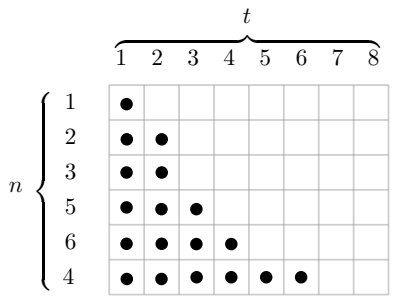

Fig. 3. A grid for the initial dataset with $n = 6$ individuals across $t \in \{1, 2, \ldots, 8\}$ time points from Fig. 2, where the time points for censored events are adjusted, setting $\tau_2 = 2$ and $\tau_4 = 6$ and getting $(\tau'_1, \tau'_2, \tau'_3, \tau'_4, \tau'_5, \tau'_6)^T = (1, 2, 2, 6, 3, 4)^T$, to hypothesize the potential event registrations if not censored.

Once the censoring is arranged and we set Υ'_i for $\forall i \in \{1, 2, \ldots, n\}$, we can enumerate a survival function for the

given group of n individuals (and the given censoring arrangement) using an expected survival $\mathbb{E}(S')$ for the group, i.e., a "surface" below a curve enveloping black dots in grid as we defined before and plot in Fig. 1. To be more specific,

$$\hat{\mathbb{E}}(S') = \frac{1}{n}\sum_{i=1}^{n}\tau'_i, \tag{2}$$

using the same notation as so far. Let's assume two groups with indices 1 and 2 and observed expected survivals, following formula (2), as $\hat{\mathbb{E}}(S'_1)$ and $\hat{\mathbb{E}}(S'_2)$. Now, we can finally calculate p-value, conditional of the given censoring and enabling any further statistical inference, as follows,

$$\begin{aligned}
&p\text{-value}_{\text{censoring}} = \\
&= \{p\text{-value} \mid \text{censoring}\} = \\
&= P\left(\begin{array}{l}\text{getting data at least as extreme} \\ \text{as the observed}\end{array} \,\middle|\, H_0, \text{censoring}\right) = \\
&= P\left(|\mathbb{E}(S_1) - \mathbb{E}(S_2)| \geq |\hat{\mathbb{E}}(S_1) - \hat{\mathbb{E}}(S_2)| \,\middle|\, \text{censoring}\right) = \\
&= P\left(|\mathbb{E}(S'_1) - \mathbb{E}(S'_2)| \geq |\hat{\mathbb{E}}(S'_1) - \hat{\mathbb{E}}(S'_2)|\right),
\end{aligned} \tag{3}$$

thus, in other words, we calculate p-value for a given censoring, denoted as p-value$_{\text{censoring}}$, as a probability of observing a scenario with an absolute difference between expected survivals as least as extreme as the observed absolute difference of survivals. Since censoring arrangements may differ, we get various p-values from formula (3) stored in vector $\boldsymbol{p}$-value$_{\text{censoring}}$, that may create an interval of p-values as follows,

$$\begin{aligned}
p\text{-value} \in \big\langle\, &\min\{\boldsymbol{p}\text{-value}_{\text{censoring}}\}, \\
&\max\{\boldsymbol{p}\text{-value}_{\text{censoring}}\}\big\rangle.
\end{aligned} \tag{4}$$

Assuming a confidence level $\alpha \in (0,1)$ as an acceptable probability of first type error rate, i.e., false rejection of null hypothesis H_0 which is in fact true, there are two cases crucial for statistical inference, either

$$\max\{\boldsymbol{p}\text{-value}_{\text{censoring}}\} \leq \alpha,$$

resulting in a strong null hypothesis H_0 rejection, and

$$\min\{\boldsymbol{p}\text{-value}_{\text{censoring}}\} \leq \alpha,$$

leading to a weak null hypothesis H_0 rejection.

B. Approaches to p-value calculation for the proposed method of survival functions' comparison

In this section, we introduce a computationally exhaustive approach and Monte Carlo simulation-based approach on calculation of p-value following formula (3), and discuss their asymptotic time complexity.

Computationally exhaustive approach for p-value calculation. Within the computationally intensive approach, we work out formula (3) for p-value calculation. Firstly, we realize that the censoring arrangement for an entry grid of n individuals and k time points is fully defined by $\boldsymbol{\Upsilon}' = (\Upsilon'_1, \Upsilon'_2, \ldots, \Upsilon'_n)^T = (\tau'_1, \tau'_2, \ldots, \tau'_n)^T$ as comes from formula (1). Assuming the total numbers of all scenarios for both groups, including all possible censoring arrangements $\boldsymbol{\Upsilon}'_1$ and $\boldsymbol{\Upsilon}'_2$, are on grids $n_1 \times k$ and $n_2 \times k$ finite, respectively, we can work out formula (3) as follows,

$$\begin{aligned}
&p\text{-value}_{\text{censoring}} = \\
&= \{p\text{-value} \mid \text{censoring}\} = \\
&= P\left(\begin{array}{l}\text{getting data at least as extreme} \\ \text{as the observed}\end{array} \,\middle|\, H_0, \text{censoring}\right) = \\
&= P\left(\begin{array}{l}\text{getting data at least as extreme} \\ \text{as the observed}\end{array} \,\middle|\, H_0, \boldsymbol{\Upsilon}'_1, \boldsymbol{\Upsilon}'_2\right) = \\
&= P\left(|\mathbb{E}(S_1) - \mathbb{E}(S_2)| \geq |\hat{\mathbb{E}}(S_1) - \hat{\mathbb{E}}(S_2)| \,\middle|\, \boldsymbol{\Upsilon}'_1, \boldsymbol{\Upsilon}'_2\right) = \\
&= P\left(|\mathbb{E}(S'_1) - \mathbb{E}(S'_2)| \geq |\hat{\mathbb{E}}(S'_1) - \hat{\mathbb{E}}(S'_2)|\right) = \\
&= \frac{1}{|\{\boldsymbol{S}_1\}| \cdot |\{\boldsymbol{S}_2\}|} \cdot \\
&\cdot \sum_{\forall s \in \{\boldsymbol{S}_1\}} \sum_{\forall \sigma \in \{\boldsymbol{S}_2\}} \mathbb{1}_{\{|\mathbb{E}(S'_1) - \mathbb{E}(S'_2)| \geq |\hat{\mathbb{E}}(S'_1) - \hat{\mathbb{E}}(S'_2)|\}},
\end{aligned} \tag{5}$$

where $\{\boldsymbol{S}_1\}$ and $\{\boldsymbol{S}_2\}$ are sets of all possible scenarios for first and second group, respectively, and $\mathbb{1}_{\{\beta\}}$ is an identifier function, so

$$\mathbb{1}_{\{\beta\}} = \begin{cases} 1, & \text{if } \beta \text{ is true,} \\ 0, & \text{if } \beta \text{ is false.} \end{cases} \tag{6}$$

The numbers of all scenarios, $|\{\boldsymbol{S}_1\}|$ and $|\{\boldsymbol{S}_2\}|$ are straightforward and can be assessed using stars-and-bars theorem. Assuming grids containing n_1 and n_2 individuals, respectively, and k time points as in Fig. 1, and a fact that for increasing index i, time $\tau_i \in \{0, 1, \ldots, k\}$ of event or censoring forms a non-decreasing sequence as in Fig. 3, it is

$$|\{\boldsymbol{S}_g\}| = \binom{n_g + k}{k} = \binom{n_g + k}{n_g} = \frac{(n_g + k)!}{n_g!\; k!}, \tag{7}$$

where $g \in \{1, 2\}$. Formula (7) enables us to investigate asymptotic time complexity of p-value from formula (5). As we can see in Algorithm 1, total counts of scenarios $|\{\boldsymbol{S}_1\}|$ and $|\{\boldsymbol{S}_2\}|$ (in the fraction part of formula (5)) are calculated asynchronously (line 1 in Algorithm 1), so an asymptotic time complexity of their calculation with respect to formula (7), using Bachmann–Landau logic [9] and unit time steps for basic arithmetic operations, is

$$\Theta(n_g{+}k{-}1{+}n_g{-}1{+}k{-}1) \approx \Theta(2n_g{+}2k) \approx \Theta(n_g{+}k), \tag{8}$$

thus, in total, the fraction part of formula (5) has got asymptotic time complexity $\Theta(\dagger)$ where

$$\Theta(\dagger) \overset{(8)}{\approx} \Theta(n_1 + k) + \Theta(n_2 + k) \approx \Theta(n_1 + n_2 + 2k). \tag{9}$$

On the other hand, the count of scenarios with absolute difference of expected survivals greater than or equal to the

observed difference is examined in the summation part of formula (5) exhaustively step by step, considering the total number of combinations of scenarios equal to $|\{\boldsymbol{S}_1\}| \cdot |\{\boldsymbol{S}_2\}|$ (lines 2–9 in Algorithm 1). Within each step, condition $|\mathbb{E}(S_1') - \mathbb{E}(S_2')| \geq |\hat{\mathbb{E}}(S_1') - \hat{\mathbb{E}}(S_2')|$ is checked (lines 5–7 in Algorithm 1) – while the part $|\hat{\mathbb{E}}(S_1') - \hat{\mathbb{E}}(S_2')|$ is once pre-calculated (so we can ignore its complexity within the loop of steps), the difference $|\mathbb{E}(S_1') - \mathbb{E}(S_2')|$ takes $n_1 + n_2$ unit times per each check, i.e., per each step, as comes from formula (2). Thus, asymptotic time complexity $\Theta(\ddagger)$ of the summation part of formula (5) is

$$\begin{aligned}
\Theta(\ddagger) &\overset{(7)}{\approx} \Theta\left\{\binom{n_1+k}{k}\binom{n_2+k}{k}(n_1+n_2)\right\} \leq \\
&\leq \Theta\left\{\binom{n_1+k}{\frac{n_1+k}{2}}\binom{n_2+k}{\frac{n_2+k}{2}}(n_1+n_2)\right\} \approx \\
&\approx \Theta\left\{\frac{(n_1+k)!}{\frac{n_1+k}{2}!\ \frac{n_1+k}{2}!}\frac{(n_2+k)!}{\frac{n_2+k}{2}!\ \frac{n_2+k}{2}!}(n_1+n_2)\right\} \approx \\
&\approx \Theta\left\{\frac{(n_1+k)!\ (n_2+k)!}{\left(\frac{n_1+k}{2}!\right)^2\left(\frac{n_2+k}{2}!\right)^2}(n_1+n_2)\right\} \approx \\
&\approx \Theta\left\{\frac{\left(\frac{n_1+k}{2}\right)^{n_1+k}\left(\frac{n_2+k}{2}\right)^{n_2+k}}{\left(\left(\frac{n_1+k}{4}\right)^{\frac{n_1+k}{2}}\right)^2\left(\left(\frac{n_2+k}{4}\right)^{\frac{n_2+k}{2}}\right)^2}(n_1+n_2)\right\} \approx \\
&\approx \Theta\left\{\frac{\left(\frac{n_1+k}{2}\right)^{n_1+k}\left(\frac{n_2+k}{2}\right)^{n_2+k}}{\left(\frac{n_1+k}{4}\right)^{n_1+k}\left(\frac{n_2+k}{4}\right)^{n_2+k}}(n_1+n_2)\right\} \approx \\
&\approx \Theta\left\{2^{n_1+k}\cdot 2^{n_2+k}\cdot(n_1+n_2)\right\} \approx \\
&\approx \Theta\left\{2^{n_1+n_2+2k}\cdot(n_1+n_2)\right\}. \qquad (10)
\end{aligned}$$

Algorithm 1: Calculation of p-value$_{\text{censoring}}$ using formula (5)

Result: Calculation of p-value$_{\text{censoring}}$ based on scenario sets $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$

1 calculate $|\{\boldsymbol{S}_1\}|$ and $|\{\boldsymbol{S}_2\}|$ using formula (7);
2 initialize count $c = 0$;
3 **forall** $s \in \{\boldsymbol{S}_1\}$ **do**
4 **forall** $\sigma \in \{\boldsymbol{S}_2\}$ **do**
5 **if** $|\mathbb{E}(S_1') - \mathbb{E}(S_2')| \geq |\hat{\mathbb{E}}(S_1') - \hat{\mathbb{E}}(S_2')|$ **then**
6 $c = c + 1$;
7 **end**
8 **end**
9 **end**
10 p-value$_{\text{censoring}} = \frac{c}{|\{\boldsymbol{S}_1\}|\cdot|\{\boldsymbol{S}_2\}|}$;
11 **return** *p-value$_{censoring}$*

Putting things together, calculation p-value$_{\text{censoring}}$ from formula (5) takes the following time $\Theta(\bullet)$,

$$\begin{aligned}
\Theta(\bullet) &\approx \Theta(\dagger) + \Theta(\ddagger) \overset{(9,10)}{\approx} \\
&\overset{(9,10)}{\approx} \Theta\left\{n_1+n_2+2k+\binom{n_1+k}{k}\binom{n_2+k}{k}(n_1+n_2)\right\}, \qquad (11)
\end{aligned}$$

and also

$$\begin{aligned}
\Theta(\bullet) &\approx \Theta(\dagger) + \Theta(\ddagger) \overset{(9,10)}{\leq} \\
&\overset{(9,10)}{\leq} \Theta\left\{n_1+n_2+2k+2^{n_1+n_2+2k}\cdot(n_1+n_2)\right\}. \qquad (12)
\end{aligned}$$

Considering formula (1) a $|\mathcal{C}|$ individuals that are censored, in theory, in any time point $\{t_1, t_2, \ldots, t_k\}$, the calculation of p-value$_{\text{censoring}}$ using formula (5) might be repeated at maximum $k^{|\mathcal{C}|}$, thus the asymptotic time complexity could be at maximally

$$\Theta(\bullet) \overset{(12)}{\leq} k^{|\mathcal{C}|}\cdot\Theta\left\{n_1+n_2+2k+2^{n_1+n_2+2k}\cdot(n_1+n_2)\right\}. \qquad (13)$$

Monte Carlo approach for p-value calculation. Using formula (5), we cannot necessary consider every possible scenario, but can randomly select a subset $\mathcal{M}$ of a joint scenarios' set $\{\boldsymbol{S}_1 \cup \boldsymbol{S}_2\}$, so $\mathcal{M} \subseteq \{\boldsymbol{S}_1 \cup \boldsymbol{S}_2\}$. Then, assuming the same conditions, p-value considering a given censoring setting can be estimated similarly as in formula (5),

$$\hat{p}\text{-value}_{\text{censoring}} = \frac{1}{|\mathcal{M}|}\cdot\sum_{\forall m\in\mathcal{M}}\mathbb{1}_{\left\{|\mathbb{E}(S_1')-\mathbb{E}(S_2')|\geq|\hat{\mathbb{E}}(S_1')-\hat{\mathbb{E}}(S_2')|\right\}}, \qquad (14)$$

using the same mathematical notation as above. Investigating Algorithm 2, asymptotic time complexity of Monte Carlo approach could be straightforwardly estimated. The condition $|\mathbb{E}(S_1') - \mathbb{E}(S_2')| \geq |\hat{\mathbb{E}}(S_1') - \hat{\mathbb{E}}(S_2')|$, which takes $n_1 + n_2$ time units, is checked $\mathcal{M}$ times, thus,

$$\Theta(\bullet) \approx |\mathcal{M}|\cdot(n_1+n_2), \qquad (15)$$

and considering the censoring, it is

$$\Theta(\bullet) \leq k^{|\mathcal{C}|}\cdot|\mathcal{M}|\cdot(n_1+n_2). \qquad (16)$$

IV. Simulation Study

In this simulation study, we compared the traditional log-rank test with our proposed method, using both exhaustive and Monte Carlo approaches for p-value calculation. The methods were tested on numerous pairs of survival curves assumed to be equivalent to assess the first type error rate, i.e., the frequency of falsely rejecting the null hypothesis. Since the proposed method is non-parametric and robust, we focused on evaluating its first type error rate rather than its power. All computations were performed in R statistical language [10].

Algorithm 2: Calculation of $\hat{p}$-value$_{\text{censoring}}$ using formula (14)

Result: Estimate of $\hat{p}$-value$_{\text{censoring}}$ using a Monte Carlo approach based on a subset $\mathcal{M}$ of scenarios

$\mathcal{M} \leftarrow$ randomly select a subset from $\{\boldsymbol{S}_1 \cup \boldsymbol{S}_2\}$;
initialize count $c = 0$;
forall $m \in \mathcal{M}$ **do**
  if $|\mathbb{E}(S_1') - \mathbb{E}(S_2')| \geq |\hat{\mathbb{E}}(S_1') - \hat{\mathbb{E}}(S_2')|$ **then**
    $c = c + 1$;
  end
end
$\hat{p}$-value$_{\text{censoring}} = \frac{c}{|\mathcal{M}|}$;
return $\hat{p}$-*value*$_{censoring}$

We generated pairs of groups of varying size of $n \in \{10, 11, \ldots, 100\}$ individuals so that their survival functions follow a negatively exponential survival function,

$$S(t) = P(T \leq t) = e^{-\lambda t}, \tag{17}$$

where λ varied in range of $\langle 0.04, 0.06 \rangle$. Sizes of the groups randomly differed between the repetitions of the simulation. A total of $m = 1000$ pairs of groups with survival functions were generated, and each pair was analyzed under different levels of censoring, set at 10 %, 20 %, 30 %, and 40 %. The occurrence of p-values as intervals, which comes from the present censoring and formula (4), either containing or does not containing the significance level $\alpha = 0.05$ was recorded, summarizing the first type error rates as frequencies when the p-value interval's maximum is below $\alpha = 0.05$, see Table II.

TABLE II
FIRST TYPE ERROR RATES FOR THE LOG-RANK TEST AND PROPOSED METHODS AT VARYING LEVELS OF CENSORING.

		proposed method	
censoring level	log-rank test	exhaustive approach	Monte Carlo
10 %	0.055	0.041	0.040
20 %	0.053	0.039	0.042
30 %	0.052	0.043	0.041
40 %	0.050	0.038	0.039

Simulations show that the first type error rate is consistently lower for the proposed method, whether using exhaustive or Monte Carlo p-value calculations. This reduction persists across all levels of censoring, highlighting the method's robustness compared to the log-rank test, especially at higher censoring levels.

V. CONCLUSION

This study introduces a novel, assumption-minimal, non-parametric method for comparing survival functions. Utilizing computationally exhaustive and Monte Carlo simulations for p-value calculation, the method consistently shows lower first type error rates than the log-rank test across various levels of censoring (10 % to 40 %). While the approach involves high asymptotic time complexity during p-value estimation, especially with exhaustive calculations, its reduced first type error rate offers an alternative for survival data analysis, potentially suitable for integration into statistical software.

VI. ACKNOWLEDGEMENT

This research is supported by grant F4/50/2023 from the Internal Grant Agency of the Prague University of Economics and Business.

REFERENCES

[1] N. Mantel, "Evaluation of survival data and two new rank order statistics arising in its consideration," *Cancer Chemotherapy Reports*, vol. 50, no. 3, pp. 163–170, 1966.
[2] F. Kong, "Robust covariate-adjusted logrank tests," *Biometrika*, vol. 84, no. 4, pp. 847–862, Dec. 1997.
[3] R. Song, M. R. Kosorok, and J. Cai, "Robust covariate-adjusted log-rank statistics and corresponding sample size formula for recurrent events data," *Biometrics*, vol. 64, no. 3, pp. 741–750, Dec. 2007.
[4] R. Peto and J. Peto, "Asymptotically efficient rank invariant test procedures," *Journal of the Royal Statistical Society. Series A (General)*, vol. 135, no. 2, p. 185, 1972.
[5] S. Yang and R. Prentice, "Improved logrank-type tests for survival data using adaptive weights," *Biometrics*, vol. 66, no. 1, pp. 30–38, Apr. 2009.
[6] C. Li, "Doubly robust weighted log-rank tests and renyi-type tests under non-random treatment assignment and dependent censoring," *Statistical Methods in Medical Research*, vol. 28, no. 9, pp. 2649–2664, Jul. 2018.
[7] L. Štěpánek, F. Habarta, I. Malá, and L. Marek, "Analysis of asymptotic time complexity of an assumption-free alternative to the log-rank test," in *Proceedings of the 2020 Federated Conference on Computer Science and Information Systems*, ser. FedCSIS 2020. IEEE, Sep. 2020, p. 453–460.
[8] ——, *Reducing the First-Type Error Rate of the Log-Rank Test: Asymptotic Time Complexity Analysis of An Optimized Test's Alternative*. Springer International Publishing, Dec. 2021, p. 281–302.
[9] D. E. Knuth, "Big omicron and big omega and big theta," *ACM Sigact News*, vol. 8, no. 2, pp. 18–24, 1976.
[10] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2023. [Online]. Available: https://www.R-project.org/