A Non-parametric Approach to Evaluating the Point of Treatment Time-Lag Effect

Kristine Gierz[1], Kayoung Park [1], Peihua Qiu[2]

Abstract

In general, the change point problem considers inference of a change in distribution for a set of time-ordered observations. This has applications in a large variety of fields, and can also apply to survival data. In survival analysis, most existing methods compare two treatment groups for the entirety of the study period. Some treatments may take a length of time to show effects in subjects. This has been called the time-lag effect in the literature, and in cases where time-lag effect is considerable, such methods may not be appropriate to detect significant differences between two groups. In this paper, we propose a novel non-parametric approach for estimating the point of treatment time-lag effect by using an empirical divergence measure. Theoretical properties of the estimator are studied. The results from the simulated data and real data example support our proposed method.

Keywords: Change point analysis, Lag effect, Non-parametric statistics, Survival analysis, Treatment time-lag effect

1 Introduction

In essence, the problem faced in this paper is a change point analysis problem - we are trying to estimate the time of treatment time-lag effect, which can be interpreted as the time point at which the survival curves of two groups change distribution. There has been a large variety of research into the change point analysis problem for different applications including financial modeling, bioinformatics, and signal processing [17, 23, 31]. Overall, the problem concerns a change in distribution for a set of time-ordered observations. This is usually seen as a problem in univariate and multivariate time series data. There are both parametric and non-parametric methods of analysis, with parametric analysis having the necessity to assume underlying distributions. While change point

[1]Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, U.S.A.

[2]Department of Biostatistics, University of Florida, Gainesville, FL 32611, U.S.A.

analysis can be considered in terms of time series data, it is useful to us, as this also describes survival data, with an added possibility for censoring present in the data.

Survival analysis is a method for analyzing time-to-event data, where survival probabilities and times are sometimes presented in tables or graphs in a time-ordered fashion with indications as to which observations are censored or truncated. A major area of survival analysis includes testing for differences between groups by comparing hazard rate functions or, equivalently, survival probability functions [18]. Many procedures have been developed that non-parametrically test for differences between treatment groups. These procedures are applications of the Wilcoxon rank sum test, the most notable of which is the Log-rank test first proposed by Nathan Mantel (1966) [20]. There have been adjustments proposed that are appropriate in a variety of situations, including different weighting functions that can give weight to earlier or later observations. These include, but are not limited to, the Breslow, Fleming-Harrington, Gehan, Peto-Peto, and Tarone and Ware tests [3, 12, 10, 21, 25, 32]. There have also been methods proposed to deal with cases that have multiple or crossing hazard functions [6, 14, 26, 27, 19]. Additionally, there have been a few proposals to deal with cases that specifically have survival probabilities, that are initially quite similar between groups, and then differ [8, 24, 34]. There have even been semi-parametric and parametric approaches to estimating the time-lag effect using change point methods [2, 5].

While there has been much research into testing whether survival is significantly different between groups for censored and truncated data, less work has been done to test for differences when the two groups are similar up to some point (say, τ) and differ afterwards, specifically in terms of estimating the time point at which the two groups begin to differ. In literature, this time has been known as the treatment time-lag point [8, 24, 34]. This type of analysis can be important in many areas - parts reliability, treatment effect, etc. - since it can be crucial to not only know if two groups differ but also at what point in time some treatment begins to take effect.

Zucker and Lagatos (1990) proposed weighted log-rank statistics for comparing two survival curves when there is a time lag [34]. Dinse et al. (1993) proposed an estimate of the time-lag point based on Kaplan-Meier estimates of survival [8]. Park and Qiu (2018) suggested a semi-parametric model to estimate the time-lag point by using maximization of log partial likelihood [24]. There is also some research into this area using the terminology "change-point analysis." Chen and Baron (2014) reviewed some MLE methods as well as introducing some least-square estimation of the Cox proportional hazard model [5]. More recently, Brazzale et al. (2018) proposed a non-parametric method to estimate the time point of change in a single survival curve, based on fitting a stump regression to p values for testing hazards rates over small time intervals [2].

We will propose a way to estimate the time-lag point non-parametrically by adapting an approach for change point analysis of multivariate time series data [22, 30]. Specifically, we want a point estimate $\widehat{\tau}$ for τ, where τ is the time point

of change between two survival curves that have different distributions. In our case, we would like the survival curves to be quite similar up until $t = \tau$, and differ after that time point. Since this is a non-parametric approach, we only say that the two survival curves have different distributions but can not assume specifically what those distributions are (i.e. $F_1 \neq F_2$ where F_1 and F_2 are some unknown distributions). The paper will be structured as follows. In Section 2, we give an overview of the method and some theoretical properties. In Section 3, we describe the overall data simulation methods, and give simulation study results. In Section 4, we give a real data example to show an application of the method. Finally, Section 5 gives some concluding remarks and suggestions for further research. The proofs for some theoretical results are given in Appendix A.

2 Proposed method

2.1 Background information and notation

For our method, we consider n independent subjects. If we let X_i be the event time (or, equivalently, survival time) for subject i and C_i as the censoring time, then we have $t_i = \min(X_i, C_i)$ as the observed event times. This means that if the data is subject to right-censoring, t_i is observed instead of X_i. We then have a censoring indicator δ_i such that

$$\delta_i = \begin{cases} 1 & \text{if } X_i \leq C_i, \\ 0 & \text{if } X_i > C_i. \end{cases}$$

We now observe pairs (t_i, δ_i) for $i = 1, \ldots, n$. Assume we have $t_1 < t_2 < \cdots < t_T$ as the distinct event times in the pooled sample of $k = 1, 2$ groups. At time t_i, we observe d_{ij} events in the j^{th} sample out of y_{ij} individuals at risk, $i = 1, \ldots, T$. One of the most commonly used estimates for the survival function was proposed by Kaplan and Meier [16]. They proposed a survival estimate such that

$$\widehat{S}_k(t) = \prod_{i:t_i<t} \left(1 - \frac{d_{ij}}{y_{ij}}\right)^{\delta_i},$$

which is a step function with jumps at observations t_i for which $\delta_i = 1$. This estimate is non-parametric and can be applied in the presence of censoring. No assumptions are required for the probability distribution other than the independence between the survival and censoring variables.

2.2 Proposed Estimator

Since we will be comparing our method to a previous non-parametric method for finding the point of time-lag effect that was proposed by Dinse et al., it is

important to note that the method proposed by Dinse et al. uses the statistic

$$D(t) = \frac{\widehat{S}_1(t) - \widehat{S}_2(t)}{\sqrt{\widehat{V}_1(t) + \widehat{V}_2(t)}},$$

where, for $k = 1, 2$, $\widehat{V}_k(t) = \widehat{S}_k(t)^2 \sum_{t_i \leq t} \frac{d_{ik}}{y_{ik}(y_{ik} - d_{ik})}$ is the variance of $\widehat{S}_k(t)$ estimated by Greenwood's formula. Then, the proposed estimate of τ by Dinse et al. is $\sup\{t : t \leq t_T, D(t) \leq z_{1-\alpha}\}$ where $z_{1-\alpha}$ is the $(1-\alpha)^{th}$ percentile of the standard normal distribution [8]. The method proposed by Park and Qiu uses a semi-parametric method to estimate τ which utilizes a model that is a generalization of the conventional Cox proportional hazards model and maximum partial likelihood estimation [24].

In our method, we will be using a divergence measure based on the Euclidean distance between the two distributions to calculate the estimate $\widehat{\tau}$ [22, 30]. In order to be most accurate, we should trim the data set so that we are only including time points until one of the groups reaches a survival probability estimate of 0 (i.e one of the groups has a number of subjects at risk of 0). If this does not occur before the final event time recorded, we will include all time points.

For random variables $X, Y \in \mathbb{R}$, let a primed variable X' be an independent copy of X and primed variable Y' be an independent copy of Y. Now, suppose $X, X' \overset{iid}{\sim} F_1$ and $Y, Y' \overset{iid}{\sim} F_2$ and that X, X', Y and Y' are mutually independent. As proposed by Székely and Rizzo (2005), if $E|X|^\alpha, E|Y|^\alpha < \infty$ for some fixed constant $\alpha \in (0, 2)$, one Euclidean distance divergence measure can be defined as

$$\mathcal{E}(X, Y; \alpha) = 2E|X - Y|^\alpha - E|X - X'|^\alpha - E|Y - Y'|^\alpha.$$

For our proposed method, let us define the random vector $\boldsymbol{Z} = \{Z_1, \ldots, Z_T\}$ where $Z_l = S_1(t_l) - S_2(t_l)$ for $l = 1, \ldots, T$ where $S_1(t_l)$ and $S_2(t_l)$ are the survival probabilities at time point l of groups 1 and 2, respectively. Then, we can define vectors $\boldsymbol{X_\tau} = \{Z_i : i = 1, \ldots, \tau\}$ of length τ and $\boldsymbol{Y_\tau} = \{Z_j : j = \tau + 1, \ldots, T\}$ of length $T - \tau$. These are two independent iid samples such that $E|\boldsymbol{X_\tau}|^\alpha$, $E|\boldsymbol{Y_\tau}|^\alpha < \infty$ for some $\alpha \in (0, 2)$. Matteson and James used the measure proposed by Székely and Rizzo to define a empirical divergence measure based on U-statistics [22, 30]. Based on this method, and with Z_l and the random vectors $\boldsymbol{X_\tau}$ and $\boldsymbol{Y_\tau}$ defined as above, the empirical divergence measure we suggest is:

$$\widehat{\mathcal{E}}(\boldsymbol{X_\tau}, \boldsymbol{Y_\tau}; \alpha) = \frac{2}{\tau(T-\tau)} \sum_{i=1}^{\tau} \sum_{j=\tau+1}^{T} |X_i - Y_j|^\alpha - \binom{\tau}{2}^{-1} \sum_{1 \leq i < k \leq \tau} |X_i - X_k|^\alpha$$
$$- \binom{T-\tau}{2}^{-1} \sum_{(\tau+1) \leq j < k \leq T} |Y_j - Y_k|^\alpha.$$

If the assumptions hold, $\widehat{\mathcal{E}}(\boldsymbol{X_\tau}, \boldsymbol{Y_\tau}; \alpha) \overset{a.s.}{\to} \mathcal{E}(X, Y; \alpha)$ as $T \to \infty$ by Lévy's Continuity Theorem and the Strong Law of Large Number for U-statistics as

proven by Hoeffding, and properties of stochastic integrals for censored data as proven by Gill [11, 13].

To estimate the point of treatment time-lag effect, define

$$\widehat{\mathcal{Q}}(\boldsymbol{X}_{\boldsymbol{\tau}}, \boldsymbol{Y}_{\boldsymbol{\tau}}; \alpha) = \frac{\tau(T-\tau)}{T}\widehat{\mathcal{E}}(\boldsymbol{X}_{\boldsymbol{\tau}}, \boldsymbol{Y}_{\boldsymbol{\tau}}; \alpha),$$

as the scaled sample measure of divergence. It is then possible to estimate the time point of treatment time-lag effect by

$$\widehat{\tau} = \underset{\tau}{\operatorname{argmax}}\ \widehat{\mathcal{Q}}(\boldsymbol{X}_{\boldsymbol{\tau}}, \boldsymbol{Y}_{\boldsymbol{\tau}}; \alpha).$$

This statistic gives a consistent approach for estimating the treatment time-lag point by adapting clustering change point methods for multivariate time series. For all calculations and simulations in this paper, we will set $\alpha = 1$ for simplicity.

2.3 Theoretical Properties

We now give the assumptions and theorems showing the strong consistency of the estimator proposed in the previous section. For full proofs, please see the Appendix A.

Assumption 1. *Begin by assuming that we have a heterogeneous sequence of independent observations from two different distributions. Then, let* $\eta \in (0,1)$ *signify the fraction of observations belonging to one of the distributions such that* $Z_1, \ldots, Z_{\lfloor \eta T \rfloor} \sim F_1$ *and* $Z_{\lfloor \eta T \rfloor + 1}, \ldots, Z_T \sim F_2$ *for every sample of size* T*. Let* $r = \lfloor \eta T \rfloor$ *and* $s = T - r$*. Let* η *be bounded away from 0 and 1 such that* $r, s \to \infty$ *as* $T \to \infty$*. Also, let* $\mu_1^\alpha = E|X - X'|^\alpha$*,* $\mu_2^\alpha = E|Y - Y'|^\alpha$*, and* $\mu_{12}^\alpha = E|X - Y|^\alpha$*. Here,* $X, X' \overset{iid}{\sim} F_1$ *and* $Y, Y' \overset{iid}{\sim} F_2$ *and* X, X', Y, Y' *are all mutually independent. Further, suppose that* $E(|X|^\alpha + |Y|^\alpha) < \infty$ *for some* $\alpha \in (0,2)$*. Therefore,* $\mu_1^\alpha, \mu_2^\alpha, \mu_{12}^\alpha, \mathcal{E}(X, Y; \alpha) < \infty$*. Finally, let* $\{\delta_T\}$ *be a sequence of positive numbers such that* $\delta_T \to 0$ *and* $T\delta_T \to \infty$ *as* $T \to \infty$*.*

Lemma 1. *If Assumption 1 holds:*

$$\sup_{\eta \in [\delta_T, 1-\delta_T]} \left| \binom{T}{2}^{-1} \sum_{i<j} |Z_i - Z_j|^\alpha - [\eta^2 \mu_1^\alpha + (1-\eta)^2 \mu_2^\alpha + 2\eta(1-\eta)\mu_{12}^\alpha] \right| \overset{a.s.}{\to} 0,$$

as $T \to \infty$*.*

The proof follows from the Strong Law of Large Numbers for U-statistics, as well as the triangle inequality and properties of stochastic integrals for censored data [11, 13, 22].

Theorem 1. *Suppose Assumption 1 holds. Let* $\widehat{\tau}_T$ *be the point estimate of treatment time-lag point for a pooled sample with* T *distinct survival times. For* T *large enough,* $\delta_T < 1/2$ *and* $\eta \in [\delta_T, 1 - \delta_T]$*. Further, for each* $\epsilon > 0$ *we have:*

$$P\left(\lim_{T \to \infty} \left| \eta - \frac{\widehat{\tau}_T}{T} \right| < \epsilon \right) = 1.$$

This proves the almost sure convergence and strong consistency of the estimator. If we wish to consider specific rates of convergence, there must be additional information available about the distribution of the estimators. This, in turn, depends on the unknown distribution of the data [22].

3 Simulation Study

All computing for the methods presented was done in `R` [28]. We begin by simulating survival times from a Weibull distribution without time-invariant covariates:

$$\lambda(t|g) = \lambda\nu t^{\nu-1}\exp\{\beta\mathbb{I}(t>\tau)g\}, \tag{1}$$

where λ is the scale parameter, ν is the shape parameter, $\mathbb{I}(\cdot)$ is an indicator function, g is the group indicator (1 if the i^{th} subject is in the treatment group, and 0 otherwise), τ is the time-lag point, and β is a regression coefficient vector. In (1), we assume $\lambda = 0.5$, $\nu = 1.5$, $\beta = 1$, and $\tau = 1$ and also $n_1 = n_2 = 50$, $n_1 = n_2 = 250$, or $n_1 = n_2 = 500$ (i.e the number of subjects in each group is equal). In this paper, we simulate survival times by using results from Austin (2012) [1]. We also discretize the survival times in order to make results more realistic in the sense that there will be more than one event per time point. Additionally, in practice most survival times are not measured on a continuous scale [33]. In the following simulations, we have rounded to one decimal place. In order to simulate data that is subject to random right-censoring, we simulate survival times one at a time and also simulate a corresponding random censoring time from $U(a, b)$ where a is the minimum observed time and b is the maximum observed time. Once a desired number of censored observations has been created, we simulate only uncensored survival times until the sample size has been reached. We consider two different censoring rates of 20% and 40% in the simulation.

Table 1: Mean point estimate (PE), bias, and mean square error (MSE) of the estimate for τ in model (1). In the table, CR denotes the censoring rate.

Sample size	Method	20% CR			40% CR		
		PE	Bias	MSE	PE	Bias	MSE
100	Proposed	1.109	0.109	0.055	1.124	0.124	0.062
	Park and Qiu	0.890	-0.110	0.096	0.849	-0.151	0.112
	Dinse et al.	1.668	0.668	0.874	1.672	0.672	0.875
500	Proposed	1.077	0.077	0.012	1.111	0.111	0.019
	Park and Qiu	0.877	-0.123	0.023	0.865	-0.135	0.031
	Dinse et al.	1.134	0.134	0.275	1.102	0.102	0.202
1000	Proposed	1.065	0.065	0.008	1.096	0.096	0.013
	Park and Qiu	0.891	-0.109	0.014	0.880	-0.120	0.019
	Dinse et al.	1.052	0.052	0.155	1.045	0.045	0.195

The results for this simulation with 1000 replications are shown in Table 1

at different sample sizes for the Dinse et al. method, the Park and Qiu method, and the proposed method. The mean point estimate along with bias and mean square error values of the estimated time-lag point are summarized in Table 1. From the table, we can see that overall, the proposed method has a lower MSE than the other methods, even in cases where the bias is slightly larger. The proposed estimator also consistently performs better with larger samples sizes and with lower censoring rates, while this is not always true for the Dinse et al. estimator. In general, we find that the estimate for τ is over-estimated using the Dinse et al. method, and is slightly under-estimated using the Park and Qiu method. It is also of note that sometimes the Dinse et al. method does not give an estimate, in which case there was no possible solution in that replication. In order to find an estimate for τ, we remove these values and calculate the estimate from the remaining values found.

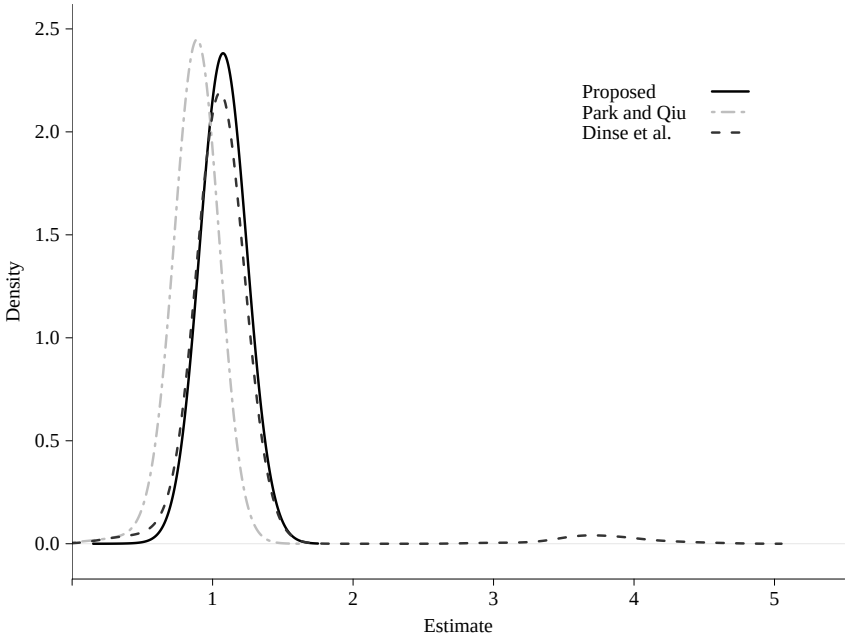

Figure 1: Density curves of the estimated time-lag points by the proposed, Park and Qiu, and Dinse et al. methods for 1000 replications, with the sample size of 500 and the censoring rate of 20% in model (1).

In the case when the sample size of 500 and the censoring rate of 20% in model (1), the density curves of the estimated time-lag points by the proposed, Park and Qiu, and Dinse et al. methods are shown in Figure 1. From the plots in the figure, it can be clearly seen that the results by Dinse et al. are generally more spread out, while the results from both proposed and Park and Qiu are typically much closer to each other and the true value of τ. The Dinse et al. results also show a cluster of time points that are quite over-estimated, which

indicates bi-modality instead of the normal distribution originally assumed by the authors [8].

In order to stay true to real data applications, we also simulate data including time-invariant covariates from the following more general model:

$$\lambda(t|g) = \lambda_0(t) \exp\{\beta \mathbb{I}(t > \tau) g + \gamma_1 v_1 + \gamma_2 v_2\}. \tag{2}$$

For example, we simulate covariate v_1 (gender) from a Binomial($n = 1$, $p = 0.5$) distribution and v_2 (age) from a U($a = 1$, $b = 25$) distribution rounded down to the year. We will show with these simulations that the addition of covariates does not largely change the results of the simulations. Further, we assume that $\lambda_0(t) = \lambda = 0.1$, $\beta = 1.25$, $\gamma_1 = 0.25$, $\gamma_2 = 0.1$, $\tau = 3$, and maximum allowed time of 10. The sample sizes are set to be $n_1 = n_2 = 100$, $n_1 = n_2 = 250$, or $n_1 = n_2 = 500$ and the censoring rates are 20% or 40%. Table 2 presents the results about the mean point estimate, bias, and mean square error of the estimate for τ in model (2). We can see in the table that the results still seem reasonable, and overall are consistent with the previous simulation results with no time-invariant covariates. The proposed method is comparable to the Park and Qiu method. Although the method of Park and Qiu performs somewhat better with the added information from covariates, this intuitively seems correct since the method is semi-parametric, and the proposed method does not use any of the information from the covariates in calculating the estimate. Also, the model used to simulate the data is the model used in the methods of Park and Qiu, so these results are reasonable. The Dinse et al. method still generally over-estimate, with the bias and MSE now being much larger than the previous simulation results.

Table 2: Mean point estimate (PE), bias, and mean square error (MSE) of the estimate for τ in model (2). In the table, CR denotes the censoring rate.

Sample size	Method	20% CR			40% CR		
		PE	Bias	MSE	PE	Bias	MSE
200	Proposed	3.210	0.210	0.646	3.269	0.269	0.559
	Park and Qiu	2.740	-0.260	0.423	2.674	-0.326	0.572
	Dinse et al.	6.617	3.617	17.373	6.370	3.370	15.927
500	Proposed	3.316	0.316	0.219	3.333	0.333	0.202
	Park and Qiu	2.805	-0.195	0.134	2.775	-0.225	0.170
	Dinse et al.	6.508	3.508	19.835	5.922	2.922	16.407
1000	Proposed	3.275	0.275	0.110	3.326	0.326	0.140
	Park and Qiu	2.848	-0.152	0.048	2.814	-0.186	0.075
	Dinse et al.	6.104	3.104	18.966	5.359	2.359	14.396

The density curves in Figure 2 show that the results for the data including covariates looks similar to the results that does not include covariates, but we can now much more obviously see the bi-modality of the Dinse et al. estimator. We see that this parameterization gives results that are approximately as accurate as previous simulations. This indicates that the addition of covariates

does not have a very large effect on the results of the method in estimating the point of treatment time-lag effect with the proposed method.

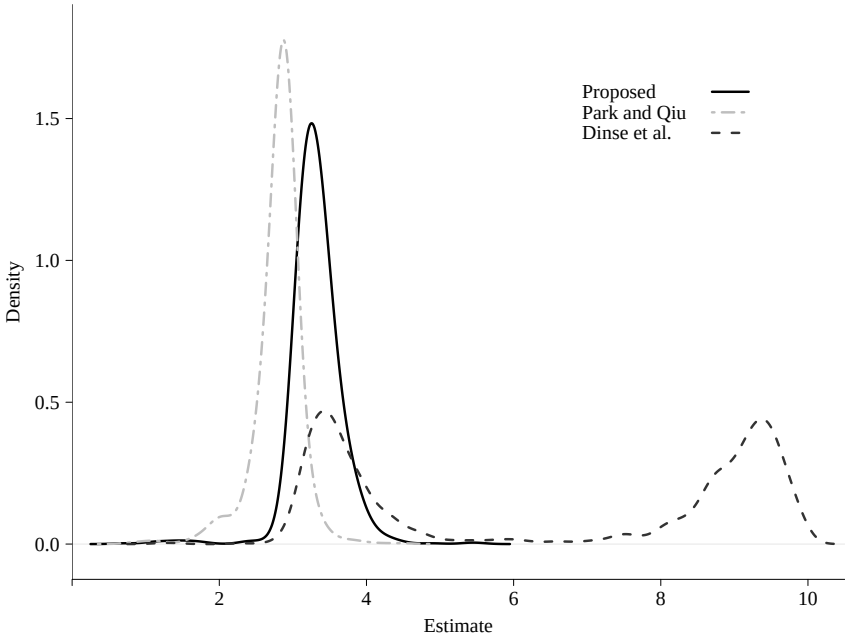

Figure 2: Density curves of the estimated time-lag points by the proposed, Park and Qiu, and Dinse et al. methods for 1000 replications, with the sample size of 500 and the censoring rate of 20% in model (2).

4 Application

In order to show an application of the proposed method, we will use a data set from the Veteran's Administration Lung Cancer Trial on patients with advanced, inoperable lung cancer who were treated with chemotherapy. The data is taken from Kalbfleisch and Prentice [15]. The variables available in the full data set are treatment (standard or test), cell type (squamous, small cell, adeno, large), survival (in days), status (dead or censored), a Karnofsky score as a measure of general performance, months from diagnosis, age in years, and a prior therapy indicator. If we split the data set into two groups depending on whether the patient received standard or test treatment, we can implement the methods shown in the prior section to estimate the time point of change between the two groups.

It may be of specific interest to see when patients who have had prior treatment begin to see a treatment lag effect between the standard treatment and the test treatment. We can subset this data set and view only subjects in the

study who have had previous treatment. In this case, we would have 40 patients in the study with an approximately 8% censoring rate. In this case, the number of subjects in the standard treatment group is 21 and 19 in the test group.

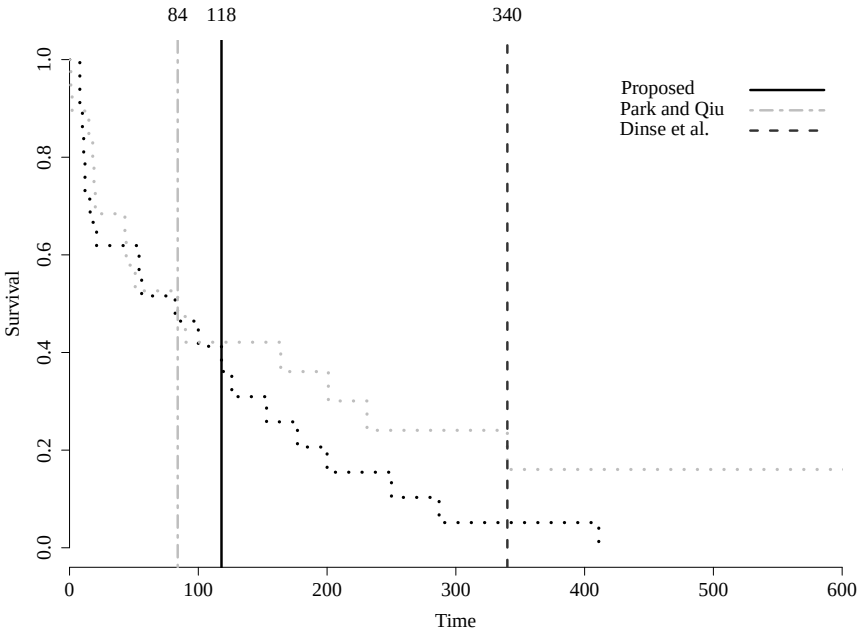

Figure 3: Kaplan-Meier curves for the distribution of time to lung cancer according to two groups; the black dotted line denotes the standard treatment group and the grey dotted line denotes the test treatment group. Estimated time-lag points by our proposed method (solid line), the Park and Qiu method (grey two-dashed line), and the Dinse et al. method (dashed line).

Figure 3 shows the Kaplan-Meier curves for the distribution of time to lung cancer according to two groups; the black dotted line denotes the standard treatment group and the grey dotted line denotes the test treatment group. There seems to be treatment lag-effect between the two groups. It is clear that no distinction can be made between two groups until around time point 100, and they show different patterns right after the time point. Estimated time-lag points by our proposed method (solid line), the Park and Qiu method (grey two-dashed line), and the Dinse et al. method (dashed line) are also presented in the figure. From the proposed method, we find an estimate for τ of 118. From the Park and Qiu method, we find an estimate of 84, and finally from the Dinse et al. method, we find an estimate of 340. The results of these estimates initially seem consistent with the results from the simulation study. The Park and Qiu estimator is, perhaps, slightly under-estimated while the proposed estimator is slightly over-estimated and the Dinse et al. estimator is in this case quite far from the time point of treatment lag-effect. This is consistent with the trends

in over- and under-estimation found in the simulation study.

In order to find a confidence interval for the estimator, we use the standard bootstrap method as initially suggested by Efron [9, 29]. We sample the observed (t_i, δ_i) with replacement and with stratification to account for the two treatment groups using the package `boot` [4, 7]. From 1000 replications, the 95% percentile bootstrap confidence interval for the estimate for τ is $(2.0, 177.0)$. When looking at the plot of the survival curves, this makes sense since the sample size is very small, many of the events take place early on in the study and there are several early time points that early on that could be seen as the "true" value of τ.

5 Conclusions

While there has been much research into change point analysis, there has been little research on time-lag effects. Some authors have used change point methods to find the time point of change in one group, but there have been few non-parametric methods considering two groups [2, 5, 24]. It should also be noted that non-parametric and semi-parametric methods that have already been proposed for survival analysis applications do not seem to be readily extendable to either identifying the time point of change between two groups, or to identify the time point of change non-parametrically. Throughout this paper, we have presented a novel non-parametric method for estimating the time point of treatment lag-effect using change point methods [22]. Some theoretical properties of strong consistency of the proposed estimator are shown.

From the simulation study, we found that our method using the change point tends to give more accurate results than the previously proposed Dinse et al. estimator in several different cases and simulation settings, and gives results that are generally comparable to the method suggested by Park and Qiu. We see that the distribution of the estimator seems to be empirically satisfactory, especially when compared to previously suggested non-parametric methods (primarily the Dinse et al. method). When compared to the methods of Park and Qiu, we see that the proposed estimator performs better in terms of bias and MSE when there are no covariates included, and gives reasonable results in the case where the information from covariates is included. This makes sense since the previous method of Park and Qiu is semi-parametric and includes covariate information [24].

Using a real data set for VA lung cancer data, we see that the estimator proposed here performs well when compared to the estimator proposed by Dinse et al., and gives results that are consistent with the simulation study in terms of the estimation bias [8, 15, 24]. The 95% percentile confidence interval using standard bootstrap methods is quite wide, and while the results do seem to agree with the graph visually, there are some potential areas to improve this research. Because survival data is somewhat complex in structure, and the sample size of the application data set is quite small this may be a case where the standard bootstrap fails and the percentile bootstrap confidence interval may not be appropriate to use [9, 29]. In the future, it would be useful to use

some alternative method that may be more suited to the structure of the data.

Overall, the method we present in this paper gives results that are reasonable both in a simulation setting and with real data application. Despite the concerns of the validity of the bootstrap methods used, the empirical results suggest that this non-parametric estimator found from change point methods is applicable and appropriate for use to analyze survival data.

References

[1] P. C. Austin, *Generating survival times to simulate Cox proportional hazards models with time-varying covariates*, Statistics in Medicine, 31 (2012), p. 3946–3958.

[2] A. R. Brazzale, H. Küchenhoff, S. Krügel, T. S. Schiergens, H. Trentzsch, and W. Hartl, *Nonparametric change point estimation for survival distributions with a partially constant hazard rate*, Lifetime Data Analysis, 25 (2019), pp. 301–321.

[3] N. Breslow, *A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship*, Biometrika, 57 (1970), pp. 579–594.

[4] A. Canty and B. D. Ripley, *boot: Bootstrap R (S-Plus) Functions*, 2017.

[5] X. Chen and M. Baron, *Change-Point Analysis of Survival Data with Application in Clinical Trials*, Open Journal of Statistics, 4 (2014), pp. 663–677.

[6] M. Y. Cheng, P. Qiu, X. Tan, and D. Tu, *Confidence intervals for the first crossing point of two hazard functions*, Lifetime Data Anal, 4 (2009), pp. 441 – 454.

[7] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and Their Applications*, Cambridge University Press, Cambridge, 1997.

[8] G. E. Dinse, W. W. Piegorsch, and D. D. Boos, *Confidence Statements about the Time Range Over Which Survival Curves Differ*, Journal of the Royal Statistical Society. Series C (Applied Statistics), 42 (1992), pp. 21–30.

[9] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, no. 57 in Monographs on Statistics and Applied Probability, Chapman & Hall/CRC, Boca Raton, Florida, USA, 1993.

[10] E. A. Gehan, *A generalized Wilcoxon test for comparing arbitrarily singly-censored samples*, Biometrika, 52 (1965), pp. 203–24.

[11] R. D. Gill, *Censoring and Stochastic Integrals*, Mathematical Centre Tracts: Amsterdam, 1980.

[12] D. P. Harrington and T. R. Fleming, *A Class of Rank Test Procedures for Censored Survival Data*, Biometrika, 69 (1982), pp. 553 – 566.

[13] W. Hoeffding, *The Strong Law of Large Numbers for U-statistics*, tech report 302, North Carolina State University, Dept. of Statistics, 1961.

[14] H. Huang, Z. Chen, and P. Qiu, *Comparison of multiple hazard rate functions*, Biometrics, 72 (2016), pp. 39–45.

[15] D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, Wiley, New York, 1980.

[16] E. L. Kaplan and P. Meier, *Nonparametric Estimation from Incomplete Observations*, Journal of the American Statistical Association, 53 (1958), pp. 457–481.

[17] A. Kim, C. Marzban, D. Percival, and W. Stuetzie, *Using Labeled Data to Evaluate Change Detectors in a Multivariate Streaming Environment*, Signal Processing, 12 (2009), pp. 2529–2536.

[18] J. P. Klein and M. L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data*, Springer, 2nd ed., 2003.

[19] K. Liu, P. Qiu, and J. Sheng, *Comparing two crossing hazard rates by cox proportional hazards modelling*, Stat Med, 2 (2007), pp. 375–391.

[20] N. Mantel, *Evaluation of survival data and two new rank order statistics arising in its consideration*, Cancer Chemotherapy Reports, 3 (1966), pp. 163–170.

[21] N. Mantel, N. R. Bohidar, and J. L. Ciminera, *Mantel-Haenszel analyses of litter-matched time to response data, with modifications for recovery of interlitter information*, Cancer Research, 37 (1977), pp. 3863–3868.

[22] D. S. Matteson and N. A. James, *A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data*, Journal of the American Statistical Association, 109 (2014), pp. 334 – 345.

[23] V. M. Muggeo and G. Adelfio, *Efficient Change Point Detection for Genomic Sequences of Continuous Measurements*, Bioinformatics, 27 (2011), pp. 161–166.

[24] K. Park and P. Qiu, *Evaluation of the treatment time-lag effect for survival data*, Lifetime Data Anal, 24 (2018), pp. 310–327.

[25] R. Peto and J. Peto, *Asymptotically Efficient Rank Invariant Test Procedures*, Journal of the Royal Statistical Society. Series A (General), 135 (1972), pp. 185–207.

[26] P. Qiu, H. Huang, and Z. Chen, *An improved two-stage procedure to compare hazard curves*, Journal of Statistical Computation and Simulation, 87 (2017), pp. 1877–1886.

[27] P. Qiu and J. Sheng, *A two-stage procedure for comparing hazard rate functions*, Journal of the Royal Statistical Society. Series B (Statistical Methodology), 70 (2008), pp. 191–208.

[28] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018.

[29] M. L. Rizzo, *Statistical Computing with R*, Computer Science and Data Analysis, Chapman & Hall /CRC, 2008.

[30] G. J. Székely and M. L. Rizzo, *Hierarchical Clustering via Joint Between-Within Distances: Extending Ward's Minimum Variance Method*, Journal of Classification, 22 (2005), pp. 151–183.

[31] M. Talih and N. Hengartner, *Structural Learning With Time-Varying Components: Tracking the Cross-Section of Financial Time Series*, Journal of the Royal Statistical Society, 67 (2005), pp. 151–183.

[32] R. E. Tarone and J. Ware, *On distribution-free tests for equality of survival distributions*, Biometrika, 64 (1977), pp. 156–60.

[33] G. Tutz and M. Schmid, *Modeling Discrete Time-to-Event Data*, Springer International Publishing, 1st ed., 2016.

[34] D. M. Zucker and E. Lakatos, *Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment*, Biometrika, 77 (1990), pp. 853–64.

Appendix A

Lemma 1. *If Assumption 1 holds,*

$$\sup_{\eta\in[\delta_T,1-\delta_T]}\left|\binom{T}{2}^{-1}\sum_{i<j}|Z_i-Z_j|^\alpha-[\eta^2\mu_1^\alpha+(1-\eta)^2\mu_2^\alpha+2\eta(1-\eta)\mu_{12}^\alpha]\right|\xrightarrow{a.s.}0,\ \textit{as } T\to\infty.$$

Proof. Let $\epsilon^*>0$. Pick $\epsilon>0$ so that $\epsilon^3+\epsilon^2(2+3\mu_1^\alpha)+\epsilon<\epsilon^*$. Define sets $A_1=\{(i,j)|i<j;Z_i,Z_j\sim F_1\}$, $A_2=\{(i,j)|Z_i\sim F_1,Z_j\sim F_2\}$, and $A_3=\{(i,j)|i<j;Z_i,Z_j\sim F_2\}$. Then, let M_1,M_2, and M_3 be the number of elements in A_1,A_2, and A_3, respectively. Note that these sets are disjoint.

By the Strong Law of Large Numbers for U-Statistics (Hoeffding, 1961), there exists an $N_1\in\mathbb{N}$ s.t. whenever $M_1>N_1$

$$\left|\binom{M_1}{2}^{-1}\sum_{A_1}|Z_i-Z_j|^\alpha-\mu_1^\alpha\right|<\epsilon.$$

We can similarly define $N_2,N_3\in\mathbb{N}$. Then, there also exists an $N_4\in\mathbb{N}$ such that, for $T>N_4$, $\frac{1}{T-1}<\epsilon/2$.

Let $N=\max\{N_1,N_2,N_3,N_4\}$ s.t. N is large enough that $\delta_T<1/2$. Then, $\forall\ T\delta_T>N$ and $\forall\ \eta\in[\delta_T,1-\delta_T]$ it is true that $M_1=\lfloor\eta T\rfloor>N_1$, $M_2=\lfloor\eta T\rfloor(T-\lfloor\eta T\rfloor)>N_2, M_3=(T-\lfloor\eta T\rfloor)>N_3$, and also that each of $|\frac{r}{T}-\eta|,|\frac{r-1}{T-1}|,|\frac{s}{T}-(1-\eta)|$, and finally $|\frac{s-1}{T-1}-(1-\eta)|$ are less than ϵ.

It is then true that

$$\begin{aligned}\binom{T}{2}^{-1}\sum_{A_1}|Z_i-Z_j|^\alpha&=\frac{2}{T(T-1)}\sum_{A_1}|Z_i-Z_j|^\alpha\\&=\frac{2}{r(r-1)}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha\\&=\binom{r}{2}^{-1}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha.\end{aligned}$$

Also,

$$\begin{aligned}&\left|\frac{r}{T}-\eta\right|\left|\frac{r-1}{T-1}-\eta\right|<\epsilon^2\\\implies&\left|\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)-\eta\left(\frac{r}{T}-\frac{r-1}{T-1}\right)+\eta^2\right|<\epsilon^2\\\implies&\left|\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)+\eta^2\right|<\epsilon^2+\eta\left(\frac{r}{T}-\frac{r-1}{T-1}\right)<\epsilon^2+2\eta\epsilon\quad\because\epsilon>0\text{ is arbitrary}\\\implies&\left|\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)-\eta^2\right|<\left|\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)+\eta^2\right|<\epsilon^2+2\eta\epsilon\quad\because r,T\geq1\end{aligned}$$

Then, we can rearrange the inequalities so that

$$\left|\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)-\eta^2\right|\left|\binom{r}{2}^{-1}\sum_{A_1}|Z_i-Z_j|^\alpha-\mu_1^\alpha\right|<\epsilon^3+2\eta\epsilon^2$$

$$\implies \left|\binom{r}{2}^{-1}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha-\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\mu_1^\alpha-\eta^2\binom{r}{2}^{-1}\sum_{A_1}|Z_i-Z_j|^\alpha+\eta^2\mu_1^\alpha\right|<\epsilon^3+2\eta\epsilon^2$$

$$\implies \left|\binom{r}{2}^{-1}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha+\eta^2\mu_1^\alpha\right|<\epsilon^3+2\eta\epsilon^2+\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\mu_1^\alpha+\eta^2\binom{r}{2}^{-1}\sum_{A_1}|Z_i-Z_j|^\alpha$$

$$<\epsilon^3+2\eta\epsilon^2+\epsilon^2\mu_1^\alpha+\eta^2\epsilon<\epsilon^3+2\eta\epsilon^2+\mu_1^\alpha\epsilon^2(1+2\eta)+\eta^2\epsilon$$

$$\implies \left|\binom{r}{2}^{-1}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha-\eta^2\mu_1^\alpha\right|<\left|\binom{r}{2}^{-1}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha+\eta^2\mu_1^\alpha\right|$$

$$<\epsilon^3+\epsilon^2(2\eta+(1+2\eta)\mu_1^\alpha)+\eta^2\epsilon$$

$$<\epsilon^3+\epsilon^2(2+3\mu_1^\alpha)+\epsilon.$$

So, $T>N$ implies that

$$P\left(\left|\binom{r}{2}^{-1}\left(\frac{r}{T}\right)\left(\frac{r-1}{T-1}\right)\sum_{A_1}|Z_i-Z_j|^\alpha-\eta^2\mu_1^\alpha\right|<\epsilon^*\right)=1 \quad \text{for } \epsilon^*>0.$$

We can apply similar reasoning to the sets defined by A_2 and A_3. Then, using the triangle inequality, we complete the proof since $\epsilon^*>0$ is arbitrary, and we have uniform convergence.

Theorem 1. *Suppose Assumption 1 holds. Let $\widehat{\tau}_T$ be the point estimate of treatment time-lag effect for a pooled sample with T distinct survival times. For T large enough, $\delta_T<1/2$ and $\eta\in[\delta_T,1-\delta_T]$. Further, $\forall\epsilon>0$*

$$P\left(\lim_{T\to\infty}\left|\eta-\frac{\widehat{\tau}_T}{T}\right|<\epsilon\right)=1.$$

Proof. Let T be such that $\eta\in[\delta_T,1-\delta_T]$. Then, for any $\tilde{\eta}\in[\delta_T,1-\delta_T]$, let $\tilde{r}=\lfloor\tilde{\eta}T\rfloor$ and $\tilde{s}=T-\tilde{r}$. Then, $\boldsymbol{X}_{\tilde{\boldsymbol{r}}}=\{Z_1,\ldots,Z_{\tilde{r}}\}$ and $\boldsymbol{Y}_{\tilde{\boldsymbol{r}}}=\{Z_{\tilde{r}+1},\ldots,Z_T\}\,\forall\,T$. Then

$$\begin{aligned}\widehat{\mathcal{E}}(\boldsymbol{X}_{\tilde{\boldsymbol{r}}},\boldsymbol{Y}_{\tilde{\boldsymbol{r}}};\alpha)&\overset{a.s.}{\to}\left(\frac{\eta}{\tilde{\eta}}\mathbb{I}(\tilde{\eta}\geq\eta)+\frac{1-\eta}{1-\tilde{\eta}}\mathbb{I}(\tilde{\eta}<\eta)\right)^2\mathcal{E}(X,Y;\alpha)\\&=h(\tilde{\eta};\eta)\mathcal{E}(X,Y;\alpha),\end{aligned}$$

as $T\to\infty$, uniformly in $\tilde{\eta}$. The maximum of $h(\tilde{\eta};\eta)$ is attained when $\tilde{\eta}=\eta$. We also see that

$$\frac{1}{T}\widehat{\mathcal{Q}}(\boldsymbol{X}_{\tilde{\boldsymbol{r}}},\boldsymbol{Y}_{\tilde{\boldsymbol{r}}};\alpha)\overset{a.s.}{\to}\tilde{\eta}(1-\tilde{\eta})h(\tilde{\eta};\eta)\mathcal{E}(X,Y;\alpha),$$

as $T \to \infty$, uniformly in $\tilde{\eta}$. Additionally, the maximum of $\tilde{\eta}(1-\tilde{\eta})h(\tilde{\eta};\eta)$ is also attained when $\tilde{\eta} = \eta$. Now, define

$$\widehat{\tau}_T = \underset{\tau \in \{\lceil T\delta_T \rceil, \ldots, \lfloor T(1-\delta_T) \rfloor\}}{\operatorname{argmax}} \widehat{\mathcal{Q}}(\boldsymbol{X}_{\boldsymbol{\tau}}, \boldsymbol{Y}_{\boldsymbol{\tau}}; \alpha),$$

and the interval

$$\widehat{\Gamma}_T = \underset{\tilde{\eta} \in [\delta_T, 1-\delta_T]}{\operatorname{argmax}} \widehat{\mathcal{Q}}(\boldsymbol{X}_{\tilde{\boldsymbol{r}}}, \boldsymbol{Y}_{\tilde{\boldsymbol{r}}}; \alpha).$$

Then, we can see that $\frac{\widehat{\tau}_T}{T} \in \widehat{\Gamma}_T$. Since

$$\frac{1}{T}\widehat{\mathcal{Q}}\left(\boldsymbol{X}_{\widehat{\boldsymbol{\tau}}_{\boldsymbol{T}}/\boldsymbol{T}}, \boldsymbol{Y}_{\widehat{\boldsymbol{\tau}}_{\boldsymbol{T}}/\boldsymbol{T}}; \alpha\right) > \frac{1}{T}\widehat{\mathcal{Q}}\left(\boldsymbol{X}_{\boldsymbol{\eta}}, \boldsymbol{Y}_{\boldsymbol{\eta}}; \alpha\right) - o(1),$$

we have that

$$\frac{1}{T}\widehat{\mathcal{Q}}\left(\boldsymbol{X}_{\widehat{\boldsymbol{\tau}}_{\boldsymbol{T}}/\boldsymbol{T}}, \boldsymbol{Y}_{\widehat{\boldsymbol{\tau}}_{\boldsymbol{T}}/\boldsymbol{T}}; \alpha\right) \geq \eta(1-\eta)h(\eta;\eta)\mathcal{E}(X,Y;\alpha) - o(1),$$

by the almost sure uniform convergence shown previously. Letting $\widehat{\eta} = \frac{\widehat{\tau}_T}{T}$, it follows that

$$\begin{aligned} 0 &\leq \eta(1-\eta)h(\eta;\eta)\mathcal{E}(X,Y;\alpha) - \widehat{\eta}(1-\widehat{\eta})h(\widehat{\eta};\eta)\mathcal{E}(X,Y;\alpha) \\ &\leq \frac{1}{T}\widehat{\mathcal{Q}}(\boldsymbol{X}_{\widehat{\boldsymbol{\eta}}}, \boldsymbol{Y}_{\widehat{\boldsymbol{\eta}}}; \alpha) - \widehat{\eta}(1-\hat{\eta})h(\widehat{\eta};\eta)\mathcal{E}(X,Y;\alpha) + o(1). \end{aligned}$$

This tends to 0 as $T \to \infty$. For every $\epsilon > 0$, $\exists\ \epsilon^*$ such that

$$\tilde{\eta}(1-\tilde{\eta})h(\tilde{\eta};\eta)\mathcal{E}(X,Y;\alpha) < \eta(1-\eta)h(\eta;\eta)\mathcal{E}(X,Y;\alpha) - \epsilon^* \quad \forall\ \tilde{\eta} \text{ with } |\tilde{\eta}-\eta| \geq \epsilon.$$

Therefore,

$$\begin{aligned} P\left(\lim_{T\to\infty} |\widehat{\eta}_T - \eta| \geq \epsilon\right) &\leq P\left(\lim_{T\to\infty} \widehat{\eta}_T(1-\widehat{\eta}_T)h(\widehat{\eta}_T;\eta)\mathcal{E}(X,Y;\alpha) < \eta(1-\eta)h(\eta;\eta)\mathcal{E}(X,Y;\alpha) - \epsilon^*\right) \\ &= 0. \end{aligned}$$

This proves the claim of uniform convergence and strong consistency of the estimator. To specifically consider the rates of convergence, we need additional information about the distribution of estimators which, in turn, depends on the distribution of the data (which is considered to be unknown or arbitrary).