# A New Correlation Coefficient for Right-Censored Survival Data

**Abstract**

This paper concerns accessing the association or dependence between the covariates and right-censored survival outcome. The main challenge lies in the presence of partially unobservable time-to-events due to the reasons such as the end of the study follow-up. A novel nonparametric and robust correlation coefficient is proposed by incorporating all information on the data. It neither constrains the distributions of covariates and survival outcomes, nor involves tuning parameters. Furthermore, it can effectively detect non-linear or non-monotonic relationships between covariates and right-censored outcome, even under heavy censoring. More importantly, the correlation coefficient is shown to converge to a new dependence measure that has rather desirable properties of being between 0 and 1, and being 0 or 1 if and only if the two random variables are independent or one is a measurable function of the other. To our knowledge, we are the first to devise a correlation coefficient for right-censored data with a limit satisfying all these appealing properties. The coefficient is also shown to be asymptotically normal. Building upon the coefficient, we construct permutation tests of independence. Extensive empirical studies show its good performance in detecting various complicated dependencies.

*Keywords:* Chatterjee's rank correlation; Nonparametric correlation; Right-censoring.

# 1 Introduction

The task of evaluating or identifying statistical associations between the covariates and a survival outcome is of great interest in clinical studies and biomedical sciences. For example, in clinical trials, one may wish to rank the importance of covariates for the time-to-event or detect the relevant variables to the failure time, such as the age, gender or any other measurements. Right-censored data appear frequently in survival analysis. The main challenge lies in the presence of censoring, where a portion of the survival outcome remains unobserved except for a lower bound, typically due to loss to follow-up or withdrawal from the study.

The correlation coefficient is a powerful tool to measure the dependence between the covariates and the response, which has been extensively applied in independence testing and high-dimensional variable screening. In cases where the data is completely observed, the most well-known classical coefficients are Pearson's correlation coefficient, Spearman's $\rho$ and Kendall's $\tau$. But, they are not effective to detect non-monotonic associations (Chatterjee, 2021). This motivates the development of new coefficients, including the maximal correlation coefficient (Breiman and Friedman, 1985), maximal information coefficient (Reshef et al., 2011), distance correlation (Székely et al., 2007; Székely and Rizzo, 2007), and kernel-based methods (Gretton et al., 2005, 2008), among many others. Recently, an ingenious rank-based correlation was introduced by Chatterjee (2021), which has quickly caught much attention in statistical community due to, e.g., its simplicity and ability to detect non-monotonic or nonlinear associations. However, for incompletely observed data, direct application of these coefficients is not feasible.

In the context of right-censored data, current approaches are particularly designed for testing the independence between the covariate and the survival outcome. These methods usually entails a semiparametric model setting; see, e.g., Cox (1972); Gray (1992); Zucker and Karr (1990). The most commonly used method is the Cox proportional hazard model by Cox (1972), which assumes a linear structure of the covariates on the log-hazard function. Nevertheless, in some cases these assumptions can be easily violated. For instance, when studying the relationship between body mass index and mortality, the hazard are observed to be U-shaped or V-shaped (Bhaskaran et al., 2018; Zajacova and Burgard, 2012).

One may resort to nonparametric approaches, which are still relatively scarce in the literature. Peng and Fine (2008) used the idea of continuously thresholding to assess the effect of a continuous covariate on multistate survival data. Ma and Mao (2019) further proposed the Fisher exact scanning method for testing and identifying dependency between two scalar random variables; see others, e.g., Le et al. (1994), McKeague et al. (1995). Extending these methods to the more general case, where the covariates belong to $\mathbb{R}^d$, is much challenging since many of them rely on dichotomizing the covariate.

Although advancements in non-parametric methods for multivariate covariates are limited, the rapid development of the distance correlation (Székely et al., 2007; Székely and Rizzo, 2007) in recent decades has sparked new interest in this issue. For instance, Rindt et al. (2021) used the optimal transport to transform the censored dataset into an uncensored one, and then conducted independence testing using a kernel-based criteria. However, this comes with the risk of significant information loss in the original data. Fernández et al. (2023) proposed a kernel-based weighted log rank tests of independence between the multivariate covariates and the failure time. Edelmann et al. (2022) constructed a distance covariance for right-censored data based on the inverse-probability-of-censoring weighting (IPCW). These methods, yet, often have one of the following limitations. First, the computation burden is at least $O(n^2)$ for the distance covariance based methods. Second, some of them may not robust to outliers, since, for example, their calculations usually require finite moment conditions of the covariates or the time-to-event. Third, tuning parameters, such as the kernel smoothing bandwidth, are involved. On top of that, as most of them are mainly tailored to independence testing, the population value of their statistics may lack certain desirable properties as a correlation coefficient.

In this article, we propose a fully nonparametric correlation coefficient for right-censored survival data, beyond simply testing for independence, which is possible to detect any dependencies between covariates and the survival outcome. This coefficient possesses virtues mainly in the following aspects: i) it is tuning-free and needs no assumptions on the distributions of covariates or survival outcomes and is robust to heavy-tailed outliers; ii) it can effectively detect non-linear or non-monotonic associations, even under heavy censoring; iii) it is much easier to calculate, requiring only approximately linear computational time. Additionally, it does not require complicated estimators for the unknown distribution of outcomes or the conditional densities under right-censoring; iv) more importantly, the new coefficient converges to a limit that has rather desirable properties of being between 0 and 1, and being 0 or 1 if and only if the two random variables are independent or one is a measurable function of the other. To our knowledge, our proposed coefficient is the first that accommodates right-censored data with a limit satisfying all these properties. This limit is referred to as the *dependence measure* (in population). The form of this measure is flexible enough to allow researchers to design ad-hoc coefficients according to their specific needs. Moreover, the new coefficient is shown to be asymptotically normal. As an important application, we construct the permutation test of independence. Empirical studies demonstrate its good performance.

Beyond independence testing, statistical community is also interested in developing model-free impact index between one scalar covariate and the time-to-event to be used in high-dimensional feature screening. He et al. (2013) proposed an IPCW-based marginal quantile-adaptive regression variable screening for high-dimensional data. Song et al. (2014) built up a Kendall's $\tau$ type of screening criterion using the technique of IPCW, among many others (Li et al., 2016; Chen et al., 2018; Pan et al., 2019; Edelmann et al., 2020). Nonetheless, many of them are not applicable for multivariate covariates. On the other hand, it is not clear that whether their estimators are suitable as a correlation coefficient for the covariate and survival outcome, as no theory ensures convergence to a dependence measure with the aforementioned desired properties. In this sense, our proposed correlation has the potential to enhance variable screening accuracy.

The remainder of the paper proceeds as follows. Section 2 gives the new dependence measure. Sections 3.1 and 3.2 present the correlation coefficient for right-censored data and its asymptotic properties. Section 3.3 discusses the connection to Chatterjee's rank correlations. The values of the correlation under strictly increasing transformations is shown in Section 3.4. The application of the new correlation on independence testing is presented in Section 3.5. Simulation studies and real data analysis are presented in Section 4 and 5, respectively. Technical proofs are provided in the supplementary material.

# 2 A Measure of Dependence

Let $Y$ be a random variable in $\mathbb{R}$ with the law $\mu$ and $X$ be an $\mathbb{R}^d$-valued random vector. To measure the dependence between $Y$ and $X$, we propose the following measure of dependence in population:

$$\xi(X, Y) = \frac{\displaystyle\int \mathrm{Var}\{\mathrm{E}\,[\,\mathbf{1}(Y \geq t)|\, X]\}\, dW(t)}{\displaystyle\int \mathrm{Var}\{\mathbf{1}(Y \geq t)\}\, dW(t)}, \tag{2.1}$$

where $\mathbf{1}(\cdot)$ is the indicator function and $W$ is a finite measure on the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ is the Borel $\sigma$-algebra on $\mathbb{R}$. Let $\mathcal{S}_\mu$ and $\mathcal{S}_W$ be the support of $\mu$ and $W$, respectively. Recall that the support of a measure $\nu$ is the set of all $x \in \mathbb{R}$ such that any open ball containing $x$ has strictly $\nu$-measure.

Under the following assumption, we show that the dependence measure (2.1) possesses the appealing properties in Theorem 2.1.

**Assumption 2.1.** $\mathcal{S}_\mu \subseteq \mathcal{S}_W$. In addition, for any $t \in \mathbb{R}$, if $\mu(\{t\}) > 0$, we must have $W(\{t\}) > 0$.

**Theorem 2.1.** *Suppose Assumption 2.1 holds and $Y$ is not almost surely a constant. Then, we have*

*(i)* $0 \leq \xi(X, Y) \leq 1$*;*

*(ii) (Consistency.)* $\xi(X, Y) = 0$ *if and only if $X$ and $Y$ are independent;*

*(iii) (Perfect dependency.)* $\xi(X, Y) = 1$ *if and only if $Y = h(X)$ almost surely, for some measurable function of $h(\cdot)$.*

Assumption 2.1 specifies the relationship between the supports of $\mu$ and the measure $W$, where the former is required to be a subset of the latter. It is relatively weak as claimed in Remark 2.2. Theorem 2.1 states that as long as $W$ satisfies the weak Assumption 2.1, the new dependence measure remains consistent and can also reflect the perfect dependency. In contrast, it is unclear yet that whether the distance covariance is able to detect perfect dependence.

Now, we provide an overview of our motivation and offer some insights into the idea behind the dependence measure (2.1) in the following two remarks.

**Remark 2.1** (*Motivations and insights on (2.1)*)**.** If we take $W(\cdot) = \mu(\cdot)$, then (2.1) is exactly the dependence measure introduced by Dette et al. (2013), defined as:

$$\eta(X,Y) = \frac{\displaystyle\int \mathrm{Var}\{\mathrm{E}\left[\mathbf{1}(Y \geq t) \middle| X\right]\} \, d\mu(t)}{\displaystyle\int \mathrm{Var}\{\mathbf{1}(Y \geq t)\} \, d\mu(t)}. \tag{2.2}$$

When $Y$ is a continuous random variable, one can write $d\mu(t) = f_Y(t)dt$, where $f_Y(t)$ is the density of $Y$. Therefore, $f_Y(t)$ can be regarded as a weight function for all possible values of $Y$. However, $\mu(t)$ is unknown in practice and thus is usually replaced by its empirical counterpart as Chatterjee (2021). In the presence of missing or censored data, approximating $\mu(t)$ becomes much more complex and time-consuming, leading to an estimator that carries a heavy computational burden or yield lower accuracy. To avoid this issue and broaden the applicability of (2.2), we consider to replace $\mu$ with a pre-specified measure $W$ that satisfies Assumption 2.1.

**Remark 2.2** (*Discussion on the measure $W$*)**.** The form of $W$ can vary depending on specific cases, as long as it satisfies Assumption 2.1. However, in real applications, the relationship between the supports of $\mu$ and $W$ is unknown or unclear. For a continuous random variable $Y$, it is easy to verify that taking $dW(t) = w(t)dt$ with $w(t) > 0$ guarantees Assumption 2.1. For simplicity, one may take $W$ to be a probability measure and we define $F_W(t) = W((-\infty, t])$, for every $t \in \mathbb{R}$. If $F_W(t)$ is differentiable, then there exits one $w(t) \geq 0$ such that $dW(t) = w(t)dt$, where $w(t)$ is the density of $F_W(t)$ with respect to the Lebesgue measure. Further, if $w(t) > 0$, for all $t \in \mathbb{R}$, then Theorem 2.1 holds. Specifically, $F_W(t)$ can be commonly used distributions, such as the normal, Cauchy, or $t$ distributions as in Section 4.

# 3 A New Correlation Under Right-Censoring

## 3.1 Construction of the Correlation

Let $(X, T, C)$ be a collection of random variables taking values on $\mathbb{R}^d \times \mathbb{R}_+ \times \mathbb{R}_+$, where $T$ is the survival outcome of interest, $C$ is a censoring variable, and $X$ is a $d$-dimensional covariate vector with $d \geq 1$. In practice, due to the end of study follow-up or withdrawal from the study, we can not directly observe samples from $(X, T, C)$. Instead, only samples from $(X, Y, \delta)$ are observable, where $Y = \min(T, C)$ and $\delta = \mathbf{1}(T \leq C)$. For ease of exposition, we assume that the censoring variable $C$ is independent of $T$ and $X$. In the presence of right-censoring, discarding all the censored observations would lead to an inefficient estimator under heavy censoring. On the other hand, completely ignoring the censored information by regarding the censored ones as the survival time is also inappropriate. Therefore, the main challenge lies in how to effectively utilize the censored observations with $\delta_i = 0$. Our goal is to construct a new correlation coefficient that incorporates all the available data in a straightforward manner.

Let $\{(X_i, Y_i, \delta_i)\}_{i=1}^n$ being independent and identically distributed (i.i.d.) samples from $(X, Y, \delta)$. Here and after, we take $dW(t) = S_c^2(t)w(t)dt$, where $S_c(t) = P(C \geq t)$ and $w(t)$ is a pre-determined weight function satisfying $\int_{\mathbb{R}_+} w(t)dt < \infty$. This setting of measure $W$ is tailored for the right-censored survival data; see details in Remark 3.3. Let $N(i)$ index the nearest neighbor (NN) of $X_i$ among $\{X_j\}_{j=1}^n$ under the Euclidean metric on $\mathbb{R}^d$, with ties breaking at random. Then, $(X_{N(i)}, Y_{N(i)}, \delta_{N(i)})$ is the observation with $X_{N(i)}$ being the NN of $X_i$. To measure the association between

$T$ and $X$, we define the following coefficient:

$$\hat{\xi}_n(X,T) = \frac{\int_{\mathbb{R}_+} Q_n(t)w(t)dt - \int_{\mathbb{R}_+} G_n(t)^2 w(t)dt}{\int_{\mathbb{R}_+} \hat{S}_c(t)G_n(t)w(t)dt - \int_{\mathbb{R}_+} G_n(t)^2 w(t)dt}$$
$$:= \frac{T_{n1} - T_{n2}}{T_{n3} - T_{n2}}, \tag{3.1}$$

where $Q_n(t) = n^{-1}\sum_{i=1}^n \mathbf{1}(Y_i \geq t)\mathbf{1}(Y_{N(i)} \geq t)$, $G_n(t) = n^{-1}\sum_{i=1}^n \mathbf{1}(Y_i \geq t)$, and $\hat{S}_c(t)$ is the Kaplan-Meier estimator of $S_c(t)$, defined as

$$\hat{S}_c(t) = \prod_{j=1}^n \left(1 - \frac{1}{\sum_{i=1}^n \mathbf{1}(Y_i \geq Y_j)}\right)^{(1-\delta_j)\mathbf{1}(Y_j \leq t)}.$$

Define "$0/0 = 0$". To better illustrate the idea behind (3.1), we first deduce

$$\begin{aligned} T_{n1} &= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}_+} \frac{\mathbf{1}(Y_i \geq t)}{S_c(t)} \cdot \frac{\mathbf{1}(Y_{N(i)} \geq t)}{S_c(t)} S_c(t)^2 w(t)dt \\ &= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}_+} \eta_i(t)\eta_i^*(t)S_c(t)^2 w(t)dt. \\ &= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}_+} \eta_i(t)\eta_i^*(t)dW(t), \end{aligned}$$

where $\eta_i(t) = \mathbf{1}(Y_i \geq t)/S_c(t)$ and $\eta_i^*(t) = \mathbf{1}(Y_{N(i)} \geq t)/S_c(t)$. Similarly, we can rewrite

$$T_{n2} = \int_{\mathbb{R}_+} \bar{\eta}_n(t)^2 dW(t), \quad T_{n3} = \int_{\mathbb{R}_+} \bar{\eta}_n(t)\Delta_n(t)dW(t),$$

where $\bar{\eta}_n(t) = n^{-1}\sum_{i=1}^n \eta_i(t)$ and $\Delta_n(t) = \hat{S}_c(t)/S_c(t)$. Then, the coefficient $\hat{\xi}_n(X,T)$ in (3.1) can be rewritten as

$$\hat{\xi}_n(X,T) = \frac{\int_{\mathbb{R}_+} n^{-1}\sum_{i=1}^n \eta_i(t)\eta_i^*(t)dW(t) - \int_{\mathbb{R}_+} \bar{\eta}_n(t)^2 dW(t)}{\int_{\mathbb{R}_+} \bar{\eta}_n(t)\Delta_n(t)dW(t) - \int_{\mathbb{R}_+} \bar{\eta}_n(t)^2 dW(t)}. \tag{3.2}$$

Note that

$$\mathrm{E}[\eta_i(t)] = \frac{\mathrm{E}[\mathbf{1}(Y_i \geq t)]}{S_c(t)} = \frac{\mathrm{E}[\mathbf{1}(T_i \geq t)\mathbf{1}(C_i \geq t)]}{S_c(t)} = \mathrm{E}[\mathbf{1}(T_i \geq t)]. \tag{3.3}$$

Hence, our proposal in (3.1) or (3.2) can be viewed as a moment method of the measure (2.1), where we leverage a NN-based estimator to approximate the conditional probability. In other words, (3.2) indeed measures the dependence between $X$ and $T$ in the presence of right-censoring.

**Remark 3.1** (*Computational complexity of* $\hat{\xi}_n(X,T)$)**.** One significant advantage of $\hat{\xi}_n(X,T)$ lies in its ease of implementation. Our fully nonparametric approach avoids estimating conditional densities, characteristic functions, or mutual information, eliminating the need for smoothing or tuning parameters as required by Dette et al. (2013). So, the computational time is mainly spent on performing nearest neighbor searches using Euclidean distance or ranking, both of which have a time complexity of $O(n\log(n))$; see Sharma et al. (2019). Hence, the total computational complexity of $\hat{\xi}_n(X,T)$ is around $O(n\log(n))$. While in practice, additional computation cost may be incurred by numerical integrations. The simulation results in Table 8 illustrate that the computational time of $\hat{\xi}_n(X,T)$ is approximately linear.

**Remark 3.2.** In principle, one can take $W$ to be an empirical process as that in (Chatterjee, 2021; Azadkia and Chatterjee, 2021). For instance, one may use $W_1(t) = n^{-1}\sum_{i=1}^n \mathbf{1}(Y_i \leq t)$ and $dW_2(t) = J(t)dN(t)/\bar{Y}(t)$, where $\bar{Y}(t) = \sum_{j=1}^n \mathbf{1}(Y_j \geq t)$, $J(t) = \mathbf{1}(\bar{Y}(t) > 0)$ and $N(t) = \sum_{i=1}^n \delta_i \mathbf{1}(Y_i \leq t)$. Here, $W_1(t)$ is the empirical distribution function of $Y$, and $W_2(t)$ is the Nelson-Aalen estimator of the cumulative hazard rate function. In these cases, the measure $W$ is related to the samples, which we leave it for future research.

### 3.2 Asymptotic Properties of $\hat{\xi}_n(X,T)$

We suppose the $T$ is not almost surely a constant. Define $\widetilde{G}(t) = P(T \geq t)$, $\widetilde{G}_X(t) = P(T \geq t|X)$. Note that $G_X(t) = P(Y \geq t|X) = P(T \geq t|X)P(C \geq t) = \widetilde{G}_X(t)S_c(t)$. For $dW(t) = S_c^2(t)w(t)dt$, the measure $\xi(X,T)$ in (2.1) can be written as

$$\begin{aligned}
\xi(X,T) &= \frac{\int_{\mathbb{R}_+} \mathrm{E}[\widetilde{G}_X^2(t)]dW(t) - \int_{\mathbb{R}_+} \widetilde{G}(t)^2 dW(t)}{\int_{\mathbb{R}_+} \widetilde{G}(t)dW(t) - \int_{\mathbb{R}_+} \widetilde{G}(t)^2 dW(t)} \\
&= \frac{\int_{\mathbb{R}_+} \mathrm{E}[G_X^2(t)]w(t)dt - \int_{\mathbb{R}_+} G(t)^2 w(t)dt}{\int_{\mathbb{R}_+} G(t)S_c(t)w(t)dt - \int_{\mathbb{R}_+} G(t)^2 w(t)dt} \\
&:= \frac{T_{01} - T_{02}}{T_{03} - T_{02}}. \qquad (3.4)
\end{aligned}$$

Denote $x \wedge y = \min\{x,y\}$. Let $F_W(t) = \int_0^t w(s)ds$, $T_{01}^* = \mathrm{E}[F_W(Y_1 \wedge Y_{N(1)})]$ and $\xi_n^*(X,T) = (T_{01}^* - T_{02})/(T_{03} - T_{02})$. Denote the joint distribution of $(X,T,C)$ by $F_{X,T,C}$. Then, the following theorem shows that the correlation $\hat{\xi}_n(X,T)$ is consistent and asymptotically normal under certain conditions.

**Theorem 3.1.** *Suppose $\int_{\mathbb{R}_+} w(t)dt < \infty$ and $w(t) \geq 0$ for all $t \in \mathbb{R}_+$. For $dW(t) = S_c^2(t)w(t)dt$, we have that*

*(i) (Consistency.) as $n \to \infty$,*

$$\hat{\xi}_n(X,T) \xrightarrow{p} \xi(X,T);$$

*(ii) (Asymptotic normality.) when $F_{X,T,C}$ is continuous, under assumptions (A1)-(A3) in the supplementary material, $n^{1/2}(\hat{\xi}_n(X,T) - \xi_n^*(X,T))$ converges to a zero-mean normal distribution, where, for some fixed constant $\beta \geq 0$,*

$$\xi_n^*(X,T) = \xi(X,T) + O\left(\frac{(\log n)^{d+\beta+1+\mathbf{1}(d=1)}}{n^{1/d}}\right). \qquad (3.5)$$

Theorem 3.1 shows that $\hat{\xi}_n(X,T)$ converges to the dependence measure $\xi(X,T)$ defined in (2.1) or equivalently (3.4). When Assumption 2.1 is satisfied, for example, in the case that the support of $C$ covers that of $T$ and $w(t) > 0$ for all $t \in \mathbb{R}_+$, the appealing properties as listed in Theorem 2.1 hold.

On the other hand, note that $\xi_n^*(X,T)$ is a constant sequence converging to the dependence measure $\xi(X,T)$. In particular, when $X$ and $T$ are independent, $\xi_n^*(X,T) = 0$. Hence, the result in Theorem 3.1 (ii) reduces to that $n^{1/2}\hat{\xi}_n(X,T)$ converges to a zero-mean normal distribution. Under the dependence of $X$ and $T$, similar convergence rates to those in (3.5) are shown in Proposition 1.1 of Lin and Han (2022) and Theorem 4.1 of Azadkia and Chatterjee (2021). When $d = 1$, it is direct to see that $\xi_n^*(X,T) = \xi(X,T) + o(n^{-1/2})$. As a result, $n^{1/2}(\hat{\xi}_n(X,T) - \xi(X,T))$ converges to a zero-mean normal distribution. Moreover, we note that proving the asymptotic normality for Chatterjee's types of rank correlations as (3.6) and (3.7) is non-trivial, even under the independence of $X$ and $T$; see Azadkia and Chatterjee (2021); Lin and Han (2022); Deb et al. (2020) for more details.

**Remark 3.3.** The setting $dW(t) = S_c^2(t)w(t)dt$ can avoid the case that $S_c(t)$ or $\hat{S}_c(t)$ appears in the numerator, which would easily yield a value close to zero and render the associated term unbounded. In addition, the weight function $w(t)$ provides additional flexibility for capturing specific survival patterns of interest. For example, when examining the impact of early survival, one may set $w(t)$ as a decreasing function with respect to $t$. In real application, to ensure Assumption 2.1, we usually utilize $w(t) > 0$ for all $t \in \mathbb{R}_+$, as suggested in Remark 2.2.

**Remark 3.4.** Let $\mathcal{S}_T$ and $\mathcal{S}_C$ be the supports of $T$ and $C$, respectively. Note that the case of $\mathcal{S}_C \subset \mathcal{S}_T$ corresponds to the situation where the maximal survival time of an individual is longer than the follow-up, which is often the case in clinical or observational studies. In this scenario, the limit in Theorem 3.1 reflects the dependence between $T$ and $X$ during the follow-up period.

### 3.3 Connections to Chatterjee's Rank Correlations

For fully observable data, Azadkia and Chatterjee (2021) and Chatterjee (2021) proposed the following two rank correlation coefficients, respectively:

$$\hat{\eta}_n(X,T) = \frac{\sum_{i=1}^n (n \min\{R_{N(i)}, R_i\} - L_i^2)}{\sum_{i=1}^n L_i(n-L_i)}, \quad \text{for } d \geq 1, \tag{3.6}$$

$$\hat{\eta}_n^*(X,T) = 1 - \frac{n\sum_{i=1}^n |R_{\tilde{N}(i)} - R_i|}{2\sum_{i=1}^n L_i(n-L_i)}, \quad \text{for } d = 1, \tag{3.7}$$

where $R_i$ is the rank of $T_i$, $L_i$ is the number of $j$ such that $T_j \geq T_i$, and $N(i)$, $\tilde{N}(i)$ index the NN of $X_i$ among $\{X_j\}_{j=1}^n$ and the right NN of $X_i$ among $\{X_j\}_{j=1}^n$ when $d = 1$.

When the censoring variable $C = \infty$, there are no censored observations. In this scenario, all survival outcomes or failure times are completely observed, namely, $Y_i = T_i$ and $\delta_i = 1$ for all $i = 1, \cdots, n$. It follows that the Kaplan-Meier estimator $\hat{S}_c(t) = 1$. In this case, set $dW(t) = dF_n(t)$, where $F_n(t) = n^{-1}\sum_{i=1}^n \mathbf{1}(T_i \leq t)$. By some simple algebra, we have

$$n^2 \sum_{i=1}^n \{F_n(T_i) \wedge F_n(T_{N(i)})\} = n\sum_{i=1}^n \{R_i \wedge R_{N(i)}\}, \quad \text{and}$$
$$n^3 \int_{\mathbb{R}_+} G_n(t)^2 dF_n(t) = = \sum_{i=1}^n L_i^2, \quad n^3 \int_{\mathbb{R}_+} G_n(t) dF_n(t) = n\sum_{i=1}^n L_i.$$

Therefore, in the case of no censoring, our coefficient $\hat{\xi}_n(X,T)$ reduces to $\hat{\eta}_n(X,T)$ in (3.6).

On the other hand, for $d = 1$, when $T$ and $X$ have a continuous joint distribution function, the correlation coefficient $\hat{\eta}_n^*(X,T)$ in (3.7) can be simplified to:

$$\hat{\eta}_n^*(X,T) = 1 - \frac{3}{n^2-1}\sum_{i=1}^{n} |R_{\tilde{N}(i)} - R_i|, \tag{3.8}$$

where we replace the NN index $N(i)$ of $i$ in (3.1) with its right NN index $\tilde{N}(t)$. Similar to Remark 3 in Lin and Han (2023), one can easily show that $\hat{\xi}_n(X,T)$ reduces to (3.8) with an asymptotically ignorable small term. More precisely,

$$\begin{aligned}\hat{\xi}_n(X,T) &= \hat{\eta}_n^*(X,T) + \frac{3}{n^2-1}\sum_{i=1}^{n}(R_{\tilde{N}(i)} - R_i)\\ &= \hat{\eta}_n^*(X,T) + O(n^{-1}),\end{aligned}$$

where the last equality holds by noting that $\{R_i\}_{i=1}^n$ is permutation of $\{1,\dots,n\}$. Then, for any $i,j \in \{1,\cdots,n\}$, the difference between $R_i$ and $R_j$ is at most $n-1$, resulting in $|\sum_{i=1}^n (R_{\tilde{N}(i)} - R_i)| \le n-1$.

### 3.4 Under Strictly Increasing Transformations

Suppose that $h(\cdot)$ is a strictly increasing mapping from $\mathbb{R}_+$ to $\mathcal{Y}$, a Borel set of $\mathbb{R}$. Denote the transformed data as $T^* = h(T)$ and $C^* = h(C)$. Now, we figure out the relationship between the coefficients before and after the transformation.

Suppose that $w(t)$ is a density function with $w(t) > 0$ for all $t \in \mathbb{R}_+$. Define $F_Z(x) = P(Z \le x) = \int_{-\infty}^{x} w(t)dt$ for all $x \in \mathbb{R}$, where $Z$ is a random variable mapping from some sample space to $\mathbb{R}_+$ with distribution function $F_Z(x)$. Then, given $w(t)$, we rewritten the original coefficient $\hat{\xi}_n(X,T)$ as a function of $F_Z$:

$$\hat{\xi}_n(X,T,F_Z) = \frac{\int_{\mathbb{R}_+} Q_n(t)dF_Z(t) - \int_{\mathbb{R}_+} G_n(t)^2 dF_Z(t)}{\int_{\mathbb{R}_+}\{\hat{S}_c(t) - G_n(t)\}G_n(t)dF_Z(t)}.$$

Additionally, for any $t \in \mathbb{R}_+$, $F_Z(t) = P(Z \le t) = P(h(Z) \le h(t))$. Here, we define $Z^* = h(Z)$ to be a new random variable from $\mathbb{R}_+$ to $\mathcal{Y}$ with distribution function $F_{Z^*}(u) = P(Z^* \le u) = F_Z(h^{-1}(u))$, for all $u \in \mathcal{Y}$, where $h^{-1}(u)$ denotes the preimage of $u$.

Then, by noting that $Y^* = T^* \wedge C^* = h(Y)$, we have

$$\begin{aligned}\int_{\mathbb{R}_+} Q_n(t)dF_Z(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}_+} \mathbf{1}(Y_i \ge t)\mathbf{1}(Y_{N(i)} \ge t)dF_Z(t)\\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}_+} \mathbf{1}\{Y_i^* \ge h(t)\}\mathbf{1}\{Y_{N(i)}^* \ge h(t)\}dF_Z(t)\\ &= \int_{t\in\mathcal{Y}} Q_n^*(t)dF_{Z^*}(t),\end{aligned}$$

where $Q_n^*(t) = n^{-1}\sum_{i=1}^{n}\mathbf{1}(Y_i^* \ge t)\mathbf{1}(Y_{N(i)}^* \ge t)$. Moreover, let $\hat{S}_c^*(t)$ to be the Kaplan-Meier estimator of $S_c^*(t) = P(C^* \ge t)$. One can easily get $\hat{S}_c(t) = \hat{S}_c^*(h(t))$. Therefore, the coefficient after the transformation $h$ is

$$\begin{aligned}\hat{\xi}_n(X, T^*, F_{Z^*}) &= \frac{\int_{t\in\mathcal{Y}} Q_n^*(t)dF_{Z^*}(t) - \int_{t\in\mathcal{Y}} G_n^*(t)^2 dF_{Z^*}(t)}{\int_{t\in\mathcal{Y}}\{\hat{S}_c^*(t) - G_n^*(t)\}G_n^*(t)dF_{Z^*}(t)},\\ &= \hat{\xi}_n(X, T, F_Z), \qquad (3.9)\end{aligned}$$

where $G_n^*(t) = n^{-1}\sum_{i=1}^{n}\mathbf{1}(Y_i^* \ge t)$. This implies that the correlation coefficients before and after a strictly increasing transformation of both $T$ and $C$ are equivalent, up to different well-defined probability measures as (2.1). In fact, since $dF_{Z^*}(u)/du = w(h^{-1}(u))dh^{-1}(u)/du$, we have $\int_{\mathbb{R}} dF_{Z^*}(u) = \int_{\mathbb{R}_+} w(t)dt < \infty$. This allows for the application of the proposed coefficient to a broader range of random variables under right-censoring, instead of being limited to survival outcomes.

**Remark 3.5.** In particular, when the data are generated from linear transformation models, we have $h(t) = \log(t)$. According to (3.9), given $w^*(t) = w(e^t)e^t$, we have

$\hat{\xi}_n(X, \log(T), w^*) = \hat{\xi}_n(X, T, w)$. For instance, if we take $w(t)$ to be the density of log-normal, then $w^*(t)$ is the density of normal distribution with the same parameter. Similarly, when we set $w(t)$ to be the density of log-$t$ distribution, $w^*(t)$ is exactly the associated density of $t$-distribution with the same degree of freedom. In addition, it is clear that $w(t) > 0$, for all $t \in \mathbb{R}_+$, implies that $w^*(t) > 0$ for all $t \in \mathbb{R}$.

## 3.5 An Application to Permutation Test

When one is interested in detecting dependence between covariates and survival times, we consider the following hypotheses:

$$H_0 : \xi(X, T) = 0, \qquad \text{v.s.} \qquad H_1 : \xi(X, T) \neq 0. \tag{3.10}$$

Given the asymptotic results in Theorem 3.1, since the unknown covariance is hard to estimate, we suggest constructing a permutation independence test between $X$ and $T$ for (3.10).

Given the number of permutations to be $B$, in each round $b \in \{1, \ldots, B\}$, we draw a permuted sample from the uniform distribution over all possible permutations on $\{1, \ldots, n\}$, denoted by $\{(X_i^{(b)}, Y_i, \delta_i)\}_{i=1}^n$. We then evaluate $\hat{\xi}_n^{(b)}(X, T)$ for each permutation as follows:

$$\hat{\xi}_n^{(b)}(X, T) = \frac{\int_{\mathbb{R}_+} Q_n^{(b)}(t) w(t) dt - \int_{\mathbb{R}_+} G_n(t)^2 w(t) dt}{\int_{\mathbb{R}_+} \{\hat{S}_c(t) - G_n(t)\} G_n(t) w(t) dt}, \tag{3.11}$$

where $Q_n^{(b)}(t) = n^{-1} \sum_{i=1}^n \mathbf{1}(Y_i \geq t) \mathbf{1}(Y_{N_b(i)} \geq t)$ with $X_{N_b(i)}$ being the NN of $X_i^{(b)}$ for pairs $(X_i^{(b)}, Y_i, \delta_i)$ and $(X_{N_b(i)}^{(b)}, Y_{N_b(i)}, \delta_{N_b(i)})$ in the permutation samples. Under $H_0$, $\hat{\xi}_n^{(b)}(X, T)$ will have the same distribution as $\hat{\xi}_n(X, T)$ because of the independence between $X$ and $T$ and the distribution-freeness of the proposed correlation coefficient.

Then, the $p$-value of the permutation test is given by the empirical proportion

$$p_v = \frac{1}{B+1}\left[1 + \sum_{b=1}^{B} \mathbf{1}\left\{\hat{\xi}_n^{(b)}(X,T) \geq \hat{\xi}_n(X,T)\right\}\right].$$

For a given significant level $\alpha \in (0,1)$, the permutation test is defined as $T_{\alpha,B} = \mathbf{1}\{p_{value} \leq \alpha\}$. The results of extensive simulation studies indicate satisfactory performance of the proposed permutation test.

# 4 Simulation Studies

In this section, we conduct simulations to evaluate the empirical performance of the proposed censored correlation coefficient.

We first investigate the finite sample performance of the proposed new correlations. We consider the relationships between the covariates and the failure time are both generalize linear and nonlinear, respectively. The corresponding data generating processes (DGP) are presented as follows. The sample sizes are set to be 200, 400 or 800, respectively, with 1000 repetitions for each.

**Example 4.1** (Linear transformation model)**.** We consider the data are generated from

$$\mathbf{DGP1}: \quad \log(T) = 2X_1 + 2X_2 + 4X_3 + \epsilon,$$

where the covariates $X = (X_1, \cdots, X_p)^T$ follows the multivariate normal distribution with mean $\mathbf{0}$ and the covariance matrix $\Sigma = (\sigma_{ij})_{p\times p}$ with $p = 3$, $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho,\ i \neq j$, $\epsilon \sim N(0, 5^2)$ independent with $X$. In this DGP, we consider different values of $\rho = 0, 0.4,$ and $0.8$. The logarithm of censoring time $C$ is generated from a mixture of the normal and uniform distributions: $N(0,4) - N(1,1) + cU(0,3)$, where the positive constant $c$ are adjusted to ensure the censoring rate (CR) is around 30%.

**Example 4.2** (Nonlinear models)**.** The samples are generated from the following two models:

$$\begin{aligned} \mathbf{DGP2}: &\quad T = X_1^2 + \epsilon, \\ \mathbf{DGP3}: &\quad T = -2\log(X_2) + \epsilon, \end{aligned}$$

where the covariates $X_1$ and $X_2$ are drawn from standard normal $N(0,1)$ and the uniform distribution $U(0,1)$, respectively. The error $\epsilon$ follows from the chi-squared distribution with two degrees of freedom, denoted by $\chi^2(2)$. In the above two DGPs, the censoring variable $C$ is assumed to follow the mixture distribution of gamma $c_0\text{Gamma}(k,\theta)$ and uniform distributions, namely, $c_0\text{Gamma}(k,\theta) + c_1 U(0,1)$. Here, the positive constants $c_0$ and $c_1$ are adjusted flexibly to ensure that the censoring rates are approximately 15%.

In each scenario, we utilize various forms of $w(t)$ to calculate the dependence measure. These include the exponential distribution with rate one (Exp(1)), the chi-squared distribution with the degree of freedom three ($\chi^2(3)$) and the log-type of distributions with density defined as $\tilde{f}(x) = f(\log(x))/x$, where $f(\cdot)$ is set to be the density of normal distributions, uniform distribution, the Cauchy distribution and the $t$ distribution, respectively. In contrast, we also set $w(t)$ to be the true density of survival time $f_T(t)$, though it is usually unknown in practice. Tables 1, 2, 3 present the true values (True) of associated dependence measures with different $w(t)$, along with the biases (Bias) and standard deviations (SD) of the proposed correlation coefficients. The smoothed densities of the estimated correlation coefficients based on different $w(t)$ are depicted in Figures 1-3.

The results presented in Tables 1, 2, 3 indicate that the bias and standard deviations exhibit stability and decrease as the sample size $n$ increases. Notably, in each scenario, the values of calculated dependence measures based on different $w(t)$ are very similar. Furthermore, as observed in Tables 1 and 2, stronger dependence between covariates and $T$ leads to larger values of the dependence measure. Importantly, the variances of the proposed correlation coefficient remain stable across different cases and are not sensitive to the choice of $w(t)$. The main observation from Figures 1-3 is that the empirical density of $\sqrt{n}\hat{\xi}_n(X,T)$ fits well with a normal distribution.

Next, we evaluate the performance of the proposed permutation tests based on the new correlation coefficients with different $w(t)$, including the normal distribution, the uniform distribution and the logistic distribution with various parameters. The number of permutations is 500. In contrast, we also perform the IPCW-based distance correlation ($\mathrm{Dcor}_{\mathrm{IPCW}}$) permutation test in (Edelmann et al., 2022) for the censored data using the R package dcortools.

Let $\tilde{Y}$ be the transformed survival time. The error $\epsilon$ generated from the standard normal $N(0,1)$ is independent with covariates. Let the values of $\lambda$ vary from 0 to 1. Now, we consider the data generated from the following DGPs:

**DGP4** (Gaussian): $\tilde{Y} = \lambda(-3X_1 - 0.4X_2 + 2X_3) + 2\epsilon$, where the covariates $X = (X_1, \cdots, X_p)^T$ follows the multivariate normal distribution with mean $\mathbf{0}$ and the covariance matrix $\Sigma = (\sigma_{ij})_{p\times p}$ with $p = 3$, $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.4$. The censoring variable $C$ is generated from the Gaussian mixture model $N(2.5, 0.3^2) - 2N(\mu, 0.6^2)$.

**DGP5** (Sinusoid): $\tilde{Y} = 3\lambda\cos(2\pi X) + 0.5\epsilon$, where $X$ follows the uniform distribution $U(-1,1)$. The censoring variable $C$ is generated from the Gaussian mixture model $N(3, 0.5^2) - c_0N(1, 0.1^2)$.

**DGP6** (W-shaped): $\tilde{Y} = 10\lambda(|X+1.5|\mathbf{1}\{X<0\} + |X-1.5|\mathbf{1}\{X \geq 0\}) + \epsilon_0$, where $X \sim U(-3,3)$ and $\epsilon_0$ follows logistic distribution $logit(0,1)$. The censoring variable $C$ is generated from the Gaussian mixture model $2N(\mu, 0.5^2) - N(1, 0.1^2)$.

**DGP7**: Let $f$ be a step function taking values $-5, 6, -5, 4$ in the intervals $[-1,-0.5)$, $[-0.5, 0)$, $[0, 0.5)$, and $[0.5, 1]$, respectively. We generate $Y$ from the model $Y = 2.8\lambda\{f(X_1) + \exp(2.5X_2 - 1)\} + \epsilon_0$, where $X_1 \sim U(-1,1)$, $X_2 \sim Beta(0.5, 0.5)$ and $\epsilon_0$ follows logistic distribution $logit(2,4)$. The censoring variable $C$ is generated from the Gaussian mixture model $N(0,4^2) - N(5,1) + 0.5N(\mu, 1)$.

For each DGP, 500 data sets with sample size $n = 200$ and 400 are analyzed. Tables 4-7 report empirical rejection rates for the permutation tests based on our proposed correlation using different $w(t)$ and the IPCW-based distance correlation with significant level 0.05. The censoring rates (CR) are around 45% and 65%, respectively.

There are mainly three observations from Tables 4-7. (i) All the seven tests have type I error rates fairly close to the significant level, even for the scenarios with heavy censoring. (ii) The new tests are much powerful than the distance-correlation-based test when the data is sinusoidal, W-shaped and piecewise constant as shown in DGPs 5-7. Moreover, it seems that the $\mathrm{Dcor}_{\mathrm{IPCW}}$ tests has no power for data drawn from DGP5 and DGP6 when the censoring rate is around 65%. (iii) The power of our tests based on different $w(t)$ appears to be of comparable magnitude, albeit with slight distinctions. The observation (ii) coincides with the performance of Chatterjee's correlation, which excels at handling oscillatory signals but may exhibit inferior performance with linear correlations.

At last, we compare the computation times for each permutation tests for DGP4 with $\rho = 0.4$ and the censoring rate (CR) around 10% and 40% with different sample sizes. The machine we perform the experiments on is equipped with AMD EPYC 7H12 64-Core Processor and a 640GB memory. Tables 8 present all the computational times. We see that the computation times of the proposed tests tend to increase linearly with sample size, while that of the IPCW-based distance correlation test seemingly exhibits explosive growth. In addition, we observe that the computation time varies with $w(t)$ but seemingly not relies on the censoring rate.

Table 1: Estimation results of $\hat{\xi}_n(X, T)$ for DGP1 with $\sigma = 5$ and various $w(t)$

| | | | $n = 200$ | | $n = 400$ | | $n = 800$ | |
|---|---|---|---|---|---|---|---|---|
| $\rho$ | $w(t)$ | True | Bias | SD | Bias | SD | Bias | SD |
| 0 | $f_T(t)$ | 0.3023 | -0.0153 | 0.0789 | -0.0055 | 0.0569 | 0.0029 | 0.0374 |
| | Exp(1) | 0.3238 | -0.0199 | 0.0950 | -0.0132 | 0.0633 | -0.0085 | 0.0459 |
| | $\chi^2(3)$ | 0.3240 | -0.0235 | 0.0959 | -0.0123 | 0.0692 | -0.0107 | 0.0474 |
| | log-$U(-50, 50)$ | 0.2862 | 0.0076 | 0.0855 | 0.0063 | 0.0583 | 0.0047 | 0.0407 |
| | log-$N(0, 1)$ | 0.3247 | -0.0221 | 0.0948 | -0.0156 | 0.0635 | -0.0077 | 0.0461 |
| | log-$N(-1, 50^2)$ | 0.2863 | 0.0077 | 0.0803 | 0.0068 | 0.0585 | 0.0065 | 0.0405 |
| | log-Cauchy$(0, 1)$ | 0.3207 | -0.0192 | 0.0907 | -0.0105 | 0.0633 | -0.0061 | 0.0430 |
| | log-$t(2)$ | 0.3228 | -0.0178 | 0.0933 | -0.0136 | 0.0621 | -0.0073 | 0.0449 |
| | log-$t(3)$ | 0.3236 | -0.0194 | 0.0918 | -0.0149 | 0.0642 | -0.0068 | 0.0447 |
| 0.4 | $f_T(t)$ | 0.3979 | -0.0085 | 0.0737 | -0.0050 | 0.0491 | 0.0017 | 0.0365 |
| | Exp(1) | 0.4204 | -0.0171 | 0.0854 | -0.0079 | 0.0607 | -0.0045 | 0.0424 |
| | $\chi^2(3)$ | 0.4207 | -0.0118 | 0.0927 | -0.0093 | 0.0665 | -0.0056 | 0.0457 |
| | log-$U(-50, 50)$ | 0.3804 | 0.0088 | 0.0725 | -0.0074 | 0.0529 | 0.0061 | 0.0353 |
| | log-$N(0, 1)$ | 0.4211 | -0.0132 | 0.0884 | -0.0063 | 0.0612 | -0.0048 | 0.0446 |
| | log-$N(-1, 50^2)$ | 0.3809 | 0.0098 | 0.0739 | 0.0080 | 0.0516 | 0.0071 | 0.0353 |
| | log-Cauchy$(0, 1)$ | 0.4174 | -0.0140 | 0.0845 | -0.0082 | 0.0585 | -0.0026 | 0.0424 |
| | log-$t(2)$ | 0.4194 | -0.0154 | 0.0834 | -0.0081 | 0.0613 | -0.0035 | 0.0418 |
| | log-$t(3)$ | 0.4202 | -0.0129 | 0.0840 | -0.0066 | 0.0624 | -0.0040 | 0.0423 |
| 0.8 | $f_T(t)$ | 0.4624 | 0.0020 | 0.0681 | 0.0033 | 0.0476 | 0.0023 | 0.0335 |
| | Exp(1) | 0.4847 | -0.0050 | 0.0813 | -0.0042 | 0.0579 | -0.0036 | 0.0390 |
| | $\chi^2(3)$ | 0.4850 | -0.0095 | 0.0858 | -0.0069 | 0.0631 | -0.0018 | 0.0445 |
| | log-$U(-50, 50)$ | 0.4444 | 0.0094 | 0.0678 | 0.0088 | 0.0489 | 0.0066 | 0.0334 |
| | log-$N(0, 1)$ | 0.4853 | -0.0078 | 0.0869 | -0.0065 | 0.0601 | -0.0020 | 0.0424 |
| | log-$N(-1, 50^2)$ | 0.4450 | 0.0081 | 0.0642 | 0.0079 | 0.0473 | 0.0070 | 0.0334 |
| | log-Cauchy$(0, 1)$ | 0.4819 | -0.0083 | 0.0814 | -0.0059 | 0.0560 | -0.0018 | 0.0411 |
| | log-$t(2)$ | 0.4838 | -0.0052 | 0.0794 | -0.0046 | 0.0576 | -0.0027 | 0.0401 |
| | log-$t(3)$ | 0.4845 | -0.0063 | 0.0810 | -0.0051 | 0.0581 | -0.0027 | 0.0412 |

Table 2: Estimation results of $\hat{\xi}_n(X,T)$ for DGP1 with $\sigma = 10$ and various $w(t)$

| | | | $n = 200$ | | $n = 400$ | | $n = 800$ | |
|---|---|---|---|---|---|---|---|---|
| $\rho$ | $w(t)$ | True | Bias | SD | Bias | SD | Bias | SD |
| 0 | $f_T(t)$ | 0.1107 | -0.0075 | 0.0909 | -0.0046 | 0.0625 | 0.0038 | 0.0437 |
| | Exp(1) | 0.1235 | -0.0160 | 0.1004 | -0.0082 | 0.0712 | -0.0064 | 0.0495 |
| | $\chi^2(3)$ | 0.1236 | -0.0136 | 0.0985 | -0.0088 | 0.0717 | -0.0056 | 0.0486 |
| | log-$U(-50, 50)$ | 0.1025 | 0.0041 | 0.0928 | 0.0026 | 0.0654 | 0.0015 | 0.0445 |
| | log-$N(0, 1)$ | 0.1237 | -0.0113 | 0.1024 | -0.0037 | 0.0759 | -0.0049 | 0.0525 |
| | log-$N(-1, 50^2)$ | 0.1028 | -0.0043 | 0.0925 | 0.0028 | 0.0665 | 0.0025 | 0.0454 |
| | log-Cauchy$(0, 1)$ | 0.1221 | -0.0101 | 0.0995 | -0.0073 | 0.0685 | -0.0036 | 0.0468 |
| | log-$t(2)$ | 0.1231 | -0.0115 | 0.1000 | -0.0083 | 0.0724 | -0.0036 | 0.0477 |
| | log-$t(3)$ | 0.1234 | -0.0130 | 0.0983 | -0.0087 | 0.0696 | -0.0044 | 0.0498 |
| 0.4 | $f_T(t)$ | 0.1668 | -0.0063 | 0.0869 | -0.0046 | 0.0627 | 0.0035 | 0.0434 |
| | Exp(1) | 0.1839 | -0.0115 | 0.1010 | -0.0043 | 0.0730 | -0.0039 | 0.0502 |
| | $\chi^2(3)$ | 0.1840 | -0.0107 | 0.0997 | -0.0049 | 0.0709 | -0.0037 | 0.0509 |
| | log-$U(-50, 50)$ | 0.1556 | 0.0094 | 0.0891 | 0.0060 | 0.0643 | 0.0037 | 0.0441 |
| | log-$N(0, 1)$ | 0.1841 | -0.0105 | 0.0989 | -0.0041 | 0.0695 | -0.0034 | 0.0503 |
| | log-$N(-1, 50^2)$ | 0.1561 | 0.0034 | 0.0897 | 0.0031 | 0.0668 | 0.0038 | 0.0456 |
| | log-Cauchy$(0, 1)$ | 0.1821 | -0.0088 | 0.0979 | -0.0046 | 0.0687 | -0.0039 | 0.0469 |
| | log-$t(2)$ | 0.1833 | -0.0083 | 0.1011 | -0.0067 | 0.0697 | 0.0012 | 0.0483 |
| | log-$t(3)$ | 0.1837 | -0.0087 | 0.1028 | -0.0064 | 0.0716 | -0.0018 | 0.0474 |
| 0.8 | $f_T(t)$ | 0.2133 | -0.0062 | 0.0859 | 0.0047 | 0.0604 | 0.0024 | 0.0409 |
| | Exp(1) | 0.2331 | -0.0080 | 0.0966 | -0.0034 | 0.0684 | -0.0013 | 0.0496 |
| | $\chi^2(3)$ | 0.2333 | -0.0078 | 0.0977 | -0.0048 | 0.0725 | -0.0033 | 0.0525 |
| | log-$U(-50, 50)$ | 0.1991 | 0.0094 | 0.0904 | 0.0052 | 0.0632 | 0.0051 | 0.0427 |
| | log-$N(0, 1)$ | 0.2334 | -0.0053 | 0.1023 | -0.0024 | 0.0687 | -0.0012 | 0.0492 |
| | log-$N(-1, 50^2)$ | 0.2002 | 0.0063 | 0.0894 | 0.0056 | 0.0618 | 0.0053 | 0.0442 |
| | log-Cauchy$(0, 1)$ | 0.1821 | -0.0085 | 0.0951 | -0.0042 | 0.0677 | -0.0032 | 0.0475 |
| | log-$t(2)$ | 0.2311 | -0.0078 | 0.0968 | -0.0036 | 0.0681 | -0.0016 | 0.0457 |
| | log-$t(3)$ | 0.2330 | -0.0082 | 0.0972 | -0.0060 | 0.0710 | -0.0038 | 0.0488 |

# 5 Real Data Analysis

We apply our proposed methods to analyse the AIDS Clinical Trials Group Study 175 dataset downloaded from the UCI Machine Learning Repository. The dataset was created to evaluate the efficacy and safety of various AIDS treatments for totally 2139 individuals. Each one represents a health record of a patient diagnosed with AIDS in the United States. In our analysis, we consider the time to failure or censoring as the event-time, which was observed, $\delta = 1$, for 24.35% of the patients. We focus on the covariates Age (age in years), Weight (weight in kg), Race (0=white, 1=non-white) and CD4 counts at baseline and (20+/−5) weeks (cd40, cd420).

Table 3: Estimation results of $\hat{\xi}_n(X,T)$ for Example 2 with various $w(t)$

| Model | $n$ | | $n=200$ Bias | SD | $n=400$ Bias | SD | $n=800$ Bias | SD |
|---|---|---|---|---|---|---|---|---|
| DGP2 | $f_T(t)$ | 0.2110 | -0.0085 | 0.0733 | -0.0039 | 0.0517 | -0.0019 | 0.0355 |
| | Exp(1) | 0.1784 | -0.0059 | 0.0862 | -0.0034 | 0.0603 | -0.0026 | 0.0410 |
| | $\chi^2(4)$ | 0.2186 | -0.0082 | 0.0734 | -0.0044 | 0.0521 | -0.0035 | 0.0370 |
| | log-$U(-50,50)$ | 0.1843 | -0.0063 | 0.0798 | -0.0014 | 0.0549 | 0.0006 | 0.0398 |
| | log-$N(0,1)$ | 0.1861 | -0.0058 | 0.0849 | -0.0036 | 0.0597 | -0.0031 | 0.0375 |
| | log-$N(-1,50^2)$ | 0.1891 | -0.0079 | 0.0835 | -0.0032 | 0.0556 | 0.0003 | 0.0421 |
| | log-Cauchy$(0,1)$ | 0.1809 | -0.0060 | 0.0854 | -0.0033 | 0.0597 | -0.0011 | 0.0407 |
| | log-$t(2)$ | 0.1858 | -0.0100 | 0.0808 | -0.0047 | 0.0565 | -0.0028 | 0.0421 |
| | log-$t(3)$ | 0.1860 | -0.0095 | 0.0800 | -0.0053 | 0.0593 | -0.0016 | 0.0402 |
| DGP3 | $f_T(t)$ | 0.4940 | -0.0068 | 0.0598 | -0.0053 | 0.0419 | -0.0027 | 0.0307 |
| | Exp(1) | 0.4500 | -0.0050 | 0.0703 | -0.0049 | 0.0517 | -0.0010 | 0.0340 |
| | $\chi^2(4)$ | 0.5053 | -0.0059 | 0.0657 | -0.0049 | 0.0456 | -0.0026 | 0.0315 |
| | log-$U(-50,50)$ | 0.4589 | -0.0074 | 0.0690 | -0.0054 | 0.0463 | -0.0028 | 0.0331 |
| | log-$N(0,1)$ | 0.4604 | -0.0060 | 0.0687 | -0.0027 | 0.0500 | 0.0006 | 0.0339 |
| | log-$N(-1,50^2)$ | 0.4665 | -0.0057 | 0.0670 | -0.0032 | 0.0488 | 0.0011 | 0.0332 |
| | log-Cauchy$(0,1)$ | 0.4600 | -0.0060 | 0.0683 | -0.0044 | 0.0486 | -0.0018 | 0.0344 |
| | log-$t(2)$ | 0.4583 | -0.0049 | 0.0691 | -0.0025 | 0.0477 | 0.0004 | 0.0325 |
| | log-$t(3)$ | 0.4590 | -0.0043 | 0.0693 | -0.0028 | 0.0485 | -0.0009 | 0.0326 |

Table 4: Rejection frequencies of the tests for DGP4 (Guassian)

| | | | The proposed permutation tests with $w(t)\sim$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| CR | $\lambda$ | $n$ | $U(-5,5)$ | $U(-50,50)$ | $N(0,2)$ | $N(-1,50)$ | logit$(0,2)$ | logit$(1,3)$ | $\text{Dcor}_{\text{IPCW}}$ |
| 45% | 0 | 200 | 0.052 | 0.052 | 0.054 | 0.056 | 0.054 | 0.048 | 0.060 |
| | | 400 | 0.044 | 0.044 | 0.038 | 0.048 | 0.040 | 0.040 | 0.054 |
| | 0.3 | 200 | 0.138 | 0.110 | 0.140 | 0.124 | 0.148 | 0.148 | 0.330 |
| | | 400 | 0.216 | 0.200 | 0.214 | 0.208 | 0.226 | 0.218 | 0.460 |
| | 0.6 | 200 | 0.630 | 0.560 | 0.598 | 0.588 | 0.620 | 0.626 | 0.694 |
| | | 400 | 0.892 | 0.816 | 0.874 | 0.822 | 0.874 | 0.868 | 0.876 |
| | 0.9 | 200 | 0.952 | 0.922 | 0.920 | 0.926 | 0.950 | 0.952 | 0.868 |
| | | 400 | 1.000 | 0.996 | 0.996 | 0.994 | 1.000 | 1.000 | 0.962 |
| 65% | 0 | 200 | 0.060 | 0.054 | 0.046 | 0.054 | 0.048 | 0.056 | 0.036 |
| | | 400 | 0.042 | 0.046 | 0.054 | 0.052 | 0.054 | 0.040 | 0.040 |
| | 0.3 | 200 | 0.102 | 0.106 | 0.084 | 0.110 | 0.104 | 0.104 | 0.148 |
| | | 400 | 0.148 | 0.120 | 0.136 | 0.138 | 0.126 | 0.134 | 0.188 |
| | 0.6 | 200 | 0.322 | 0.264 | 0.248 | 0.264 | 0.274 | 0.274 | 0.360 |
| | | 400 | 0.498 | 0.444 | 0.412 | 0.450 | 0.476 | 0.482 | 0.524 |
| | 0.9 | 200 | 0.702 | 0.682 | 0.518 | 0.692 | 0.638 | 0.682 | 0.544 |
| | | 400 | 0.938 | 0.914 | 0.794 | 0.928 | 0.890 | 0.926 | 0.726 |

We first transform the event-time $Y$ by $\log(Y)$ for the ease of computation, and then we calculate the correlation coefficients and implement independence testing between $\log(Y)$ and covariates $X$. Clearly, $Y$ and $X$ are independent if and only if

Table 5: Rejection frequencies of the tests for DGP5 (sinusoid)

| | | | The proposed permutation tests with $w(t) \sim$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| CR | $\lambda$ | $n$ | $U(-5,5)$ | $U(-50,50)$ | $N(0,2)$ | $N(-1,50)$ | logit$(0,2)$ | logit$(1,3)$ | $\text{Dcor}_{\text{IPCW}}$ |
| 45% | 0 | 200 | 0.044 | 0.038 | 0.038 | 0.040 | 0.046 | 0.042 | 0.048 |
| | | 400 | 0.046 | 0.044 | 0.044 | 0.048 | 0.044 | 0.050 | 0.056 |
| | 0.2 | 200 | 0.348 | 0.308 | 0.352 | 0.330 | 0.338 | 0.344 | 0.144 |
| | | 400 | 0.458 | 0.438 | 0.492 | 0.454 | 0.488 | 0.478 | 0.216 |
| | 0.4 | 200 | 0.932 | 0.926 | 0.958 | 0.936 | 0.946 | 0.950 | 0.142 |
| | | 400 | 0.996 | 0.996 | 0.998 | 0.996 | 0.998 | 0.996 | 0.368 |
| | 0.6 | 200 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.118 |
| | | 400 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.230 |
| 65% | 0 | 200 | 0.052 | 0.040 | 0.040 | 0.046 | 0.046 | 0.044 | 0.028 |
| | | 400 | 0.054 | 0.048 | 0.054 | 0.052 | 0.054 | 0.050 | 0.054 |
| | 0.2 | 200 | 0.118 | 0.110 | 0.126 | 0.122 | 0.128 | 0.118 | 0.070 |
| | | 400 | 0.128 | 0.124 | 0.156 | 0.132 | 0.158 | 0.158 | 0.096 |
| | 0.4 | 200 | 0.344 | 0.340 | 0.410 | 0.362 | 0.380 | 0.378 | 0.086 |
| | | 400 | 0.568 | 0.560 | 0.638 | 0.572 | 0.596 | 0.602 | 0.110 |
| | 0.6 | 200 | 0.718 | 0.718 | 0.752 | 0.736 | 0.746 | 0.756 | 0.106 |
| | | 400 | 0.952 | 0.948 | 0.968 | 0.960 | 0.958 | 0.960 | 0.094 |

Table 6: Rejection frequencies of the tests for DGP6 (W-shape)

| | | | The proposed permutation tests with $w(t) \sim$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| CR | $\lambda$ | $n$ | $U(-5,5)$ | $U(-50,50)$ | $N(0,2)$ | $N(-1,50)$ | logit$(0,2)$ | logit$(1,3)$ | $\text{Dcor}_{\text{IPCW}}$ |
| 45% | 0 | 200 | 0.042 | 0.042 | 0.048 | 0.050 | 0.046 | 0.050 | 0.042 |
| | | 400 | 0.042 | 0.048 | 0.048 | 0.052 | 0.044 | 0.038 | 0.048 |
| | 0.2 | 200 | 0.184 | 0.170 | 0.204 | 0.170 | 0.212 | 0.178 | 0.070 |
| | | 400 | 0.324 | 0.292 | 0.342 | 0.292 | 0.328 | 0.344 | 0.066 |
| | 0.4 | 200 | 0.766 | 0.716 | 0.670 | 0.726 | 0.716 | 0.768 | 0.096 |
| | | 400 | 0.966 | 0.948 | 0.928 | 0.946 | 0.954 | 0.966 | 0.170 |
| | 0.6 | 200 | 1.000 | 0.986 | 0.920 | 0.986 | 0.984 | 0.992 | 0.100 |
| | | 400 | 1.000 | 1.000 | 0.998 | 1.000 | 1.000 | 1.000 | 0.236 |
| 65% | 0 | 200 | 0.054 | 0.052 | 0.054 | 0.056 | 0.060 | 0.058 | 0.050 |
| | | 400 | 0.050 | 0.050 | 0.056 | 0.052 | 0.058 | 0.058 | 0.048 |
| | 0.2 | 200 | 0.094 | 0.088 | 0.090 | 0.092 | 0.098 | 0.100 | 0.052 |
| | | 400 | 0.130 | 0.128 | 0.152 | 0.128 | 0.140 | 0.138 | 0.076 |
| | 0.4 | 200 | 0.308 | 0.288 | 0.334 | 0.310 | 0.334 | 0.338 | 0.072 |
| | | 400 | 0.472 | 0.442 | 0.492 | 0.456 | 0.466 | 0.476 | 0.074 |
| | 0.6 | 200 | 0.752 | 0.678 | 0.712 | 0.688 | 0.732 | 0.752 | 0.082 |
| | | 400 | 0.956 | 0.928 | 0.936 | 0.922 | 0.946 | 0.950 | 0.082 |

$\log(Y)$ and $X$ are independent. Moreover, according to Section 3.4, the logarithm transformation would not essentially influence the correlation coefficient between the event-time and the covariates. In addition, by utilizing the mean ($\hat{\mu} = 6.684$) and standard deviation ($\hat{\sigma} = 0.514$) of $\log(Y)$, it is reasonable to incorporate this informa-

Table 7: Rejection frequencies of the tests for DGP7

| | | | The proposed permutation tests with $w(t) \sim$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| CR | $\lambda$ | $n$ | $U(-5,5)$ | $U(-50,50)$ | $N(0,2)$ | $N(-1,50)$ | logis$(0,2)$ | logis$(1,3)$ | $\text{Dcor}_{\text{IPCW}}$ |
| 45% | 0 | 200 | 0.052 | 0.038 | 0.052 | 0.052 | 0.052 | 0.048 | 0.046 |
| | | 400 | 0.034 | 0.044 | 0.040 | 0.046 | 0.048 | 0.048 | 0.062 |
| | 0.3 | 200 | 0.392 | 0.286 | 0.388 | 0.292 | 0.390 | 0.382 | 0.222 |
| | | 400 | 0.628 | 0.454 | 0.646 | 0.456 | 0.638 | 0.626 | 0.334 |
| | 0.6 | 200 | 0.982 | 0.912 | 0.986 | 0.922 | 0.980 | 0.982 | 0.454 |
| | | 400 | 1.000 | 0.996 | 1.000 | 0.996 | 1.000 | 1.000 | 0.652 |
| | 0.9 | 200 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.630 |
| | | 400 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.744 |
| 65% | 0 | 200 | 0.056 | 0.050 | 0.056 | 0.056 | 0.054 | 0.060 | 0.040 |
| | | 400 | 0.048 | 0.034 | 0.052 | 0.044 | 0.046 | 0.042 | 0.044 |
| | 0.3 | 200 | 0.140 | 0.110 | 0.134 | 0.122 | 0.132 | 0.140 | 0.110 |
| | | 400 | 0.234 | 0.170 | 0.220 | 0.172 | 0.212 | 0.218 | 0.182 |
| | 0.6 | 200 | 0.472 | 0.358 | 0.412 | 0.370 | 0.454 | 0.460 | 0.182 |
| | | 400 | 0.720 | 0.588 | 0.680 | 0.598 | 0.716 | 0.730 | 0.308 |
| | 0.9 | 200 | 0.638 | 0.606 | 0.526 | 0.624 | 0.632 | 0.682 | 0.312 |
| | | 400 | 0.912 | 0.884 | 0.812 | 0.898 | 0.906 | 0.936 | 0.486 |

Table 8: Run times (in sec) for permutation tests of independence over 500 replications, $\lambda = 0.3$, $\rho = 0.4$ in DGP4

| | | The proposed permutation tests with $w(t) \sim$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CR | $n$ | $U(-5,5)$ | $U(-50,50)$ | $N(0,2)$ | $N(-1,50)$ | logis$(0,2)$ | logis$(1,3)$ | $\text{Dcor}_{\text{IPCW}}$ |
| 10% | 100 | 3.156 | 4.711 | 1.076 | 2.214 | 1.088 | 1.366 | 0.121 |
| | 500 | 3.408 | 5.082 | 1.166 | 2.388 | 1.172 | 1.504 | 3.707 |
| | 1000 | 3.726 | 5.650 | 1.309 | 2.665 | 1.315 | 1.668 | 18.512 |
| | 2000 | 4.329 | 6.609 | 1.554 | 3.121 | 1.556 | 1.962 | 70.307 |
| | 5000 | 5.954 | 9.102 | 2.196 | 4.335 | 2.193 | 2.714 | 648.279 |
| | 10000 | 8.977 | 13.935 | 3.345 | 6.610 | 3.347 | 4.178 | $>$ 40mins |
| 30% | 100 | 3.125 | 4.725 | 1.064 | 2.188 | 1.064 | 1.347 | 0.073 |
| | 500 | 3.342 | 5.077 | 1.168 | 2.384 | 1.169 | 1.502 | 2.061 |
| | 1000 | 3.601 | 5.528 | 1.278 | 2.604 | 1.298 | 1.661 | 10.532 |
| | 2000 | 4.224 | 6.651 | 1.550 | 3.129 | 1.547 | 1.952 | 47.041 |
| | 5000 | 5.865 | 9.272 | 2.212 | 4.402 | 2.215 | 2.757 | 391.068 |
| | 10000 | 8.639 | 13.868 | 3.346 | 6.602 | 3.340 | 4.162 | $>$ 25mins |

tion into the normal distribution to obtain an alternative version of $w(t)$ that takes into account prior knowledge on the distribution of the event-time.

The $p$-values and correlation coefficients for various tests of independence are displayed in Table 9. We observe that all the proposed tests for independence between Weight and event-time consistently yield $p$-values below 0.05, accompanied by large

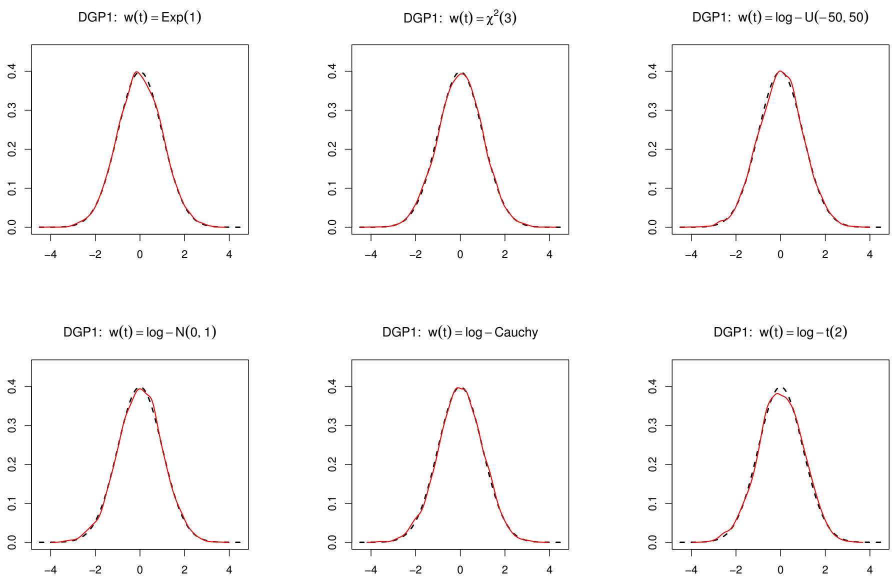


Figure 1: Density of $N(0,1)$ (dashed lines) and empirical densities (solid lines) of $\hat{\xi}_n(X,T)$ for different $w(t)$ in DPG1 with $\sigma = 5$, $\rho = 0.4$, $n = 1000$ and 5000 replications.

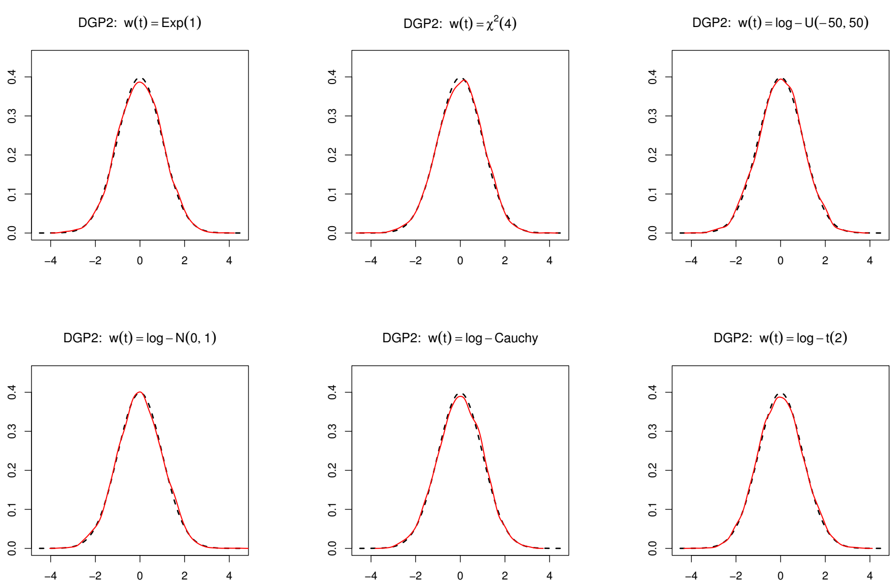


Figure 2: Density of $N(0,1)$ (dashed lines) and empirical densities (solid lines) of $\hat{\xi}_n(X,T)$ for different $w(t)$ in DPG2 with $n = 1000$ and 5000 replications.

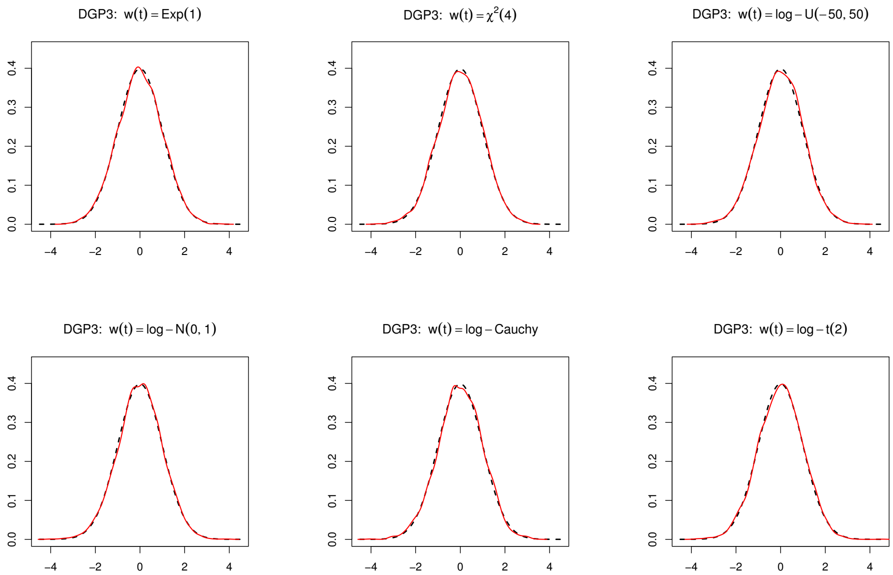


Figure 3: Density of $N(0,1)$ (dashed lines) and empirical densities (solid lines) of $\hat{\xi}_n(X,T)$ for different $w(t)$ in DPG3 with $n = 1000$ and 5000 replications.

values of correlation coefficients exceeding 0.56. This coincides with the cumulative hazard curses present in Figure 4, where the curves are not ordered by weights. In contrast, the distance correlation test $Dcor_{IPCW}$ suggests that there is no significant relationship between the event-time and the variable Weight. Furthermore, all the tests reject the null hypothesis and indicate that the event-time is associated with covariates (cd40, cd420), where changes in CD4 levels over time provide insight into the patient's immune status and response to treatment. At last, all the tests suggest the association between the event-time and age or race is not significant at level 0.05.

Table 9: Correlation coefficients (Corrs) and $p$-values obtained by various tests for AIDS Clinical Trials data

| Covariates | | The proposed permutation tests with $w(t) \sim$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | $U(0,10)$ | $N(\hat{\mu},\hat{\sigma}^2)$ | $N(-1,50)$ | logis$(0,2)$ | logis$(1,3)$ | Dcor$_{\mathrm{IPCW}}$ |
| Age | Corrs | 0.503 | 0.328 | 0.503 | 0.598 | 0.550 | 0.005 |
| | $p$-values | 0.232 | 0.295 | 0.251 | 0.210 | 0.218 | 0.156 |
| Weight | Corrs | 0.659 | 0.568 | 0.659 | 0.702 | 0.677 | -0.004 |
| | $p$-values | 0.014* | 0.030* | 0.020* | 0.008* | 0.012* | 0.608 |
| Race | Corrs | 0.021 | 0.007 | 0.027 | 0.027 | 0.025 | -0.001 |
| | $p$-values | 0.407 | 0.441 | 0.339 | 0.363 | 0.365 | 0.940 |
| cd40 | Corrs | 0.465 | 0.503 | 0.461 | 0.431 | 0.455 | 0.075 |
| | $p$-values | 0.202 | 0.128 | 0.168 | 0.251 | 0.218 | 0.002* |
| cd420 | Corrs | 0.506 | 0.558 | 0.505 | 0.472 | 0.491 | 0.160 |
| | $p$-values | 0.168 | 0.084 | 0.150 | 0.257 | 0.186 | 0.002* |
| (cd40, cd420) | Corrs | 0.332 | 0.332 | 0.337 | 0.312 | 0.326 | 0.152 |
| | $p$-values | 0.002* | 0.002* | 0.002* | 0.002* | 0.002* | 0.002* |

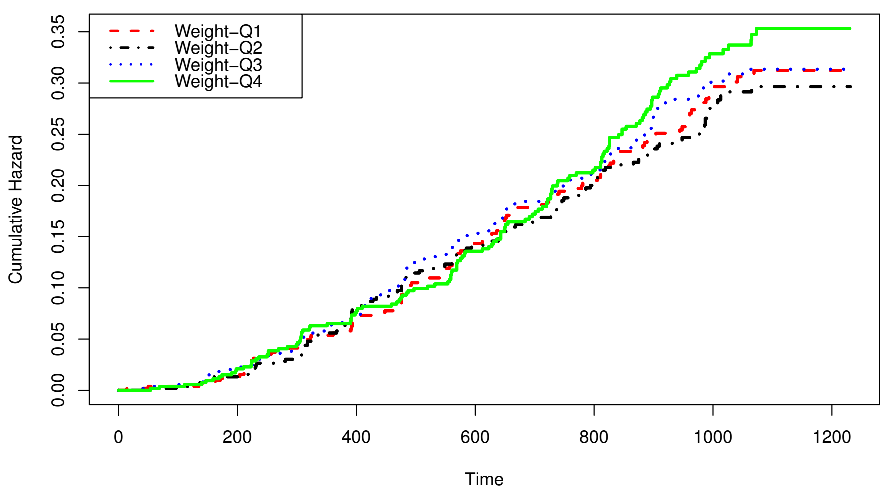


Figure 4: Nelson-Aalen estimates of the cumulative hazard curve for 25% (Weight-Q1), 50% (Weight-Q2), 75% (Weight-Q3) and 100% (Weight-Q4) quantiles of the variable Weight.

## Supplementary Material

The supplementary material contains all technical proofs.

# References


Azadkia, M., and Chatterjee, S. (2021), "A Simple Measure of Conditional Dependence", *The Annals of Statistics*, 49, 3070–3102.

Bhaskaran, K., dos-Santos-Silva, I., Leon, D. A., Douglas, I. J., and Smeeth, L. (2018), "Association of BMI with Overall and Cause-Specific Mortality: A Population-Based Cohort Study of 3.6 Million Adults in the UK", *The lancet Diabetes & endocrinology*, 6, 944–953.

Breiman, L., and Friedman, J. H. (1985), "Estimating Optimal Transformations for Multiple Regression and Correlation", *Journal of the American Statistical Association*, 80, 580–598.

Chatterjee, S. (2021), "A New Coefficient of Correlation", *Journal of the American Statistical Association*, 116, 2009–2022.

Chen, X., Chen, X., and Wang, H. (2018), "Robust Feature Screening for Ultra-high Dimensional Right Censored Data via Distance Correlation", *Computational Statistics & Data Analysis*, 119, 118–138.

Cox, D. R. (1972), "Regression Models and Life-Tables", *Journal of Royal Statistical Society, Series B*, 34, 187–220.

Deb, N., Ghosal, P., and Sen, B. (2020), "Measuring Association on Topological Spaces Using Kernels and Geometric Graphs", Available at arXiv:2010.01768.

Dette, H., Siburg, K. F., and Stoimenov, P. A. (2013), "A Copula-Based Nonparametric Measure of Regression Dependence", *Scandinavian Journal of Statistics*, 40, 21–41.

Edelmann, D., Hummel, M., Hielscher, T., Saadati, M., and Benner, A. (2020), "Marginal Variable Screening for Survival Endpoints", *Biometrical Journal*, 62, 610–626.

Edelmann, D., Welchowski, T., and Benner, A. (2022), "A Consistent Version of Distance Covariance for Right-Censored Survival Data and its Application in Hypothesis Testing", *Biometrics*, 78, 867–879.

Fernández, T., Gretton, A., Rindt, D., and Sejdinovic, D. (2023), "A Kernel Log-Rank Test of Independence for Right-Censored Data", *Journal of the American Statistical Association*, 118, 925–936.

Gray, R. J. (1992), "Flexible Methods for Analyzing Survival Data Using Splines, With Applications to Breast Cancer Prognosis", *Journal of the American Statistical Association*, 87, 942–951.

Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005), "Measuring Statistical Dependence With Hilbert–Schmidt Norms", In *International Conference on Algorithmic Learning Theory*, Berlin: Springer, pp. 63–77.

Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B., and Smola, A. (2008), "A Kernel Statistical Test of Independence", *Advances in Neural Information Processing Systems*, 585–592.

He, X., Wang, L., and Hong, H. G. (2013), "Quantile-Adaptive Model-Free Variable Screening for High-Dimensional Heterogeneous Data", *The Annals of Statistics*, 41, 342–369.

Le, C. T., Grambsch, P. M., and Louis, T. A. (1994), "Association Between Survival Time and Ordinal Covariates", *Biometrics*, 50, 213–219.

Li, J., Zheng, Q., Peng, L., and Huang, Z. (2016), "Survival Impact Index and Ultrahigh-Dimensional Model-Free Screening With Survival Outcomes", *Biometrics*, 72, 1145–1154.

Lin, Z., and Han, F. (2023), "On Boosting the Power of Chatterjee's Rank Correlation", *Biometrika*, 110, 283–299.

Lin, Z., and Han, F. (2022), "Limit Theorems of Chatterjee's Rank Correlation", Available at arXiv:2204.08031.

Ma, L., and Mao, J. (2019), "Fisher Exact Scanning for Dependency", *Journal of the American Statistical Association*, 114, 245–258.

McKeague, I. W., Nikabadze, A. M., and Sun, Y. Q. (1995), "An Omnibus Test for Independence of a Survival Time from a Covariate", *The Annals of Statistics*, 23, 450–475.

Pan, W., Wang, X., Xiao, W., and Zhu, H. (2019), "A Generic Sure Independence Screening Procedure", *Journal of the American Statistical Association*, 114, 928–937.

Peng, L., and Fine, J. P. (2008), "Nonparametric Tests for Continuous Covariate Effects With Multistate Survival Data", *Biometrics*, 64, 1080–1089.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., and Sabeti, P. (2011), "Detecting Novel Associations in Large Datasets", *Science*, 334, 1518–1524.

Rindt, D., Sejdinovic, D., and Steinsaltz, D. (2021), "A Kernel-And Optimal Transport-Based Test of Independence Between Covariates and Right Censored Lifetimes", *The International Journal of Biostatistics*, 17, 331–348.

Sharma, N., Chakrabarti, A., and Balas, V. E. (2019), "Data Management, Analytics and Innovation", *Proceedings of ICDMAI*, 1.

Song, R., Lu, W., Ma, S., and Jeng, X. J. (2014), "Censored Rank Independence Screening for High-Dimensional Survival Data", *Biometrika*, 101, 799–814.

Székely, G. J., and Rizzo, M. L. (2009), "Brownian Distance Covariance", *The Annals of Applied Statistics*, 3, 1236–1265.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007), "Measuring and Testing Dependence by Correlation of Distances", *The Annals of Statistics*, 35, 2769–2794.

Zajacova, A., and Burgard, S. A. (2012), "Shape of the BMI-Mortality Association by Cause of Death, Using Generalized Additive Models: NHIS 1986–2006", *Journal of Aging and Health*, 24, 191–211.

Zucker, D. M., and Karr, A. F. (1990), "Nonparametric Survival Analysis With Time-Dependent Covariate Effects: a Penalized Partial Likelihood Approach", *The Annals of Statistics*, 18, 329–353.