Scandinavian Journal of Statistics, Vol. 37: 680–700, 2010
doi: 10.1111/j.1467-9469.2010.00697.x

Inverse Probability of Censoring Weighted *U*-statistics for Right-Censored Data with an Application to Testing Hypotheses

SOMNATH DATTA

Department of Bioinformatics and Biostatistics, University of Louisville

DIPANKAR BANDYOPADHYAY

Division of Biostatistics and Epidemiology, Medical University of South Carolina

GLEN A. SATTEN

Division of Reproductive Health, Centers for Disease Control and Prevention

ABSTRACT. A right-censored version of a U-statistic with a kernel of degree $m \geq 1$ is introduced by the principle of a mean preserving reweighting scheme which is also applicable when the dependence between failure times and the censoring variable is explainable through observable covariates. Its asymptotic normality and an expression of its standard error are obtained through a martingale argument. We study the performances of our U-statistic by simulation and compare them with theoretical results. A doubly robust version of this reweighted U-statistic is also introduced to gain efficiency under correct models while preserving consistency in the face of model mis-specifications. Using a Kendall's kernel, we obtain a test statistic for testing homogeneity of failure times for multiple failure causes in a multiple decrement model. The performance of the proposed test is studied through simulations. Its usefulness is also illustrated by applying it to a real data set on graft-versus-host-disease.

Key words: doubly robust, inverse probability of censoring weighted, Kaplan–Meier, Kendall's τ, right-censoring, U-statistics

1. Introduction

U-statistics (Hoeffding, 1948) are fundamental objects in the theory and methods of statistics. They are generalizations of sample means and cover a whole range of estimators, estimating functions (for each value of the parameter) and test statistics. An even greater collection of estimators and test statistics are approximable by U-statistics or generalized U-statistics.

We start with a brief introduction of a U-statistic based on uncensored (complete) data. Consider a collection of independent random variables (vectors) $X_1, \ldots, X_n$ each with a common distribution function F on $\mathcal{X} (= \Re^p$ for some $p \geq 1)$. Let $h: \mathcal{X}^m \to \mathbb{R}$ be a measurable function, referred to as a kernel of degree (or order) $m (\leq n)$, that is symmetric in its m arguments and $Eh^2(X_1, \ldots, X_m) < \infty$. Let P_{nm} be the collection of ordered m indices out of $\{1, \ldots, n\}$; that is, $P_{nm} = \{\underline{i}: (i_1, \ldots, i_m) \in \mathbb{N}^m : 1 \leq i_1 < \ldots i_m \leq n\}$. A U-statistic based on the kernel h is obtained by averaging the summands $h(X_{i_1}, \ldots, X_{i_m}), \underline{i} \in P_{nm}$, as

$$U = \binom{n}{m}^{-1} \sum_{\underline{i} \in P_{nm}} h(X_{i_1}, \ldots, X_{i_m}). \tag{1}$$

Clearly, U is an unbiased estimator of the functional

$$\theta_F = Eh(X_1, \ldots, X_m) = \int_{\mathcal{X}^m} h(x_1, \ldots, x_m) \, \mathrm{d}F(x_1) \ldots \mathrm{d}F(x_m). \tag{2}$$

Asymptotic properties of these statistics can be found in Bickel & Lehmann (1979), Randles & Wolfe (1979), Serfling (1980), Sen (1981) and Lee (1990), among others.

The rest of the article is organized as follows. In section 2, we introduce our censored U-statistic under independent right censoring. We also extend its definition to the case of dependent right censoring when there are observable covariates that explain the dependence between the true failure times T_i^*, and the censoring times C_i. A doubly robust (DR) version of the censored U-statistic is also proposed to improve efficiency and preserve consistency under certain model mis-specifications. Section 3 presents results from a number of simulation studies with varying kernels to validate our large-sample theory results. In section 4, we apply our approach, specialized to a Kendall's τ statistic for right-censored data, to obtain a test of homogeneity in a multiple decrement model. The test is also applied on a real data set. The main body of the article concludes with a discussion in section 5. We defer the proof of asymptotic normality of the censored U-statistic to the Appendix.

2. IPCW U-statistics for right censored data

In survival analysis or analysis of time-to-event data, individuals are often subject to right censoring. Under the generally accepted framework of random (or independent) censoring (one minus), the Kaplan–Meier estimator provides a consistent estimator of the common marginal distribution function F. Therefore, one approach to extending U-statistics to right-censored data (under random or independent censoring) is to integrate (average) the kernel h with respect to the product of Kaplan–Meier estimators of the failure time distribution. Such statistics are generally referred to as the Kaplan–Meier U-statistics (Akritas, 1986; Gijbels & Veraverbeke, 1991; Stute & Wang, 1993a,b; Stute, 1995; Bose & Sen, 1999, 2002). Thus far, asymptotic normality of a Kaplan–Meier U-statistic of order two has been established in the literature (Bose & Sen, 2002), which is given by

$$\hat{U} = \frac{\sum_{1\leq i<j\leq n} h(T_i, T_j) W_i W_j}{\sum_{1\leq i<j\leq n} W_i W_j}, \tag{3}$$

where W_i is the mass assigned to the ith right-censored failure time T_i $(=T_i^* \wedge C_i)$ by a Kaplan–Meier estimator. As can be seen from Bose & Sen (2002), one faces extensive asymptotic calculations with this form of censored U-statistic.

In this article, we propose an inverse probability of censoring weighted (IPCW) approach to define U-statistics for right-censored data, where each summand calculated using uncensored data is reweighted using the inverse of the survival function of the censoring random variable. Using results of Satten & Datta (2001) (also see Datta, 2005) we can show that numerically it is the same as a Kaplan–Meier U-statistic for continuous data when the largest observation is not a censored observation [or it is given the remaining mass in (3) if it is a censored observation]. However, there are several advantages of the IPCW approach. The resulting censored U-statistic has a relatively simple martingale representation which allows demonstration of asymptotic normality as well as giving a simple variance formula for a general kernel of arbitrary order (whereas existing asymptotic normality results for the Kaplan–Meier U-statistic are available only for a second-order kernel). It is also simple to extend it for dependent censoring explainable by observed covariables. With this approach (see section 2.1), one can also define U-statistics that include additional random variables X as well as the censored failure times T. Section 4 provides an example of this. Kaplan–Meier faces a difficulty in this case if, for some tied observed failure times T, the corresponding X values are not equal, as it is unclear to distribute the mass among the corresponding Xs.

The IPCW approach has been popularized for survival data through the work of Koul *et al.* (1981), Robins & Rotnitzky (1992), Robins (1993), Satten *et al.* (2001), Satten & Datta (2001, 2004), Rotnitzky & Robins (2005) and so on. Finally, we note that Schisterman & Rotnitzky (2001) have considered reweighted U-statistics for missing data. Although our approach results in statistics that appear similar, it is not a special case of their work as they do not consider survival data. The coarsening mechanisms for missing data and right-censored data are not related. Consequently, the formulas for variance estimates and the mathematical tools and details for establishing the theoretical properties are vastly different.

2.1. Independent censoring

First, for the simplicity of exposition, we assume independent right censoring leading to i.i.d. data $\{T_i = T_i^* \wedge C_i,\ \delta_i = I(T_i^* \le C_i),\ X_i = X_i^* \delta_i\}$, $1 \le i \le n$. Here T^* are the true failure times and X^* are the variables of interest which are used to calculate the U-statistic and the product $X_i^* \delta_i$ needs to be interpreted as a scaler multiplication if X_i^* is vector-valued. Most often X^* will equal T^* but the generality allowed in this formulation could be useful for certain applications (see, e.g. section 4; also see Yao *et al.*, 1998). Note that we only observe T^* and X^* for those individuals with $\delta = 1$. Throughout this subsection, we assume the following independent censoring model

$$\begin{aligned}
&\lim_{dt \to 0} \frac{P\{t \le C < t + dt \mid T \ge t, T^*, X^*\}}{dt}, \\
&= \lim_{dt \to 0} \frac{P\{t \le C < t + dt \mid T \ge t\}}{dt} = \lambda_c(t),
\end{aligned} \tag{4}$$

where λ_c denotes the hazard rate for C, which is assumed to have an absolutely continuous distribution. An important special case of the independent censorship model is the so-called random censorship model under which censoring time C is assumed to be independent of (T^*, X^*). In the next subsection, we briefly describe how it can be extended to more general censoring models where the dependence between failure and censoring times are carried through a set of observed covariables.

Following the mean-preserving reweighting approach of Koul *et al.* (1981) or Datta (2005), we could define a censored U-statistic based on kernel h as:

$$U_c = \frac{1}{\binom{n}{m}} \sum_{\underline{i} \in P_{nm}} \frac{h(X_{i_1}, \ldots, X_{i_m}) \prod_{\ell \in \underline{i}} \delta_\ell}{\prod_{\ell \in \underline{i}} K_c(T_\ell -)}, \tag{5}$$

where the notation $\ell \in \underline{i}$ is used here and afterwards to indicate that ℓ is one of the integers $\{i_1, \ldots, i_m\}$, and K_c is the survival function of the censoring variable C. It is easy to see that U_c itself is a U-statistic based on the triplets (X_i, T_i, δ_i), $1 \le i \le n$; its kernel $\mathcal{H}$ is of order m and is given by

$$\mathcal{H}((X_{i_1}, T_{i_1}, \delta_{i_1}), \ldots, (X_{i_m}, T_{i_m}, \delta_{i_m})) = \frac{h(X_{i_1}, \ldots, X_{i_m}) \prod_{\ell \in \underline{i}} \delta_\ell}{\prod_{\ell \in \underline{i}} K_c(T_\ell -)}. \tag{6}$$

As the distribution of censoring time is continuous, $K_c(t-) = K_c(t)$; however, we prefer to use the left limit because it leads to an appropriate estimator to deal with tied data in practice. It is rudimentary to check that, under the random censorship model, U_c is mean preserving as

$$E\left(\frac{h(X_{i_1},\dots,X_{i_m})\prod_{\ell\in\underline{i}}\delta_\ell}{\prod_{\ell\in\underline{i}}K_c(T_{\ell}-)}\right)=E\left\{\frac{h(X^*_{i_1},\dots,X^*_{i_m})}{\prod_{j=1}^m K_c(T^*_{i_j}-)}\prod_{j=1}^m E\left(I(T^*_{i_j}\le C_{i_j})\,\Big|\,T^*_{i_j},X^*_{i_j}\right)\right\}$$
$$=E\left(\frac{h(X^*_{i_1},\dots,X^*_{i_m})}{\prod_{j=1}^m K_c(T^*_{i_j}-)}\prod_{j=1}^m K_c(T^*_{i_j}-)\right)$$
$$=E(h(X^*_{i_1},\dots,X^*_{i_m}))=\theta, \tag{7}$$

provided $K_c(T_i^*)>0$, for each i, with probability 1. For general independent censoring, a martingale argument as in Satten *et al.* (2001) can be used to show the mean preserving property of U_c.

As K_c is unknown, it needs to be estimated by the Kaplan–Meier estimator, where the role of censored and failed observations are reversed. Substituting this estimator $\hat{K}_c$ into (5) we get the following statistic that is calculable from right-censored data

$$\hat{U}=\frac{1}{\binom{n}{m}}\sum_{\underline{i}\in P_{nm}}\frac{h(X_{i_1},\dots,X_{i_m})\prod_{\ell\in\underline{i}}\delta_\ell}{\prod_{\ell\in\underline{i}}\hat{K}_c(T_{\ell}-)}. \tag{8}$$

We refer to this statistic an IPCW U-statistic based on censored data although, like a Kaplan–Meier U-statistic, it is technically not a U-statistic as $\hat{K}_c$ uses data from all individuals.

The weighted average form makes the asymptotic analysis of the U-statistic for right-censored data possible. For our theoretical analysis, we introduce the following counting process notation. Let $N_i^c(t)=I(T_i\le t,\delta_i=0)$ be the counting process of censoring for the ith individual and $M_i^c(t)=N_i^c(t)-\int_0^t Y_i(u)\lambda_c(u)\,\mathrm{d}u$ be the associated martingale, where λ_c is the censoring hazard and $Y_i(t)=I(T_i\ge t)$. Let $N^c=\sum_{i=1}^n N_i^c$ and $Y=\sum_{i=1}^n Y_i$. Also, let $\bar{n}$ be the subdistribution function of the pair (X^*,T^*) corresponding to $\delta=1$, that is,

$$\bar{n}(x,t)=P(X_1\le x,T_1\le t,\delta_1=1),\quad x\in\mathcal{X},\ \ t\ge 0.$$

Let

$$y(t)=P(T_1\ge t)$$

and

$$w(t)=\frac{1}{y(t)}\int_{\mathcal{X}\times[0,\infty)}\frac{h_1(x)}{K_c(u-)}I(u>t)\,\mathrm{d}\bar{n}(x,u),\quad t\ge 0, \tag{9}$$

where

$$h_1(x)=E(h(x,X_2^*,\dots,X_m^*)\,|\,X_1^*=x),\quad x\in\mathcal{X}.$$

Assume that

$$Eh^2(X_1^*,\dots,X_m^*)<\infty,$$

$$\int_{\mathcal{X}\times[0,\infty)}\frac{h_1^2(x)}{K_c^2(u-)}\,\mathrm{d}\bar{n}(x,u)<\infty\ \text{ and }\int_{[0,\infty)}w^2(t)\lambda_c(t)\,\mathrm{d}t<\infty.$$

With these assumptions, we can establish theorem 1 which governs the asymptotic distribution of an IPCW U-statistic. A proof of the theorem is given in appendix A.1.

Theorem 1. *As* $n \to \infty$,

$$\sqrt{n}(\hat{U} - \theta) \xrightarrow{d} N(0, \sigma^2),$$

where

$$\sigma^2 = m^2 \mathrm{var}\left(\frac{h_1(X_1)\delta_1}{K_c(T_1-)} + \int w(t)\,\mathrm{d}M_1^c(t)\right). \tag{10}$$

Certainly, the theorem is of interest when $\sigma^2 > 0$. We do not pursue the so-called degenerate case in this article. An estimate of σ^2 is given by (see appendix A.2)

$$\hat{\sigma}^2 = \frac{m^2}{n-1}\sum_{i=1}^{n}(V_i - \bar{V})^2, \tag{11}$$

where

$$V_i = \frac{\hat{h}_1(X_i)\delta_i}{\hat{K}_c(T_i-)} + \hat{w}(T_i)(1-\delta_i) - \sum_{j=1}^{n}\frac{\hat{w}(T_j)I(T_i \geq T_j)(1-\delta_j)}{Y(T_j)},$$

and the formulas for $\hat{h}_1$ and $\hat{w}$ are given in appendix A.2.

2.2. Dependent censoring

In many applications, one may observe failure time data along with a collection of covariables (fixed or time varying) $\boldsymbol{Z} = \{\boldsymbol{Z}(t) : t \geq 0\}$, which carry the dependence between the failure and censoring times. We assume that the $\boldsymbol{Z}$ are observable for all individuals (including those whose failures are censored) at least up to the time T. If covariables $\boldsymbol{Z}$ explain all the dependence between the failure and censoring times, then the following technical condition on the hazard for censoring can be imposed:

$$\begin{aligned}&\lim_{\mathrm{d}t \to 0}\frac{P\{t \leq C < t + \mathrm{d}t \mid \boldsymbol{Z}(u), 0 \leq u < t, T \geq t, T^*, X^*\}}{\mathrm{d}t}\\ &= \lim_{\mathrm{d}t \to 0}\frac{P\{t \leq C < t + \mathrm{d}t \mid \boldsymbol{Z}(u), 0 \leq u < t, T \geq t\}}{\mathrm{d}t}.\end{aligned} \tag{12}$$

We denote this censoring hazard by $\lambda_c(t) = \lambda_c(t \mid \bar{Z}(t))$, where $\bar{\boldsymbol{Z}}(t) = \sigma(\boldsymbol{Z}(u), 0 \leq u < t)$, and denote the corresponding integrated hazard by $\Lambda_c(t)$, that is $\Lambda_c(t) = \int_0^t \lambda_c(s)\,\mathrm{d}s$. Under this setting, K_c can be defined as $K_c(t) = \exp\{-\Lambda_c(t)\}$, which technically does not have a survival interpretation in general. Also, for simplicity of exposition, we will assume that $\{T_i^*, X_i^*, C_i, \mathbf{Z}_i\}$ are i.i.d. With additional technical conditions, this assumption can be relaxed.

A flexible hazard model increases the chance of obtaining an estimate of K_c that is close to its true value. One such model is Aalen's (1980, 1989) linear hazard model where the censoring hazard is expressed as:

$$\lambda_c(t) = \sum_{k=0}^{p}\beta_k(t)W_k(t), \tag{13}$$

where the $W_k(t) = \phi_k(Z(u), 0 \leq u < t)$ are predictable functions of the time-dependent covariates, each $\beta_k(t)$ is an unknown function, and where we assume that the first component $W_0(t) \equiv 1$. We let $B_k(t) = \int_0^t \beta_k(s)\,\mathrm{d}s$ and let $W_i(t)$ be the vector $(W_{i0}(t), W_{i1}(t), \ldots, W_{ip}(t))^T$ corresponding to the ith subject; then Aalen's estimator of the vector $\boldsymbol{B}(t) = (B_0(t), \ldots, B_p(t))^T$ is given by

$$\hat{\boldsymbol{B}}(t) = \sum_{i=1}^{n} I(T_i \le t)(1-\delta_i)\boldsymbol{A}^{-1}(t)\boldsymbol{W}_i(t), \tag{14}$$

where the matrix $\boldsymbol{A}(t)$ is given by

$$\boldsymbol{A}(t) = \sum_{i=1}^{n} I(T_i \ge t)\boldsymbol{W}_i(t)\boldsymbol{W}_i^T(t). \tag{15}$$

For subject i, this leads to

$$\begin{aligned}\hat{\Lambda}_c^i(t) &= \sum_{k=0}^{p} \int_0^t W_{ik}(t)\,\mathrm{d}\hat{\boldsymbol{B}}_k(t)\\ &= \sum_{j=1}^{n} I(T_j \le t)(1-\delta_j)\boldsymbol{W}_i^T(T_j)\hat{\boldsymbol{A}}^{-1}(T_j)\boldsymbol{W}_j(T_j), \quad t \le T_i.\end{aligned} \tag{16}$$

The formula for the censored U-statistic (8) remains intact; the only change needed is to use $\hat{K}_c(T_i-) = \exp\{-\hat{\Lambda}_c^i(T_i-)\}$, which will be based on the subject-specific covariate history. We will denote it by $\hat{U}_{\mathrm{mod}}$ to indicate that its definition has been modified to incorporate dependent right censoring. A similar approach was taken in Satten *et al.* (2001) for extending the Kaplan–Meier in the case of dependent censoring. It can be checked using a martingale argument similar to that in Satten *et al.* (2001) that U_{mod} defined in this way (using K_c) remains unbiased for θ even if T and C are not independent but (12) holds.

A martingale representation leading to the asymptotic normality of $\hat{U}_{\mathrm{mod}}$ can be established along the lines of Datta & Satten (2002). As before, $M_i^c(t) = N_i^c(t) - \int_0^t I[T_i \ge u]\,\mathrm{d}\Lambda_c^i(u)$ is the subject-specific zero mean martingale associated with the counting process of censoring times $N_i^c(t) = I[T_i \le t, \delta_i = 0]$, $1 \le i \le n$ where, by (12), the σ-algebra can be taken to be $\mathcal{F}_{i,t}^c = \sigma(\{N_i^c(s), Z_i(s), 0 \le s \le t; T_i^*, X_i^*\})$. Let $\mathbb{M}^c(t)$ be a vector with ith component given by $M_i^c(t)$ and let $\tilde{\mathbb{W}}(t)$ be a matrix with ith column equalling $\tilde{\boldsymbol{W}}_i(t) = I(T_i \ge t)\boldsymbol{W}_i(t)$. Assume that the following quantities exist (and are based on finite elements) for each $t \ge 0$:

$$\gamma^T(t) = E\frac{h_1(X_1)\delta_1 I(T_1 > t)\boldsymbol{W}_1^T(t)}{K_c(T_1-)},$$

$$\mathbf{a}(t) = E\tilde{\boldsymbol{W}}_1(t)\tilde{\boldsymbol{W}}_1^T(t),$$

$$\sigma_2^2 = \operatorname*{plim}_{n\to\infty} n^{-1}\int_0^\infty \boldsymbol{\Delta}_n^T(t)\mathrm{diag}\Big(I[T_1 \ge t]\lambda_c^1(t), \ldots, I[T_n \ge t]\lambda_c^n(t)\Big)\Delta_n(t)\,\mathrm{d}t,$$

with

$$\boldsymbol{\Delta}_n^T(t) = \boldsymbol{\eta}^T(t) - \gamma^T(t)\,\mathbf{a}(t)\tilde{\mathbb{W}}(t),$$

and

$$\boldsymbol{\eta}^T(t) = \left(\frac{h_1(X_1^*)}{K_c(t)}, \ldots, \frac{h_1(X_n^*)}{K_c(t)}\right),$$

where $\operatorname*{plim}_{n\to\infty}$ stands for in probability limit.

We also need technical conditions for the weak convergence of $\hat{B}_k$ obtained from Aalen's linear model; see, for example, theorem VII.4.1 of Andersen *et al.* (1993). The asymptotic representation in theorem 2 provides the asymptotic distribution of the modified IPCW U-statistic using Aalen's weights under dependent censoring that is explained by covariates.

Theorem 2. *As* $n \to \infty$,

$$\sqrt{n}(\hat{U}_{\text{mod}} - \theta) = \frac{m}{\sqrt{n}} \sum_{i=1}^{n} \left(h_1(X_i^*) - \theta\right) + \frac{m}{\sqrt{n}} \int_0^\infty \boldsymbol{\Delta}^T(t)\,\mathrm{d}\mathbb{M}^c(t) + o_p(1),$$

and hence

$$\sqrt{n}(\hat{U}_{\text{mod}} - \theta) \xrightarrow{d} N(0, \sigma^2_{\text{mod}}),$$

where $\sigma^2_{\text{mod}} = m^2\{\mathrm{var}(\mathrm{h}_1(\mathrm{X}_1^*)) + \sigma_2^2\}$.

Although it is not immediately obvious, it is possible to see from the proof of theorem 2 (appendix A.3) that the martingale representation in theorem 2 agrees with that in theorem 1 when the censoring hazard does not depend on $\boldsymbol{Z}$.

2.3. *A doubly-robust censored U-statistic*

Generally speaking, although estimators based on inverse probability of censoring are asymptotically unbiased, they may not be efficient (i.e. not achieving the lower bound on variance in the class of regular estimators) under various parametric and semi-parametric models. To achieve a reduction in variance, we propose a doubly robust (DR) version of the IPCW U-statistic that involves additional modelling (i.e. that of the failure times given the covariates). For more on DR, see Robins *et al.* (1999; section 7) and the monograph by van der Laan & Robins (2003). Taking the cue from estimation of a mean, for example, a kernel with $m=1$ (see, e.g. Rotnitzky & Robins, 2005), we propose a DR estimator as:

$$\begin{aligned}
U_{\text{DR}} = \frac{1}{\binom{n}{m}} \sum_{P_{nm}} \Bigg\{ & \frac{h(X_{i_1}, \ldots, X_{i_m}) \prod_{\ell \in \underline{i}} \delta_\ell}{\prod_{\ell \in \underline{i}} K_c(T_\ell -)} \\
& + \sum_{\ell \in \underline{i}} \int_0^\infty \frac{E\left(h(X_{i_1}^*, \ldots, X_{i_m}^*) \mid \bar{Z}_\ell(u_\ell), T_\ell^* \geq u_\ell\right)}{K_c(u_\ell)} \,\mathrm{d}M_\ell^c(u_\ell) \\
& - \sum_{\ell_1 < \ell_2 \in \underline{i}} \int_0^\infty \int_0^\infty \frac{1}{K_c(u_{\ell_1}) K_c(u_{\ell_2})} \\
& \times E(h(X_{i_1}^*, \ldots, X_{i_m}^*) \mid \bar{\boldsymbol{Z}}_{\ell_1}(u_{\ell_1}), T_{\ell_1}^* \geq u_{\ell_1}, \bar{\boldsymbol{Z}}_{\ell_2}(u_{\ell_2}), T_{\ell_2}^* \geq u_{\ell_2}) \,\mathrm{d}M_{\ell_1}^c(u_{\ell_1}) \,\mathrm{d}M_{\ell_2}^c(u_{\ell_2}) \\
& + \cdots + \cdots - \frac{(-1)^m}{\binom{n}{m}} \sum_{P_{nm}} \int_0^\infty \cdots \int_0^\infty \frac{1}{\prod_{\ell \in i} K_c(u_\ell)} \\
& \times E(h(X_{i_1}^*, \ldots, X_{i_m}^*) \mid \bar{\boldsymbol{Z}}_{i_1}(u_{i_1}), T_{i_1}^* \geq u_{i_1}, \ldots, \bar{\boldsymbol{Z}}_{i_m}(u_{i_m}), T_{i_m}^* \geq u_{i_m}) \prod_{\ell \in \underline{i}} \mathrm{d}M_\ell^c(u_\ell) \Bigg\}.
\end{aligned} \tag{17}$$

Using (A.5) in Satten *et al.* (2001), we get the following representations:

$$U_{\mathrm{DR}} = U^* + \frac{1}{\binom{n}{m}} \sum_{P_{nm}} \Bigg[- \sum_{\ell \in \underline{i}} \int_0^\infty \left\{ \frac{h(X_{i_1}^*, \ldots, X_{i_m}^*) - E(h(X_{i_1}^*, \ldots, X_{i_m}^*) | \bar{\boldsymbol{Z}}_\ell(u_\ell), T_\ell^* \geq u_\ell)}{K_c(u_\ell)} \right\} \mathrm{d}M_\ell^c(u_\ell)$$
$$+ \sum_{\ell_1 < \ell_2 \in \underline{i}} \int_0^\infty \int_0^\infty \frac{1}{K_c(u_{\ell_1}) K_c(u_{\ell_2})}$$
$$\times \left\{ h(X_{i_1}^*, \ldots, X_{i_m}^*) - E\left(h(X_{i_1}^*, \ldots, X_{i_m}^*) \,|\, \bar{\boldsymbol{Z}}_{\ell_1}(u_{\ell_1}), T_{\ell_1}^* \geq u_{\ell_1}, \bar{\boldsymbol{Z}}_{\ell_2}(u_{\ell_2}), T_{\ell_2}^* \geq u_{\ell_2} \right) \right\}$$
$$\times \mathrm{d}M_{\ell_1}^c(u_{\ell_1}) \mathrm{d}M_{\ell_2}^c(u_{\ell_2}) + \cdots + (-1)^m \int_0^\infty \cdots \int_0^\infty \frac{1}{\prod_{\ell \in i} K_c(u_\ell)}$$
$$\times \left\{ h(X_{i_1}^*, \ldots, X_{i_m}^*) - E\left(h(X_{i_1}^*, \ldots, X_{i_m}^*) \,|\, \bar{\boldsymbol{Z}}_{i_1}(u_{i_1}), X_{i_1}^* \geq u_{i_1}, \ldots, \bar{\boldsymbol{Z}}_{i_m}(u_{i_m}), X_{i_m}^* \geq u_{i_m} \right) \right\}$$
$$\times \prod_{\ell \in i} \mathrm{d}M_\ell^c(u_\ell) \Bigg], \tag{18}$$

and

$$U_{\mathrm{mod}} = U^* + \frac{1}{\binom{n}{m}} \sum_{P_{nm}} \Bigg\{ - \sum_{\ell \in \underline{i}} \int_0^\infty \frac{h(X_{i_1}^*, \ldots, X_{i_m}^*)}{K_c(u_\ell)} \mathrm{d}M_\ell^c(u_\ell)$$
$$+ \sum_{\ell_1 < \ell_2 \in \underline{i}} \int_0^\infty \int_0^\infty \frac{h(X_{i_1}^*, \ldots, X_{i_m}^*)}{K_c(u_{\ell_1}) K_c(u_{\ell_2})} \mathrm{d}M_{\ell_1}^c(u_{\ell_1}) \mathrm{d}M_{\ell_2}^c(u_{\ell_2})$$
$$+ \cdots + (-1)^m \int_0^\infty \cdots \int_0^\infty \frac{h(X_{i_1}^*, \ldots, X_{i_m}^*)}{\prod_{\ell \in \underline{i}} K_c(u_\ell)} \prod_{\ell \in \underline{i}} \mathrm{d}M_\ell^c(u_\ell) \Bigg\}, \tag{19}$$

where U^* is the U-statistic based on complete data and U_{mod} is given by (5) with $K_c(t) = \exp\{-\Lambda_c(t)\}$ defined using (12).

This estimator is called DR because $E(U_{\mathrm{DR}}) = \theta$, if either (i) the model for λ_c is correctly specified or (ii) the model for (X^*, T^*) given $\bar{\boldsymbol{Z}}(\cdot)$ is correctly specified. Furthermore, when both models are correct

$$A - \mathrm{var}(\sqrt{n} U_{DR})$$
$$= m^2 \left[\mathrm{var}(h_1(X_1^*)) + E \int_0^\infty \frac{E\left\{ h_1(X_1^*) - E\left(h(X_1^*) \,|\, \bar{\boldsymbol{Z}}_1(u_1), T_1^* \geq u_1 \right) \right\}^2}{K_c^2(u_1)} I(T_1 \geq u_1) \lambda_c^1(u_1) \mathrm{d}u_1 \right]$$
$$\leq m^2 \left\{ \mathrm{var}(h_1(X_1^*)) + E \int_0^\infty \frac{E\left(h_1(X_1^*) \right)^2}{K_c^2(u_1)} I(T_1 \geq u_1) \lambda_c^1(u_1) \mathrm{d}u_1 \right\}$$
$$= A - \mathrm{var}(\sqrt{n} U_{\mathrm{mod}}).$$

Once again, in practice, we need to substitute estimators of K_c and $E(h(X^*_{i_1},\ldots,X^*_{i_m})\mid \bar{Z}_{i_1}(u_{i_1}), T^*_{i_1}\geq u_{i_1},\ldots,\bar{Z}_{i_r}(u_{i_r}), T^*_{i_r}\geq u_{i_r})$, for $1\leq r\leq m$, in (17). This makes the final estimator fairly complex and computationally challenging. A full investigation of the resulting estimator is beyond the scope of this article. However, see the following and also the next section for an illustrative example.

As an example, consider a kernel h of order 2 and assume that the dependence between failure times $X^*=T^*$ and the censoring times C is explained by a fixed binary covariate $Z=G$. One can calculate the necessary quantities by fitting a semi-parametric regression model, such as the Cox model, to the failure as well as the censoring hazards. In this case, a semi-parametrically estimated kernel of the DR U-statistic is given by

$$
\begin{aligned}
\hat{h}_{\mathrm{DR}}(i,j)=&\frac{h(X_i,X_j)\delta_i\delta_j}{\hat{K}_{c,G_i}(X_i-)\hat{K}_{c,G_j}(X_j-)}+\left[\frac{1}{\hat{K}_{c,G_j}(X_j)}\left\{\sum_{\tau_k\geq X_j}\hat{h}_1(\tau_k)\frac{\Delta\hat{S}_{G_j}(\tau_k)}{\hat{S}_{G_j}(X_j)}\right\}(1-\delta_j)\right.\\
&\left.-\sum_{\tau\leq X_j}\frac{1}{\hat{K}_{c,G_j}(\tau_l)}\left\{\sum_{\tau_k\geq\tau_l}\hat{h}_1(\tau_k)\frac{\Delta\hat{S}_{G_j}(\tau_k)}{\hat{S}_{G_j}(\tau_l)}\right\}\Delta\hat{\Lambda}_{c,G_j}(\tau_l)\right]\\
&+\left[\frac{1}{\hat{K}_{c,G_i}(X_i)}\left\{\sum_{\tau_k\geq X_i}\hat{h}_1(\tau_k)\frac{\Delta\hat{S}_{G_i}(\tau_k)}{\hat{S}_{G_i}(X_i)}\right\}(1-\delta_i)\right.\\
&\left.-\sum_{\tau\leq X_i}\frac{1}{\hat{K}_{c,G_i}(\tau_l)}\left\{\sum_{\tau_k\geq\tau_l}\hat{h}_1(\tau_k)\frac{\Delta\hat{S}_{G_i}(\tau_k)}{\Delta\hat{S}_{G_i}(\tau_l)}\right\}\hat{\Lambda}_{c,G_i}(\tau_l)\right]\\
&-\left[\frac{1}{\hat{K}_{c,G_i}(X_i)\hat{K}_{c,G_j}(X_j)}\left\{\sum_{\tau_k\geq X_i,\tau_l\geq X_j}h(\tau_k,\tau_l)\frac{\Delta\hat{S}_{G_i}(\tau_k)}{\hat{S}_{G_i}(X_i)}\frac{\Delta\hat{S}_{G_i}(\tau_l)}{\hat{S}_{G_j}(X_j)}\right\}(1-\delta_i)(1-\delta_j)\right.\\
&-\frac{1}{\hat{K}_{c,G_i}(X_i)}\left\{\sum_{\tau_l\leq X_j}\frac{1}{\hat{K}_{c,G_j}(\tau_l)}\left(\sum_{\tau_k\geq X_i,\tau_{l'}\geq\tau_l}h(\tau_k,\tau_{l'})\right.\right.\\
&\left.\left.\times\frac{\Delta\hat{S}_{G_i}(\tau_k)}{\hat{S}_{G_i}(X_i)}\frac{\Delta\hat{S}_{G_j}(\tau_{l'})}{\hat{S}_{G_j}(\tau_l)}\right)\Delta\hat{\Lambda}_{c,G_i}(\tau_l)\right\}(1-\delta_i)\\
&-\frac{1}{\hat{K}_{c,G_j}(X_j)}\left\{\sum_{\tau_k\leq X_i}\frac{1}{\hat{K}_{c,G_i}(\tau_k)}\left(\sum_{\tau_l\geq X_j,\tau_{k'}\geq\tau_k}h(\tau_{k'},\tau_l)\right.\right.\\
&\left.\left.\times\frac{\Delta\hat{S}_{G_j}(\tau_l)}{\hat{S}_{G_j}(X_j)}\frac{\Delta\hat{S}_{G_i}(\tau_{k'})}{\hat{S}_{G_i}(\tau_k)}\right)\Delta\hat{\Lambda}_{c,G_j}(\tau_k)\right\}(1-\delta_j)\\
&+\sum_{\tau_k\leq X_i,\tau_l\leq X_j}\frac{1}{\hat{K}_{c,G_i}(\tau_k)\hat{K}_{c,G_j}(\tau_l)}\left\{\sum_{\tau_{k'}\geq\tau_k,\tau_{l'}\geq\tau_l}h(\tau_{k'},\tau_{l'})\frac{\Delta\hat{S}_{G_i}(\tau_{k'})}{\hat{S}_{G_i}(\tau_k)}\frac{\Delta\hat{S}_{G_j}(\tau_{l'})}{\hat{S}_{G_j}(\tau_l)}\right\}\\
&\left.\times\Delta\hat{\Lambda}_{c,G_i}(\tau_k)\Delta\hat{\Lambda}_{c,G_j}(\tau_l)\right],
\end{aligned}
\tag{20}
$$

where $\hat{h}_1$ is as in theorem 2; τ_k are the ordered event times in the pooled sample; G_i is the group for the ith observation; $\hat{S}_G$ is an estimator for failure times group G and $\Delta\hat{S}_G(\tau)$ is the size of the jump of $\hat{S}_G$ at time τ; $\Delta\hat{\Lambda}_{c,G}(\tau)$ is the jump of an estimate of censoring hazard in group G at time τ and $\hat{K}_{c,G}$ is the estimator of the survival function of the censoring times in group G. In our simulation example (see the next section), we use the Cox model to estimate these quantities.

It is interesting to note that, for the case of a single fixed binary covariate, Aalen's model is indeed a non-parametric model and the resulting estimated correction term for the DR estimator will be zero.

3. Simulation studies

To validate the performance of the IPCW U-statistic (8) and its asymptotic variance formulas in moderate to large samples, we conducted simulation studies based on right-censored data with two different choices of the kernel. Another simulation is used to study the various versions of the U-statistic when dependent censoring is present.

Example 1. We consider a third-order kernel (Heffernan, 1997) given by $h(x_1, x_2, x_3) = (2x_1 - x_2 - x_3)(-x_1 + 2x_2 - x_3)(-x_1 - x_2 + 2x_3)/6$. The corresponding U-statistic estimates the third central moment of $F, \mu_3(F) = \int (x-\mu)^3 \mathrm{d}F(x)$, where F is the distribution of T^* and μ is the population mean. We generate T^* from a Weibull distribution with the shape parameter 2.3 and scale parameter 2.0. The censoring times C were generated from a Weibull distribution with shape 1.5 and scale 4, leading to 25 per cent censoring, or a Weibull with shape 0.3 and scale 5.0, leading to 50 per cent censoring, respectively. We choose sample sizes $n = 50$, 200 and 1000 and used 5000 replicate datasets for each sample size. The theoretical third central moment of T^* was 0.2479, which is compared with the mean of our IPCW U-statistic. We also compare the empirical variance of the censored U-statistic with the estimated asymptotic variance obtained using (11). Finally, we also study the performance of the 95 per cent asymptotic confidence intervals for μ_3, $\hat{U} \pm 1.96(\hat{\sigma}_{\hat{U}})/\sqrt{n}$. All these are reported in Table 1.

We can conclude from Table 1 that the empirical answers do match their population counterparts and the performance of the censored U-statistic and its variance formula appears to be satisfactory. It should be noted, however, that large sample sizes are needed for the asymptotic and empirical results to agree, especially, the coverage of the confidence interval, when there is heavy censoring. This is not unexpected as the third central moment is very dependent on tail behaviour of the distribution which is not well estimated when censoring is present. At a given sample size, the performance worsens with higher censoring rate as expected.

Example 2. Next we consider the kernel $h(x_1, x_2) = I(\log(x_1 x_2) \leq 0)$; $x_1, x_2 > 0$. The corresponding U-statistic estimates

$$\theta(F) = P\{\log(X_1^*) + \log(X_2^*) \leq 0\}.$$

In the full data case, the U-statistic calculates the average number of pairs (X_i^*, X_j^*) for which $\log(X_i^*) + \log(X_j^*) \leq 0$, and can be used as a test for investigating whether the median of distribution of the observations $\log(X^*)$ is located at zero. This statistic is called a one-sample Wilcoxon statistic and is equivalent to the signed-rank statistic (both being based on log-transformed failure times). In this example, we generate $X^* = T^*$ from a log-normal distribution with parameters 0 and 1 and censoring times C from a log-normal distribution with parameters (0.9539, 1) and (0, 1) to represent about 25 per cent and 50 per cent censoring, respectively. As before, we choose sample sizes $n = 50$, 200 and 1000 and used 5000 replicate data sets for each sample size. The empirical average of our censored U-statistic is compared with the population $\theta = 0.5$ in Table 2. As before, we also compare the empirical variance of the censored U-statistic with the estimated asymptotic variance obtained using

Table 1. *Estimation of population third central moment μ_3 from censored data using a U-statistic with kernel $h(x_1, x_2, x_3) = (2x_1 - x_2 - x_3)(-x_1 + 2x_2 - x_3)(-x_1 - x_2 + 2x_3)/6$; the true population μ_3 is 0.25. The results in the table are based on 5000 replicates*

Sample size (n)	Censoring rate (%)	Average $\hat{\mu}_3$	Average estimated variance of $\hat{\mu}_3$ (times n)	Empirical variance of $\hat{\mu}_3$ (times n)	Empirical coverage of 95% CI
50	25	0.23	2.56	3.63	0.77
	50	0.21	3.17	4.84	0.73
200	25	0.24	4.16	4.43	0.88
	50	0.24	4.70	5.20	0.85
1000	25	0.25	4.43	4.49	0.93
	50	0.25	5.15	5.17	0.91

CI, confidence interval.

Table 2. *Wilcoxon one-sample statistics based on log-transformed censored data. The true θ is 0.5. The results in the table are based on 5000 replicates*

Sample size (n)	Censoring rate (%)	$\hat{\theta}$	Average estimated variance of $\hat{\theta}$ (times n)	Empirical variance of $\hat{\theta}$ (times n)	Empirical coverage of 95% CI
50	25	0.50	0.35	0.35	0.93
	50	0.49	0.50	0.52	0.93
200	25	0.50	0.36	0.37	0.95
	50	0.50	0.50	0.52	0.94
1000	25	0.50	0.34	0.36	0.95
	50	0.50	0.50	0.52	0.94

CI, confidence interval.

(11) and the empirical coverage of a 95 per cent confidence interval constructed using the large sample theory.

Once again, we conclude that the empirical answers match the theoretical counterparts reasonably well. Although the same general trend as in the previous example was observed, the censored U-statistic and the large sample methods seem to work very well for this example even for sample size as small as 50. This is possibly because of the choice of the kernel which is not sensitive to large values of its arguments.

Example 3. This example illustrates the performance of a censored U-statistic when dependent censoring is present through a covariate. We use the same kernel $h(x_1, x_2) = I(\log(x_1 x_2) \leq 0)$ as in example 2 and a binary covariate (a group indicator).

We generate equal numbers of samples in each of the two groups. The failure times T^* in group 1 were taken from an exponential distribution with rate parameter 1 whereas T^* in group 2 were taken from an exponential distribution with rate parameter 0.2. The censoring times in each group were independently generated using the same group-specific distributions leading to 50 per cent censoring. Note that censoring and failure times are only independent conditional on group membership. The true θ in this case is approximately 0.390.

In this situation, a naive U-statistic (8) that uses an overall Kaplan–Meier weight $\hat{K}_c$ will be inconsistent and asymptotically biased. We must use the modified form as described in

Table 3. *Comparison of bias and variance of four U-statistics based on the kernel $h(x_1, x_2) = I(\log(x_1 x_2) \le 0)$ in the presence of dependent censoring induced by a binary covariate. The failure distribution, as well as the censoring distribution in each group was exponential but with different rate parameters. The censoring rate was 50 per cent. The reported variances are multiplied by the sample size. The true θ is* 0.39

	U-statistics				
n	Full data (using uncensored data)	Naive (assumes independent censoring)	Modified (weights estimated using Aalen; non-parametric)	Modified (weights using estimated Cox; semi-parametric)	Doubly robust (uses Cox model for failure and censoring hazards)
Mean					
50	0.39	0.32	0.36	0.35	0.36
200	0.39	0.33	0.38	0.38	0.38
1000	0.39	0.33	0.38	0.39	0.39
Variance					
50	0.23	0.29	0.33	0.34	0.30
200	0.23	0.30	0.33	0.37	0.30
1000	0.24	0.30	0.33	0.41	0.31

section 2.2 for achieving consistency or asymptotic unbiasedness. One way to do this in this context is to use Kaplan–Meier of censoring inverse weights calculated separately for the two groups which is also equivalent to using the Aalen's linear model to estimate the censoring hazard using the group indicator as a covariate. In Table 3, we describe this as the modified *U*-statistic with non-parametric weights. A second modified *U*-statistic is also computed for this simulation using weights obtained from a Cox model for the censoring hazards and using $\hat{K}_{c,i} = \exp\{-\hat{\Lambda}_c^i\}$ from the model-based estimates. We also calculate a DR *U*-statistic by estimating the required quantities using the Cox model to both censoring as well as failure distributions with a binary covariate. Note that these are correct models for the previously simulated event times. As explained before, in this example, the correction term for a doubly-robust estimator will be zero if one attempts to use the Aalen's model to estimate the required quantities. We added the performance of the full (complete) *U*-statistic (based on T^*) in Table 3 as a benchmark.

In Table 3, we report the empirical mean and variance of these *U*-statistics for samples of size is $n = 50$, 200 and 1000, each based on 2000 replicate data sets. The results show that the naive *U*-statistic has the largest bias as expected. The performance of the modified *U*-statistic using the Cox weights was slightly worse than that using the Aalen's weights. The DR estimator had a low bias and a reduced variance compared with the modified *U*-statistic using the Cox weights.

4. An application to testing hypotheses

We apply our IPWC *U*-statistic asymptotic normality results to test the homogeneity of the conditional survival functions in a multiple decrement model with two failure types, thereby extending the test by Dewan *et al.* (2004) to the case of right-censored data. Consider a multiple decrement model where individuals fail at time T^* owing to one of the two possible causes $d^* = 1$ or 2. Let the conditional probability functions be given by

$$\Phi_j(t) = P(d^* = j \mid T^* > t) = \frac{S_j(t)}{S(t)}, \quad j = 1, 2, \ t \ge 0, \tag{21}$$

where $S_j(t)=P(T^*>t,\ d^*=j)$ is the sub-survival function for cause j and $S(t)=P(T^*>t)$ is the survival function owing to all causes. We consider the problem of testing the null hypothesis

$$H_0:\Phi_1(t)/\Phi_2(t)=\phi, \text{ a constant (in } t)$$

against an alternative

$$H_1:\Phi_1(t)/\Phi_2(t) \text{ is not a constant.}$$

Note that when H_0 holds, ϕ equals the ratio of probabilities of death owing to the two causes $P(d^*=1)/P(d^*=2)$.

Here we consider a test based on the concept of concordance and discordance (as in Kendall's τ), originally proposed by Dewan *et al.* (2004). Kendall's τ has become a cornerstone kernel of the U-statistic measuring deviations among concordances and discordances. Brown *et al.* (1974), Weier & Basu (1980) and Oakes (1982) proposed estimation under right censoring but none of the estimators were consistent when the true value of τ equals zero. Wang & Wells (2000) estimated τ using a suitable bivariate survival estimator into the integral form that defines τ. Betensky & Finkelstein (1999) considered estimation of τ in bivariate interval-censored data.

Let $X_i^*=(T_i^*,d_i^*), i=1,\dots,n$ be the failure time and failure cause data obtained from n individuals. A pair $(T_{i_1}^*,d_{i_2}^*)$ and $(T_{i_2}^*,d_{i_2}^*)$ is a concordant pair if $T_{i_1}^*>T_{i_2}^*, d_{i_1}^*=2, d_{i_2}^*=1$ or $T_{i_1}^*<T_{i_2}, d_{i_1}^*=1, d_{i_2}^*=2$, and is a discordant pair otherwise. Consider a U-statistic of bivariate observations (T^*,d^*) with kernel

$$\psi(X_{i_1}^*,X_{i_2}^*)=\begin{cases} 1, & \text{if } T_{i_1}^*>T_{i_2}^*, d_{i_1}^*=2, d_{i_2}^*=1 \text{ or } T_{i_1}^*<T_{i_2}^*,\ d_{i_1}^*=1, d_{i_2}^*=2 \\ -1, & \text{if } T_{i_1}^*>T_{i_2}^*, d_{i_1}^*=1, d_{i_2}^*=2 \text{ or } T_{i_1}^*<T_{i_2}^*,\ d_{i_1}^*=2, d_{i_2}^*=1 \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

Dewan *et al.* (2004) showed that, under H_0, the corresponding U-statistic is asymptotically normal with zero mean and estimated asymptotic variance $(4/3)\hat{\phi}(1-\hat{\phi})$, where $\hat{\phi}=n^{-1}\sum_{i=1}^n I(d_i^*=1)$ is the proportion of samples that fail as a result of cause 1.

In most real applications with multiple decrement data, however, the failure times are right-censored and the aforegiven test cannot be used. We can easily extend the definition of the test statistic to include right-censored failure times following our approach. As before, let C be the censoring time and let $\delta=I(T^*\le C)$. Then, (8) leads to the following test statistic based on observed data:

$$\hat{U}_\tau=\binom{n}{2}^{-1}\sum_{1\le i<j\le n}\psi(X_i,X_j)\frac{\delta_i\delta_j}{\hat{K}_c(T_i-)\hat{K}_c(T_j-)}, \tag{23}$$

where $\hat{K}_c$ is the Kaplan–Meier estimator of the survival function for censoring times.

By theorem 1, we see that under $H_0, n^{1/2}\hat{U}_\tau$ is asymptotically normal with zero mean and a variance that can be estimated by (11). We would reject the null hypothesis in favour of a two-sided alternative if $|n^{1/2}\hat{U}_\tau/\hat{\sigma}_{\hat{U}_\tau}|>z_{1-\alpha/2}$, where α is the nominal level and z are standard normal percentiles. To assess the large sample performance of this test for censored data, we conducted a simulation study using sample sizes of $n=200$, 500 and 1500 and a nominal level of $\alpha=0.05$. The power was empirically computed by the proportion of rejection out of 2000 Monte Carlo replications. Two levels of censoring were used along with full data test for which we have used the asymptotic formula of Dewan *et al.* (2004). The data generation scheme is described next.

4.1. *Simulating event times from a bivariate distribution*

Consider a unit being exposed to two risks and let the notional (or latent) lifetimes of the unit under the two risks be denoted by T^1 and T^2. In general, T^1 and T^2 are dependent and, being lifetimes, they should be non-negative. In this simulation setup, we only observe (T, δ), where $T = \min(T^1, T^2)$ is the failure time and $\delta = 2 - I(T^1 \leq T^2)$ is the cause of failure. We generate $(\tilde{T}^1, \tilde{T}^2)$ from a bivariate normal distribution with mean vector $(0, (1-a))$ and variance–covariance matrix $((1, \rho a), (\rho a, a^2))$, where $a = 1(0.1)1.5$. Finally, we let $T^1 = \exp(\tilde{T}^1)$ and $T^2 = \exp(\tilde{T}^2)$. The censoring times were generated from log-normal distributions with variance parameter 1 and mean parameters 0.954 and 0 leading to light (8–17%) and moderately heavy (28–45%) censoring, respectively.

4.2. *Results*

First, we report the empirical sizes for our test along with the full data test of Dewan *et al.* (2004). As can be seen from Table 4, the empirical sizes are marginally inflated in a few cases for heavy censoring and smaller sample sizes but overall, the asymptotic distribution seems to be reasonably effective.

Figure 1 displays an array of plots illustrating the power curves as a function of the alternative parameter a which was defined earlier. Note that $a = 1$ corresponds to the null hypothesis. For clarity of presentation we only show the results for sample sizes $n = 200$ and 500. In each figure, the solid curves correspond to the full data test. Also the corresponding curve for a larger sample size ($n = 500$) is above the curve for the lower sample size ($n = 200$) as expected. The performance of the censored data test is fairly close to the full data test when the censoring level is low and the sample size is large. The performance worsens with a higher rate of censoring as expected.

4.3. *A real data example*

A randomized, prospective study was conducted by the Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins between October 1997 and April 2002 to evaluate the antitumour activity of autologous graft-versus-host disease (GVHD) induced with cyclosporine, and augmented by the administration of γ-interferon and interleukin-2 in patients

Table 4. *Empirical sizes of our U-statistics-based test in a simulation experiment. The nominal size was* 0.05. *The estimates are based on* 2000 *replicates* (*SE* < 0.005)

		Censoring level		
ρ	n	Light	Moderately heavy	Uncensored[1]
0.5	200	0.067	0.070	0.052
	500	0.061	0.067	0.051
	1500	0.059	0.063	0.049
0	200	0.063	0.072	0.053
	500	0.058	0.070	0.052
	1500	0.048	0.055	0.051
−0.5	200	0.055	0.067	0.048
	500	0.045	0.058	0.043
	1500	0.047	0.049	0.052

[1]Dewan *et al.* (2004).

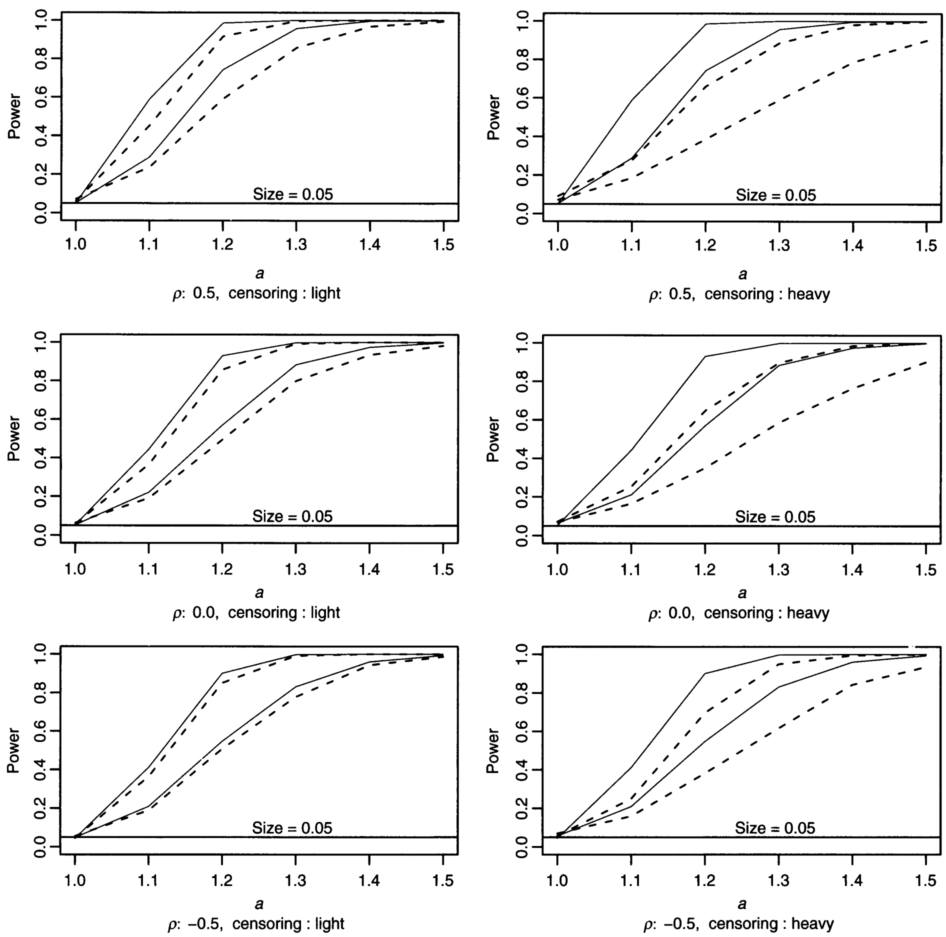

Fig. 1. Power curves of U-statistic-based tests in a simulation study. Curves for $n = 500$ are above the curves for $n = 200$. The solid curve corresponds to the benchmark test without censoring and the dashed curve corresponds to our U-statistic-based test for right-censored data.

with chemorefractory Hodgkin's and aggressive non-Hodgkin's lymphomas (Bolaños-Meade *et al.*, 2007). A total of 51 patients received treatments after bone marrow transplantation (BMT). Two types of events were observed namely, relapse (R) and treatment-related mortality (TRM). However, there is also right-censored data from persons who were still alive at the completion of the study and whose (eventual) cause of failure was not available. Of the 51 patients in the study, 33 (65%) experienced relapse, 8 (16%) experienced death as a result of TRM and the remaining 10 (19%) were right-censored. The survival times of relapse patients ranged from 0.37 to 62.97 months and that of the TRM patients from 0.9 to 6.7 months.

As indicated before, it was observed that patients who had TRM will die very soon after BMT whereas those dying of relapse usually survive a few months after BMT. We wanted to see if this empirical difference was statistically significant, in spite of the small sample size. We label TRM as cause 1 and relapse as cause 2 in the aforesaid formulation. The value of the test statistic U_τ (based on concordances and discordances) given in (23) was 0.11 with an estimated standard error of 0.051. Thus, the value of the Z-statistic was 2.14 which is significant at the 5 per cent level.

5. Discussion

U-statistics are standard tools in mathematical statistics. They have been extensively studied for i.i.d. complete (i.e. uncensored) data in the literature and are generally covered in graduate-level text books on non-parametric statistics. All previous attempts to generalize these statistics to the right-censored data take the approach of defining them as Kaplan–Meier integrals of the underlying kernel. By taking a reweighting approach, we have been able to establish the asymptotic distribution for an IPCW U-statistic with an arbitrary kernel. Further, we have extended these results to include situations where dependent censoring can be explained by measured covariates. Finally, we have considered the use of DR U-statistics to increase efficiency.

A motivating application for this work was a Kendall's test recently introduced by Dewan *et al.* (2004) with uncensored data. This test examines the homogeneity of failure times among various failure types in a multiple decrement model. As right censoring is often present with failure time data its practical utility would be enhanced if a right-censored version were available. We obtain this as an easy application to our general U-statistic methodology. It is anticipated that similar tests can be applied in a variety of other settings to compare two event times.

As mentioned before, a large number of statistics are approximate or generalized U-statistics and it is conceivable that the methods in this manuscript can be extended to a broader setting. Another possible direction for future research will be to develop versions of U-statistics for other types of censored data problems, notably current status and interval-censored data.

Acknowledgements

The first author's research was supported in part by a grant from the United States National Science Foundation. Bandyopadhyay acknowledges support of a University of Georgia Graduate School Dissertation Completion Fellowship. The authors thank Javier Bolaños-Meade, Georgia B. Vogelsang and Elizabeth Garrett-Mayer for generously providing the BMT data analysed in section 4. The authors thank two anonymous referees and the editors for their constructive suggestions. The findings and conclusions in this report are those of the authors, and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

References

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical statistics and probability, lecture notes on mathematical statistics and probability, 2* (eds W. Klonecki, A. Kozek & J. Rosiskii), 1–25. Springer, New York.

Aalen, O. O. (1989). A linear regression model for the analysis of lifetimes. *Stat. Med.* **8**, 907–925.

Akritas, M. B. (1986). Empirical processes associated with V-statistics and a class of estimators under random censoring. *Ann. Statist.* **14**, 619–637.

Andersen, P. K., Borgan, O., Gill, R. D. & Keiding, N. (1993). *Statistical models based on counting processes*. Springer, New York.

Betensky, R. A. & Finkelstein, D. M. (1999). An extension of Kendall's coefficient of concordance to bivariate interval censored data. *Stat. Med.* **18**, 3101–3109.

Bickel, P. J. & Lehmann, E. L. (1979). Descriptive statistics for nonparametric models IV. Spread. In *Contributions to statistics, Hajek memorial volume* (ed. J. Jurečková), 33–40. Academia, Prague.

Bolaños-Meade, J., Garrett-Mayer, E., Luznik, L., Anders, V., Webb, J., Fuchs, E., Huff, C., Matsui, W., Borrello, I. & Brodsky, R. (2007). Induction of autologous graft-versus-host disease: results of a randomized prospective clinical trial in patients with poor risk lymphoma. *Biol. Blood Marrow Transplant.* **13**, 1185–1191.

Bose, A. & Sen, A. (1999). The strong law of large numbers for Kaplan–Meier *U*-statistics. *J. Theoret. Probab.* **12**, 181–200.
Bose, A. & Sen, A. (2002). Asymptotic distribution of the Kaplan–Meier *U*-statistics. *J. Multivariate Anal.* **83**, 84–123.
Brown, B. W., Hollander, M. & Korwar, R. M. (1974). Nonparametric tests of independence for censored data, with applications to heart transplant studies. In *Reliability and biometry* (eds F. Proschan & R. J. Serfling), 327–354. SIAM, Philadelphia.
Datta, S. (2005). Estimating the mean life time using right censored data. *Statist. Methodol.* **2**, 65–69.
Datta, S. & Satten, G. A. (2002). Estimation of integrated transition hazards and stage occupation probabilities for non-Markov systems under stage dependent censoring. *Biometrics* **58**, 792–802.
Dewan, I., Deshpande, J. V. & Kulathinal, S. B. (2004). On testing dependence between time to failure and cause of failure via conditional probabilities. *Scand. J. Statist.* **31**, 79–91.
Gijbels, I. & Veraverbeke, N. (1991). Almost sure asymptotic representation for a class of functionals of the Kaplan–Meier estimator. *Ann. Statist.* **19**, 1457–1470.
Heffernan, P. M. (1997). Unbiased estimation of central moments by using *U*-statistics. *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* **59**, 861–863.
Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293–325.
Koul, H., Susarla, V. & Van Ryzin, J. (1981). Regression analysis of randomly right censored data. *Ann. Statist.* **9**, 1276–1288.
van der Laan, M. J. & Robins, J. M. (2003). *Unified methods for censored longitudinal data and causality*. Springer Verlag, New York.
Lee, A. J. (1990). *U-statistics, theory and practice*. Marcel Dekker, New York.
Oakes, D. (1982). A concordance test for independence in the presence of censoring. *Biometrics* **38**, 451–455.
Randles, R. H. & Wolfe, D. A. (1979). *Introduction to the theory of nonparametric statistics*. Wiley, New York.
Robins, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the American Statistical Association, Biopharmaceutical Section*, 24–33. ASA, Alexandria.
Robins, J. M. & Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS epidemiology – methodological issues* (eds N. Jewell, K. Dietz, & V. Farewell), 297–331. Birkhauser, Boston.
Robins, J. M., Rotnitzky, A. & van der Laan, M. J. (1999). Discussion of 'On Profile Likelihood' by S. A. Murphy and A. W. van der Vaart. *J. Amer. Statist. Assoc.* **95**, 477–482.
Rotnitzky, A. & Robins, J. (2005). Inverse probability weighted estimation in survival analysis. In *Encyclopedia of Biostatistics*, 2nd edn (eds P. Amitage & T. Colton), 2619–2625. Wiley, New York.
Satten, G. A. & Datta, S. (2001). The Kaplan–Meier Estimator as an inverse-probability-of-censoring weighted average. *Amer. Statist.* **55**, 207–210.
Satten, G. A. & Datta, S. (2004). Marginal analyses of multistage data. In *Advances in survival analysis, handbook of statistics*, 23 (eds N. Balakrishnan & C. R. Rao), 559–574. Elsevier, Amsterdam.
Satten, G. A., Datta, S. & Robins, J. M. (2001). An estimator for the survival function when data are subject to dependent censoring. *Statist. Probab. Lett.* **54**, 397–403.
Schisterman, E. & Rotnitzky, A. (2001). Estimation of the mean of a *K*-sample *U*-statistic with missing outcomes and auxiliaries. *Biometrika* **88**, 713–725.
Sen, P. K. (1981) *Sequential nonparametrics: invariance principles and statistical inference*. Wiley, New York.
Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. Wiley, New York.
Stute, W. (1995). The bias of Kaplan–Meier integrals. *Scand. J. Statist.* **21**, 475–484.
Stute, W. & Wang, J. L. (1993a). The strong law under random censorship. *Ann. Statist.* **21**, 1591–1607.
Stute, W. & Wang, J. L. (1993b). Multi-sample *U*-statistics for censored data. *Scand. J. Statist.* **20**, 369–374.
Wang, W. & Wells, M. T. (2000). Estimation of Kendall's under censoring. *Statist. Sinica* **10**, 1199–1215.
Weier, D. R. & Basu, A. P. (1980). An investigation of Kendall's tau modified for censored data with applications. *J. Statist. Plann. Inference* **4**, 381–390.
Yao, Q., Wei, L. J. & Hogan, J. (1998). Analysis of incomplete repeated measurements with dependent follow-up times. *Biometrika* **85**, 139–149.

Ying, Z. (1989). A note on the asymptotic properties of the product-limit estimator on the whole line. *Statist. Probab. Lett.* **7**, 311–314.

Received December 2008, in final form January 2010

Somnath Datta, Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY 40202, USA.
E-mail: somnath.datta@louisville.edu

Appendix

A.1. Proof of theorem 1

By repeated triangulation, we get

$$\sqrt{n}(\hat{U}-\theta)=\sqrt{n}(U-\theta)$$
$$-\sum_{j=1}^{m}\sqrt{n}\frac{1}{\binom{n}{m}}\sum_{\underline{i}}\frac{(\hat{K}_c(T_{i_j}-)-K_c(T_{i_j}-))h(X_{i_1},\ldots,X_{i_m})\prod_{\ell\in\underline{i}}\delta_\ell}{\hat{K}_c(T_{i_j}-)K_c(T_{i_j}-)\prod_{\ell<j}\hat{K}_c(T_{i_\ell}-)\prod_{\ell'>j}K_c(T_{i_{\ell'}}-)}. \tag{24}$$

Using weak convergence of $\sqrt{n}\left(\hat{K}_c(\cdot)-K_c(\cdot)\right)$, it is possible to replace the $\hat{K}_c$ in the denominator of (24) by its in probability limit K_c, together with an added error term that is no more than $o_p(1)$. For simplicity of presentation, we show this for the case of $m=1$. The arguments in the general case are similar. One can express the difference

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\hat{K}_c(T_i-)-K_c(T_i-))h(X_i)\delta_i}{\hat{K}_c(T_i-)K_c(T_i-)}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\hat{K}_c(T_i-)-K_c(T_i-))h(X_i)\,\delta_i}{K_c^2(T_i-)}$$

as $\int R_n(t)h(x)\,\mathrm{d}W_n(x,t)$, where W_n is the empirical sub-distribution of the pairs $\{X_i,T_i\}$ corresponding to $\delta_i=1$ and R_n converges weakly to the zero process which follows from the asymptotic theory for the Kaplan–Meier estimator (see, e.g. Ying, 1989). Therefore,

$$\left|\int R_n(t)h(x)\,\mathrm{d}W_n(x,t)\right|\leq\left(\int R_n^2(t)\,\mathrm{d}W_n(x,t)\right)^{1/2}\left(\int h_n^2(x)\,\mathrm{d}W_n(x,t)\right)^{1/2}$$
$$\leq\sup_{t\geq 0}|R_n^2(t)|\left(n^{-1}\sum_{i=1}^{n}h_n^2(X_i)\right)^{1/2}\xrightarrow{P}0.$$

Next note that for any u,

$$\sqrt{n}\left(\hat{K}_c(u-)-K_c(u-)\right)=\sqrt{n}\left(\mathrm{e}^{-\hat{\Lambda}_c(u-)}-\mathrm{e}^{-\Lambda_c(u-)}\right)+o_p(1),$$

where Λ_c is the cumulative censoring hazard and $\hat{\Lambda}_c$ is its Nelson–Aalen estimator for censoring,

$$=-\sqrt{n}\,K_c(u-)\left(\hat{\Lambda}_c(u-)-\Lambda_c(u-)\right)+o_p(1),$$

by the delta method.

Thus,

$$\sqrt{n}(\hat{U}-\theta)=\sqrt{n}(U-\theta)$$
$$+\sum_{j=1}^{m}\sqrt{n}\frac{1}{\binom{n}{m}}\sum_{\underline{i}}\frac{h(X_{i_1},\ldots,X_{i_m})\prod_{\ell\in\underline{i}}\delta_\ell}{\prod_{\ell\in\underline{i}}K_c(T_\ell-)}\left(\hat{\Lambda}_c(T_{i_j}-)-\Lambda_c(T_{i_j}-)\right)+o_p(1)\cdot \tag{25}$$

Recall that U is a U-statistic with kernel $\mathcal{H}$ given in (6). By projection arguments (and L_2 calculations) as in Hoeffding's (1948) decomposition (25) equals

$$\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{H}_1^c(X_i, T_i, \delta_i)+\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\frac{h_1(X_i)\delta_i}{K_c(T_i-)}\left(\hat{\Lambda}_c(T_i-)-\Lambda_c(T_i-)\right)+o_p(1), \tag{26}$$

where

$$\begin{aligned}\mathcal{H}_1^c(x_1, t_1, \delta_1)&=E(\mathcal{H}(X_1, T_1, \delta_1; X_2, T_2, \delta_2; \ldots; X_m, T_m, \delta_m)\,|\,X_1=x_1, T_1=t_1, \delta_1)-\theta,\\ &=h_1(x_1)\delta_1/K_c(t_1-)-\theta,\end{aligned}$$

with $h_1(x_1)=E(h(x_1, X_2^*, \ldots, X_m^*)\,|\,X_1^*=x_1)$. See, for example, Serfling (1980) for the first term. Note that the difference between the second terms in (25) and (26) is bounded by

$$m\left\{n^{-1}\sum_{i=1}^{n}\left(\frac{h_1(X_i)\delta_i}{K_c(T_i-)}-\frac{1}{\binom{n-1}{m-1}}\sum_{j_1<\cdots<j_{m-1}\in\{1,\ldots,n\}\backslash\{i\}}\frac{h(X_i, X_{j_1}\ldots, X_{j_{m-1}})\delta_i, \delta_{j_1}\ldots, \delta_{j_{m-1}}}{K_c(T_i-)K_c(T_{j_1}-)\cdots K_c(T_{j_{m-1}}-)}\right)^2\right\}^{1/2}$$

$$\times\sup_{t\geq 0}\sqrt{n}|\hat{\Lambda}_c(t)-\Lambda_c(t)|.$$

Now the second part is $O_p(1)$ as before and the first part is $o_p(1)$ as it converges to zero in L_2. To see this, note that

$$\begin{aligned}&En^{-1}\sum_{i=1}^{n}\left(\frac{h_1(X_i)\delta_i}{K_c(T_i-)}-\frac{1}{\binom{n-1}{m-1}}\sum_{j_1<\cdots<j_{m-1}\in\{1,\ldots,n\}\backslash\{i\}}\frac{h(X_i, X_{j_1}\ldots, X_{j_{m-1}})\delta_i, \delta_{j_1}\ldots, \delta_{j_{m-1}}}{K_c(T_i-)K_c(T_{j_1}-)\cdots K_c(T_{j_{m-1}}-)}\right)^2\\ &=E\left(\frac{h_1(X_1)\delta_1}{K_c(T_1-)}-\frac{1}{\binom{n-1}{m-1}}\sum_{j_1<\cdots<j_{m-1}\in\{2,\ldots,n\}}\frac{h(X_1, X_{j_1}\ldots, X_{j_{m-1}})\,\delta_1, \delta_{j_1}\ldots, \delta_{j_{m-1}}}{K_c(T_1-)K_c(T_{j_1}-)\cdots K_c(T_{j_{m-1}}-)}\right)^2\\ &=E\left[E\left\{\left(\frac{h_1(X_1)\delta_1}{K_c(T_1-)}-\frac{1}{\binom{n-1}{m-1}}\right.\right.\right.\\ &\qquad\left.\left.\left.\times\sum_{j_1<\cdots<j_{m-1}\in\{2,\cdots,n\}}\frac{h(X_1, X_{j_1}\ldots, X_{j_{m-1}})\delta_1, \delta_{j_1}\cdots, \delta_{j_{m-1}}}{K_c(T_1-)K_c(T_{j_1}-)\ldots K_c(T_{j_{m-1}}-)}\right)^2\,\middle|\,X_1, T_1, \delta_1\right\}\right].\end{aligned}$$

Now, given X_1, T_1, δ_1, the second term inside the parentheses is a U-statistic whose mean equals the first term. Therefore, this is $O(n^{-1})$.

The expression in (26) further equals

$$\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{H}_1^c(X_i, T_i, \delta_i)+\frac{m}{n}\sum_{i=1}^{n}\frac{h_1(X_i)\delta_i}{K_cT_i-)}\int_0^{T_i-}\frac{\mathrm{d}\{\bar{M}_n^c(s)\}}{y(s)}+o_p(1), \tag{27}$$

by martingale representation of the Nelson–Aalen estimator (see, e.g. Andersen *et al.*, 1993), where $M_i^c(t)=N_i^c(t)-\int_0^t Y_i(u)\,\mathrm{d}\Lambda_c(u)$ is the martingale of the censoring process defined with respect to the appropriate filtration, $N_i^c(t)=I(T_i\leq t, \delta_i=0)$ is the counting process of the censored data, $Y_i(t)=I(T_i\geq t)$ is the appropriate 'at-risk' process, $\bar{M}_n^c=n^{-1/2}\sum M_i^c$ and $y(t)=EY_i(t)$.

Next, note that the expression in (27) equals

$$\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{H}_1^c(X_i,T_i,\delta_i)+m\int_{\mathcal{X}\times[0,\infty)}\frac{h_1(x)}{K_c(u-)}\left\{\int_0^{u-}\frac{\mathrm{d}\left(\bar{M}_n^c(s)\right)}{y(s)}\right\}\mathrm{d}\bar{n}(x,u)+o_p(1), \tag{28}$$

by linearization of a U-statistic of degree 2 (see, e.g. Serfling, 1980).

Now, by Fubini's theorem, (28) can be expressed as:

$$\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{H}_1^c(X_i,T_i,\delta_i)+\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\infty}\left(\frac{1}{y(s)}\int_{\mathcal{X}\times[0,\infty)}I(u>s)\frac{h_1(x)}{K_c(u-)}\times\mathrm{d}\bar{n}(x,u)\right)\mathrm{d}M_i^c(s)+o_p(1),$$

$$=\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{h_1(X_i)\delta_i}{K(T_i-)}+\int_0^{\infty}w(s)\,\mathrm{d}M_i^c(s)-\theta\right)+o_p(1), \tag{29}$$

where w is given in (9).

Therefore, we have, as $n\to\infty$,

$$\sqrt{n}(\hat{U}-\theta)\xrightarrow{d}N(0,\sigma^2),$$

where

$$\sigma^2=m^2\,\mathrm{var}\left(\frac{h_1(T_1)\delta_1}{K_c(T_1-)}+\int w(s)\,\mathrm{d}M_1^c(s)\right).$$

This proves the theorem.

A.2. Estimation of σ^2

Note that

$$\int w(s)\,\mathrm{d}\hat{M}_i^c(s)=w(T_i)(1-\delta_i)-\int\frac{w(s)I(T_i\geq s)}{Y(s)}\mathrm{d}N^c(s)$$

$$=w(T_i)(1-\delta_i)-\int_0^{\infty}\frac{w(s)I(T_i\geq s)}{Y(s)}\,\mathrm{d}\left(\sum_{j=1}^{n}N_j^c(s)\right)$$

$$=w(T_i)(1-\delta_i)-\sum_{j=1}^{n}\int_0^{\infty}\frac{w(T_j)I(T_i\geq T_j)}{Y(T_j)}\,\mathrm{d}N_j^c(s)$$

$$=w(T_i)(1-\delta_i)-\sum_{j=1}^{n}\frac{w(T_j)I(T_i\geq T_j)(1-\delta_j)}{Y(T_j)}.$$

Using the reweighting principle, we define estimators of $h_1$and w by

$$\hat{h}_1(u)=n^{-(m-1)}\sum_{1\leq i_2,\ldots,i_m\leq n}h(u,X_{i_2},\ldots,X_{i_m})\frac{\delta_{i_2}\ldots\delta_{i_m}}{\hat{K}_c(T_{i_2}-)\ldots\hat{K}_c(T_{i_m}-)}$$

and

$$\hat{w}(s)=\frac{1}{Y(s)}\int\frac{\hat{h}_1(x)}{\hat{K}_c(u-)}I(u>s)\,\mathrm{d}\left(\sum I(X_i\leq x,T_i\leq u)\right),$$

$$=\frac{1}{Y(s)}\sum_{i=1}^{n}\frac{\hat{h}_1(X_i)\delta_i}{\hat{K}_c(T_i-)}I(T_i>s).$$

Finally, we can estimate σ^2 by

$$\hat{\sigma}^2=\frac{m^2}{n-1}\sum_{i=1}^{n}(V_i-\bar{V})^2,$$

where

$$V_i = \frac{\hat{h}_1(X_i)\delta_i}{\hat{K}_c(T_i-)} + \hat{w}(T_i)(1-\delta_i) - \sum_{j=1}^{n} \frac{\hat{w}(T_j)I(T_i \geq T_j)(1-\delta_j)}{Y(T_j)}.$$

A.3. Proof of theorem 2

We proceed with the triangulation step as in theorem 1 leading to

$$\sqrt{n}(\hat{U}_{\text{mod}} - \theta) = \frac{m}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{h_1(X_i)\delta_i}{K_c(T_i-)} - \theta \right)$$

$$+ \sum_{j=1}^{m} \sqrt{n} \frac{1}{\binom{n}{m}} \sum_{\underline{i}} \frac{h(X_{i_1}, \ldots, X_{i_m}) \prod_{\ell \in \underline{i}} \delta_\ell}{\prod_{\ell \in \underline{i}} K_c(T_\ell -)} \left(\hat{\Lambda}_c^i(T_{i_j}-) - \Lambda_c^i(T_{i_j}-) \right) + o_p(1). \tag{30}$$

Unlike theorem 1, we do not have a simple i.i.d. sum representation in the general case; instead, we resort to martingale arguments. To that end, we express the first term as:

$$\frac{m}{\sqrt{n}} \sum_{i=1}^{n} (h_1(X_i^*) - \theta) - \frac{m}{\sqrt{n}} \sum_{i=1}^{n} h_1(X_i^*) \int_0^\infty \frac{1}{K_c(u)} \mathrm{d}M_i^c(u), \tag{31}$$

using (A.5) of Satten *et al.* (2001).

The second term is (using 7.4.7 of Andersen *et al.*, 1993):

$$\sum_{j=1}^{m} \sqrt{n} \frac{1}{\binom{n}{m}} \sum_{\underline{i}} \frac{h(X_{i_1}, \ldots, X_{i_m}) \prod_{\ell \in i_0} \delta_\ell}{\prod_{\ell \in \underline{i}} K_c(T_\ell -)} \int_0^{T_{i_j}-} W_{i_j}^T(s) A^{-1}(s) \tilde{W}(s) \mathrm{d}\mathbb{M}^c(s)$$

$$= \frac{1}{\sqrt{n}} \int_0^\infty \sum_{j=1}^{m} \frac{1}{\binom{n}{m}} \sum_{\underline{i}} \frac{h(X_{i_1}, \ldots, X_{i_m}) \prod_{\ell \in i_0} \delta_\ell}{\prod_{\ell \in \underline{i}} K_c(T_\ell -)}$$

$$\times I(T_{i_j} > s) W_{i_j}^T(s) \left(n^{-1} A(s)\right)^{-1} \tilde{W}(s) \mathrm{d}\mathbb{M}^c(s),$$

$$= \frac{m}{\sqrt{n}} \int_0^\infty \gamma^T(s) \left(\mathbf{a}(s)\right)^{-1} \tilde{W}(s) \mathrm{d}\mathbb{M}^c(s) + o_p(1), \tag{32}$$

where

$$\gamma^{\mathrm{T}}(s) = \operatorname*{plim}_{n\to\infty} \frac{1}{\binom{n}{m}} \sum_{\underline{i}} \frac{m^{-1} \left(\sum_{j=1}^{m} I(T_{i_j} > s) W_{i_j}^T(s) \right) h(X_{i_1}, \ldots, X_{i_m}) \prod_{\ell \in \underline{i}} \delta_\ell}{\prod_{\ell \in \underline{i}} K_c(T_\ell -)}.$$

We now obtain the martingale representation by combining (30)–(32). The asymptotic normality follows from the martingale representation using the Rebolledo's central limit theorem (see, e.g. theorem II.5.1 in Andersen *et al.*, 1993), where the variance expression is obtained using proposition II.4.1 of Andersen *et al.* (1993).