Ann Inst Stat Math (2017) 69:271–302
DOI 10.1007/s10463-015-0537-6

CrossMark

U-statistics with conditional kernels for incomplete data models

Ao Yuan[1] · **Mihai Giurcanu**[2] · **George Luta**[1] · **Ming T. Tan**[1]

Received: 26 March 2014 / Revised: 18 June 2015 / Published online: 23 August 2015

Abstract For incomplete data models, the classical U-statistic estimator of a functional parameter of the underlying distribution cannot be computed directly since the data are not fully observed. To estimate such a functional parameter, we propose a U-statistic using a substitution estimator of the conditional kernel given the observed data. This kernel estimator is obtained by substituting the non-parametric maximum likelihood estimator for the underlying distribution function in the expression of the conditional kernel. We study the asymptotic properties of the proposed U-statistic for several incomplete data models, and in a simulation study, we assess the finite sample performance of the Mann–Whitney U-statistic with conditional kernel in the current status model. The analysis of a real-world data set illustrates the application of the proposed methods in practice.

Keywords U-statistics · Censored data · Incomplete data models · Non-parametric MLE

✉ Ao Yuan
ay312@georgetown.edu

Mihai Giurcanu
giurcanu@ufl.edu

George Luta
gl77@georgetown.edu

Ming T. Tan
mtt34@georgetown.edu

1 Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, 4000 Reservoir Road, NW, Washington, DC 20057, USA

2 Department of Statistics, University of Florida, McCarty C, Gainesville, Florida 32611, USA

1 Introduction

Since the pioneering work of Hoeffding (1948), the U-statistics have been actively studied and used in data analysis due to their wide range of applications. They are used for estimation of functional parameters involving multiple observations of the distribution of interest and they are closely related to the V-statistics proposed by von Mises (1947). Hoeffding (1961) established fundamental properties of the U-statistics, Berk (1966) discovered their reverse martingale property, and Sen (1974) made a number of contributions on this topic. Gregory (1977) obtained the asymptotic distribution of degenerate U-statistics of order two, result which was further generalized for degenerate U-statistics of arbitrary order by Janson (1979) and Rubin and Vitale (1980), among others. Borovskikh (1986) extended the U-statistics theory for Hilbert space valued parameters. A detailed review and major historical developments of U-statistics theory can be found in Korolyuk and Borovskikh (1994).

Censored data models with medical applications have been extensively studied by many researchers, such as Bennet (1983), Heagerty and Zheng (2005), Tsiatis (2006), and Therneau and Grambsch (2010), among others. Akritas (1986) studied a class of V-statistics under random censorship, Bose and Sen (1999, 2002) studied the asymptotic properties of U-statistics using the Kaplan–Meier estimator for right censored observations, Datta et al. (2010) introduced a U-statistic for right censored data via inverse probability weighting and studied its asymptotic properties, Schisterman and Rotnitzky (2001) studied the mean of a k-sample U-statistic with missing outcomes, and Tressou (2006) studied non-parametric modeling of left censored data in food risk assessment. More recently, Kowalski and Tu (2007) collected a number of results and modern applications of U-statistics including applications to missing data models, Hu and Degrutolla (2011) proposed a U-statistic method for longitudinal biomarkers, and Ma et al. (2010) investigated a U-statistic approach to modeling the Cronbach's alpha coefficient with missing data.

For incomplete data models, the classical U-statistic estimator of a functional parameter of the underlying distribution cannot be computed directly since the data are not fully observed. In this paper, we propose a U-statistic using partially observed data based on a substitution estimator of the conditional kernel. This kernel estimator is obtained by substituting the non-parametric maximum likelihood estimator of the underlying distribution function in the expression of the conditional kernel given the observed data. We study the asymptotic properties of the proposed U-statistic for several incomplete data models, such as the type I (current status) interval censoring model, the type II interval censoring model, the double censoring model, the convolution model, and the multiplicative censoring model, respectively. In a simulation study, we assess the finite sample performance of the non-parametric bootstrap (Efron 1979) in conjunction with the Mann–Whitney U-statistic with conditional kernel in the current status model, and the analysis of a real-world data set illustrates the application of the proposed methods in practice. We have created a package UStat (Giurcanu et al. 2015) in the R language (R Core Team 2014) in which we have implemented the bootstrap methods in conjunction with the Mann–Whitney U-statistic with conditional kernel for the current status data model.

We close this section with an outline. In Sect. 2, we make a short review on the properties of the U-statistics for fully observed data models and introduce the U-statistics with conditional kernels for incomplete data models. In Sect. 3, we present the theoretical properties of the proposed U-statistics for several incomplete data models. In Sect. 4, we present the results of a simulation study of the finite sample performance of the Mann–Whitney U-statistic for the current status model, and in Sect. 5, we apply the proposed methods on a real-world data set consisting of the time to lung tumor onset for two groups of mice. The proofs of the theorems are deferred to an "Appendix".

2 U-Statistics for complete and incomplete data models

2.1 Complete data models

A complete (fully observed) data model is given by a random sample of n i.i.d. observations $\mathscr{X} = \{X_1, \ldots, X_n\}$ from an unknown distribution function (cdf) F. The parameter of interest is given by

$$\theta = \mathrm{E}\big(h(\boldsymbol{X})\big) = \int h(\boldsymbol{x})\,\mathrm{d}F_m(\boldsymbol{x}),$$

where $\boldsymbol{X} = (X_1, \ldots, X_m)^T \in \mathbb{R}^m$, $h : \mathbb{R}^m \mapsto \mathbb{R}$ is a known *kernel* function, $\boldsymbol{x} = (x_1, \ldots, x_m)^T \in \mathbb{R}^m$, $F_m(\boldsymbol{x}) = \prod_{i=1}^m F(x_i)$ is the cdf of $\boldsymbol{X}$, $\mathbb{R}$ is the set of real numbers, and m is the *degree* of the functional parameter θ, with $m < n$. We assume that the kernel h is permutation symmetric in its m arguments (any kernel can be always substituted with a symmetric kernel). The U-statistic with kernel h (Hoeffding 1948) is defined as

$$\hat{U} = \left(C_n^m\right)^{-1} \sum_{\boldsymbol{i} \in D_{n,m}} h(\boldsymbol{X}_{\boldsymbol{i}}),$$

where

$$D_{n,m} = \big\{\boldsymbol{i} = (i_1, \ldots, i_m) : 1 \le i_1 < \cdots < i_m \le n\big\},$$

$C_n^m = n!/(m!(n-m)!)$ is the binomial coefficient, and $\boldsymbol{X}_{\boldsymbol{i}} = (X_{i_1}, \ldots, X_{i_m})^T \in \mathbb{R}^m$ for $\boldsymbol{i} = (i_1, \ldots, i_m) \in D_{n,m}$. It is known that $\hat{U}$ is a minimum variance unbiased estimator of θ (see Serfling 1980, p. 176).

Let $h_c : \mathbb{R}^c \to \mathbb{R}$, $h_c^0 : \mathbb{R}^c \to \mathbb{R}$, and $\tilde{h}_c : \mathbb{R}^c \to \mathbb{R}$ for $c = 1, \ldots, m$, be given by

$$\begin{aligned}
h_c(x_1, \ldots, x_c) &= \mathrm{E}\big(h(X_1, \ldots, X_m)|X_1 = x_1, \ldots, X_c = x_c\big),\\
h_c^0(x_1, \ldots, x_c) &= h_c(x_1, \ldots, x_c) - \theta,\\
\tilde{h}_1(x_1) &= h_1^0(x_1),\\
\tilde{h}_2(x_1, x_2) &= h_2^0(x_1, x_2) - \tilde{h}_1(x_1) - \tilde{h}_1(x_2),
\end{aligned}$$

and

$$\tilde{h}_c(x_1, \ldots, x_c) = h_c^0(x_1, \ldots, x_c) - \sum_{\boldsymbol{i} \in D_{c,1}} \tilde{h}_1(\boldsymbol{x_i}) - \cdots - \sum_{\boldsymbol{i} \in D_{c,c-1}} \tilde{h}_{c-1}(\boldsymbol{x_i}),$$

where $\boldsymbol{x_i} = (x_{i_1}, \ldots, x_{i_j})^T \in \mathbb{R}^j$ for $\boldsymbol{i} = (i_1, \ldots, i_j) \in D_{c,j}$ and $j = 1, \ldots, c-1$. Note that $\mathrm{E}_F(\tilde{h}_c) = 0$, where $\mathrm{E}_F(\tilde{h}_c) = \mathrm{E}\big(\tilde{h}_c(X_1, \ldots, X_c)\big)$. The functions $\tilde{h}_1, \ldots, \tilde{h}_m$ are called the *canonical forms* of the kernel h. If $\mathrm{Var}_F(\tilde{h}_1) \neq 0$, then $\hat{U}$ is called nondegenerate. If $\mathrm{Var}_F(\tilde{h}_1) = \cdots = \mathrm{Var}_F(\tilde{h}_{r-1}) = 0$ and $\mathrm{Var}_F(\tilde{h}_r) \neq 0$, where $1 < r \leq m$, then $\hat{U}$ is called degenerate, with degeneracy of *order* r.

Hoeffding (1948, p. 300) showed that the variance of $\hat{U}$ is given by

$$\mathrm{Var}(\hat{U}) = \left(C_n^m\right)^{-1} \sum_{c=1}^{m} C_m^c C_{n-m}^{m-c} \eta_c^2,$$

where $\eta_c^2 = \mathrm{E}_F(\tilde{h}_c^2)$. Hoeffding (1961) further showed that $\hat{U}$ admits the following representation

$$\hat{U} - \theta = \sum_{c=1}^{m} C_m^c U_{nc}, \quad \text{where } U_{nc} = \left(C_n^c\right)^{-1} \sum_{\boldsymbol{i} \in D_{n,c}} \tilde{h}_c(\boldsymbol{X_i}).$$

If $\hat{U}$ is nondegenerate, then

$$n^{1/2}(\hat{U} - \theta) \xrightarrow{\mathrm{d}} J_1(\tilde{h}_1),$$

where $J_1(\tilde{h}_1) \sim N(0, m^2 \eta_1^2)$ and $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution. Note that in this case, the rate of convergence of $\hat{U}$ to θ is $n^{1/2}$ and the limiting distribution is normal. If $\hat{U}$ is degenerate of order $r > 1$, then the limiting distribution of $\hat{U}$ is not normal, but rather a Gaussian chaos (see, e.g., Koroljuk and Borovskikh 1988; van der Vaart 1998); specifically,

$$n^{r/2}(\hat{U} - \theta) \xrightarrow{\mathrm{d}} J_r(\tilde{h}_r),$$

where $J_r(\tilde{h}_r)$ is defined in "Appendix".

2.2 Incomplete data models

An incomplete data model is given by a random sample of n i.i.d. observations $\mathscr{Y} = \{Y_1, \ldots, Y_n\}$ which contains information about F but is not an i.i.d. sample from F. In this case, our interest is to estimate the parameter θ using the sample $\mathscr{Y}$. Usually, the underlying cdf F can be estimated by $\hat{F}$ using $\mathscr{Y}$, in which case, one may directly estimate θ using a *substitution* estimator given by

$$\tilde{\theta} = \int h(\boldsymbol{x}) \, \mathrm{d}\hat{F}_m(\boldsymbol{x}),$$

where $\hat{F}_m(\boldsymbol{x}) = \prod_{i=1}^m \hat{F}(x_i)$. Note that when $\hat{F}$ is the empirical cdf, then $\tilde{\theta}$ is just a V-statistic, while for incomplete data models, $\hat{F}$ is not the empirical cdf, and in that case, $\tilde{\theta}$ is not a V-statistic. It is known that although the U- and V-statistics are asymptotically equivalent up to a term of order $O(n^{-1})$ almost surely, their limiting distributions are not the same if the kernel is degenerate (see, e.g., Leucht 2012).

For incomplete data models, we propose a U-statistic defined in terms of a conditional expectation of the kernel $h(\boldsymbol{X})$ given the observable vector $\boldsymbol{Y}$, that is,

$$H(\boldsymbol{y}) = \mathrm{E}\big(h(\boldsymbol{X})|\boldsymbol{Y} = \boldsymbol{y}\big), \tag{1}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^T$ and $\boldsymbol{y} = (y_1, \ldots, y_m)^T$. Since the conditional kernel $H(\boldsymbol{y}) = H(\boldsymbol{y}|F)$ depends on the underlying F, it is thus unknown. However, we use the non-parametric maximum likelihood estimator (NPMLE) $\hat{F}$ of F and plug it into (1) to obtain a substitution estimator of $H(\boldsymbol{y})$, which we denote by $\hat{H}(\boldsymbol{y}) = H(\boldsymbol{y}|\hat{F})$. Then, the U-statistic with conditional kernel is defined as follows

$$\hat{U}(\hat{F}) = \big(C_n^m\big)^{-1} \sum_{\boldsymbol{i} \in D_{n,m}} \hat{H}(\boldsymbol{Y_i}),$$

where $\boldsymbol{Y_i} = (Y_{i_1}, \ldots, Y_{i_m})^T$ for $\boldsymbol{i} = (i_1, \ldots, i_m) \in D_{n,m}$. Note that the kernel of $\hat{U}(\hat{F})$ is random, and thus, it changes on random samples. Since we use $\hat{H}(\boldsymbol{y})$ in the expression of $\hat{U}(\hat{F})$, we expect that the convergence rate of $\hat{U}(\hat{F})$ to θ may be slower than that of $\hat{U}$ obtained under the complete data model; for details on the asymptotic behavior of $\hat{U}(\hat{F})$ for several incomplete data models, see the next section.

3 Theoretical results

In this section, we study the asymptotic properties of $\hat{U}(\hat{F})$ for some classical incomplete data models, such as the type I (current status) interval censoring model, the type II interval censoring model, the double censoring model, the convolution model, and the multiplicative censoring model.

3.1 Type I interval censoring model

The type I interval censoring model, also called the *current status* model, can be described as follows. Let $X \sim F$ and $T \sim G$ be two independent random variables, where F and G are unknown cdfs on the set of positive real numbers $\mathbb{R}^+$. Let $\Delta = I_{(X<T)}$ be a censoring indicator and $Y = (T, \Delta)$, where I_A denotes the indicator function of an event A. Assume that F and G have densities f and g with respect to the Lebesgue measure on $\mathbb{R}$ and that $0 < \Pr(\Delta = 1) < 1$. The available data consist of a random sample of n i.i.d. observations $\mathscr{Y} = \{Y_1, \ldots, Y_n\}$ from the distribution of Y, where $Y_i = (T_i, \Delta_i)$ for $i = 1, \ldots, n$. The density–mass function of the distribution of Y is given by

$$p_{F,G}(y) = F(t)^{\delta}\big(1 - F(t)\big)^{1-\delta} g(t), \quad y = (t, \delta) \in \mathbb{R}^+ \times \{0, 1\}.$$

The log-likelihood function of F, omitting the terms only containing $g(t)$, is

$$\hat{\ell}(F) = \sum_{i=1}^{n} \Big(\Delta_i \log\big(F(T_i)\big) + (1 - \Delta_i) \log\big(1 - F(T_i)\big)\Big).$$

We write the parameter of interest θ as follows

$$\theta = \mathrm{E}\big(h(\boldsymbol{X})\big) = \mathrm{E}\big[\mathrm{E}\big(h(\boldsymbol{X})|\boldsymbol{Y}\big)\big] = \mathrm{E}\big(H(\boldsymbol{Y})\big),$$

where $H(\boldsymbol{y}) = \mathrm{E}\big(h(\boldsymbol{X})|\boldsymbol{Y} = \boldsymbol{y}\big)$. Note that

$$\begin{aligned} H(\boldsymbol{y}) &= H(\boldsymbol{y}|F) \\ &= \int h(\boldsymbol{x}) \prod_{j=1}^{m} \left(\delta_j \frac{I_{[0,t_j)}(x_j)\, F(\mathrm{d}x_j)}{F(t_j)} + (1-\delta_j) \frac{I_{[t_j,\infty)}(x_j)\, F(\mathrm{d}x_j)}{1-F(t_j)}\right). \end{aligned} \tag{2}$$

Since F is unknown, then $H(\boldsymbol{y})$ is unknown, and thus, it needs to be estimated. To this end, let $\hat{H}(\boldsymbol{y}) = H(\boldsymbol{y}|\hat{F})$ be a substitution estimator of $H(\boldsymbol{y})$ obtained by substituting $\hat{F}$ for F in (2), where $\hat{F}$ is the NPMLE of F; that is,

$$\hat{H}(\boldsymbol{y}) = \int h(\boldsymbol{x}) \prod_{j=1}^{m} \left(\delta_j \frac{I_{[0,t_j)}(x_j) \hat{F}(\mathrm{d}x_j)}{\hat{F}(t_j)} + (1 - \delta_j) \frac{I_{[t_j,\infty)}(x_j) \hat{F}(\mathrm{d}x_j)}{1 - \hat{F}(t_j)}\right).$$

The U-statistic with conditional kernel estimator of θ is thus given by

$$\hat{U}(\hat{F}) = \big(C_n^m\big)^{-1} \sum_{\boldsymbol{i} \in D_{n,m}} \hat{H}(\boldsymbol{Y_i}).$$

The asymptotic properties of $\hat{U}(\hat{F})$ depend on the asymptotic properties of $\hat{F}$, results which have been studied by Groeneboom and Wellner (1992). Specifically, let $\{\mathbb{B}(t) : t \in \mathbb{R}\}$ be a two-sided Brownian motion originating at zero, i.e., $\{\mathbb{B}(t) : t \in \mathbb{R}\}$ is a mean zero Gaussian process on $\mathbb{R}$ with independent increments, $\mathbb{B}(0) = 0$ almost surely, and the increments $\mathbb{B}(s) - \mathbb{B}(t) \sim N(0, |s - t|)$. Groeneboom and Wellner (1992) showed that for all $t > 0$, we have

$$n^{1/3}\big(\hat{F}(t) - F(t)\big) \xrightarrow{\mathrm{d}} A(t)Z, \tag{3}$$

where

$$Z = \mathrm{argmin}_{t\in\mathbb{R}} \left\{\mathbb{B}(t) + t^2\right\} \quad \text{and} \quad A(t) = \left(\frac{4F(t)\big(1 - F(t)\big) f(t)}{g(t)}\right)^{1/3}. \tag{4}$$

The random variable Z follows the Chernoff distribution whose properties have been analyzed by Groeneboom and Wellner (2001).

The asymptotic properties of $\hat{U}(\hat{F})$ depend also on the first and higher order Hadamard derivatives of $H(\boldsymbol{y}|F)$. Specifically, let $H^{(j)}(\boldsymbol{y}|F)$ denote the jth order Hadamard derivative of $H(\boldsymbol{y}|F)$ with respect to F in $\ell^\infty(\mathbb{R}^+)$, where $\ell^\infty(\mathbb{R}^+)$ is the space of bounded functions defined on $\mathbb{R}^+$ equipped with the uniform norm. In particular, the first order Hadamard derivative of $H(\boldsymbol{y}|F)$ at $\alpha \in \ell^\infty(\mathbb{R}^+)$ (see "Appendix" for more details) is given by

$$H^{(1)}(\boldsymbol{y}|F;\alpha)=\sum_{k=1}^{m}\int h(\boldsymbol{x})\prod_{j\neq k}^{m}\left(\delta_j\frac{I_{[0,t_j)}(x_j)F(\mathrm{d}x_j)}{F(t_j)}+(1-\delta_j)\frac{I_{[t_j,\infty)}(x_j)F(\mathrm{d}x_j)}{1-F(t_j)}\right)$$
$$\times\left(\delta_k\frac{I_{[0,t_k)}(x_k)\big(F(t_k)\alpha(\mathrm{d}x_k)-\alpha(t_k)F(\mathrm{d}x_k)\big)}{F(t_k)^2}\right.$$
$$\left.+(1-\delta_k)\frac{I_{[t_k,\infty)}(x_k)\big[\big(1-F(t_k)\big)\alpha(\mathrm{d}x_k)+\alpha(t_k)F(\mathrm{d}x_k)\big]}{\big(1-F(t_k)\big)^2}\right). \tag{5}$$

We say that $H(\boldsymbol{y}|F)$ is of *rank* k at $\alpha \in \ell^\infty(\mathbb{R}^+)$ if it is k times Hadamard differentiable at F and

$$\mathrm{E}\big(H^{(j)}(\boldsymbol{Y}|F;\alpha)\big)=0 \quad \text{for } j=1,\ldots,k-1 \quad \text{and} \quad 0\neq \mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F;\alpha)\big)<\infty.$$

Let $\tilde{H}_c(\boldsymbol{y})$ be the canonical forms of $H(\boldsymbol{y})$, r be the order of $H(\boldsymbol{y})$, and k be the rank of $H(\boldsymbol{y}|F)$ at A, where A is given by (4). Theorem 1 shows the asymptotic properties of $\hat{U}(\hat{F})$; the regularity conditions and the proof are provided in "Appendix". Note that when $1=r<k$, then the convergence rate of $\hat{U}(\hat{F})$ is $n^{1/2}$ and the limiting distribution is normal.

Theorem 1 *(i) Assume that conditions* (A1)–(A2) *in "Appendix" hold. Then,*

$$\hat{U}(\hat{F})\xrightarrow{a.s.}\theta.$$

(ii) Assume that conditions (A1)–(A4) *in "Appendix" hold. Then*

$$n^s\big(\hat{U}(\hat{F})-\theta\big)\xrightarrow{d}\begin{cases}J_r(\tilde{H}_r), & s=r/2 \quad \textit{if} \quad 1\leq r<2k/3;\\ (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y})|F,A)\big)Z^k, & s=k/3 \quad \textit{if} \quad 1\leq k<3r/2.\end{cases}$$

3.2 Type II interval censoring model

Let $(X,U,V)\in\mathbb{R}^+\times(\mathbb{R}^+)^2$, where $X\sim F$ and $(U,V)\sim G$ are independent, with $U<V$ almost surely. Assume that F and G have densities f and g with respect to the Lebesgue measures on $\mathbb{R}$ and $\mathbb{R}^2$, respectively. Let further $\Delta=I_{(X<U)}$, $\Gamma=I_{(U\leq X<V)}$, $Y=(U,V,\Delta,\Gamma)$, and assume that $0<\Pr(\Delta=1)<1$ and

$0 < \Pr(\Gamma = 1) < 1$. For the type II interval censoring model, the available data consist of a random sample of n i.i.d. observations $\mathscr{Y} = \{Y_1, \ldots, Y_n\}$ from the distribution of Y, where $Y_i = (U_i, V_i, \Delta_i, \Gamma_i)$ for $i = 1, \ldots, n$. This model is described in Example 1.6 of Groeneboom and Wellner (1992, p. 5). The density–mass function of the distribution of Y is given by

$$p_{F,G}(y) = F(u)^{\delta}\big(F(v) - F(u)\big)^{\gamma}\big(1 - F(v)\big)^{1-\delta-\gamma} g(u, v),$$

where $y = (u, v, \delta, \gamma) \in \mathbb{R}^+ \times \mathbb{R}^+ \times \{0, 1\} \times \{0, 1\}$. Groeneboom and Wellner used an iterative convex minorant algorithm to compute the NPMLE $\hat{F}$ of F, and they determined the limiting distribution of $\hat{F}$ using the following hypothesis: starting with the true underlying cdf F, the iterative convex minorant algorithm will give at the first iteration an estimator which is asymptotically equivalent to the NPMLE. Groeneboom and Wellner (1992, Theorem 5.3, p. 100) showed that

$$(n \log n)^{1/3}\big(\hat{F}(t) - F(t)\big) \xrightarrow{\text{d}} A(t)Z, \tag{6}$$

where Z has the Chernoff distribution and

$$A(t) = \left(\frac{6f(t)^2}{g(t,t)}\right)^{1/3}. \tag{7}$$

Since $\{\Delta = 0; \Gamma = 0\} = \{X \geq V\}$, $\{\Delta = 0; \Gamma = 1\} = \{U \leq X < V\}$, and $\{\Delta = 1; \Gamma = 0\} = \{X < U\}$, then

$$F(x|Y = y) = \begin{cases} F(x)/\big(1 - F(v)\big)I_{(v \leq x)} & \text{if } \delta = 0 \text{ and } \gamma = 0, \\ F(x)/\big(F(v) - F(u)\big)I_{(u \leq x < v)} & \text{if } \delta = 0 \text{ and } \gamma = 1, \\ F(x)/F(u)I_{(x<u)} & \text{if } \delta = 1 \text{ and } \gamma = 0. \end{cases}$$

Similarly to the current status data model, we write the parameter of interest θ as follows

$$\theta = \mathrm{E}\big(h(\boldsymbol{X})\big) = \mathrm{E}\big[\mathrm{E}\big(h(\boldsymbol{X})|\boldsymbol{Y}\big)\big] = \mathrm{E}\big(H(\boldsymbol{Y})\big),$$

where

$$\begin{aligned} H(\boldsymbol{y}) &= H(\boldsymbol{y}|F) \\ &= \int h(\boldsymbol{x}) \prod_{j=1}^{m} \Bigg((1-\delta_j)(1-\gamma_j)\frac{I_{[v_j,\infty)}(x_j)F(\mathrm{d}x_j)}{1 - F(v_j)} \\ &\quad + \gamma_j(1-\delta_j)\frac{I_{[u_j,v_j)}(x_j)F(\mathrm{d}x_j)}{F(v_j) - F(u_j)} + \delta_j(1-\gamma_j)\frac{I_{[0,u_j)}(x_j)F(\mathrm{d}x_j)}{F(u_j)}\Bigg). \end{aligned} \tag{8}$$

Since F is unknown, we substitute the NPMLE $\hat{F}$ for F in (8) to obtain a substitution estimator of the conditional kernel $\hat{H}(\boldsymbol{y}) = H(\boldsymbol{y}|\hat{F})$, that is,

$$\hat{H}(\boldsymbol{y}) = \int h(\boldsymbol{x}) \prod_{j=1}^{m} \left((1-\delta_j)(1-\gamma_j) \frac{I_{[v_j,\infty)}(x_j)\hat{F}(\mathrm{d}x_j)}{1-\hat{F}(v_j)} \right.$$
$$\left. + \gamma_j(1-\delta_j) \frac{I_{[u_j,v_j)}(x_j)\hat{F}(\mathrm{d}x_j)}{\hat{F}(v_j)-\hat{F}(u_j)} + \delta_j(1-\gamma_j) \frac{I_{[0,u_j)}(x_j)\hat{F}(\mathrm{d}x_j)}{\hat{F}(u_j)} \right).$$

The first-order Hadamard derivative of $H(\boldsymbol{y}|F)$ at $\alpha \in \ell^\infty(\mathbb{R}^+)$ (for more details, see "Appendix" is given by

$$H^{(1)}(\boldsymbol{y}|F,\alpha) = \sum_{k=1}^{m} \int h(\boldsymbol{x}) \prod_{j\neq k}^{m} \left((1-\delta_j)(1-\gamma_j) \frac{I_{[v_j,\infty)}(x_j)F(\mathrm{d}x_j)}{1-F(v_j)} \right.$$
$$\left. + \gamma_j(1-\delta_j) \frac{I_{[u_j,v_j)}(x_j)F(\mathrm{d}x_j)}{F(v_j)-F(u_j)} + \delta_j(1-\gamma_j) \frac{I_{[0,u_j)}(x_j)F(\mathrm{d}x_j)}{F(u_j)} \right)$$
$$\times \left((1-\delta_k)(1-\gamma_k) \frac{I_{[v_k,\infty)}(x_k)\big[\big(1-F(v_k)\big)\alpha(\mathrm{d}x_k)+\alpha(v_k)F(\mathrm{d}x_k)\big]}{\big(1-F(v_k)\big)^2} \right.$$
$$+ \gamma_k(1-\delta_k) \frac{I_{[u_k,v_k)}(x_k)\big[\big(F(v_k)-F(u_k)\big)\alpha(\mathrm{d}x_k)-\big(\alpha(v_k)-\alpha(u_k)\big)F(\mathrm{d}x_k)\big]}{\big(F(v_k)-F(u_k)\big)^2}$$
$$\left. + \delta_k(1-\gamma_k) \frac{I_{[0,u_k)}(x_k)\big(F(u_k)\alpha(\mathrm{d}x_k) - \alpha(u_k)F(\mathrm{d}x_k)\big)}{F^2(u_k)} \right). \tag{9}$$

Theorem 2 shows the asymptotic properties of $\hat{U}(\hat{F})$ for the type II interval censoring model and its proof is provided in "Appendix". Note that when $1 = r < k$, then the rate of convergence of $\hat{U}(\hat{F})$ is $n^{1/2}$ and the limiting distribution is normal.

Theorem 2 *(i) Assume that* (A1)–(A2) *in "Appendix" hold. Then*

$$\hat{U}(\hat{F}) \xrightarrow{a.s.} \theta.$$

(ii) Assume that, in addition to conditions (A1)–(A3), *condition* (B1) *in "Appendix" holds, where A is given by* (7). *Then*

$$a_n(\hat{U}(\hat{F}) - \theta) \xrightarrow{d} \begin{cases} J_r(\tilde{H}_r), & a_n = n^{r/2} \; if\, 1 \le r < 2k/3; \\ (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F,A)\big)Z^k, & a_n = (n\log n)^{k/3 if\, 1\le k\le 3r/2}. \end{cases}$$

3.3 Double censoring model

The double censoring model was studied by Turnbull (1974) and Tsai and Crowley (1985), among others. Let $(X, U, V) \in \mathbb{R}^+ \times (\mathbb{R}^+)^2$, where $X \sim F$ and $(U, V) \sim G$ are independent, with $U < V$ almost surely, and let f and g be the densities of F and G, respectively. Let $Y = \big(Z, \Delta, \Gamma\big)$, where $Z = (X \vee U) \wedge V$, $\Delta = I_{(X<U)}$, and $\Gamma = I_{(U\le X<V)}$. For the double censoring model, the data consist of a random sample of n i.i.d. observations $\mathscr{Y} = \{Y_1, \ldots, Y_n\}$ from the distribution of Y, where

$Y_i = \big(Z_i, \Delta_i, \Gamma_i\big)$ and $Z_i = (X_i \vee U_i) \wedge V_i$ for $i = 1, \ldots, n$. The density–mass function $p_{F,G}$ of the distribution of Y is given by

$$p_{F,G}(y) = \big(M(z) f(z)\big)^{\gamma} \big(F(z) g_U(z)\big)^{\delta} \big[\big(1 - F(z)\big) g_V(z)\big]^{1-\gamma-\delta}, \quad y = (z, \delta, \gamma),$$

where $M(z) = \Pr\big(U \le z < V\big) = G_U(z) - G_V(z)$, G_U and G_V are the (marginal) cdfs of U and V, and g_U and g_V are the (marginal) densities of U and V, respectively.

Let $\mathscr{Z}_1 = \{Z_1, \ldots, Z_{n_1}\}$ be the observed values of Z_i for which $(\Delta_i, \Gamma_i) = (0, 1)$ and we may be tempted to use $\mathscr{Z}_1$ to estimate θ using a classical U-statistic given by

$$\hat{U}_1 = \big(C^m_{n_1}\big)^{-1} \sum_{\boldsymbol{i} \in D_{n_1,m}} h(\boldsymbol{Z_i}),$$

where $n_1 = \sum_{i=1}^n \Gamma_i$. However, $\mathscr{Z}_1$ is not an i.i.d. sample from F but rather an i.i.d. sample from the conditional distribution of X given $(U \le X < V)$. This has important consequences on the asymptotic behavior of $\hat{U}_1$. Specifically, using classical U-statistics theory, we have $\hat{U}_1 \xrightarrow{\text{a.s.}} \theta_1$, where

$$\theta_1 = \mathrm{E}\big(h(\boldsymbol{X})|(\boldsymbol{U} \le \boldsymbol{X} < \boldsymbol{V})\big).$$

If the kernel h is of order 1, then

$$n_1^{1/2}(\hat{U}_1 - \theta_1) \xrightarrow{\text{d}} N(0, m^2 \tilde{\eta}_1^2),$$

where

$$\tilde{\eta}_1^2 = \mathrm{E}\big(\tilde{h}_1(\boldsymbol{X})^2|(\boldsymbol{U} \le \boldsymbol{X} \le \boldsymbol{V})\big).$$

Note that, in this case, although it is asymptotically normal, $\hat{U}_1$ converges to $\theta_1 \ne \theta$, and thus, it is inconsistent.

When $m = 1$, we have

$$\mathrm{E}\big(h(X)|Y = y\big) = \int h(x) \left(\delta \frac{I_{(x<z)}}{F(z)} F(\mathrm{d}x) + \gamma\, \zeta_z(\mathrm{d}x) + (1 - \delta - \gamma) \frac{I_{(z<x)}}{1 - F(z)} F(\mathrm{d}x)\right),$$

where $y = (z, \delta, \gamma)$ and $\zeta_z(ds)$ is the Heaviside function, with $\int_0^\infty h(s)\zeta_z(ds) = h(z)$. When $m > 1$, we have

$$\begin{aligned} H(\boldsymbol{y}|F) = \mathrm{E}\big(h(\boldsymbol{X})|\boldsymbol{Y} = \boldsymbol{y}\big) = \int h(\boldsymbol{x}) \prod_{j=1}^m \Bigg(&\delta_j \frac{I_{(x_j<z_j)}}{F(z_j)} F(\mathrm{d}x_j) \\ &+ \gamma_j\, \zeta_{z_j}(\mathrm{d}x_j) + (1 - \delta_j - \gamma_j) \frac{I_{(z_j<x_j)}}{1 - F(z_j)} F(\mathrm{d}x_j)\Bigg). \end{aligned} \tag{10}$$

Since F in the expression of $H(\boldsymbol{y}|F)$ at (10) is unknown, we use the NPMLE $\hat{F}$ of F to obtain a substitution estimator of the conditional kernel $\hat{H}(\boldsymbol{y}) = H(\boldsymbol{y}|\hat{F})$:

$$\hat{H}(\boldsymbol{y}) = \int h(\boldsymbol{x}) \prod_{j=1}^{m} \left(\delta_j \frac{I_{(x_j<z_j)}}{\hat{F}(z_j)} \hat{F}(\mathrm{d}x_j) + \gamma_j \zeta_{z_j}(\mathrm{d}x_j) + (1-\delta_j-\gamma_j) \frac{I_{(z_j<x_j)}}{1-\hat{F}(z_j)} \hat{F}(\mathrm{d}x_j) \right).$$

The Hadamard derivative of $H(\boldsymbol{y}|F)$ at $\alpha \in \ell^\infty(\mathbb{R}^+)$ (for more details, see "Appendix" is given by

$$\begin{aligned} H^{(1)}(\boldsymbol{y}|F;\alpha) = \sum_{k=1}^{m} \int h(\boldsymbol{x}) \prod_{j\neq k} &\left(\delta_j \frac{I_{(x_j<z_j)}}{F(z_j)} F(\mathrm{d}x_j) + \gamma_j \zeta_{z_j}(\mathrm{d}x_j) \right. \\ &\left. + (1-\delta_j-\gamma_j) \frac{I_{(z_j<x_j)}}{1-F(z_j)} F(\mathrm{d}x_j) \right) \\ &\times \left(\delta_k \frac{I_{(x_k<z_k)}\left[F(z_k)\alpha(\mathrm{d}x_k) - \alpha(z_k)F(\mathrm{d}x_k)\right]}{F(z_k)^2} + \gamma_k \zeta_{z_k}(\mathrm{d}x_k) \right. \\ &\left. + (1-\delta_k-\gamma_k) \frac{I_{(z_k<x_k)}\left[\left(1-F(z_k)\right)\alpha(\mathrm{d}x_k) + \alpha(z_k)F(\mathrm{d}x_k)\right]}{\left(1-F(z_k)\right)^2} \right). \end{aligned} \tag{11}$$

Tsai and Crowley (1985) studied the asymptotic properties of the NPMLE $\hat{F}$ of F. They showed that

$$\sup_{L\le t\le R} |\hat{F}(t) - F(t)| \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad n^{1/2}(\hat{F} - F) \xrightarrow{\text{d}} S \quad \text{in } \ell^\infty([L,R]), \tag{12}$$

where $0 < L < R < \infty$ and $S = \{S(t) : t \geq 0\}$ is a mean zero Gaussian process on $[0,\infty)$ whose covariance function is not given in explicit form. Let $\mathrm{E}_{\boldsymbol{Y}}\left(H^{(k)}(\boldsymbol{Y}|F;S)\right)$ denote the expectation with respect to the distribution of $\boldsymbol{Y}$ only. Theorem 3 shows the asymptotic behavior of $\hat{U}(\hat{F})$ and its proof is provided in "Appendix". Note that when $1 = r < k$ or $1 = k < r$, then the rate of convergence of $\hat{U}(\hat{F})$ is $n^{1/2}$ and the limiting distribution is normal.

Theorem 3 *(i) Assume that* (A1)–(A2) *in "Appendix" hold. Then*

$$\hat{U}(\hat{F}) \xrightarrow{a.s.} \theta.$$

(ii) Assume that in addition to conditions (A1)–(A2)*, the condition* (C1) *in "Appendix" holds. Then,*

$$n^s(\hat{U}(\hat{F}) - \theta) \xrightarrow{d} \begin{cases} J_r(\tilde{H}_r), & s = r/2 \text{ if } 1 \leq r < k; \\ (1/k!)\,\mathrm{E}_{\boldsymbol{Y}}\left(H^{(k)}(\boldsymbol{Y}|F;S)\right), & s = k/2 \text{ if} 1 \leq k < r. \end{cases}$$

3.4 The convolution model

Let $(X, W) \in \mathbb{R} \times \mathbb{R}$, where $X \sim F$, $W \sim G$, X and W are independent, and let $Y = X + W$. We assume that F is unknown and that G is known with density g. For the convolution model, the data consist of a random sample of n i.i.d. observations $\mathscr{Y} = \{Y_1, \ldots, Y_n\}$ from the distribution of Y. The density of the distribution of Y is given by

$$p_{F,G}(y) = \int g(y - x)\, F(\mathrm{d}x).$$

As before, we write the parameter of interest θ as follows:

$$\theta = \mathrm{E}\big[\mathrm{E}\big(h(\boldsymbol{X})|\boldsymbol{Y}\big)\big] = \mathrm{E}\big(H(\boldsymbol{Y}|F)\big),$$

where

$$H(\boldsymbol{y}|F) = \frac{\int h(\boldsymbol{x})g(\boldsymbol{y}-\boldsymbol{x})\, F(d\boldsymbol{x})}{\int g(\boldsymbol{y}-\boldsymbol{x})\, F(d\boldsymbol{x})} \quad \text{and} \quad g(\boldsymbol{y}-\boldsymbol{x})\, F(d\boldsymbol{x}) = \prod_{i=1}^{m} g(y_i - x_i)\, F(\mathrm{d}x_i). \tag{13}$$

The substitution estimator $\hat{H}(\boldsymbol{y})$ of $H(\boldsymbol{y}|F)$ is obtained by substituting the NPMLE $\hat{F}$ of F in (13); that is,

$$\hat{H}(\boldsymbol{y}) = \frac{\int h(\boldsymbol{x})g(\boldsymbol{y}-\boldsymbol{x})\, \hat{F}(d\boldsymbol{x})}{\int g(\boldsymbol{y}-\boldsymbol{x})\, \hat{F}(d\boldsymbol{x})} \quad \text{where } g(\boldsymbol{y}-\boldsymbol{x})\, \hat{F}(d\boldsymbol{x}) = \prod_{i=1}^{m} g(y_i - x_i)\, \hat{F}(\mathrm{d}x_i).$$

The Hadamard derivative of $H(\boldsymbol{y}|F)$ at $\alpha \in \ell^{\infty}(\mathbb{R})$ is given by (for more details, see "Appendix"

$$\begin{aligned} H^{(1)}(\boldsymbol{y}|F;\alpha) = \sum_{k=1}^{m} \int h(\boldsymbol{x}) \Bigg(\prod_{j \neq k} \frac{g(y_j - x_j)F(\mathrm{d}x_j)}{\int g(y_j - x_j)\, F(\mathrm{d}x_j)} \Bigg) \\ \times g(y_k - x_k) \frac{\int g(y_k - u)\, F(du)\alpha(\mathrm{d}x_k) - \int g(y_k - u)\alpha(du)F(\mathrm{d}x_k)}{\big(\int g(y_k - u)F(du)\big)^2}. \end{aligned} \tag{14}$$

The asymptotic properties of the NPMLE $\hat{F}$ of F for the convolution model have been derived by Groeneboom and Wellner (1992). Groeneboom and Wellner (1992, Theorem 5.4) showed that

$$n^{1/3}\big(\hat{F}(y) - F(y)\big) \xrightarrow{\mathrm{d}} A(y)Z, \tag{15}$$

where

$$A(y) = \left(4f(y) \sum_{j=0}^{k} h(y + a_j)\big(g(a_j) - g(a_j-)\big)^{-1} \right)^{1/3}, \tag{16}$$

and Z has the Chernoff distribution. Theorem 4 shows the asymptotic properties of $\hat{U}(\hat{F})$ and its proof is provided in "Appendix". Note that when $1 = r < k$, then the rate of convergence of $\hat{U}(\hat{F})$ is $n^{1/2}$ and the limiting distribution is normal.

Theorem 4 *(i) Assume that conditions* (A1)–(A2) *in "Appendix" hold. Then*

$$\hat{U}(\hat{F}) \xrightarrow{a.s.} \theta.$$

(ii) Assume that, in addition to conditions (A1)–(A3)*, with A given by* (16)*, the condition* (D1) *in "Appendix" holds. Then*

$$n^s\big(\hat{U}(\hat{F}) - \theta\big) \xrightarrow{d} \begin{cases} J_r(\tilde{H}_r), & s = r/2 \textit{ if } 1 \le r < 2k/3; \\ (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F;\,A)\big)Z^k, & s = k/3 \textit{ if } 1 \le k < 3r/2. \end{cases}$$

3.5 Multiplicative censoring model

The multiplicative censoring model has been studied by Vardi (1989) and Vardi and Zhang (1992), among others. Let $(X, W) \in \mathbb{R}^+ \times \mathbb{R}^+$, with $X \sim F$, F is unknown with density f, $W \sim G$, $G = p\xi_1 + (1-p)U(0,1)$, $0 < p < 1$ is a known (mixing) proportion, ξ_1 is the point mass at 1, and X and W are independent. Let $T = XW$, $\Delta = I_{(W=1)}$, and $Y = (T, \Delta)$; hence, $T = X$ with probability p and $T = XW$ with probability $1-p$. For the multiplicative censoring model, the data consist of a random sample of n i.i.d. observations $\mathscr{Y} = \{Y_1, \ldots, Y_n\}$ from the distribution of Y. The density–mass function of the distribution of Y is given by

$$p_{F,G}(y) = \big(pf(t)\big)^{\delta}\left((1-p)\int_t^{\infty} x^{-1} f(x)\,\mathrm{d}x\right)^{1-\delta}, \quad y = (t, \delta).$$

If $\Delta = 1$, we observe X, and thus, we can use the classical U-statistic estimator of θ using only the subsample $\mathscr{X}_1 = \{X_1, \ldots, X_{n_1}\}$ for which $\Delta_i = 1$, where $n_1 = \sum_{i=1}^n \Delta_i$, given by

$$\hat{U}_1 = \big(C_{n_1}^m\big)^{-1} \sum_{\boldsymbol{i} \in D_{n_1,m}} h(\boldsymbol{X_i}).$$

Using classical U-statistics theory, we have $\hat{U}_1 \xrightarrow{\text{a.s.}} \theta$, and if $h(\boldsymbol{x})$ is a non-degenerate kernel, then

$$n^{1/2}(\hat{U}_1 - \theta) \xrightarrow{\mathrm{d}} N\big(0, p^{-1}m^2\eta_1^2\big).$$

Note that $\hat{U}_1$ uses only the data from the subsample $\mathscr{X}_1$ for which $\Delta_i = 1$, and thus, since the observations for which $\Delta_i = 0$ are totally ignored, we expect that the U-statistic with conditional kernel is more efficient than $\hat{U}_1$. To derive the expression of the conditional kernel, note first that, when $m = 1$, we have

$$H(y|F) = \mathrm{E}\big(h(X)|Y = y\big) = \int h(x)\left(\delta\zeta_t(\mathrm{d}x) + (1-\delta)\frac{I_{(t<x)}x^{-1}\,F(\mathrm{d}x)}{\int_t^\infty x^{-1}\,F(\mathrm{d}x)}\right),$$

where $\zeta_t(x)$ is the Heaviside function, with $\int h(x)\zeta_t(\mathrm{d}x) = h(t)$. When $m > 1$, then

$$\begin{aligned} H(\boldsymbol{y}|F) &= \mathrm{E}(h(\boldsymbol{X})|\boldsymbol{Y} = \boldsymbol{y}) \\ &= \int h(\boldsymbol{x})\prod_{j=1}^m\left(\delta_j\zeta_{t_j}(\mathrm{d}x_j) + (1-\delta_j)\frac{I_{(t_j<x_j)}x_j^{-1}\,F(\mathrm{d}x_j)}{\int_{t_j}^\infty u^{-1}\,F(du)}\right). \end{aligned} \tag{17}$$

As before, we substitute the NPMLE $\hat{F}$ for F in (17) to obtain the substitution kernel estimator $\hat{H}(\boldsymbol{y})$, that is

$$\hat{H}(\boldsymbol{y}) = \int h(\boldsymbol{x})\prod_{j=1}^m\left(\delta_j\zeta_{t_j}(\mathrm{d}x_j) + (1-\delta_j)\frac{I_{(t_j<x_j)}x_j^{-1}\,\hat{F}(\mathrm{d}x_j)}{\int_{t_j}^\infty u^{-1}\,\hat{F}(du)}\right).$$

The Hadamard derivative of $H(\boldsymbol{y}|F)$ at $\alpha \in \ell^\infty(\mathbb{R}^+)$ is given by

$$\begin{aligned} H^{(1)}(\boldsymbol{y}|F;\alpha) = \sum_{k=1}^m \int h(\boldsymbol{x})\prod_{j\neq k}\left(\delta_j\zeta_{t_j}(\mathrm{d}x_j) + (1-\delta_j)\frac{I_{(t_j<x_j)}x_j^{-1}\,F(\mathrm{d}x_j)}{\int_{t_j}^\infty u^{-1}\,F(du)}\right) \\ \times(1-\delta_k)\frac{I_{(t_k<x_k)}x_k^{-1}\big(\alpha(\mathrm{d}x_k)\int_{t_k}^\infty u^{-1}F(du) - F(\mathrm{d}x_k)\int_{t_k}^\infty u^{-1}\alpha(du)\big)}{\big(\int_{t_k}^\infty u^{-1}F(du)\big)^2}. \end{aligned} \tag{18}$$

Vardi and Zhang (1992) studied the asymptotic properties of the NPMLE $\hat{F}$ of F. They showed that, if $p > 0$, then

$$\sup_{t\geq 0}|\hat{F}(t) - F(t)| \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad n^{1/2}(\hat{F} - F) \xrightarrow{\text{d}} S \quad \text{in } D_0[0,\infty), \tag{19}$$

where $S = \{S(t) : t \geq 0\}$ is a zero mean Gaussian process on $D_0[0,\infty)$, $D_0[0,\infty)$ is the Banach space of all functions defined on $[0,\infty)$ which are right-continuous with left limits, converge to 0 at ∞, and are equal to 0 at 0. Theorem 5 shows the asymptotic properties of the U-statistic with conditional kernel for the multiplicative censoring model and its proof is provided in "Appendix". Note that when $1 = r < k$ or $1 = k < r$, then the rate of convergence of $\hat{U}(\hat{F})$ is $n^{1/2}$ and the limiting distribution is normal.

Theorem 5 *(i) Assume that conditions* (A1)–(A2) *in "Appendix" hold. Then*

$$\hat{U}(\hat{F}) \xrightarrow{a.s.} \theta.$$

(ii) *Assume that conditions* (A1)–(A2) *in "Appendix" hold, and that the conditions for* (19) *hold. Then,*

$$n^s(\hat{U}(\hat{F}) - \theta) \xrightarrow{d} \begin{cases} J_r(\tilde{H}_r), & s = r/2 \text{ if } 1 \le r < k; \\ (1/k!)\,\mathrm{E}_{\boldsymbol{Y}}\big(H^{(k)}(\boldsymbol{Y}|F; S)\big), & s = k/2 \text{ if } 1 \le k < r. \end{cases}$$

4 Simulation study

In this section, we present the results of a simulation study to describe the finite sample properties of the Mann–Whitney U-statistic with conditional kernel for the current status model. We first contrast the finite sample properties of the Mann–Whitney U-statistic with conditional kernel for the current status data model with the Mann–Whitney U-statistic for the complete data model in terms of the bias, standard error, and root mean squared error. We also present the finite sample properties of the non-parametric bootstrap confidence intervals and tests of hypotheses in conjunction with the Mann–Whitney U-statistic with conditional kernel for the current status model. The simulations were performed in the statistical environment R (R Core Team 2014), the NPMLE estimator of F is calculated using the R package fdrtool (Klaus and Strimmer 2014). We implemented these methods in an R package UStat (Giurcanu et al. 2015) which is available from the authors upon request. The routines that compute the Mann–Whitney U-statistic with conditional kernel have been implemented in the C language and compiled into a shared library to increase the speed of computations. We used the Sweave utility of R for reproducibility of the simulation study and of the data analysis of Sect. 5.

As a data generating process for the current status model, we consider the family of exponential distributions. Specifically, a synthetic data set consists of two independent i.i.d. samples $\mathscr{Y}^{(1)} = \{Y_1^{(1)}, \ldots, Y_m^{(1)}\}$ and $\mathscr{Y}^{(2)} = \{Y_1^{(2)}, \ldots, Y_n^{(2)}\}$ from the distributions of $Y^{(1)} = (T^{(1)}, \Delta^{(1)})$ and $Y^{(2)} = (T^{(2)}, \Delta^{(2)})$, respectively, where $\Delta^{(1)} = I_{(X^{(1)} < T^{(1)})}$ and $\Delta^{(2)} = I_{(X^{(2)} < T^{(2)})}$, with $X^{(1)} \sim \mathrm{Exp}(\lambda_1)$, $T^{(1)} \sim \mathrm{Exp}(\eta_1)$, $X^{(2)} \sim \mathrm{Exp}(\lambda_2)$, $T^{(2)} \sim \mathrm{Exp}(\eta_2)$, and $\mathrm{Exp}(\lambda)$ is the Exponential distribution with rate parameter λ. The kernel of the Mann–Whitney U-statistic (see, e.g., van der Vaart 1998 p. 166) is given by

$$h\big(x^{(1)}, x^{(2)}\big) = I_{\left(x^{(1)} \le x^{(2)}\right)}.$$

The parameter of interest θ is thus given by

$$\begin{aligned}\theta &= \Pr\big(X^{(1)} \le X^{(2)}\big)\\ &= \int I_{(x^{(1)} \le x^{(2)})} \lambda_1 \exp\big(-\lambda_1 x^{(1)}\big) \lambda_2 \exp\big(-\lambda_2 x^{(2)}\big)\, dx^{(1)} dx^{(2)}\\ &= \int_0^\infty \big[1 - \exp\big(-\lambda_1 x^{(2)}\big)\big] \lambda_2 \exp\big(-\lambda_2 x^{(2)}\big)\, dx^{(2)}\\ &= \lambda_1/(\lambda_1 + \lambda_2).\end{aligned}$$

For the complete data model, a synthetic data set consists of two independent i.i.d. samples $\mathscr{X}^{(1)} = \{X_1^{(1)}, \ldots, X_m^{(1)}\}$ and $\mathscr{X}^{(2)} = \{X_1^{(2)}, \ldots, X_n^{(2)}\}$ from the distributions of $X^{(1)}$ and $X^{(2)}$, respectively; in this case, the Mann–Whitney U-statistic is given by

$$\hat{\theta} = (mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{(X_i^{(1)} \leq X_j^{(2)})}.$$

Let $\hat{F}^{(1)}$ and $\hat{F}^{(2)}$ be the NPMLEs of $F^{(1)} = \mathrm{Exp}(\lambda_1)$ and $F^{(2)} = \mathrm{Exp}(\lambda_2)$, respectively. The Mann–Whitney U-statistic with conditional kernel is given by

$$\hat{U}\big(\hat{F}^{(1)}, \hat{F}^{(2)}\big) = (mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \hat{H}\big(Y_i^{(1)}, Y_j^{(2)}\big),$$

where

$$\begin{aligned} \hat{H}\big(y^{(1)}, y^{(2)}\big) = \int h\big(x^{(1)}, x^{(2)}\big) & \Bigg(\delta^{(1)} \frac{I_{[0,t^{(1)})}(x^{(1)}) \hat{F}^{(1)}(dx^{(1)})}{\hat{F}^{(1)}(t^{(1)})} \\ & + \big(1 - \delta^{(1)}\big) \frac{I_{[t^{(1)},\infty)}(x^{(1)}) \hat{F}^{(1)}(dx^{(1)})}{1 - \hat{F}^{(1)}(t^{(1)})} \Bigg) \\ & \times \Bigg(\delta^{(2)} \frac{I_{[0,t^{(2)})}(x^{(2)}) \hat{F}^{(2)}(dx^{(2)})}{\hat{F}^{(2)}(t^{(2)})} \\ & + \big(1 - \delta^{(2)}\big) \frac{I_{[t^{(2)},\infty)}(x^{(2)}) \hat{F}^{(2)}(dx^{(2)})}{1 - \hat{F}^{(2)}(t^{(2)})} \Bigg), \end{aligned} \tag{20}$$

$y^{(1)} = \big(t^{(1)}, \delta^{(1)}\big)$, and $y^{(2)} = \big(t^{(2)}, \delta^{(2)}\big)$. Let $\hat{M}^{(1)}$ and $\hat{M}^{(2)}$ be the greatest convex minorants of the piecewise linear functions that connect the set of points

$$\left\{ \left(i, \sum_{j=1}^{i} \Delta_{(j)}^{(1)} \right); i = 1, \ldots, m \right\} \quad \text{and} \quad \left\{ \left(i, \sum_{j=1}^{i} \Delta_{(j)}^{(2)} \right); i = 1, \ldots, n \right\},$$

respectively, where $\Delta_{(j)}^{(1)}$ and $\Delta_{(j)}^{(2)}$ are the censoring indicators corresponding to the jth order statistic $T_{(j)}^{(1)}$ and $T_{(j)}^{(2)}$. Let further $\nu_1^{(1)}, \ldots, \nu_{m_0}^{(1)}$ and $\nu_1^{(2)}, \ldots, \nu_{n_0}^{(2)}$ be the knots of the $\hat{M}^{(1)}$ and $\hat{M}^{(2)}$, respectively. Then, the NPMLE $\hat{F}^{(1)}$ and $\hat{F}^{(2)}$ of $F^{(1)}$ and $F^{(2)}$ are the left derivatives of $\hat{M}^{(1)}$ and $\hat{M}^{(2)}$ (see, e.g., Groeneboom and Wellner 1992). Hence, we have

$$\hat{F}^{(1)}\big(dx^{(1)}\big) = \sum_{i=1}^{m_0} p_i^{(1)} \xi_{\nu_i^{(1)}}(x^{(1)}) \quad \text{with } p_i^{(1)} = \hat{F}^{(1)}(\nu_i^{(1)}) - \hat{F}^{(1)}(\nu_i^{(1)}-)$$

and

$$\hat{F}^{(2)}\big(dx^{(2)}\big) = \sum_{j=1}^{n_0} p_j^{(2)} \xi_{\nu_j^{(2)}}(x^{(2)}) \quad \text{with } p_j^{(2)} = \hat{F}^{(2)}(\nu_j^{(2)}) - \hat{F}^{(2)}(\nu_j^{(2)}-),$$

where $\hat{F}^{(1)}(\mathrm{d}x)$ and $\hat{F}^{(2)}(\mathrm{d}x)$ are the probability mass functions of $\hat{F}^{(1)}$ and $\hat{F}^{(2)}$, $f(x-)$ is the left limit of f at x, and ξ_x is the point mass at x. Therefore, by (20), we obtain

$$\hat{H}\big(y^{(1)}, y^{(2)}\big) = \sum_{i=1}^{m_0}\sum_{j=1}^{n_0} h\big(\nu_i^{(1)}, \nu_j^{(2)}\big)\left(\delta^{(1)} \frac{I_{[0,t^{(1)})}(\nu_i^{(1)}) p_i^{(1)}}{\hat{F}^{(1)}(t^{(1)})} + \big(1-\delta^{(1)}\big)\frac{I_{[t^{(1)},\infty)}(\nu_i^{(1)}) p_i^{(1)}}{1-\hat{F}^{(1)}(t^{(1)})}\right)$$
$$\times\left(\delta^{(2)} \frac{I_{[0,t^{(2)})}(\nu_j^{(2)}) p_j^{(2)}}{\hat{F}^{(2)}(t^{(2)})} + \big(1-\delta^{(2)}\big)\frac{I_{[t^{(2)},\infty)}(\nu_j^{(2)}) p_j^{(2)}}{1-\hat{F}^{(2)}(t^{(2)})}\right).$$

Table 1 (p. 14) shows the simulation estimates of the bias, standard error (SE), and root mean squared error (RMSE) of the Mann–Whitney U-statistic for complete data (C) and the Mann–Whitney U-statistic with conditional kernel for the current status data model (I), with $\lambda_1 = \lambda_2 = \eta_1 = \eta_2 = 1$, for sample sizes $n = 100, 200, 400$; the number of simulated samples is $S = 1000$. Note that, as expected, the RMSEs of $\hat{\theta}$ are smaller than those of $\hat{U}\big(\hat{F}^{(1)}, \hat{F}^{(2)}\big)$ for all sample sizes under consideration.

We next describe the results of the simulation study for the bootstrap confidence interval estimation. To this end, let $\gamma_{\alpha,\mathrm{E}}$ and $\gamma_{\alpha,\mathrm{S}}$ be the α-quantiles of $\mathscr{L}\big(\hat{U}\big(\hat{F}^{(1)}, \hat{F}^{(2)}\big) - \theta\big)$ and $\mathscr{L}\big(|\hat{U}\big(\hat{F}^{(1)}, \hat{F}^{(2)}\big) - \theta|\big)$, where $\mathscr{L}\big(\hat{U}\big(\hat{F}^{(1)}, \hat{F}^{(2)}\big) - \theta\big)$ denotes the distribution of $\hat{U}\big(\hat{F}^{(1)}, \hat{F}^{(2)}\big) - \theta$. The $(1-\alpha)100\,\%$ equal-tailed confidence interval for θ (see, e.g., Hall 1992) is given by

$$\mathscr{I}_{1,n} = \Big(\hat{U}\big(\hat{F}^{(1)}, \hat{F}^{(2)}\big) - \gamma_{1-\alpha/2,\mathrm{E}}, \hat{U}\big(\hat{F}^{(1)}, \hat{F}^{(2)}\big) - \gamma_{\alpha/2,\mathrm{E}}\Big), \tag{21}$$

and the $(1-\alpha)100\,\%$ symmetric confidence interval for θ is given by

$$\mathscr{I}_{2,n} = \Big(\hat{U}\big(\hat{F}^{(1)}, \hat{F}^{(2)}\big) - \gamma_{1-\alpha,\mathrm{S}}, \hat{U}\big(\hat{F}^{(1)}, \hat{F}^{(2)}\big) + \gamma_{1-\alpha,\mathrm{S}}\Big). \tag{22}$$

Table 1 Simulation estimates of the bias, the standard error (SE), and the root mean squared error (RMSE) of the Mann–Whitney statistic for complete data model (C) and for the Mann–Whitney U-statistic with conditional kernel for the current status data model (I), with $\lambda_1 = \lambda_2 = \eta_1 = \eta_2 = 1$, for sample sizes $n = 100, 200, 400$, respectively; the number of simulated samples is $S = 1000$

n	BIAS(C)	SE(C)	RMSE(C)	BIAS(I)	SE(I)	RMSE(I)
100	0.000	0.041	0.041	0.017	0.079	0.081
200	0.001	0.028	0.028	0.012	0.051	0.052
400	−0.000	0.020	0.020	0.005	0.035	0.036

Table 2 Simulation estimates of the coverage probability (CPr), the average length (ML), and the standard error of the length (SEL) of the bootstrap equal-tailed (ET) and symmetric (S) confidence intervals for θ, at nominal confidence levels (Lev) $1-\alpha = 0.90, 0.95, 0.99$, and sample sizes $n = 100, 200, 400$

n	Lev	CPr(ET)	CPr(S)	ML(ET)	ML(S)	SEL(ET)	SEL(S)
100	0.90	0.907	0.950	0.299	0.311	0.036	0.042
100	0.95	0.957	0.984	0.361	0.376	0.046	0.053
100	0.99	0.987	0.999	0.489	0.521	0.073	0.089
200	0.90	0.914	0.952	0.194	0.202	0.018	0.021
200	0.95	0.961	0.982	0.233	0.243	0.022	0.027
200	0.99	0.989	0.998	0.311	0.328	0.037	0.044
400	0.90	0.900	0.900	0.128	0.133	0.009	0.012
400	0.95	0.946	0.946	0.153	0.159	0.011	0.014
400	0.99	0.987	0.987	0.202	0.211	0.018	0.022

The rate parameters of the exponential distributions generating the data for the current status model are $\lambda_1 = \lambda_2 = \eta_1 = \eta_2 = 1$, the number of simulated samples is $S = 1000$, and for each simulated sample, the number of bootstrap resamples is $B = 999$

Since $\gamma_{\alpha,\mathrm{E}}$ and $\gamma_{\alpha,\mathrm{S}}$ are unknown, we use the non-parametric bootstrap (Efron 1979) to estimate them. A non-parametric bootstrap resample is obtained by combining the with replacement random samples from each sample $\mathscr{Y}^{*(1)}$ and $\mathscr{Y}^{*(2)}$. The bootstrap confidence intervals are obtained by replacing the bootstrap estimates of the quantiles for the theoretical values in (21) and (22). Table 2 (p. 14) shows the simulation estimates of the coverage probability (CPr), the average length (ML), and the standard error of the length (SEL) of the bootstrap equal-tailed (ET) and symmetric (S) confidence intervals for θ at nominal confidence levels (Lev) $1-\alpha = 0.90, 0.95, 0.99$, and sample sizes $n = 100, 200, 400$. The number of simulated samples is $S = 1000$, and for each simulated sample, the number of bootstrap resamples is $B = 999$. We can see that the empirical coverage probabilities of the bootstrap equal-tailed confidence intervals are very close to the nominal levels for all sample sizes under consideration, and that the bootstrap symmetric confidence intervals are slightly conservative for smaller sample sizes ($n = 100$ and $n = 200$).

We finally describe the results of the simulation study for hypothesis testing of θ in conjunction with the Mann–Whitney U test statistic with conditional kernel for the current status model. Let us suppose that we want to test the null hypothesis $\mathrm{H}_0 : \theta = \theta^0$ versus the alternative hypothesis $\mathrm{H}_a : \theta \neq \theta^0$. For example, if we are interested to test if the survival functions of two treatment groups are different, then $\theta^0 = 1/2$. Let $T_n = \hat{U}\big(\hat{F}^{(1)}, \hat{F}^{(2)}\big) - \theta^0$, let $\mathscr{L}_0\big(T_n\big)$ denote the null distribution of T_n, and let $L_n(t)$ denote the cdf of $\mathscr{L}_0(T_n)$. The p-value of the Mann–Whitney U test statistic with conditional kernel is given by

$$\text{p-value} = 2 \times \min\big\{L_n(T_n), 1 - L_n(T_n)\big\}, \tag{23}$$

and thus, the test rejects H_0 at α nominal level (type I error rate) if p-value $\leq \alpha$. Since $\mathscr{L}_0(T_n)$ is unknown, we use the non-parametric bootstrap to estimate the null distribution of T_n; the bootstrap p-value is obtained by substituting the bootstrap estimator for the null distribution of T_n in (23). A non-parametric bootstrap resample is obtained by combining the with replacement random samples from each sample $\mathscr{Y}^{*(1)}$ and $\mathscr{Y}^{*(2)}$. The bootstrap version of T_n is defined in such a way that the null hypothesis holds in the bootstrap world irrespective of whether the null hypothesis holds or not in the population. For common pitfalls in devising bootstrap tests of hypotheses, see, e.g., Becher et al. (1993). Thus, the bootstrap p-value is given by

$$\widehat{\text{p-value}} = 2 \times \min\left\{\hat{L}_n(T_n), 1 - \hat{L}_n(T_n)\right\},$$

where $\hat{L}_n(t)$ is the (conditional) cdf of

$$T_n^* = \hat{U}^*\left(\hat{F}^{(*1)}, \hat{F}^{(*2)}\right) - \hat{U}\left(\hat{F}^{(1)}, \hat{F}^{(2)}\right)$$

given the combined sample $\mathscr{Y} = \left\{\mathscr{Y}^{(1)}, \mathscr{Y}^{(2)}\right\}$. Table 3 (p. 15) shows the empirical rejection rates of the Mann–Whitney U test under H_0 at nominal levels (Lev) $\alpha = 0.10, 0.05, 0.01$, and sample sizes $n = 100, 200, 400$. The number of simulated samples is $S = 1000$, and for each simulated sample, the number of bootstrap resamples is $B = 999$.

5 Mice data set

In this section, we analyze the mice data set (Hoel and Walburg 1972) using the Mann–Whitney U-statistic with conditional kernel for the current status data model. The data set has been recently analyzed by Choi et al. (2013). The mice data set consists of age-to-death with lung tumor of 144 RFM mice exposed to 300 R X radiation for two groups of mice: the conventional group, which consists of 96 mice kept in conventional environment, and the germ-free group, which consists of 48 mice that were kept in a germ-free environment. The event of interest is the time to tumor onset (in days), variable which is not directly observable. Specifically, the available data consist of the sacrifice time for each mouse and a censoring indicator of whether or not the mouse has lung tumor at the time of sacrifice. It is of interest to test if the time to lung tumor onset is different among the two treatment groups, and we will be using the Mann–

Table 3 Empirical rejection rates of the bootstrap Mann–Whitney U test with conditional kernel under $\mathrm{H}_0 : \theta = 1/2$ at nominal levels (Lev) $\alpha = 0.10, 0.05, 0.01$, and sample sizes $n = 100, 200, 400$

n	Lev = 0.10	Lev = 0.05	Lev = 0.01
100	0.087	0.042	0.007
200	0.079	0.042	0.010
400	0.098	0.050	0.010

The rate parameters of the exponential distributions generating the data for the current status model are $\lambda_1 = \lambda_2 = \eta_1 = \eta_2 = 1$, the number of simulated samples is $S = 1000$, and for each simulated sample, the number of bootstrap resamples is $B = 999$

Table 4 Bootstrap lower (LCL) and upper (UCL) confidence limits of the equal-tailed (ET) and symmetric (S) confidence intervals for θ at nominal confidence levels (Lev) $1-\alpha = 0.90, 0.95, 0.99$

Lev (%)	LCL(ET)	UCL(ET)	LCL(S)	UCL(S)
90	0.12	0.57	0.16	0.61
95	0.04	0.61	0.10	0.67
99	−0.08	0.70	−0.03	0.79

The number of bootstrap resamples is $B = 1999$

Whitney U-statistic with conditional kernel in conjunction with the non-parametric bootstrap confidence intervals and test of hypotheses described in Sect. 4 to test this hypothesis.

The value of the Mann–Whitney U test statistic with conditional kernel for the mice data is $\hat{U}(\hat{F}^{(1)}, \hat{F}^{(2)}) = 0.384$. Table 4 (p. 16) shows the bootstrap equal-tailed (ET) and symmetric (S) lower (LCL) and upper (UCL) confidence limits for θ at nominal confidence levels (Lev) $1-\alpha = 0.90, 0.95, 0.99$. The number of bootstrap resamples is $B = 1999$. Since the 95 % bootstrap equal-tailed and symmetric confidence intervals for θ contain 0.5, we conclude that we do not have statistical evidence that the two treatment groups have different survival functions (at $\alpha = 0.05$ nominal level). This claim is further supported by the bootstrap Mann–Whitney U test with conditional kernel; indeed, the bootstrap p-value of the Mann–Whitney U test with conditional kernel is 0.27, and thus, we again find that there is not statistical evidence that the survival functions of the two treatment groups are significantly different at $\alpha = 0.05$ nominal level.

Appendix: Proof of Theorem 1

We first introduce some notation to describe the limit distribution of degenerate U-statistics following closely van der Vaart (1998, Sect. 12.3). Let $\{\mathbb{G}(f); f \in L_2(\mathbb{R}, \mathscr{B}, F)\}$ be an F-Brownian bridge process, i.e., a mean zero Gaussian process indexed by functions in $L_2(\mathbb{R}, \mathscr{B}, F)$ with covariance function $\mathrm{E}(\mathbb{G}(f)\mathbb{G}(g)) = \mathrm{E}_F(fg) - \mathrm{E}_F(f)\,\mathrm{E}_F(g)$. Let $H_j(x)$ be the jth (non-normalized) Hermite polynomial, that is, the monic polynomial of degree j such that $\int H_i(x)H_j(x)\phi(x)\,\mathrm{d}x = 0$ for all $i \neq j \geq 1$, where $\phi(z)$ is the standard normal density function. Let $\{f_k; k \geq 1\}$ be an orthonormal basis in $L_2(\mathbb{R}, \mathscr{B}, F)$, where $f_1 = 1$. Then the functions $\{f_{k_1} \times \cdots \times f_{k_r}; k_1, \ldots, k_r \geq 1\}$ form an orthonormal basis in $(\mathbb{R}^r, \mathscr{B}^r, F^r)$, where $r \geq 1$. For any $K \in L_2(\mathbb{R}^r \mathscr{B}^r, F^r)$, let

$$\langle K, f_{k_1} \times \cdots \times f_{k_r}\rangle = \int K(x_1, \ldots, x_r) f_{k_1}(x_1) \ldots f_{k_r}(x_r)\mathrm{d}x_1 \ldots \mathrm{d}x_r$$

denote the inner product of K with the basis functions. If the kernel h of a degenerate U-statistic is of order $r > 1$, then by Theorem 12.10 of van der Vaart (1998), we have:

$$n^{r/2}\big(\hat{U}(\hat{F}) - \theta\big) \xrightarrow{\mathrm{d}} J_r(\tilde{h}_r),$$

where

$$J_r(\tilde{h}_r) = \sum_{k=(k_1,\ldots,k_r)\in\mathbb{N}^r} \langle h_r, f_{k_1} \times \cdots \times f_{k_r}\rangle \prod_{i=1}^{d(k)} H_{a_i(k)}\big(\mathbb{G}(\psi_i(k))\big),$$

where $\mathbb{N}$ is the set of positive integers, $\psi_1(k), \ldots, \psi_{d(k)}(k)$ are the different elements in $\{f_{k_1}, \ldots, f_{k_r}\}$ and $a_i(k)$ is the number of times $\psi_i(k)$ occurs among $\{f_{k_1}, \ldots, f_{k_r}\}$.

Regularity conditions

We assume the following regularity conditions:

(A1) $\mathrm{E}\big(h^2(\boldsymbol{X})\big) < \infty$.

(A2) The mapping $H(\boldsymbol{y}|F) : \mathbb{R}^d \times \ell^\infty(\mathbb{R}^+) \mapsto \mathbb{R}$ is k-times Hadamard continuously differentiable at F uniformly in $\boldsymbol{y}$, i.e.,

$$\sup_{\boldsymbol{y}} \left| H(\boldsymbol{y}|F + s_n\alpha_n) - H(\boldsymbol{y}|F) - \sum_{j=1}^{k}(1/j!)H^{(j)}(\boldsymbol{y}|F, s_n\alpha) \right| = o\big(|s_n|^k\big)$$

for all sequences $s_n \to 0$, $\|\alpha_n - \alpha\| \to 0$, with $s_n \in \mathbb{R}$ and $\alpha, \alpha_n \in \ell^\infty(\mathbb{R}^+)$, $n \geq 1$.

(A3) $A \in \ell^\infty(\mathbb{R}^+)$.

We first derive the expression of the Hadamard derivative of $H(\boldsymbol{y}|F)$ for the current status model. For fixed $\boldsymbol{y}$, the mapping $H(\boldsymbol{y}|\cdot) : \ell^\infty(\mathbb{R}^+) \mapsto \mathbb{R}$ is Hadamard differentiable at $F \in \ell^\infty(\mathbb{R}^+)$ if there exists a continuous linear map $H^{(1)}(\boldsymbol{y}|F, \cdot) : \ell^\infty(\mathbb{R}^+) \mapsto \mathbb{R}$, which is called the Hadamard derivative of $H(\boldsymbol{y}|F)$, such that

$$\frac{H(\boldsymbol{y}|F + s_n\alpha_n) - H(\boldsymbol{y}|F)}{s_n} \to H^{(1)}(\boldsymbol{y}|F, \alpha) \quad \text{as } n \to \infty$$

for all sequences $s_n \to 0$, $\|\alpha_n - \alpha\| \to 0$, with $s_n \in \mathbb{R}$ and $\alpha_n \in \ell^\infty(\mathbb{R}^+)$ for all n. By (2), we have

$$\begin{aligned}
\frac{H(\boldsymbol{y}|F + s_n\alpha_n) - H(\boldsymbol{y}|F)}{s_n} &= \sum_{k=1}^{m} \int h(\boldsymbol{x}) \\
\prod_{j\neq k}^{m} \Bigg(\delta_j \frac{I_{[0,t_j)}(x_j)\,\big(F + \zeta_j^k s_n\alpha_n\big)(\mathrm{d}x_j)}{\big(F + \zeta_j^k s_n\alpha_n\big)(t_j)} &+ (1-\delta_j)\frac{I_{[t_j,\infty)}(x_j)\,\big(F + \zeta_j^k s_n\alpha_n\big)(\mathrm{d}x_j)}{1 - \big(F + \zeta_j^k s_n\alpha_n\big)(t_j)} \Bigg) \\
\times s_n^{-1} \Bigg[\Bigg(\delta_k \frac{I_{[0,t_k)}(x_k)\,\big(F + s_n\alpha_n\big)(\mathrm{d}x_k)}{\big(F + s_n\alpha_n\big)(t_k)} &+ (1-\delta_k)\frac{I_{[t_k,\infty)}(x_k)\,\big(F + s_n\alpha_n\big)(\mathrm{d}x_k)}{1 - \big(F + s_n\alpha_n\big)(t_k)} \Bigg) \\
- \Bigg(\delta_k \frac{I_{[0,t_k)}(x_k)\, F(\mathrm{d}x_k)}{F(t_k)} &+ (1-\delta_k)\frac{I_{[t_k,\infty)}(x_k)\, F(\mathrm{d}x_k)}{1 - F(t_k)} \Bigg) \Bigg],
\end{aligned}$$

where $\zeta_j^k = 0$ for $j < k$ and $\zeta_j^k = 1$ for $j > k$. The first term in the product above tends to

$$\left(\delta_j \frac{I_{[0,t_j)}(x_j)\, F(\mathrm{d}x_j)}{F(t_j)} + (1-\delta_j)\frac{I_{[t_j,\infty)}(x_j)\, F(\mathrm{d}x_j)}{1-F(t_j)}\right) \quad \text{as } n \to \infty,$$

and the second term tends to

$$\left(\delta_k \frac{I_{[0,t_k)}(x_k)\big(F(t_k)\alpha(\mathrm{d}x_k) - \alpha(t_k)F(\mathrm{d}x_k)\big)}{F(t_k)^2} \right.$$
$$\left. +(1-\delta_k)\frac{I_{[t_k,\infty)}(x_k)\big[\big(1-F(t_k)\big)\alpha(\mathrm{d}x_k) + \alpha(t_k)F(\mathrm{d}x_k)\big]}{\big(1-F(t_k)\big)^2}\right),$$

for all $j = 1, \ldots, m$, as claimed at (5).

For part (ii) of Theorem 1, we need the following additional condition:

(A4) $n^{1/3}(\hat{F} - F)$ is asymptotically tight in $\ell^\infty(\mathbb{R}^+)$.

Proof (Theorem 1) Groeneboom and Wellner (1992, p. 79) showed that $\|\hat{F} - F\| \xrightarrow{\text{a.s.}} 0$, where $\|\alpha\|$ is the uniform norm of $\alpha \in \ell^\infty(\mathbb{R}^+)$. By condition (A2), it follows that

$$\sup_{\boldsymbol{y}} |H(\boldsymbol{y}|\hat{F}) - H(\boldsymbol{y}|F)| \xrightarrow{\text{a.s.}} 0.$$

Hence, $\hat{U}(\hat{F}) - \hat{U}(F) \xrightarrow{\text{a.s.}} 0$. Since $\hat{U}(F) \xrightarrow{\text{a.s.}} \theta$ by classical U-statistics theory, it follows that $\hat{U}(\hat{F}) \xrightarrow{\text{a.s.}} 0$, as stated at (i).

Note that $H^{(j)}(\boldsymbol{y}|F, Az) = H^{(j)}(\boldsymbol{y}|F, A)z^j$ for $z \in \mathbb{R}$ and $j = 1, \ldots, k$. Since $H(\boldsymbol{y}|F)$ is k times Hadamard differentiable at F, by the Functional Delta Theorem (van der Vaart 1998 Theorem 20.8) and condition (A4), it follows that

$$\sup_{\boldsymbol{y}} \left| n^{k/3}\big(H(\boldsymbol{y}|\hat{F}) - H(\boldsymbol{y}|F)\big) - \sum_{j=1}^{k} (1/j!) H^{(j)}\big(\boldsymbol{y}|F, n^{1/3}(\hat{F} - F)\big) \right| = o_P(1),$$

where $o_P(1)$ denotes convergence in probability to 0. We thus obtain the following asymptotic expansion of $n^{k/3}\big(\hat{U}(\hat{F}) - \hat{U}(F)\big)$:

$$n^{k/3}\big(\hat{U}(\hat{F}) - \hat{U}(F)\big) = \big(C_n^m\big)^{-1} \sum_{\boldsymbol{i} \in D_{n,m}} \sum_{j=1}^{k} (1/j!) H^{(j)}\big(\boldsymbol{Y}_{\boldsymbol{i}}|F, n^{1/3}(\hat{F} - F)\big) + o_P(1).$$

Since the rank of $H(\boldsymbol{y})$ at A is k, then $\mathrm{E}\big(H^{(j)}(\boldsymbol{Y}|F, A)\big) = 0$ for all $j < k$. Hence, for all $z_n \to z \in \mathbb{R}$,

$$\big(C_n^m\big)^{-1} \sum_{\boldsymbol{i} \in D_{n,m}} \sum_{j=1}^{k} (1/j!) H^{(j)}(\boldsymbol{Y}_{\boldsymbol{i}}|F, Az_n) \xrightarrow{\text{a.s.}} (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F, A)\big) z^k.$$

By condition (A3) and (3), it follows that

$$n^{1/3}(\hat{F} - F) \xrightarrow{\mathrm{d}} AZ \quad \text{in } \ell^{\infty}(\mathbb{R}^{+}), \tag{24}$$

where $A \in \ell^{\infty}(\mathbb{R}^{+})$ is given by (4) and Z has a Chernoff distribution. By (24), Slutsky's theorem, and continuous mapping theorem in $\ell^{\infty}(\mathbb{R}^{+})$ (see, e.g., van der Vaart 1998 Chapter 18), we obtain

$$n^{k/3}\big(\hat{U}(\hat{F}) - \hat{U}(F)\big) \xrightarrow{\mathrm{d}} (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F, A)\big)Z^{k}. \tag{25}$$

Since, $H(\boldsymbol{y})$ is a kernel of order r, by classical U-statistics theory, then

$$n^{r/2}(\hat{U}(F) - \theta) \xrightarrow{\mathrm{d}} J_r(\tilde{H}_r). \tag{26}$$

If $1 \leq r < 2k/3$, then by (25) we obtain $n^{r/2}\big(\hat{U}(\hat{F}) - \hat{U}(F)\big) = o_P(1)$. Thus, by (25) and (26), it follows that

$$\begin{aligned} n^{r/2}(\hat{U}(\hat{F}) - \theta) &= n^{r/2}\big(\hat{U}(\hat{F}) - \hat{U}(F)\big) + n^{r/2}\big(\hat{U}(F) - \theta\big) \\ &= n^{r/2}\big(\hat{U}(F) - \theta\big) + o_P(1) \xrightarrow{\mathrm{d}} J_r(\tilde{H}_r), \end{aligned}$$

as claimed at (ii). If $1 \leq k < 3r/2$, then by (26), we have $n^{k/3}(\hat{U}(F) - \theta) = o_P(1)$. Thus, by (25) and (26), it follows that

$$\begin{aligned} n^{k/3}\big(\hat{U}(\hat{F}) - \theta\big) &= n^{k/3}\big(\hat{U}(\hat{F}) - \hat{U}(F)\big) + n^{k/3}\big(\hat{U}(F) - \theta\big) \\ &= n^{k/3}\big(\hat{U}(\hat{F}) - \hat{U}(F)\big) + o_P(1) \xrightarrow{\mathrm{d}} (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F, A)\big)Z^{k}, \end{aligned}$$

as claimed at (ii). This completes the proof of the theorem. □

Proof of Theorem 2

We first obtain the expression of the Hadamard derivative for the type II interval censoring model. Using the notation used for the current status model, by (8), we have

$$\begin{aligned} \frac{H(\boldsymbol{y}|F + s_n a_n) - H(\boldsymbol{y}|F)}{s_n} &= \sum_{k=1}^{m} \int h(\boldsymbol{x}) \prod_{j \neq k}^{m} \\ \Bigg((1 - \delta_j)(1 - \gamma_j) &\frac{I_{[v_j, \infty)}(x_j)\big(F + \zeta_j^k s_n a_n\big)(\mathrm{d}x_j)}{1 - \big(F + \zeta_j^k s_n a_n\big)(v_j)} \\ + \gamma_j(1 - \delta_j) &\frac{I_{[u_j, v_j)}(x_j)\big(F + \zeta_j^k s_n a_n\big)(\mathrm{d}x_j)}{\big(F + \zeta_j^k s_n a_n\big)(v_j) - \big(F + \zeta_j^k s_n a_n\big)(u_j)} \end{aligned}$$

$$
\begin{aligned}
&+\delta_j(1-\gamma_j)\frac{I_{[0,u_j)}(x_j)\big(F+\zeta_j^k s_n a_n\big)(\mathrm{d}x_j)}{\big(F+\zeta_j^k S_n a_n\big)(u_j)}\Bigg)\\
&\times s_n^{-1}\Bigg[\Bigg((1-\delta_k)(1-\gamma_k)\frac{I_{[v_k,\infty)}(x_k)\big(F+s_n a_n\big)(\mathrm{d}x_k)}{1-\big(F+s_n a_n\big)(t_k)}\\
&\quad+\gamma_k(1-\delta_k)\frac{I_{[u_k,v_k)}(x_k)\big(F+s_n a_n\big)(\mathrm{d}x_k)}{\big(F+s_n a_n\big)(v_k)-\big(F+s_n a_n\big)(u_k)}\\
&+\delta_k(1-\gamma_k)\frac{I_{[0,u_k)}(x_k)\big(F+\zeta_k s_n a_n\big)(\mathrm{d}x_k)}{\big(F+\zeta_k S_n a_n\big)(u_k)}\Bigg)\\
&\Bigg((1-\delta_k)(1-\gamma_k)\frac{I_{[v_k,\infty)}(x_k)\,F(\mathrm{d}x_k)}{1-F(t_k)}+\gamma_k(1-\delta_k)\frac{I_{[u_k,v_k)}(x_k)\,F(\mathrm{d}x_k)}{F(v_k)-F(u_k)}\\
&\quad+\delta_k(1-\gamma_k)\frac{I_{[0,u_k)}(x_k)F(\mathrm{d}x_k)}{F(u_k)}\Bigg)\Bigg].
\end{aligned}
$$

The first term in the product above tends to

$$
\begin{aligned}
&\Bigg((1-\delta_j)(1-\gamma_j)\frac{I_{[v_j,\infty)}(x_j)\,F(\mathrm{d}x_j)}{1-F(v_j)}+\gamma_j(1-\delta_j)\frac{I_{[u_j,v_j)}(x_j)\,F(\mathrm{d}x_j)}{F(v_j)-F(u_j)}\\
&\quad+\delta_j(1-\gamma_j)\frac{I_{[0,u_j)}(x_j)F(\mathrm{d}x_j)}{F(u_j)}\Bigg)\quad \text{as } n\to\infty
\end{aligned}
$$

and the second term tends to

$$
\begin{aligned}
&(1-\delta_k)(1-\gamma_k)\frac{I_{[v_k,\infty)}(x_k)\big[\big(1-F(v_k)\big)\alpha(\mathrm{d}x_k)+\alpha(v_k)F(\mathrm{d}x_k)\big]}{\big(1-F(v_k)\big)^2}\\
&+\gamma_k(1-\delta_k)\frac{I_{[u_k,v_k)}(x_k)\big[\big(F(v_k)-F(u_k)\big)\alpha(\mathrm{d}x_k)-\big(\alpha(v_k)-\alpha(u_k)\big)F(\mathrm{d}x_k)\big]}{\big(F(v_k)-F(u_k)\big)^2}\\
&+\delta_k(1-\gamma_k)\frac{I_{[0,u_k)}(x_k)\big(F(u_k)\alpha(\mathrm{d}x_k)-\alpha(u_k)F(\mathrm{d}x_k)\big]}{F^2(u_k)}\quad \text{as } n\to\infty
\end{aligned}
$$

for all $j=1,\ldots,m$, as claimed at (9).

For the part (ii) of Theorem 2, we need the following additional condition:

(B1) $f(t)>0$, $g(t,t)>0$, $g(t,\cdot)$ is left continuous at t, and $\big(n\log(n)\big)^{1/3}(\hat F-F)$ is asymptotically tight in $\ell^\infty(\mathbb{R}^+)$.

Proof (Theorem 2) Groeneboom and Wellner (1992) showed that $\|\hat F-F\|\xrightarrow{\text{a.s.}}0$. By condition (A2), similarly to the proof of Theorem 1, it follows that $\hat U(\hat F)-\hat U(F)\xrightarrow{\text{a.s.}}0$. Since $\hat U(F)\xrightarrow{\text{a.s.}}\theta$ by classical U-statistics theory, it follows that $\hat U(\hat F)\xrightarrow{\text{a.s.}}\theta$, as claimed at (i).

Since $H(\boldsymbol{y}|F)$ is k times Hadamard differentiable at F, by the Functional Delta Theorem and condition (B1), it follows that

$$\sup_{\boldsymbol{y}} \Big| \big(n\log(n)\big)^{k/3}\big(H(\boldsymbol{y}|\hat{F}) - H(\boldsymbol{y}|F)\big) \\ - \sum_{j=1}^{k}(1/j!)H^{(j)}\big(\boldsymbol{y}\big|F, n^{1/3}\log(n)(\hat{F}-F)\big)\Big| = o_P(1).$$

Hence

$$\big(n\log(n)\big)^{k/3}\big(\hat{U}(\hat{F}) - \hat{U}(F)\big) \\ = \big(C_n^m\big)^{-1}\sum_{\boldsymbol{i}\in D_{n,m}}\sum_{j=1}^{k}(1/j!)H^{(j)}\big(\boldsymbol{Y_i}|F, \big(n\log(n)\big)^{1/3}(\hat{F}-F)\big) + o_P(1).$$

Since the rank of $H(\boldsymbol{y})$ at A is k, then $\mathrm{E}\big(H^{(j)}(\boldsymbol{Y}|F, A)\big) = 0$ for all $j < k$. Hence, for all $z_n \to z \in \mathbb{R}$,

$$\big(C_n^m\big)^{-1}\sum_{\boldsymbol{i}\in D_{n,m}}\sum_{j=1}^{k}(1/j!)H^{(j)}(\boldsymbol{Y_i}|F, Az_n) \xrightarrow{\text{a.s.}} (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F, A)\big)z^k.$$

By condition (B1) and (6), it follows that

$$\big(n\log(n)\big)^{1/3}(\hat{F} - F) \xrightarrow{\text{d}} AZ \quad \text{in } \ell^{\infty}(\mathbb{R}^+) \tag{27}$$

where $A \in \ell^{\infty}(\mathbb{R}^+)$ is given by (7) and Z has a Chernoff distribution. By (27), Slutsky's theorem, and continuous mapping theorem in $\ell^{\infty}(\mathbb{R}^+)$, we obtain

$$\big(n\log(n)\big)^{k/3}\big(\hat{U}(\hat{F}) - \hat{U}(F)\big) \xrightarrow{\text{d}} (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F, A)\big)Z^k. \tag{28}$$

Since $H(\boldsymbol{y})$ is a kernel of order r, by classical U-statistics theory, we have

$$n^{r/2}(\hat{U}(F) - \theta) \xrightarrow{\text{d}} J_r(\tilde{H}_r). \tag{29}$$

If $1 \le r < 2k/3$, then by (28) we have $n^{r/2}(\hat{U}(\hat{F}) - \hat{U}(F)) = o_P(1)$. Thus, by (28) and (29), it follows that $n^{r/2}(\hat{U}(\hat{F}) - \theta) \xrightarrow{\text{d}} J_r(\tilde{H}_r)$, as claimed at (ii). If $1 \le k < 3r/2$, then by (29), we have $\big(n\log(n)\big)^{k/3}(\hat{U}(F) - \theta) = o_P(1)$. Thus, by (28), it follows that $\big(n\log(n)\big)^{k/3}(\hat{U}(\hat{F}) - \theta) \xrightarrow{\text{d}} (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F, A)\big)Z^k$, as claimed at (ii). This completes the proof of the theorem. □

Proof of Theorem 3

We first derive the expression of the Hadamard derivative for the double censoring model. By (10), for all sequences $s_n \to s$ and $\|\alpha_n - \alpha\| \to 0$, as before we have

$$\frac{H(\boldsymbol{y}|F + s_n\alpha_n) - H(\boldsymbol{y}|F)}{s_n} = \sum_{k=1}^{m} \int h(\boldsymbol{x})$$

$$\prod_{j\neq k}^{m}\Bigg(\delta_j \frac{I_{(x_j<z_j)}\,\big(F+\zeta_j^k s_n\alpha_n\big)(\mathrm{d}x_j)}{\big(F+\zeta_j^k s_n\alpha_n\big)(z_j)} + \gamma_j\zeta_{z_j}(\mathrm{d}x_j)$$

$$+(1-\delta_j-\gamma_j)\frac{I_{(z_j<x_j)}\,\big(F+\zeta_j^k s_n\alpha_n\big)(\mathrm{d}x_j)}{1-\big(F+\zeta_j^k s_n\alpha_n\big)(z_j)}\Bigg)$$

$$\times s_n^{-1}\Bigg[\Bigg(\delta_k \frac{I_{(x_k<z_k)}\,\big(F+s_n\alpha_n\big)(\mathrm{d}x_k)}{\big(F+s_n\alpha_n\big)(z_k)}$$

$$+\gamma_k\zeta_{z_k}(\mathrm{d}x_k) + (1-\delta_k-\gamma_k)\frac{I_{(z_k<x_k)}\,\big(F+s_n\alpha_n\big)(\mathrm{d}x_k)}{1-\big(F+s_n\alpha_n\big)(z_k)}\Bigg)$$

$$-\left(\delta_k\frac{I_{(x_k<z_k)}\,F(\mathrm{d}x_k)}{F(z_k)} + \gamma_k\zeta_{z_k}(\mathrm{d}x_k) + (1-\delta_k-\gamma_k)\frac{I_{(z_k<x_k)}\,F(\mathrm{d}x_k)}{1-F(z_k)}\right)\Bigg].$$

The first term in the product above tends to

$$\left(\delta_j\frac{I_{(x_j<z_j)}\,F(\mathrm{d}x_j)}{F(z_j)} + \gamma_j\zeta_{z_j}(\mathrm{d}x_j) + (1-\delta_j-\gamma_j)\frac{I_{(z_j<x_j)}\,F(\mathrm{d}x_j)}{1-F(z_j)}\right) \quad \text{as } n\to\infty$$

and the second term tends to

$$\Bigg(\delta_k\frac{I_{(x_k<z_k)}\big(F(t_k)\alpha(\mathrm{d}x_k) - \alpha(z_k)F(\mathrm{d}x_k)\big)}{F(z_k)^2}$$

$$+(1-\delta_k-\gamma_k)\frac{I_{(z_k<x_k)}\big[\big(1-F(z_k)\big)\alpha(\mathrm{d}x_k)\alpha(t_k)F(\mathrm{d}x_k)\big]}{\big(1-F(z_k)\big)^2}\Bigg) \quad \text{as } n\to\infty,$$

for all $j = 1, \ldots, m$, as claimed at (11).

For the part (ii) of Theorem 3, we need the following additional condition:

(C1) For all $0 < L < R < \infty$, then $F(L) > 0$ and $F(R) < 1$; $G_U(\cdot)$ and $G_V(\cdot)$ have no common discontinuity points; the operator $H_4^{*'}$ given in Corollary 5.5 of Tsai and Crowley (1985) is invertible; and $n^{1/2}(\hat{F} - F)$ is asymptotically tight in $\ell^\infty(\mathbb{R}^+)$.

Proof (Theorem 3) Tsai and Crowley (1985) showed that $\hat{F}(t) \xrightarrow{\text{a.s.}} F(t)$ for all $t \in \mathbb{R}^+$. Since $F(t)$ is continuous, then $\|\hat{F} - F\| \to 0$ (see, e.g., van der Vaart 1998 Problem 1, p. 339). By condition (A2), similarly to the proof of Theorem 1, it follows that $\hat{U}(\hat{F})-\hat{U}(F) \xrightarrow{\text{a.s.}} 0$. Since $\hat{U}(F) \xrightarrow{\text{a.s.}} \theta$ by classical U-statistics theory, it follows that $\hat{U}(\hat{F}) \xrightarrow{\text{a.s.}} \theta$, as claimed at (i).

Since $H(\boldsymbol{y}|F)$ is k times Hadamard differentiable at F, by the Functional Delta Theorem and condition (C1), we obtain

$$\sup_{\boldsymbol{y}} \left| n^{k/2}\big(H(\boldsymbol{y}|\hat{F}) - H(\boldsymbol{y}|F)\big) - \sum_{j=1}^{k}(1/j!)H^{(j)}\big(\boldsymbol{y}|F, n^{1/2}(\hat{F}-F)\big) \right| = o_P(1).$$

Hence,

$$n^{k/2}\big(\hat{U}(\hat{F}) - \hat{U}(F)\big) = \big(C_n^m\big)^{-1} \sum_{\boldsymbol{i}\in D_{n,m}} \sum_{j=1}^{k}(1/j!)H^{(j)}(\boldsymbol{Y}_{\boldsymbol{i}}|F, n^{1/2}(\hat{F}-F)\big) + o_P(1).$$

Since the rank of $H(\boldsymbol{y})$ is k, then $\mathrm{E}\big(H^{(j)}(\boldsymbol{Y}|F,S)\big) = 0$ for all $j < k$. Hence, for all $s_n \to s \in \mathbb{R}$,

$$\big(C_n^m\big)^{-1} \sum_{\boldsymbol{i}\in D_{n,m}} \sum_{j=1}^{k}(1/j!)H^{(j)}(\boldsymbol{Y}_{\boldsymbol{i}}|F, s_n) \xrightarrow{\text{a.s.}} (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F,s)\big).$$

By condition (C1) and (12), it follows that

$$n^{1/2}(\hat{F}-F) \xrightarrow{\text{d}} S \quad \text{in } \ell^\infty(\mathbb{R}^+), \tag{30}$$

where S is given by (12). By (30), Slutsky's theorem, and continuous mapping theorem in $\ell^\infty(\mathbb{R}^+)$, we have

$$n^{k/2}\big(\hat{U}(\hat{F}) - \hat{U}(F)\big) \xrightarrow{\text{d}} (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F,S)\big). \tag{31}$$

Since $H(\boldsymbol{y})$ is a kernel of order r, then

$$n^{r/2}\big(\hat{U}(F) - \theta\big) \xrightarrow{\text{d}} J_r(\tilde{H}_r). \tag{32}$$

If $1 \le r < k$, then by (31) we have $n^{r/2}(\hat{U}(\hat{F}) - \hat{U}(F)) = o_P(1)$. Thus, by (31) and (32), it follows that $n^{r/2}\big(\hat{U}(\hat{F}) - \theta\big) \xrightarrow{\text{d}} J_r(\tilde{H}_r)$, as claimed at (ii). If $1 \le k < r$, then by (32), we have $n^{k/2}(\hat{U}(F) - \theta) = o_P(1)$. Thus, by (31), it follows that $n^{k/2}\big(\hat{U}(\hat{F}) - \theta\big) \xrightarrow{\text{d}} (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F,S)\big)$, as claimed at (ii). This completes the proof of the theorem. $\square$

Proof of Theorem 4

We first derive the expression of the Hadamard derivative for the convolution model. As before, for all sequences $s_n \to 0$ and $\|\alpha_n - \alpha\| \to 0$,

$$\frac{H(\boldsymbol{y}|F + s_n\alpha_n) - H(\boldsymbol{y}|F)}{s_n} = \sum_{k=1}^{m} \int h(\boldsymbol{x}) \left(\prod_{j \neq k} \frac{g(y_j - x_j)(F + \zeta_j^k s_n \alpha_n)(\mathrm{d}x_j)}{\int g(y_j - u)(F + \zeta_j^k s_n \alpha_n)(du)} \right)$$
$$\times s_n^{-1} \left(\frac{g(y_k - x_k)(F + s_n\alpha_n)(\mathrm{d}x_k)}{\int g(y_k - u)(F + s_n\alpha_n)(du)} - \frac{g(y_k - x_k)F(\mathrm{d}x_k)}{\int g(y_k - u)F(du)} \right).$$

The first term of the product above tends to

$$\prod_{j \neq k} \frac{g(y_j - x_j)F(\mathrm{d}x_j)}{\int g(y_j - u)F(du)} \quad \text{as } n \to \infty$$

and the second term tends to

$$g(y_k - x_k) \frac{\int g(y_k - u)F(du)\alpha(\mathrm{d}x_k) - \int g(y_k - u)\alpha(du)F(\mathrm{d}x_k)}{\left(\int g(y_k - u)F(du)\right)^2} \quad \text{as } n \to \infty,$$

for all $k = 1, \ldots, m$, as claimed at (14).

For part (ii) of Theorem 4, we consider the following additional condition:

(D1) Assume that q is a right-continuous and decreasing density function on $[0, \infty)$ having only a finite number of discontinuity points at $a_0 = 0 < a_1 < \cdots < a_k < \infty$; that the derivative $q'(x)$ of $q(x)$ exists for all $x \neq a_1, \ldots, a_k$ and satisfies $\int_0^\infty q'(x)^2/q(x)\,\mathrm{d}x < \infty$; that $q'(x)$ is bounded and continuous on (a_{i-1}, a_i) for $i = 1, \ldots, k+1$, where $a_{k+1} = \infty$; and that $n^{1/2}(\hat{F} - F)$ is asymptotically tight in $\ell^\infty(\mathbb{R}^+)$.

Proof (Theorem 4) Groeneboom and Wellner (1992) showed that $\hat{F}(x) \xrightarrow{\text{a.s.}} F(x)$ for all $x \geq 0$. As in the proof of Theorem 3, this result continues to hold in the uniform norm, that is $\|\hat{F} - F\| \xrightarrow{\text{a.s.}} 0$. By condition (A2), similarly to the proof of Theorem 1, it follows that $\hat{U}(\hat{F}) - \hat{U}(F) \xrightarrow{\text{a.s.}} 0$. Since $\hat{U}(F) \xrightarrow{\text{a.s.}} \theta$ by classical U-statistics theory, it follows that $\hat{U}(\hat{F}) \xrightarrow{\text{a.s.}} \theta$, as claimed at (i).

Since $H(\boldsymbol{y}|F)$ is k times Hadamard differentiable at F, by the Functional Delta Theorem and condition (D1), it follows that

$$\sup_{\boldsymbol{y}} \left| n^{k/3}\left(H(\boldsymbol{y}|\hat{F}) - H(\boldsymbol{y}|F)\right) - \sum_{j=1}^{k} (1/j!) H^{(j)}\left(\boldsymbol{y}\middle|F, n^{1/3}(\hat{F} - F)\right) \right| = o_P(1).$$

Hence

$$n^{k/3}\big(\hat{U}(\hat{F})-\hat{U}(F)\big)=\big(C_n^m\big)^{-1}\sum_{\boldsymbol{i}\in D_{n,m}}\sum_{j=1}^{k}(1/j!)H^{(j)}\big(\boldsymbol{Y_i}|F,n^{1/3}(\hat{F}-F)\big)+o_P(1).$$

Since the rank of $H(\boldsymbol{y})$ is k, then $\mathrm{E}\big(H^{(j)}(\boldsymbol{Y}|F,A)\big)=0$ for all $j<k$. Hence, for all $z_n\to z\in\mathbb{R}$,

$$\big(C_n^m\big)^{-1}\sum_{\boldsymbol{i}\in D_{n,m}}\sum_{j=1}^{k}(1/j!)H^{(j)}(\boldsymbol{Y_i}|F,Az_n)\xrightarrow{\text{a.s.}}(1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F,A)\big)z^k.$$

By condition (D1) and (15), it follows that

$$n^{1/3}(\hat{F}-F)\xrightarrow{\text{d}}AZ\quad\text{in }\ell^{\infty}(\mathbb{R}^+),\tag{33}$$

where $A\in\ell^{\infty}(\mathbb{R}^+)$ is given by (16) and Z has a Chernoff distribution. By (33), Slutsky's theorem, and continuous mapping theorem in $\ell^{\infty}(\mathbb{R}^+)$, we obtain

$$n^{k/3}\big(\hat{U}(\hat{F})-\hat{U}(F)\big)\xrightarrow{\text{d}}(1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F,A)\big)Z^k.\tag{34}$$

Since $H(\boldsymbol{y})$ is a kernel of order r, by classical U-statistics theory,

$$n^{r/2}\big(\hat{U}(F)-\theta\big)\xrightarrow{\text{d}}J_r(\tilde{H}_r).\tag{35}$$

If $1\le r<2k/3$, then by (34) we have $n^{r/2}(\hat{U}(\hat{F})-\hat{U}(F))=o_P(1)$. Thus, by (34) and (35), it follows that $n^{r/2}\big(\hat{U}(\hat{F})-\theta\big)\xrightarrow{\text{d}}J_r(\tilde{H}_r)$, as claimed at (ii). If $1\le k<3r/2$, then by (35), we have $n^{k/3}(\hat{U}(F)-\theta)=o_P(1)$. Thus, by (34), it follows that $n^{k/3}\big(\hat{U}(\hat{F})-\theta\big)\xrightarrow{\text{d}}(1/k!)\,\mathrm{E}\big[H^{(k)}(\boldsymbol{Y}|F,A)\big]Z^k$, as claimed at (ii). This completes the proof of the theorem. □

Proof of Theorem 5

We first derive the expression of the Hadamard derivative for the multiplicative censoring model. As before, for all sequences $s_n\to 0$ and $\|\alpha_n-\alpha\|\to 0$,

$$\frac{H(\boldsymbol{y}|F+s_n\alpha_n)-H(\boldsymbol{y}|F)}{s_n}=\sum_{k=1}^{m}\int h(\boldsymbol{x})\prod_{j\neq k}$$
$$\left(\delta_j\zeta_{t_j}(\mathrm{d}x_j)+(1-\delta_j)\frac{I_{(t_j<x_j)}x_j^{-1}(F+\zeta_j^k s_n\alpha_n)(\mathrm{d}x_j)}{\int_{t_j}^{\infty}u^{-1}(F+\zeta_j^k s_n\alpha_n)(du)}\right)$$

$$\times\, s_n^{-1}\left(\delta_k\zeta_{t_k}(\mathrm{d}x_k)+(1-\delta_k)\frac{I_{(t_k<x_k)}x_k^{-1}(F+s_n\alpha_n)(\mathrm{d}x_k)}{\int_{t_k}^{\infty}u^{-1}(F+s_n\alpha_n)(du)}-\delta_k\zeta_{t_k}(\mathrm{d}x_k)\right.$$
$$\left.-(1-\delta_k)\frac{I_{(t_k<x_k)}x_k^{-1}F(\mathrm{d}x_k)}{\int_{t_k}^{\infty}u^{-1}F(du)}\right).$$

The first term of the product above tends to

$$\prod_{j\neq k}\left(\delta_j\zeta_{t_j}(\mathrm{d}x_j)+(1-\delta_j)\frac{I_{(t_j<x_j)}x_j^{-1}F(\mathrm{d}x_j)}{\int_{t_j}^{\infty}u^{-1}F(du)}\right)\quad \text{as } n\to\infty,$$

and the second term tends to

$$(1-\delta_k)\frac{I_{(t_k<x_k)}x_k^{-1}\big(\alpha(\mathrm{d}x_k)\int_{t_k}^{\infty}u^{-1}F(du)-F(\mathrm{d}x_k)\int_{t_k}^{\infty}u^{-1}\alpha(du)\big)}{\big(\int_{t_k}^{\infty}u^{-1}F(du)\big)^2}\quad \text{as } n\to\infty,$$

for all $k=1,\dots,m$, as claimed at (18).

Proof (Theorem 5) Tsai and Crowley (1985) showed that $\|\hat{F}-F\|\xrightarrow{\text{a.s.}}0$. By condition (A2), similarly to the proof of Theorem 1, it follows that $\hat{U}(\hat{F})-\hat{U}(F)\xrightarrow{\text{a.s.}}0$. Since $\hat{U}(F)\xrightarrow{\text{a.s.}}\theta$ by classical U-statistics theory, it follows that $\hat{U}(\hat{F})\xrightarrow{\text{a.s.}}\theta$, as claimed at (i).

Since $H(\boldsymbol{y}|F)$ is k times Hadamard differentiable at F, by the Functional Delta Theorem, it follows that

$$\sup_{\boldsymbol{y}}\left|n^{k/2}\big(H(\boldsymbol{y}|\hat{F})-H(\boldsymbol{y}|F)\big)-\sum_{j=1}^{k}(1/j!)H^{(j)}\big(\boldsymbol{y}|F,n^{1/2}(\hat{F}-F)\big)\right|=o_P(1).$$

Thus,

$$n^{k/2}\big(\hat{U}(\hat{F})-\hat{U}(F)\big)=\big(C_n^m\big)^{-1}\sum_{\boldsymbol{i}\in D_{n,m}}\sum_{j=1}^{k}(1/j!)H^{(j)}(\boldsymbol{Y_i}|F,n^{1/2}(\hat{F}-F)\big)+o_P(1).$$

Since the rank of $H(\boldsymbol{y})$ is k, then $\mathrm{E}\big(H^{(j)}(\boldsymbol{Y}|F,S)\big)=0$ for all $j<k$. Hence, for all $s_n\to s\in\mathbb{R}$,

$$\big(C_n^m\big)^{-1}\sum_{\boldsymbol{i}\in D_{n,m}}\sum_{j=1}^{k}(1/j!)H^{(j)}(\boldsymbol{Y_i}|F,s_n)\xrightarrow{\text{a.s.}}(1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F,s)\big).$$

By (19), Slutsky's theorem, and continuous mapping theorem in $\ell^{\infty}(\mathbb{R}^+)$, it follows that

$$n^{k/2}\big(\hat{U}(\hat{F})-\hat{U}(F)\big)\xrightarrow{\text{d}}(1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F,S)\big). \tag{36}$$

Since $H(\boldsymbol{y})$ is a kernel of order r, by classical U-statistics theory, we have

$$n^{r/2}\big(\hat{U}(F) - \theta\big) \xrightarrow{\mathrm{d}} J_r(\tilde{H}_r). \tag{37}$$

If $1 \leq r < k$, then by (36) we have $n^{r/2}(\hat{U}(\hat{F}) - \hat{U}(F)) = o_P(1)$. Thus, by (36) and (37), it follows that $n^{r/2}\big(\hat{U}(\hat{F}) - \theta\big) \xrightarrow{\mathrm{d}} J_r(\tilde{H}_r)$, as claimed at (ii). If $1 \leq k < r$, then by (37), we have $n^{k/2}(\hat{U}(F) - \theta) = o_P(1)$. Thus, by (36), it follows that $n^{k/2}\big(\hat{U}(\hat{F}) - \theta\big) \xrightarrow{\mathrm{d}} (1/k!)\,\mathrm{E}\big(H^{(k)}(\boldsymbol{Y}|F, S)\big)$, as claimed at (ii). This completes the proof of the theorem. □

References

Akritas, M. G. (1986). Empirical processes associated with V-statistics and a class of estimators under random censoring. *The Annals of Statistics*, *14*, 619–637.

Becher, H., Hall, P., Wilson, S. R. (1993). Bootstrap hypothesis testing procedures. *Biometrics*, *49*, 1268–1272.

Bennet, S. (1983). Log-logistic regression models for survival data. *Journal of the Royal Statistical Society Series C* (*Applied Statistics*), *32*, 165–171.

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Annals of Mathematical Statistics*, *37*, 51–58.

Borovskikh, Y. V. (1986). *Theory of U-Statistics in Hilbert spaces*. Kiev: Institute of Mathematics, Ukrainean Academy of Science.

Bose, A., Sen, A. (1999). The strong law of large numbers for Kaplan-Meier U-statistics. *Journal of Theoretical Probability*, *12*, 181–200.

Bose, A., Sen, A. (2002). Asymptotic distribution of the Kaplan-Meier U-statistics. *Journal of Multivariate Analysis*, *83*, 84–123.

Choi, B. Y., Fine, J. P., Brookhart, M. A. (2013). Predictable confidence intervals for current status data. *Statistics in Medicine*, *32*, 1419–1428.

Core Team, R. (2014). *R*: *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Datta, S., Bandyopadhyay, D., Satten, G. A. (2010). Inverse probability of censoring weighted U-statistics for right-censored data with and application to testing hypotheses. *Scandinavian Journal of Statistics*, *37*, 680–700.

Efron, B. (1979). Bootstrap methods: Another look at Jackknife. *The Annals of Statistics*, *7*, 1–26.

Giurcanu, M., Yuan, A., Luta, G., Tan, M. (2015). UStat: The Mann-Whitney U-statistic with conditional kernel. *R package version*, 1.

Gregory, G. (1977). Large sample theory for U-statistics and tests of fit. *Annals of Statistics*, *5*, 110–123.

Groeneboom, P., Wellner, J. (1992). *Information bounds and nonparametric maximum likelihood estimation*. Basel: Birkháuser Verlag.

Groeneboom, P., Wellner, J. (2001). Computing Chernoff's distribution. *Journal of Computational and Graphical Statistics*, 388–400.

Hall, P. (1992). *The bootstrap and edgeworth expansion*. New York: Springer.

Heagerty, P., Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics*, *61*, 92–105.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, *19*, 293–325.

Hoeffding, W. (1961). The strong law of large numbers for U-statistics. *Institute of Statistics Mimeo Series*, *302*, 1–10.

Hoel, D. G., Walburg, H. E. (1972). Statistical analysis of survival experiments. *Journal of the National Cancer Institute*, *49*, 361–372.

Hu, C., Degrutolla, V. (2011). Recursive partitioning of resistant mutations for longitudinal markers based on a U-type score. *Biostatistics*, *12*, 750–762.

Janson, S. (1979). *The asymptotic distribution of degenerate U-statistics*. Preprint No 5, Department of Mathematics, University of Uppsala.

Klaus, B., Strimmer, K. (2014). Fdrtool: Estimation of (local) false discovery rates and higher criticism. *R package version*, *1*(2), 12.

Koroljuk, V. S., Borovskikh, I. V. (1988). Asymptotic theory of U-statistics. *Institute of Mathematics, Academy of Sciences of the Ukrainean SSR*, *40*, 169–182.

Korolyuk, V. S., Borovskikh, I. V. (1994). *Theory of U-statistics*. Dodrecht: Kluwer.

Kowalski, J., Tu, X. M. (2007). *Modern applied U-statistics*. Kiev: Wiley.

Leucht, A. (2012). Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency. *Bernoulli*, *18*(2), 552–585.

Ma, Y., Valle, A. G., Zhang, A., Tu, X. M. (2010). A U-statistics based approach for modelling Cronbach coefficient alpha within a longitudinal data setting. *Statistics in Medicine*, *29*, 659–670.

Rubin, H., Vitale, R. A. (1980). Asymptotic distribution of symmetric statistics. *Annals of Statistics*, *8*, 165–170.

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. (1974). Almost sure behavior of U-statistics and von Mises' differentiable statistical functions. *The Annals of Statistics*, *2*, 387–395.

Serfling, R. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

Therneau, T., Grambsch, P. (2010). *Modelling survival data*: *Extending the Cox model*. New York: Springer.

Tressou, J. (2006). Nonparametric modelling of the left censorship of analytical data in food risk assessment. *Journal of the American Statistical Association*, *101*, 1377–1386.

Tsai, W. Y., Crowley, J. (1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. *Annals of Statistics*, *13*, 1317–1334.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. New York: Springer.

Turnbull, B. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association*, *69*, 169–173.

van der Vaart, A. W. (1998). *Asymptotic statistics*. New York: Springer.

Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika*, *76*, 751–761.

Vardi, Y., Zhang, C. H. (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *Annals of Statistics*, *20*, 1022–1039.

von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Annals of Mathematical Statistics*, *18*, 309–348.