The Annals of Statistics
2017, Vol. 45, No. 5, 1988–2015
DOI: 10.1214/16-AOS1516

ASYMPTOTIC THEORY OF GENERALIZED ESTIMATING EQUATIONS BASED ON JACK-KNIFE PSEUDO-OBSERVATIONS

BY MORTEN OVERGAARD, ERIK THORLUND PARNER AND JAN PEDERSEN[1]

Aarhus University

A general asymptotic theory of estimates from estimating functions based on jack-knife pseudo-observations is established by requiring that the underlying estimator can be expressed as a smooth functional of the empirical distribution. Using results in p-variation norms, the theory is applied to important estimators from time-to-event analysis, namely the Kaplan–Meier estimator and the Aalen–Johansen estimator in a competing risks model, and the corresponding estimators of restricted mean survival and cause-specific lifetime lost. Under an assumption of completely independent censorings, this allows for estimating parameters in regression models of survival, cumulative incidences, restricted mean survival, and cause-specific lifetime lost. Considering estimators as functionals and applying results in p-variation norms is apparently an excellent way of studying the asymptotics of such estimators.

1. Introduction. The pseudo-observation method was introduced in Andersen, Klein and Rosthøj (2003) as a way to perform regression analysis when modeling, for example, state occupation probabilities in multi-state settings in time-to-event analysis. It is an alternative to the popular Cox model that models the hazard functions, but results in a complicated model for the state occupation probabilities. The pseudo-observation method has particularly been considered in the competing risks setting, where the state occupation probabilities are cause-specific incidences and survival probability, which are estimated by the Aalen–Johansen estimator and Kaplan–Meier estimator, respectively, but the method has also been used for modeling the restricted mean survival that is not a state occupation probability.

The pseudo-observation method is suited for a situation in which we are interested in the effect of covariates, Z, on a variable, V, that we for some reason may not always be able to observe entirely, for example, due to missingness, censoring or other coarsening of the data. We are interested in modeling the effect of Z on V by

$$\mathrm{E}(V \mid Z) = \mu(\beta_0; Z) \tag{1.1}$$

Received October 2015; revised August 2016.

[1]Supported by the Danish Council for Independent Research (Grant DFF-4002-00003).

MSC2010 subject classifications. Primary 62N02; secondary 62F12, 62J12.

Key words and phrases. Pseudo-values, functional differentiability, von Mises expansion, pseudo-observation method, p-variation, U-statistics.

for some mean function μ and true, but unknown parameter vector β_0. Typically, $\mu(\beta_0; Z) = \mu(\beta_0^T Z)$ is the inverse of the link function in a generalized linear model setup. Estimating β_0 is usually done by solving an estimating equation of the type

$$\sum_{k=1}^{n} A(\beta; Z_k)\big(V_k - \mu(\beta; Z_k)\big) = 0, \tag{1.2}$$

but this is not possible when the i.i.d. sample $V_1, \dots, V_n$ is not fully observed. The pseudo-observation method presumes the expected value $\theta = \mathrm{E}(V)$ can be estimated reasonably well from a sample, $X_1, \dots, X_n$, of some observable random variable X, which will likely include information on V when V is observed. Letting $\hat{\theta}_n$ denote the estimate of θ based on the entire sample and letting $\hat{\theta}_n^{(k)}$ denote the similar estimate based on the sample $X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n$, that is, leaving out X_k, the jack-knife pseudo-observation of the potentially unobserved V_k is defined as

$$\hat{\theta}_{n,k} = n\hat{\theta}_n - (n-1)\hat{\theta}_n^{(k)}. \tag{1.3}$$

Doing this for all k, we obtain a sample of pseudo-observations, $\hat{\theta}_{n,1}, \dots, \hat{\theta}_{n,n}$. Estimates of β_0 are then obtained by solving an estimating equation of the type

$$\sum_{k=1}^{n} A(\beta; Z_k)\big(\hat{\theta}_{n,k} - \mu(\beta; Z_k)\big) = 0, \tag{1.4}$$

which is just (1.2) with V_k substituted for the pseudo-observation $\hat{\theta}_{n,k}$. The pseudo-observations are an intricate transformation of the original dataset and are likely to be correlated, and as such it is not clear by results from the standard methods that this is a reasonable estimating procedure.

Andersen, Klein and Rosthøj (2003) conjectured that an estimating equation like (1.4) in the generalized linear model setup would yield consistent, asymptotically normal estimates with a variance that could be consistently estimated by the ordinary sandwich variance estimator. A proof of the conjectures of Andersen, Klein and Rosthøj (2003) in the competing risks setting was presented by Graw, Gerds and Schumacher (2009), based on a so-called von Mises expansion of the estimator, and relying on an assumption that the censoring mechanism is independent of event time, event type and covariates. We will call this completely independent censorings later on.

In the paper of Jacobsen and Martinussen (2016), the proof of Graw, Gerds and Schumacher (2009) was, however, found to be lacking in the handling of some remainder terms. Jacobsen and Martinussen (2016) also used a von Mises expansion, but only studied the survival probability, that is, pseudo-observations based on the Kaplan–Meier estimator, and found that while all other conclusions of Graw, Gerds and Schumacher (2009) remain true in that setting, the variance of the estimator cannot in general be consistently estimated by the ordinary sandwich variance estimator, which instead leads to conservative confidence bounds and tests. In the context of pseudo-observations, however, many von Mises expansions are used and it is then important to handle the remainder terms with great care.

In this paper, we consider the pseudo-observation method in a general framework. This involves considering estimators as functionals and applying concepts from functional analysis. Functional analytic results stated in Dudley and Norvaiša (2011) allow us to study the pseudo-observations rigorously in this framework. We argue that we will have asymptotic normality and consistency of parameter estimates in a quite general setting under certain conditions. By appealing to a p-variation approach, we will, by results stated in Dudley and Norvaiša (1999), see that this is specifically the case when the pseudo-observations are based on the Kaplan–Meier estimator, the Aalen–Johansen estimator in the competing risks setting, and the estimators of restricted mean and cause-specific lost lifetime. We will also see that the usual sandwich variance estimator generally is not consistent, and we propose a modification.

In Section 2, we will discuss how many estimators can be considered functionals mapping a sample average to an estimate and briefly review some of the theory on functionals and their differentiability. In Section 3, the main result, a general asymptotic theory for the pseudo-observation method, is presented based on the assumption that the estimator, and thereby functional, under study is sufficiently well behaved. In Section 4, we consider various important regression models from time-to-event analysis in the competing risks setting, and we will see how the parameters can be estimated using the pseudo-observation method and how the necessary assumptions can be fulfilled.

2. Estimators as functionals.

A reasonable estimator $\hat{\theta}(\cdot)$ of a parameter θ based on an i.i.d. sample $X_1, \dots, X_n$, can likely be considered a function of the empirical distribution of the X_i's. For instance, take a look at the log-likelihood function when X_i has density f,

$$l_n(\theta) = \sum_{i=1}^{n} \log f(\theta; X_i) = n \int \log f(\theta; x) \,\mathrm{d}F_n(x), \tag{2.1}$$

where F_n is the empirical distribution. Any solution to $l_n(\theta) = 0$ will depend on the data only through F_n. In other words, a reasonable estimator is given by

$$\hat{\theta}_n = \hat{\theta}(F_n), \tag{2.2}$$

where $\hat{\theta}(\cdot)$ is a map from the set of distributions to the set of parameters.

In the following, we treat an estimating map like $\hat{\theta}$ from (2.2) by considering it a functional from a Banach space to the parameter space. Since the actual empirical distribution is not necessarily a convenient starting point, we introduce the following general setup in which we allow F_n to be a more general sample average. In our applications, this sample average will be some linear transformation of the empirical distribution, and we want to avoid the technicalities of this transformation.

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, $(\mathcal{X}, \mathcal{A})$ a measurable space and $(\mathbf{D}, \|\cdot\|)$ a Banach space. Consider a map $\delta_{(\cdot)}: \mathcal{X} \to \mathbf{D}$ and an i.i.d. sample $X_1, \dots, X_n$ defined on $(\Omega, \mathcal{F}, \mathrm{P})$ with values in $\mathcal{X}$. We define the *sample average* by

$$F_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \in \mathbf{D}. \tag{2.3}$$

When $(F_n)_{n\geq 1}$ has a limit in $\mathbf{D}$, we will not hesitate to call it F.

EXAMPLE 2.1. In our applications in competing risks scenarios in time-to-event analysis, an observation is a pair $X = (\tilde{T}, \tilde{\Delta})$ of the time of exit $\tilde{T}$ and the status at the time of exit $\tilde{\Delta}$. We can take $\mathcal{X} = \mathbb{R}^2$ or $\mathcal{X} = \mathbb{R}_+ \times \{0, \dots, d\}$ (the possible statuses are 0 for censored and $1, \dots, d$ for competing events). The actual empirical distribution function of a sample $X_1, \dots, X_n$ is not a convenient starting point, but the following transformation is: For an observation $x = (\tilde{t}, \tilde{\delta}) \in \mathcal{X}$ define $Y_x(s) = 1(\tilde{t} \geq s)$, an indication of still being at risk at time s, and $N_{x,j}(s) = 1(\tilde{t} \leq s, \tilde{\delta} = j)$ for $j = 0, \dots, d$, indications for having exited with a specific type of event or censoring before time s. We let $\delta_{(\cdot)}$ be given by $\delta_x = (Y_x, N_{x,0}, \dots, N_{x,d})^T$ such that $F_n = \frac{1}{n}\sum \delta_{X_i}$ is a vector of step functions. There are different choices of Banach space $\mathbf{D}$ with different properties in this case, for example, a function space with a supremum norm or with a norm based on the entry-wise total variations or perhaps something in between. We will return to this choice later on. The candidate for the limit of F_n is of course $F = (H, H_0, \dots, H_d)^T$, where $H(s) = \mathrm{P}(\tilde{T} \geq s)$ is the probability of being at risk at time s and $H_j(s) = \mathrm{P}(\tilde{T} \leq s, \tilde{\Delta} = j)$ for $j = 0, \dots, d$ are the probabilities of having exited with a specific type of event or censoring before time s.

In the setting above, the estimating map, $\hat{\theta}$, can be considered a functional from $\mathbf{D}$ to the parameter space, which we later for simplicity will assume is $\mathbb{R}$, another Banach space. For such a functional, the concept of Fréchet differentiability generalizes the ordinary concept of differentiability. Details on functional differentiability and useful results for our setting can be found in the supplement [Overgaard, Parner and Pedersen (2017)]. Here, we give a short introduction.

Let $\mathbf{D}$ and $\mathbf{E}$ be Banach spaces and consider an open subset $W \subseteq \mathbf{D}$ and a functional $\phi : W \to \mathbf{E}$. Then ϕ is said to be (Fréchet) differentiable at $f \in W$ if a continuous, linear map $\phi'_f: \mathbf{D} \to \mathbf{E}$ exists such that

$$\|\phi(f+h) - \phi(f) - \phi'_f(h)\|_{\mathbf{E}} = o(\|h\|_{\mathbf{D}}), \tag{2.4}$$

and it is differentiable in W if it is differentiable at all $f \in W$. This notation is inspired by van der Vaart (1998). We usually call $\phi'_f(h)$ the derivative of ϕ at f in the direction of h. The derivative, $\phi': f \mapsto \phi'_f$ is itself a functional from W

to $L^1(\mathbf{D}, \mathbf{E})$, where $L^1(\mathbf{D}, \mathbf{E})$ is the space of linear, continuous maps from $\mathbf{D}$ to $\mathbf{E}$, which is a Banach space when endowed with the operator norm. Higher-order differentiability and continuous differentiability (C^k) is then defined in terms of this functional.

Assuming ϕ is our estimating functional, that is, $\hat{\theta}_n = \phi(F_n)$, and that it is differentiable at a given $F \in \mathbf{D}$, the corresponding (first-order) influence function is defined by

$$\dot{\phi}(x) = \phi'_F(\delta_x - F), \qquad x \in \mathcal{X}. \tag{2.5}$$

Note that due to linearity of ϕ'_F we have $\phi'_F(F_n - F) = \frac{1}{n}\sum_i \dot{\phi}(X_i)$. The following example shows how a candidate of the influence function can be found in one situation.

EXAMPLE 2.2. Consider the setting of Example 2.1 and suppose we have decided on a Banach space $\mathbf{D}$ consisting of elements $h = (h_*, h_0, \dots, h_d)$ that are vectors of functions, including $\delta_x = (Y_x, N_{x,0}, \dots, N_{x,d})$. Consider the functional $\psi : \mathbf{D} \to \mathbb{R}$ given by $\psi(h) = \int_0^t \frac{1(h_*(s)>0)}{h_*(s)} \, \mathrm{d}h_1(s)$. Note that

$$\psi(F) = \int_0^t \frac{1}{H(s)} \, \mathrm{d}H_1(s) = \Lambda_1(t), \tag{2.6}$$

the cumulative hazard for cause 1 at time t, if $H(t) > 0$, and

$$\psi(F_n) = \int_0^t \frac{1(Y(s) > 0)}{Y(s)} \, \mathrm{d}N_1(s) = \hat{\Lambda}_1(t), \tag{2.7}$$

the Nelson–Aalen estimate of Λ_1 at time t. Here, we use the notation $Y(s) = \#\{i : \tilde{T}_i \geq s\}$ and $N_1(s) = \#\{i : \tilde{T}_i \leq s, \tilde{\Delta}_i = 1\}$ as in Andersen et al. (1993). Taking the limit for $u \downarrow 0$ of $u^{-1}(\psi(h + ug) - \psi(h))$ for $h, g \in \mathbf{D}$ and ignoring the indicator in the numerator, we obtain

$$\int_0^t \frac{1}{h_*(s)} \, \mathrm{d}g_1(s) - \int_0^t \frac{g_*(s)}{h_*(s)^2} \, \mathrm{d}h_1(s), \tag{2.8}$$

which is the candidate for $\psi'_h(g)$ when ψ is differentiable. The corresponding candidate for the influence function $\dot{\psi}(x) = \psi'_F(\delta_x - F)$ is

$$\begin{aligned} &\int_0^t \frac{1}{H(s)} \, \mathrm{d}(N_{x,1} - H_1)(s) - \int_0^t \frac{Y_x(s) - H(s)}{H(s)^2} \, \mathrm{d}H_1(s) \\ &\quad = \int_0^t \frac{1}{H(s)} \, \mathrm{d}N_{x,1}(s) - \int_0^t \frac{Y_x(s)}{H(s)} \, \mathrm{d}\Lambda_1(s) = \int_0^t \frac{1}{H(s)} \, \mathrm{d}M_{x,1}(s), \end{aligned} \tag{2.9}$$

where $M_{x,1} = N_{x,1} - \int_0^{(\cdot)} Y_x(s) \, \mathrm{d}\Lambda_1(s)$. This matches the canonical gradient found by James (1997). In Section 4, we will introduce a framework where a functional like ψ will be differentiable in the sense of (2.4).

In similarity to an ordinary Taylor approximation, an estimator on the form $\phi(F_n)$ for a C^1 functional ϕ can be approximated by $\phi(F) + \phi'_F(F_n - F)$. This idea was first presented by von Mises (1947) and is often called the von Mises method. Reeds (1976) elaborated on this method focusing on Hadamard differentiable functionals. The idea is that if $\dot{\phi}$ can be shown to be a reasonable function, the asymptotic behavior of $\phi'_F(F_n - F) = \frac{1}{n}\sum_i \dot{\phi}(X_i)$ is given by a central limit theorem and a law of large numbers. This method is also related to the functional delta method; see, for example, Gill (1989).

If ϕ is two times differentiable, we can similarly introduce the second-order influence function given by

$$\ddot{\phi}(x_1, x_2) = \phi''_F(\delta_{x_1} - F, \delta_{x_2} - F), \qquad x_1, x_2 \in \mathcal{X}, \tag{2.10}$$

where ϕ''_F is the second-order derivative of ϕ at F. The second-order derivative is a bilinear, continuous map from $\mathbf{D}^2$ to $\mathbb{R}$, see the supplement [Overgaard, Parner and Pedersen (2017)] for more details. It is also symmetric in its arguments by Theorem 5.27 of Dudley and Norvaiša (2011), the Schwarz theorem. Note that due to (2.3) and the bilinearity of ϕ''_F, we have $\phi''_F(F_n - F, F_n - F) = \frac{1}{n^2}\sum_i\sum_j \ddot{\phi}(X_i, X_j)$. In Graw, Gerds and Schumacher (2009) the idea of analyzing the asymptotic behavior of the pseudo-observations by approximating $\phi(F_n)$ with

$$\begin{aligned} &\phi(F) + \phi'_F(F_n - F) + \frac{1}{2}\phi''_F(F_n - F, F_n - F) \\ &\quad = \theta + \frac{1}{n}\sum_i \dot{\phi}(X_i) + \frac{1}{2}\frac{1}{n^2}\sum_i\sum_j \ddot{\phi}(X_i, X_j) \end{aligned} \tag{2.11}$$

was presented in the special case of the Aalen–Johansen estimator for competing risks models. We will also use this second-order von Mises expansion, but we will keep track of the remainder term by using the integral representation in (1.10) of the supplement. As it turns out, a careful analysis of the remainder term (see Proposition 3.1) is essential for obtaining asymptotic normality of the estimating function, and thereby also the consistency and asymptotic normality of the parameter estimates from the estimating equation in (1.4) as described in Theorem 3.4.

A consequence of studying random elements of general Banach spaces is the fact that measurability is a more complicated matter on such a space. Often the map $x \mapsto \delta_x$ will not be Borel measurable in applications. But we do not require this kind of measurability in our approach. We require Borel measurability of $x \mapsto \dot{\phi}(x)$ and $(x_1, x_2) \mapsto \ddot{\phi}(x_1, x_2)$. A natural assumption to work under is that

$$(x_1, \dots, x_n) \mapsto \phi\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right) \qquad \text{is Borel measurable for any } n \in \mathbb{N}. \tag{2.12}$$

The measurability is with respect to the product σ-algebra $\mathcal{A}^n$ on $\mathcal{X}^n$, and the assumption will be fulfilled for any reasonable estimating functional since the assumption means measurability of the estimator. Specifically, it will be fulfilled for the estimators that we consider in Section 4. The estimators in Section 4 are all related to the Nelson–Aalen estimator of (2.7) in Example 2.2 which is seen to depend measurably on the observations through Y evaluated at the event times given by jumps in N_1. We argue in the supplement [Overgaard, Parner and Pedersen (2017)] that (2.12) yields measurability of $x \mapsto \dot{\phi}(x)$ and $(x_1, x_2) \mapsto \ddot{\phi}(x_1, x_2)$ under the assumption that F is in fact a limit of $(\frac{1}{n}\sum_{i=1}^n \delta_{x_i})$ in $\mathbf{D}$ for some choice of sequence (x_i). This limit assumption is implied by the assumptions of Proposition 3.1 below, which we are going to work under. To keep things simple, we will also assume that

$$\|\delta_X\| \quad \text{and} \quad \|F_n - F\| \qquad \text{are random variables,} \tag{2.13}$$

although most statements concerning these variables could be made in terms of outer probability with slight modification. Both $\|\delta_X\|$ and $\|F_n - F\|$ are random variables in our applications as we shall see.

3. Pseudo-observations and asymptotics. To formalize the setting from the Introduction, we consider a triple $(V, X, Z) \in \mathbb{R} \times \mathcal{X} \times \mathcal{Z}$ of random variables, where V is not entirely observable, X is observable, and Z represents covariates. For simplicity, we consider a real-valued V, but the following results extend to a vector-valued V. We model $\mathrm{E}(V|Z)$ by $\mu(\beta_0; Z)$ with an unknown parameter vector $\beta_0 \in \mathbb{R}^q$ for some appropriate function μ, and we want to estimate β_0. Consider i.i.d. copies $(V_1, X_1, Z_1), \ldots, (V_n, X_n, Z_n)$ of (V, X, Z), where $(X_1, Z_1), \ldots, (X_n, Z_n)$ is the observed sample. We will assume that we have an estimator of $\theta = \mathrm{E}(V) \in \mathbb{R}$ that can be considered a function of a sample average $F_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i} \in \mathbf{D}$, as in (2.3), for some choice of $\delta_{(\cdot)}$ and Banach space $\mathbf{D}$, that is, that the estimates are $\hat{\theta}_n = \phi(F_n)$ for some functional ϕ from (an appropriate subset of) $\mathbf{D}$ to the real line.

With the setting introduced above, we will next give a definition of the pseudo-observations, which will replace the V's in the estimating equation. We will then decompose the pseudo-observations into their essential parts and remainder parts. When the remainder terms vanish appropriately, the essential parts lead to a central limit theorem that is useful for the estimating functions that we are going to consider. Under standard regularity conditions, this leads to the consistency and asymptotic normality of the corresponding estimates of β_0.

Let the empirical distribution leaving out the kth observation be denoted by

$$F_n^{(k)} = \frac{1}{n-1}\sum_{i \neq k} \delta_{X_i}. \tag{3.1}$$

Then the leave-k-out estimate is similarly

$$\hat{\theta}_n^{(k)} = \phi\big(F_n^{(k)}\big). \tag{3.2}$$

In this case the jack-knife pseudo-observations are defined as

$$\hat{\theta}_{n,k} = n\hat{\theta}_n - (n-1)\hat{\theta}_n^{(k)} = n\phi(F_n) - (n-1)\phi\big(F_n^{(k)}\big), \tag{3.3}$$

for $k = 1, \dots, n$.

When ϕ is C^2, that is, two times continuously differentiable, we can approximate each of $\phi(F_n)$ and $\phi(F_n^{(k)})$ by a second-order Taylor expansion as in (1.10) of the supplement [Overgaard, Parner and Pedersen (2017)]. This allows for a decomposition of a pseudo-observation into an essential part and a remainder,

$$\hat{\theta}_{n,k} = \hat{\theta}^*_{n,k} + R_{n,k}. \tag{3.4}$$

The remainder, $R_{n,k}$, will correspond to remainder terms and some second-order terms from the Taylor expansions. How fast, if at all, this remainder converges to 0 will depend on the convergence of $\|F_n - F\|$ in similarity to an ordinary Taylor expansion. A precise statement is found in Proposition 3.1 below. As we shall see later on in (3.42), a sufficient condition for the remainder to be ignorable in our estimating procedure is that (3.8) below holds for $\lambda = \frac{1}{4}$, and we can always have that number in mind. Also, our applications use the simpler (3.5) version of condition (a) since $\|\delta_x\|$ is related to the number of state transitions in our setting, and there will only be one state transition in the competing risks setting, which gives us an upper bound on $\|\delta_x\|$. The (3.6) version of condition (a) will be useful in the general multi-state setting, where such a bound may not exist, but where the probability of having many transitions drops fast.

PROPOSITION 3.1. *Assume an $F \in \mathbf{D}$ exists such that*

(a) *there is a $\lambda \in [\frac{1}{4}, \frac{1}{2})$ and a $c > 0$ such that*

$$\|F_n - F\| = o_{\mathrm{P}}(n^{-\lambda}) \quad \text{and} \quad \|\delta_x\| \le c \qquad \text{for all } x \in \mathcal{X}, \tag{3.5}$$

or there is a $\lambda \in [\frac{1}{4}, \frac{1}{2})$ and a $\xi \in (0, \frac{1}{2} - \lambda]$ such that

$$\|F_n - F\| = o_{\mathrm{P}}(n^{-(\lambda + \frac{\xi}{2})}) \quad \text{and} \quad \lim_{y \to \infty} y^{\frac{1}{\xi}} \mathrm{P}(\|\delta_X\| > y) = 0, \tag{3.6}$$

(b) *ϕ is C^2 in a neighborhood of F and the second-order derivative of ϕ is Lipschitz continuous in a neighborhood of F.*

Consider a decomposition of the pseudo-observation $\hat{\theta}_{n,k}$ into $\hat{\theta}^*_{n,k} + R_{n,k}$ with the essential part

$$\hat{\theta}^*_{n,k} = \phi(F) + \phi'_F(\delta_{X_k} - F) + \phi''_F(\delta_{X_k} - F, F_n^{(k)} - F). \tag{3.7}$$

Then

$$n^{2\lambda} \max_k |R_{n,k}| \to 0 \tag{3.8}$$

in probability for $n \to \infty$.

PROOF. Consider the expansions

$$
\begin{aligned}
\phi(F_n) &= \phi(F) + \phi'_F(F_n - F) + \frac{1}{2}\phi''_F(F_n - F, F_n - F) \\
&\quad + \int_0^1 (1-s)\big(\phi''_{F_{n,s}} - \phi''_F\big)(F_n - F, F_n - F)\,\mathrm{d}s
\end{aligned}
\tag{3.9}
$$

and

$$
\begin{aligned}
\phi\big(F_n^{(k)}\big) &= \phi(F) + \phi'_F\big(F_n^{(k)} - F\big) + \frac{1}{2}\phi''_F\big(F_n^{(k)} - F, F_n^{(k)} - F\big) \\
&\quad + \int_0^1 (1-s)\big(\phi''_{F_{n,s}^{(k)}} - \phi''_F\big)\big(F_n^{(k)} - F, F_n^{(k)} - F\big)\,\mathrm{d}s,
\end{aligned}
\tag{3.10}
$$

where $F_{n,s} = F + s(F_n - F)$ and $F_{n,s}^{(k)} = F + s(F_n^{(k)} - F)$, which are just versions of (1.10) of the supplement and will be valid with high probability for large n due to (a).

Using the two expansions of (3.9) and (3.10), we see that the 0th-order derivative term of $n\phi(F_n) - (n-1)\phi(F_n^{(k)})$ is $\phi(F)$, and that the first-order derivative term, by linearity of ϕ'_F, is

$$
n\phi'_F(F_n - F) - (n-1)\phi'_F\big(F_n^{(k)} - F\big) = \phi'_F(\delta_{X_k} - F).
\tag{3.11}
$$

Both are included in the the essential part of the pseudo-observation.

As for the second-order derivative terms, note the identity

$$
F_n^{(k)} - F = (F_n - F) + \frac{1}{n-1}(F_n - \delta_{X_k}),
\tag{3.12}
$$

such that

$$
\begin{aligned}
&\frac{n}{2}\phi''_F(F_n - F, F_n - F) - \frac{n-1}{2}\phi''_F\big(F_n^{(k)} - F, F_n^{(k)} - F\big) \\
&\quad = \frac{1}{2}\phi''_F(F_n - F, F_n - F) + \frac{1}{2}\frac{1}{n-1}\phi''_F(\delta_{X_k} - F_n, \delta_{X_k} - F_n) \\
&\qquad + \phi''_F\big(F - F_n, F_n^{(k)} - F\big) + \phi''_F\big(\delta_{X_k} - F, F_n^{(k)} - F\big)
\end{aligned}
\tag{3.13}
$$

and

$$
\begin{aligned}
&n\big(\phi''_{F_{n,s}} - \phi''_F\big)(F_n - F, F_n - F) \\
&\quad - (n-1)\big(\phi''_{F_{n,s}^{(k)}} - \phi''_F\big)\big(F_n^{(k)} - F, F_n^{(k)} - F\big) \\
&\quad = \big(\phi''_{F_{n,s}} - \phi''_F\big)(F_n - F, F_n - F) \\
&\qquad + (n-1)\big(\phi''_{F_{n,s}} - \phi''_{F_{n,s}^{(k)}}\big)(F_n - F, F_n - F) \\
&\qquad + \big(\phi''_{F_{n,s}^{(k)}} - \phi''_F\big)(F_n - F, \delta_{X_k} - F_n) \\
&\qquad + \big(\phi''_{F_{n,s}^{(k)}} - \phi''_F\big)\big(\delta_{X_k} - F_n, F_n^{(k)} - F\big),
\end{aligned}
\tag{3.14}
$$

with only the last term from (3.13) included in the essential part and all other terms going into the remainder. To sum up, the remainder is

$$
\begin{aligned}
R_{n,k} = {}& \frac{1}{2}\phi''_F(F_n - F, F_n - F) + \frac{1}{2}\frac{1}{n-1}\phi''_F(\delta_{X_k} - F_n, \delta_{X_k} - F_n) \\
& + \phi''_F\big(F - F_n, F_n^{(k)} - F\big) \\
& + \int_0^1 (1-s)\big(\phi''_{F_{n,s}} - \phi''_F\big)(F_n - F, F_n - F)\,\mathrm{d}s \\
& + \int_0^1 (1-s)(n-1)\big(\phi''_{F_{n,s}} - \phi''_{F_{n,s}^{(k)}}\big)(F_n - F, F_n - F)\,\mathrm{d}s \\
& + \int_0^1 (1-s)\big(\phi''_{F_{n,s}^{(k)}} - \phi''_F\big)(F_n - F, \delta_{X_k} - F_n)\,\mathrm{d}s \\
& + \int_0^1 (1-s)\big(\phi''_{F_{n,s}^{(k)}} - \phi''_F\big)\big(\delta_{X_k} - F_n, F_n^{(k)} - F\big)\,\mathrm{d}s.
\end{aligned} \tag{3.15}
$$

Using the assumptions and in particular the Lipschitz continuity of the second-order derivative, there exists a constant $K > 0$ such that for large n with high probability

$$
\begin{aligned}
|R_{n,k}| \le {}& \frac{K}{2}\|F_n - F\|^2 + \frac{1}{2}\frac{K}{n-1}\Big(2\max_i \|\delta_{X_i}\|\Big)^2 \\
& + K\|F - F_n\|\big\|F_n^{(k)} - F\big\| \\
& + K \sup_{s\in[0,1]} \|F_{n,s} - F\|\|F_n - F\|^2 \\
& + K(n-1) \sup_{s\in[0,1]} \big\|F_{n,s} - F_{n,s}^{(k)}\big\|\|F_n - F\|^2 \\
& + 2K \max_i \|\delta_{X_i}\| \sup_{s\in[0,1]} \big\|F_{n,s}^{(k)} - F\big\|\|F_n - F\| \\
& + 2K \max_i \|\delta_{X_i}\| \sup_{s\in[0,1]} \big\|F_{n,s}^{(k)} - F\big\|\big\|F_n^{(k)} - F\big\|.
\end{aligned} \tag{3.16}
$$

Since

$$
F_{n,s} - F_{n,s}^{(k)} = \frac{s}{n-1}(\delta_{X_k} - F_n), \qquad s \in [0,1], \tag{3.17}
$$

we have

$$
\big\|F_{n,s} - F_{n,s}^{(k)}\big\| \le \frac{2}{n-1}\max_i \|\delta_{X_i}\| \tag{3.18}
$$

and

$$
\big\|F_{n,s}^{(k)} - F\big\| \le \frac{2}{n-1}\max_i \|\delta_{X_i}\| + \|F_n - F\|, \tag{3.19}
$$

and so we see that $\max_k |R_{n,k}|$ is of the order

$$O_{\mathrm{P}}\biggl(\Bigl(1+\max_i \|\delta_{X_i}\|\Bigr)\|F_n - F\|^2 + \frac{1}{n}\max_i \|\delta_{X_i}\|^2\bigl(1+\|F_n - F\|\bigr)\biggr) + O_{\mathrm{P}}\biggl(\frac{1}{n^2}\max_i \|\delta_{X_i}\|^3\biggr). \tag{3.20}$$

Under (3.6), we have for any $\varepsilon > 0$ that $\mathrm{P}(\frac{1}{n^\xi}\max_i \|\delta_{X_i}\| > \varepsilon) \le n\,\mathrm{P}(\|\delta_X\| > \varepsilon n^\xi) \to 0$ for $n \to \infty$ by subadditivity. In other words, $\max_i \|\delta_{X_i}\| = o_{\mathrm{P}}(n^\xi)$ for $n \to \infty$. Under (3.5), $\max_i \|\delta_{X_i}\|$ is bounded by c. In both cases, the convergence order in (3.20) is put together at least $n^{-2\lambda}$, and the result follows. □

In (3.42) below, it is important that the convergence of $|R_{n,k}|$ is uniform in k. As can be seen in the proof, essentially the inequality in (3.16), the Lipschitz continuity of the second-order derivative allows us to obtain that uniformity from the uniform closeness of $F_n^{(k)}$ to F_n in (3.18) and thereby to F in (3.19) under our assumptions. As noted by Jacobsen and Martinussen (2016) in the case of the Kaplan–Meier functional, the term $\phi_F''(\delta_{X_k} - F, F_n^{(k)} - F)$ is generally not small enough to be ignored, that is, we should not consider it a part of the remainder term, if the remainder term is to satisfy (3.8).

The challenge posed by Proposition 3.1 is a balancing act. The norm we consider should be strong enough that the functionals of importance are sufficiently smooth [condition (b)], but weak enough that $\|F_n - F\|$ converges sufficiently fast [condition (a)].

EXAMPLE 3.2. Consider the setting of Example 2.1. We now want to decide on which Banach space, $\mathbf{D}$, we are going to consider F_n a member of. A first attempt could be the space of uniformly bounded vector functions endowed with the supremum norm. Both F_n and F are certainly members. But this supremum norm approach is too weak to make the functionals we are interested in sufficiently smooth [e.g., even if f_0 and g_0 are sufficiently well behaved that $\int f_0\,\mathrm{d}g_0$ is meaningful, the functional $(f, g) \mapsto \int f\,\mathrm{d}g$ can be difficult to define in a supremum norm neighborhood of (f_0, g_0) and it will definitely not be continuous]. A second attempt could be to consider the space of functions of bounded variation (in each coordinate) with the corresponding total variation norm. But this approach is too strong since F_n does not converge to F in total variation norm. To succeed in the challenge of the balancing act posed by Proposition 3.1, we should likely be looking for something in between. And as it turns out, the p-variation approach, lodged between the supremum norm approach and the total variation norm approach, solves the problem. The p-variation of a function $f\colon [a, b] \to \mathbb{R}$ is defined by $v_p(f; [a, b]) = \sup \sum_{i=1}^m |f(x_{i-1}) - f(x_i)|^p$ where the supremum is over $m \in \mathbb{N}$ and points $x_0 < x_1 < \cdots < x_m$ in the interval $[a, b]$. We consider only $p \in [1, 2)$ in the following. The space of functions of bounded p-variation, $\mathcal{W}_p([a, b])$ or in short $\mathcal{W}_p$, is a Banach space when endowed with the p-variation norm, $\|f\|_{[p]} = v_p(f; [a, b])^{\frac{1}{p}} + \|f\|_\infty$, where $\|\cdot\|_\infty$ is the supremum norm. We see that F_n and F can be considered members of $\mathcal{W}_p^{d+2}$. The product space $\mathcal{W}_p^{d+2}$ is a Banach space when endowed with a norm that is the sum of the entry-wise p-variation norms. This norm is still denoted by $\|\cdot\|_{[p]}$. We argue in the supplement that

$$\|F_n - F\|_{[p]} = O_P\big(n^{\frac{1-p}{p}} (\log\log n)^{\frac{1}{2}}\big) \tag{3.21}$$

for $1 \leq p < 2$ based on a similar result for one-dimensional distribution functions stated in, for example, Dudley and Norvaiša (1999). Note that $\|\delta_X\|$ and $\|F_n - F\|$ are random variables [i.e., that (2.13) is met] since both δ_X and $F_n - F$ consist of one left-continuous entry and $d + 1$ right-continuous entries, whose p-variations can be obtained by suprema over the rationals. Since $\|\delta_x\|_{[p]} \leq 2(d + 2)$ for any x (and any p), we can establish that (3.5) in (a) of Proposition 3.1 is met for $\lambda = \frac{1}{4}$ if we take $\frac{4}{3} < p < 2$. Since we can take λ arbitrarily close to $\frac{1}{2}$, if we take p close to 2, the generality of Proposition 3.1 offers the extra piece of information that the remainders vanish uniformly fairly rapidly for a functional satisfying condition (b). The supplement [Overgaard, Parner and Pedersen (2017)] offers more details on p-variation, referencing Dudley and Norvaiša (2011), and we will also return to this in Section 4.

Note that the assumption of Proposition 3.1 that the functional is C^2 with a derivative that is Lipschitz continuous in a neighborhood of F is implied if ϕ is C^3 according to Proposition 1.4 of the supplement [Overgaard, Parner and Pedersen (2017)]. The functionals we are considering in Section 4 are C^k for any k, so this assumption is automatically met.

The essential part in (3.7) can also be written

$$\hat{\theta}^*_{n,k} = \phi(F) + \dot{\phi}(X_k) + \frac{1}{n-1} \sum_{i \neq k} \ddot{\phi}(X_k, X_i) \tag{3.22}$$

in terms of the influence functions. Under the assumption in (b), we have for some $K > 0$ that $|\dot{\phi}(x)| = |\phi'_F(\delta_x - F)| \leq K\|\delta_x - F\| \leq K(\|\delta_x\| + \|F\|)$, and $|\ddot{\phi}(x_1, x_2)| = |\phi''_F(\delta_{x_1} - F, \delta_{x_2} - F)| \leq K(\|\delta_{x_1}\| + \|F\|)(\|\delta_{x_2}\| + \|F\|)$, such that $\dot{\phi}(X)$, $\ddot{\phi}(X, x)$, and $\ddot{\phi}(X_1, X_2)$ have mth moment whenever $\|\delta_X\|$ has mth moment [and $\ddot{\phi}(X, X)$ has mth moment whenever $\|\delta_X\|$ has $(2m)$th moment]. Under (3.5), $\|\delta_X\|$ has any moment, and under (3.6) $\|\delta_X\|$ has mth moment for all $m < \frac{1}{\xi}$, which can be seen by using the formula $\mathrm{E}(\|\delta_X\|^m) = m \int_0^\infty y^{m-1} \mathrm{P}(\|\delta_X\| > y)\, \mathrm{d}y$. This means that the influence functions are at least integrable under the assumptions of

Proposition 3.1, and under the same assumptions, they are necessarily centered in the sense that

$$\mathrm{E}(\dot{\phi}(X)) = 0, \tag{3.23}$$

$$\mathrm{E}(\ddot{\phi}(X, x)) = 0 \qquad \text{for all } x \in \mathcal{X}, \tag{3.24}$$

as, for some $K > 0$, $|\frac{1}{n}\sum_i \dot{\phi}(X_i)| = |\phi'_F(F_n - F)| \leq K\|F_n - F\| \to 0$ in probability for $n \to \infty$, and similarly for $\ddot{\phi}$.

As a preliminary result to handle estimating equations using the jack-knife pseudo-observations we prove the following central limit theorem.

THEOREM 3.3. *Make the assumptions of Proposition 3.1 with respect to a functional ϕ and consider the essential parts of the corresponding pseudo-observations as in (3.7). For some $q \in \mathbb{N}$, let $A \colon \mathcal{Z} \to \mathbb{R}^q$ be a function such that $A(Z)$ is a column vector which has finite second moment if ϕ meets condition (3.5) or has finite rth moment for an $r > \frac{2}{1-2\xi}$ if ϕ meets condition (3.6). Define*

$$U_n^* = \sum_{k=1}^{n} A(Z_k)\big(\hat{\theta}_{n,k}^* - \mathrm{E}(\hat{\theta}_{n,k}^* | Z_k)\big). \tag{3.25}$$

Then

$$\frac{1}{\sqrt{n}} U_n^* \xrightarrow{d} N(0, \Sigma), \tag{3.26}$$

where

$$\Sigma = \mathrm{Var}\big(A(Z)\big(\dot{\phi}(X) - \mathrm{E}(\dot{\phi}(X)|Z)\big) + h_1(X)\big) \tag{3.27}$$

is the variance of the asymptotic distribution with $h_1(x) = \mathrm{E}(A(Z)\ddot{\phi}(X, x))$. Alternatively, the variance can be expressed as

$$\Sigma = \mathrm{E}\big(h(X_1, Z_1, X_2, Z_2) h(X_1, Z_1, X_3, Z_3)^T\big), \tag{3.28}$$

where

$$\begin{aligned} h(x_1, z_1, x_2, z_2) = {} & \sum_{i=1}^{2} A(z_i)\big(\dot{\phi}(x_i) - \mathrm{E}(\dot{\phi}(X)|Z = z_i)\big) \\ & + \big(A(z_1) + A(z_2)\big)\ddot{\phi}(x_1, x_2). \end{aligned} \tag{3.29}$$

PROOF. This result is proved by recognizing U_n^* as a second-order U-statistic and applying results that apply to those. This is essentially the course taken in Jacobsen and Martinussen (2016) to prove asymptotic normality of the estimating function in the case of the Kaplan–Meier estimator.

Since $\mathrm{E}(\hat{\theta}^*_{n,k}|Z_k = z) = \phi(F) + \mathrm{E}(\dot{\phi}(X)|Z = z)$ by (3.22) and (3.24), we are looking at

$$\sum_{k=1}^{n} A(Z_k)\big(\dot{\phi}(X_k) - \mathrm{E}\big(\dot{\phi}(X_k)|Z_k\big) + \phi''_F\big(\delta_{X_k} - F, F_n^{(k)} - F\big)\big)$$

$$(3.30)\qquad = \sum_{k=1}^{n} A(Z_k)\big(\dot{\phi}(X_k) - \mathrm{E}\big(\dot{\phi}(X_k)|Z_k\big)\big) + \frac{1}{n-1}\sum_{k=1}^{n}\sum_{j\neq k} A(Z_k)\ddot{\phi}(X_k, X_j),$$

which can also be expressed as

$$(3.31)\qquad n\frac{1}{\binom{n}{2}}\sum_{k=1}^{n}\sum_{j<k}\frac{1}{2}h(X_i, Z_i, X_j, Z_j).$$

The factor n aside, this is a U-statistic of order 2. Note that $\ddot{\phi}(X_k, X_j)$, $\dot{\phi}(X_k)$, and hence $\mathrm{E}(\dot{\phi}(X_k)|Z_k)$, have rth moment for any $r < \frac{1}{\xi}$ according to the moment considerations earlier on when condition (3.6) is met, and that they have any moment when (3.5) is met. Under the moment conditions on $A(Z)$, a Hölder inequality gives us that the U-statistic above has finite second moment. As the mean of the U-statistic is $\mathrm{E}(h(X_1, Z_1, X_2, Z_2)) = 0$, the result with the variance expression from (3.28) follows from a multivariate version of Theorem 12.3 of van der Vaart (1998). Note that

$$(3.32)\qquad \begin{aligned} &\mathrm{E}\big(h(X_1, Z_1, X_2, Z_2)h(X_1, Z_1, X_3, Z_3)^T|X_1 = x, Z_1 = z\big) \\ &\quad = \mathrm{E}\big(h(x, z, X, Z)\big)\,\mathrm{E}\big(h(x, z, X, Z)\big)^T \end{aligned}$$

and

$$(3.33)\qquad \begin{aligned} &\mathrm{E}\big(h(x, z, X, Z)\big) \\ &\quad = A(z)\big(\dot{\phi}(x) - \mathrm{E}\big(\dot{\phi}(X)|Z = z\big)\big) \\ &\qquad + \mathrm{E}\big(A(Z)\big(\dot{\phi}(X) - \mathrm{E}\big(\dot{\phi}(X)|Z\big)\big)\big) \\ &\qquad + A(z)\,\mathrm{E}\big(\ddot{\phi}(X, x)\big) + \mathrm{E}\big(A(Z)\ddot{\phi}(X, x)\big) \\ &\quad = A(z)\big(\dot{\phi}(x) - \mathrm{E}\big(\dot{\phi}(X)|Z = z\big)\big) + h_1(x), \end{aligned}$$

using (3.24), such that the variance expression in (3.27) matches the one in (3.28). □

We are going to consider estimating functions of the type

$$(3.34)\qquad U_n(\beta) = \sum_k A(\beta; Z_k)\big(\hat{\theta}_{n,k} - \mu(\beta; Z_k)\big),$$

where μ is the function that models $\mathrm{E}(V|Z)$, that is,

$$\mu(\beta_0; z) = \mathrm{E}(V|Z = z), \tag{3.35}$$

for a q-dimensional parameter β_0 as described in the beginning of this section and where A is a q-dimensional function.

We are now ready to state our main theorem on consistency and asymptotic normality of estimates obtained from the pseudo-observation method. Other than the assumptions of Proposition 3.1, the assumption of (3.36) below is the key. It establishes a close connection between the estimating functional and V.

THEOREM 3.4. *Make the assumptions of Proposition 3.1 with respect to a functional ϕ. Consider U_n in (3.34) with pseudo-observations, $\hat{\theta}_{n,k}$, based on estimates, $\hat{\theta}_n = \phi(F_n)$, from that functional. Assume that*

$$\mathrm{E}\big(\dot{\phi}(X)|Z = z\big) = \mathrm{E}(V|Z = z) - \phi(F), \tag{3.36}$$

and that the following regularity conditions are met:

1. *$\mu(\cdot; z)$ and $A(\cdot; z)$ are continuously differentiable for (almost) all $z \in \mathcal{Z}$,*
2. *$A(\beta_0, Z)$ has finite second moment if condition (3.5) is met or has finite rth moment for an $r > \frac{2}{1-2\xi}$ if condition (3.6) is met,*
3. *$\frac{\partial}{\partial\beta} A(\beta; Z)\mu(\beta; Z)$ and $A(\beta; Z)\frac{\partial}{\partial\beta}\mu(\beta; Z)$ are dominated integrable in a neighborhood of β_0,*
4. *$|\frac{\partial}{\partial\beta} A(\beta; Z)|^r$ is dominated integrable in a neighborhood of β_0 for $r = 1$ if ϕ meets condition (3.5) or for an $r > \frac{1}{1-\xi}$ if ϕ meets condition (3.6),*
5. *the matrix*

$$M = \mathrm{E}\bigg(A(\beta_0; Z)\frac{\partial}{\partial\beta}\mu(\beta; Z)\Big|_{\beta=\beta_0}\bigg) \tag{3.37}$$

has full rank.

Then for every n an estimator $\hat{\beta}_n$ exists such that $U_n(\hat{\beta}_n) = 0$ with a probability tending to 1 for $n \to \infty$. Moreover, $\hat{\beta}_n \to \beta_0$ in probability and

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N\big(0, M^{-1}\Sigma\big(M^{-1}\big)^T\big) \tag{3.38}$$

as $n \to \infty$, with

$$\Sigma = \mathrm{Var}\big(A(\beta_0; Z)\big(\phi(F) + \dot{\phi}(X) - \mu(\beta_0; Z)\big) + h_1(X)\big), \tag{3.39}$$

where $h_1(x) = \mathrm{E}(A(\beta_0; Z)\ddot{\phi}(X, x))$.

PROOF. Let us first realize that $\frac{1}{\sqrt{n}} U_n(\beta_0) \xrightarrow{d} N(0, \Sigma)$. Using the decomposition from Proposition 3.1, consider

$$U_n^* = \sum_{k=1}^{n} A(\beta_0; Z_k)\big(\hat{\theta}_{n,k}^* - \mathrm{E}\big(\hat{\theta}_{n,k}^*|Z_k\big)\big). \tag{3.40}$$

Using (3.22) and (3.24),

$$
\begin{aligned}
\mathrm{E}(\hat{\theta}^*_{n,k} \mid Z_k = z) &= \phi(F) + \mathrm{E}(\dot{\phi}(X) \mid Z = z) \\
&= \phi(F) + \mathrm{E}(V \mid Z = z) - \phi(F) = \mu(\beta_0; z)
\end{aligned}
\tag{3.41}
$$

under the assumption in (3.36). Thus, we see that

$$
\begin{aligned}
\left| \frac{1}{\sqrt{n}} U_n(\beta_0) - \frac{1}{\sqrt{n}} U_n^* \right| &= \frac{1}{\sqrt{n}} \left| \sum_{k=1}^n A(\beta_0; Z_k)(\hat{\theta}_{n,k} - \hat{\theta}^*_{n,k}) \right| \\
&\le \sqrt{n} \max_k |R_{n,k}| \frac{1}{n} \sum_k |A(\beta_0; Z_k)| = o_{\mathrm{P}}(n^{\frac{1}{2} - 2\lambda}),
\end{aligned}
\tag{3.42}
$$

as $\max_k |R_{n,k}| = o_{\mathrm{P}}(n^{-2\lambda})$ by Proposition 3.1 under the assumptions, and $\frac{1}{n} \sum_k |A(\beta_0; Z_k)| = O_{\mathrm{P}}(1)$ by the regularity assumptions. Since $\frac{1}{2} - 2\lambda \le 0$, $\frac{1}{\sqrt{n}} U_n(\beta_0)$ and $\frac{1}{\sqrt{n}} U_n^*$ are asymptotically equivalent, and the asymptotic normality of $\frac{1}{\sqrt{n}} U_n(\beta_0)$ follows from Theorem 3.3.

With the asymptotic normality of the estimating function, the theorem is a rather standard statement. See, for example, Sørensen (1999) for further details. Note that

$$
\begin{aligned}
\frac{\partial}{\partial \beta} \left(\frac{1}{n} U_n(\beta) \right) &= \frac{1}{n} \sum_{k=1}^n \frac{\partial}{\partial \beta} (A(\beta; Z_k)) (\hat{\theta}_{n,k} - \mu(\beta; Z_k)) \\
&\quad - \frac{1}{n} \sum_{k=1}^n A(\beta; Z_k) \frac{\partial}{\partial \beta} \mu(\beta; Z_k),
\end{aligned}
\tag{3.43}
$$

and that according to a uniform law of large numbers for U-statistics [a law of large numbers for U-statistics can be found in Problem 15 of Chapter 12 of van der Vaart (1998)], this will converge (almost surely) to $-M$ uniformly on balls shrinking towards $\{\beta_0\}$ as $n \to \infty$ under the regularity conditions since we can again approximate $\hat{\theta}_{n,k}$ by $\hat{\theta}^*_{n,k}$ uniformly well and $\mathrm{E}(\hat{\theta}^*_{n,k} \mid Z_k) = \mu(\beta_0; Z_k)$ as established above. □

The matrix M is easily estimated by an empirical mean,

$$
\hat{M} = \frac{1}{n} \sum_{i=1}^n A(\hat{\beta}_n; Z_i) \frac{\partial}{\partial \beta} \mu(\beta; Z_i) \Big|_{\beta = \hat{\beta}_n}
\tag{3.44}
$$

at the estimate $\hat{\beta}_n$. If a consistent estimate $\hat{\Sigma}$ of Σ can be obtained, the sandwich variance estimator $\hat{M}^{-1} \hat{\Sigma} \hat{M}^{-1}$ estimates $M^{-1} \Sigma M^{-1}$ consistently. The usual estimator of Σ, the one suggested by Andersen, Klein and Rosthøj (2003),

$$
\frac{1}{n} \sum_{k=1}^n A(\hat{\beta}_n; Z_k) A(\hat{\beta}_n; Z_k)^T (\hat{\theta}_{n,k} - \mu(\hat{\beta}_n; Z_k))^2
\tag{3.45}
$$

can be seen to converge in probability to $\mathrm{Var}(A(\beta_0; Z)(\phi(F)+\dot{\phi}(X)-\mu(\beta_0; Z)))$, leaving out the $h_1(X)$ part of (3.39), and as such it is in general a biased estimate of Σ. Equation (3.39) suggests estimating Σ by

$$\begin{aligned}(3.46)\quad &\frac{1}{n}\sum_{k=1}^{n}\big(A(\hat{\beta}_n; Z_k)\big(\phi(F_n)+\phi'_{F_n}(\delta_{X_k}-F_n)-\mu(\hat{\beta}_n; Z_k)\big)+\hat{h}_1(X_k)\big)\\ &\qquad\times\big(A(\hat{\beta}_n; Z_k)\big(\phi(F_n)+\phi'_{F_n}(\delta_{X_k}-F_n)-\mu(\hat{\beta}_n; Z_k)\big)+\hat{h}_1(X_k)\big)^T,\end{aligned}$$

where $\hat{h}_1(x)=\frac{1}{n}\sum_{j=1}^{n}A(\hat{\beta}_n; Z_j)\phi''_{F_n}(\delta_x-F_n,\delta_{X_j}-F_n)$. This estimator is of an abstract nature. But since ϕ, ϕ', and ϕ'' will be known, or can be derived, in applications, it certainly can be obtained in practice.

4. Applications in time-to-event analysis. We are going to have a look at different pseudo-observation-based analyses that can be studied in our framework, and we will see how they fit the conditions of Theorem 3.4.

We will consider the time-to-event setting with competing risks as described in Example 2.1. The pseudo-observations are in all cases based on the sample of $X=(\tilde{T},\tilde{\Delta})$, a pair of the exit time $\tilde{T}$ and the status at exit $\tilde{\Delta}$. We imagine such an observation has come about due to a censoring time C censoring the actual event time T and event type Δ, that is, that $\tilde{T}=T\wedge C$ and $\tilde{\Delta}=\Delta 1(T\le C)$. Recall that we consider $\delta_x=(Y_x,N_{x,0},\dots,N_{x,d})^T$ such that $F_n=\frac{1}{n}\sum\delta_{X_i}$ is a vector of step functions. The expected limit of F_n is $F=(H,H_0,\dots,H_d)^T$, where $H(s)=\mathrm{P}(\tilde{T}\ge s)$ and $H_j(s)=\mathrm{P}(\tilde{T}\le s,\tilde{\Delta}=j)$. We use the notation $S(s)=\mathrm{P}(T>s)$, $G(s)=\mathrm{P}(C>s)$, $F_j(s)=\mathrm{P}(T\le s,\Delta=j)$ for $j=1,\dots,d$, and $\Lambda_j(s)=\int_0^s\frac{1}{H(u)}\,\mathrm{d}H_j(u)$ and $M_{x,j}(s)=N_{x,j}(s)-\int_0^s Y_x(u)\,\mathrm{d}\Lambda_j(u)$ for $j=0,\dots,d$. In our applications, we will be interested in a particular timepoint $t>0$ and it will suffice to let all these functions be defined on the interval $[0,t]$. We will assume:

1. Continuity of F. This implies continuity of S, G and Λ_j, etc.
2. Positivity of $H(t)$, that is, $H(t)>0$ such that it is possible to remain at risk up to time t.
3. Completely independent censorings in the sense that C is independent of (T,Δ,Z).

The last assumption is perhaps the most restrictive and it will be emphasized in the following. The assumption was key in Graw, Gerds and Schumacher (2009).

We will consider the Banach space $\mathbf{D}=\mathcal{W}_p^{d+2}([0,t])$ for a $p\in(\frac{4}{3},2)$ as described in Example 3.2, and we will see that the conditions of Proposition 3.1 are met for our functionals of interest. As argued in the example, condition (a) [version (3.5)] is met in this setting.

In the supplement [Overgaard, Parner and Pedersen (2017)], we have listed several differentiable maps between $\mathcal{W}_p$-type spaces based on Dudley and Norvaiša (2011), including:

1. $f \mapsto \frac{1}{f}$ defined on an open set of functions in $\mathcal{W}_p$ bounded away from 0. Its first-order derivative at f is $g \mapsto -\frac{g}{f^2}$.
2. $(f, g) \mapsto \int_0^{(\cdot)} f(s)\,\mathrm{d}g(s)$ (from $\mathcal{W}_p^2$ to $\mathcal{W}_p$), the integration operator, with derivative at (f, g) being $(u, v) \mapsto \int_0^{(\cdot)} u(s)\,\mathrm{d}g(s) + \int_0^{(\cdot)} f(s)\,\mathrm{d}v(s)$. Note that the ordinary Lebesgue–Stieltjes integrals may not be well defined (when the integrator is not of bounded variation). Integrals should be considered central Young integrals.
3. $f \mapsto \prod_0^{(\cdot)}(1 + \mathrm{d}f(s))$ (from $\mathcal{W}_p$ to $\mathcal{W}_p$), the product integral operator. The derivative at a continuous f is $g \mapsto e^f(g(\cdot) - g(0))$.

Also, a continuous, linear map ϕ from $\mathcal{W}_p$, for example, $f \mapsto f_- \in \mathcal{W}_p$, mapping to the left-continuous version, and $f \mapsto f(t) \in \mathbb{R}$, a coordinate projection, is differentiable of any order with first-order derivative given by $\phi'_f(g) = \phi(g)$. Using the chain rule, we can use these basic differentiable functionals to establish differentiability of any order of a range of important functionals.

EXAMPLE 4.1. Consider again the Nelson–Aalen functional from Example 2.2. In the setting introduced in this section, the functional, given by $\psi(t; h) = \int_0^t \frac{1(h_*(s)>0)}{h_*(s)}\,\mathrm{d}h_1(s)$ for $h = (h_*, h_0, \dots, h_d)^T \in \mathcal{W}_p^{d+2}$, is differentiable of any order in a neighborhood of F (under the positivity assumption) since it is a composition of differentiable operations. Using the chain rule (see the supplement) and the derivatives of the elementary functionals stated above, the first-order derivative is seen to be given by

$$\psi'_h(t; g) = \int_0^t \frac{1}{h_*(s)}\,\mathrm{d}g_1(s) - \int_0^t \frac{g_*(s)}{h_*(s)^2}\,\mathrm{d}h_1(s) \tag{4.1}$$

for h close to F. This matches the expression stated in Example 2.2. The first-order influence function is thus given by $\dot{\psi}(t; x) = \int_0^t \frac{1}{H(s)}\,\mathrm{d}M_{x,1}(s)$. We even have that the functional given by $\psi(h) = \int_0^{(\cdot)} \frac{1(h_*(s)>0)}{h_*(s)}\,\mathrm{d}h_1(s)$, mapping into $\mathcal{W}_p([0, t])$, is differentiable with the first-order derivative, as in (4.1), given by $\psi'_h(g) = \int_0^{(\cdot)} \frac{1}{h_*(s)}\,\mathrm{d}g_1(s) - \int_0^{(\cdot)} \frac{g_*(s)}{h_*(s)^2}\,\mathrm{d}h_1(s)$.

So, in each of our examples of application below, the procedure will be as follows. Interest will be in estimating parameters in a regression model $\mathrm{E}(V|Z) = \mu(\beta_0; Z)$ for some V based on the potentially unobserved (T, Δ). We will find an estimator of $\mathrm{E}(V)$ that can be considered a functional in the p-variation setting, differentiable of any order. Then an estimating equation based on the jack-knife pseudo-observations of that estimator and the particular model given by μ can be considered. Based on Theorem 3.4, solving the estimating equation will yield reasonable estimates of β_0 if the conditions are met.

As condition (b) of Proposition 3.1 is met for a functional that is differentiable of any order, we will have that Proposition 3.1 is met. We also need to have the regularity conditions of Theorem 3.4 fulfilled. Often $\mu(\beta; Z)$ is on the form $\mu(\beta^T Z)$ and μ is the inverse of a standard link function (identity, log, logit, cloglog, etc.), and A is given by $A(\beta; Z) = \mu'(\beta^T Z)Z$ (or some variation of it). The regularity conditions needed in Theorem 3.4 are then met if the covariates are reasonably well behaved: they should not be colinear and they should fulfil some moment condition. These assumptions are standard in a regression analysis. The remaining condition of Theorem 3.4, the condition in (3.36), is then the real hurdle and has to be checked for each functional and V we consider.

4.1. *Modeling a survival probability.* In time-to-event analysis, assessing the effect of covariates on the probability of survival may well be of interest. So, we let $V = 1(T > t)$ for some t of interest, and we want to estimate parameters in a regression model of the type (3.35) of $\mathrm{E}(V|Z) = \mathrm{P}(T > t|Z)$. In this section, we will see that this can be done using estimating equations of the type (3.34) based on jack-knife pseudo-observations from the Kaplan–Meier estimator of $\mathrm{E}(V) = \mathrm{P}(T > t)$ under standard regularity conditions on the model and under the assumption of completely independent censorings. This fact was proven by Jacobsen and Martinussen (2016). Here, the Kaplan–Meier functional serves as an example of use of our framework and as a building-block for the subsequent section.

To begin with, we will see what the Kaplan–Meier functional looks like in the p-variation setting. In similarity to Example 4.1, let ψ denote the all-cause Nelson–Aalen functional,

$$\psi(h) = \sum_{i=1}^{d} \int_0^{(\cdot)} \frac{1(h_*(s) > 0)}{h_*(s)} \, \mathrm{d}h_i(s), \tag{4.2}$$

and define a functional χ by

$$\chi(h) = \prod_0^{(\cdot)} \bigl(1 - \psi(\mathrm{d}s; h)\bigr). \tag{4.3}$$

We see that $\chi(F) = S$, the survival function, while $\chi(F_n) = \hat{S}$ is the Kaplan–Meier estimate, that is, χ is the Kaplan–Meier functional. As in Example 4.1, ψ is seen to be differentiable of any order in a neighborhood of F, and since the product integral is differentiable of any order, the same applies to χ in a neighborhood of F. Using the continuity of $\psi(F)$, the first-order derivative of χ at F in direction g is seen to be

$$\chi'_F(g) = -\chi(F)\psi'_F(g) = -S \cdot \sum_{i=1}^{d} \biggl(\int \frac{1}{H} \, \mathrm{d}g_i - \int \frac{g_*}{H^2} \, \mathrm{d}H_i \biggr) \tag{4.4}$$

by the chain rule.

Since we are interested in estimates of $S(t)$, our estimating functional is ϕ given by

$$\phi(h) = \chi(t; h), \tag{4.5}$$

where $\chi(t; h)$ is the function $\chi(h)$ evaluated at time t. In the estimating equation, the pseudo-observations will be based on the estimator given by ϕ. The functional ϕ is differentiable of any order since this is the case for χ. The first-order derivative at F is given by $\phi'_F(g) = \chi'_F(t; g)$. Let $M_x = \sum_{i=1}^{d} M_{x,i}$. Based on the expression for the Nelson–Aalen influence function, the influence function of ϕ can be expressed as

$$\dot{\phi}(x) = -S(t) \int_0^t \frac{1}{H(s)} \, \mathrm{d}M_x(s). \tag{4.6}$$

This is similar to the expression in Jacobsen and Martinussen (2016). As a consequence of their Proposition 1, we have that

$$\mathrm{E}\big(\chi'_F(s; \delta_X - F) | Z\big) = S_Z(s) - S(s), \qquad s \in [0, t] \tag{4.7}$$

and specifically that $\mathrm{E}(\dot{\phi}(X)|Z) = S_Z(t) - S(t)$, under the assumption of completely independent censorings, where $S_Z(s) = \mathrm{P}(T > s|Z)$ defines the conditional survival function given covariates. This means that when $V = 1(T > t)$, the condition in (3.36) is met using the functional ϕ. This condition was the remaining assumption for Theorem 3.4 to be applicable in a standard setting. In conclusion, we are able to model and estimate the effect of covariates on survival using pseudo-observations based on the Kaplan–Meier functional under the assumption of completely independent censorings.

The pseudo-observation method for a survival probability based on the Kaplan–Meier estimator has been examined in Klein et al. (2007) on real and simulated data. In their Section 5, they estimated the odds ratio of disease-free survival for leukemia patients given an autotransplantation relative to leukemia patients given an allotransplantation using an estimating equation based on the logit link function, that is, with $\mu(\beta, Z) = \frac{\exp(\beta^T Z)}{1+\exp(\beta^T Z)}$. Our results indicate that this is a reasonable estimating procedure if the condition of completely independent censorings is met.

4.2. *Modeling a cause-specific cumulative incidence.* When study participants are at risk of multiple causes of death or other types of mutually exclusive events, looking at the risk of having one specific type of event and how this risk depends on covariates is often of interest. In other words, for some timepoint $t > 0$, we consider a regression model of $\mathrm{P}(T \leq t, \Delta = 1|Z) = \mathrm{E}(V|Z)$ with $V = 1(T \leq t, \Delta = 1)$ (we assume, without loss of generality, that event 1 is of interest) as in (3.35). We aim at estimating the regression parameters. The risk $\mathrm{P}(T \leq t, \Delta = 1)$, the cumulative incidence of event 1, is usually estimated using the Aalen–Johansen estimator. In this section, we shall see that if we assume completely independent censorings and consider a reasonable model and estimating equation, all assumptions of Theorem 3.4 are met in this setting and we can expect the pseudo-observation approach to work.

It seems the Aalen–Johansen estimator in competing risks is most easily studied as an *inverse probability of censoring weighting* functional. Let ψ now denote the Nelson–Aalen functional for the cumulative censoring hazard, $\psi(h) = \int_0^{(\cdot)} \frac{1(h_*(s)>0)}{h_*(s)} \,\mathrm{d}h_0(s)$, and let χ denote the Kaplan–Meier functional for the censoring, $\chi(h) = \prod_0^{(\cdot)}(1-\psi(\mathrm{d}s;h))$. Now, consider the functional γ given by

$$\gamma(h) = \int_0^{(\cdot)} \frac{1(\chi(s-;h)>0)}{\chi(s-;h)} \,\mathrm{d}h_1(s). \tag{4.8}$$

This leads to $\gamma(F) = \int_0^{(\cdot)} \frac{1}{G(s)} \,\mathrm{d}H_1(s) = \int_0^{(\cdot)} S(s)\,\mathrm{d}\Lambda_1(s) = F_1$ under the assumption of completely independent censorings, and similarly $\gamma(F_n) = \int_0^{(\cdot)} \frac{1}{\hat{G}(s-)} \times \frac{\mathrm{d}N_1(s)}{n} = \int_0^{(\cdot)} \hat{S}(s-)\,\mathrm{d}\hat{\Lambda}_1(s) =: \hat{F}_1$, the Aalen–Johansen estimate of the cause 1-specific cumulative incidence function. In the last part, the identity $\frac{Y(s)}{n} = \hat{S}(s-)\hat{G}(s-)$, cf. p. 36 of Gill (1980), was used where $\hat{S}$ and $\hat{G}$ are the Kaplan–Meier estimates for survival and censoring, respectively.

We see that γ is a composition of functionals that are differentiable of any order in the p-variation setting described earlier and, therefore, itself differentiable of any order. Based on the chain rule, its first-order derivative at F in direction g is

$$\begin{aligned}\gamma'_F(g) &= \int_0^{(\cdot)} \frac{1}{\chi(s-;F)} \,\mathrm{d}g_1(s) - \int_0^{(\cdot)} \frac{\chi'_F(s-;g)}{\chi(s-;F)^2} \,\mathrm{d}H_1(s) \\ &= \int \frac{1}{G(s)} \,\mathrm{d}g_1(s) + \int \frac{\psi'_F(s-;g)}{G(s)} \,\mathrm{d}H_1(s),\end{aligned} \tag{4.9}$$

as $\chi'_F(g) = -G\psi'_F(g)$ as was seen in the previous section.

Let $F_{Z,1}(t) := \mathrm{P}(T \le t, \Delta = 1 | Z)$. Then the following applies.

PROPOSITION 4.2. *We have*

$$\mathrm{E}\big(\gamma'_F(s;\delta_X - F)|Z\big) = F_{Z,1}(s) - F_1(s), \qquad s \in [0,t], \tag{4.10}$$

under the assumption of completely independent censorings.

PROOF. A proof of this result was essentially given in Graw, Gerds and Schumacher (2009) in the proof of their Lemma 2. In the supplement Overgaard, Parner and Pedersen (2017), a detailed martingale-based proof is given. □

Since we are specifically interested in the cumulative incidence at time t, our estimating functional is ϕ given by

$$\phi(h) = \gamma(t;h), \tag{4.11}$$

where $\gamma(t; h)$ is the function $\gamma(h)$ evaluated at time t. When estimating parameters in a regression model of the cause 1-specific cumulative incidence using an estimating equation like (3.34), the pseudo-observations may be based on the estimator given by ϕ. The functional ϕ is clearly differentiable of any order in the p-variation setting, since this is the case for γ. So, the conditions of Proposition 3.1 are met for this functional in the p-variation setting. We have $\phi'_F(g) = \gamma'_F(t; g)$ and $\dot{\phi}(x) = \gamma'_F(t; \delta_x - F)$, or specifically

$$\begin{aligned}\dot{\phi}(x) = {}& \int_0^t \frac{1}{G(s)} \, \mathrm{d}N_{x,1}(s) - F_1(t) \\ & + \int_0^t \frac{1}{G(s)} \int_0^{s-} \frac{1}{H(u)} \, \mathrm{d}M_{x,0}(u) \, \mathrm{d}H_1(s).\end{aligned} \tag{4.12}$$

This is equivalent to the expression used in Graw, Gerds and Schumacher (2009).

Proposition 4.2 shows that the condition (3.36) is met for ϕ and $V = 1(T \leq t, \Delta = 1)$. By Theorem 3.4, we can conclude that we are able to model and estimate the effect of covariates on the cause 1-specific cumulative incidence at time t under the assumption of completely independent censorings. Of course, other causes or combinations of causes could be considered just as well. The application of the pseudo-observation method for the Aalen–Johansen estimator in a competing risks scenario has been examined in various papers, among others, Klein (2006) and Klein and Andersen (2005). Both papers are in fact focusing on using more than one timepoint at once and are not strictly covered by our results, which are focusing on the situation with one timepoint of interest, but this generalization can certainly be made. Klein and Andersen (2005) applied the method to data on events after bone marrow transplantation, focusing on estimating equations based on the logit and complementary log-log link functions. Klein (2006) advocates the use of additive models of the cause-specific cumulative incidences. With one timepoint of interest, this can be done using an estimating equation based on the identity link function [i.e., with $\mu(\beta^T Z) = \beta^T Z$],

$$\sum_{k=1}^{n} Z_k \big(\hat{\theta}_{n,k} - \beta^T Z_k\big) = 0, \tag{4.13}$$

where $\hat{\theta}_{n,k}$ are the pseudo-observations based on the Aalen–Johansen estimator for the type of event of interest. Of course, the implied model poses restrictions on the distribution of the covariates for $\beta_0^T Z$ to be a reasonable model of the probability $\mathrm{P}(T \leq t, \Delta = 1 | Z)$. The β_0 parameters are interpreted as risk differences in this model.

4.3. *Modeling a restricted mean survival time.* Estimating the effect of covariates on the survival time can be of interest, but is often hard to do since censorings are likely to hinder observation of the larger survival times, leaving the tail of the distribution ill-determined. As an alternative, the restricted survival time, $T \wedge t$ for some t, is sometimes considered instead. Letting $V = T \wedge t$, we consider a regression model of $\mathrm{E}(V|Z) = \mathrm{E}(T \wedge t|Z)$ as in (3.35). In the following, we will see how and when regression parameters can be estimated using the pseudo-observation approach of solving (3.34).

The restricted mean, $\mathrm{E}(T \wedge t)$, is obtained by integrating the survival function up to time t, that is, $\mathrm{E}(T \wedge t) = \int_0^t S(s)\,\mathrm{d}s$, and it is estimated by plugging in the Kaplan–Meier estimate for S, that is, estimated by $\int_0^t \hat{S}(s)\,\mathrm{d}s$. We calculate pseudo-observations, $\hat{\theta}_{n,k}$, based on this estimator for use in an estimating equation like (3.34). By letting χ denote the Kaplan–Meier functional from Section 4.1, the restricted mean functional, which will be our estimating functional in this section, is given by

$$\phi(h) = \int_0^t \chi(s; h)\,\mathrm{d}s. \tag{4.14}$$

Since χ is differentiable of any order in the p-variation setting (and $[0, t] \ni s \mapsto s$ is in $\mathcal{W}_p$) its composition with the integration operator is also differentiable of any order. In other words: the restricted mean functional, ϕ, is differentiable of any order in the p-variation setting, and so we see that the assumptions of Proposition 3.1 are met in this setting. Its first-order derivative at F is given by

$$\phi'_F(g) = \int_0^t \chi'_F(s; g)\,\mathrm{d}s, \tag{4.15}$$

and its influence function is given by $\dot{\phi}(x) = \int_0^t \chi'_F(s; \delta_x - F)\,\mathrm{d}s$. Remembering the result of (4.7) that holds when assuming completely independent censorings, and considering $V = T \wedge t$, this means

$$\mathrm{E}(\dot{\phi}(X)|Z) = \int_0^t S_Z(s)\,\mathrm{d}s - \int_0^t S(s)\,\mathrm{d}s = \mathrm{E}(V|Z) - \mathrm{E}(V), \tag{4.16}$$

such that the condition (3.36) is met when using this functional. In conclusion, Theorem 3.4 applies under the assumption of completely independent censorings if we consider reasonable A and μ, and thus we are able to model and estimate the effect of covariates on the restricted mean survival using the pseudo-observation method. The pseudo-observation method based on the restricted mean estimator was examined on real and simulated data in Andersen, Hansen and Klein (2004) using an estimating equation based on the identity link function. Their simulation study used completely independent censorings and showed only tiny amounts of bias on their β estimates which is in agreement with our results.

4.4. *Modeling an expected cause-specific lost lifetime.* Consider $V = (t - T \wedge t)1(\Delta = 1)$, then V is the amount of lifetime lost due to cause 1 before time t. The paper of Andersen (2013) proposed that a regression analysis on such a V may be useful and suggested using the pseudo-observation approach since V may well not be observable due to censorings. The expectation is $\mathrm{E}(V) = \int_0^t F_1(s)\,\mathrm{d}s$, which is estimated by $\int_0^t \hat{F}_1(s)\,\mathrm{d}s$, where $\hat{F}_1$ is the Aalen–Johansen estimate for cause 1. We now want to study an estimating equation based on jack-knife pseudo-observations of this estimator. Specifically, we want to show that Theorem 3.4 applies when the pseudo-observations are based on this estimator. The first step is to consider the estimator as a functional in a reasonable setting. Letting γ denote the cause 1-specific Aalen–Johansen functional from Section 4.2, this cause-specific lost lifetime estimator can be considered as a functional given by

$$\phi(h) = \int_0^t \gamma(s; h)\,\mathrm{d}s. \tag{4.17}$$

As was the case for the similar restricted mean functional, this cause-specific lost lifetime functional is differentiable of any order in the p-variation setting since γ is differentiable of any order in this setting. The assumptions of Proposition 3.1 are therefore met. The first-order derivative at F is given by

$$\phi'_F(g) = \int_0^t \gamma'_F(g)\,\mathrm{d}s \tag{4.18}$$

and the influence function is given by $\dot{\phi}(x) = \int_0^t \gamma'_F(s; \delta_x - F)\,\mathrm{d}s$. Still considering $V = (t - T \wedge t)1(\Delta = 1)$, we now see that under the assumption of completely independent censorings Proposition 4.2 applies and

$$\mathrm{E}(\dot{\phi}(X)|Z) = \int_0^t (F_{Z,1}(s) - F_1(s))\,\mathrm{d}s = \mathrm{E}(V|Z) - \mathrm{E}(V). \tag{4.19}$$

Thus, the condition in (3.36) is met for this functional. And so Theorem 3.4 applies, assuming the regularity conditions are met, thereby allowing us to model and estimate the effect of covariates on the cause specific lost lifetime. In Andersen (2013), this was done using an estimation equation based on the identity link function in the context of the bone marrow transplantation study.

5. A numerical example. The asymptotic properties of certain estimation procedures proven in the previous sections do not guarantee reasonable finite sample properties. Let us in this section take a look at one example of the behavior of a simple pseudo-observation based estimating procedure on a modestly sized sample, including parameter estimation and variance estimation using the proposed variance estimate in (3.46) and the usual variance estimate in (3.45). To do this, we simulate data from the following competing risks scheme. We consider a covariate Z such that $\mathrm{P}(Z = 0) = \mathrm{P}(Z = 1) = 0.5$. We want two competing risks and we are interested in a model of $F_{Z,1}(1)$ where $F_{Z,j}(s) = \mathrm{P}(T \le s, \Delta = j|Z)$. Event times and types follow the probabilities:

$$F_{Z,1}(s) = (0.05 + 0.70Z)s, \qquad F_{Z,2}(s) = (0.03 + 0.12Z)s, \tag{5.1}$$

for $s \in [0, 1]$. It follows that Z has a massive impact on $F_{Z,1}$ and $F_{Z,2}$; see the upper right of Figure 1 for an illustration. An exponentially distributed censoring

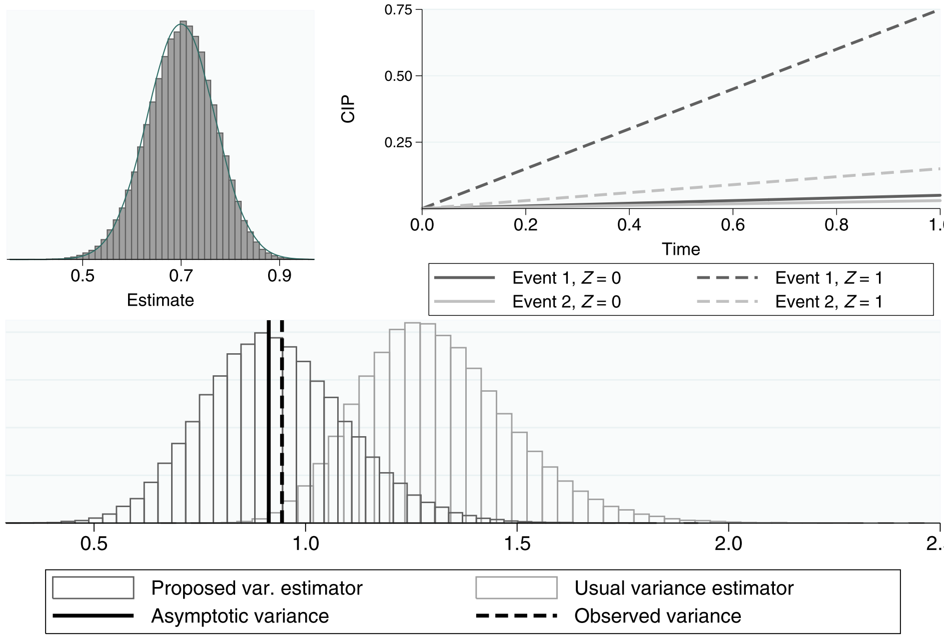

FIG. 1. *Upper left*: *Distribution of parameter estimates. Upper right*: *The true underlying cumulative incidence proportions for each event and depending on the covariate Z. Bottom*: *Distributions of variance estimates with the asymptotical variance and the observed variance among the parameter estimates plotted.*

time, C, is introduced such that we expect to be hindered in observing the variable $V = 1(T \leq 1, \Delta = 1)$ with probability 0.5, that is, such that $\mathrm{P}(C < T \wedge 1) = 0.5$ (in this scenario, it corresponds to a censoring intensity close to 1). We generate a sample of $n = 200$ observations, and we are interested in estimating parameters in the model $F_{Z,1}(1) = \beta_0^T(1, Z)^T = \beta_{00} + \beta_{01} Z$, that is, with true parameters $\beta_{00} = 0.05$ and $\beta_{01} = 0.70$. For that reason, we compute pseudo-observations, $(\hat{\theta}_{n,k})$, of the partially observed V_k's based on the Aalen–Johansen estimator, and we use them in an estimating equation:

$$\sum_{k=1}^{n} \begin{pmatrix} 1 \\ Z_k \end{pmatrix} \left(\hat{\theta}_{n,k} - \beta^T \begin{pmatrix} 1 \\ Z_k \end{pmatrix} \right) = 0 \tag{5.2}$$

to estimate β_0 by the solution $\hat{\beta}_n = (\hat{\beta}_{n,0}, \hat{\beta}_{n,1})$. This scheme ought to work well (at least for large n) according to the results from Section 4.2. We consider the estimate $\hat{\beta}_{n,1}$ of special interest and have focused on this in the following. The variance of $\hat{\beta}_{n,1}$ can be estimated based on the usual, but biased, variance estimate in (3.45) and based on the proposed variance estimate in (3.46). Note that the second-order derivative of the Aalen–Johansen functional can be found by using the differentiability results stated in the supplement. Using the proposed variance estimator is then only a question of insertion of the observed F_n. Similarly, the asymptotic variance can be numerically approximated using (3.39) since the true F is known.

We replicated the experiment 100,000 times and stored the parameter estimate and variance estimates for each sample of size 200. The results are summarized by Figure 1. We see that the distribution of estimates seems normal. The sample mean is 0.6998 and the sample standard deviation is 0.0687, leaving no indication of bias from the true 0.70. The proposed variance estimator seems centered around the asymptotic variance, but is on average slightly lower than the observed variance. The usual variance estimator seems to be biased upwards. This is all reflected in the coverages of corresponding 95 % confidence intervals. Considering coverage as a Bernoulli $(0/1)$ experiment, the true coverage probability can be estimated by 94.0% (93.9%–94.2%) for the proposed variance estimator and by 97.8% (97.7%–97.9%) for the usual variance estimator.

The properties of the proposed variance estimator is a topic of ongoing research.

6. Discussion. We have provided a general framework for the study of jack-knife pseudo-observations and we have extended the result of consistency and asymptotic normality of the β_0-estimates from Jacobsen and Martinussen (2016) to a much more general setting. The framework appears to suit inverse probability of censoring weighted estimators like the Kaplan–Meier and Aalen–Johansen estimators well.

We have focused only on real-valued pseudo-observations, but the approach can be generalized to handle vector-valued pseudo-observations, for example, when considering more timepoints than one simultaneously in the Kaplan–Meier case, if the estimating function is modified accordingly. The approach can likely be generalized to process-valued pseudo-observations if a functional central limit theorem is used appropriately.

The p-variation concept works well for counting processes as in our examples, and it is likely suited to also handle more general Markov multi-state models. In other settings, for example, when multi-dimensional covariates are used for the estimation of θ, other tools may be necessary in order to have the relevant sample averages contained in appropriate Banach spaces.

The p-variation approach draws on results from Dudley and Norvaiša (1999, 2011) that treat basic functionals in p-variation that can be composed into the functionals we consider in examples. We believe these works pave the way for the use of functional analysis in many statistical applications like this, and that they can find much more use in this field than they have found so far.

SUPPLEMENTARY MATERIAL

Supplement to "Asymptotic theory of generalized estimating equations based on jack-knife pseudo-observations" (DOI: 10.1214/16-AOS1516SUPP;

.pdf). The supplement contains an overview of the theory of differentiable functionals, some details on the p-variation setting, a note on the measurability of the influence functions, and a detailed proof of Proposition 4.2.

REFERENCES

ANDERSEN, P. K. (2013). Decomposition of number of life years lost according to causes of death. *Stat. Med.* **32** 5278–5285. MR3141373

ANDERSEN, P. K., HANSEN, M. G. and KLEIN, J. P. (2004). Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Anal.* **10** 335–350. MR2137286

ANDERSEN, P. K., KLEIN, J. P. and ROSTHØJ, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika* **90** 15–27. MR1966547

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York. MR1198884

DUDLEY, R. M. and NORVAIŠA, R. (1999). *Differentiability of Six Operators on Nonsmooth Functions and p-Variation*. *Lecture Notes in Math.* **1703**. Springer, Berlin. MR1705318

DUDLEY, R. M. and NORVAIŠA, R. (2011). *Concrete Functional Calculus*. Springer, New York. MR2732563

GILL, R. D. (1980). *Censoring and Stochastic Integrals*. *Mathematical Centre Tracts* **124**. Mathematisch Centrum, Amsterdam. MR596815

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method. I. *Scand. J. Stat.* **16** 97–128. MR1028971

GRAW, F., GERDS, T. A. and SCHUMACHER, M. (2009). On pseudo-values for regression analysis in competing risks models. *Lifetime Data Anal.* **15** 241–255. MR2510866

JACOBSEN, M. and MARTINUSSEN, T. (2016). A note on the large sample properties of estimators based on generalized linear models for correlated pseudo-observations. *Scand. J. Stat.* **43** 845–862.

JAMES, L. F. (1997). A study of a class of weighted bootstraps for censored data. *Ann. Statist.* **25** 1595–1621. MR1463566

KLEIN, J. P. (2006). Modelling competing risks in cancer studies. *Stat. Med.* **25** 1015–1034. MR2225189

KLEIN, J. P. and ANDERSEN, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* **61** 223–229. MR2135864

KLEIN, J. P., LOGAN, B., HARHOFF, M. and ANDERSEN, P. K. (2007). Analyzing survival curves at a fixed point in time. *Stat. Med.* **26** 4505–4519. MR2410057

OVERGAARD, M., PARNER, E. T and PEDERSEN, J. (2017). Supplement to "Asymptotic theory of generalized estimating equations based on jack-knife pseudo-observations." DOI:10.1214/16-AOS1516SUPP.

REEDS, J. A. III (1976). On the definition of Von Mises functionals. Ph.D. thesis, Harvard University, ProQuest LLC, Ann Arbor, MI. MR2940746

SØRENSEN, M. (1999). On asymptotics of estimating functions. *Braz. J. Probab. Stat.* **13** 111–136. MR1803041

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Stat.* **18** 309–348. MR0022330

M. OVERGAARD
E. T. PARNER
DEPARTMENT OF PUBLIC HEALTH
AARHUS UNIVERSITY
BARTHOLINS ALLÉ 2—BYGN. 1260
DK-8000 AARHUS C
DENMARK
E-MAIL: moov@ph.au.dk
parner@ph.au.dk

J. PEDERSEN
DEPARTMENT OF MATHEMATICS
AARHUS UNIVERSITY
NY MUNKEGADE 118
DK-8000 AARHUS C
DENMARK
E-MAIL: jan@math.au.dk