Published in final edited form as:
J Multivar Anal. 2012 April ; 106: 147–166. doi:10.1016/j.jmva.2011.11.007.

Root-*n* estimability of some missing data models

Ao Yuan[a,*], **Jinfeng Xu**[b], and **Gang Zheng**[c]

[a]Howard University, Washington, DC 20059, USA

[b]National University of Singapore, Singapore 117546, Singapore

[c]National Institutes of Health, Bethesda, MD 20892, USA

Abstract

It is known that in many missing data models, for example, survival data models, some parameters are root-*n* estimable while the others are not. When they are, their limiting distributions are often Gaussian and easy to use. When they are not, their limiting distributions, if exists, are often non-Gaussian and difficult to evaluate. Thus it is important to have some preliminary assessments of the root-*n* estimability in these models. In this article, we study this problem for four missing data models: two-point interval censoring, double censoring, interval truncation, and a case-control genetic association model. For the first three models, we identify some parameters which are not root-*n* estimable. For some root-*n* estimable parameters, we derive the corresponding information bounds when they exist. Also, as the Cox regression model is commonly used for such data, we give asymptotic efficient information for these regression parameters. For the case-control genetic association model, we compute the asymptotic efficient information and relative efficiency, in relation to that of the full data, when only the case-control status data are available, as is often the case in practice.

Keywords

Information operator; Missing data model; Root-*n* estimability; Score operator

1. Introduction

Missing data are often encountered in practice, such as in various types of censoring or truncation data, and case-control status data. In many such cases we need to estimate infinite-dimensional parameters or their functionals. It is known that some parameters in some missing data models cannot be estimated with the usual rate of $\sqrt{n}$, either because the estimators are not consistent or because they are consistent with a rate slower than $\sqrt{n}$, and in such case the limiting distribution is often non-normal and difficult to evaluate. For example, in the convolution model and some interval censoring models, the distribution functions are not root-*n* estimable, the corresponding nonparametric maximum-likelihood estimates are consistent with rate $n^{1/3}$, and the asymptotic limit is non-Gaussian (see, for example, [9]). In these cases, the limiting behavior of the estimators are non-standard, and difficult to use in practice. For the estimation of infinite-dimensional parameters, although there is no general criterion, no simple and clear-cut rule for rate- $\sqrt{n}$ estimability in the existing statistics theory, it is still of meaning to have some preliminary results, using a combination of existing tools, regarding the estimability of some commonly used missing

*Corresponding author. yuanao@hotmail.com (A. Yuan).

data models, so that the practitioners can have some prior knowledge in the investigation with such models. Most missing data models have been extensively studied (for example, [16,29,7,8,28,10,25,14,24,33,35,21,5,15,32]; and many more). Some parameters in many such models are root-n estimable and have been well studied; many of them are efficient in that their asymptotic variances achieve the information lower bounds. But in many other cases even the corresponding information bounds are not straightforward to compute. On the other hand, although many functionals of missing data models are not root-n estimable, the corresponding regression parameter(s) are often root-n efficiently estimable. For example, Huang [13] established such a result for the interval censoring model, although the nonparametric maximum-likelihood estimate of the underlying hazard function is consistent with rate $n^{1/3}$. Indeed, many smooth functionals of infinite-dimensional parameters are root-n efficient estimable, even though such parameters themselves may not be so. An excellent survey of works in this field is given in [2], hereafter BKRW.

Here, we study the mentioned problem for four missing data models: interval censoring (such data can be found in practice, e.g. in AIDS research [17]); double censoring, which has been studied by Turnball [29], Chang and Yang [4], Gu and Zhang [10], and Sun [26] amongst others; interval truncation, which is an extension of the one-sided truncation model, and can be found applications in economics and social studies [12]; and a case-control genetic association model [20,23,34]. The first three models are survival models, while the last one is not. Although some results for one-point interval censoring, double censoring, and one-point truncation exist in the literature, to our best knowledge, a formal treatment of the above four models regarding the identification of root-n estimability has not been undertaken. Here, for the first three models, we identify some parameters which are not root-n estimable, and, for root-n estimable parameters, we derive the corresponding information bounds when they exist. Also, as Cox's regression model is commonly used for such data, we obtain the asymptotic efficient information for the regression parameters. For the case-control genetic association model, we compute the asymptotic information and relative efficiency when only the case-control status data are available, as is often the case in practice.

Our results show that, for the interval censoring model and the double censoring model, the root-n estimability depends on the form of the functionals. For the interval truncation model, under rather general conditions any smooth functional is root-n estimable. For the case-control genetic association model, which can be regarded as a missing data model with only case-control status, we considered two scenarios with the underlying distribution known or unknown, both with presence of covariates. We calculated the asymptotic efficient information for the association parameter and the asymptotic relative efficiency of using only the case-control status data as compared to that for the complete data.

2. Root-*n* estimability for some missing data models

Let the original data be $X^0 \sim Q_\theta \in \mathcal{Q} = \{Q_\theta: \theta \in \Theta\}$. In many cases, we cannot directly observe X^0; instead, we observe $X = T(X^0)$ for some measurable map T. The induced model for X is $\mathcal{P} = \{P = Q_\theta T^{-1}: \theta \in \Theta\}$. A parameter $\nu = \nu(P)$ is *root-n estimable* (see BKRW) if there is an estimator ν_n based on n independent observations $X_1, \ldots, X_n$ such that

$$\lim_{M\to\infty} \limsup_{n\to\infty} P(\sqrt{n}|\nu_n - \nu| \geq M)=0.$$

Apparently, when $\sqrt{n}(\nu_n - \nu)$ converges weakly, ν is rate- $\sqrt{n}$ estimable. Denote $L_2^0(Q)=\{g: \int g(x)Q(dx)=0, \int g^2(x)Q(dx)<\infty\}$; similarly for $L_2^0(P)$. The *score operator* $\mathbf{i}: L_2^0(Q) \to L_2^0(P)$ is defined as (see [9])

$$\mathbf{i}a=E(a(X^0)|T(X^0)=x), \quad \forall a \in L_2^0(Q).$$

It is a bounded linear operator and has an adjoint $\mathbf{i}^*: L_2^0(P) \to L_2^0(Q)$ determined by

$$\langle b, \mathbf{i}a\rangle_{L_2(P)}=\langle \mathbf{i}^* b, a\rangle_{L_2(Q)}, \quad \forall a \in L_2^0(Q), \ \ b \in L_2^0(P),$$

where $\langle b, g\rangle_{L2(P)} = \int b(x)g(x)P(dx)$ and $\langle b, g\rangle_{L2(Q)} = \int b(x)g(x)Q(dx)$. Also, $\mathbf{i}^*b=E(b(X)|X^0), \forall b \in L_2^0(P)$. The *information operator* is defined as $\mathbf{i}^*\mathbf{i}: L_2^0(Q) \to L_2^0(Q)$. The score operator (and information operator) defined above for missing data models are the same as those for parameters as given in BKRW, though the definitions there look different. Latter on, we will see that the information operator given above can be used to assess the estimability of parameters in missing data models.

In the parametric case, bounded differentiability of log-likelihood implies the existence of Fisher information or an information bound. In the nonparametric (infinite-dimensional parameters) case, commonly used derivatives include Gâteaux, Hadamard (compactly), pathwise, Hellinger, and Fréchet differentials. Often, Gâteaux differentiability is too weak (even a discontinuous functional can be Gâteaux differentiable), the Fréchet differential is too strong (many commonly used functionals do not have Fréchet differentiability), and the Hadamard differential is stronger than Gâteaux and weaker than Fréchet and is considered more appropriate to use in most statistical problems. The Hellinger differential is a special form of the Fréchet differential, and the pathwise differential is a special form of the Hadamard differential and is often used for semiparametric and nonparametric models. Thus, for nonparametric models, the existence of an information bound, hence plausible root-n estimability, relies on the pathwise differentiability of the models. Since in nonparametric models the ‘parameters’ to be estimated are often implicit functionals of the unknown distribution, pathwise differentiability (with respect to the distribution) are often obtained using van der Vaart’s [31] composite differentiability result. For missing data models, a commonly used result is that if the information operator $\mathbf{i}^*\mathbf{i}$ is boundedly invertible, then any pathwise differentiable functional (parameter) has an information bound. Although the existence of an information bound does not guarantee the existence of a rate-$\sqrt{n}$ consistent estimator [22], in many such cases there do exist rate- $\sqrt{n}$ asymptotic normal estimators which achieve the information bound, i.e., the existence of an efficient estimator. In fact, it follows from the work of LeCam [18] that regularity of the model together with existence of a $\sqrt{n}$-consistent preliminary estimator guarantees the existence of an efficient estimator; see also Theorem 2.5.2 in BKRW. Also, as in Theorem 5.2.3 in BKRW, in the presence of joint $\sqrt{n}$-consistency of the estimator and score, pathwise differentiability is almost equivalent to regularity. When $\mathbf{i}^*\mathbf{i}$ is not boundedly invertible, then some functional(s) of the distribution do not have an information bound; thus, by the convolution theory [11], these functional(s) cannot be estimated at rate $\sqrt{n}$. Although, for infinite-dimensional parameters, there is no clear cut rule for $\sqrt{n}$-estimability, in summary there are some preliminary guides.

- For a given model Q, if the information operator $\mathbf{i}^*\mathbf{i}$ is boundedly invertible, then, for any (pathwise) differentiable functional v of Q, an information bound exists; thus there is a Gaussian distribution as the optimal weak limit (with zero mean and the corresponding information bound as the variance) for any regular estimator of v, in the sense of a convolution decomposition, suggesting that v may be root-n estimable with a Gaussian weak limit. This plus some other existing results allows us to investigate whether it is root-n estimable.
- If $\mathbf{i}^*\mathbf{i}$ is not boundedly invertible, then many functionals of Q have no information bounds; these functionals are not estimable at rate $n^{1/2}$, and their weak limits, if they exist, are often non-Gaussian. In this case, however, by combining additional results, we may identify some possible specific functionals which are root-n estimable.

Below, we investigate the aforementioned four missing data models for the rate- $\sqrt{n}$ estimability for some of their parameters, and compute their information bounds or asymptotic information when they exist.

Interval censoring with two points (case II)

In this model, $X^0 = (Z, U, V) \in R^+ \times (R^+)^2$, where $Z \sim F$ and $(U, V) \sim H$ are independent, and $U < V$ a.s. H has a density h with respect to the Lebesque measure λ on $(R^+)^2$. We observe $X = T(X^0) = (U, V, 1_{[Z\le U]}, 1_{[Z\le V]}) := (U, V, \delta, \gamma)$, with $0 < P(\delta = 1) < 1$ and $0 < P(\gamma = 1) < 1$. This model is in Example 1.6 of Groeneboom and Weller [9]. Here, $Q = F \times H$, and the distribution $P_{F,H}$ of (U, V, δ, γ) has density/mass function

$$p_{F,H}(x)=p_{F,H}(u,v,\delta,\gamma)=F^{\delta}(u)(F(v)-F(u))^{\gamma(1-\delta)}(1-F(v))^{1-(\delta\vee\gamma)}h(u,v)$$

with respect to $\lambda \times$ (counting measure) on $(R^+)^2 \times \{0, 1\}^2$. We want to know whether some smooth functional of (F, H) is rate-$n^{1/2}$ estimable based on observations $X_1, \ldots, X_n$ i.i.d. with X. In this model, our main interest is to estimate parameters which are functionals of F. For this, let $\mathbf{i}_1 : L_2^0(F) \to L_2^0(Q)$ be the score operator for F. We have the following.

Theorem 1—For the model of interval censoring with two points, $\mathbf{i}_1^*\mathbf{i}_1$ is not boundedly invertible.

Thus, by our discussion above, many parameters in the interval censoring model with two points are not root-n estimable. We want to know the root-n estimability of some specific parameters (or functionals of F). Since (U, V) are directly observed, any reasonably smooth functional of H is rate- $\sqrt{n}$ estimable. Now, we consider functionals of F, for example the parameter $F(z)$, the distribution function of Z at z, or the parameter

$$v=\int g(z)F(dz), \qquad \text{with } g(z)=E(a(U,V,\delta,\gamma)|Z=z) \text{ for some } a(u,v,\delta,\gamma)\in L_2(P_{F,H}).$$

In particular, we consider parameters of the following form, assuming its existence:

$$\mu=E(g(Z)), \qquad g(z)=\int_0^z r(v)dv, \qquad r(\cdot) \text{ given.}$$

Thus, if $r(v) = kv^{k-1}$, μ is the k-th moment of F. These parameters are functionals of F. Note that $g(z)$ can also be written in the form $g(z) = E[b(U, V, \delta, \gamma)|Z = z]$ for some b given in Theorem 2(iii). Since Z is not directly observable, the estimability of these parameters is not obvious.

On the other hand, its is known that, for many missing data models, the finite-dimensional regression parameter(s) in the Cox [6] model are root-n efficiently estimable, although the underlying hazard function is not root-n estimable. Huang [13] constructed root-n efficient estimate, for the case I interval censoring model, for the regression parameters, although the underlying hazard function estimator is consistent at a slower rate of $n^{1/3}$. Some other such examples can be found in BKRW. Here, assuming root-n estimability, we give the asymptotic information for such parameters in this case, and the corresponding information bound is the inverse of the asymptotic information. Let W be the covariate(s), and let θ be the regression parameter(s) in the Cox model:

$$\Lambda(z|w)=\exp(\theta' w)\Lambda(z) \quad \text{or} \quad \overline{F}(z|w)=\overline{F}(z)^{\exp(\theta' w)}$$

where $\Lambda(z)=\int_{-\infty}^{z} dF/(1-F(z-))$ is the hazard function and $\bar{F}(z) = 1 - F(z)$.

Let $h(u, v)$ be the density of $H(u, v)$, and let $h_1(u)$ be its U-margin. We have the following.

Theorem 2—For the model of interval censoring with two points, the following hold.

- **i.** *$F(z)$ is not pathwise differentiable with respect to $P_{F,H}$, and hence not root-n estimable for each z.*
- **ii.** *v is pathwise differentiable with respect to $P_{F,H}$.*
- **iii.** *Assume that $h_1(\cdot)r(\cdot)/h(\cdot, \cdot) \in L_2(P_{F,H})$. Then μ is pathwise differentiable with respect to $P_{F,H}$. Then the efficient influence function for estimating μ is*

$$\dot{\mu}^*(1)=\tilde{I}(1)=b(u,v,\delta,\gamma)-b_1(u,v),$$

and the information bound $I(\mu)$ for estimating μ, for H known or unknown, is given by

$$I(\mu)=\iint \Big(b(u,v,0,0)(b(u,v,0,0)-2b_1(u,v))(1-F(v))+b_1^2(u,v)\Big)H(du,dv),$$

where $b(u, v, \delta, \gamma) = h_1(u)r(v)(1 - \delta)(1 - \gamma)/h(u, v)$ and $b_1(u, v) = E(b(U, V, \delta, \gamma)|(U, V) = (u, v))$. When the data $\{Z_i: i = 1, \ldots, n\}$ are observable, $I(\mu)$ reduces to $I(\mu) = E(g^2(Z)) - \mu^2$. If, further, H is known, then μ is rate-$\sqrt{n}$ estimable.

- **iv.** *The asymptotic information $i(\theta|f)$ for estimating θ, in the presence of nuisance f, in the above Cox model is*

$$i(\theta|f)=E_p\left\|l_{\theta|f}(U,V,\delta,\gamma,W)\right\|^2,$$

where P is the joint distribution of (U, V, δ, γ) and W,

$$l_{\theta|f}(u,v,\delta,\gamma,w)=\exp(\theta' w)\left(\delta\frac{\overline{F}(u|w)\Lambda(u)(w-R(u))}{1-\overline{F}(u|w)}-(1-(\delta\vee\gamma)\Lambda(v)(w-R(v)))\right.$$
$$\left.+\gamma(1-\delta)\frac{\overline{F}(u|w)\Lambda(u)(w-R(u))-\overline{F}(v|w)\Lambda(v)(w-R(v))}{\overline{F}(v|w)-\overline{F}(u|w)}\right),$$

$R(s) = E(O(W, S)W|S = s)/E(O(W, S)|S = s)$, *and* $O(W, S) = \exp(2\theta' W)\bar{F}(S|W)W/(1-\bar{F}(S|W))$.

In (iii) above, if we set $r(\cdot) \equiv 1$, $h(u, v) = h_1(u)h_2(v)$ and $\delta \equiv 0$, then $\tilde{I}(1)$ and $I(\mu)$ reduce to their counterparts for the indicator censoring model as in Example 5.4.1 (BKRW).

In the above interval censoring model (case II), if we set $V \equiv \infty$ (or $\gamma \equiv 1$), and define $0^0 = 1$, then we get the interval censoring model (case I), and the efficient score $l_{\theta|f}$, and the asymptotic information $i(\theta|f)$ of Theorem 2(iv) reduces to that in [13], in which the corresponding notations are $i_\theta^*(x)$ and $I(\theta)$.

Double censoring

In this model, $X^0 = (Z, U, V) \in R^3$, $U < V$ a.s. We observe the survival time Z, with distribution function $F(\cdot)$, when it falls between U and V, i.e. we observe $X = T(X^0) = ((Z \vee U) \wedge V, 1_{[Z\le U]}, 1_{[U<Z\le V]}) = (Y, \delta, \gamma)$. Here, $Z \perp (U, V)$, and (U, V) has joint distribution function $H(\cdot, \cdot)$. Here, (δ, γ) can only take values (0, 0), (0, 1), and (1, 0). The density $p_{F,H}$ for X is

$$p_{F,H}(x)=p_{F,H}(y,\delta,\gamma)=(M(y)f(y))^{\gamma}(F(y)h_U(y))^{\delta}((1-F(y))h_V(y))^{1-\gamma-\delta},$$

where $M(y) = P(U < y \le V) = H_U(y) - H(y, y)$, H_U is the marginal distribution of U, and h_U and h_V are the marginal densities of U and V. This model was studied by Turnball [29] and Tsai and Crowley [28], among others. When $(\delta, \gamma) = (0, 0), (0,1)$, and $(1, 0)$, we observe V, Z, and U, respectively. Intuitively, differentiable functionals of the marginal distributions of V, Z, and U are estimable at rate $\sqrt{n}$ as long as (δ, γ) can take all three of the above values with positive probabilities, but not for functionals of the joint distribution H, since we cannot observe (V, U) jointly. In fact, Chang and Yang [4] showed the strong consistency of the marginal empirical survival functions of Z, U, and V based on the observed data of $(Z \vee U) \wedge V$. Chang [3] showed the rate- $\sqrt{n}$ weak consistency of the empirical distribution and survival function for Z. But root-n consistent estimates of functionals of the joint H or (F, H) have not been seen.

Let $\mathbf{i}$ be the score operator for (F, H), and let $\mathbf{i}_1$ and $\mathbf{i}_2$ be those for F and G. We give a formal justification of the above conjectures as follows.

Theorem 3—For the double censoring model, the following hold.

- **i.** *When $Q = (F, H)$ is unknown, the information operator $\mathbf{i}^*\mathbf{i}$ is not boundedly invertible.*
- **ii.** *When H is known and F is unknown, assume that* $\inf_y M(y) > 0$. *Then the information operator $\mathbf{i}_1^*\mathbf{i}_1$ is boundedly invertible.*
- **iii.** *When F is known and H is unknown, then the information operator $\mathbf{i}_2^*\mathbf{i}_2$ is not boundedly invertible.*

Result (ii) in the above was proved in Example 6.6.5 in BKRW in a different way. Next, we show that the joint distribution function $H(\cdot, \cdot)$ of (U, V) is not rate- $\sqrt{n}$ estimable, but functionals of F alone are, such as the survival function $S(t) = 1-F(t)$, and the hazard function $\Lambda(t)=\int_{-\infty}^{t}\frac{dF}{1-F_-}$, with $F_-(t) = F(t-)$.

We also consider the Cox regression model in the double censoring model. We show the regression parameters are root-*n* estimable in this case, and derive the asymptotic efficient information.

Theorem 4—For the double censoring model, the following hold.

i. H is not root-n estimable.

ii. *Assume that* $M^{-2}(\cdot) \in L_2(P_{F,H})$. *Then, for each fixed* $t \in R$, $S(t)$ *is pathwise differentiable with respect to* $P_{F,H}$ *and is rate-* $\sqrt{n}$ *estimable, with efficient influence function, for H known or unknown,*

$$\dot{S}^*(t)(Y,\delta,\gamma)=\frac{1_{(t,\infty)}(Y)(1-\delta)\gamma}{M(Y)}-(1-\gamma-2\delta+2\delta\gamma)\int_{-\infty}^{Y}\frac{1_{(t,\infty)}(z)}{M(z)}F(dz),$$

and the information bound for $S(t)$ *is*

$$I_S(t)=\int\frac{1_{(t,\infty)}(z)}{M(z)}F(dz)+\int R_S^2(t\vee u,v)H(du,dv),$$

where $R_S(a,b)=\int_a^b M^{-1}(z)F(dz)$. *When* $\{Z_i: i = 1, \ldots, n\}$ *is fully observable,* I_S *reduces to* $E(1_{(t,\infty)}(Z) - 1 + F(t))^2$.

iii. *Under assumptions in* (ii), *for each fixed* $t \in R$, $\Lambda(t)$ *is pathwise differentiable with respect to* $P_{F,H}$ *and is rate-* $\sqrt{n}$ *estimable, with efficient influence function, for H known or unknown, as*

$$\dot{\Lambda}^*(t)(Y,\delta,\gamma)=\frac{1_{(t,\infty)}(Y)(1-\delta)\gamma}{M(Y)(1-F(Y-))}-(1-\gamma-2\delta+2\delta\gamma)\int_{-\infty}^{Y}\frac{1_{(t,\infty)}(z)}{M(z)(1-F(z-))}F(dz),$$

and the information bound for $\Lambda(t)$ *is*

$$I_\Lambda(t)=\int\frac{1_{(t,\infty)}(z)}{M(z)(1-F(z-))}F(dz)+\int R_\Lambda^2(t\vee u,v)H(du,dv),$$

where $R_\Lambda(a,b)=\int_a^b M^{-1}(z)(1-F(z-))^{-1}F(dz)$.

iv. *The regression parameters* θ *are root-n estimable with asymptotic information, in the presence of nuisance f,*

$$i(\theta|f)=E_P\left\|l_{\theta|f}\right\|^2,$$

where $l_{\theta|f}$ *is given in the proof.*

The asymptotic variance of the survival function estimator by Chang [3] is evaluated via a system of integral equations, not in closed form. Whether it achieves the information lower bound in Theorem 4(ii) is not easy to check.

Interval truncation

Suppose that $X^0 = (Z, U, V)$ is a vector of independent random variables with marginal distributions F, G_1, and G_2 on $R^+ = [0, \infty)$ and corresponding densities f, g_1, and g_2. Here, $Q = F \times G_1 \times G_2$. Denote $G = G_1 \times G_2$ and $g(u, v) = g_1(u)g_2(v)$. We observe (U, Z, V) only if $U < Z < V$, i.e. $X = T(X^0) = (U, Z, V)1_{(U<Z<V)}$. This is an extension of the commonly used truncation model in which only Z and V are involved. For identifiability, we assume that $G_1^{-1}(0+) \le F^{-1}(0+) \le G_2^{-1}(0+)$ and $G_1^{-1}(1) \le F^{-1}(1) \le G_2^{-1}(1)$ (cf. Example 6.4.1 in BKRW). The density function $P_{F,G}$ for the observed data has density

$$p_{F,G}(u, z, v) = \alpha^{-1} f(z) g(u, v) 1_{(u<z<v)},$$

where $\alpha = P(U < Z < V) = \int\int [F(v) - F(u)]G(du, dv)$. Naturally, we assume that $0 < \alpha < 1$ so that this model is of meaning. Obviously, any smooth functional of $P_{F,G}$ is root-n estimable, since (U, Z, V) are directly observable; we give a formal justification of this. We compute the information bound for some parameters in this model. Let $M(s) = \alpha^{-1}G_1(s)\bar{G}_2(s)(1 - F_-(s)) = \alpha^{-1}P(U < s < V \wedge Z)$, and let $\Lambda(t) = \int_0^t dF/(1 - F_-)$ be the hazard function for F. We also give the asymptotic information for estimating θ in the Cox regression model in this case.

Theorem 5

- **i.** *For the interval truncation model,* $\mathbf{i}^*\mathbf{i}$ *is boundedly invertible, and hence any smooth functionals of* $P_{F,G}$ *on* $\{(z, u, v): u < z < v\}$ *are rate-* $\sqrt{n}$ *estimable, as long as* $\alpha > 0$.
- **ii.** *Assume that* $E_F(M^{-2}(Z)) < \infty$. *Then the efficient influence function for estimating* $\Lambda(t)$ *is*

$$\tilde{l}(u, z, v) = \frac{1_{(0 \le z \le t)}}{M(z)} - \alpha \int_0^t \frac{1_{(u<s<z\wedge v)}}{G_1(s)\bar{G}_2(s)(1 - F_-(s))^2} F(ds),$$

and the information bound is $I = E_{P_{F,G}} \tilde{l}^2(U, Z, V)$.

- **iii.** *The regression parameter* θ *in the Cox model with interval truncation is root-n estimable, with asymptotic efficient information* $i(\theta|f) = E_P\|l_{\theta|f}\|^2$, *where* $l_{\theta,f} = l_\theta(\theta, f) - l_f(\theta, f)(a^*)$, *which are given in the proof.*

In (ii) above, if we take $U \equiv 0$, then $\tilde{l}(u, z, v)$ reduces to its counterpart for the random truncation model, as given in Example 6.4.1 (BKRW).

Case-control genetic association study

The original data are $X^0 = (Y|Z, G)$, $Y \sim F$ with density f. Z is the covariate and G the genotype of Y: typically G takes one of three possible codes $g_0 = AA$, $g_1 = Aa$, and $g_2 = aa$, where A and a are the two alleles at the gene locus. The genotype is valued by some function $c(\cdot)$ which depends on the mode of inheritance. For the recessive, additive, and dominant modes, we set $(c(g_0), c(g_1), c(g_2)) = (0, 0, 1)$, $(0,1/2,1)$, and $(0, 1, 1)$, respectively (e.g. as in

[34]). In practice, we often only observe $X = T(X^0) = (1_{(Y \le Y_0)}|Z, G) := (\delta|Z, G)$ for some pre-specified threshold value Y_0, which is known to the experimenter but often unknown to the user of the data. Here, $\delta = 0$ corresponding to control, and $\delta = 1$ corresponding to case. This model is related to the censoring models. A retrospective model is appropriate for such data, but a simpler prospective model is equivalent [20], and is often used. A prospective logistic model is commonly used for the mass function of $\delta|(Z, G)$:

$$P(\delta=0|Z,G;\alpha,\beta)=\frac{1}{1+\exp\{\alpha' Z+\beta U\}}, \quad \text{with } U=\sum_{j=0}^{2} 1_{(G=g_j)} c(g_j),$$

where β is the log odds ratio, and $\beta = 0$ implies no association of the underlying gene locus to case. For such data, it is known that the regression parameters (α, β) are root-n estimable based on either the original data $(Y|Z, G)$ or only the case-control status data $(\delta|Z, G)$. But it is not obvious if many other functionals of F are root-n estimable based on $(\delta|Z, G)$. Also, we want to know how much information is lost when we only observe $(\delta|Z, G)$ compared to if we can observe the original data $(Y|Z, G)$, so we assume a semiparametric model with an unknown location family F. Thus, if the original data are observed, the regression model $Y = \alpha'Z + \beta U + \varepsilon$, $\varepsilon \sim f$ will be used, i.e. the data is from a location family with conditional density

$$f(y|Z,G;\alpha,\beta)=f(y-\mu), \quad \text{with } \mu=\alpha' Z+\beta U.$$

When only $(\delta|Z, G)$ are observed, the mass function for $(\delta|Z, G)$ is

$$p_f(\delta):=p(\delta|Z,G;\alpha,\beta)=(F(Y_0))^{1-\delta}(1-F(Y_0))^{\delta},$$

where $F(Y_0):=F(Y_0|Z,G;\alpha,\beta)=P(\delta=0|Z,G;\alpha,\beta)=\int_{-\infty}^{Y_0} f(y-\mu)dy$.

The logistic specification of $P(\delta = 0|Z, G; \alpha, \beta)$ corresponds to the logistic distribution

$$F_0(y)=\frac{1}{1+\exp\{-(y-\tilde{\mu})\}}, \quad \text{or} \quad f_0(y|Z,G;\alpha,\beta)=\frac{\exp\{-(y-\tilde{\mu})\}}{(1+\exp\{-(y-\tilde{\mu})\})^2},$$

where $\tilde{\mu} = Y_0 + \alpha'Z + \beta U$. In general, we assume f (or F) to be unknown as an infinite-dimensional nuisance parameter, along with the nuisance parameter α, and β is the parameter of interest.

We use $i(\beta|\alpha, f)$ to denote the asymptotic information in the model $f(\cdot|Z, G; \alpha, \beta)$ for estimating β in the presence of nuisance (α, f). The corresponding information bound is $I = i^{-1}(\beta|\alpha, f)$, and so we only compute the former. Similarly, let $i(\beta|\alpha, p_f)$ be the asymptotic information in the model $p_f(\cdot|Z, G; \alpha, \beta)$. The information loss due to using $p_f(\cdot|Z, G; \alpha, \beta)$ instead of $f(\cdot|Z, G; \alpha, \beta)$ is

$$i(\beta|\alpha,f)-i(\beta|\alpha,p_f).$$

Likewise, let $i(\beta|\alpha, f_0)$ and $i(\beta|\alpha, p_{f_0})$ be the asymptotic information under the corresponding models f and p_f with $f = f_0$. Let $\rho_\beta(\cdot - \mu) = \partial f^{1/2}(\cdot - \mu)/\partial\beta = (1/2)f^{-1/2}\partial f/\partial\beta = (1/2)f^{1/2}\,\partial \log f/\partial\beta$, similarly for $\rho_\alpha(\cdot)$; $\|f\|_\lambda^2=\int f(x)\lambda(dx)$, $\|f\|^2 = \int f(x)dx$ for λ to be the Lebesque measure, $\rho_{\beta|\alpha}=\rho_\beta - \langle\rho_\beta,\rho_\alpha'\rangle_\lambda \|\rho_\alpha\|_\lambda^{-2}\rho_\alpha, \rho_{\beta|\alpha,\mu=0}(\cdot)$ denote $\rho_{\beta|\alpha}(\cdot - \mu)$ with $\mu = 0$. Let $\tilde{\rho}_\beta$, $\tilde{\rho}_\alpha$ and $\tilde{A}$ be the corresponding differentials under model $p_f, \tilde{\rho}_{\beta|\alpha}=\tilde{\rho}_\beta - \langle\tilde{\rho}_\beta,\tilde{\rho}_\alpha'\rangle_N \|\tilde{\rho}_\alpha\|_N^{-2}\tilde{\rho}_\alpha$, with N the counting measure on $\{0, 1\}$. The specific values of ρ_β, ρ_α, $\tilde{\rho}_\beta$ and $\tilde{\rho}_\alpha$ for f known or unknown are given in the proof of Theorem 6.

Theorem 6

i. *For the model $p_f(\delta)$, $\mathbf{i}^*\mathbf{i}$ is not boundedly invertible. Hence many functionals of F are not root-n estimable based on the data* $(\delta|Z, G)$.

ii. *Assume that the support of f does not depend on (α, β). Then the asymptotic information for estimating β under f and p_f, with f unknown, is*

$$i(\beta|\alpha,f)=4\|\rho_{\beta|\alpha}-\rho_{\beta|\alpha,\mu=0}\|^2,\quad i(\beta|\alpha,p_f)=4\|\tilde{\rho}_{\beta|\alpha}-\tilde{A}(h^*)\|_N^2,$$

where

$$\tilde{A}(h^*)(\delta)=\frac{p^{1/2}(1)\tilde{\rho}_{\beta|\alpha}(1)-p^{1/2}(0)\tilde{\rho}_{\beta|\alpha}(0)}{p_f^{1/2}(0)p_f^{1/2}(1)}(1-\delta-F(T_0))p_f^{1/2}(\delta).$$

iii. *Under the assumption in* (ii), *when $f = f_0$ is known, the asymptotic information is*

$$i(\beta|\alpha,f_0)=U^2(I-Z'(ZZ')^{-1}Z)^2/3,\quad i(\beta|\alpha,p_{f_0})=U^2(I-Z'(ZZ')^{-1}Z)^2q(\mu),$$

where $q(\mu) = \exp(\mu)/(1+\exp(\mu))^2$. Thus the information loss using p_{f_0} *instead of* f_0 *is* $U^2(I - Z'(ZZ')^{-1}Z)^2(1/3-q(\mu))$, *and the asymptotic relative efficiency is* $r(\mu) = 3q(\mu)$, *with* $0 < r(\mu) \le 3/4$, $r(\mu) = 3/4$ *iff* $\mu = 0$.

Acknowledgments

This work is supported in part by the National Center for Research Resources at NIH grant 2G12RR003048 (Yuan), and National University of Singapore, Grant R-155-000-112-112 (Xu).

References

1. Begun JM, Hall WJ, Huang W, Weller JA. Information and asymptotic efficiency in parametric–nonparametric models. Annals of Statistics. 1983; 11:432–452.
2. Bickel, PJ.; Klaassen, CA.; Ritov, Y.; Wellner, JA. Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press; Baltimore, Maryland: 1993.
3. Chang MN. Weak convergence of a self-consistent estimator of the survival function with doubly censored data. Annals of Statistics. 1990; 18:391–404.
4. Chang MN, Yang GL. Strong consistency of a nonparametric estimator of the survival function with doubly censored data. Annals of Statistics. 1987; 15:1536–1547.
5. Chen K, Guo S, Sun L, Wang JL. Global partial likelihood for nonparametric proportional hazard models. Journal of the American Statistical Association. 2010; 105:750–760.

6. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society: Series B. 1972; 34:187–220.
7. Gill, RD. Math Centre Tract. Vol. 124. Mathematisch Centrum; Amsterdam: 1980. Censoring and Stochastic Integrals.
8. Gill RD. Large sample behavior of the product-limit estimator on the whole line. Annals of Statistics. 1983; 11:49–58.
9. Groeneboom, P.; Wellner, J. Information Bounds and Nonparametric Maximum Likelihood Estimation. Birkhäuser Verlag; Basel: 1992.
10. Gu MG, Zhang CH. Asymptotic properties of self-consistent estimators based on doubly censored data. Annals of Statistics. 1993; 21:501–519.
11. Hájek J. A characterization of limiting distributions of regular estimates. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete. 1970; 14:323–330.
12. Heckman J. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Economic and Social Measurement. 1976; 5(4):475–492.
13. Huang J. Efficient estimation for the proportional hazards model with interval censoring. Annals of Statistics. 1996; 24:540–568.
14. Huang J, Wellner JA. Asymptotic normality of the NPMLE of the linear functionals for interval censored data, case 1. Statistica Neerlandica. 1995; 49:153–163.
15. Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High-dimensional variable selection for survival data. Journal of the American Statistical Association. 2010; 105:205–217.
16. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association. 1958; 53:457–481.
17. Kim MY, De Gruttola VG, Lagakos SW. Analyzing doubly censored data with covariates, with application to AIDS. Biometrics. 1993; 49:13–22. [PubMed: 8513098]
18. LeCam L. On the assumptions used to prove asymptotic normality of maximum likelihood estimates. Annals of Mathematical Statistics. 1970; 41:802–828.
19. LeCam L, Yang GL. On the preservation of local asymptotic normality under information loss. Annals of Statistics. 1988; 16:483–520.
20. Prentice R, Pyke R. Logistic disease incidence models and case-control studies. Biometrika. 1979; 66:403–411.
21. Ren J-J. Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data. Annals of Statistics. 2008; 36:147–166.
22. Ritov Y, Bickel P. Achieving information bounds in non and semiparametric models. Annals of Statistics. 1990; 18:925–938.
23. Song K, Elston RC. A powerful method of combining measures of association and Hardy–Weinberg disequilibrium for fine-mapping in case-control studies. Statistics in Medicine. 2006; 25:105–126. [PubMed: 16220513]
24. Stute W. The central limit theorem under random censorship. Annals of Statistics. 1995; 23:422–439.
25. Stute W, Wang JL. The strong law under random censorship. Annals of Statistics. 1993; 21:1591–1607.
26. Sun, J. The Statistical Analysis of Interval-Censored Failure Time Data. Springer Science + Business Inc; 2006.
27. Tricomi, FG. Integral Equations. Dover Publications; New York: 1985.
28. Tsai W-Y, Crowley J. A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. Annals of Statistics. 1985; 13:1317–1334.
29. Turnball BW. Nonparametric estimation of a survivorship function with doubly censored data. Journal of American Statistical Association. 1974; 69:169–173.
30. van der Vaart, A. CWI Tract. Vol. 44. Centrum voor Wiskunde en Information; Amsterdam: 1988. Statistical Estimation in Large Parameter Spaces.
31. van der Vaart A. On differentiable functionals. Annals of Statistics. 1991; 19:178–205.

32. Xia Y, Zhang D, Xu J. Dimension reduction and semiparametric estimation of survival models. Journal of the American Statistical Association. 2010; 105:278–290.
33. Yang S. Censored median regression using weighted empirical survival and hazard functions. Journal of American Statistical Association. 1999; 94:137–145.
34. Zheng, G.; Yang, Y.; Zhu, X.; Elston, RC. Analysis of Genetic Association Studies. Springer; 2012. in press
35. Zhou M. Empirical likelihood analysis of the rank estimator for the censored accelerated failure time model. Biometrika. 2005; 92:492–498.

Appendix

Proof of Theorem 1

We only prove the result for H known; then the result is automatic for H unknown. Thus we identify the tangent space as $\dot{Q}=L_2^0(Q)=L_2^0(F)$. To evaluate the score operator, we need the conditional density $f(z|u, v, \delta, \gamma)$ of $Z|T(X^0) = Z|X = x$. As $\{(\delta, \gamma) = (0, 0)\} = \{(Z > V)\}$, $\{(\delta, \gamma) = (0, 1)\} = \{(U < Z \le V)\}$ and $\{(\delta, \gamma) = (1, 1)\} = \{(Z \le U)\}$, we have $F(z|u, v, 0, 0) = P(Z \le z|Z > v) = 1_{z>v} F(z)/(1 - F(v))$, $F(z|u, v, 0, 1) = 1_{(u<z\le u)}F(z)/(F(v) - F(u))$ and $F(z|u, v, 1, 1) = 1_{(z\le u)}F(z)/F(u)$. Thus

$$f(z|u,v,\delta,\gamma)dz=\delta\frac{1_{(z\le u)}F(dz)}{F(u)}+\gamma(1-\delta)\frac{1_{(u<z\le v)}F(dz)}{F(v)-F(u)}+(1-(\delta\vee\gamma))\frac{1_{(z>v)}F(dz)}{1-F(v)}.$$

Since h is known, the score operator $\mathbf{i} = \mathbf{i}_1$ is given by, $\forall a \in L_2^0(F)$,

$$\begin{aligned}(\mathbf{i}_1a)(x)&=(\mathbf{i}_1a)(u,v,\delta,\gamma)=E\left(a(Z)|T(X^0)=(u,v,\delta,\gamma)\right)=E(a(Z)|(\delta,\gamma))\\&=\delta\frac{\int_0^u a(z)F(dz)}{F(u)}+\gamma(1-\delta)\frac{\int_u^v a(z)F(dz)}{F(v)-F(u)}+(1-(\delta\vee\gamma))\frac{\int_v^\infty a(z)F(dz)}{1-F(v)}.\end{aligned}\qquad\text{(A.1)}$$

Its adjoint $\mathbf{i}_1^*$ satisfies $\langle b,\mathbf{i}_1a\rangle_{L_2(P_{F,H})}=\langle\mathbf{i}_1^*b,a\rangle_{L_2(F)}$, $\forall(a,b)\in(L_2^0(F),L_2^0(P_{F,H}))$. Since

$$\begin{aligned}\langle b,\mathbf{i}_1a\rangle_{L_2(P_{F,H})}&=\int b(x)\left(\delta\frac{\int_0^u a(z)F(dz)}{F(u)}+\gamma(1-\delta)\frac{\int_u^v a(z)F(dz)}{F(v)-F(u)}\right.\\&\left.+(1-(\delta\vee\gamma))\frac{\int_v^\infty a(z)F(dz)}{1-F(v)}\right)p_{F,H}(x)dx=\int\left(b(u,v,1,1)\int_0^u a(z)F(dz)\right.\\&\left.+(b(u,v,0,1)\int_u^v a(z)F(dz)+b(u,v,0,0)\int_0^\infty a(z)F(dz)\right)H(du,dv)\\&=\int\left(\int\left[b(u,v,1,1)1_{(u>z)}+b(u,v,0,1)1_{(u\le z<v)}+b(u,v,0,0)1_{(v<z)}\right]H(du,dv)\right)a(z)F(dz),\end{aligned}$$

i.e., $\mathbf{i}_1^*$ is given by, $\forall b \in L_2^0(P_{F,H})$,

$$(\mathbf{i}_1^*b)(z)=\int\left[b(u,v,1,1)1_{(u>z)}+b(u,v,0,1)1_{(u\le z<v)}+b(u,v,0,0)1_{(v<z)}\right]H(du,dv).\qquad\text{(A.2)}$$

As in [9, p. 8] or [30], or [19], we also have $(\mathbf{i}_1^* b)(z)=E_{P_{F,H}}(b(X)|Z=z), \forall b \in L_2^0(P_{F,H})$, so $E_F[(\mathbf{i}_1^* b)(Z)]=E_{P_{F,H}}(b(X))=0$, i.e. we have automatically $\mathbf{i}_1^* b \in L_2^0(F)$. Now, we have, $\forall a \in L_2^0(F)$,

$$\begin{aligned}
(\mathbf{i}_1^*\mathbf{i}_1 a)(z)=\mathbf{i}_1^*[(\mathbf{i}_1 a)(x)]=\mathbf{i}_1^*\left[\delta\frac{\int_0^u a(t)F(dt)}{F(u)}+\gamma(1-\delta)\frac{\int_u^v a(t)F(dt)}{F(v)-F(u)}+(1-(\delta\vee\gamma))\frac{\int_v^\infty a(t)F(dt)}{1-F(v)}\right]\\
=\iint\left[\frac{\int_0^u a(t)F(dt)}{F(u)}1_{(u>z)}+\frac{\int_u^v a(t)F(dt)}{F(v)-F(u)}1_{(u\le z<v)}+\frac{\int_v^\infty a(t)F(dt)}{1-F(v)}1_{(v<z)}\right]H(du,dv)\\
=\int_z^\infty\int\frac{\int_0^u a(t)F(dt)}{F(u)}H(du,dv)+\int_0^z\int_z^\infty\frac{\int_u^v a(t)F(dt)}{F(v)-F(u)}H(du,dv)+\int\int_0^z\frac{\int_v^\infty a(t)F(dt)}{1-F(v)}H(du,dv)\\
=\int\left[\int_{z\vee t}^\infty\int\frac{H(du,dv)}{F(u)}+\int_0^{z\wedge t}\int_{z\vee t}^\infty\frac{dH(u,v)}{F(v)-F(u)}+\int\int_0^{z\wedge t}\frac{H(du,dv)}{1-F(v)}\right]a(t)F(dt)\\
:=\int K(z,t)a(t)F(dt).
\end{aligned}$$

Thus solving $\mathbf{i}_1^*\mathbf{i}_1 a=b$ in $a \in L_2^0(F)$ for any given $b \in L_2^0(F)$ amounts to solving a Fredholm equation of the first kind $b(z) = \int K(z, t)a(t)F(dt)$ for all z, with kernel $K(\cdot, \cdot)$, and $K(z, t)$ is not symmetrical nor of the form $K(z - t)$. It is known (see, for example, [27]) that such an equation is extremely ill conditioned: it has no solution for many functions b, i.e., $\mathbf{i}_1^*\mathbf{i}_1$ is not boundedly invertible.

Proof of Theorem 2

(i) Denote $F(z) = \Psi(f) = \int 1_{[0,z]}(x)f(x)dx = \int 1_{[0,z]}(x)dF(x)$. Then, by Proposition A.5.2 in BKRW, $F(z)$ is pathwise differentiable with respect to F, with adjoint pathwise differential valued at 1 as $\dot{\Psi}^*(x) := (\dot{\Psi}^* 1)(x) = 1_{[0,z]}(x) - F(z)$, which is a discontinuous function. On the other hand, in (A.2) we see that $\mathbf{i}_1^* b$ is a continuous function for each fixed $b \in L_2^0(P_{F,H})$; thus $\dot{\Psi}^* \notin R(\mathbf{i}_1^*)$, the range of $\mathbf{i}_1^*$, and, by Theorem 3.1 in van der Vaart [31] (or Theorem 5.4.1 in BKRW), $F(z)$ is not pathwise differentiable with respect to $P_{F,H}$.

(ii) Denote $\nu = \Psi_1(f) = \int g(z)F(dz) = \int E(a(U, V, \delta, \gamma)|Z = z)F(dz)$. The score operator $\mathbf{i}_1$ and its adjoint $\mathbf{i}_1^*$ are given in (A.1) and (A.2). Note that $\mathbf{i}_1^*: L_2^0(P_{F,H}) \to L_2^0(F)$ can also be written as $(\mathbf{i}_1^* b)(z)=E(b(U,V,\delta,\gamma)|Z=z), \forall b \in L_2^0(P_{F,H})$, and by the expression of $(\mathbf{i}_1 b)(\cdot)$, it is a continuous function in z. Similarly as in the proof of (i), we have $(\dot{\Psi}_1^* 1)(z)=g(z)-Eg(Z)=E(a(U,V,\delta,\gamma)|Z=z)-Ea(U,V,\delta,\gamma)=\mathbf{i}_1^*(a-Ea)(z)$. As an operator $\dot{\Psi}_1^*: R^*=R \to L_2^0(F), \dot{\Psi}_1^*(c)=c(g(\cdot)-Eg(Z))=\mathbf{i}_1^*(ca-cEa)(z)$, $\forall c \in R$, and it is continuous, $ca-cEa \in L_2^0(P_{F,H})$; thus $R(\dot{\Psi}_1^*) \subset R(\mathbf{i}_1^*) \subset R(\mathbf{i}^*)$, and so, by Theorem 5.4.1 in BKRW, ν is pathwise differentiable with respect to $P_{F,H}$.

(iii) Let

$$b(u,v,\delta,\gamma)=\frac{h_1(u)r(v)(1-\delta)(1-\gamma)}{h(u,v)}.$$

Then $b(u, v, \delta, \gamma) \in L_2(P_{F,H})$. Using the expression for $(\mathbf{i}_1^* b)(z)$ given in the proof of Theorem 1, we have

$$E(b(u,v,\delta,\gamma)|Z=z)=(\mathbf{i}_1^* b)(z)=\int_0^z r(v)dv:=g(z),$$

and, by (ii), $\mu = E(g(Z)) = E(b(U, V, \delta, \gamma))$ is pathwise differentiable with respect to $P_{F,H}$. In particular, if $r(v) = kv^{k-1}$, μ is the k-th moment of F. To evaluate the information bound for either case, we need the following.

Lemma

If $\dot{\mathbf{P}} = L_2^0(P)$, and the score operator $\mathbf{i}:L_2^0(Q) \to L_2^0(P)$ is given by $(\mathbf{i}a)(x) = E(a(X^0)|X = x)$, $\forall a \in L_2^0(Q)$, then $\dot{\mathbf{P}} = \overline{R(\mathbf{i})}$, the closure of the range of $\mathbf{i}$.

Proof

Apparently, $(\mathbf{i}a)(\cdot) \in L_2^0(P), \forall a \in L_2^0(Q)$; thus $\overline{R(\mathbf{i})}=\overline{\{\mathbf{i}a:a \in L_2^0(Q)\}} \subset \overline{L_2^0(P)}= \dot{\mathbf{P}}$. On the other hand, $\forall b \in \dot{\mathbf{P}} = L_2^0(P), a(x^0):=E(b(X)|X^0=x^0)=(\mathbf{i}b)(x^0) \in L_2^0(Q)$; thus $\dot{\mathbf{P}} \subset \overline{\{\mathbf{i}a:a \in L_2^0(Q)\}}$.

Assume H unknown. We will show that the information bound is the same for H known or not. The score operator for H, $\mathbf{i}_2:\dot{\mathbf{P}}_2=L_2^0(H) \to L_2^0(P_{F,H})= \dot{\mathbf{P}}$, and its adjoint, $\mathbf{i}_2^*:L_2^0(P_{F,H}) \to L_2^0(H)$, are given by

$$\begin{aligned}(\mathbf{i}_2 b)(x)=E(b(U,V)|T(X^0)=(u,v,\delta,\gamma))=b(u,v), \quad \forall b \in L_2^0(H),\\ \text{and} \quad (\mathbf{i}_2^* a)(u,v)=E[a(U,V,\delta,\gamma)|(U,V)=(u,v)], \quad \forall a \in L_2^0(P_{F,H}).\end{aligned}$$

Note that R is self-dual. As in the proof of (ii), the adjoint pathwise derivative $\dot{\Psi}_1^*:R^*=R \to L_2^0(F)$ of $\mu = \Psi_1(f)$ with respect to F is

$$\dot{\Psi}_1^*(c)=c(g(\cdot)-\mu), \quad \forall c \in R.$$

In our case, F is unrestricted, so $\dot{\mathbf{P}}_1=L_2^0(F)$, and by the lemma, $\dot{\mathbf{P}}_1=\overline{R(\mathbf{i}_1)}$. Similarly, $\dot{\mathbf{P}}_2=\overline{R(\mathbf{i}_2)}$. Let $\mathbf{i}_{12}=\Pi(\mathbf{i}_1|\dot{\mathbf{P}}_2^{\perp}), \mathbf{i}_{21}=\Pi(\mathbf{i}_2|\dot{\mathbf{P}}_1^{\perp})$, and $\mathbf{i}_{12}^*$ and $\mathbf{i}_{21}^*$ be their adjoints, respectively. Since $R(\dot{\Psi}_1^*) \subset R(\mathbf{i}_1^*)$ as in the proof of (ii), note we will show $R(\mathbf{i}_{21}^*)=R(\mathbf{i}_2^*)$, and by Corollary 5.5.1 in BKRW, thus, $\mu \in R^*$, the unique efficient influence operator $\tilde{I}:R^* \to L_2^0(P_{F,H})$ for estimation of μ is the solution of the operator equations

$$\begin{cases} \dot{\Psi}_1^*=\mathbf{i}_{12}^*\tilde{I} \\ 0=\mathbf{i}_2^*\tilde{I}. \end{cases} \tag{A.3}$$

In fact, as in the proof of Corollary 5.5.1 in BKRW (p. 218), we have $\dot{\mathbf{P}}_2^{\perp}=R(\mathbf{i}_2)^{\perp}=N(\mathbf{i}_2^*)$, the null space of $\mathbf{i}_2^*$. Using the expression for $\mathbf{i}_1^*$ given in (A.2), we get

$$(\mathbf{i}_1^*\mathbf{i}_2 b)(z)=Eb(U,V)=0, \quad \forall z \in R^+, \forall b \in L_2^0(H),$$

which is equivalent to $R(\mathbf{i}_1) \perp R(\mathbf{i}_2)$ and to $(\mathbf{i}_2^*\mathbf{i}_1 a)(u,v)=0, \forall (u,v) \in (R^+)^2, \forall a \in L_2^0(F)$. This implies that $\dot{\mathrm{P}}_1 \perp \dot{\mathrm{P}}_2$, and the information bound for estimating μ is the same for H known or unknown.

Also, $\mathbf{i}_1 \in N(\mathbf{i}_2^*)$; consequently, $\mathbf{i}_{12}=\Pi(\mathbf{i}_1|N(\mathbf{i}_2^*))=\mathbf{i}_1$ and $\mathbf{i}_{12}^*=\mathbf{i}_1^*$. Similarly, $\mathbf{i}_{21}^*=\mathbf{i}_2^*$, and $R(\mathbf{i}_{12}^*)=R(\mathbf{i}_2^*)$. Now, (A.3) becomes

$$\begin{cases} \dot{\Psi}_1^*=\mathbf{i}_1^*\tilde{I} \\ 0=\mathbf{i}_2^*\tilde{I}. \end{cases} \tag{A.4}$$

Take $\tilde{I}(c) = c(b(U, V, \delta, \gamma) - b_1(U, V))$, with $b_1(u, v) = E[b(U, V, \delta, \gamma)|(U, V) = (u, v)]$. Then $\tilde{I}(c) \in \dot{\mathbf{P}} = L_2^0(P_{F,H})$, $\forall c \in R^* = R$. Since $E(b_1(U, V)|Z = z) = Eb_1(U, V) = Eb(U, V, \delta, \gamma) = \mu$, we have

$$\mathbf{i}_1^*\tilde{I}(c)=E_{P_{F,H}}(\tilde{I}(c)|Z=z)=c(g(z)-\mu)=\dot{\Psi}_1^*(c), \quad \forall c \in R.$$

Also, $\mathbf{i}_2^*\tilde{I}(c)=E_{P_{F,H}}[\tilde{I}(c)|(U,V))=0$; thus $\tilde{I}$ is the solution of (A.4), the efficient influence function for estimating μ is

$$\dot{\mu}^*(1)=\tilde{I}(1)=b(u,v,\delta,\gamma)-b_1(u,v),$$

and the information bound is

$$I(\mu)=E_{P_{F,H}}(\tilde{I}^2(1))=\int\int \left(b(u,v,0,0)(b(u,v,0,0)-2b_1(u,v))(1-F(v))+b_1^2(u,v)\right)H(du,dv).$$

When h is known, so is b, and $\mu_n=n^{-1}\sum_{i=1}^n b(U_i,V_i,\delta_i,\gamma_i)$ is a rate- $\sqrt{n}$ unbiased estimator of μ, so it is rate- $\sqrt{n}$estimable.

When $\{Z_i: i = 1, \dots, n\}$ is observed, the estimate of μ should be based on $\{Z_i: i = 1, \dots, n\}$, and $\mu_e = E(b(U, V, \delta, \gamma)) \in R$ defines an extension of μ from $\mathcal{P}_1 = \{F \in \mathcal{F}\}$ to $\mathcal{P} = \{P_{F,H}, F \in \mathcal{F}, H \in \mathcal{H}\}$, where $\mathcal{F}$ and $\mathcal{H}$ are some collections of distribution functions. Denote $\dot{\mu}_e: \dot{\mathbf{P}} \to R$ the pathwise differential operator of μ_e, where $\dot{\mathbf{P}}$ is the tangent space of model $P_{F,H}$. Obviously, the adjoint $\dot{\mu}_e^*: R \to \dot{\mathbf{P}}$ of $\dot{\mu}_e$ valued at 1 is $\dot{\mu}_e^*(x):=(\dot{\mu}_e^* 1)(x)=b(u,v,\delta,\gamma)-\mu$. Let $\Pi(y|x)$ be the projection of y on to the space spanned by x, and let $\dot{\mathbf{P}}_1$ be the tangent space for F in the model $P_{F,H}$. The efficient influence function for estimating μ in the model F is

$$\tilde{I}=\Pi(\dot{\mu}_e^*|\dot{\mathbf{P}}_1)=\dot{\mu}_e^*-\Pi(\dot{\mu}_e^*|\dot{\mathbf{P}}_1^{\perp}).$$

Since $\dot{\mathbf{P}}_1=\overline{R(\mathbf{i}_1)}$, we have $\dot{\mathbf{P}}_1^{\perp}=N(\mathbf{i}_1^*)=\{h:h\in L_2^0(P_{F,H}),\mathbf{i}_1^* h=0\}$. Since $\dot{\mu}_e^* - \mathbf{i}_1^*\dot{\mu}_e^* \in L_2^0(P_{F,H})$ and $\mathbf{i}_1^*(\dot{\mu}_e^* - \mathbf{i}_1^*\dot{\mu}_e^*)=E(\dot{\mu}_e^*|Z=z)-E[E(\dot{\mu}_e^*|Z=z)|Z=z)=0, \dot{\mu}_e^* - \mathbf{i}_1^*\dot{\mu}_e^* \in N(\mathbf{i}_1^*)$, and we have

$$\Pi(\dot{\mu}_e^*|\dot{\mathbf{P}}_1^{\perp})=\Pi(\dot{\mu}_e^*|N(\mathbf{i}_1^*))=\dot{\mu}_e^* - \mathbf{i}_1^*\dot{\mu}_e^*=b(u,v,\delta,\gamma)-g(z),$$
$$\text{so}\quad \tilde{I}=\dot{\mu}_e^*-(b(u,v,\delta,\gamma)-g(z))=g(z)-\mu,$$

and $I(\mu)=E(\tilde{I}^2)=E(g(Z)-\mu)^2=Eg^2(Z)-\mu^2$.

(iv) Given the covariate w, the density for X is

$$p_{F,H}(x|w)=\left(1-\overline{F}(u)^{\exp(\theta' w)}\right)^{\delta}\left(\overline{F}(u)^{\exp(\theta' w)}-\overline{F}(v)^{\exp(\theta' w)}\right)^{\gamma(1-\delta)}\overline{F}(v)^{(1-(\delta\vee\gamma))\exp(\theta' w)}h(u,v),$$

with log-likelihood

$$l(\theta,f)=\delta\log\left(1-\overline{F}(u)^{\exp(\theta' w)}\right)+\gamma(1-\delta)\log\left(\overline{F}(u)^{\exp(\theta' w)}-\overline{F}(v)^{\exp(\theta' w)}\right)+(1-(\delta\vee\gamma))\exp(\theta' w)\log\overline{F}(v)+\log h(u,v).$$

Let $l_\theta(\theta,f)$ be the partial derivative of $l(\theta,f)$ with respect to θ, and let $l_f(\theta,f)(a)$ be the Hadamard differential of $l(\theta,f)$ with respect to f in the direction af (for $a\in L_2^0(F)$. Note that $\Lambda(u)=-\log\bar{F}(u)$. Then

$$l_\theta(\theta,f)=\left(\delta\frac{\overline{F}(u|w)\Lambda(u)}{1-\overline{F}(u|w)}-\gamma(1-\delta)\frac{\overline{F}(u|w)\Lambda(u)-\overline{F}(v|w)\Lambda(v)}{\overline{F}(u|w)-\overline{F}(v|w)}-(1-(\delta\vee\gamma))\Lambda(v)\right)\exp(\theta' w)w,$$

and

$$l_f(\theta,f)(a)=\left(-\delta\frac{\overline{F}(u|w)\int_u^\infty adF}{\overline{F}(u)(1-\overline{F}(u|w))}+\gamma(1-\delta)\frac{\overline{F}(v)\overline{F}(u|w)\int_u^\infty adF-\overline{F}(u)\overline{F}(v|w)\int_v^\infty adF}{\overline{F}(u)\overline{F}(v)(\overline{F}(u|w)-\overline{F}(v|w))}+(1-(\delta\vee\gamma))\frac{\int_v^\infty adF}{\overline{F}(v)}\right)\exp(\theta' w).$$

The efficient score for estimating θ in the presence of nuisance f is

$$l_{\theta|f}=l_\theta(\theta,f)-l_f(\theta,f)(a^*),$$

where, with k=dim(w), $a^*=(a_1^*,\ldots,a_k^*)'$, $l_f(\theta,f)(a^*)=(l_f(\theta,f)(a_1^*),\ldots,l_f(\theta,f)(a_k^*))'$, we have that a^* satisfies, in the componentwise sense,

$$
\begin{aligned}
&l_\theta(\theta,f)-l_f(\theta,f)(a^*)\perp l_f(\theta,f)(a),\quad \forall a\in L_2^0(F)\text{ or}\\
&0=-\langle l_\theta(\theta,f)-l_f(\theta,f)(a^*),l_f(\theta,f)(a)\rangle_P\\
&=E_{(U,V)}\left(E_W\left\{\exp(2\theta' W)\left[\frac{\Lambda(V)\overline{F}(V)W+\int_V^\infty a^*dF}{\overline{F}(V)}\frac{\int_V^\infty adF}{\overline{F}(V)}\overline{F}(V|W)\right.\right.\right.\\
&+\frac{\overline{F}(V)\overline{F}(U|W)[\Lambda(U)\overline{F}(U)W+\int_U^\infty a^*dF]-\overline{F}(U)\overline{F}(V|W)[\Lambda(V)\overline{F}(V)W+\int_V^\infty a^*dF]}{\overline{F}(U)\overline{F}(V)(\overline{F}(U|W)-\overline{F}(V|W))}\\
&\times\frac{\overline{F}(V)\overline{F}(U|W)\int_U^\infty adF-\overline{F}(U)\overline{F}(V|W)\int_V^\infty adF}{\overline{F}(U)\overline{F}(V)}\\
&\left.\left.\left.+\frac{\overline{F}^2(U|W)[\Lambda(U)\overline{F}(U)W+\int_U^\infty a^*dF]}{\overline{F}(U)(1-\overline{F}(U|W))}\frac{\int_U^\infty adF}{\overline{F}(U)}\right]\middle|(U,V)\right\}\right).
\end{aligned}
$$

Set $E_W\{\cdots|(U,V)=(u,v)\}=0$ for all $u<v$ and let $u\nearrow v$. Noting that the second summand in $E_W\{\cdots|(U,V)=(u,v)\}$ tends to zero (vector), we have

$$
0=E_W\left\{\exp(2\theta' W)\left[\frac{\Lambda(V)\overline{F}(V)W+\int_V^\infty a^*dF}{\overline{F}(V)}\overline{F}(V|W)+\frac{\overline{F}^2(V|W)(\Lambda(V)\overline{F}(V)W+\int_V^\infty a^*dF)}{\overline{F}(V)(1-\overline{F}(V|W))}\right]\middle|V=v\right\}\frac{\int_v^\infty adF}{\overline{F}(v)},\quad \forall a\in L_2^0(F),\forall v,
$$

and so

$$
0=E_W\left\{\exp(2\theta' W)\left[\frac{\Lambda(V)\overline{F}(V)W+\int_V^\infty a^*dF}{\overline{F}(V)}\overline{F}(V|W)+\frac{\overline{F}^2(V|W)(\Lambda(V)\overline{F}(V)W+\int_V^\infty a^*dF)}{\overline{F}(V)(1-\overline{F}(V|W))}\right]\middle|V=v\right\},\quad \forall v,
$$

or, with division in the componentwise sense,

$$
\begin{aligned}
&\int_v^\infty a^*dF=-\Lambda(v)\overline{F}(v)\frac{E\left\{\exp(2\theta' W)\frac{\overline{F}(V|W)}{1-\overline{F}(V|W)}W|V=v\right\}}{E\left\{\exp(2\theta' W)\frac{\overline{F}(V|W)}{1-\overline{F}(V|W)}|V=v\right\}}\\
&:=-\Lambda(v)\overline{F}(v)\frac{E(O(W,V)W|V=v)}{E(O(W,V)|V=v)}=-\Lambda(v)\overline{F}(v)R(v).
\end{aligned}
$$

Similarly, in the above computation with $V=v\searrow U=u$, we have

$$
\int_u^\infty a^*dF=-\Lambda(u)\overline{F}(u)\frac{E(O(W,U)W|U=u)}{E(O(W,U)|U=u)}=-\Lambda(u)\overline{F}(u)R(u).
$$

Plugging $\int_u^\infty a^*dF$ and $\int_v^\infty a^*dF$ in the expression of $l_f(\theta,f)(a)$ with a replaced by a^*, we get the expression for $l_f(\theta,f)(a^*)$, and hence for $l_{\theta|f}$.

Note that, in the above computations, if we set $\gamma\equiv 1$ and $v=\infty$, then we have the interval censoring model case I, and the expressions for $l_\theta(\theta,f)$, $l_f(\theta,f)$ and $\int_u^\infty a^*dF$ are exactly the same as in [13].

Proof of Theorem 3

Let $h_{U|V}(u|v)$ be the conditional density of $U|V = v$, accordingly, for $h_{V|U}(v|u)$.

(i) We first find the conditional density $f(z, u, v|x) = f(z, u, v|y, \delta, \gamma)$ of $X^0|T(X^0) = x$. If $(\delta, \gamma) = (0, 0)$,

$$f(z,u,v|y,\delta,\gamma)=f(z,u,v|Z>V,V=y)=\frac{1_{(z>y,v\geq y)}f(z)h_{U|V}(u|y)}{1-F(y)},$$

if $(\delta, \gamma) = (0, 1)$,

$$f(z,u,v|y,\delta,\gamma)=f(z,u,v|U<Z\leq V,Z=y)=\frac{1_{(u<y\leq v)}h(u|v)}{H_U(y)-H(y,y)},$$

and if $(\delta, \gamma) = (1, 0)$,

$$f(z,u,v|y,\delta,\gamma)=f(z,u,v|Z\leq U,U=y)=\frac{1_{(z\leq y,u\leq y)}f(z)h_{V|U}(v|y)}{F(y)},$$

or

$$f(z,u,v|x)=(1-\delta-\gamma)\frac{1_{(z>y,v\geq y)}f(z)h_{U|V}(u|y)}{1-F(y)}+\gamma\frac{1_{(u<y\leq v)}h(u,v)}{H_U(y)-H(y,y)}+\delta\frac{1_{(z\leq y,u\leq y)}f(z)h_{V|U}(v|y)}{F(y)}.$$

So $\mathbf{i}$ is given by, $\forall a\in L_2^0(Q)$,

$$\begin{aligned}(\mathbf{i}a)(x)=E(a(X^0)|T(X^0)=x)=E(a(Z,U,V)|y,\delta,\gamma)\\=(1-\delta-\gamma)\int\int\tfrac{1_{(z>y,v\geq y)}}{1-F(y)}a(z,u,v)h_{U|V}(u|y)F(dz)du\\+\gamma\int\int\tfrac{1_{(u<y\leq v)}}{H_U(y)-H(y,y)}a(y,u,v)H(du,dv)+\delta\int\int\tfrac{1_{(z\leq y,u\leq y)}}{F(y)}a(z,y,v)h_{V|U}(v|y)F(dz)dv.\end{aligned}$$

Since $\forall b \in L_2(P_{F,H})$ and $a \in L_2(Q)$, using the expression of $p_{F,H}$, $\mathbf{i}^*$ satisfies

$$\begin{aligned}\langle\mathbf{i}^*b,a\rangle_{L_2(Q)}=\langle b,\mathbf{i}a\rangle_{L_2(P_{F,H})}\\=\iiint\left(1_{(v<z)}b(v,0,0)+1_{(u<z\leq v)}b(z,0,1)+1_{(u>z)}b(u,1,0)\right)a(z,u,v)F(dz)H(du,dv),\end{aligned}$$

we have, $\forall b \in L_2(P_{F,H})$,

$$(\mathbf{i}^*b)(z,u,v)=1_{(v<z)}b(v,0,0)+1_{(u<z\leq v)}b(z,0,1)+1_{(u>z)}b(u,1,0).$$

Now, we have

$$(\mathbf{i}^*\mathbf{i}a)(z,u,v)=\mathbf{i}^*((\mathbf{i}a)(y,\delta,\gamma))=1_{(v<z)}\iint\frac{1_{(t>v)}}{1-F(v)}a(t,r,v)F(dt)H_{U|V}(dr|v)$$
$$+1_{(u<z\le v)}\iint\frac{1_{(r<z\le s)}}{H_U(z)-H(z,z)}a(z,r,s)H(dr,ds)+1_{(u>z)}$$
$$\times\iint\frac{1_{(u\ge t)}}{F(u)}a(t,u,s)F(dt)H_{V|U}(ds|u).$$

This is a linear combination of three Fredholm equations of the first kind for $a \in L_2^0(Q)$, each with a kernel non-symmetric nor of the form $K(s, t) = K(s-t)$. That is, solving a in the equation $\mathbf{i}^*\mathbf{i}a = b$ for given b is solving $b(s)=\sum_{j=1}^{3}\int K_j(s,t)F(dt)$; as before, each of which cannot be solved for many functions b, and hence $\mathbf{i}^*\mathbf{i}$ is not boundedly invertible.

(ii) In this case, we identify $\dot{Q}=L_2^0(F)$, and denote the score operator for F as $\mathbf{i}_1$. Now, $\forall a \in L_2^0(F)$,

$$(\mathbf{i}_1a)(y,\delta,\gamma)=E(a(Z)|T(X^0)=x)=E(a(Z)|y,\delta,\gamma)$$
$$=(1-\delta-\gamma)\int\frac{1_{(z>y)}}{1-F(y)}a(z)F(dz)+\gamma a(y)+\delta\int\frac{1_{(z\le y)}}{F(y)}a(z)F(dz).$$

For $\mathbf{i}_1^*$, $\forall b \in L_2^0(P_{F,H})$ and $a \in L_2^0(F)$, we have

$$\langle\mathbf{i}_1^*b,a\rangle_{L_2(F)}=\langle b,\mathbf{i}_1a\rangle_{L_2(P_{F,H})}=\int b(y,0,0)(\mathbf{i}_1a)(y,0,0)p_{F,H}(y,0,0)dy$$
$$+\int b(y,0,1)(\mathbf{i}_1a)(y,0,1)p_{F,H}(y,0,1)dy+\int b(y,1,0)(\mathbf{i}_1a)(y,1,0)p_{F,H}(y,1,0)dy$$
$$=\int\left(\int_{-\infty}^{z}b(y,0,0)H_V(dy)+b(z,0,1)M(z)+\int_z^{\infty}b(y,1,0)H_U(dy)\right)a(z)F(dz),$$

which gives

$$(\mathbf{i}_1^*b)(z)=\int_{-\infty}^{z}b(y,0,0)H_V(dy)+b(z,0,1)M(z)+\int_z^{\infty}b(y,1,0)H_U(dy), \quad \text{(A.5)}$$

and $\mathbf{i}_1^*\mathbf{i}_1$ is given by the operation, $\forall a \in L_2^0(F)$,

$$(\mathbf{i}_1^*\mathbf{i}_1a)(z)=\mathbf{i}_1^*[(\mathbf{i}_1a)(y,\delta,\gamma)]=\int_{-\infty}^{z}\int_y^{\infty}\frac{a(t)}{1-F(y)}F(dt)H_V(dy)+M(z)a(z)+\int_z^{\infty}\int_{-\infty}^{y}\frac{a(t)}{F(y)}F(dt)H_U(dy)$$
$$=M(z)a(z)+\int\left(\int_{-\infty}^{z\wedge t}\frac{1}{1-F(y)}H_V(dy)+\int_{t\vee z}^{\infty}\frac{1}{F(y)}H_U(dy)\right)a(t)F(dt)$$
$$:=M(z)a(z)+\int K(z,t)a(t)F(dt).$$

Since $M(z) \neq 0$, for given $b \in L_2^0(F)$, solving $a \in L_2^0(L)$ from $\mathbf{i}_1^*\mathbf{i}_1a=b$ is a Fredholm equation of the second kind, which admits a unique solution given by a Liouville–Neumann series; see pp. 288–289 in BKRW for more details. Thus, $\mathbf{i}_1^*\mathbf{i}_1$ is boundedly invertible in this case.

Or, alternatively, to show that $\mathbf{i}_1^*\mathbf{i}_1$ is boundedly invertible, we only need to show the bounded invertiability for $A:=(\mathbf{i}_1^*\mathbf{i}_1)^{1/2}$. Since $A^* = A$, by Corollary A.1.2 in BKRW, A is boundedly invertible if and only if $||Aa|| \geq \varepsilon\, ||a||$, or equivalently $||Aa||^2 \geq \varepsilon\, ||a||^2$, $\forall a \in L_2^0(F)$, for some $\varepsilon > 0$. In fact, assuming that $\inf_y M(y) = \varepsilon > 0$, we have

$$\begin{aligned}
\left\|Aa\right\|^2 &= \langle Aa, Aa\rangle_{L_2(F)} = \langle a, A^*Aa\rangle_{L_2(F)} = \langle a, \mathbf{i}_1^*\mathbf{i}_1 a\rangle_{L_2(F)} = \int M(z)a^2(z)F(dz) \\
&+ \int a(z)\int_{-\infty}^{z} \frac{\int_y^\infty a(t)F(dt)}{1-F(y)} H_V(dy)F(dz) + \int a(z)\int_z^\infty \frac{\int -\infty^y a(t)F(dt)}{F(y)} H_U(dy)F(dz) \\
&= \int M(z)a^2(z)F(dz) + \int \frac{\left(\int_y^\infty a(t)F(dt)\right)^2}{1-F(y)} H_V(dy) + \int \frac{\left(\int_{-\infty}^y a(t)F(dt)\right)^2}{F(y)} H_U(dy) \\
&\geq \int M(z)a^2(z)F(dz) \geq \varepsilon \int a^2(z)F(dz) = \varepsilon\left\|a\right\|^2.
\end{aligned}$$

(iii) In this case, we identify $\dot{Q} = L_2^0(H)$, and the score operator for H as $\mathbf{i}_2$. Then, $\forall a \in L_2^0(H)$,

$$\begin{aligned}
(\mathbf{i}_2 a)(y,\delta,\gamma) &= E(a(U,V)|y,\delta,\gamma) = (1-\delta-\gamma) \iint \frac{1_{(z>y)}}{1-F(y)} a(u,y)h_{U|V}(u|y)duF(dz) \\
&+ \gamma \iint \frac{1_{(u<y\leq v)}}{M(y)} a(u,v)H(du,dv) + \delta \iint \frac{1_{(z\leq y)}}{1-F(y)} a(y,v)h_{V|U}(v|y)dvF(dz) \\
&= (1-\delta-\gamma)\int a(u,y)H_{U|V}(du|y) + \gamma \iint \frac{1_{(u<y\leq v)}}{M(y)} a(u,v)H(du,dv) \\
&+ \delta \int a(y,v)H_{V|U}(dv|y).
\end{aligned}$$

Also, $\forall b \in L_2^0(P_{F,H})$ and $a \in L_2^0(H)$,

$$\begin{aligned}
\langle \mathbf{i}_2^* b, a\rangle_{L_2(H)} &= \langle b, \mathbf{i}_2 a\rangle_{L_2(H)} = \int b(y,0,0)\int a(u,y)H_{U|V}(du|y)(1-F(y))H_V(dy) \\
&+ \int b(y,1,0) \iint \frac{1_{(u<y\leq v)}}{M(y)} a(u,v)H(du,dv)M(y)F(dy) \\
&+ \int b(y,1,0)\int a(y,v)H_{V|U}(dv|y)F(y)H_U(dy) \\
&= \iint \left(b(v,0,0)(1-F(v)) + \int_u^v b(y,0,1)F(dy) + b(u,1,0)F(u)\right) a(u,v)H(du,dv),
\end{aligned}$$

and we have

$$(\mathbf{i}_2^* b)(u,v) = b(v,0,0)(1-F(v)) + \int_u^v b(y,0,1)F(dy) + b(u,1,0)F(u), \tag{A.6}$$

and $\mathbf{i}_2^*\mathbf{i}_2$ is given by, $\forall a \in L_2^0(H)$,

$$\begin{aligned}
(\mathbf{i}_2^*\mathbf{i}_2 a)(u,v) &= \mathbf{i}_2^*[(\mathbf{i}_2 a)(y,\delta,\gamma)] = (1-F(v))\int a(r,v)H_{U|V}(dr|v) \\
&+ \int\int\int_u^v \frac{1_{(r<y\leq s)}}{M(y)} F(dy)a(r,s)H(dr,ds) + F(u)\int a(u,s)H_{V|U}(ds|u),
\end{aligned}$$

which is a linear combination of three Fredholm equations of the first kind, each with a non-symmetrical kernel $K(r, s)$ and $K(r, s) \neq K(r - s)$, and consequently, as before, $\mathbf{i}_2^*\mathbf{i}_2$ is not boundedly invertible.

Proof of Theorem 4

(i) We only need to show the conclusion when F is known. For each fixed (s, t), let $\Psi_{s,t}(H) = H(s, t) = \int 1_{(-\infty,s]\times(-\infty,t]}(u, v)H(du, dv)$. As before, the adjoint pathwise differential with respect to H (evaluated at 1) is $\dot{\Psi}^*_{s,t}(u,v)=1_{(-\infty,s]\times(-\infty,t]}(u,v)-H(s,t)$. By the expression of $\mathbf{i}_2^* b$ in (A.6),

$$R(\mathbf{i}_2^*)=\left\{b(v,0,0)(1-F(v))+\int_u^v b(y,0,1)F(dy)+b(u,1,0)F(u):b\in L_2^0(P_{F,H})\right\},$$

i.e., any function in $R(\mathbf{i}_2^*)$ has three-component decomposition, with the first component a function of v alone, the third a function of u alone, and apparently, there is no $b(y,0,1)\in L_2^0(P_{F,H})$ such that $\int_u^v b(y,0,1)F(dy)=\dot{\Psi}^*_{s,t}(u,v)$ for all (u, v). Thus $\dot{\Psi}^*_{s,t}\notin R(\mathbf{i}_2^*)$, so $\Psi_{s,t}$ is not pathwise differentiable with respect to $P_{F,H}$, and hence not rate- $\sqrt{n}$ estimable.

(ii) Assume H unknown. For fixed s, let $\Psi_s = S(s) = \int 1_{(s,\infty)}(z)F(dz)$. Then Ψ_s is pathwise differentiable with respect to F, with adjoint pathwise differential

$$\Psi_s^*(c)(\cdot)=c(1_{(s,\infty)}(\cdot)-1+F(s)).$$

In (A.5), take $b(\cdot, 0, 0) \equiv b(\cdot, 1, 0) \equiv 0$ and $b(y,0,1)=\Psi_s^*(c)(y)/M(y)$. Then $b(\cdot,\cdot,\cdot)\in L_2^0(P_{F,H})$ and $(\mathbf{i}_1^* b)(z)=\Psi_s^*(c)(z)$; i.e., $\Psi_s^*(c)(\cdot)\in R(\mathbf{i}_1^*)$.

Since $Z \perp (U, V)$, we have $\mathbf{i}_1^*\mathbf{i}_2 a=E[E(a(U,V)|(Y,\delta,\gamma)|Z)]=E[a(U,V)|Z]=0, \forall a\in L_2^0(H)$. As in the proof of Theorem 2(iii), we have $R(\mathbf{i}_1) \perp R(\mathbf{i}_2)$ and $\dot{\mathbf{P}}_1\perp \dot{\mathbf{P}}_2$, and, consequently, the information bound is the same for H known or unknown, and $\mathbf{i}_{12}:=\Pi(\mathbf{i}_1|\dot{\mathbf{P}}_2^{\perp})=\mathbf{i}_1$ and the corresponding adjoint $\mathbf{i}_{12}^*=\mathbf{i}_1^*$; similarly, $\mathbf{i}_{21}:=\Pi(\mathbf{i}_2|\dot{\mathbf{P}}_1^{\perp})=\mathbf{i}_2$ and the corresponding adjoint $\mathbf{i}_{21}^*=\mathbf{i}_2^*$. So, by Corollary 5.5.1 in BKRW, the unique efficient influence operator $\tilde{I}: R^* \to \dot{\mathbf{P}}$ for estimating $S(s)$ is the solution of the equations

$$\begin{cases}\dot{\Psi}_s^*=\mathbf{i}_1^*\tilde{I}\\ 0=\mathbf{i}_2^*\tilde{I}.\end{cases} \tag{A.7}$$

To solve $\tilde{I}$ in (A.7), define $g(y, \delta, \gamma) \in L_2(P_{F,H})$ as

$$g(y,\delta,\gamma)=\frac{(1-\delta)\gamma 1_{(s,\infty)}(y)}{M(y)}.$$

Then, using the expression for $p_{F,H}(x)$, we have $E_{P_{F,H}}\, g(Y, \delta, \gamma) = 1 - F(s)$, by (A.5), $(\mathbf{i}_1^* g)(z)=E_{P_{F,H}}(g(Y,\delta,\gamma)|Z=z)=1_{(s,\infty)}(z)$; and, by (A.6), $(\mathbf{i}_2^* g)(u,v)=E_{P_{F,H}}[g(Y,\delta,\gamma)|(u,v))=\int_u^v(1_{(s,\infty)}(y)/M(y))F(dy)$. Since F and H are unknown and unrestricted, $\dot{\mathbf{P}}=L_2^0(P_{F,H})$. Now, let

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\tilde{I}(c)=c(g(Y,\delta,\gamma)-(\mathbf{i}_2^*g)(U,V))=c\left(\frac{1_{(s,\infty)}(Y)(1-\delta)\gamma}{M(Y)}-\int_U^V\frac{1_{(s,\infty)}(y)}{M(y)}F(dy)\right)$$
$$=c\left(\frac{1_{(s,\infty)}(Y)(1-\delta)\gamma}{M(Y)}-(1-\gamma-2\delta+2\delta\gamma)\int_{-\infty}^Y\frac{1_{(s,\infty)}(z)}{M(z)}F(dz)\right).$$

Then, $\tilde{I}(c)\in\dot{\mathbf{P}}=L_2^0(P_{F,H})$, $\forall c\in R^*=R$. The last equality above holds as $V=Y\,1_{(\delta,\gamma)=(0,0)}$ and $U=Y1_{(\delta=1)}$, so $\int_U^V=\int_{-\infty}^V-\int_{-\infty}^U=\left(\int_{-\infty}^Y-(\delta(1-\gamma)+\gamma(1-\delta))\int_{-\infty}^Y\right)-\left(\int_{-\infty}^Y-(1-\delta)\int_{-\infty}^Y\right)=(1-\gamma-2\delta+2\delta\gamma)\int_{-\infty}^Y$.

Since, $\forall c\in R^*$, we have

$$\mathbf{i}_1^*\tilde{I}(c)(\cdot)=c(\mathbf{i}_1^*g)(\cdot)-cE(E[g(Y,\delta,\gamma)|(U,V)]|Z)=c(1_{(s,\infty)}(\cdot)-1+F(s))=\dot{\Psi}_s^*(c)(\cdot),$$
$$\text{and }\mathbf{i}_2^*\tilde{I}(c)=c(E[g(Y,\delta,\gamma)|(U,V)]-E[g(Y,\delta,\gamma)|(U,V)])=0,$$

thus $\tilde{I}$ is the solution of (A.7), and the efficient influence function for estimating $S(t)$ is $\tilde{S}^*(t)=\tilde{I}(1)$. Note that $P((\delta,\gamma)=(0,1))=P(U<Z\le V)=\int[F(v)-F(u)]H(du,dv)$. The information bound for $S(t)$ is

$$I_S(t)=E_{P_{F,H}}(\dot{S}^*(t))^2=\int_t^\infty\frac{1}{M(z)}F(dz)+\int R_S^2(t\vee u,v)H(du,dv),$$

where $R_S(a,b)=\int_a^bM^{-1}(z)F(dz)$. As in the proof of Theorem 2(iii), when Z_i: $i=1,\ldots,n\}$ is fully observable, $I_S(t)$ reduces to the usual form.

Finally, since the Z_i are partially observed, any pathwise differentiable functional of F is rate-$\sqrt{n}$ estimable.

(iii) Denote $\Psi_t=\Psi_F(t)=\Lambda(t)=\int\frac{1_{(-\infty,t)}(s)F(ds)}{1-F(s-)}$. Then Ψ_t is pathwise differentiable with respect to F, with adjoint pathwise differential $\dot{\Psi}_t^*(c)=c(\frac{1_{(-\infty,t)}(z)}{1-F(z-)}-\Lambda(t))$. As before, the efficient influence operator $\tilde{I}$ for $\Lambda(t)$ is the solution of the equations

$$\begin{cases}\dot{\Psi}_t^*=\mathbf{i}_1^*\tilde{I}\\ 0=\mathbf{i}_2^*\tilde{I}.\end{cases}\tag{A.8}$$

For this, take

$$\tilde{I}(c)=c\left(\frac{1_{(-\infty,t)}(Y)(1-\delta)\gamma}{M(Y)(1-F(Y-))}-\mathbf{i}_2^*\left[\frac{1_{(-\infty,t)}(Y)(1-\delta)\gamma}{M(Y)(1-F(Y-))}\right]\right).$$

It is easy to see that $\tilde{I}:R^* \to L_2^0(P_{F,H})$ is the solution of (A.8). Also, by (A.6), which holds also for $b \in L_2(P_{F,H})$, $\mathbf{i}_2^*[\frac{1_{(-\infty,t)}(Y)(1-\delta)\gamma}{M(Y)(1-F(Y-))}]=E[\frac{1_{(-\infty,t)}(Y)(1-\delta)\gamma}{M(Y)(1-F(Y-))}|(U,V)]=\int_U^V \frac{1_{(-\infty,t)}(z)F(dz)}{M(z)(1-F(z-))}$. Thus the efficient influence function $\dot{\Lambda}^*(t) = \tilde{I}(1)$ for $\Lambda(t)$ is

$$\begin{aligned}\dot{\Lambda}^*(t)&=\frac{1_{(-\infty,t)}(Y)(1-\delta)\gamma}{M(Y)(1-F(Y-))} - \int_U^V \frac{1_{(-\infty,t)}(z)F(dz)}{M(z)(1-F(z-))}\\ &=\frac{1_{(-\infty,t)}(Y)(1-\delta)\gamma}{M(Y)(1-F(Y-))} - (1-\gamma-2\delta+2\delta\gamma)\int_{-\infty}^Y \frac{1_{(-\infty,t)}(z)F(dz)}{M(z)(1-F(z-))},\end{aligned}$$

and the information bound for $\Lambda(t)$ is

$$I_\Lambda(t)=E_{P_{F,H}}(\dot{\Lambda}^*(t))^2=\int \frac{1_{(-\infty,t)}(z)F(dz)}{M(z)(1-F(z-))^2}+\int R_\Lambda^2(t\vee u,v)H(du,dv).$$

(iv) Since Z is observable with non-zero probability in this model, the root-n estimability of θ follows from Kac's result. The conditional density of $Z|W$ is

$$f(z|w)=\exp(\theta' w)\overline{F}(z)^{\exp(\theta' w)}\frac{f(z)}{\overline{F}(z)},$$

and the conditional density/mass function for $X|W = (Y, \delta, \gamma)|W$ is

$$p_{F,H}(x|w)=\left(M(y)\exp(\theta' w)\overline{F}(y)^{\exp(\theta' w)}\frac{f(z)}{\overline{F}(z)}\right)^\gamma\left((1-\overline{F}(y)^{\exp(\theta' w)})h_U(y)\right)^\delta\left(\overline{F}(y)^{\exp(\theta' w)}h_V(y)\right)^{1-\gamma-\delta},$$

with log-likelihood, up to a function of (y, δ, γ),

$$l(\theta,f)=\gamma[\theta' w+\exp(\theta' w)\log\overline{F}(y)]+\delta\log(1-\overline{F}(y)^{\exp(\theta' w)})+(1-\gamma-\delta)\exp(\theta' w)\log\overline{F}(y).$$

Similarly as in the proof of Theorem 2(iv), recalling that $\log \bar{F}(y) = -\Lambda(y)$, we have

$$l_\theta(\theta,f)=\left(\gamma-\exp(\theta' w)\Lambda(y)+\delta\frac{1}{1-\overline{F}(y|w)}\exp(\theta' w)\Lambda(y)\right)w,$$

and, $\forall a \in L_2^0(F)$,

$$l_f(\theta,f)(a)=\exp(\theta' w)\left(1-\delta\frac{1}{1-\overline{F}(y|w)}\right)\frac{\int_y^\infty a dF}{\overline{F}(y)}.$$

Let $a^* \in (L_2^0(F))^k (k=\dim(w))$ satisfy

$$l_\theta(\theta, f) - l_f(\theta, f)(a^*) \perp l_f(\theta, f)(a), \quad \forall a \in L_2^0(F).$$

As before,

$$0=E_P\{[l_\theta(\theta, f)-l_f(\theta, f)(a^*)]l_f(\theta, f)(a)\}=E_Y(E\{[l_\theta(\theta, f)-l_f(\theta, f)(a^*)]l_f(\theta, f)(a)|Y\}),$$

and set $E\{[l_\theta(\theta, f) - l_f(\theta, f)(a^*)]l_f(\theta, f)(a)|Y\} = 0$ for all $a \in L_2^0(F)$ and Y. After some computation, we get

$$A(y):=\int_y^\infty a^* dF=-\frac{M(y)f(y)a_1(y)-\overline{F}(y)\Lambda(y)h_V(y)a_2(y)-\overline{F}(y)\Lambda(y)h_U(y)a_3(y)}{M(y)f(y)b_1(y)-\overline{F}(y)h_V(y)b_2(y)-\overline{F}(y)h_U(y)b_3(y)},$$

where

$$\begin{gathered} a_1(y)=E\left(\exp(2\theta' W)\overline{F}(Y|W)(1-\exp(\theta' W)\Lambda(y))W|Y=y\right), \\ a_2(y)=E\left(\exp(2\theta' W)\overline{F}(Y|W)W|Y=y\right), \quad a_3(y)=E\left(\exp(2\theta' W)\tfrac{\overline{F}^2(Y|W)W}{1-\overline{F}(Y|W)}|Y=y\right), \\ b_1(y)=E\left(\exp(3\theta' W)\overline{F}(Y|W)|Y=y\right), \quad b_2(y)=E\left(\exp(2\theta' W)\overline{F}(Y|W)|Y=y\right), \end{gathered}$$

and

$$b_3(y)=E\left(\exp(2\theta' W)\frac{\overline{F}^2(Y|W)}{1-\overline{F}(Y|W)}|Y=y\right).$$

As before, $l_{\theta|f} = l_\theta(\theta,f) - l_f(\theta,f)(a^*)$, $l_f(\theta, f)(a^*)=\exp(\theta' w)\left(1-\delta\frac{1}{1-\overline{F}(y|w)}\right)\frac{A(y)}{1-F(y)}$, and the asymptotic information for estimating θ in the presence of nuisance f is $i(\theta|f)=\left\|l_{\theta|f}\right\|_P^2$.

Proof of Theorem 5

(i) We have, $\forall a \in L_2(Q)$,

$$(\mathbf{i}a)(z, u, v)=E(a(Z, U, V)|U=u<Z=z<V=v)=a(z, u, v)1_{(u<z<v)}.$$

By the equality $\langle \mathbf{i}a, b\rangle_{L_2(P_{F,G})} = \langle a, \mathbf{i}^* b\rangle_{L_2(Q)}$, $\forall a \in L_2(Q)$ and $b \in L_2(P_{F,G})$, we get

$$(\mathbf{i}^* b)(z,u,v) = \frac{1}{\alpha} b(z,u,v) 1_{(u<z<v)}, \quad \forall b \in L_2(P_{F,G})$$

and

$$(\mathbf{i}^* \mathbf{i} a)(z,u,v) = \frac{1}{\alpha} a(z,u,v) 1_{(u<z<v)}, \quad \forall a \in L_2(Q).$$

Since $0 < \alpha < 1$, $\mathbf{i}^*\mathbf{i}$ is boundedly invertible on $\{(z, u, v): u < z < v\}$. Since $\alpha > 0$, the full data (Z_i, U_i, V_i) are partially observable. Hence, by Kac's result, any smooth functionals of $P_{F,G}$ are $\sqrt{n}$-estimable.

(ii) Let $F^*(\cdot)$ be the marginal distribution of Z under the joint model $P_{F,G}$, and let $\bar{G}_2(z) = 1 - G_2(z)$. Then

$$F^*(z) = P(Z \le z) = \alpha^{-1} \int_0^z G_1(x) \bar{G}_2(x) F(dx);$$

thus $dF^*(z) = \alpha^{-1} G_1(z) \bar{G}_2(z) F(dz)$. Let $M(z) = \alpha^{-1} G_1(z) \bar{G}_2(z)(1 - F_-(z)) = \alpha^{-1} P(U < z < V)$; then

$$\Lambda(t) := \Psi(P_{F,G}) = \int \frac{1_{(0 \le z \le t)}}{M(z)} dF^*(z) = \iiint \frac{1_{(0 \le z \le t)}}{M(z)} P^*(du, dz, dv) := \Psi_e(P^*),$$

where P^* is the distribution of (U, Z, V) under the sample space $\{(u, z, v): u < z < v\}$ and with margin F^*, so $\Psi_e(P^*)$ defines an extension of $\Lambda(t)$ from $\mathbf{P} = \{P_{F,G}: F \in \mathcal{F}, G \in \mathcal{G}\}$ to $\mathbf{M} = \{P^*: \int\int P^*(du, \cdot, dv) = F^*(\cdot)\}$. Obviously, $\Psi_e(P^*)$ is pathwise differentiable with respect to P^*, with adjoint pathwise derivative at 1

$$\begin{aligned} \dot{\Psi}_e^*(1)(u,z,v) &= \frac{1_{(0 \le z \le t)}}{M(z)} - \int_0^t \frac{1_{(0<s<z\wedge v)}}{M^2(s)} F^*(ds) \\ &= \int_0^{t\wedge u} \frac{1}{M^2(s)} F^*(ds) + \frac{1_{(0 \le z \le t)}}{M(z)} - \int_0^{t\wedge z \wedge v} \frac{1}{M^2(s)} F^*(ds). \end{aligned}$$

To see this, for fixed (F, G), denote P_0 the corresponding distribution in $\mathbf{P} \subset \mathbf{M}$. We first write $\Psi_e(P_{F,G})$ in terms of E_{P_0}. Let $S \sim F$. Then

$$\begin{aligned} \Psi_e(P_0) &= E_{P_0}\left(\frac{1_{(0 \le S \le t)}}{M(S)}\right) = E_{P_0}\left[E_F\left(\frac{1_{(0 \le S \le t)}}{M(S)} | U < S < Z \wedge V\right)\right] \\ &= E_{P_0}\left[\alpha^{-1} \int_0^t \frac{1_{(U<s<Z\wedge V)}}{M^2(s)} F(ds)\right] = E_{P_0}\left[\int_0^t \frac{1_{(U<s<Z\wedge V)}}{M^2(s)} F^*(ds)\right]. \end{aligned}$$

Now, let $\dot{\Psi}_e(P_0)(h) = \langle \dot{\Psi}_e^*(1), h \rangle_{P_0}$ for $h \in \dot{\mathbf{M}}$, and let $P_\eta \in \mathbf{M}$ be any path of distributions passing through P_0 as $\eta \to 0$. Then, by the same way as in the proof of Proposition A.5.2 in BKRW, we have

$$\Psi_e(P_\eta) - \Psi_e(P_0)=\eta\dot{\Psi}_e(P_0)(h)+o(\eta), \quad \forall h \in \mathbf{M};$$

i.e., $\dot{\Psi}_e(P_0)(h)$ is the pathwise differential of Ψ_e at P_0 in $\mathbf{M}$, and $\dot{\Psi}_e^*(1)(u, z, v)$ is its adjoint at 1.

The adjoint pathwise derivative at 1 of $\Psi(P_{F,G})$ at P_0 in $\mathbf{P}$ is the projection

$$\dot{\Psi}^*(1)=\Pi(\dot{\Psi}_e^*(1)|\,\dot{\mathbf{P}}).$$

Denote the score operators $\mathbf{i}_1:\dot{G}_1 \to L_2^0(P_{F,G})$, $\mathbf{i}_2: \dot{F} \times\dot{G}_2 \to L_2^0(P_{F,G})$ and $\mathbf{i}: \dot{F} \times \dot{G}\to L_2^0(P_{F,G})$. Then

$$\mathbf{i}(a,b)=\mathbf{i}_1a+\mathbf{i}_2b, \quad a(\cdot) \in \dot{G}_1=L_2^0(G_1), \quad b(\cdot,\cdot) \in\dot{F} \times\dot{G}_2=L_2^0(F \times G_2),$$

where

$$(\mathbf{i}_1a)(u,z,v)=a(u) - E_{P_{F,G}}a(u)=a(u) - \int adG_1^*, \quad \text{and}$$
$$(\mathbf{i}_2b)(u,z,v)=b(z,v) - E_{P_{F,G}}b(Z,V)=b(z,v) - \int bdP_{F,G_2}^*,$$

with G_1^* and P_{F,G_2}^* the corresponding margins under $P_{F,G}$. By the same way as on p. 243, BKRW, we have

$$\dot{\mathbf{P}}\supset L_2^0(G_1^*)+L_2^0(P_{F,G_2}^*).$$

Let $R=E_{G_1^*}[\int_0^{t\wedge U}M^{-2}(s)F^*(ds)]$. Since $E_{P^*}(\dot{\Psi}^*(1)(U, Z, V)) = 0$, we get

$$\dot{\Psi}_e^*(1)(u,z,v)=\left(\int_0^{t\wedge u}\frac{1}{M^2(s)}F^*(ds) - R\right)+\left(\frac{1_{(0\le z\le t)}}{M(z)} - \int_0^{t\wedge z\wedge v}\frac{1}{M^2(s)}F^*(ds)+R\right)$$
$$:=a(u)+b(z,v);$$

then $a \in L_2^0(G_1^*), b \in L_2^0(P_{F,G_2}^*)$. Thus $\dot{\Psi}_e^*(1) \in\dot{\mathbf{P}}$, so

$$\dot{\Psi}^*(1)=\Pi(\dot{\Psi}_e^*(1)|\,\dot{\mathbf{P}})=\dot{\Psi}_e^*(1),$$

and the efficient influence function for estimating $\Lambda(t)$ is

$$\tilde{I}_\Lambda(t)=\dot{\Psi}^*(1)=\frac{1_{(0\le z\le t)}}{M(z)}-\alpha\int_0^t \frac{1_{(u<s<s\wedge v)}}{G_1(s)\overline{G}_2(s)(1-F_-(s))^2}F(ds).$$

(iii) The root-n estimability of θ follows the same reasoning as that in the proof of Theorem 4(iv). Let $q(w)$ be the density of W. The density of (u, z, v, w) is

$$\begin{aligned}p(u,z,v,w)=&p_{F,H}(x|w)q(w)=\alpha^{-1}(\theta-w)f(z|w)g(u,v)1_{(u<z<v)}q(w)\\ =&\alpha^{-1}(\theta,w)\exp(\theta' w)\overline{F}(z)^{\exp(\theta' w)}\tfrac{f(z)}{\overline{F}(z)}g(u,v)1_{(u<z<v)}q(w),\end{aligned}$$

where $\alpha(\theta, w) = P(U < Z < V|w) = \int \bar{F}(s|w)G_1 - G_2)(ds)$, with $\bar{F}(u|w) = \bar{F}(u)^{\exp(\theta' w)}$ and G_1 and G_2 the marginal distribution functions of U and V, respectively. The log-likelihood, up to a function of (z, u, v, w), is

$$l(\theta,f)=\theta' w+\exp(\theta' w)\log\overline{F}(z)-\log\left(\int\overline{F}(s)^{\exp(\theta' w)}(G_1-G_2)(ds)\right).$$

We have

$$\begin{aligned}l_\theta(\theta,f)=&\left(1-\exp(\theta' w)\Lambda(z)\tfrac{\exp(\theta' w)}{\alpha(\theta,w)}\int\Lambda(s)\overline{F}(s|w)(G_1-G_2)(ds)\right)w,\\ l_f(\theta,f)(a)=&\exp(\theta' w)\left(\frac{\int_z^\infty adF}{\overline{F}(z)}-\alpha^{-1}(\theta,w)\int\overline{F}(s|w)\frac{\int_s^\infty adF}{\overline{F}(s)}(G_1-G_2)(ds)\right).\end{aligned}$$

Since $l_f(\theta,f)(\cdot):L_2^0(F)\to L_2^0(P)$, we have $E_P(l_f(\theta,f)(a))=0$, for all $a\in L_2^0(F)$. This gives

$$\alpha^{-1}(\theta,w)\int\overline{F}(s|w)\frac{\int_s^\infty adF}{\overline{F}(s)}(G_1-G_2)(ds)=E\left(\frac{\int_Z^\infty adF}{\overline{F}(Z)}\middle|W=w\right),$$

so

$$l_f(\theta,f)(a)=\exp(\theta' w)\left(\frac{\int_z^\infty adF}{\overline{F}(z)}-E\left(\frac{\int_Z^\infty adF}{\overline{F}(Z)}\middle|W=w\right)\right).$$

Since $0 = E_P(l_f(\theta, f)(a)) = E[E(l_f(\theta, f)(a)|W)] = E[E(l_f(\theta, f)(a)|Z)]$, and noting that the covariate W is independent of Z, we have

$$E\left(\exp(\theta' w)E\left(\frac{\int_Z^\infty adF}{\overline{F}(Z)}\middle|W\right)\right)=E(\exp(\theta' W)|Z)E\left(\frac{\int_Z^\infty adF}{\overline{F}(Z)}\right)=E(\exp(\theta' W))E\left(\frac{\int_Z^\infty adF}{\overline{F}(Z)}\right),$$

which is possible only if $E(\frac{\int_Z^\infty adF}{\overline{F}(Z)}|W)=E\left(\frac{\int_Z^\infty adF}{\overline{F}(Z)}\right)$. Now, we have

$$l_f(\theta, f)(a)=\exp(\theta' w)\left(\frac{\int_z^\infty adF}{\overline{F}(z)} - E\left(\frac{\int_Z^\infty adF}{\overline{F}(Z)}\right)\right).$$

As before, we have to find $a^* \in (L_2^0(F))^k$ such that, $\forall a \in L_2^0(F)$,

$$\begin{aligned}0=&E_P[(l_\theta(\theta, f) - l_f(\theta, f)(a^*))l_f(\theta, f)(a)]\\ =&E\left(E\left[(l_\theta(\theta, f) - l_f(\theta, f)(a^*))\exp(\theta' W)|Z\right]\left(\frac{\int_Z^\infty adF}{\overline{F}(Z)} - E\left(\frac{\int_Z^\infty adF}{\overline{F}(Z)}\right)\right)\right),\end{aligned}$$

which gives

$$0=E\left[(l_\theta(\theta, f) - l_f(\theta, f)(a^*))\exp(\theta', W)|Z\right],$$

and after some computation we get

$$\frac{\int_z^\infty a^* dF}{\overline{F}(z)}=\frac{b(\theta, z)}{E(\exp(\theta' W)|Z=z)},$$

where

$$b(\theta, z)=\Lambda(z)E(\exp(2\theta' W)W|Z=z)-E(\exp(\theta' W)W|Z=z)-\int E\left(\frac{\exp(2\theta' W)\overline{F}(s|W)W}{\alpha(\theta,W)}\Big|Z=z\right)\Lambda(s)(G_1-G_2)(ds).$$

This gives the expression for $l_{\theta|f} = l_\theta(\theta, f) - l_f(\theta, f)(a^*)$.

Proof of Theorem 6

(i) As before, we first compute the conditional density $f(y|\delta, Z, G)$. We have

$$F(y|\delta=0, Z, G)=P(Y \le y|Y>Y_0;Z, G)=\frac{F(y) - F(Y_0)}{1 - F(Y_0)}, \quad F(y|\delta=1, Z, G)=\frac{F(y \wedge Y_0)}{F(Y_0)},$$

which give

$$f(y|\delta, Z, G)=(1-\delta)\frac{f(y)}{1-F(Y_0)}1_{(Y_0<y)}+\delta\frac{f(y)}{F(Y_0)}1_{(y\leq Y_0)}.$$

We get, $\forall a \in L_2(F)$,

$$(\mathbf{i}a)(\delta)=(1-\delta)\int_{Y_0}^{\infty}\frac{a(y)}{1-F(Y_0)}F(dy)+\delta\int_{-\infty}^{Y_0}\frac{a(y)}{F(Y_0)}F(dy),$$

and the relationship $\langle \mathbf{i}a, b\rangle_N = \langle a, \mathbf{i}^*b\rangle$, $\forall a \in L_2(F)$ and $b \in L_2(P_f)$ gives

$$\begin{aligned}(\mathbf{i}a)(0)b(0)p_f(0)+(\mathbf{i}a)(1)b(1)p_f(1)&=b(0)p_f(0)\int_{Y_0}^{\infty}\tfrac{a(y)}{1-F(Y_0)}F(dy)+b(1)p_f(1)\int_{-\infty}^{Y_0}\tfrac{a(y)}{F(Y_0)}F(dy)\\&=\int a(y)(\mathbf{i}^*b)(y)F(dy),\end{aligned}$$

so we have

$$(\mathbf{i}^*b)(y)=\frac{b(0)p_f(0)}{1-F(Y_0)}1_{(y>Y_0)}+\frac{b(1)p_f(1)}{F(Y_0)}1_{(y\leq Y_0)},$$

and

$$\begin{aligned}(\mathbf{i}^*\mathbf{i}a)(y)&=\int\left(\tfrac{F(Y_0)1_{(t>Y_0)}}{(1-F(Y_0))^2}1_{(y>Y_0)}+\tfrac{(1-F(Y_0))1_{(t\leq Y_0)}}{F^2(Y_0)}1_{(y\leq Y_0)}\right)a(t)F(dt)\\&:=\int K(y,t)a(t)F(dt).\end{aligned}$$

As before, the above Fredholm equation of the first kind is not generally solvable; hence $\mathbf{i}^*\mathbf{i}$ is not boundedly invertible.

(ii) Here, β is not a known functional form of f, so the method of pathwise differential does not apply directly. Instead, we use the method of Hellinger differential as in [1]. We first compute $i(\beta|\alpha, f)$. Denote $g(\cdot|Z, U; \alpha, \beta, f) = [f^{1/2}(\cdot|Z, U; \alpha, \beta)]^2$. Recall that for densities g and f with respect to dominating measures λ and ν, with $g_n = g(\cdot|Z, U; \alpha_n, \beta_n, f_n)$, $\alpha_n = \alpha + n^{-1/2}a + o(n^{-1/2})$, $\beta_n = \beta + n^{-1/2}b + o(n^{-1/2})$ and $\left\|n^{1/2}(f_n^{1/2}-f^{1/2})-h\right\|_\nu \to 0$, g is Hellinger differentiable at (α, β, f) if there exist ρ_α with $\|\rho_\alpha\|_\lambda \in L_2(\lambda)$, $\rho_\beta \in L_2(\lambda)$ and $A: L_2(\nu) \to L_2(\lambda)$ such that

$$\frac{\left\|g_n^{1/2}-g^{1/2}-(\rho_\alpha(\alpha_n-\alpha)+\rho_\beta(\beta_n-\beta)+A(f_n^{1/2}-f^{1/2}))\right\|_\lambda}{\left\|\alpha_n-\alpha\right\|+|\beta_n-\beta|+\left\|f_n{}^{1/2}-f^{1/2}\right\|_\nu}\to 0,$$

and then its Hellinger differential at (α, β, f) in the direction (a, b, h) is

$$\eta = a\rho_\alpha + b\rho_\beta + A(h).$$

Specifically, $\rho_\alpha = (1/2)g^{-1/2}\partial g/\partial\alpha = (1/2)g^{1/2}\partial \log g/\partial\alpha$, and similarly for ρ_β and $A(h) = (1/2)g^{-1/2}\,\Delta(g, f^{1/2}, h) = (1/2)g^{1/2}\,\Delta\,(\log g, f^{1/2}, h)$, where $\Delta\,(g, f^{1/2}, h)$ is the Gateaux differential of g with respect to $f^{1/2}$ in the direction h.

As in Remark 3.2 in Begun et al. [1], the asymptotic information for estimating β in the presence of nuisance (α, f) is

$$i(\beta|\alpha, f) = 4\left\|\rho_{\beta|\alpha} - A(h^*)\right\|_\lambda,$$

where

$\rho_{\beta|\alpha} = \rho_\beta - \langle \rho_\beta, \rho'_\alpha\rangle_\lambda \left\|\rho_\alpha\right\|_\lambda^{-2}\rho_\alpha, \left\|\rho\alpha\right\|_\lambda^2 = \langle \rho_\alpha, \rho'_\alpha\rangle_\lambda, h^* \in \mathscr{H} := \{h \in L_2(\nu): \left\|n^{1/2}(f_n^{1/2} - f^{1/2}) - h\right\|_\nu \to 0\}$
, such that

$$\rho_{\beta|\alpha} - A(h^*) \perp A(h), \quad \forall h \in \mathscr{H}.$$

Let A^* be the adjoint of A. When A^*A is invertible, $h^* = ((A^*A)^{-1}A^*)(\rho_{\beta|\alpha})$.

Note that the condition $\left\|n^{1/2}(f_n^{1/2} - f^{1/2}) - h\right\|_\nu \to 0$ implies that h is of the form $h(\cdot - \mu)$ when f takes the form $f(\cdot - \mu)$. This gives $A(h)(\cdot) = (1/2)g^{-1/2}\,\Delta(g, f^{1/2}, h) = h(\cdot - \mu)$. In our case, $\lambda = \nu =$, the Lebesque measure, and A^* is determined by $\langle Ah, r\rangle_\lambda = \langle h, A^*r\rangle_\nu$ or $\int h(x - \mu)r(x)dx = \int h(x)(A^*r)(x)dx$, $\forall h \in L_2(\lambda)$, $r \in L_2(\nu)$. Thus $(A^*r)(x) = r(x + \mu)$, and A^*A is the identity. Consequently, $h^* = A^*(\rho_{\beta|\alpha})$, and

$$i(\beta|\alpha, f) = 4\left\|\rho_{\beta|\alpha} - \rho_{\beta|\alpha,\mu=0}\right\|_\lambda^2.$$

Now, we compute $i(\beta|\alpha, p_f)$. With $f'(y) = df(y)/dy$, we have

$$l_\beta(\delta) := \frac{\partial \log p_f(\delta)}{\partial\beta} = -\frac{U(1-\delta-F(Y_0))}{F(Y_0)(1-F(Y_0))}\int^{Y_0} f'(y-\mu)dy \quad \text{and}$$
$$l_\alpha(\delta) := \frac{\partial \log p_f(\delta)}{\partial\alpha} = -\frac{Z(1-\delta-F(Y_0))}{F(Y_0)(1-F(Y_0))}\int^{Y_0} f'(y-\mu)dy.$$

Then, with N being the counting measure on $\{0, 1\}$,

$$\tilde{\rho}_\beta=(1/2)p_f^{1/2}(\delta)l_\beta(\delta), \quad \tilde{\rho}_\alpha=(1/2)p_f^{1/2}(\delta)l_\alpha(\delta), \quad \tilde{\rho}_{\beta|\alpha}=\tilde{\rho}_\beta-\langle\tilde{\rho}_\beta,\tilde{\rho}_\alpha'\rangle_N \left\|\tilde{\rho}_\alpha\right\|_N^{-2}\tilde{\rho}_\alpha,$$
$$\tilde{A}(h)(\delta)=(1-\delta-F(Y_0))p_f^{1/2}(\delta)\frac{\int 1_{(y\le Y_0-\mu)}f^{1/2}(y)h(y)dy}{p_f(0)p_f(1)}, \quad \forall h\in L_2(F),$$
$$\tilde{A}^*(r)(x)=\frac{p_f^{1/2}(1)r(0)-p_f^{1/2}(0)r(1)}{[p_f(0)p_f(1)]^{1/2}}f^{1/2}(x)1_{(x\le Y_0-\mu)}, \quad \forall r\in L_2(P_f),$$
$$(\tilde{A}^*\tilde{A})(h)(x)=f^{1/2}(x)1_{(x\le Y_0-\mu)}\frac{\int 1_{(y\le Y_0-\mu)}f^{1/2}(y)h(y)dy}{p_f(0)p_f(1)},$$

which is not invertible. So

$$i(\beta|\alpha,p_f)=4\left\|\tilde{\rho}_{\beta|\alpha}-\tilde{A}(h^*)\right\|_N^2,$$

with $h^*\in\mathscr{H}$ determined by

$$\tilde{\rho}_{\beta|\alpha}-\tilde{A}(h^*)\perp\tilde{A}(h), \quad \forall h\in\mathscr{H}, \quad \text{or} \quad 0=<\tilde{\rho}_{\beta|\alpha}-\tilde{A}(h^*),\tilde{A}(h)>_N, \quad \forall h\in\mathscr{H}\subset L_2(F).$$

The last condition above gives

$$\int 1_{(y\le Y_0-\mu)}f^{1/2}(y-\mu)h^*(y)dy=p_f^{1/2}(0)p_f(1)\tilde{\rho}_{\beta|\alpha}(0)-p_f(0)p^{1/2}(1)\tilde{\rho}_{\beta|\alpha}(1),$$

which yields

$$h^*(x-\mu)=\frac{p_f^{1/2}(0)p_f(1)\tilde{\rho}_{\beta|\alpha}(0)-p_f(0)p^{1/2}(1)\tilde{\rho}_{\beta|\alpha}(1)}{F(Y_0)}f^{1/2}(x-\mu),$$

and so

$$\tilde{A}(h^*)(\delta)=\frac{p^{1/2}(1)\tilde{\rho}_{\beta|\alpha}(1)-p^{1/2}(0)\tilde{\rho}_{\beta|\alpha}(0)}{p_f^{1/2}(0)p_f^{1/2}(1)}(1-\delta-F(Y_0))p_f^{1/2}(\delta).$$

(iii) In this case,

$$\rho_\beta=\frac{U\exp(-(y-\tilde{\mu})/2)(1-\exp(-(y-\tilde{\mu})))}{2(1+\exp(-(y-\tilde{\mu})))^2}, \quad \rho_\alpha=\frac{Z\exp(-(y-\tilde{\mu})/2)(1-\exp(-(y-\tilde{\mu})))}{2(1+\exp(-(y-\tilde{\mu})))^2},$$
$$\langle\rho_\beta,\rho_\alpha'\rangle=\frac{UZ'}{4}\int_0^\infty\frac{(1-t)^2}{(1+t)^4}dt, \quad \left\|\rho_\alpha\right\|^2=\frac{ZZ'}{4}\int_0^\infty\frac{(1-t)^2}{(1+t)^4}dt,$$
$$\text{and} \quad i(\beta|\alpha,f_0)=4\left\|\rho_{\beta|\alpha}\right\|^2=U^2(I-Z'(ZZ')^{-1}Z)^2\int_0^\infty\frac{(1-t)^2}{(1+t)^4}dt=U^2(I-Z'(ZZ')^{-1}Z)^2/3.$$

Likewise, with N being the counting measure on $\{0, 1\}$,

$$\tilde{\rho}_\beta=\frac{U(\exp(\mu))^{\delta/2}(\delta+(\delta-1)\exp(\mu))}{2(1+\exp(\mu))^{3/2}}, \quad \tilde{\rho}_\alpha=\frac{Z(\exp(\mu))^{\delta/2}(\delta+(\delta-1)\exp(\mu))}{2(1+\exp(\mu))^{3/2}},$$
$$\langle\tilde{\rho}_\beta,\tilde{\rho}_\alpha'\rangle_N=\frac{UZ'}{4}\frac{\exp(\mu)}{(1+\exp(\mu))^2}, \quad \left\|\tilde{\rho}_\alpha\right\|_N^2=\frac{ZZ'}{4}\frac{\exp(\mu)}{(1+\exp(\mu))^2},$$
$$\text{and} \quad i(\beta|\alpha,p_{f_0})=4\left\|\tilde{\rho}_{\beta|\alpha}\right\|_N^2=U^2(I-Z'(ZZ')^{-1}Z)^2\frac{\exp(\mu)}{(1+\exp(\mu))^2}.$$

The corresponding information bounds under f_0 and p_{f0} are $i^{-1}(\beta|\alpha, f_0)$ and $i^{-1}(\beta|\alpha, p_{f0})$, and the asymptotic relative efficiency is

$$r(\mu)=\frac{i^{-1}(\beta|\alpha,f_0)}{i^{-1}(\beta|\alpha,p_{f_0})}=3q(\mu), \quad \text{with } q(\mu)=\frac{\exp(\mu)}{(1+\exp(\mu))^2}.$$

Note that $q(\cdot)$ is increasing on $(-\infty, 0]$ and decreasing on $(0, \infty)$, with $q(0) = 1/4$, so $\sup_\mu r(\mu) = 3q(0) = 3/4$.