Interval Censored Survival Data: A Review of Recent Progress

Jian Huang
Jon A. Wellner

ABSTRACT We review estimation in interval censoring models, including nonparametric estimation of a distribution function and estimation of regression models. In the nonparametric setting, we describe computational procedures and asymptotic properties of the nonparametric maximum likelihood estimators. In the regression setting, we focus on the proportional hazards, the proportional odds and the accelerated failure time semiparametric regression models. Particular emphasis is given to calculation of the Fisher information for the regression parameters. We also discuss computation of the regression parameter estimators via profile likelihood or maximization of the semiparametric likelihood, distributional results for the maximum likelihood estimators, and estimation of (asymptotic) variances. Some further problems and open questions are also reviewed.

1. Introduction: interval censoring models.

Interval censored data arises when a failure time T can not be observed, but can only be determined to lie in an interval obtained from a sequence of examination times. Kongerud and Samuelsen (1991) and Samuelsen and Kongerud (1993) report two studies on respiratory symptoms and asthmatic symptoms among Norwegian aluminum workers. In these studies, the time to the development of respiratory symptoms or asthmatic symptoms is only known to be between two health examinations. Other examples of interval censored data in animal carcinogenicity and epidemiology studies can be found in Hoel and Walburg (1972), Finkelstein and Wolfe (1985), Finkelstein (1986), and Self and Grossman (1986). Diamond, McDonald and Shah (1986), and Diamond and McDonald (1991) contain examples of interval censored data from demography studies. Closely related censoring schemes also arise in AIDS studies, see for example, De Gruttola and Lagakos (1989), Shiboski and Jewell (1992), Jewell, Malani, and Vittinghoff (1994), etc.

We now briefly describe the three types of interval-censored data considered in this review. Let T (or T_i) be the unobservable failure time.

1.1. "Case 1" interval censoring or current status data.

Suppose that U is an "examination" or "observation" time. Then suppose that an observation consists of the random vector (δ, U) where $\delta = 1_{[T \le U]}$, or (δ, U, Z) if a vector Z of covariates is also available. The only knowledge about the "failure time" T is whether it has occurred before U or not. Such data is substantially different from right-censored data. In a right censorship model, the observed data is $(\min(T, Y), 1_{\{T \le Y\}}, Z)$ where Y is a "censoring time" and the probability $P\{T \le Y\}$ of observing the survival time T exactly is positive. But with current status data, we are not able to observe the exact value of the survival time at all, just $1_{\{T \le Y\}}$ (or $1_{\{T \le U\}}$).

The earliest work on nonparametric likelihood estimation (NPMLE) with current status data goes back to Ayer, Brunk, Ewing, Reid and Silverman (1955) and Van Eeden (1956, 1957). These authors introduced the pool-adjacent-violators algorithm to compute the NPMLE of a distribution function. Groeneboom (1987) established asymptotic properties of the NPMLE. See also Groeneboom and Wellner (1992) and Huang and Wellner (1995a).

In a semiparametric setting, Huang (1994, 1995, 1996) showed that the maximum likelihood estimator (MLE) for the regression parameter of the proportional hazards or proportional odds regression model with "case 1" interval censoring is asymptotically normal and efficient. Rossini and Tsiatis (1996) proved asymptotic normality and efficiency of sieve estimators for the proportional odds regression model.

There is an enormous amount of literature in econometrics on the binary choice model. This model can be thought as a linear regression model with unknown error distribution under "case 1" interval censoring. The first consistent estimator, the maximum score estimator, was introduced by Manski (1975, 1985). Kim and Pollard (1990) derived its asymptotic distribution. See also Cosslett (1983, 1987), Chamberlain (1986), Han (1987), Horowitz (1992), and Klein and Spady (1993) for related work. In particular, the estimator proposed by Klein and Spady (1993) is asymptotically efficient under appropriate conditions.

1.2. "Case 2" and "case k" interval censoring.

With "case 2" interval censored data, we only know that T has occurred either within some random time interval, or before the left end point of the time interval, or after the right end point of the time interval. More precisely, suppose that there are two examination (or observation) times U and V, the data observed is:

$$(\delta_1, \delta_2, U, V, Z) = (1_{\{T \le U\}}, 1_{\{U < T \le V\}}, U, V, Z).$$

"Case k" interval censoring arises when there are k examination times per subject. This is a generalization of "case 2" interval censoring; see e.g.

Wellner (1996).

Estimation of the proportional hazards model (Cox, 1972) and the proportional odds regression model was considered by Huang and Wellner (1995b) and Huang and Rossini (1995), respectively. They showed that the MLEs of the regression parameters in both models are asymptotically normal and efficient, even though the MLEs of the baseline cumulative hazard function or odds function only have $n^{1/3}$-rates of convergence.

1.3. A general interval censoring scheme.

Suppose that

$$0 < Y_{i,1} < Y_{i,2} < \cdots < Y_{i,n_i} < \infty$$

are ordered examination times for the ith patient, $i = 1, \dots, n$. Denote $Y_i = (Y_{i,1}, Y_{i,2}, \cdots, Y_{i,n_i})$. Let T_i be the ith patient's unobservable failure time. Computationally, it is convenient to reduce the general interval censoring to "case 2" interval censoring by considering three possibilities: (i) the failure occurred before the first examination time. Denote $U_i = Y_{i,1}$ and let $V_i = Y_{i,2}$. Let $\delta_{1i} = 1_{[T_i \le U_i]}$ and $\delta_{2i} = 1_{[U_i < T_i \le V_i]}$. Then $\delta_{1i} = 1$ and $\delta_{2i} = 0$. (ii) T_i is known to be bracketed between a pair of examination times $(Y_{i,L}, Y_{i,R})$, where $Y_{i,L}$ is the last examination time preceding T_i and $Y_{i,R}$ is the first examination time following T_i. Denote $U_i = Y_{i,L}$ and $V_i = Y_{i,R}$. Define δ_{1i} and δ_{2i} as in (i). Then $\delta_{1i} = 0$ and $\delta_{2i} = 1$. (iii) At the last examination, the failure did not occur. Then $\delta_{1i} = 0$ and $\delta_{2i} = 0$. The effective observations are

$$(\delta_{1i}, \delta_{2i}, U_i, V_i), \quad i = 1, \dots, n.$$

Turnbull (1976) derived self-consistency equations for a very general censoring scheme which includes interval censoring as a special case. This yields an EM algorithm for computing the NPMLE. Groeneboom's Iterative Convex Minorant algorithm can be used for computing the NPMLE. As suggested in Groeneboom and Wellner (1992), the Iterative Convex Minorant (ICM) algorithm is considerably faster than the EM algorithm, especially when the sample size is large. Finkelstein (1986) and Rabinowitz, Tsiatis and Aragon (1995) considered estimation in the proportional hazards model (Cox, 1972), and in the linear regression model, with general interval censoring, respectively. Large sample properties of their estimators are still unknown.

Computationally, the general interval censoring scheme can be reduced to "case 2" interval censoring. The estimation approach described in section 4 for "case 2" interval censoring works for general interval censoring. The distributional results do *not* carry over to the general case although they can be easily extended to "case k" interval censoring.

The organization of this paper is as follows. In section 2, computation and distributional results for the nonparametric maximum likelihood estimator $\widehat{F}_n$ (NPMLE) of a distribution function with interval censored are discussed. Section 3 contains recent results on maximum likelihood estimation of three widely used regression models (the proportional hazards, the proportional odds, and the accelerated failure time models) with "case 1" interval censored data. Section 4 contains results on estimation in these three regression models with "case 2" interval censored data We emphasize that, although the distributional results do not carry over to general interval censored data, the estimation procedures *do* carry over. In this section, consistency of the maximum likelihood estimators and information calculation in the accelerated failure time model are new results as far as we know. Section 5 contains brief discussions on further problems. Some proofs, including most of the information calculations, are included in section 6. Throughout, we emphasize the importance of information calculations in the (semiparametric) regression models we consider. These calculations not only provide benchmarks for comparison of estimators, but also help to understand the structure of the models which can facilitate proofs of the distributional results for maximum likelihood estimators and can lead to natural families of estimating equations.

2. Nonparametric likelihood estimation.

We assume that the examination times are independent of the failure time and that their distribution is independent of the distribution function of the failure time. With these conditions, the joint densities and the likelihood functions are:

(1) "Case 1" interval censoring: the joint density of a single observation $X = (\delta, U)$ is

$$p(x) = F(u)^{\delta}(1 - F(u))^{1-\delta}h(u),$$

where $h(u)$ is the density of U. The log-likelihood function of a random sample of size n is (up to an additive term not involving F)

$$l_n(F) = \sum_{i=1}^{n}\{\delta_i \log F(U_i) + (1 - \delta_i)\log(1 - F(U_i)\}.$$

(2) "Case 2", and general interval censoring: the joint density of a single observation $X = (\delta_1, \delta_2, U, V)$ of a random sample of size n is

$$p(x) = F(u)^{\delta_1}[F(v) - F(u)]^{\delta_2}(1 - F(v))^{\delta_3}h(u, v), \tag{2.1}$$

where $\delta_3 = 1 - \delta_1 - \delta_2$ and $h(u, v)$ is the joint density of (U, V). The log-likelihood function is

$$l_n(F) = \sum_{i=1}^{n}\{\delta_{1i}\log F(U_i) + \delta_{2i}\log(F(V_i) - F(U_i)) + \delta_{3i}\log(1 - F(V_i))\},$$

where $\delta_{3i} = 1 - \delta_{1i} - \delta_{2i}$. Notice that for general interval censoring, the meaning of $(\delta_{1i}, \delta_{2i}, U_i, V_i)$ is described in section 1. Since the likelihood functions $l_n(F)$ in both (1) and (2) depend on F only through its values at the observation times U_i (or (U_i, V_i) in the case of (2)), our convention is that the NPMLE $\widehat{F}_n$ of F in either (1) or (2) is a piecewise constant function with jumps only at the observation points. Although "case 1" interval censoring can be regarded as a special case of "case 2" interval censoring by taking $U_i = 0$ (or $V_i = \infty$), maximization of $l_n(F)$ given in (2) is substantially more difficult.

2.1. Computation of $\widehat{F}_n$.

For "case 1" interval-censored data, the NPMLE $\widehat{F}_n$ that maximizes $l_n(F)$ of (1) can be obtained as follows:

(i) Order the examination times: $U_{(1)} \leq U_{(2)} \leq ... \leq U_{(n)}$ and relabel δ_i accordingly to obtain $\delta_{(1)}, \ldots, \delta_{(n)}$.

(ii) Plot $(i, \sum_{j=1}^{i} \delta_{(j)})$, $i = 1, \ldots, n$.

(iii) Form the Greatest Convex Minorant (GCM) G^* of the points in step (ii).

(iv) $\widehat{F}_n(U_{(i)}) =$ left-derivative of G^* at i, $i = 1, \ldots, n$.

Equivalently, $\widehat{F}_n$ can be expressed by the max-min formula:

$$\widehat{F}_n(U_{(i)}) \;=\; \max_{j \leq i} \min_{k \geq i} \frac{\sum_{m=j}^{k} \delta_{(m)}}{k - j + 1}. \tag{2.2}$$

However, for "case 2" or the general interval censoring, there is no closed form expression for $\widehat{F}_n$. We describe the iterative convex minorant algorithm and the EM algorithm for computing $\widehat{F}_n$. Let P_n denote the empirical measure of a sample of size n from the "case 2" density (2.1);

$$P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{(\underline{\delta}_i, U_i, V_i)}.$$

For any distribution function F and $t \geq 0$, denote

$$\begin{aligned} W_F(t) \;&=\; \int_{u \leq t} \left\{ \frac{\delta_1}{F(u)} - \frac{\delta_2}{F(v) - F(u)} \right\} d\, P_n \\ &\quad + \int_{v \leq t} \left\{ \frac{\delta_2}{F(v) - F(u)} - \frac{\delta_3}{1 - F(v)} \right\} d\, P_n, \end{aligned} \tag{2.3}$$

$$\begin{aligned} G_F(t) \;&=\; \int_{u \leq t} \left\{ \frac{\delta_1}{F(u)^2} + \frac{\delta_2}{(F(v) - F(u))^2} \right\} d\, P_n \\ &\quad + \int_{v \leq t} \left\{ \frac{\delta_2}{(F(v) - F(u))^2} + \frac{\delta_3}{(1 - F(v))^2} \right\} d\, P_n, \end{aligned} \tag{2.4}$$

and

$$V_F(t) = W_F(t) + \int_{[0,\ t]} F(s)dG(s)\,. \tag{2.5}$$

Let J_{1n} be the set of examination times U_i such that T_i either belongs to $[0, U_i]$ or $(U_i, V_i]$, and let J_{2n} be the set of examination times V_i such that T_i either belongs to $(U_i, V_i]$ or (V_i, ∞). Furthermore, let $J_n = J_{1n} \cup J_{2n}$ and let $T_{(j)}$ be the jth order statistic of the set J_n.

PROPOSITION 2.1. *(Groeneboom, 1991) Let $T_{(1)}$ correspond to an observation U_i such that $\delta_{1i} = 1$, and let the largest order statistic $T_{(m)}$ correspond to an observation V_i such that $\delta_{1i} = \delta_{2i} = 0$. Then $\widehat{F}_n$ is the NPMLE of F_0 if and only if $\widehat{F}_n$ is the left derivative of the convex minorant of the self-induced cumulative sum diagram formed by the points*

$$P_{(j)} = (G_{\widehat{F}_n}(T_{(j)}),\ V_{\widehat{F}_n}(T_{(j)})), \qquad j = 1, 2, \ldots, m$$

and $P_{(0)} = (0, 0)$.

With this characterization, Groeneboom introduced the *iterative convex minorant algorithm* to compute $\widehat{F}_n$. See Jongbloed (1995a,b) for modifications of the iterative convex minorant algorithm which always converge. Zhan and Wellner (1995) have adapted Jongbloed's argument to the related double censoring model; see also Wellner and Zhan (1996).

It can be verified that the NPMLE $\widehat{F}_n$ satisfies the self-consistency equation:

$$\begin{aligned} \widehat{F}_n(t) &= E_{\widehat{F}_n}[F_n(t) \mid \delta_i, \gamma_i, U_i, V_i, i = 1, \ldots, n] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_{1i}\frac{\widehat{F}_n(U_i \wedge t)}{\widehat{F}_n(U_i)} + \delta_{2i}\frac{\widehat{F}_n(V_i \wedge t) - \widehat{F}_n(U_i \wedge t)}{\widehat{F}_n(V_i) - \widehat{F}_n(U_i)}\right. \\ &\qquad\qquad \left. +\delta_{3i}\frac{\widehat{F}_n(t) - \widehat{F}_n(V_i \wedge t)}{1 - \widehat{F}_n(V_i)}\right\} \end{aligned} \tag{2.6}$$

where F_n is the (unobservable) empirical distribution function of $T_1, \ldots, T_n$. This was first derived by Turnbull (1976). However, this equation does not characterize the NPMLE $\widehat{F}_n$. That is, there may exist other distribution functions different from $\widehat{F}_n$ that satisfy (2.6).

In general, $\widehat{F}_n$ has no closed form expression. In spite of this, Groeneboom (1991) characterized the NPMLE for case 2 interval censored data and developed a fast algorithm (the iterative convex minorant algorithm) for computing the NPMLE. Aragón and Eberly (1992) and Jongbloed (1995a,b) proposed modifications of the iterative convex minorant algorithm, and Jongbloed (1995a,b) shows that his modified algorithm always converges. See Groeneboom and Wellner (1992), pages 69 - 73 and Jongbloed (1995a,b).

2.2. Distributional results.

The asymptotic distribution of $\widehat{F}_n$ with "case 1" interval censoring was established by Groeneboom (1987). See Groeneboom and Wellner (1992), page 89.

THEOREM 2.1. *(Groeneboom, 1987) Let G be the distribution function of U and $g = G'$ be the corresponding density function. Let t_0 be such that $0 < F(t_0), G(t_0) < 1$, and let F_0 and G be differentiable at t_0 with strictly positive derivatives $f(t_0)$ and $g(t_0)$, respectively. Furthermore, let $\widehat{F}_n$ be the NPMLE of F_0. Then we have, as $n \to \infty$,*

$$n^{1/3}\{\widehat{F}_n(t_0) - F_0(t_0)\} \to_d 2c_1(t_0)Z,$$

where Z is the last time where standard two-sided Brownian motion minus the parabola $y(t) = t^2$ reaches its maximum and

$$c_1(t_0) = \{[f(t_0)F_0(t_0)(1 - F_0(t_0)]/2g(t_0)\}^{1/3}.$$

With "case 2" interval-censored data, consistency of $\widehat{F}_n$ was proved by Groeneboom and Wellner (1992) and by Van de Geer (1993). Finding the asymptotic distribution of $\widehat{F}_n$ with "case 2" interval-censored data appears to be a much harder problem. Groeneboom (1991) conjectured that, with a different rate of convergence $(n \log n)^{-1/3}$ and a different scaling constant, the limiting distribution of the NPMLE takes the same form under the assumption that the joint density h of U and V is strictly positive along the diagonal $u = v$. He also showed that an approximation $\widehat{F}_n^{(1)}$ to the NPMLE $\widehat{F}_n$ has this limiting distribution (marginally at a fixed point t_0). However, this conjecture is still not proved. See also Groeneboom and Wellner (1992), page 100. Wellner (1995) proposed to explore alternative hypotheses under which the joint density h of U and V converges to zero as $u \leq v$ approach the diagonal $u = v$; in particular, the hypothesis (C2) below holds if the joint distribution H of (U, V) puts zero mass on some small strip bordering the diagonal $u = v$. Under such hypotheses, he proved a result about the "one-step" approximation to the NPMLE analogous to Groeneboom's theorem (see theorem 2.2 below), and made a corresponding conjecture about the asymptotic distribution of the NPMLE. Although conjecture of Groeneboom (1991) is still unproved, Groeneboom (1996) has succeeded in verifying Wellner's (1995) conjecture in the strict separation case.

Here we will state the version of the result for the "one-step approximation" $\widehat{F}_n^{(1)}$ at a point t_0 under the hypotheses of Wellner (1995). The hypotheses needed are as follows:

Assumption (C1): the support of F_0 is an interval $[0, \tau]$ where $\tau < \infty$, and $t_0 \in (0, \tau)$.

Define functions k_1 and k_2 by

$$k_1(u) = \int_u^\tau \frac{h(u,v)}{F_0(v) - F_0(u)} dv, \quad \text{and} \quad k_2(v) = \int_0^v \frac{h(u,v)}{F_0(v) - F_0(u)} du.$$

We also define functions $k_1(u,\epsilon)$ and $k_2(v,\epsilon)$ which isolate the contributions to k_1 and k_2 by values of v and u respectively near the diagonal $u = v$:

$$k_1(u;\epsilon) \equiv \int_u^\tau \frac{h(u,v)}{F_0(v) - F_0(u)} 1_{[1/(F_0(v)-F_0(u))>\epsilon]} dv$$

and

$$k_2(v;\epsilon) \equiv \int_0^v \frac{h(u,v)}{F_0(v) - F_0(u)} 1_{[1/(F_0(v)-F_0(u))>\epsilon]} du\,.$$

Assumption (C2): For each $\epsilon > 0$ and $i = 1, 2$

$$\alpha \int_{(t_0, t_0 + t/\alpha]} k_i(u; \epsilon\alpha) du \to 0 \qquad \text{as} \qquad \alpha \to \infty. \tag{2.7}$$

Assumption (C3): (a) $F_0(t)$ and $H(u,v)$ have densities $f_0(t)$ and $h(u,v)$ with respect to Lebesgue measure on R and R^2, respectively; (b) $h(u,v)$ has bounded partial derivatives on the support of (U,V). Let $h_1(u)$ and $h_2(v)$ be the marginal densities of U and V, respectively.

Assumption (C4): $0 < F_0(t_0) < 1$ and $0 < H(t_0,t_0) < 1$.

THEOREM 2.2. *(Wellner, 1995). Suppose that assumptions (C1) to (C4) hold. Suppose that f_0, h_1, h_2, k_1 and k_2 are continuous at t_0 and $f_0(t_0) > 0$. Then,*

$$n^{1/3}\{\widehat{F}_n^{(1)}(t_0) - F_0(t_0)\} \to_d 2c_2(t_0)Z,$$

where Z is the last time where standard two-sided Brownian motion $W(t)$ minus the parabola $y(t) = t^2$ reaches its maximum,

$$c_2(t_0) = \{f_0(t_0)/2\xi(t_0)\}^{1/3},$$

and

$$\xi(t_0) = \frac{h_1(t_0)}{F_0(t_0)} + k_1(t_0) + k_2(t_0) + \frac{h_2(t_0)}{1 - F_0(t_0)}. \tag{2.8}$$

From the preceding discussion it is clear that for case 1 interval censored failure time data and examination times U with positive density h, the NPMLE $\widehat{F}_n$ does *not* satisfy the central limit theorem with the usual $n^{1/2}$-rate of convergence, and the use of observed information to construct confidence intervals or confidence bands for F_0 do not have large sample justifications. On the other hand, the general bootstrap theory of Politis and Romano (1994) does apply to yield confidence intervals for $F_0(t_0)$ as a consequence of theorem 2.1; see their example 2.1.1, page 2035, for a problem also involving a convergence rate of $n^{1/3}$. In view of the conjectures by Groeneboom and Wellner concerning $\widehat{F}_n$ with case 2 interval censored data, these remarks very likely apply to case 2 as well.

It should also be noted that with case 1 data the sequence of stochastic processes $\{n^{1/3}(\widehat{F}_n(t) - F_0(t)) : \ 0 \le t \le \tau\}$ (or the corresponding processes with scaling $(n \log n)^{1/3}$ for case 2 data under Groeneboom's hypotheses) is *not tight* in $D[0, \tau]$ and do *not converge weakly as processes*. If $0 < t_1 < t_2 < \tau$ are fixed points with the hypotheses of theorem 2.1 holding at both t_1 and t_2, then, much as in density estimation problems, it can be shown that

$$n^{1/3}(\widehat{F}_n(t_1) - F_0(t_1), \widehat{F}_n(t_2) - F_0(t_2)) \to_d (c_1 Z_1, c_2 Z_2)$$

where Z_1 and Z_2 are independent, and this precludes the possibility of a tight limit process. See Groeneboom (1985) for a study of the (local) dependence structure of this type of process in a closely related problem.

To complete the picture concerning the NPMLE $\widehat{F}_n$, consider case 1 interval censored data when the observation times U *fail to have a continuous density* h. Instead, suppose that U has a discrete distribution H with

$$h_j \equiv P(U = u_j), \qquad j = 1, \dots, d$$

where we suppose $0 < u_1 < \ldots < u_d < \tau$. Then it quickly becomes clear that we can only estimate the distribution function F at the points u_i, $i = 1, \dots, d$: note that the U_i's in the formula for the NPMLE given by (2.2) only take values in the set $\{u_1, \dots, u_d\}$. The NPMLE $\widehat{F}_n(u_j)$ is a consistent estimator of $F_0(u_j)$, and in fact it is $\sqrt{n}$–consistent. This follows by re-expressing the NPMLE $\widehat{F}_n(u_j)$ as a monotonization of the simple binomial estimators obtained via (unconstrained) maximization of the resulting likelihood (see Ayer et al. (1955), Van Eeden (1956), or Robertson, Wright, and Dykstra (1988), example 1.5.1, page 32), and then using the continuous mapping theorem as in Chernoff (1954). It seems that similar remarks also apply to the NPMLE $\widehat{F}_n$ with case 2 or even "case k" type data (the binomial problems become trinomial or multinomial with $k+1$ cells, and the "weights" in the monotonization problem depend on the solution itself, but the continuous mapping theorem still carries the argument).

3. Regression models with interval censoring, case 1.

Several regression models can be viewed as special cases of the transformation model. This model postulates that the conditional distribution $F(t|z)$ of T given the covariate $Z = z$ satisfies

$$g(F(t|z)) = h(t) + \theta' z, \tag{3.1}$$

where g is a specified function, $h(t)$ is an unknown increasing function and θ is a d-dimensional regression parameter. If we take $g(s) = \log[-\log(1-s)], 0 < s < 1$, then (3.1) results in the famous proportional hazards model proposed by Cox (1972); in this case the model is more commonly written in terms of the cumulative hazard function as

$$\Lambda(t|z) = \Lambda(t)e^{\theta' z}, \tag{3.2}$$

where Λ is the unknown baseline cumulative hazard function.

If we take $g(s) = \text{logit}(s) \equiv \log[s/(1-s)], 0 < s < 1$, then we get the proportional odds regression model:

$$\text{logit}[F(t|z)] = \text{logit}[F(t)] + \theta' z, \tag{3.3}$$

where $F(t) \equiv F(t|0)$ is the baseline distribution function. Let $\alpha(t) = \text{logit}\, F(t)$, the baseline monotone increasing log-odds function. The proportional odds regression model is an interesting alternative to the proportional hazards model, and might be appropriate when the proportional hazards assumption is not satisfied. This model has been used by several authors in analyzing survival data; for right-censored data, see Bennett (1983), Pettitt (1984) and Parzen (1993), while for "case 1" interval censored data, see Dinse and Lagakos (1983).

Another important model which is closely related to (3.1) is the accelerated failure time regression model:

$$\log T = Z'\theta + \varepsilon, \tag{3.4}$$

where the distribution function F of ε is completely unspecified and where $\log T$ can be replaced by other more appropriate known monotone functions. In terms of conditional distributions, this model can be written as

$$F(\log T|Z) = F(\log T - Z'\theta).$$

There is an enormous amount of literature on statistical inference for the accelerated failure time model with right-censored data, but much less for this model with interval censoring. Much of the existing literature seems to be in connection with the "binary choice" model in econometrics: see e.g. Cosslett (1983, 1987), Chamberlain (1986), and Manski (1985). For further review and the relationship with interval censoring see Huang and Wellner (1996), and Huang (1993).

Throughout we assume the following basic assumptions:

(A1) The (unobservable) failure time is independent of the examination times given the covariates.

(A2) The joint distribution of the examination times and the covariates are independent of the parameters of interest.

For consistency of the maximum likelihood estimators, the following identifiability condition is needed.

(A3) (a) The distribution of Z is not concentrated on any proper affine subspace of R^d (i.e. of dimension $d-1$ or smaller). (b) Z is bounded, that is, there exists $Z_0 > 0$ such that $P(|Z| \le Z_0) = 1$, where $|\cdot|$ denotes the Euclidean norm in R^d.

For distributional results of the maximum likelihood estimators, we need the following further regularity conditions.

(A4) $F_0(0) = 0$. Let $\tau_{F_0} = \inf\{t : F_0(t) = 1\}$. The support of U is an interval $S[U] = [\tau_0, \tau_1]$, and $0 < \tau_0 \le \tau_1 < \tau_{F_0}$.

(A5) F_0 has a strictly positive and continuous density on $S[U]$, and the joint distribution function $H(u, z)$ of (U, Z) has bounded second order (partial) derivative with respect to u.

Under assumptions (A1) and (A2), the joint density of $X = (\delta, U, Z)$ is

$$p(x) = F(u|z)^{\delta}(1 - F(u|z))^{1-\delta} h(u, z),$$

where $h(u, z)$ is the joint density of (U, Z). So for an independent sample $(\delta_i, U_i, Z_i), i = 1, \ldots, n$ with the same distribution as (δ, U, Z), the general form of the log-likelihood function is, up to an additive constant,

$$l_n = \sum_{i=1}^{n} \{\delta_i \log F(U_i|Z_i) + (1 - \delta_i) \log(1 - F(U_i|Z_i))\}.$$

A distinct feature of "case 1" interval censoring is that, in the regression problems we discuss below, the efficient score function and the Fisher information for the regression parameter have explicit expressions. However, there are no explicit expressions for these quantities with "case 2" interval censoring in general. Consequently, properties of various estimators (in particular, maximum likelihood estimators) are better understood for regression models with "case 1" interval censoring than the corresponding problems with "case 2" interval censoring.

3.1. Proportional hazards model.

In this case, it is convenient to parametrize the model in terms of the regression parameter and the baseline cumulative hazard function. The joint density function for the proportional hazards model with case 1 data is

$$p_{\theta,\Lambda}(x) = p_{\theta,\Lambda}(\delta, u, z) = (1 - e^{-\Lambda(u)e^{\theta' z}})^{\delta} e^{-(1-\delta)\Lambda(u)e^{\theta' z}} h(u, z).$$

It follows that the log-likelihood function is (up to an additive term not involving (θ, Λ)) is:

$$l_n(\theta, \Lambda) = \sum_{i=1}^{n} \left\{ \delta_i \log[1 - \exp(-\Lambda(U_i) e^{\theta' Z_i})] - (1 - \delta_i)\Lambda(U_i) e^{\theta' Z_i} \right\}.$$

3.1.1. Information for θ.

It is well known that in most parametric models and many semiparametric models (such as the Cox model with right censoring), we can estimate the finite-dimensional parameter at $\sqrt{n}$-convergence rate and asymptotically efficiently. A necessary condition is that we must have positive Fisher information. For "case 1" interval censored data, it is not clear a priori that the information is in fact positive. Actually, from the results on estimation of the distribution in the nonparametric setting, the Fisher information for the baseline cumulative hazard function is zero. Therefore, it is useful to calculate the information for the regression parameter. Positive information will suggest that it is possible to estimate θ at the $n^{1/2}$-rate of convergence as in a regular parametric model, even though it is impossible to achieve the same thing for the baseline cumulative hazard function. These comments also apply to the corresponding estimation problem with "case 2" or general interval censoring.

Define the functions

$$Q(u,\delta,z) = \delta \frac{\overline{F}(u \mid z)}{1-\overline{F}(u \mid z)} - (1-\delta), \tag{3.5}$$

and

$$O(u \mid z) = E[Q^2(U,\delta,Z) \mid U=u, Z=z] = \frac{\overline{F}(u \mid z)}{1-\overline{F}(u \mid z)}. \tag{3.6}$$

THEOREM 3.1. *Suppose that assumptions (A1) to (A5) are satisfied. Then:*
(a) The efficient score function for θ is

$$l_\theta^*(x) \quad = \quad e^{\theta' Z} Q(u,\delta,z)\Lambda(u)\left\{ z - \frac{E[(Ze^{2\theta' Z})O(U \mid Z) \mid U=u]}{E[(e^{2\theta' Z})O(U \mid Z) \mid U=u]} \right\}.$$

(b) The information for θ is

$$I(\theta) = E[l_\theta^*(X)]^{\otimes 2} = E\left\{ R(U,Z)\left[Z - \frac{E(ZR(U,Z) \mid U)}{E(R(U,Z) \mid U)} \right]^{\otimes 2} \right\}, \tag{3.7}$$

where $a^{\otimes 2} = aa'$ for any column vector $a \in R^d$, and $R(u,z) = \Lambda^2(u \mid Z)O(u \mid z)$.

The proof of theorem 3.1 will be given in section 6.

3.1.2. Distributional results.

Let $S_0(t)$ be the baseline survival function and let $\widehat{S}_n(t) = \exp(-\widehat{\Lambda}_n(t))$ be the corresponding estimator. Let G be the distribution function of U.

THEOREM 3.2. *(Consistency) Suppose that either (i) conditions (A1), (A2) and (A3) hold, or, (ii) condition (A1), (A2) and (A3a) holds and that the parameter space Θ of θ is bounded. Then*

$$\widehat{\theta}_n \to_{a.s} \theta_0, \quad \textit{and} \quad \widehat{S}_n(t) \to_{a.s} S_0(t) \quad G- \textit{ almost surely }.$$

Two interesting special cases of this theorem are: (i) If U is discrete, then $\widehat{S}_n(t) \to_{a.s} S_0(t)$ at all the mass points of U. (ii) If U has a continuous distribution function whose support contains the support of S_0, then

$$\sup_{0 \le t < \infty} |\widehat{S}_n(t) - S_0(t)| \to_{a.s} 0.$$

This implies for any finite $M > 0$,

$$\sup_{0 \le t \le M} |\widehat{\Lambda}_n(t) - \Lambda_0(t)| \to_{a.s} 0.$$

Although consistency is obtained with minimal assumptions, to obtain rate of convergence and asymptotic normality, further conditions (A4) and (A5) are needed.

THEOREM 3.3. *(Asymptotic normality and efficiency with "case 1" interval censoring) Suppose that θ_0 is an interior point of Θ and that assumptions (A1) to (A5) are satisfied. Then*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d N(0, I(\theta_0)^{-1}),$$

where $I(\theta_0)$ is the generalized Fisher information given in (3.7) which takes into account that the baseline cumulative hazard function Λ_0 is unknown. However,

$$\left\{ \int_{\tau_0}^{\tau_1} (\widehat{\Lambda}_n(u) - \Lambda_0(u))^2 dG(u) \right\}^{1/2} = O_p(n^{-1/3}).$$

Proofs of theorems 3.2 and 3.3 are given in Huang (1996).

3.1.3. Variance estimation.

By Theorem 3.3, $I^{-1}(\theta_0)/n$ is the asymptotic variance-covariance matrix for $\widehat{\theta}_n$. This provides one way to obtain a consistent estimator of the variance of $\widehat{\theta}_n$. Recall in the expression for $I(\theta_0)$, $R(u,z)$ is defined to be

$$R(u,z) = \Lambda^2(u \mid z) O(u \mid z) = e^{2z'\theta_0} \Lambda_0^2(u) \frac{\exp(-\Lambda_0(u) e^{z'\theta_0})}{1 - \exp(-\Lambda_0(u) e^{z'\theta_0})}.$$

Denote

$$\widehat{R}_n(u,z) = e^{2z'\widehat{\theta}_n} \widehat{\Lambda}_n^{\,2}(u) \frac{\exp(-\widehat{\Lambda}_n(u) e^{z'\widehat{\theta}_n})}{1 - \exp(-\widehat{\Lambda}_n(u) e^{z'\widehat{\theta}_n})}. \tag{3.8}$$

Let

$$\mu_1(u) = E(R(u,Z)|U = u), \qquad \text{and} \qquad \mu_2(u) = E(ZR(u,Z)|U = u).$$

Then when we have obtained reasonable estimators $\mu_{1n}(u)$ and $\mu_{2n}(u)$ for $\mu_1(u)$ and $\mu_2(u)$, we can estimate $I(\theta_0)$ by

$$\widehat{I}_n(\widehat{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{R}_n(U_i, Z_i)\left[Z_i - \frac{\mu_{2n}(U_i)}{\mu_{1n}(U_i)}\right]^{\otimes 2}\right\}. \tag{3.9}$$

The hard work is to estimate $\mu_1(u)$ and $\mu_2(u)$. Also, θ_0 and the rest of R need to be estimated. When Z is a continuous covariate vector, $\mu_1(u) \equiv E(R(u,Z)|U=u)$ can be approximated by $E(\widehat{R}_n(u,Z)|U=u)$. Then we can estimate $E(\widehat{R}_n(u,Z)|U=u)$ by nonparametric regression approaches, see, e.g. Stone (1977).

When Z is a categorical covariate, the above nonparametric smoothing procedure does not work well because of the discrete nature of the values of $\widehat{R}_n(u,z)$. Here we consider the simplest case when Z is a dichotomous variable indicating two treatment groups, that is, Z only takes values 0 or 1 with $P\{Z=1\}=\gamma$ and $P\{Z=0\}=1-\gamma$. Thus $E[ZR(Z,U)|U=u] = R(1,u)P\{Z=1|U=u\} = R(1,u)f_1(u)\gamma/f(u)$, and

$$\begin{aligned} E[R(Z,U)|U=u] &= R(1,u)P\{Z=1|U=u\} + R(0,u)P\{Z=0|U=u\} \\ &= \frac{R(1,u)f_1(u)\gamma}{f(u)} + \frac{R(0,u)f_0(u)(1-\gamma)}{f(u)}, \end{aligned}$$

where $f_1(u)$ is the conditional density of Y given $Z=1$, $f_0(u)$ is the conditional density of Y given $Z=0$, and $f(u)$ is the marginal density of Y. Notice that we only need to estimate the ratio the ratio of the two conditional expectations. First we can estimate γ by the total number of subjects in the treatment group with $Z=1$ divided by the sample size. Let $\widehat{f}_{1n}(u)$ be a kernel density estimator of $f_1(u)$, and $\widehat{f}_{0n}(u)$ be a kernel density estimator of $f_0(u)$. Then a natural estimator of $E[ZR(Z,U)|U=u]/E[R(Z,U)|U=u]$ is

$$\widehat{\mu}_n(u) = \frac{\widehat{R}_n(1,u)\widehat{f}_{1n}(u)\widehat{\gamma}_n}{\widehat{R}_n(1,u)\widehat{f}_{1n}(u)\widehat{\gamma}_n + \widehat{R}_n(0,u)\widehat{f}_{0n}(u)(1-\widehat{\gamma}_n)}.$$

Here $\widehat{R}_n(u,z)$ is defined in (3.8). With a proper choice of the bandwidth and kernel in estimation of $f_1(u)$ and $f_0(u)$, the above estimator is consistent, see e.g., Silverman (1986). Hence a reasonable estimator of $I(\theta_0)$ is:

$$\widehat{I}_n(\widehat{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{R}_n(U_i,Z_i)(Z_i - \widehat{\mu}_n(U_i))^2\right\}. \tag{3.10}$$

In the special case when Y and Z are independent, the above nonparametric smoothing is not necessary. In this case, we have

$$I(\theta_0) = E\left\{R(U,Z)\left[Z - \frac{E_Z(ZR(U,Z))}{E_Z R(U,Z)}\right]^{\otimes 2}\right\},$$

where E_Z means expectation with respect to Z. We can simply estimate $I(\theta_0)$ by

$$\widehat{I}_n(\widehat{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{R}_n(U_i, Z_i)\left[Z_i - \frac{\sum_{j=1}^{n} Z_j \widehat{R}_n(U_i, Z_j))}{\sum_{j=1}^{n}\widehat{R}_n(U_i, Z_j)}\right]^{\otimes 2}\right\}. \quad (3.11)$$

Two alternative approaches based on the observed Fisher information or the curvature of the profile likelihood function can also be used. These will be discussed in section 4.1.3.

3.2. Proportional odds model.

We parametrize the model in terms of the regression parameter and the baseline log-odds function $\alpha(t) = \text{logit}[F(t)]$. Then with conditions (A1) and (A2), the joint density function of the observations is

$$p_{\theta,\alpha}(x) = p_{\theta,\alpha}(\delta, u, z) = e^{\delta(\alpha(t)+\theta' z)}(1 + e^{\alpha(t)+\theta' z})^{-1} h(u, z).$$

It follows that the log-likelihood function is (up to an additive term not depending on (θ, α))

$$l_n(\theta, \alpha) = \sum_{i=1}^{n}\left\{\delta_i(\alpha(U_i) + \theta' Z_i) - \log(1 + \exp(\alpha(U_i) + \theta' Z_i))\right\}. \quad (3.12)$$

The maximum likelihood estimator is the $(\widehat{\theta}_n, \widehat{\alpha}_n)$ that maximizes $l_n(\theta, \alpha)$ with the constraint that $\widehat{\alpha}_n$ is a nondecreasing function.

3.2.1. Information for θ.

The following result on the information bound for estimation of θ is given in Rossini and Tsiatis (1996).

THEOREM 3.4. *Suppose that conditions (A1) to (A5) are satisfied. Then:*
(a) The efficient score function for θ is

$$\dot{l}_\theta^*(x) = (\delta - E(\delta|U = u, Z = z))\left(z - \frac{E(Z\, Var(\delta|U, Z)|U = u)}{E(\,Var(\delta|U, Z)|U = u)}\right)$$

(b) The information for θ is

$$I(\theta) = E\left[(\delta - E(\delta|U, Z))^2\left(Z - \frac{E(Z\, Var(\delta|U, Z)|U)}{E(\,Var(\delta|U, Z)|U)}\right)^{\otimes 2}\right]. \quad (3.13)$$

3.2.2. Distribution results.

THEOREM 3.5. *(Asymptotic normality) Suppose that conditions (A1)-(A5) stated earlier hold and that θ_0 is an interior point of Θ. Then*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \rightarrow_d N(0, I(\theta_0)^{-1}),$$

where $I(\theta_0)$ is the Fisher information given in (3.13). However,

$$\left\{ \int [\widehat{\alpha}_n(u) - \alpha_0(u)]^2 dG(u) \right\}^{1/2} = O_p(n^{-1/3}).$$

The proof of this result is given in Huang (1995). Although a boundedness restriction is imposed on $\widehat{\alpha}_n$ in that paper, it can be shown that this restriction can be removed; the first author intends to do this elsewhere.

3.2.3. Variance estimation.

This can be done similarly as in section 3.1.3 by using the explicit form of the asymptotic variance of $\widehat{\theta}_n$ given in terms of $I(\theta)$ in (3.13). Two alternative estimators are based on observed Fisher information or the curvature of the profile likelihood function in a neighborhood of $\widehat{\theta}_n$. We will discuss this further in section 4.1.3.

3.3. The accelerated failure time model.

For simplicity, let T be the logarithm or other appropriate transformation of the failure time and assume that the same transformation has been made for the examination time. So the model is

$$T = Z'\theta + \varepsilon,$$

where $\varepsilon \sim F_0$ unspecified. Suppose that ε is independent of Z and the examination times. Then the density of one observation is

$$p_{\theta,F}(x) = p_{\theta,F}(\delta, u, z) = F(u - \theta' z)^{\delta}(1 - F(u - \theta' z))^{1-\delta} h(u, z);$$

hence the log-likelihood function for (θ, F) is (up to an additive term)

$$l_n(\theta, F) = \sum_{i=1}^{n} \log\{\delta_i F(U_i - \theta' Z_i) + (1 - \delta_i)(1 - F(V_i - \theta' Z_i))\}.$$

The maximum (profile) likelihood estimator of (θ_0, F_0) is the $(\widehat{\theta}_n, \widehat{F}_n)$ that maximizes $l_n(\theta, F)$ with the restriction that $\widehat{F}_n$ is a (sub)distribution function. As proposed by Cosslett (1983), $(\widehat{\theta}_n, \widehat{F}_n)$ can be computed via the maximum profile likelihood approach:

(i) For any fixed θ, maximize $l_n(\theta; F)$ with respect to F under the constraint that F is a distribution function. Denote the resulting maximizer by $F_n(\cdot;\theta)$.

(ii) Substituting $F_n(\cdot;\theta)$ back into $l_n(\theta; F)$, yields the profile likelihood $l_n(\theta; F_n(\cdot;\theta))$. Then the maximum profile likelihood estimator $\widehat{\theta}_n$ is any value of θ that maximizes $l_n(\theta; F_n(\cdot;\theta))$ (assuming that it exists).

(iii) A natural estimator of F is $F_n(\cdot;\widehat{\theta}_n)$.

3.3.1. Information for θ.

The following is the result concerning the efficient score and the information bound for θ in this model.

THEOREM 3.6. *Let $k(s) = E(Z|U - Z'\theta = s)$. Suppose that k is bounded. Then the efficient score for θ is*

$$l^*_\theta(x) = f(u - z'\theta)[z - E(Z|U - Z'\theta = y - z'\theta)]\left[\frac{1-\delta}{1-F(u-z'\theta)} - \frac{\delta}{F(u-z'\theta)}\right]. \tag{3.14}$$

Moreover, the information for estimation of θ is:

$$I(\theta) = E\left\{\frac{f(U-Z'\theta)^2}{F(U-Z'\theta)(1-F(U-Z'\theta))}[Z - E(Z|U-Z'\theta)]^{\otimes 2}\right\}. \tag{3.15}$$

The proof of theorem 3.6 will be given in section 6.

3.3.2. Distributional results.

Consistency of $(\widehat{\theta}_n, \widehat{F}_n(\cdot;\widehat{\theta}_n))$ can be shown as in Cosslett (1983); alternatively, consistency can be shown as in theorem 4.7.

We have not been able to establish distributional results for $\widehat{\theta}_n$. It is not clear whether or not $\widehat{\theta}_n$ has a normal limiting distribution with $n^{1/2}$-rate of convergence since $l_n(\theta, \widehat{F}_n(\cdot;\theta))$ is not a "smooth" function of θ. This is a challenging open problem.

3.4. Computation.

We describe in detail an approach for computing the maximum likelihood estimators in the proportional hazards model.

Unlike in the nonparametric setting, there is no closed form expression for $\widehat{\Lambda}_n$. We describe an algorithm which is also applicable to general interval censoring.

In principle, computation of $(\widehat{\theta}_n, \widehat{\Lambda}_n)$ can be accomplished by maximizing (4.2) jointly with respect to (θ, Λ), or, by using a profile likelihood approach, maximizing over Λ for all fixed values of $\theta \in \Theta$ first to obtain $\widehat{\Lambda}_n(\cdot,\theta)$, and then maximizing the profile log-likelihood function $l_n(\theta, \widehat{\Lambda}_n(\cdot,\theta))$ over θ to find $\widehat{\theta}_n$ and hence $\widehat{\Lambda}_n = \widehat{\Lambda}_n(\cdot,\widehat{\theta}_n)$. For low dimensional θ, Groeneboom's iterative convex minorant algorithm (see Groeneboom and Wellner (1992), pages 69 - 74) is sufficiently fast to implement this profile likelihood approach for computing $(\widehat{\theta}_n, \widehat{\Lambda}_n)$. For higher dimensional θ we propose the following iterative algorithm for computing $(\widehat{\theta}_n, \widehat{\Lambda}_n)$. Let $\theta^{(0)}$ be an initial guess and set $k = 0$.

Step (i). Maximize $l_n(\theta^{(k)}, \Lambda)$ with respect to Λ to obtain $\Lambda^{(k)}$.

Step (ii). Maximize $l_n(\theta, \Lambda^{(k)})$ with respect to θ. Set $k = k + 1$, and let $\theta^{(k)}$ be the maximizer. Go back to (i). Repeat (i) and (ii) until convergence.

It can be verified that for any fixed θ, $l_n(\theta, \Lambda)$ is concave in Λ, and for any fixed Λ, $l_n(\theta, \Lambda)$ is concave in θ. So steps (i) and (ii) are well defined concave maximization problems. Since each iteration increases the likelihood function, the algorithm converges. A much stronger conclusion holds as stated in the following proposition.

Proposition 3.1. *(1) The function $l_n(\theta, \Lambda)$ has a unique maximizer $(\widehat{\theta}_n, \widehat{\Lambda}_n)$; (2) Starting from any point $\theta^{(0)}$, the algorithm produces a sequence $(\theta^{(k)}, \Lambda^{(k)})$ converging to $(\widehat{\theta}_n, \widehat{\Lambda}_n)$.*

Proof. Reparametrize $l_n(\theta, \Lambda)$ in terms of θ and $\phi = \log \Lambda$. Let

$$\begin{aligned} l_n(\theta, \phi) &= \sum_{i=1}^{n} \{\delta_i \log[1 - \exp(-\exp(\theta' Z_i + \phi(U_i)))] \\ &\qquad - (1 - \delta_i)\exp(\theta' Z_i + \phi(U_i))\}. \end{aligned}$$

Then maximizing $l_n(\theta, \Lambda)$ with respect to (θ, Λ) with monotonicity constraints on Λ is equivalent to maximizing $l_n(\theta, \phi)$ with respect to (θ, ϕ) with monotonicity constraints on ϕ. Since functions $\log[1 - \exp(-\exp(x))]$ and $-\exp(x)$ are concave, $l_n(\theta, \phi)$ is concave in (θ, ϕ) (This follows from Theorem 5.7 of Rockafellar (1970), page 38). Thus a local maximizer of $l_n(\theta, \phi)$ will also be a global maximizer. This implies a local maximizer of $l_n(\theta, \Lambda)$ will also be a global maximizer because of the one-to-one correspondence between (θ, Λ) and (θ, ϕ). The uniqueness of the global maximizer follows since $l_n(\theta, \phi)$ is strictly concave on the product of the space of θ and the reduced space of ϕ. Here the reduced space of ϕ means the space of the distinct values of $\phi(U_{(1)}), \ldots, \phi(U_{(n)})$, since the monotonicity constraints will force $\phi(U_{(i)}), i = 1, \ldots, n$ to be blockwise constant. □

To implement step (i), we use the iterative convex minorant algorithm introduced by Groeneboom. To describe it in the present setting, we first introduce some notation. Let

$$q(u', z, \theta, \Lambda) = \frac{\exp(-\exp(\theta' z)\Lambda(u'))}{1 - \exp(-\exp(\theta' z)\Lambda(u'))}.$$

Define the processes W_Λ, G_Λ and V_Λ by

$$W_\Lambda(u) = \int_{u' \in [0,u]} \left\{ 1_{\{t \le u'\}} q(u', z, \theta, \Lambda) - 1_{\{t > u'\}} \exp(\theta' z) \right\} dQ_n(t, u', z),$$

$$G_\Lambda(u) = \int_{u' \in [0,u]} \left\{ 1_{\{t \le u'\}} q(u', z, \theta, \Lambda) - 1_{\{t > u'\}} \exp(\theta' z) \right\}^2 dQ_n(t, u', z),$$

and

$$V_\Lambda(u) = W_\Lambda(u) + \int_{[0,u]} \Lambda(u') dG_\Lambda(u'), \qquad u \ge 0,$$

where Q_n is the empirical measure of the (unobservable) points (T_i, U_i, Z_i), $i = 1, \ldots, n$.

THEOREM 3.7. *For any fixed* θ*, suppose that* $\delta_{(1)} = 1, \delta_{(n)} = 0$*. then* $\Lambda_n(\cdot;\theta)$ *maximizes* $l_n(\theta, \Lambda)$ *if and only if* $\Lambda_n(\cdot;\theta)$ *is the left derivative of the greatest convex minorant of the "self-induced" cumulative sum diagram, consisting of the points*

$$P_{(j)} = \left(G_{\Lambda_n(\cdot;\theta)}(U_{(j)}), V_{\Lambda_n(\cdot;\theta)}(U_{(j)}) \right), \qquad j = 1, \cdots, n,$$

and the origin $(0,0)$*.*

The cumulative sum diagram is simply the linear interpolation of the points $P_{(0)}, P_{(1)}, \ldots, P_{(m)}$, and the greatest convex minorant is the greatest convex function that is below this linear interpolation.

The proof of Theorem 3.7 is completely analogous to that of Proposition 1.4 of Groeneboom and Wellner (1992). This theorem gives an iterative procedure to compute $\Lambda_n(\cdot;\theta)$ for any fixed θ. It proceeds as follows. Suppose $\Lambda^{(k)}(\cdot;\theta)$ is obtained at the kth iteration; then $\Lambda^{(k+1)}(\cdot;\theta)$ is computed as the left derivative of the convex minorant of the cumulative sum diagram, consisting of the points

$$(G_{\Lambda^{(k)}(\cdot;\theta)}(U_{(j)}), V_{\Lambda^{(k)}(\cdot;\theta)}(U_{(j)})), \qquad j = 1, \ldots, n;$$

and the origin $(0,0)$.

In step (ii), the Newton-Raphson method can be used. Specifically, for any fixed Λ, let

$$s_1(\theta) = (\partial/\partial\theta) l_n(\theta, \Lambda).$$

By concavity, the solution to $s_1(\theta) = 0$ is the unique maximizer of $l_n(\theta, \Lambda)$ (for fixed Λ).

The approach described above can be used for computing $(\widehat{\theta}_n, \widehat{\alpha}_n)$ in the proportional odds model. This is because $l_n(\theta, \alpha)$ is a concave function of (θ, α), so proposition 3.1 holds for the proportional odds model. To see this, verify that $-\log(1 + \exp(x))$ is a concave function. Thus the concavity of $l_n(\theta, \alpha)$ follows from Theorem 5.7 of Rockafellar (1970), page 38, or it can be verified directly.

4. Regression models with interval censoring, case 2.

Again we assume conditions (A1) to (A3) stated in the beginning of section 3. The joint density of $X = (\delta_1, \delta_2, \delta_3, U, V, Z)$ where $\delta_i \in \{0,1\}$ for $i =$

$1, 2, 3$ and $\delta_1 + \delta_2 + \delta_3 = 1$ is

$$p(x) = F(u|z)^{\delta_1}[F(v|z) - F(u|z)]^{\delta_2}(1 - F(v|z))^{\delta_3} h(u, v, z); \tag{4.1}$$

here h is the joint density function of (U, V, Z) with respect to the product of Lebesgue measure on R^2 and a fixed measure μ on R^d. The log-likelihood likelihood function of an independent sample $(\delta_{1i}, \delta_{2i}, \delta_{3i}, U_i, V_i, Z_i)$, $i = 1, \ldots, n$ with the same distribution as $(\delta_1, \delta_2, \delta_3, U, V, Z)$ is, up to an additive term not depending on $F(\cdot|Z)$,

$$\begin{aligned} l_n &= \sum_{i=1}^n \{\delta_{1i} \log F(U_i|Z_i) \\ &\qquad + \delta_{2i} \log[F(V_i|Z_i) - F(U_i|Z_i)] + \delta_{3i} \log(1 - F(V_i|Z_i))\}. \end{aligned}$$

For proofs of the distributional results of maximum likelihood estimators discussed below, we also need the following regularity conditions.

(B4) (a) There exists a positive number η such that $P(V - U \geq \eta) = 1$; (b) the union of the support of U and V is contained in an interval $[\tau_0, \tau_1]$, where $0 < \tau_0 < \tau_1 < \infty$.

(B5) F_0 has strictly positive and bounded continuous derivative on $[\tau_0, \tau_1]$.

(B6) The conditional density $g(u, v|z)$ of (U, V) given Z has bounded partial derivatives with respect to u and v. The bounds of these partial derivatives do not depend on z.

4.1. The proportional hazards model.

In the case of the proportional hazards model, the log-likelihood function for the regression parameter θ and the baseline cumulative hazard function Λ is

$$\begin{aligned} l_n(\theta, \Lambda) &= \sum_{i=1}^n \Big\{\delta_{1i} \log(1 - \exp(-\Lambda(U_i)e^{\theta' Z_i})) \\ &\quad + \delta_{2i} \log[\exp(-\Lambda(U_i)e^{\theta' Z_i}) - \exp(-\Lambda(V_i)e^{\theta' Z_i})] \\ &\quad - \delta_{3i}\Lambda(V_i)e^{\theta' Z_i}\Big\}. \end{aligned} \tag{4.2}$$

The maximum likelihood estimator is then the $(\widehat{\theta}_n, \widehat{\Lambda}_n)$ that maximizes $l_n(\theta, \Lambda)$ under the constraint that $\widehat{\Lambda}_n$ is a nonnegative and nondecreasing function.

4.1.1. Information for θ.

Let $x = (\delta_1, \delta_2, \delta_3, u, v, z)$. Denote the log-likelihood function for one observation by

$$\begin{aligned} l(x; \theta, \Lambda) &= \delta_1 \log\{1 - \exp(-\Lambda(u)e^{\theta' z})\} \\ &\quad + \delta_2 \log[\exp(-\Lambda(u)e^{\theta' z}) - \exp(-\Lambda(v)e^{\theta' z})] \\ &\quad - \delta_3\Lambda(v)e^{\theta' z}\}. \end{aligned} \tag{4.3}$$

Let f and F be the density and distribution corresponding to Λ, and let f_s be a one-dimensional smooth curve through f, where the smoothness is with respect to s. Denote $a = \frac{\partial}{\partial s} \log f_s \big|_{s=0}$. Then $a \in L_2^0(F) \equiv \{a : \int a dF = 0 \text{ and } \int a^2 dF < \infty\}$. Let

$$h = \frac{\partial}{\partial s} \Lambda_s \bigg|_{s=0} = \frac{\int_0^{\cdot} a dF}{1 - F} = \frac{-\int_{\cdot}^{\infty} a dF}{1 - F}; \tag{4.4}$$

it follows from Hardy's inequality that $h \in L_2(F)$; see Bickel, Klaassen, Ritov, and Wellner (1993), page 423 [hereafter referred to as BKRW (1993)].

The score function for θ is

$$\dot{l}_\theta(x) = \frac{\partial}{\partial \theta} l(x; \theta, \Lambda),$$

and the score operator for Λ is

$$\dot{l}_\Lambda a(x) = \frac{\partial}{\partial s} l(x; \theta, \Lambda_s) \bigg|_{s=0}.$$

Explicit expressions for $\dot{l}_\theta(x)$ and $\dot{l}_\Lambda a(x)$ can be obtained by carrying out the differentiation. The score operator $\dot{l}_\Lambda$ maps $L_2^0(F)$ to $L_2^0(P)$, where P is the joint probability measure of $(\delta_1, \delta_2, U, V, Z)$ and $L_2^0(P)$ is defined similarly as $L_2^0(F)$. Let $\dot{l}_\Lambda^T : L_2^0(P) \to L_2^0(F)$ be the adjoint operator of $\dot{l}_\Lambda$, i.e., for any $a \in L_2^0(F)$ and $b \in L_2^0(P)$, define h as in (4.7),

$$\langle b, \dot{l}_\Lambda a \rangle_P = \langle \dot{l}_\Lambda^T b, a \rangle_F \ ,$$

where $\langle \cdot, \cdot \rangle_P$ and $\langle \cdot, \cdot \rangle_F$ are the inner products in $L_2^0(P)$ and $L_2^0(F)$, respectively. We need to find a_* such that $\dot{l}_\theta - \dot{l}_\Lambda(a_*)$ is orthogonal to $\dot{l}_\Lambda(a)$ in $L_2^0(P)$. This amounts to solving the following normal equation:

$$\dot{l}_\Lambda^T \dot{l}_\Lambda(a_*) = \dot{l}_\Lambda^T \dot{l}_\theta. \tag{4.5}$$

The value of $\dot{l}_\Lambda^T$ at any $b \in L_2^0(P)$ can be computed by

$$\dot{l}_\Lambda^T b(t) = E[b(X)|T = t] = E_Z E[b(X)|T = t, Z];$$

see e.g. BKRW (1993), pages 271-272, or Groeneboom and Wellner (1992), pages 8 and 9.

The proof of the following theorem is deferred to the appendix.

THEOREM 4.1. *For $\theta \in R$, under conditions (A1)-(A3) and (B4)-(B6), equation (4.5) has a unique solution a_*. Moreover, the corresponding function h_* (given by (4.4) with a replaced by a_*) has a bounded derivative. In general, for $\theta \in R^d$, $h_* \in L_2(F)^d$ is a d-dimensional vector and each component has a bounded derivative. The efficient score for θ is*

$$l_\theta^*(x) = \dot{l}_\theta(x) - \dot{l}_\Lambda a_*(x),$$

and the information bound for θ *is*

$$I(\theta) = E[l^*_\theta(X)]^{\otimes 2},$$

where $a^{\otimes 2} = aa'$ *for any column vector* $a \in R^d$. *Under conditions (A1)–(A5),* $I(\theta)$ *is a positive definite matrix with finite entries. The efficient influence function is*

$$\widetilde{l}_\theta(x) = I(\theta)^{-1} l^*_\theta(x).$$

4.1.2. Distributional results.

Let G be the joint distribution function of (U, V) and let G_1 and G_2 be the marginal distribution functions of U and V, respectively. The following assumption is needed in the consistency result.

Theorem 4.2. *(Consistency) Under the same conditions as in theorem 3.2, we have*

$$\widehat{\theta}_n \to_{a.s} \theta_0, \quad \textit{and} \quad \widehat{S}_n(t) \to_{a.s} S_0(t) \quad (G_1 + G_2) - \ \textit{almost everywhere}\ .$$

Two interesting special cases of this theorem are: (i) If both U and V are discrete, this implies $\widehat{S}_n(t) \to_{a.s} S_0(t)$ at all the mass points of U and V. (ii) If at least one of U and V has a continuous distribution function whose support contains the support of S_0, then

$$\sup_{0 \le t < \infty} |\widehat{S}_n(t) - S_0(t)| \to_{a.s} 0.$$

This implies for any finite $M > 0$,

$$\sup_{0 \le t \le M} |\widehat{\Lambda}_n(t) - \Lambda_0(t)| \to_{a.s} 0.$$

With regularity conditions (B4) to (B6), we have the following theorem.

Theorem 4.3. *(Asymptotic normality and efficiency) Suppose that* θ_0 *is an interior point of the bounded set* Θ *and that conditions (A1)-(A3) and (B4)-(B6) hold. Then*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d N(0, I(\theta_0)^{-1}),$$

where $I(\theta_0)$ *is the generalized Fisher information for* θ *taking into account that the baseline cumulative hazard function* Λ_0 *is unknown; hence* $\widehat{\theta}_n$ *is asymptotically normal and efficient. For* $\widehat{\Lambda}_n$*,*

$$||\widehat{\Lambda}_n - \Lambda_0||_2 = O_p(n^{-1/3})\,.$$

Proofs of theorems 4.2 and 4.3 can be found in Huang and Wellner (1995b). Notice that in that paper, a boundedness restriction is imposed on

$\widehat{\Lambda}_n$. We recently found that such restriction can be removed with additional arguments. These will appear in the revision of that paper.

It is noteworthy that, with both "case 1" and "case 2" interval censoring, although the nonparametric component $\widehat{\Lambda}_n$ only has $n^{1/3}$ rate of convergence, the parametric component can be estimated at $n^{1/2}$ rate. With "case 1" interval censoring, Theorem 3.3 provides an explicit form of the asymptotic variance which can be used to estimate the variance of $\widehat{\theta}_n$. With "case 2" interval censoring, although theorem 4.3 does not help in estimating the variance of $\widehat{\theta}_n$ because of the implicitness of $I(\theta)$, it lends support for the use of the observed Fisher information or the curvature of the profile likelihood to estimate the variance of $\widehat{\theta}_n$.

We emphasize that imposition of the conditions for theorems 3.3 and 4.3 is for mathematical rigor of the proofs. In applications, implementation of the estimation procedure described in section 3.1 does not require that these conditions be satisfied. (Of course, the independent censorship assumption (A1) is needed for our likelihood based approach to be sensible.) Although heuristic arguments suggest that conditions (A3)-(A4) and (B4)-(B6) can be weakened, at present, we have not yet been able to prove these two theorems without these conditions. On the other hand, these conditions are not too restrictive in many practical situations. For example, assumption (B4) means that there is a positive time interval between two examination times. This is often the case since two examination times are usually separated by a positive time interval. We believe that asymptotic normality of $\widehat{\theta}_n$ in Theorems 3.3 and 4.3 continue to hold without(A3) or (B4). However, the proofs will be considerably more difficult.

4.1.3. Variance estimation.

With the "case 2" or general interval censoring, there is no close form expression for the asymptotic variance of $\widehat{\theta}_n$. Direct estimation appears to be difficult. We suggest two approaches which are straightforward extensions of two well-known methods used for parametric models. Let $l_n(\widehat{\theta}_n, \widehat{\Lambda}_n)$ be the log-likelihood function of $\widehat{\theta}_n$ and the *distinct* values of $\widehat{\Lambda}_n$. Let

$$\widehat{\Sigma}_{11} = \frac{\partial^2}{\partial \widehat{\theta}_n^2} l_n(\widehat{\theta}_n, \widehat{\Lambda}_n), \quad \widehat{\Sigma}_{12} = \frac{\partial^2}{\partial \widehat{\theta}_n \partial \widehat{\Lambda}_n} l_n(\widehat{\theta}_n, \widehat{\Lambda}_n), \quad \widehat{\Sigma}_{22} = \frac{\partial^2}{\partial \widehat{\Lambda}_n^2} l_n(\widehat{\theta}_n, \widehat{\Lambda}_n),$$

and

$$\widehat{\Sigma}_n = \begin{pmatrix} \widehat{\Sigma}_{11} & \widehat{\Sigma}_{12} \\ \widehat{\Sigma}_{12}' & \widehat{\Sigma}_{22} \end{pmatrix}.$$

Let $\widehat{\Sigma}^{11}$ be the upper left corner corresponding to $\widehat{\Sigma}_{11}$ of the inverse of $\widehat{\Sigma}_n$. Then $\widehat{\Sigma}^{11}$ can be used as an estimator of the variance-covariance matrix of $\widehat{\theta}_n$. It is well known that

$$\widehat{\Sigma}^{11} = (\widehat{\Sigma}_{11} - \widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1}\widehat{\Sigma}_{12}')^{-1}.$$

Computing the inverse of Σ_{22}^{-1} is not as bad as it looks even for large sample size n. $\widehat{\Lambda}_n$ is blockwise constant, and by the large sample properties of $\widehat{\Lambda}_n$ (Theorem 4.3) presented in the next section, the number of distinct values of $\widehat{\Lambda}_n$ is on the order of $n^{1/3}$. So the dimension of Σ_n is in the order of $O(n^{1/3})$.

The second approach is via profile likelihood: for low dimensional θ, it is sometimes more convenient to compute the curvature of the profile likelihood function $l_n(\theta, \widehat{\Lambda}_n(\cdot, \theta))$. The inverse of the curvature can be used as an estimator of the variance covariance matrix of $\widehat{\theta}_n$. Limited simulations and the example in section 4 suggest that $l_n(\theta, \widehat{\Lambda}_n(\cdot, \theta))$ is probably a smooth function of θ. But a rigorous proof of this smoothness is not yet available. In general, we can interpolate $l_n(\theta, \widehat{\Lambda}_n(\cdot, \theta))$ in a neighborhood of $\widehat{\theta}_n$ by a quadratic function. The second derivative of this quadratic function can be used as an approximation to the curvature of $l_n(\theta, \widehat{\Lambda}_n(\cdot, \theta))$.

It should be noted, however, that the asymptotic validity of both of the above approaches remains to be verified.

4.2. The proportional odds model.

In the case of the proportional odds model, the density of one observation is as in (4.1) with

$$F(t|z) = \exp(\alpha(t) + \theta' z)/(1 + \exp(\alpha(t) + \theta' z)). \qquad (4.6)$$

Hence the log-likelihood function of the regression parameter θ and the baseline log-odds function α is

$$\begin{aligned} l_n(\theta, \alpha) = \sum_{I=1}^{n} \{ & \delta_{1i} \log F(U_i|Z_i) \\ & + \delta_{2i} \log[F(V_i|Z_i) - F(U_i|Z_i)] + \delta_{3i} \log[1 - F(V_i|Z_i)]\} \end{aligned}$$

where the conditional distribution function $F(t|z)$ is given as in (4.6). The maximum likelihood estimator is the $(\widehat{\theta}_n, \widehat{\alpha}_n)$ that maximizes $l_n(\theta, \alpha)$ under the constraint that $\widehat{\alpha}_n$ is a nondecreasing function.

4.2.1. Information for θ.

The log-likelihood function for one observation is, up to an additive constant not dependent on (θ, α),

$$l(x, \theta, \alpha) = \delta_1 \log F(u|z) + \delta_2 \log[F(v|z) - F(u|z)] + \delta_3 \log[1 - F(v|z)],$$

with $F(t|z) = \exp(\alpha(t) + \theta' z)/(1 + \exp(\alpha(t) + \theta' z))$. The score function $\dot{l}_\theta(x)$ for θ is simply the vector of partial derivatives of $l(x, \theta, \alpha)$ with respect to θ. In the following, we carry out the calculation for $\theta \in R$; the general case $\theta \in R^d$ only requires repeating the same calculation for each component of $\dot{l}_\theta(x)$.

Recall that $\alpha(t) = \text{logit } F(t) = \log F(t) - \log(1 - F(t))$, and let f be the density corresponding to F with f_s a one-dimensional smooth curve through f. Denote $a = \frac{\partial}{\partial s} \log f_s\big|_{s=0}$. Then $a \in L_2^0(F) \equiv \{a : \int a dF = 0 \text{ and } \int a^2 dF < \infty\}$. Let

$$h = \frac{\partial}{\partial s}\alpha_s\bigg|_{s=0} = \frac{\int_0^{\cdot} adF}{F} - \frac{\int_{\cdot}^{\infty} adF}{1-F} = \int_0^{\cdot} adF\left\{\frac{1}{F} + \frac{1}{1-F}\right\} \tag{4.7}$$

since $\int_0^\infty adF = 0$. The score operator for α, $\dot{l}_\alpha$, maps $L_2^0(F)$ to $L_2^0(P)$, where P is the joint probability measure of $(\delta_1, \delta_2, \delta_3, U, V, Z)$ and $L_2^0(P)$ is defined similarly as $L_2^0(F)$. This can be computed as

$$\dot{l}_\alpha a(x) = \frac{\partial}{\partial s} l(x, \theta, \alpha_s)\bigg|_{s=0}. \tag{4.8}$$

Let $\dot{l}_\alpha^T : L_2^0(P) \to L_2^0(F)$ be the adjoint operator of $\dot{l}_\alpha$, i.e., for any $a \in L_2^0(F)$ and $b \in L_2^0(P)$, define h as in (4.7),

$$\langle b, \dot{l}_\alpha a\rangle_P = \langle \dot{l}_\alpha^T b, a\rangle_F \ ,$$

where $\langle \cdot,\cdot\rangle_P$ and $\langle \cdot,\cdot\rangle_F$ are inner products in $L_2^0(P)$ and $L_2^0(F)$, respectively. We need to find a_* such that $\dot{l}_\theta - \dot{l}_\alpha(a_*)$ is orthogonal (in $L_2^0(P)$) to $\dot{l}_\alpha(a)$ for all a in $L_2^0(F)$. This amounts to solving the following normal equation:

$$\dot{l}_\alpha^T \dot{l}_\alpha(a_*)(t) = \dot{l}_\alpha^T \dot{l}_\theta(t). \tag{4.9}$$

THEOREM 4.4. *For $\theta \in R$, under conditions (A1)–(A3) and (B4)-(B6), equation (4.9) has a unique solution a_*. Moreover, the corresponding function h_* (given by (4.7) with a replaced by a_*) has bounded derivative. In general, for $\theta \in R^d$, h_* is a d-dimensional vector of functions in $L_2^0(F)$ and each component has a bounded derivative. The efficient score for θ is*

$$l_\theta^*(x) = \dot{l}_\theta(x) - \dot{l}_\alpha a_*(x).$$

The information bound for θ is

$$I(\theta) = E[l_\theta^*(X)^{\otimes 2}].$$

$I(\theta)$ is a positive definite matrix with finite entries. The efficient influence function is

$$\tilde{l}_\theta(x) = I(\theta)^{-1} l_\theta^*(x).$$

The proof of this theorem is given in Huang and Rossini (1996).

4.2.2. Distributional results.

Consistency of $(\widehat{\theta}_n, \widehat{\alpha}_n)$ can be shown similarly to theorem 4.2. For brevity, we only state the distributional results for $\widehat{\theta}_n$ which can be proved in the same way as theorem 4.3. See Huang and Wellner (1995b).

THEOREM 4.5. *(Asymptotic normality and efficiency) Under the same conditions as in theorem 4.3,*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d N(0, I^{-1}(\theta_0)).$$

Hence $\widehat{\theta}_n$ is asymptotically normal and efficient.

4.2.3. Variance estimation.

This can be done similarly as in the proportional hazards model with "case 2" interval censoring, by using the observed Fisher information or the curvature of the profile likelihood function in a neighborhood of $\widehat{\theta}_n$.

4.3. The accelerated failure time model.

The log-likelihood function is, up to an additive constant,

$$\begin{aligned} l_n(\theta, F) &= \sum_{i=1}^{n} \{\delta_{1i} \log F(U_i - \theta' Z_i) \\ &\quad + \delta_{2i} \log[F(V_i - \theta' Z_i) - F(U_i - \theta' Z_i)] \\ &\quad + \delta_{3i} \log(1 - F(V_i - \theta' Z_i))\}. \end{aligned}$$

The maximum (profile) likelihood estimator of (θ_0, F_0) is the $(\widehat{\theta}_n, \widehat{F}_n)$ that maximizes $l_n(\theta, F)$. $(\widehat{\theta}_n, \widehat{F}_n)$ can be computed as in the "case 1" interval censoring, using the maximum profile likelihood approach of Cosslett (1983).

Asymptotic distribution theory for $\widehat{\theta}_n$ appears to be a more difficult problem than the corresponding one in the proportional hazards model, and is apparently still unknown. The main difficulty is that two parameters θ and F are "bundled" together: the result is that $\widehat{F}_n(\cdot; \theta)$ and consequently also the profile likelihood function are not smooth functions of θ. This model provides an example in which estimation of the variance-covariance matrix of $\widehat{\theta}_n$ by observed information fails since $l_n(\theta, F)$ is not a twice-differentiable function of θ.

4.3.1. Information.

For "case 2" interval censoring, as far as we know, the information calculation for θ has not been carried out and no efficient estimator of θ has been constructed. (For some work on inefficient estimators based on natural families of estimating equations, see Rabinowitz, Tsiatis, and Aragon (1995). Below we show that the information $I(\theta)$ for θ is positive. This suggests that it is possible to construct efficient estimators for θ. However, in contrast to the "case 1" setting, $I(\theta)$ does not have an explicit expression. The least favorable direction which defines the efficient score and $I(\theta)$ is determined by an integral equation with no closed form solution in general. This equation is similar to the one encountered in calculating the information for smooth functionals of the distribution function in the nonparametric setting (with no covariates). This problem is solved by Geskus and Groeneboom (1996a,b,c). Their approach and results can be used in the present setting. It should be noted that although the information calculations for this model with "case 1" interval censoring is quite straightforward, the same problem with "case 2" interval censored data is considerably more difficult. It involves an integral equation with a singular kernel (in general), and hence the Fredholm theory of integral equations cannot be applied directly.

We need the following conditions to carry out the information calculations. These conditions are similar (and slightly stronger for simplicity) to those given in Geskus and Groeneboom (1996c) in calculating the information for smooth functionals of the distribution function in the nonparametric setting. Let $U_1 = U - Z'\theta$ and $V_1 = V - Z'\theta$, and let h_1 be the joint density of (U_1, V_1). Let h_{11} and h_{12} be the marginal densities of U_1 and V_1, respectively. Finally, let

$$
\begin{aligned}
h_2(u_1, v_1) &= E\{Z|U_1 = u_1, V_1 = v_1\} h_1(u_1, v_1) \\
&= \int z h(u_1 + z'\theta, v_1 + z'\theta, z) d\mu(z).
\end{aligned}
$$

Assumption (D1): the union of the supports of U_1 and V_1 is a bounded interval $[-\tau_1, \tau_2]$, where $0 < \tau_1, \tau_2 < \infty$.

Assumption (D2): F has a continuous density f which is bounded on $[-\tau_1, \tau_2]$.

Assumption (D3): $h_1(u_1, v_1)$ is continuous and has continuous partial derivatives which are bounded, uniformly over $-\tau_1 \le u_1 \le v_1 \le \tau_2$. In addition, $h_{11}(s) + h_{12}(s) > 0$ for all $s \in [-\tau_1, \tau_2]$.

Assumption (D4): $h_2(u_1, v_1)$ is continuous and uniformly bounded over $-\tau_1 \le u_1 \le v_1 \le \tau_2$.

THEOREM 4.6. *Suppose that assumptions (D1) to (D4) hold. Then the efficient score for θ is*

$$
l_\theta^*(x) = \dot{l}_\theta(x) - \dot{l}_f a_*(x), \tag{4.10}
$$

where $\dot{l}_\theta$ and $\dot{l}_f a$ are defined by (6.9) and (6.10) below, and a_* is the unique solution to equation (6.12). The information bound for θ is:

$$
I(\theta) = E[l_\theta^*(X)]^{\otimes 2}. \tag{4.11}
$$

In section 6 we show how this theorem can be reduced to the situation considered by Geskus and Groeneboom (1996c) (or to Geskus and Groeneboom (1996a, 1996b)).

4.3.2. Consistency.

We prove a consistency result for the accelerated life model with "case 2" interval censored data; again the generalization to "case k" is straightforward.

THEOREM 4.7. *(i) The true value of the regression parameter is $\theta_0 \in \Theta$, where Θ is a bounded subset of R^d.*
(ii) The distribution of Z is not concentrated on a hyperplane.
(iii) F_0 is continuous. Then

$$\widehat{\theta}_n \rightarrow_{a.s.} \theta_0, \tag{4.12}$$

and

$$F_n(t; \widehat{\theta}_n) \rightarrow_{a.s.} F_0(t) . \tag{4.13}$$

for almost all t except on a set with $G_1 + G_2$-measure zero where G_1 and G_2 are probability measures corresponding to the marginal distribution functions G_1 and G_2 of U and V, respectively. In particular, if F_0 and at least one of G_1 and G_2 are continuous, then

$$\sup_{-\infty<t<\infty} |F_n(t; \widehat{\theta}_n) - F_0(t)| \rightarrow_{a.s.} 0.$$

The proof of this theorem is given in section 6.

4.4. Computation.

We focus on the proportional hazards model. Computation of the maximum likelihood estimator in the proportional odds model can be done in a similar way. For one or two-dimensional θ, the maximum profile likelihood approach can be used to compute $(\widehat{\theta}_n, \widehat{\Lambda}_n)$. For high-dimensional θ, the iterative algorithm described in section 3.4 is needed. The likelihood function with "case 2" interval censoring enjoys similar properties as the likelihood function for "case 1" data. Specifically, for any fixed θ, $l_n(\theta, \Lambda)$ is concave in Λ, and for any fixed Λ, $l_n(\theta, \Lambda)$ is concave in θ. So steps (i) and (ii) of the algorithm described in section 3.4 are again well-defined concave maximization problems, and the algorithm converges. Proposition 3.1 continues to hold in the "case 2" interval censoring setting.

To implement step (i), we use the iterative convex minorant algorithm introduced by Groeneboom (1991). To describe it in the present setting, we first introduce some notation. Define functions a_i, $i = 1, 2, 3$, by

$$a_1(x; \Lambda) = \frac{e^{\theta' z} \exp(-\Lambda(u) e^{\theta' z})}{1 - \exp(-\Lambda(u) e^{\theta' z})},$$

$$a_2(x; \Lambda) = \frac{e^{\theta' z} \exp(-\Lambda(u) e^{\theta' z})}{\exp(-\Lambda(u) e^{\theta' z}) - \exp(-\Lambda(v) e^{\theta' z})},$$

and

$$a_3(x;\Lambda) = \frac{e^{\theta' z}\exp(-\Lambda(v)e^{\theta' z})}{\exp(-\Lambda(u)e^{\theta' z}) - \exp(-\Lambda(v)e^{\theta' z})}.$$

Let

$$\begin{aligned} W_\Lambda(t) &= \sum_{i=1}^n \{\delta_{1i}a_1(X_i,\Lambda) - \delta_{2i}a_2(X_i,\Lambda)\}1_{[U_i \le t]} \\ &\quad + \sum_{i=1}^n \{\delta_{2i}a_3(X_i,\Lambda) - \delta_{3i}e^{\theta' Z_i}\}1_{V_i \le t]}, \end{aligned}$$

$$\begin{aligned} G_\Lambda(t) &= \sum_{i=1}^n \{\delta_{1i}a_1^2(X_i,\Lambda) + \delta_{2i}a_2^2(X_i,\Lambda)\}1_{[U_i \le t]} \\ &\quad + \sum_{i=1}^n \{\delta_{2i}a_3^2(X_i,\Lambda) + \delta_{3i}e^{2\theta' Z_i}\}1_{V_i \le t]}, \end{aligned}$$

and

$$V_\Lambda(t) = W_\Lambda(t) + \int_{[0,t]} \Lambda(s)dG_\Lambda(s).$$

Then $\widehat{\Lambda}_n$ is the left derivative of the greatest convex minorant of the self-induced cumulative sum diagram formed by the points $P_0 = (0,\ 0)$ and

$$P_j = (G_{\widehat{\Lambda}_n}(Y_{(j)}),\ V_{\widehat{\Lambda}_n}(Y_{(j)})), \qquad j = 1, \ldots, m.$$

With this characterization, the iterative convex minorant algorithm can again be used in step (i). Finally, the Newton-Raphson method can be used in step (ii).

5. Further problems.

1. Regularity conditions. In the regression setting, it would be very desirable to relax conditions (A4) and (A5) for case 1 data, and conditions (B4) to (B6) for case 2 data. It appears that removing the positive separation hypothesis (B4a) is the most difficult task. Recently, Groeneboom and Geskus (1996) succeeded in proving asymptotic normality and efficiency of the NPMLE of a smooth functional of the distribution function in the nonparametric setting (with no covariates) in the case when the two examination times U and V are arbitrarily close and the joint density $h(u,v)$ is strictly positive along the diagonal $u = v$. The techniques developed by Groeneboom and Geskus (1996) can probably be extended to regression problems.

2. The accelerated failure time model with interval censoring. Although with case 1 data, efficient estimators have been constructed by Klein and Spady (1993), it seems that no efficient estimators with case 2 or general interval-censored data have been constructed in the literature. It is shown in theorem 4.6 that the Fisher information for θ with case 2 data is positive under appropriate conditions. This suggests that it is possible to construct efficient estimators of θ. We are not able to prove that the maximum profile likelihood estimator is asymptotically normal and efficient. The main difficulty is that the profile likelihood function is not smooth in θ. We are investigating regularized (or penalized) profile maximum likelihood estimation approaches with case 2 data. It seems likely that these approaches will yield efficient estimators.

3. Estimation of $I(\theta)$ and Confidence Sets for θ. In the case 2 or the general interval censoring, we lack an explicit formula for the information $I(\theta)$ and do not know how to estimate $I(\theta)$ directly. In section 4.1.3 we suggested the use of observed information or the curvature of the profile likelihood function as estimators of the information and hence of the asymptotic variance of the estimators. It is reasonable to conjecture is that these two approaches provide consistent estimators of $I(\theta)$, but this remains to be proved. For some progress in this direction, see Murphy and van der Vaart (1996a, 1996b).

4. Testing Hypotheses with interval-censored data. It would be desirable to be able to test hypotheses about θ; for example, consider testing $H : c'(\theta - \theta_0) = 0$ versus $K : c'(\theta - \theta_0) \neq 0$ for fixed vectors θ_0 and c. Furthermore, inversion of the family of likelihood ratio tests of the hypotheses $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$ provides confidence sets for θ in the usual way. [Note that this method apparently circumvents direct estimation of the asymptotic variances of estimators.] Likelihood ratio type tests for semiparametric models have recently been studied by Murphy and Van der Vaart (1996). Roughly their result asserts that the natural likelihood ratio test for the finite-dimensional parameter θ in a semiparametric model "works as would be expected from finite-dimensional parametric" theory if the corresponding maximum likelihood estimator $\widehat{\theta}_n$ is asymptotically normal and efficient. Murphy and Van der Vaart (1995) consider several examples, including the proportional hazards model with case 1 interval censored data (see our section 3.1 and Huang (1996a)). It remains to implement their proposal in a range of interval censoring models and to compare the resulting confidence intervals with those obtained by other methods, including those suggested in sections 3.1.3 and 4.1.3.

Score tests seem feasible and useful in all of the interval censoring regression models considered here. Tests of this type have been previously suggested in the literature (see e.g. Hoel and Walburg (1972) and Peto et al. (1980)), but apparently have not yet been studied thoroughly and still lack justification from the current viewpoint. Fay (1996) considered a class of rank invariant tests for interval-censored data which appears to be closely related to score tests. The methods of Huang (1994) and (1996) seem to apply directly to these natural score tests, and we intend to pursue this elsewhere.

Another interesting problem concerns nonparametric tests for two or k populations with interval-censored data. We are currently investigating test statistics generated by score functions based on certain regression models; see e.g. Huang, Wang, and Wellner (1996).

5. Regression models with time-dependent covariates. In this review we have restricted attention to regression models with covariates *not* depending on time. Time-dependent covariates also often arise in practice. Extension of the maximum likelihood approach considered in this article seems feasible and potentially useful. It may also be possible to extend the methods for proving distributional properties with time-independent covariates to the case of time-dependent covariates.

6. Small sample comparisons and refinements. We have emphasized maximum likelihod procedures in this review. It remains to carry out thorough small-sample comparisons of these methods with alternative procedures involving sieved maximum likelihood or smoothing methods. For a bit of work in this direction in the case of the proportional odds model, see Rossini and Tsiatis (1996). [Rossini and Tsiatis (1996) concluded that "Nonetheless, we did find that the automatic choice of intervals resulting by the use of the NPMLE behaved well, for the most part, over a variety of models and sample sizes."]

6. Appendix.

Proof of Theorem 3.1. Without loss of generality, we only prove the theorem for $Z \in R$. The general case can be proved similarly. We first compute the score function for θ and F. The score function for θ is simply the derivative of the log-likelihood with respect to θ, that is,

$$\dot{l}_\theta(x) = z e^{\theta z} \Lambda(u) \left[\delta \frac{\overline{F}(u|z)}{1 - \overline{F}(u|z)} - (1-\delta) \right].$$

Now suppose $\mathcal{F}_0 = \{F_\eta, |\eta| < 1\}$ is a regular parametric sub-family of $\mathcal{F} = \{F : F \ll \mu, \mu = \text{Lebesgue measure}\}$. Set $\frac{\partial}{\partial \eta} \log f_\eta(t)|_{\eta=0} = a(t)$, then $a \in L_2^0(F)$ and $\frac{\partial}{\partial \eta} \overline{F}_\eta(t)|_{\eta=0} = \int_t^\infty a dF$. The score operator for f is

$$\dot{l}_f(a)(x) \quad = \quad e^{\theta z} \frac{\int_u^\infty a dF}{\overline{F}(u)} \left[-\delta \frac{\overline{F}(u \mid z)}{1 - \overline{F}(u \mid z)} + (1-\delta) \right].$$

Let $Q(u, \delta, z)$ be defined by (3.5), then

$$\dot{l}_\theta(x) = z e^{\theta z} \Lambda(u) Q(u, \delta, z) \qquad \text{and} \qquad \dot{l}_f(a)(x) = -e^{\theta z} \frac{\int_u^\infty a dF}{\overline{F}(u)} Q(u, \delta, z).$$

With conditions (A2) and (A3), $\dot{l}_\theta$ is square integrable, and for any $a \in L_2(F)$, $\dot{l}_f(a)$ is square integrable. To calculate the efficient score $\dot{l}_\theta^*$ for θ, we need to find a function a_* so that

$$\dot{l}_\theta - \dot{l}_f a_* \perp \dot{l}_f a \text{ for all } a \in L_2^0(F), \tag{6.1}$$

that is, $E(\dot{l}_\theta - \dot{l}_f a_*)(\dot{l}_f a) = 0$. for all $a \in L_2^0(F)$. Let $r(z) = e^z$. Some calculation yields

$$\begin{aligned} -E(\dot{l}_\theta - \dot{l}_f a_*)(\dot{l}_f a) &= E_U \left\{ \frac{\int_U^\infty a dF}{\overline{F}(U)} E[r(2\theta Z) O(U \mid Z)[Z\Lambda(U) \right. \\ &\qquad \left. + \frac{\int_U^\infty a_* dF}{\overline{F}(U)}] \mid U] \right\}, \end{aligned}$$

where $O(U|Z)$ is defined by (3.6). Let

$$E\left\{ r(2\theta Z) O(U \mid Z)[Z\Lambda(U) + \frac{\int_U^\infty a_* dF}{\overline{F}(U)}] \mid U \right\} = 0,$$

then

$$\Lambda(U) E[r(2\theta Z) O(U \mid Z) Z \mid U] = -\frac{\int_U^\infty a_* dF}{\overline{F}(U)} E[r(2\theta Z) O(U \mid Z) \mid U],$$

thus with a_* determined by

$$\int_u^\infty a_* dF = -\frac{\Lambda(u)\overline{F}(u) E[Zr(2\theta Z) O(U \mid Z) | U = y]}{E[r(2\theta Z) O(U \mid Z) \mid U = y]},$$

(6.1) holds. Notice that a_* is only determined on the support of U. However, $\dot{l}_f a_*$ is a square integrable function with expectation zero. So the efficient score function for θ is

$$\begin{aligned} \dot{l}_\theta^*(x) &= \dot{l}_\theta(x) - (\dot{l}_f a_*)(x) \\ &= r(\theta z) Q(y, \delta, z) \Lambda(u) \left\{ z - \frac{E[(Zr(2\theta Z) O(U \mid Z) \mid U = y]}{E[(r(2\theta Z) O(U \mid Z) \mid U = y]} \right\}. \end{aligned}$$

The information for θ is

$$I(\theta) = E[\dot{l}_\theta^*(X)]^2 = E\left\{ R(U, Z) \left[Z - \frac{E(ZR(U, Z) \mid U)}{E(R(U, Z) \mid U)} \right]^2 \right\}, \tag{6.2}$$

where $R(U, Z) = \Lambda^2(U \mid Z) O(U \mid Z)$. □

Proof of Theorem 3.6. For simplicity, we will prove (3.14) and (3.15) for $\theta \in R$. The generalization to $\theta \in R^d$ is straightforward.

We first compute the score function for θ and F. The score function for θ is simply the derivative of the log-likelihood with respect to θ. That is,

$$\dot{l}_\theta(x) = -\delta\frac{f(u-z'\theta)z}{F(u-z'\theta)} + (1-\delta)\frac{f(u-z'\theta)z}{1-F(u-z'\theta)}.$$

Now suppose $\mathcal{F}_0 = \{F_\eta, |\eta| < 1\}$ is a regular parametric sub-family of $\mathcal{F} = \{F : F \ll \mu, \mu = \text{Lebesgue measure}\}$. Set

$$\frac{\partial}{\partial\eta}\log f_\eta(t)\mid_{\eta=0} = a(t).$$

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

$$\frac{\partial}{\partial\eta}F_\eta(t)|_{\eta=0} = \int_{-\infty}^t a dF, \quad \frac{\partial}{\partial\eta}\overline{F}_\eta(t)|_{\eta=0} = -\int_{-\infty}^t a dF.$$

The score operator for f is:

$$\dot{l}_f(a)(x) = \delta\frac{\int_{-\infty}^{u-z'\theta} a(t)dF(t)}{F(u-z'\theta)} - (1-\delta)\frac{\int_{-\infty}^{u-z'\theta} a(t)dF(t)}{1-F(u-z'\theta)}.$$

To calculate the information for θ in this semiparametric model, we follow the general theory of Bickel, Klaassen, Ritov, and Wellner (1993). We first need to compute the efficient score function $\dot{l}_\theta^*$ for θ. Geometrically, $\dot{l}_\theta^*$ can be interpreted as the residual of $\dot{l}_\theta$ projected in the space spanned by $\dot{l}_f a$, where $a \in L_2^0(F) = \{a : \int a dF = 0 \text{ and } \int a^2 dF < \infty\}$. Thus we need to find a function a_* with $\int a_* dF = 0$ so that $\dot{l}_\theta^* = \dot{l}_\theta - \dot{l}_f a_* \perp \dot{l}_f a$ for all $a \in L_2^0(F)$. That is

$$E(\dot{l}_\theta - \dot{l}_f a_*)(\dot{l}_f a) = 0 \tag{6.3}$$

for all $a \in L_2^0(F)$. Now we proceed to find a_* such that (6.3) is true. We have

$$\begin{aligned}
&\dot{l}_\theta(x) - (\dot{l}_f a_*)(x) \\
&\quad = \left[f(u-z'\theta)z + \int_{-\infty}^{u-z'\theta} a_*(t)dF(t)\right]\left[\frac{1-\delta}{1-F(u-z'\theta)} - \frac{\delta}{F(u-z'\theta)}\right].
\end{aligned}$$

Thus

$$\begin{aligned}
&-E((\dot{l}_\theta - \dot{l}_f a_*)(\dot{l}_f a)) \\
&\quad = E\left\{\left[\frac{1-\delta}{1-F(U-Z'\theta)} - \frac{\delta}{F(U-Z'\theta)}\right]^2\right. \\
&\qquad \left.\times\left[f(U-Z'\theta)Z + \int_{-\infty}^{U-Z'\theta} a_*(t)dF(t)\right]\int_{-\infty}^{U-Z'\theta} a(t)dF(t)\right\}
\end{aligned}$$

$$= E\left\{\left[f(U-Z'\theta)Z + \int_{-\infty}^{U-Z'\theta} a_*(t)dF(t)\right] \times \left[\frac{1}{1-F(U-Z'\theta)} + \frac{1}{F(U-Z'\theta)}\right]\int_{-\infty}^{U-Z'\theta} a(t)dF(t)\right\}.$$

So, with $U_1 \equiv U - Z'\theta$, we can set

$$E\left[\left(f(U_1)Z + \int_{-\infty}^{U_1} a_*(t)dF(t)\right)|U_1\right] = 0$$

to ensure that (6.3) is true. We obtain a_* by solving the following equation:

$$f(s)E(Z|U - Z'\theta = s) + \int_{-\infty}^{s} a_*(t)dF(t) = 0.$$

In other words, we can choose any a_* that satisfies the above equation. In particular, if $f(s)$ and $k(s) = E(Z|U - Z'\theta = s)$ are differentiable, and $f(s) > 0$ for any $s \in R$, then we have an explicit expression for a_*:

$$a_*(s) = -k'(s) - k(s)\frac{f'(s)}{f(s)}.$$

By the assumptions, we have

$$\int a_*(t)dF(t) = \lim_{s\to\infty}\int_{-\infty}^{s} a_*(t)dF(t) = \lim_{s\to\infty} f(s)E(Z|U - Z'\theta = s) = 0.$$

It follows that the efficient score function for θ is

$$\begin{aligned} l^*_\theta(x) &= \dot{l}_\theta(x) - \dot{l}_f a_*(x) \\ &= f(u - z'\theta)[z - E(Z|U - Z'\theta = u - z'\theta)] \\ &\qquad \times \left[\frac{1-\delta}{1-F(u-z'\theta)} - \frac{\delta}{F(u-z'\theta)}\right]. \end{aligned}$$

The information for θ is

$$\begin{aligned} I(\theta) &= E[\dot{l}^*_\theta(X)]^2 \\ &= E\left\{\left[\frac{f(U-Z'\theta)^2}{F(U-Z'\theta)(1-F(U-Z'\theta))}\right][Z - E(Z|U-Z'\theta)]^2\right\}. \end{aligned}$$

Hence the information for θ is positive unless $Z = E(Z|U - Z'\theta)$ with probability one. □

Proof of Theorem 4.1. Recall that a_1, a_2 and a_3 are defined in section 4.4. By conditions (A3) and (A4), a_1, a_2 and a_3 are positive functions of

(u, v, z). To make formulas shorter, we will drop the arguments of a's in the following. Let $x = (\delta, \gamma, u, v, z)$, and let $l(x; \theta, \Lambda)$ be the logarithm of the density function of X. The score function for θ is

$$\dot{l}_\theta(x) = \frac{\partial}{\partial \theta} l(x; \theta, \Lambda) = z e^{\theta' z}\{\delta \Lambda(u) a_1 - \gamma[\Lambda(u) a_2 - \Lambda(v) a_3] - (1 - \delta - \gamma)\Lambda(v)\}.$$

The score operator for Λ is

$$\begin{aligned} \dot{l}_\Lambda a(x) &= \left.\frac{\partial}{\partial s} l(x; \theta, \Lambda_s)\right|_{s=0} \\ &= e^{\theta' z}\{\delta h(u) a_1 - \gamma[h(u) a_2 - h(v) a_3] - (1 - \delta - \gamma) h(v)\} \end{aligned}$$

where $h \equiv h[a]$ is defined by (4.4). In the following, we carry out the calculation for $\theta \in R$. For the general case $\theta \in R^d$, we only need to repeat the same calculation for each component of $\dot{l}_\theta(x)$.

The score operator $\dot{l}_\Lambda$ maps $L_2^0(F)$ to $L_2^0(P)$, where P is the joint probability measure of $(\delta, \gamma, U, V, Z)$ and $L_2^0(P)$ is defined similarly as $L_2^0(F)$. Let $\dot{l}_\Lambda^T : L_2^0(P) \to L_2^0(F)$ be the adjoint operator of $\dot{l}_\Lambda$, i.e., for any $a \in L_2^0(F)$ and $b \in L_2^0(P)$, define h as in (4.7),

$$\langle b, \dot{l}_\Lambda a\rangle_P = \langle \dot{l}_\Lambda^T b, a\rangle_F \ ,$$

where $\langle \cdot, \cdot\rangle_P$ and $\langle \cdot, \cdot\rangle_F$ are the inner products in $L_2^0(P)$ and $L_2^0(F)$, respectively. We need to find a_* such that $\dot{l}_\theta - \dot{l}_\alpha(a_*)$ is orthogonal to $\dot{l}_\alpha(a)$ in $L_2^0(P)$. This amounts to solving the following normal equation:

$$\dot{l}_\Lambda^T \dot{l}_\Lambda(a_*) = \dot{l}_\Lambda^T \dot{l}_\theta. \tag{6.4}$$

We know that

$$\dot{l}_\Lambda^T \dot{l}_\Lambda(a)(t) = E[\dot{l}_\Lambda(a)(X)|T = t] = E_Z E[\dot{l}_\Lambda(a)(X)|T = t, Z];$$

see e.g. BKRW (1993), pages 271-272, or Groeneboom and Wellner (1992), pages 8 and 9. By conditions (A1) and (A4),

$$\begin{aligned} & E[\dot{l}_\Lambda(a)(X)|T = t, Z = z] \\ = & \int_{u=t}^{\tau_1} \int_{v=u+\eta}^{\tau_1} e^{\theta' z} h(u) a_1 g(u, v|z) dv du \\ & - \int_{u=\tau_0}^{t} \int_{v=t}^{\tau_1} e^{\theta' z} (h(u) a_2 - h(v) a_3) g(u, v|z) 1_{[v-u \geq \eta]} dv du \\ & - \int_{u=\tau_0}^{t} \int_{v=u+\eta}^{t} e^{\theta' z} h(v) g(u, v|z) du dv, \end{aligned}$$

where $g(u, v|z)$ is the conditional density of (U, V) given Z. Let

$$b_1(u, v) = E_Z[e^{\theta' Z} a_1(u, v, Z) g(u, v|Z)],$$

$$b_2(u,v) = E_Z[e^{\theta' Z} a_2(u,v,Z) g(u,v|Z)],$$
$$b_3(u,v) = E_Z[e^{\theta' Z} a_3(u,v,Z) g(u,v|Z)],$$

and

$$b_4(u,v) = E_Z[e^{\theta' Z} g(u,v|Z)].$$

By the their definition, all the functions A_j and B_j are positive functions, and

$$b_2(u,v) = b_3(u,v) + b_4(u,v). \tag{6.5}$$

We calculate

$$\begin{aligned} L(t) &\equiv \dot{l}_\Lambda^T \dot{l}_\Lambda(a)(t) \\ &= \int_{u=t}^{\tau_1} \int_{v=u+\eta}^{\tau_1} h(u) b_1(u,v) dv du \\ &\quad - \int_{u=\tau_0}^{t} \int_{v=t}^{\tau_1} [h(u) b_2(u,v) - h(v) b_3(u,v)] 1_{[v-u \geq \eta]} dv du \\ &\quad - \int_{u=\tau_0}^{t} \int_{v=u+\eta}^{t} h(v) b_4(u,v) dv du; \end{aligned}$$

recall that h is defined in terms of a by (4.4). Let

$$c_1(u,v) = E_Z[Z e^{\theta' Z} a_1(u,v,Z) g(u,v|Z)],$$
$$c_2(u,v) = E_Z[Z e^{\theta' Z} a_2(u,v,Z) g(u,v|Z)],$$
$$c_3(u,v) = E_Z[Z e^{\theta' Z} a_3(u,v,Z) g(u,v|Z)],$$

and

$$c_4(u,v) = E_Z[Z e^{\theta' Z} g(u,v|Z)].$$

Then further calculation yields

$$\begin{aligned} R(t) \equiv \dot{l}_\Lambda^T \dot{l}_\theta(t) &= \int_{u=t}^{\tau_1} \int_{v=u+\eta}^{\tau_1} \Lambda(u) c_1(u,v) dv du \\ &\quad - \int_{u=\tau_0}^{t} \int_{v=t}^{\tau_1} [\Lambda(u) c_2(u,v) - \Lambda(v) c_3(u,v)] 1_{[v-u \geq \eta]} dv du \\ &\quad - \int_{u=\tau_0}^{t} \int_{v=u+\eta}^{t} \Lambda(v) c_4(u,v) dv du. \end{aligned}$$

Let

$$b(t) = \int_{t+\eta}^{\tau_1} b_1(t,x) dx + \int_{t+\eta}^{\tau_1} b_2(t,x) dx + \int_{\tau_0}^{t-\eta} b_3(x,t) dx + \int_{\tau_0}^{t-\eta} b_4(x,t) dx.$$

After some straightforward calculations, the derivative of $L(t)$ is

$$L'(t) = -b(t) h(t) + \int_{\tau_0}^{t-\eta} h(x) b_2(x,t) dx + \int_{t+\eta}^{\tau_1} h(x) b_3(t,x) dx.$$

Similarly, let

$$c(t) = \int_{t+\eta}^{\tau_1} c_1(t,x)dx + \int_{t+\eta}^{\tau_1} c_2(t,x)dx + \int_{\tau_0}^{t-\eta} c_3(x,t)dx + \int_{\tau_0}^{t-\eta} c_4(x,t)dx.$$

The derivative of $R(t)$

$$r(t) \equiv R'(t) = -c(t)\Lambda(t) + \int_{\tau_0}^{t-\eta} \Lambda(x)c_2(x,t)dx + \int_{t+\eta}^{\tau_1} \Lambda(x)c_3(t,x)dx.$$

By conditions (A3)–(A5), r has a bounded derivative r' on $[\tau_0, \tau_1]$. So equation (6.4) reduces to

$$-b(t)h(t) + \int_{\tau_0}^{t-\eta} h(x)b_2(x,t)dx + \int_{t+\eta}^{\tau_1} h(x)b_3(t,x)dx = r(t). \qquad (6.6)$$

By conditions (A3) and (A4), $\inf_{\tau_0 \le t \le \tau_1} b(t) > 0$. Let $d(t) = -r(t)/b(t)$ and

$$K(t,x) = [b_2(x,t)1_{[\tau_0 \le x \le t-\eta]} + b_3(t,x)1_{[t+\eta \le x \le \tau_1]}]/b(t).$$

We obtain a Fredholm integral equation of the second kind,

$$h_*(t) - \int K(t,x)h_*(x)dx = d(t).$$

Since $K(t,x)$ is a bounded kernel (K being a L_2 kernel suffices), by lemma 6.1 below and the classical results on Fredholm integral equations (Kanwal 1971), Sections 4.2 and 4.3; or Kress (1989), Chapter 4), there exists a resolvent $\Gamma(t,x)$ (completely determined by K) such that

$$h_*(t) = d(t) + \int \Gamma(t,x)d(x)dx. \qquad (6.7)$$

From equations (6.6) and (6.7), we can derive properties of h_*. By (6.7), h_* is bounded on $[\tau_0, \tau_1]$. By (6.6), this implies h_* is continuous. This in turn implies h_* is differentiable. Since b is bounded away from zero, the partial derivative of $b_2(x,t)$ with respect to t is bounded on $[\tau_0, t-\eta]$, the partial derivative of $b_3(t,x)$ with respect to t is bounded on $[t+\eta, \tau_1]$, and the derivative of r is bounded, it follows that the derivative of h_* is bounded. □

The following lemma is used in the proof of theorem 4.1.

Lemma 6.1. *If*

$$h(t) - \int K(t,x)h(x)dx = 0,$$

then $h(t) \equiv 0$ *on* $[\tau_0, \tau_1]$.

Proof of Lemma 6.1. According to the definition of K, the equation can be written as

$$\begin{aligned}
&\int_{t+\eta}^{\tau_1} B_3(t,x)h(x)dx + \int_{\tau_0}^{t-\eta} B_2(x,t)h(x)dx \\
&= h(t)\left[\int_{t+\eta}^{\tau_1} B_1(t,x)dx + \int_{t+\eta}^{\tau_1} B_2(t,x)dx \right. \\
&\qquad \left. + \int_{\tau_0}^{t-\eta} B_3(x,t)dx + \int_{\tau_0}^{t-\eta} B_4(x,t)dx\right]. \qquad (6.8)
\end{aligned}$$

If there exists a point x_0 such that $h(x_0) > 0$, let y be the point in $[\tau_0, \tau_1]$ such that $h(u) = \sup_{\tau_0 \le x \le \tau_1} h(x)$, then by equation (6.5),

$$\begin{aligned}
&\int_{u+\eta}^{\tau_1} h(x)B_3(u,x)dx + \int_{\tau_0}^{u-\eta} h(x)B_2(x,y)dx \\
&\le h(u)\left[\int_{u+\eta}^{\tau_1} B_3(u,x)dx + \int_{\tau_0}^{u-\eta} B_2(x,y)dx\right] \\
&< h(u)\left[\int_{u+\eta}^{\tau_1} B_2(u,x)dx + \int_{\tau_0}^{u-\eta} B_2(x,y)dx\right] \\
&\le h(u)\left[\int_{u+\eta}^{\tau_1} B_1(u,x)dx + \int_{u+\eta}^{\tau_1} B_2(u,x)dx + \int_{\tau_0}^{u-\eta} B_2(x,y)dx\right] \\
&\le h(u)\left[\int_{u+\eta}^{\tau_1} B_1(u,x)dx + \int_{u+\eta}^{\tau_1} B_2(u,x)dx \right. \\
&\qquad \left. + \int_{\tau_0}^{u-\eta} B_3(x,y)dx + \int_{\tau_0}^{u-\eta} B_4(x,y)dx\right].
\end{aligned}$$

This is a contradiction to equation (6.8). Since if h satisfies (6.8), $-h$ also satisfies (6.8), so there can not be a point x_0 such that $h(x_0) < 0$. □

Proof of Theorem 4.6. For simplicity, we will carry out the computation for one-dimensional $\theta \in R$. For $\theta \in R^d$, we only need to repeat the same calculation for each component of θ.

The score function for θ is simply the derivative of the log-likelihood with respect to θ. That is,

$$\begin{aligned}
\dot{l}_\theta(x) &= \left(-\delta_1 \frac{f(u-z'\theta)}{F(u-z'\theta)} - \delta_2 \frac{f(v-z'\theta) - f(u-z'\theta)}{F(v-z'\theta) - F(u-z'\theta)}\right. \\
&\qquad \left. + \delta_3 \frac{f(v-z'\theta)}{1-F(v-z'\theta)}\right) z. \qquad (6.9)
\end{aligned}$$

Now suppose $\mathcal{F}_0 = \{F_\eta, |\eta| < 1\}$ is a regular parametric sub-family of $\mathcal{F} = \{F : F \ll \mu, \mu = \text{Lebesgue measure}\}$. Set

$$\frac{\partial}{\partial \eta} \log f_\eta(t) \mid_{\eta=0} = a(t).$$

Then $a \in L_2^0(F) \equiv \{a : \int a dF = 0,$ and $\int a^2 dF < \infty$, and, with $\overline{F} \equiv 1 - F$,

$$\frac{\partial}{\partial \eta} F_\eta(t)|_{\eta=0} = \int_{-\infty}^t a dF, \qquad \frac{\partial}{\partial \eta} \overline{F}_\eta(t)|_{\eta=0} = -\int_{-\infty}^t a dF.$$

Let $\phi(t) = \int_{-\infty}^t a(s) dF(s)$. The score operator for f is:

$$\dot{l}_f(a)(x) = \delta_1 \frac{\phi(u - z'\theta)}{F(u - z'\theta)} + \delta_2 \frac{\phi(v - z'\theta) - \phi(u - z'\theta)}{F(v - z'\theta) - F(u - z'\theta)} - \delta_3 \frac{\phi(v - z'\theta)}{1 - F(v - z'\theta)}. \tag{6.10}$$

The score operator $\dot{l}_f$ maps $L_2^0(F)$ to $L_2^0(P)$, where P is the joint probability measure of $(\delta_1, \delta_2, U, V, Z)$ and $L_2^0(P)$ is defined similarly as $L_2^0(F)$. Let $\dot{l}_f^T : L_2^0(P) \to L_2^0(F)$ be the adjoint operator of $\dot{l}_f$, i.e., for any $a \in L_2^0(F)$ and $b \in L_2^0(P)$,

$$\langle b, \dot{l}_f a \rangle_P = \langle \dot{l}_f^T b, a \rangle_F \ ,$$

where $\langle \cdot, \cdot \rangle_P$ and $\langle \cdot, \cdot \rangle_F$ are the inner products in $L_2^0(P)$ and $L_2^0(F)$, respectively. We need to find a_* such that $\dot{l}_\theta - \dot{l}_f(a_*)$ is orthogonal to $\dot{l}_\alpha(h)$ in $L_2^0(P)$. This amounts to solving the following normal equation:

$$\dot{l}_f^T \dot{l}_f(a_*) = \dot{l}_f^T \dot{l}_\theta. \tag{6.11}$$

The value of $\dot{l}_f^T$ at any $b \in L_2^0(P)$ can be computed by

$$\begin{aligned} \dot{l}_f^T b(t) &= E[b(X)|T = t] \\ &= E_Z E[b(X)|T = t, Z] = E_Z E[b(X)|T - Z'\theta = t - Z'\theta, Z]; \end{aligned}$$

see e.g. BKRW (1993), pages 271-272, or Groeneboom and Wellner (1992), pages 8 and 9.

Let h_1 be the joint density of (U_1, V_1) where $U_1 = U - Z'\theta, V_1 = V - Z'\theta$. Let $s = t - z'\theta$. Then, with $\phi(t) \equiv \int_{-\infty}^t a dF$ as before,

$$\begin{aligned} \dot{l}_f^T \dot{l}_f a(t) = &\int_{u_1=t}^\infty \int_{v_1=u_1}^\infty \frac{\phi(u_1)}{F(u_1)} h_1(u_1, v_1) dv_1 du_1 \\ &+ \int_{u_1=-\infty}^t \int_{v_1=t}^\infty \frac{\phi(v_1) - \phi(u_1)}{F(v_1) - F(u_1)} h_1(u_1, v_1) dv_1 du_1 \\ &- \int_{v_1=-\infty}^t \int_{u_1=-\infty}^{v_1} \frac{\phi(v_1)}{1 - F(v_1)} h_1(u_1, v_1) du_1 dv_1. \end{aligned}$$

Similarly, let

$$\begin{aligned} h_2(u_1, v_1) &= \int z h(u_1 + z'\theta, v_1 + z'\theta, z) d\mu(z) \\ &= E\{Z | U_1 \equiv U - Z'\theta = u_1, V_1 \equiv V - Z'\theta = v_1\} h_1(u_1, v_1)\,. \end{aligned}$$

Then

$$\dot{l}_f^T \dot{l}_\theta(t) = -\int_{u_1=t}^{\infty}\int_{v_1=u_1}^{\infty} \frac{f(u_1)}{F(u_1)} h_2(u_1,v_1)dv_1 du_1 - \int_{u_1=-\infty}^{t}\int_{v_1=t}^{\infty} \frac{f(v_1)-f(u_1)}{F(v_1)-F(u_1)} h_2(u_1,v_1)dv_1 du_1 + \int_{v_1=-\infty}^{t}\int_{u_1=-\infty}^{v_1} \frac{f(v_1)}{1-F(v_1)} h_2(u_1,v_1)du_1 dv_1 \,.$$

Let h_{11} and h_{12} be the marginal densities of h_1. Using these two expressions and differentiating both sides of (6.11), we find, after multiplying across by $-d$, that

$$\begin{aligned} \phi(t) \quad &+ \quad d(t)\left[\int_{-\infty}^{t} \frac{\phi(t)-\phi(u)}{F(t)-F(u)} h_1(u,t)du - \int_{t}^{\infty} \frac{\phi(u)-\phi(t)}{F(u)-F(t)} h_1(t,u)du\right] \\ &= \quad -\, d(t)g(t) \end{aligned} \tag{6.12}$$

where

$$d(t) = \frac{F(t)(1-F(t))}{h_{11}(t)[1-F(t)] + h_{12}(t)F(t)},$$

$g(t)$ is the derivative of $\dot{l}_f^T \dot{l}_\theta(t)$, and $g(t)F(t)(1-F(t))$ is given by

$$g(t)F(t)\overline{F}(t) = \{h_{21}(t)[1-F(t)] + h_{22}(t)F(t)\}f(t) + F(t)\overline{F}(t)\left[\int_{u=-\infty}^{t} \frac{\phi(t)-\phi(u)}{F(t)-F(u)} h_2(u,t)du - \int_{u=t}^{\infty} \frac{\phi(u)-\phi(t)}{F(u)-F(t)} h_2(t,u)du\right]$$

with $h_{21}(u_1) = \int_{u_1}^{\infty} h_2(u_1,v_1)dv_1$ and $h_{22}(v_1) = \int_{-\infty}^{v_1} h_2(u_1,v_1)du_1$.

Equation (6.12) has exactly the same structure as the integral equation involved in computing the information for a smooth functional of the distribution function studied in Geskus and Groeneboom (1996a,b,c).

In particular, Geskus and Groeneboom (1996c) showed that (6.12) has a unique solution $\phi_*(s)$ and that the Radon-Nikodyn derivative $d\phi_*/dF$ is a.e.$-[F]$ bounded under hypotheses which are implied by our assumptions (D1) - (D4); see their Theorem 2.2 and Corollary 2.1. This implies (6.11) has a unique solution. □

Proof of Theorem 4.7. We apply Theorem 3.1 of Van de Geer (1993). Let

$$p(x;\theta,F) \quad = \quad F(u-\theta' z)^{\delta_1}[F(v-\theta' z) - F(u-\theta' z)]^{\delta_2}(1-F(v-\theta' z))^{\delta_3}.$$

Denote the class of functions

$$\mathcal{H} = \{p(x; \theta, F) : (\theta, F) \in \Theta \times \mathcal{F}\}.$$

Clearly $\mathcal{H}$ is uniformly bounded by 1. Here we change notation from section 4 slightly by letting μ be the product of counting measure on $\{(0,0), (0,1), (1,0)\}$ and the joint measure induced by the joint distribution H of (U, V, Z). Then

$$\int p(x; \theta_0, F_0) d\mu(x) \leq 1.$$

Furthermore, the class $\mathcal{H}$ is a VC-hull class; see Dudley (1987), or Van der Vaart and Wellner (1996), section 2.6.3 and lemma 2.6.19, page 145. Thus the entropy condition of Theorem 3.1 of Van de Geer (1993) are satisfied. This implies

$$\int \left(\sqrt{p(x; \widehat{\theta}_n, \widehat{F}_n)} - \sqrt{p(x; \theta_0, F_0)} \right)^2 d\mu(x) \to_{a.s.} 0.$$

This in turn implies,

$$\int \left(\sqrt{\widehat{F}_n(u - \widehat{\theta}_n{}' z; \widehat{\theta}_n)} - \sqrt{F_0(u - \theta_0' z)} \right)^2 dG(u, v, z) \to_{a.s.} 0.$$

Changing variables in the integration and integration over v yields

$$\int \left(\sqrt{\widehat{F}_n(s; \widehat{\theta}_n)} - \sqrt{F_0(u + \widehat{\theta}_n - \theta_0' z)} \right)^2 dG_{13}(s + \widehat{\theta}_n{}' z, z) \to_{a.s.} 0 \quad (6.13)$$

where G_{13} is the marginal distribution of (U, Z). Since Θ is bounded, for any subsequence of $\widehat{\theta}_n$, we can find a further subsequence converging to $\theta_* \in \overline{\Theta}$, the closure of Θ. On the other hand, by Helly's selection theorem, for any subsequence of $F_n(\cdot; \widehat{\theta}_n)$, we can find a further subsequence converging in distribution to some subdistribution function F_*; i.e. pointwise convergence at continuity points of F_*. Apparently, we can choose the convergent subsequence of $\widehat{\theta}_n$ and the convergent subsequence of $F_n(\cdot; \widehat{\theta}_n)$ so that they have the same indices. Without causing confusion, we assume that $\widehat{\theta}_n$ converges to θ_* and that $F_n(\cdot; \widehat{\theta}_n)$ converges to F_*. To prove the theorem, it suffices to prove that $\theta_* = \theta_0$ and $F_* = F_0$. By continuity of F_0 and (6.13), F_* and θ_* satisfy

$$F_*(u) = F_0(u + \theta_*' z - \theta_0' z) \quad \text{for } G-\text{almost all } (u, z).$$

Thus

$$\theta_*' z = \theta_0' z \quad \text{and} \quad F_*(u) = F_0(u) \quad G-\text{almost all } (u, z).$$

Condition (ii) implies $\theta_* = \theta_0$. (4.12) follows. Similarly, we have

$$F_*(v) = F_0(v) \quad G_2 - \text{almost all } v,$$

and (4.13) follows. □

ACKNOWLEDGEMENTS: The authors owe thanks to Piet Groeneboom and Tony Rossini for many helpful discussions about interval censoring. We also owe thanks to John Crowley for pointing out several references in connection with testing with interval censored data. The research of Jon A. Wellner was supported in part by National Science Foundation grant DMS-9306809 and by NIAID grant 2R01 AI291968-04.

REFERENCES

Aragón, J. and Eberly D. (1992). On convergence of convex minorant algorithms for distribution estimation with interval-censored data. *J. of Computational and Graphical Stat.* **1**, 129-140.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Stat. in Medicine* **2**, 273 - 277.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore.

Chamberlain, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *J. Econometrics* **32**, 189-218.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25**, 573 - 578.

Cosslett, S. R. (1983) Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica* **51**, 765-782.

Cosslett, S. R. (1987). Efficiency bounds for distribution-free estimators of the binary choice and censored regression models. *Econometrica* **55**, 559-585.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society. B* **34**, 187-220.

De Gruttola, V. and Lagakos, S. W. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* **45**, 1-11.

Diamond, I. D. and McDonald, J. W. (1991). Analysis of current status data. In *Demographic Applications of Event History Analysis*, eds. J. Trussell, R. Hankinson, and J. Tilton, Oxford, U.K., Oxford University Press, 231-252.

Diamond, I. D. and McDonald, J. W. and Shah, I. H. (1986). Proportional hazards models for current status data: application to the study of differentials in age at weaning in Pakistan. *Demography* **23**, 607-620.

Dinse, G. E. and Lagakos, S. W. (1983). Regression analysis of tumour prevalence data. *Applied Statistics* **32**, 236 - 248. (Corr. **33**, 79-80.)

Fay, M. (1996). Rank invariant tests for interval censored data under the grouped continuous model. *Biometrics* **52**, to appear.

Finkelstein, D. M. and Wolfe R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41**, 933-945.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42**, 845-854.

Geskus, R. and Groeneboom, P. (1996a). Asymptotically optimal estimation of smooth functionals for interval censoring, part 1. *Statistica Neerl.* **50**, to appear. Technical Report, Delft University of Technology.

Geskus, R. and Groeneboom, P. (1996b). Asymptotically optimal estimation of smooth functionals for interval censoring, part 2. *Statistica Neerl.* **50**, to appear. Technical Report, Delft University of Technology.

Geskus, R. and Groeneboom, P. (1996c). Asymptotically optimal estimation of smooth functionals for interval censoring, case 2. Technical Report, Delft University of Technology. Submitted to *Ann. Statist.*

Gill, R. D. and Levit, B. Y. (1995). Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli* **1**, 59 - 79.

Groeneboom, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II, Lucien M. LeCam and Richard A. Olshen eds.

Groeneboom, P. (1987). Asymptotics for interval censored observations. Technical Report 87-18, Department of Mathematics, University of Amsterdam.

Groeneboom, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution. *Technical Report 378*, Department of Statistics, Stanford University.

Groeneboom, P. (1996). Inverse problems in statistics. Proceedings of the St. Flour Summer School in Probability, 1994. *Lecture Notes in Math.*, to appear. Springer Verlag, Berlin.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* DMV Seminar Band 19, Birkhäuser, Basel.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model. *J. Econometrics*, **35**, 303-316.

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

Huang, J. (1993). Maximum scored likelihood estimation of a linear regression with interval censored data. *Technical Report* **256**, Department of Statistics, University of Washington.

Huang, J. (1995). Maximum Likelihood Estimation for Proportional Odds Regression Model with Current Status Data. *Analysis of Censored Data*, IMS Lecture Notes - Monograph Series **27**, 129 - 146. H.L. Koul and J. V. Deshpande, editors. IMS, Hayward.

Huang, J. (1996). Efficient estimation for the Cox model with interval censoring. *Ann. Statist.* **24**, 540 - 568.

Huang, J. and Rossini, A. J. (1995). Sieve estimation for the proportional odds failure-time regression model with interval censoring. Technical Report, Department of Statistics, University of Iowa. Submitted to *J. Amer. Statist. Assoc.*.

Huang, J., Wang, K., and Wellner, J. A. (1996). Nonparametric tests with interval-censored data. Submitted to *J. Amer. Statist. Assoc.*.

Huang, J. and Wellner, J. A. (1995a). Asymptotic Normality of the NPMLE of Linear Functionals for Interval Censored Data, Case 1. *Statistica Neerlandica* **49**, 153 - 163.

Huang, J. and Wellner, J. A. (1995b). Efficient estimation for the proportional hazards model with "case 2" interval censoring. *Technical Report No.* **290**, Department of Statistics, University of Washington, Seattle. Submitted to *Biometrika*.

Huang, J. and Wellner, J. A. (1996). Regression models with interval censoring. To appear, *Proceedings of the Kolmogorov Seminar, Euler Mathematics Institute,* St. Petersburg, Russia. *Technical Report* **261**, Department of Statistics, University of Washngton.

Jewell, N. P. , Malani, H. M. and Vittinghoff, E. (1994) Nonparametric estimation for a form of doubly censored data, with application to two problems in AIDS. *Journal of the American Satistical Association* **89**, 7-18.

Jongbloed, G. (1995a). *Three Statistical Inverse Problems.* Unpublished Ph.D. dissertation, Delft University of Technology, Delft.

Jongbloed, G. (1995b). The iterative convex minorant algorithm for nonparametric estimation. Submitted to *J. of Computational and Graphical Stat.*

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18**, 191-219.

Klein, R. W. and Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica* **61**, 387-421.

Kongerud, J. and Samuelsen, S.O. (1991). A longitudinal study of respiratory symptoms in aluminum potroom workers. *American Review of Respiratory Diseases,* **144**, 10-16.

Kress, R. (1989). *Linear Integral Equations.* Springer-Verlag, Berlin.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *J. Econometrics* **3**, 205-228.

Manski, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *J. Econometrics* **27**, 313-333.

Murphy, S. A. and Van der Vaart, A. W. (1996a). Semiparametric likelihood ratio inference. Technical Report, Pennsylvania State University.

Murphy, S. A. and Van der Vaart, A. W. (1996b). Observed information in semiparametric models. Preprint.

Peto, R., Pike, M., Day, N., Gray, R., Lee, P., Parish, S., Peto, J., Richards, S., and Wahrendorf, J. (1980). Guidelines for simple, sensitive significance tests for carcinogenic effects in long term animal experiments. In *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, Supplement 2, Long Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal*, pages 311 - 346. International Agency for Research in Cancer, Lyon.

Parzen, M. (1993). *Computational Issues in Robust Regression.* Ph. D. dissertation, Harvard University.

Pettitt, A. N. (1984). Proportional odds models for survival data and estimates using ranks. *Applied Statistics* **33**, 169 - 175.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22**, 2031 - 2050.

Rabinowitz, D., Tsiatis, A., and Aragon, J. (1995). Regression with interval censored data. *Biometrika* **82**, 501 - 514.

Robertson, T., Wright, F. T., Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* Wiley, New York.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton University Press.

Rossini, A. J. and Tsiatis, A. A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *J. Amer. Statist. Assoc.* **91**, 713 - 721.

Samuelsen, S. O. and Kongerud, J. (1993). Evaluation of applying interval censoring on longitudinal data on Asthmatic symptoms. *Statistical Research Report* **2**, Institute of Mathematics, University of Oslo.

Self, S. G. and Grossman, E. A. (1986). Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics* **42**, 521-530.

Shiboski, S. C. and Jewell, N. P. (1992). Statistical analysis of the time dependence of HIV infectivity based on partner study data. *J. Amer. Statist. Assoc.* **87**, 360 - 372.

Silverman, B. M. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York.

Stone, C. J. (1977). Consistent nonparametric regression (with discussion). *Ann. Statist.* **5**, 595-645.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist. Soc. B* **38**, 290 - 295.

Van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21**, 14-44.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer Verlag, New York.

Van Eeden, C. (1956). Maximum likelihood estimation of ordered probabilities. *Proc. Kon. Ned. Akad. Wet.* **A59**, 444 - 455.

Van Eeden, C. (1957). Maximum likelihood estimation of partially ordered or completely ordered probabilities, I. *Proc. Kon. Ned. Akad. Wet.* **A60**, 128 - 136.

Wellner, J. A. (1995). Interval censoring case 2: alternative hypotheses. *Analysis of Censored Data*, IMS Lecture Notes - Monograph Series **27**, 271 - 291. H.L. Koul and J. V. Deshpande, editors. IMS, Hayward.

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the NPMLE from censored data. *J. Amer. Statist. Assoc.* **92**, to appear.

Zhan, Y. and Wellner, J. A. (1995). Double censoring: characterization and computation of the nonparametric maximum likelihood estimator. *Technical Report No.* **292**, Department of Statistics, University of Washington.

Department of Statistics
and Acturial Science
University of Iowa
Iowa City, Iowa 52242
U.S.A.
e-mail: jian@stat.uiowa.edu

University of Washington
Statistics
Box 354322
Seattle, Washington 98195-4322
U.S.A.
e-mail: jaw@stat.washington.edu