

Maximum Likelihood Estimation for Proportional Odds Regression Model with Current Status Data

Jian Huang ¹

Dept. of Statistics & Actuarial Sci., Univ. of Iowa, Iowa City, IA 52242

Abstract The maximum likelihood estimator (MLE) for the semiparametric proportional odds regression model with current status data is studied. It is shown that the MLE for the regression parameter is asymptotically normal and asymptotically efficient, even though the MLE for the baseline log-odds function only converges at $n^{1/3}$ rate.

1. Introduction. The proportional odds regression model is an interesting alternative to the widely used Cox's (1972) proportional hazards regression model. This model has been used by several authors in analyzing survival data, see for example, Bennett (1983), Dinse and Lagakos (1983), Pettitt (1984) and Parzen (1993). It specifies that

$$\text{logit}F(t|z) = \text{logit}F_0(t) + \beta'z, \quad (1.1)$$

where $\beta \in R^d$ is the regression parameter, $F(t|z)$ is the probability that the failure time is less than or equal to t given that the value of the covariate Z is z , i.e, $F(t|z) = P(T \leq t|Z = z)$, and $F_0(t) \equiv F(t|0)$ is the baseline distribution function. The logit function is defined by $\text{logit}(x) = \log(x/(1-x))$ for $0 < x < 1$. For simplicity, denote $\alpha(t) = \text{logit}F_0(t)$. $\alpha(t)$ can be interpreted as the baseline log-odds function, and is a monotone increasing function since $F_0(t)$ is increasing. In comparison, the Cox model can be written as

$$\log(-\log(1 - F(t|z))) = \log(-\log(1 - F_0(t))) + \beta'z.$$

In model (1.1), the logit function is used as the link function, while in the Cox model, $\log(-\log)$ is the link. The proportional odds regression model

¹Research supported in part by a postdoctoral fellowship at the Fred Hutchinson Cancer Research Center.

AMS 1991 subject classifications. Primary: 62G05, 62E20; secondary 62G20, 62P99.

Key words and phrases. Asymptotic normality, efficiency, current status data, interval censoring, information, maximum likelihood estimator, proportional odds regression model, semiparametrics.

resembles the logistic regression for binary data. For a survey of connection between binary response models and survival models, see Doksum and Gasko (1990).

In this paper, we study maximum likelihood estimation of the semiparametric proportional odds regression model with current status data. With current status data, it is only known whether the failure event has occurred before or after a censoring time Y . Thus the observable variable is

$$X = (Y, \delta, Z) \in R^+ \times \{0, 1\} \times R^d,$$

where $\delta = 1_{\{T \leq Y\}}$ indicating whether T has occurred or not. Current status data is also called “case 1” interval censored data by Groeneboom and Wellner (1992). They studied properties of the nonparametric maximum likelihood estimators of a distribution function with current status data and more general “case 2” interval censored data.

Current status data arises naturally in many applications. For example, it arises in animal tumorigenicity experiments, see, e.g., Hoel and Walburg (1972), Dinse and Lagakos (1983), Finkelstein and Wolfe (1985), and Finkelstein (1986). It also arises in HIV and AIDS studies, see, for example, Shiboski and Jewell (1992) and Jewell, Malani and Vittinghoff (1994). For applications in demographic studies, see Diamond, McDonald and Shah (1986) and Diamond and McDonald (1991).

There has been much interest in studying regression models with current status data. Recent works include Robinowitz, Tsiatis and Aragon (1995) and Huang (1994) among others. For a survey of regression models with interval censored data, see Huang and Wellner (1993). The enormous amount of work on binary choice model in the econometrics literature is closely related to the regression models with current status data, see Klein and Spady (1993) and the references therein.

Rossini (1994), and Rossini and Tsiatis (1994) first studied estimation of the semiparametric proportional odds regression model with current status data. Their approach is to use a step function as an approximation to the baseline log-odds function α and carry out an maximum “approximate” likelihood estimation procedure. They showed that their estimator for the regression parameter β is asymptotically normal and asymptotically efficient. However, the resulting estimator for α in general is not an increasing function as α is. Their approach can be regarded as a sieve estimation procedure, where the “sieve” consists of all the step functions over a finite interval of examination time, with the number of jump points depending on the sample size n and increasing to infinity with an appropriate rate. It is conceivable that other types of sieves can also be used. With this approach, one needs to specify the sieve à priori.

Here we apply the approach of Huang (1994) in studying maximum

likelihood estimation of the Cox model with current status data. We show that the MLE for β is also asymptotically normal and efficient. Moreover, the MLE of α can be taken as an increasing function and converges with $n^{-1/3}$ rate. It is shown in Gill and Levit (1993) that this rate is optimal for estimating a distribution function with current status data with minimal smoothness assumptions.

In the following, we first define the MLE $(\hat{\beta}_n, \hat{\alpha}_n)$ for (β, α) . The results are stated in section 3. Proofs are put together in section 5. Section 4 contains a brief discussion on diagnostics of the proportional odds models. In proving our main Theorem 3.3, we apply theorem 6.1 of Huang (1994) for MLE's for a class of semiparametric models. This theorem asserts that under certain regularity conditions, the MLE of the finite dimensional parameter has \sqrt{n} -convergence rate and is asymptotically normal, and moreover it achieves the asymptotic efficiency bound even though the MLE for the infinite dimensional parameter has a convergence rate slower than \sqrt{n} . Several technical lemmas that are used in section 5 are included in the appendix.

2. Maximum likelihood estimators of β and α . The goal of this section is to define and characterize the maximum likelihood estimator $(\hat{\beta}_n, \hat{\alpha}_n)$ of (β_0, α_0) for a finite sample size n , where β_0 and α_0 are the "true" regression parameter and baseline logit function. We will also denote the "true" baseline distribution function as F_0 . The characterization is in terms of the score function for β , and makes use of the monotonicity constraints, since the baseline logit function $\alpha(s)$ is an increasing function.

Throughout the rest of the discussion, we assume that T and Y are independent given Z . For a single observation $X = (Y, \delta, Z)$, under model (1.1),

$$E(\delta|Y = y, Z = z) = \frac{\exp(\alpha(y) + \beta'z)}{1 + \exp(\alpha(y) + \beta'z)}, \quad (2.2)$$

so the probability density function is

$$p_{\beta, \alpha}(x) = \frac{\exp(\delta(\alpha(y) + \beta'z))}{1 + \exp(\alpha(y) + \beta'z)} h(y, z).$$

$h(y, z)$ is the joint density of (Y, Z) and we assume that it does not involve (β, α) . The log-likelihood function is, up to a constant,

$$l(\beta, \alpha; x) = \delta(\alpha(y) + \beta'z) - \log(1 + \exp(\alpha(y) + \beta'z)). \quad (2.3)$$

Let $(Y_1, \delta_1, Z_1), \dots, (Y_n, \delta_n, Z_n)$ be an i.i.d. sample distributed according to p_{β_0, α_0} . Then the log-likelihood for the sample is, up to an additive constant,

$$l_n(\beta, \alpha) = \sum_{i=1}^n \{\delta_i(\alpha(Y_i) + \beta'Z_i) - \log(1 + \exp(\alpha(Y_i) + \beta'Z_i))\}. \quad (2.4)$$

Let $Y_{(1)}, \dots, Y_{(n)}$ be the order statistics of Y_1, \dots, Y_n ; that is, $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$. Let $\delta_{(i)}, Z_{(i)}$ correspond to $Y_{(i)}$, i.e., if $Y_{(i)} = Y_j$, then $\delta_{(i)} = 1_{\{T_j \leq Y_j\}}$ and $Z_{(i)} = Z_j$. Let $\alpha_{(i)} = \alpha(Y_{(i)})$. Since only the values of α at $Y_{(i)}$'s matter in the log-likelihood function, to avoid ambiguity, we will take the maximum likelihood estimator $\hat{\alpha}_n$ of α_0 as the right continuous increasing step function with jump points at $Y_{(i)}$ and values $\hat{\alpha}_n(Y_{(i)}), i = 1, \dots, n$.

Since the baseline odds function is a nondecreasing function, it is natural to require its estimator to be nondecreasing. Hence we require $\hat{\alpha}_{(1)} \leq \hat{\alpha}_{(2)} \leq \dots \leq \hat{\alpha}_{(n)}$. Let $\Theta \subset R^d$ be the finite dimensional parameter space of β . We assume that Θ is a bounded (convex) subset of R^d .

Suppose the support of the unobservable failure time T is $[0, \tau_{F_0}]$, where $\tau_{F_0} = \inf\{t : F_0(t) = 1\}$. For the censoring variable Y , we suppose its support $[l_y, u_y]$ is strictly contained in the support of T , i.e., $0 < l_y \leq u_y < \tau_{F_0}$. Since $0 < F_0(l_y) \leq F_0(u_y) < 1$, the baseline odds function $\alpha_0(t) = \text{logit} F_0(t) = \log[F_0(t)/(1 - F_0(t))]$ is finite on $[l_y, u_y]$. We assume that $-M_0 < \alpha_0(t) < M_0$ on $[l_y, u_y]$ for some known large positive number M_0 . It makes sense to require the estimator of α_0 be bounded between $-M_0$ and M_0 . So the maximum likelihood estimator of β_0 and α_0 is the $\hat{\beta}_n$ and $\hat{\alpha}_n$ corresponding to $(\hat{\alpha}_{(1)}, \dots, \hat{\alpha}_{(n)})$ that maximizes

$$\phi(\beta, \tilde{x}) = \sum_{i=1}^n \left\{ \delta_{(i)}(x_i + \beta' Z_{(i)}) - \log(1 + \exp(x_i + \beta' Z_{(i)})) \right\} \tag{2.5}$$

subject to $\beta \in \Theta$ and the monotonicity constraints

$$-M_0 \leq x_1 \leq x_2 \leq \dots \leq x_n \leq M_0.$$

Since $-\log(1 + \exp(x))$ is a concave function, it follows by Theorem 5.7 of Rockafellar (1970), page 38, or it can be verified directly that $\phi(\beta, \tilde{x})$ is a concave function jointly in (β, \tilde{x}) . So for any sample size n , this maximization problem is well defined and has a unique solution.

To characterize the solutions to the maximization problem (2.5), first consider the following closely related problem. Let $(\tilde{\alpha}_n(Y_{(1)}), \dots, \tilde{\alpha}_n(Y_{(n)}))$ be the solution to maximizing

$$\phi(\hat{\beta}_n, \tilde{x}) = \sum_{i=1}^n \left\{ \delta_{(i)}(x_i + \hat{\beta}_n' Z_{(i)}) - \log(1 + \exp(x_i + \hat{\beta}_n' Z_{(i)})) \right\} \tag{2.6}$$

subject to

$$x_1 \leq x_2 \leq \dots \leq x_n.$$

Without the restriction that x_i 's are bounded by $-M_0$ and M_0 , the solution can be unbounded. For example, if for some $1 \leq k < n$, $\delta_{(1)} = \dots = \delta_{(k)} = 0$, then to maximize (2.6) without violating the

constraints, $\tilde{\alpha}_n(Y_{(1)}) = \dots = \tilde{\alpha}_n(Y_{(k)}) = -\infty$. On the other hand, if for some $1 < k \leq n$, $\delta_{(k)} = \dots = \delta_{(n)} = 1$, then to maximize (2.6) without violating the constraints, $\tilde{\alpha}_n(Y_{(k)}) = \dots = \tilde{\alpha}_n(Y_{(n)}) = \infty$. This will make the solution to (2.6) inconsistent at the points l_Y and u_Y . Technically, without compactness of the parameter space, it seems difficult to prove consistency in any appropriate sense. However, for the solution to (2.5), even though it is not consistent at l_Y and u_Y , we will be able to show that $\hat{\alpha}_n$ is consistent in $L_2(Q_Y)$, where Q_Y is the marginal probability measure of Y . Let $\tau_{(1)} < \dots < \tau_{(m)}$ be the jump points of $\tilde{\alpha}_n(y)$, and $\tau_{(0)} = l_Y, \tau_{(m+1)} = u_Y$. Then $\tilde{\alpha}_n(\tau_j)$ is the solution to the equation

$$\sum_{\tau_j \leq Y_i < \tau_{j+1}} \left\{ \delta_i - \frac{\exp(x + \hat{\beta}_n' Z_{(i)})}{1 + \exp(x + \hat{\beta}_n' Z_{(i)})} \right\} = 0, \quad j = 1, \dots, m. \quad (2.7)$$

This can be proved exactly the same way as in the proof of Proposition 1.2 of Groeneboom and Wellner (1992).

Let t_1 be the first jump point of $\hat{\alpha}_n$ such that $\hat{\alpha}_n(t_1) \geq -M_0$. Let t_m be the first jump point of $\hat{\alpha}_n$ such that $\hat{\alpha}_n(t_m) \geq M_0$. Then by concavity of the function ϕ defined in (2.5), we have

$$\hat{\alpha}_n(y) = \begin{cases} -M_0 & \text{if } y < t_1 \\ \tilde{\alpha}_n(y) & \text{if } t_1 \leq y < t_m \\ M_0 & \text{if } y \geq t_m \end{cases} .$$

3. Main results. In this section, we state our main results. The proofs are put together in section 4.

3.1. Information calculation The following result on the information bound for estimation of β is given in Rossini and Tsiatis (1994). For a detailed treatment of information calculation in a semiparametric model, see chapter 3 of Bickel, Klaassen, Ritov and Wellner (1993).

THEOREM 3.1. *Suppose that:*

(i) *The covariate Z has bounded support; i.e, there exists z_0 such that $|Z| \leq z_0$ with probability one.*

(ii) *The support of the distribution of the censoring variable Y is strictly contained in the support of the distribution of T .*

Then:

(a) *The efficient score function for β is*

$$l_{\beta}^*(x) = (\delta - E(\delta|Y = y, Z = z)) \left(z - \frac{E(Z \text{Var}(\delta|Y, Z)|Y = y)}{E(\text{Var}(\delta|Y, Z)|Y = y)} \right)$$

(b) The information for β is

$$\begin{aligned}
 I(\beta) &= E[i_{\beta}^*(X)]^{\otimes 2} \\
 &= E \left[(\delta - E(\delta|Y = y, Z = z))^2 \left(Z - \frac{E(Z \text{Var}(\delta|Y, Z)|Y)}{E(\text{Var}(\delta|Y, Z)|Y)} \right)^{\otimes 2} \right],
 \end{aligned}
 \tag{3.8}$$

where $a^{\otimes 2} = aa'$ for any column vector $a \in R^d$.

3.2. Consistency and rate of convergence As we have shown, for each fixed sample size n , $(\hat{\beta}_n, \hat{\alpha}_n)$ is well defined. The following theorem asserts the consistency of $\hat{\beta}_n$ and consistency of $\hat{\alpha}_n$ on the support of Y .

THEOREM 3.2. (Consistency) Suppose that:

- (i) The finite dimensional parameter space Θ is a bounded subset of R^d .
- (ii) $F_0(0) = 0$. Let $\tau_{F_0} = \inf\{t : F_0(t) = 1\}$. The support of Y is an interval $S[Y] = [l_Y, u_Y]$, and $0 \leq l_Y \leq u_Y < \tau_{F_0}$.
- (iii) There exists z_0 such that $|Z| \leq z_0$ with probability one. Moreover, for any $\beta \neq \beta_0$, the probability $P\{\beta'Z \neq \beta_0'Z\} > 0$.

Then

$$\hat{\beta}_n \rightarrow_{a.s.} \beta_0,$$

and

$$\int_{S[Y]} |\hat{\alpha}_n(y) - \alpha_0(y)|^2 dQ_Y(y) \rightarrow_{a.s.} 0,$$

where Q_Y is the marginal probability measure of the censoring variable Y .

Define the distance d on $R \times \Phi$ as follows:

$$d((\alpha_1, \beta_1), (\alpha_2, \beta_2)) = |\beta_1 - \beta_2| + \|\alpha_1 - \alpha_2\|_2,$$

where $|\beta - \beta_0|$ is the Euclidean distance in R^d , and $\|\alpha_1 - \alpha_2\|_2 = [\int (\alpha_1(y) - \alpha_2(y))^2 dQ_Y(y)]^{1/2}$.

Applying Lemma 6.1, Theorem 3.2.1 and Lemma 3.2.2 of Van der Vaart and Wellner (1995), we can prove the following result.

THEOREM 3.3. (Rate of convergence) Suppose that conditions (i)-(iii) of theorem 3.2 are satisfied. Furthermore, $S[Y]$ is strictly contained in the support of F_0 , i.e., $0 < l_Y < u_Y < \tau_{F_0}$.

Then:

$$d((\alpha_0, \beta_0), (\hat{\alpha}_n, \hat{\beta}_n)) = O_p(n^{-1/3}).$$

The overall rate of convergence is dominated by $\hat{\alpha}_n$, which agrees with the convergence rate of the NPMLE of a distribution function studied by Groeneboom and Wellner (1992). In the following, we will show that the convergence rate of $\hat{\beta}_n$ can be refined to achieve \sqrt{n} .

3.3. Asymptotic normality and efficiency We now state the main theorem, which asserts that, under appropriate regularity conditions, the maximum likelihood estimator $\hat{\beta}_n$ satisfies a central limit theorem and is asymptotically efficient.

THEOREM 3.4. (*Asymptotic normality*) *Suppose that conditions of theorem 3.3 are satisfied. Furthermore, suppose that:*

- (i) β_0 is an interior point of Θ .
- (ii) The cumulative hazard function α_0 has strictly positive derivative on $S[Y]$.
- (iii) The function

$$h^*(y) = \frac{E[Z \text{Var}(\delta|Y, Z)|Y = y]}{E[\text{Var}(\delta|Y, Z)|Y = y]}$$

has bounded derivative on $S[Y]$. (This d -dimensional function comes from the information calculation.)

Then

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = I(\beta_0)^{-1} \sqrt{n} P_n \dot{l}_{\beta_0}^*(x) + o_p(1) \rightarrow_d N(0, I(\beta_0)^{-1}),$$

where $\dot{l}_{\beta_0}^*(x)$ is the efficient score defined in Theorem 3.1, and $I(\beta_0)$ is the information.

Since $\hat{\beta}_n$ is asymptotically linear with efficient influence function, and the likelihood function is Hellinger differentiable with respect to (β, α) , it is asymptotically efficient in the sense that any regular estimator has asymptotic variance matrix no less than that of $\hat{\beta}_n$. We do not go into the details here, but refer the reader to Van der Vaart (1991), and Bickel et al. (1993), chapter 3.

4. Model diagnostics. Although this paper is mainly concerned with estimation based on a specified proportional odds regression model, a referee has suggested that some discussions on motivations to specify the proportional odds model rather than the more common proportional hazards model or diagnostics would be helpful. For the former, we refer the reader to Bennet (1983) for discussions and examples. Basically, the proportional odds model allows for covariate effects that are not necessarily multiplicative on the baseline hazard.

Model diagnostics can be most easily done in an informal fashion. First consider the simplest case of Z being a one dimensional covariate only taking value 0 or 1. We can first compute the nonparametric maximum likelihood estimates (NPMLE) $\hat{F}_n(t|1)$ and $\hat{F}_n(t|0)$ of the distributions for the two groups separately. Computation and properties of NPMLE with

current status data are discussed in Groeneboom and Wellner (1992). If the proportional odds model is reasonable, then the plot of $\text{logit} \widehat{F}_n(t|1) - \text{logit} \widehat{F}_n(t|0)$ against t should be roughly a horizontal straight line. Any serious departure from this would indicate that the proportional odds model is probably not adequate. This informal approach can be generalized to the case when Z is a vector of categorical covariates. For any fixed z_1 and z_2 , the NPMLE's $\widehat{F}_n(t|z_1)$ and $\widehat{F}_n(t|z_2)$ can be computed from the observations corresponding to z_1 and z_2 , respectively; and then the difference $\widehat{F}_n(t|z_1) - \widehat{F}_n(t|z_2)$ can be plotted against t . For continuous covariates, some form of grouping or discretization is needed.

Formal diagnostics can be done based on embedding the proportional odds model into a more general model, for example, the γ -logit model described in Doksum and Gasko (1990). This model contains an extra parameter γ and includes both the proportional odds (when $\gamma = 1$) and the proportional hazards model (when $\gamma = 0$) as special cases. The goodness of fit of the proportional odds model can be checked by testing $\gamma = 1$. Discussion of details is beyond the scope of the present paper.

5. Proofs.

In this section, we prove Theorems 3.2, 3.3, and 3.4.

PROOF OF THEOREM 3.2. Let $x = (\delta, y, z)$ and

$$f_n(x) = \delta(\widehat{\alpha}_n(y) + \widehat{\beta}_n' z) - \log(1 + \exp(\widehat{\alpha}_n(y) + \widehat{\beta}_n' z)).$$

Similarly define f_0 , but replace $\widehat{\beta}_n$ and $\widehat{\alpha}_n$ by β_0 and α_0 , respectively. Recall that $(\widehat{\beta}_n, \widehat{\alpha}_n)$ maximizes $l_n(\beta, \alpha)$ subject to the constraint that $\widehat{\alpha}_n$ is increasing and $|\widehat{\alpha}_n(y)| \leq M_0$ for all n and $y \in S[Y]$. Let $X_i = (\delta_i, Y_i, Z_i)$, $i = 1, \dots, n$. Since $|\alpha_0(y)| < M_0$ for all $y \in S[Y]$,

$$\frac{1}{n} \sum_{i=1}^n f_n(X_i) \geq \frac{1}{n} \sum_{i=1}^n f_0(X_i)$$

Let P_n denote the empirical measure of $(\delta_1, Y_1, Z_1), \dots, (\delta_n, Y_n, Z_n)$. We can write this inequality as

$$P_n f_n(x) \geq P_n f_0(x) \tag{5.9}$$

Let the sample space Ω be the space of all infinite sequences $(\delta_1, Y_1, Z_1), (\delta_2, Y_2, Z_2), \dots$, endowed with the usual σ -algebra generated by the product topology on $\prod_1^\infty (\{0, 1\} \times R^2)$ and the product measure \mathbf{P} . By Lemma 6.1 in the appendix and the bracketing Glivenko-Cantelli theorem, there exists a set $A \in \Omega$ with $\mathbf{P}(A) = 1$ such that for every $\omega \in A$,

$$\int f_n(x) d(P_n - P) \rightarrow 0,$$

where P is the joint probability measure of (δ_1, Y_1, Z_1) .

Now fix an arbitrary $\omega \in A$. For this ω , write $\hat{\beta}_n = \hat{\beta}_n(\omega)$ and $\hat{\alpha}_n(\cdot) = \hat{\alpha}_n(\cdot, \omega)$. Since Θ is bounded, for any subsequence of $\hat{\beta}_n$ we can find a further subsequence converging to $\beta_* \in \bar{\Theta}$, the closure of Θ . Moreover, by Helly's selection theorem, for any subsequence of $\hat{\alpha}_n$, we can find a further subsequence converging to some increasing function α_* . Choose the convergent subsequence of $\hat{\beta}_n$ and the convergent subsequence of $\hat{\alpha}_n$ so that they have the same indices, and assume that $\hat{\beta}_n$ converges to β_* and that $\hat{\alpha}_n$ converges to $\alpha_*(\cdot)$.

By the bounded convergence theorem, we have

$$\int f_n(x)dP \rightarrow \int f_*(x)dP$$

where f_* is similarly defined as f_0 .

By (5.9), and using the strong law of large numbers for the right hand side of (5.9),

$$\int f_*(x)dP \geq \int f_0(x)dP$$

By the Kullback-Leibler inequality, the left hand side is less than or equal to the right hand side, with equality holds if and only if

$$\alpha_*(y) + \beta'_*z = \alpha_0(y) + \beta'_0z \quad P_G - \text{a.s.},$$

where P_G is the probability measure induced by the joint distribution function G of (Y, Z) . Condition (iii) then implies

$$\beta_* = \beta_0 \tag{5.10}$$

and

$$\alpha_*(y) = \alpha_0(y) \quad Q_Y - \text{a.s.}$$

Since both $\hat{\alpha}_n(y)$ and $\alpha_*(y)$ are bounded for every y , by the bounded convergence theorem, it follows that

$$\int |\hat{\alpha}_n(y) - \alpha_0(y)|^2 dQ_Y(y) \rightarrow 0. \tag{5.11}$$

Since (5.10) and (5.11) hold for any $\omega \in A$ with $\mathbf{P}(A) = 1$, the proof is completed. \square

PROOF OF THEOREM 3.3. Using a Taylor expansion, it can be verified that

$$El(\beta, \alpha, X) - El(\beta_0, \alpha_0, X) \leq -Cd^2((\alpha_0, \beta_0), (\alpha, \beta))$$

for some constant $C > 0$. Thus

$$\sup_{\eta/2 \leq d((\alpha_0, \beta_0), (\alpha, \beta)) \leq \eta} (El(\beta, \alpha, X) - El(\beta_0, \alpha_0, X)) \leq -C\eta^2/4.$$

By Lemma 6.1, Remark 6.1, and Lemma 3.2.2 of Van der Vaart and Wellner (1995),

$$\begin{aligned} E^* \sup_{d((\alpha_0, \beta_0), (\alpha, \beta)) \leq \eta} |\sqrt{n}(P_n - P)(l(\beta, \alpha, x) - l(\beta_0, \alpha_0, x))| \\ = O(1)\eta^{1/2} \left(1 + \frac{\eta^{1/2}}{\eta^2\sqrt{n}}M\right). \end{aligned}$$

Let

$$\phi_n(\eta) = \eta^{1/2} \left(1 + \frac{\eta^{1/2}}{\eta^2\sqrt{n}}M\right).$$

Then $\phi_n(\eta)/\eta$ is a decreasing function, and it is easy to verify that

$$n^{2/3}\phi_n(n^{-1/3}) = O(\sqrt{n})$$

for n large. Furthermore, by Theorem 3.2, $\widehat{\beta}_n$ is consistent and $\widehat{\alpha}_n$ is consistent in $L^2(Q_Y)$. Hence the conditions of Theorem 3.2.1 of Van der Vaart and Wellner (1995) are satisfied. This implies

$$d((\alpha_0, \beta_0), (\widehat{\alpha}_n, \widehat{\beta}_n)) = O_p(n^{-1/3}).$$

□

PROOF OF THEOREM 3.4. We prove the main theorem by verifying the conditions of Theorem 6.1 of Huang (1994).

For the proportional odds regression model under interval censoring,

$$l(\beta, \alpha; x) = \delta(\alpha(Y) + \beta'Z) - \log(1 + \exp(\alpha(Y) + \beta'Z)).$$

The partial derivative of $l(\beta, \alpha; x)$ with respect to β is

$$l_1(\beta, \alpha; x) = \delta z - zA(y, z),$$

where

$$A(y, z) = \frac{\exp(\alpha(y) + \beta'y)}{1 + \exp(\alpha(y) + \beta'z)}.$$

Since $(\widehat{\beta}_n, \widehat{\alpha}_n)$ is the MLE,

$$S_{1n}(\widehat{\beta}_n, \widehat{\alpha}_n) \equiv P_n l_1(\widehat{\beta}_n, \widehat{\alpha}_n; x) = 0.$$

For \mathbf{h}^* defined in condition (v), define

$$l_2(\beta, \alpha; x)[\mathbf{h}^*] = \mathbf{h}^*(y)(\delta - A(y, z))$$

We now show that

$$S_{2n}(\hat{\beta}_n, \hat{\alpha}_n)[\mathbf{h}^*] \equiv P_n l_2(\beta, \alpha; x)[\mathbf{h}^*] = o_p(n^{-1/2}).$$

Since α_0 is a strictly increasing continuous function, its inverse α_0^{-1} is well defined. Let $\xi_0 = \mathbf{h}^* \circ \alpha_0^{-1}$, i.e., the composition of h_* on the inverse of α_0 . Then ξ_0 is well defined on the range of α_0 . Since $\xi(\hat{\alpha}_n(y))$ is a right continuous step function and has exactly the same jump points as $\hat{\alpha}_n(y)$, by the characterization of $\hat{\alpha}_n$,

$$\frac{1}{n} \sum_{t_1 \leq Y_i < t_m} \left\{ \mathbf{h}^*(\alpha_0^{-1}(\hat{\alpha}_n(Y_i)))(\delta_i - A_n(Y_i, Z_i)) \right\} = 0,$$

where

$$A_n(y, z) = \frac{\exp(\hat{\alpha}_n(Y_i) + \hat{\beta}_n' Z_i)}{1 + \exp(\hat{\alpha}_n(Y_i) + \hat{\beta}_n' Z_i)}.$$

Recall that t_1 and t_m are defined in the last paragraph of section 2. For $l_Y \leq y < t_1$, $\hat{\alpha}_n(y) = -M_0$. In addition, $\hat{\beta}_n$ is bounded, and Z has bounded support. It follows that

$$\begin{aligned} & \frac{1}{n} \left| \sum_{l_Y \leq Y_i < t_1} \left\{ \mathbf{h}^*(\alpha_0^{-1}(\hat{\alpha}_n(Y_i)))(\delta_i - A_n(Y_i, Z_i)) \right\} \right| \\ & \leq \frac{1}{n} \sum_{l_Y \leq Y_i < t_1} \left| \mathbf{h}^*(\alpha_0^{-1}(\hat{\alpha}_n(Y_i)))(\delta_i - A_n(Y_i, Z_i)) \right| \\ & \leq C \frac{1}{n} \sum_{i=1}^n 1_{[l_Y \leq Y_i < t_1]}, \end{aligned}$$

for some constant C not dependent on n . Now write

$$\frac{1}{n} \sum 1_{[l_Y \leq Y_i < t_1]} = P_n 1_{[l_Y \leq Y_i < t_1]} = (P_n - P) 1_{[l_Y \leq Y_i < t_1]} + P 1_{[l_Y \leq Y_i < t_1]}.$$

The first term is $o_p(n^{-1/2})$. This follows from Lemma 4.1 of Pollard (1989), because the class of indicator functions of intervals is a VC-subgraph class and that $P 1_{[l_Y \leq Y < t_1]} \rightarrow_p 0$. The second term is equal to

$$Q_Y(t_1) - Q_Y(l_Y) = O_p(n^{-2/3}). \tag{5.12}$$

Here we also use Q_Y to denote the distribution function of Y . This is because, by Theorem 3.3,

$$\int_{l_Y}^{t_1} (\hat{\alpha}_n(y) - \alpha_0(y))^2 dQ_Y(y) = O_p(n^{-2/3}).$$

However, on the interval $[l_Y, t_1)$, $\hat{\alpha}_n(y) = -M_0$, and $\alpha_0(y) > -M_0$, thus

$$(-M_0 - \alpha_0(l_Y))^2(Q_Y(t_1) - Q_Y(l_Y)) \leq \int_{l_Y}^{t_1} (M_0 + \alpha_0(y))^2 dG_Y(y) = O_p(n^{-2/3}).$$

So equation (5.12) holds. It follows that

$$\frac{1}{n} \sum_{l_Y \leq Y_i < t_1} \left\{ \mathbf{h}^*(\alpha_0^{-1}(\hat{\alpha}_n(Y_i)))(\delta_i - A_n(Y_i, Z_i)) \right\} = o_p(n^{-1/2}).$$

Similarly,

$$\frac{1}{n} \sum_{t_m \leq Y_i \leq u_Y} \left\{ \mathbf{h}^*(\alpha_0^{-1}(\hat{\alpha}_n(Y_i)))(\delta_i - A_n(Y_i, Z_i)) \right\} = o_p(n^{-1/2}).$$

So we have proved that

$$\frac{1}{n} \sum_{i=1}^n \left\{ \mathbf{h}^*(\alpha_0^{-1}(\hat{\alpha}_n(Y_i)))(\delta_i - A_n(Y_i, Z_i)) \right\} = o_p(n^{-1/2}).$$

Write this in terms of empirical process notation, we have

$$P_n \left\{ \mathbf{h}^*(\alpha_0^{-1}(\hat{\alpha}_n(y)))(\delta - A_n(y, z)) \right\} = o_p(n^{-1/2}). \quad (5.13)$$

Furthermore, since \mathbf{h}^* is differentiable and α_0 has strictly positive derivative, ξ_0 has bounded derivative. So noticing $\mathbf{h}^* = \mathbf{h}^* \circ \alpha_0^{-1} \circ \alpha_0 = \xi_0 \circ \alpha_0$, we have

$$\begin{aligned} & P_n \{ \mathbf{h}^*(y)(\delta - A_n(y, z)) \} \\ &= P_n \{ [\xi_0 \circ \alpha_0(y) - \xi_0 \circ \hat{\alpha}_n(y)](\delta - A_n(y, z)) \} + o_p(n^{-1/2}) \\ &= (P_n - P) \{ [\xi_0 \circ \alpha_0(y) - \xi_0 \circ \hat{\alpha}_n(y)](\delta - A_n(y, z)) \} \\ &\quad + P \{ [\xi_0 \circ \alpha_0(y) - \xi_0 \circ \hat{\alpha}_n(y)](\delta - A_n(y, z)) \} + o_p(n^{-1/2}) \end{aligned}$$

To show that the first term is of the order $o_p(n^{-1/2})$, let

$$\psi(x; \beta, \alpha) = (\xi_0 \circ \alpha_0(y) - \xi_0 \circ \alpha(y))(\delta - A(y, z)).$$

For any $\eta > 0$, define the class of functions

$$\Psi(\eta) = \{ \psi(x; \beta, \alpha) : |\beta - \beta_0| + \|\alpha - \alpha_0\| \leq \eta, \text{ and } \alpha \in \Phi \}, \quad (5.14)$$

where

$$\Phi = \{ \alpha : \alpha \text{ is increasing, and } -M_0 \leq \alpha(y) \leq M_0 < \infty \text{ for all } y \in S[Y] \}, \quad (5.15)$$

and M_0 is a positive constant. It is verified in Lemma 6.2, that for any probability measure Q , the $L_2(Q)$ ε -entropy number for the class $\Psi(\eta)$ is

in the order of $1/\varepsilon$, and hence $\Psi(\eta)$ is a Donsker class. In addition, since $\sup_{\psi \in \Psi(\eta)} P\psi(x; \beta, \alpha)^2 \rightarrow 0$ as $\eta \rightarrow 0$, by Lemma 4.1 of Pollard (1989), it follows that

$$\sup_{\psi \in \Psi(Cn^{-1/3})} (P_n - P)\psi(x; \beta, \alpha) = o_p(n^{-1/2}).$$

This implies the first term is of order $o_p(n^{-1/2})$. For the second term, by equation (2.2), the Cauchy-Schwarz inequality and Theorem 3.3, we have for some finite constant $C > 0$,

$$\begin{aligned} & |P\{[\xi_0 \circ \alpha_0(y) - \xi_0 \circ \hat{\alpha}_n(y)](\delta - A_n(y, z))\}| \\ &= \left| P \left\{ [\xi_0 \circ \alpha_0(y) - \xi_0 \circ \hat{\alpha}_n(y)] \frac{\exp(\alpha_0(y) + \beta'_0 z) - \exp(\hat{\alpha}_n(y) + \hat{\beta}'_n z)}{1 + \exp(\hat{\alpha}_n(y) + \hat{\beta}'_n z)} \right\} \right| \\ &\leq C \{P[\xi_0 \circ \alpha_0(y) - \xi_0 \circ \hat{\alpha}_n(y)]^2\}^{1/2} \\ &\quad \times \{P[\exp(\alpha_0(y) + \beta'_0 z) - \exp(\hat{\alpha}_n(y) + \hat{\beta}'_n z)]^2\}^{1/2} \\ &= O_p(n^{-2/3}). \end{aligned}$$

Thus the first assumption of Theorem 6.1 of Huang (1994) is verified.

Now we verify Conditions 1 — 5 listed there. By Theorem 3.3, Condition 1 is satisfied with $\gamma = 1/3$. The information calculation asserts that Condition 2 holds. To verify Condition 3, consider the following two classes of functions,

$$\{l_1(\beta, \alpha; x) - l_1(\beta_0, \alpha_0; x) : |\beta - \beta_0| \leq \eta, \|\alpha - \alpha_0\| \leq \eta\},$$

$$\{l_2(\beta, \alpha; x)[\mathbf{h}^*] - l_2(\beta_0, \alpha_0; x)[\mathbf{h}^*] : |\beta - \beta_0| \leq \eta, \|\alpha - \alpha_0\| \leq \eta\},$$

where η is near 0. It can be proved as in Lemma 6.2 below that the entropy numbers for the above two classes are of order $1/\eta$. This implies that these two classes are Donsker, and hence Condition 3 is satisfied. As in Huang (1994), it can be verified using a straightforward Taylor expansion that Condition 4 is satisfied with

$$\dot{S}_{11}(\beta_0, \alpha_0) = -El_1(\beta_0, \alpha_0; x)l_1^T(\beta_0, \alpha_0; x),$$

$$\dot{S}_{12}(\beta_0, \alpha_0)[\mathbf{h}^*] = \dot{S}_{21}^T(\beta_0, \alpha_0)[\mathbf{h}^*] = -El_1(\beta_0, \alpha_0; x)l_2^T(\beta_0, \alpha_0; x)[\mathbf{h}^*],$$

$$\dot{S}_{22}(\beta_0, \alpha_0)[h_1, h_2] = -El_2'(\beta_0, \alpha_0; x)[h_1]l_2'(\beta_0, \alpha_0; x)[h_2],$$

and $\lambda = 2$. So $\lambda\gamma = 2 \times 1/3 > 1/2$. Condition 5 is satisfied because the information $I(\beta_0)$ is finite and positive. Thus the result follows from the cited theorem. \square

Acknowledgements. I started this work while I was a postdoctoral fellow at the Fred Hutchinson Cancer Research Center. I thank Ruth Etzioni,

Steve Self and Emily White for their encouragement. I also thank Tony Rossini for his helpful comments on an earlier version of this paper, and Aad van der Vaart and Jon Wellner for making their unpublished manuscript available to me. Thanks also go to two referees for their constructive suggestions. One referee brings my attention to the paper by Doksum and Gasko (1990).

6. Appendix. Define the class of log-likelihood functions $l(\beta, \alpha)$ defined by (2.3):

$$\mathcal{H} = \{l(\beta, \alpha) : \beta \in B(\beta_0, \eta), \alpha \in \Phi\}, \tag{6.16}$$

where Φ is defined by equation (5.15), $B(\beta_0, \eta)$ is an η -ball around β_0 , and $\eta > 0$ is any fixed positive number.

For any probability measure Q , define $L_2(Q) = \{f : \int f^2 dQ < \infty\}$. Let $\|\cdot\|_2$ be the usual L_2 norm, i.e., $\|f\|_2 = (\int f^2 dQ)^{1/2}$. For any subclass \mathcal{F} of $L_2(Q)$, define the bracketing number $N_{[]}(\varepsilon, \mathcal{F}, L_2(Q)) = \min\{m : \text{there exist } f_1^L, f_1^U, \dots, f_m^L, f_m^U \text{ such that for each } f \in \mathcal{F}, f_i^L \leq f \leq f_i^U \text{ for some } i, \text{ and } \|f_i^U - f_i^L\|_2 \leq \varepsilon\}$. Let

$$J_{[]}(\eta, \mathcal{F}, \|\cdot\|_2) = \int_0^\eta \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(Q))} .d\varepsilon, \tag{6.17}$$

be the bracketing integral of the class of functions \mathcal{F} .

LEMMA 6.1. *Let \mathcal{H} be defined by (6.16), and suppose that Z has bounded support. Then there exists a constant $C > 0$ such that*

$$\sup_Q N_{[]}(\varepsilon, \mathcal{H}, L_2(Q)) \leq C(1/\varepsilon^d)e^{1/\varepsilon}, \quad \text{for all } \varepsilon > 0,$$

where d is the dimension of β . Hence for ε small enough, we have

$$\sup_Q \log N_{[]}(\varepsilon, \mathcal{H}, L_2(Q)) \leq C \frac{1}{\varepsilon}.$$

Here Q runs through the class of all probability measures.

PROOF. The proof is similar to the proof of Lemma 3.1 of Huang (1994), and is omitted. \square

REMARK 6.1. From this lemma, the bracketing integral for the class \mathcal{H} is

$$J_{[]}(\eta, \mathcal{H}, L_2(Q)) = O(1) \int_0^\eta \sqrt{1/\varepsilon} d\varepsilon = O(\eta^{1/2})$$

for η close to zero.

LEMMA 6.2. For any $\eta > 0$, define the class of functions

$$\Psi(\eta) = \{\psi(x; \beta, \alpha) : |\beta - \beta_0| + \|\alpha - \alpha_0\|_2 \leq \eta, \text{ and } \alpha \in \Phi\},$$

where ψ is defined in (5.14), and Φ is defined in (5.15). Then the L_2 covering number $N(\varepsilon, \Psi, L_2(Q))$ of Ψ

$$\sup_Q N(\varepsilon, \Psi, L_2(Q)) \leq \text{constant} \cdot (1/\varepsilon^d) \exp(1/\varepsilon).$$

Hence for ε close to zero, the entropy number

$$\sup_Q \log N(\varepsilon, \Psi, L_2(Q)) \leq \text{constant} \cdot \frac{1}{\varepsilon}.$$

Here Q runs through all probability measures. This implies that $\Psi(\eta)$ is a Donsker class.

PROOF. The proof is similar to the proof of Lemma 7.1 of Huang (1994), and hence is omitted. \square

REFERENCES

- 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.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B* **34**, 187-220.
- 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, pp. 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. in *Demography*, **23**, 607-620.
- Dinse, G. E. and Lagakos, S. W. (1983). Regression analysis of tumor prevalence data (corr: V33, P79-80), *Applied Statistics* **32**, 236-248.
- Doksum, A. A. and Gasko, M. (1990). On a correspondence between models in binary regression analysis and in survival analysis. *International Statistical Review* **58**, 243-252.

- 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.
- Gill, R. D. and Levit, B. Y. (1993) Applications of the van Tree inequality: a Bayesian Cramér-Rao bound. Tech. Rep. #733, University of Utrecht (To appear in Journal of Bernoulli Society, 1995, 1 No. 1).
- Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. DMV Seminar Band 19, Birkhäuser, Basel.
- Hoel, D. G. and Walburg, H. E. (1972). Statistical analysis of survival experiments. *J. of the National Cancer Institute* **49**, 361-372.
- Huang, J. (1994). Efficient estimation for the Cox model with interval censoring. *Technical report No. 256*, Dept. of Stat. Univ. of Washington.
- Huang, J. and Wellner, J. A. (1993). Regression models with interval censoring. To appear, *Proceedings of the Kolmogorov Seminar, Euler Mathematics Institute*, St. Petersburg, Russia.
- 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 Statistical Association* **89**, 7-18.
- Klein, R. W. and Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica* **61**, 387-421.
- Parzen, M. (1993). *Computational Issues in Robust Regression*. Ph. D. thesis, Harvard University.
- Pettitt, A. N. (1984). Proportional odds models for survival data and estimates using ranks. *Applied Statistics* **33**, 169-175.
- Pollard, D. (1989). Asymptotics via empirical processes (with discussion). *Statistical Science* **4**, 341-366.
- Rabinowitz, D., Tsiatis, A., and Aragon, J. (1995). Regression with interval censored data. *Biometrika*, to appear.
- 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. (1994). *Regression Models for Interval Censored Data*. Ph. D. thesis, Department of Biostatistics, Harvard School of Public Health.
- Rossini, A. J. and Tsiatis, A. A. (1994). A semiparametric proportional odds regression model for the analysis of current status data. Technical Report, Department of Biostatistics, Harvard School of Public Health.
- 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.
- Vaart, A. W. van der (1991). *On differentiable functionals*. *Ann. Statist.* **19**, 178-204.
- Vaart, A. W. van der and Wellner, J. A. (1995). *Weak Convergence and Empirical Processes*. To appear.

