

NONLINEAR CENSORED REGRESSION

Winfried Stute

Giessen University

Abstract: In the standard accelerated lifetime model, log lifetime is, up to noise, a linear function of a vector of covariables. In the present paper this model is extended to admit general nonlinear functional relationships. A Weighted Least-Squares estimator for the unknown true parameter is proposed. Consistency and asymptotic normality are shown when the lifetimes are subject to censoring but the covariables are known. The accuracy of the procedure is demonstrated in a small sample simulation study under various degrees of censoring.

Key words and phrases: Nonlinear regression, random censorship, weighted least-squares.

1. Introduction and Main Result

Assume that we observe a sample of independent identically distributed random vectors (X_i, Y_i) in $d + 1$ -dimensional Euclidean space, $1 \leq i \leq n$, defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Denote with $m(x) = E(Y_1 | X_1 = x)$ their common regression function. Under a linear model assumption, the admissible m 's are of the form $m(x) = \langle \theta, x \rangle = \theta_1 x_1 + \dots + \theta_d x_d$, where θ ranges in a suitable parameter set Θ and the "true parameter" θ_0 is unknown. Linearity may be very restrictive and not appropriate for modelling the dependence structure between X and Y in various situations. At the other extreme, starting with Nadaraya (1964) and Watson (1964), there has been much interest in the estimation of m in a completely nonparametric framework. This requires some amount of smoothing of the data and may lead to a less precise fit than in a parametric setup.

A compromise between a linear and a nonparametric approach which may be flexible enough to model various data structures is given by the nonlinear regression model. See Seber and Wild (1989) for a comprehensive treatment of the subject. Now, the admissible m 's are of the form $m(x) = f(x, \theta_0)$, where $f(x, \theta_0)$ is a known function, θ_0 is a p -dimensional parameter, and p and d may be different. In other words, Y may be written as

$$Y = f(X, \theta_0) + \varepsilon \text{ with } E(\varepsilon | X) = 0. \quad (1.1)$$

If $m(x) = f(x, \theta_0)$, the condition $E(\varepsilon | X) = 0$ is automatically satisfied and more flexible than the assumption that the error variable ε has expectation zero and

is independent of X . Actually, in many heteroscedastic models, ε is given as $\varepsilon = \sigma(X)\tilde{\varepsilon}$, where $\tilde{\varepsilon}$ is a centered random variable independent of X and σ is a given scale function, which in addition may depend on an unknown parameter. The parameter θ_0 may be estimated by the Least-Squares Method or any of its robust or weighted modifications. For nonrandom design Jennrich (1969) proved strong consistency of the Least-Squares Estimator and derived its limit distribution. See also Malinvaud (1970). More recent notable contributions are Wu (1981) and Richardson and Bhattacharyya (1986).

In the present paper we consider a nonlinear regression model in the context of survival analysis. Now the variable Y_i may be viewed as the lifetime, or a monotone transformation of it, of an individual, while X_i is a vector of covariables to be sampled at the entry into, or in the course of a follow-up study. Since Y_i is not always available, standard methods which require knowledge of all Y 's are not applicable. Under random censorship, rather than a random variable Y_i , one observes Z_i together with an indicator δ_i such that $Z_i = Y_i$ if and only if $\delta_i = 1$. In censored regression two models which have often been discussed in the literature are the accelerated lifetime model and the Cox proportional hazards model; see, for example, Kalbfleisch and Prentice (1981). In the first model it is assumed that $Y = \log$ lifetime satisfies (1.1) with ε independent of X . In most cases, $f(x, \theta) = \langle \theta, x \rangle$. In the Cox proportional hazards model, the hazard function factors into a baseline function which only depends on time and another factor which only depends on the covariables. Doksum and Gasko (1990) pointed out that, rather than (1.1), the Cox model is a particular example of a transformation model. Nonproportional hazards models were considered by Carter, Wamper and Stablein (1983) in modelling the impact of drug combinations on survival in cancer chemotherapy. Their discussion was in terms of hazard functions rather than regression. Since there is a one-to-one correspondence between the two, each of their models could be rephrased so as to become a model equation for m . Equation (1.1) is more appropriate than the Cox Model if one wants to describe the functional relationship between log-lifetime and the covariable vector X rather than modelling the conditional survival and hazard functions. In the (log-)linear case this model has been studied in a couple of papers. See Andersen, Borgan, Gill and Keiding (1993, p.581) for a brief review. A statistical analysis of a general nonlinear regression model under censorship, however, still seems to be missing.

In the linear case, a standing assumption throughout the literature was that, modulo further regularity conditions, $Z = \min(Y, C)$ and $\delta = 1_{\{Y \leq C\}}$, where C is a censoring variable such that Y and C are independent conditionally on X . Buckley and James (1979) initiated the study of a modified Least-Squares Estimator of θ_0 in which the observable, possibly censored, Z_i were replaced

by their conditional expectations w.r.t. Y_i . This approach required estimation of the error distribution via Kaplan-Meier. For a criticism of the state of art at about the early nineties, see Ying (1993). Ritov (1990) and Tsiatis (1990) studied extensions of R -estimators of θ_0 under censorship, but under regularity conditions which are difficult to check in practice. Fyngenson and Ritov (1994) and Akritas, Murphy and LaValley (1995) considered extensions of the Sen-Theil estimator of θ_0 to the linear censored regression model. The advantage of their approach was that the resulting estimating equation is monotone so that the estimator is essentially unique. The limit variance is complicated and requires some smoothing. All of these approaches do not seem to allow for simple generalizations to the nonlinear case.

In the present paper we adopt and extend a methodology of Stute (1993) designed to estimate the joint distribution of (X, Y) when Y is subject to censoring and X is observable. There a weighted Least-Squares Estimator was shown to be strongly consistent when f in (1.1) is linear. See also Zhou (1992). In the present paper we obtain consistency and asymptotic normality of a weighted Least-Squares Estimator for a general smooth f . See Theorems 1.1 and 1.2. Proofs will be deferred to the Appendix. In Section 2 we provide simulation results on a particular bivariate regression model.

To make things more rigorous some notation will be required. Neglecting the covariables for a moment, denote with F the unknown distribution function of the Y 's. The nonparametric Maximum-Likelihood Estimator of F is given by the time-honoured Kaplan-Meier (1958) estimator \hat{F}_n . Following Stute and Wang (1993), \hat{F}_n may be written as a weighted sum of Dirac-measures:

$$\hat{F}_n = \sum_{i=1}^n W_{in} D_{Z_{i:n}},$$

where

$$W_{in} = \frac{\delta_{[i:n]}}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{[j:n]}}.$$

Here $Z_{1:n} \leq \dots \leq Z_{n:n}$ are the Z -order statistics and $\delta_{[i:n]}$ is the δ associated with $Z_{i:n}$. A detailed study of the W 's in connection with the Strong Law of Large Numbers under censoring has been carried out in Stute and Wang (1993). In the presence of covariables, Stute (1993) extended \hat{F}_n so as to obtain a consistent estimator of the joint distribution function $F^0(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ of (X, Y) , namely

$$\hat{F}_n^0(x, y) = \sum_{i=1}^n W_{in} 1_{\{X_{[i:n]} \leq x, Z_{i:n} \leq y\}}.$$

It was also shown that the Weighted Least-Squares Estimator minimizing

$$\sum_{i=1}^n W_{in} \left[Z_{i:n} - \langle \theta, X_{[i:n]} \rangle \right]^2$$

is strongly consistent. For a general f we are led to consider

$$S_n(\theta) = \sum_{i=1}^n W_{in} \left[Z_{i:n} - f(X_{[i:n]}, \theta) \right]^2$$

and then take any (measurable) minimizer θ_n of S_n as an estimator of θ_0 . Note that S_n incorporates all Z 's and not just those bounded away from the least upper bound of the support of the Z 's, as is done in most work on Kaplan-Meier. Our θ_n should not be confused with the Least Squares Estimator proposed by Miller (1976) in the linear case. That method was based on the Kaplan-Meier weights of the residuals rather than the observed Z 's.

We now state and discuss the conditions needed for Theorem 1.1. First note that W_{in} , and hence $S_n(\theta)$, are well-defined without any assumption on the censoring mechanism. As such, for the computational part, we do not have to assume that $Z = \min(Y, C)$ where C is a proper censoring variable. Actually, as was pointed out in Stute (1996b), almost sure and distributional convergence of Kaplan-Meier integrals carries through if only the $(Z_i, \delta_i), 1 \leq i \leq n$, are i.i.d. Independence of Y and C is just needed to identify the limit. Tsiatis (1975), among others, showed the fully nonparametric character of the censorship model in that the target quantities cannot be consistently estimated from a set of censored data without assuming an identifiability condition, similarly in the context of the present paper. The parameter θ_n minimizing $S_n(\theta)$ will turn out to converge with probability one and in distribution (when properly standardized) under no assumption on the censoring mechanism. Conditions (i) and (ii) below are needed for identifiability. Let $Z = \min(Y, C)$ where

- (i) Y and C are independent,
- (ii) $\mathbb{P}(Y \leq C | X, Y) = \mathbb{P}(Y \leq C | Y)$.

Furthermore assume that

- (iii) $EY^2 < \infty$,
- (iv) Θ is compact,
- (v) $f(x, \theta)$ is continuous in θ for each x ,
- (vi) $f^2(x, \theta) \leq M(x)$ for some integrable function M ,
- (vii) $L(\theta, \theta_0) = E \{f(X, \theta) - f(X, \theta_0)\}^2 > 0$ for $\theta \neq \theta_0$.

As mentioned earlier, independence of Y and C is a standard assumption in the random censorship model when no covariables are present. There is no reason to replace this widely accepted assumption by conditional independence when

covariables are present. Condition (ii) may be interpreted as a Markov property of the vector (X, Y, δ) . In terms of Turnbull (1992), (X, Y, δ) constitutes a “serial system” as opposed to the “parallel system”, which typically appears in multivariate lifetime estimation.

A vector (X, Y, C) satisfying (i) and (ii) may be generated in the following way: take (X, Y) from a given joint distribution or, as in our context, through a given regression model (plus noise). In order to determine C and hence Z , generate a Bernoulli random variable δ with

$$\mathbb{P}(\delta = 1 | X = x, Y = y) = 1 - G(y-),$$

where G is the distribution function of C . Given $Y = y$ and $\delta = 1$, C is taken from G restricted (and normalized) to $[y, \infty)$, while for $\delta = 0$, C is distributed according to G restricted to $[0, y)$. It is easily seen that C has distribution G and is independent of Y . Hence (i) and (ii) are satisfied. This algorithm puts no obvious restriction on the joint distribution of X and C . Of course, (ii) is also satisfied if C is independent of (X, Y) . This assumption underlies our simulation results reported in Section 2.

The second moment assumption (iii) is always needed in Least-Squares Estimation. It may be neglected if we consider robustified versions of θ_n . Assumptions (iv) and (v) guarantee that a minimizer θ_n exists. (iv) and (vi) are not needed for linear f 's, while (v) is evidently true in this case. If $f(x, \cdot)$ admits a continuous extension to the compactification of Θ , (iv) becomes superfluous. See Richardson and Bhattacharyya (1986) for a discussion of this issue. Condition (vi) together with (v) and dominated convergence guarantee that all relevant integrals (such as L) are continuous in θ . It is also needed to show that S_n converges with probability one uniformly in $\theta \in \Theta$. Finally, (vii) guarantees that θ_0 may be identified from a sample of (X, Y) 's. It holds true for linear f 's if X has a finite second moment and is not concentrated on a hyperplane.

Technically, consistency and asymptotic normality of θ_n will be shown by first considering a linear approximation of θ_n and then applying a Strong Law of Large Numbers and a Central Limit Theorem for Kaplan-Meier integrals $\int \varphi d\hat{F}_n^0$, respectively, for proper vector-valued functions φ . For example, consider

$$S_n(\theta) = \int \varphi(x, y) \hat{F}_n^0(dx, dy)$$

with $\varphi(x, y) = \{y - f(x, \theta)\}^2$. Recall G and introduce H , the d.f. of the observed Z 's. By (i), $1 - H = (1 - F)(1 - G)$. Let $\tau_H = \inf\{x : H(x) = 1\}$ be the least upper bound for the support of H . Similarly for F and G . Clearly, there will be no data beyond τ_H . So, if $\int \varphi dF^0$ is a parameter of interest, the best we can hope

for is to consistently estimate the integral restricted to $y \leq \tau_H$. More precisely, the theorem in Stute (1993) asserts that with probability one

$$\lim_{n \rightarrow \infty} \int \varphi d\hat{F}_n^0 = \int_{\{Y < \tau_H\}} \varphi(X, Y) d\mathbb{P} + 1_{\{\tau_H \in A\}} \int_{\{Y = \tau_H\}} \varphi(X, \tau_H) d\mathbb{P},$$

where A is the set of H atoms, possibly empty. For $S_n(\theta)$ we get

$$\begin{aligned} \lim_{n \rightarrow \infty} S_n(\theta) &= \int_{\{Y < \tau_H\}} \{Y - f(X, \theta)\}^2 d\mathbb{P} + 1_{\{\tau_H \in A\}} \int_{\{Y = \tau_H\}} \{\tau_H - f(X, \theta)\}^2 d\mathbb{P} \\ &= \int_{\{Y < \tau_H\}} \{\varepsilon + f(X, \theta_0) - f(X, \theta)\}^2 d\mathbb{P} \\ &\quad + 1_{\{\tau_H \in A\}} \int_{\{Y = \tau_H\}} \{\varepsilon + f(X, \theta_0) - f(X, \theta)\}^2 d\mathbb{P}. \end{aligned} \tag{1.2}$$

For most lifetime distributions considered in the literature $\tau_F = \infty = \tau_G$ and therefore $\tau_H = \infty$. In such a situation the above limit is, because of (1.1), $\text{Var}(\varepsilon) + L(\theta, \theta_0)$ which, by (vii), attains its minimum at $\theta = \theta_0$. Stute and Wang (1993) contains a more detailed discussion of situations when the limit of Kaplan-Meier integrals equals the target. If the limit does not happen to coincide with $\text{Var}(\varepsilon) + L(\theta, \theta_0)$, the identifiability function (vii) needs to be replaced by the more general (1.2). For notational convenience we assume throughout without further mention that the limit is $\text{Var}(\varepsilon) + L(\theta, \theta_0)$. To demonstrate the wide applicability of our results, however, the censoring distribution in our simulation study will have compact support included in that of Y , so that (1.2) applies.

Theorem 1.1. *Let θ_n be a minimizer of S_n . Then under (i) – (vii), with probability one,*

$$\lim_{n \rightarrow \infty} \theta_n = \theta_0$$

and

$$\lim_{n \rightarrow \infty} S_n(\theta_n) = \sigma^2 \equiv \text{Var}(\varepsilon).$$

For asymptotic normality, further smoothness assumptions are needed. Consider, in place of (v), the condition

(viii) $f(x, \cdot)$ is twice continuously differentiable.

The matrix

$$\Omega = \left\{ \int \frac{\partial f(X, \theta_0)}{\partial \theta_r} \frac{\partial f(X, \theta_0)}{\partial \theta_s} d\mathbb{P} \right\}_{1 \leq r, s \leq p} \tag{1.3}$$

becomes part of the limit covariance matrix of θ_n , as usual. A change due to censoring comes up in another matrix Π to be defined below. Introduce the sub-distribution functions $\tilde{H}^{11}(x, y) = \mathbb{P}(X \leq x, Z \leq y, \delta = 1)$ and $\tilde{H}^0(y) = \mathbb{P}(Z \leq$

$y, \delta = 0$). These functions may be consistently estimated by their empirical counterparts. Furthermore, put

$$\gamma_0(y) = \exp \left\{ \int_{-\infty}^{y^-} \frac{\tilde{H}^0(dz)}{1 - H(z)} \right\}$$

and, for every real-valued function φ ,

$$\begin{aligned} \gamma_1^\varphi(y) &= \frac{1}{1 - H(y)} \int 1_{\{y < w\}} \varphi(x, w) \gamma_0(w) \tilde{H}^{11}(dx, dw), \\ \gamma_2^\varphi(y) &= \int \int \frac{1_{\{v < y, v < w\}} \varphi(x, w) \gamma_0(w)}{\{1 - H(v)\}^2} \tilde{H}^0(dv) \tilde{H}^{11}(dx, dw). \end{aligned}$$

Note that for a continuous H , $\gamma_0 = (1 - G)^{-1}$, see Stute and Wang (1993). Stute (1996a) then showed that under weak moment assumptions on φ , in probability,

$$\begin{aligned} \int \varphi d(\hat{F}_n^0 - F^0) &= n^{-1} \sum_{i=1}^n \{ \varphi(X_i, Z_i) \gamma_0(Z_i) \delta_i + \gamma_1^\varphi(Z_i) (1 - \delta_i) - \gamma_2^\varphi(Z_i) \} \\ &\quad + o(n^{-1/2}) \equiv n^{-1} \sum_{i=1}^n \psi_i^\varphi + o(n^{-1/2}). \end{aligned} \tag{1.4}$$

The ψ 's are independent and identically distributed with expectation zero. In this paper we shall have to consider several φ 's at the same time, namely

$$\varphi_r(x, y) = \{y - f(x, \theta_0)\} \frac{\partial f(x, \theta_0)}{\partial \theta_r}, \quad 1 \leq r \leq p.$$

Set

$$\mathbb{I} = (\sigma_{rs})_{1 \leq r, s \leq p} \text{ with } \sigma_{rs} = \text{cov}(\psi_1^{\varphi_r}, \psi_1^{\varphi_s}). \tag{1.5}$$

If $\gamma_0, \gamma_1^{\varphi_r}$ and $\gamma_2^{\varphi_r}$ were known \mathbb{I} could be estimated by the sample covariance \mathbb{I}_n of the ψ 's. In practice, of course, they are not. Note, however, that each of the γ 's is a function of the H 's. Replacing these by their empirical counterparts, we arrive at estimated γ 's. Inserting these into \mathbb{I}_n we obtain an estimator of \mathbb{I} which can be computed from the data. An alternative method to estimate \mathbb{I} would be an adaptation of the Jackknife to the regression case, as studied for ordinary Kaplan-Meier integrals in Stute (1996b). This will be done in detail in the example of Section 2.

We finally have to state the assumptions for (1.4). Put

$$T(w) = \int_0^{w^-} \frac{G(dy)}{\{1 - H(y)\}\{1 - G(y)\}}.$$

Assume that for $1 \leq r \leq p$,

$$(ix) \int \{\varphi_r(X, Z)\gamma_0(Z)\delta\}^2 \mathbb{P} < \infty$$

and

$$(x) \int |\varphi_r(X, Z)|T^{1/2}(Z)d\mathbb{P} < \infty.$$

Condition (ix) guarantees that the first term in the expansion (1.4) has a finite second moment. Assumption (x) is mainly to control the bias of a Kaplan-Meier integral. An extensive discussion of this issue may be found in Stute (1994). In particular, it was found that the bias may decrease to zero at any rate $n^{-\delta}$, $\delta < \frac{1}{2}$. Consequently, we cannot expect (1.4) without a first moment condition pushing the bias below $n^{-1/2}$. The function T is related to the variance process of the empirical cumulative hazard function for the censored data. Consult Stute (1995) for further discussion.

Theorem 1.2. *Assume that θ_n is an interior point of Θ . Then, under (i) – (x),*

$$n^{1/2}(\theta_n - \theta_0) \rightarrow \mathcal{N}(0, \Omega^{-1} \prod \Omega^{-1}) \quad \text{in distribution,}$$

where Ω and \prod are given in (1.3) and (1.5).

Observe that if there is no censoring, all δ 's equal 1 and $\gamma_2^\varphi \equiv 0$ so that the expansion (1.4) collapses to the sample mean of $\varphi(X_i, Y_i)$, $1 \leq i \leq n$. Thus, if in addition ε is independent of X , \prod becomes $\sigma^2\Omega$ so that the limit variance simplifies to the familiar $\sigma^2\Omega^{-1}$. Also the function T is identically zero so that (x) is trivially true. Condition (ix) reduces to $\int \varphi_r^2(X, Y)d\mathbb{P} < \infty$. In the general situation, under continuity, the integral (ix) equals

$$\int_{\{Y < \tau_H\}} \varphi_r^2(X, Y)/\{1 - G(Y)\}d\mathbb{P}.$$

Summarizing, we see that consistency holds under no extra conditions on the censoring mechanism while asymptotic normality requires some extra (weak) moment conditions guaranteeing that in the right tails the censoring variables do not dominate the variables of interest.

2. Simulation Study

To demonstrate the validity of our results for finite sample size, we report on a small simulation study with

$$f(x, \theta) = (\theta_1 + \theta_2)^{-1} \exp(\theta_1 x_1 + \theta_2 x_2)$$

as the underlying admissible nonlinear regression functions. The sample (X_i, Y_i) , $1 \leq i \leq n$, was generated as follows: each $X_i = (X_i^1, X_i^2)$ consisted of two independent random variables from the uniform distribution on $[0, 3]$. The noise

variable ε_i was chosen to be standard normal and independent of X_i . The true parameter was set equal to $\theta_0 = (\theta_{01}, \theta_{02})^t = (0.5, 0.3)^t$ so that

$$Y_i = \frac{5}{4} \exp(0.5X_i^1 + 0.3X_i^2) + \varepsilon_i, \quad 1 \leq i \leq n.$$

The censoring distribution G was equal to the uniform distribution on $[0,20]$, $[0,12]$ and $[0,10]$, respectively, corresponding to light, medium and heavy censoring. C_i was taken to be independent of (X_i, Y_i) .

For the minimization of S_n we applied a modified Gauss-Newton algorithm implemented in the National Physical Laboratory Algorithms Library (LSFDN2). Theorem 1.2 will serve as a basis for the computation of confidence regions. The components of the matrix Ω were replaced by the sample means of

$$\left\{ \frac{\partial f(X_i, \theta_n)}{\partial \theta_r} \frac{\partial f(X_i, \theta_n)}{\partial \theta_s} \right\}, \quad 1 \leq i \leq n.$$

Estimation of Π was already discussed in the previous section. Sample sizes were $n = 30, 50$ and 100 . Confidence intervals for θ_{01} and θ_{02} were computed via Scheffé's method. Finally, for each n , 1000 replications were performed and their actual coverage frequencies compared with the nominal coverage level 0.95.

In Table 1 we present the average values of $\theta_n = (\theta_{n1}, \theta_{n2})^t$ in four different situations.

Table 1. Average values of θ_{n1} and θ_{n2} . Target values are $\theta_{01} = 0.5$ and $\theta_{02} = 0.3$.

	$n = 30$	$n = 50$	$n = 100$
no censoring			
θ_{n1}	0.493	0.496	0.499
θ_{n2}	0.304	0.303	0.299
weak censoring			
θ_{n1}	0.493	0.493	0.497
θ_{n2}	0.308	0.301	0.301
medium censoring			
θ_{n1}	0.478	0.481	0.489
θ_{n2}	0.294	0.294	0.296
heavy censoring			
θ_{n1}	0.470	0.477	0.482
θ_{n2}	0.286	0.288	0.294

It becomes clear that the bias slightly increases as censoring becomes more and more substantial. Table 2 presents the associated limit covariance matrix $\Omega^{-1}\Pi\Omega^{-1}$ of $n^{1/2}(\theta_n - \theta_0)$ together with its estimators for different sample sizes and degrees of censoring. Again it turns out that the approximations work well already for small to moderate sample size.

Table 2. Estimated and limit covariance matrices of $n^{1/2}(\theta_n - \theta_0)$.

	$n = 30$	$n = 50$	$n = 100$	$n = \infty$
no censoring	0.032 -0.019 -0.019 0.040	0.034 -0.021 -0.021 0.041	0.035 -0.021 -0.021 0.043	0.031 -0.020 -0.020 0.043
weak censoring	0.050 -0.021 -0.021 0.059	0.057 -0.026 -0.026 0.062	0.055 -0.027 -0.027 0.065	0.040 -0.012 -0.012 0.067
medium censoring	0.076 -0.004 -0.004 0.098	0.087 -0.011 -0.011 0.098	0.078 -0.021 -0.021 0.103	0.072 0.002 0.002 0.133
heavy censoring	0.118 0.011 0.011 0.131	0.116 0.000 0.000 0.129	0.119 -0.005 -0.005 0.154	0.088 0.009 0.009 0.172

In Table 3 we present the actual coverage probabilities for θ_{01} and θ_{02} for Scheffé confidence intervals derived from Theorem 1.2.

Table 3. 95% Confidence Intervals: Coverage Frequencies.

	coverage frequencies of	$n = 30$	$n = 50$	$n = 100$
no censoring	θ_{01}	0.97	0.96	0.96
	θ_{02}	0.96	0.96	0.96
weak censoring	θ_{01}	0.95	0.99	0.98
	θ_{02}	0.91	0.94	0.95
medium censoring	θ_{01}	0.90	0.90	0.94
	θ_{02}	0.92	0.93	0.96
heavy censoring	θ_{01}	0.90	0.92	0.92
	θ_{02}	0.90	0.97	0.96

The actual coverage frequencies are already acceptable under weak to moderate censoring for sample size $n = 30$. To get a visual impression we now present 100 replicates of 95% confidence intervals (CI) for $\theta_{01} = 0.5$ under medium censoring. Due to loss of information, the average lengths of the intervals will slightly increase for heavier censorship.

Appendix. Proofs

Proof of Theorem 1.1. As already mentioned in Section 1, the theorem in Stute (1993) asserts that for each $\theta \in \Theta$ and with probability one,

$$\lim_{n \rightarrow \infty} S_n(\theta) = \int_{\{Y < \tau_H\}} \{Y - f(X, \theta)\}^2 d\mathbb{P} + 1_{\{\tau_H \in A\}} \int \{\tau_H - f(X, \theta)\}^2 d\mathbb{P},$$

which under our standard assumption equals $\sigma^2 + L(\theta_0, \theta)$. Use compactness of Θ , continuity of $f(x, \cdot)$ and (vi) to show that with probability one, $S_n(\theta)$

converges uniformly in θ . See, for example, Theorem 2 in Jennrich (1969). Now, let θ' be any limit point of (θ_n) . Along a subsequence (n_j) we have $\theta_{n_j} \rightarrow \theta'$. By uniform convergence and continuity, $S_{n_j}(\theta_{n_j}) \rightarrow \sigma^2 + L(\theta_0, \theta')$. Since also $S_{n_j}(\theta_0) \rightarrow \sigma^2$ and, by construction of θ_{n_j} , $S_{n_j}(\theta_{n_j}) \leq S_{n_j}(\theta_0)$, we obtain $\sigma^2 + L(\theta_0, \theta') \leq \sigma^2$ and therefore $L(\theta_0, \theta') = 0$. This proves $\theta_0 = \theta'$. Finally, $S_n(\theta_n) \rightarrow \sigma^2$ with probability one.

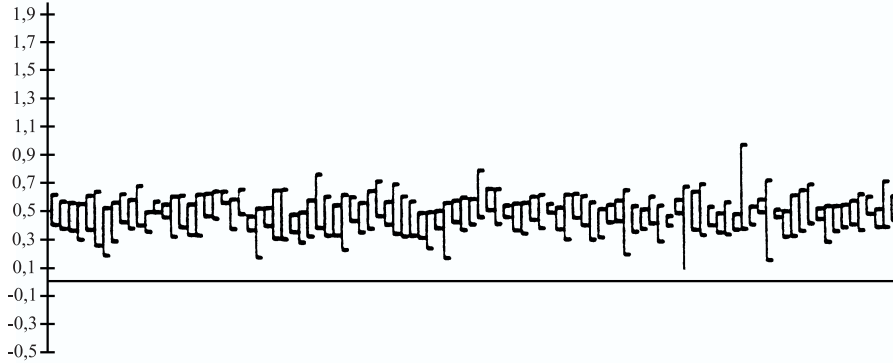


Figure 1. 100 95% CI's for $\theta_{01} = 0.5$, $n = 30$ and medium censoring.

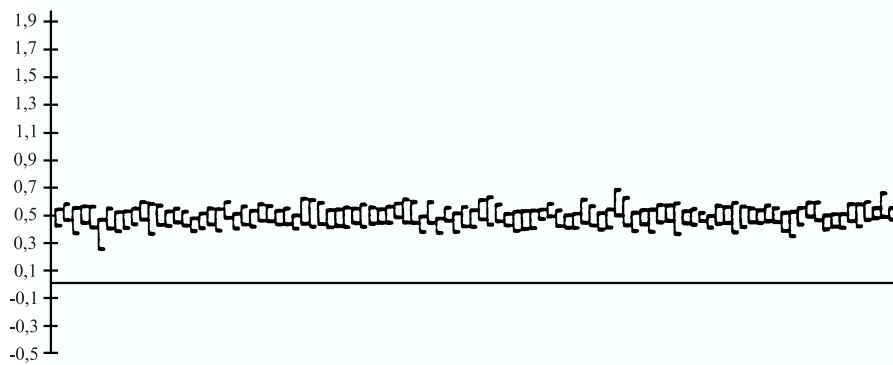


Figure 2. 100 95% CI's for $\theta_{01} = 0.5$, $n = 100$ and medium censoring.

Proof of Theorem 1.2. Since θ_n is an interior point of Θ , there exists some θ_n^1 between θ_n and θ_0 such that

$$n^{1/2}(\theta_n - \theta_0) = -A_n^{-1}n^{1/2}\frac{\partial S_n(\theta_0)}{\partial \theta} \tag{A.1}$$

with

$$A_n = \frac{\partial^2 S_n(\theta_n^1)}{\partial \theta \partial \theta'}.$$

Put

$$b_n = n^{1/2}\frac{\partial S_n(\theta_0)}{\partial \theta}.$$

We first show asymptotic normality of b_n . Actually,

$$b_n = -2n^{1/2} \sum_{i=1}^n W_{in} \left\{ Z_{i:n} - f(X_{[i:n]}, \theta_0) \right\} \frac{\partial f(X_{[i:n]}, \theta_0)}{\partial \theta}$$

is a particular example of a Kaplan-Meier integral as studied in Stute (1996a). Setting

$$\varphi(x, y) = \{y - f(x, \theta_0)\} \frac{\partial f(x, \theta_0)}{\partial \theta},$$

we get

$$b_n = -2n^{1/2} \int \varphi(x, y) \hat{F}_n^0(dx, dy).$$

Note that φ is a vector-valued function. Write $\varphi = (\varphi_1, \dots, \varphi_p)$. From Theorem 1.1 in Stute (1996a) we obtain the representation in probability

$$\begin{aligned} \int \varphi_r(x, y) \hat{F}_n^0(dx, dy) &= n^{-1} \sum_{i=1}^n \varphi_r(X_i, Z_i) \gamma_0(Z_i) \delta_i \\ &\quad + n^{-1} \sum_{i=1}^n \{ \gamma_1^{\varphi_r}(Z_i)(1 - \delta_i) - \gamma_2^{\varphi_r}(Z_i) \} + o(n^{-1/2}), \end{aligned}$$

where $\gamma_1^{\varphi_r}$ and $\gamma_2^{\varphi_r}$ already appeared in Section 1. The second sum consists of independent identically distributed random variables with zero mean. As to the first sum, note that because of $E(\varepsilon_1 | X_1) = 0$,

$$E\{ \varphi_r(X_1, Z_1) \gamma_0(Z_1) \delta_1 \} = E \left[\{ Y_1 - f(X_1, \theta_0) \} \frac{\partial f(X_1, \theta_0)}{\partial \theta_r} \right] = 0.$$

Recalling the definition of \mathbb{I} in (1.5), the multivariate Central Limit Theorem yields

$$b_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, 4 \mathbb{I}) \quad \text{as } n \rightarrow \infty. \quad (\text{A.2})$$

We now study

$$\begin{aligned} A_n &= 2 \sum_{i=1}^n W_{in} \frac{\partial f(X_{[i:n]}, \theta_n^1)}{\partial \theta} \frac{\partial f(X_{[i:n]}, \theta_n^1)}{\partial \theta'} \\ &\quad - 2 \sum_{i=1}^n W_{in} \left\{ Z_{i:n} - f(X_{[i:n]}, \theta_n^1) \right\} \frac{\partial^2 f(X_{[i:n]}, \theta_n^1)}{\partial \theta \partial \theta'}. \end{aligned}$$

Apply a uniform version of the Strong Law of Large Numbers of Stute (1993) and continuity to show that the first sum converges to Ω . Similarly, the second goes to

$$\int \varepsilon \frac{\partial^2 f(X, \theta_0)}{\partial \theta \partial \theta'} d\mathbb{P} = 0.$$

The assertion of Theorem 1.2 therefore follows from (A.1) and (A.2).

References

- Akritas, M. G., Murphy, S. A. and LaValley, M. P. (1995). The Theil-Sen estimator with doubly censored data and applications to astronomy. *J. Amer. Statist. Assoc.* **90**, 170-177.
- Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York-Berlin-Heidelberg.
- Buckley, J. and James, I. R. (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.
- Carter, W. H., Wamper, G. L. and Stablein, D. M. (1983). *Regression Analysis of Survival Data in Cancer Chemotherapy*. Marcel Dekker, New York.
- Doksum, K. A. and Gasko, M. (1990). On a correspondence between models in binary regression analysis and in survival analysis. *Internat. Statist. Rev.* **58**, 243-252.
- Fyngenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data. *Ann. Statist.* **22**, 732-746.
- Jennrich, R. I. (1969). Asymptotic properties of non-linear least-squares estimators. *Ann. Math. Statist.* **40**, 633-643.
- Kalbfleisch, J. D. and Prentice, R. L. (1981). *The Statistical Analysis of Failure Time Data*. John Wiley, New York.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.
- Malinvaud, E. (1970). The consistency of nonlinear regressions. *Ann. Math. Statist.* **41**, 956-969.
- Miller, R. G. (1976). Least squares regression with censored data. *Biometrika* **63**, 449-464.
- Nadaraya, E. A. (1964). On estimating regression. *Theor. Probab. Appl.* **9**, 141-142.
- Richardson, G. D. and Bhattacharyya, B. B. (1986). Consistent estimators in nonlinear regression for a noncompact parameter space. *Ann. Statist.* **14**, 1591-1596.
- Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18**, 303-328.
- Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. John Wiley, New York.
- Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multiv. Analysis* **45**, 89-103.
- Stute, W. (1994). The bias of Kaplan-Meier integrals. *Scand. J. Statist.* **21**, 475-484.
- Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.* **23**, 422-439.
- Stute, W. (1996a). Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.* **23**, 461-471.
- Stute, W. (1996b). The Jackknife estimate of variance of a Kaplan-Meier integral. *Ann. Statist.* **24**, 2679-2704.
- Stute, W. and Wang, J. L. (1993). The strong law under random censorship. *Ann. Statist.* **21**, 1591-1607.
- Tsiatis, A. A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Nat. Acad. Sci. USA* **72**, 20-22.
- Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354-372.
- Turnbull, B. W. (1992). Multivariate failure time analysis: a discussion of papers by Oakes; Pons, Kaddour and de Turckheim; and Prentice and Cai. In *Survival Analysis: State of the Art* (Edited by J. P. Klein and P. K. Goel), 407-414. Kluwer Academic Publishers.
- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26**, 359-372.
- Wu, C. F. (1981). Asymptotic theory of nonlinear least-squares estimation. *Ann. Statist.* **9**, 501-513.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76-99.

Zhou, M. (1992). M-estimation in censored linear models. *Biometrika* **79**, 837-841.

Mathematical Institute, University of Giessen, Arndtstr. 2, D-35392 Giessen, Germany.

E-mail: winfried.stute@math.uni-giessen.de

(Received April 1997; accepted February 1999)