JOURNAL OF MULTIVARIATE ANALYSIS 45, 89–103 (1993)

# Consistent Estimation under Random Censorship When Covariables Are Present

WINFRIED STUTE*

*Math. Institute, University of Giessen, Giessen, Germany*

Assume that $(X_i, Y_i)$, $1 \leqslant i \leqslant n$, are independent $(p+1)$-variate vectors, where each $Y_i$ is at risk of being censored from the right and $X_i$ is a vector of observable covariables. We introduce a $(p+1)$-dimensional extension of the Kaplan–Meier estimator and show its consistency. Also a general strong law for Kaplan–Meier integrals is proved, which, e.g., may be utilized to prove consistency of a new regression parameter estimator under random censorship. 

## 1. INTRODUCTION AND MAIN RESULTS

Suppose that $Y_1, ..., Y_n$ are independent observations from some unknown distribution function (d.f.) $F$ on the real line. A nonparametric efficient estimator of $F$ is then given by the empirical d.f.

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{Y_i \leqslant y\}}, \qquad y \in \mathbb{R}.$$

When analyzing lifetime data it is common that due to censoring effects not all of the $Y$'s are available. To be precise, rather than $Y_i$, one observes

$$Z_i = \min(Y_i, C_i) \qquad \text{and} \qquad \delta_i = 1_{\{Y_i \leqslant C_i\}},$$

where $C_1, ..., C_n$ are independent copies from some unknown censoring d.f. $G$ such that $Y_i$ is independent of $C_i$ for each $1 \leqslant i \leqslant n$. Here $\delta_i$ is an indicator of whether $Y_i$ has been observed or not. The nonparametric analogue of $F_n$ then becomes the time-honoured Kaplan–Meier estimator $\hat{F}_n$ (cf. Shorack and Wellner [24]):

$$1 - \hat{F}_n(y) = \prod_{i=1}^{n} \left[ 1 - \frac{\delta_{[i:n]}}{n-i+1} \right]^{1_{\{Z_{i:n} \leqslant y\}}}.$$

Received November 8, 1991; revised July 28, 1992.

AMS 1990 subject classifications: 62G05, 62G30; 60F15.

Key words and phrases: random censorship, covariables, consistency.

* Work done while on leave at MSRI Berkeley. Supported by the "Deutsche Forschungsgemeinschaft" and NSF Grant DMS 850 5550.



0047-259X/93 $5.00

Here $Z_{1:n} \leqslant \cdots \leqslant Z_{n:n}$ are the ordered $Z$-values, where ties within lifetimes or within censoring times are ordered arbitrarily and ties among lifetimes and censoring times are treated as if the former precedes the later. $\delta_{[i:n]}$ is the concomitant of the $i$th order statistic, i.e., the $\delta$ paired with $Z_{i:n}$.

In this paper we consider the situation when the available data are of the form $(Z_i, \delta_i, X_i)$, $1 \leqslant i \leqslant n$. Here $X_i$ is a $p$-variate vector of covariables paired with the possibly unobserved $Y_i$. It may contain discrete components (as sex, age at entry into study) as well as continuous components (such as measurements on blood). Typically $X_i$ is correlated with $Y_i$. It will then be of outstanding interest to study the joint distribution of $X_1$ and $Y_1$:

$$F^0(x, y) = \mathbb{P}(X_1 \leqslant x, Y_1 \leqslant y), \qquad x \in \mathbb{R}^p, \; y \in \mathbb{R}.$$

Here $X_1 \leqslant x$ is defined coordinatewise.

So far, under random censoring, the existing literature mainly focused on two special cases:

(a) bivariate censorship

(b) censoring in semiparametric models.

Under bivariate censoring the (real-valued) variable $X$ is viewed as the failure time of the second component of a unit rather than a covariable of $Y$. Censoring of both $X$ and $Y$ may occur, e.g., when the unit is removed from the study before both components have been observed to fail. For identifiability reasons it had to be assumed that the (bivariate) censoring vector is independent of $(X, Y)$. In particular, the censoring variable for $Y$ is also independent of $X$. Much work has then been done to find a representation of $F^0$ in terms of estimable quantities. See, e.g., Langberg and Shaked [15] and Dabrowska [8]. Properties of the corresponding estimators may be found in Campbell and Földes [6], Campbell [5], Horváth [10], Burke [3], Lo and Wang [17], Dabrowska [8] and Tsai *et al.* [27]. It is noted that apart from the last none of these estimators is a proper d.f. on $\mathbb{R}^{p+1}$. The Tsai *et al.* [27] estimate requires nonparametric smoothing, since their representation of $F^0$ contains the conditional d.f. of $Y$ given $X$.

Under (b), the vector $X$ is always observable and may be correlated with the censoring mechanism. On the other hand, the dependence between $Y$ and $X$ is assumed to be of a particular semiparametric type. The most famous examples are the Cox-proportional hazards model and the accelerated failure time regression model. See, e.g., Kalbfleisch and Prentice [12], Lawless [16], Ritov [21], Tsiatis [28, 29].

In this paper no such model will be postulated. As such the dependence structure between $X$, $Y$, and $C$ may be quite arbitrary. Also no assumption as to continuity etc. will be required. For the censoring mechanism it will be assumed that

(i) $Y$ and $C$ are independent and $F$ and $G$ have no jumps in common.

Independence is a widely accepted assumption. The jump condition does not exclude discontinuities of $F$ and $G$ at distinct points. As said before, we also wish to allow for dependencies between $X$ and $C$. For identifiability reasons the following condition will be required (write $\mathbb{P}$ for the underlying probability measure):

(ii) $\mathbb{P}(Y \leqslant C \mid X, Y) = \mathbb{P}(Y \leqslant C \mid Y)$.

In other words, (ii) says that given the time of death, the covariables do not provide any further information as to whether censoring will take place or not. This is a convenient way to remind you of the uneasy fact that once $Y$ is known, things which had been considered to be of some importance in your life then become irrelevant. Condition (ii) is satisfied iff $\delta$ and $X$ are independent conditionally on $Y$. Of course, the latter in turn holds true if $C$ is independent of $(X, Y)$, as was postulated under (a).

In this paper, our first goal will be to look for a multivariate extension $\hat{F}_n^0(x, y)$ of the Kaplan–Meier estimator, when the lifetime is at risk of being censored but the covariables are available. In order not to lose the optimality properties of the univariate Kaplan–Meier estimate, $\hat{F}_n^0(x, y)$ should satisfy (in an obvious notation)

$$\hat{F}_n(y) = \hat{F}_n^0(\infty, y), \qquad \infty = (\infty, ..., \infty).$$

Hence the multivariate extension should have Kaplan–Meier as its last marginal. Moreover, $\hat{F}_n^0(x, y)$ should be a proper distribution function on $\mathbb{R}^{p+1}$, possibly defective due to the fact that Kaplan–Meier may attach mass less than one to the real line when the largest $Z$ is censored. We do not spend as much attention to the remaining marginals. This may lead to some loss of efficiency if $\hat{F}_n^0$ is, e.g., utilized to estimate the d.f. of $X_1$. Note, however, that our main concern is about quantities involving both $X$ and $Y$. As such it seems legitimate to put more emphasis on that part of the sample which contains less information.

Another application we have in mind is to construct completely nonparametric measures of association for $X$ and $Y$. The simplest example is the covariance matrix of $(X, Y)$. Finally, we shall be aiming at constructing new estimates for a regression parameter when $(X, Y)$ (resp. $(\ln Y, X)$) happens to satisfy a linear model, which are easy to implement and which are consistent under minimal distributional assumptions.

To cover all the examples at the same time we find it convenient to introduce the weights

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

attached to $Z_{i:n}$ under Kaplan–Meier. Also let $\varphi$ be any real-valued function defined on the Euclidean space $\mathbb{R}^{p+1}$. Set

$$S_n = \sum_{i=1}^{n} W_{in} \varphi(X_{[i:n]}, Z_{i:n}), \tag{1.1}$$

where $X_{[i:n]}$ is the concomitant vector associated with $Z_{i:n}$. There is no question that $S_n$ is well defined when the $Z$'s have no ties. As mentioned before, when ties (among the observed failures) are present, the definition of $\hat{F}_n$ is self-adjusted in that the mass attributed to a tied observation equals the sum of $W$'s attached to each of its replicates irrespective of their order. In the presence of covariables (1.1) needs to be further clarified. In this case ties will be broken up by randomization through independent uniform random variables. We already mention here that this mechanism will be coherent with the arguments required in the proofs to be presented in the second section of this paper.

1.1. EXAMPLE. Putting $\varphi = 1_{(-\infty, x] \times (-\infty, y]}$, we are led to

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

This will be our extension of the univariate Kaplan–Meier estimate to the multivariate setup. Obviously $\hat{F}_n^0$ enjoys all the properties we had postulated before.

1.2. EXAMPLE. Assume that $X$ is univariate (if not take any univariate subcomponent of $X$) and put

$$\varphi_1(x, z) = zx, \quad \varphi_2(x, z) = z, \quad \varphi_3(x, z) = x, \quad \varphi_4(x, z) = z^2, \quad \varphi_5(x, z) = x^2.$$

Denote with $S_n^i$, $1 \leq i \leq 5$, the corresponding $S$-quantities. Combination of these yields estimates for the covariance and correlation of $(X, Y)$.

More examples will be presented after the following theorem. It shows that $S_n$ converges with probability one and in the mean under minimal assumptions on $\varphi$. To describe the limit, a little bit of notation is needed.

Denote with $H$ the d.f. of the observable $Z$'s, and let

$$\tau_H = \inf\{x: H(x) = 1\}$$

be the least upper bound for the support of $H$. Finally we denote with $A$ the (possibly empty) set of jumps (atoms) of $H$.

THEOREM. *Under* (i) *and* (ii), *assume that* $\varphi(X, Y)$ *is integrable. Then with probability one and in the mean*

$$\lim_{n \to \infty} S_n = \int_{\{Y < \tau_H\}} \varphi(X, Y)\, d\mathbb{P} + 1_{\{\tau_H \in A\}} \int_{\{Y = \tau_H\}} \varphi(X, \tau_H)\, d\mathbb{P}. \quad (1.2)$$

1.3. *Remark.* In terms of $\hat{F}_n^0(x, y)$, we have

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

From (1.2) we may infer conditions under which $S_n$ is a consistent estimator for

$$\int \varphi(X, Y)\, d\mathbb{P} = \int \varphi(x, y)\, F^0(dx, dy),$$

the quantity of interest. For this, let $\tau_F$ and $\tau_G$ be defined similarly to $\tau_H$. Of course, $\tau_H = \min(\tau_F, \tau_G)$ by independence of $Y$ and $C$. If $\tau_F < \tau_G$, (1.2) implies that $S_n$ is consistent for any $\varphi$. Since $Y$ is real-valued, consistency also holds whenever $\tau_G = \infty$ irrespective of whether $\tau_F$ is finite or not. If $\tau_G < \tau_F$ consistency does not hold in general (and cannot be obtained by any other estimator) since relevant information about $F$ on $(\tau_G, \tau_F]$ will always be cut off due to censoring. From a purely mathematical point of view, $\tau_F = \tau_G < \infty$ is the most interesting case. Equation (1.2) then nicely features that consistency heavily depends on the local structure of $F$ and $G$ at their common endpoint.

1.4. *Remark.* A convenient way of rewriting (1.2) is to introduce the d.f. (in an obvious notation)

$$\hat{F}^0(x, y) = \begin{cases} F^0(x, y) & \text{if } y < \tau_H \\ F^0(x, \tau_H -) + 1_{\{\tau_H \in A\}} F^0(x, \{\tau_H\}) & \text{if } y \geq \tau_H. \end{cases}$$

Then (1.2) gives

$$\lim_{n \to \infty} \int \varphi(x, y)\, \hat{F}_n^0(dx, dy) = \int \varphi(x, y)\, \hat{F}^0(dx, dy).$$

Application of the theorem to $\varphi$ as considered in Example 1.1 yields the pointwise convergence of $\hat{F}_n^0$ to $\hat{F}^0$ at a single $(x, y)$. A standard uniformity argument leads to the following corollary, which establishes Glivenko–Cantelli convergence of $\hat{F}_n^0$ to $\hat{F}^0$.

1.5. COROLLARY. *Under* (i) *and* (ii), *with probability one*

$$\sup_{x, y} |\hat{F}_n^0(x, y) - \hat{F}^0(x, y)| \to 0 \qquad \text{as} \quad n \to \infty.$$

Another example is obtained by considering the "mean residual lifetime function of $Y$ subject to constraints on $X$." For this, let $\mathscr{R}$ be an arbitrary (measurable) subset of $\mathbb{R}^p$. Set, for $y \geqslant 0$,

$$\gamma(y) = \gamma(y, \mathscr{R}) = \begin{cases} \int_{\{Y > y, X \in \mathscr{R}\}} (Y - y)\, d\mathbb{P}/\mathbb{P}(Y > y, X \in \mathscr{R}), \\ \qquad \text{if} \quad \text{the denominator is positive} \\ 0, \qquad \text{otherwise} \end{cases} \tag{1.3}$$

(assuming of course that $Y$ has a finite mean).

Informally, $\gamma(y)$ is the mean residual lifetime of a person who has already reached age $y$ subject to the condition that the covariable vector describing his status belongs to $\mathscr{R}$. For example, $\mathscr{R}$ may consist of all vectors with 1 in their first coordinate, symbolizing the fact that a patient had obtained a special treatment (a 0 indicating no treatment). We may also wish to let $\mathscr{R}$ vary so that $\gamma$ becomes a function of $y$ and $\mathscr{R}$.

The theorem implies that the numerator and the denominator may be consistently estimated under conditions which have been discussed in Remark 1.3 above. For this set

$$\gamma_n^1(y, \mathscr{R}) = \sum_{i=1}^{n} W_{in}(Z_{i:n} - y)\, 1_{\{Z_{i:n} > y, X_{[i:n]} \in \mathscr{R}\}},$$

$$\tau_n^2(y, \mathscr{R}) = \sum_{i=1}^{n} W_{in}\, 1_{\{Z_{i:n} > y, X_{[i:n]} \in \mathscr{R}\}}$$

and

$$\gamma_n(y, \mathscr{R}) = \gamma_n^1(y, \mathscr{R})/\gamma_n^2(y, \mathscr{R}) \qquad \text{(whenever defined)}.$$

1.6. Corollary. *Under* (i), (ii), *and* (*e.g.*) $\tau_F < \tau_G$ (*resp.* $\tau_F = \tau_G = \infty$), *we have*

$$\lim_{n \to \infty} \gamma_n(y, \mathscr{R}) = \gamma(y, \mathscr{R}) \qquad \textit{with probability one.}$$

The assertion of the corollary may be extended to uniform convergence if the class of sets $(y, \infty) \times \mathscr{R}$ is a universal uniformity class as considered by Stute [25], provided that the denominators in (1.3) are bounded away from zero. Universal uniformity classes are obtained, e.g., if for $\mathscr{R}$ we take all rectangles, balls, or halfspaces.

The next example treats rank statistics for measuring the amount of association between $X$ and $Y$. For notational convenience take $p = 1$. Let

$J$ be any continuous score function on the unit square $[0, 1] \times [0, 1]$. Denote with $\hat{F}_n^{01}$ and $\hat{F}_n^{02}$ the marginals of $\hat{F}_n^0$, and set

$$T_n = \iint J(\hat{F}_n^{01}(x), \hat{F}_n^{02}(y))\, \hat{F}_n^0(dx, dy).$$

$T_n$ is a statistic which has been designed, e.g., to test the hypothesis

$$H_0 : X \text{ is independent of } Y$$

and which is asymptotically distribution-free under $H_0$ (assuming continuity of the underlying d.f.). For uncensored data there is a huge literature on $T_n$. See, e.g., Bhuchongkul [1] and Ruymgaart [23]. For censored data we refer to Dabrowska [7], Gombay [9], and Pons [20], who investigated tests of independence under bivariate censoring.

To the best of our knowledge, if censoring is allowed to depend on the vector of covariables, no completely nonparametric approach has been analyzed. Corollary 1.7 below treats almost sure convergence of $T_n$.

1.7. COROLLARY. *Under the assumptions of Corollary* 1.6, *suppose that* $\hat{F}^0$ *is continuous. Then, under* $H_0$,

$$\lim_{n \to \infty} T_n = \int_0^1 \int_0^1 J(u, v)\, du\, dv \qquad \textit{with probability one.}$$

*Proof.* Follows from 1.5 and the (uniform) continuity of $J$, upon utilizing the fact that under $H_0$ and by continuity of $\hat{F}^0$,

$$\iint J(\hat{F}^{01}(x), \hat{F}^{02}(y))\, \hat{F}^0(dx, dy) = \int_0^1 \int_0^1 J(u, v)\, du\, dv. \quad \blacksquare$$

If $H_0$ is not true, the limit also exists but may depend on $\hat{F}^0$ in terms of the so-called copula function.

Unlike in our previous examples it will now be assumed that $(X, Y)$ (resp. a known monotone transformation of $Y$ such as $\ln Y$) satisfies a linear regression model:

$$Y_i = X_i^t \beta + \varepsilon_i, \qquad 1 \leqslant i \leqslant n.$$

We shall assume without further mentioning that $\mathbb{E}[\varepsilon \mid X] = 0$, dependencies being allowed otherwise. We now propose a new estimator $\beta_n$ of the regression vector $\beta$ and show its consistency, subject to censoring of the $Y$'s. $\beta_n$ coincides with the LSE if there is no censoring. It is easily implemented once the weights have been computed. We only mention that robust versions of $\beta_n$ may also be available.

Now, define the matrices

$$M_{1n}(i,j)=\sum_{k=1}^{n} W_{kn}X^i_{[k:n]}X^j_{[k:n]}, \qquad 1\leqslant i,j\leqslant p$$

$$M_{2n}(i,s)=W_{sn}X^i_{[s:n]}, \qquad 1\leqslant i\leqslant p,\ 1\leqslant s\leqslant n$$

and

$$\tilde{Z}_n=(Z_{1:n},...,Z_{n:n})^t,$$

where superscript t denotes transposition and $X^i_{[s:n]}$ is the $i$th coordinate of $X_{[s:n]}$. Put

$$\beta_n=M_{1n}^{-1}M_{2n}\tilde{Z}_n.$$

It is easily seen that $\beta_n$ minimizes the function

$$\beta\to\sum_{i=1}^{n} W_{in}[Z_{i:n}-X^t_{[i:n]}\beta]^2;$$

i.e., $\beta_n$ is a weighted LSE. The weights $\{W_{in}: 1\leqslant i\leqslant n\}$ do not satisfy the assumptions which are usually required in weighted least squares estimation. In such a situation our theorem will be crucial to get consistency of $\beta_n$.

1.8. COROLLARY. *Under the assumptions of Corollary* 1.6, *suppose that* $\mathbb{E}[XX^t]$ *exists and is regular. Then we have with probability one*

$$\lim_{n\to\infty}\beta_n=\beta.$$

*Proof.* Apply the theorem to obtain with probability one

$$M_{1n}\to\mathbb{E}[XX^t] \quad\text{and}\quad M_{2n}\tilde{Z}_n\to\mathbb{E}[XX^t]\beta. \quad ∎$$

There exists an extensive literature on censored linear regression. Basic references are Miller [18], Buckley and James [2], Koul *et al.* [13], James and Smith [11], and Lai and Ying [14]. In most of this work the covariables are nonrandom. The corresponding estimators need to be computed iteratively and consistency is obtained under particular model assumptions.

Cadarso Suárez [4] contains, among other things, a comprehensive simulation study for various estimators. To demonstrate the validity of our approach, we compare $\beta_n$ with the Miller and Buckley–James estimators. The underlying model is

$$Y=3+X^1+X^2+\varepsilon$$

$$C=3+X^1+X^2+\tilde{\varepsilon},$$

where $X^1$, $X^2$ are independent uniform [0, 1]-random variables and $\varepsilon$, $\tilde{\varepsilon}$ are independent $\mathcal{N}(0, \sigma^2)$ and $\mathcal{N}(\Delta, \sigma^2)$ variables, respectively, being also independent of $(X^1, X^2)$. In Table I we present, under various sampling situations, the multivariate MSEs based on 200 replications. Rather than $\Delta$, the percentage of censoring is given.

A particular feature of the underlying model is that the censoring variable has the same regression function on $X$ as $Y$. Under such an assumption Miller's estimate is known to work pretty well. Our simulation results indicate that $\beta_n$ outperforms the Buckley–James estimator in most situations and is better than Miller's estimator for a small $\sigma^2$.

TABLE I

Multivariate MSEs

| $\sigma^2$ | $\beta_n$ | Miller | Buckley-James |
|---|---|---|---|
| | $n = 25$, percentage of censoring 16% | | |
| 0.1 | 0.01620 | 0.02581 | 0.27376 |
| 0.5 | 0.45635 | 0.40053 | 0.69927 |
| 1 | 1.85398 | 1.57567 | 1.95569 |
| | $n = 25$, percentage of censoring 33% | | |
| 0.1 | 0.02210 | 0.03944 | 0.36190 |
| 0.5 | 0.61857 | 0.52823 | 0.95236 |
| 1 | 2.53757 | 2.04109 | 2.63116 |
| | $n = 25$, percentage of censoring 50% | | |
| 0.1 | 0.03250 | 0.06487 | 0.97369 |
| 0.5 | 0.85625 | 0.74603 | 1.71338 |
| 1 | 3.49725 | 2.87294 | 4.39953 |
| | $n = 50$, percentage of censoring 16% | | |
| 0.1 | 0.00658 | 0.01607 | 0.02457 |
| 0.5 | 0.17687 | 0.16842 | 0.23412 |
| 1 | 0.73469 | 0.65419 | 0.70311 |
| | $n = 50$, percentage of censoring 33% | | |
| 0.1 | 0.00999 | 0.02078 | 0.05981 |
| 0.5 | 0.26777 | 0.25393 | 0.30879 |
| 1 | 1.15074 | 0.97448 | 1.00613 |
| | $n = 50$, percentage of censoring 50% | | |
| 0.1 | 0.01562 | 0.03738 | 0.11376 |
| 0.5 | 0.38158 | 0.35452 | 0.40308 |
| 1 | 1.35116 | 1.33385 | 1.24980 |

## 2. PROOFS

A traditional and convenient way to analyze the large sample behavior of the Kaplan–Meier estimator is to reduce the analysis, via the exp–ln transformation, to the study of the cumulative hazard function. This approach works satisfactorily if $y$ is restricted to $y \leqslant T$, with $T < \tau_H$. Due to the unstable behavior of the cumulative hazard function estimator in the tails of $H$, things get worse (in proofs) if one wants to establish strong uniform convergence on $[0, \tau_H]$. What is more, since for statistical applications the SLLN is needed for a whole bunch of $\varphi$'s and not just for indicators $\varphi = 1_{[0,t]}$, a reduction to the cumulative hazard function seems hopeless in the general case. In Stute and Wang [26] a new approach has been proposed which avoids such transformations. Rather we were able to treat sums of the form (cf. (1.1))

$$S_n = \sum_{i=1}^{n} W_{in} \varphi(Z_{i:n}) = \int \varphi \, d\hat{F}_n \tag{2.1}$$

directly for an arbitrary (integrable) $\varphi$ by analyzing in detail the joint probabilistic structure of the $W$'s and $\varphi(Z)$'s.

As will be clear now, the present paper constitutes an extension of that work to the effect, that also covariables may be included in the analysis. In order to save space, we shall feel free to adopt facts (appropriately modified) from Stute and Wang [26] whenever it will be convenient.

A nice feature about (2.1) is (apart from the many ugly ones) the fact that $S_n$ may be written as a function of the $Z$-order statistics and their $\delta$-concomitants. This is a property which both (1.1) and (2.1) have in common, provided we are willing to include $X_{[i:n]}$ in the concomitant (vector) of $Z_{i:n}$. The probabilistic structure of order statistics and their concomitants is described in the following lemma. Part (a) of it, under further regularity assumptions, is due to Yang [30]. Since in our applications the concomitants may (and will!) contain discrete components (e.g., $\delta$), the more general version is needed for our purposes.

2.1. LEMMA. *Let $(Z_i, D_i)$, $1 \leqslant i \leqslant n$, be independent random vectors from the same $1+m$-variate d.f., with conditional d.f.*

$$m(y \mid z) = \mathbb{P}(D_i \leqslant y \mid Z_i = z).$$

*Let $Z_{1:n} \leqslant \cdots \leqslant Z_{n:n}$ denote the ordered values of the $Z$'s, and let $D_{[i:n]}$ be the ith concomitant vector paired with $Z_{i:n}$. Then*

(a) *Conditionally on $Z_{1:n} \leqslant \cdots \leqslant Z_{n:n}$, the concomitants are independent with the same conditional distributional structure as the pair $(Z_1, D_1)$, i.e.,*

$$\mathbb{P}(D_{[i:n]} \leqslant y \mid Z_{i:n} = z) = m(y \mid z).$$

(b) *If the Z's have a continuous d.f., then the vector* $(Z_{i:n}, D_{[i:n]})_{1\leqslant i\leqslant n}$ *is independent of the vector of Z-ranks.*

*Proof.* Lemma 2.1 may be proved similarly to the corresponding lemma in Stute and Wang [26]. Just replace the bivariate version of Rosenblatt's [22] lemma by its $(1+m)$-dimensional analogue. ∎

Recall that

$$S_n=\sum_{i=1}^{n} W_{in}\varphi(X_{[i:n]}, Z_{i:n}),$$

where for $1\leqslant i\leqslant n$,

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

and $X_{[i:n]}$ is the vector of covariables associated with $Z_{i:n}$. In proofs we shall apply Lemma 2.1 to $D_{[i:n]}\equiv(X_{[i:n]}, \delta_{[i:n]})$ and $m=p+1$. Put, for $n\geqslant 1$,

$$\mathscr{F}_n=\sigma(Z_{i:n}, D_{[i:n]}, 1\leqslant i\leqslant n, Z_{n+1}, D_{n+1}, ...).$$

Clearly, $S_n$ is adapted to $\mathscr{F}_n$, and $\mathscr{F}_n\downarrow\mathscr{F}_\infty=\bigcap_{n\geqslant 1}\mathscr{F}_n$. We shall consider the case $\varphi\geqslant 0$ and $H$ continuous first.

2.2. LEMMA. *For* $\varphi\geqslant 0$ *and continuous* $H$, $\{S_n, \mathscr{F}_n\}_{n\geqslant 1}$ *is a reverse-time supermartingale, i.e.,*

$$\mathbb{E}(S_n\mid\mathscr{F}_{n+1})\leqslant S_{n+1}.$$

*Proof.* Utilizing Lemma 2.1, we may follow the lines of Lemma 2.2 in Stute and Wang [26] to show that for some (random) quantities $a_{1,n+1}, ..., a_{n+1,n+1}$,

$$\mathbb{E}(S_n\mid\mathscr{F}_{n+1})=\sum_{i=1}^{n+1} a_{i,n+1}\varphi(X_{[i:n+1]}, Z_{i:n+1}).$$

It has been shown there that

$$a_{i,n+1}=W_{i,n+1}\qquad\text{for}\quad 1\leqslant i\leqslant n$$

and

$$a_{n+1,n+1}\leqslant W_{n+1,n+1}.$$

Since $\varphi\geqslant 0$, the proof is complete. ∎

With Lemma 2.2 at hand, we are in the position to apply powerful convergence theorems for reverse-time supermartingales. See, e.g., Proposition V-3-11 in Neveu [19]. Moreover, since $\mathscr{F}_\infty = \bigcap_{n\geq 1} \mathscr{F}_n$ is trivial by the Hewitt–Savage 0–1 law, the limit $S$ is constant with probability one. To determine $S$, we put

$$m(t) = \mathbb{P}(\delta = 1 \mid Z = t),$$

$$\varphi_n(t) = \prod_{i=1}^{n} \left(1 + \frac{1 - m(Z_{i:n})}{n-i+1}\right)^{1\{Z_{i:n} < t\}},$$

and

$$g_n(t) = \mathbb{E}\varphi_n(t); \qquad g_0(t) \equiv 1.$$

Finally, set

$$\tilde{\varphi}(z) = \mathbb{E}[\varphi(X, Z)\,\delta \mid Z = z].$$

2.3. Lemma. *Under the assumptions of Lemma 2.2, we have*

$$\mathbb{E}S_n = \mathbb{E}[\tilde{\varphi}(Z)\, g_{n-1}(Z)].$$

*Proof.* Let $R_{jn}$ denote the rank of $Z_j$ among $Z_1, ..., Z_n$. It follows from Lemma 2.1 that

$$\begin{aligned}
\mathbb{E}S_n &= \mathbb{E}\left[\sum_{i=1}^{n} W_{in}\varphi(X_{[i:n]}, Z_{i:n})\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \frac{1}{n-i+1}\,\mathbb{E}\left[\varphi(X_{[i:n]}, Z_{i:n})\,\delta_{[i:n]}\right.\right. \\
&\qquad \left.\left.\times \prod_{j=1}^{i-1}\left(\frac{n-j}{n-j+1}\right)^{\delta_{[j:n]}} \,\middle|\, Z_{1:n}, ..., Z_{n:n}\right]\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \frac{1}{n-i+1}\,\tilde{\varphi}(Z_{i:n}) \prod_{j=1}^{i-1}\left(1 - \frac{m(Z_{j:n})}{n-j+1}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \frac{\tilde{\varphi}(Z_{i:n})}{n} \prod_{j=1}^{i-1}\left(1 + \frac{1-m(Z_{j:n})}{n-j}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \frac{\tilde{\varphi}(Z_i)}{n} \prod_{j=1}^{n}\left(1 + \frac{1-m(Z_j)}{n-R_{jn}}\right)^{1\{Z_j < Z_i\}}\right] \\
&= \mathbb{E}\left[\tilde{\varphi}(Z_1) \prod_{j=1}^{n}\left[1 + \frac{1-m(Z_j)}{n-R_{jn}}\right]^{1\{Z_j < Z_1\}}\right].
\end{aligned}$$

Since $R_{jn} = R_{j,n-1}$ on $\{Z_j < Z_1\}$, the result follows easily by conditioning on $Z_1$. ∎

It has also been shown in Stute and Wang [26] that

$$g_n(t) \uparrow \frac{1}{1-G(t)} \qquad \text{for each } t \text{ such that } H(t) < 1. \tag{2.2}$$

2.4. LEMMA. *For a continuous $H$ and any integrable $\varphi(X, Y)$, we have*

$$S_\infty := \lim_{n \to \infty} \mathbb{E} S_n = \int_{\{Y < \tau_H\}} \varphi(X, Y)\, d\mathbb{P}.$$

*Proof.* Assume $\varphi \geqslant 0$ w.l.o.g. The general case is obtained by decomposing $\varphi$ into its positive and negative part. Apply (2.2), Lemma 2.3, and the monotone convergence theorem to obtain

$$\begin{aligned} S_\infty &= \int_{\{Z < \tau_H\}} \tilde{\varphi}(Z)/(1-G(Z))\, d\mathbb{P} \\ &= \int_{\{Y < \tau_H\}} \varphi(X, Y)\, 1_{\{Y \leqslant C\}}/(1-G(Y))\, d\mathbb{P}. \end{aligned}$$

Since $G$ along with $H$ is also continuous, conditioning on $(X, Y)$ together with (ii) leads to the assertion of Lemma 2.4. ∎

We are now in the position to give the

*Proof of the Theorem.* As indicated earlier, for a continuous $H$ the proof may be established by combining Lemma 2.2, Lemma 2.4, and Proposition V-3-11 in Neveu [19] to obtain with probability one and in the mean:

$$\lim_{n \to \infty} S_n = \lim_{n \to \infty} \mathbb{E} S_n = S = S_\infty.$$

The case of an arbitrary not necessarily continuous $H$ may be easily traced back to the continuous (in fact uniform) case by incorporating the randomization mentioned in the Introduction. To determine the limit just apply the theorem to

$$\varphi^0(x, u) = \varphi(x, H^{-1}(u)), \qquad m^0(u) = m \circ H^{-1}(u)$$

and

$$\tilde{\varphi}^0(u) = \mathbb{E}[\varphi^0(X, U)\, \delta \mid U = u] = \tilde{\varphi} \circ H^{-1}(u), \qquad 0 < u < 1.$$

Derivation of $S$ then becomes straightforward. ∎

## ACKNOWLEDGMENT

I am very grateful to Carmen Cadarso Suárez for her assistance with the simulations.

## REFERENCES

[1] BHUCHONGKUL, S. (1964). A class of nonparametric tests for independence in bivariate populations. *Ann. Math. Statist.* **35** 138–149.
[2] BUCKLEY, J., AND JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66** 429–436.
[3] BURKE, M. D. (1988). An Almost Sure Approximation of a Multivariate Product-Limit Estimator under Random Censorship. *Statist. and Decisions* **6** 89–108.
[4] CADARSO SUÁREZ, C. M. (1990). *Nuevos Aportes a la Regresión Paramétrica y no Paramétrica con Datos Censurados.* Ph.D. thesis, University of Santiago de Compostela, Spain.
[5] CAMPBELL, G. (1982). Asymptotic properties of several nonparametric multivariate distribution function estimators under random censorship. In *Survival Analysis* (J. Crowley and R. A. Johnson, Eds.), pp. 243–256. IMS, Hayward, CA.
[6] CAMPBELL, G., AND FÖLDES, A. (1982). Large-sample properties of nonparametric bivariate estimators with censored data. In *Nonparametric Statistical Inference* (B. V. Gnedenko, M. L. Puri, and I. Vincze, Eds.), pp. 103–122. North-Holland, Amsterdam.
[7] DABROWSKA, D. M. (1986). Rank tests for independence for bivariate censored data. *Ann. Statist.* **14** 250–264.
[8] DABROWSKA, D. M. (1988). Kaplan–Meier estimate on the plane. *Ann. Statist.* **16** 1475–1489.
[9] GOMBAY, E. (1988). A test of independence for the coordinates of bivariate censored data. *J. Multivariate Anal.* **25** 223–240.
[10] HORVÁTH, L. (1983). The rate of strong uniform consistency for the multivariate product-limit estimator. *J. Multivariate Anal.* **13** 202–209.
[11] JAMES, I. R., AND SMITH, P. J. (1984). Consistency results for linear regression with censored data. *Ann. Statist.* **12** 590–600.
[12] KALBFLEISCH, J. D., AND PRENTICE, R. L. (1981). *The Statistical Analysis of Failure Time Data.* Wiley, New York.
[13] KOUL, H. L., SUSARLA, V., AND VAN RYZIN, J. (1982). Multi-step estimation of regression coefficients in a linear model with censored survival data. In *Survival Analysis* (J. Crowley and R. A. Johnson, Eds.), pp. 86–100. IMS, Hayward, CA.
[14] LAI, T. L., AND YING, Z. (1991). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19** 531–556.
[15] LANGBERG, N. A., AND SHAKED, M. (1982). On the identifiability of multivariate life distribution functions. *Ann. Probab.* **10** 773–779.
[16] LAWLESS, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* Wiley, New York.
[17] LO, S. H., AND WANG, J. L. (1989). I.I.D. Representations for the bivariate product limit estimators and the bootstrap versions. *J. Multivariate Anal.* **28** 211–226.
[18] MILLER, R. G. (1976). Least squares regression with censored data. *Biometrika* **63** 449–464.
[19] NEVEU, J. (1975). *Discrete-Parameter Martingales.* North-Holland, Amsterdam/Oxford.
[20] PONS, O. (1986). A test of independence between two censored survival times. *Scand. J. Statist.* **13** 173–185.

[21] RITOV, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18** 303–328.

[22] ROSENBLATT, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.* **23** 470–472.

[23] RUYMGAART, F. H. (1973). *Asymptotic Theory of Rank Tests for Independence.* Tracts, Vol. 43. Math. Centre, Amsterdam.

[24] SHORACK, G. R., AND WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[25] STUTE, W. (1976). On a generalization of the Glivenko–Cantelli theorem. *Z. Wahrsch. Verw. Gebiete* **35** 167–175.

[26] STUTE, W., AND WANG, J. L. (1993). The strong law under random censorship. *Ann. Statist.* **21**, in press.

[27] TSAI, W. Y., LEURGANS, S., AND CROWLEY, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.* **14** 1351–1365.

[28] TSIATIS, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **8** 93–108.

[29] TSIATIS, A. A. (1990). Estimation regression parameters using linear rank tests for censored data. *Ann. Statist.* **18** 354–372.

[30] YANG, S. S. (1977). General distribution theory of the concomitants of order statistics. *Ann. Statist.* **5** 996–1002.

683 45 1-8