

ASYMPTOTIC NORMALITY OF KOUL-SUSARLA-VAN RYZIN ESTIMATOR USING  
COUNTING PROCESS

by

MAI ZHOU

Department of Statistics  
University of North Carolina  
Chapel Hill, NC 27599-3260  
USA

**Abstract**

To estimate the covariates in a linear model based on censored survival data, Koul, Susarla & Van Ryzin (1981) proposed a generalization of the ordinary least squares estimator. This paper uses counting process and martingales techniques to provide a new proof of the asymptotic normality of the estimator. We represent the estimator as a martingale plus a high order term. This approach allows us to relax some assumptions made by Koul, Susarla & Van Ryzin. We also find that the asymptotic variance is better than they thought it was.

**Key Words and Phrases:** Censored data, linear regression, counting processes, central limit theorem, asymptotic variance.

**Section 0. Introduction**

Suppose the patients' survival times (or its logarithm)  $T_i$  are random variables, independent of each other and follow the linear model

$$T_i = \alpha + \beta x_i + \epsilon_i \quad i = 1, 2, \dots, n. \quad (0.1)$$

where  $x_i$ 's are known and  $\epsilon_i$ 's are (iid) r.v.'s with zero mean finite variance. Let  $F_i(t) = P(T_i < t)$ .  $(\alpha, \beta)$  are covariates, (e.g. age, indicator of treatment, sex, etc.) to be estimated. We only observe the pair

$$Z_i = \min\{T_i, Y_i\}, \quad \delta_i = [T_i \leq Y_i]$$

where  $Y_i$ 's are iid censoring times, r.v.'s independent of  $\epsilon_i$  and  $P(Y_i < t) = G(t)$ .

Koul, Susarla & Van Ryzin (1981) suggested that one use the following to estimate  $\alpha, \beta$  based on  $\{(Z_i, \delta_i)_{i=1}^n\}$ ,

$$\hat{\beta} = \sum b_{ni} \frac{\delta_i Z_i}{1 - \hat{G}(Z_i)} I_{[Z_i \leq M_n]} \quad (0.2)$$

$$\hat{\alpha} = \sum a_{ni} \frac{\delta_i Z_i}{1 - \hat{G}(Z_i)} I_{[Z_i \leq M_n]}, \quad (0.3)$$

where  $b_{ni} = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$ ,  $a_{ni} = \frac{1}{n} - \bar{x} b_{ni}$ ,  $\bar{x} = \frac{1}{n} \sum X_i$ ;

and  $\hat{G}$  is a Kaplan-Meier (1958) product-limit type estimator of the distribution  $G(t)$  of  $Y$ 's.  $M_n$  is a sequence of constants ( $\rightarrow \infty$ ) satisfy certain conditions. This estimator is particularly easy to implement on computer, it can use the

existing package with a slight modification.

Notice if  $\delta_i \equiv 1$ ,  $\hat{G} \equiv 0$  (no censoring) and  $M_n = \infty$ , the estimator (0.2), (0.3) reduces to the usual least squares estimator.

In order to make inference on  $\alpha$ ,  $\beta$ , we not only need a consistent estimate, but their distributions or asymptotic distributions. Thus a central limit theorem for  $\hat{\alpha}$ ,  $\hat{\beta}$  are important. The proof of Asymptotic Normality in their (1981) paper Koul, Susarla & Van Ryzin used a U-statistic argument, which boils down to approximate the estimator by a U-statistic and then applying Hoeffding (1948). Our approach uses the martingale structure of the counting processes associated with the estimator and represents the estimator as an martingale plus a term tending to zero. This approach naturally involves time and can be adapted for sequential analysis of the model. The truncation sequence  $M_n$  in (0.2), (0.3), which were "constants depending on the unknown  $F_i(t)$  and  $G(t)$  in a complicated way" and "further guidelines and experience are needed in the proper choice of  $M_n$ " (Miller & Halpern (1982), Gill (1983)) are replaced here by some (observable) order statistic of the  $Z_i$ 's. In addition, our discussion on the asymptotic variance reveals that Koul, Susarla & Van Ryzin (1981) formula (3.7) needs an extra  $n$  factor in the second term there. Thus their asymptotic variance estimator need to be fixed. Besides, this fact suggest some interesting consequences that can be explored to improve the estimation of covariate  $(\alpha, \beta)$ .

Section 1 below contains additional notation and some counting process results that are useful later.

Section 2 contains the asymptotic normality theorem and its proof, but the treatment of high order term is deferred to section 4, so that one can better concentration on the martingale representation part of the proof. Extension to multiple regression is also discussed.

Section 3 takes a closer look at the asymptotic variance derived in section 2, and points out a correction on a formula in Koul, Susarla & Van Ryzin (1981) and remarked some consequences thereof.

**Section 1. Notation and Counting Processes.**

We now introduce some additional notation and establish some simple facts that will be useful later.

For  $i = 1, 2, \dots, n$ ; let

$$1 - H_i(t) = P(Z_i \geq t) = [1 - F_i(t)][1 - G(t)]$$

And let

$$1 - \hat{H}_i(t) = I_{[Z_i \geq t]}; \quad y^+(t) = \sum_{i=1}^n I_{[Z_i \geq t]} = \Sigma(1 - \hat{H}_i); \quad \text{and } T^n = \max_i \{Z_i\} \quad (1.4)$$

Also, let

$$\Lambda_i^+(t) = \int_{[0, t]} \frac{dH_i(s)}{1 - H_i(s-)}; \quad \Lambda_i^D(t) = \int_{[0, t]} \frac{dF_i(s)}{1 - F_i(s-)}; \quad \Lambda^C(t) = \int_{[0, t]} \frac{dG(s)}{1 - G(s-)}$$

It is well known that the three processes

$$M_i^+(t) = I_{[Z_i < t]} - \int_0^t I_{[Z_i > s]} d \Lambda_i^+(s) \quad (1.5)$$

$$M_i^D(t) = I_{[Z_i < t; \delta_i = 1]} - \int_0^t I_{[Z_i > s]} d \Lambda_i^D(s) \quad (1.6)$$

and

$$M_i^C(t) = I_{[Z_i < t; \delta_i = 0]} - \int_0^t I_{[Z_i > s]} d \Lambda^C(s) \quad (1.7)$$

are all square integrable martingales with respect to the filtration

$$\mathcal{F}_s = \sigma\{Z_k I_{[Z_k \leq s]}; \delta_k I_{[Z_k \leq s]}; k = 1, 2, \dots, n.\}$$

and

$$\langle M_i^D \rangle(t) = \int_0^t I_{[Z_i > s]} d \Lambda_i^D(s); \quad \langle M_i^C \rangle(t) = \int_0^t I_{[Z_i > s]} d \Lambda^C(s). \quad (1.8)$$

Clearly  $M_i^+ = M_i^D + M_i^C$ , since  $\Lambda_i^+ = \Lambda_i^D + \Lambda_i^C$ . And we define  $M_c^+ = \sum M_i^C$ .

The Kaplan-Meier estimator  $\hat{G}(t)$  can be defined as

$$\hat{G}(t) = 1 - \prod_{s \leq t} \left( 1 - \frac{\Delta N_c^+(s)}{Y^+(s)} \right),$$

where

$$N_c^+(s) = \sum_{i=1}^n I_{[Z_i < s; \delta_i = 0]}.$$

Thus the processes

$$\frac{\hat{G}(t) - G(t)}{1 - G(t)} = \int_0^{t \wedge T^n} \frac{1 - \hat{G}}{1 - G} \frac{1}{Y^+(s)} dM_c^+(s); \quad t \in [0, T^n] \quad \text{and} \quad (1.9)$$

$$\frac{\hat{H}_i(t) - H_i(t)}{1 - H_i(t)} = \int_0^t \frac{1}{1 - H_i} dM_i^+(s) \quad \text{for } t \text{ such that } 1 - H_i(t) > 0 \quad (1.10)$$

are martingales. Notice that the right hand side of (1.9) is well defined even for  $t > T^n$ , a constant equal to its value at  $T^n$ , we shall use this definition to extend the domain of the martingale so that both martingales are now defined on  $[0, \infty)$  which brings some convenience later.

Lemma 0.1 The cross variation process of the martingales (1.6) and (1.7) is zero, i.e.  $\langle M_i^D, M_i^C \rangle = 0$ .

*Proof:* Notice that  $M_i^D + M_i^C$  is also a martingale - a compensated counting process with intensity  $\int_0^t I_{[Z_i > s]} d(\Lambda_i^D + \Lambda_i^C)$ . Thus,  $MA = (M_i^D + M_i^C)^2 - \int_0^t I_{[Z_i > s]} d(\Lambda_i^D + \Lambda_i^C)$  is a martingale. With a little algebra, we see that

$$\begin{aligned} MA &= (M_i^D)^2 + (M_i^C)^2 + 2M_i^D M_i^C - \int_0^t I_{[Z_i > s]} d\Lambda_i^D - \int_0^t I_{[Z_i > s]} d\Lambda_i^C \\ &= [(M_i^D)^2 - \int_0^t I_{[Z_i > s]} d\Lambda_i^D] + [(M_i^C)^2 - \int_0^t I_{[Z_i > s]} d\Lambda_i^C] + 2M_i^D M_i^C \end{aligned}$$

Hence, we have

$$2 M_i^{D,C} = MA - [(M_i^D)^2 - \int_0^t I_{[Z_i > s]} d \Lambda_i^D] - [(M_i^C)^2 - \int_0^t I_{[Z_i > s]} d \Lambda_i^C]$$

which shows, in view of (1.8), that  $M_i^{D,C}$  is a martingale, an equivalent statement that their cross variation is zero. ■

### Section 2. Asymptotic Normality

Since the estimates (0.2), (0.3) are truncated at  $M_n$ , it is natural to expect that the best we can do is to estimate  $\alpha, \beta$  up to  $M_n$ , and the correct centering quantity are thus

$$\beta^* = \sum b_{ni} \int_{-\infty}^{M_n} t dF_i(t) \tag{2.0}$$

$$\alpha^* = \sum a_{ni} \int_{-\infty}^{M_n} t dF_i(t)$$

notice if we replace  $M_n$  by  $\infty$ , we get exact  $\alpha$  and  $\beta$ .

Theorem 2.1. Under the conditions that will be specified at the end of this paper, we have

$$\sqrt{n} \begin{bmatrix} \hat{\alpha} - \alpha^* \\ \hat{\beta} - \beta^* \end{bmatrix} \xrightarrow{\mathcal{D}} N(\underline{0}, \Sigma).$$

where  $\Sigma = \sigma_{ij}$

$$\sigma_{11} = \sigma_{\alpha}^2 = (2.5) \text{ below but with } b_{ni} \ b_{nj} \text{ replaced by } a_{ni}, \ a_{nj}.$$

$$\sigma_{22} = \sigma_{\beta}^2 = (2.5) \text{ below.}$$

$$\text{and } \sigma_{12} = \sigma_{21} = \lim n \sum_i b_{ni} a_{ni} \int \left[ \frac{t}{1-G(t-)} - \frac{h_i(t)}{1-H_i(t)} \right]^2 [1-H_i(t)] \frac{dF_i}{1-F_i}$$

$$+ \lim n \sum_i \int \left[ \frac{\sum b_{nj} h_j(t)}{\sum (1-H_j)} - \frac{b_{ni} h_i(t)}{1-H_i(t)} \right] \left[ \frac{\sum a_{nj} h_j(t)}{\sum (1-H_j)} - \frac{a_{ni} h_i(t)}{1-H_i(t)} \right] [1-H_i(t)] \frac{dG}{1-G}.$$

Proof: We only illustrate the proof for  $\hat{\beta}$  alone, since the proof for joint normality of  $(\hat{\alpha}, \hat{\beta})$  are similar. Observe that

$$y_i^* = \frac{Z_i \delta_i}{1-\hat{G}(Z_i^-)} = \int \frac{t}{1-\hat{G}(t-)} d I_{[Z_i \leq t, \delta_i=1]}, \quad (2.1)$$

so the KSV estimator (0.2) of  $\beta$  can be written as (take  $\hat{G}$  to be the Kaplan-Meier estimator).

$$\hat{\beta} = \sum b_{ni} \frac{Z_i \delta_i}{1-\hat{G}(Z_i^-)} = \sum b_{ni} \int \frac{t}{1-\hat{G}(t-)} d I_{[Z_i \leq t, \delta_i=1]}$$

If we need to truncate at  $M_n$ , as KSV did, we can simply set the upper limit of the integral to  $M_n$ .

Since  $\beta^* = \sum b_{ni} \int_0^{M_n} t dF_i$ , we have

$$\hat{\beta} - \beta^* = \sum b_{ni} \left\{ \int \frac{t}{1-\hat{G}(t-)} d I_{[Z_i \leq t, \delta_i=1]} - \int t dF_i \right\}.$$

Now plus and minus  $\int \frac{t}{1-\hat{G}(t-)} I_{[Z_i > t]} \frac{dF_i}{1-F_i}$  in the {...} we get

$$\begin{aligned} \hat{\beta} - \beta^* &= \sum b_{ni} \left\{ \int \frac{t}{1-\hat{G}(t-)} d I_{[Z_i \leq t, \delta_i=1]} - \int \frac{t}{1-\hat{G}(t-)} I_{[Z_i > t]} \frac{dF_i}{1-F_i} \right. \\ &\quad \left. + \int \frac{t}{1-\hat{G}(t-)} I_{[Z_i > t]} \frac{dF_i}{1-F_i} - \int t dF_i \right\} \\ &= \sum b_{ni} \left\{ \int \frac{t}{1-\hat{G}(t-)} [d I_{[Z_i \leq t, \delta_i=1]} - I_{[Z_i > t]} \frac{dF_i}{1-F_i}] \right. \\ &\quad \left. + \int \left[ \frac{I_{[Z_i > t]}}{[1-\hat{G}(t-)] (1-F_i)} - 1 \right] t dF_i \right\} \\ &= \sum b_{ni} \int \frac{t}{1-\hat{G}(t-)} dM_i^D(t) \end{aligned}$$

$$+ \sum b_{ni} \int \left[ \frac{1-G(t)}{1-\hat{G}(t-)} \frac{1-\hat{H}_i}{1-H_i} - 1 \right] t dF_i \quad (2.2)$$

Now, because

$$\frac{1-\hat{H}_i}{1-H_i} = 1 + \frac{H_i - \hat{H}_i}{1-H_i}, \quad \text{and} \quad \frac{1-G(t)}{1-\hat{G}(t-)} = 1 + \frac{\hat{G}(t-)-G(t)}{1-\hat{G}(t-)},$$

it is easily seen that

$$\begin{aligned} \left[ \frac{1-G(t)}{1-\hat{G}(t-)} \frac{1-\hat{H}_i}{1-H_i} - 1 \right] &= \frac{H_i - \hat{H}_i}{1-H_i} + \frac{\hat{G}(t-)-G(t)}{1-\hat{G}(t-)} + (\text{a high order term}) \\ &= \frac{H_i - \hat{H}_i}{1-H_i} + \frac{\hat{G}(t-)-G(t)}{1-\hat{G}(t-)} + (\text{two high order terms}) \end{aligned} \quad (2.3)$$

now plug this into (2.2) and ignore the high order terms, we get

$$\begin{aligned} \hat{\beta} - \beta^* &\approx \sum b_{ni} \int \frac{t}{1-\hat{G}(t-)} dM_i^D(t) \\ &+ \sum b_{ni} \int \left[ \frac{H_i - \hat{H}_i}{1-H_i} + \frac{\hat{G}(t-)-G(t)}{1-\hat{G}(t-)} \right] t dF_i \end{aligned}$$

If we let  $h_i(t) = \int_t^\infty s dF_i(s)$ , we have  $d h_i(t) = -t dF_i(t)$ , and integration by parts in the second term above, we see that

$$\begin{aligned} \int \frac{H_i - \hat{H}_i}{1-H_i} t dF_i &= \int \frac{H_i - \hat{H}_i}{1-H_i} (-) d h_i = -h_i(t) \frac{H_i - \hat{H}_i}{1-H_i} \Big|_{-\infty}^{+\infty} + \int h_i d \frac{H_i - \hat{H}_i}{1-H_i} \\ &= \int h_i \frac{-1}{1-H_i} dM_i^+(+) \quad \text{by (1.10),} \end{aligned}$$

similarly

$$\int \frac{\hat{G}_- - G}{1-G_-} t dF_i = -h_i(t) \frac{\hat{G}_- - G}{1-G_-} \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} h_i(t) d \frac{\hat{G}_- - G}{1-G_-}$$

$$= \int_{-\infty}^{+\infty} h_i(t) \frac{1-\hat{G}_-}{1-G_-} \frac{1}{y^+(t)} d M_C^+(t)$$

So, we have

$$\begin{aligned} \hat{\beta}-\beta &\approx \sum b_{ni} \int \frac{t}{1-\hat{G}(t-)} d M_i^D(t) \\ &+ \sum b_{ni} \int h_i \frac{-1}{1-H_i} d M_i^+(t) + \sum b_{ni} \int h_i(t) \frac{1-\hat{G}_-}{1-G_-} \frac{1}{y^+(t)} d M_C^+(t) \\ &= \sum b_{ni} \int \frac{t}{1-\hat{G}(t-)} d M_i^D(t) - \sum b_{ni} \int \frac{h_i}{1-H_i} d M_i^+(t) \\ &+ \int (\sum b_{ni} h_i(t)) \frac{1-\hat{G}_-}{1-G_-} \frac{1}{y^+(t)} d M_C^+(t) \\ &= \sum_{i=1}^n b_{ni} \int \frac{t}{1-\hat{G}(t-)} d M_i^D(t) - \sum_{i=1}^n b_{ni} \int \frac{h_i}{1-H_i} d M_i^+(t) \\ &+ \sum_{i=1}^n \int (\sum_j b_{nj} h_j(t)) \frac{1-\hat{G}_-}{1-G_-} \frac{1}{y^+(t)} d M_i^C(t) \\ &= \sum_{i=1}^n b_{ni} \int \left[ \frac{t}{1-\hat{G}(t-)} + \frac{h_i(t)}{1-H_i(t)} \right] d M_i^D(t) \\ &+ \sum_{i=1}^n \int \left[ \left( \sum_j b_{nj} h_j(t) \right) \frac{1-\hat{G}_-}{1-G_-} \frac{1}{y^+(t)} - \frac{b_{ni} h_i(t)}{1-H_i(t)} \right] d M_i^C(t) \quad (2.4) \end{aligned}$$

The quadratic variation process of (1.4) ( $\approx \langle \hat{\beta}-\beta \rangle$ ) can easily be computed

$$\begin{aligned} \langle 1.4 \rangle \quad \langle \hat{\beta}-\beta \rangle &\approx \sum_i b_{ni}^2 \int \left[ \frac{t}{1-\hat{G}(t-)} - \frac{h_i(t)}{1-H_i(t)} \right]^2 I_{[T_i > t]} \frac{d F_i(t)}{1-F_i(t)} \\ &+ \sum_i \int \left[ \left( \sum_j b_{nj} h_j(t) \right) \frac{1-\hat{G}_-}{1-G_-} \frac{1}{y^+(t)} - \frac{b_{ni} h_i(t)}{1-H_i(t)} \right]^2 I_{[T_i > t]} \frac{d G}{1-G} \end{aligned}$$

The limit of the above quadratic variation process is the asymptotic variance.

Of course we need to check the 'Lingdeberg' condition here to assure the

asymptotic normality, see e.g. Andersen & Borgan (1985), it can be done similar to Zhou (1988). Also we need to show the high order terms are negligible under certain mild tail conditions. For details see section 4. And so we have a central limit theorem, and the asymptotic variance are

$$\begin{aligned} \lim n \langle \hat{\beta} - \beta \rangle &= \lim n \sum_i b_{ni}^2 \int \left( \frac{t}{1-G(t-)} - \frac{h_i(t)}{1-H_i(t)} \right)^2 [1-H_i(t)] \frac{dF_i}{1-F_i} \\ &+ \lim n \sum_i \int \left[ \left( \sum_j b_{nj} h_j(t) \right) \frac{1}{\sum_j (1-H_j)} - \frac{b_{ni} h_i(t)}{1-H_i(t)} \right]^2 [1-H_i(t)] \frac{dG}{1-G} \end{aligned} \quad (2.5)$$

There is no difficulty in extending Theorem 2.1 to multiple regression case. Since the estimator is  $\hat{\beta} = (X^T X)^{-1} X^T y^*$ , and its  $i$ th component  $\hat{\beta}_i = \sum_j u_{ij} y_j^*$  is again a linear combination of  $y_j^*$ 's defined in (2.1). So the same proof also works.

### Section 3. Asymptotic Variance

To write the asymptotic variance in another form, we develop the big square in the second term of (2.5),  $(a-b)^2 = a^2 + b^2 - 2ab$ , we find that (1)

and (3) combines to give

$$\begin{aligned} \lim n \langle \hat{\beta} - \beta \rangle &= \lim n \sum_i b_{ni}^2 \int \left( \frac{t}{1-G(t-)} - \frac{h_i(t)}{1-H_i(t)} \right)^2 [1-H_i(t)] \frac{dF_i}{1-F_i} \\ &+ \lim n \int \sum \frac{b_{ni}^2 h_i^2(t)}{1-H_i(t)} \frac{dG}{1-G} - \lim n \int \frac{[\sum_j b_{nj} h_j(t)]^2}{\sum_j (1-H_j)} \frac{dG}{1-G} \\ &= \lim n \sum_{i=1}^n b_{ni}^2 \int \left( \frac{t}{1-G(t-)} - \frac{h_i(t)}{1-H_i(t)} \right)^2 [1-H_i] \frac{dF_i}{1-F_i} \end{aligned}$$

$$\begin{aligned}
 & + \lim n \sum_{i=1}^n b_{ni}^2 \int \frac{h_i^2(t)}{[1-H_i(t)]^2} (1-F_i) dG \\
 & - \lim n \int \frac{[\sum_j b_{nj} h_j(t)]^2}{\sum_j (1-H_j)} \frac{dG}{1-G}. \tag{3.1}
 \end{aligned}$$

The first two positive terms of (3.1) are just the variance of

$\sqrt{n} \sum b_{ni} \frac{\delta_i T_i}{1-G(T_i)}$  as the next Lemma shows. Notice the third term is negative! and is a little different from Koul, Susarla & Van Ryzin's formulas, as they had an excess  $\frac{1}{n}$  term there.

Lemma 3.1:

$$\begin{aligned}
 \text{Var} \frac{\delta_i Z_i}{1-G(Z_i)} &= \int \left( \frac{t}{1-G(t-)} - \frac{H_i(t)}{1-H_i} \right) (1-F_i) \frac{dF_i}{1-F_i} \\
 &+ \int \frac{h_i^2(t)}{[1-H_i(t)]^2} (1-F_i) dG.
 \end{aligned}$$

Proof: We first observe that the mean of the quantity is

$$E \frac{\delta_i Z_i}{1-G(Z_i)} = \int t dF_i(t) = \alpha + \beta x_i$$

and so

$$\text{Var} \left[ \frac{\delta_i Z_i}{1-G(Z_i)} \right] = E \left[ \frac{\delta_i Z_i}{1-G(Z_i)} - \int t dF_i \right]^2$$

but  $\frac{\delta_i Z_i}{1-G(Z_i)} = \int \frac{t}{1-G(t)} dI_{[Z_i \leq t, \delta_i=1]}$ , see (2.1). Thus

$$\begin{aligned}
 \left[ \frac{\delta_i Z_i}{1-G(Z_i)} - \int t dF_i \right] &= \int \frac{1}{1-G(t)} dI_{[Z_i \leq t, \delta_i=1]} - \int t dF_i \\
 &= \int \frac{1}{1-G(t)} \left[ dI_{[Z_i \leq t, \delta_i=1]} - I_{[T_i > t]} \frac{dF_i}{1-F_i} \right] \tag{3.2}
 \end{aligned}$$

$$\begin{aligned}
 & + \int \frac{t}{1-G(t-)} I_{[Z_i > t]} \frac{d F_i}{1-F_i} - t dF_i \\
 = & \int \frac{t}{1-G(t-)} d M_i^D + \int \left[ \frac{I_{[Z_i > t]}}{(1-G)(1-F_i)} - 1 \right] t dF_i \\
 = & \int \frac{t}{1-G} d M_i^D + \int \left[ \frac{1-\hat{H}_i}{1-H_i} - 1 \right] t dF_i \\
 = & \int \frac{t}{1-G} d M_i^D + \int \frac{H_i - \hat{H}_i}{1-H_i} t dF_i \tag{3.3}
 \end{aligned}$$

if we let  $h_i(t) = \int_t^\infty s d F_i(s)$  then  $d h_i = -t dF_i(t)$  and (3.3) become

$$= \int \frac{t}{1-G(t-)} d M_i^D + \int \frac{H_i - \hat{H}_i}{1-H_i} (-) d h_i(t). \tag{3.4}$$

Integration by parts on the second term of (3.4) we have

$$- \int \frac{H_i - \hat{H}_i}{1-H_i} d h_i(t) = \left[ -h_i(t) \frac{H_i - \hat{H}_i}{1-H_i} \Big|_{-\infty}^{+\infty} + \int h_i(t) d \frac{H_i - \hat{H}_i}{1-H_i} \right] = \int h_i(t) d \frac{H_i - \hat{H}_i}{1-H_i},$$

Plug this into (3.4), we have

$$\begin{aligned}
 (3.4) \quad & = \int \frac{t}{1-G(t-)} d M_i^D + \int h_i(t) d \frac{H_i - \hat{H}_i}{1-H_i} \\
 & = \int \frac{t}{1-G(t-)} d M_i^D + \int h_i(t) \frac{-1}{1-H_i(t)} d M_i^+(t) \\
 & = \int \frac{t}{1-G} d M_i^D - \int \frac{h_i}{1-H_i} d M_i^D - \int \frac{h_i}{1-H_i} d M_i^C(t) \\
 & = \int \left( \frac{t}{1-G(t-)} - \frac{h_i}{1-H_i} \right) d M_i^D(t) - \int \frac{h_i}{1-H_i} d M_i^C(t), \tag{3.5}
 \end{aligned}$$

in view of (1.10).

The quadratic variation process of the above martingale is now easy to

compute and the expected value of the quadratic variation process gives the desired variance:

$$\begin{aligned} \text{Var} = E \langle 3.5 \rangle &= \int \left( \frac{t}{1-G} - \frac{h_i}{1-H_i} \right)^2 (1-H_i) \frac{d F_i}{1-F_i} \\ &+ \int \frac{h_i^2}{[1-H_i]^2} (1-H_i) \frac{d G}{1-G} \\ &= \int \left( \frac{t}{1-G} - \frac{h_i}{1-H_i} \right)^2 (1-G) d F_i + \int \frac{h_i^2}{[1-H_i]^2} (1-F_i) d G. \end{aligned}$$

■

The negative term in (3.1) is, in general, nonzero. To see this, notice  $[\sum_j b_{nj} h_j(-\infty)]^2 = \beta^2$  and it is a continuous nonnegative function of  $t$  while the other part of the integral

$$\int \frac{n}{\sum_j (1-H_j)} \frac{d G(t)}{1-G(t)} \geq \int \frac{dG}{1-G} = -\log(1-G(\infty))$$

Unless  $G(t)$  is completely flat at places where  $[\sum b_{nj} h_j(t)]^2$  is positive, the term is going to be nonzero. Thus the variance estimator proposed by Koul, Susarlar & Van Ryzin (1981) (4.8) need to include an extra negative term. We suggest to use

$$-n \int \frac{[\sum b_{nj} \hat{h}_j(t)]^2 dN_c^+(t)}{[y^+(t)-1] y^+(t)}$$

where

$$N_c^+(t) = \sum_{i=1}^n I_{[Z_i < t, \delta_i = 0]}$$

$$\hat{h}_j(t) = \int_t^\infty \text{sd} \left( \frac{I_{[Z_j > s]}}{1-\hat{G}(s)} \right). \quad \left( \frac{0}{0} \doteq 0 \right)$$

to estimate the negative part of the asymptotic variance (3.1).

REMARK: The first two positive terms in the asymptotic variance (3.1) can be thought as the asymptotic variance of an estimator with known censoring distribution (Lemma 3.1), (0.2) with a true  $G(t)$  instead of  $\hat{G}(t)$  therein. Estimation of  $G(t)$  actually make the estimator of  $\beta$  better. Using this fact to modify the estimator (0.2), (0.3) so that to result a larger negative part is possible and interesting. For some discussions see Fyngson & Zhou (1988).

#### Section 4. High Order Term

We show in this section that the high order terms in the proof of Theorem 2.1 are negligible.

Recall from section 2, (2.3) the high order term is

$$\sum b_{ni} \int_{-\infty}^{M_n} \left[ \frac{H_i - \hat{H}_i}{1 - H_i} \cdot \frac{\hat{G}(t-) - G(t)}{1 - \hat{G}(t-)} + \frac{\hat{G}(t-) - G(t)}{1 - \hat{G}(t-)} \frac{\hat{G}(t-) - G(t-)}{1 - G(t-)} \right] t dF_i(t) \quad (4.1)$$

$$= \int_{-\infty}^{M_n} \left[ \sum b_{ni} \frac{H_i - \hat{H}_i}{1 - H_i} f_i(t) \right] \frac{\hat{G}(t-) - G(t)}{1 - \hat{G}(t-)} t dt +$$

$$\int_{-\infty}^H \frac{\hat{G}(t-) - G(t)}{1 - \hat{G}(t-)} \frac{\hat{G}(t-) - G(t-)}{1 - G(t-)} t d(\sum b_{ni} F_i(t)), \quad (4.2)$$

we need to show  $\sqrt{n} \times (4.2) = o_p(1)$ .

The following Lemmas are useful.

Lemma 4.1 If  $x_1, x_2, \dots, x_n$  are independent random variables with  $P(X_i < t) = F_i(t)$  and  $f_{ni}$  are constants such that  $\sum f_{ni}^2 \rightarrow 0$  as  $n \rightarrow \infty$ ;  $\epsilon > 0$ ; then for those  $n$  such that  $\sum f_{ni}^2 \leq \frac{\epsilon^2}{2}$ , we have

$$P(\sup_t |\sum f_{ni} [I_{[X_i < t]} - F_i(t)]| > \epsilon) \leq 8(n+1) \exp\left(-\frac{\epsilon^2}{32 \sum f_{ni}^2}\right).$$

Furthermore, if  $f_{ni}$  are functions of  $t$ ,  $f_{ni}(t)$ ; we have the following: If  $f_{ni}(t)$  are of bounded variation with  $V_{-\infty}^{\infty} f_{ni}(t) \leq K < \infty$ , and  $\sup_t \sum f_{ni}^2(t) \rightarrow 0$  as  $n \rightarrow \infty$  then for those  $n$  such that  $\sup_t \sum f_{ni}^2(t) \leq \frac{\epsilon^2}{2}$ , we have

$$P(\sup_t |\sum f_{ni}(t)[I_{[X_i < t]} - F_i(t)]| > \epsilon) \leq 8 C_{\epsilon}(n) \exp\left(-\frac{\epsilon^2}{128 \sup_t \sum f_{ni}^2(t)}\right)$$

with  $C_{\epsilon}(n) = \frac{16K}{\epsilon} n^2 + n + 1$ .

Proof: See Zhou (1988).

Lemma 4.2

$$\sup_{t \leq M_n} \frac{1-G(t)}{1-\hat{G}(t)} = O_p(1) \quad \sup_{t \leq M_n} \frac{\hat{G}(t)-G(t)}{1-\hat{G}(t)} = O_p(1)$$

provided any one of the following condition is satisfied:

- (a)  $1-G(\infty) > 0$
- (b) The design is random, i.e.  $X_i$  are iid  $D(t)$  independent of  $\epsilon_i$ ,  $\text{Var}(X_i) < \infty$ .
- (c)  $M_n = Z_{(k_n)} \uparrow \tau$ , such that

$$\int_{-\infty}^{M_n} \frac{d \Lambda^c(t)}{\frac{1}{n} \sum (1-H_i(t))} = O_p(n).$$

Proof: See Zhou (1988).

Lemma 4.3 If  $h(t)$  is a real function, such that  $h(t) = h_1(t)-h_2(t)$  where  $h_i(t)$  are nonnegative nonincreasing, functions such that

$$\int_{-\infty}^{+\infty} h_i^2(t) \frac{dG}{(1-G)^2(1-F)} < \infty .$$

Then the processes

$$h(t \wedge T_n) Z_n(t \wedge T_n); \quad \int^{t \wedge T_n} h(s) d Z_n(s); \quad \int^{t \wedge T_n} Z_n(s) d h(s)$$

converge jointly in  $D[0, \tau_G]$  in distribution to

$$h(t) B(C(t)), \quad \int^t h(s) d B(C(s)); \quad \int^t B(C(s)) d h(s)$$

and  $h(t) B(C(t)) \rightarrow 0$  a.s. as  $t \rightarrow \tau_G$ . Where

$$Z_n(t) = \sqrt{n} \frac{\hat{G}(t)}{1-G(t)}$$

$B(\cdot)$  is a standard Brownian motion, and

$$C(t) = \int_{-\infty}^t \frac{d G}{(1-G)^2(1-F)}$$

Proof: This is Gill (1983) Theorem 2.1 with a generalization on  $h$  function. For details see Zhou (1986). ■

It is easy to see that for a fixed finite  $\tau > 0$ , the two integrals in (4.2) but restricted to  $\tau$ , i.e.

$$[\sqrt{n} \int_{-\tau}^{\tau} (*) d(*), \text{ are } o_p(1) \text{ under the assumptions made.}$$

To show that the tail part of the integrals are  $o_p(1)$  is more involved. Let us look at second term of (4.2) first.

If we assume

$$\lim_{\tau \rightarrow \infty} \limsup_n \int_{\tau}^{\infty} \frac{[t \sum b_{ni} f_i(t)]^2}{\frac{1}{n} \sum (1-F_i)} \frac{d G}{(1-G)^2} = 0 \quad (4.3)$$

then by Lemma 4.3,  $\forall \epsilon > 0$ , we can choose  $\tau$  large to make

$$P\left[\sqrt{n} \int_{\tau}^{\infty} \left| \frac{\hat{G}(t-) - G(t-)}{1-G(t-)} t \sum b_{ni} f_i(t) \right| dt > \epsilon\right]$$

arbitrary small, in view of Lemma 4.2, this proves the upper tail is  $o_p(1)$ . The lower tail is easier, we need only to assume

$$\int_{-\infty}^{\tau} |t \sum b_{ni} f_i(t)| dt < \infty. \quad (4.4)$$

As to the first term of (4.2), similar to Zhou (1988), but here in addition to

$$\frac{1}{\sum b_{ni}^2} \sum b_{ni}^2 H_i(s)[1-H_i(s)] \rightarrow K(s, t); s < t \quad (4.5)$$

$$\int \frac{1}{[1-G]^2} d\left[\frac{1}{\sum b_{ni}^2} \sum b_{ni}^2 H_i(s)\right] < \infty \quad (4.6)$$

we also assume

$$\left| \frac{f_i(t) t}{1-F_i(t)} \right| \leq M \quad \text{for each } i \text{ and all } t, \quad (4.7)$$

then, (4.5), (4.6) imply that  $\frac{1}{\sqrt{\sum b_i^2}} \sum b_i \frac{H_i(t) - \hat{H}_i(t)}{1-G(t)}$  converge weakly to a

stretch out Browning bridge. Now (4.7), Lemma 4.2 guarantees that

$$\left| \frac{f_i(t)}{1-F_i(t)} \frac{\hat{G}(t) - G(t)}{1-\hat{G}(t-)} t \right| = o_p(1) \quad (4.8)$$

thus the whole term is  $o_p(1)$ .

Conditions:

(a1)  $0 < m \leq n \times \sum b_{ni}^2 \leq M < \infty$ , for large  $n$ .

(a2) the variance-covariance  $\sigma_{ij}$  are well defined, non infinite and non zero.

(a3)  $\max_i n b_{ni}^2 \rightarrow 0$ ,  $\max_i n a_{ni}^2 \rightarrow 0$ , as  $n \rightarrow \infty$ .

(a4)  $\sup_{i, t} E[\epsilon_i - t | \epsilon_i > t] < \infty$ .

(b1)  $M_n$  are stopping times such that  $\lim M_n = \lim T_n$  and one of the conditions of Lemma 4.2 hold.

(b2) (4.3) (4.4) (4.5) (4.6) (4.7) hold.

## REFERENCES

- AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701-726.
- ANDERSEN, P. K. and BORGAN, O. (1985). Counting process models for life history data: A review (with discussion). *Scand. J. Statist.* **12**, 97-158.
- ANDERSEN, P. K., BORGAN, O., GILL, R. and KEIDING, N. (1982). Linear non-parametric tests for comparison of counting processes, with applications to censored survival data. *Int. Stat. Rev.* **50**, 219-258.
- ANDERSEN, P. K. and GILL, R. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.
- BILLINGSLEY, P. (1968). *Weak Convergence of Probability Measures*. Wiley, New York.
- BUCKLEY, J. and JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.
- CHATTERJEE, S. and McLEISH, D. L. (1986). Fitting linear regression models to censored data by least squares and maximum likelihood methods. *Commun. Statist. - Theory Meth.* **15(11)**, 3227-3243.
- COX, D. R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc. B* **34**, 187-220.
- FYGENSON, M. and ZHOU, M. (1988). Analysis of linear regression models with censoring. *Preprint*.
- GILL, R. (1980). *Censoring and Stochastic Integrals*. Mathematical Centre Tracts 124. Mathematisch Centrum, Amsterdam.
- GILL, R. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11**, 49-58.
- HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **16**, 293-325.
- JACOBSON, M. (1982). *Statistical Analysis of Counting Processes*. Lecture Notes in Statistics. Springer, New York.
- KAPLAN, E. and MEIER, P. (1958). Non-parametric estimator from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

- KOUL, H., SUSARLA, V. and VAN RYZIN, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9**, 1276-1288.
- LEURGANS, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74**, 301-309.
- MILLER, R. G. and HALPERN, J. (1982). Regression with censored data. *Biometrika* **69**, 521-531.
- POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.
- SHORACK, G. R. and WELLNER, J. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- SUSARLA, V., TSAI W. Y. and VAN RYZIN, J. (1984). A Buckley-James-type estimator for the mean with censored data. *Biometrika* **71**, 624-625.
- SUSARLA, V. and VAN RYZIN, J. (1980). Large sample theory for an estimator of the mean survival time from censored samples. *Ann. Statist.* **8**, 1002-1016.
- WANG, J. G. (1987). A note on the uniform consistency of the Kaplan-Meier estimator. *Ann. Statist.* **15**, 1313-1316.
- ZHENG, Z. (1984). Regression with randomly censored data. *Ph. D. thesis*. Columbia University.
- ZHOU, M. (1986). Some nonparametric two sample tests with censored data. *Ph. D. thesis*. Columbia University.
- ZHOU, M. and VAN RYZIN, J. (1987). Difference of means test with censored data. *Submitted to J. Amer. Statist. Assoc.*
- ZHOU, M. (1988). Asymptotic normality of 'synthetic data' regression estimator for censored survival data. *Preprint*.