1. Introduction.

Let $x_1^0, \cdots, x_n^0$ be a sequence of positive random variables, usually representing survival times of n patients in a clinical trial. Let $z_1, \cdots, z_n$ be their corresponding covariates. The Cox (1972, 1975) regression model has been used widely to make inference on the relationship between the distribution of x^0's and the z's. More specifically, for a given z, assume that the hazard function of x^0 satisfies the proportional hazard model

$$\lambda(t|z) = \lambda_0(t)\exp(\beta_0 z), \tag{1.1}$$

where λ_0 is called the baseline hazard function. In practice, x^0's are not directly observable. We observe (x_i, δ_i), $i = 1, \cdots, n$, where $x_i = \min(x_i^0, c_i)$, $\delta_i = 1_{[x_i = x_i^0]}$ and c_i is the censoring variable. Assume that the censoring is non-informative, Cox proposed to consider the partial likelihood

$$\tilde{L}_n(\beta) = \prod_{i=1}^{n}\left[\frac{\exp(\beta z_i)}{\sum_{j=1}^{n}\exp(\beta z_j)1_{[x_j \geq x_i]}}\right]^{\delta_i}$$

for inference on β_0. If $\hat{\beta}$ is a number such that $\tilde{L}_n(\hat{\beta}) = \sup_\beta \tilde{L}_n(\beta)$, then we shall expect $\hat{\beta}$ to behave like a maximum likelihood estimate.

A large amount of literature has been devoted to show the consistency and the asymptotic normality of $\hat{\beta}$ and other related estimates, notably, Tsiatis (1981), Andersen and Gill (1982), Naes (1982), Bailey (1983) among others. For a nice introduction, see also Gill (1984) or Andersen and Borgan (1985).

By considering a slightly modified version of $\tilde{L}_n$, or considering

$$L_n(\beta) = \prod_{i=1,\cdots,n;\, x_i \leq T}\left[\frac{\exp(\beta z_i)}{\sum_{j=1}^{n}\exp(\beta z_j)1_{[x_j \geq x_i]}}\right]^{\delta_i}, \tag{1.2}$$

where T satisfies $\Pr\{x_i \geq T\} > 0$ and assuming that $y_i = (z_i, x_i, \delta_i)$, $i = 1, \cdots, n$ be i.i.d. and some other regularity conditions, Tsiatis (1981) has shown that there exits a sequence $\hat{\beta}$ such that $\hat{\beta}$ converges to β_0 almost surely and that $\sqrt{n}(\hat{\beta} - \beta_0)$ converges in distribution to a normal variable with mean 0 and finite variance. The variance can be estimated by $(I_n(\hat{\beta}))^{-1}$, where I_n is minus the second derivative of $\log L_n$. He also showed that the nonparametric MLE of the baseline cumulative hazard function based on $\beta = \hat{\beta}$ is uniformly consistent and converges weakly.

It was vaguely known that the speed of the convergence of the distribution $\sqrt{n}(\hat{\beta} - \beta_0)$ to normality is $O(n^{-1/2})$. If this is true, does a one-term Edgeworth expansion hold for this convergence? If Edgeworth expansion holds, what is the constant involved here? How should we estimate it? We provide positive answers to these questions

in this article in the case when β is one dimensional and assume essentially Tsiatis's i.i.d. assumption. We will only deal with model (1.2).

Recent work by Hall (1988) and others have shown that the Edgeworth expansion results can be used to show the bootstrap procedure is actually second order correct for estimating the distribution of mean based on i.i.d. observations. Babu and Singh (1984) have indicated that the studentization enables us to obtain more accurate bootstrap estimates of the distribution of statistical functions. Helmers (1991) has shown similar results for a Studentized U-statistics. We also investigate the Edgeworth expansion for the Studentized version of $\hat{\beta} - \beta$, or $(I_n(\hat{\beta}))^{1/2}(\hat{\beta} - \beta)$ and show correct ways for approximating the sampling distribution for both original and Studentized versions of the statistics.

We also pay attention to the statistic $U_n(\beta_0)$, the first derivative of $\log L_n$ at value β_0 (The score statistics). This statistic plays a central role in the testing problem $H_0 : \beta = \beta_0$ vs $H_1 : \beta \neq \beta_0$. If $\beta_0 = 0$, then $U_n(\beta_0)$ reduces to the popular log-rank statistics. The Studentized version $(I_n(\beta_0))^{-1/2}U_n(\beta_0)$ and $(I_n(\hat{\beta}))^{-1/2}U_n(\beta_0)$ will also be considered. As a consequence of our result in Section 2, the normal approximation for the log-rank statistic in a 50%-50% randomized design clinical trial is automatically second order correct (Example 2.1). Even though this may not be true if the design is not 50%-50%, we can achieve the second order correctness by analytic correction or bootstrap.

The method employed in this paper is first to approximate the relevant statistics by U-statistics and then use the known results of the Edgeworth expansion for U-statistics (Bickel, Götze and Van Zwet, 1986). In the proof, we shall use some results concerning the moments which have been carried out in a technical report by Gu (1992). This report is available from the author.

Section 2 contains Edgeworth expansion results for the relevant statistics. In section 3, we first propose a simple analytic procedure which achieves the second order correction and then propose a bootstrap procedure which achieves the second order correction automatically. Section 4 contains the proof of the Edgeworth expansion results in Section 2 and Section 5 contains the proof of results in Section 3.

Before closing this section, we shall say something about (1.2). Even though Tsiatis (1981) has argued the practicality of likelihood L_n, we emphasize that it is not only easy to handle mathematically, it is robust in some sense. From the data analysis point of view, large values of x_i are not as reliable as moderate values since they might be produced by the effects not under the control of the study. Model (1.2) amounts to taking any observation larger than T as censored at T. The idea here is very similar to the idea of truncated mean. We shall expect to lose some efficiency if the data are not contaminated, but gain on the reliability of the procedure and get much better precision on the confidence level we usually report.

2. Edgeworth Expansion for Statistics Related to Cox Regression Model.

As in the last section and throughout this article, we shall denote $U_n(\beta)$ as the first derivative of $\log L_n(\beta)$, $I_n(\beta)$ as minus the second derivative of $\log L_n(\beta)$ and $\hat{\beta}$ as a number which maximizes $L_n(\beta)$. In applications concerning Cox regression model, it is important to know the asymptotic behavior of the statistics $U_n(\beta_0)$, $U_n(\beta_0)/\sqrt{I_n(\beta_0)}$, $U_n(\beta_0)/\sqrt{I_n(\hat{\beta})}$, $\hat{\beta}-\beta_0$ and $\sqrt{I_n(\hat{\beta})}(\hat{\beta}-\beta_0)$. Under some regularity conditions, the asymptotic distributions of these statistics are known to be normal after proper normalizations (seeTsiatis (1981), Andersen and Gill (1982), Naes (1982), Bailey (1983)). In this section, we provide one-term Edgeworth Expansion for these statistics.

Assumption 2.1 *z is bounded uniformly.*

Assumption 2.2 *The range of consideration for β is bounded.*

Assumption 2.3 *$y_i = (z_i, x_i^0, c_i)$ are i.i.d., given z_i, x_i^0 and c_i are independent and (1.1) holds with continuous λ_0.*

Assumption 2.4

$$\sigma^2 = \int_0^T \left(\alpha_2(t) - \frac{\alpha_1^2(t)}{\alpha_0(t)}\right)\lambda_0(t)dt > 0, \tag{2.1}$$

where T is the same as in (1.2) and $\alpha_k(t)$ for $k = 0, 1, 2$ are given by

$$\alpha_k(t) = E\left[z^k e^{\beta_0 z} 1[x \geq t]\right]. \tag{2.2}$$

σ^2 defined above is the asymptotic variance of $U_n(\beta_0)/\sqrt{n}$. Assumption 2.4 ensures that the statistics are non-degenerate.

Theorem 2.1 *(Edgeworth expansion for $U_n(\beta_0)$ and related statistics). Under Assumptions 2.1 - 2.4,*

$$\begin{aligned}
&\sup_x |\Pr(U_n(\beta_0) \leq x n^{1/2}\sigma) - \tilde{F}_n(x)| = o(n^{-1/2}),\\
&\sup_x \left|\Pr\left(U_n(\beta_0)/\sqrt{I_n(\beta_0)} \leq x\right) - E_n(x)\right| = o(n^{-1/2}),\\
&\sup_x \left|\Pr\left(U_n(\beta_0)/\sqrt{I_n(\hat{\beta})} \leq x\right) - \tilde{E}_n(x)\right| = o(n^{-1/2}),\\
&\sup_x \left|\Pr\left(n^{1/2}\sigma(\hat{\beta}-\beta_0) \leq x\right) - G_n(x)\right| = o(n^{-1/2}),\\
&\sup_x \left|\Pr\left(\sqrt{I_n(\hat{\beta})}(\hat{\beta}-\beta_0) \leq x\right) - E_n(x)\right| = o(n^{-1/2}),
\end{aligned} \tag{2.3}$$

where

$$\begin{aligned}
\tilde{F}_n(x) &= \Phi(x) - \frac{\kappa_3}{6} n^{-1/2}\phi(x)(x^2-1), \\
E_n(x) &= \Phi(x) + \left(\frac{\Delta}{6\sigma^3}(x^2-1) + \frac{\zeta}{2\sigma^3}x\right) n^{-1/2}\phi(x), \\
\tilde{E}_n(x) &= \Phi(x) - \left(\frac{\Delta}{3\sigma^3}(x^2-1) + \frac{\Delta-\zeta}{2\sigma^3}x\right) n^{-1/2}\phi(x), \qquad (2.4)\\
G_n(x) &= \Phi(x) - \left(\frac{\kappa_3}{6}(x^2-1) + \frac{\Delta-2\zeta}{2\sigma^3}x\right) n^{-1/2}\phi(x),
\end{aligned}$$

where the quantities σ^2 is given by (2.1), κ_3, ζ, and Δ are given by

$$\begin{aligned}
\kappa_3 &= -\frac{\Delta}{\sigma^3} + 3\frac{\zeta}{\sigma^3}, \\
\zeta &= \Delta_A - \Delta_B, \qquad (2.5)\\
\Delta &= -\int_0^T \left(\alpha_3(t) - 3\frac{\alpha_2(t)\alpha_1(t)}{\alpha_0(t)} + 2\frac{\alpha_1^3(t)}{\alpha_0^2(t)}\right)\lambda_0(t)dt,
\end{aligned}$$

and

$$\begin{aligned}
\Delta_A &= -\int_0^T \left(\tilde{\alpha}_3(t) - 2\frac{\tilde{\alpha}_2(t)\alpha_1(t)}{\alpha_0(t)} + \frac{\tilde{\alpha}_1(t)\alpha_1^2(t)}{\alpha_0^2(t)}\right)\Lambda_0(t)\lambda_0(t)dt. \qquad (2.6)\\
\Delta_B &= -\int_0^T \left(\tilde{\alpha}_2(t) - 2\frac{\tilde{\alpha}_1(t)\alpha_1(t)}{\alpha_0(t)} + \frac{\tilde{\alpha}_0(t)\alpha_1^2(t)}{\alpha_0^2(t)}\right)\Lambda_1(t)\lambda_0(t)dt.
\end{aligned}$$

where for $k = 1, 2, 3$, $\alpha_k(t)$ is given by (2.2) and $\tilde{\alpha}_k(t)$ and $\Lambda_k(t)$ are defined by

$$\begin{aligned}
\tilde{\alpha}_k(t) &= E\left[z^k e^{2\beta_0 z} 1[x \geq t]\right], \\
\Lambda_k(t) &= \int_0^t \frac{\alpha_k(t)}{\alpha_0(t)}\lambda_0(t)dt.
\end{aligned}$$

Remark 2.1. Assumptions 2.1-2.4 are not in the most general form. For example, with more refined argument, only some moment assumptions on z would be enough. On the other hand, Assumptions 2.1-2.4 are usually satisfied in applications.

Remark 2.2. The explicit expressions of the functions on the right of (2.4) can be written in terms of the constant σ^2, Δ, Δ_A and Δ_B only. The constants κ_3, and ζ are intermediaries.

REMARK 2.3. Theorem 2.1 proves proves two things: (i) the normal approximation to the distributions of the statistics involved is $O(n^{-1/2})$ (Analogue to Berry-Essen theorem); (ii) the speed of convergence improved to $o(n^{-1/2})$ if we know the quantities σ^2, Δ, Δ_A and Δ_B.

Example 2.1. In a randomized clinical trial, with z_i taking on 0 (control) and 1 (treatment) and $\beta_0 = 0$ (no treatment effect), x_i is independent of z_i. We have $\tilde{\alpha}_k(t) = \alpha_k(t)$. Assuming that $E[z_i] = p$, then $\alpha_k(t)/\alpha_0(t) = p$, for $k = 1, 2, 3$. It is easy to see that

$$\begin{aligned}
\Delta &= -(p - 3p^2 + 2p^3) \int_0^T \alpha_0(t)\lambda_0(t)dt \\
\Delta_A &= -(p - 2p^2 + p^3) \int_0^T \alpha_0(t)\Lambda_0(t)\lambda_0(t)dt \\
\Delta_B &= -(p - p^2)p \int_0^T \alpha_0(t)\Lambda_0(t)\lambda_0(t)dt
\end{aligned} \tag{2.7}$$

If $p = 1/2$, then $\kappa_3 = \Delta = \zeta = 0$. In this case, the functions on the right of (2.4) reduce to the standard normal distribution and the normal approximations is automatically $o(n^{-1/2})$. This may be part of the reason that a 50%-50% randomized design is more popular than a non-symmetric design in some clinical trials.

The proof of Theorem 2.1 is deferred to Section 4.

3. Analytic and Bootstrap Approximations.

When σ^2, Δ, Δ_A and Δ_B are unknown, Theorem 2.1 cannot be directly applied. We can consistently estimate these constants. Let us define for $k = 0, 1, 2, 3$

$$\begin{aligned}
\hat{\alpha}_k(t, \beta) &= \sum_{i=0}^{n} \xi_{i,k}(t, \beta)/n, \qquad \hat{\tilde{\alpha}}_k(t, \beta) = \frac{1}{n}\sum_{i=1}^{n} \tilde{\xi}_{i,k}(t, \beta)/n, \\
\hat{\Lambda}_k(t) &= \sum_{i=1}^{n} \int_0^t \frac{\hat{\alpha}_k(t, \hat{\beta})dN_i(t)}{n\hat{\alpha}_0^2(t, \hat{\beta})},
\end{aligned}$$

where

$$\begin{aligned}
\xi_{i,k}(t, \beta) &= z_i^k e^{\beta z_i} 1_{[x_i \geq t]}, \qquad \tilde{\xi}_{i,k}(t, \beta) = z_i^k e^{2\beta z_i} 1_{[x_i \geq t]}. \\
N_i(t) &= 1[x_i \leq t, \delta_i = 1], \quad i = 1, \cdots, n.
\end{aligned} \tag{3.1}$$

In the definition of σ^2, Δ, Δ_A and Δ_B in (2.2), (2.5) and (2.6), replace $\alpha_k(t)$ with $\hat{\alpha}_k(t, \hat{\beta})$, $\tilde{\alpha}_k(t)$ with $\hat{\tilde{\alpha}}_k(t, \hat{\beta})$ and $\Lambda_k(t)$ with $\hat{\Lambda}_k(t)$ and then denote these constants as

$\hat{\sigma}^2$, $\hat{\Delta}$, $\hat{\Delta}_A$ and $\hat{\Delta}_B$. Also define $\hat{\zeta} = \hat{\Delta}_A - \hat{\Delta}_B$, and $\hat{\kappa}_3 = \left(-\hat{\Delta} + 3\hat{\zeta}\right)/\sigma^3$. Then we have parallel to Theorem 2.1:

Theorem 3.1 *Under Assumption 2.1 - 2.4,*

$$\sup_x \left| \Pr\left(U_n(\beta_0) \le x n^{1/2}\sigma_g\right) - \hat{\tilde{F}}_n(x)\right| = o(n^{-1/2}), \quad a.s.$$
$$\sup_x \left| \Pr\left(U_n(\beta_0)/\sqrt{I_n(\beta_0)} \le x\right) - \hat{E}_n(x)\right| = o(n^{-1/2}), \quad a.s.$$
$$\sup_x \left| \Pr\left(U_n(\beta_0)/\sqrt{I_n(\hat{\beta})} \le x\right) - \hat{\tilde{E}}_n(x)\right| = o(n^{-1/2}), \quad a.s. \tag{3.2}$$
$$\sup_x \left| \Pr\left(n^{1/2}\sigma(\hat{\beta} - \beta_0) \le x\right) - \hat{G}_n(x)\right| = o(n^{-1/2}), \quad a.s.$$
$$\sup_x \left| \Pr\left(\sqrt{I_n(\hat{\beta})}(\hat{\beta} - \beta_0) \le x\right) - \hat{E}_n(x)\right| = o(n^{-1/2}), \quad a.s.$$

where $\hat{F}_n(x)$, $\hat{E}_n(x)$, $\hat{\tilde{E}}_n(x)$, and $\hat{G}_n(x)$ are defined same way as that of $F_n(x)$, $E_n(x)$, $\tilde{E}_n(x)$, and $G_n(x)$ in (2.4) with Δ replaced by $\hat{\Delta}$, ζ by $\hat{\zeta}$, and κ_3 by $\hat{\kappa}_3$.

Another way of getting good estimates of the sampling distribution is by the method of bootstrap. Let $\{y_1^*, \cdots, y_n^*\}$ be an i.i.d. sample obtained by sampling with replacement from $\{y_1, y_2, \cdots, y_n\}$. Consider the likelihood

$$L_n^*(\beta) = \prod_{i=1,\cdots,n; x_i^* \le T} \left[\frac{\exp(\beta z_i^*)}{\sum_{j=1}^n \exp(\beta z_j^*) 1_{[x_j^* \ge x_i^*]}}\right]^{\delta_i^*}, \tag{3.3}$$

where $y_i^* = (z_i^*, x_i^*, \delta_i^*)$, $i = 1, \cdots, n$. The corresponding first derivative of $\log L_n^*(\beta)$ is

$$U_n^*(\beta) = \sum_{i=0}^n \int_0^T \left(z_i^* - \frac{\hat{\alpha}_1^*(t,\beta)}{\hat{\alpha}_0^*(t,\beta)}\right) dN_i^*(t),$$

and minus the second derivative is

$$I_n^*(\beta) = \sum_{i=0}^n \int_0^T \left(\frac{\hat{\alpha}_2^*(t,\beta)}{\hat{\alpha}_0^*(t,\beta)} - \frac{\hat{\alpha}_1^{*2}(t,\beta)}{\hat{\alpha}_0^{*2}(t,\beta)}\right) dN_i^*(t),$$

where $N_i^*(t) = 1_{[} x_i^* \le t, \delta^* = 1]$, $\hat{\alpha}_k^*(t,\beta) = \sum_{i=0}^n \xi_{i,k}^*(t,\beta)/n$, and $\xi_{i,k}^*(t,\beta) = z_i^{*k} e^{\beta z_i^*} 1_{[} x_i^* \ge t]$. Let $\hat{\beta}^*$ be a number which maximizes $L_n^*(\beta)$ in (3.3), then we have the following

Theorem 3.2 *Under the Assumptions 2.1-2.4, then almost surely*

$$\begin{aligned}
&\sup_x \left|\Pr{}_*\left(U_n^*(\hat\beta) \le x\sqrt{I_n(\hat\beta)}\right) - \Pr\left(U_n(\beta_0) \le x n^{1/2}\sigma\right)\right| = o(n^{-1/2}),\\
&\sup_x \left|\Pr{}_*\left(U_n^*(\hat\beta)/\sqrt{I_n^*(\hat\beta)} \le x\right) - \Pr\left(U_n(\beta_0)/\sqrt{I_n(\beta_0)} \le x\right)\right| = o(n^{-1/2}),\\
&\sup_x \left|\Pr{}_*\left(U_n^*(\hat\beta)/\sqrt{I_n^*(\hat\beta^*)} \le x\right) - \Pr\left(U_n(\beta_0)/\sqrt{I_n(\hat\beta)} \le x\right)\right| = o(n^{-1/2}), \quad (3.4)\\
&\sup_x \left|\Pr{}_*\left(\sqrt{I_n(\hat\beta)}(\hat\beta^* - \hat\beta) \le x\right) - \Pr\left(\sqrt{n}\sigma(\hat\beta - \beta_0) \le x\right)\right| = o(n^{-1/2}),\\
&\sup_x \left|\Pr{}_*\left(\sqrt{I_n^*(\hat\beta^*)}(\hat\beta^* - \hat\beta) \le x\right) - \Pr\left(\sqrt{I_n(\hat\beta)}(\hat\beta - \beta_0) \le x\right)\right| = o(n^{-1/2}),
\end{aligned}$$

where $\Pr{}_*()$ *is the bootstrap probability conditional on the observations* $y_1, \cdots, y_n$.

Remark 3.1. Theorem 3.1 and 3.2 can be used to construct confidence intervals. For example, let x_α be the α-th percentile of the function $\hat{\tilde{G}}_n$ in (3.2), then an approximate 95% confidence interval for β_0 is $IT_n = (\hat\beta - x_{.975}/\sqrt{I_n(\hat\beta)}, \hat\beta - x_{.025}/\sqrt{I_n(\hat\beta)})$. From Theorem 3.1, we have $\Pr\{\beta_0 \in IT_n\} = .95 + o(n^{-1/2})$. Note that in general the error of the coverage probability of confidence interval based on normal approximation is $O(n^{-1/2})$. To see the use of Theorem 3.2, let y_α^* be the α-th percentile of the function $F_*(y) = \Pr{}_*\left(U_n^*(\hat\beta)/\sqrt{I_n^*(\hat\beta)} \le y\right)$; then a one sided, approximate 5% type one error test for H_0: $\beta = \beta_0$ can be constructed by rejecting H_0 if $U_n(\beta_0)/\sqrt{I_n(\beta_0)} \ge y_{.95}^*$. As above, the type one error should be closer to 5% than in the case we reject H_0 if $U_n(\beta_0)/\sqrt{I_n(\beta_0)} \ge 1.96$. Note that if $\beta_0 = 0$, $U_n(\beta_0)/\sqrt{I_n(\beta_0)}$ is the log-rank statistics.

Remark 3.2. As first pointed out by Babu and Singh (1984) in the case of estimating mean, $\sqrt{I_n(\hat\beta)}$ in the first and fourth expression of (3.4) cannot be replaced by $\sqrt{I_n(\beta_0)}$. Similar facts for the general U-statistics are pointed out by Helmers (1991). See also Hartigan (1986). For the same reason, $\sqrt{I_n^*(\hat\beta^*)}$ in the second, third and last expressions cannot be replaced by $\sqrt{I_n(\beta_0)}$, or $\sqrt{I_n(\hat\beta)}$.

Remark 3.3. The $I_n^*(\hat\beta)$ appearing in (3.4), can be replaced by $\tilde I_n^*(\hat\beta)$ and $I_n^*(\hat\beta^*)$ can be replaced by $\tilde I_n^*(\hat\beta^*)$ where

$$\tilde I_n^*(\beta) = \sum_{i=1}^n \int_0^T \left(\frac{\hat\alpha_2^*(t,\beta)}{\hat\alpha_0^*(t,\hat\beta)} - \frac{\hat\alpha_1^{*2}(t,\beta)}{\hat\alpha_0^{*2}(t,\beta)}\right) dN_i(t). \tag{3.5}$$

This amounts to bootstraping the variance by keeping the original counting processes and the sampling only affects $\hat{\alpha}_k(t,\beta)$. Let

$$\tilde{I}_n(\beta) = \int_0^T \left(\hat{\alpha}_2(t,\beta) - \frac{\hat{\alpha}_1^2(t,\beta)}{\hat{\alpha}_0(t,\beta)} \right) \frac{\alpha_0(t)}{\hat{\alpha}_0(t,\beta)} \lambda_0(t) dt. \tag{3.6}$$

The above fact can be proved by noting that if we change $I_n(\beta_0)$ to $\tilde{I}_n(\beta_0)$ and $I_n(\hat{\beta})$ to $\tilde{I}_n(\hat{\beta})$ in (2.3), the distributions E_n , $\tilde{E}_n$ and $\tilde{G}_n$ will not change.

Theorem 3.2 proves that the bootstrap distribution is more accurate than the normal approximation. The proofs of Theorem 3.1 and Theorem 3.2 are deferred to Section 5.

4. Proof of Theorem 2.1.

As has been indicated in Section 1, the method employed in proving Theorem 2.1 is first to approximate the relevant statistics V_n by U-statistics U_n' and then use the known results of the Edgeworth expansion for U_n' (Bickel, Götze and Van Zwet,1986). The lemma which makes this to work is the following

Lemma 4.1 *Suppose U_n' and V_n be two random variables with* $\Pr(|U_n' - V_n| \geq o(1)) = o(n^{-1/2})$, *$U_n'$ satisfy*

$$\sup_x |\Pr(U_n' \leq x n^{1/2}\sigma_g) - F_n(x)| = o(n^{-1/2}), \tag{4.1}$$

where F_n is an absolutely continuous distribution function with bounded derivative. Then

$$\sup_x |\Pr(V_n \leq x n^{1/2}\sigma_g) - F_n(x)| = o(n^{-1/2}). \tag{4.2}$$

The next three lemmas are the approximation lemmas. Let us first define some notations. Define $A_i(t) = \int_0^t \xi_{i,0}(t,\beta_0)\lambda_0(t)dt$, where $\xi_{i,0}(t,\beta)$ has been defined in (3.1), and set $M_i(t) = N_i(t) - A_i(t)$. Then it is well known that $M_i(t)$, $i = 1, \cdots, n$ are independent martingales with $\langle M_i \rangle(t) = A_i(t)$ (see, for example, Gill, 1980).

Lemma 4.2 *Under Assumptions 2.1-2.3,*

$$\begin{aligned} U_n(\beta_0) &= \sum_{i=1}^n g(y_i) + \frac{1}{n}\sum_{i<j}\psi(y_i,y_j) + R_{1,n}, & (4.3)\\ I_n(\beta_0) &= n\sigma^2 + \sum_{i=1}^n h(y_i) + R_{2,n}, & (4.4) \end{aligned}$$

where for any $p \geq 2$, $E\left|R_{1,n}\right|^p = O(n^{-p/2})$, $E\left|R_{2,n}\right|^p = O(1)$ *and*

$$\begin{aligned}
g(y_i) &= \int_0^T \left(z_i - \frac{\alpha_1(t)}{\alpha_0(t)}\right) dM_i(t), \qquad h(y_i) = h_1(y_i) + h_2(y_i), \\
h_1(y_i) &= \int_0^T \left(\alpha_2(t) - \frac{\alpha_1^2(t)}{\alpha_0(t)}\right) \frac{dM_i(t)}{\alpha_0(t)}, \qquad h_2(y_i) = \int_0^T \rho_i(t)\lambda_0(t)dt, \\
\rho_i(t) &= \xi_{i,2}(t) - \alpha_2(t) - 2\frac{\alpha_1(t)}{\alpha_0(t)}(\xi_{i,1}(t) - \alpha_1(t)) + \frac{\alpha_1^2(t)}{\alpha_0^2(t)}(\xi_{i,0}(t) - \alpha_0(t)), \\
\psi(y_i, y_j) &= \psi_0(y_i, y_j) + \psi_0(y_j, y_i), \qquad \psi_0(y_i, y_j) = -\int_0^T \pi_i(t) dM_j(t), \qquad (4.5) \\
\pi_i(t) &= \frac{1}{\alpha_0(t)}(\xi_{i,1}(t) - \alpha_1(t)) - \frac{\alpha_1(t)}{\alpha_0^2(t)}(\xi_{i,0}(t) - \alpha_0(t)).
\end{aligned}$$

Define $\Delta_n(\beta)$ as the third derivative of $\log L_n(\beta)$ and

$$\begin{aligned}
\sigma^2(\beta) &= \int_0^T \left(\alpha_2(t,\beta) - \frac{\alpha_1^2(t,\beta)}{\alpha_0(t,\beta)}\right) \frac{\alpha_0(t)}{\alpha_0(t,\beta)} \lambda_0(t)dt, \\
\Delta(\beta) &= -\int_0^T \left(\frac{\alpha_3(t,\beta)}{\alpha_0(t,\beta)} - 3\frac{\alpha_2(t,\beta)\alpha_1(t,\beta)}{\alpha_0^2(t,\beta)} + 2\frac{\alpha_1^3(t,\beta)}{\alpha_0^3(t,\beta)}\right) \alpha_0(t)\lambda_0(t)dt.
\end{aligned}$$

Then we have

Lemma 4.3 *Under Assumptionsdoes 2.1-2.3, we have*

$$\begin{aligned}
I_n(\beta) &= n\sigma^2(\beta) + R_{3,n}(\beta), \\
\Delta_n(\beta) &= n\Delta(\beta) + R_{4,n}(\beta),
\end{aligned}$$

where for any $\epsilon > 0$, $\Pr\left(\sup_\beta \left|R_{j,n}\right| > n^\epsilon\right) = o(n^{-1/2})$, $j = 3, 4$.

Lemma 4.4 *Under Assumption 2.1-2.4,*

$$\begin{aligned}
(I_n(\beta_0))^{-1/2} U_n(\beta_0) &= \sum_{i=1}^n \frac{g(y_i)}{n^{1/2}\sigma} + \sum_{i<j} \frac{\tilde{\psi}(y_i, y_j)}{n^{3/2}\sigma} - \frac{\zeta}{2n^{1/2}\sigma^3} + R_{5,n}, \qquad (4.6) \\
n^{1/2}\sigma\left(\hat{\beta} - \beta_0\right) &= \sum_{i=1}^n \frac{g(y_i)}{n^{1/2}\sigma} + \sum_{i<j} \frac{\Psi(y_i, y_j)}{n^{3/2}\sigma} + \frac{\Delta/2 - \zeta}{n^{1/2}\sigma^3} + R_{6,n}, \qquad (4.7)
\end{aligned}$$

where for any $\epsilon > 0$, $\Pr\left(\left|R_{j,n}\right| > n^{-1+\epsilon}\right) = o(n^{-1/2})$, $j = 5, 6$ *and*

$$\begin{aligned}
\tilde{\psi}(y_i, y_j) &= \psi(y_i, y_j) - \frac{1}{2\sigma^2}\left(g(y_i)h(y_j) + g(y_j)h(y_i)\right), \\
\Psi(y_i, y_j) &= \psi(y_i, y_j) - \frac{1}{\sigma^2}\left(g(y_i)h(y_j) + g(y_j)h(y_i)\right) + \frac{\Delta}{\sigma^4} g(y_i)g(y_j),
\end{aligned}$$

Lemma 4.5 *Under Assumptions 2.1-2.3, for any* $\eta > 0$

$$\hat{\alpha}_k(t, \hat{\beta}) - \hat{\alpha}_k(t) = \alpha_{k+1}(t) \sum_{i=1}^{n} \frac{g(y_i)}{n\sigma^2} + R_{7,n}, \tag{4.8}$$

where $R_{7,n}$ *satisfies the same property as* $R_{5,n}$ *does in Lemma 4.3.*

Lemma 4.6 *Under Assumptions 2.1-2.4,*

$$\begin{aligned} I_n(\hat{\beta}) &= n\sigma^2 + \sum_{i=1}^{n} \left(h(y_i) - \frac{\Delta}{\sigma^2} g(y_i) \right) + R_{8,n}, & (4.9) \\ (I_n(\hat{\beta}))^{-1/2} U_n(\beta_0) &= \sum_{i=1}^{n} \frac{g(y_i)}{n^{1/2}\sigma} + \sum_{i<j} \frac{\psi^*(y_i, y_j)}{n^{3/2}\sigma} + \frac{\Delta - \zeta}{2n^{1/2}\sigma^3} + R_{9,n}, & (4.10) \\ (I_n(\hat{\beta}))^{1/2} (\hat{\beta} - \beta_0) &= \sum_{i=1}^{n} \frac{g(y_i)}{n^{1/2}\sigma} + \sum_{i<j} \frac{\tilde{\psi}(y_i, y_j)}{n^{3/2}\sigma} - \frac{\zeta}{2n^{1/2}\sigma^3} + R_{10,n}. & (4.11) \end{aligned}$$

where for any $\epsilon > 0$, $\Pr(|R_{8,n}| > n^{\epsilon}) = o(n^{-1/2})$, $R_{9,n}$ *and* $R_{10,n}$ *satisfy the same property as* $R_{5,n}$ *dose in Lemma 4.3 and*

$$\psi^*(y_i, y_j) = \psi(y_i, y_j) - \frac{1}{2\sigma^2} \left(g(y_i)h(y_j) + g(y_j)h(y_i) \right) + \frac{\Delta}{\sigma^4} g(y_i) g(y_j), \tag{4.12}$$

Proof of Lemma 4.2. Under model (1.1), $U_n(\beta_0)$ can be written as a martingale (see, for example, Andersen and Borgan, 1985), or

$$U_n(\beta_0) = \sum_{i=0}^{n} \int_0^T \left(z_i - \frac{\hat{\alpha}_1(t)}{\hat{\alpha}_0(t)} \right) dM_i(t). \tag{4.13}$$

Rewrite it as

$$U_n(\beta_0) = \sum_{i=0}^{n} g(y_i) - \int_0^T \left(\frac{\hat{\alpha}_1(t)}{\hat{\alpha}_0(t)} - \frac{\alpha_1(t)}{\alpha_0(t)} \right) dM_i(t)$$

and use the two terms Taylor expansion for the integrands of the last term, we have

$$U_n(\beta_0) = \sum_{i=0}^{n} g(y_i) + \sum_{i \neq j} \psi_0(y_i, y_j)/n + \sum_{i=1}^{n} \psi_0(y_i, y_i)/n + R_{n,1},$$

where $\psi_0(y_i, y_j)$ is defined in (4.5) and

$$R_{n,1} = \sum_{i=1}^{n} \int_0^T \left(\frac{(\hat{\alpha}_0(t) - \alpha_0(t))^2 \hat{\alpha}_1(t)}{\hat{\alpha}_0(t)\alpha_0^2(t)} - \frac{(\hat{\alpha}_0(t) - \alpha_0(t))(\hat{\alpha}_1(t) - \alpha_1(t))}{\hat{\alpha}_0(t)\alpha_0(t)} \right) dM_i(t).$$

Since $n(\hat{\alpha}_0(t) - \alpha_0(t))$ is a sum of i.i.d. variables, $R_{n,1}/n$ can be written as a U-statistics of degree three with degenerate kernel plus some higher order terms. An extended version of the inequality of Dharmadhikari, Fabian and Jogdeo (1968) shows that $E\left|R_{1,n}\right|^p = O(n^{-p/2})$. If p is an integer, we can use the inequality due to Serfling (1980) (also see Lee, 1990).

To prove (4.4), we write

$$I_n(\beta_0) - n\sigma^2 = A_1 + A_2 + A_3, \tag{4.14}$$

where

$$\begin{aligned} A_1 &= \int_0^T \left(\alpha_2(t) - \frac{\alpha_1^2(t)}{\alpha_0(t)}\right) \frac{dM_i(t)}{\hat{\alpha}_0(t)}, \\ A_2 &= \int_0^T \left(\hat{\alpha}_2(t) - \alpha_2(t) - \frac{\hat{\alpha}_1^2(t)}{\hat{\alpha}_0(t)} + \frac{\alpha_1^2(t)}{\alpha_0(t)}\right) \lambda_0(t)dt, \\ A_3 &= \int_0^T \left(\hat{\alpha}_2(t) - \alpha_2(t) - \frac{\hat{\alpha}_1^2(t)}{\alpha_0(t)} + \frac{\alpha_1^2(t)}{\alpha_0(t)}\right) \frac{dM_i(t)}{\hat{\alpha}_0(t)}. \end{aligned} \tag{4.15}$$

A_1 can be written as

$$\sum_{i=1}^n h_1(y_i) + R^{(1)}, \tag{4.16}$$

where

$$R^{(1)} = \sum_{i=1}^n \int_0^T \left(\alpha_2(t) - \frac{\alpha_1^2(t)}{\alpha_0(t)}\right) \left(\frac{1}{\hat{\alpha}_0(t)} - \frac{1}{\alpha_0(t)}\right) dM_i(t). \tag{4.17}$$

By an argument similar to the Proof of (4.3), we can show that $E\left|R^{(1)}\right|^p = O(1)$. For A_2, using two terms Taylor expansion of x^2/y shows that

$$A_2 = \sum_{i=1}^n h_2(y_i) + R^{(2)} \tag{4.18}$$

where

$$R^{(2)} = -\int_0^T \frac{(\hat{\alpha}_1(t)\alpha_0(t) - \alpha_1(t)\hat{\alpha}_0(t))^2}{\hat{\alpha}(t)\alpha_0^2(t)} \lambda_0(t)dt. \tag{4.19}$$

Once again, the moment inequality for the U-statistics with degenerate kernel can be used to show $E\left|R^{(2)}\right|^p = O(1)$ and $E\left|A_3\right|^p = O(1)$. The proof is finished by setting $R_{2,n} = R^{(1)} + R^{(2)} + A_3$.

Proof of Lemma 4.3. We note that

$$\begin{aligned} I_n(\beta) &= \sum_{i=1}^{n} \int_0^T \left(\hat{\alpha}_2(t,\beta) - \frac{\hat{\alpha}_1^2(t,\beta)}{\hat{\alpha}_0(t,\beta)} \right) \frac{dN_i(t)}{\hat{\alpha}_0(t,\beta)}, \\ \Delta_n(\beta) &= -\sum_{i=1}^{n} \int_0^T \left(\frac{\hat{\alpha}_3(t,\beta)}{\hat{\alpha}_0(t,\beta)} - 3\frac{\hat{\alpha}_2(t,\beta)\hat{\alpha}_1(t,\beta)}{\hat{\alpha}_0^2(t,\beta)} + 2\frac{\hat{\alpha}_1^3(t,\beta)}{\hat{\alpha}_0^3(t,\beta)} \right) dN_i(t). \end{aligned}$$

The proof is similar to the approximation lemmas for stochastic integrals of empirical-type processes developed in Lai and Ying (1988). We omit the details.

Proof of Lemma 4.4. For given $\epsilon > 0$, choose $\epsilon_j > 0$, $j = 1, \cdots, 5$ such that $\sum \epsilon_j < \epsilon$. Let

$$\begin{aligned} E_n &= \Big\{ |R_{1,n}| < n^{-1/2+\epsilon_1};\ |R_{2,n}| < n^{\epsilon_2};\ \sup_\beta |R_{j,n}(\beta)| < n^{1/2+\epsilon_j}, j = 3,4; \\ &\qquad \left| \sum_{i=1}^{n} g(y_i) \right| < n^{1/2+\epsilon_5} \Big\} \end{aligned} \tag{4.20}$$

Then by Lemmas 4.2 and 4.3, the fact that $g(y_i)$ are mean zero bounded variables (therefore $E|\sum_{i=1}^n g(y_i)|^p = O(n^{p/2})$) and Markov's inequality, we have $\Pr(E_n) = o(n^{-1/2})$. On the set E_n, a simple calculation shows

$$\begin{aligned} & (I_n(\beta_0))^{-1/2} U_n(\beta_0) \\ = & \sum_{i=1}^{n} \frac{g(y_i)}{n^{1/2}\sigma} + \sum_{i<j} \frac{\psi(y_i,y_j)}{n^{3/2}\sigma} - \frac{1}{2n^{3/2}\sigma^3} \sum_{i=1}^{n} g(y_i) \sum_{j=1}^{n} h(y_j) + O(n^{-1+\epsilon}). \end{aligned} \tag{4.21}$$

The third term of (4.21) can be written as

$$-\frac{\zeta}{2n^{1/2}\sigma^3} - \sum_{i<j} \frac{g(y_i)h(y_j) + g(y_j)g(y_i)}{2n^{3/2}\sigma^3} + R^{(3)}, \tag{4.22}$$

where $R^{(3)} = 2n^{-3/2}\sigma^{-3}\sum_{i=1}^n (g(y_i)h(y_i) - \zeta)$. It is easy to see (since $\zeta = E[g(y_1)h(y_1)]$, see Gu, (1992)) $E|R^{(3)}|^p = O(n^{-p})$ for any $p > 0$ and therefore $\Pr\left(\left|R^{(3)}\right| \ge n^{-1+\epsilon}\right) = o(n^{-1/2})$ Inserting (4.22) into (4.21) we get (4.6).

For the proof of (4.7), we first note that Assumptions 2.2 and 2.4 imply that $\sigma^2(\beta)$ is bounded below from zero since $\sigma^2(\beta)$ is continuous in β and

$$\sigma^2(\beta) = \int_0^T E\left[\left(z_1 - \frac{\alpha_0(t,\beta)}{\alpha_1(t,\beta)} \right)^2 \xi_{1,0}(t) \right] \frac{\alpha_0(t)}{\alpha_0(t,\beta)} \lambda_0(t) dt = 0$$

implies that z_1 is almost surely constant, which contradicts Assumption 2.4.

From $0 = U_n(\beta_0) - I_n(\beta^*)(\hat{\beta} - \beta_0)$, the discussion above and Lemma 4.3, we have that on the set E_n as defined in (4.20), $\hat{\beta} - \beta_0 = O(n^{-1/2+\epsilon})$. A two term Taylor expansion gives

$$0 = U_n(\beta_0) - I_n(\beta_0)(\hat{\beta} - \beta_0) + \Delta_n(\beta^*)(\hat{\beta} - \beta_0)^2/2, \tag{4.23}$$

so that on the set E_n

$$\begin{aligned} \hat{\beta} - \beta_0 &= \frac{U_n(\beta_0)}{I_n(\beta_0)} + O(n^{-1+3\epsilon}) \\ &= \sum_{i=1}^{n} \frac{g(y_i)}{n\sigma^2} + O(n^{-1+3\epsilon}). \end{aligned} \tag{4.24}$$

In (4.23), use (4.3) for $U_n(\beta_0)$, (4.4) for $I_n(\beta_0)$, $n\Delta + \Delta_n'(\beta^{**})$ for $\Delta_n(\beta^*)$, and (4.24) for $(\hat{\beta} - \beta_0)^2$ and noting that $\Delta_n'(\beta) = O(n)$, we shall get on E_n

$$\begin{aligned} \hat{\beta} - \beta_0 &= \frac{1}{n\sigma^2} \left[\sum_{i=1}^{n} g(y_i) + \frac{1}{n} \sum_{i<j} \psi(y_i, y_j) \right. \\ &\quad \left. + \frac{n\Delta}{2} \left(\frac{g(y_i)}{n\sigma^2} \right)^2 - \frac{1}{n\sigma} \sum_{i=1}^{n} g(y_i) \sum_{j=1}^{n} h(y_j) \right] + O(n^{-1+3\epsilon}), \end{aligned}$$

which is equivalent to (4.7) since $\zeta = E[g(y_1)h(y_1)]$.

Proof of Lemma 4.5. We have

$$\begin{aligned} \hat{\alpha}_k(t, \hat{\beta}) - \hat{\alpha}_k(t) &= \frac{1}{n} \sum_{i=1}^{n} z_i^k (e^{\hat{\beta} z_i} - e^{\beta_0 z_i}) 1_{[x_i \ge t]} \\ &= \frac{\hat{\beta} - \beta_0}{n} \sum_{i=1}^{n} z_i^{k+1} e^{\beta_0 z_i} 1_{[x_i \ge t]} + O\left((\hat{\beta} - \beta_0)^2 \right) \end{aligned} \tag{4.25}$$

by Taylor expansion. The rest of the proof follows easily from (4.25) and (4.24).

Proof of Lemma 4.6. Simple algebra shows

$$I_n(\hat{\beta}) - n\sigma^2 = \sum_{i=1}^{n} h_1(y_i) + V_1 + V_2,$$

where

$$\begin{aligned} V_1 &= \sum_{i=1}^{n} \int_0^T \left[\left(\frac{\hat{\alpha}_2(t, \hat{\beta})}{\hat{\alpha}_0(t, \hat{\beta})} - \frac{\hat{\alpha}_1^2(t, \hat{\beta})}{\hat{\alpha}_0^2(t, \hat{\beta})} \right) - \left(\frac{\alpha_2(t)}{\alpha_0(t)} - \frac{\alpha_1^2(t)}{\alpha_0^2(t)} \right) \right] dM_i(t), \\ V_2 &= n \int_0^T \left[\left(\hat{\alpha}_2(t, \hat{\beta}) - \frac{\hat{\alpha}_1^2(t, \hat{\beta})}{\hat{\alpha}_0(t, \hat{\beta})} \right) \frac{\hat{\alpha}_0(t)}{\hat{\alpha}_0(t, \hat{\beta})} - \left(\alpha_2(t) - \frac{\alpha_1^2(t)}{\alpha_0(t)} \right) \right] \lambda_0(t) dt. \end{aligned}$$

Similar to the proof of Lemma 4.2, for any $\epsilon > 0$, we have $\Pr\left(|V_1| > n^{\epsilon}\right) = o(n^{-1/2})$ and V_2/n can be written as

$$\begin{aligned} & \int_0^T \left((\hat{\alpha}_2(t,\hat{\beta}) - \alpha_2(t)) + \frac{\alpha_2(t)}{\alpha_0(t)} \left[(\hat{\alpha}_0(t) - \alpha_0(t)) - (\hat{\alpha}_0(t,\hat{\beta}) - \alpha_0(t)) \right] \right) \lambda_0(t)dt \\ - & \int_0^T 2\frac{\alpha_2(t)}{\alpha_0(t)} (\hat{\alpha}_1(t,\hat{\beta}) - \alpha_1(t))\lambda_0(t)dt \qquad\qquad (4.26) \\ - & \int_0^T \frac{\alpha_1^2(t)}{\alpha_0^2(t)} \left[(\hat{\alpha}_0(t) - \alpha_0(t)) - 2(\hat{\alpha}_0(t,\hat{\beta}) - \alpha_0(t)) \right] \lambda_0(t)dt + R^{(4)} \end{aligned}$$

where $\Pr\left(\left|R^{(4)}\right| > n^{-1+\epsilon}\right) = o(n^{-1/2})$. Applying (4.25), after some simplification (details are omitted), (4.26) becomes $-\sum g(y_i)\Delta/n\sigma^2 + \sum h_2(y_i)/n + R^{(5)}$, where $R^{(5)}$ satisfies the same property as $R^{(4)}$. The proof of (4.9) is complete if we set $R_{8,n} = V_1 + nR^{(5)}$.

The proof of (4.10) and (4.11) is a direct application of (4.9), (4.3), (4.7) and arguments similar to the proof of Lemma 4.4. We omit the details.

Proof of Theorem 2.1. By Lemma 4.1, Lemma 4.2 and the discussion preceding it, we need to show that $\Pr(|R_{1,n}| \geq o(1)) = o(n^{-1/2})$ and $U_n' = \sum_{i=1}^n g(y_i) + \frac{1}{n}\sum_{i<j}\psi(y_i, y_j)$ satisfy (4.1) with F_n as $\tilde{F}_n$ in (2.4). The former is based on Markov's inequality and the later is guaranteed by the Edgeworth expansion theorem for U-statistics (see, for example, Bickel, Götze and Van Zwet, 1986). $g(y_1)$ is non-lattice since λ_0 is continuous. In Gu (1992), it has been shown

$$\kappa_3 = \sigma^{-3}E[g^3(y_1)] + 3\sigma^{-3}E[g(y_1)g(y_2)\psi(y_1,y_2)] = -\frac{\Delta}{\sigma^3} + 3\frac{\zeta}{\sigma^3}.$$

To show the second expression of (2.3), we use the same argument with (4.6) of Lemma 4.4. For the third expression of (2.3), we use (4.10) of Lemma 4.6. For the fourth expression of (2.3), we use (4.7) of Lemma 4.4. For the fifth expression of (2.3), we use (4.11) of Lemma 4.6. The argument is similar and we omit the details.

Remark 4.1. The result on the Edgeworth expansion for Studentized U-statistics (Helmers, 1991) cannot be applied to show the Edgeworth expansion of the Studentized statistics $U_n(\beta_0)/\sqrt{I_n(\beta_0)}$, $U_n(\beta_0)/\sqrt{I_n(\hat{\beta})}$ or $\sqrt{I_n(\hat{\beta})}(\hat{\beta} - \beta_0)$ since the variance estimator is not the jacknife estimator.

5. Proof of Theorem 3.1 and 3.2.

Proof of Theorem 3.1. From Theorem 2.1, we only have to show $\sup_x \left|\tilde{F}_n(x) - \hat{\tilde{F}}_n(x)\right| = o(n^{-1/2})$, $\cdots$, almost surely. which requires to show $\hat{\sigma}^2 - \sigma^2$, $\hat{\sigma}_A^2 - \sigma_A^2$, *etc.*, $\hat{\Delta}_B - \Delta_B$ are $o(1)$ almost surely. The latter is guaranteed by their definitions.

Proof of Theorem 3.2. Denote

$$M_i^*(t) = N_i^*(t) - \int_0^t \xi_{i,0}^*(s,\hat{\beta})\frac{dN(s)}{n\hat{\alpha}_0(s,\hat{\beta})} = N_i^*(t) - \int_0^t \xi_{i,0}^*(s,\hat{\beta})d\hat{\Lambda}_0(s).$$

where $N(t) = \sum_{i=1}^n N_i(t)$. Then it is easy to check that $M_i^*(t)$ is a mean zero random variable with respect to resampling probability $\Pr_*$. Parallel to (4.13), we can show that

$$U_n^*(\hat{\beta}) = \sum_{i=0}^n \int_0^T \left(z_i^* - \frac{\hat{\alpha}_1^*(t,\hat{\beta})}{\hat{\alpha}_0^*(t,\hat{\beta})}\right) dM_i^*(t)$$

since

$$\sum_{i=0}^n \int_0^T \left(z_i^* - \frac{\hat{\alpha}_1^*(t,\beta)}{\hat{\alpha}_0^*(t,\beta)}\right) \xi_{i,0}^*(s,\hat{\beta})d\hat{\Lambda}_0(s) = 0.$$

Parallel to Lemma 4.2, we can show that

$$U_n^*(\hat{\beta}) = \sum_{i=0}^n g_n(y_i^*) + \frac{1}{n}\sum_{i<j}\psi_n(y_i^*, y_j^*) + O^*(n^{-1/2+\eta}),$$

where $O^*(a_n)$ means that there exists a constant C such that $\Pr_*(O^*(a_n) \geq Ca_n) \leq o(n^{-1/2})$, almost surely

$$g_n(y_i^*) = \int_0^T \left(z_i^* - \frac{\hat{\alpha}_1(t,\hat{\beta})}{\hat{\alpha}_0(t,\hat{\beta})}\right) dM_i^*(t) \tag{5.1}$$

and ψ_n is similarly defined with $\xi_{i,k}(t)$ replaced by $\xi_{i,k}^*(t,\hat{\beta})$, $\alpha_k(t)$ replaced by $\hat{\alpha}_k(t,\hat{\beta})$ and M_i replaced by M_i^* in the definition of ψ (see (4.5)). A simple calculation yields $E^*[g_n(y_i^*)] = 0$ (which uses $U_n(\hat{\beta}) = 0$). Similarly, $E^*[\psi_n(y_i^*, y_j^*)|y_i^*] = 0$. We will have results similar to Lemma 4.2-4.6 for the bootstrapped counterpart in Section 4.

The $g_n(y_i)$ is the counterpart of $g(y_i)$ in Sections 2 and 4. Instead of using the result on Edgeworth expansion for U-statistics of degree two (as we did in Section 4, we need to check that the distribution of $g_n(y_1^*)$ is almost surely non-lattice when n is large), we can show the statements in (3.4) directly. We first show that for any fixed a,

$$\Theta_n = \sup_{|t|\leq a}\left|E^*[e^{itg_n(y_1^*)}] - E[e^{itg(y_1)}]\right| \to 0, \quad \text{almost surely.} \tag{5.2}$$

The proof is similar to the proof of (4.8) of Helmers (1991). We first write

$$\begin{aligned}
\Theta_n &= \sup_{|t|\le a}\left| n^{-1}\sum_{i=1}^n \left\{ e^{itg_n(y_i)} - e^{itg(y_i)} \right\} \right| \\
&\quad + \sup_{|t|\le a}\left| n^{-1}\sum_{i=1}^n e^{itg(y_i)} - E[e^{itg(y_1)}] \right| = \Theta_{n,1} + \Theta_{n,2}.
\end{aligned}$$

$\Theta_{n,2} \to 0$ is a consequence of a theorem of Feuerverger and Mureika (1977) and

$$\begin{aligned}
\Theta_{n,1} &\le an^{-1}\sum_{i=1}^n |g_n(y_i) - g(y_i)| \\
&\le an^{-1}\sum_{i=1}^n \left| \int_0^T \left(\frac{\hat\alpha_1(t,\hat\beta)}{\hat\alpha_0(t,\hat\beta)} - \frac{\alpha_1(t)}{\alpha_0(t)} \right) dN_i(t) \right| \qquad (5.3) \\
&\quad + an^{-1}\sum_{i=1}^n \left| \int_0^T \left(z_i - \frac{\hat\alpha_1(t,\hat\beta)}{\hat\alpha_0(t,\hat\beta)} \right) \frac{\xi_{i,0}(t,\hat\beta)}{n\hat\alpha_0(t,\hat\beta)} dN(t) - \int_0^T \left(z_i - \frac{\alpha_1(t)}{\alpha_0(t)} \right) \xi_{i,0}\lambda_0(t)dt \right|
\end{aligned}$$

The first term on the right goes to 0 almost surely since $\sup_t \left|\hat\alpha_k(t,\hat\beta) - \alpha_k(t)\right| \to 0$ almost surely for $k = 0, 1$ and the second term is bounded by

$$\begin{aligned}
& an^{-1}\sum_{i=1}^n \left| \int_0^T \left[\left(z_i - \frac{\hat\alpha_1(t,\hat\beta)}{\hat\alpha_0(t,\hat\beta)} \right) \frac{\xi_{i,0}(t,\hat\beta)}{n\hat\alpha_0(t,\hat\beta)} - \left(z_i - \frac{\alpha_1(t)}{\alpha_0(t)} \right) \frac{\xi_{i,0}(t)}{n\hat\alpha_0(t)} \right] dN(t) \right| \\
&+ an^{-1}\sum_{i=1}^n \left| \sum_{j=1}^n \int_0^T \left(z_i - \frac{\alpha_1(t)}{\alpha_0(t)} \right) \xi_{i,0}(t) \left(\frac{1}{n\hat\alpha_0(t)} - \frac{1}{n\alpha_0(t)} \right) dM_j(t) \right| \\
&+ an^{-1}\sum_{i=1}^n \left| \sum_{j=1}^n \int_0^T \left(z_i - \frac{\alpha_1(t)}{\alpha_0(t)} \right) \frac{\xi_{i,0}(t)}{n\alpha_0(t)} dM_j(t) \right|
\end{aligned}$$

where the first and second term go to 0 for the same reason as the first term of (5.3) and $\sup_{t,i} |\xi_{i,0}(t,\hat\beta) - \xi_{i,0}(t)| \to 0$ almost surely. The third term goes to zero almost surely by the same argument containing the last three displayed equations of Helmers (1991). So we have proved (5.2).

Once we have (5.2), we can follow the arguments as in the proof of Lemma 2.1 of Bickel, Götze and Van Zwet (1986) (noting that only the value of $E[e^{itg(y_1)}]$ for $|t| \le a$ for some fixed a is used in the proof) to show that

$$\int_{-a\sqrt{n}}^{a\sqrt{n}} \left| \frac{\varphi_n(t) - \tilde\varphi_n(t)}{t} \right| = o(n^{-1/2}), \quad \text{almost surely}$$

where $\varphi_n(t) = E^* \exp\left\{ it \left(\sum_{i=1}^n g_n(y_i^*) + \sum_{i<j} \psi_n(y_i^*, y_j^*)/n \right) / n^{-1/2}\sigma^* \right\}$, σ^{*2} is the variance of $g_n(y_1^*)$ under the resampling probability (similar to the proof of (5.2), we can

show that $\sigma^{*2} \to \sigma^2$, almost surely) and $\tilde{\varphi}_n(t)$ is the characteristic function corresponds to $\tilde{F}_n$ in the case of proving the first expression of (3.4). The proof is finished by Esseen's smoothing lemma (Feller, 1971, page 538), Lemma 4.1 and Theorem 2.1. The proof of the rest of the expressions in (3.4) is similar with the U-statistic approximation at hand. We omit the remaining details.

References

Andersen, P. K. and Gill, D. R. (1982). Cox's regression model for counting processes. Ann. Statist., 10, 1100-1120.

Andersen, P. K. and Borgan, O. (1985). Counting process model for life history data: A review. Scand. J. Statist., 12, 97-158.

Babu, G. and Singh, K. (1984). On one-term Edgeworth expansion correction by Efron's bootstrap. Sankhyá, 46, Series A. Part 2, 219-232.

Bailey, K. (1983). The asymptotic joint distribution of regression and survival parameter in the Cox Regression Model. Ann. Statist., 11, 39-48.

Bickel, P. J., Götze, F. and Van Zwet, W. R. (1986). The Edgeworth expansion for U-statistics of degree two. Ann. Statist., 14, 1463-1484.

Cox, D. R. (1972). Regression model and life tables (with discussion). J. Roy. Statist. Soc. Ser. B, 34, 187-220.

Cox, D. R. (1975). Partial likelihood. Biometrika, 62, 269-276.

Dharmadhikari, S. W., Fabian, V. and Jogdeo, K. (1968). Bounds on the moments of martingales. *Ann. Math. Statist.* **39**, 1719-1723.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications* 2, 2nd ed. Wiley, New York.

Feuerverger, A. and Mureika, R. A. (1977). The empirical characteristic function and its applications. Ann. Statist., 5, 88-97.

Gill, R.D. (1980). *Censoring and Stochastic Integrals.* Mathematisch Centrum Tract 124, Amsterdam.

Gill, R.D. (1984). Understanding Cox's Regression Model: A Martingale Approach. J. Amer. Statist. Assoc. 79, 441-447.

Gu, M. (1992). Some Moments Concerning the Cox regression Model. Report from the Dept. of Math. and Stat., McGill University 92-01.

Hall, P. (1988). Theoretical comparison of Bootstrap confidence intervals. Ann. Statist., 16, 927-953.

Hartigan, J. A. (1986). Comment on " Bootstrap Methods for measures of Statistical Accuracy", by Efron and Tibshirani, Statistical Science, Vol. 1, no. 1, 75-77.

Helmers, R. (1991). On the Edgeworth expansion and the bootstrap approximation for a Studentized U-statistic. Ann. Statist., 19, 470-484.

Lai, T. L. and Ying, Z. (1988). Stochastic integrals of empirical-type processes with applications to censored regression. J. Multivariate Anal., 27, 334-358.

Lee, A. J. (1990). *U-statistics, Theory and Practice.* Marcel Dekker, New York.

Naes, T. (1982). The asymptotic distribution of the estimator for the regression parameter in Cox's regression model. Scand. J. Statist., 9, 107-115.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

Tsiatis, A. A. (1981). A large sample study of the Cox's regression model. Ann. Statist., 9, 93-108.

Department of Mathematics and Statistics
McGill University
Burnside Hall, 805 Sherbrooke St. W.
Montréal, P.Q. H3A 2K6

On the Edgeworth Expansion and Bootstrap Approximation for the Cox Regression Model under Random Censorship

Minggao Gu[1]
McGill University

Abstract

In this paper, we establish the one-term Edgeworth expansion for various statistics related to Cox semi-parametric regression model when the covariate is one dimensional and the observations are i.i.d. We show that the Bootstrap approximation method is second order correct. The second order correct estimates of the sampling distribution can be obtained without Monte-Carlo simulation. We pay special attention to the studentized version of the statistics and show that their distributions are different from those of the original statistics to the order of $n^{-1/2}$.

Abbreviated Title. Edgeworth Expansion for Cox Model.
AMS 1980 subject classification. Primary 62E20, 62G05; Secondary 62P10.
Key Words and Phrases. Asymptotic representation, Bootstrap, Cox regression model, Edgeworth expansion, Martingale, Survival data.

[1] Partially supported by NSERC and FCAR.