Statistica Sinica **19** (2009), 969-979

RANDOM WEIGHTING AND EDGEWORTH EXPANSION FOR THE NONPARAMETRIC TIME-DEPENDENT AUC ESTIMATOR

Chin-Tsang Chiang, Shao-Hsuan Wang and Hung Hung

National Taiwan University

Abstract: A confidence region for the time-dependent area under the receiver operating characteristic curve (AUC) can be constructed based on the asymptotic normality of a non-parametric estimator. In numerical studies, it was found that the performance of the normal approximated confidence interval is dramatically affected by small sample size and high censoring rate. To improve the accuracy of coverage probabilities as well as interval estimators, the random weighted bootstrap distribution and the Edgeworth expansion with remainder term $o(n^{-1/2})$ are proposed to approximate the sampling distribution of the estimator. The asymptotic properties of random weighted bootstrap analogue and the one-term Edgeworth expansion are developed in this article. The usefulness of the proposed procedures are confirmed by a class of simulations with different sample sizes and censoring rates. Moreover, our methods are demonstrated using the ACTG 175 data.

Key words and phrases: AUC, Edgeworth expansion, Kaplan-Meier estimator, normal approximation, random weighted bootstrap, survival data, U-statistic.

1. Introduction

In the ACTG study 175, the research objective is to evaluate the predictive ability of CD4 biomarker on patient's survival time from the entry date to the date of AIDS diagnosis or death measured in weeks. For the ith individual, let Y_i and T_i denote the diagnostic test and the time of disease or death, $i = 1, \cdots, n$. One of the most popular scalar measures to evaluate the performance of Y to the disease or vital status $\{T \leq t\}$ is the time-dependent AUC $\theta_t = P(Y_i > Y_j | T_i \leq t, T_j > t)$ for $i \neq j$. Note that the definition of θ_t is a natural extension of traditional AUC with the cases and controls being defined over time. Since patients might be lost to follow-up or drop-out during the study period, we present the statistical inferences on θ_t based on the censored survival data $\{(X_i, \delta_i, Y_i)\}_{i=1}^n$, where $X_i = min\{T_i, C_i\}$ is the last observed time, with C_i being the censoring time and $\delta_i = I(X_i = T_i)$ the censoring status. Let $D_{it}^{*c} = I(X_i > t)$, $\phi_{ij} = I(Y_i > Y_j)$, and $S_C(t)$ be the survival distribution of the censoring time C. By assumption (**A1**: C and (Y, T) are independent) and the property $E(D_{it}^{*c}/S_C(t)) = E(I(T_i > t))$, we propose a simple and easily computed estimator

$$\widehat{\theta}_t = \frac{\sum_{i \neq j}(\widehat{S}_C(t) - D_{it}^{*c})D_{jt}^{*c}\phi_{ij}}{\sum_{i \neq j}(\widehat{S}_C(t) - D_{it}^{*c})D_{jt}^{*c}} \tag{1.1}$$

for θ_t, where $\widehat{S}_C(t)$ is the non-parametric Kaplan-Meier estimator of $S_C(t)$. An alternative estimation approach can refer to the recursive estimation procedure of Chambless and Diao (2006) at the observed failure times. It was detected in Web Figure 1 that both estimators have very similar performance under complete failure time data or type I censoring data. The recursive estimator was further found to have substantially large bias and variance in regions of sparse failure times, although its variance was relatively small in regions where failure times are dense.

The asymptotic normality of $\widehat{\theta}_t$ and the consistency of a variance estimator enable us to construct an approximate confidence interval for θ_t (See the Web Appendix A). To improve the performance of the interval for θ_t under small sample size and high censoring rate, we construct two alternative confidence intervals based on the random weighted bootstrap approximation and the one-term Edgeworth expansion. Tu and Zheng (1991) found that the random weighted bootstrap confidence interval has better coverage probability than the naive bootstrap one, especially when the sample size is small and the sampling distribution of estimator is non-normal. Chiang, James and Wang (2005) found that the weighted bootstrap and the naive bootstrap yielded very similar results in the recurrent event data setting, but the computational speed of the weighted bootstrap procedure is better than that of naive bootstrap one. In this article, the random weighted bootstrap distribution is used to estimate the sampling distribution of $\widehat{\theta}_t$. Since the estimator $\widehat{\theta}_t$ and the corresponding bootstrap analogue are U-statistics with dependent random quantities caused by the estimator $\widehat{S}_C(t)$ in the kernel function, the consistency of the weighted bootstrap approximation to the sampling distribution of $\widehat{\theta}_t$ is derived via first using an appropriate U-statistic approximation based on independent random quantities. For the sampling distribution of a U-statistic with independent random quantities in the kernel function, Callaert, Janssen and Veraverbeke (1980) and Bickel, Götze and Van Zwet (1986) established the corresponding Edgeworth expansion. Callaert and Veraverbeke (1981) studied the order of normal approximation for a studentized U-statistic. Based on a studentized U-statistic, Helmers (1991) provided the one-term Edgeworth expansion to improve the coverage probability of confidence interval. By extending this approach to the studentized statistic of $\widehat{\theta}_t$, the one-term Edgeworth expansion is established.

The rest of this paper is organized as follows. The random weighted bootstrap approximation and the one-term Edgeworth expansion of the sampling distribution of $\widehat{\theta}_t$ are separately established in Section 2. In Section 3, the performance and the comparison of different procedures are investigated through a class of simulations with different settings on the sample size and the censoring rate. Our proposed methods are further applied to the ACTG 175 data in Section 4. Finally, a brief discussion is provided in Section 5. Web Appendices and Figure referenced in Sections 1-2 are available at the website `http://www.stat.sinica.edu.tw/statistica`.

2. Random Weighting and Edgeworth Expansion

In this section, a generalized bootstrap procedure is used to construct an approximate confidence interval for θ_t. The Edgeworth expansion procedure is further proposed to improve the coverage probability of the normal approximate confidence interval under small sample size and high censoring rate.

2.1. Weighted bootstrap procedure

Let $D_n = \{(X_i, \delta_i, Y_i)\}_{i=1}^n$ represent the censored survival data. Independent of D_n, the random variables $\xi_1, \cdots, \xi_n$ are independently generated from a common distribution with mean $\mu = E(\xi_1)$, variance $\sigma^2 = Var(\xi_1)$, satisfying $P(\xi_1 = 0) < 1$. The weighted bootstrap analogue of $\widehat{\theta}_t$ is defined as

$$\widehat{\theta}_t^w = \frac{\sum_{i\neq j} w_i w_j (\widehat{S}_C^w(t) - D_{it}^{*c}) D_{jt}^{*c} \phi_{ij}}{\sum_{i\neq j} w_i w_j (\widehat{S}_C^w(t) - D_{it}^{*c}) D_{jt}^{*c}}, \tag{2.1}$$

where $w_i = \xi_i / (\sum_{j=1}^n \xi_j)$ and $\widehat{S}_C^w(t) = \int_0^t \widehat{S}_C^w(u^-) dN_\cdot^w(u)/R_\cdot^w(u)$, with $N_\cdot^w(t) = \sum_{i=1}^n w_i I(X_i \leq t, \delta_i = 0)$ and $R_\cdot^w(t) = \sum_{i=1}^n w_i I(X_i \geq t)$. In the next theorem, we show that the conditional distribution of $\rho n^{1/2}(\widehat{\theta}_t^w - \widehat{\theta}_t)$ is a consistent estimator of the distribution of $n^{1/2}(\widehat{\theta}_t - \theta_t)$.

Theorem 2.1. *Suppose that assumptions* (A1)−(A2) *are satisfied. Then*

$$\sup_{x\in R} |P(\rho n^{1/2}(\widehat{\theta}_t^w - \widehat{\theta}_t) \leq x | D_n) - P(n^{1/2}(\widehat{\theta}_t - \theta_t) \leq x)| \xrightarrow{p} 0 \text{ as } n \to \infty, \tag{2.2}$$

where $\rho = \mu/\sigma$ is a scale factor modification for the variability in the weights.

Proof. See the proof in the Web Appendix B.

Theorem 2.1 forms the basis of the random weighted bootstrap procedure and an approximate $(1-\alpha)$, confidence interval for θ_t is constructed via

$$(\widehat{\theta}_t - \rho q^*_{1-\alpha/2}(\widehat{\theta}_t^w - \widehat{\theta}_t), \widehat{\theta}_t - \rho q^*_{\alpha/2}(\widehat{\theta}_t^w - \widehat{\theta}_t)), \tag{2.3}$$

where $q_p^*(\cdot)$ denotes the 100pth percentile of B weighted bootstrap estimators $(\widehat{\theta}_t^w - \widehat{\theta}_t), 0 < \alpha < 1$.

Remark. The construction of naive bootstrap confidence interval is mainly based on the B bootstrap estimators $(\widehat{\theta}_t^n - \widehat{\theta}_t)$ with the bootstrap analogue $\widehat{\theta}_t^n$ being computed from the bootstrap sample, in which the entire measurements of each subject are re-sampled with replacement from D_n.

2.2. Edgeworth expansion procedure

A simple decomposition shows that

$$(\widehat{\theta}_t - \theta_t) = \frac{U_{nt}^{(1)} + \eta_t(\widehat{S}_C(t) - S_C(t)) + U_{nt}^{(2)}(\widehat{S}_C(t) - S_C(t))}{V_{nt}}, \tag{2.4}$$

where $U_{nt}^{(1)} = \sum_{i \neq j} H_{ijt}^{(1)}/[n(n-1)]$ with $H_{ijt}^{(1)} = [(S_C(t) - D_{it}^{*c})D_{jt}^{*c}(\phi_{ij} - \theta_t) + (S_C(t) - D_{jt}^{*c})D_{it}^{*c}(\phi_{ji} - \theta_t)]/2$. Here, $H_{ijt}^{(1)}$ has conditional mean $h_{it}^{(1)} = E(H_{ijt}^{(1)}| D_{it}^*, Y_i)$ such that $E(h_{it}^{(1)}) = 0$ and $\sigma_{1t}^2 = E(h_{it}^{(1)2})$, $U_{nt}^{(2)} = \sum_{i \neq j}(D_{jt}^{*c}(\phi_{ij} - \theta_t) - \eta_t)/[n(n-1)]$, $\eta_t = E(D_{jt}^{*c}(\phi_{ij} - \theta_t))$, and $V_{nt} = \sum_{i \neq j}(\widehat{S}_C(t) - D_{it}^{*c})D_{jt}^{*c}/[n(n-1)]$. It was established by Chang (1991) that $(-\ln \widehat{S}_C(t) - \Lambda_C(t))$ can be approximated by a U-statistic $U_{nt}^{(0)}$ with symmetric kernel function $H_{ijt}^{(0)}$ and conditional mean $h_{it}^{(0)} = E(H_{ijt}^{(0)}|X_i, \delta_i)$, such that $E(h_{it}^{(0)}) = 0$ and $\sigma_{0t}^2 = E(h_{it}^{(0)2})$. In Lemma 2.2 (Appendix B in the Supplementary Document), the U-statistic representation of $(\widehat{S}_C(t) - S_C(t))$ is shown to be $U_{nt}^{(km)} = \sum_{i \neq j} H_{ijt}^{(km)}/[n(n-1)]$ with $H_{ijt}^{(km)} = -S_C(t)(H_{ijt}^{(0)} - 2h_{it}^{(0)}h_{jt}^{(0)})$. Thus, $(\widehat{\theta}_t - \theta_t)$ in (2.4) can be further expressed as

$$(\widehat{\theta}_t - \theta_t) = \frac{U_{nt}^{(s)} + \eta_t r_{nt}^{(km)} + (U_{nt}^{(km)} + r_{nt}^{(km)})U_{nt}^{(2)}}{V_{nt}}, \tag{2.5}$$

where $U_{nt}^{(s)}$ is a U-statistic with a kernel $H_{ijt}^{(s)} = H_{ijt}^{(1)} + \eta_t H_{ijt}^{(km)}$ that has conditional mean $h_{it}^{(s)} = E(H_{ijt}^{(s)}|X_i, \delta_i, Y_i)$ and variance $\sigma_t^{(s)2}$, and $r_{nt}^{(km)} = (\widehat{S}_C(t) - S_C(t)) - U_{nt}^{(km)}$.

Let $\widehat{\sigma}_{nt}^2$ be the sample variance of $p_i = n\widehat{U}_{nt}^{(s)} - (n-1)\widehat{U}_{nt(-i)}^{(s)}, i = 1, \cdots, n$. Here, $\widehat{U}_{nt}^{(s)}$ is defined as

$$\widehat{U}_{nt}^{(s)} = \frac{\sum_{i \neq j}(\widehat{S}_C(t) - D_{it}^{*c})D_{jt}^{*c}(\phi_{ij} - \widehat{\theta}_t)}{n(n-1)}.$$

$\widehat{U}_{nt(-i)}^{(s)}$ is the Jackknife estimator computed as $\widehat{U}_{nt}^{(s)}$ with the measurements of the ith subject being deleted. In the construction of confidence region for θ_t, the asymptotic variance $\sigma_t^2 = 4\sigma_t^{(s)2}/\nu_t^2$ of $\widehat{\theta}_t$ is proposed to be estimated by a consistent estimator $\widehat{\sigma}_{nt}^2/V_{nt}^2$. The studentized estimator $\widehat{\theta}_t^{(s)} = n^{1/2}V_{nt}(\widehat{\theta}_t - \theta_t)/\widehat{\sigma}_{nt}$ of $\widehat{\theta}_t$ is then derived as

$$\widehat{\theta}_t^{(s)} = \frac{n^{1/2}[U_{nt}^{(s)} + \eta_t r_{nt}^{(km)} + (U_{nt}^{(km)} + r_{nt}^{(km)})U_{nt}^{(2)}]}{\widehat{\sigma}_{nt}}. \tag{2.6}$$

The one-term Edgeworth expansion of the distribution $F_n^{(s)}(x) = P(\widehat{\theta}_t^{(s)} \le x)$ in the following theorem is established to be

$$\begin{aligned}\widehat{F}_n^{(s)}(x) = \Phi(x) + \frac{n^{-1/2}\phi(x)}{\sigma_t^{(s)}}\Big[&\frac{(2x^2+1)E(h_{1t}^{(s)3}) + 3(x^2+1)E(h_{1t}^{(s)}h_{2t}^{(s)}H_{12t}^{(s)})}{6\sigma_t^{(s)2}}\\ &-\frac{2(x^2+1)S_C(t)E[h_{1t}^{(s)}h_{1t}^{(0)}]E[h_{1t}^{(s)}h_{1t}^{(\eta)}]}{\sigma_t^{(s)2}} - 2S_C(t)E(h_{1t}^{(0)}h_{1t}^{(\eta)})\Big],\end{aligned}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are separately the cumulative distribution function and the density function of a standard normal, and $h_{it}^{(\eta)} = [E(D_{jt}^{*c}(\phi_{ij} - \theta_t) + D_{it}^{*c}(\phi_{ji} - \theta_t)|D_{it}^*, Y_i)]/2 - \eta_t$.

Theorem 2.2. *Suppose that assumptions* (A1)$-$(A2) *are satisfied. Then*

$$\sup_x |F_n^{(s)}(x) - \widehat{F}_n^{(s)}(x)| = o(n^{-1/2}) \text{ as } n \to \infty. \tag{2.7}$$

Proof. See the proof in the Web Appendix B.

Based on the one-term Edgeworth expansion for $F_n^{(s)}(x)$ and a $n^{1/2}$-consistent estimator of $\widehat{F}_n^{(s)}(x)$, an alternative confidence interval for θ_t can be constructed. The estimator for $\widehat{F}_n^{(s)}(x)$ mainly substitutes the unknown moments and parameters by the corresponding sample moments and consistent estimators. An approximate $(1-\alpha)$ confidence interval for θ_t is then constructed via

$$(\widehat{\theta}_t - \frac{n^{-1/2}\widehat{\sigma}_{nt}}{V_{nt}}u_{\alpha/2}^-, \widehat{\theta}_t + \frac{n^{-1/2}\widehat{\sigma}_{nt}}{V_{nt}}u_{\alpha/2}^+), \tag{2.8}$$

where $u_{\alpha/2}^{\pm}$ is computed using the inverting formula of Cornish and Fisher (1937) as

$$\begin{aligned}z_{\alpha/2} \pm \frac{n^{-1/2}V_{nt}}{\widehat{\sigma}_{nt}}[&\frac{V_{nt}^2(2z_{\alpha/2}^2+1)\widehat{E}(h_{1t}^{(s)3}) + 3V_{nt}^2(z_{\alpha/2}^2+1)\widehat{E}(h_{1t}^{(s)}h_{2t}^{(s)}H_{12t}^{(s)})}{6\widehat{\sigma}_{nt}^2}\\ &-\frac{2(z_{\alpha/2}^2+1)\widehat{S}_C(t)\widehat{E}[h_{1t}^{(s)}h_{1t}^{(0)}]\widehat{E}[h_{1t}^{(s)}h_{1t}^{(\eta)}]}{\widehat{\sigma}_{nt}^2} - 2\widehat{S}_C(t)\widehat{E}(h_{1t}^{(0)}h_{1t}^{(\eta)})],\end{aligned}$$

where $z_p = \Phi^{-1}(1-p)$ and $\widehat{E}(\cdot)$ is the sample moment of $E(\cdot)$. The normal approximate confidence interval for θ_t is established as (2.8) with $u^-_{\alpha/2}$ and $u^-_{\alpha/2}$ being replaced with $z_{\alpha/2}$ and $z_{1-\alpha/2}$, respectively.

3. Simulation Study

The performance of the proposed procedures for the construction of confidence region is investigated through a class of simulations. In our setting, Y was generated from a standard normal distribution. Conditioning on $Y = y$, the failure time T was specified from a survival distribution $S_T(t|y) = \exp(-0.02t\exp(0.5y))$. The time-dependent AUC was computed by using the formula of Chambless and Diao (2006) as

$$\theta_t = \frac{E((1-S(t|Y_i))S(t|Y_j)I(Y_j < Y_i))}{E(1-S(t|Y_j))E(S(t|Y_j))}, i \neq j. \tag{3.1}$$

Independent of (Y, T), the censoring time C was set to follow an exponential distribution with different scale parameters which lead to different censoring rates.

Based on the above design, the censored data $\{(X_i, \delta_i, Y_i)\}_{i=1}^n$ were repeatedly generated 1,000 times with the sample sizes of 250 and 500, and the censoring rates of 30%, and 50%. The constructed confidence regions based on the normal approximation (NA), random weighted bootstrap (RWB), and Edgeworth expansion (EE) methods were evaluated at selected time points from 5 to 145. In our simulation process, the naive bootstrap (NB) method for the construction of confidence region was also provided. For the random weighted bootstrap confidence regions, the exchangeable random weights were investigated based on the independent and identically distributed Uniform (0,1) and $Gamma(4, 1)$ random variables. The corresponding random weighting methods are denoted here by RWBU and RWBG.

As can be seen from Table 3.1, the NA method produces a confidence interval in which the lower bound is generally larger than the 0.25 quantile of 1,000 estimates $\widehat{\theta}_t$, especially for the small sample size. Although the lower confidence limit of EE was closer to the 0.25 qunatile of estimates, the constructed upper limit was relatively lower than the 0.975 quantile of the estimates. This can be explained by the fact that the length of confidence interval computed based on the one-term EE is same as that of the NA confidence interval. Table 3.1 further reveals that the bootstrap confidence intervals were very close to the 0.95 quantile intervals of 1,000 estimates at the selected time points. Table 3.2 summarizes the empirical coverage probabilities of the 0.95 confidence intervals of the NA, EE, NB, RWBU, and RWBG procedures. The probabilities are generally around

Table 3.1. The 0.95 quantile intervals (Q.I.) of estimates, the averages of 0.95 NA, EE, NB, RWB1, and RWB2 confidence intervals with two sample sizes (n) and two censoring rates ($c.r.$).

method	Q.I.	NA	EE	NB	RWB1	RWB2
time			$n = 250$	$c.r. = 30\%$		
5	(0.512,0.778)	(0.510,0.780)	(0.504,0.774)	(0.503,0.783)	(0.506,0.781)	(0.508,0.776)
15	(0.566,0.751)	(0.565,0.752)	(0.562,0.749)	(0.563,0.751)	(0.564,0.751)	(0.564,0.749)
25	(0.588,0.753)	(0.585,0.757)	(0.582,0.754)	(0.584,0.756)	(0.584,0.756)	(0.584,0.754)
65	(0.611,0.812)	(0.622,0.814)	(0.615,0.807)	(0.620,0.812)	(0.619,0.811)	(0.619,0.809)
125	(0.613,0.913)	(0.637,0.910)	(0.612,0.886)	(0.623,0.910)	(0.626,0.908)	(0.632,0.897)
135	(0.613,0.936)	(0.643,0.926)	(0.614,0.897)	(0.623,0.926)	(0.629,0.924)	(0.638,0.911)
145	(0.599,0.942)	(0.646,0.940)	(0.615,0.908)	(0.624,0.939)	(0.628,0.938)	(0.641,0.923)
			$n = 250$	$c.r. = 50\%$		
5	(0.486,0.814)	(0.476,0.807)	(0.472,0.803)	(0.468,0.809)	(0.471,0.806)	(0.473,0.803)
15	(0.536,0.777)	(0.537,0.773)	(0.535,0.771)	(0.537,0.775)	(0.536,0.771)	(0.536,0.769)
25	(0.551,0.781)	(0.555,0.781)	(0.553,0.779)	(0.557,0.783)	(0.554,0.779)	(0.555,0.777)
65	(0.548,0.859)	(0.561,0.862)	(0.549,0.850)	(0.562,0.863)	(0.556,0.857)	(0.559,0.851)
125	(0.347,0.980)	(0.594,0.929)	(0.570,0.905)	(0.602,0.915)	(0.574,0.916)	(0.601,0.898)
135	(0.325,0.986)	(0.614,0.908)	(0.595,0.889)	(0.628,0.901)	(0.599,0.896)	(0.623,0.881)
145	(0.325,0.992)	(0.641,0.883)	(0.626,0.869)	(0.659,0.885)	(0.630,0.871)	(0.649,0.860)
			$n = 500$	$c.r. = 30\%$		
5	(0.552,0.746)	(0.551,0.740)	(0.548,0.737)	(0.548,0.739)	(0.549,0.739)	(0.549,0.737)
15	(0.593,0.720)	(0.592,0.723)	(0.590,0.721)	(0.591,0.722)	(0.591,0.722)	(0.590,0.721)
25	(0.613,0.728)	(0.611,0.732)	(0.610,0.730)	(0.610,0.731)	(0.610,0.730)	(0.610,0.730)
65	(0.649,0.787)	(0.651,0.787)	(0.648,0.784)	(0.650,0.786)	(0.650,0.785)	(0.649,0.785)
125	(0.669,0.873)	(0.680,0.878)	(0.665,0.863)	(0.673,0.875)	(0.675,0.876)	(0.675,0.871)
135	(0.670,0.891)	(0.682,0.892)	(0.665,0.875)	(0.674,0.890)	(0.677,0.890)	(0.677,0.884)
145	(0.664,0.908)	(0.686,0.908)	(0.665,0.886)	(0.675,0.905)	(0.678,0.906)	(0.680,0.898)
			$n = 500$	$c.r. = 50\%$		
5	(0.526,0.761)	(0.531,0.762)	(0.529,0.760)	(0.529,0.765)	(0.531,0.762)	(0.531,0.761)
15	(0.573,0.737)	(0.574,0.740)	(0.573,0.739)	(0.575,0.742)	(0.574,0.740)	(0.574,0.739)
25	(0.593,0.748)	(0.591,0.750)	(0.590,0.749)	(0.592,0.751)	(0.591,0.749)	(0.591,0.749)
65	(0.609,0.823)	(0.612,0.824)	(0.605,0.818)	(0.613,0.826)	(0.610,0.822)	(0.610,0.819)
125	(0.506,0.957)	(0.595,0.942)	(0.563,0.910)	(0.580,0.930)	(0.573,0.936)	(0.598,0.917)
135	(0.451,0.978)	(0.620,0.933)	(0.596,0.909)	(0.613,0.922)	(0.600,0.927)	(0.625,0.909)
145	(0.447,0.987)	(0.643,0.925)	(0.624,0.906)	(0.644,0.915)	(0.625,0.919)	(0.650,0.902)

the nominal level of 0.95 when the sample size is moderate and the censoring rate is low. It is indicated from this table that EE, NB, and RWBU intervals had better coverage than NA in regions of sparse failure times, especially when the sample size was small and the censoring rate was high. For the random weighted bootstrap intervals, an appropriate specification of random weights becomes a key factor in the improvement of coverage probability for the NA interval. The computation times for constructing the EE, RW, and NB confidence intervals in each simulated data were 0.89, 7.22, and 55.29 seconds with the sample size of

Table 3.2. The empirical coverage probabilities of 0.95 NA, EE, NB, RWB1, and RWB2 confidence intervals with two sample sizes (n) and two censoring rates ($c.r.$).

method		NA		EE		NB		RWB1		RWB2	
n		250	500	250	500	250	500	250	500	250	500
$c.r.$	time										
30%	5	0.943	0.933	0.947	0.938	0.945	0.934	0.943	0.927	0.938	0.935
	15	0.953	0.956	0.955	0.957	0.956	0.949	0.950	0.949	0.949	0.953
	25	0.956	0.961	0.959	0.963	0.956	0.954	0.957	0.961	0.951	0.954
	65	0.934	0.941	0.939	0.946	0.937	0.942	0.934	0.946	0.931	0.939
	125	0.896	0.920	0.904	0.937	0.909	0.929	0.903	0.920	0.894	0.920
	135	0.866	0.916	0.882	0.930	0.886	0.924	0.882	0.922	0.867	0.919
	145	0.837	0.906	0.841	0.929	0.862	0.928	0.859	0.918	0.838	0.916
50%	5	0.953	0.940	0.948	0.941	0.948	0.933	0.943	0.933	0.944	0.934
	15	0.950	0.953	0.947	0.952	0.943	0.946	0.941	0.944	0.940	0.943
	25	0.946	0.956	0.943	0.955	0.941	0.951	0.940	0.953	0.938	0.958
	65	0.942	0.942	0.946	0.944	0.942	0.940	0.938	0.931	0.944	0.939
	125	0.631	0.788	0.655	0.802	0.654	0.818	0.654	0.818	0.621	0.783
	135	0.545	0.675	0.560	0.689	0.565	0.702	0.567	0.696	0.533	0.666
	145	0.445	0.591	0.456	0.594	0.450	0.621	0.452	0.625	0.427	0.580

250, and 3.21, 26.81, and 136.40 seconds with the sample size of 500 on a Pentium IV with 3.4 GHz of CPU and 1.0GB of RAM. Note that the random weighted bootstrap procedures were more computationally efficient than the naive bootstrap procedure.

4. Application to ACTG 175 Data

The data analyzed are from the ACTG study 175. A total of 2467 HIV-1-infected patients, whose CD4 cell counts range from 200 to 500 cells cu/mm during December 1991 and October 1992, was recruited. Patients with lower CD4 counts usually have a higher risk of AIDS or death. In the data analysis, let Y be a strictly decreasing transformation of CD4 counts, so that θ_t can reasonably be expected to fall within the range of 0.5 to 1. Details of the analyzed data can be found in Hammer et al. (1996).

We can see from Table 4.1 that the estimated AUCs increased from week 60 and stayed around 0.8, which indicates a high accuracy of classification, after week 70. The approximate 0.95 confidence intervals based on the NA, EE, NB, RWBU, and RWBG methods are also provided at the selected weeks. It is further indicated from this table that the RWBG confidence intervals have shorter interval lengths than other confidence intervals though expected length might not be a good criterion for the accuracy of confidence intervals. The EE confidence intervals are very similar to the NB and RWBU ones. For the sake of

Table 4.1. The 0.95 NA, EE, NB, RWBU, and RWBG confidence intervals at the selected weeks.

week	$\widehat{\theta}_t$	NA	EE	NB	RWBU	RWBG
70	0.829	(0.648,1.000)	(0.658,1.000)	(0.640,1.000)	(0.634,1.000)	(0.656,1.000)
80	0.861	(0.699,1.000)	(0.710,1.000)	(0.692,1.000)	(0.699,1.000)	(0.700,1.000)
90	0.863	(0.716,1.000)	(0.726,1.000)	(0.712,1.000)	(0.710,1.000)	(0.718,1.000)
100	0.802	(0.682,0.921)	(0.689,0.927)	(0.692,0.916)	(0.679,0.911)	(0.681,0.908)
110	0.796	(0.691,0.900)	(0.697,0.906)	(0.705,0.908)	(0.701,0.919)	(0.715,0.915)
120	0.797	(0.702,0.891)	(0.707,0.897)	(0.707,0.882)	(0.691,0.900)	(0.704,0.881)
130	0.798	(0.712,0.883)	(0.717,0.888)	(0.712,0.875)	(0.709,0.873)	(0.710,0.857)
140	0.800	(0.717,0.882)	(0.722,0.882)	(0.715,0.876)	(0.714,0.874)	(0.716,0.868)
150	0.781	(0.706,0.856)	(0.710,0.861)	(0.705,0.851)	(0.692,0.852)	(0.702,0.847)
160	0.810	(0.718,0.901)	(0.726,0.909)	(0.731,0.886)	(0.722,0.901)	(0.734,0.892)

computational cost, the EE and RWBU methods are recommended for practical implementation.

5. Concluding Remarks

Based on survival data, an approximate confidence interval for θ_t can be constructed via the asymptotic normality of $\widehat{\theta}_t$ and a consistent estimator of the asymptotic variance of $\widehat{\theta}_t$. Our numerical studies suggest that the empirical coverage probabilities are not close to the nominal level, and the constructed confidence intervals are not accurate enough for data with small sample size and high censoring rate. We establish alternative confidence regions for θ_t based on the random weighted bootstrap distribution and the one-term Edgeworth expansion.

In our simulations, we find that the adequacy of random weighted bootstrap confidence intervals relies on the specification of random weights. This was investigated by James (1997) for the Kaplan-Meier and cumulative hazard estimators, and by Chiang, James and Wang (2005) for the occurrence rate function with censored data using random weights $(w_1, \cdots, w_n)$ having a Dirichlet distribution with parameters $(4, \cdots, 4)$ for better approximations. However, the weights generated from uniform ξ_i's did relatively well in our numerical studies. Currently, there is no standard criterion for selecting random weights in the data setting. We found that the one-term Edgeworth expansion procedure improved the coverage probability but not the accuracy of interval estimators. A higher-order Edgeworth expansion should be used in the construction of more accurate confidence intervals.

Acknowledgements

The first author's research was partially supported by the National Science Council grant 97-2118-M-002-020-MY2 (Taiwan). We would like to thank the referee for valuable comments. We also grateful to Professor Xiao Song for sharing the ACTG 175 data.

References

Bickel, P. J., Götze, F. and Van Zwet, W. R. (1986). The Edgeworth expansion for U-statistics of degree 2. *Ann. STatist.* **19**, 470-484.

Callaert, H., Janssen, P. and Veraverbeke, N. (1980). An Edgeworth expansion for U-statistics. *Ann. STatist.* **8**, 299-312.

Callaert, H. and Veraverbeke, N. (1981). The order of the normal approximation for a studentized U-statistic. *Ann. STatist.* **9**, 194-200.

Chambless, L. E. and Diao, G. (2006). Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statist. Medicine* **25**, 3474-3486.

Chang, M. N. and Rao, P. V. (1989). Berry-Esseen bound for the Kaplan-Meier estimator. *Comm. Statist. Theory Methods* **18**, 4647-4664.

Chang, M. N. (1991). Edgeworth expansion for the Kaplan-Meier estimator. *Comm. Statist. Theory Methods* **20**, 2479-2494.

Chiang, C. T., James, L. F. and Wang, M. C. (2005). Random weighted bootstrap method for recurrent events with informative censoring. *Lifetime Data Analysis* **11**, 489-509.

Cornish, E. A. and Fisher, R. A. (1937). Moments and cumulants in the specification of distributions. *Review of the International Statistical Institute* **5**, 307-320.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Process and Survival Analysis.* Wiley, New York.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S. and Merigan, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *The New England Journal of Medicine* **335**, 1081-1090.

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

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293-325.

Hoeffding, W. (1961). The strong law of large numbers for U-statistics. University of North Carolina Institute of Statistics Mimeo Series. No. 302.

James, L. F. (1997). A study of a class of weighted bootstraps for censored data. *Ann. STatist.* **25**, 1595-1621.

Janssen, P. (1994). Weighted bootstrapping of U-statistics. *J. Statist. Plann. Inference* **38**, 31-42.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

Malevich, T. L. and Abdalimov, B. (1979). Large deviation probabilities for U-statistics. *Theory Probab. Appl.* **24**, 215-219.

Pollard, D. (1990). *Empirical Processes: Theory and Applications. Institute of Mathematical Statistics*, Hayward, California.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

Tu, D. and Zheng, Z. (1991). Random weighting: another approach to approximate the unknown distributions of pivotal quantities. *J. Combinatorics, Information and System Sciences* **16**, 249-270.

Department of Mathematics, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan.

E-mail: chiangct@ntu.edu.tw

Department of Mathematics, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan.

E-mail: warmsay@yahoo.com.tw

Department of Mathematics, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan.

E-mail: double.red@gmail.com

(Received March 2008; accepted October 2008)