Sheppard's correction for grouping in Cox's proportional hazards model

Ian W. McKeague
Florida State University

Mei-Jie Zhang*
Medical College of Wisconsin

Abstract

Cox's proportional hazards model is often fit to *grouped* survival data, i.e. occurrence/exposure data over given time intervals and covariate strata. We derive a Sheppard correction for the bias in the grouped data analogue of Cox's maximum partial likelihood estimator. This is done via a large sample theory in which the covariate strata and time intervals shrink as the sample size increases.

1 Introduction

In many estimation settings, data are grouped prior to their statistical analysis. Grouping may be unavoidable, as with round-off error, or done intentionally, e.g. to economize on data transmission and storage, to reduce computation, to protect the privacy of individual records, or to account for the limitations of a measurement instrument. Moreover, some large data sets are publically released only in grouped form, as discussed by Haitovsky (1973, 1983).

It is important to be able to assess estimation bias caused by grouping and to correct it if necessary. The bias can be severe, irrespective of sample size; e.g., for a parametric model and grouping intervals with equi-spaced limits, the bias of the approximate maximum likelihood estimator (in which observations are approximated by interval midpoints) is of order $O(w^2)$, where w is the interval width. A 'Sheppard correction' can be used to reduce the bias to order $O(w^3)$, see Lindley (1950). Sheppard corrections have been obtained in some other contexts by Haitovsky (1973), Don (1981), Dempster and Rubin (1983), and Kolassa and McCullagh (1990).

The purpose of the present paper is to obtain a Sheppard correction for Cox's (1972) proportional hazards model. This popular model specifies the conditional hazard function

*Research partially supported by National Cancer Institute Grant 1 R01 CA54706-03.

[1] *AMS 1991 subject classifications.* Primary: 62F10; secondary: 62G99

[2] *Key words and Phrases.* Occurrence/exposure data, asymptotic bias, survival analysis, counting processes, martingale methods.

of the survival time of an individual as $\lambda(t|z) = \lambda_0(t)\exp(\beta_0 z)$, where z is a covariate, λ_0 is an unknown baseline hazard function and β_0 is a regression parameter. (For notational simplicity, we assume that the covariate is one-dimensional and non-time dependent). Grouped data in this setting are occurrence/exposure data for cells determined by time intervals and covariate strata, see, e.g., Breslow (1986), Preston et al. (1987) and Selmer (1990).

Our main result, stated in Section 2, shows how grouping disturbs the asymptotic behavior of the maximum partial likelihood estimator of β_0. An estimator of the Sheppard correction is provided in Section 3, and its performance is assessed through a simulation study in Section 4. The proof of the main result is given in Section 5.

2 Correction for grouping

Let (X, C, Z) be random variables such that the survival time X and the censoring time C are conditionally independent given the covariate Z. Denote $\delta = 1_{\{X \le C\}}$ and $T = X \wedge C$. The ungrouped data consist of n independent replicates (T_i, δ_i, Z_i) of (T, δ, Z). Cox's maximum partial likelihood estimator $\hat\beta$ is obtained by maximizing

$$L(\beta) = \prod_{i=1}^{n} \left\{ \frac{e^{\beta Z_i}}{\sum_{k \in \mathcal{R}_i} e^{\beta Z_k}} \right\}^{\delta_i}$$

where $\mathcal{R}_i$ is the set of individuals observed to be at risk at time T_i-. Under suitable regularity conditions (see Andersen and Gill, 1982), $\sqrt{n}(\hat\beta - \beta_0) \xrightarrow{\mathcal{D}} N(0, V)$, where V^{-1} is consistently estimated by $-n^{-1}\partial U(\hat\beta)/\partial\beta$ and U is the partial likelihood score function $U(\beta) = \partial \log L(\beta)/\partial\beta$.

The grouped data based estimator $\hat\beta_g$ is obtained by maximizing the following approximation to the partial likelihood:

$$L_g(\beta) = \prod_{r,j} \left\{ \frac{e^{\beta z_j}}{\sum_k Y_{rk} e^{\beta z_k}} \right\}^{N_{rj}}$$

where the product is over the grouping cells, the sum is over the covariate strata, and z_j is the midpoint of the jth covariate stratum. Here Y_{rj} and N_{rj} are, respectively, the total time at risk (exposure) and the number of observed failures (occurrence) in the rjth grouping cell $\mathcal{C}_{rj} = \mathcal{T}_r \times \mathcal{I}_j$. We assume that the time intervals $\mathcal{T}_r$ are of equal length $l = l_n \to 0$, and the covariate strata $\mathcal{I}_j$ are of equal length $w = w_n \to 0$. This estimator has been studied by Kalbfleisch and Prentice (1973), Holford (1976), Prentice and Gloeckler (1978), Breslow (1986), Hoem (1987), Selmer (1990), and Huet and Kaddour (1994). It can be interpreted as the maximum likelihood estimator in a Poisson regression model, see Laird and Olivier (1981).

We obtain the following Sheppard correction for $\hat\beta_g$:

$$\Delta = -\tfrac{V}{12} \iint e^{\beta_0 z}\{z - \bar z(\beta_0, t)\}\{l^2 \dot\lambda_0(t)\dot F'(t,z) + w^2 \beta_0 \lambda_0(t) F''(t,z)\}\, dt dz,$$

where the double integral is over the region covered by the cells used in grouping the data, $\bar{z}(\beta,t) = s^{(1)}(\beta,t)/s^{(0)}(\beta,t)$, $Y(t) = 1_{\{T\geq t\}}$ and $F(t,z) = P(T \geq t, Z \leq z)$. Here $s^{(k)}(\beta,t) = E\{Y(t)Z^k e^{\beta Z}\}$, and $\dot{F}, F'$ denote the partial derivatives of F with respect to t and z, respectively. The various derivatives implicit in Δ are assumed to exist and to be continuous. Two mild conditions, (C1) and (C2) in Section 5, are also assumed to hold. A somewhat simpler expression for Δ is given in Section 3.

The Sheppard correction Δ is justified by the following result.

Theorem 2.1 $\hat{\beta} = \hat{\beta}_g + \Delta + O_P\{l^3 + w^3 + (l + w + c_n)n^{-1/2} + c_n^2 n^{-1}\}$ *where* $c_n = \frac{w^2}{\sqrt{l}} + \frac{1}{\sqrt{nl}}$.

This result indicates that the Sheppard correction is potentially useful when the 'higher order' terms above are of order $o_P(l^2 + w^2)$, e.g., if $l \asymp w \asymp n^{-\alpha}$, where $0 < \alpha < 2/5$. In applications this means that the number of grouping intervals should be small relative to the sample size. To implement the Sheppard correction, a grouped data based estimator $\hat{\Delta}$ of Δ is required, see Section 3. It would be worthwhile using the corrected estimator $\hat{\beta}_c = \hat{\beta}_g + \hat{\Delta}$ in place of $\hat{\beta}_g$ if the *relative* Sheppard correction $\sqrt{n}\Delta/\sqrt{V}$ is appreciable. We recommend correcting $\hat{\beta}_g$ if the estimated relative Sheppard correction exceeds .2 in absolute value, as is the case with the examples considered in Section 4.

Our result also allows the usual asymptotic normality of $\hat{\beta}$ to be extended to $\hat{\beta}_c$; we only need that the higher order terms above are $o_P(n^{-1/2})$ (e.g., $l \asymp w \asymp n^{-\alpha}$, where $1/6 < \alpha < 1$), and that $\hat{\Delta} - \Delta = o_P(n^{-1/2})$, which is a mild condition, see Section 3. Then

$$\sqrt{n}(\hat{\beta}_c - \beta_0) \xrightarrow{\mathcal{D}} N(0, V).$$

This allows the construction of confidence intervals (say) for β_0 centered on $\hat{\beta}_c$, provided there are sufficiently many grouping intervals relative to the sample size.

The optimal rate of convergence for the uncorrected $\hat{\beta}_g$ is attained when the squared bias and variance are of the same order, or if $l \sim w \sim n^{-1/4}$, in which case $\sqrt{n}(\hat{\beta}_g - \beta_0) \xrightarrow{\mathcal{D}} N(\mu, V)$, where the asymptotic bias $\mu = -\Delta$ with $l = w = 1$ in Δ.

3 Estimation of Δ

We first express Δ in a form more suitable for estimation. The follow-up period is taken to be $[0,1]$; any individual that survives beyond the end of follow-up is censored. Also assume that the covariate Z takes values in $[0,1]$. Some elementary calculus shows that

$$\Delta = -\tfrac{V}{12}\left(\tfrac{1}{2}l^2\Delta_1 + w^2\beta_0\Delta_2\right),$$

where

$$\Delta_1 = \int_0^1 \{\bar{z}(\beta_0,t) - \bar{z}(2\beta_0,t)\}s^{(0)}(2\beta_0,t)\lambda_0^2(dt), \qquad \Delta_2 = \psi(1) - \psi(0) - P(\delta = 1)$$

and

$$\psi(z) = \int_0^1 \{z - \bar{z}(\beta_0, t)\} e^{\beta_0 z} \lambda_0(t) F'(t, z)\, dt.$$

It follows from the expression for Δ_1 that if there is only minor variation in the baseline hazard λ_0 over the follow-up period, then a correction for grouping in the time domain would not be necessary. Use Holford's (1976) grouped data based estimator of λ_0:

$$\hat{\lambda}_0(t) = \frac{\sum_j N_{rj}}{\sum_j Y_{rj} e^{\hat{\beta}_g z_j}} \quad \text{for } t \in \mathcal{T}_r.$$

We recommend inspection of a plot of $\hat{\lambda}_0$ to assess the variation in λ_0 over the follow-up period.

A grouped data based estimator of $s^{(k)}(\beta, t)$ is given by $S_g^{(k)}(\beta, t) = n^{-1} \sum_j z_j^k Y_{rj} e^{\beta z_j}$ at $t \in \mathcal{T}_r$, see Lemma 5.1(ii). We may estimate $F'(t, z)$, at $(t, z) \in \mathcal{C}_{rj}$, by $Y_{rj}/(nwl)$. These estimators can be plugged into Δ_1 and ψ, replacing each integral by a sum of terms, where for Δ_1 the terms involve the increment in $\hat{\lambda}_0^2$ from one time interval $\mathcal{T}_r$ to the next. The last term in Δ_2 is consistently estimated by $\int_0^1 S_g^{(0)}(\hat{\beta}_g, t)\hat{\lambda}_0(t)\, dt$. A consistent grouped data based estimator of V^{-1} is given by $\hat{V}_g^{-1} = -n^{-1} \partial U_g(\hat{\beta}_g)/\partial \beta$, where U_g is the grouped data version of U, see Lemma 5.5(ii).

This leads to consistent estimators of Δ_1 and Δ_2, and consequently to an estimator $\hat{\Delta}$ such that $\hat{\Delta} - \Delta = o_P(n^{-1/2})$, assuming that $l \asymp w \asymp n^{-\alpha}$, where $1/4 \le \alpha < 1$.

4 Numerical results

We have carried out a Monte Carlo experiment to evaluate the performance of the estimated Sheppard correction $\hat{\Delta}$. This is a limited simulation study, but it suggests that the Sheppard correction is adequate to compensate for the grouping bias that would occur in typical applications.

We used $\beta_0 = 3$ and a linear baseline hazard function $\lambda_0(t) = bt$, with $b = 1, 3$. The covariate was uniformly distributed on $[0, 1]$. The censoring time was independent of both the survival time and the covariate, and exponentially distributed with parameter values 0.35 and 1.25, for $b = 1, 3$ respectively. The follow-up intervals were taken as $[0, 1]$ and $[0, .6]$, respectively. In each case, this gave a censoring rate of about 30%, including about 12% that were still at risk at the end of follow-up. We used equal numbers of time periods and covariate strata. There were 1000 samples in each simulation run.

In Table 1 we report the mean Sheppard correction and the (normalized) mean difference between $\hat{\beta}$ and $\hat{\beta}_g$. The normalization used here was the 'standard error' $\sigma/\sqrt{n}$, where $\sigma^2 = E\hat{V}_g$. We also report observed levels of the Wald tests of the null hypothesis that $\beta_0 = 3$, based on $\hat{\beta}_g$, $\hat{\beta}_c$, and $\hat{\beta}$, against the two-sided alternative.

The Sheppard correction has removed most of the grouping bias (compare the fifth and sixth columns of Table 1). Moreover, it has restored the levels of the hypothesis tests to be much closer to the level of the analogous continuous data tests (compare the last three columns of Table 1). Although the effect of the grouping in this example is modest—less than half a standard error—the Sheppard correction is expected to continue to perform adequately in cases where the bias is more pronounced.

Table 1: Monte Carlo estimates of the mean Sheppard correction and the the (normalized) mean difference between $\hat{\beta}$ and $\hat{\beta}_g$; observed levels of (nominal 5%) Wald tests of $\beta_0 = 3$ based on $\hat{\beta}_g$, $\hat{\beta}_c$, and $\hat{\beta}$ are labeled P_g, P_c and P_0, respectively.

b	n	# strata	$E\hat{\Delta}$	$\sqrt{n}E\hat{\Delta}/\sigma$	$\sqrt{n}E(\hat{\beta}-\hat{\beta}_g)/\sigma$	P_g	P_c	P_0
1	100	3	0.210	0.413	0.484	0.082	0.085	0.048
	500	5	0.097	0.436	0.454	0.056	0.057	0.039
	1000	6	0.069	0.445	0.464	0.086	0.070	0.058
3	100	3	0.203	0.397	0.537	0.085	0.067	0.053
	500	5	0.099	0.441	0.466	0.081	0.061	0.055
	1000	6	0.071	0.450	0.497	0.084	0.060	0.050

5 Proof of Theorem 2.1

The follow-up period is taken to be $[0, 1]$. We need two mild conditions:

(C1) There exists a (compact) neighborhood $\mathcal{B}$ of β_0 such that, for all t and $\beta \in \mathcal{B}$,

$$s^{(1)}(\beta, t) = \frac{\partial}{\partial\beta} s^{(0)}(\beta, t), \quad s^{(2)}(\beta, t) = \frac{\partial^2}{\partial\beta^2} s^{(0)}(\beta, t).$$

(C2) The functions $s^{(k)}$ are Lipschitz, $s^{(0)}$ is bounded away from zero on $\mathcal{B} \times [0, 1]$, and

$$V^{-1} = \int_0^1 v(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t)\, dt$$

is positive, where $v = s^{(2)}/s^{(0)} - (s^{(1)}/s^{(0)})^2$.

As in Andersen and Gill (1982), hereafter AG, we have $\hat{\beta} = \beta_0 + U(\beta_0)/I(\beta^*)$, where $I(\beta) = -\partial U(\beta)/\partial\beta$ and β^* is on the line segment between $\hat{\beta}$ and β_0. Similarly, in the grouped data case we have $\hat{\beta}_g = \beta_0 + U_g(\beta_0)/I_g(\beta_g^*)$ where the β_g^* is on the line segment between $\hat{\beta}_g$ and β_0, and U_g and I_g are the grouped data versions of U and I. Thus, as $U(\beta_0) = O_P(n^{1/2})$ by AG (p.1106), some simple algebra shows that

$$\hat{\beta} = \hat{\beta}_g + VA + O_P\{|A|(C + D) + (B + C + D)n^{-1/2}\}, \tag{5.1}$$

where $A = \{U(\beta_0) - U_g(\beta_0)\}/n$ and

$$B = \left|\frac{1}{I(\beta^*)/n} - V\right|, \quad C = \left|\frac{1}{I(\beta_g^*)/n} - V\right|, \quad D = \left|\frac{1}{I(\beta_g^*)/n} - \frac{1}{I_g(\beta_g^*)/n}\right|.$$

We shall examine the various terms in (5.1) through a series of lemmas.

Adopting the notation of AG, let $S^{(k)}(\beta,t) = n^{-1}\sum_{i=1}^n Z_i^k Y_i(t) e^{\beta Z_i}$ for $k = 0, 1, 2$, where $0^0 = 1$. We denote by $\bar{h}(\beta, t)$ the piecewise constant approximation to a function $h(\beta, t)$ obtained by averaging over each time interval $\mathcal{T}_r$. Note that $S_g^{(k)} = \bar{S}_a^{(k)}$, where

$$S_a^{(k)}(\beta,t) = \frac{1}{n}\sum_{i,j} Y_i(t) 1_{\{Z_i \in \mathcal{I}_j\}} z_j^k e^{\beta z_j}.$$

Lemma 5.1 For $k = 0, 1, 2$, uniformly in t and $\beta \in \mathcal{B}$,
(i) $|S^{(k)}(\beta,t) - s^{(k)}(\beta,t)| = O_P(n^{-1/2})$;
(ii) $|S_g^{(k)}(\beta,t) - s^{(k)}(\beta,t)| = O_P(l + w + n^{-1/2})$;
(iii) $E\sup_\beta\{S_g^{(k)}(\beta,t) - \bar{S}^{(k)}(\beta,t)\}^2 = O(w^4 + wn^{-1})$.

Proof We sketch the proof. (i) can be proved using a functional central limit theorem of Hahn (1978) extended to two-parameter processes via the convergence criteria of Bickel and Wichura (1971). (ii) follows from (i), the Lipschitz condition on $s^{(k)}$, and since $S_a^{(k)} = S^{(k)} + O(w)$ uniformly in t and $\beta \in \mathcal{B}$. Part (iii) is proved by Taylor expanding the function $z \mapsto z^k e^{\beta z}$ about each z_j, so as to express $\bar{S}^{(k)}$ as the sum of $S_g^{(k)}$ and two 'higher order' terms. The means of these terms can be expressed as integrals involving $f = F'$, and, by Taylor expanding f about each z_j, are seen to be of order $O(w^2)$. The variances of the two terms are found to be of order $O(wn^{-1})$. □

Lemma 5.2 $\hat{\beta}_g$ is consistent.

Proof Define $X(\beta) = n^{-1}\log\{L(\beta)/L(\beta_0)\}$, and define $X_g(\beta)$ similarly in the grouped data case. In terms of the counting processes $N_i(t) = 1_{\{T_i \le t, \delta_i = 1\}}$ and $\bar{N} = \sum_{i=1}^n N_i$ we have

$$\begin{aligned} |X(\beta) - X_g(\beta)| \;\le\; & \frac{1}{n}\left|\sum_{r,j,i}\int_{\mathcal{T}_r}(\beta-\beta_0)(Z_i - z_j)1_{\{Z_i\in\mathcal{I}_j\}}dN_i(u)\right| \\ & + \frac{1}{n}\int_0^1\left|\log\left(\frac{S^{(0)}(\beta,u)}{S^{(0)}(\beta_0,u)}\right) - \log\left(\frac{S_g^{(0)}(\beta,u)}{S_g^{(0)}(\beta_0,u)}\right)\right| d\bar{N}(u). \end{aligned}$$

The first term on the r.h.s. is bounded above by

$$|\beta-\beta_0|\sup_{i,j}|(Z_i - z_j)1_{\{Z_i\in\mathcal{I}_j\}}| \cdot \frac{1}{n}\bar{N}(1) \xrightarrow{P} 0,$$

since the width of $\mathcal{I}_j$ is $w_n \to 0$. If $\beta \in \mathcal{B}$ the second term tends in probability to zero by continuity of log, Lemma 5.1 (i) and (ii), and the assumption that $s^{(0)}$ is bounded away from zero on $\mathcal{B}\times[0,1]$. Thus $|X(\beta) - X_g(\beta)| \xrightarrow{P} 0$ if $\beta \in \mathcal{B}$. The result now follows using the argument of AG (Lemma 3.1). □

Lemma 5.3 $A = \{U(\beta_0) - U_g(\beta_0)\}/n = \Delta V^{-1} + O_P\{l^3 + w^3 + (l + w + c_n)n^{-1/2}\}$.

Proof In terms of the martingales $M_i(t) = N_i(t) - \int_0^t Y_i(u)\lambda_0(u)e^{\beta_0 Z_i}\,du$ and $\bar{M} = \sum_{i=1}^n M_i$ we write A as

$$\frac{1}{n}\sum_{i,j}\int_0^1 (Z_i - z_j)\,1_{\{Z_i \in \mathcal{I}_j\}} dM_i(u) \tag{5.2}$$

$$+\frac{1}{n}\int_0^1 \left\{\frac{S_g^{(1)}(\beta_0, u)}{S_g^{(0)}(\beta_0, u)} - \frac{S^{(1)}(\beta_0, u)}{S^{(0)}(\beta_0, u)}\right\} d\bar{M}(u) \tag{5.3}$$

$$-\frac{1}{n}\sum_{r,i,j}\int_{\mathcal{T}_r} z_j e^{\beta_0 Z_i} 1_{\{Z_i \in \mathcal{I}_j\}} Y_i(u)\lambda_0(u)\,du \tag{5.4}$$

$$+\frac{1}{n}\sum_r \frac{S_g^{(1)}(\beta_0, t_r)}{S_g^{(0)}(\beta_0, t_r)}\sum_{i,j}\int_{\mathcal{T}_r} e^{\beta_0 Z_i} 1_{\{Z_i \in \mathcal{I}_j\}} Y_i(u)\lambda_0(u)\,du, \tag{5.5}$$

where t_r is the midpoint of $\mathcal{T}_r$. Standard martingale theory gives that (5.2) is of order $O_P(wn^{-1/2})$. Next, $|(5.3)|$ is bounded by

$$\frac{1}{n}\left|\int_0^1 \left\{\frac{S^{(1)}(\beta_0, u)}{S^{(0)}(\beta_0, u)} - \frac{\bar{s}^{(1)}(\beta_0, u)}{\bar{s}^{(0)}(\beta_0, u)}\right\} d\bar{M}(u)\right| \tag{5.6}$$

$$+\frac{1}{n}\sum_r \left|\frac{\bar{s}^{(1)}(\beta_0, t_r)}{\bar{s}^{(0)}(\beta_0, t_r)} - \frac{\bar{S}^{(1)}(\beta_0, t_r)}{\bar{S}^{(0)}(\beta_0, t_r)}\right| |\bar{M}_r| \tag{5.7}$$

$$+\frac{1}{n}\sum_r \left|\frac{\bar{S}^{(1)}(\beta_0, t_r)}{\bar{S}^{(0)}(\beta_0, t_r)} - \frac{S_g^{(1)}(\beta_0, t_r)}{S_g^{(0)}(\beta_0, t_r)}\right| |\bar{M}_r|, \tag{5.8}$$

where $\bar{M}_r$ is the increment of $\bar{M}$ over $\mathcal{T}_r$. Using the Lipschitz property of $s^{(k)}$, the Cauchy–Schwarz inequality, the uniform boundedness of $E\{S^{(0)}(\beta_0, u)\}^{-4}$, and since $s^{(0)}(\beta_0, u)$ is bounded away from zero, it can be shown that the second moment of the integrand in (5.6) is of order $O_P(n^{-1}+l^2)$ uniformly in u. The integrand is also predictable, so martingale theory gives (5.6) $= O_P(n^{-1} + ln^{-1/2})$. Martingale theory also gives $E\bar{M}_r^2 = O(nl)$ uniformly in r, so that, using Lemma 5.1 (i), we find (5.7) $= O_P(n^{-1}l^{-1/2})$. Lemma 5.1 and the Cauchy–Schwarz inequality give (5.8) $= O_P(c_n n^{-1/2})$. Combining terms gives (5.3) $= O_P\{(l + c_n)n^{-1/2}\}$.

Next, Taylor expanding $z \mapsto e^{\beta_0 z}$ about each z_j and λ_0 about each t_r, gives

$$(5.4) = -\frac{1}{n}\sum_{i,r,j} z_j e^{\beta_0 z_j}\int_{\mathcal{T}_r} \gamma_{rj}(u) 1_{\{Z_i \in \mathcal{I}_j\}} Y_i(u)\,du + O_P(l^3 + w^3),$$

where

$$\gamma_{rj}(u) = \lambda_0(t_r) + (u - t_r)\dot{\lambda}_0(t_r) + \tfrac{1}{2}(u - t_r)^2\ddot{\lambda}_0(t_r) + \lambda_0(u)\left\{\beta_0(Z_i - z_j) + \tfrac{1}{2}\beta_0^2(Z_i - z_j)^2\right\}.$$

There is a similar decomposition for (5.5). Note that the leading terms involving $\lambda_0(t_r)$ in these decompositions cancel. By examining means and variances of the remaining terms, Taylor expanding $f = F'$ about each t_r or each z_j, and cancelling the terms involving β_0^2 or $\ddot{\lambda}_0$, we find that

$$(5.4) + (5.5) = \Delta V^{-1} + O_P\{l^3 + w^3 + (l+w)n^{-1/2}\}.$$

Combining the above rates for (5.2)–(5.5) completes the proof. □

Lemma 5.4 $\{I(\beta) - I_g(\beta)\}/n = O_P\{l + w^2 + (1 + c_n)n^{-1/2}\}$ uniformly over $\beta \in \mathcal{B}$.

Proof Using Lemma 5.1 and the uniform boundedness of $s^{(0)}$ away from 0, we find that

$$\begin{aligned}\sup_\beta |I(\beta) - I_g(\beta)|/n &\le O_P(1) \sum_{k=0}^{2} \Big\{ \sum_r \sup_\beta \Big| S_g^{(k)}(\beta, t_r) - \bar{S}^{(k)}(\beta, t_r)\Big| \bar{N}(\mathcal{T}_r)/n \\ &\qquad + \sup_{t,\beta} \Big|\bar{S}^{(k)}(\beta, t) - S^{(k)}(\beta, t)\Big| \bar{N}(1)/n\Big\} \\ &= O_P\{w^2 + (\sqrt{w} + c_n)n^{-1/2}\} + O_P(l + n^{-1/2}),\end{aligned}$$

where the Cauchy–Schwarz inequality was used for the first term, and the Lipschitz property of $s^{(k)}$ for the second term. □

Lemma 5.5

(i) $I(\beta^*)/n = V^{-1} + O_P(n^{-1/2})$

(ii) $I(\beta_g^*)/n = V^{-1} + O_P\{l^2 + w^2 + (1 + c_n)n^{-1/2}\}$.

Proof Part (i) is proved by inspecting (3.1) in AG, and use of Lemma 5.1 (i), the rate $\beta^* = \beta_0 + O_P(n^{-1/2})$ given by Theorem 3.2 of AG, and the Lipschitz property of v. For part (ii), first note that β_g^* is consistent for β_0 by Lemma 5.2. Also, from a grouped data version of AG's (3.1) and Lemma 5.1 (ii), we have $I(\beta_g^*)/n \xrightarrow{P} V^{-1}$. Thus, $\hat{\beta}_g$ converges at rate

$$\hat{\beta}_g - \beta_0 \;=\; U_g(\beta_0)/I_g(\beta_g^*) = O_P\{l^2 + w^2 + (1 + c_n)n^{-1/2}\}, \tag{5.9}$$

by Lemma 5.3, since $U(\beta_0)/n = O_P(n^{-1/2})$ and $\Delta = O(l^2 + w^2)$. Once more using the grouped data version of AG's (3.1), but with the rate (5.9) applied to β_g^*, proves part (ii). □

The proof of Theorem 2.1 is completed by using (5.1), applying Lemma 5.3 to A, Lemma 5.5 to B and C, Lemma 5.4 to D, and simplifying.

References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100–1120.

Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Statist.* **42** 1656–1670.

Breslow, N. E. (1986). Cohort analysis in epidemiology, in A. C. Atkinson and S. E. Fienberg, eds., *A Celebration of Statistics: the ISI Centenary Volume*. Springer-Verlag, New York, 109–143.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. B.* **34** 187–220.

Dempster, A. P. and Rubin, D. B. (1983). Rounding error in regression: The appropriateness of Sheppard's corrections. *J. Roy. Statist. Soc. B.* **45** 51–59.

Don, F. J. H. (1981). A note on Sheppard's corrections for grouping and maximum likelihood estimation. *J. Mult. Anal.* **11** 452–458.

Hahn, M. G. (1978). Central limit theorems in $D[0,1]$. *Z. Wahrsch. Verw. Gebiete* **44** 89–102.

Haitovsky, Y. (1973). *Regression estimation from grouped observations*. Hafner Press, New York.

Haitovsky, Y. (1983). Grouped data. *Encyclopedia of Statistical Sciences* **3** 527–536 (1983), Eds. N.L. Johnson and S. Kotz, John Wiley, New York.

Hoem, J. M. (1987). Statistical analysis of a multiplicative model and its application to the standardization of vital rates: a review. *Int. Statist. Rev.* **55** 119–152.

Holford, T. R. (1976). Life tables with concomitant information. *Biometrics* **32** 587–597.

Huet, S. and Kaddour, A. (1994). Maximum likelihood estimation in survival analysis with grouped data on censored individuals and continuous data on failures. *Appl. Statist.* **43** 325–333.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60** 267–278.

Kolassa, J. E. and McCullagh, P. (1990). Edgeworth series for lattice distributions. *Ann. Statist.* **18** 981–985.

Laird, N. and Olivier, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *J. Amer. Statist. Assoc.* **76** 231–240.

Lindley, D. V. (1950). Grouping corrections and maximum likelihood equations. *Proc. Camb. Phil. Soc.* **46** 106–110.

Prentice R. L. and Gloeckler L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* **34** 57–67.

Preston, D. L., Kato, H., Kopecky, K. J., and Fujita, M. S. (1987). Life Span Study Report 10, Part I, Cancer Mortality among A-Bomb Survivors in Hiroshima and Nagasaki, 1950–82. *Radiat. Res.* **111** 151–178.

Selmer, R. (1990). A comparison of Poisson regression models fitted to multiway summary tables and Cox's survival model using data from a blood pressure screening in the city of Bergen, Norway. *Statistics in Medicine* **9** 1157–1165.