Statistica Sinica **5**(1995), 793-804

GROUP SEQUENTIAL METHODS FOR SURVIVAL DATA USING PARTIAL LIKELIHOOD SCORE PROCESSES WITH COVARIATE ADJUSTMENT

Minggao Gu and Zhiliang Ying

McGill University and University of Illinois

Abstract. A general Cox-type partial likelihood score process for staggered entry data with covariate adjustment is shown to be asymptotically equivalent to a Gaussian process with independent increments, regardless of whether or not the covariates being adjusted for are independent of the covariates of primary interest. The approximation yields new and simple group sequential tests as well as repeated confidence intervals that effectively incorporate information from ancillary concomitant variables. A recursive formula is derived for computing discrete boundary values when the parameter of interest is multidimensional. A prostatic cancer data set is implemented to illustrate usefulness of the new approach. Results of simulation studies with moderate sample sizes are reported, showing that the group sequential tests with covariate adjustment perform markedly well in terms of efficiency improvement and bias reduction.

Key words and phrases: Censoring, covariate adjustment, Gaussian process, group sequential test, independent increments, martingale, proportional hazards regression, repeated confidence intervals, survival data.

1. Introduction

In his seminal work, Cox (1972, 1975) proposed the proportional hazards regression model and its associated partial likelihood for the analyses of survival data. He demonstrated that many basic identities, crucial to the asymptotic theory for the usual maximum likelihood estimation, also hold for the partial likelihood function. Rigorous theoretical developments of large sample properties were subsequently provided by Tsiatis (1981) and Andersen and Gill (1982). Parallel results for general relative risk models were due to Prentice and Self (1983).

In many clinical trials and epidemiological studies, individuals are recruited sequentially and interim analyses are conducted to monitor their progress. An important step in applying Cox's model to these situations is to characterize limiting behavior of the partial likelihood score process calculated over the entire study period. Despite this seemingly complicated setup, Tsiatis (1982), Sellke and Siegmund (1983), Slud (1984) and Gu and Lai (1991) were able to reveal, in

increasing generality, a very simple underlying structure, i.e., the score process so calculated with a univariate concomitant variable converges to a time-rescaled Brownian motion process. This important characteristic of the limiting distribution legitimates use of many available sequential boundaries. Furthermore, these boundaries can easily be discretized to form more realistic group sequential boundaries with more or less the same operating characteristics; (cf. Pocock (1977), O'Brien and Fleming (1979) and Lan and DeMets (1983)). These characteristics also facilitate the use of repeated confidence intervals; (cf. Jennison and Turnbull (1989)).

Tsiatis, Rosner and Tritchler (1985) found that efficiency in a group sequential test may be substantially increased when other prognostic factors are available. Their idea to construct more efficient tests was to first estimate parameters for those factors and then replace them in the sequential testing statistics by their estimators. A key assumption in their paper is that the covariate component being tested is independent of all other components being adjusted for. They argued that under this assumption, the test statistics they considered are asymptotically equivalent to the ones without replacements, thereby making the Brownian approximations, developed by earlier authors cited in the preceding paragraph, valid. A similar finding for the accelerated life model under the same independence assumption was obtained by Lin (1992). While it is reasonable to assume the independence between the treatment variable and other baseline variables in a randomized clinical trial, such an assumption does impose a serious limitation. In the words of Meier (1983), "A major exclusion is the common situation in which allocation to treatment is not made at random, so that statistical comparability of treatment groups cannot be assumed. In such studies, extensive adjustment for baseline variables is not simply an option with merits and limitations to be weighed, but an absolute necessity."

We study, in this article, the partial likelihood score process adjusting for ancillary covariate components, which may depend on the covariate components of primary interest. A key result of our findings is that, when covariate adjustment is made in a proper way, such a process can still be approximated by a Gaussian process with independent increments. Thus, in the case that only one covariate component is of primary interest, the process is approximately a time-rescaled Brownian motion. In this regard, many available sequential and group sequential boundaries may be used to construct desirable tests. We also demonstrate how the property of independent increments can facilitate computation of a general group sequential boundary, especially when the process is multidimensional. Comparisons with tests that do not adjust for ancillary covariates are made through a prostatic cancer data set as well as simulation studies.

2. Notation and Main Theoretical Results

Consider a typical clinical trial with patients being recruited sequentially. Suppose patient i enters the trial at calendar time τ_i and either dies at $\tau_i + T_i$, or is censored at $\tau_i + C_i$. Associated with each i are two time-dependent covariate processes. Denote $Z_i(p \times 1)$ to be the covariate vector of primary interest, usually indicating the type of treatment patient i receives, and $W_i(q \times 1)$ to be other measurements we intend to adjust. Define $X_i(t) = \max\{\min(T_i, C_i, t - \tau_i), 0\}$ and $\Delta_i(t) = I\{T_i \leq \min(t - \tau_i, C_i)\}$. Thus, the observed data at calendar time t consist of $X_i(t)$, $\Delta_i(t)$ and $\{Z_i(u), W_i(u); u \leq X_i(t)\}$.

The proportional hazards model with a general relative risk form assumes

$$\text{pr}\{s \leq T_i \leq s + ds | T_i \geq s,\ Z_i(u),\ W_i(u); u \leq s\} = h\{\gamma' Z_i(s),\ \beta' W_i(s)\}\lambda_0(s)ds,$$

where h is a prespecified smooth link function, γ and β are unknown parameter vectors and λ_0 is an arbitrary baseline hazard rate function. We shall call γ the primary parameter and β the ancillary parameter. The special case of the exponential link $h\{\gamma' Z_i(s),\ \beta' W_i(s)\} = \exp\{\gamma' Z_i(s) + \beta' W_i(s)\}$ is the well-known Cox regression model. The general relative risk form was studied in detail by Prentice and Self (1983), who argued that in some situations link functions other than the exponential form are more suitable. A particularly useful example is the linear intensity function model $h\{\gamma' Z_i(s),\ \beta' W_i(s)\} = 1 + \gamma' Z_i(s) + \beta' W_i(s)$.

It will be assumed throughout the sequel that censoring is noninformative in the sense that given Z_i and W_i, T_i and C_i are independent. Under this assumption, the partial likelihood function (Cox (1975), Prentice and Self (1983)) calculated at calendar time t takes the form

$$L_n(t; \gamma, \beta) = \prod_{i=1}^{n} \left[\frac{h\{\gamma' Z_i(X_i(t)),\ \beta' W_i(X_i(t))\}}{\sum_{j=1}^{n} h\{\gamma' Z_j(X_i(t)),\ \beta' W_j(X_i(t))\} I\{X_j(t) \geq X_i(t)\}} \right]^{\Delta_i(t)}.$$

Differentiating the log partial likelihood function with respect to γ and β gives rise to two score processes

$$S_n(t; \gamma, \beta) = \frac{\partial}{\partial \gamma} \log\, L_n(t; \gamma, \beta), \quad U_n(t; \gamma, \beta) = \frac{\partial}{\partial \beta} \log\, L_n(t; \gamma, \beta).$$

To test hypothesis $H_0 : \gamma = \gamma_0$, one might use $S_n(t; \gamma_0, \beta_0)$ with an appropriate normalizing factor if β_0 were known. In practice, β_0 is unknown but can be estimated from the available data at t. We recommend using $\hat{\beta}_t$ defined by

$$U_n(t; \gamma_0, \hat{\beta}_t) = 0. \tag{1}$$

The role and motivation of (1) for estimation of the ancillary parameter β_0 will be discussed in Section 5. It is seen there that $\hat{\beta}_t$ is crucial to the simple variance-covariance structure of our statistic. Arising from this is the score

process $S_n(t;\gamma_0,\hat{\beta}_t)$ upon which we will concentrate. The usefulness of this process depends on its distributional characteristics. In Appendix, we will derive the key asymptotic result, i.e., under H_0, $\{n^{-1/2}S_n(t;\gamma_0,\hat{\beta}_t), t \geq 0\}$ converges to a Gaussian process with independent increments and variance-covariance function $J(t) = \sigma_{z,z}(t) - \sigma_{z,w}(t)\sigma^{-1}_{w,w}(t)\sigma'_{z,w}(t)$, where $\sigma_{z,z}(t)$ is the limit of

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^t Z_{h,i}(s;t,\gamma_0,\beta_0)Z'_{h,i}(s;t,\gamma_0,\beta_0)I\{X_i(t)\geq s\}h(\gamma_0'Z_i(s),\ \beta_0'W_i(s))d\Lambda_0(s), \tag{2}$$

and, $\sigma_{z,w}(t)$ and $\sigma_{w,w}(t)$ are similarly defined with Z and Z' replaced respectively by Z and W', and W and W'. Here $Z_{h,i}(s;t,\gamma,\beta)$ and $W_{h,i}(s;t,\gamma,\beta)$ are the derivatives of $\log h\{\gamma'Z_i(s),\ \beta'W_i(s)\} - \log\sum_{j=1}^n h\{\gamma'Z_j(s),\ \beta'W_j(s)\}I\{X_j(t)\geq s\}$ with respect to γ and β.

In the special case $p=1$, $n^{-1/2}S_n(t;\gamma_0,\hat{\beta}_t)$ converges to a time-rescaled Brownian motion $B(J(t))$, where B denotes the standard Brownian motion process. Moreover, under contiguous alternatives $H_a : \gamma = \gamma_0 + \mu/\sqrt{n}$, the normalized score process $n^{-1/2}S_n(t;\gamma_0,\hat{\beta}_t)$ converges to $B(J(t)) + J(t)\mu$. In other words, under the new time scale J, the score process is approximately a Brownian motion with a linear drift.

The approximation to $S_n(t;\gamma_0,\hat{\beta}_t)$ by a Gaussian process with independent increments generalizes Sellke and Siegmund (1983) to accommodate covariate adjustment. It also extends the findings by Tsiatis et al. (1985), who assumed that (a) the two covariate vectors Z and W are independent; (b) the link function $h(\gamma'Z,\ \beta'W)$ can be factorized to $h_1(\gamma'Z)h_2(\beta'W)$; (c) $\gamma_0 = 0$. To see this, note that under their assumptions, $\sigma_{z,w} = 0$ and, therefore, $n^{-1/2}S_n(t;\gamma_0,\hat{\beta}_t)$ has the same limiting distribution as $n^{-1/2}S_n(t;\gamma_0,\beta_0)$.

The variance-covariance function $J(t)$ can be estimated consistently. We recommend using $\hat{J}(t) = \hat{\sigma}_{z,z}(t) - \hat{\sigma}_{z,w}(t)\hat{\sigma}^{-1}_{w,w}(t)\hat{\sigma}'_{z,w}(t)$, where $\hat{\sigma}_{z,z}$ $\hat{\sigma}_{z,w}$ and $\hat{\sigma}_{w,w}$ are obtained by substituting β_0 and $\Lambda_0(s)$ in (2) by $\hat{\beta}_t$ and

$$\hat{\Lambda}_0(s;t) = \sum_{i=1}^{n}\frac{\Delta_i(t)I\{X_i(t)\leq s\}}{\sum_{j=1}^n h\{\gamma_0'Z_j(X_i(t)),\ \hat{\beta}_t'W_j(X_i(t))\}I\{X_j(t)\geq X_i(t)\}}.$$

The proof for the consistency of $\hat{J}(t)$ is outlined in Appendix.

3. Repeated Significance Tests

In the preceding section we have derived the limiting distribution of the adjusted score process $n^{-1}S_n(t;\gamma_0,\hat{\beta}_t)$. For $p=1$, it is distributed as a time-rescaled Brownian motion $B(J(t))$ under the null hypothesis, and as $B(J(t))$ plus a linear drift $J(t)\mu$ under contiguous alternatives. Thus, many well-studied

sequential boundaries for testing a drift in Brownian motion can be applied. For example, one may use the simple parallel boundary to test $H_0 : \gamma = \gamma_0$ versus $H_a : \gamma < \gamma_0$. This leads to stopping rules $\tau = \inf\{t : n^{-1/2} S_n(t; \gamma_0, \hat{\beta}_t) \notin (a, b)\}$. On the other hand, when repeated significance tests are desirable, we can use stopping rules $\tau = \inf\{t \geq t_0 : |S_n(t; \gamma_0, \hat{\beta}_t)|/\{n\hat{J}(t)\}^{1/2} \geq b$, or $n\hat{J}(t) \geq m\}$ with a prespecified $m > 0$ to get curved boundaries. Note that by taking $v = \hat{J}(t)$, $S_n(t; \gamma_0, \hat{\beta}_t)/\{n\hat{J}(t)\}^{1/2}$ behaves like $B(v)/v^{1/2}$. Details on various sequential boundaries for Brownian motion can be found in Siegmund (1985, pp. 36-43, 70-89).

Instead of continuously monitoring the trial, we may consider more realistic group sequential problems. Suppose the number of looks K can be specified in advance. Then we can determine successively the calendar times $t_1 = \inf\{t > t_{i-1} : \hat{J}(t) > mi/K\}$, $i = 1, \ldots, K$, to review the trial, where t_0 and m are fixed constants. Since t_i defined in this way give equal information increments, the boundaries of Pocock (1977) and O'Brien and Fleming (1979) can be applied by specifying exit probabilities accordingly. On the other hand, if K cannot be determined in advance but there is a continuous boundary with stopping rule, say τ, available, then it is natural to follow the proposal of Lan and DeMets (1983). This is done by calculating $\alpha(t) = \text{pr}(\tau > t)$ and by using it to obtain discrete boundary values. The method proposed by Slud and Wei (1982) may also be used to obtain desirable tests when the monitoring times t_i and the upper bound K are predetermined.

When $p > 1$, it is prohibitively difficult to construct continuous boundaries for testing $H_0 : \gamma = \gamma_0$, even when $\gamma_0 = 0$; (cf. Siegmund (1985)). Nonetheless, boundaries for group sequential tests can be evaluated in principle by applying, for example, Slud and Wei (1982). In practice, the evaluation process involves $K \times p$ multiple integration, which we found to be difficult even when $K \times p$ is moderate, say, 10. We now describe a method that avoids high dimensional integration by exploiting the Markovian nature of $S_n(t; \gamma_0, \hat{\beta}_t)$. To test H_0 against $H_a : \gamma \neq \gamma_0$ with levels α_k at t_k, $k = 1, \ldots, K$, so that the overall level is $\alpha = \alpha_1 + \cdots + \alpha_K$, consider the score process $\hat{V}_n(t) = S_n'(t; \gamma_0, \hat{\beta}_t)\{n\hat{J}(t)\}^{-1} S_n(t, \gamma_0, \hat{\beta}_t)$. Denote J_k to be $n\hat{J}(t_k)$ and ξ_k to be normal random variables with $E\xi_k = 0$ and $E(\xi_k \xi_{k^*}') = J_k, k \leq k^*$. The boundary values $d_1, \ldots, d_k$ of Slud and Wei are defined successively through

$$\alpha_k = \text{pr}(V_1 \leq d_1, \ldots, V_{k-1} \leq d_{k-1}, V_k > d_k), \quad k = 1, \ldots, K, \tag{3}$$

where $V_k = \xi_k' J_k^{-1} \xi_k$. Note that direct computation of (3) involves $k \times p$ dimensional integration. To avoid doing this, we propose to first compute

$$f_k(x)dx = \text{pr}\{V_1 \leq d_1, \ldots, V_{k-1} \leq d_{k-1}, J_k^{-1/2} \xi_k \in [x, x + dx)\}. \tag{4}$$

In Appendix, we shall make use of the Markovian property of $\{\xi_k\}$ to derive the key recursion

$$f_k(x) = \int_{\|y\|^2 \le d_{k-1}} N(J_k^{-1/2} J_{k-1}^{1/2} y, I - J_k^{-1/2} J_{k-1} J_k^{-1/2}; x) f_{k-1}(y) dy, \qquad (5)$$

where $N(\mu, \Sigma; x)$ denotes the p-variate normal density at x with mean μ and covariance matrix Σ. But (5) involves only p-dimensional integration, regardless of k. Once f_k is computed, we proceed to determine d_k via $\int_{\|x\|^2 \ge d_k} f_k(x) dx = \alpha_k$.

4. Numerical Studies

To illustrate the usefulness of the foregoing results, we reexamine the prostatic cancer data given in Byar (1985). The data set was collected from a randomized clinical trial for comparing treatments at four different levels, placebo, .2 mg diethylstilbestrol (DES), 1 mg and 5 mg, for prostatic cancer patients in stages 3 and 4, which represent, respectively, without or with evidence of distant metastasis. Clearly, which stage a cancer patient is in is strongly correlated to his/her survival. We therefore expect efficiency improvement if the stage effect is incorporated. In our analysis, we excluded all other prognostic factors that are also listed in Byar (1985) for the sake of simplicity. Cox's proportional hazards model

$$\lambda(t|Z, W) = \exp(\gamma Z + \beta W) \lambda_0(t)$$

was used with Z indicating the treatments and W the cancer stages. As it turns out, there is (log)nonlinearity of the treatment effect between 1.0 mg and 5.0 mg. So we combined the two levels into one, i.e., Z takes three values instead of four.

Table 1 reports two sequential tests, one with adjustment and the other without, for the hypothesis of no treatment effect against the two-sided alternatives that treatment effect does exist. Both tests use standardized Cox's score processes with boundaries calculated via the method of Slud and Wei (1982). Discretization is done at the end of each year from 1968 to 1974. We omitted 1973 partly because there are few deaths during that period and partly to facilitate computation. The overall type-one error is $\alpha = .05$, which is evenly distributed over the six looks. By comparing the third and the last columns of Table 1, we see that the boundary is crossed at the fourth look with a substantial margin of overshoot. Thus the null hypothesis would be rejected at the end of 1971 if the test with covariate adjustment as described were used. On the other hand from Table 1, the boundary is crossed at the fifth look, with a smaller margin. So the null hypothesis would also be rejected, but with a one-year delay, if no covariate adjustment is made.

Table 1. Sequentially calculated values of number of death, standardized score statistic, variance and boundary for Prostatic Cancer Data

Time	Death	Std. statistic	Variance	Boundary*
Dec., 68	22	1.982 (1.998)**	3.628 (3.727)	2.638 (2.638)
Dec., 69	53	2.229 (1.972)	10.14 (10.27)	2.588 (2.587)
Dec., 70	88	2.280 (1.833)	17.51 (17.80)	2.519 (2.519)
Dec., 71	109	2.886 (2.309)	21.63 (22.04)	2.401 (2.401)
Dec., 72	123	3.003 (2.427)	24.27 (25.00)	2.291 (2.299)
Dec., 74	130	3.140 (2.548)	25.77 (26.30)	2.183 (2.178)

*The overall type-one error is $\alpha = .05$, which is distributed evenly over the six looking times.

**Figures inside ( ) are calculated using the unadjusted log rank statistic.

Table 2. Summary of type-one errors and powers of simulation results for the tests ($\alpha = .05$) based on adjusted and unadjusted partial likelihood score processes

		corr = 0		corr = .3		corr = .5	
β	γ	Adjusted	Unadj.	Adjusted	Unadj.	Adjusted	Unadj.
0	0	.055	.054	.053	.054	.053	.054
	1	.765	.775	.701	.775	.507	.775
.5	0	.056	.053	.051	.097	.055	.141
	1	.778	.780	.715	.875	.521	.921
1	0	.052	.052	.051	.166	.053	.312
	1	.790	.760	.722	.931	.529	.978
2	0	.050	.052	.051	.373	.050	.760
	1	.795	.668	.737	.976	.540	.999
3	0	.049	.052	.050	.578	.050	.956
	1	.799	.557	.741	.989	.544	1.000

To further investigate the finite sample behavior of the proposed tests and compare their efficiency with tests based on unadjusted partial likelihood score processes, we have conducted some simulation studies with sample size $n = 100$. The results are presented in Table 2. To ease computation, both the primary covariate X and the ancillary covariate W were taken to be univariate with

uniform $[0,1]$. The correlation between X and W was taken as 0, .3 and .5 and the nuisance parameter, β, was taken as 0, .5, 1, 2 and 3. Given X and W, the survival time T is exponentially distributed with hazard rate $\lambda(X,W) = \exp(\gamma X + \beta W)$. The censoring time C was independently generated from an exponential distribution with hazard rate $\lambda_c = \exp(-1.5)$. The entry time τ is uniformly distributed over $[0,4]$. Three interim analyses were conducted at times 2, 3.5 and 5. Pocock-type boundaries were used throughout for $\gamma = 0$ versus $\gamma > 0$. In each entry, the top figure stands for the rejection rate under the null hypothesis (type-one error) while the bottom one is the rejection rate under alternative $\gamma = 1$ (power). 20,000 simulations were generated for each combination of parameters; and both adjusted and unadjusted statistics were calculated from the same sequence of the generated data when the parameters are the same. When there is no correlation, the asymptotic theory indicates that both adjusted and unadjusted log rank tests are valid. The figures under corr $= 0$ show that the type-one errors are quite close to the target value .05. Their powers are compatible for $\beta = 0$ and $\beta = .5$. As β increases, the adjusted test becomes significantly more powerful. On the other hand, when the two covariate variables are correlated, the unadjusted test is no longer valid unless $\beta = 0$. This is because even if $\gamma = 0$, X could still be related to the survival time through W. Indeed, figures under corr $= .3$ and corr $= .5$ show that their type-one errors can deviate from .05 quite substantially. Nonetheless, the adjusted test still gives approximately correct type-one errors and reasonably good powers.

5. Discussion

As pointed out in Section 2, use of $\hat{\beta}_t$ defined by (1) is essential for process $S_n(t;\gamma_0,\hat{\beta}_t)$ to have independent increments. For example, if one uses $\hat{\beta}_t^*$, where $(\hat{\gamma}_t^*,\hat{\beta}_t^*)$ solves $U_n(t;\hat{\gamma}_t^*,\hat{\beta}_t^*) = 0$ and $S_n(t;\hat{\gamma}_t^*,\hat{\beta}_t^*) = 0$, then it is easy to show that the resulting statistic $S_n(t;\gamma_0,\hat{\beta}_t^*)$ in general need not have independent increments.

Another application of the score process with covariate adjustment is to obtain a sequence of repeated confidence intervals (Jennison and Turnbull (1989), Lai (1984)) at $t_1, t_2, \ldots$, say, for γ by inverting intervals for $S_n(t_k;\gamma,\hat{\beta}_{t_k})$, $k = 1,2,\ldots$, where $\hat{\beta}_t$ solves $U_n(t;\gamma,\hat{\beta}_t) = 0$. Although direct inversion is computationally complex, a rather simple approximation can be used to obtain the confidence intervals. In fact, it is easy to show that $S_n(t;\gamma,\hat{\beta}_t)$ is asymptotically equivalent to $D_n(t;\hat{\gamma}_t^*,\hat{\beta}_t^*)(\gamma - \hat{\gamma}_t^*)$, where

$$D_n(t;\gamma,\beta) = \frac{\partial}{\partial\gamma}S_n(t;\gamma,\beta) - \frac{\partial}{\partial\beta}S_n'(t;\gamma,\beta)\Big[\frac{\partial}{\partial\beta}U_n(t;\gamma,\beta)\Big]^{-1}\frac{\partial}{\partial\gamma}U_n(t;\gamma,\beta).$$

Thus, we can choose appropriate constants $c_{n,k}$ to get confidence intervals $\hat{\gamma}^*_{t_k} \pm c_{n,k} D_n^{-1}(t_k; \hat{\gamma}^*_{t_k}, \hat{\beta}^*_{t_k})$. This is in analogy with a proposal by Harrington (1989) on repeated confidence intervals for the regression parameter when there is no covariate adjustment.

For the prostatic cancer data, the repeated confidence intervals at the six looking times specified in Table 1 are found to be $(-2.38, .38)$, $(-1.50, .12)$, $(-1.15, .05)$, $(-1.13, -.09)$, $(-1.08, -.15)$ and $(-1.05, -.19)$, respectively. Here the spending error probabilities are chosen to be equal to .05/6 and thus the overall coverage probability is .95. Note that the fourth interval is the first one to exclude 0, in agreement with the testing results in Table 1.

The idea of sequentially testing a parameter in presence of a nuisance parameter using the likelihood ratio statistic was first proposed by Whitehead (1978, 1983). A more explicit description of his method can be found in Siegmund (1985, pp. 63-66).

While the proposed method was applied only to a data set from a clinical trial, we believe interesting applications can also be found in epidemiological studies, where dependency among covariate components is common. Research on applying our findings and catering them towards special needs of those studies are certainly desirable.

The theory developed in Section 2 is based upon the assumption that probability relation between covariates and survival times are correctly modeled. In practice, some kind of model misspecification is unavoidable. Lin and Wei (1989) discussed the robustness issue of Cox's regression model in a nonsequential setting and proposed a robust variance estimator that is valid even when the model is misspecified. It is possible to apply their approach to get a similar alternative variance estimator for $S_n(t; \gamma_0, \hat{\beta}_t)$. Many theoretical and methodological problems in this respect, however, remain to be addressed.

Since the purpose of this work is to show sequentially computed Cox's score process with covariate adjustment has (asymptotically) independent increment and to demonstrate the importance of this feature, the technical developments given in Appendix are somewhat sketchy. Fully rigorous proofs, including tightness of multidimensional Cox's score with staggered entry, are quite involved and will be reported elsewhere.

Acknowledgement

This research was supported by the National Sciences and Engineering Research Council of Canada (for Gu) and the National Science Foundation and the National Security Agency (for Ying). The authors thank the referee for suggesting the simple construction of repeated confidence intervals.

Appendix

Derivation of the limiting distribution of $n^{-1/2}S_n(t;\gamma_0,\hat{\beta}_t)$. Let $A_n(t;\gamma,\beta)$ and $B_n(t;\gamma,\beta)$ be the derivative functions respectively of $U_n(t;\gamma,\beta)$ and $S_n(t;\gamma,\beta)$ with respect to β. By taking the Taylor expansions at β_0, we have

$$0 = U_n(t;\gamma_0,\hat{\beta}_t) = U_n(t;\gamma_0,\beta_0) + nA_n(t;\gamma_0,\beta_0)(\hat{\beta}_t - \beta_0) + o_p(n^{1/2}),$$
$$S_n(t;\gamma_0,\hat{\beta}_t) = S_n(t;\gamma_0,\beta_0) + nB_n(t;\gamma_0,\beta_0)(\hat{\beta}_t - \beta_0) + o_p(n^{1/2}),$$

where $o_p(1)$ is uniform in t. Standard algebraic and probabilistic calculations as given in Prentice and Self (1983) yield $A_n(t;\gamma_0,\beta_0) = \sigma_{w,w}(t)+o_p(1)$, $B_n(t;\gamma_0,\beta_0) = \sigma_{z,w}(t) + o_p(1)$. Therefore,

$$n^{-1/2}S_n(t;\gamma_0,\hat{\beta}_t) = n^{-1/2}S_n(t;\gamma_0,\beta_0) + n^{-1/2}\sigma_{z,m}(t)\sigma^{-1}_{w,w}(t)U_n(t;\gamma_0,\beta_0) + o_p(1),$$

which is certainly tight and asymptotically Gaussian. To verify the covariance structure of the limiting process, note that

$$n^{-1/2}S_n(t;\gamma_0,\beta_0) = \sum_{i=1}^{n} n^{-1/2}\int_0^t Z_{h,i}(s;t,\gamma_0,\beta_0)dM_i(s;t),$$

where $M_i(s;t) = I\{X_i(t)\le s\}\Delta_i(t) - \int_0^s I\{X_i(t)\ge u\}d\Lambda_0(u)$. For fixed $t\le t^*$, $M_i(\cdot;t)$ and $M_i(\cdot;t^*)$ are martingales with respect to an appropriate σ-filtration. Evaluating the predictable covariation between the two martingale integrals $n^{-1/2}$ $S_n(t;\gamma_0,\beta_0)$ and $n^{-1/2}S_n(t^*;\gamma_0,\beta_0)$ gives

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^t Z_{h,i}(s;t,\gamma_0,\beta_0)Z_{h,i}(s;t^*,\gamma_0,\beta_0)h_i(s;\gamma_0,\beta_0)I\{X_i(t)\ge s\}d\Lambda_0(s)$$
$$= \sigma_{z,z}(t) + o_p(1).$$

This and other similar variance-covariance calculations show that the asymptotic covariance between $n^{-1/2}S_n(t;\gamma_0,\hat{\beta}_t)$ and $n^{-1/2}S_n(t^*;\gamma_0,\hat{\beta}^*_t)$ is $J(t) = \sigma_{z,z}(t) - \sigma_{z,w}(t)\sigma^{-1}_{w,w}(t)\sigma_{w,z}(t)$. Hence, the limit of $n^{-1/2}S_n(t;\gamma_0,\hat{\beta}_t)$ has independent increments with J as its variance-covariance function.

Proof of consistency of $\hat{J}(t)$. It suffices to show that $\hat{\sigma}_{z,z}(t),\hat{\sigma}_{z,w}(t)$ and $\hat{\sigma}_{w,w}(t)$ converge respectively to $\sigma_{z,z}(t),\sigma_{z,w}(t)$ and $\sigma_{w,w}(t)$, which are straightforward consequences of $\hat{\beta}_t\to\beta_0$ and $\hat{\Lambda}_0(s;t)\to\Lambda_0(s)$.

Derivation of Equation (5). For $k>1$, by definition

$$f_k(x)dx = \int_{\|y\|^2\le d_{k-1}} \mathrm{pr}\{V_1\le d_1,\ldots,V_{k-2}\le d_{k-2}, J^{-1/2}_{k-1}\xi_{k-1}\in[y,y+dy),$$
$$J^{-1/2}_k\xi_k\in[x,x+dx)\}$$
$$= \int_{\|y\|^2\le d_{k-1}} \mathrm{pr}\{J^{-1/2}_k\xi_k\in[x,x+dx)|J^{-1/2}_{k-1}\xi_{k-1}=y\}f_{k-1}(y)dy,$$

where the second equality holds by the strong Markov property. Since the conditional density of $J_k^{-1/2}\xi_k$ at x given $\xi_{k-1} = J_{k-1}^{1/2}y$ is the first term of the integrand of (7), we have the proof.

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.

Byar, D. P. (1985). Prognostic variables for survival in a randomized comparison of treatments for prostatic cancer. In Data (Edited by D. F. Andrews and A. M. Herzberg), 261-262, Springer-Verlag, New York.

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

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

Gu, M. G. and Lai, T. L. (1991). Weak convergence of time-sequential censored rank statistics with applications to sequential testing in clinical trials. Ann. Statist. **19**, 1403-1433.

Harrington, D. (1989). Discussion of 'Interim analyses: the repeated confidence interval approach' by C. Jennison and B. W. Turnbull. J. Roy. Statist. Soc. Ser.B **51**, 346-347.

Jennison, C. and Turnbull, B. W. (1989). Interim analyses: The repeated confidence interval approach. J. Roy. Statist. Soc. Ser.B **51**, 305-361.

Lai, T. L. (1984). Incorporating scientific, ethical and economic considerations into the design of clinical trials in the pharmaceutical industry: A sequential approach. Comm. Statist. Theory Methods **13**, 2355-2368.

Lan, K. K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. Biometrika **70**, 659-663.

Lin, D. Y. (1992). Sequential log rank tests adjusting for covariates with the accelerated life model. Biometrika **79**, 523-529.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. J. Amer. Statist. Assoc. **84**, 1074-1078.

Meier, P. (1983). Statistical analysis of clinical trials. In Clinical Trials (Edited by S. H. Shapiro and T. A. Louis), 155-189, Marcel Dekker, New York.

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. Biometrics **35**, 549-556.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. Biometrika **64**, 191-199.

Prentice, R. L. and Self, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. Ann. Statist. **11**, 804-813.

Sellke, T. and Siegmund, D. (1983). Sequential analysis of the proportional hazards model. Biometrika **70**, 315-326.

Siegmund, D. (1985). Sequential Analysis: Tests and Confidence Intervals. Springer-Verlag, New York.

Slud, E. V. (1984). Sequential linear rank tests for two-sample censored survival data. Ann. Statist. **12**, 551-571.

Slud, E. and Wei, L. J. (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic. J. Amer. Statist. Assoc. **77**, 862-868.

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

Tsiatis, A. A. (1982). Repeated significance testing for a general class of statistics used in censored survival analysis. J. Amer. Statist. Assoc. **77**, 855-861.

Tsiatis, A. A., Rosner, G. L. and Tritchler, D. L. (1985). Group sequential tests with censored survival data adjusting for covariates. Biometrika **72**, 365-373.

Whitehead, J. (1978). Large sample sequential methods with application to the analysis of 2×2 contingency tables. Biometrika **65**, 351-356.

Whitehead, J. (1983). The Design and Analysis of Sequential Clinical Trials. Ellis Horwood Limited and John Wiley, New York.

Department of Mathematics and Statistics, McGill University, Montreal, Quebec, Canada H3A 2K6.

Department of Statistics , Hill Center, Busch Campus, Rutgers University, New Brunswick, NJ 08909, U.S.A.

(Received July 1993; accepted September 1994)