On fitting the proportional hazards model to data from complex surveys

Susana Rubin Bleuer[1]

Abstract

We use the weighted sample partial likelihood score (SPLS) function to fit the proportional hazards regression model to survey data with complex sampling designs. The sample maximum partial likelihood estimator is the solution of the sample partial likelihood score function. Many authors applied this method to fit survival survey data. Binder (1992) dealt with inference on the descriptive census population parameter, that is, design-based inference on the maximum partial likelihood estimate that could be calculated had a census been taken on the finite population. Lin (2000) gave a formal justification of Binder's method under the super-population approach and dealt with inference on the model parameter. Neither Binder nor Lin provided conditions for the respective asymptotic results to hold. Rubin-Bleuer (2003) uses Lin's (2000) set up of the super-population approach and develops counting process methodology for a joint design-model space to obtain, under stated sufficient model and design conditions, a rigorous proof of Binder's approximation of the SPLS. In this paper I prove the weak convergence of the SPLS process and the asymptotic normality of the sample maximum partial likelihood estimator in a formally expressed joint design-model space.

Key words: proportional hazards, complex survey data, counting processes.

1. Introduction

Suppose we want to explore the relationship between the length of spells of unemployment and the covariates education and gender. The Cox (1972) proportional hazards regression model (PHM) provides a method for studying the effects of primary covariates, such as education, on failure times (end of the spells), while adjusting for other variables (e.g., identifiable regional characteristics). If we assume that no covariates vary with time and let $S(t \mid X) = 1 - P(T \le t \mid X)$ be the conditional survival function of the failure time T associated with an r-dimensional vector of covariates X, then the conditional hazard function (or instantaneous conditional failure rate) is defined by

$$\lambda(t \mid X) = \lim_{h \to 0} h^{-1} P(t \le T < t + h \mid T \ge t, X).$$

The PHM specifies that the conditional hazard rate $\lambda(t \mid X)$ of the failure time T satisfies

$$\lambda(t \mid X) = \lambda_0(t) \cdot \exp(\beta' \cdot X),$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and β is an r-dimensional vector valued regression parameter pertaining to the log hazard ratio.

Most methods of survival analysis were developed for independent observations, which we call the "census case" from now on.

The census case

We denote the failure time by T, subject to right censoring given by C. Let $\tilde{T} = \min(T, C)$, $\delta = I(\tilde{T} = T)$ and $Y(t) = I(\tilde{T} \ge t)$, where $I(\cdot)$ is the indicator function. The data consists of $(\tilde{T}_k, \delta_k, X_k)$, $k = 1,..., N$, independent not necessarily identically distributed random variables defined on a probability space $(\Omega, \Im, P)$. Under the PHM β_0 can be estimated from the census partial likelihood score (CPLS) function

$$U(\beta) = \sum_{k=1}^{N} \delta_k \left\{ X_k - \frac{S^{(1)}(\beta, \tilde{T}_k)}{S^{(0)}(\beta, \tilde{T}_k)} \right\}, \qquad (1.1)$$

where the S-functions are sums of independent random vectors:

$$S^{(0)}(\beta, t) = \frac{1}{N} \sum_{k=1}^{N} Y_k(t) \cdot e^{\beta' \cdot X_k}$$ and

$$S^{(1)}(\beta, t) = \frac{1}{N} \sum_{k=1}^{N} X_k \cdot Y_k(t) \cdot e^{\beta' \cdot X_i}.$$

The solution to $U(\beta) = 0$ yields β_N, the maximum partial likelihood estimator of the model parameter β_0. We call β_N the census parameter. Under certain model conditions, the expression $\sqrt{N}(\beta_N - \beta_0)$ is asymptotically normal with zero mean and covariance matrix that can be consistently estimated by $N\, I^{-1}(\beta) = -N\{\partial U / \partial \beta\}^{-1}(\beta_N)$ (Andersen and Gill, 1982).

The survey data case

The data consist of a sequence of units selected from the finite population

$$(\tilde{T}_{hik}, \delta_{hik}, X_{hik}), \quad k = 1,..., n_{hi}, \quad i = 1,..., n_h,$$

1 Susana Rubin-Bleuer, Statistics Canada, Ottawa, Canada, K1H-0T6, rubisus@statcan.ca

$h = 1,...,L, \quad n = n_1 + ... + n_L,$

where some units could be selected more than once, according to the sampling design. For without replacement designs, the data consist of a subset of the finite population.

Fitting the PHM to survey data poses difficulties because survey data consist of dependent observations and is often subject to selection bias due to unequal selection probabilities (see for example, Pfeffermann, 1993). As a result, the usual asymptotic theory does not apply. In order to analyze survey data, and do inference for the parameters of the model, survey samplers think of it as the result of a two-phase procedure (an approach introduced by Hartley and Sielken in 1975), where the infinite population (also called super-population) generates a finite population in the first phase, and the sample is selected from the finite population in the second phase. The finite population could have been completely observed, had we taken a census.

We assume a general stratified without replacement, two-stage design $p_d(s)$ on a finite population obtained from independent failure and censoring times and independent covariates.
We consider a sample estimator of the Census Partial Likelihood Score function, which we call the Sample Partial Likelihood Score (SPLS) function:

$$\hat{U}(\beta) = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{k=1}^{N_{hi}} \delta_{hik} \left\{ X_{hik} - \frac{\hat{S}^{(1)}(\beta, \tilde{T}_{hik})}{\hat{S}^{(0)}(\beta, \tilde{T}_{hik})} \right\} \frac{I_{hik}(s)}{\pi_{hik}}, \tag{1.2}$$

where
(i) I_{hik} are the sample selection indicators and π_{hik} is the inclusion probability for the hik unit of the finite population, that is $p_d(I_{hik} = 1) = \pi_{hik}$;

(ii) the $\hat{S}$-functions are the Horvitz-Thompson sample estimators (Sarndal et al, 1992, p.43) of the census S-functions:

$$\hat{S}^{(0)}(\beta, t) = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{k=1}^{N_{hi}} I(\tilde{T}_{hik} \geq t) e^{\beta' \cdot X_{hik}} \frac{I_{hik}(s)}{\pi_{hik}}$$

and

$$\hat{S}^{(1)}(\beta, t) = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{k=1}^{N_{hi}} I(\tilde{T}_{hik} \geq t) e^{\beta' \cdot X_{hik}} \frac{I_{hik}(s)}{\pi_{hik}}.$$

The solution $\hat{\beta}_N$ of the estimating equation $\hat{U}(\beta) = 0$ is called the Sample Maximum Partial Likelihood estimator (SMPLE. The problem is to determine the inferential properties of the SMPLE.

Three previous papers applied the PHM to data from complex surveys: the first work (Binder, 1992) dealt with inference on the census population parameter β_N; a second paper by (Lin, 2000) dealt with inference on the infinite population (or model) parameter β_0; and a third study on the subject (Rubin-Bleuer, 2003) provided a theoretical justification of an approximation property used by Binder and Lin for their work. Binder (1992) proposed a method of fitting proportional hazards models to survey data from complex designs, based on asymptotic theory in the design probability space. His method provides inference on the "descriptive" census estimator β_N that would be completely known if all the values of the finite population were known. It does not assume a super-population model and it is entirely based on a fixed finite population from which the sample is selected.

Binder (1992) assumes the following approximation to the SPLS function and bases his proof of asymptotic normality of the SMPLE on this:

$$\frac{\hat{U}(\beta)}{\sqrt{N}} = \frac{1}{\sqrt{N}} \sum_{hik} \delta_{hik} \left\{ X_{hik} - \frac{S^{(1)}(\beta, \tilde{T}_{hik})}{S^{(0)}(\beta, \tilde{T}_{hik})} \right\} \cdot \frac{I_{hik}(s)}{\pi_{hik}}$$
$$+ o_{P_d}(1),$$

where p_d in the above equation denotes design probability and $o_{p_d}(1)$ denotes a term that goes to zero in design probability as the sample size goes to infinity. Thus, each term in the approximation is the product of a population value times a sample indicator, and hence the sample partial likelihood score function is asymptotically equivalent to the estimator of a population total. One can then apply existing results on the asymptotic normality of the estimator of a population total to obtain the limiting distribution, in design probability, of the SPLS function and the estimator $\hat{\beta}_N$ of β_N
.

Lin (2000) proposed a method to do inference on the model parameter β_0. He works with the super-population approach of Hartley and Sielken (1975) and showed how the sample maximum partial likelihood estimator $\hat{\beta}_N$, proposed by Binder (1992), can provide inference for the model parameter β_0 if the corresponding variance accounts for both the design and the model variability. Lin (2000) stated that both, the SPLS function and the sample maximum partial likelihood estimator $\hat{\beta}_N$ are asymptotically normal provided that certain sample processes were tight. However, he did not provide neither design nor model conditions under which those sample processes are tight.

Rubin-Bleuer (2003) used the super-population approach working on a joint design-model space in a formal way, and developed counting process methodology for this joint design-model space to obtain a proof of Binder′s and Lin′s conjecture on the approximation of the SPLS function. For a given realization $\omega \in \Omega$ in the model space and a given sample s, the normalized SPLS function is approximated as

$$\frac{\hat{U}(\beta, s, \omega)}{\sqrt{N}} = \frac{1}{\sqrt{N}} \sum_{hik} \delta_{hik} \left\{ X_{hik}(\omega) - \frac{S^{(1)}(\beta, \tilde{T}_{hik}(\omega))}{S^{(0)}(\beta, \tilde{T}_{hik}(\omega))} \right\} \cdot \frac{I_{hik}(s)}{\pi_{hik}}$$

$+o_{P^*}(1)$,

where the subscript P^* in the equation denotes a probability that embraces both the model and the design randomizations. The result requires the assumption of a super-population model which would span the finite population.

In this paper, we obtain a rigorous proof of the weak convergence of the SPLS process in the joint design-model space, the consistency of the maximum partial likelihood estimator $\hat{\beta}_N$ and its asymptotic normality about the model parameter β_0. Sufficient model and design conditions are stated for these results to hold.

The proofs, which follow the same reasoning and techniques of those by Andersen and Gill (1982), Andersen and Borgan (1985) and Naes (1982) for similar results on the census case, are omitted in this paper due to lack of space, but we do present the basic ideas behind the assumptions and the proof in Remarks 4.1 and 4.2 respectively. For a detailed version, we refer the reader to the paper of the same name published by the *Technical Report Series of the Laboratory for Research in Statistics and Probability, Carleton University-University of Ottawa.*

The most serious restriction we have to impose is that the data used must be restricted to an interval $0 \le t \le \tau$. Even if the distribution of the censoring variables has finite support $0 \le t \le C_0$ the value τ would have to be strictly smaller than C_0 and should be chosen in advance. Hence we may not be able to use all the data.

However, in the census case the weak convergence results extend to $\tau = \limsup_{N\to\infty} (\text{support}\{Y(t)\})$, in which case all the data can be used to build the Census Partial Likelihood process. In the immediate future we plan to extend it for the survey case.

In Section 2, I define the proportional hazards super-population model for independent failure and censoring times, and a general stratified without replacement, two-stage design on a finite population obtained from a realization of the super-population. I also formally express the joint design-model space used by Lin (2000) as a "product space" containing both the model and the design probability spaces. In Section 3, I establish notation following the standard for both counting process theory and survey design, and state the approximation result for the partial likelihood score process mentioned above. In Section 4, I state the weak convergence of the partial likelihood score process as well as the asymptotic normality of the maximum partial likelihood estimator, along with the sufficient model and design conditions.

2. The model and the design

2.1 The model

We assume that the vector of covariates X does not depend on time. As a result, simpler model conditions are used for the results we seek. We assume a model of failure times T subject to right censoring denoted by C, defined on a probability space $(\Omega, \Im, P)$. E_m and V_m denote, respectively, the expectation and variance in the space $(\Omega, \Im, P)$. The model is defined by:

$$(\tilde{T}_k^N, \delta_k^N, X_k^N) \quad k = 1,..., N, \qquad (2.1)$$

which are independent triplets of

a) r-dimensional covariates $X_k^N \quad k = 1,..., N$,

b) censored failure times $\tilde{T}_k^N = T_k^N \wedge C_k^N$, where failure time and censoring time variables T_k^N and C_k^N are assumed conditionally independent given X_k^N, $k = 1,..., N$, and

c) indicators of whether a failure time is actually being observed or not, $\delta_k^N = I\left(\tilde{T}_k^N = T_k^N\right), \; k = 1,..., N$.

The Cox (1972) proportional hazards model specifies that the hazard rate $\lambda_k^N(t)$ (or instantaneous failure rate) of the failure time T_k^N satisfies

$$\lambda_k^N(t) = \lambda_0(t)\exp(\beta_0' X_k^N) \quad k = 1,..., N, \qquad (2.2)$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, with absolutely continuous survival function $S_0(t) = 1 - F_0(t)$ and β_0 is an r-dimensional vector valued regression parameter. Thus, even though the failure times T_k^N are not necessarily identically distributed, they share the same baseline hazard function.

2.2 The design

Let us now consider a general stratified, without replacement, two-stage design on a finite population obtained from independent failure and censoring times and independent covariates. For consistency of notation, we re-index the finite population units into L strata, N_h primary sampling units (*psu's*) within stratum h, $h = 1,..., L$, and N_{hi} secondary sampling units within each *psu* i, $i = 1,..., N_h$. We also set $N = \sum_{h=1}^{L} N_h$. Thus, for an outcome $\omega \in \Omega$ of the super-population, the finite population is represented by

$$\left(\tilde{T}_{hik}^N(\omega), \delta_{hik}^N(\omega), X_{hik}^N(\omega)\right),$$

$k = 1,...,N_{hi}$, $i = 1,...,N_h$, $h = 1,...,L$. We let π_{hik}^N denote the probability that the unit hik is selected to sample. For simplicity we omit the superscript N in the notation of the inclusion probabilities. The sample selection indicators are defined by $I_{hik}(s) = 1$, if unit $hik \in s$, and $I_{hik}(s) = 0$ otherwise, $k = 1,..., N_{hi}$, $i = 1,..., N_h$, $h = 1,..., L$.

Denote by S_N the collection of all possible samples under the sample scheme, by $C(S_N)$ be the collection of subsets of S_N, and let p_{dN} be a sampling probability distribution defined on $C(S_N)$. The design space is given by the triplet $(S_N,\ C(S_N),\ p_{dN})$. In the following, E_d and V_d denote, respectively, the expectation and variance with respect to the sampling design.

3. The joint design-model space

The product space determined by the proportional hazards model and the sampling design is given by $(\Omega \times S_N,\ \Im \times C(S_N),\ P_{d,m})$ with probability measure defined in the elementary rectangles by:

$$P_{d,m}(s \times F) = p_{dN}(s) \cdot P(F),\ \ s \in C(S_N),\ \ F \in \Im .$$

see Rubin-Bleuer & Schiopu-Kratina (2002), Example 4.2 for a description of the product space where the model probability is conditional to the prior information. We will use tools of counting process theory adapted to the product space as in Rubin-Bleuer (2003).

3. A representation of the SPLS

In the following, we use the standard notation for counting processes and sample estimators (see, for example, Fleming and Harrington, 1991, chapters 4 and 8 and Särndal et al, 1992 respectively; see Särndal et al, p. 167 for the definition of consistency). From now on, for simplicity of notation, whenever there is no room for confusion, we omit the superscript N.

With $I(A)$ denoting the indicator function of the set A, let

$$\eta(t) = \sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} \eta_{hik}(t),\ \ \eta_{hik}(t) = I(T_{hik} \le t)\delta_{hik}$$

denote the counting process, which is the number of failed uncensored observations by time t. Let the number of units at risk at time t be

$$Y(t) = \sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} Y_{hik}(t),\ \ Y_{hik}(t) = I(\tilde{T}_{hik} \ge t).$$

The symbol $\otimes 2$ denotes the outer product of the vector within brackets (i.e. $X^{\otimes 2} = X \cdot X'$). Also we write $X^{\otimes 1} = X$ and $X^{\otimes 0} = 1$. We define $S^{(j)}$ respectively a scalar, an r-dimensional vector and an $r \times r$ dimensional matrix, and the $S_\pi^{(j)}$ also respectively a scalar, an r-dimensional vector and an $r \times r$ dimensional matrix, and their Horvitz-Thompson sample estimators:

$$S^{(j)}(\beta,t) = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} X_{hik}^{\otimes j} \cdot Y_{hik}(t) \cdot e^{\beta' \cdot X_{hik}},$$

$$\hat{S}^{(j)}(\beta,t) = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} \frac{I_{hik}(s)}{\pi_{hik}} X_{hik}^{\otimes j} \cdot Y_{hik}(t) \cdot e^{\beta' \cdot X_{hik}},$$

$$S_\pi^{(j)}(\beta,t) = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} \frac{1}{\pi_{hik}} X_{hik}^{\otimes j} \cdot Y_{hik}(t) \cdot e^{\beta' \cdot X_{hik}},$$

and

$$\hat{S}_\pi^{(j)}(\beta,t) = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} \frac{I_{hi}(s)}{\pi_{hik}^2} X_{hik}^{\otimes j} \cdot Y_{hik}(t) \cdot e^{\beta' \cdot X_{hik}},$$

$j = 0,1,2.$

Also let $e(\beta,t) = \dfrac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}$

and $\hat{e}(\beta,t) = \dfrac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}$.

We define the sample partial likelihood score vector by:

$$\hat{U}(\beta,t) = \sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} \int_0^t \{X_{hik} - \hat{e}(\beta,u)\} d\eta_{hik}(u) \frac{I_{hik}(s)}{\pi_{hik}}.$$

The SPLS function of equation (1.2) is a special case of the SPLS process with $\hat{U}(\beta) = \hat{U}(\beta,\infty)$.

The process $\hat{U}(\beta,t)$ has a martingale representation with respect to the filtration $\{\Im_t^{d,m} = C(S_N) \times \Im_t,\ t \ge 0\}$ where $\Im_t$ is the sigma field defined by the failure and censoring indicators (see Rubin-Bleuer, 2003):

$$\hat{U}(\beta,t) = \sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} \int_0^t \{X_{hik} - \hat{e}(\beta,u)\} dM_{hik}(u) \frac{I_{hik}(s)}{\pi_{hik}}$$

where

$$M_{hik}(t) = \eta_{hik}(t) - \int_0^t Y_{hik}(u)\exp^{\beta' X_{hik}} \cdot \lambda_0(u)du .$$

We state Theorem 3.1 and Corollary 3.1 shown in Rubin-Bleuer (2003).

Theorem 3.1 (Rubin-Bleuer, 2003). We assume the proportional hazards model, the sampling design stated in Section 2, and the following:

(A.1) The covariate vectors X_{hik} are time-invariant and bounded: $\sup_{h,i,k}\left|X_{hik}\right| \le B \ \ a.s. \ \ as \ \ N \to \infty$.

(A.2) There exists a neighborhood Λ of β_0 and, respectively, scalar, vector and matrix functions $s^{(0)}$, $s^{(1)}$ and $s^{(2)}$ defined on $\Lambda \times \lceil 0, \tau \rceil$ such that for $j = 0,1,2,$ and for $0 \le t \le \tau, \ \beta \in \Lambda$ we have:

$$s^{(j)}(\beta,t) = \lim_{N\to\infty} \frac{1}{N} \sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} E_m\left\{ X_{hik}^{\otimes j} \cdot Y_{hik}(t) \cdot e^{\beta' X_{hik}} \right\},$$

and

$$s^{(j)}(\beta,t+) = \lim_{N\to\infty} \frac{1}{N} \sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} E_m\left\{ X_{hik}^{\otimes j} \cdot Y_{hik}(t+) \cdot e^{\beta' X_{hik}} \right\}$$

(A.3) The sample estimators $\hat{S}^{(j)}$, $j = 0,1,2,$ are the Horvitz-Thompson sample estimators.

(A.4) $C_0: \ f = \lim_n n/N > 0 \ \ as \ \ n \to \infty$.

$$C_1: \ \lim_N \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} \frac{1}{\pi_{hik}} < \infty \ .$$

Then the following holds:

1) $S^j(\beta,t) \xrightarrow{P} s^j(\beta,t)$, and

$S^j(\beta,t+) \xrightarrow{P} s^j(\beta,t+), \ \ as \ \ N \to \infty, \ \ j = 0,1,2,$

for each $t \ge 0$,

2) $\sup_{0\le t\le\tau, \ \beta\in\Lambda} | \ S^j(\beta,t) - s^j(\beta,t) \ | \xrightarrow{P} 0 \ \ as \ \ N \to \infty, \ \ j = 0,1,2,$

for each $t \ge 0$, and

3) $\sup_{0\le t\le\tau, \ \beta\in\Lambda} | \ \hat{S}^j(\beta,t) - S^j(\beta,t) \ | \xrightarrow{P_{d.m}} 0 \ \ \ j = 0,1,2 \ \ as \ \ n \to \infty$.

Corollary 3.1 (Rubin-Bleuer, 2003). Approximation for the sample partial likelihood score process under the proportional hazards model. We consider the SPLS as a process in the product space, where both the sample s and the outcome $\omega \in \Omega$ of the model variables are random. We assume that the hazard function is integrable in $0 \le t \le \tau$. Under the conditions of Theorem 3.1 we have:

$$\frac{\hat{U}(\beta,t)}{\sqrt{N}} = \frac{1}{\sqrt{N}} \sum_{hik} 1_0^t \{X_{hik} - e(\beta,u)\} dM_{hik}(u) \frac{I_{hik}(s)}{\pi_{hik}} = o_{P_{d,m}}(1)$$

independently of t.

4. Weak convergence of the SPLS

We denote by $D[0,\tau]$ the space of right continuous functions on the interval $0 \le t \le \tau$ which have finite left-hand limits at each point of $0 \le t \le \tau$. We write $D[0,\tau]^r = D[0,\tau] \times D[0,\tau] \times ... \times D[0,\tau]$.

Theorem 4.1. Assume the conditions of Theorem 3.1 (A.1-A.4) and the following:

(A.5) The time τ is such that $\int_0^\tau \lambda_0(t)dt < \infty$.

(A.6) There exists a neighborhood Λ of β_0 and, respectively, scalar, vector and matrix functions $s_\pi^{(0)}$, $s_\pi^{(1)}$ and $s_\pi^{(2)}$ defined on $\Lambda \times [0,\tau]$ such that for $j = 0,1,2,$ and for $0 \le t \le \tau, \ \beta \in \Lambda$ we have:

$$s_\pi^{(j)}(\beta,t) = \lim_{N\to\infty} \frac{1}{N} \sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} E_m\left\{ X_{hik}^{\otimes j} \cdot Y_{hik}(t) \cdot e^{\beta' X_{hik}} \right\} / \pi_{hik} .$$

and

$$s_\pi^{(j)}(\beta,t+) = \lim_{N\to\infty} \frac{1}{N} \sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} E_m\left\{ X_{hik}^{\otimes j} \cdot Y_{hik}(t+) \cdot e^{\beta' X_{hik}} \right\} / \pi_{hik}$$

(A.7) $\liminf_{N\to\infty} \frac{1}{N} \sum_{hik} P(C_{hik} \ge \tau) \ \rangle \ 0$.

(A.8) The limiting covariance matrix of the SPLS, $\Sigma_\pi(\beta_0,\tau) = \lim_{N\to\infty} V_{d,m}(\frac{1}{\sqrt{N}}\hat{U}(\beta,\tau))$ is positive definite.

Then

1. The normalized vector sample partial likelihood score process (SPLS) $\{\frac{1}{\sqrt{N}}\hat{U}(\beta,t)): 0 \le t \le \tau\}$ whose value at time t is

$$\frac{\hat{U}(\beta_0,t)}{\sqrt{N}} = \frac{1}{\sqrt{N}} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{k=1}^{N_{hi}} \frac{I_{hik}(s)}{\pi_{hik}} \int_0^t \{X_{hik} - \hat{e}(\beta_0,u)\} d\eta_{hik}(u),$$

converges weakly in $D[0,\tau]^r$ to mean zero r-dimensional Gaussian process such that
a) each component process has independent increments, and
b) the covariance function at t for components ℓ and ℓ' is

$$\Sigma_\pi(\beta_0,t)_{\ell,\ell'} = \int_0^t v_\pi(\beta_0,u)\lambda_0(u)du,$$

$$v_\pi(\beta_0,u) = \left\{ s^{(2)}_{\pi,\ell,\ell'} - \frac{s^{(1)}_{\pi,\ell} \cdot s^{(1)'}_{\ell'}}{s^{(0)}} - \frac{s^{(1)}_{\ell} \cdot s^{(1)'}_{\pi,\ell'}}{s^{(0)}} - \frac{s^{(1)}_{\ell} \cdot s^{(1)'}_{\ell'}}{s^{(0)2}} s^{(0)}_\pi \right\}$$

where all the s-functions in the integrand above are evaluated at (β_0,u).

2. For each t, $0 \le t \le \tau$, $\hat{\beta}_N(t)$, the solution of $\hat{U}(\beta,t) = 0$, is a consistent estimator of β_0.

3. Let $I(\beta,t) = \frac{-1}{N}(\partial\hat{U}/\partial\beta)(\beta,t)$,

$\Sigma_m(\beta,t) = \int_0^t v(\beta,u)s^{(0)}(\beta,u)\lambda_0(u)du$, with

$v(\beta_0,t) = \frac{s^{(2)}s^{(0)} - s^{(1)} \cdot s^{(1)'}}{(s^{(0)})^2}$. Then

$\sup_{0\le t\le\tau} \left| I(\hat{\beta}_N(t),t) - \Sigma_m(\beta_0,t) \right| \xrightarrow{P_{d,m}} 0$ as $n \to \infty$.

We note that in the census case, the expectation of the information matrix $-\frac{1}{N}(\partial\hat{U}/\partial\beta)$ coincides with the limiting covariance matrix of the census partial likelihood score process. However in the survey data case, the information matrix converges to the same limit $\Sigma_m(\beta,t)$ as that of the census case, but this is not the limiting covariance matrix of the sample partial likelihood score process:

$$I(\beta_0,t) = \lim_{N\to\infty} -\frac{1}{N}(\partial\hat{U}/\partial\beta) \neq \Sigma_\pi(\beta_0,t).$$

Remark 4.1 (On the assumptions)

In the census case, there exist results for more general model assumptions than those we present here. We work with somewhat restricted conditions (i.e., continuous failure times, covariates constant over time and uniformly bounded, and conditionally independent failure and censoring times) in order to stress and to concentrate on the added complexity of the survey process. However, we trust that most practical situations occurring with survey data, fall under these conditions.

The model conditions required for Theorems 3.1 and 4.1 are not too different from the conditions imposed by Fleming and Harrington (1991), Theorem 8.4.1 in their simplified version of weak convergence in the census case. A new family of weighted S_π-functions is introduced, because they appear in the calculation of the variance-covariance matrix of the Sample Partial Likelihood Score process. We do require point-wise convergence of the expected values of the S- and S_π-functions because they are sums of independent, but not identically distributed random vectors and matrices, and thus this requirement together with independence yields the point-wise convergence of the S- and S_π-functions by the Weak Law of Large Numbers. In the Fleming and Harrington (1991) version (Theorem 8.4.1) the assumption is of identically distributed random triplets $(\tilde{T}_k^N, \delta_k^N, X_k^N)$ $k = 1,...,N$, which automatically yields point-wise convergence of the random vectors by the Strong Law of Large Numbers.
Also, Condition (A.7) is used to show that the function $s^{(0)}(\beta,t)$ is bounded away from zero in the set $\Lambda \times [0,\tau]$. If the censoring random variables were identically distributed we would only need that for some hik, $P(C_{hik} \ge \tau) \rangle 0$.

We also require design-based conditions, that is, the sample estimators are the Horvitz-Thompson estimators (for a definition see Särndal et al., 1992). This condition may be relaxed to admitting any sample estimators that are asymptotically design unbiased and design consistent. Under with replacement designs the requirement would be that the sample estimators be (asymptotically) design unbiased and design consistent. The two other design conditions are given by (A.4). Condition C_0 in (A.4) ensures that the relationship between the sample and the population sizes (and its impact on the statistics considered) remains the same as we increase the population size towards infinity. Thus $N \to \infty$ if and only if $n \to \infty$. Condition C_1 in (A.4) means that as $N \to \infty$ the sizes are approximately of the same magnitude.

Remark 4.2 (On the proof)

As we said in the introduction, the most serious restriction we have to impose is that the data used must be restricted to an interval $0 \le t \le \tau$ and hence we may not be able to use all the data.

Even though the asymptotic theory developed by Andersen and Gill (1982) for the census case is intended for quite general sequences of probability spaces, we do not apply it directly. The Sample Partial Likelihood process "lives" in a joint design-model space, and the design element induces some stochastic dependencies that cannot be accommodated by the results for independent random vectors. Thus we work from "scratch" using the Central Limit Theorem for Martingales, and most other tools developed by Andersen and Gill (1982) subsequently used by Fleming and Harrington (1991).

An interesting property is worth remarking. The covariance matrix of the SPLS process is simple to calculate. Indeed, even though the two-stage without replacement design introduces dependency in the sampled units, and thus in the joint design-model space, the super-population triplets

$(\tilde{T}_{hik}, \delta_{hik}, X_{hik}),\ \ k = 1,..., N_{hi},$

$i = 1,..., N_h,\ \ h = 1,..., L,$

remain stochastically independent in the model space. By Corollary 3.1, the covariance matrix of the normalized SPLS process is equivalent of the covariance matrix of a process of the form:

$$\frac{1}{\sqrt{N}}\tilde{U}(\beta, t) = \frac{1}{\sqrt{N}}\sum_{hik}\frac{I_{hik}(s)}{\pi_{hik}}\int_0^t \{X_{hik} - e(\beta, u)\}dM_{hik}(u)\cdot$$

If we set

$Z(hik, h'i'k') = \frac{1}{N}(X_{hik} - e(\beta, u))(X_{h'i'k'} - e(\beta, u))'$, the variance with respect to the model and the design has cross product terms of the form

$$\int_0^t E_d\left\{\frac{I_{hik}I_{h'i'k'}}{\pi_{hik}\pi_{h'i'k'}}\right\}E_m\{Z(hik, h'i'k')d\langle M_{hik}, M_{h'i'k'}\rangle(u)\},$$

and due to the stochastic independence of the census units, the predictable co-variation processes between two different units are zero:

$\langle M_{hik}, M_{h'i'k'}\rangle(u) = 0$ if $h \neq h',\ \ i \neq i'$ or $k \neq k',$

hence the cross-product terms in the variance of the sample processes become zero.

Finally, we note that contrary to Lin's supposition with respect to the proof of asymptotic normality of the maximum sample partial likelihood estimator, we do not require Binder's approximation proved in Corollary 3.1 for obtaining the asymptotic results in Theorem 4.1 nor in Corollary 4.1. We prove the weak convergence of $\{\frac{1}{\sqrt{N}}\hat{U}(\beta, t)) : 0 \le t \le \tau\}$ directly.

Corollary 4.1

Under the same conditions of Theorem 4.1, the solution $\hat{\beta}_N(t)$ be the solution of the SPLS estimating equation evaluated at t, which we call maximum sample partial likelihood estimator, is asymptotically normal and

$$\sqrt{N}\left(\hat{\beta}_N(t) - \beta_0\right) \quad \Rightarrow$$

$N(0, I^{-1}(\beta_0, t)\Sigma_\pi(\beta_0, t)I^{-1}(\beta_0, t))$

where

$I(\beta_0, t) = \int_0^\tau v(\beta_0, t)s^{(0)}(\beta_0, t)\lambda_0(t)dt$,

and $\Sigma_\pi(\beta_0, t)$ is as defined in 1 b) above.

5. References

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

Andersen, P. K. and Gill, R. D. (1982), Cox′s regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-20.

Billingsley, P.(1979), *Probability and Measure*, Wiley, New York.

Binder, D. A.(1992), Fitting Cox′s proportional hazards models from survey data. *Biometrika* **79**, 139-147.

Chung, K. L. (1974). *A course in probability theory*. Academic Press, New York and London, second ed.

Cox, D.A. (1992). Regression models and life-tables (with Discussion). *J.R. Statist. Soc*. B **34**, 187-220.

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

Hartley, H.O.& Sielken, R. L.(1975), A "super-population viewpoint" for finite population sampling, *Biometrics* **31**, 411-422

Lin, D.Y.(2000), On fitting Cox′s proportional hazards models to survey data. *Biometrika* 8**7**, 37-47.

Naes, T. (1982). The asymptotic distribution of the estimator for the regression parameter in Cox's regression model. *Scandinavian J. Statist.* **9**, 107-15.

Pfeffermann, D.(1993), The role of Sampling weights when modeling survey data. *International Statistical Review* **61**, 317-337.

Rubin Bleuer, S. (2004). On fitting the proportional hazards model to data from complex surveys. *Technical Report Series of the Laboratory for Research in Statistics and Probability Carleton University-University of Ottawa (to appear).*

Rubin Bleuer, S. (2003 a).On the weak convergence of the sample empirical distribution process. *Statistics Canada Working Paper Series BSMD – 2003 -003E.*

Rubin Bleuer, S. (2003 b). An approximation of the partial likelihood score in a joint design-model space. *Proceedings of the Survey Methods Section, Statistical Society of Canada Annual Meeting, 2003.*

Shao, J. (1999), *Mathematical Statistics*, Springer-Verlag, New York.

Rubin Bleuer, S. and Schiopu Kratina, I. (2002), On the two-phase framework for joint model and design-based inference. *Technical Report Series of the Laboratory for Research in Statistics and Probability, Number 382, Carleton University-University of Ottawa.*

Rubin Bleuer, S. (2001). A test for survival distributions using data from a complex sample. *Proceedings of the Survey Methods Section, SSC Annual meeting, 2001,*103-110.

Särndal, C-E, Swensson, B. and Wretman, J. (1992), *Model Assisted Survey Sampling*, Springer-Verlag, New York.