SSC Annual Meeting, June 2003
Proceedings of the Survey Methods Section

AN APPROXIMATION OF THE PARTIAL LIKELIHOOD SCORE IN A JOINT DESIGN-MODEL SPACE

Susana Rubin Bleuer[1]

ABSTRACT

An approximation to the Sample Partial Likelihood Score (SPLS) in design probability was first proposed by Binder (1992) and then considered by Lin (2000), in order to derive asymptotic theory for the parameters of the proportional hazards regression model. However they did not provide conditions under which this property held. In this paper I use Lin's (2000) set up of the super-population approach and develop counting process methodology for a joint design-model space, to obtain, under sufficient model and design conditions, a rigorous proof of the approximation to the SPLS proposed by Binder (1992).

KEY WORDS: Complex survey data, counting process , joint design-model space, proportional hazards model, tightness.

RÉSUMÉ

Une approximation au Score de vraisemblance partielle de l'échantillon (SVPE) utilisant les probabilités provenant du plan de sondage a initialement été proposée par Binder (1992), et considérée par Lin (2000) pour dériver la théorie asymptotique des paramètres d'un modèle de risques proportionnels. Ils n'ont cependant pas fourni les conditions nécessaires pour l'existence de cette propriété. Dans cet article, j'utilise l'approche d'une super-population de Lin (2000) pour développer la méthodologie d'un processus de comptage pour une espace mixte, tenant compte du plan de sondage et du modèle, pour obtenir, sous certaines conditions pour le modèle et pour le plan de sondage, une preuve de l'approximation au SVPE proposée par Binder (1992)

MOTS CLÉS : Données provenant d'enquêtes à plan complexe; espace mixte tenant compte du plan de sondage et du modèle; modèle de risques proportionnels; processus de comptage; robustesse

1. INTRODUCTION

Suppose we want to explore the relationship between the length of spells of unemployment and 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). The PHM specifies that the hazard rate $\lambda(t)$ (or instantaneous failure rate) of the failure time T associated with a vector of possible time varying covariates X satisfies

$$\lambda(t) = \lambda_0(t) \cdot \exp(\beta^{'} \cdot X(t)),$$

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.

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

---

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

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

where

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

The solution to $U(\beta) = 0$ yields the census parameter β_N. Under mild conditions, $\sqrt{N}(\beta_N - \beta_0)$ is asymptotically normal with zero mean, where β_0 is the true value of β, and with a covariance matrix that can be consistently estimated by $N\ I^{-1}(\beta) = -N\{\partial U / \partial \beta\}^{-1}(\beta_N)$, (Andersen and Gill, 1982, also Fleming and Harrington, 1991, Theorems 8.2.1 and 8.3.2, 1991, p. 290-299, for non-identically distributed censored failure times.)

Fitting the PHM to survey data poses difficulties because survey data are often subject to selection bias due to unequal selection probabilities or consist of dependent observations due to clustering (Pfeffermann, 1993). As a result, the usual asymptotic theory does not apply. Two different methods were proposed for the application of the PHM to data from complex surveys: one developed by Binder (1992) for inference on the census population parameters β_N and the other by Lin (2000) for inference on the infinite population (or model) parameter β_0.

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. He considered the census Partial Likelihood Score (PLS) function $U(\beta)$. Once a sample of size n is obtained from the finite (census) population, Binder defined the Sample Partial Likelihood Score (SPLS) function as a design unbiased estimator of the Census Partial Likelihood Score function (CPLS):

$$\hat{U}(\beta) = \sum_k \frac{1}{\pi_k} \delta_k \left\{ X_k(\widetilde{T}_k) - \frac{\hat{S}^{(1)}(\beta, \widetilde{T}_k)}{\hat{S}^{(0)}(\beta, \widetilde{T}_k)} \right\} \cdot I_k(s),$$

where the $I_k(s)$ are sample selection indicators and the π_k are the probabilities of selection, $k = 1,...,N$. The estimator $\hat{\beta}_N$ of the census parameter β_N is obtained as the solution of $\hat{U}(\beta) = 0$. Binder (1992) conjectured the following approximation to the SPLS function:

$$\frac{\hat{U}(\beta)}{\sqrt{N}} = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \delta_k \left\{ X_k(\widetilde{T}_k) - \frac{S^{(1)}(\beta, \widetilde{T}_k)}{S^{(0)}(\beta, \widetilde{T}_k)} \right\} \cdot I_k(s) / \pi_k + o_{p_d}(1), \tag{1.1}$$

where p_d in the above equation denotes design probability and $o_{p_d}(1)$ denotes that the term goes to zero in design probability as the sample size goes to infinity. The approximation (1.1) enables the application of 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. This method does not assume a super-population model and is entirely based on a fixed finite population from which the sample is observed. This 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.

Lin (2000) followed the super-population approach introduced by Hartley and Sielken (1975): the finite population is assumed to be a random sample of size N drawn from the infinite population following a specified model, and the sample of size n obtained from the finite population via a sampling design in the second phase of a two-phase sampling process. Lin (2000) showed how the sample 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 randomizations. Lin (2000) viewed the SPLS function as a stochastic process depending on the model and the sample design. The approximation to the SPLS takes the form

$$\frac{\hat{U}(\beta,s,\omega)-U(\beta,\omega)}{\sqrt{N}} = \frac{1}{\sqrt{N}}\sum_{k=1}^{N}\delta_k\left\{X_k(\tilde{T}_k)-\frac{S^{(1)}(\beta,\tilde{T}_k)}{S^{(0)}(\beta,\tilde{T}_k)}\right\}\cdot\left(I_k(s)/\pi_k-1\right)+o_{P*}(1)\cdot \quad (1.2)$$

The subscript $P*$ in equation (1.2) denotes a probability that embraces both the model and the design randomizations. Lin (2000) also described some asymptotic properties of the SPLS function stating, among other results, Binder's approximation of the sample partial likelihood score function, which would hold subject to certain sample processes being tight. However, he did not provide design or model conditions under which those sample processes are tight.

In this paper, I use the super-population approach of Lin (2000) and counting process methodology for a joint design-model space, to obtain a rigorous proof of the approximation (1.1) to the SPLS function, under sufficient model and design conditions. Even though the result is ultimately in design probability, in order to obtain it, I need to assume a super-population spanning the finite population from which the sample is selected. The asymptotic normality (in design probability) of the sample partial likelihood score function and the estimator $\hat{\beta}_N$ follow from the arguments of Binder (1992).

In Section 2, I define a stratified super-population model and design, and formally express the joint design-model space used by Lin (2000) as a "product space" containing both the model and the design probability spaces (see also Rubin Bleuer and Ioana Schiopu Kratina, 2002). In Section 3, I further develop the counting process methodology stated in Rubin Bleuer (2001) for the design-model space, which is used to derive the variance of sample processes with a martingale representation. Many survival processes are linear functions of the number of units at risk at time t, or of the individual indicators of whether a unit is at risk or not at time t (decreasing in t). Sample survival processes are sample estimators of these functions and tightness often follows from the convergence in sup norm of these functions. In Section 4, I present Lemma 4.1 which states the sup-norm convergence in probability of a sequence of bounded monotonic random functions. One corollary to Lemma 4.1 is the Glivenko-Cantelli theorem for the joint design-model space (sup-norm convergence of the sample empirical distribution functions, see Rubin Bleuer, 2003), and this in turn yields tightness of some weighted log-rank statistics in such a space (Rubin Bleuer, 2001). Of interest to this article is another consequence of the lemma, Theorem 4.1, the sup-norm convergence of functions of the risk indicators involved in the partial likelihood score under the proportional hazards model. I apply the counting process approach to the "product space" to obtain the approximation of the sample partial likelihood score function as a corollary of this property.

Though it is beyond the scope of this paper, this methodology can be further applied to obtain the weak convergence of the SPLS process in the joint design-model space envisaged by Lin(2000), and the asymptotic normality $\hat{\beta}_N$ (as an estimator of β_0) in the product space.

2. THE MODEL AND THE DESIGN

2.1 The model

In this paper, the main result assumes that the covariates X do not depend on time. As a result, simpler model conditions are used for the results we seek. We assume a stratified super-population of censored failure times T subject to right censoring denoted by C, with L strata and N_h units in stratum h, $h=1,...,L$, $N=\sum_{h=1}^{L}N_h$, defined on a probability space $(\Omega,\ \Im,\ P)$. It is defined by the N random vectors (see Rubin-Bleuer & Schiopu-Kratina (2002) for a definition of the super-population in this context):

$$(\tilde{T}_{hi}^{N},\delta_{hi}^{N},X_{hi}^{N})\ \ i=1,...,N_h,\quad h=1,...,L, \qquad (2.1)$$

which are independent triples of

a) r-dimensional covariates X_{hi}^{N} $i=1,...,N_h$, $h=1,...,L$,

b) censored failure times $\tilde{T}_{hi}^N = T_{hi}^N \wedge C_{hi}^N$, where failure time and censoring time variables T_{hi}^N and C_{hi}^N are assumed conditionally independent given X_{hi}^N, and

c) indicators of whether a failure time is actually being observed or not, $\delta_{hi}^N = I\left(\tilde{T}_{hi}^N = T_{hi}^N\right)$.

The failure times T_{hi}^N are assumed identically distributed within strata, but not across strata. The censoring times C_{hi}^N are not necessarily identically distributed within and across strata.

The Cox (1972) proportional hazards model specifies that the hazard rate $\lambda_{hi}^N(t)$ (or instantaneous failure rate) of the failure time T_{hi}^N satisfies

$$\lambda_{hi}^N(t) = \lambda_0(t) \cdot \exp(\beta_0{}' \cdot X_{hi}^N), \tag{2.2}$$

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_{hi}^N are not necessarily identically distributed across strata, they share the same baseline hazard function.

Let $\eta_{hi}(t) = I(T_{hi}^N \le t)\delta_{hi}$, which takes the value 1 if unit hi is uncensored and failed before time t, and 0 otherwise. Also, let $\eta_{hi}^C(t) = I(C_{hi}^N \le t)(1 - \delta_{hi})$, and let

$$\eta(t) = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \eta_{hi}(t) \tag{2.3}$$

denote the counting process, which is the number of failed uncensored observations by time t. We use the notation $\eta(t)$ for a counting process, instead of the usual $N(t)$ because in this study N denotes the size of the finite population. Let $Y_{hi}(t) = I(\tilde{T}_{hi}^N \ge t), i = 1,..., N_h, h = 1,..., L$ be the indicator of whether unit hi is at risk at time t and let the number of units at risk at time t be $Y(t) = \sum_{h=1}^{L} \sum_{i=1}^{N_h} Y_{hi}(t)$. (2.4)

We omit the superscript N in the processes $\eta(\cdot)$, $Y(\cdot)$ and martingales $M(\cdot)$, $M_{hi}(\cdot)$ defined below, for simplicity of notation. The following definitions, notation, counting process set-up, and basic results can be found in Fleming and Harrington (1991) chapters 4 and 8. 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 $X^{\otimes 0} = 1$. E_m and V_m denote, respectively, the expectation and variance in the model probability space $(\Omega, \Im, P)$. Let $S^{(j)}$, $j = 0,1,2$ be respectively, a scalar, an r-dimensional vector and an $r \times r$ dimensional matrix defined by:

$$S^{(0)}(\beta,t) = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} Y_{hi}(t) \cdot e^{\beta' \cdot X_{hi}^N} = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} X_{hi}^{\otimes 0} \cdot Y_{hi}(t) \cdot e^{\beta' \cdot X_{hi}^N}$$

$$S^{(1)}(\beta,t) = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} X_{hi}^N \cdot Y_{hi}(t) \cdot e^{\beta' \cdot X_{hi}^N} = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} X_{hi}^{\otimes 1} \cdot Y_{hi}(t) \cdot e^{\beta' \cdot X_{hi}^N}, \tag{2.5}$$

$$S^{(2)}(\beta,t) = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} X_{hi}^N X_{hi}'^N \cdot Y_{hi}(t) \cdot e^{\beta' X_{hi}^N} = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} X_{hi}^{\otimes 2} \cdot Y_{hi}(t) \cdot e^{\beta' \cdot X_{hi}^N}.$$

Also let $e(\beta,t) = \dfrac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}$. (2.6)

The process $M_{hi}(t) = \eta_{hi}(t) - \int_0^t Y_{hi}(u)\exp\left\{\beta' X_{hi}^N\right\} \cdot \lambda_0(u)du$ is a martingale with respect to $\{\Im_t : t \geq 0\}$(see Fleming and Harrington (1991) chapter 4) . We define a score process and work with results pertaining to the process rather than with the score statistic, since they are more general. The partial likelihood score (PLS) process is defined by

$$U^N(\beta,t) = \sum_{h=1}^{L}\sum_{i=1}^{N_h}\int_0^t \left\{X_{hi}^N - e(\beta,u)\right\} d\eta_{hi}(u) = \sum_{h=1}^{L}\sum_{i=1}^{N_h}\int_0^t \left\{X_{hi}^N - e(\beta,u)\right\} dM_{hi}(u) = \sum_{h=1}^{L}\sum_{i=1}^{N_h} U_{hi}^N(\beta,t). \quad (2.7)$$

Hence, the PLS function is $U(\beta) = U(\beta,\infty)$ (Fleming and Harrington, 1991). Thus $U^N(\beta,t)$ has the martingale representation needed for the counting process approach: with an appropriate filtration and simple regularity conditions, it is a stochastic integral of a predictable process with respect to a martingale and $E_m\left\{U^N(\beta,t)\right\} = 0,\ 0 \leq t \leq \infty.$.

2.2 The design

Let us now consider a stratified one-stage design with expected sample size equal to n, defined on a finite population that is an outcome of the super-population defined above: $\left(\tilde{T}_{hi}^N(\omega), \delta_{hi}^N(\omega), X_{hi}^N(\omega)\right)$, $i = 1,..., N_h$, $h = 1,..., L$. The prior information is given by values $\{z_{hi}, i = 1,..., N_h, h = 1,..., L\}$ that are known at the time of the design. These values could be outcomes of super-population random vectors originally associated with the failure and censoring times given above (see Rubin-Bleuer & Schiopu-Kratina (2002) Example 3.1, for a description of the super-population model with prior information given before it is built). Let S_N be the collection of all possible samples under the sample scheme, $C(S_N)$ be the collection of subsets of S_N, and let p_{dN} be a sampling probability distribution defined on $C(S_N)$,

$$p_{dN}(s,z) = p_{dN}\left(s, z_{11},..., z_{1N_1},..., z_{hi},..., z_{LN_L}\right).$$

We consider here a few of the more common sampling designs along with conditions on the designs as the finite population size $N \to \infty$. An overview of these designs can be found in Särndal et al (1992). Under stratified simple random sample without replacement (SRSWOR), Poisson or πps one-stage designs, we let $\pi_{hi} = \pi_{hi}(z)$ denote the probability that unit i in stratum h is selected to sample. Let $I_{hi}(s) = 1$ if $i \in s$, $I_{hi}(s) = 0$ otherwise, $i = 1,..., N_h$, $h = 1,..., L$. Thus I_{hi} follows a binomial distribution:

$$I_{hi} \sim B(1, \pi_{hi}) \quad i = 1,..., N_h, \quad h = 1,..., L. \quad (2.8)$$

Under a probability proportional to size with replacement (pps) sampling scheme with units "sizes" z_{hi} and selection probabilities $p_{hi} = z_{hi} / \sum_{l=1}^{N_h} z_{hl}$, we define $0 \leq I_{hi}(s) \leq n$, $i = 1,..., N_h, h = 1,..., L$ as the number of times unit i in stratum h is selected to the sample. Thus the vector $\left(I_{h1}(s),..., I_{hN_h}(s)\right)$ follows a multinomial distribution:

$$\left(I_{h1}(s),..., I_{hN_h}(s)\right) \sim MN\left(n_h, p_{h1},..., p_{hN_h}\right), \quad h = 1,..., L. \quad (2.9)$$

Remark 2.1 Traditional notation under the proportional hazards model use S for the random functions and their deterministic limits: $S^{(j)}(\beta,t,\omega) \to s^{(j)}(\beta,t)$ $j = 0,1,2$. For $\omega \in \Omega$ the scalar, vector and matrix functions $S^{(j)}(\beta,t)$, $j = 0,1,2$ are finite population parameters. Their sample estimators of depend on the sample $s \in S_N$: $\hat{S}^{(j)}(\beta,t,s,\omega)$ $j = 0,1,2$. The use s to denote an outcome of the sample design is also a well known convention in survey theory, and we will do so here with the caveat that the sample s not be confused with the deterministic limit functions $s^{(j)}(\beta,t)$ $j=0,1,2$. Design-consistent estimators of the S functions (Särndal et al (1992), p.167) are given by:

$$\hat{S}^{(j)}(\beta,t)=\frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{1}{\pi_{hi}}X_{hi}^{\otimes j}\cdot Y_{hi}(t)\cdot e^{\beta'\cdot X_{hi}^{N}}I_{hi}(s),\quad j=0,1,2. \tag{2.10}$$

where the $I_{hi}(s)/\pi_{hi}$ are as in (2.8) or in (2.9); for pps with replacement , we replace π_{hi} by np_{hi} and recall that in this case $0\le I_{hi}(s)\le n$ follows a different sample distribution. The sample partial likelihood score vector has the form:

$$\hat{U}^{N}(\beta,t)=\frac{1}{\sqrt{N}}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{1}{\pi_{hi}}\int_0^t\left\{X_{hi}^{N}-\hat{e}(\beta,u)\right\}d\eta_{hi}(u)\cdot I_{hi}(s), \text{ with } \hat{e}(\beta,t)=\hat{S}^{(1)}(\beta,t)/\hat{S}^{(0)}(\beta,t).$$

$\hat{U}^{N}(\beta,t)$ has also a martingale representation: indeed, the expression denoted by $\hat{A}(\beta,t)$ is equal to zero:

$$\frac{1}{\sqrt{N}}\hat{A}^{N}(\beta,t)=\frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{1}{\pi_{hi}}\int_0^t\left\{X_{hi}^{N}-\hat{e}(\beta,u)\right\}\cdot Y_{hi}(u)\cdot e^{\beta' X_{hi}^{N}}\lambda_0(u)du\cdot I_{hi}(s)=\int_0^t(\hat{S}^{(1)}-\frac{\hat{S}^{(1)}}{\hat{S}^{(0)}}\cdot\hat{S}^{(0)})\cdot\lambda_0(u)\cdot du=0\,.$$

We subtract the zero expression $\hat{A}^{N}(\beta,t)$ from $\hat{U}^{N}(\beta,t)$ to obtain:

$$\hat{U}^{N}(\beta,t)=\frac{1}{\sqrt{N}}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{1}{\pi_{hi}}\int_0^t\left\{X_{hi}^{N}-\hat{e}(\beta,u)\right\}dM_{hi}(u)\cdot I_{hi}(s)=\sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{1}{\pi_{hi}}\hat{U}_{hi}(\beta,t)I_{hi}(s),$$

where $\hat{U}_{hi}^{N}(\beta,t)=\int_0^t\left\{X_{hi}^{N}-\hat{e}(\beta,u)\right\}dM_{hi}(u)$. (2.11)

In the following E_d and V_d denote, respectively, the expectation and variance with respect to the sampling design. We assume the following regularity conditions on the sampling design:

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

C_1 : $\lim_N \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{1}{np_{hi}}<\infty$ for stratified pps one stage designs.

C_2 : $\lim_N \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{1}{\pi_{hi}}<\infty$ for stratified SRSWOR, Poisson or πps one-stage designs.

Remark 2.2 Condition C_0 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 $n\to\infty$. For stratified SRSWOR designs, condition C_2 follows directly from C_0. For pps , Poisson and πps designs conditions C_1 and C_2 mean, respectively, that as $N\to\infty$, the sizes are of the same magnitude.

3. COUNTING PROCESS THEORY IN THE PRODUCT 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:

$$P_{d,m}(s\times F)=p_{dN}(s,z_{11},...,z_{LN_L})\cdot P(F),\ \ s\in C(S_N),\ \ F\in\Im;.$$

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. Now we define the tools of counting process theory in the product space.

3.1.1 Filtrations

Martingale and counting process theory is developed on a stochastic basis, that is, a probability space with a filtration. A filtration is an increasing family of right continuous sub-σ-algebras. The filtration corresponding to the counting process $\eta\ (t)$ in the proportional hazards model is $\Im_t = \sigma(\eta_{hi}(u),\ \eta_{hi}^{C}(u), i \le N_h, h \le L,\ 0 \le u \le t)$. We define here a stochastic basis for the product space. The family of sub-σ-algebras defined by $\{\Im_t^{d,m} = C(S_N) \times \Im_t,\ \ t \ge 0\}$ is a filtration; it is increasing and right continuous because $\{\Im_t,\ \ t \ge 0\}$ is so.

3.1.2 Sample Counting Processes and Martingales

We define the sample counting process $\hat{\eta}(t,s,\omega) = \sum_{h=1}^{L}\sum_{i=1}^{N_h} \eta_{hi}(t,\omega) \cdot I_{hi}(s)/\pi_{hi}$. $\hat{\eta}$ is a counting process in the product space with respect to the filtration $\left\{\Im_t^{d,m} : t \ge 0\right\}$, since each term is the product of a counting process with respect to the original filtration $\left\{\Im_t\ : t \ge 0\right\}$ in $(\Omega, \Im, P)$ and the factor I_{hi} which is $C(S_N)$-measurable.

If $M_{hi},\ \ i = 1,\ldots,N_h,\ \ h = 1,\ldots,L$ are martingales in $(\Omega, \Im, P)$ with respect to $\{\Im_t\ : t \ge 0\}$ (and hence with respect to $\left\{\Im_t^{d,m} : t \ge 0\right\}$ as well), we define the martingale sample process by:

$$\hat{M}\ \ (t,s,\omega) = \sum_{h=1}^{L}\sum_{i=1}^{N_h} M_{hi}(t,\omega) \cdot I_{hi}(s)/\pi_{hi}$$

Hence $\hat{M}$ is a martingale with respect to the filtration $\left\{\Im_t^{d,m} : t \ge 0\right\}$ in the product space. We can easily verify the three necessary conditions (see Definition 1.2.8, Fleming and Harrington, 1991, p. 22): each term is a martingale since

1) $M_{hi}(t,\omega) I_{hi}(s)/\pi_{hi}$ is adapted to $\{\Im_t^{d,m} : t \ge 0\}$,
2) $E_{d,m}(| M_{hi}(t,\omega) \cdot I_{hi}(s)/\pi_{hi} |) < \infty$ for all $t < \infty$, and
3) $E_{d,m}(M_{hi}(t+u) \cdot I_{hi}(s)/\pi_{hi} \mid \Im_t^{d,m}) = (I_{hi}(s)/\pi_{hi}) E_{d,m}(M_{hi}(t+u) \mid \Im_t) = (I_{hi}(s)/\pi_{hi}) M_{hi}(t)$ for all $t \ge 0,\ \ u \ge 0$.

3.1.3 Predictable co-variation processes

In order to complete the counting process tools for the product space, we need to define either multivariate counting processes $\{\eta_{hi}(t),\ \ i = 1,\ldots,N_h,\ \ h = 1,\ldots,L\}$,i.e., counting processes such that not two of them jump at the same time, or processes where $\{\Delta\eta_{hi}(t),\ \ i = 1,\ldots,N_h,\ \ h = 1,\ldots,L\}$ are independent random variables. Under either of these two assumptions, we obtain that the following predictable co-variation sample processes are zero:

$$\langle M_{hi} I_{hi}(s)/\pi_{hi}, M_{h'j} I_{h'j}(s)/\pi_{h'j}\rangle(t) = 0,\ \ h \ne h'\ \ or\ \ i \ne j,\ \ i = 1,\ldots,N_h,\ \ \ j = 1,\ldots,N_{h'},\ \ h,h' = 1,\ldots,L,\ t \ge 0$$

(see Fleming and Harrington(1992) Lemma 2.6.1 p. 81). This property is crucial for calculating co-variances of the stochastic integrals under consideration, and for applying the Central Limit Theorem for Martingales.

In the product space, if the $\{\eta_{hi}(t),\ \ i = 1,\ldots,N_h,\ \ h = 1,\ldots,L\}$ are multivariate counting processes, then the sampling processes $\{\eta_{hi}(t) I_{hi}(s)/\pi_{hi},\ \ i = 1,\ldots,N_h,\ \ h = 1,\ldots,L\}$ are also multivariate counting processes. However, the second condition, namely that $\{\Delta\eta_{hi}(t)I_{hi}(s)/\pi_{hi},\ \ i=1,...,N_h,\ \ h=1,...,L\}$ are independent random variables, does not hold because $\{\eta_{hi}(t)I_{hi}(s)/\pi_{hi},\ \ i=1,...,N_h,\ \ h=1,...,L\}$ are not independent in the product space. Instead we have:

Lemma 3.1 If the counting processes $\{\eta_{hi}(t),\ \ i=1,...,N_h,\ \ h=1,...,L\}$ are stochastically independent in the model, and we define $M_{hi}(t)=\eta_{hi}(t)-A_{hi}$ where A_{hi} is the compensator for η_{hi} (see Fleming and Harrington, 1991, p.38 for definition of a compensator), then the predictable co-variation sample-processes satisfy:

$$\langle M_{hi}I_{hi}(s)/\pi_{hi}, M_{h'j}I_{h'j}/\pi_{h'j}\rangle(t)=0,\ \ h\neq h'\ \ or\ \ i\neq j,\ \ i=1,...,N_h,\ \ j=1,...,N_{h'},\ \ h,h'=1,...,L,\ \ t\geq 0$$

Proof: We follow the argument of Lemma 2.6.1 in Fleming and Harrington (1992), taking into consideration that calculations are with respect to the product space filtration $\{\Im_t^{d,m}=C(S_N)\times\Im_t,\ \ t\geq 0\}$.

4. THE LEMMA

Lemma 4.1 Let $\{G_N(t):-\infty<t<\infty\}$ be a sequence of random functions defined on probability spaces $(\Omega_N,\Im_N,P_N)$ with sample paths that are monotonic bounded functions. In addition, they are right-continuous if non-decreasing, and left-continuous if non-increasing. Let $\{g(t):-\infty<t<\infty\}$ be the non-stochastic bounded monotonic limit in probability of $\{G_N(t):-\infty<t<\infty\}$, such that $G_N(t)-g(t)\to 0\ \ in\ \ P_N$ for all t and either $G_N(t-)-g(t-)\to 0\ \ in\ \ P_N$ for all t if non-decreasing, or $G_N(t+)-g(t+)\to 0\ \ in\ \ P_N$ for all t if non-increasing.

Then $\sup\limits_t |G_N(t)-g(t)|\to 0\ \ in\ \ P_N$.

Proof: The proof is nearly identical to that used in the proof of the Glivenko-Cantelli Theorem for super-populations (cf. Billingsley, P.(1979)) and only diverges at the end, where we are required to prove convergence in probability (P_N). See also the comment given by Fleming and Harrington p. 305 for the proof for almost sure convergence.

Theorem 4.1 We assume the proportional hazards model stated in Section 2 and the following:

i) The covariate vectors X_{hi} are constant (in time) and bounded: $\sup\limits_{h,i}\left|X_{hi}^N\right|\leq K\ \ a.s.\ \ as\ \ N\to\infty$.

ii) There exists a neighborhood $\aleph$ of β_0 and, respectively, scalar, vector and matrix functions $s^{(0)}$, $s^{(1)}$ and $s^{(2)}$ defined on $\aleph\times[0,\infty)$ such that for $j=0,1,2,$ and for $t\geq 0,\ \ \beta\in\aleph$ such that

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

iii) $\lim\limits_{N\to\infty}E\left\{S^{(j)}(\beta,t+)\right\}=s^{(j)}(\beta,t+), t\geq 0$.

iv) The sample estimators $\hat{S}^{(j)}$, $j=0,1,2,$ are design-consistent.

v) Conditions C_0 and C_1 or C_2 hold, depending on the design.

Remark 4.1 Note that the model conditions above are not too different from the conditions imposed on the model by Fleming and Harrington, 1991, Theorem 8.4.1, for the partial likelihood score function to converge. The extra conditions (ii) and (iii) stem from the slight generalization of that theorem to the case of a stratified super-population which is based on independent, not identically distributed random triples $(\widetilde{T}_{hi}^N,\delta_{hi}^N,X_{hi}^N)$, $i=1,...,N_h,\ \ h=1,...,L$. A result for covariates varying with time and survival curves that are not absolutely continuous, as is done in classical survival theory, could follow if we impose stricter model conditions (see, for example, Theorem 8.2.1, Fleming and Harrington (1991)).

We have:

1) $S^j(\beta,t) \to s^j(\beta,t)$, and $S^j(\beta,t+) \to s^j(\beta,t+)$, *a.s. as* $N \to \infty$, $j = 0,1,2,$ for each $t \geq 0$,

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

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

Proof: We use the Strong Law of Large Numbers for independent, non-identically distributed random vectors as stated in Shao (1999), Theorem 1.14, p.46, then follow the argument used in Theorem 8.4.1, Fleming and Harrington, 1991, and the fact that sample estimators of the $S^{(j)}$, $j = 0,1,2,$ are design-consistent.

Remark 4.2 Note that the probability in statement 3) is the product space probability ($P_{d,m}$), and not the design probability: we have to make use of both, the model convergence (P) assumed for the S averages, and the design consistency (p_{dN}) of the sample processes, both stated in the hypothesis. Thus we do need to assume a super-population behind the design to obtain the theorem.

Corollary 4.1 Approximation for the sample partial likelihood score function 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 $0 \leq u \leq t,\ t \in (0,\infty)$. Under the conditions of Theorem 4.1 and conditions C_0 and C_1 or C_2 depending on the design, we have, uniformly in $t,\ 0 \leq t \leq \infty$,

$$\frac{1}{\sqrt{N}} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \hat{U}_{hi}^N(\beta,t) = \frac{1}{\sqrt{N}} \sum_{h=1}^{L} \sum_{i=1}^{N_h} U_{hi}^N(\beta,t) \cdot I_{hi}(s)/\pi_{hi} + o_{P_{d,m}}(1),$$

where the $U_{hi}^N(\beta,t)$ and the $\hat{U}_{hi}^N(\beta,t)$ are defined in (2.7) and (2.11) respectively. In particular,

$$\frac{1}{\sqrt{N}}\hat{U}(\beta) = \frac{1}{\sqrt{N}}\hat{U}^N(\beta,\infty) = \frac{1}{\sqrt{N}} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \frac{I_{hi}(s)}{\pi_{hi}} \int_0^\infty \left\{X_{hi}^N - e(\beta,u)\right\} dM_{hi}(u) + o_{P_{d,m}}(1).$$

Furthermore, convergence to zero in $P_{d,m}$ implies convergence to zero in design probability for this particular product space (see Rubin-Bleuer and Schiopu Kratina (2002), Example 4.2 and Remark 4.3), and hence the approximation is also in design probability.

Proof: $\frac{1}{\sqrt{N}} \sum_{h=1}^{L}\sum_{i=1}^{N_h} U_{hi}^N(\beta,t) - \frac{1}{\sqrt{N}} \sum_{h=1}^{L}\sum_{i=1}^{N_h} U_{hi}^N(\beta,t) \cdot I_{hi}(s)/\pi_{hi} = \frac{1}{\sqrt{N}} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \frac{I_{hi}(s)}{\pi_{hi}} \int_0^t \{\frac{\hat{S}^1(\beta,u)}{\hat{S}^0(\beta,u)} - \frac{S^1(\beta,u)}{S^0(\beta,u)}\} dM_{hi}(u) \cdot$

We show that the double sum above converges to zero for all $t \leq \infty$ in the probability of the product space. We use the theory of martingales to calculate the variance in the product space. By Theorem 4.1, we have the point-wise and sup-norm convergence of the S-averages to non-stochastic functions. Moreover, $s^{(0)}(\beta,t)$ is bounded away from zero on $\aleph \times [0,\tau]$ by the same arguments used in theorem 8.4.1. of Fleming and Harrington, 1991. Thus, both $S^{(0)}(\beta,t)$ and $\hat{S}^{(0)}(\beta,t)$ are bounded away from zero and $\sup_{0 \leq u < \infty} |\{\frac{\hat{S}^1(\beta,u)}{\hat{S}^0(\beta,u)} - \frac{S^1(\beta,u)}{S^0(\beta,u)}\}| = o_{P_{d,m}}(1)$ as $n \to \infty$.

We have:

$$V(\beta,t)=V_{d,m}\left(\frac{1}{\sqrt{N}}\sum_h\sum_i\int_0^t\left\{\frac{\hat{S}^1(\beta,u)}{\hat{S}^0(\beta,u)}-\frac{S^1(\beta,u)}{S^0(\beta,u)}\right\}d\left\{M_{hi}(u)\frac{I_{hi}(s)}{\pi_{hi}}\right\}\right)$$

$$=E_{d,m}\left(\int_0^t\left\{\frac{\hat{S}^1(\beta,u)}{\hat{S}^0(\beta,u)}-\frac{S^1(\beta,u)}{S^0(\beta,u)}\right\}^2\sum_{h=1}^{L}\frac{1}{N}\sum_{i=1}^{N_h}\frac{I_{hi}(s)}{\pi_{hi}^2}d\langle M_{hi},M_{hi}\rangle\right)\le o\;(1)\cdot\int_0^t\frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{1}{\pi_{hi}}E_m(Y_{hi}(u)e^{\beta'\cdot X_{hi}^N})\lambda_0(u)du,$$

by Theorem 2.6.2 Fleming and Harrington (1992) p.82, the Dominated Convergence Theorem and Theorem 4.1. Note that $o(1)\to 0\;\; as\;\; n\to\infty$ independently of t. On the other hand,

$$\frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{1}{\pi_{hi}}E_m(Y_{hi}(u)e^{\beta'\cdot X_{hi}^N})\le e^{|\beta|K}\frac{1}{N}\sum_h\sum_i\frac{1}{\pi_{hi}}E_m(P(T_{hi}^N\ge u\mid X_{hi}^N)\cdot P(C_{hi}^N\ge u\mid X_{hi}^N))$$

and $\quad P(T_{hi}^N\ge u\mid X_{hi}^N)\cdot P(C_{hi}^N\ge u\mid X_{hi}^N)=S_0(u)^{\exp(\beta_0\cdot X_{hi}^N)}\cdot P(C_{hi}^N\ge u\mid X_{hi}^N)\le S_0(u).$

Thus for all t, $0\le t\le\infty$, we have,

$$V(\beta,t)\le o(1)\cdot\int_0^t\frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{1}{\pi_{hi}}S_0(u)\lambda_0(u)du\le o(1)\cdot\left(\frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{1}{\pi_{hi}}\right).$$

The only condition on the design we require for this approximation, apart of design-consistent estimators is, depending on the design, either condition C_1 or C_2, which state that the last factor is bounded as N increases.

REFERENCES

Andersen, P. K.& 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.

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.

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

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.

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

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.*

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

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