A Logistic Regression Model for Hazard: Asymptotic Results

Author(s): Elja Arjas and Pentti Haara

Source: *Scandinavian Journal of Statistics*, 1987, Vol. 14, No. 1 (1987), pp. 1-18

Published by: Wiley on behalf of Board of the Foundation of the Scandinavian Journal of Statistics

Stable URL: https://www.jstor.org/stable/4616044

Scand J Statist 14: 1–18, 1987

A Logistic Regression Model for Hazard: Asymptotic Results

ELJA ARJAS and PENTTI HAARA

University of Oulu

ABSTRACT. A dynamic form of the discrete time logistic regression model is considered as a means to analyse complicated failure time data. A characteristic property of this approach is that the events recorded in the data are always treated in the model in the order in which they occurred in real time, without first aligning them according to some particular "basic time measurement". Parametric modelling is used throughout. This paper deals with asymptotic results. The key theorems concern the asymptotic normality of the estimated regression coefficients and the limiting behaviour of the empirical score process as observation time tends to infinity.

Key words: discrete failure time, covariate, censoring, martingale CLT, ML estimation, asymptotic normality, score process, Brownian bridge

1. Introduction

In recent years, the dominant role among regression models of hazard has been played by the *semiparametric* model of Cox (1972) and its extensions. An unspecified time-dependent *baseline hazard*, common to all individuals, is assumed to act multiplicatively on a *relative risk function*, and is then suppressed in the estimation of the regression coefficients. The estimation is based on considering relative risks within risk sets, formed by aligning individuals according to some particular measurement of time, such as age or time since diagnosis. A drawback of such alignment is that it usually changes the natural sequencing of events in real time. In particular, it can destroy the natural martingale structure of the real time counting process models (cf. Sellke & Siegmund (1983) and Arjas (1985a)).

We have, in Arjas & Haara (1984), advocated what might be called *a real-time approach to hazard regression*, arguing that if a hazard model uses time-dependent covariates, all dependence on time-related quantities can actually be accommodated into such covariates. Then, instead of postulating the existence of a common unspecified multiplicative baseline hazard, a function of some particular measurement of time, one attempts to model how an individual's hazard at a certain (real) time t depends on "the currently prevailing conditions for survival". Frequently, some of such conditions are best expressed in terms of conveniently chosen time readings, such as age, time from diagnosis, time from treatment, or indeed, calendar time. Several time readings may be needed simultaneously for a realistic description. Suitably chosen functions of these readings can then be listed as covariates, among other factors that are thought to be relevant to the individual's survival.

Arjas & Haara (1984) showed that, under general conditions, the real-time approach leads to likelihood expressions of a common form. On the other hand, this general likelihood formula has too little structure for immediate statistical application such as parameter estimation. To facilitate such application, a concrete proposal for a regression model was made in Arjas (1985b). That paper incorporated into the real-time approach two features that have practical rather than conceptual or theoretical motivation. First a *discrete time parameter* was used, which, together with a natural conditional independence assumption, removes all difficulties concerning tied failure times. Second, a logistic regression model with binomial response was used as the primary statistical tool, leading to unproblematic numerical routines in ordinary ML estimation. Finally, an example was analysed to illustrate the method.

Here we continue to study this logistic regression model, concentrating on asymptotic results. In section 2 we recall the statistical model in Arjas (1985b) and explain its basic martingale structure. A more careful exposition of the statistical ideas is presented in appendix 1. Section 3 contains an asymptotic normality result of the estimated regression coefficients (theorem 1) and some corollaries. The proof of theorem 1 is contained in section 4. Section 5 discusses the limiting behaviour of the empirical score, viewed as a stochastic process, linking it with the Brownian bridge (theorem 2).

2. The statistical model

As mentioned in the introduction, we consider a discrete time index, and then simply use the positive integers $t \geqslant 1$. In practice we let the time unit depend directly on how time is measured in the experimental data. Nearly all survival data are reported by using a day as time unit. In such a case we simply identify the time index value t with the tth day of the study. We make the corresponding choice if, for example, a week, a month, or a year, is used in the reporting. No further approximation, e.g. by grouping survival times in fixed subintervals, is assumed here (cf. Kalbfleisch & Prentice (1980, pp. 36–38 and 98 ff.)). Neither do we connect in any way the number of model parameters to the discretization. (A well-known example of such a connection is the assumption of piecewise constant baseline hazards on the fixed subintervals, where each subinterval adds one parameter to the model.)

The individuals included in the study are indexed by j, $j \geqslant 1$. We define the risk indicators $Y_j(t-1)$, $j, t \geqslant 1$, by

$$Y_j(t-1)=\begin{cases}1 & \text{if individual } j \text{ is at risk at } t\\ 0 & \text{otherwise}\end{cases} \tag{2.1}$$

and the failure indicators $\Delta N_j(t), j, t \geqslant 1$, by

$$\Delta N_j(t)=\begin{cases}1 & \text{if individual } j \text{ at risk at } t \text{ fails}\\ 0 & \text{otherwise.}\end{cases} \tag{2.2}$$

Denote by $R(t-1)=\{j \geqslant 1: Y_j(t-1)=1\}$ the *risk set* at t. The size of the risk set,

$$\text{card } R(t-1)=\sum_{j\geqslant 1} Y_j(t-1),$$

is assumed to be finite for all $t \geqslant 1$.

We do not assume that all individuals are present at time 0; they may enter and leave the risk set many times, and they may also "fail" many times.

Suppose then, that, for every individual j and time t such that $Y_j(t-1)=1$, the investigator knows the value of a p vector $\mathbf{Z}_j(t-1)=\{Z_{j1}(t-1), \ldots, Z_{jp}(t-1)\}$ of the relevant covariates. We shall view the value of $\Delta N_j(t)$ as the outcome of a Bernoulli experiment, where the probability of the event $\{\Delta N_j(t)=1\}$ depends on $\mathbf{Z}_j(t-1)$. For convenience, we may assume that $\mathbf{Z}_j(t-1)=\mathbf{0}$ whenever $Y_j(t-1)=0$.

Let $(\Omega, \mathscr{F})$ be a measurable space in which the variables $Y_j(t-1)$, $\mathbf{Z}_j(t-1)$ and $\Delta N_j(t)$ are defined. Let $\mathscr{F}_0$ be the σ-field representing "initial information"; often $\mathscr{F}_0$ is the trivial field. Then the σ-fields $\mathscr{F}_t$ and $\mathscr{G}_{t-1}$, $t \geqslant 1$, defined inductively by

$$\begin{aligned}\mathscr{G}_{t-1}&=\mathscr{F}_{t-1} \vee \sigma[R(t-1), \{\mathbf{Z}_j(t-1); j \in R(t-1)\}],\\ \mathscr{F}_t&=\mathscr{G}_{t-1} \vee \sigma\{\Delta N_j(t); j \in R(t-1)\},\end{aligned} \tag{2.3}$$

represent the experimental history registered up to time t, $\mathscr{F}_t$ including and $\mathscr{G}_{t-1}$ excluding the failures at t.

Consider then a partially specified statistical model $\{\mathbf{P}^{\boldsymbol{\beta}};\ \boldsymbol{\beta}\in\mathbf{R}^p\}$ (cf. section 5 in Arjas & Haara (1984)) for the observation process $[R(v-1),\ \{\Delta N_j(v),\ \mathbf{Z}_j(v-1);\ j\in R(v-1)\}]_{v\geqslant 1}$. In particular, we assume that the likelihood function corresponding to data up to time t depends on $\boldsymbol{\beta}$, the parameter of interest, through the factor

$$L_t^{\boldsymbol{\beta}}=\prod_{v\leqslant t}\prod_{j\in R(v-1)}\mathbf{P}^{\boldsymbol{\beta}}\{\Delta N_j(v)=\Delta n_j(v)|\mathscr{G}_{v-1}\},\tag{2.4}$$

where $\{\Delta n_j(v),\ j\in R(v-1),\ v\leqslant t\}$ are the observed values of the variables (2.1) and, for any $j\in R(v-1)$,

$$\log\frac{\mathbf{P}^{\boldsymbol{\beta}}\{\Delta N_j(v)=1|\mathscr{G}_{v-1}\}}{\mathbf{P}^{\boldsymbol{\beta}}\{\Delta N_j(v)=0|\mathscr{G}_{v-1}\}}=\boldsymbol{\beta}'\mathbf{Z}_j(v-1).\tag{2.5}$$

The assumptions leading to this logistic regression model for binary data were presented in Arjas (1985b). In order to make this paper more self-contained, we restate the assumptions, and a discussion, in appendix 1.

Note the formal similarity of this model and the discrete time model in Cox (1972) if one in the later sets $\lambda_0(t)=1$. However, setting $\lambda_0(t)=1$ is a way of specifying the baseline hazard completely. Here we use a different likelihood, one that does not suppress $\lambda_0(t)$ by using relative risk estimates. On the other hand, our model is similar to the continuous time model of Borgan (1984) in the sense that both are fully parametric.

As is well known, an alternative way of expressing (2.5) is to write

$$\mathbf{P}^{\boldsymbol{\beta}}\{\Delta N_j(v)=1|\mathscr{G}_{v-1}\}=Y_j(v-1)\cdot L\{\boldsymbol{\beta}'\mathbf{Z}_j(v-1)\},\tag{2.6}$$

where $L(x)=\{1+\exp(-x)\}^{-1}$ is the inverse of the logit function. Thus, $Y_j(v-1)=0$ implies that $\Delta N_j(v)=0$. Also recall the convention that $\mathbf{Z}_j(v-1)=\mathbf{0}$ whenever $Y_j(v-1)=0$.

The logarithmic likelihood corresponding to (2.4) and (2.5) gets a familiar form in logistic regression,

$$l(\boldsymbol{\beta},\ t)\underset{\text{def}}{=}\log L_t^{\boldsymbol{\beta}}=\sum_{v\leqslant t}\sum_j Y_j(v-1)[\Delta N_j(v)\cdot\boldsymbol{\beta}'\mathbf{Z}_j(v-1)-g\{\boldsymbol{\beta}'\mathbf{Z}_j(v-1)\}],\tag{2.7}$$

where we denote $g(x)=\log\{1+\exp(x)\}$. The ML estimate $\hat{\boldsymbol{\beta}}_t$ can be determined in the standard way by solving $(\partial/\partial\boldsymbol{\beta})l(\boldsymbol{\beta},\ t)=\mathbf{0}$ for $\boldsymbol{\beta}$. For a practical illustration of this, see Arjas (1985b).

Before considering the asymptotic results we show briefly how the present framework gives rise to a number of interesting martingales and martingale-related processes. These results will be used later in the proofs. Defining, for $j\geqslant 1$, $\Delta M_j^{\boldsymbol{\beta}}(v)$ by

$$\Delta M_j^{\boldsymbol{\beta}}(v)=\Delta N_j(v)-\mathbf{P}^{\boldsymbol{\beta}}\{\Delta N_j(v)=1|\mathscr{G}_{v-1}\},\tag{2.8}$$

we immediately see that

$$\mathbf{E}^{\boldsymbol{\beta}}\{\Delta M_j^{\boldsymbol{\beta}}(v)|\mathscr{G}_{v-1}\}=0,\qquad v\geqslant 1.\tag{2.9}$$

Therefore, each sequence $\Delta M_j^{\boldsymbol{\beta}}(v)$, $v\geqslant 1$, is a $(\mathbf{P}^{\boldsymbol{\beta}},\ \mathscr{G}_t)$ martingale difference, and each partial sum process

$$M_j^{\boldsymbol{\beta}}(t)=\sum_{v\leqslant t}\Delta M_j^{\boldsymbol{\beta}}(v),\qquad t\geqslant 1,\tag{2.10}$$

is a (locally bounded) $(\mathbf{P}^{\boldsymbol{\beta}},\ \mathscr{G}_t)$ martingale. This is the fundamental martingale related to the counting process

$$N_j(t)=\sum_{v\leqslant t}\Delta N_j(v),\qquad t\geqslant 1.$$

By differentiating in (2.7), we find the familiar form of the score function in the logistic regression model,

$$\mathbf{U}(\boldsymbol{\beta}, t) = \underset{\text{def}}{\frac{\partial}{\partial \boldsymbol{\beta}}} l(\boldsymbol{\beta}, t) = \sum_{v \leqslant t} \sum_{j \geqslant 1} \mathbf{Z}_j(v-1) \Delta M_j^{\boldsymbol{\beta}}(v), \qquad t \geqslant 1, \tag{2.11}$$

which again is a (p-vector valued) $(\mathbf{P}^{\boldsymbol{\beta}}, \mathcal{G}_t)$ locally square-integrable martingale.

Also the second derivative of the log-likelihood function (2.7) gets a martingale related interpretation: We have from (2.6) that

$$\Delta \langle M_j^{\boldsymbol{\beta}} \rangle(v) \underset{\text{def}}{=} \operatorname{Var}^{\boldsymbol{\beta}} \{\Delta M_j^{\boldsymbol{\beta}}(v) | \mathcal{G}_{v-1}\} = Y_j(v-1) V\{\boldsymbol{\beta}' \mathbf{Z}_j(v-1)\}, \tag{2.12}$$

where $V(x) = L(x)\{1 - L(x)\}$ is a notation for binomial variance when the logit is given by x. Similarly, using the conditional independence of the variables $\Delta N_j(v)$ and $\Delta N_h(v)$ for $j \neq h$, we obtain

$$\Delta \langle M_j^{\boldsymbol{\beta}}, M_h^{\boldsymbol{\beta}} \rangle(v) \underset{\text{def}}{=} \operatorname{Cov}^{\boldsymbol{\beta}} \{\Delta M_j^{\boldsymbol{\beta}}(v), \Delta M_h^{\boldsymbol{\beta}}(v) | \mathcal{G}_{v-1}\} = 0. \tag{2.13}$$

Thus, the martingales $\{M_j^{\boldsymbol{\beta}}(t)\}_{t \geqslant 1}$ and $\{M_h^{\boldsymbol{\beta}}(t)\}_{t \geqslant 1}$, $j \neq h$, are *orthogonal*. By a straightforward calculation we obtain, for $1 \leqslant i, k \leqslant p$,

$$\begin{aligned}
\Delta \langle U_i(\boldsymbol{\beta}, \cdot), U_k(\boldsymbol{\beta}, \cdot) \rangle(v) &= \Delta \left\langle \frac{\partial}{\partial \beta_i} l(\boldsymbol{\beta}, \cdot), \frac{\partial}{\partial \beta_k} l(\boldsymbol{\beta}, \cdot) \right\rangle(v) \\
&= \operatorname{Cov}^{\boldsymbol{\beta}} \left\{ \sum_{j \geqslant 1} Z_{ji}(v-1) \Delta M_j^{\boldsymbol{\beta}}(v), \sum_{j \geqslant 1} Z_{jk}(v-1) \Delta M_j^{\boldsymbol{\beta}}(v) | \mathcal{G}_{v-1} \right\} \qquad \text{(by (2.11))} \\
&= \sum_{j \geqslant 1} Z_{ji}(v-1) Z_{jk}(v-1) \operatorname{Var}^{\boldsymbol{\beta}} \{\Delta M_j^{\boldsymbol{\beta}}(v) | \mathcal{G}_{v-1}\} \qquad \text{(by (2.13))} \\
&= \sum_{j \geqslant 1} Z_{ji}(v-1) Z_{jk}(v-1) V\{\boldsymbol{\beta}' \mathbf{Z}_j(v-1)\} \qquad \text{(by (2.12))}.
\end{aligned} \tag{2.14}$$

Note that the above calculation is formally correct only when $\mathbf{U}$ is assumed square integrable. We feel that the elementary calculation (2.14) demonstrates the covariance structure of the model better than the rigorous proof using localizing stopping times (for which we do not need extra integrability assumptions). Differentation in (2.11) gives easily

$$\frac{\partial^2}{\partial \beta_i \partial \beta_k} l(\boldsymbol{\beta}, t) = - \sum_{1 \leqslant v \leqslant t} \sum_j Z_{ji}(v-1) Z_{jk}(v-1) V\{\boldsymbol{\beta}' \mathbf{Z}_j(v-1)\}, \qquad t \geqslant 1. \tag{2.15}$$

Thus, the ($p \times p$ matrix valued) covariance process of the score, evaluated at t,

$$\langle \mathbf{U}(\boldsymbol{\beta}, \cdot) \rangle(t) \underset{\text{def}}{=} \left\{ \sum_{v \leqslant t} \Delta \left\langle \frac{\partial}{\partial \beta_i} l(\boldsymbol{\beta}, \cdot), \frac{\partial}{\partial \beta_k} l(\boldsymbol{\beta}, \cdot) \right\rangle(v) \right\}_{1 \leqslant i, k \leqslant p},$$

satisfies

$$\langle \mathbf{U}(\boldsymbol{\beta}, \cdot) \rangle(t) = - \frac{\partial^2}{\partial \boldsymbol{\beta}^2} l(\boldsymbol{\beta}, t). \tag{2.16}$$

From now on we denote this matrix by $\mathbf{I}(\boldsymbol{\beta}, t)$.

It is interesting to compare the above martingale analysis with the works of Sellke & Siegmund (1983) and Slud (1984): considering sequential testing in the case of staggered entry, they indexed counting variables with both calendar time and age. Fixing one of the two variables, and assuming a number of conditional independence properties concerning the individual entry times, covariates, and latent life and censoring times, they obtained a martingale property in the other.

After these preliminaries we go into the detailed discussion of the asymptotic results.

3. Asymptotic normality results

Apart from trivial cases, the exact distribution of the ML estimate $\hat{\boldsymbol{\beta}}_t$ is not known. Therefore, asymptotic results are needed to approximate this distribution. The main result of this section is the asymptotic normality of $\hat{\boldsymbol{\beta}}_t$ as $t \to \infty$ (theorem 1). As a corollary, we obtain a chi-squared limit for the likelihood ratio statistic (corollary 3).

To prove asymptotic normality, we use the now well-established approach which is based on the fundamental martingale CLT of Rebolledo (1978, 1980), and was first used by Gill (1980). We follow closely the treatment of Andersen & Gill (1982), the main differences being: (a) we consider a discrete time parametric model whereas Andersen & Gill consider a continuous time semiparametric model based on relative risks; (b) to obtain asymptotic results, we let the observation time t approach infinity, while Andersen & Gill consider a fixed time interval, letting the number of counted individuals go to infinity; (c) our conditions for consistency and asymptotic normality, and the needed proofs, are somewhat simpler. It is also of interest to compare our model and proof to those in Borgan (1984). Like ours, Borgan's model is fully parametric. However, being in continuous time and in using a fixed observation interval, it also has much in common with the approach of Andersen & Gill.

Our motives for (a) were already explained at the beginning of section 2. We now elaborate on (b) a little.

In most follow-up studies involving human data, entry is staggered. As the time span for observation becomes longer, typically a larger number of individuals will be included in the study, and the amount of information increases. This is the kind of situation we have in mind when formulating the conditions (C1)–(C3) below. It can be contrasted with the more common formulation for asymptotics, where the number of individuals, all assumed to be present at $t=0$, is directly let go to infinity while the time interval for observation remains fixed at a finite value. This latter framework is natural in laboratory experiments with animals. With human data, however, if entry is staggered, an alignment of individuals is first necessary before the limit can be considered (cf. our comments in section 1).

Having now two different approaches to asymptotics, it is of interest to ask: "Can the approach in which $t \to \infty$ be easily modified to also apply in the case where the observation period is fixed?" The answer is "Yes": a key requirement in the following proofs is to have the number of non-zero terms in the score martingale (2.11) tend to infinity. Each such term corresponds to an individual at risk during a unit time, and so the number of terms coincides with the well-known "total time on test" quantity. Obviously, this total time tends to infinity with the number of individuals studied, even though the observation period may remain fixed. Moreover, the total time can be used in an obvious way as time index when writing an expression for the score, still preserving its martingale property. (Instead of a double sum, one then obtains a simple sum.) If the conditions (C1)–(C3) below hold for this modification, then will also, of course, the theorems. The same reasoning holds in section 5 as well. We stress here that, in spite of perhaps changing the time index for score, the role of real time in our statistical model remains the same.

First we need some new notation. Let $\boldsymbol{\beta}_0$ be the fixed "true value" of the parameter and let $\hat{\boldsymbol{\beta}}_t$ be the ML estimate corresponding to the data up to time t. For simplicity we write $\mathbf{P}$, $\mathbf{E}$ and $\Delta M_j(t)$ in place of $\mathbf{P}^{\boldsymbol{\beta}_0}$, $\mathbf{E}^{\boldsymbol{\beta}_0}$ and $\Delta M_j^{\boldsymbol{\beta}_0}(t)$. We denote by $\xrightarrow[\mathbf{P}]{}$ the stochastic convergence and by $\xrightarrow[\mathscr{D}]{}$ the convergence in distribution w.r.t. $\mathbf{P}$. Sometimes we write $X_t = X + 0_{\mathbf{P}}(1)$ meaning $X_t \xrightarrow[\mathbf{P}]{} X$. In the sequel, all convergence statements hold as t approaches infinity through the values 1, 2, ...

For a matrix $\mathbf{A}$ or a vector $\mathbf{a}$, we denote $\|\mathbf{A}\| = \max_{i,j} |a_{ij}|$ and $\|\mathbf{a}\| = \max |a_i|$. For a function (random or non-random) $x(t) = x_t$, defined for $t \geq 1$, we denote

$$x_t(u) = x([ut]), \qquad 0 \leq u \leq 1, \quad t \geq 1, \tag{3.1}$$

where $x_0 = 0$ and $[ut]$ is the integer part of ut. Thus, $x_t(\cdot)$ can be viewed as the first t points of the sequence $x(\cdot)$, time transformed into the unit interval. Note that (3.1) transforms a $(\mathbf{P}, \mathscr{G}_t)$ martingale $M(t)$, $t \geq 1$, into a $(\mathbf{P}, \mathscr{G}_{[ut]})$ martingale $M_t(u)$, $0 \leq u \leq 1$.

We call a sequence c_t, $t \geq 1$, of positive numbers *stable* if $c_t \uparrow + \infty$ and the limit

$$\gamma(u) = \lim_{t \to \infty} c_t^{-1} c_t(u), \qquad 0 \leq u \leq 1, \tag{3.2}$$

exists and is continuous at $u = 1$ and $u = 0$. Trivially, γ is monotone with $\gamma(0) = 0$ and $\gamma(1) = 1$. In fact, γ is a continuous bijection of the interval [0, 1] (continuity at 1 implies that γ is multiplicative). The convergence in (3.2) is then uniform over $0 \leq u \leq 1$. Note that the sequence $c_t = t^\alpha$, $t \geq 1$, is stable with $\gamma(u) = u^\alpha$, $0 \leq u \leq 1$, $\alpha > 0$.

We define the important processes

$$\begin{aligned} S(\boldsymbol{\beta}, t) &= \sum_{v \leq t} \sum_{j \geq 1} Y_j(v-1) g\{\boldsymbol{\beta}' \mathbf{Z}_j(v-1)\}, \\ \mathbf{A}(\boldsymbol{\beta}, t) &= \sum_{v \leq t} \sum_{j \geq 1} Y_j(v-1) \mathbf{Z}_j(v-1) L\{\boldsymbol{\beta}' \mathbf{Z}_j(v-1)\}, \qquad t \geq 1, \end{aligned} \tag{3.3}$$

establishing easily the relationships

$$\begin{aligned} \mathbf{A}(\boldsymbol{\beta}, t) &= \frac{\partial}{\partial \boldsymbol{\beta}} S(\boldsymbol{\beta}, t), \\ \mathbf{I}(\boldsymbol{\beta}, t) &= \frac{\partial^2}{\partial \boldsymbol{\beta}^2} S(\boldsymbol{\beta}, t). \end{aligned} \tag{3.4}$$

The technical conditions, named after Andersen & Gill (1982), that are needed for the asymptotic normality are the following:

(*C1*) ("*Asymptotic stability*"). There exist a stable sequence c_t, $t \geq 1$, a neighbourhood $\mathscr{B}$ of $\boldsymbol{\beta}_0$, and a bounded continuous matrix valued function $\boldsymbol{\Sigma}$ defined on $\mathscr{B}$ such that

$$c_t^{-1} \mathbf{I}(\boldsymbol{\beta}, t) \underset{\mathbf{P}}{\to} \boldsymbol{\Sigma}(\boldsymbol{\beta}) \tag{3.5}$$

uniformly over $\boldsymbol{\beta} \in \mathscr{B}$. Further, there exist a scalar s_0 and a vector $\mathbf{a}_0$ such that

$$c_t^{-1} S(\boldsymbol{\beta}_0, t) \underset{\mathbf{P}}{\to} s_0$$

and

$$c_t^{-1} \mathbf{A}(\boldsymbol{\beta}_0, t) \underset{\mathbf{P}}{\to} \mathbf{a}_0.$$

(*C2*) ("*Lindeberg condition*"). The random variable

$$\hat{Z} \underset{\text{def}}{=} \sup_{v \geq 1} \sum_{j \geq 1} Y_j(v-1) \|\mathbf{Z}_j(v-1)\|$$

is a.s. finite.

(*C3*) ("*Asymptotic regularity*"). The matrix $\boldsymbol{\Sigma}_0 \underset{\text{def}}{=} \boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$ is positive definite.

In the formulation discussed at the beginning of this section, where t represents the total time on test, it is natural to let $c_t = t$. Then obviously $\gamma(u) = u$.

Our first main result is then:

Theorem 1

(Asymptotic normality of $\hat{\boldsymbol{\beta}}_t$)

Suppose that conditions (C1)–(C3) are satisfied. Then, as $t \to +\infty$,

$$c_t^{1/2}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0) \underset{\mathscr{D}}{\to} N(\mathbf{0}, \boldsymbol{\Sigma}_0^{-1}). \tag{3.6}$$

The proof of this theorem is given in section 4. Here are some simple consequences of the theorem or of its proof.

From Lemmas 3 and 5 we have:

Corollary 1

(Consistency of covariance matrix estimator)

$$c_t^{-1} \boldsymbol{I}(\hat{\boldsymbol{\beta}}_t, t) \underset{\mathbf{P}}{\to} \boldsymbol{\Sigma}_0. \tag{3.7}$$

Sometimes it may be of interest to use, in theorem 1, a *random* normalizing sequence C_t, $t \geqslant 1$. For example, such a sequence could be given by

$$C_t = \sum_{v \leqslant t} \sum_{j \geqslant 1} Y_j(v-1),$$

the total number of "days at risk" up to time t. This is possible if

$$c_t^{-1} C_t \underset{\mathbf{P}}{\to} c \tag{3.8}$$

for some constant $c \in (0, \infty)$. For completeness, we formulate the corresponding normality result below.

Corollary 2

Suppose that (C1)–(C3) and (3.8) hold. Then

$$C_t^{1/2}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0) \underset{\mathscr{D}}{\to} N(\mathbf{0}, c\boldsymbol{\Sigma}_0^{-1}). \tag{3.9}$$

As a third corollary we derive the asymptotic distribution of likelihood ratio statistic.

Corollary 3

Under conditions (C1)–(C3),

$$2\{l(\hat{\boldsymbol{\beta}}_t, t) - l(\boldsymbol{\beta}_0, t)\} \underset{\mathscr{D}}{\to} \chi_p^2. \tag{3.10}$$

Proof. Taylor expanding $l(\hat{\boldsymbol{\beta}}_t, t)$ around $\boldsymbol{\beta}_0$, and using (4.22) and (4.23) in section 4, we get

$$2\{l(\hat{\boldsymbol{\beta}}_t, t) - l(\boldsymbol{\beta}_0, t)\} = c_t^{1/2}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0)' \boldsymbol{\Sigma}_0 \{c_t^{1/2}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0)\} + o_{\mathbf{P}}(1).$$

The result then follows from theorem 1 and the continuous mapping theorem. □

4. The proof of theorem 1

The proof is through a sequence of lemmas. We start by a technical result concerning uniform convergence, go on by proving the consistency of $\hat{\boldsymbol{\beta}}_t$, and finally consider the asymptotic

normality. As mentioned earlier, the steps of the proof are modelled after Andersen & Gill (1982), but our proof is often simpler. Conditions (C1)–(C3) are assumed to hold.

Lemma 1

First recall the notation in (3.1). There exists a continuous monotone function γ*:* $[0, 1] \to [0, 1]$ *with* $\gamma(0)=0, \gamma(1)=1$*, such that*

$$c_t^{-1} \boldsymbol{I}_t(\boldsymbol{\beta}, u) \xrightarrow[P]{} \gamma(u)\boldsymbol{\Sigma}(\boldsymbol{\beta}) \tag{4.1}$$

uniformly on $\mathscr{B} \times [0, 1]$.

Proof. By (3.2) and (3.5), the sequence

$$c_t^{-1}\mathbf{I}_t(\boldsymbol{\beta}, u) = \{c_t^{-1} c_t(u)\} \cdot \{c_{[ut]}^{-1}\mathbf{I}(\boldsymbol{\beta}, [ut])\}, \qquad t \geqslant 1, \tag{4.2}$$

has a limit (in probability) of the form (4.1) uniformly over $\boldsymbol{\beta} \in \mathscr{B}$ for each $0 \leqslant u \leqslant 1$. (We adopt the convention $0/0 = 0$ in (4.2).) To prove that the convergence is uniform over $(\boldsymbol{\beta}, u) \in \mathscr{B} \times [0, 1]$ we note that the family $\gamma(\cdot)\boldsymbol{\Sigma}(\boldsymbol{\beta})$, $\boldsymbol{\beta} \in \mathscr{B}$, is equicontinuous on [0, 1] by (C1). The sample paths of the diagonal elements $c_t^{-1}\{\mathbf{I}_t(\boldsymbol{\beta}, u)\}_{kk}$, $0 \leqslant u \leqslant 1$, are monotone for each k, t and $\boldsymbol{\beta}$ so that the uniformity holds for these elements. (Choose a finite partition $0 = u_0 < u_1 < \cdots < u_l = 1$ of the interval [0, 1] such that the increase in $\gamma(\cdot)\boldsymbol{\Sigma}(\boldsymbol{\beta})$ on any subinterval $[u_{i-1}, u_i]$ is below a given small constant $\varepsilon > 0$ uniformly over $\boldsymbol{\beta} \in \mathscr{B}$. Note that the convergence is uniform over $\mathscr{B} \times \{u_0, \ldots, u_l\}$.) For the off-diagonal elements, the uniformity follows from this by using the Cauchy criterion for convergence and then applying Cauchy–Schwarz inequality on the terms (2.14). □

Lemma 2

There exist differentiable functions $s: \mathscr{B} \to \boldsymbol{R}$ *and* $\boldsymbol{a}: \mathscr{B} \to \boldsymbol{R}^p$ *such that*

$$\begin{aligned} c_t^{-1} S(\boldsymbol{\beta}, t) &\xrightarrow[P]{} s(\boldsymbol{\beta}), \\ c_t^{-1} A(\boldsymbol{\beta}, t) &\xrightarrow[P]{} \boldsymbol{a}(\boldsymbol{\beta}), \end{aligned} \tag{4.3}$$

uniformly over $\boldsymbol{\beta} \in \mathscr{B}$*. Furthermore,*

$$\begin{aligned} \boldsymbol{a}(\boldsymbol{\beta}) &= \frac{\partial}{\partial \boldsymbol{\beta}} s(\boldsymbol{\beta}), \\ \boldsymbol{\Sigma}(\boldsymbol{\beta}) &= \frac{\partial^2}{\partial \boldsymbol{\beta}^2} s(\boldsymbol{\beta}). \end{aligned} \tag{4.4}$$

Proof. See Appendix 2. □

Note that $\{\gamma(u) - \gamma(u')\}\boldsymbol{\Sigma}_0$ is positive definite for each pair $0 \leqslant u' < u \leqslant 1$ and thus $u \mapsto \gamma(u)\boldsymbol{\Sigma}_0$ can appear as a covariance function of a p-dimensional Gaussian martingale on [0, 1].

Lemma 3

(*Consistency of* $\hat{\boldsymbol{\beta}}_t$)

We have

$$\hat{\boldsymbol{\beta}}_t \xrightarrow[P]{} \boldsymbol{\beta}_0. \tag{4.5}$$

Proof. From (2.7) we see that the log-likelihood $l_t(\boldsymbol{\beta}, 1) = l(\boldsymbol{\beta}, t)$ is a concave function of $\boldsymbol{\beta}$ with a (with probability tending to 1) unique maximum at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_t$ (by the definition of $\hat{\boldsymbol{\beta}}_t$). Now consider the process

$$k_t(\boldsymbol{\beta}, u) \underset{\text{def}}{=} \sum_{v=1}^{[ut]} \sum_{j \geqslant 1} Y_j(v-1)[\boldsymbol{\beta}'\mathbf{Z}_j(v-1)L\{\boldsymbol{\beta}_0'\mathbf{Z}_j(v-1)\} - g\{\boldsymbol{\beta}'\mathbf{Z}_j(v-1)\}]$$

$$= \boldsymbol{\beta}'\mathbf{A}_t(\boldsymbol{\beta}_0, u) - S_t(\boldsymbol{\beta}, u), \qquad 0 \leqslant u \leqslant 1. \tag{4.6}$$

Then, for each $\boldsymbol{\beta} \in \mathscr{B}$,

$$c_t^{-1}\{l_t(\boldsymbol{\beta}, u) - k_t(\boldsymbol{\beta}, u)\} = c_t^{-1} \sum_{v=1}^{[ut]} \sum_{j \geqslant 1} Y_j(v-1)\boldsymbol{\beta}'\mathbf{Z}_j(v-1)\Delta M_j(v), \qquad 0 \leqslant u \leqslant 1,$$

is a local square integrable $(\mathbf{P}, \mathscr{G}_{[ut]})$ martingale with variance process

$$\langle c_t^{-1}\{l_t(\boldsymbol{\beta}, \cdot) - k_t(\boldsymbol{\beta}, \cdot)\}\rangle(u)$$

$$= c_t^{-2} \sum_{v=1}^{[ut]} \sum_{j \geqslant 1} Y_j(v-1)\{\boldsymbol{\beta}'\mathbf{Z}_j(v-1)\}^2 \Delta\langle M_j\rangle(v) \qquad \text{(by (2.13))}$$

$$= c_t^{-1}\boldsymbol{\beta}'\left[c_t^{-1} \sum_{v=1}^{[ut]} \sum_{j \geqslant 1} Y_j(v-1)\mathbf{Z}_j(v-1)^{\otimes 2} V\{\boldsymbol{\beta}_0'\mathbf{Z}_j(v-1)\}\right]\boldsymbol{\beta} \qquad \text{(by (2.12))}$$

$$= c_t^{-1}\boldsymbol{\beta}'\{c_t^{-1}\mathbf{I}_t(\boldsymbol{\beta}_0, u)\}\boldsymbol{\beta} \qquad \text{(by (2.14), (2.15))} \tag{4.7}$$

In (4.7) we have used the matrix notation $\mathbf{a}^{\otimes 2} = \{a_h a_k\}_{1 \leqslant h, k \leqslant p}$ and the fact that $\mathbf{x}'\mathbf{a}^{\otimes 2}\mathbf{x} = (\mathbf{x}'\mathbf{a})^2$ for any $\mathbf{a}, \mathbf{x} \in \mathbf{R}^p$. On the other hand, by (4.6) and lemma 2, it follows that for each $\boldsymbol{\beta} \in \mathscr{B}$

$$c_t^{-1}k_t(\boldsymbol{\beta}, 1) \underset{\mathbf{P}}{\to} \boldsymbol{\beta}'\mathbf{a}(\boldsymbol{\beta}_0) - s(\boldsymbol{\beta}) \underset{\text{def}}{=} l_\infty(\boldsymbol{\beta}, 1), \tag{4.8}$$

while by (4.7) and (C1),

$$c_t\langle c_t^{-1}\{l_t(\boldsymbol{\beta}, \cdot) - k_t(\boldsymbol{\beta}, \cdot)\}\rangle(1) \underset{\mathbf{P}}{\to} \boldsymbol{\beta}'\boldsymbol{\Sigma}_0\boldsymbol{\beta}. \tag{4.9}$$

But using the inequality of Lenglart (see, e.g., Andersen & Gill (1982, p. 1115)), together with (4.8) and (4.9), implies that

$$c_t^{-1}l_t(\boldsymbol{\beta}, 1) \underset{\mathbf{P}}{\to} l_\infty(\boldsymbol{\beta}, 1) \tag{4.10}$$

for each $\boldsymbol{\beta} \in \mathscr{B}$. Further, by (4.4) and (4.8)

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} l_\infty(\boldsymbol{\beta}, 1) &= \mathbf{a}(\boldsymbol{\beta}_0) - \mathbf{a}(\boldsymbol{\beta}) \\ \frac{\partial^2}{\partial \boldsymbol{\beta}^2} l_\infty(\boldsymbol{\beta}, 1) &= -\boldsymbol{\Sigma}(\boldsymbol{\beta}). \end{aligned} \tag{4.11}$$

Since the matrix $\boldsymbol{\Sigma}(\boldsymbol{\beta})$ is positive semidefinite for each $\boldsymbol{\beta} \in \mathscr{B}$ and $\boldsymbol{\Sigma}_0$ is positive definite, the claim follows by exactly the same convex analysis arguments as were used in the proof of lemma 3.1 in Andersen & Gill (1982). □

We now turn to the asymptotic normality of $\hat{\boldsymbol{\beta}}_t$. In the usual way, let $D[0, 1]$ be the space of real-valued right continuous functions with left limits, endowed with Skorohod topology. Let $\mathbf{G}(u) = \{G_1(u), \ldots, G_p(u)\}$, $0 \leqslant u \leqslant 1$, be a Gaussian martingale with covariance function

$\mathrm{Cov}(\mathbf{G})=\gamma(\cdot)\boldsymbol{\Sigma}_0$, defined on some stochastic basis $(\Omega', \mathscr{F}', (\mathscr{F}'_u)_{0\leqslant u\leqslant 1}, \mathbf{P}')$ satisfying the usual conditions. We denote by $I(A)$ the indicator of a set $A\subset\Omega$.

Lemma 4

The sequence of processes $c_t^{-1/2}\boldsymbol{U}_t(\boldsymbol{\beta}_0, \cdot)$, $t\in\boldsymbol{N}$, *converges in distribution (in the product space* $(D[0, 1])^p$*) towards* $\boldsymbol{G}$. *In particular*

$$c_t^{-1/2}\boldsymbol{U}_t(\boldsymbol{\beta}_0, 1) \underset{\mathscr{D}}{\to} \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_0), \tag{4.12}$$

where $\boldsymbol{N}$ *stands for the normal distribution on* $\boldsymbol{R}^p$.

Proof. First we show that

$$c_t^{-1/2}U_{ti}(\boldsymbol{\beta}_0, \cdot) \underset{\mathscr{D}}{\to} G_i, \qquad 1\leqslant i\leqslant p. \tag{4.13}$$

We have

$$\langle c_t^{-1/2}U_{ti}(\boldsymbol{\beta}_0, \cdot)\rangle(u)=c_t^{-1}I_{tii}(\boldsymbol{\beta}_0, u) \underset{\mathbf{P}}{\to} \gamma(u)\boldsymbol{\Sigma}_{0ii} \tag{4.14}$$

for $0\leqslant u\leqslant 1$, by lemma 1. Further the formulas

$$\bar{I}^{\varepsilon}_{ti}(v-1)=\underset{\text{def}}{I}\left[\left\{\sum_{j\geqslant 1} Y_j(v-1)|Z_{ji}(v-1)|\geqslant c_t^{1/2}\varepsilon\right\}\right], \qquad v\geqslant 1,$$

$$\bar{U}^{\varepsilon}_{ti}(\boldsymbol{\beta}_0, u)=\underset{\text{def}}{\sum_{v=1}^{[ut]}}\sum_{j\geqslant 1} Y_j(v-1)Z_{ji}(v-1)\bar{I}^{\varepsilon}_{ti}(v-1)\Delta M_j(v), \qquad 0\leqslant u\leqslant 1,$$

define the ε-decomposition

$$c_t^{-1/2}\mathbf{U}_{ti}(\boldsymbol{\beta}_0, \cdot)=c_t^{-1/2}\bar{\mathbf{U}}^{\varepsilon}_{ti}(\boldsymbol{\beta}_0, \cdot)+c_t^{-1/2}\{\mathbf{U}_{ti}(\boldsymbol{\beta}_0, \cdot)-\bar{\mathbf{U}}^{\varepsilon}_{ti}(\boldsymbol{\beta}_0, \cdot)\}$$

of the martingale $c_t^{-1/2}U_{ti}(\boldsymbol{\beta}_0, \cdot)$ in the sense of Gill (1980, pp. 16–17). Clearly, for arbitrary $\delta>0$,

$$\begin{aligned}&\mathbf{P}\{\langle c_t^{-1/2}\bar{U}^{\varepsilon}_{ti}(\boldsymbol{\beta}_0, \cdot)\rangle(u)>\delta\}\\ &\quad\leqslant\mathbf{P}\left[\bigcup_{v=1}^{\infty}\{\bar{I}^{\varepsilon}_{ti}(v-1)>0\}\right]\\ &\quad\leqslant\mathbf{P}(\hat{Z}\geqslant c_t^{1/2}\varepsilon)\to 0, \qquad 0\leqslant u\leqslant 1,\end{aligned} \tag{4.15}$$

by (C2). Now (4.13) follows from (4.14), (4.15) and Rebolledo's martingale CLT (a suitable form is, e.g., in Gill (1980, theorem 2.4.1)).

The second part of the proof consists of deriving the claim of the lemma from (4.13). The above mentioned martingale CLT will again be applied, now in a similar manner as in the proof of theorem 3.5 in Rebolledo (1978, pp. 39–40).

We need some new notation. Let $\mathbf{h}=(h_1, \ldots, h_p)$ be a vector of simple step functions h_i: $[0, 1]\to\mathbf{R}$. There is a finite partition $0=t_0<t_1<\cdots<t_{m+1}=1$ of $[0, 1]$ such that each h_i, $1\leqslant i\leqslant p$, remains constant on every subinterval $[t_k, t_{k+1})$, $0\leqslant k\leqslant m$. Define, for $0\leqslant u\leqslant 1$,

$$H_{t,i}(u)=c_t^{-1/2}\int_0^u h_i(s)\,dU_{ti}(\boldsymbol{\beta}_0, s),$$

$$H_i(u)=\int_0^u h_i(s)\,dG_i(s),$$

$$H_t(u)=\sum_{i=1}^{p} H_{t,i}(u), \quad H(u)=\sum_{i=1}^{p} H_i(u).$$

By (4.13) and lemma A3 in Rebolledo (1978, p. 67), it suffices to prove that $H_t(1) \underset{\mathscr{D}}{\to} H(1)$. In fact, we shall prove

$$H_t \underset{\mathscr{D}}{\to} H \tag{4.16}$$

in $D[0, 1]$.

Now for given $0 \leqslant u \leqslant 1$,

$$\langle H_t\rangle(u)=\sum_{i,k=1}^{p}\langle H_{ti}, H_{tk}\rangle(u)=\sum_{i,k=1}^{p}\int_0^u h_i(s)h_k(s)\, d\{c_t^{-1}I_{tik}(\boldsymbol{\beta}_0, s)\}$$

and

$$\langle H\rangle(u)=\sum_{i,k=1}^{p}\langle H_i, H_k\rangle(u)=\sum_{i,k=1}^{p}\int_0^u h_i(s)h_k(s)\boldsymbol{\Sigma}_{0ik}\, d\gamma(s)$$

and merely linear combinations of the values the processes $c_t^{-1}I_{tik}(\boldsymbol{\beta}_0, \cdot)$ and $\gamma(\cdot)\boldsymbol{\Sigma}_{0ik}$, $0 \leqslant i, k \leqslant p$, taken at points $t_0, t_1, \ldots, t_{m+1}$. Thus lemma 1, and the fact that $\gamma(\cdot)\boldsymbol{\Sigma}_0$ is deterministic, together imply

$$\langle H_t\rangle(u) \underset{\mathbf{P}}{\to} \langle H\rangle(u), \qquad 0 \leqslant u \leqslant 1. \tag{4.17}$$

Denote

$$\|\mathbf{h}\| = \sup_{\text{def } 0\leqslant u\leqslant 1} \|\mathbf{h}(u)\|,$$

and

$$\bar{I}_t^{\varepsilon}(v-1)=I\left[\left\{p\|\mathbf{h}\|\sum_{j\geqslant 1} Y_j(v-1)\|\mathbf{Z}_j(v-1)\| \geqslant c_t^{1/2}\varepsilon\right\}\right],$$

$$\bar{H}_{ti}^{\varepsilon}(u)=\underset{\text{def}}{c_t^{-1/2}}\int_0^u h_i(s)\bar{I}_t^{\varepsilon}([st]-1)\, dU_{ti}(\boldsymbol{\beta}_0, s)$$

$$=c_t^{-1/2}\sum_{v=1}^{[ut]}\sum_{j\geqslant 1} Y_j(v-1)h_i\left(\frac{v}{t}\right)Z_{ji}(v-1)\bar{I}_t^{\varepsilon}(v-1)\Delta M_j(v),$$

$$\bar{H}_t^{\varepsilon}(u)=\underset{\text{def}}{\sum_{i=1}^{p}}\bar{H}_{ti}^{\varepsilon}.$$

It is then easy to see that $H_t=\bar{H}_t^{\varepsilon}+(H_t-\bar{H}_t^{\varepsilon})$ is an ε-decomposition for the martingale H_t. An argument similar to (4.15) shows that

$$\langle \bar{H}_{ti}^{\varepsilon}, \bar{H}_{tk}^{\varepsilon}\rangle(u)=c_t^{-1}\sum_{v=1}^{[ut]}\bar{I}_t^{\varepsilon}(v-1)h_i\left(\frac{v}{t}\right)h_k\left(\frac{v}{t}\right)\sum_{j\geqslant 1} Z_{ji}(v-1)Z_{jk}(v-1)\Delta\langle M_j\rangle(v) \underset{\mathbf{P}}{\to} 0, \qquad 0\leqslant u\leqslant 1,$$

for all $1 \leqslant i, k \leqslant p$. Clearly this implies

$$\langle \bar{H}_t^{\varepsilon}\rangle(u)=\sum_{i,k=1}^{p}\langle \bar{H}_{ti}^{\varepsilon}, \bar{H}_{tk}^{\varepsilon}\rangle(u) \underset{\mathbf{P}}{\to} 0, \qquad 0\leqslant u\leqslant 1. \tag{4.18}$$

Now (4.16) follows from (4.17), (4.18) and Rebolledo's theorem. □

The next lemma is slightly more general than is needed here for proving theorem 1. We need this more general formulation later in the proof of lemma 7.

Lemma 5

Let $\boldsymbol{\beta}_t$: $[0, 1]\times\Omega\to\boldsymbol{R}^p$ *and* λ_t: $[0, 1]\times\Omega\to[0, 1]$, $t\geqslant 1$, *be random functions such that the following conditions hold:*

$$\|\boldsymbol{\beta}_t(u)-\boldsymbol{\beta}_0\|\leqslant\|\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta}_0\|, \qquad 0\leqslant u\leqslant 1, \quad t\geqslant 1, \tag{4.19a}$$

and

$$\lambda_t(u)\xrightarrow[\mathbf{P}]{}\lambda(u) \tag{4.19b}$$

uniformly in u, $0\leqslant u\leqslant 1$, *where* λ: $[0, 1]\to[0, 1]$ *is non-random. Then*

$$c_t^{-1}\boldsymbol{I}_t\{\boldsymbol{\beta}_t(u), \lambda_t(u)\}\xrightarrow[\mathbf{P}]{}\gamma\{\lambda(u)\}\boldsymbol{\Sigma}_0$$

uniformly in u, $0\leqslant u\leqslant 1$.

Proof. It suffices to show that

$$\Delta_{t,1}=\sup_{\text{def } 0\leqslant u\leqslant 1}\|c_t^{-1}\mathbf{I}_t\{\boldsymbol{\beta}_t(u), \lambda_t(u)\}-\gamma\{\lambda_t(u)\}\boldsymbol{\Sigma}\{\boldsymbol{\beta}_t(u)\}\|\xrightarrow[\mathbf{P}]{}0$$

and

$$\Delta_{t,2}=\sup_{\text{def } 0\leqslant u\leqslant 1}\|\gamma\{\lambda_t(u)\}\boldsymbol{\Sigma}\{\boldsymbol{\beta}(u)\}-\gamma\{\lambda(u)\}\boldsymbol{\Sigma}_0\|\xrightarrow[\mathbf{P}]{}0.$$

Let $\varepsilon>0$. First, by lemmas 1 and 3 and assumption (4.19a),

$$\mathbf{P}(\Delta_{t,1}>\varepsilon)\leqslant\mathbf{P}(\hat{\boldsymbol{\beta}}_t\notin\mathscr{B})+\mathbf{P}(\Delta_{t,1}>\varepsilon, \hat{\boldsymbol{\beta}}_t\in\mathscr{B})$$

$$\leqslant\mathbf{P}(\hat{\boldsymbol{\beta}}_t\notin\mathscr{B})+\mathbf{P}\left\{\sup_{\boldsymbol{\beta}\in\mathscr{B};0\leqslant u\leqslant 1}\|c_t^{-1}\mathbf{I}_t(\boldsymbol{\beta}, u)-\gamma(u)\boldsymbol{\Sigma}(\boldsymbol{\beta})\|>\varepsilon\right\}\to 0.$$

Second, there exist $\delta_1>0$ and $\delta_2>0$ such that $\|\gamma(u')\boldsymbol{\Sigma}(\boldsymbol{\beta})-\gamma(u)\boldsymbol{\Sigma}_0\|<\varepsilon$ whenever $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|<\delta_1$ and $|u-u'|<\delta_2$ (use the triangle inequality). Consequently, by using the assumptions (4.19a–b) and lemma 3,

$$\mathbf{P}(\Delta_{t,2}>\varepsilon)\leqslant\mathbf{P}(\|\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta}_0\|\geqslant\delta_1)+\mathbf{P}\left\{\sup_{0\leqslant u\leqslant 1}|\lambda_t(u)-\lambda(u)|\geqslant\delta_2\right\}$$

$$+\mathbf{P}\left\{\Delta_{t,2}>\varepsilon, \|\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta}_0\|<\delta_1, \sup_{0\leqslant u\leqslant 1}|\lambda_t(u)-\lambda(u)|<\delta_2\right\}$$

$$=\mathbf{P}(\|\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta}_0\|\geqslant\delta_1)+\mathbf{P}\left\{\sup_{0\leqslant u\leqslant 1}|\lambda_t(u)-\lambda(u)|\geqslant\delta_2\right\}\to 0.$$

□

Now we can give *the proof of theorem 1*: Taylor expanding $\mathbf{U}_t(\boldsymbol{\beta}, 1)$ around $\boldsymbol{\beta}_0$ we get

$$\mathbf{U}_t(\boldsymbol{\beta}, 1)-\mathbf{U}_t(\boldsymbol{\beta}_0, 1)=-\mathbf{I}_t(\boldsymbol{\beta}_t^*, 1)(\boldsymbol{\beta}-\boldsymbol{\beta}_0), \tag{4.20}$$

where $\boldsymbol{\beta}_t^*$ is on the line segment between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$. Inserting $\hat{\boldsymbol{\beta}}_t$ in (4.20) we get

$$c_t^{-1/2}\mathbf{U}_t(\boldsymbol{\beta}_0, 1)=\{c_t^{-1}\mathbf{I}_t(\boldsymbol{\beta}_t^*, 1)\}\{c_t^{1/2}(\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta}_0)\}, \tag{4.21}$$

since $\mathbf{U}_t(\hat{\boldsymbol{\beta}}_t, 1)=\mathbf{0}$. Now $\|\boldsymbol{\beta}_t^*-\boldsymbol{\beta}_0\|\leqslant\|\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta}_0\|$, and then

$$c_t^{-1}\mathbf{I}_t(\boldsymbol{\beta}_t^*, 1)\xrightarrow[\mathbf{P}]{}\boldsymbol{\Sigma}_0 \tag{4.22}$$

by lemma 5. Using (4.12), (4.21) and theorem 10.1 in Billingsley (1961) we can write

$$c_t^{-1/2}\mathbf{U}_t(\boldsymbol{\beta}_0, 1)=\boldsymbol{\Sigma}_0 c_t^{1/2}(\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta}_0)+0_{\mathbf{P}}(1). \tag{4.23}$$

The claim (3.6) now follows from lemma 4. □

5. Asymptotic behaviour of the empirical score process

A key step in proving the asymptotic normality of the ML estimator $\hat{\boldsymbol{\beta}}_t$ above was to show that the score process $\mathbf{U}_t(\boldsymbol{\beta}_0, \cdot)$ corresponding to the correct parameter value, normalized by c_t, is asymptotically a Gaussian martingale (lemma 4). But the score process itself can be of independent interest. For example, Sim (1981) and Whitehead (1983) used the score process for sequential testing in the case of independent sampling. Another possibility is to use the score process as an indicator of goodness-of-fit. For applications of this, see Wei (1984) (to test the proportionality of hazards in a two-sample case), and McLeish (1984) (to test a transition matrix in an aggregate Markov chain model). It is this latter type of use we shall be concerned with in this section. Also recall from the beginning of section 3 the possibility of using the total time on test as the time parameter.

We consider the process $\mathbf{w}'\mathbf{U}_t(\boldsymbol{\beta}, \cdot)$, evaluated at the MLE $\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_t$, as $t\to\infty$. Here $\mathbf{w}$ can be any p-vector such that $\|\mathbf{w}\|>0$. Typically, $\mathbf{w}$ could be one of the unit vectors. Thus the unknown parameter value $\boldsymbol{\beta}_0$ in $\mathbf{U}_t(\boldsymbol{\beta}_0, \cdot)$ is here replaced by a function of the data. Similarly, we change the normalizing constants c_t, $t\geqslant 1$, used in lemma 4 into the empirical variance $\mathbf{w}'\mathbf{I}_t(\hat{\boldsymbol{\beta}}_t, 1)\mathbf{w}$, $t\geq 1$ (cf. lemma 5). Finally, by making a data dependent time change on the argument u, we can derive a "functional statistic" with known standard limit, in the sense of convergence in distribution in $D[0, 1]$.

Before proceeding further, note that, by the definition of $\hat{\boldsymbol{\beta}}_t$, we have $\mathbf{U}_t(\hat{\boldsymbol{\beta}}_t, 1)=\mathbf{0}$. Thus, although $\hat{\boldsymbol{\beta}}_t$ is consistent as $t\to\infty$, one cannot expect $\mathbf{w}'\mathbf{U}_t(\hat{\boldsymbol{\beta}}_t, \cdot)$ to behave in the limit exactly like $\mathbf{w}'\mathbf{U}_t(\boldsymbol{\beta}_0, \cdot)$. Instead, the property $\mathbf{w}'\mathbf{U}_t(\hat{\boldsymbol{\beta}}_t, 1)=0$ makes the limit to be a *Brownian bridge* (up to a time change).

We then go on by discussing the needed time change. Define the function $\Psi_t(\boldsymbol{\beta}, \cdot)$: $[0, 1]\to[0, 1]$ by

$$\Psi_t(\boldsymbol{\beta}, u)=\frac{\mathbf{w}'\mathbf{I}_t(\boldsymbol{\beta}, u)\mathbf{w}}{\mathbf{w}'\mathbf{I}_t(\boldsymbol{\beta}, 1)\mathbf{w}}, \qquad 0\leqslant u\leqslant 1. \tag{5.1}$$

Clearly, $\Psi_t(\boldsymbol{\beta}, \cdot)$ is right continuous and non-decreasing with $\Psi_t(\boldsymbol{\beta}, 0)=0$ and $\Psi_t(\boldsymbol{\beta}, 1)=1$. Let $\Psi_t^{-1}(\boldsymbol{\beta}, \cdot)$ be the right continuous inverse function of $\Psi_t(\boldsymbol{\beta}, \cdot)$, i.e.

$$\Psi_t^{-1}(\boldsymbol{\beta}, u)=\inf\{0\leqslant s\leqslant 1 | \Psi_t(\boldsymbol{\beta}, s)>u\}. \tag{5.2}$$

We get directly from (4.1) that

$$\Psi_t(\boldsymbol{\beta}, u)\underset{\mathbf{P}}{\to}\gamma(u) \tag{5.3}$$

uniformly in $\mathscr{B}\times[0, 1]$ (by restricting the neighbourhood $\mathscr{B}$ if necessary), and further

$$\Psi_t^{-1}(\boldsymbol{\beta}, u)\underset{\mathbf{P}}{\to}\gamma^{-1}(u) \tag{5.4}$$

uniformly in $\mathscr{B}\times[0, 1]$.

We denote by W, W^0 and i, respectively, the Wiener process, the Brownian bridge and the identity mapping $(i(u)=u)$ on $[0, 1]$. For stochastic processes on $[0, 1]$, unless otherwise stated, the convergence "$\underset{\mathscr{D}}{\to}$" or "$\underset{\mathbf{P}}{\to}$" takes place in $D[0, 1]$ endowed with Skorohod topology (or in $D[0, 1]\times D[0, 1]$ for bivariate processes).

Consider then the weighted empirical score corresponding to the ML estimate $\hat{\boldsymbol{\beta}}_t$, time-changed according to $\Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, \cdot)$ and standardized by the empirical variance. Let

$$B_t(u) \underset{\text{def}}{=} \{\mathbf{w}'\mathbf{I}_t(\hat{\boldsymbol{\beta}}_t, 1)\mathbf{w}\}^{-1/2}\mathbf{w}'\mathbf{U}_t\{\hat{\boldsymbol{\beta}}_t, \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, u)\}, \qquad 0 \leqslant u \leqslant 1. \tag{5.5}$$

Our second main result is then:

Theorem 2
Under conditions (C1)–(C3),

$$B_t \underset{\mathscr{D}}{\to} W^0. \tag{5.6}$$

Proof of theorem 2. Write

$$B_t(u) = B_t^0(u) + \{B_t(u) - B_t^0(u)\}, \tag{5.7}$$

where

$$B_t^0(u) \underset{\text{def}}{=} \{\mathbf{w}'\mathbf{I}_t(\hat{\boldsymbol{\beta}}_t, 1)\mathbf{w}\}^{-1/2}[\mathbf{w}'\mathbf{U}_t\{\boldsymbol{\beta}_0, \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, u)\} - u\mathbf{w}'\mathbf{U}_t(\boldsymbol{\beta}_0, 1)]. \tag{5.8}$$

We show below, in lemmas 6 and 7, that

$$B_t^0(\cdot) \underset{\mathscr{D}}{\to} W(\cdot) - i(\cdot)W(1) \tag{5.9}$$

and

$$\sup_{0 \leqslant u \leqslant 1} |B_t(u) - B_t^0(u)| \underset{\mathbf{P}}{\to} 0. \tag{5.10}$$

Since $W(u) - uW(1)$, $0 \leqslant u \leqslant 1$, is the Brownian bridge, the claim (5.6) follows from (5.7), (5.9) and (5.10) by applying theorem 4.1 in Billingsley (1968) to B_t^0 and B_t. □

Lemma 6
The convergence in (5.9) holds.

Proof. We start by showing that

$$c_t^{-1/2}\mathbf{w}'\mathbf{U}_t\{\boldsymbol{\beta}_0, \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, \cdot)\} \underset{\mathscr{D}}{\to} (\mathbf{w}'\boldsymbol{\Sigma}_0\mathbf{w})^{1/2}W(\cdot). \tag{5.11}$$

By the argument in section 17 of Billingsley (1968) it suffices to show that jointly

$$\{c_t^{-1/2}\mathbf{w}'\mathbf{U}_t(\boldsymbol{\beta}_0, \cdot), \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, \cdot)\} \underset{\mathscr{D}}{\to} [(\mathbf{w}'\boldsymbol{\Sigma}_0\mathbf{w})^{1/2}W\{\gamma(\cdot)\}, \gamma^{-1}(\cdot)].$$

This follows (by theorem 4.4 in Billingsley (1968)) if

$$c_t^{-1/2}\mathbf{w}'\mathbf{U}_t(\boldsymbol{\beta}_0, \cdot) \underset{\mathscr{D}}{\to} [(\mathbf{w}'\boldsymbol{\Sigma}_0\mathbf{w})^{1/2}W\{\gamma(\cdot)\}] \tag{5.12}$$

and

$$\Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, \cdot) \underset{\mathbf{P}}{\to} \gamma^{-1}(\cdot). \tag{5.13}$$

But (5.12) follows from lemma 4, and (5.13) is implied by (5.4) (for uniform convergence implies convergence in the Skorohod topology).

Then consider (5.9). By (3.7) we get

$$c_t^{-1}\mathbf{w}'\mathbf{I}_t(\hat{\boldsymbol{\beta}}_t, 1)\mathbf{w} \underset{\mathbf{P}}{\to} \mathbf{w}'\boldsymbol{\Sigma}_0\mathbf{w}. \tag{5.14}$$

Again using theorem 4.4 in Billingsley (1968), together with (5.11), implies that

$$\{\mathbf{w}'\mathbf{I}_t(\hat{\boldsymbol{\beta}}_t, 1)\mathbf{w}\}^{-1/2}\mathbf{w}'\mathbf{U}_t\{\boldsymbol{\beta}_0, \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, \cdot)\} \xrightarrow[\mathscr{D}]{} W(\cdot). \tag{5.15}$$

Finally, use the continuous mapping theorem for $h: D[0, 1] \to D[0, 1]$, defined by

$$(h \circ x)(u) = x(u) - ux(1), \qquad x \in D[0, 1],$$

establishing (5.9). □

Lemma 7

The convergence in (5.10) holds.

Proof. First observe that (cf. (4.20))

$$\mathbf{U}_t\{\boldsymbol{\beta}_0, \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, u)\} - \mathbf{U}_t\{\hat{\boldsymbol{\beta}}_t, \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, u)\} = \mathbf{I}_t\{\boldsymbol{\beta}^*_{t,u}, \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, u)\}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0) \tag{5.16a}$$

and

$$\mathbf{U}_t(\boldsymbol{\beta}_0, 1) - \mathbf{U}_t(\hat{\boldsymbol{\beta}}_t, 1) = \mathbf{I}_t(\boldsymbol{\beta}_t^{**}, 1)(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0) \tag{5.16b}$$

where $\boldsymbol{\beta}^*_{t,u}$ and $\boldsymbol{\beta}_t^{**}$ are on the line segment joining $\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}_t$. Therefore, and since $\mathbf{U}_t(\hat{\boldsymbol{\beta}}_t, 1) = \mathbf{0}$,

$$\begin{aligned} &c_t^{-1/2}\{\mathbf{w}'\mathbf{I}_t(\hat{\boldsymbol{\beta}}_t, 1)\mathbf{w}\}^{1/2}\{B_t^0(u) - B_t(u)\} \\ &\quad = c_t^{-1/2}[[\mathbf{w}'\mathbf{U}_t\{\boldsymbol{\beta}_0, \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, u)\} - \mathbf{w}'\mathbf{U}_t\{\hat{\boldsymbol{\beta}}_t, \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, u)\}] - \{u\mathbf{w}'\mathbf{U}_t(\boldsymbol{\beta}_0, 1) - u\mathbf{w}'\mathbf{U}_t(\hat{\boldsymbol{\beta}}_t, 1)\}] \\ &\quad = \mathbf{w}'[c_t^{-1}\mathbf{I}_t\{\boldsymbol{\beta}^*_{t,u}, \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, u)\} - uc_t^{-1}\mathbf{I}_t(\boldsymbol{\beta}_t^{**}, 1)] \cdot c_t^{1/2}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0). \end{aligned}$$

The claim then follows from (i)–(iv) below:

(i) $c_t^{1/2}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0) \xrightarrow[\mathscr{D}]{} \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_0^{-1})$ (theorem 1);

(ii) $c^{-1/2}\{\mathbf{w}'\mathbf{I}_t(\hat{\boldsymbol{\beta}}_t, 1)\mathbf{w}\}^{1/2} \xrightarrow[\mathbf{P}]{} (\mathbf{w}'\boldsymbol{\Sigma}_0\mathbf{w})^{1/2}$ ((3.7));

(iii) $uc_t^{-1}\mathbf{I}_t(\boldsymbol{\beta}_t^{**}, 1) \xrightarrow[\mathbf{P}]{} u\boldsymbol{\Sigma}_0$ uniformly in u, $0 \leqslant u \leqslant 1$

(apply lemma 5 for $\lambda_t(\cdot) = \lambda(\cdot) = 1$ and $\boldsymbol{\beta}_t(\cdot) = \boldsymbol{\beta}_t^{**}$);

(iv) $c_t^{-1}\mathbf{I}_t\{\boldsymbol{\beta}^*_{t,u}, \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, u)\} \xrightarrow[\mathbf{P}]{} u\boldsymbol{\Sigma}_0$ uniformly in u, $0 \leqslant u \leqslant 1$

(apply lemma 5 for $\lambda_t(\cdot) = \Psi_t^{-1}(\hat{\boldsymbol{\beta}}_t, \cdot)$, $\lambda(\cdot) = \gamma^{-1}(\cdot)$ and $\boldsymbol{\beta}_t(u) = \boldsymbol{\beta}^*_{t,u}$).

Acknowledgements

This work was done while the first author was visiting the Fred Hutchinson Cancer Research Center in Seattle, Washington. The work was supported in part by the National Institutes of Health Grants 5R01-GM28314 and 5R01-GM-24472. The work of the second author was supported by the Finnish Academy. We are grateful to Richard Gill for pointing out a flaw in our original proof of lemma 4.

Appendix 1: Detailed model assumptions

Consider a statistical model $\{\mathbf{P}^{\vartheta}; \vartheta \in \Theta\}$ for the observation process $[R(v-1), \{\Delta N_j(v), \mathbf{Z}_j(v-1); j \in R(v-1)\}]_{v \geqslant 1}$ and a $\mathbf{P}^{\vartheta}$ likelihood which corresponds to data collected up to time t, $t \geqslant 1$.

Suppose that the parameter ϑ can be represented in the form $\vartheta = (\vartheta_1, \vartheta_2)$, where ϑ_1 is the parameter of interest and ϑ_2 is a nuisance parameter. Typically, we think of ϑ_1 as parametrizing the conditional distribution of the variables $\Delta N_j(v)$, conditioned on $\mathscr{G}_{v-1}$, and of ϑ_2 as the

parameter associated with the conditional law of the variables $R(v)$ and $\mathbf{Z}_j(v)\{j \in R(v)\}$, given $\mathscr{F}_{v-1}$. The full likelihood corresponding to the observed values $[r(v-1), \{\Delta n_j(v), \mathbf{z}_j(v-1); j \in r(v-1)\}; v \leqslant t]$ can then be expressed as the product of two terms, viz. as

$$\prod_{v \leqslant t} \mathbf{P}^{\vartheta}\{R(v-1)=r(v-1), \mathbf{Z}_j(v-1)=\mathbf{z}_j(v-1); j \in r(v-1) | \mathscr{F}_{v-1}\}$$
$$\cdot \prod_{v \leqslant t} \mathbf{P}^{\vartheta}\{\Delta N_j(v)=\Delta n_j(v); j \in r(v-1) | \mathscr{G}_{v-1}\}. \qquad \text{(A.1)}$$

Following Cox (1975), the second factor can be called a partial likelihood. Ordinary ML estimation of ϑ_1, the parameter of interest, can be done by considering that factor alone provided that the following condition holds:

Assumption 1. (i) For each $v \geqslant 1$, the conditional $\mathbf{P}^{\vartheta}$ distribution of $[R(v-1), \{\mathbf{Z}_j(v-1); j \in R(v-1)\}]$ given $\mathscr{F}_{v-1}$ does not depend on ϑ_1; and (ii) for each $v \geqslant 1$, the conditional $\mathbf{P}^{\vartheta}$ distribution of $\{\Delta N_j(v); j \in R(v-1)\}$, given $\mathscr{G}_{v-1}$, does not depend on ϑ_2.

Of course, the validity of assumption 1 depends on the model $\{\mathbf{P}^{\vartheta}; \vartheta \in \Theta\}$. Actual verification of this assumption would require that the model were fully specified, including the probability law of the censoring mechanism and possible random covariates. This is usually not done explicitly. However, part (ii) of assumption 1 becomes obvious if the censoring times and the covariates are fixed, or random but $\mathscr{F}_0$-measurable. More generally, we can consider (ii) to be valid if the censoring is non-informative about ϑ_1 and the covariates are external (cf. Kalbfleisch & Prentice (1980)). For internal covariates more caution is needed: if (i) is not met, also the first factor in (A.1) can depend on ϑ_1, and then using only the second factor in the maximization is a potential source of bias. Finally, it seems that part (ii) in assumption 1 can always be met in practice by making a convenient choice for ϑ_1, the parameter of interest.

For continuous time version of assumption 1, see Arjas & Haara (1984).

Our next assumption imposes an independence condition between the individuals and simplifies, in particular, the handling of ties.

Assumption 2. For each $v \geqslant 1$, and $\vartheta \in \Theta$, the random variables $\{\Delta N_j(v); j \geqslant 1\}$ are conditionally $\mathbf{P}^{\vartheta}$ independent given $\mathscr{G}_{v-1}$.

This assumption is likely to hold in practice if there are no multiple failures of common cause, or if such failures can occur but the background variable causing the failure can be included as a covariate.

Under assumptions 1 and 2, the likelihood function (A.1) depends on ϑ_1 only through the factor

$$\prod_{v \leqslant t} \prod_{j \in r(v-1)} \mathbf{P}^{\vartheta}\{\Delta N_j(v)=\Delta n_j(v) | \mathscr{G}_{v-1}\}. \qquad \text{(A.2)}$$

On the other hand, because of assumption 1 (ii), this expression does not depend on ϑ_2.

It remains to specify the conditional probabilities in (A.2). Our next assumption guarantees that all relevant information in $\mathscr{G}_{v-1}$, when used as a condition for the probability of $\{\Delta N_j(v)=\Delta n_j(v)\}$, is actually contained in the p-vector $\mathbf{Z}_j(v-1)$ and the indicator $Y_j(v-1)$.

Assumption 3. For all $v, j \geqslant 1$, and $\vartheta \in \Theta$, $\Delta N_j(v)$ and $\mathscr{G}_{v-1}$ are conditionally $\mathbf{P}^{\vartheta}$ independent given $Y_j(v-1)$ and $\mathbf{Z}_j(v-1)$.

As a last step, we specify the conditional probabilities according to the logistic regression model for binomial response. We also change the notation of the parameter, writing $\boldsymbol{\beta}=(\beta_1, \ldots, \beta_p)'$ instead of ϑ_1 and $\mathbf{P}^{\boldsymbol{\beta}}$ instead of $\mathbf{P}^{\vartheta}$.

Assumption 4. For all (j, v) such that $j \in R(v-1)$,

$$\log \frac{\mathbf{P}^{\beta}\{\Delta N_j(v)=1 \mid \mathcal{G}_{v-1}\}}{\mathbf{P}^{\beta}\{\Delta N_j(v)=0 \mid \mathcal{G}_{v-1}\}} = \boldsymbol{\beta}' \mathbf{Z}_j(v-1).$$

Appendix 2: The proof of lemma 2

Lemma 2 is an immediate consequence of the lemma below. Here we denote the derivative of a differentiable mapping f by f' and the norm in an n-dimensional real space by $\|\cdot\|_n$.

Lemma A

Let $B = B(b, r)$ be an open ball with centre b and radius r in $\mathbf{R}^p$. Let $(X_t)_{t \geq 1}$ be a sequence of $\mathbf{R}^n$-valued mappings of $B \times \Omega$ such that

(a) *$\omega \mapsto X_t(y, \omega)$ is $\mathcal{F}$-measurable for all $y \in B$ and $t \geq 1$;*

(b) *$y \mapsto X_t(y, \omega)$ is continuously differentiable in B for $\mathbf{P}$-almost all $\omega \in \Omega$ and all $t \geq 1$;*

(c) *there exist points $y_0 \in B$ and $f_0 \in \mathbf{R}^n$ such that $X_t(y_0) \xrightarrow[\mathbf{P}]{} f_0$;*

(d) *there exists a continuous $n \times p$ matrix valued mapping $f^{(1)}$ such that $X_t'(y) \xrightarrow[\mathbf{P}]{} f^{(1)}(y)$ uniformly in B.*

Then there exists a differentiable mapping $f: B \to \mathbf{R}^n$ such that

$$X_t(y) \xrightarrow[\mathbf{P}]{} f(y) \text{ uniformly in } B, \tag{A.1}$$

and

$$f' = f^{(1)}. \tag{A.2}$$

Proof. The continuity assumptions guarantee that all functions of ω below are $\mathcal{F}$-measurable. The assumptions (c) and (d) imply the existence of a sequence $0 < t_1 < t_2 < \cdots$ such that

$$\sup_{z \in B} \| X_{t_k}'(z) - f^{(1)}(z) \|_{pn} \to 0 \qquad \mathbf{P}\text{-a.s.} \tag{A.3}$$

and

$$\| X_{t_k}(y_0) - f_0 \|_n \to 0 \qquad \mathbf{P}\text{-a.s.} \tag{A.4}$$

Now, separately for each $\omega \in \Omega$ such that (A.3) and (A.4) hold, we can apply Dieudonne (1969, 8.6.3) and get that there exists a mapping $X_\infty : B \times \Omega \to \mathbf{R}^n$ satisfying

$$\sup_{z \in B} \| X_{t_k}(z) - X_\infty(z) \|_n \to 0 \qquad \mathbf{P}\text{-a.s.};$$

$$z \mapsto X_\infty(z) \text{ is differentiable} \qquad \mathbf{P}\text{-a.s.};$$

$$X_\infty' = f^{(1)} \qquad \mathbf{P}\text{-a.s.}$$

Because necessarily $X_\infty(y_0) = f_0$ $\mathbf{P}$-a.s. by (A.4) we see that $X_\infty = f \cap$ $\mathbf{P}$-a.s. for some differentiable $f: B \to \mathbf{R}^n$.

It remains to establish (A.1). Outside a $\mathbf{P}$-null set we have by (b) and the Mean Value Theorem (Dieudonne (1969, 8.5.4)) that

$$\begin{aligned} &\| X_t(y) - X_{t+h}(y) - \{X_t(y_0) - X_{t+h}(y_0)\} \|_n \\ &\quad \leq \| y - y_0 \|_p \sup_{z \in B} \| \{X_t(z) - X_{t+h}(z)\}' \|_{pn} \leq 2r \sup_{z \in B} \| X_t'(z) - X_{t+h}'(z) \|_{pn} \end{aligned}$$

for all $y \in B$, $t, h \geqslant 1$. Therefore, for any $\varepsilon > 0$

$$\mathbf{P}\left\{\sup_{y \in B} \| X_t(y) - X_{t+h}(y) \|_n > \varepsilon\right\}$$

$$\leqslant \mathbf{P}\left\{\| X_t(y_0) - X_{t+h}(y_0) \|_n > \frac{\varepsilon}{2}\right\} + \mathbf{P}\left\{\sup_{z \in B} \| X'_t(z) - X'_{t+h}(z) \|_{pn} > \frac{\varepsilon}{4r}\right\}.$$

This, together with the assumptions, implies (A.1). □

In order to obtain lemma 2, we apply lemma A to the sequence $c_t^{-1}\mathbf{A}_t$, $t \geqslant 1$ (with $n = p$) and thereafter to the sequence $c_t^{-1}S_t$, $t \geqslant 1$ (with $n = 1$).

References

Albert, A. & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71,** 1–10.

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

Arjas, E. (1985a). Contribution to the discussion on the paper by Andersen and Borgan. *Scand. J. Statist.* **12,** 150–153.

Arjas, E. (1985b). Stanford heart transplantation data revisited: a real time approach. In *Modern statistical methods in chronic disease epidemiology.* Wiley, New York.

Arjas, E. & Haara, P. (1984). A marked point process approach to censored failure data with complicated covariates. *Scand. J. Statist.* **11,** 193–209.

Billingsley, P. (1961). *Statistical inference for Markov processes.* University of Chicago Press, Chicago.

Billingsley, P. (1968). *Convergence of probability measures.* Wiley, New York.

Borgan, Ø. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Statist.* **11,** 1–16.

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

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

Dieudonne, J. (1969). *Foundations of modern analysis.* Academic Press, New York.

Gill, R. D. (1980). *Censoring and stochastic integrals.* MC Tracts, Mathematical Centre, Amsterdam.

Haberman, S. (1974). *The analysis of frequency data.* University of Chicago Press, Chicago.

Kalbfleisch, J. D. & Prentice, R. L. (1980). *The statistical analysis of failure time data.* Wiley, New York.

McLeish, D. L. (1984). Estimation for aggregate models: the aggregate Markov chain. *Canad. J. Statist.* **12,** 265–282.

Rebolledo, R. (1978). Sur les applications de la théorie des martingales a l'étude statistique d'une famille de processus ponctuels. *Lecture Notes in Mathematics* **636,** 27–70. Springer, Berlin.

Rebolledo, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete* **51,** 269–286.

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

Sim, D. A. (1981). A sequential score test. Ph.D. thesis, University of Washington.

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

Wei, L. J. (1984). Testing goodness-of-fit for proportional hazards model with censored observations. *J. Amer. Statist. Assoc* **79,** 649–652.

Whitehead, J. (1983). *The design and analysis of sequential clinical trials.* Wiley, New York.

Received October 1985, in final form September 1986

Elja Arjas, University of Oulu, Department of Applied Mathematics and Statistics, SF-90570 Oulu 57, Finland