PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 14, Fasc. 1 (1993), pp. 115–124

A PROCESS BASED ON STRATIFICATION FOR COX'S REGRESSION MODEL

BY

PAWEŁ MARZEC (WROCŁAW)

Abstract. A process based on stratification to check the proportional hazards assumption in a general Cox's regression model is considered. The stratification is based on a nonrandom choice of strata for individuals together with a random one generated by the behaviour of covariates at failure times. A formal description for the asymptotic performance of the process is given and some statistical applications are established.

1. Introduction. We are interested in the goodness of fit inference for Cox's [6] regression model of a general form of Andersen and Gill [3]. This model has been extensively studied in the statistical literature (see, e.g., [2], [3] and references therein). A class of procedures has been proposed to check the assumption of constant proportionality between two hazard rates, a special case of Cox's model. They include a method based on total time given by Gill and Schumacher [8] and a test of fit based on the score process proposed by Wei [18]. The graphical methods for assessing the goodness of fit in the model using covariates are discussed in Andersen [1], Arjas [4] and Crowley and Storer [7], whereas the procedures based on "partial residuals" and "score residuals" are proposed by Schoenfeld [15] and Thernau et al. [16], respectively. For other procedures see also Gray [9], Marzec and Marzec [11], McKeague and Utikal [12], Moreau et al. [13].

In the present paper we stratify data with respect to a nonrandom choice of the subset of the whole set of individuals under observation, and simultaneously a random one which is generated by the behaviour of covariates at failure times. Next we define the process based on this stratification to check the proportional hazards assumption. It should be mentioned that recently in [11] the problem of nonrandom stratification together with asymptotic behaviour of Arjas's [4] type processes has been considered. Now a random stratification based on covariates leads to the newly defined process that also has, under the assumed model of proportional hazards, the trajectories of "martingale type". This is a starting point for all further considerations, and therefore it should be noted that general ideas of this approach, to deal with stratified processes, were firstly proposed and explored by Arjas [4]. Some brief comments about the adaptation of the graphical method described in [4] to our process are given in Section 2. Consequently, the result obtained concerning the asymptotic behaviour of the process of Section 2 also concerns the asymptotic behaviour of Arjas's plots arising from a correctly specified model. Some goodness of fit tests based on this process are proposed in Section 3. All proofs are given in the Appendix.

2. Definitions and model assumptions. Let $(\Omega, \mathscr{F}, \mathscr{P})$ be a complete probability space and let $\{\mathscr{F}_t\colon t \in [0, \tau]\}$ be an increasing right-continuous family of sub-σ-algebras of $\mathscr{F}$. Local martingales, predictable processes, etc. are defined with reference to these sub-σ-algebras. Following Andersen and Gill [3], let $N = (N_1, \ldots, N_n)$, $n \geqslant 1$, be the multivariate counting process defined so that N_i counts failures on the i-th subject at times $t \in [0, \tau]$. Thus N has components N_i which are right-continuous step functions, zero at time zero, with jumps of size $+1$ only, such that no two components jump simultaneously. Assume each $N_i(\tau)$ to be almost surely finite. Following [3] we suppose that N_i for $i = 1, \ldots, n$ has random intensity process λ_i, i.e.

$$M_i(t) = N_i(t) - \int_0^t \lambda_i(s)\,ds, \qquad t \in [0, \tau], \tag{1}$$

is a local square integrable martingale, of the Cox's regression form

$$\lambda_i(t) = Y_i(t)\exp[\beta_0^{\mathrm{T}} Z_i(t)]\,\lambda_0(t), \qquad t \in [0, \tau]. \tag{2}$$

Here β_0 is a p-vector of unknown regression coefficients, λ_0 is an arbitrary unspecified baseline hazard function, Y_i is a predictable $\{0, 1\}$-valued process indicating that the i-th individual is at risk when $Y_i = 1$, and Z_i is a p-variate column vector process which is assumed to be predictable and locally bounded.

The *maximum partial likelihood estimator* $\hat{\beta}$ of β_0 is defined as the solution of the equation

$$U(\beta, \tau) = 0, \tag{3}$$

where

$$U(\beta, t) = n^{-1/2} \sum_{i=1}^{n} \int_0^t \left\{ Z_i(s) - \frac{\sum_{i=1}^{n} Y_i(s) Z_i(s) \exp[\beta^{\mathrm{T}} Z_i(s)]}{\sum_{i=1}^{n} Y_i(s)\exp[\beta^{\mathrm{T}} Z_i(s)]} \right\} dN_i(s), \qquad t \in [0, \tau]. \tag{4}$$

We use the estimator $\hat{\beta}$ in the inference for goodness of fit in Cox's proportional hazards model (2).

Given $I \subset \{1, \ldots, n\}$ and $z \in \boldsymbol{R}^p$, let us define the process

$$(5) \qquad H_I(\hat{\beta}, z, t) = \sum_{i \in I} \sum_{j \in I} \int_0^t \Bigg[I\{Z_i(s) \in J_z\} dN_j(s) - \frac{I\{Z_i(s) \in J_z\} Y_j(s) \exp[\hat{\beta}^{\mathrm{T}} Z_j(s)]}{\sum_{k=1}^{n} Y_k(s) \exp[\hat{\beta}^{\mathrm{T}} Z_k(s)]} d\bar{N}(s) \Bigg], \qquad t \in [0, \tau],$$

where $J_z \subset \boldsymbol{R}^p$ is a finite interval containing z, and $\bar{N} = \sum_{k=1}^{n} N_k$.

The process $H_I(\hat{\beta}, z, \cdot)$ makes a direct comparison between observed and expected failure frequencies, as estimated from the model for individuals in stratum I with covariates in J_z. Since, under the correct model, $H_I(\beta_0, z, \cdot)$ is a local martingale, $H_I(\hat{\beta}, z, \cdot)$ expresses a balance between the "suitable restricted" actual count of failures in stratum I and a corresponding estimated collective cumulative hazard. The above argument is used in our statistical inference in Section 3.

It should be noted that the graphical method of Arjas [4] can also be applied to the process of the form (5). To observe this take a stratum I and a finite interval J_z. Then the x-coordinates should count these individuals from stratum I for which the corresponding covariates evaluated at the observed failure times of I belong to J_z. The y-coordinates should mark the values of the corresponding estimated collective cumulative hazard. Obviously, under the assumed Cox's model one can expect the approximately linear graphs with slopes close to one (see [4] for a detailed discussion of a graphical plotting technique).

Note that the process $H_I(\beta_0, z, \cdot)$ is ad hoc robust to covariate outliers in stratum I due to its direct dependence on the bounded interval J_z which censors "large" covariate observations (see, e.g., [14]). Consequently, the same remark concerns the above-mentioned graphs and tests of Section 3.

We study the asymptotic properties of the process of the form (5) when the measure of J_z tends to zero as the size of I and n tend to infinity. Let $\{1, \ldots, n\}$ be stratified into q strata $I_1, \ldots, I_q$. Moreover, let $z_1, \ldots, z_r$ denote a finite set of distinct covariate levels. We use the following notation:

$$(6) \qquad S_k^{(j)}(\beta, t) = n_k^{-1} \sum_{l \in I_k} Z_l(t)^j Y_l(t) \exp[\beta^{\mathrm{T}} Z_l(t)], \quad j = 0, 1, 2,$$

$$Q_k(z, t) = [n_k w]^{-1} \sum_{l \in I_k} I\{Z_l(t) \in J_z\},$$

$$\bar{N}_k = \sum_{l \in I_k} N_l, \quad \bar{M}_k = \sum_{l \in I_k} M_l, \quad k \in Q,$$

where $Q = \{1, \ldots, q\}$, and also write $R = \{1, \ldots, r\}$. Here n_k denotes the size of I_k, w a measure of J_z, and Z^0, Z^1, Z^2 mean 1, Z, ZZ^{T}, respectively. For

$z = z_i$ we use the notation w_i. If $I_k = \{1, \ldots, n\}$, then $S_k^{(j)}$, $\bar{N}_k$ and $\bar{M}_k$ are denoted by $S^{(j)}$, $\bar{N}$ and $\bar{M}$, respectively. We also write H_k in place of H_I for $I = I_k$ in (5). Moreover, $(b_{ki}: Q \times R)$ means a qr-variate vector $(b_{11}, \ldots, b_{q1}, \ldots, b_{1r}, \ldots, b_{qr})$.

We let the assumptions (A)–(D) of [3] which guarantee the asymptotic normality of Cox's estimator $\hat{\beta}$ be satisfied throughout the paper. They include the assumption of the sums $S^{(0)}(\beta, t)$, $S^{(1)}(\beta, t)$ and $S^{(2)}(\beta, t)$ converging uniformly for $t \in [0, \tau]$ and β in a neighbourhood of β_0 to functions $s^{(0)}(\beta, t)$, $s^{(1)}(\beta, t)$ and $s^{(2)}(\beta, t)$, respectively, in probability.

Under (6) we make for $k \in Q$ and $i \in R$ the following

ASSUMPTIONS.

E: $w_i \to 0$, $n_k/n \to p_k$, $p_k \in (0, 1)$, $a_{ki}/n^{1/2} \to \infty$ as $n \to \infty$, where $a_{ki} = n_k w_i n^{1/2}$.

F: $S_k^{(0)}(\beta, t)$ and $S_k^{(1)}(\beta, t)$ converge uniformly for $t \in [0, \tau]$ and β in some neighbourhood of β_0 to functions $s_k^{(0)}(\beta, t)$ and $s_k^{(1)}(\beta, t)$, respectively, in probability. The functions $s_k^{(0)}$ and $s_k^{(1)}$ satisfy the same regularity conditions as those for $s^{(0)}$ and $s^{(1)}$ which are given in (D) of [3].

G: $Q_k(z_i, t)$ converges in probability to a nonzero bounded function $q_k(z_i, t)$, uniformly for $t \in [0, \tau]$.

The conditions (F) and (G) include the asymptotic stability and regularity assumptions.

3. Asymptotic properties and goodness of fit tests. Under the assumption of Section 2 we can state the following

THEOREM 3.1. *$H_n(\cdot) = (a_{ki}^{-1} H_k(\hat{\beta}, z_i, \cdot): Q \times R)$ converges weakly as $n \to \infty$ to a qr-variate continuous Gaussian process $\Gamma(\cdot) = (\Gamma_k(z_i, \cdot): Q \times R)$ with zero mean and covariance function defined by*

$$(7) \quad \operatorname{Cov}[\Gamma_k(z_i, t), \Gamma_l(z_j, u)] = \int_0^t q_k(z_i, s) q_l(z_j, s) [I\{k = l\} - p_l s_l^{(0)}(\beta_0, s)/s^{(0)}(\beta_0, s)] p_k s_k^{(0)}(\beta_0, s) \lambda_0(s) ds - v_k(z_i, t)^{\mathrm{T}} \Sigma^{-1} v_l(z_j, u),$$

where $u, t \in [0, \tau]$, $t \leqslant u$, $k, l \in Q$, $i, j \in R$. Here

$$(8) \quad v_k(z, t) = p_k \int_0^t q_k(z, s)[s_k^{(1)}(\beta_0, s) - s_k^{(0)}(\beta_0, s) s^{(1)}(\beta_0, s)/s^{(0)}(\beta_0, s)] \lambda_0(s) ds,$$

$$\Sigma = \int_0^\tau [s^{(2)}(\beta_0, s) - s^{(1)}(\beta_0, s) s^{(1)}(\beta_0, s)^{\mathrm{T}}/s^{(0)}(\beta_0, s)] \lambda_0(s) ds.$$

Note that in the case where (N_i, Y_i, Z_i) are i.i.d. replicates of (N, Y, Z), say, the covariance function given in (7) takes a simpler form. Then we have $s_k^{(m)} = s^{(m)}$, $m = 0, 1$, $k \in Q$, and conclude that the process $a_{ki}^{-1} H_k(\hat{\beta}, z_i, \cdot)$ converges weakly to the Gaussian limit process

$$W\Big(p_k(1-p_k)\int_0^{(\cdot)} q_k^2(z_i, s)s^{(0)}(\beta_0, s)\lambda_0(s)ds\Big), \qquad k\in Q,\ i\in R,$$

where W is a standard Brownian motion. This means that the asymptotic randomness of a single Arjas's graph relates to the time transformed Brownian motion.

Theorem 3.1 gives the limiting distribution of the process H_n under model conditions. The model should be rejected if H_n is significantly different from zero, as measured by some "suitably chosen" functional. We present some formal statistical applications in the sequel.

Let

$$(9)\qquad \Sigma_n = n^{-1}\int_0^{\tau}[S^{(2)}(\hat\beta, s)/S^{(0)}(\hat\beta, s)-S^{(1)}(\hat\beta, s)S^{(1)}(\hat\beta, s)^{\mathrm{T}}/S^{(0)}(\hat\beta, s)^2]d\bar N(s),$$

$$T_{ki} = n^{-1}\int_0^{\tau} Q_k^2(z_i, s)\frac{n_k S_k^{(0)}(\hat\beta, s)}{nS^{(0)}(\hat\beta, s)}\left[1-\frac{n_k S_k^{(0)}(\hat\beta, s)}{nS^{(0)}(\hat\beta, s)}\right]d\bar N(s),$$

$$V_k(\beta, z, t) = [n_k/n]\int_0^{t} Q_k(z, s)\left[\frac{S_k^{(1)}(\beta, s)}{S^{(0)}(\beta, s)}-\frac{S_k^{(0)}(\beta, s)S^{(1)}(\beta, s)}{S^{(0)}(\beta, s)^2}\right]n^{-1/2}d\bar N(s),$$

$$U_{kilj}(t, u, x, y) = n^{-1}[V_k(\hat\beta, z_i, u)-V_k(\hat\beta, z_i, t)]^{\mathrm{T}}\Sigma_n^{-}[V_l(\hat\beta, z_j, y)-V_l(\hat\beta, z_j, x)],$$

$$W_{kilj}(t, u) = n^{-1}\int_t^{u} Q_k(z_i, s)Q_l(z_j, s)\left[I\{k=l\}-\frac{n_l S_l^{(0)}(\hat\beta, s)}{nS^{(0)}(\hat\beta, s)}\right]d\bar N_k(s) - U_{kilj}(t, u, t, u),$$

where $k, l\in Q$, $i, j\in R$, $t, u, x, y\in[0, \tau]$. Here Σ_n^{-} means, if necessary, a generalized inverse of Σ_n.

Under the above notation, for $k\in Q$, $i\in R$ we have

COROLLARY 3.2. *If* $s_k^{(m)} = s^{(m)}$, $m = 0, 1$, *then*

$$\max\{|a_{ki}^{-1}H_k(\hat\beta, z_i, t)|:\ t\in[0, \tau]\} * T_{ki}^{-1/2}$$

is asymptotically distributed as $\max\{|W(t)|:\ t\in[0, 1]\}$, *where* W *is the standard Brownian motion.*

Note that the test statistic of Corollary 3.2 is based on a suitably normed maximal distance between the counted failures and the corresponding estimated collective cumulative hazard for individuals in stratum I_k with covariates in J_{z_i}. Consequently, it has an intuitive appeal to the graphical method since it measures the maximal value of the observed "trend" of the graph (see [4] and [11]). A derivation of the distribution of the limit variable can be found in [5], and a table in [17].

Now we present some χ^2-type statistics based on the process H_n. Let $0 = t_0 < t_1 < \ldots < t_m = \tau$ be a partition of $[0, \tau]$ into m intervals, let h_{ki} and g_r, $k \in Q$, $i \in R$, $r \in \{1, \ldots, m\}$ be given numbers with $h_{ki} \neq 0$ and $g_r \neq 0$ for some k, i, r.

Let

$$\mathscr{X}_{ki} = \sum_r g_r a_{ki}^{-1} [H_k(\hat{\beta}, z_i, t_r) - H_k(\hat{\beta}, z_i, t_{r-1})],$$

$$\mathscr{X} = \sum_{k,i} h_{ki} a_{ki}^{-1} H_k(\hat{\beta}, z_i, \tau).$$

Under (9) and the above notation we have

COROLLARY 3.3. *The random variables*

$$\mathscr{X}_{ki}^2 \{\sum_r g_r^2 W_{kiki}(t_{r-1}, t_r) + 2 \sum_{s<r} g_s g_r U_{kiki}(t_r, t_{r-1}, t_{s-1}, t_s)\}^{-1}, \qquad k \in Q,\ i \in R,$$

and

$$\mathscr{X}^2 \{ \sum_{k,i,l,j} h_{ki} h_{lj} W_{kilj}(0, \tau)\}^{-1}$$

are asymptotically χ^2 distributed with one degree of freedom.

Note that, in view of Theorem 3.1, if $s_k^{(m)}$ or $s_l^{(m)}$ equals $s^{(m)}$, $m = 0, 1$, then U_{kilj} can be taken to be zero. This implies that in the case where (N_i, Y_i, Z_i) are observations of independent replicates of (N, Y, Z) the test statistics of the above corollary can be of a simpler form.

From a practical point of view the time axis should be divided into m intervals $[t_{r-1}, t_r]$ which will contain approximately the same number of observations $\sum \int I\{Z_j(s) \in J_{z_i}\} dN_l(s)$, where the sum and the integral are over $j, l \in I_k$ and $s \in [t_{r-1}, t_r]$, respectively. Some further practical way to determine weights g_r of the test statistic based on $\mathscr{X}_{ki}$ can be based on the following considerations. Obviously, by plotting the t-time coordinates against $H_k(\hat{\beta}, z_i, t)$ one can expect, under the assumed model, to obtain the trajectory oscillating round about zero. Otherwise, there will typically be groups of intervals $[t_{r-1}, t_r]$ for which the corresponding expected estimated cumulative hazards are systematically too high or too low to match to the data. Consequently, by considering

$$\Delta_r = a_{ki}^{-1} [H_k(\hat{\beta}, z_i, t_r) - H_k(\hat{\beta}, z_i, t_{r-1})], \qquad r \in \{1, \ldots, m\},$$

we can choose only the time intervals for which the corresponding Δ_r shows the largest deviations from zero. Then we can determine the corresponding weights g_r to obtain $g_r \Delta_r$ of the same positive sign. In view of the standardized factor of the test statistic based on $\mathscr{X}_{ki}$ the choice of weights g_r restricted to the set $\{-1, 0, +1\}$ seems to be practically useful. A similar practical procedure can also be used to determine weights h_{ki} of the test statistic based on $\mathscr{X}$.

Let $\hat{B}_n$ and B denote $(qr \times qr)$-matrices with elements of the form $W_{kilj}(0, \tau)$ given by (9) and $\mathrm{Cov}[\Gamma_k(z_i, \tau), \Gamma_l(z_j, \tau)]$ given by (7), respectively.

COROLLARY 3.4. *$H_n(\tau)^{\mathrm{T}} \hat{B}_n^- H_n(\tau)$ is asymptotically χ^2 distributed with* $\mathrm{rank}(B)$ *degrees of freedom.*

In a general situation the exact value of $\mathrm{rank}(B)$ seems extremely hard to obtain. However, when (N_i, Y_i, Z_i) are i.i.d. replicates of (N, Y, Z), where Z is a time independent random vector, the problem simplifies. Then $\mathrm{rank}(B)$ equals $q-1$. To observe this note that, in view of condition G of the Assumptions, the function $q_k(z, t)$ is now equal to $q(z)$, say, $k \in Q$. Consequently, by (7) we infer that $\mathrm{Cov}[\Gamma_k(z_i, \tau), \Gamma_l(z_j, \tau)]$ is of the form

$$q(z_i)q(z_j)p_k(1-p_k)\int s^{(0)}(\beta_0, s)\lambda_0(s)ds \qquad \text{for } k = l,$$

and

$$-q(z_i)q(z_j)p_k p_l \int s^{(0)}(\beta_0, s)\lambda_0(s)ds \qquad \text{for } k \neq l,$$

where the integral is over $[0, \tau]$, $k, l \in Q$, $i, j \in R$. It can be shown that $\mathrm{rank}(B)$ equals the rank of the matrix with the elements $p_{kl} = p_k(I\{k = l\} - p_l)$, $k, l \in Q$. Thus the assertion follows.

Now we present some verbal discussion concerning the power of the tests of Corollaries 3.2–3.4. First observe that in view of (5) all these tests are based on the quantities of the following form:

$$\int \sum_I I\{Z_j \in J_z\} \sum_{I \neq I_k} [\delta(I_k)dN_I - \delta(I)dN_{I_k}],$$

where

$$\delta(I_k) = [\sum_{I_k} Y_j \exp(\hat{\beta}^{\mathrm{T}} Z_j)] * [\sum_1^n Y_j \exp(\hat{\beta}^{\mathrm{T}} Z_j)]^{-1},$$

N_{I_k} corresponds to the notation of $\bar{N}_k$, and the integral is over $(s, t]$, say, with $0 \leqslant s < t \leqslant \tau$. This means that for these departures from the assumed model of proportional hazards where there are some differences between the strata $\{I\}$, as far as the distribution of (N_i, Y_i, Z_i) is concerned, the power of the considered tests should be relatively high. Then the covariates with distributions concentrated round about the levels $\{z\}$ should also affect the power. Otherwise, the gain in power is expected to be slight and only the appropriate choice of $\{z\}$ can influence the higher power values. This, however, concerns the problem of choice of a stratum generated by the levels of covariate space.

Some practical remark about this looks like the following (see also [20]). The partition of the range of a covariate Z based on the finite number of the covariate levels $\{z\}$ can be constructed as follows. For each coordinate Z_k of Z which is of a continuous type, one can choose the same finite number of the one-dimensional levels $\{z_k\}$ and the corresponding intervals $\{J_k\}$, with $z_k \in J_k$, so that they contain approximately the same large number of coordinate observations $(Z_i)_k$, $i = 1, \ldots, n$, evaluated at the observed failure times. On the other hand, each discrete coordinate of Z, e.g. that of a qualitative type as an indicator of sex, treatment group, etc. (see [10]), can determine the corresponding level coordinate. In particular, if the range of the covariate Z is discrete, one may want to use each value of Z as a partition. This leads to the choice of $\{z\}$ and $\{J\}$. It should be mentioned that the ways of choice of the stratification based on the strata $\{I\}$ are discussed in [4] (see also [11]).

APPENDIX

Proof of Theorem 3.1. Fix $k \in Q$ and $i \in R$. By using the mean value Lagrange theorem, in view of (1), (5) and (6), we obtain

$$H_k(\hat{\beta}, z_i, t) = n_k w_i \{\int_0^t Q_k(z_i, s)\{d\bar{M}_k(s) - n_k S_k^{(0)}(\beta_0, s)/[nS^{(0)}(\beta_0, s)]\, d\bar{M}(s)\}$$
$$- V_k(\bar{\beta}, z_i, t)^{\mathrm{T}} n^{1/2}(\hat{\beta} - \beta_0)\},$$

where V_k is given by (9), and $\bar{\beta}$ is on the line segment between β_0 and $\hat{\beta}$. First observe that $n^{-1/2} V_k(\bar{\beta}, z_i, t)$ converges in probability to $v_k(z_i, t)$, given by (8), uniformly for $t \in [0, \tau]$. This follows by the asymptotic stability and regularity conditions and the fact that, by Lenglart's inequality (see [19]), $n^{-1}\bar{N}(\tau)$ is bounded in probability whereas

$$\int_0^t q_k(z_i, s)\left[\frac{s_k^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s)} - \frac{s_k^{(0)}(\beta_0, s)s^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s)^2}\right] n^{-1}\, d\bar{M}(s)$$

converges in probability to zero, uniformly for $t \in [0, \tau]$. Moreover, in view of [3], $n^{1/2}(\hat{\beta} - \beta_0) - \Sigma^{-1} U(\beta_0, \tau)$ converges in probability to zero, where U is given by (4). Consequently, the process

$$H_n(\cdot) = (a_{ki}^{-1} H_k(\hat{\beta}, z_i, \cdot)\colon\ Q \times R)$$

is asymptotically equivalent to the process $\bar{H}_n(\cdot)$ with the components

$$n^{-1/2} \int_0^{(\cdot)} Q_k(z_i, s)\{d\bar{M}_k(s) - n_k S_k^{(0)}(\beta_0, s)/[nS^{(0)}(\beta_0, s)]\, d\bar{M}(s)\}$$
$$- v_k(z_i, \cdot)\Sigma^{-1} U(\beta_0, \tau), \qquad k \in Q,\ i \in R.$$

Consider a $(qr+p)$-variate local square integrable martingale $G_n(\cdot)$ with the components

$$n^{-1/2} \int_0^{(\cdot)} Q_k(z_i, s)\{d\bar{M}_k(s) - n_k S_k^{(0)}(\beta_0, s)/[nS^{(0)}(\beta_0, s)]\, d\bar{M}(s)\}, \qquad k \in Q,\ i \in R,$$

and the last one is the vector equal to $U(\beta_0, \cdot)$ given by (4). It can be shown that G_n satisfies the assumptions of Theorem I.2 of [3], and consequently this process converges weakly to a $(qr+p)$-variate Gaussian martingale with continuous sample paths and zero mean. Let $\bar{G}_n(\cdot)$ be the process obtained from $G_n(\cdot)$ so that only the last component of $G_n(\cdot)$, i.e. $U(\beta_0, \cdot)$ is replaced here by $U(\beta_0, \tau)$. The process $\bar{H}_n$ itself may be written as $f(\bar{G}_n)$, where f is the function from $D^{qr}[0, \tau]\times \boldsymbol{R}^p$ to $D^{qr}[0, \tau]$ defined by

$$f(x, u) = \left(x_{ki} - v_k(z_i)\Sigma^{-1}u\colon\ Q\times R\right)$$

for $x\in D^{qr}[0, \tau]$, $u\in \boldsymbol{R}^p$, and $v_k(z_i)$, Σ given by (8). Obviously, f is continuous on $C^{qr}[0, \tau]\times \boldsymbol{R}^p$ relative to the product Skorokhod topology. Consequently, in view of Theorem 5.1 of [5], the process $\bar{H}_n$ converges weakly to a continuous qr-variate Gaussian process Γ, say. Obviously, Γ has zero mean. A direct calculation brings a formula for the covariance function of Γ of the form (7), which completes the proof.

P r o o f o f C o r o l l a r y 3.2. It follows from Theorem 3.1 that the process $a_{ki}^{-1}H_k(\hat{\beta}, z_i, \cdot)$ converges weakly to $W(A(\cdot))$, where W is the standard Brownian motion and

$$A(t) = p_k(1-p_k)\int_0^t q_k^2(z_i, s)s^{(0)}(\beta_0, s)\lambda_0(s)ds, \qquad t\in[0, \tau].$$

In view of the consistency of $\hat{\beta}$ and the Assumptions, the integrand of T_{ki} given in (9) converges in probability to $p_k(1-p_k)q_k^2(z_i, s)$, uniformly in $s\in[0, \tau]$. By applying Lenglart's inequality one can infer that T_{ki} is the consistent estimate of $A(\tau)$. Consequently, the scale-change property of the Brownian motion completes the proof.

P r o o f o f C o r o l l a r y 3.3. By the same arguments as those we used to deal with T_{ki} in the preceding proof we can show that $U_{kiki}(u, t, x, y)$ given by (9) is the consistent estimate of $\mathrm{Cov}[\Gamma_k(z_i, u)-\Gamma_k(z_i, t), \Gamma_k(z_i, y)-\Gamma_k(z_i, x)]$ obtained from (7), where $x<y\leqslant t<u$. Similarly, $W_{kiki}(t, u)$ given in (9) converges in probability to the above-mentioned covariance, but with $x = t$ and $y = u$. Quite analogously we conclude that $W_{kilj}(0, \tau)$ is the consistent estimate of $\mathrm{Cov}[\Gamma_k(z_i, \tau), \Gamma_l(z_j, \tau)]$. Thus Theorem 3.1 completes the proof.

P r o o f o f C o r o l l a r y 3.4. First note that $\hat{B}_n$ is the consistent estimate of B. Consequently, by Theorem 3.1, $H_n(\tau)^{\mathrm{T}}\hat{B}_n^{-}H_n(\tau)$ converges weakly to $\Gamma(\tau)^{\mathrm{T}}B^{-}\Gamma(\tau)$. This completes the proof.

REFERENCES

[1] P. K. A n d e r s e n, *Testing goodness of fit of Cox's regression and life model*, Biometrics 38 (1982), pp. 67–77.

[2] – and O. B o r g a n, *Counting process models for life history data*, Scand. J. Statist. 12 (1985), pp. 97–158.

[3] P. K. Andersen and R. D. Gill, *Cox's regression model for counting process: a large sample study*, Ann. Statist. 10 (1982), pp. 1100–1120.
[4] E. Arjas, *A graphical method for assessing goodness of fit in Cox's proportional hazards model*, J. Amer. Statist. Assoc. 83 (1988), pp. 204–212.
[5] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York 1968.
[6] D. R. Cox, *Regression models and life tables (with discussion)*, J. Roy. Statist. Soc. Ser. B 34 (1972), pp. 187–220.
[7] J. Crowley and B. E. Storer, *Comment on 'A reanalysis of the Stanford Heart Transplant Data', by M. Aitkin, N. Laird, and B. Francis*, J. Amer. Statist. Assoc. 78 (1983), pp. 277–281.
[8] R. D. Gill and M. Schumacher, *A simple test for the proportional hazards assumption*, Biometrika 74 (1987), pp. 289–300.
[9] R. Gray, *Some diagnostic methods for Cox regression models through hazard smoothing*, Biometrics 46 (1990), pp. 93–102.
[10] S. W. Lagakos and D. Schoenfeld, *Properties of proportional hazards score tests under misspecified regression models*, ibidem 40 (1984), pp. 1037–1048.
[11] L. Marzec and P. Marzec, *Goodness of fit inference based on stratification in Cox's regression model*, Scand. J. Statist. 20 (1993), in print.
[12] I. W. McKeague and K. J. Utikal, *Goodness-of-fit tests for additive hazards and proportional hazards models*, ibidem 18 (1991), pp. 177–195.
[13] T. Moreau, J. O'Quigley and M. Mesbah, *A global goodness of fit statistic for the proportional hazards model*, Appl. Statist. 34 (1985), pp. 212–218.
[14] S. Samuels, *Robustness of survival estimators*, Ph. D. Thesis, Department of Biostatistics, University of Washington, 1978.
[15] D. Schoenfeld, *Partial residuals for the proportional hazards regression model*, Biometrika 69 (1982), pp. 239–241.
[16] T. M. Therneau, P. M. Grambsch and T. R. Fleming, *Martingale-based residuals for survival models*, ibidem 77 (1990), pp. 147–160.
[17] J. E. Walsh, *Handbook of Nonparametric Statistic*, Van Nostrand, Princeton 1962.
[18] L. J. Wei, *Testing goodness of fit for proportional hazards model with censored observations*, J. Amer. Statist. Assoc. 79 (1984), pp. 649–652.
[19] E. Lenglart, *Relation de domination entre deux processus*, Ann. Inst. H. Poincaré 13 (1977), pp. 171–179.
[20] D. Schoenfeld, *Chi-squared goodness-of-fit tests for proportional hazards regression model*, Biometrika 67 (1980), pp. 145–153.

Mathematical Institute
University of Wrocław
pl. Grunwaldzki 2/4
50-384 Wrocław, Poland

Received on 25.3.1992;
revised version on 19.4.1993