JOURNAL OF MULTIVARIATE ANALYSIS 33, 125-142 (1990)

Asymptotic Inference for Multiplicative Counting Processes Based on One Realization

ÅKE SVENSSON*

Institute of Actuarial Mathematics and Mathematical Statistics, University of Stockholm, Sweden

Communicated by the Editors

It is assumed that we observe one realization of an r dimensional counting process with intensities that are products of predictable and observable weight processes, a common function of time, and predictable functions that depend on an unknown parameter θ. Given that the realization brings increasing information on θ as the observed time grows asymptotic results are proved for the distributions of parameter estimates, certain test statistics for parametric hypothesis, and goodness-of-fit tests.

1. INTRODUCTION

Let $N(t)=(N_1(t), ..., N_r(t))$ be a r-dimensional ($r \geqslant 2$) counting process defined on the probability space $(\Omega, \mathscr{A}, P)$ and adapted to the filtration $\mathscr{A}_t$, $t \in [0, \infty[$. This implies that the components, N_i, are increasing integer valued and right continuous functions with jumps of size 1 only, starting with $N_i(0)=0$. Furthermore, two component processes cannot jump at the same time.

We will consider models where the counting process is assumed to have a predictable intensity $\lambda(\theta, t)=(\lambda_1(\theta, t), ..., \lambda_r(\theta, t))$ of the form

$$\lambda_i(\theta, t)=e^{h_i(\theta, t)} Y_i(t)\, \alpha(t) \tag{1.1}$$

for $i=1, ..., r$. Here Y_i is an observable and predictable (weight) process, α

Received September 19, 1988; revised October 3, 1989.

AMS 1980 subject classifications: primary 62M09.

Key words and phrases: counting processes, Cox regression model, goodness-of-fit tests, martingale limit theorems.

* This work was supported by the Swedish Natural Science Council, NFR.

0047-259X/90 $3.00

is an unknown nonnegative (nuisance) parameter function, $\theta = (\theta_1, ..., \theta_m)$ is a m-dimensional parameter taking an unknown value in some open subset Γ of R^m, and $h_i(\theta, \cdot)$ are for any $\theta \in \Gamma$ predictable processes. All measurability assumptions are to be understood relative to the filtration $\mathscr{A}_t$. It is in the following assumed that $\int_0^u \lambda_i(\theta, t)\, dt < \infty$ for all $i = 1, ..., r$ and $u < \infty$.

There is some arbitrariness in the formulation of the model. The model given by (1.1) can also be obtained with $\tilde{h}_i(\theta, t) = h_i(\theta, t) - a_i(t) - b(\theta, t)$, $\tilde{Y}_i(t) = Y_i(t)\, e^{a_i(t)}$, and $\tilde{\alpha}(t) = \alpha(t)\, e^{b(\theta, t)}$. This reformulation will not influence the estimates of θ. We will in the following use the normalization $h_r(\theta, t) = 0$. This can always can be obtained by choosing $b(\theta, t) = h_r(\theta, t)$.

We will study inference about the parameter θ using *one* realization of the process during a finite time interval $[0, u]$. We will derive asymptotic results as the observation time $u \to \infty$. The special case $h_i(\theta, t) = \theta_i$ has been studied in [11]. The present generalization is made in order to motivate tests of parametrical hypothesis and goodness-of-fit.

A model given by (1.1) is of semi-parametrical type. It has the same structure as a Cox regression model commonly used in survival analysis. There is a large literature on inference in such models in the asymptotic case where an increasing number of independent copies of the process are observed during a finite time interval (cf. [2]). The observation plan that is studied here, where the observation time increases, is different and has to be treated by other methods. We will, however, use the partial likelihood, introduced in [4], as a basis for the inference.

2. Assumptions and Notations

We will assume that we have observed the counting process N and the weight process Y during the finite time interval $[0, u]$. Our concern is to estimate the structural parameter θ. The presence of the nuisance parameter function α complicates matters since it describes an underlying and unknown jump intensity that is common for the τ component process. The information about the parameter θ will have to be derived from observations not on when a jump occurs but of which of the r processes jumps when a jump occurs. We will base the inference on the partial likelihood:

$$L(\theta, u) = \prod_i \prod_{t \leqslant u} (e^{h_i(\theta, t)} Y_i(t))^{dN_i(t)} \Big/ \prod_{t \leqslant u} \left(\sum_{i=1}^{r} e^{h_i(\theta, t)} Y_i(t) \right)^{\sum_{i=1}^{r} N_i(t)}. \quad (2.1)$$

We will use the normalization $h_r(\theta, t) = 0$.

As in ordinary likelihood estimation we will maximize (2.1) by differentiating the logarithm of the partial likelihood. The estimate, $\hat{\theta}(u)$ will then solve the equation

$$M(\gamma, u) = \int_0^u dG(\gamma, t)\, H(\gamma, t) = 0, \tag{2.2}$$

where $(r' = r - 1)$

$$H(\gamma, t) = [\partial h_i(\gamma, t)/\partial \gamma_j]_{i=1,r'}^{j=1,m} = [h_{ij}(\gamma, t)]_{i=1,r'}^{j=1,m},$$

$$p_i(\gamma, t) = e^{h_i(\gamma, t)} Y_i(t) \Big/ \sum_{j=1}^{r} e^{h_j(\gamma, t)} Y_j(t),$$

$$dG_i(\gamma, t) = dN_i(t) - p_i(\gamma, t)\, d\bar{N}(t),$$

and

$$dG(\gamma, t) = (dG_1(\gamma, t), ..., dG_{r'}(\gamma, t)).$$

(Here $\bar{N}(t) = \sum_{i=1}^{r} N_i(t)$.)

Observe that $G(\theta, \cdot)$ is a martingale relative the filtration $\mathscr{A}_t$, $t \in [0, \infty[$. We will assume that

(A1) $h(\theta, t) = (h_1(\theta, t), ..., h_{r'}(\theta, t))$ is differentiable in θ and $\int_0^u |h_{ij}(\theta, t)|^3\, \bar{\lambda}(\theta, t)\, dt < \infty$ for all $i = 1, ..., r'$, $j = 1, ..., m$, and $u < \infty$, with $\bar{\lambda}(\theta, t) = \sum_{i=1}^{r} \lambda_i(\theta, t)$.

If this assumption is satisfied then $M(\theta, \cdot)$ is a martingale relative the filtration $\mathscr{A}_t$, $t \in [0, \infty[$. This fact will be essential for the results that follow.

Before proceeding we will introduce some further notations:

$$\Pi(\gamma, t) = [\delta_i^j p_i(\gamma, t) - p_i(\gamma, t)\, p_j(\gamma, t)]_{i=1,r'}^{j=1,r'},$$

$$W(\gamma, u) = \int_0^u H'(\gamma, t)\, \Pi(\gamma, t)\, H(\gamma, t)\, \bar{\lambda}(\gamma, t)\, dt,$$

$$T(\gamma, u) = \int_0^u H'(\gamma, t)\, \Pi(\gamma, t)\, H(\gamma, t)\, d\bar{N}(t),$$

and

$$U(\gamma, u) = \left[\sum_{k=1}^{r} \int_0^u \left(h_{ki} - \sum_r h_{ri} p_r \right) \left(h_{kj} - \sum_r h_{rj} p_r \right) dN_k \right]_{i=1,m}^{j=1,m}.$$

Here $W(\theta, u)$ is the predictable quadratic variation $\langle M(\theta, \cdot) \rangle (u)$ of the martingale $M(\theta, u)$. The matrix T does not depend on the nuisance parameter α and will be shown to approximate W asymptotically. This is also the case for the matrix U. Observe that $U(\theta, u)$ equals the optional quadratic variation $[M(\theta, \cdot)](u)$. Here we have chosen to work with the matrix T instead of with U even if similar results can be proved for both matrices.

Since we rely on only one realization of the process to do inference about θ it is important that this realization carry sufficent information about the parameter in order that the estimate behave asymptotically regularly. This may very well be the case for some realizations whereas for other realizations the accumulation of information will be too small or irregular. It is of interest not only to prove the asymptotic consistency of an estimator but also of finding its asymptotic distribution and asymptotic properties of different kind of test statistics. We will be able to prove consistency, asymptotic mixed normal distribution for the maximum partial likelihood estimate, and asymptotic χ^2 distribution of different test statistics at least for some realizations. The more complicated the result the more is demanded of the structure of the information contained in the realization. We will define subsets of the sample space for whose realizations we can prove asymptotic results of different kinds.

We will denote the smallest eigenvalue of a matrix A with $\mathbf{A}$, and the largest eigenvalue with $\bar{A}$. Let $F_h^a \in \mathscr{A}$ be a subset of Ω such that:

(i) $\mathbf{W}(\theta, u) \to \infty$ and $\mathbf{W}(\theta, u)/\bar{W}(\theta, u)$ is bounded away from 0 as $u \to \infty$,

(ii) for all $\varepsilon > 0$ there exists an $u_0 < \infty$ such that

(a) $\int_0^u dG(\theta, t)(H(\gamma, t) - H(\theta, t))\, W^{-1}(\theta, u)| \leqslant \varepsilon\rho^a$ and

(b) $|\int_0^u (dG(\gamma, t) - dG(\theta, t))\, H(\gamma, t)\, W^{-1}(\theta, u) - (\gamma - \theta)| \leqslant \varepsilon\rho^a$ if $|\gamma - \theta| \leqslant \rho$ and $u \geqslant u_0$.

The specification of the sets F_h^a are given in a technical form that is suited to the proofs given below. It is, of course, possible to make more explicit assumptions in special cases. Observe that the larger the value of a, the stricter the assumption for the set F_h^a.

3. Consistency

The theory developed below will rely on strong results on convergence of martingales. In Section 7 we state a number of such results in a form convenient for the purposes of this paper.

THEOREM 3.1. *If* (A1) *holds then* $\hat{\theta}(u) \to \theta$ *a.s. on* F_h^0.

Proof. We will use the equality:

$$dG(\gamma, t)\, H(\gamma, t) = dG(\theta, t)\, H(\theta, t) + dG(\theta, t)(H(\gamma, t) - H(\theta, t)) + (dG(\gamma, t) - dG(\theta, t))\, H(\gamma, t). \tag{3.1}$$

According to Theorem 7.1 and condition (i),

$$M(\theta, u)\, W^{-1}(\theta, u) = \int_0^u dG(\theta, t)\, H(\theta, t)\, W^{-1}(\theta, t) \to 0$$

as $u \to \infty$ on F_h^0.

Looking trajectorywise we find that for realizations in F_h^0 we can for any $\varepsilon > 0$ find a u_1 such that

$$|M(\theta, u)\, W^{-1}(\theta, u)| \leqslant \varepsilon/4$$

if $u > u_1$. According to (ii), we can, when $|\gamma - \theta| \leqslant \varepsilon$, find a u_0 such that

$$\left| \int_0^u dG(\theta, t)(H(\gamma, t) - H(\theta, t))\, W^{-1}(\theta, u) \right| \leqslant \varepsilon/4$$

and

$$\left| \int_0^u (dG(\gamma, t) - dG(\theta, t))\, H(\gamma, t)\, W^{-1}(\theta, u) + (\gamma - \theta) \right| \leqslant \varepsilon/4.$$

If $u > \max(u_0, u_1)$ then

$$\int_0^u dG(\gamma, t)\, H(\gamma, t)\, W^{-1}(\theta, u)(\gamma - \theta) < 0$$

when $|\gamma - \theta| = \varepsilon$. This implies that there is a solution of $M(\gamma, u)\, W^{-1}(\theta, u) = 0$ and thus of (2.2) such that $|\hat{\theta}(u) - \theta| \leqslant \varepsilon$ (cf. [1]). This proves the theorem. ∎

4. Asymptotic Distribution

When deriving the asymptotic distribution of the estimate, $\hat{\theta}(u)$, we have to strengthen the conditions. First we will make a general assumption

(A2) There exists a non-random sequence $b(u)$ and an $\mathscr{A}$-measurable $m \times m$-matrix η such that $W(\theta, u)/b(u) \to \eta$ in probability as $u \to \infty$.

Theorem 4.1. *If* (A1)–(A2) *are satisfied then* $(\hat{\theta}(u) - \theta)\, W^{1/2}(\theta, u)$ *is asymptotically* $N(0, I_m)$ *distributed conditionally on* $F_h^1 \cap \{\eta > 0\}$.

Proof. Using the equality (3.1) with $\gamma = \hat{\theta}(u)$, we find that

$$0 = M(\theta, u)\, W^{-1/2}(\theta, u) + \int_0^u dG(\theta, t)(H(\hat{\theta}(u), t) - H(\theta, t))\, W^{-1/2}(\theta, u)$$
$$+ \left(\int_0^u (dG(\hat{\theta}(u), t) - dG(\theta, t))\, H(\hat{\theta}(u), t)\, W^{-1}(\theta, u) - (\hat{\theta}(u) - \theta) \right)$$
$$\times W^{1/2}(\theta, u) - (\hat{\theta}(u) - \theta)\, W^{1/2}(\theta, u).$$

Looking at the second term of the right-hand side of this equation we find that on F_h^1 for any $\varepsilon > 0$ and u sufficiently large,

$$\left| \int_0^u dG(\theta, t)(H(\hat{\theta}(u), t) - H(\theta, t))\, W^{-1/2}(\theta, u) \right|$$
$$\leqslant \left| \int_0^u dG(\theta, t)(H(\hat{\theta}(u), t) - H(\theta, t))\, W^{-1}(\theta, u) \right| \bar{W}^{1/2}(\theta, u)$$
$$\leqslant \varepsilon |(\hat{\theta}(u) - \theta)\, W^{1/2}(\theta, u)|\, (\bar{W}(\theta, u)/\mathbf{W}(\theta, u))^{1/2}.$$

This implies that the second term is asymptotically dominated by $(\hat{\theta}(u) - \theta)\, W^{1/2}(\theta, u)$ on F_h^1. By a similar argument this also holds for the third term of (4.1). Thus $M(\theta, u)\, W^{-1/2}(\theta, u)$ and $(\hat{\theta}(u) - \theta)\, W^{1/2}(\theta, u)$ are asymptotically equivalent on F_h^1. From Theorem 7.2 it follows that the first of these variables is asymptotically $N(0, I_m)$ distributed on $F_h^1 \cap \{\eta > 0\}$. ∎

The practical use of Theorem 4.1 is impeded by the fact that the normation $W(\theta, u)$ does depend on unknown parameters. We may substitute θ with $\hat{\theta}(u)$ but $W(\hat{\theta}(u), u)$ still depends on α. We will here prefer not to estimate the nuisance function α directly but instead use the approximation $T(\hat{\theta}(u), u)$. To prove that this is permissible we need another lemma and one further assumption.

LEMMA 4.2. *$T(\theta, u)\, W^{-1}(\theta, u) \to I_m$ (the $m \times m$ unit matrix) a.s. on $\{\eta > 0\}$ as $u \to \infty$.*

Proof. Let H_i be the ith column of the matrix H. Then the (ij)th element of the matrix $T(\theta, u) - W(\theta, u)$ equals

$$Q_{ij}(u) = \int_0^u H_i'(\theta, t)\, \Pi(\theta, t)\, H_j(\theta, t)(d\bar{N}(t) - \bar{\lambda}(t)\, dt).$$

This is a martingale with predictable quadratic variation:

$$\langle Q_{ij}(u)\rangle = \int_0^u (H_i(\theta,t)\,\Pi(\theta,t)\,H_j'(\theta,t))^2\,\tilde{\lambda}(\theta,t)\,dt$$

$$\leqslant \int_0^u (H_i(\theta,t)\,\Pi(\theta,t)\,H_i'(\theta,t)\,H_j(\theta,t)\,\Pi(\theta,t)\,H_j'(\theta,t))\,\tilde{\lambda}(\theta,t)\,dt$$

$$\leqslant \sqrt{W_{ii}(\theta,u)\,W_{jj}(\theta,u)}.$$

Thus on $\{\eta>0\}$, $\langle Q_{ij}\rangle(u)/b(u)$ is asymptotically bounded by $\sqrt{n_{ii}\eta_{jj}}$. If $\langle Q_{ij}\rangle(\infty)=\infty$ then, according to Theorem 7.1, $Q_{ij}(u)/\langle Q_{ij}\rangle(u)\to 0$ a.s. which implies that also $Q_{ij}(u)/b(u)\to 0$ a.s. as $u\to\infty$. If $\langle Q_{ij}\rangle(\infty)<\infty$ then $Q_{ij}(u)$ has a finite bound a.s. and $Q_{ij}(u)/b(u)\to 0$ a.s., since $b(u)\to\infty$. Thus $(T(\theta,u)-W(\theta,u))/b(u)\to 0$ a.s. on $\{\eta>0\}$ as $u\to\infty$. ∎

With the assumption

(A3) For any sequence $\varepsilon(u)\to 0$ as $u\to\infty$ $\sup_{|\gamma-\theta|\leqslant\varepsilon(u)}|T(\gamma,u)-T(\theta,u)|/b(u)\to 0$,

it follows, with the use of Lemma 4.2, that if $\gamma(u)\to\theta$ then $T(\gamma(u),u)T^{-1}(\theta,u)\to I_r$ a.s. on $\{\eta>0\}$ as $u\to\infty$. We thus have the following corollary:

COROLLARY 4.3. *If* (A1)–(A3) *are satisfied then* $(\hat{\theta}(u)-\theta)\,T^{1/2}(\hat{\theta}(u),u)$ *is asymptotically* $N(0,I_m)$ *distributed conditionally on* $F_h^1\cap\{\eta>0\}$.

5. TESTS OF PARAMETRICAL HYPOTHESES

We will be interested in testing if the model (1.1) can be simplified in such a way that θ can be assumed to vary in a space of lower dimension than m. We will specify this hypothesis as

$$\theta=\beta S, \tag{5.1}$$

where $\beta=(\beta_1,...,\beta_s)$ is a s-dimensional parameter and S is a $s\times m$ matrix. This hypothesis has an alternative formulation in the form of linear restrictions on θ. Let R be a $m\times(m-s)$ matrix whose rows are orthogonal to the columns of S. We can restate the hypothesis (5.1) as

$$\theta R=0. \tag{5.2}$$

All expressions will be greatly simplified if the matrix R is rotated so that $R'R=I_{m-s}$. We will in the following assume that this is the case. There is, of course, no loss of generality. If the hypothesis is true we can estimate β as a solution, $\hat{\beta}(u)$ of

$$M(\delta S, u)S' = \int_0^u dG(\delta S, t)\, H(\delta S, t)\, S' = 0. \tag{5.3}$$

From the previous sections it follows that the estimate $\hat{\beta}(u)$ is a consistent estimate conditionally on F_h^0 and that both $(\hat{\beta}(u)-\beta)(SW(\beta S, u)S')^{1/2}$ and $(\hat{\beta}(u)-\beta)(ST(\hat{\beta}(u)S, u)S')^{1/2}$ are asymptotically $N(0, I_s)$ distributed conditionally on F_h^1 if (A1)–(A3) are satisfied.

There are several general methods available to test hypotheses of type (5.1) or (5.2). We can consider the ratio test based on the partial likelihoods and reject the hypothesis if

$$Q_R(u) = 2\ln\,(L(\hat{\theta}(u), u)/L(\hat{\beta}(u)S, u))$$

is too large. The score test is based on the deviance of $M(\hat{\beta}(u)S, u)R$ from 0. A third test, of Wald type, may be based on the deviance of $\hat{\theta}(u)R$ from 0. In addition to the likelihood ratio test we thus define the two test statistics:

$$Q_S(u) = M(\hat{\beta}(u)S, u)\, RR'T^{-1}(\hat{\beta}(u)S, u)\, RR'M'\ (\hat{\beta}(u)S, u)$$

and

$$Q_W(u) = \hat{\theta}(u)\, RR'T(\hat{\beta}(u)S, u)\, RR'\hat{\theta}'(u).$$

THEOREM 5.1. *$Q_S(u)$ and $Q_W(u)$ are asymptotically equivalent and χ^2 distributed with $m-s$ d.f. conditionally on $F_h^1 \cap \{\eta > 0\}$.*

Proof. The parameter θ can be written in the form $\theta = \beta S + \varepsilon R'$, where $\beta \in \mathbf{R}^s$ and $\varepsilon \in \mathbf{R}^{m-s}$. Due to condition (ii) we can find a function $K(u) \to 0$ as $u \to \infty$ and

$$\begin{aligned}&|(M(\hat{\theta}(u), u) - M(\hat{\beta}(u)S, u))\, W^{-1}(\theta, u) + (\hat{\theta}(u) - \hat{\beta}(u)S)|\\ &\quad \leqslant K(u)\, |\hat{\theta}(u) - \hat{\beta}(u)S|.\end{aligned}$$

Since $W^{-1/2}(\theta, u)RR' = RR'W^{-1/2}(\theta, u)$, it follows that

$$\begin{aligned}&|(M(\hat{\theta}(u), u) - M(\hat{\beta}(u)S, u))\, RR'W^{-1/2}(\theta, u)\\ &\qquad + (\hat{\theta}(u) - \hat{\beta}(u)S)\, W^{1/2}(\theta, u)RR'|\\ &\quad = |-M(\hat{\beta}(u)S, u)\, RR'W^{-1/2}(\theta, u) + \hat{\theta}(u)\, W^{1/2}(\theta, u)\, RR'|\\ &\quad \leqslant K(u)\, |(\hat{\theta}(u) - \hat{\beta}(u)S)\, W^{1/2}(\theta, u)|\ (\bar{W}(\theta, u)/\mathbf{W}(\theta, u))^{1/2}\, |RR'|.\end{aligned}$$

Using Lemma 4.2 and assumptions (A2) and (A3) we see that $Q_S(u)$ and $Q_W(u)$ are asymptotically equivalent on F_h^1.

It remains to find the asymptotic distribution of the two test statistics. According to Corollary 4.3, $M(\hat{\beta}(u)S, u)\, RR'T^{-1/2}(\hat{\theta}(u), u)$ is asymptotically $N(0, RR')$ distributed conditionally on F_h^1. Thus (denoting a generalized inverse of the matrix A with A^{-})

$$M(\hat{\beta}(u)S, u)\, RR'T^{-1/2}(\hat{\theta}(u), u)(RR'RR')^{-}\, T^{-1/2}(\hat{\theta}(u), u)\, RR'M(\hat{\beta}(u)S, u) \tag{5.4}$$

is asymptotically χ^2 distributed with $\operatorname{rank}(RR') = m - s$ degrees of freedom. (Observe that the expression (5.4) is invariant of rotations of the matrix R). Since we have assumed that $R'R = I_{m-s}$ the $(m-s)$-dimensional unit matrix is a generalized inverse of RR'. This implies that $Q_S(u)$ is asymptotically χ^2 distributed with $m - s$ degrees of freedom conditionally on $F_h^1 \cap \{\eta > 0\}$. ∎

THEOREM 5.2. *$Q_R(u)$, $Q_S(u)$, and $Q_W(u)$ are asymptotically equivalent and χ^2 distributed with $m - s$ d.f. conditionally on $F_h^a \cap \{\eta > 0\}$ for $a > 1$.*

Proof. First consider

$$\ln(L(\hat{\theta}(u), u)/L(\theta, u)) = \int_0^1 (\hat{\theta}(u) - \theta)\, M'(\theta + b(\hat{\theta}(u) - \theta), u)\, db.$$

It is easy to see that on F_h^a this variable is asymptotically equivalent to $(\hat{\theta}(u) - \theta)\, M'(\theta, u) - (\hat{\theta}(u) - \theta)\, W(\theta, u)(\hat{\theta}(u) - \theta)'/2$ and (using a result in the proof of Theorem 4.1) to $M(\theta, u)\, W^{-1}(\theta, u)\, M'(\theta, u)/2$.

In the same way $\ln(L(\hat{\beta}(u)S, u)/L(\theta, u))$ and $M(\theta, u)\, S'(SW(\theta, u)S')^{-1} SM'(\theta, u)/2$ are asymptotically equivalent on F_h^a.

On F_h^a, $M(\theta, u)(I_m + S'(SW(\theta, u)S')^{-1}SW(\theta, u))\, W^{-1/2}(\theta, u)$ and $M(\hat{\beta}(u)S, u)\, W^{-1/2}(\theta, u)$ are asymptotically equivalent. Combining these three results it follows that $Q_R(u)$ is asymptotically equivalent to

$$M(\hat{\beta}(u)S, u)(W^{-1}(\theta, u) - S'(SW(\theta, u)\, S')^{-1}S)\, M(\hat{\beta}(u)S, u). \tag{5.5}$$

Using the fact that $R'R = I_{m-s}$, it follows with elementary matrix algebra that (5.5) equals $M(\hat{\beta}(u)S, u)\, RR'W^{-1}(\theta, u)\, RR'M'(\hat{\beta}(u)S, u)$, which is asymptotically equivalent to $Q_S(u)$. This completes the proof. ∎

6. Goodness-of-Fit Tests

6.1. *General Case*

We shall consider a general method of constructing goodness-of-fit tests for a model given by (1.1). Let $K(\theta, t)$ be a predictable $r' \times k$ matrices. Under some restrictions on K,

$$Q(\theta, u) = \int_0^u dG(\theta, t)\, K(\theta, t) \tag{6.1}$$

is a k-dimensional square integrable martingale with expectation 0. It seems natural to base a goodness-of-fit test on this martingale property and reject the model if $Q(\theta, u)$ is too far away from 0. Since θ is unknown, it has to be estimated from the realization. We will therefore base the test on the observed balue of $Q(\hat{\theta}(u), u)$.

It is convenient to regard a goodness-of-fit test as a test of a parametrical hypothesis in an extended model. Let $\bar{\theta} = (\theta_1, ..., \theta_m, \theta_{m+1}, ..., \theta_{m+k})$ and define the functions

$$g_i(\bar{\theta}, t) = h_i(\theta, t) + \sum_{j=1}^{k} \theta_{j+m} K_{ij}(\theta, t), \tag{6.2}$$

$i = 1, ..., r'$ and $g_r(\bar{\theta}, t) = 0$. If we assume that the counting process N has intensities of the form

$$\lambda_i(\bar{\theta}, t) = \exp\{ g_i(\bar{\theta}, t)\}\, Y_i(t)\, \alpha(t), \tag{6.3}$$

$i = 1, ..., r$, then a test of the parametrical hypothesis $\theta_{m+1} = \cdots = \theta_{m+k} = 0$ is a test of the original model when $g_i(\bar{\theta}, t) = h_i(\theta, t)$ holds. This parametrical hypothesis can be expressed in the form $\bar{\theta}R = 0$, where $R'R = I_k$. Of the three test statistics discussed in the previous section, the score statistic seems to be the most appropriate for a goodness-of-fit test, since it only depends on the estimate of θ derived under the null hypothesis. We need, however, an expression for the matrix $T(\theta, u)$ in the extended model. If the hypothesis holds (i.e., $\bar{\theta} = (\theta_1, ..., \theta_m, 0, ..., 0)$ then

$$T(\bar{\theta}, u) = \begin{pmatrix} \tilde{T}(\theta, u) & C(\theta, u) \\ C'(\theta, u) & V(\theta, u) \end{pmatrix}, \tag{6.4}$$

where $\tilde{T}(\theta, u) = \int_0^u H'(\theta, t)\, \Pi(\theta, t)\, H(\theta, t)\, d\bar{N}(t)$, $C(\theta, u) = \int_0^u H'(\theta, t)\, \Pi(\theta, t)\, K(\theta, t)\, d\bar{N}(t)$, and $V(\theta, u) = \int_0^u K'(\theta, t)\, \Pi(\theta, t)\, K(\theta, t)\, d\bar{N}(t)$. Theorem 5.1 implies

THEOREM 6.1. *If the extended model* (6.3) *satisfies* (A1)–(A3)) *then*

$$Q(\hat\theta(u), u)[V(\hat\theta(u), u) - C'(\hat\theta(u), u)\,\tilde T^{-1}(\hat\theta(u), u)\,C(\hat\theta(u), u)]^{-1}Q'(\hat\theta(u), u) \tag{6.5}$$

is asymptotically χ^2 *distributed with k d.f. if the model is true conditionally on* $F_g^1 \cap \{\eta > 0\}$.

Proof. Let S be the $m \times (m+k)$ matrix of the form $(I_m, 0)$ (where 0 is a matrix with only zeros). Then we can identify $Q(\hat\theta(u), u)$ with $M(\hat\theta(u)S, u)$ used in the score test statistic. The matrix $RR'T^{-1}(\hat\theta(u)S, u)\,R'R$ is the lower right $k \times k$ submatrix of $T^{-1}(\hat\theta(u), u)$. This matrix equals by elementary matrix algebra the inverse of $V(\hat\theta(u), u) - C'(\hat\theta(u), u)\,\tilde T^{-1}(\hat\theta(u), u)\,C(\hat\theta(u), u)$. ▮

6.2. *A Special Test*

We will in this section consider a particular goodness-of-fit test based on the statistic

$$\tilde Q(\theta, u) = \int_0^u dG(\theta, t)\,\tilde K(\theta, t) \tag{6.6}$$

where $\tilde K(\theta, t) = (k_1(\theta, t), ..., k_{r'}(\theta, t))$ and

$$k_i(\theta, t) = \begin{cases} \ln(Y_i(t)/Y_r(t)) + h_i(\theta, t) & \text{if } \; Y_i(t)\,Y_r(t) \neq 0 \\ \ln(Y_i(t)) + h_i(\theta, t) & \text{if } \; 0 = Y_r(t) \neq Y_i(t) \\ -\ln(Y_r(t)) + h_i(\theta, t) & \text{if } \; 0 = Y_i(t) \neq Y_r(t) \\ h_i(\theta, t) & \text{if } \; Y_i(t) = Y_r(t) = 0. \end{cases}$$

In the previous section we saw that this goodness-of-fit test can be considered as a test of the parametrical hypothesis $v = 1$ in the model

$$\lambda_i(\theta, t) = e^{vh_i(\theta, t)}Y_i^v(t)\,\alpha(t). \tag{6.7}$$

The alternatives can perhaps be considered as a kind of Lehmann alternatives on the intensity level.

The goodness-of-fit test based on (6.6) is of interest for several reasons. First, it can be seen as an analogue of the conditional exact test derived for multiplicative Poisson models in [10]. Second, it can be derived by an argument that in essence is similar to one used in [3]. The logarithm of the partial likelihood equals

$$\int_0^u \sum_{i=1}^r (\ln(Y_i(t)) + h_i(\theta, t))\,dN_i(t) - \int_0^u \ln\left(\sum_{i=1}^r e^{h_j(\theta, t)}Y_j(t)\right) d\bar N(t).$$

This is a jump process with the compensator

$$\int_0^u \sum_{i=1}^r (\ln(Y_i(t)) + h_i(\theta, t))\, p_i(\theta, t)\, \bar{\lambda}(\theta, t)\, dt$$
$$- \int_0^u \ln\left(\sum_{i=1}^r e^{h_j(\theta, t)} Y_j(t) \right) \bar{\lambda}(\theta, t)\, dt.$$

Taking the difference between these two expressions and substituting the unknown θ with its estimate $\hat{\theta}(u)$ and integrating with $d\bar{N}(t)$ instead of $\bar{\lambda}(\hat{\theta}(u), t)\, dt$, we obtain $\tilde{Q}(\hat{\theta}(u), u)$.

The asymptotic properties of the test statistic can be derived from Theorem 6.1. In this special case we have

$$C_i(\theta, u) = \int_0^u \left(\sum_j H_{ij}(\theta, t)\, p_j(\theta, t)\, k_j(\theta, t) \right.$$
$$\left. - \sum_j H_{ij}(\theta, t)\, p_j(\theta, t) \sum_j p_j(\theta, t)\, k_j(\theta, t) \right) d\bar{N}(t),$$

$i = 1, ..., m$, and

$$V(\theta, u) = \int_0^u \left(\sum_j p_j(\theta, t)\, k_j^2(\theta, t) - \left(\sum_j p_j(\theta, t)\, k_j(\theta, t) \right)^2 \right) d\bar{N}(t).$$

7. Two Theorems on the Convergence of Martingales

In this section we will state some results for convergence of martingales that are used in the previous sections. No proofs are given here. Proofs for more general settings can be found in, e.g., [6 or 9]. The versions quoted here can also be found in [12].

In this section N is an r dimensional counting process defined on $(\Omega, \mathscr{A}, P)$ and adapted to the filtration $\mathscr{A}_t$, $t \in [0, \infty[$. Let $\mathscr{A}_\infty = \bigcup_{t>0} \mathscr{A}_t$. The elements of N have continuous intensities $\lambda_i(t)$, $i = 1, ..., r$. We will give some results on the asymptotic behaviour of the m-dimensional martingale

$$G(u) = \int_0^u [dN(t) - \lambda(t)\, dt]\, \zeta(t)$$

as $u \to \infty$, where $\zeta(t)$ is an $r \times m$ matrix whose elements are predictable functions. The predictable quadratic variation of G is

$$W(u) = \left[\sum_{i=1}^r \int_0^u \zeta_{ij}(t)\, \zeta_{ik}(t)\, \lambda_i(t)\, dt \right]_{j=1,m}^{k=1,m}.$$

We will assume that $\mathbf{W}(u)/\bar{W}(u)$ is bounded away from 0, and that $\int_0^u |\zeta_{ij}^3(t)|\ \bar{\lambda}(t) < \infty$ for all i and j and $u < \infty$.

THEOREM 7.1. *$G(u) W^{-1}(u) \to 0$ a.s. on the set $\{W(\infty) = \infty\}$.*

Proof. In case $m = 1$, G will be a scalar martingale. For this case the theorem follows from a result in [8]. For the general case it will, by the same argument, hold that $G_i(u)/W_{ii}(u) \to 0$ a.s. as $u \to \infty$ on the set $\{W_{ii}(\infty) = \infty\}$. Now

$$|G(u)\ W^{-1}(u)| \leqslant \left|\sum_i G_i(u)/W_{ii}(u)\right| \bar{W}(u)/\mathbf{W}(u)$$

and the theorem follows immediately. ▮

THEOREM 7.2. *If there is a sequence $b(u) \to \infty$, and*

$$W(u)/b(u) \to \eta$$

in probability as $u \to \infty$, where η is a $\mathscr{A}_\infty$ measurable random matrix, then

$$G(u)\ W^{-1/2}(u) \to U^*$$

$\mathscr{A}_\infty$-mixing conditionally on any $\mathscr{A}_\infty$-measurable subset of $\{\eta > 0\}$, where U^* is $N(0, I_m)$ distributed.

Proof. This is a straightforward application of Theorem 5.2 in [12]. (Observe that with $\sigma_u = \inf\{t;\ \bar{W}(t) \geqslant \sqrt{b(u)}\}$, $W(\sigma_u)/b(u) \to 0$ as $u \to \infty$). Similar convergence theorems can also be found in [5, 6, 7, or 9]. ▮

8. THREE EXAMPLES

The aim of the following three examples is to illustrate how estimates and tests look in some simple cases. It would be possible to give exact conditions when the assumptions (A1)-((A3) are satisfied. Since this would be a very technical, and not very informative, exercise we have chosen only to indicate what is needed.

8.1. *A Model for Time-Inhomogeneous Poisson Processes*

Assume that we, according to some model, are to observe r independent Poisson processes with intensities,

$$\lambda_i(\theta, t) = e^{\theta_i} Y_i(t)\ \alpha(t),$$

$i = 1, ..., r$. Here the non-random functions Y_i and α describe timeinhomogenity of the processes. The weight-functions Y_i's are observable and α, that is common for all processes, is not. The partial ML estimate of θ will then solve the equations,

$$N_i(u) = \int_0^u p_i(\gamma, t)\, d\bar{N}(t),$$

$i = 1, ..., r$. If the Y_i's and α behave "reasonably" it follows from the theory developed above that the estimate is consistent and that it, after a random normation, is asymptotically normally distributed. Such a behaviour could, e.g., be that all weight-functions and α are uniformly bounded away from 0 and ∞. In such cases $\{\eta > 0\}$ and F_h^0 equal the entire sample space with the possible exclusion of a null set. There will also be needed some asymptotic regularity in order that assumption (A3) shall be satisfied.

We can apply the special goodness-of-fit test suggested in Section 6.2. According to this test we shall reject the model if the observed value of

$$Q(u) = \sum_{i=1}^{r} \int_0^u \ln(Y_i(t))(dN_i(t) - p_i(\hat{\theta}(u), t)\, d\bar{N}(t))$$

is not sufficiently close to zero. Denote by

$$\begin{aligned} D(u) = \int_0^u \left(p_i(\hat{\theta}(u), t)(\ln(Y_i(t)))^2 - \sum_{j=1}^{r} p_j(\hat{\theta}(u), t)(\ln(Y_j(t)))^2 \right) d\bar{N}(t) \\ - \sum_{i=1}^{r} \left(\left(\int_0^u p_i(\hat{\theta}(u), t) \right. \right. \\ \left. \left. \times \left(\ln(Y_i(t)) - \sum_{j=1}^{r} p_j(\hat{\theta}(u), t)\, Y_j(t) \right) d\bar{N}(t) \right) \middle/ N_i(u). \right. \end{aligned}$$

We shall reject the model if $Q^2(u)/D(u)$ exceeds some suitable percentile of the χ^2 distribution with one degree of freedom. This test exploits variations in the relative weights $Y_i(t)/Y_j(t)$ $i, j = 1, ..., r$. In order that the asymptotic results shall hold there has to be a substantial such variation.

8.2. *A Birth-and-Death Process*

Let $N_1(u)$ count the number of births and $N_2(u)$ count the number of deaths during $[0, u]$ is a linear birth-and-death process. If the process starts with one individual at time $u = 0$ then there will be $F(u) = 1 + N_1(u) - N_2(u)$ individuals at time u. Assume that the intensities can be written as

$$\lambda_i(t) = e^{\theta_i} F(t-)\, \alpha(t), \qquad i = 1, 2,$$

with the usual normation $\theta_2 = 0$. Here the unknown function α describes a common time-dependent variation in the birth and death intensities and e^{θ_1} is the ratio of the birth intensity and the death intensity. Such a process may, with positive probability, die out in finite time. If this is the case we will, after some time, not observe any more births or deaths; i.e., we will not get any more information usable to estimate θ_1. There will thus not be any estimate with "good" asymptotic properties. However, if $\theta_1 > 0$ and $A(u) = \int_0^u \alpha(t)\,dt \to \infty$ as $u \to \infty$ there will, according to well-known theory, be realizations which live on and generate new information as the observation time grows. It is possible to show that $W(\theta, u)/A(u) \to \eta$ as $u \to \infty$, where η is some random variable. If we apply the results obtained above, we can show that for realizations in the set $\{\eta > 0\}$ the estimate

$$\hat{\theta}_1(u) = \ln(N_1(u)/N_2(u))$$

is consistent. With $S(u) = \int_0^u (N_1(t) + N_2(t))\,dt$ it also follows that

$$(N_1(u) + N_2(u)/\sqrt{(N_1(u)\,N_2(u)\,S(u))}(\ln(N_1(u)/N_2(u)) - \theta_1)$$

is asymptoticallly $N(0, 1)$ distributed conditionally on $\{\eta > 0\}$.

8.3. *A Model for Analysing the Effect of a Traffic Safety Measure*

We will suggest (a simplistic) model for the study of certain kinds of road traffic safety measures. Suppose that we simultaneously observe traffic accidents in r road crossings. The accidents in the ith crossing will occur according to a counting process whose intensity depends on (a) the safety standard of the individual crossing (described by a parameter θ_i), (b) the amount of traffic passing through the crossing (measured by $Y_i(t)$), and (c) by a time dependent parameter ($\alpha(t)$) common to all crossings describing seasonal variations, weather conditions, etc. If an accident occurs, some sort of safety measure is taken in order to decrease the accident intensity. We will assume that the measure changes (hopefully lowers) the intensity by a factor e^{ν} and that it is effective for a time period of length a. Police supervision may be a safety measure having an effect of this kind.

The accident intensity of the ith crossing is

$$\lambda_i(t) = e^{\theta_i + \nu J_i(t)}\, Y_i(t)\, \alpha(t),$$

where

$$J_i(t) = \begin{cases} 1 & \text{if} \quad N_i(t) > N_i(t-a) \\ 0 & \text{if} \quad N_i(t) = N_i(t-a). \end{cases}$$

After renorming the model we can write

$$\lambda_i(t) = e^{\theta_i + \nu(J_i(t) - J_r(t))} Y_i(t)\, \alpha(t) \tag{8.1}$$

and assume that $\theta_r = 0$.

The model given by (8.1) is of the type studied in this paper. Let

$$p_i(\theta, \nu, t) = e^{\theta_i + \nu J_i(t)} Y_i(t) \Big/ \sum_{j=1}^{r} e^{\theta_j + \nu J_j(t)} Y_j(t), \qquad i = 1, ..., r'.$$

Then the ML-estimates of θ and ν shall solve the equations:

$$\int_0^u (dN_i(t) - p_i(\gamma, \delta, t)\, d\bar{N}(t)) = 0, \qquad i = 1, ..., r',$$

$$\sum_{i=1}^{r'} \int_0^u (dN_i(t) - p_i(\gamma, \delta, t)\, d\bar{N}(t))(J_i(t) - J_r(t)) = 0.$$

The last of these equations is trivially identical to

$$\sum_{i=1}^{r} \int_0^u (dN_i(t) - p_i(\gamma, \delta, t)\, d\bar{N}(t))\, J_i(t) = 0.$$

To be able to use the theoretical results above we have to verify the very abstract assumptions (A1)–(A3). In the present example we could perhaps assume that $Y_i(t)\,\alpha(t)$ has small variation in time and that $b(u) = u$ would be a convenient norming constant. We would also expect that the set $\{\eta > 0\}$ is the entire sample space. It is easily verified that all assumptions are satisfied in the simplest case of all, namely when $Y_i(t) \equiv Y_i > 0$, $\alpha(t) \equiv \alpha > 0$, and $\nu = 0$.

It may in a situation of this kind be of special interest to test the hypothesis $\nu = 0$. If this hypothesis holds then the safety measure has no effect on the accident process. To illustrate how such a test can look we will consider only the simple case where all Y_i-processes are constant. In this case the functions $p_i(\theta, t)$ will not depend on time t. Under this hypothesis, the ML estimates of the θ parameters will solve the simple equations:

$$N_i(u) = e^{\theta_i} Y_i \Big/ \sum_{i=1}^{r} e^{\theta_j} Y_j$$

$i = 1, ..., r'$. In fact,

$$p_i(\hat{\theta}(u), t) = N_i(u)/\bar{N}(u).$$

The score test will depend on the deviance of

$$\sum_{i=1}^{r} (S_i(u) - N_i(u)\, \bar{S}_i(u)/\bar{N}(u))$$

from 0. Here $S_i(u)$ is the number of accidents in the ith road crossing when the safety measure is in force (i.e., when an accident has occured less than a time units before) and $\bar{S}_i(u)$ is the total number of accidents in all crossing during this time. All counts are made up till time u. This seems a very natural test statistic. In order to calculated the score test, e.g., one now has to find the lower right element of the inverse of the $r \times r$ matrix $T(\hat{\theta}(u), t)$. Trivial but long calculations yield that this element can be written as

$$D(u) = \sum_{i=1}^{r} N_i(u)\, \bar{S}_i(u)/\bar{N}(u) - \sum_{i=1}^{r} \sum_{j=1}^{r} N_i(u)\, N_j(u)\, \bar{S}_{ij}(u)/\bar{N}^2(u)$$
$$- \sum_{i=1}^{r} \left(N_i(u)\, \bar{S}_i(u)/\bar{N}(u) - N_i(u) \sum_{j=1}^{r} N_j(u)\, \bar{S}_j(u)/\bar{N}^2(u) \right)^2 \Big/ N_i(u),$$

where $\bar{S}_{ij}(u)$ is the total number of accidents in all crossings during times when both the ith and the jth crossing are subject to the safety measure. (Of course, $\bar{S}_{ii}(u) = \bar{S}_i(u)$).

In this case the score test rejects the hypothesis of no effect of the measure if the test statistic

$$\left(\sum_{i=1}^{r} (S_i(u) - N_i(u)\, \bar{S}_i(u)/\bar{N}(u)) \right)^2 \Big/ D(u)$$

is larger than some suitable percentile in the χ^2 distribution with one degree of freedom.

REFERENCES

[1] AITCHINSON, J., AND SILVEY, S. D. (1958). Maximum likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29**, 813–828.

[2] ANDERSEN, P. K., AND GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100–1125.

[3] COX, D. R. (1962). Further results on tests of separate families of hypotheses. *J. Roy. Statist. Soc. Ser. B* **24**, 406–424.

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

[5] FEIGIN, P. D. (1985). Stable convergence of semimartingales. *Stochastic Process. Appl.* **19**, 125–134.

[6] JACOD, J., AND SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin/Heidelberg.

[7] HUTTON, J. E., AND NELSON, P. I. (1984). *A Mixing and Stable Central Limit Theorem for Continuous Time Martingales.* Kansas State University Technical Report 42.

[8] LEPINGLE, D. (1978). Sur le comportement asymptotique des martingales locales. In *Lecture Notes in Math.*, Vol. 649, pp. 148–161, Springer-Verlag, Berlin/New York.

[9] LIPTSER, P. S., AND SHIRYAEV, A. N. (1986). *Theoriya Martingalov.* Nauka, Moscow. [Russian]

[10] SVENSSON, Å. (1981). On a goodness-of-fit test for multiplicative Poisson models. *Ann. Statist.* **9**, 697–704.

[11] SVENSSON, Å. (1990). Goodness-of-fit test and estimation in some counting process models with multiplicative structure. *Ann. Statist.* **17**.

[12] SVENSSON, Å. (1990). Asymptotic estimation in counting processes with parametric intensities based on one realization. *Scand. J. of Statist.* (to appear).