International Statistical Review, **51** (1983), pp. 165–174. Longman Group Limited/Printed in Great Britain

An Extension of Cox's Regression Model

Søren Johansen

Institute of Mathematical Statistics, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

Summary

It is shown how one can construct a model for a jump process depending on an arbitrary intensity measure with the property that if the measure is absolutely continuous it reduces to Cox's regression model for survival data. The model has the property that the maximum likelihood estimator of the parameters are Cox's estimate for the regression parameter and the Nelson–Aalen estimate for the measure. Cox's partial likelihood for the regression parameter becomes a partially maximized likelihood and the model has a property corresponding to S-ancillarity which explains the partial likelihood.

Key words: Cox's regression model; Multiplicative intensity models; S-ancillarity; Nonparametric maximum likelihood.

1 Introduction

The starting point of this investigation is the Cox model for survival data which can be described as follows. The life length of n individuals are distributed independently with a hazard function or an intensity $e^{\beta' z_i(t)}\lambda(t)$ where $\lambda(.)$ and $\beta \in R^k$ are unknown parameters whereas $z_i(t) \in R^k$ is a vector of known cofactors for the ith individual at time t.

The likelihood function becomes

$$L = \prod_i \left[e^{\beta' z_i(t_i)} \lambda(t_i) \exp\left(-\int_0^{t_i} e^{\beta' z_i(u)} \lambda(u)\, du \right) \right].$$

Now define $Y_i^\beta(u) = e^{\beta' z_i(u)} 1\{u \leq t_i\}$, where $1\{.\}$ denotes an indicator function, and $Y^\beta(u) = \sum_i Y_i^\beta(u)$; then

$$\begin{aligned} L &= \left[\prod_i Y_i^\beta(t_i) \lambda(t_i) \right] \exp\left(-\int_0^\infty Y^\beta(u)\lambda(u)\, du \right) \\ &= \prod_i \frac{Y_i^\beta(t_i)}{Y^\beta(t_i)} \left[\prod_i Y^\beta(t_i)\lambda(t_i) \right] \exp\left(-\int_0^\infty Y^\beta(u)\lambda(u)\, du \right). \end{aligned}$$

Cox suggested to use the first factor

$$L_c(\beta) = \prod_i \frac{Y_i^\beta(t_i)}{Y^\beta(t_i)} = \prod_i \left[e^{\beta' z_i(t_i)} \Big/ \left(\sum_{j : t_j \geqq t_i} e^{\beta' z_j(t_i)} \right) \right] \tag{1.1}$$

for estimating β.

The second factor can be used to find an estimate for λ as follows. Let us assume that λ is piecewise constant on intervals of length ε and let I_i be the interval which contains t_i. Then we clearly find a maximum value of the second factor by setting

$$\lambda(u) = 0, \quad u \notin \bigcup_i I_i.$$

Thus we are left with

$$\prod_i \left[Y^{\beta}(t_i)\lambda(t_i) \exp\left(-\int_{I_i} Y^{\beta}(u)\, du\, \lambda(t_i)\right)\right]$$

since λ is constant on I_i. This expression maximizes for

$$\lambda(t_i) = \left[\int_{I_i} Y^{\beta}(u)\, du\right]^{-1} \simeq [\varepsilon Y^{\beta}(t_i)]^{-1},$$

which will tend to ∞ for $\varepsilon \to 0$. If we consider the measure $\Lambda(A) = \int \lambda(u)\, du$, where the integral is over A, we get that, as $\varepsilon \to 0$, this measure approaches a discrete measure with mass $1/Y^{\beta}(t_i)$ at the point t_i.

This heuristic argument, which is due to Bay & Mac (1981), suggests that the maximum likelihood estimate of β and Λ are found by first maximizing $L_c(\beta)$ to find $\hat{\beta}$ and then use

$$\hat{\Lambda}(A) = \sum_{t_i \in A} 1/Y^{\hat{\beta}}(t_i). \tag{1.2}$$

Breslow, in the discussion of Cox's paper (1972), has suggested a different estimate Λ^* which corresponds to a λ which is piecewise constant between jump points. Then Λ^* is absolutely continuous and coincides with $\hat{\Lambda}$ on intervals of the form $]t_i, t_{i+1}]$; see also Breslow (1974).

For both situations the value of the maximized second factor is independent of β which gives the result that $L_c(\beta)$ is a partially maximized likelihood function or likelihood profile. Various arguments have been given for considering the likelihood function L_c (Cox, 1975; Oakes, 1981). The method of deriving maximum likelihood estimates is a universal method which often gives a natural estimate that has been derived by heuristic means.

It is my opinion that it is of interest to ask the question of finding a model for which a given estimator is the maximum likelihood estimator.

The purpose of this paper is thus to find out if the above heuristic argument can be made precise. It turns out that it can, but in a slightly different model for Poisson variables where β and Λ appear in a natural way. This model which I shall call a regression model for Poisson variables depends on the parameters β and Λ, where Λ is an arbitrary measure on R_+. The maximum likelihood estimates of β and Λ are the estimates $\hat{\beta}$ and $\hat{\Lambda}$ above. Moreover, the model enjoys a property similar to S-ancillarity, which can also be used to argue that one should use $L_c(\beta)$ to estimate β even though it is estimated at the same time as the infinitely many parameters Λ. Finally the model has the property that if Λ is restricted to be of the form $\int \lambda(u)\, du$ then the model reduces to the Cox model.

The main idea in the construction of the model is taken from the relation between the Cox model and the multiplicative intensity model for a counting process formulated by Aalen (1978). Aalen considered a counting process N whose predictable compensator is of the form

$$\int_0^t \lambda(u) Y(u)\, du,$$

where $\lambda(.)$ was an unknown function and $Y(.)$ a predictable process depending on the history of N. For details see §§ 2, 3.

If we consider the number of deaths N in the Cox model, we have a counting process with intensity $Y^{\beta}(t)\lambda(t)$. Thus for a fixed β we have a multiplicative Aalen model and the estimate (1.2) is just the Nelson–Aalen estimate of $\int \lambda(u)\, du$.

Thus we first solve the problem of defining a multiplicative intensity model allowing for

$\int \lambda(u)\, du$ to be replaced by an arbitrary measure Λ, and then use this model to construct the one we were looking for.

With this brief survey of the motivation and methodology we then present a summary of some of the results of Jacod (1975) on multivariate counting processes in § 2 and define the general Aalen model in § 3. In § 4 we derive the maximum likelihood estimate in this model and in § 5 we then define the regression model for Poisson variables and find the maximum likelihood estimate and show that it has the required properties.

Finally, we shall discuss briefly the relation of this model to other work in this area and point out some problems with the proposed model.

2. Multivariate counting processes and the Poisson process

We consider the stochastic process X with the properties $X(0)=0$, $X(.)$ nondecreasing and integer valued with a finite number of jumps in a finite interval. Let (T_n, X_n), for $n=1,2,\ldots$, denote the jump points and the jump sizes respectively. We shall work with the process on $[0, 1]$ and assume for simplicity that $EX(1)<\infty$.

We can then define the multivariate counting process $\{N_x\}$ by

$$N_x(t)=\sum_n 1\{T_n \leqslant t, X_n = x\}, \quad x \in \mathbb{N},$$

which counts the number of jumps of size x up to time t. Now let Ω be the space of all sample paths of these multivariate counting processes i.e. the space of all possible realizations, with the nondecreasing σ-fields

$$\mathscr{F}_t=\sigma\{N_x(u), u \leqslant t, x \in \mathbb{N}\}.$$

We define the univariate counting process $N=\sum_x N_x$ and note that $X=\sum_x xN_x$.

Finally, we need the concept of a predictable process Y as a function $\Omega \times R_+ \rightarrow R$ which is measurable with respect to the σ-field generated by the functions $Z: \Omega \times R_+ \rightarrow R$ with the properties that $Z(., t)$ is measurable $\mathscr{F}_t$ and $Z(\omega, .)$ is left-continuous.

The results which we shall need in the following have all been taken from Jacod (1975), who discusses multivariate counting processes with a more general index than the integers which are relevant for our purpose.

First we need the compensator (or predictable projection) of the process $\{N_x\}$. This is defined as a collection $\{\nu_x\}$ of nonnegative random measures on R_+ with the properties that $\nu_x[0, t]$ is predictable for each x, $\sum_x \nu_x\{t\} \leqslant 1$ and $(N_x(t)-\nu_x[0, t])$, for $0 \leqslant t \leqslant 1$, is a martingale with respect to the σ-fields $\mathscr{F}_t$ for $0 \leqslant t \leqslant 1$. An explicit form for ν_x can be given in terms of the distributions of $\{T_i, X_i\}$, for $i=1,2,\ldots$. If these are sufficiently smooth one can show that

$$\nu_x[0, t]=\int_0^t \alpha_x(u)\, du$$

where

$$\alpha_x(t+)=\lim_{h \downarrow 0} \frac{1}{h} P(N_x(t+h)-N_x(t)=1 \mid \mathscr{F}_t)=\lim_{h \downarrow 0} \frac{1}{h} E(N_x(t+h)-N_x(t) \mid \mathscr{F}_t).$$

The process is called the intensity process of N_x. One should think of the compensator $\nu_x[0, t]$ as representing what is known about $N_x(t)$ just before time t. Thus in a sense $\nu_x[0, t]$ represents the systematic part of $N_x(t)$ given the past. It is a result that in order to define the whole process $\{N_x\}$ it is enough to specify the process $\{\nu_x\}$. We shall need to calculate densities of one probability measure P' with respect to P expressed in terms of the compensators $\{\nu'_x\}$ and $\{\nu_x\}$.

Let $\nu=\sum_x \nu_x$ and $\nu'=\sum_x \nu'_x$ and assume that $\nu'_x(A)=\int Y\,d\nu_x$ where the integral is over A, and where Y is some predictable function such that $\nu'\{t\}\leq 1$ and $\nu'\{t\}=1$ if $\nu\{t\}=1$. It then follows from Proposition 4.3 and Theorem 5.1 of Jacod (1975) that P' has density with respect to P and that the restriction to the σ-field $\mathscr{F}_t$ has the form

$$\left.\frac{dP'}{dP}\right|_{\mathscr{F}_t}=\prod_{u\leq t}\prod_x\left(\frac{d\nu'_x}{d\nu_x}\right)^{N_x(du)}\frac{\{\prod_{u\leq t}(1-\nu'(du))^{1-N(du)}\}}{\{\prod_{u\leq t}(1-\nu(du))^{1-N(du)}\}},\tag{2.1}$$

where we have used the product integral notation for a measure ν on the interval $[0, 1]$

$$\prod_{[0,t]}(1-\nu(du))=\prod_{u\leq t}(1-\nu_s\{u\})\exp(-\nu_c[0,t]),$$

where $\nu=\nu_s+\nu_c$ is the decomposition of ν into its atomic and continuous part. For the notion of product integral see Dollard & Friedman (1979).

Thus if we can find a multivariate counting process, we can construct new ones by specifying their density with respect to the old one by the above expression.

As the basic process we shall use the Poisson process. Consider a Poisson process $X(t)$, $t\in R_+=[0,\infty[$ on some probability space $(\Omega, \mathscr{F}, P)$. Let $\Lambda\in\mathscr{M}^+(R_+)$ denote a nonnegative Borel measure on $\mathscr{B}(R_+)$, the Borel sets of R_+, with the property that it is finite on bounded sets.

We shall assume that $X(0)=0$, $X(t)$ has independent increments, is piecewise constant, right-continuous and that the distribution of $X(t)-X(s)$ is Poisson with parameter $\Lambda(]s,t])$. We define the multivariate counting process $\{N_x\}$ as before.

The predictable compensator for the counting process $\{N_x\}$ with respect to the family $\{\mathscr{F}_t\}$ is given by

$$\nu_x[0,t]=\int_{[0,t]}\frac{\Lambda(du)^x}{x!}e^{-\Lambda(du)}\quad(x\in\mathbb{N}).\tag{2.2}$$

This notation means that

$$\nu_1[0,t]=\sum_{u\leq t}\Lambda_s\{u\}e^{-\Lambda_s\{u\}}+\Lambda_c[0,t],\tag{2.3}$$

$$\nu_x[0,t]=\sum_{u\leq t}\frac{\Lambda_s\{u\}^x}{x!}e^{-\Lambda_s\{u\}}\quad(x=2,3,\ldots).\tag{2.4}$$

We also define

$$\nu_0[0,t]=\sum_{u\leq t}e^{-\Lambda_s\{u\}}.\tag{2.5}$$

We then find the compensator ν for $N=\sum_x N_x$ as

$$\nu[0,t]=\sum_x\nu_x[0,t]=\sum_{u\leq t}(1-e^{-\Lambda_s(u)})+\Lambda_c[0,t];$$

hence $\nu_s\{u\}=1-e^{-\Lambda_s\{u\}}$ and $\nu_c=\Lambda_c$, and we find the result

$$\begin{aligned}\prod_{u\leq t}(1-\nu(du))&=\prod_{u\leq t}(1-\nu_s(du))\exp\left(-\int_0^t\nu_c(du)\right)\\&=\exp\left(-\sum_{u\leq t}\Lambda_s(u)-\int_0^t\nu_c(du)\right)=\exp(-\Lambda[0,t]).\end{aligned}\tag{2.6}$$

Note also that the compensator for $X=\sum_x xN_x$ is given by $\sum_x x\nu_x[0,t]=\Lambda[0,t]$.

3 The extended Aalen model

Consider first a counting process N which has compensator $\nu[0, t] = \int \alpha(u)\, du$, where the integral is over $[0, t]$. Aalen (1978) formulated the following multiplicative intensity model by specifying the dependence of α on the parameter $\lambda(.)$ and the past $Y_0(.)$ by

$$\alpha(u) = \lambda(u) Y_0(u), \tag{3.1}$$

where $\lambda(.)$ is an unknown nonnegative function such that $\int \lambda(u)\, du < \infty$, where the integral is over $[0, t]$, and Y_0 is predictable with respect to N.

He suggested estimating $\int \lambda(u) 1\{Y_0(u) > 0\}\, du$, where the integral is over $[0, t]$, by the estimator

$$\int_0^t \frac{N(du)}{Y_0(u)} = \sum_{T_i \leq t} \frac{1}{Y_0(T_i)}. \tag{3.2}$$

The properties of this estimator are investigated in Aalen (1978). Note that the estimate of $\int \lambda$ is only possible on the random set where $\{Y_0 > 0\}$, since if $Y_0 = 0$, then the process N is constant and does not contain information on λ.

The estimate of $\int \lambda$ is a discrete measure and it is therefore tempting to try to extend the model by allowing $\int \lambda(u)\, du$ to be replaced by the measure Λ. This causes the difficulty that at points where $\Lambda\{t\} > 0$, multiple jumps may occur. This is overcome by letting Λ give rise to a multivariate counting process where N_x counts the number of times x jumps occur simultaneously. It turns out that $X = \sum_x x N_x$, and not $N = \sum_x N_x$ has the compensator $\int Y_0\, d\Lambda$.

This is accomplished as follows. First we take a function Y which is predictable and bounded on $[0, t]$ for all t. We then specify the density with respect to the Poisson process with intensity measure Λ by defining the compensator for the new process to be

$$\nu'_x[0, t] = \int_0^t \frac{(Y(u)\Lambda(du))^x}{x!} \exp(-Y(u)\Lambda(du)) \quad (x \in \mathbb{N}), \tag{3.3}$$

with a similar interpretation as (2.3) and (2.4). We then find as before

$$\nu'_s\{u\} = 1 - \exp(-Y(u)\Lambda_s\{u\}), \quad \nu'_c[0, t] = \int_0^t Y(u)\nu_c(du),$$

$$\prod_{u \leq t} (1 - \nu'(du)) = \exp\left(-\int_0^t Y(u)\Lambda(du)\right),$$

and that X has compensator $\sum_x x\nu_x[0, t] = \int Y(u)\Lambda(du)$, where the inegral is over $[0, t]$. We obtain the derivative

$$\frac{d\nu'_x}{d\nu_x}(t) = Y(t)^x \exp(-(Y(t) - 1)\Lambda(dt))$$

and hence the density from (2.1)

$$\left.\frac{dP'}{dP}\right|_{\mathscr{F}_t} = \prod_{u \leq t} \prod_x [Y(u)^x \exp(-(Y(u) - 1)\Lambda(du))]^{N_x(du)} \frac{\prod_{u \leq t} (1 - \nu'(du))^{1 - N(du)}}{\prod_{u \leq t} (1 - \nu(du))^{1 - N(du)}}.$$

Now we reduce the expression as follows:

$$\prod_{u \leq t} \prod_x \exp(-\Lambda(du) N_x(du)) \prod_{u \leq t} (1 - \nu(du))^{1 - N(du)}$$

$$= \prod_{u \leq t} \exp(-\Lambda_s\{u\} N(du)) \prod_{u \leq t} (1 - \nu_s\{u\})^{1 - N(du)} \exp\left(-\int_0^t \nu_c(du)\right).$$

But $1-\nu_s\{u\}=\exp(-\Lambda_s\{u\})$ which shows that it reduces to $\exp(-\Lambda[0,t])$. A similar expression for ν' reduces to $\exp(-\int Y(u)\Lambda(du))$, where the integral is over $[0,t]$, and hence we find the density

$$\left.\frac{dP'}{dP}\right|_{\mathscr{F}_t}=\prod_{u\leq t} Y(u)^{X(du)}\exp\left(-\int_0^t (Y(u)-1)\Lambda(du)\right). \tag{3.4}$$

Let the probability measure P' determined by (3.4) be denoted by P_Λ. Now we can formulate the general multivariate intensity model for the jump process X or the multivariate counting process $\{N_x\}$ as the family of probability measures

$$\{P_\Lambda, \Lambda\in\mathscr{M}^+(R_+)\}. \tag{3.5}$$

In this model the compensator for X is $\int Y(u)\Lambda(du)$, where the integral is over $[0,t]$. Note that if Λ is restricted to be of the form $\int \lambda(u)\,du$ and $Y=Y_0$, then the model (3.5) reduces to (3.1). In this sense (3.4) is an extension of the model suggested by Aalen.

A different but more intuitive formulation of the model for X is the following. The distribution of $X(du)$ given $\mathscr{F}_u-$ is the Poisson distribution with a parameter depending on the past in the form $Y(u)\Lambda(du)$.

The reason for choosing the Poisson distribution is that it gives the right result in the end, that is the estimate for Λ, and the reason that it does so, is to be found in the reduction of the density (2.1) to (3.4) by the choice of ν'_x in (3.3). There seems to be no obvious way of specifying ν'_x such that N would get the compensator $\int Y(u)\Lambda(du)$.

The above intuitive formulation of the process can of course also be given for a univariate counting process N, by saying that the distribution of $N(du)$ given $\mathscr{F}_u-$ is a two point distribution with $\nu(du)$ specifying the probability that $N(du)=1$. This also explains the condition $\nu\{u\}\leq 1$, and the result that the compensator determines the whole process.

In general one can use this approach to define a dynamical exponential family specifying $X(du)$ given $\mathscr{F}_u-$ to have a distribution in some exponential family with a mean depending on the past and the parameter in some way. Note that each increment will depend on a new parameter, hence the 'number of parameters' will increase to infinity with time. The term dynamical refers to the fact that time enters into the model building and in the analysis in an essential way.

Heyde & Feigin (1975) and Feigin (1976, 1981) have introduced the notion of a conditional exponential family. For the situation considered here of a jump process, they discuss the case where the process of jumps is a Markov chain and the conditional distributions are given by members of an exponential family with a parameter of fixed dimension.

We do not use the full strength of the results of Jacod, since we are only working with integer valued processes, but it appears that the formulation via the multivariate counting processes is a convenient tool for the discussion of the processes and the likelihood function in the situation where the dependence on the past is more complicated.

4 The maximum likelihood estimator in the extended Aalen model

Let $X=\{X(t), t\in[0,1]\}$ have M jumps of size $X_1,\ldots,X_M$, at times $0<t_1<t_2<\ldots<t_M\leq 1$. We want to estimate Λ in the model $\{P_\Lambda, \Lambda\in\mathscr{M}^+(R_+)\}$.

We shall use the method of maximum likelihood as discussed by Kiefer & Wolfowitz (1956), which in this case amounts to finding a $\hat\Lambda$ such that

$$P_{\hat\Lambda}(x)\geq P_\Lambda(x),\quad \Lambda\in\mathscr{M}^+(R_+).$$

The natural guess is the Nelson–Aalen estimator

$$\hat{\Lambda}(A)=\int\frac{X(du)}{Y(u)}=\sum\frac{X_i}{Y(t_i)},\tag{4.1}$$

where the integral is over $A\cap\{Y>0\}$, and the sum is over $t_i\in A\cap\{Y>0\}$. We find the following expression for the probability

$$\begin{aligned}P_\Lambda(x)&=\left[\prod_{i=1}^{m}\frac{(Y(t_i)\Lambda\{t_i\})^{X_i}}{X_i!}\right]\exp\left(-\int_0^1 Y(u)\Lambda(du)\right)\\&=\left[\prod_{i=1}^{m}\frac{(Y(t_i)\Lambda\{t_i\})^{X_i}}{X_i!}\exp\left(-Y(t_i)\Lambda\{t_i\}\right)\right]\exp\left(-\int_B Y(u)\Lambda(du)\right),\end{aligned}$$

where $B=[0,1]\backslash\{t_1,\ldots,t_m\}$. On the set $\{Y=0\}$, this function does not depend on Λ, but on the set $\{Y>0\}$ we find that $P_\Lambda(x)$ is maximized for $\Lambda\{t_i\}=X_i/Y(t_i)$ and

$$\Lambda[\{Y>0\}\cap[0,1]\backslash\{t_1,\ldots,t_m\}]=0,$$

which is precisely $\hat{\Lambda}$.

Thus the Nelson–Aalen estimator is the maximum likelihood estimator in the extended Aalen model. The statistical properties of $\hat{\Lambda}$ have been investigated when Λ is absolutely continuous by Aalen (1978).

5 A regression model for the dynamical Poisson model

We shall now formulate a regression model for the jump processes considered in § 3. We let X_i $(i=1,\ldots,n)$ be independent jump processes observed on $[0,1]$, specified by the compensators

$$\int_0^t Y_i^\beta(u)\Lambda(du)=\int_0^t e^{\beta' z_i(du)}1\{X_i(u^-)=0\}\Lambda(du);$$

hence the process is killed just after the first jump. The parameters are Λ, the unknown underlying intensity measure, and $\beta\in R^k$ the regression coefficient which specifies how the covariates $z_i(.)$ influence the intensity. Let $P_i(\Lambda,\beta)$ denote the distribution of X_i, and let $\otimes P_i(\Lambda,\beta)$, for $i=1,\ldots,n$, denote their joint distribution; we thus consider the model

$$\left\{\bigotimes_{i=1}^{n}P_i(\Lambda,\beta),\ \Lambda\in\mathcal{M}^+(R_+),\ \beta\in R^k\right\}.\tag{5.1}$$

If Λ has the form $\int\lambda(u)\,du$ then the jump processes X_i will in fact be counting processes and describe the death of an individual at the time of the jump. The integrated hazard function of this death time will be

$$\int_0^t e^{\beta' z_i(u)}\lambda(u)\,du,$$

showing that we have the Cox model for survival data if Λ is absolutely continuous. In this sense we have an extension of the Cox model. We shall now derive the maximum likelihood estimates of this model. The probability of an outcome of the processes $X_i(u)$, $u\in[0,1]$ $(i=1,\ldots,n)$ then has the form, see (3.4),

$$\begin{aligned}L&=\prod_i\left\{\prod_{u\leq 1}\frac{(Y_i^\beta(u)\Lambda(du))^{X_i(du)}}{X_i(du)!}\exp\left(-\int_0^1 Y_i^\beta(u)\Lambda(du)\right)\right\}\\&=\prod_{u\leq 1}\prod_i\left(\frac{Y_i^\beta(u)}{\sum_j Y_j^\beta(u)}\right)^{X_i(du)}\Big/X_i(du)!\prod_{u\leq 1}(Y^\beta(u)\Lambda(du))^{X(du)}\exp\left(-\int_0^1 Y^\beta(u)\Lambda(du)\right),\end{aligned}\tag{5.2}$$

where $X=\sum_i X_i$ and $Y^\beta=\sum_i Y_i^\beta$.

For fixed β the last factor is exactly the likelihood in the general Aalen model and thus gives the estimate

$$\hat{\Lambda}_\beta(A) = \int_A \frac{X(du)}{Y^\beta(u)},$$

which inserted into (5.2) gives

$$\max_\Lambda L(\Lambda, \beta) = L_c(\beta) \prod_{u \le 1} \frac{X(du)^{X(du)}}{\Pi_i X_i(du)!} \exp\left(-\int_0^1 X(du)\right),$$

where $L_c(\beta)$ is given by (1.1). Thus the maximum likelihood estimate of β is found by maximizing $L_c(\beta)$.

Hence we have seen that $(\hat{\beta}, \hat{\Lambda})$ are the estimates of β and Λ discussed in § 1, and in this extended model the estimation procedure can be explained by the method of maximum likelihood without having to invoke the notion of partial likelihood.

The model has an extra property, however, which perhaps accounts for the fact that the estimation of β is useful even though it is estimated simultaneously with the infinitely many parameters Λ. Recall that in a statistical problem with parameters (β, τ) we call the statistic $t(x)$ S-ancillary for β if the parameters (β, τ) vary in a product space and if the likelihood function factorizes as follows:

$$f(x, \beta, \tau) = g(t(x), \tau) h(x \mid t(x), \beta);$$

see Barndorff-Nielsen (1978, p. 208).

Unfortunately it is not true that $X = \sum_i X_i$ is S-ancillary for β in the parametrization (β, Λ) but, if we consider the experiment $\{X_i(du)\}$, for $i = 1, \ldots, n$, given $\mathcal{F}_u-$, then we have n independent Poisson distributions with parameters

$$Y_i^\beta(u)\Lambda(du) = e^{\beta' z_i(u)} 1\{X_i(u^-) = 0\}\Lambda(du),$$

and a likelihood function of the form

$$L(\beta, \Lambda(du)) = \frac{\exp(\beta' \sum_i z_i(u) X_i(du)) \Lambda(du)^{\sum_i X_i(du)}}{\Pi_i X_i(du)!} \prod_i 1\{X_i(u^-) = 0\} \exp\left(-\sum_i Y_i^\beta(u)\Lambda(du)\right).$$

It is seen that $U = \sum_i z_i(u) X_i(du)$ and $X(du) = \sum_i X_i(du)$ are sufficient. We introduce the mixed parametrization (β, τ) with

$$\tau(du) = E(X(du) \mid \mathcal{F}_u-) = \sum_i Y_i^\beta(u)\Lambda(du).$$

Then it is easily seen that $\beta, \tau(du)$ vary in a product space and, since the likelihood function factorizes into

$$\frac{\tau(du)^{X(du)}}{X(du)!} e^{-\tau(du)} \frac{X(du)!}{\prod_i X_i(du)!} \prod_i \left(\frac{Y_i^\beta}{Y^\beta}\right)^{X_i(du)},$$

it follows that $X(du)$ is S-ancillary for β. Thus at any instant of time u we have that the innovation $\{X_i(du)\}$ in the experiment conditional on the past $\mathcal{F}_u-$ has the property that $X(du) = \sum_i X_i(du)$ is S-ancillary for β. This leads naturally to conditioning on $X(du)$, when making inference on β, and hence to the factorizing which gives the partial likelihood L_c.

The succesive conditioning is the argument that Cox used to define the partial likelihood. The present problem has the extra structure of successive S-ancillarity which is a specific instance of the general principles formulated by Cox, that (i) no omitted factor should contain important information about β, and that (ii) the partial likelihood itself should only depend on β (Cox, 1975, p. 270).

6 Discussion

We have constructed a model for an integer valued jump process where the increment given the past is Poisson distributed with a parameter depending in a specified way on the past.

If the parameter Λ is a discrete measure on the points $(1, 2, \ldots)$ say, then this model has been considered by Andersen & Rudemo (1980).

If Λ is absolutely continuous the model reduces to the Cox model, and in this sense is an extension of the Cox model. Cox (1972) and Kalbfleisch & Prentice (1980, pp. 35–38) have discrete versions of the Cox model, but they give different estimators of the parameters; see also Jacobsen (1982). Although the model is an extension, it is not obvious that the new measures we get for Λ discrete correspond in a natural way to the underlying phenomenon that the Cox model describes, since the individual sample paths may have a jump of size larger than one.

Thus the model should not be considered an alternative to the discrete version suggested by Cox and Kalbfleisch & Prentice.

The fact that the model considered is the natural one in which $\hat{\beta}$ and $\hat{\Lambda}$ should be considered, indicates that $\hat{\Lambda}$ is not the natural estimate in the Cox model. It is instructive to consider the simplest possible Cox model, namely for $\beta = 0$. In this case we have independently and identically distributed waiting times T_i, and the natural extension is to consider their distribution F to be an arbitrary probability on R_+.

In this alternative extension it is easy to see that $\hat{F}$ is the empirical distribution and hence that the estimator of the survivor function is $\hat{G} = 1 - \hat{F}$. Since we can define the intensity measure $\Lambda(A) = \int F(du)/(1 - F(u^-))$, where the integral is over A, we have that $\Lambda\{t\} \leq 1$, and that

$$G(t) = 1 - F(t) = \prod_{[0,t]} (1 - \Lambda(du)).$$

This shows that in the natural extension we have

$$\hat{G}(t) = \prod_{[0,t]} (1 - \hat{\Lambda}(du)) = \prod_{[0,t]} \left(1 - \frac{N(du)}{n - N(u^-)}\right) = 1 - \hat{F}(t).$$

In the extended Poisson model we get that the survivor function is

$$G(t) = \exp\left(-\int_0^t \Lambda(du)\right),$$

and hence that the estimate is

$$\hat{G}(t) = e^{-\hat{\Lambda}(t)} \neq \prod_{[0,t]} (1 - \hat{\Lambda}(dt)).$$

The reason for bringing in this property of the model is to point out the essential difficulty in the extension. If we want the extension of the counting process to be a counting process then we must be sure that the atoms in the intensity measure are bounded by 1. The reason for this is that a counting process is also a dynamical exponential family, namely a family of two point measures, where $\Lambda(dt)$ is the probability of success given the past. For $\beta = 0$ it is easy to consider this extension but for $\beta \neq 0$ the constraint imposed on the atoms of the intensity measure is

$$e^{\beta_i'(t)}\Lambda\{t\}\leq 1, \tag{6.1}$$

which gives a strange relation between the parameters. By considering a Poisson extension we avoid this and the atom (6.1) appears as a mean value instead which does not impose any constraints on the parameters.

Acknowledgements

I want to thank Martin Jacobsen for many critical remarks and Niels Keiding, Hans Bay and Flemming Mac for stimulating discussions. The paper has improved from critical comments by the referee.

References

Aalen, O.O. (1978). Non-parametric inference for a family of counting processes. *Ann. Statist.* **6,** 701–726.
Andersen, S. & Rudemo, M. (1980). Private communication.
Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. New York: Wiley.
Bay, H. & Mac, F. (1981). Thesis. University of Copenhagen.
Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30,** 89–99.
Cox, D.R. (1972). Regression models and life tables (with discussion). *J.R. Statist. Soc.* B **34,** 187–220.
Cox, D.R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.
Dollard, J.D. & Friedman, C.N. (1979). *Product Integration*. London: Addison-Wesley.
Feigin, P.D. (1976). Maximum likelihood estimation for continuous time stochastic processes. *Adv. Appl. Prob.* **8,** 712–736.
Feigin, P.D. (1981). Conditional exponential families and a representation theorem for asymptotic inference. *Ann. Statist.* **9,** 597–603.
Heyde, C.C. & Feigin, P.D. (1975). On efficiency and exponential families in stochastic process estimation. In *Statistical Distributions in Scientific Work*, **1,** Ed. G.P. Patil, S. Kotz and J. Ord, pp. 227–240. Boston: Riedel.
Jacobsen, M. (1982). Maximum likelihood estimation in the multiplicative intensity model. Preprint. University of Copenhagen.
Jacod, J. (1975). Multivariate point processes: predictable projection, Radon–Nikodym derivatives, representation of martingales. *Z. Wahr. verw. Geb.* **31,** 235–253.
Kalbfleisch, J.D. & Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
Keifer, J. & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27,** 887–906.
Oakes, D. (1981). Survival times: Aspects of partial likelihood. *Int. Statist. Rev.* **49,** 235–264.

Résumé

On construit un modèle pour un processus des sauts à valeurs entiers dépendant d'une mesure sur R_+. Si cette mesure est absolument continue, le modèle coincide avec le modèle de régression de Cox pour les observations des durées de vie. L'estimateur de vraisemblance des coéfficients de régression est l'estimateur de Cox, et l'estimateur de la mesure est l'estimateur de Aalen et Nelson. La fonction de vraisemblance maximée par la mesure donne la vraisemblance partielle de Cox.

[*Paper received September* 1982, *revised January* 1983]