# Cox's factoring of regression model likelihoods for continuous time processes *

Priscilla E. Greenwood† Wolfgang Wefelmeyer‡

**Abstract**

Cox showed that the likelihood of regression models for discrete-time processes factors into a partial likelihood and a product of conditional laws for the covariates, given the history. Jacod constructed a partial likelihood for continuous-time regression models in terms of the predictable characteristics of the response process. Here we prove a factorization of the likelihood, analogous to Cox's, assuming both the response and the covariates to be semimartingales. The result is useful for counting process regression modeling and inference, and also for regression involving continuous processes and diffusions with jumps.

## 1 Introduction

Suppose we observe a response process $X$ and a vector $V$ of covariate processes. A regression model specifies how the history of $X$ and $V$ affects the evolution of the response. In discrete time one models the conditional densities $p_n^\vartheta(x)$ of $X_n$ given the past observations $X_1, \dots, X_{n-1}$ and $V_1, \dots, V_{n-1}$. Cox (1975) suggested basing inference about the parameter $\vartheta$ on the *partial likelihood*

$$\overline{Z}^\vartheta = \prod_{n=1}^N p_n^\vartheta(X_n), \tag{1.1}$$

and showed that the full likelihood factors as

$$Z = \overline{Z}^\vartheta Z_*, \quad Z_* = \prod_{n=1}^N p_{*n}(X_n, V_n), \tag{1.2}$$

*Work supported by NSERC, Canada.

†University of British Columbia, Department of Mathematics, 121-1984 Mathematics Road, Vancouver, B. C., Canada V6T 1Z2.

‡Universität - Gesamthochschule Siegen, Fachbereich 6 Mathematik, Hölderlinstr. 3, 57068 Siegen, Germany.

[1] *AMS 1991 subject classifications.* Primary: 62A10, 62M99; secondary: 62B15, 62J02.

[2] *Key words and phrases.* Partial likelihood, partially specified model, survival analysis, semimartingale.

where $p_{*n}(x,v)$ is the conditional density of $V_n$ given the observations before time $n$ and $X_n = x$. The notation is chosen for consistency with later sections. It is natural to number the covariates $V_1, \dots, V_{n-1}$ so that they are in the past of $X_n$. The second factor may be completely unspecified, or it may depend on $\vartheta$ and on further parameters. If the second factor does not depend on $\vartheta$, the partial likelihood contains all the information about $\vartheta$. Note that $p_{*n}$ varies independently of $p_n^\vartheta$. Oakes (1981) reviews applications of the partial likelihood in survival analysis. Wong (1986) gives an optimality result for estimators based on the partial likelihood. Slud and Kedem (1994) and Møller and Sørensen (1994) analyse specific models. We emphasize that the partial likelihood (1.1) is different from Cox's (1972) partial likelihood for the proportional hazards model, which is a factor of the partial likelihood considered here. The construction of partial likelihoods is discussed extensively in Kalbfleisch and Prentice (1980) and Arjas (1989).

Does Cox's factorization have a version for continuous-time processes? Regression models involving such processes have become prominent in statistics. For example, survival regression models involving time-dependent covariates and censoring are conveniently described by counting processes. Numerous applications are discussed in the monographs by Fleming and Harrington (1991) and Andersen *et al.* (1993). In these applications, $X_t$ is a counting process with intensity $\lambda_t^\vartheta$ depending on the history of both $X$ and a vector $V$ of covariate processes. The first example of such a counting process regression model is Aalen's (1980) additive risk model, with intensity $\lambda_t = \alpha_t^T V_{t-}$. Another example is Andersen and Gill's (1982) version of Cox's (1972) proportional hazards model, with intensity $\lambda_t = C_{t-}\alpha_t \exp(\beta^T V_{t-})$, where $V_t$ together with the censoring process $C_t$ form the covariate process. Gill (1985) suggested using as a *partial likelihood*

$$\overline{Z}_t^\vartheta = \prod_{T_n \le t} \lambda_{T_n}^\vartheta \exp\left(-\int_0^t \lambda_s^\vartheta ds\right),$$

where $T_n$ are the successive jump times of $X$. This continuous-time partial likelihood, like the discrete-time partial likelihood above, has the form of the full likelihood of $X$ except for the dependence on $V$.

A partial likelihood for semimartingales $X$ was defined by Jacod (1987). Again it has the form of the full likelihood except for the dependence on $V$. For multivariate point processes and diffusion processes, Slud (1992) approximates the partial likelihood by discrete-time partial likelihoods.

When does the partial likelihood contain all the information about $\vartheta$? When and how can one use additional information about the model? For discrete-time processes, both questions are answered by Cox's factorization (1.2) of the likelihood. Gill (1985) has given a heuristic derivation of the factorization for multivariate point processes in terms of product integrals; see now Andersen *et al.* (1993, p. 107). We obtain such a factorization for continuous-time processes in full generality, with $X$ and $V$ semimartingales. Our explicit description of the second factor makes it possible to decide when optimal inference can be based on the partial likelihood. Having the second factor explicitly, we can now also give a nonasymptotic justification for the efficiency concept which we introduced earlier, Greenwood and Wefelmeyer (1990). A different, asymptotic, justification is in Greenwood and Wefelmeyer (1992).

## 2 Factoring the likelihood

This section is organized as follows. First we recall some notation from the general theory of processes, and give Jacod's construction of the partial likelihood for semimartingales and Jacod and Mémin's representation of the full likelihood in terms of the predictable characteristics. For a response process $X$ and a vector $V$ of covariate processes, we describe the consistency relations between the characteristics of $X$ and those of $(X,V)$. We find a 'parametrization' of the full model in terms of the characteristics of $X$ and of additional predictable processes not involving them. The Theorem expresses the second factor of the likelihood in terms of these additional processes.

We need the following notation from the theory of semimartingales. A more detailed description and associated results are given in the monograph of Jacod and Shiryaev (1987). Let $\mathcal{P}$ denote the predicable $\sigma$-field on $\Omega \times [0,\infty)$. If $X$ is a semimartingale and $b$ a predictable process, we write $b \cdot X_t$ for the stochastic integral $\int_0^t b_s dX_s$, and $X^c$ for the continuous martingale part of $X$. If $\nu$ is a random measure on $[0,\infty)\times\mathbf{R}^d$ and $Y$ a $\mathcal{P}\otimes\mathcal{B}^d$-measurable function, we write $Y * \nu_t$ for the stochastic integral $\int_0^t \int Y(s,x)\nu(ds,dx)$.

We observe a real-valued cadlag response process $X$ and a $d$-dimensional cadlag covariate process $V$ on a finite time interval in $[0,\infty)$. They generate the filtration $(\mathcal{F}_t)_{t\geq 0}$. The random jump measure of $X$ is

$$\mu^X(dt,dx) = \sum_{s:\ \Delta X_s \neq 0} \varepsilon_{(s,\Delta X_s)}(dt,dx).$$

Similarly, the random jump measure of $(X,V)$ is

$$\mu(dt,dx,dv) = \sum_{s:\ (\Delta X_s,\Delta V_s)\neq 0} \varepsilon_{(s,\Delta X_s,\Delta V_s)}(dt,dx,dv).$$

Likelihoods of continuous-time processes are conveniently written with respect to a base measure in the model. We fix $\vartheta$ and $\vartheta'$ and introduce the corresponding partial and full likelihoods (likelihood ratios). The labels $\vartheta$ and $\vartheta'$ are dropped. First we recall how the characteristics of $X$ change under an absolutely continuous change of measure. Consider two probability measures $P$ and $P'$ under which $X$ is a semimartingale, with characteristics $(\overline{B},\overline{C},\overline{\nu})$ and $(\overline{B}',\overline{C}',\overline{\nu}')$, respectively, with respect to a truncation function, say $\overline{h}$. Assume that $P'_t \ll P_t$ for $t \in [0,\infty)$, where $P_t = P \mid \mathcal{F}_t$. To keep the notation simpler, we assume that $P'_0 = P_0$. Write

$$\overline{a}_t = \overline{\nu}(\{t\}\times\mathbf{R}), \quad \overline{a}'_t = \overline{\nu}'(\{t\}\times\mathbf{R}).$$

Choose an increasing predictable process $\overline{F}$ such that $\overline{C} = \overline{c}\cdot\overline{F}$ with $\overline{c}$ a nonnegative predictable process. By a Girsanov theorem (Jacod and Shiryaev, 1987, p. 159, Theorem

3.24), there exist a $\mathcal{P} \times \mathbf{R}$-measurable function $\overline{Y}(t,x)$ and a predictable process $\overline{\beta}$ such that $P'$-a.s.,

$$\begin{aligned}\overline{B}' &= \overline{B} + \overline{c}\overline{\beta} \cdot \overline{F} + \overline{h}(x)(\overline{Y} - 1) * \overline{\nu},\\ \overline{C}' &= \overline{C},\\ \overline{\nu}'(dt,dx) &= \overline{Y}(t,x)\overline{\nu}(dt,dx).\end{aligned}$$

We recall Jacod's (1987, 1990b) construction of the partial likelihood. Introduce

$$\begin{aligned}\overline{\Sigma}^c &= \{t:\ \overline{c}\overline{\beta} \cdot \overline{F}_t < \infty\},\\ \overline{\Sigma}^d &= \{t:\ (1-\overline{Y}^{1/2})^2 * \overline{\nu}_t + \sum_{s\le t}\left((1-\overline{a}'_s)^{1/2} - (1-\overline{a}_s)^{1/2}\right)^2 < \infty\}.\end{aligned}$$

On $\overline{\Sigma} = \overline{\Sigma}^c \cap \overline{\Sigma}^d$, define the local $P$-martingale $\overline{N} = \overline{N}^c + \overline{N}^d$ with

$$\overline{N}^c = \overline{\beta} \cdot X^c, \tag{2.1}$$

$$\overline{N}^d = \left(\overline{Y} - 1 - \frac{\overline{a} - \overline{a}'}{1-\overline{a}} 1_{\{\overline{a}<1\}}\right) * (\mu^X - \overline{\nu}). \tag{2.2}$$

The *partial likelihood process* is defined on $\overline{\Sigma}$ as the Doléans exponential

$$\overline{Z}_t = \mathcal{E}(\overline{N})_t = \exp(\overline{N}_t - \tfrac{1}{2}\overline{c}\overline{\beta}^2 \cdot \overline{F}_t)\prod_{s\le t}(1+\Delta\overline{N}_s)e^{-\Delta\overline{N}_s}. \tag{2.3}$$

The partial likelihood is described through the pairs of 'parameters' $\overline{B}$, $\overline{\nu}$ and $\overline{B}'$, $\overline{\nu}'$. The change from $\overline{B}$ to $\overline{B}'$ and from $\overline{\nu}$ to $\overline{\nu}'$ is given through $\overline{\beta}$ and $\overline{Y}$. These would play the role of local parameters in asymptotic theory. We will describe the full model in such a way that the second factor of the likelihood does not involve the parameters $\overline{\beta}$, $\overline{Y}$ of the partial likelihood.

We assume that $P'_t$ is dominated by $P_t$ for $t \in [0,\infty)$, and have introduced the partial likelihood with reference to $P$ and $P'$. However, a partial likelihood can also be constructed when $P'_t$ is not dominated by $P_t$. Jacod (1990b) gives a definition which is free from $P$ and $P'$ except for the formal dependence of the stochastic integals on $P$.

In order to produce an explicit factoring, we assume that the likelihood of $(X,V)$ admits a representation in terms of its characteristics. This is not a serious restriction. Let $(X,V)$ be a semimartingale under $P$ and $P'$, with characteristics $(B,C,\nu)$ and $(B',C',\nu')$, respectively, with respect to a truncation function $h$. Write

$$a_t = \nu(\{t\} \times \mathbf{R} \times \mathbf{R}^d), \quad a'_t = \nu'(\{t\} \times \mathbf{R} \times \mathbf{R}^d).$$

As before, the Girsanov theorem allows us to write $C = c \cdot F$ with $c$ a nonnegative definite predictable $(d+1)\times(d+1)$ matrix, and we can write $P'$-a.s.,

$$\begin{aligned}B' &= B + c\beta \cdot F + h(x,v)(Y-1) * \nu,\\ C' &= C,\\ \nu'(dt,dx,dv) &= Y(t,x,v)\nu(dt,dx,dv).\end{aligned}$$

Define the random time intervals $\Sigma^c$, $\Sigma^d$ and $\Sigma$ in the same way as before, and on $\Sigma$ define the local $P$-martingale $N = N^c + N^d$ with

$$N^c = \beta \cdot (X^c, V^c), \tag{2.4}$$

$$N^d = \left( Y - 1 - \frac{a - a'}{1 - a} 1_{\{a<1\}} \right) * (\mu - \nu). \tag{2.5}$$

If all $P$-martingales have the representation property relative to $(X, V)$, the density process $Z_t = dP'_t/dP_t$ can be represented on $\Sigma$ as the Doléans exponential

$$Z_t = \mathcal{E}(N)_t = \exp(N_t - \tfrac{1}{2}\beta^T c\beta \cdot F_t) \prod_{s \le t} (1 + \Delta N_s) e^{-\Delta N_s}. \tag{2.6}$$

The result is due to Jacod and Mémin (1976) and Kabanov *et al.* (1979, 1980); see Jacod and Shiryaev (1987, p. 180, Theorem 5.19). One sees from the representation that the full likelihood is described through the pairs of parameters $B$, $\nu$ and $B'$, $\nu'$.

The characteristics of $(X, V)$ must be consistent with the characteristics of $X$. We may choose $\overline{F} = F$ and $\overline{h} = h_1$, the first component of the truncation function $h$. For the quadratic characteristics, we have

$$c'_{11} = c_{11} = \overline{c}. \tag{2.7}$$

The first component of the 'drift' characteristic of $(X, V)$ is the specified characteristic of $X$,

$$B_1 = \overline{B}, \quad B'_1 = \overline{B}'. \tag{2.8}$$

To describe consistency of $\nu$ with $\overline{\nu}$, we partition the state space as

$$\mathbf{R} \times \mathbf{R}^d = \left( (\mathbf{R} \setminus \{0\}) \times \mathbf{R}^d \right) + \left( \{0\} \times \mathbf{R}^d \right)$$

and write $\nu$ as its sum on these two sets:

$$\nu(dt, dx, dv) = \nu_{-0}(dt, dx, dv) + \nu_0(dt, dv)\varepsilon_0(dx).$$

For consistency, the marginal of $\nu_{-0}$ must be $\overline{\nu}$,

$$\nu_{-0}(dt, dx, dv) = \overline{\nu}(dt, dx)\nu_{-*}(t, x, dv), \tag{2.9}$$

and similarly for $\nu'$. Here $\nu_{-*}(t, x, dv)$ is the regular conditional jump size distribution of $V$ given that $X$ has a jump of size $x$ at time $t$ and varies independently of $\overline{\nu}$ just as $p_{*n}$ varies independently of $p_n^\vartheta$. In particular, $\nu_{-*}(t, x, \mathbf{R}^d) = 1$, and the total mass of $\mu$ at each fixed $t$ is $a(t) = a_0(t) + \overline{a}(t)$, where $a_0(t) = \nu_0(\{t\} \times \mathbf{R}^d)$.

We reparametrize the full model by the parameters $\overline{B}, \overline{\nu}$ of the partial specification and additional parameters which vary independently of them. Let $B_{(1)}$ denote the vector consisting of all components of $B$ except the first. Note that $B_{(1)}$ and $\nu_{-*}$ vary independently of $\overline{B}$ and $\overline{\nu}$. Define $\nu_*$ by

$$\nu_0(dt, dv) = (1 - \overline{a}_t)\nu_*(dt, dv), \tag{2.10}$$

and similarly for $\nu'$. In particular, $a_0 = (1 - \overline{a})a_*$. Note that $\nu_*$ equals $\nu_0$ unless both $X$ and $V$ have a positive probability of jumping at time $t$, $\overline{a}_t > 0$ and $a_{0t} > 0$. Then

$$\nu_*(\{t\}, dv) = \frac{\nu_0(\{t\}, dv)}{a_{0t}} \frac{a_{0t}}{1 - \overline{a}_t},$$

where $\nu_0(\{t\}, dv)/a_{0t}$ is the jump size distribution of $V_t 1_{\{\Delta X_t = 0\}}$, and $a_0/(1 - \overline{a}) = a_*$ is the conditional probability that $V$ jumps given $\Delta X = 0$. Because both these factors vary independently of $\overline{B}$ and $\overline{\nu}$, so does $\nu_*$. The model is now described by $\overline{B}$, $\overline{\nu}$, $B_{(1)}$, $\nu_{-*}$ and $\nu_*$.

The consistency relation (2.8) for $B$ translates into a consistency relation for the change from $B$ to $B'$,

$$\overline{c}\overline{\beta} = (c\beta)_1. \tag{2.11}$$

Since $\nu' \ll \nu$, also $\nu'_{-0} \ll \nu_{-0}$, $\nu'_0 \ll \nu_0$ and $\nu'_{-*} \ll \nu_{-*}$, $\nu'_* \ll \nu_*$. Write $Y_{-0}$, $Y_0$ and $Y_{-*}$, $Y_*$ for the corresponding relative densities. Then

$$Y(t, x, v) = Y_{-0}(t, x, v)1_{\{x \neq 0\}} + Y_0(t, v)1_{\{x = 0\}},$$

and the consistency relations (2.9) and (2.10) translate into

$$Y_{-0}(t, x, v) = \overline{Y}(t, x)Y_{-*}(t, x, v), \tag{2.12}$$

$$Y_0(t, v) = \frac{1 - \overline{a}'_t}{1 - \overline{a}_t} Y_*(t, v). \tag{2.13}$$

Partition $c$ and $\beta$ as

$$c = \begin{pmatrix} \overline{c} & c^T_{(1)1} \\ c_{(1)1} & c_{(11)} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_{(1)} \end{pmatrix}.$$

Here $\beta_{(1)}$ is the vector consisting of all components of $\beta$ except the first. Similarly, $c_{(11)}$ is obtained from $c$ by deleting the first row and column, and $c_{(1)1}$ from the first column by deleting its first element.

The factorization is expressed in terms of the processes

$$N^c_* = \beta_{(1)} \cdot (-\overline{c}^{-1} c_{(1)1} \cdot X^c + V^c), \tag{2.14}$$

$$N^d_{-*} = (Y_{-*} - 1)1_{\{x \neq 0\}} * \mu, \tag{2.15}$$

$$N^d_* = \left(Y_* - 1 - \frac{a_* - a'_*}{1 - a_*} 1_{\{a_* < 1\}}\right) 1_{\{x=0\}} * (\mu - \nu_0), \tag{2.16}$$

using the Doléans exponential (2.6) with $\Delta N$ defined by (2.23).

**Theorem** *On the random time interval* $\Sigma \cap \overline{\Sigma}$, *the likelihood process* $Z$ *factors as* $\overline{Z}Z_*$, *where* $\overline{Z}$ *is the partial likelihood, and*

$$Z_* = \mathcal{E}(N_*^c)\mathcal{E}(N_{-*}^d)\mathcal{E}(N_*^d).$$

*Here* $N_*^c$ *is a continuous local* $P$*-martingale,* $N_{-*}^d$ *and* $N_*^d$ *are purely discontinuous local* $P$*-martingales, and* $\mathcal{E}(N_*^d)$ *depends on* $\nu_0$ *only through* $\nu_*$.

A more explicit description of the factorization $Z = \overline{Z}Z_*$ is

$$\overline{Z} = \mathcal{E}(\overline{N}^c)\mathcal{E}(\overline{N}^d)$$

with factors given by (2.18) and the limit in $t$ of (2.26), and

$$Z_* = \mathcal{E}(N_*^c)\mathcal{E}(N_{-*}^d)\mathcal{E}(N_*^d)$$

with the first two factors given by (2.19) and (2.27), and the last by the limit of (2.28). Here we have

$$\mathcal{E}(N_{-*}^d + N_*^d) = \mathcal{E}(N_{-*}^d)\mathcal{E}(N_*^d)$$

because $N_{-*}^d$ and $N_*^d$ jump on different time sets, $\Delta X \neq 0$ and $\Delta X = 0$.

**Proof** By (2.6), the full likelihood on $\Sigma$ is

$$Z = \mathcal{E}(N) = \mathcal{E}(N^c)\mathcal{E}(N^d),$$

with $N^c$ and $N^d$ the continuous and purely discontinuous martingale parts of $N$, (2.4) and (2.5). We factor $\mathcal{E}(N^c)$ first.

The consistency conditions (2.7) and (2.11) give

$$\begin{aligned} \beta &= \begin{pmatrix} \overline{\beta} \\ 0 \end{pmatrix} + \begin{pmatrix} \beta_1 - \overline{\beta} \\ \beta_{(1)} \end{pmatrix} \\ &= \begin{pmatrix} \overline{\beta} \\ 0 \end{pmatrix} + \begin{pmatrix} -\overline{c}^{-1} c_{(1)1}^T \beta_{(1)} \\ \beta_{(1)} \end{pmatrix} \\ &= \begin{pmatrix} \overline{\beta} \\ 0 \end{pmatrix} + c^{-1} \begin{pmatrix} 0 \\ (c_{(11)} - \overline{c}^{-1} c_{(1)1} c_{(1)1}^T)\beta_{(1)} \end{pmatrix}, \end{aligned}$$

a sum of two vectors orthogonal with respect to the inner product $xcy$. This implies

$$\beta^T c \beta = \overline{c}\overline{\beta}^2 + \beta_{(1)}^T (c_{(11)} - \overline{c}^{-1} c_{(1)1} c_{(1)1}^T)\beta_{(1)}.$$

Hence $\overline{c}\overline{\beta}^2 < \infty$ and $\beta_{(1)}^T(c_{(11)} - \overline{c}^{-1}c_{(1)1}c_{(1)1}^T)\beta_{(1)} < \infty$ on $\Sigma^c$. Also

$$\beta \cdot (X^c, V^c) = \overline{\beta} \cdot X^c + \beta_{(1)} \cdot (-\overline{c}^{-1} c_{(1)1} \cdot X^c + V^c).$$

We obtain on $\Sigma^c$

$$\mathcal{E}(N^c) = \exp\left(\beta\cdot(X^c,V^c) - \tfrac{1}{2}\beta^T c\beta\cdot F\right) = \mathcal{E}(\overline{N}^c)\mathcal{E}(N^c_*) \tag{2.17}$$

with

$$\mathcal{E}(\overline{N}^c) = \exp(\overline{\beta}\cdot X^c - \tfrac{1}{2}\overline{c}\overline{\beta}^2\cdot F) \tag{2.18}$$

and

$$\mathcal{E}(N^c_*) = \exp\left(\beta_{(1)}\cdot(-\overline{c}^{-1}c_{(1)1}\cdot X^c + V^c) - \tfrac{1}{2}\beta^T_{(1)}(c_{(11)} - \overline{c}^{-1}c_{(1)1}c^T_{(1)1})\beta_{(1)}\cdot F\right). \tag{2.19}$$

Now we factor $\mathcal{E}(N^d)$. Assume first that the jumps of $X$ and $V$ are bounded away from 0 by $\varepsilon$. Then

$$\mathcal{E}(N^d)_t = \exp\left(-(Y-1)*\nu^c_t\right) \prod_{\substack{s\leq t\\ (\Delta X_s,\Delta V_s)\neq 0}} (1+\Delta N^d_s).$$

By the consistency relations (2.12) and (2.13),

$$Y(s,x,v) = \overline{Y}(s,x)Y_{-*}(s,x,v)1_{\{x\neq 0\}} + \frac{1-\overline{a}'_s}{1-\overline{a}_s}Y_*(s,v)1_{\{x=0\}}. \tag{2.20}$$

Since $\nu_{-*}(t,x,\mathbf{R}) = \nu'_{-*}(t,x,\mathbf{R}) = 1$,

$$\int (Y_{-*}(t,x,v)-1)\,\nu_{-*}(t,x,dv) = 0. \tag{2.21}$$

Furthermore, $1/(1-\overline{a}) = 1$ a.s. under $\nu^c_0$. Using (2.20), (2.21) and $\nu^c_{-0}(dt,dx,dv) = \nu_{-*}(t,x,dv)\overline{\nu}^c(dt,dx)$, we can compute

$$\begin{aligned}(Y-1)*\nu^c &= (\overline{Y}Y_{-*}-1)1_{\{x\neq 0\}}*\nu^c_{-0} + (Y_*-1)*\nu^c_0\\ &= (\overline{Y}-1)*\overline{\nu}^c + (Y_*-1)*\nu^c_*.\end{aligned} \tag{2.22}$$

As in Jacod and Shiryaev (1987, p. 180, 5.13),

$$\Delta N_s = (Y(s,\Delta X_s,\Delta V_s)-1)\,1_{\{(\Delta X_s,\Delta V_s)\neq 0\}} + \frac{a_s - a'_s}{1-a_s}1_{\{(\Delta X_s,\Delta V_s)=0\}}. \tag{2.23}$$

By the consistency relation (2.10), integrated on $v$,

$$1-a = 1-\overline{a}-a_0 = (1-\overline{a})\left(1-\frac{a_0}{1-\overline{a}}\right) = (1-\overline{a})(1-a_*), \tag{2.24}$$

and similarly for $a'$. With (2.20) and (2.22)–(2.24), we can write $\mathcal{E}(N^d)$ as

$$\begin{aligned}
\mathcal{E}(N^d)_t &= \exp\left(-(Y-1)*\nu^c_t\right) \prod_{\substack{s\le t\\(\Delta X_s,\Delta V_s)\neq 0}} Y(s,\Delta X_s,\Delta V_s) \prod_{\substack{s\le t\\(\Delta X_s,\Delta V_s)=0}} \frac{1-a'_s}{1-a_s} \\
&= \exp\Big(-(\overline{Y}-1)*\overline{\nu}^c_t - (Y_*-1)*\nu^c_{*t}\Big) \\
&\quad \prod_{\substack{s\le t\\ \Delta X_s\neq 0}} \overline{Y}(s,\Delta X_s) Y_{-*}(s,\Delta X_s,\Delta V_s) \\
&\quad \prod_{\substack{s\le t\\ \Delta V_s\neq 0,\ \Delta X_s=0}} \frac{1-\overline{a}'_s}{1-\overline{a}_s} Y_*(s,\Delta V_s) \prod_{\substack{s\le t\\(\Delta X_s,\Delta V_s)=0}} \frac{1-\overline{a}'_s}{1-\overline{a}_s}\frac{1-a'_{*s}}{1-a_{*s}}. \qquad (2.25)
\end{aligned}$$

From (2.23) with $N$ replaced by $\overline{N}$, we have

$$\mathcal{E}(\overline{N}^d)_t = \exp\Big(-(\overline{Y}-1)*\overline{\nu}^c_t\Big) \prod_{\substack{s\le t\\ \Delta X_s\neq 0}} \overline{Y}(s,\Delta X_s) \prod_{\substack{s\le t\\ \Delta X_s=0}} \frac{1-\overline{a}'_s}{1-\overline{a}_s}. \qquad (2.26)$$

By relation (2.21),

$$\begin{aligned}
N^d_{-*} &= (Y_{-*}-1)1_{\{x\neq 0\}}*\mu \\
&= (Y_{-*}-1)1_{\{x\neq 0\}}*(\mu-\nu)
\end{aligned}$$

is seen to be a purely discontinuous local $P$-martingale, and

$$\mathcal{E}(N^d_{-*})_t = \prod_{\substack{s\le t\\ \Delta X_s\neq 0}} Y_{-*}(s,\Delta X_s,\Delta V_s). \qquad (2.27)$$

Further, from (2.23) with $N$ replaced by $N_*$ we have

$$\mathcal{E}(N^d_*)_t = \exp\left(-(Y_*-1)*\nu^c_{*t}\right) \prod_{\substack{s\le t\\ \Delta V_s\neq 0,\ \Delta X_s=0}} Y_*(s,\Delta V_s) \prod_{\substack{s\le t\\ \Delta V_s=0,\ \Delta X_s=0}} \frac{1-a'_{*s}}{1-a_{*s}}. \qquad (2.28)$$

Applying (2.26)–(2.28) to (2.25), we obtain

$$\mathcal{E}(N^d) = \mathcal{E}(\overline{N}^d)\mathcal{E}(N^d_{-*})\mathcal{E}(N^d_*). \qquad (2.29)$$

Relations (2.17) and (2.29) give us the asserted factorization for $\varepsilon > 0$.

The assertion that the jumps of $X$ and $V$ are of size greater than $\varepsilon$ can be removed since the limit of each factor in (2.29) as $\varepsilon \to 0$ is the corresponding factor with $\varepsilon = 0$. In particular, the limit of $\mathcal{E}(N^d_*)$ in (2.28) depends on $\nu_0$ only through $\nu_*$.

# 3 Discussion

The Theorem of Section 2 allows us to evaluate the relative efficiency of statistical procedures based on certain factors of the likelihood. Here we discuss some aspects of the factorization.

*Efficiency of the partial likelihood.* The partial likelihood $\overline{Z}$ depends only on the partial specification $(\overline{B}, \overline{\nu})$. The second factor depends on $B_{(1)}$, $\nu_{-*}$ and $\nu_*$, and through the continuous martingale part $X^c$ in $\mathcal{E}(N_*^c)$ it depends also on $\overline{B}$, but not on $\overline{\beta}$. The information about parameters of the model does not depend on the parameters $(B, \nu)$ of the base measure $P$, but only on the parameters $(\beta, Y)$ of the likelihood $Z$. Therefore, if we have a parametric model for $(\overline{B}, \overline{\nu})$, the partial likelihood leads to efficient inference as long as $\beta_{(1)}$, $Y_{-*}$ and $Y_*$ do not involve the parameter of interest.

*The partial likelihood is a likelihood.* If $B_{(1)}$, $\nu_{-*}$ and $\nu_*$ do not depend on the parameter of interest,

$$B'_{(1)} = B_{(1)}, \quad \nu'_{-*} = \nu_{-*}, \quad \nu'_* = \nu_*, \tag{3.1}$$

then $\beta_2 = 0$, and the $\nu$-density of $\nu'$ is

$$Y(s, x, v) = \overline{Y}(s, x) 1_{\{x \neq 0\}} + \frac{1 - \overline{a}'_s}{1 - \overline{a}_s} 1_{\{x=0\}}.$$

Hence the representation (2.6) of the full likelihood $Z$ reduces to the partial likelihood $\overline{Z}$. For multivariate point processes, this observation is due to Arjas and Haara (1984). Their condition A is contained in (3.1). They use Jacod's (1975) representation of the full likelihood. The reduction of $Z$ to $\overline{Z}$ under (3.1) also follows from our factorization.

*The partial likelihood is a projection.* Suppose we are given a parametric model for the partial specification $\overline{B}$, $\overline{\nu}$, but leave the model completely unspecified otherwise. If the full likelihood admits a representation (2.6), our factorization shows that the partial likelihood leads to efficient inference about the parameter. If we do not have the representation (2.6), it is still possible to prove this. Following Greenwood and Wefelmeyer (1990), introduce $\overline{P}$ by $d\overline{P} = \overline{Z} dP$, and call an estimator efficient if it is efficient in this model. Greenwood and Wefelmeyer (1992) show that this is equivalent to efficiency in the full model: $\overline{P}$ is in the full model because $X$ has characteristics $\overline{B}, \overline{\nu}$ under $\overline{P}$ by the Girsanov theorem, and $\overline{P}$ is least favorable since $\overline{Z}$ is the projection of any full likelihood $Z_t = dP'_t / dP_t$ such that $X$ has characteristics $\overline{B}', \overline{\nu}'$ under $P'$, by the converse of the Girsanov theorem (Jacod and Shiryaev, 1987, p. 160, 3.28). The latter argument is already used by Jacod (1990a, see also 1990b) to prove that the partial Fisher information is smaller than the full Fisher information.

*Factoring the partial likelihood.* The partial likelihood factors as $\mathcal{E}(\overline{N}^c)\mathcal{E}(\overline{N}^d)$, with factors defined in (2.18) and (2.26). Suppose we model $X$ as $X_c + X_d$, where $X_c$ is a continuous semimartingale with characteristics $(\overline{B}_c, \overline{C}, 0)$, and $X_d$ is a pure jump process with jump characteristic $\overline{\nu}$. Then $X^c = X_c^c = X_c - \overline{B}_c$. Furthermore, the Girsanov theorem says that

$$\overline{B}_c' = \overline{B}_c + \overline{c}\overline{\beta}_c \cdot \overline{F}.$$

Hence

$$\mathcal{E}(\overline{N}^c) = \exp\left(\overline{\beta}_c \cdot (X_c - \overline{B}_c) - \tfrac{1}{2}\overline{c}\overline{\beta}_c^2 \cdot \overline{F}\right)$$

depends only on $\overline{B}_c$, and $\mathcal{E}(\overline{N}^d)$ depends only on $\overline{\nu}$. We arrive at a factorization of $\overline{Z}$ with factors depending on independently varying parameters $\overline{B}_c$ and $\overline{\nu}$.

*Fixed fixed jumps.* In most applications, the probabilities $\overline{a}_t$ and $a_{*t}$ of jumping at a fixed time $t$ do not depend on the parameter of interest, $\overline{a}_t' = \overline{a}_t$ and $a_{*t}' = a_{*t}$. For example, they are simultaneously either 1 or 0, or all 0. In this case, $Y_* = Y_0$, and (2.26) and (2.28) reduce to

$$\begin{aligned}
\mathcal{E}(\overline{N}^d)_t &= \exp\left(-(\overline{Y}-1) * \overline{\nu}_t^c\right) \prod_{\substack{s \le t \\ \Delta X_s \ne 0}} \overline{Y}(s, \Delta X_s), \\
\mathcal{E}(N_*^d)_t &= \exp\left(-(Y_0 - 1) * \nu_{*t}^c\right) \prod_{\substack{s \le t \\ \Delta V_s \ne 0,\ \Delta X_s = 0}} Y_0(s, \Delta V_s).
\end{aligned}$$

# 4 Examples

The factorization simplifies for more specific processes. We give some examples.

*Discrete-time processes.* For discrete-time processes, we have already described Cox's factorization in (1.2). It is instructive to derive a comparable factorization from our Theorem, now for likelihood *ratios* rather than likelihoods. The observations will be jumps of a semimartingale, and we will use small letters for them. Let $x_0, x_1, \ldots$ be real-valued responses, and $v_0, v_1, \ldots$ $d$-dimensional covariates. They generate a filtration $(\mathcal{F}_n)_{n \ge 0}$. As in Section 2 we assume, for simplicity, that the distribution of $(x_0, v_0)$ is known. For $n = 1, 2, \ldots$, we specify a model for the regular conditional distributions $\overline{p}_n(dx)$ of $x_n$ given $\mathcal{F}_{n-1}$. The continuous-time process $X_t = \sum_{n \le t} x_n$, $t \ge 0$, is a (special) semimartingale with characteristics $\overline{B} = 0$, $\overline{C} = 0$, and

$$\overline{\nu}(dt, dx) = \sum_{n \ge 1} \overline{\nu}_n(dx)\varepsilon_n(dt),$$

where $\overline{\nu}_n$ is defined by

$$\overline{p}_n(dx) = \overline{\nu}_n(dx) + \overline{p}_n(\{0\})\varepsilon_0(dx).$$

The distinction between $x_n = 0$ and $x_n \ne 0$ is necessary because the jump measure of $X$ does not charge $x_n = 0$. This will lead to a more complicated factorization than (1.2), namely $f_{*n}$ has the form (4.9). Note that

$$\overline{p}_n(\{0\}) = 1 - \overline{\nu}_n(\mathbf{R}) = 1 - \overline{a}_n. \tag{4.1}$$

To introduce the partial likelihood process, let $\overline{p}'_n$ be another regular conditional distribution in the model, with $\overline{p}'_n \ll \overline{p}_n$, and write $\overline{Y}_n$ for the $\overline{\nu}_n$-density of $\overline{\nu}'_n$. As in (2.26), the partial likelihood process (2.3) can be written

$$\overline{Z}_t = \mathcal{E}(\overline{N}^d)_t = \prod_{\substack{n\le t\\ x_n\ne 0}} \overline{Y}_n(x_n) \prod_{\substack{n\le t\\ x_n=0}} \frac{1-\overline{a}'_n}{1-\overline{a}_n}. \tag{4.2}$$

Now let $p_n(dx, dv)$ be a regular conditional distribution of $(x_n, v_n)$ given $\mathcal{F}_{n-1}$. Define $V_t = \sum_{n\le t} v_n$. Then $(X, V)$ is a (special) semimartingale with characteristics $B = 0$, $C = 0$, and

$$\nu(dt, dx, dv) = \sum_{n\ge 1} \nu_n(dx, dv)\varepsilon_n(dt),$$

where $\nu_n$ is defined by

$$p_n(dx, dv) = \nu_n(dx, dv) + p_n(\{(0,0)\})\varepsilon_{(0,0)}(dx, dv).$$

As in (4.1),

$$p_n(\{(0,0)\}) = 1 - \nu_n(\mathbf{R}\times\mathbf{R}^d) = 1 - a_n. \tag{4.3}$$

To introduce the full likelihood process, let $p'_n$ be another regular conditional distribution of $(x_n, v_n)$ given $\mathcal{F}_{n-1}$, and assume $p'_n \ll p_n$. With notation analogous to the above, the density process (2.6) can be written

$$Z_t = \prod_{\substack{n\le t\\ (x_n,v_n)\ne 0}} Y_n(x_n, v_n) \prod_{\substack{n\le t\\ (x_n,v_n)=0}} \frac{1-a'_n}{1-a_n}. \tag{4.4}$$

To describe the factorization, write

$$\nu_n(dx, dv) = \nu_{-0,n}(dx, dv) + \nu_{0n}(dv)\varepsilon_0(dx).$$

The consistency relations (2.9) and (2.10) are then written

$$\begin{aligned}\nu_{-0,n}(dx, dv) &= \overline{\nu}_n(dx)\nu_{-*,n}(x, dv),\\ \nu_{0n}(dv) &= (1-\overline{a}_n)\nu_{*n}(dv),\end{aligned}$$

where $\nu_{-*,n}(x, dv)$ is the conditional jump size distribution of $v_n$ given $x_n = x$. In particular (compare (2.24)),

$$1 - a_n = 1 - \overline{a}_n - a_{0n} = (1-\overline{a}_n)(1-a_{*n}). \tag{4.5}$$

Then the consistency relations (2.12) and (2.13) for the relative densities are

$$Y_{-0,n}(x, v) = \overline{Y}_n(x)Y_{-*,n}(x, v), \tag{4.6}$$

$$Y_{0n}(v) = \frac{1-\overline{a}'_n}{1-\overline{a}_n}Y_{*n}(v). \tag{4.7}$$

The factorization $Z = \overline{Z} Z_*$ can now be obtained from the Theorem and (2.26) to (2.28) or directly from (4.2) and (4.4) and the consistency relations (4.5) to (4.7), with

$$Z_{*t} = \prod_{\substack{n \le t \\ x_n \ne 0}} Y_{-*,n}(x_n, v_n) \prod_{\substack{n \le t \\ v_n \ne 0,\ x_n = 0}} Y_{*n}(v_n) \prod_{\substack{n \le t \\ v_n = 0,\ x_n = 0}} \frac{1 - a'_{*n}}{1 - a_{*n}}. \tag{4.8}$$

To compare with Cox's discrete-time factorization (1.2), write $\overline{f}_n$ for the $\overline{p}_n$-density of $\overline{p}'_n$. Then $\overline{f}_n$ equals $\overline{Y}_n$ on $\mathbf{R} \setminus \{0\}$, and with (4.1),

$$\overline{f}_n(0) = \frac{\overline{p}'_n(\{0\})}{\overline{p}_n(\{0\})} = \frac{1 - \overline{a}'_n}{1 - \overline{a}_n}.$$

Hence the partial likelihood process (4.2) is $\overline{Z}_t = \prod_{n \le t} \overline{f}_n(x_n)$. This is the partial likelihood ratio obtained from the usual discrete-time partial likelihood. Now factor

$$p_n(dx, dv) = \overline{p}_n(dx) p_{*n}(x, dv).$$

We have $p_{*n}(x, dv) = \nu_{-*,n}(x, dv)$ for $x \ne 0$, and $p_{*n}(0, dv) = \nu_{*n}(dv)$ on $\mathbf{R}^d \setminus \{0\}$. Further, $p_{*n}(0, \{0\}) = 1 - a_{*n}$. Write $f_{*n}(x, v)$ for the $p_{*n}(x, dv)$-density of $p'_{*n}(x, dv)$. By (1.2), the factor (4.8) must equal $\prod_{n \le t} f_{*n}$. Indeed,

$$f_{*n}(x, v) = \begin{cases} Y_{-*,n}(x, v), & x \ne 0, \\ Y_{*n}(v), & x = 0,\ v \ne 0, \\ \frac{1 - a'_{*n}}{1 - a_{*n}}, & x = 0,\ v = 0. \end{cases} \tag{4.9}$$

*Continuous processes.* Suppose $(X, V)$ is a continuous semimartingale, with characteristics $(B, C, 0)$. Then the factorization reduces to $Z = \overline{Z} Z_* = \mathcal{E}(\overline{N}^c)\mathcal{E}(N_*^c)$ with factors defined in (2.18) and (2.19). In this case, $X^c = X - \overline{B}$ and $V^c = V - B_{(1)}$ with

$$B = \begin{pmatrix} \overline{B} \\ B_{(1)} \end{pmatrix} = cb \cdot F, \quad b = \begin{pmatrix} b_1 \\ b_{(1)} \end{pmatrix}.$$

In particular,

$$\overline{Z} = \exp\left(\overline{\beta} \cdot (X - \overline{B}) - \tfrac{1}{2}\overline{c}\overline{\beta}^2 \cdot F\right).$$

The compensator of $\beta_{(1)} \cdot (-\overline{c}^{-1} c_{(1)1} \cdot X + V)$ is

$$(-\overline{c}^{-1} c_{(1)1}^T \beta_{(1)},\ \beta_{(1)}^T) cb \cdot F = \beta_{(1)}^T (c_{(11)} - \overline{c}^{-1} c_{(1)1} c_{(1)1}^T) b_{(1)} \cdot F.$$

Hence

$$\begin{aligned} Z_* = \exp\Big(&\beta_{(1)} \cdot (-\overline{c}^{-1} c_{(1)1} \cdot X + V) \\ &- \beta_{(1)}^T (c_{(11)} - \overline{c}^{-1} c_{(1)1} c_{(1)1}^T) b_{(1)} \cdot F \\ &- \tfrac{1}{2} \beta_{(1)}^T (c_{(11)} - \overline{c}^{-1} c_{(1)1} c_{(1)1}^T) \beta_{(1)} \cdot F\Big), \end{aligned}$$

which does not depend on $\overline{\beta}$.

*Jump processes.* Let $X$ and $V$ be pure jump processes, and let $\overline{\nu}$ and $\nu$ be the compensators of the random jump measures of $X$ and $(X, V)$, respectively. Then the factorization reduces to $Z = \overline{Z} Z_*$ with

$$\overline{Z}_t = \exp\left(-(\overline{Y} - 1) * \overline{\nu}^c_t\right) \prod_{\substack{s \leq t \\ \Delta X_s \neq 0}} \overline{Y}(s, \Delta X_s) \prod_{\substack{s \leq t \\ \Delta X_s = 0}} \frac{1 - \overline{a}'_s}{1 - \overline{a}_s}$$

and

$$\begin{aligned} Z_{*t} = & \prod_{\substack{s \leq t \\ \Delta X_s \neq 0}} Y_{-*}(s, \Delta X_s, \Delta V_s) \\ & \exp\left(-(Y_* - 1) * \nu^c_{*t}\right) \prod_{\substack{s \leq t \\ \Delta V_s \neq 0,\ \Delta X_s = 0}} Y_*(s, \Delta V_s) \prod_{\substack{s \leq t \\ \Delta V_s = 0,\ \Delta X_s = 0}} \frac{1 - a'_{*s}}{1 - a_{*s}}. \end{aligned}$$

An analogous factorization holds for multivariate point processes with more general state spaces; the partial likelihood $\overline{Z}$ is described by Arjas and Haara (1984).

*Diffusions with jumps.* Let $(X, V)$ be a diffusion with jumps, with characteristics

$$\begin{aligned} dB_t &= b_t(X_{t-}, V_{t-})dt, \\ dC_t &= c_t(X_{t-}, V_{t-})dt, \\ \nu(dt, dx, dv) &= dt K_t(X_{t-}, V_{t-}, dx, dv). \end{aligned}$$

Specify a model for the characteristics of $X$,

$$\begin{aligned} d\overline{B}_t &= \overline{b}_t(X_{t-}, V_{t-})dt, \\ d\overline{C}_t &= \overline{c}_t(X_{t-}, V_{t-})dt, \\ \overline{\nu}(dt, dx) &= dt \overline{K}_t(X_{t-}, V_{t-}, dx). \end{aligned}$$

Write

$$K_t(X_{t-}, V_{t-}, dx, dv) = K_{-0,t}(X_{t-}, V_{t-}, dx, dv) + K_{0t}(X_{t-}, V_{t-}, dv)\varepsilon_0(dx).$$

Consistency of $\nu$ with $\overline{\nu}$ gives

$$K_{-0,t}(X_{t-}, V_{t-}, dx, dv) = \overline{K}_t(X_{t-}, V_{t-}, dx) K_{-*,t}(X_{t-}, V_{t-}, x, dv).$$

Let $P'$ be another distribution with $P'_t \ll P_t$ for $t \in [0, \infty)$, so that $C' = C$. For simplicity, assume again $P'_0 = P_0$. Write

$$\begin{aligned} \beta_t &= c_t(X_{t-}, V_{t-})^{-1}(b'_t(X_{t-}, V_{t-}) - b_t(X_{t-}, V_{t-})), \\ \overline{\beta}_t &= \overline{c}_t(X_{t-}, V_{t-})^{-1}(\overline{b}'_t(X_{t-}, V_{t-}) - \overline{b}_t(X_{t-}, V_{t-})). \end{aligned}$$

As in Section 2, the consistency relation for $\beta$ and $\overline{\beta}$ is $\overline{c}_t(X_{t-}, V_{t-})\overline{\beta}_t = (c_t(X_{t-}, V_{t-})\beta_t)_1$. Write $Y_t(X_{t-}, V_{t-}, x, v)$ for the $K_t(X_{t-}, V_{t-}, dx, dv)$-density of $K'_t(X_{t-}, V_{t-}, dx, dv)$, and define $Y_0$, $\overline{Y}$ and $Y_{-*}$ correspondingly. Since $(X, V)$ does not have fixed jumps, we have $a = 0$ and hence $\overline{a} = a_* = 0$ and $Y_*(x, v) = Y_0(X_{t-}, V_{t-}, x, v)$. We suppress $(X_{s-}, V_{s-})$. The partial likelihood process is

$$\begin{aligned}\overline{Z}_t &= \exp\Big(\int_0^t \overline{\beta}_s dX_s^c - \tfrac{1}{2}\overline{c}_s\overline{\beta}_s^2 ds\Big) \\ &\quad \exp\Big(-\int_0^t \int \big(\overline{Y}_s(x) - 1\big)\overline{K}_s(dx)ds\Big) \prod_{\substack{s \le t \\ \Delta X_s \neq 0}} \overline{Y}_s(\Delta X_s).\end{aligned}$$

The likelihood factors, $Z = \overline{Z}Z_*$, where $Z_*$ is obtained from (2.19), (2.27) and (2.28) as

$$\begin{aligned}Z_{*t} &= \exp\Big(\int_0^t \beta_{(1)s}\Big(-\overline{c}_s^{-1}c_{(1)1s}dX_s^c + dV_s^c\Big) \\ &\qquad - \tfrac{1}{2}\int_0^t \beta_{(1)s}^T\Big(c_{(11)s} - \overline{c}_s^{-1}c_{(1)1s}c_{(1)1s}^T\Big)\beta_{(1)s}ds\Big) \\ &\quad \prod_{\substack{s \le t \\ \Delta X_s \neq 0}} Y_{-*,s}(\Delta X_s, \Delta V_s) \\ &\quad \exp\Big(-\int_0^t \int \Big(Y_{0s}(v) - 1\Big)K_0(dv)ds\Big) \prod_{\substack{s \le t \\ \Delta V_s \neq 0,\ \Delta X_s = 0}} Y_{0s}(\Delta V_s).\end{aligned}$$

Here $\overline{Z}$ depends on $\overline{b}$ through $\overline{\beta}$, and on $\overline{K}$ through $\overline{Y}$. The second factor $Z_*$ depends on $b$ through $\beta_{(1)}$, and on $K$ through $Y_{-*}$, $K_0$ and $Y_0$. As in the continuous processes case, we argue that the first term of $Z_*$ does not depend on $\overline{b}$ even though it contains $X^c$.

# References


Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. In: *Mathematical Statistics and Probability Theory* (W. Klonecki, A. Kozek and J. Rosiński, eds.), 1–25, Lecture Notes in Statistics 2, Springer-Verlag, New York.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

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

Arjas, E. (1989). Survival models and martingale dynamics. *Scand. J. Statist.* 16, 177–225.

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

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

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

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

Gill, R. D. (1985). Notes on product integration, likelihood and partial likelihood for counting processes, non-informative and independent censoring. See: Andersen *et al.* (1993, Chapter 2).

Greenwood, P. E. and Wefelmeyer, W. (1990). Efficiency of estimators for partially specified filtered models. *Stochastic Process. Appl.* 36, 353–370.

Greenwood, P. E. and Wefelmeyer, W. (1992). Partially specified filtered models and efficiency. *Theory Probab. Appl.* 37, 139–142.

Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 31, 235–253.

Jacod, J. (1987). Partial likelihood process and asymptotic normality. *Stochastic Process. Appl.* 26, 47–71.

Jacod, J. (1990a). Regularity, partial regularity, partial information process, for a filtered statistical model. *Probab. Theory Rel. Fields* 86, 305–335.

Jacod, J. (1990b). Sur le processus de vraisemblance partielle. *Ann. Inst. Henri Poincaré* 26, 299–329.

Jacod, J. and Mémin, J. (1976). Caractéristiques locales et conditions de continuité absolue pour les semimartingales. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 35, 1–37.

Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes.* Grundlehren der mathematischen Wissenschaften 288, Springer-Verlag, Berlin.

Kabanov, Yu., Liptser, R. S. and Shiryaev, A. N. (1979). Absolute continuity and singularity of locally absolutely continuous distributions. I. *Math. USSR Sbornik* 35, 631–680.

Kabanov, Yu., Liptser, R. S. and Shiryaev, A. N. (1980). Absolute continuity and singularity of locally absolutely continuous distributions. II. *Math. USSR Sbornik* 36, 31–58.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

Møller, J. and Sørensen, M. (1994). Statistical analysis of a spatial birth-and-death process model with a view to modelling linear dune fields. *Scand. J. Statist.* 21, 1–19.

Oakes, D. (1981). Survival times: aspects of partial likelihood (with discussion). *Internat. Statist. Rev.* 49, 235–264.

Slud, E. V. (1992). Partial likelihood for continuous-time stochastic processes. *Scand. J. Statist.* 19, 97–109.

Slud, E. V. and Kedem, B. (1994). Partial likelihood analysis of logistic regression and autoregression. *Statist. Sinica* 4, 89–106.

Wong, W. H. (1986). Theory of partial likelihood. *Ann. Statist.* 14, 88–123.