Luhm, Pruscha:

Semi-parametric Inference for Regression Models Based on Marked Point Processes

Sonderforschungsbereich 386, Paper 78 (1997)

Semi-parametric Inference for Regression Models Based on Marked Point Processes

by A. Luhm* and H. Pruscha*
University of Munich

SUMMARY. We study marked point processes (MPP's) with an arbitrary mark space. First we develop some statistically relevant topics in the theory of MPP's admitting an intensity kernel $\lambda_t(dz)$, namely martingale results, central limit theorems for both the number n of objects under observation and the time t tending to infinity, the decomposition into a local characteristic $(\lambda_t, \Phi_t(dz))$ and a likelihood approach. Then we present semi-parametric statistical inference in a class of Aalen (1975)-type multiplicative regression models for MPP's as $n \to \infty$, using partial likelihood methods. Furthermore, considering the case $t \to \infty$, we study purely parametric M-estimators.

KEYWORDS: Marked point process, intensity kernel, (locally square integrable) martingale, local characteristic, partial likelihood, M-estimator.

1 Introduction and Basic Definitions

The monography by Andersen et al. (1993) presents a kind of canonical approach to the statistical analysis of point process models. It deals with multivariate point processes where each random event carries information on the occurrence time and the type of event, the latter being from a finite set E of alternatives. The theoretical fundament to multivariate point processes was laid - among others - by Jacod (1975), Bremaud (1981) and Dellacherie & Meyer (1982). There are applications, however, where an uncountable set E (e.g., E the set of real numbers) of alternatives - now called marks - is more appropriate, see Scheike (1994a,b), Murphy (1995) and Pruscha (1997). A mathematical foundation of marked point processes (MPP's) is given by Last and Brandt (1995), but this work does not contain all tools necessary for statistical analysis.
The first goal of the present paper is to fill this gap. We present (i) results on MPP-integrals, (ii) likelihood functions of an MPP observation, (iii) central limit laws for two different situations denoted as I and II below.
These tools are then used for the asymptotic statistical inference in the case of two different kinds of data schemes. Scheme I contains n realisations over a fixed time interval [0,T], with n tending to infinity. Here the semi-parametric analysis of a wide class of Aalen (1975)-type models can be presented. Scheme II has one single realisation over a longer time interval [0,T], with T tending to infinity. Here we present a purely parametric analysis only, the work on the nonparametric part of the problem is still in progress.
The proofs are sketched only, for complete versions we refer to a forthcoming paper by Luhm.

*Mathemat. Inst. der Ludwig-Maximilians-Universität München, Theresienstr.39, D-80333 München

Unless mentioned otherwise, we suppose all random elements and thus all processes to be defined on some probability space $(\Omega, \mathcal{F}, I\!\!P)$ being equipped with a complete, right continuous filtration F. Following Brémaud (1981), we fix an arbitrary measurable space $(E, \mathcal{E})$ and define an **MPP** to be a double sequence $(\tau_k, \zeta_k)_{k \in I\!\!N}$ such that

(i) $(\tau_k)_{k \in I\!\!N_0}$ is a point process (with $\tau_0 = 0$),
(ii) $(\zeta_k)_{k \in I\!\!N}$ is a sequence of random elements in E.

The double sequence $(\tau_k, \zeta_k)_{k \in I\!\!N}$ shall be identified with its associated counting measure $N(dt \times dz)$ which is defined by

$$N(]0,t] \times A) := N_t(A) = \sum_{k=1}^{\infty} 1(\tau_k \leq t) 1(\zeta_k \in A), \quad t \geq 0, \quad A \in \mathcal{E}.$$

The filtration $F^N = (\mathcal{F}_t^N)_{t \geq 0}$ that consists of the sigma algebras

$$\mathcal{F}_t^N := \sigma(N_s(A) : \ 0 \leq s \leq t, \ A \in \mathcal{E}), \quad t \geq 0,$$

is called **internal history** of the MPP. Now let $\mathcal{P}(F)$ denote the F-predictable σ-algebra on $\Omega \times]0, \infty[$ and $\tilde{\mathcal{P}}(F) := \mathcal{P}(F) \otimes \mathcal{E}$ the F-predictable σ-algebra. Each mapping $H : \Omega \times [0, \infty[\times E \to I\!\!R$ such that

(i) $H|_{\Omega \times \{0\} \times E}$ is $\mathcal{F}_0 \otimes \mathcal{E}, \mathcal{B}$-measurable,

(ii) $H|_{\Omega \times]0,\infty[\times E}$ is $\tilde{\mathcal{P}}(F), \mathcal{B}$-measurable,

we call **F-predictable E-indexed process** or shortly FE-P.
Finally, we define an **intensity kernel** $\lambda_t(dz)$ to be a transition measure from $(\Omega \times [0, \infty[\ , \ \mathcal{F} \otimes \mathcal{B}_+)$ into $(E, \mathcal{E})$ such that $\forall\ A \in \mathcal{E}$ the point process $(N_t(A))_{t \geq 0}$ admits the F-predictable intensity $(\lambda_t(A))_{t \geq 0}$. We put $\forall\ A \in \mathcal{E}$

$$\Lambda_t(A) := \int_0^t \lambda_s(A)\ ds = \int_0^t \int_A \lambda_s(dz)\ ds, \quad t \geq 0,$$

setting especially $\lambda_t := \lambda_t(E)$ and $\Lambda_t := \Lambda_t(E)$, $t \geq 0$. Furthermore, we suppose that

$$\Lambda_t < \infty \quad I\!\!P - a.s. \quad \forall\ t \geq 0.$$

Defining

$$\Phi_t(dz) := \frac{\lambda_t(dz)}{\lambda_t}, \quad t \geq 0,$$

we obtain a probability measure on the mark space E given the history until the occurrence time t (cf. Jacod (1975), Brémaud (1981) and Last & Brandt (1995)). The pair $(\lambda_t, \Phi_t(dz))$ is then called **local characteristic**.

Throughout the paper, we denote by $|\underline{a}|$ and $|\underline{A}|$ the Euklidian norm of a vector $\underline{a}$ and a matrix $\underline{A}$, respectively; $\underline{a}^{②}$ is the matrix with (i,j)-entry $a_i a_j$. For a countable set I, we write $|I|$ for the number of elements of I, while I_d stands for the d-dimensional unit matrix, $d \in I\!\!N$.

2 Results on Marked Point Processes

2.1 Martingale Theory

As a continuation of Brémaud (1981), we consider d-dimensional stochastic processes ($d \in \mathbb{N}$) which are generated by the integration of a d-dimensional FE-P $\underline{H}(t,z)$ w.r.t. the measure $M(dt \times dz) := N(dt \times dz) - \lambda_t(dz)dt$; i.e., we deal with terms of the form

$$\underline{M}_t \;=\; \int_0^t \int_E \underline{H}(s,z)\; M(ds \times dz), \quad t \geq 0.$$

Obviously, all these processes satisfy $\underline{M}_0 = 0$.

Theorem 1. *The following implications hold:*

(a) (1) $\Rightarrow$ $\underline{M}$ *is a local martingale,*

(b) (2) $\Rightarrow$ $\underline{M}$ *is a martingale,*

(c) (3) $\Rightarrow$ $\underline{M}$ *is a uniformly integrable martingale,*

where

$$\int_0^t \int_E |H_i(s,z)|\; \lambda_s(dz)\; ds < \infty \quad \mathbb{P}-a.s. \quad \forall\, 1 \leq i \leq d \;\; \forall\, t \geq 0, \tag{1}$$

$$\mathbb{E}\left(\int_0^t \int_E |H_i(s,z)|\; \lambda_s(dz)\; ds\right) < \infty \quad \forall\, 1 \leq i \leq d \;\; \forall\, t \geq 0, \tag{2}$$

$$\mathbb{E}\left(\int_0^\infty \int_E |H_i(s,z)|\; \lambda_s(dz)\; ds\right) < \infty \quad \forall\, 1 \leq i \leq d. \tag{3}$$

Proof: As to (a),(b), see Brémaud (1981, VIII Corollary 4). Part (c) can be shown by applying (b), Brémaud (1981, VIII Theorem 3) and Kopp (1984, Theorem 3.3.8). □

Observe that all processes we deal with are of bounded variation on finite intervals. For square integrable martingales we obtain a result similar to Theorem 1, continuing Brémaud (1972) and Boel et al. (1975).

Theorem 2. **(a)** $\underline{M}$ *is a square integrable martingale* $\Longleftrightarrow$

$$\mathbb{E}\left(\int_0^\infty \int_E H_i^2(s,z)\; \lambda_s(dz)\; ds\right) < \infty \quad \forall\, 1 \leq i \leq d. \tag{4}$$

(b) $\underline{M}$ *is a martingale which is locally square integrable* $\Longleftarrow$

$$\mathbb{E}\left(\int_0^t \int_E H_i^2(s,z)\; \lambda_s(dz)\; ds\right) < \infty \quad \forall\, 1 \leq i \leq d \;\; \forall\, t \geq 0. \tag{5}$$

(c) $\underline{M}$ *is a locally square integrable martingale* $\Longleftrightarrow$

$$\int_0^t \int_E H_i^2(s,z)\; \lambda_s(dz)\; ds < \infty \quad \mathbb{P}-a.s. \quad \forall\, 1 \leq i \leq d \;\; \forall\, t \geq 0. \tag{6}$$

Proof: Part (a) can be obtained using the optional covariation process $[\underline{M}]_t = \int_0^t \int_E \underline{H}^{②}(s,z)\; N(ds \times dz)$, $t \geq 0$, which is derived according to Dellacherie & Meyer (1982, VII Theorem 36), while (c) follows from (a) generalizing the lines of Liptser & Shiryayev

(1978, Theorem 18.8). Part (b) is a direct consequence of (c) and Theorem 1(b) and was first formulated by Scheike (1994a). □

Notice that a locally square integrable martingale need not be a martingale as such. To close this section, we give an explicit formula for the characteristic $\langle \underline{M} \rangle$ of $\underline{M}$.

Theorem 3. *Let* (6) *hold. Then the characteristic of* $\underline{M}$ *is given by*

$$\langle \underline{M} \rangle_t = \int_0^t \int_E \underline{H}^{②}(s,z)\lambda_s(dz)\; ds, \quad t \geq 0. \tag{7}$$

Proof: Apply Theorem 1 (a) to the optional variation process $[\underline{M}]$ and use the uniqueness of the compensator. □

2.2 Central Limit Theorems

First we formulate a Rebolledo (1980)-type central limit theorem for the number n of objects under observation tending to infinity.

Theorem 4. *For each* $n \in I\!N$, *let* $(E^{(n)}, \mathcal{E}^{(n)})$ *be a measurable space and* $N^{(n)}(dt \times dz)$ *be an MPP on* $(\Omega, \mathcal{F}, F^{(n)}, I\!P)$ *with the intensity kernel* $\lambda_t^{(n)}(dz)$.
Let further $\underline{H}^{(n)}$ *be a d-dimensional* $F^{(n)}E^{(n)}$*-P for each* $n \in I\!N$, *fulfilling*

$$\int_0^t \int_{E^{(n)}} (H_i^{(n)}(s,z))^2\ \lambda_s^{(n)}(dz)\ ds < \infty \quad I\!P-a.s. \quad \forall\, 1 \leq i \leq d \ \ \forall\, t \geq 0. \tag{8}$$

Finally, let the following two conditions hold $\forall\ \varepsilon > 0,\ t \geq 0,\ (n \to \infty)$

$$\int_0^t \int_{E^{(n)}} |\underline{H}^{(n)}(s,z)|^2 \cdot 1(|\underline{H}^{(n)}(s,z)| \geq \varepsilon)\ \lambda_s^{(n)}(dz)\ ds \ \stackrel{I\!P}{\longrightarrow}\ 0, \tag{9}$$

$$\int_0^t \int_{E^{(n)}} (\underline{H}^{(n)}(s,z))^{②}\ \lambda_s^{(n)}(dz)\ ds \ \stackrel{I\!P}{\longrightarrow}\ G_t, \tag{10}$$

where $G_t = (g_t^{ij})_{1 \leq i,j \leq d}$ *is* $\forall\ t > 0$ *a positive definite* $d \times d$*-matrix, continuous in* $t \geq 0$, *and* $g_0^{ii} = 0 \ \ \forall\ 1 \leq i \leq d$.
Then we have the following convergence in distribution in the space $D^d[0,\infty[$:

$$\underline{M}_t^{(n)} \ \stackrel{\mathcal{D}}{\longrightarrow}\ \underline{M} \ \ (n \to \infty),$$

where $\underline{M}_t^{(n)} := \int_0^t \int_E \underline{H}^{(n)}(s,z)\ M^{(n)}(ds \times dz),\ t \geq 0$, *and* $\underline{M}$ *are locally square integrable martingales. Furthermore,* $\underline{M}$ *is Gaussian with characteristic* $\langle \underline{M} \rangle_t = G_t,\ t \geq 0$.

Proof: Apply Meister (1991, Theorem 2.60) in combination with Lindvall (1973, Theorem 3′), using Theorems 1,2,3. □

In the next theorem, we consider limits of the form $t \to \infty$ instead of $n \to \infty$. We will use a family of non-singular $d \times d$-matrices $(\Gamma_t)_{t \geq 0}$, fulfilling

(T) (*i*) $\Gamma_t \to 0$ (element-wise) as $t \to \infty$,

(*ii*) $\exists$ a family of $d \times d$-matrices $(C_t)_{t \geq 0}$ being non-singular for each $t > 0$ and continuous in $t \geq 0$, such that for each fixed $s \geq 0$ we have $\Gamma_t \Gamma_{st}^{-1} \to C_s$ as $t \to \infty$.

Theorem 5. *Let $N(dt \times dz)$ be an MPP with intensity kernel $\lambda_t(dz)$. Let $\underline{H}$ be a d-dimensional FE-P satisfying* (6) *and $(\Gamma_t)_{t\geq 0}$ be a family of non-singular $d \times d$-matrices fulfilling* (T).
Let further $\forall\, \varepsilon > 0$ and for $t \to \infty$ the following two conditions hold

$$\int_0^t \int_E |\Gamma_t \underline{H}(s,z)|^2 \cdot 1(|\Gamma_t \underline{H}(s,z)| \geq \varepsilon)\ \lambda_s(dz)\ ds \ \stackrel{I\!P}{\longrightarrow}\ 0, \tag{11}$$

$$\int_0^t \int_E (\Gamma_t \underline{H}(s,z))^{②}\ \lambda_s(dz)\ ds \ \stackrel{I\!P}{\longrightarrow}\ G, \tag{12}$$

where G is a positive definite $d \times d$-matrix. Then

$$\Gamma_t \underline{M}_t \ \stackrel{\mathcal{D}}{\longrightarrow}\ \mathcal{N}_d(0,G) \quad (t \to \infty).$$

Proof: Generalize the lines of Pruscha (1984, Theorem 2.4.9), where Rebolledo (1980) was used. □

2.3 Likelihood

Following Jacod (1975), Liptser & Shiryayev (1978), Brémaud (1981), Pruscha (1984) and Last & Brandt (1995), we present a process L, which can be interpreted as the Radon-Nikodym derivative of an MPP w.r.t. another MPP, especially a marked Poisson process, and which opens the door to the likelihood approach.
Let two probability measures $I\!P, I\!P'$ on $(\Omega, \mathcal{F})$ be given satisfying $I\!P \ll I\!P'$. If $N(dt \times dz)$ admits an $(F, I\!P')$-intensity kernel $\mu_t(dz)$, then there exists a unique (up to $I\!P$-indistinguishability) FE-P $h = h(t,z)$, called **Jacod-process**, such that $N(dt \times dz)$ admits the $(F, I\!P)$-intensity kernel

$$\lambda_t(dz) = h(t,z)\mu_t(dz), \quad t \geq 0, \quad z \in E, \tag{13}$$

(see Jacod (1975, Theorem 4.1) and Last & Brandt (1995, Theorem 10.2.1)). Considering the decompositions $(\lambda_t, \Phi_t(dz))$ and $(\mu_t, \Psi_t(dz))$ as local characteristics, i.e.,

$$\lambda_t(dz) = \lambda_t \Phi_t(dz), \qquad \mu_t(dz) = \mu_t \Psi_t(dz), \quad t \geq 0,$$

we get an FE-P $g = g(t,z)$ satisfying $\Phi_t(dz) = g(t,z)\Psi_t(dz)$ by defining

$$g(t,z) := \frac{\mu_t}{\lambda_t} h(t,z), \quad t \geq 0,\ z \in E. \tag{14}$$

Now let Π be the probability on $\mathcal{F}$ admitting a local characteristic $(1, \Psi(dz))$ independent of $t \geq 0$, satisfying $\lambda_t(dz) \ll \Psi(dz) \ \ \forall\, t \geq 0$. Then $N(dt \times dz)$ is a marked standard Poisson process (see Brémaud (1981), VIII Exercise 3), and defining $\tilde{I\!P}_\sigma := \tilde{I\!P}|_{\mathcal{F}_\sigma}$ for any probability measure $\tilde{I\!P}$ and any stopping time σ, we formulate

Theorem 6. *Let h denote the Jacod process introduced in* (13) *and $L^{(0)}$ be the density of $I\!P_0$ w.r.t. Π_0.*
If τ is an F-stopping time with $\tau < \infty$ $I\!P$- and Π-a.s., then

$$I\!P_\tau \ll \Pi_\tau \quad \textit{with} \quad \frac{dI\!P_\tau}{d\Pi_\tau} = L_\tau,$$

where the process $(L_t)_{t\geq 0}$ *is defined by*

$$L_t := L^{(0)} \left(\prod_{1\leq k\leq N_t} h(\tau_k, \zeta_k) \right) \exp\left\{ \int_0^t \int_E (1 - h(s,z))\Psi(dz)\ ds \right\}, \quad t \geq 0.$$

Proof: The Theorem can be derived from Jacod (1975, Proposition 4.3), Brémaud (1981, VIII Theorem 10) and Last & Brandt (1995, Theorems 10.2.2 and 10.2.6), following the lines of Liptser & Shiryayev (1978, Theorem 19.9) and Pruscha (1984, Theorem 2.3.1). □

Let the intensity kernel depend on some - not necessarily finite dimensional - unknown parameter θ of the form $\lambda_t(\theta, dz) = h(t,z,\theta)\Psi(dz)$. Then $L \equiv L(\theta)$ can be written as

$$L_t(\theta) = \exp\left\{ \int_0^t \int_E \log h(s,z,\theta)\ N(ds \times dz) - \Lambda_t(\theta) + R_t \right\}, \quad t \geq 0, \quad \theta \in \Theta, \tag{15}$$

where $R_t := \log L^{(0)} + t,\ t \geq 0$, does not depend on $\theta \in \Theta$.

3 Semi-parametric Multiplicative Models

3.1 A Class of Multiplicative Models

As a generalization of Andersen et al. (1993, sec.VII.2), we consider a wide class of multiplicative regression models.
Let I be a countable set and $N_i(dt \times dz)$ for each $i \in I$ be an MPP with intensity kernel $\lambda_{i,t}(dz)$ of the type

$$\begin{aligned} \lambda_{i,t}(\alpha, \beta, dz) &= \lambda_{i,t}(\alpha, \beta_1) \cdot \Phi_{i,t}(\beta_2, dz), \\ \lambda_{i,t}(\alpha, \beta_1) &= \alpha(t) \cdot r(\beta_1, X_{i,t}) \cdot Y_{i,t}, \quad t \geq 0,\ i \in I,\ z \in E, \end{aligned}$$

where (with $c, d_1, d_2 \in I\!N$, $d := d_1 + d_2$, $B = B_1 \times B_2$, $B_j \overset{open}{\subset} I\!R^{d_j}$, $j = 1,2$)

$Y_{i,t},\ i \in I,\ t \geq 0,$	is an observable, bounded, (often $\{0,1\}$-valued) nonnegative FE-P,
$\alpha:\ I\!R_+ \to]0,\infty[$	is an unknown baseline hazard function satisfying $\int_0^t \alpha(s)\ ds < \infty \quad \forall\ t \geq 0$,
$r:\ I\!R^{d_1} \times I\!R^c \to]0,\infty[$	is a known regression function,
$\beta = (\beta_1, \beta_2)^T \in B$	is an unknown d-dimensional parameter,
$X_{i,t},\ i \in I,\ t \geq 0,$	is an observable, c-dimensional F-predictable process of covariates.

By N_t we denote the superposition $\sum_{i\in I} N_{i,t},\ t \geq 0$.
Assume that there exists a probability measure $\Psi(dz)$ on E such that

$$\Phi_{i,t}(\beta_2, dz) \ll \Psi(dz) \quad \forall\ i \in I,\ t \geq 0.$$

Then $\Psi(dz)$ induces a probability measure Π on $(\Omega, \mathcal{F})$ such that the MPP $N(dt \times dz)$ on $(\Omega, \mathcal{F}, F, \Pi)$ with the (F, Π)-local characteristic $(1, \Psi(dz))$ is a marked standard Poisson process. Consequently, there exists $\forall\ i \in I$ an FE-P g_i as in (14) satisfying

$$\Phi_{i,t}(\beta_2, dz) = g_i(t, z, \beta_2)\ \Psi(dz), \quad t \geq 0,\ i \in I,\ z \in E,$$

leading to an intensity kernel of the form

$$\lambda_{i,t}(\alpha, \beta, dz) \ = \ \alpha(t) \cdot r(\beta_1, X_{i,t}) \cdot Y_{i,t} \cdot g_i(t, z, \beta_2)\ \Psi(dz), \quad t \geq 0,\ i \in I,\ z \in E.$$

Thus, the Jacod process is given by

$$h_i(t, z, \beta) = \alpha(t) \cdot r(\beta_1, X_{i,t}) \cdot Y_{i,t} \cdot g_i(t, z, \beta_2), \quad t \geq 0,\ i \in I,\ z \in E.$$

Similarly to Andersen et al. (1993, p.482), we define

$$S_t(\beta_1) = \sum_{i \in I} r(\beta_1, X_{i,t}) \cdot Y_{i,t}, \quad t \geq 0,$$

and by virtue of formula (15), with $\theta = (\alpha, \beta)$, we get

$$\begin{aligned} \log L_t(\alpha, \beta) \ &= \ \int_0^t \int_E [\log \alpha(s) + \log r(\beta_1, X_{i,s}) + \log g_i(s, z, \beta_2)]\ N_i(ds \times dz) \ - \\ &\quad - \int_0^t S_s(\beta_1) \cdot \alpha(s)\ ds + R'_t, \quad t \geq 0, \end{aligned} \tag{16}$$

where R'_t neither depends on α nor on β. In case of a multivariate point process $(N_{i,t})_{t \geq 0}$, $i \in I$ (i.e. when $|E| = 1$), with $c = d_1 = d$ and

$$r(\beta, X_{i,t}) := \exp(\beta^T X_{i,t}), \quad t \geq 0,\ i \in I,$$

we obtain the classic Cox' regression model.

3.2 Partial Log-Likelihood

Guided by the Nelson-Aalen estimator (cf. Andersen et al. (1993, p.482)), we substitute $\log \alpha(t)$ by $\log \frac{1}{S_t(\beta_1)}$ and $\alpha(t)\ dt$ by $\frac{1}{S_t(\beta_1)}\ dN_t$ in formula (16), obtaining the partial log-likelihood

$$\begin{aligned} l_t(\beta) \ &= \ \sum_{i \in I} \int_0^t \int_E [\log r(\beta_1, X_{i,s}) + \log g_i(s, z, \beta_2)]\ N_i(ds \times dz) \ - \\ &\quad - \int_0^t \log S_s(\beta_1)\ dN_s \ + \ R''_t, \quad t \geq 0, \end{aligned}$$

which only depends on the unknown parameter $\beta \in B$. We will suppose that all processes are continuously differentiable as often as needed, and that the order of summation, integration and differentiation may always be changed.

For sake of simpler notation, we define for $t \geq 0,\ i \in I,\ z \in E$ the following processes which we assume to be FE-P's.

$\underline{r}^{(1)}(\beta_1, X_{i,t}) := \frac{d}{d\beta_1} r(\beta_1, X_{i,t}),$ (d_1-dimensional vector),

$\underline{r}^{(2)}(\beta_1, X_{i,t}) := \frac{d^2}{d\beta_1 d\beta_1^T} r(\beta_1, X_{i,t}),$ ($d_1 \times d_1$-matrix),

$\underline{g}_i^{(1)}(t, z, \beta_2) := \frac{d}{d\beta_2} g_i(t, z, \beta_2),$ (d_2-dimensional vector),

$\underline{g}_i^{(2)}(t, z, \beta_2) := \frac{d^2}{d\beta_2 d\beta_2^T} g_i(t, z, \beta_2),$ ($d_2 \times d_2$-matrix),

$\underline{S}_t^{(1)}(\beta_1) := \frac{d}{d\beta_1} S_t(\beta_1)$ (d_1-dimensional vector),

$\underline{S}_t^{(2)}(\beta_1) := \frac{d^2}{d\beta_1 d\beta_1^T} S_t(\beta_1)$ ($d_1 \times d_1$-matrix).

Furthermore,

$\underline{S}_t^{(1)}(\beta) := ((\underline{S}_t^{(1)}(\beta_1))^T, 0, \ldots, 0)^T$ (d-dimensional vector),

$\underline{S}_t^{(2)}(\beta) := \begin{pmatrix} \underline{S}_t^{(2)}(\beta_1) & 0 \\ 0 & 0 \end{pmatrix}$ ($d \times d$-matrix).

Finally, we introduce for $j = 1, 2$ the vectors and matrices, respectively,

$\underline{\rho}^{(j)}(\beta_1, X_{i,t}) := \frac{\underline{r}^{(j)}(\beta_1, X_{i,t})}{r(\beta_1, X_{i,t})},$ $\underline{\gamma}_i^{(j)}(t, z, \beta_2) := \frac{\underline{g}_i^{(j)}(t,z,\beta_2)}{g_i(t,z,\beta_2)}.$

Using methods of purely parametric inference (cf. Pruscha (1984), Andersen et al. (1993, sec. VI.2.2)), we get the following three Lemmas.

Lemma 1. *Defining* $\underline{U}_t(\beta) := \frac{d}{d\beta} l_t(\beta),\ t \geq 0$, *we obtain*

$$\begin{aligned} \underline{U}_t(\beta) &= \sum_{i \in I} \int_0^t \int_E \begin{pmatrix} \underline{\rho}^{(1)}(\beta_1, X_{i,s}) \\ \underline{\gamma}_i^{(1)}(s, z, \beta_2) \end{pmatrix} N(ds \times dz) \ - \int_0^t \frac{\underline{S}_s^{(1)}(\beta)}{S_s(\beta)} dN_s \ = \\ &= \sum_{i \in I} \int_0^t \int_E \underline{K}_i(s, z, \beta) \ M_i(\alpha, \beta, ds \times dz), \quad t \geq 0, \end{aligned}$$

where $M_i(\alpha, \beta, dt \times dz) = N_i(dt \times dz) - \lambda_{i,t}(\alpha, \beta, dz)\ dt$ *and*

$$\underline{K}_i(t, z, \beta) := \begin{pmatrix} \underline{\rho}^{(1)}(\beta_1, X_{i,t}) \\ \underline{\gamma}_i^{(1)}(t, z, \beta_2) \end{pmatrix} \ - \ \frac{\underline{S}_t^{(1)}(\beta)}{S_t(\beta)}, \quad i \in I,\ t \geq 0,\ z \in E.$$

Proof: Observe that $\sum_{i \in I} \int_0^t \int_E \underline{K}_i(s, z, \beta)\ \lambda_{i,s}(\alpha, \beta, dz)\ ds = 0 \ \ \forall\, t \geq 0.$ □

Lemma 2. *Assume that condition*

$$\text{(A)} \qquad \sum_{i \in I} \int_0^t \int_E K_{i,j}^2(s, z, \beta)\ \lambda_{i,s}(\alpha, \beta, dz)\ ds < \infty \quad I\!P\text{-a.s.} \quad \forall\, t \geq 0,\ 1 \leq j \leq d$$

holds. Then

$$\langle \underline{U}(\alpha, \beta) \rangle_t \ = \ \sum_{i \in I} \int_0^t \int_E \underline{K}_i^{②}(s, z, \beta)\ \lambda_{i,s}(\alpha, \beta, dz)\ ds, \quad t \geq 0.$$

Proof: Apply Theorems 2 (c) and 3. □

For the next Lemma, we put for $t \geq 0,\ i \in I,\ z \in E,\ \beta \in B$

$$
\begin{aligned}
\underline{C}_{i,t,z}(\beta) \ &:= \ \begin{pmatrix} \underline{\rho}^{(2)}(\beta_1, X_{i,t}) - \left(\underline{\rho}^{(1)}(\beta_1, X_{i,t})\right)^{\textcircled{2}} & 0 \\ 0 & \underline{\gamma}_i^{(2)}(t,z,\beta_2) - \left(\underline{\gamma}_i^{(1)}(t,z,\beta_2)\right)^{\textcircled{2}} \end{pmatrix}, \\
\underline{D}_{i,t,z}(\beta) \ &:= \ \begin{pmatrix} \left(\underline{\rho}^{(1)}(\beta_1, X_{i,t})\right)^{\textcircled{2}} & \underline{\rho}^{(1)}(\beta_1, X_{i,t})(\underline{\gamma}_i^{(1)}(t,z,\beta_2))^T \\ \underline{\gamma}_i^{(1)}(t,z,\beta_2)(\underline{\rho}^{(1)}(\beta_1, X_{i,t}))^T & \left(\underline{\gamma}_i^{(1)}(t,z,\beta_2)\right)^{\textcircled{2}} \end{pmatrix}.
\end{aligned}
$$

Lemma 3. For the $d \times d$-matrices $\underline{W}_t(\beta) := \frac{d}{d\beta}(\underline{U}_t(\beta))^T,\ t \geq 0$, we get under (A)

$$
\begin{aligned}
\underline{W}_t(\beta) \ &= \ \sum_{i \in I} \int_0^t \int_E \underline{C}_{i,s,z}(\beta)\ N_i(ds \times dz) \ - \ \int_0^t \left[\frac{\underline{S}_s^{(2)}(\beta)}{S_s(\beta)} - \left(\frac{\underline{S}_s^{(1)}(\beta)}{S_s(\beta)} \right)^{\textcircled{2}} \right] dN_s \ = \\
&= \ \underline{w}_t^*(\alpha, \beta) \ - \ \underline{w}_t(\alpha, \beta), \quad t \geq 0,
\end{aligned}
$$

where for $t \geq 0$

$$
\begin{aligned}
\underline{w}_t^*(\alpha, \beta) \ &= \ \sum_{i \in I} \int_0^t \int_E \left\{ \underline{C}_{i,s,z}(\beta) - \left[\frac{\underline{S}_s^{(2)}(\beta)}{S_s(\beta)} - \left(\frac{\underline{S}_s^{(1)}(\beta)}{S_s(\beta)} \right)^{\textcircled{2}} \right] \right\} M_i(\alpha, \beta, ds \times dz), \\
\underline{w}_t(\alpha, \beta) \ &= \ \sum_{i \in I} \int_0^t \int_E \underline{D}_{i,s,z}(\beta)\ \lambda_{i,s}(\alpha, \beta, dz)\ ds \ - \ \int_0^t \alpha(s) \frac{(\underline{S}_s^{(1)}(\beta))^{\textcircled{2}}}{S_s(\beta)}\ ds \ = \\
&= \ \langle \underline{U}(\alpha, \beta) \rangle_t,
\end{aligned}
$$

Notice that, in contrast to $\langle \underline{U}(\alpha, \beta) \rangle$, neither $\underline{U}(\beta)$ nor $\underline{W}(\beta)$ depend on the function α. Introducing the further abbrevation

$$
\underline{R}_t(\beta) := \sum_{i \in I} \int_E \underline{D}_{i,t,z}(\beta)\ r(\beta_1, X_{i,t})\ Y_{i,t}\ g_i(t,z,\beta_2)\ \Psi(dz), \quad t \geq 0,
$$

we obtain

$$
\underline{w}_t(\alpha, \beta) \ = \ \langle \underline{U}(\alpha, \beta) \rangle_t \ = \ \int_0^t \left(\underline{R}_s(\beta) - \frac{(\underline{S}_s^{(1)}(\beta))^{\textcircled{2}}}{S_s(\beta)} \right) \alpha(s)\ ds, \quad t \geq 0.
$$

Observe that in the purely multivariate case (where $|E| = 1$ and $\beta = \beta_1$)

$$
\underline{R}_t(\beta) = \sum_{i \in I} \frac{(\underline{r}^{(1)}(\beta, X_{i,t}))^{\textcircled{2}}}{r(\beta, X_{i,t})}\ Y_{i,t}, \quad t \geq 0.
$$

In the even more special Cox' case we have $\underline{R}(\beta) \equiv \underline{S}^{(2)}(\beta)$ and $\underline{C}_{i,t,z} = 0$.

3.3 Asymptotical Inference as $n \to \infty$

Now we substitute the countable set I of the last sections by a sequence $(I^{(n)})_{n \in \mathbb{N}}$ of countable sets, where $I^{(n)}$ might label the objects under observation (i.e., $I^{(n)} = \{1, \ldots, n\}$). Accordingly, we consider sequences $((N_i^{(n)}(dt \times dz), Y_{i,t}^{(n)}, X_{i,t}^{(n)}),\ i \in I^{(n)},\ z \in E,\ t \geq 0)_{n \in \mathbb{N}}$

of the corresponding processes such that for each $i \in I^{(n)}$ the MPP $N_i^{(n)}(dt \times dz)$ admits the local characteristic $(\lambda_{i,t}^{(n)}(\alpha, \beta_1), \Phi_{i,t}^{(n)}(\beta_2, dz))$ with

$$\begin{aligned}\lambda_{i,t}^{(n)}(\alpha, \beta_1) &= \alpha(t) \cdot r(\beta_1, X_{i,t}^{(n)}) \cdot Y_{i,t}^{(n)},\\ \Phi_{i,t}^{(n)}(\beta_2, dz) &= g_i^{(n)}(t, z, \beta_2)\ \Psi^{(n)}(dz), \quad t \geq 0,\ z \in E,\ n \in I\!N.\end{aligned}$$

Within this setting we denote condition (A) by (A_n). Observe that the functions α, r and the unknown parameter β do not depend on $n \in I\!N$.
For the rest of this section, we fix an arbitrary $T > 0$ to describe the end of the observation intervall $[0, T]$. All limits in this section are taken as n tends to infinity. Now we introduce a sequence of non-singular $d \times d$-norming matrices $(\Gamma_n)_{n \in I\!N}$ satisfying

(N) $\quad (i)\ \ \Gamma_n \to 0 \quad$ (element-wise),
$\quad (ii)\ \ \exists\, C_\Gamma \in \,]0, \infty[$ such that $|\Gamma_n|^2 \cdot |I^{(n)}| \leq C_\Gamma \quad \forall\, n \in I\!N.$

Adopting the terms of the previous sections and equipping them with an additional index $n \in I\!N$ if necessary, we present a further set of conditions:

There exist mappings $s, \underline{\sigma}^{(1)}, \underline{\sigma}^{(2)}, \underline{R}^{(\infty)} :\ [0, T] \times B \to I\!R, I\!R^d, I\!R^{d\times d}, I\!R^{d\times d}$, respectively, such that as $n \to \infty$

$(\mathrm{B_n})$

$$(i) \quad \sup_{t\in[0,T],\ \beta\in B} \left| |\Gamma_n|^2 S_t^{(n)}(\beta) - s_t(\beta) \right| \xrightarrow{I\!P_\beta} 0,$$

$$(ii) \quad \sup_{t\in[0,T],\ \beta\in B} \left| \frac{\underline{S}_t^{(1,n)}(\beta)}{S_t^{(n)}(\beta)} - \underline{\sigma}_t^{(1)}(\beta) \right| \xrightarrow{I\!P_\beta} 0,$$

$$(iii) \quad \sup_{t\in[0,T],\ \beta\in B} \left| \frac{\underline{S}_t^{(2,n)}(\beta)}{S_t^{(n)}(\beta)} - \underline{\sigma}_t^{(2)}(\beta) \right| \xrightarrow{I\!P_\beta} 0,$$

$$(iv) \quad \sup_{t\in[0,T],\ \beta\in B} \left| \Gamma_n \frac{(\underline{S}_t^{(1,n)}(\beta))^{②}}{S_t^{(n)}(\beta)} \Gamma_n^T - s_t(\beta) \cdot (\underline{\sigma}_t^{(1)}(\beta))^{②} \right| \xrightarrow{I\!P_\beta} 0,$$

$$(v) \quad \sup_{t\in[0,T],\ \beta\in B} \left| \Gamma_n \underline{R}_t^{(n)}(\beta) \Gamma_n^T - \underline{R}_t^{(\infty)}(\beta) \right| \xrightarrow{I\!P_\beta} 0,$$

$(\mathrm{C_n})$ $\quad (i)\ \ s, \underline{\sigma}^{(1)}, \underline{\sigma}^{(2)}$ and $\underline{R}^{(\infty)}$ are bounded in $[0, T] \times B$,
$\quad (ii)\ \ s, \underline{\sigma}^{(1)}, \underline{\sigma}^{(2)}$ and $\underline{R}^{(\infty)}$ are continuous functions in $\beta \in B$ uniformly in $t \in [0, T]$.

$(\mathrm{D_n})$ The $d \times d$-matrix $\Sigma_t(\beta) := \int_0^t \left(\underline{R}_s^{(\infty)}(\beta) - s_s(\beta) \cdot (\underline{\sigma}_s^{(1)}(\beta))^{②} \right) \alpha(s)\ ds$ is positive definite $\forall\, t \in \,]0, T]$.

$(\mathrm{E_n})$ There exists a $\delta > 0$ such that

$$\sup_{i\in I^{(n)},\ t\in[0,T]} \int_E 1 \left(r(\beta_1, X_{i,t}^{(n)}) g_i^{(n)}(t, z, \beta_2) > \exp\left\{ -\delta \left| \begin{pmatrix} \underline{\rho}^{(1)}(\beta_1, X_{i,t}^{(n)}) \\ \underline{\gamma}_i^{(1,n)}(t, z, \beta_2) \end{pmatrix} \right| \right\} \right) \cdot$$
$$\cdot \left| \Gamma_n \begin{pmatrix} \underline{\rho}^{(1)}(\beta_1, X_{i,t}^{(n)}) \\ \underline{\gamma}_i^{(1,n)}(t, z, \beta_2) \end{pmatrix} \right| \cdot Y_{i,t}^{(n)}\ \Psi^{(n)}(dz) \xrightarrow{I\!P_\beta} 0.$$

In case of $\Gamma_n = \frac{1}{\sqrt{a_n}} I_d$ with $0 < a_n \to \infty$, the following condition is sufficient for (B_n)(i)-(v)

$$(\mathrm{B}'_\mathrm{n}) \qquad (i) \quad \sup_{t\in[0,T],\ \beta\in B} \left| \frac{1}{a_n} \underline{S}_t^{(m,n)}(\beta) - \underline{s}_t^{(m)}(\beta) \right| \xrightarrow{I\!P_\beta} 0 \quad (m = 0,1,2),$$

$$(ii) \quad \sup_{t\in[0,T],\ \beta\in B} \left| \frac{1}{a_n} \underline{R}_t^{(n)}(\beta) - \underline{R}_t^{(\infty)}(\beta) \right| \xrightarrow{I\!P_\beta} 0,$$

where $\underline{s}^{(m)}(\beta) \equiv s(\beta) \cdot \underline{\sigma}^{(m)}(\beta)$, $m = 1,2$, demanding additionally that $s.(\beta) : [0,T] \to I\!R$ is bounded away from 0. In the purely multivariate Cox' case, (B'_n)(i) contains (B'_n)(ii) with $\underline{R}^\infty(\beta) \equiv \underline{s}^{(2)}(\beta)$, see Andersen et al. (1993 p.497, condition VII.2.1).

A sequence $(\hat\beta_t^{(n)})_{n\in I\!N}$ of d-dimensional random vectors is called **consistent partial likelihood estimator** or shortly **consistent PMLE**, iff we have for $n \to \infty$

$$I\!P_\beta(|\hat\beta_t^{(n)} - \beta| < \delta,\ \ \underline{U}_t^{(n)}(\hat\beta_t^{(n)}) = 0) \ \to\ 1 \qquad \forall\ \delta > 0,$$

where $\underline{U}_t^{(n)}$ is given by Lemma 1.

Proposition 1. *Under* (A_n),(B_n)(i),(ii),(iv),(v),(C_n),(D_n),(E_n) *we have*

$$(\mathrm{U}^*_\mathrm{n}) \qquad \Gamma_n \underline{U}_t^{(n)}(\beta) \xrightarrow{\mathcal{D}_\beta} \mathcal{N}_d\,(0, \Sigma_t(\beta)) \qquad \forall\ t \in [0,T].$$

Proof: One can show that (B_n)(i),(ii), (C_n) and (E_n) imply (9), while (B_n)(iv),(v) and (D_n) yield (10). Condition (A_n) allows the application of Theorem 4 which completes the proof. □

Now we consider sequences of d-dimensional random vectors $(\beta_n^*)_{n\in I\!N}$ fulfilling

$$(\mathrm{B}^*_\mathrm{n}) \qquad \Gamma_n^{-T}(\beta_n^* - \beta), \ \ n \in I\!N, \ \ \text{is} \ \ I\!P_\beta\text{-stochastically bounded.}$$

Proposition 2. *Under* (A_n),(B_n),(C_n) *we have*

$$(\mathrm{W}^*_\mathrm{n}) \qquad -\Gamma_n \underline{W}_t^{(n)}(\beta_n^*)\Gamma_n^T \xrightarrow{I\!P_\beta} \Sigma_t(\beta) \qquad \forall\ \ t \in [0,T]$$

and $\forall$ *sequences of d-dimensional random vectors* $(\beta_n^*)_{n\in I\!N}$ *satisfying* $(\mathrm{B}^*_\mathrm{n})$.

Proof: Decompose $|-\Gamma_n \underline{W}_t^{(n)}(\beta_n^*)\Gamma_n^T - \Sigma_t(\beta)|$ similar to Andersen et al. (1993, p.500-01). □

Theorem 7. Let (A_n),(B_n),(C_n),(D_n) hold. Then there exists for each $t \in [0,T]$ a consistent PMLE $(\hat\beta_t^{(n)})_{n\in I\!N}$ for β fulfilling (B^*_n).

Proof: See Pruscha (1996 sec.VI, Satz 1.2). The basic ideas are due to Aitchison & Silvey (1958), Billingsley (1971) and Feigin (1975). □

Theorem 8. (A_n),(B_n),(C_n),(D_n) *and* (E_n) *imply* (U^*_n), (W^*_n), *and under* (U^*_n), (W^*_n) *we have for any consistent PMLE* $(\hat\beta_t^{(n)})_{n\in I\!N}$ *satisfying* (B^*_n)

$$\Gamma_n^{-T}(\hat\beta_t^{(n)} - \beta) \ \xrightarrow{\mathcal{D}_\beta}\ \mathcal{N}_d(0, \Sigma_t^{-1}(\beta)) \qquad \forall\ t \in [0,T].$$

Proof: Apply Propositions 1,2 and Pruscha (1996 sec.VI, Satz 1.6). □

Now we can estimate α as in the classical point process theory (see, e.g., Andersen et al. (1993, sec.VII.2)). First we approximate $A(t) := \int_0^t \alpha(s)\, ds$ by the Breslow estimator

$$\hat{A}_t^{(n)}(\hat{\beta}_t^{(n)}) := \int_0^t \frac{1(Y_s^{(n)} > 0)}{S_s^{(n)}(\hat{\beta}_t^{(n)})} dN_s^{(n)}, \quad t \in [0, T],\ n \in I\!N,$$

where $Y_t^{(n)} := \sum_{i \in I^{(n)}} Y_{i,t}^{(n)}$. Finally, we obtain the kernel estimator

$$\hat{\alpha}_t^{(n)} = \frac{1}{b} \sum_{k=1}^{N_t^{(n)}} K\left(\frac{t - \tau_k^{(n)}}{b}\right) \frac{1}{S_{\tau_k}^{(n)}}, \quad t \in [0, T],\ n \in I\!N,$$

where K is a kernel function and b the bandwidth.

4 Parametric Inference as $t \to \infty$

In this chapter, we deal with purely parametric problems as time t tends to infinity when there is just one object under observation, devellopping topics from Pruscha (1984), Hjort (1985) and Andersen et al. (1993).

4.1 M-estimators

We consider an MPP $N(dt \times dz)$ with intensity kernel $\lambda_t(\theta, dz)$ admitting the decomposition

$$\lambda_t(\theta, dz) = h(t, z, \theta)\Psi(dz), \quad t \geq 0,\ z \in E,$$

where $\theta \in \Theta \overset{open}{\subset} I\!R^d$, $d \in I\!N$, $\Psi(dz)$ is a probability measure on E satisfying $\lambda_t(\theta, dz) \ll \Psi(dz)\ \ \forall\, t \geq 0,\ \theta \in \Theta$, and h is the accompanying Jacod process which we assume to be continuously differentiable w.r.t θ.
An **M-estimator** $\hat{\theta}_t$ **of** $\theta \in \Theta$ is a solution of the equations $\underline{U}_t(\theta) = 0$, where

$$\underline{U}_t(\theta) := \int_0^t \int_E \underline{K}(s, z, \theta)\ M(\theta, ds \times dz), \quad t \geq 0,$$

and $M(\theta, dt \times dz) = N(dt \times dz) - \lambda_t(\theta, dz)\ dt$. For the d-dimensional process $\underline{K}$, we assume

$$\int_0^t \int_E |\underline{K}(s, z, \theta)|\ \lambda_s(\theta, dz)\ ds < \infty \quad I\!P_\theta - a.s.\ \ \forall\, t \geq 0.$$

If $\underline{K}(t, z, \theta) = \frac{\frac{d}{d\theta} h(t,z,\theta)}{h(t,z,\theta)}$, we refer to the **likelihood case**, and $\hat{\theta}_t$ is called **maximum likelihood estimator (MLE)**.
Now we state a first set of basic conditions

(A_t) (i) $\underline{K}$ has continuous first order derivatives w.r.t. θ;

the processes $\underline{K}$, $\underline{K}^{(1)} := \frac{d}{d\theta} \underline{K}^T$ and $\underline{h}^{(1)} := \frac{d}{d\theta} h$ are FE-P's.

(ii) $\underline{K}$ and h have continuous second order derivatives w.r.t. θ;

the processes $\underline{K}^{(2)} := \left(\frac{\partial}{\partial\theta_j} \underline{K}^{(1)}\right)_{1 \leq j \leq d}$ and $\underline{h}^{(2)} := \frac{d^2}{d\theta d\theta^T} h$ are FE-P's.

Note that the Jacod-process h is by definition an FE-P. The following Lemma can easily be proven using differentiation rules.

Lemma 4. **(a)** *Assuming* $(\mathrm{A}_t)(\mathrm{i})$, *we have for* $t \geq 0$

$$\underline{W}_t(\theta) \;:=\; \frac{d}{d\theta}(\underline{U}_t(\theta))^T \;=\; \underline{w}_t^*(\theta) - \underline{w}_t(\theta),$$

where

$$\underline{w}_t^*(\theta) \;:=\; \int_0^t \int_E \underline{K}^{(1)}(s,z,\theta)\; M(\theta, ds \times dz),$$

$$\underline{w}_t(\theta) \;:=\; \int_0^t \int_E \underline{K}(s,z,\theta)\frac{(\underline{h}^{(1)}(s,z,\theta))^T}{h(s,z,\theta)}\lambda_s(\theta,dz)\; ds.$$

(b) *In the likelihood case,* $(\mathrm{A}_t)(\mathrm{i})$ *implies for* $t \geq 0$

$$\underline{W}_t(\theta) \;=\; \underline{v}_t^*(\theta) - \underline{v}_t(\theta),$$

with

$$\underline{v}_t^*(\theta) \;:=\; \int_0^t \int_E \frac{\underline{h}^{(2)}(s,z,\theta)}{h(s,z,\theta)}\; M(\theta, ds \times dz),$$

$$\underline{v}_t(\theta) \;:=\; \int_0^t \int_E \frac{(\underline{h}^{(1)}(s,z,\theta))^{②}}{h^2(s,z,\theta)}\; N(ds \times dz).$$

Even in the likelihood case, we have $\underline{w}^*(\theta) \neq \underline{v}^*(\theta)$ and $\underline{w}(\theta) \neq \underline{v}(\theta)$; however, $\underline{w}(\theta)$ is the compensator of $\underline{v}(\theta)$. Both $\underline{w}^*(\theta)$ and $\underline{v}^*(\theta)$ have martingale structure.
We further define the $d \times d$-matrix

$$\underline{I}_t(\theta) \;:=\; \int_0^t \int_E \underline{K}^{②}(s,z,\theta)\; \lambda_s(\theta,dz)\; ds, \quad t \geq 0,$$

and the $d^2 \times d^2$-matrix

$$\underline{I}_t^*(\theta) \;:=\; \int_0^t \int_E (\underline{K}^{(1)}(s,z,\theta))^{②}\; \lambda_s(\theta,dz)\; ds, \quad t \geq 0.$$

For these matrices we state a further condition:

$(\mathrm{B_t})$ (i) Let the diagonal elements of $\underline{I}_t(\theta)$ be finite $I\!P_\theta$-a.s. $\forall\, t \geq 0$.
         (ii) Let the diagonal elements of $\underline{I}_t^*(\theta)$ be finite $I\!P_\theta$-a.s. $\forall\, t \geq 0$.

Note that (B_t) implies via Cauchy-Schwarz inequality that *all* the elements of the matrices are finite $I\!P_\theta$-a.s. Now Theorems 2(c) and 3 yield

Lemma 5. *Let* $(\mathrm{A}_t)(\mathrm{i})$ *hold. Then we have*
(a) $(\mathrm{B}_t)(i) \Rightarrow \underline{U}(\theta)$ *is a locally square integrable martingale with characteristic*

$$\langle \underline{U}(\theta)\rangle_t \;=\; \int_0^t \int_E \underline{K}^{②}(s,z,\theta)\; \lambda_s(\theta,dz)\; ds, \quad t \geq 0.$$

(b) $(\mathrm{B}_t)(ii) \Rightarrow \underline{w}^*(\theta)$ *is a locally square integrable martingale with characteristic*

$$\langle \underline{w}^*(\theta)\rangle_t \;=\; \int_0^t \int_E (\underline{K}^{(1)}(s,z,\theta))^{②}\, \lambda_s(\theta,dz)\; ds, \quad t \geq 0.$$

4.2 Asymptotical Inference as $t \to \infty$

As a generalization of Pruscha (1984) and Andersen et al. (1993, sec.VI.2), we present a set of conditions under which the asymptotic theory of Pruscha (1996, sec.VI.1) can be applied for M-estimators.
Considering a family of non-singular $d \times d$-matrices $(\Gamma_t)_{t\geq 0}$ satisfying (T), we formulate a further condition which has an ergodic [(i),(ii)] and a Lindeberg-like part [(iii)]. All the limits in this section are taken as $t \to \infty$.

($\mathrm{C_t}$) (*i*) $\Gamma_t \langle \underline{U}(\theta)\rangle_t \Gamma_t^T \xrightarrow{I\!P_\theta} \Sigma(\theta)$, where $\Sigma(\theta)$ is a positive definite, symmetric $d \times d$-matrix,

(*ii*) $\Gamma_t \underline{w}_t(\theta) \Gamma_t^T \xrightarrow{I\!P_\theta} B(\theta)$, where $B(\theta)$ is a positive definite $d \times d$-matrix,

(*iii*) $\mathcal{L}_t(\theta,\varepsilon) \xrightarrow{I\!P_\theta} 0 \quad \forall\, \varepsilon > 0$, where

$$\mathcal{L}_t(\theta,\varepsilon) := \int_0^t \int_E |\Gamma_t \underline{K}(s,z,\theta)|^2 \cdot 1(|\Gamma_t \underline{K}(s,z,\theta)| > \varepsilon)\ \lambda_s(\theta, dz)\ ds.$$

In the likelihood case, conditions (C_t)(i) and (C_t)(ii) are identical, with $B(\theta) = \Sigma(\theta)$.

Proposition 3. *Let conditions* (A_t)(i), (B_t)(i), (C_t)(i),(iii) *be satisfied. Then we have the following convergence in distribution*

$$(\mathrm{U_t^*}) \qquad \Gamma_t \underline{U}_t(\theta) \xrightarrow{\mathcal{D}_\theta} \mathcal{N}_d\left(0, \Sigma(\theta)\right).$$

Proof: (B_t)(i), (C_t)(iii) and (C_t)(i) yield (6), (11) and (12), respectively, such that the application of Theorem 5 completes the proof. □

To state asymptotic results for the process $\underline{W}(\theta)$, we formulate a further set of conditions

($\mathrm{D_t}$)(*i*) $\gamma_t^{kg}\gamma_t^{jh}\langle w^{*hg}(\theta)\rangle_t$, $t \geq 0$, is $I\!P_\theta$-stochastically bounded $\forall\, 1 \leq g,h,j,k \leq d$, where $\Gamma_t = (\gamma_t^{jk})_{1\leq j,k\leq d}$ and $\underline{w}_t^*(\theta) = (w_t^{*jk}(\theta))_{1\leq j,k\leq d}$.

(*ii*) There exists a neighborhood $\Theta_0 \subset \Theta$ of θ and $\exists\, M_\theta < \infty$ such that $\lim_{t\to\infty} I\!P_\theta(|\Gamma_t \tilde{R}_t(\tilde{\theta})\Gamma_t^T| < M_\theta \quad \forall\, \tilde{\theta} \in \Theta_0) = 1$,

where $\tilde{R}_t(\theta) = (\tilde{R}_t^{jk}(\theta))_{1\leq j,k\leq d}$, $\tilde{R}_t^{jk}(\theta) := \sum_{l=1}^d R_t^{jkl}(\theta)$, $R_t^{jkl}(\theta) := \frac{\partial}{\partial\theta_l} W_t^{jk}(\theta)$.

Proposition 4. (A_t), (B_t)(ii), (C_t)(ii) *and* (D_t)(i) *imply*

$$(\mathrm{W_{t,0}^*}) \qquad -\Gamma_t \underline{W}_t(\theta)\Gamma_t^T \xrightarrow{I\!P_\theta} B(\theta),$$

Proof: Apply Lemma 5(b) and use Lenglart's inequality. □

Now we consider sequences of d-dimensional random vectors $(\theta_t^*)_{t\geq 0}$ fulfilling

$$(\mathrm{B_t^*}) \qquad \Gamma_t^{-T}(\theta_t^* - \theta), \quad t \geq 0, \quad \text{is } I\!P_\theta\text{-stochastically bounded.}$$

Proposition 5. *Under conditions* (A_t), (B_t)(ii), (C_t)(ii) *and* (D_t) *we have as* $t \to \infty$

($\mathrm{W_t^*}$) $-\Gamma_t \underline{W}_t(\theta_t^*)\Gamma_t^T \xrightarrow{I\!P_\theta} B(\theta)$

$\forall$ *sequences of d-dimensional random vectors* $(\theta_t^*)_{t\geq 0}$ *satisfying* (B_t^*).

Proof: Expand $\underline{W}(\theta)$ in a Taylor series around the true parameter θ and apply Proposition 4. □

Theorem 9. (A_t), (B_t), (C_t) *and* (D_t) *imply* (U_t^*), (W_t^*), *and under* (U_t^*), (W_t^*), *the following holds:*
There exists a consistent M-estimator $(\hat{\theta}_t)_{t\geq 0}$ *for* θ *fulfilling* (B_t^*).
If there exists an estimation function $l : I\!\!R_+ \times \Theta \to I\!\!R$ *such that* $\underline{U}(\theta) = \frac{d}{d\theta} l(\theta)$, *then with* $I\!\!P_\theta$*-probability tending to one,* l_t *takes a local maximum at* $\hat{\theta}_t$.
Furthermore, we have for consistent M-estimators $(\hat{\theta}_t)_{t\geq 0}$ *fulfilling* (B_t^*)

$$\Gamma_t^{-T}(\hat{\theta}_t - \theta) \xrightarrow{\mathcal{D}_\theta} \mathcal{N}_d(0, B^{-1}(\theta)\Sigma(\theta)B^{-T}(\theta)).$$

Proof: Apply Propositions 3,5 and Pruscha (1996 sec.VI, Satz 1.2 and Satz 1.6, using ideas due to Aitchison & Silvey (1958), Billingsley (1961) and Feigin (1975)). □

References

[1] Aalen, O.O. (1975): Statistical Inference for a Family of Counting Processes. PhD thesis, University of California, Berkeley.

[2] Aitchison, J. and Silvey, S.D. (1958): Maximum Likelihood Estimation of Parameters subject to Restraints. Ann. Math. Statist. **29**, 813-828.

[3] Andersen, P.K. and Borgan, Ø. (1985): Counting Process Models for Life History Data: A Review. Scand. J. Statistics **12**, 97-158.

[4] Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993): Statistical Models Based on Counting Processes. Springer, New York.

[5] Billingsley, P. (1961): Statistical Inference for Markov Processes. The University of Chicago Press, Chicago.

[6] Boel, R., Varaiya, P. and Wong, E. (1975): Martingales on Jump Processes I, II. Siam J. Control **13**, 999-1021 and 1022-1061.

[7] Brémaud, P. (1972): A Martingale Approach to Point Processes. Electronic Res. Lab., Memo no. M-345. University of California, Berkeley.

[8] Brémaud, P. (1981): Point Processes and Queues; Martingale Dynamics. Springer, New York.

[9] Dellacherie, C. and Meyer, P.-A. (1982): Probabilities and Potential B. North-Holland, Amsterdam.

[10] Feigin, P.D. (1975): Maximum Likelihood Estimation for Stochastic Processes - a Martingale Approach. Ph. D. Thesis, Australian National University.

[11] Hjort, N. (1985): Discussion of Andersen & Borgan (1985). Scand. J. Statistics **12**, 141-150.

[12] Jacod, J. (1975): Multivariate Point Processes: Predictable Projection, Radon-Nikodym Derivatives, Representation of Martingales. Z. Wahrsch. verw. Gebiete **31**, 235-253.

[13] Kopp, P.E. (1984): Martingales and Stochastic Integrals. Cambridge University Press, Cambridge.

[14] Last, G. and Brandt, A. (1995): Marked Point Processes on the Real Line. Springer, New York.

[15] Lindvall, T. (1973): Weak Convergence of Probability Measures and Random Functions in the Function Space $D[0,\infty[$. Journal of Applied Probability **10**, 109-121.

[16] Liptser, R.S. and Shiryayev, A.N. (1977): Statistics of Random Processes I; General Theory. Springer, New York.

[17] Liptser, R.S. and Shiryayev, A.N. (1978): Statistics of Random Processes II; Applications. Springer, New York.

[18] Meister, H. (1991): Funktionale Grenzwertsätze für Martingale. Diplomarbeit am mathematischen Institut der Ludwig-Maximilians-Universität in München.

[19] Murphy, S.A. (1995): A Central Limit Theorem for Local Martingales with Applications to the Analysis of Longitudinal Data. Scand. J. Statistics 1995 **22**, 279-294. Blackwell, Cambridge (USA).

[20] Pruscha, H. (1984): Statistical Inference in Multivariate Point Processes with Parametric Intensity. Habilitationsschrift, Fachbereich Mathematik der Ludwig-Maximilians-Universität in München.

[21] Pruscha, H. (1996): Angewandte Methoden der Mathematischen Statistik. Teubner, Stuttgart.

[22] Pruscha, H. (1997): Semiparametric Estimation in Regression Models for Point Processes based on One Realization. Proc. 12th Int. Workshop Stat. Mod., Schriftenreihe d. Österr. Stat. Ges. **5**, Wien.

[23] Rebolledo, R. (1980): Central Limit Theorems for Local Martingales. Z. Wahrsch. verw. Gebiete **51**, 269-286.

[24] Scheike, T.H. (1994a): Parametric Regression for Longitudinal Data with Counting Process Measurement Times. Scand. J. Statistics 1994 **21**, 245-263. Blackwell, Cambridge (USA).

[25] Scheike, T.H. (1994b): Nonparametric Kernel Regression when the Regressor follows a Counting Process. Research Report 94/5, Univ. of Copenhagen.