CWI Tracts

Managing Editors

J.W. de Bakker (CWI, Amsterdam)
M. Hazewinkel (CWI, Amsterdam)
J.K. Lenstra (CWI, Amsterdam)

Editorial Board

W. Albers (Enschede)
P.C. Baayen (Amsterdam)
R.C. Backhouse (Groningen)
E.M. de Jager (Amsterdam)
M.A. Kaashoek (Amsterdam)
M.S. Keane (Delft)
H. Kwakernaak (Enschede)
J. van Leeuwen (Utrecht)
P.W.H. Lemmens (Utrecht)
M. van der Put (Groningen)
M. Rem (Eindhoven)
H.J. Sips (Delft)
M.N. Spijker (Leiden)
H.C. Tijms (Amsterdam)

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The CWI is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (N.W.O).

CWI Tract 71

Counting process systems
Identification and stochastic realization

P.J.C. Spreij

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

1980 Mathematics Subject Classification: 60G55, 62M05, 62F12, 93E03, 93E12.
ISBN 90 6196 388 5
NUGI-code: 811

Copyright © 1990, Stichting Mathematisch Centrum, Amsterdam
Printed in the Netherlands

Preface

This monograph is a slightly revised version of the author's thesis. The thesis has been prepared at CWI under supervision of J.H. van Schuppen, whose continuous interest and valuable advices were particularly welcome.
H. Kwakernaak, my promotor, and R.K. Boel commented on earlier drafts of this monograph. Their remarks and suggestions are also highly appreciated.

Mini Middelberg did an excellent job by typing the manuscript. Finally I thank CWI for allowing me to publish this material in the tract series.

P.J.C. Spreij

Contents

Chapter 1

Introduction

In this monograph we treat identification and realization problems for counting process systems. Counting processes are nowadays frequently used in mathematical models for phenomena that occur in fields like nuclear biology, survival analysis, optical physics, traffic flow analysis and software reliability. For instance in software reliability counts of observed failures of a computer program are registered. These counts can be used to estimate the total number of errors in the program [43,44]. An example in traffic flow analysis is given in section 1.1. Many other examples can be found in the book by SNYDER [48].

The theory of counting processes has seen a fast development over the past 15 years, stimulated by the growth of the general theory of stochastic processes, martingale theory, stochastic integration and stochastic differential equations. This theory provides a most suitable framework for studying the dynamic behaviour of counting processes. Since then many problems concerning the modelling and identification of counting processes that arise in the aforementioned fields can be solved in a satisfactory way. One reason for succesful attempts in this direction is the availability of many theoretical results for martingales like convergence theorems and central limit theorems that also play a crucial role in this monograph. Our aim is to derive results on recursive estimation and realization problems for counting process systems.

The purpose of this chapter is to introduce counting processes and related identification and realization problems on an intuitive level.

1.1 INFORMAL INTRODUCTION TO COUNTING PROCESSES

We introduce the notion of a counting process by means of an example. In the analysis of freeway traffic flows information about such flows is obtained by counting vehicles that pass a certain location at a freeway [30]. To that end a detection loop is built at that location where the time instants are registered when vehicles pass over it. Clearly those time instants exhibit an irregular behaviour, which points in the direction of a random phenomenon. More precisely, the time intervals that elapse between two succesive registrations can be viewed as random variables. One way to build a mathematical model, albeit unrealistic, is to make the assumption that those time intervals are independent and identically distributed according to an exponential distribution with parameter λ. So if, starting from an initial time 0, we denote by T_n the time when the n-th vehicle passes the detection loop, we have that all the $T_{n+1}-T_n$ are independent with the same exponential distribution and $E(T_{n+1}-T_n)=\frac{1}{\lambda}$.

Thus we see that we can anticipate the intervals $T_{n+1}-T_n$ to be short for high values of λ, whereas the opposite will happen for low values of λ. The interpretation in this example is that high values of λ reflect a high intensity of the traffic flow. A counting process $\{N_t\}$ is now defined by

$$N_t = \sum_n 1_{\{T_n \leq t\}} \tag{1.1}$$

So N_t is the number of vehicles that have passed the detection loop up to time t. In agreement with its intuitive interpretation the parameter λ is called the intensity of the counting process. The counting process that arises in this example is called the homogeneous Poisson process with parameter λ. This name is due to the fact that for all t and $h \geqslant 0$, the random variable $N_{t+h}-N_t$ is distributed as a Poisson random variable with parameter λh. If we capture the history of the process $\{N_t\}$ up to time t into a σ-algebra $\mathcal{F}_t^N$, we can show that for $h \downarrow 0$ we have the conditional probabilities

$$P(N_{t+h}-N_t = 1|\mathcal{F}_t^N) = \lambda h + \mathcal{O}(h) \tag{1.2a}$$

$$P(N_{t+h}-N_t = 0|\mathcal{F}_t^N) = 1-\lambda h + \mathcal{O}(h) \tag{1.2b}$$

The equations (1.2) again justify why λ is called the intensity of $\{N_t\}$. Indeed a high value of λ makes a new count in a small time interval $(t,t+h)$ more likely. In a more fancy way (1.2) admits an alternative description. Define the process $\{m_t\}$ by $N_t = \lambda t + m_t$. In a differential notation this can be written as

$$dN_t = \lambda dt + dm_t. \tag{1.3}$$

Then one can show that $\{m_t\}$ is a martingale. A martingale is the archetype of a stochastic process with an asystematic behaviour. [See definition 2.2.1]. The important observation is that under fairly general assumptions one can always decompose a counting process $\{N_t\}$ (the assumption that the time intervals $T_{n+1}-T_n$ are i.i.d. is not necessary any more) as

$$dN_t = \lambda_t dt + dm_t \tag{1.4}$$

The process $\{m_t\}$ arising in (1.4) is then again a martingale and this decomposition is essentially unique. However λ_t is in general not a fixed constant but a stochastic process. It is called the intensity process. The decomposition (1.4) is important because of two reasons. Often the assumption in the previous example that the $T_{n+1}-T_n$ are independent and identically distributed is untenable. So (1.4) allows a greater variety of counting processes. The second reason is that, not only is the process $\{\lambda_t\}$ unique, it also uniquely determines the distribution of $\{N_t\}$. So a way to model a counting process is to specify the intensity process. Stated otherwise one can say that modelling a counting process is equivalent to modelling the corresponding intensity process.

1.2 Modelling the intensity process

We have seen in section 1.1 that the intensity of a counting process is in general a stochastic process. In a practical situation one can often think of the intensity λ_t as a function of another stochastic process $\{X_t\}$ which has a physical meaning in the particular situation at hand, so that we can write $\lambda_t = f(t, X_t)$. In such a case the modelling of the intensity process reduces to specifying the function f. In doing this one can often distinguish between two stages, the first of which may be called structural modelling. We illustrate this with an example. Suppose that $X_t \in \mathbb{R}^d$. If the structural form of the intensity is linear we have $\lambda_t = \alpha(t)^T X_t$, where $\alpha(t) \in \mathbb{R}^d$. This is the multiplicative intensity model studied by Aalen [12,17]. A special case is obtained if $\alpha(t) = \alpha$ for all t, so $\lambda_t = \alpha^T X_t$. In another example the structural form of the intensity process might be exponential, so we can write $\lambda_t = \exp(\alpha(t)^T X_t)$ or $\lambda_t = \exp(\alpha^T X_t)$. If the structural form of the intensity process is given, one can say that one is left with a parameteric modelling problem. This can be an infinite dimensional problem, as in the case where $\lambda_t = \alpha(t)^T X_t$. The parameter is then the function α which is in general an infinite dimensional object. In the special case where $\lambda_t = \alpha^T X_t$ the parameter is an element of $\mathbb{R}^d$ which yields a finite dimensional problem. One can say that establishing the structural form of the intensity corresponds to the selection of a model class. The model class is then described by the finite or infinite dimensional parameter.

An intensity process of the form $\lambda_t = f(t, X_t)$ naturally arises in the context of stochastic system theory. Roughly speaking a stochastic system is a pair of stochastic processes, the state and the output process. The state process is always a Markov process that drives the output process. If the output process is a counting process then its intensity process is of the form $\lambda_t = f(t, X_t)$ if $\{X_t\}$ is the state process. A stochastic system with counting process output is called a counting process system. A shorthand notation for such a system is (X,N) where X stands for $\{X_t\}$ and N for $\{N_t\}$. An important example is the following. (See [20] for an application). Suppose that the state space of the Markov process $\{X_t\}$ is finite, $\{x_1,...,x_n\}$ say. Then we can define the indicator process $\{Y_t\}$ by $Y_t = [Y_{1t},...,Y_{nt}]^T$ and $Y_{it} = 1_{\{X_t = x_i\}}$. Assume that $\lambda_t = f(X_t)$. Let $c_i = f(x_i)$ and $C = [c_1,...,c_n]$. Then we can write $\lambda_t = CY_t$. Let A be the generator matrix of $\{X_t\}$. It can be shown that we can represent this model by the following equations

$$dY_t = AY_t dt + dM_t \tag{1.5a}$$

$$dN_t = CY_t dt + dm_t \tag{1.5b}$$

The processes $\{M_t\}$ and $\{m_t\}$ in (1.5) are martingales. Observe the analogy of (1.5) with Gaussian systems, where one has similar equations (1.5) for an n-dimensional Gaussian process $\{Y_t\}$ and a one dimensional process $\{N_t\}$ with $\{M_t\}$ and $\{m_t\}$ an n-dimensional and a one dimensional Brownian motion. The equations (1.5) can be considered as the structural model of the counting process. The parametric modelling problem in this case is the specification of the matrices A and C. It is noticed that a structural model can often be set up

on the basis of information that one has about the particular phenomenon that one wants to model. The values of the parameters are usually not known beforehand. The only way to obtain these is by analyzing the observed data. This leads to the problem of identification.

1.3 Identification

Speaking in loose terms one can say that identification is concerned with obtaining a model from the observed data. In the situation where one deals with counting process observations this amounts to identification of the underlying intensity process. As in modelling problems it is in principle possible to distinguish between structural and parametric identification. However if one wants to perform an identification procedure based on a single realization of the counting process, it is in general difficult if not impossible, to obtain the structural form of the intensity process. In such a situation one has to assume that the structural form of the intensity process is given, in which case the identification problem becomes a parameter identification problem. But even then one cannot always expect to identify the "true" but unknown parameter. This is for instance the case if $\lambda_t = \alpha(t)^T X_t$ where the parameter α is a function. For the identification of the function α there exist non parametric techniques for which several realizations of the same counting process are needed [12,17]. Since we will treat some identification procedures that are based on a single realization of the counting process, we will assume that the unknown parameter is finite dimensional which is for instance the case if $\lambda_t = \alpha^T X_t$, where $\alpha \in \mathbf{R}^d$. If the structural form of the intensity, or a model class, is given, one can view identification as approximate modelling. Estimating a parameter then corresponds to the selection of a model that best explains the observation according to a certain criterion. There are basically two procedures for estimating a finite dimensional parameter called off-line and on-line. Off-line procedures are applied if it is possible to collect the data before actually computing a parameter estimate. Off-line estimators can often be obtained by minimizing a suitable criterion.

A well-known example is the maximum likelihood estimator. It is known that these estimators are consistent and enjoy certain optimality properties. For counting process observations results in this direction are obtained in [19,22,23,27,32], by means of a suitable analysis of the likelihood ratio.

On-line or recursive procedures naturally arise if one is confronted with a control or a filtering problem. In a control problem one usually looks for a feedback control law in order to meet some required behaviour of the output process. This control law depends on a parameter whose value might be unknown. Hence if one wants to apply this control law at all time instants t one also needs estimates of the unknown parameter values for all t. This calls for a device that computes new estimates from previous ones and from new observations. This device usually consists of a set of stochastic differential equations. Denote by $\hat{\theta}_t$ an estimator of θ at time t. In the context of counting processes the stochastic differential equation for $\hat{\theta}_t$ is then of the following form

$$d\hat{\theta}_t = f(t,\hat{\theta}_t)dt + g(t,\hat{\theta}_{t-})dN_t \tag{1.6}$$

The interpretation of (1.6) is that $\hat{\theta}_t$ evolves according to an ordinary differential equation between the occurrence times of the counting process (then $dN_t = 0$) whereas $\hat{\theta}_t$ jumps to a new value if a count is registered ($dN_t = 1$). Similar considerations hold for filtering problems. A filtering or state estimation problem arises when one can only observe a counting process, but not the state process X_t that influences the intensity process as for instance in (1.5). A filter is then a set of stochastic differential equations that determines the conditional expectation of X_t given the past observations. Such a filter generally depends on the unknown parameter value. Hence to actually compute such a conditional expectation one again needs parameter estimates that are generated by an equation like (1.6). The combined set of stochastic differential equations that simultaneously estimate state and parameter is called an adaptive filter. One of the requirements that an adaptive filter should satisfy is that parameter estimates converge to the true parameter if the true model is in the model class. The same requirement is of course desired for any on-line estimation procedure whether or not it is part of an adaptive filter. Although we will not treat adaptive filtering problems, the above considerations at least motivate why it is important to study the convergence properties of recursive estimators.

One of the difficult problems for recursive estimators is the design of an equation like (1.6). A reason is that it is often not clear how to obtain a recursive formula starting from known off-line estimators. Equations like (1.6) can be obtained via heuristic reasoning and by making suitable approximations. In chapter 4 we give some examples. A consequence is that, even if it is intuitively clear that a certain recursive estimator defined by (1.6) is close to a known off-line estimator, it is not easy to see whether certain known properties for the off-line estimator carry over to the corresponding on-line estimator. A technical analysis is hampered by the fact that it is often not known what criterion a recursive estimator minimizes, or what estimation equation it satisfies. Hence for on-line estimators new techniques for investigation of e.g. asymptotic properties have to be developed. Since recursive estimators are defined via stochastic differential equations, stochastic Lyapunov functions can be expected to play a role. An analysis of this type is presented in chapter 4. Finally it is noticed that if one wants to prove consistency of a family of estimators $\{\hat{\theta}_t\}$ one has to be sure that the "true" value θ of the parameter is in principle identifiable. For instance if two different values θ_0 and θ'_0 give rise to the same observed process one cannot expect to get convergence of the $\{\hat{\theta}_t\}$ to θ_0. A possible way out is to redefine the parameter space such that no two different values of the parameter induce the same observed process and to be sure that all the $\hat{\theta}_t$ belong to this new parameter space.

1.4 Realization

Realization problems play a prominent role in stochastic system theory. In the context of counting process systems the basic question is the following. Given

a counting process $\{N_t\}$, can we view it as the output of a stochastic system? In the light of the discussion in section 1.2 this amounts to reformulating this question as follows. Given a counting process $\{N_t\}$ with intensity process $\{\lambda_t\}$, does there exist a Markov process $\{X_t\}$ with a small tractable state space such that λ_t is a function of t and X_t, so $\lambda_t = f(t, X_t)$? If the answer is affirmative then one says that the system (X,N) is a realization of $\{N_t\}$. Suppose now that we have a realization (X,N) where X is a finite state process such that (1.5) holds. A second question is then whether this realization is minimal. Minimality here means that we cannot find another Markov process $\bar{X}$ and a function $\bar{f}$ such that $\lambda_t = \bar{f}(\bar{X}_t)$ and such that the state space of $\bar{X}$ contains fewer elements than that of X. In order to answer this second question we need a characterization of minimality. It will be given in chapter 5 in terms of the matrices C and A appearing in (1.5). Minimality is an important concept that also plays a role in identification problems where one only has observations of the counting process whereas the state process cannot be observed. A rather trivial example clarifies this. Suppose that one wants to identify the matrix A from the observations of the counting process and suppose that $C = [\lambda,...,\lambda]$ for some $\lambda > 0$, so that $\lambda_t = CY_t \equiv \lambda$. Then clearly any Markov process X that yields a minimal realization (X,N) is such that its space is a singleton. For Markov processes X with a larger state space such that $CY_t \equiv \lambda$, the counting process contains no information whatsoever about X. Hence one is not able to identify A on the basis of the counting process observations alone. Although this example treats a degenerate case, it indicates that minimality is a prerequisite for identification of the underlying state process. Finally we notice that a characterization of minimality can be used to define the parameter space in such a way that all its elements are identifiable.

1.5 Organization of the monograph

In this section we briefly go through the contents of this monograph. In chapter 2 the relevant results from the general theory of stochastic processes are reviewed. All results except a convergence theorem of semimartingales can be found in the literature [3,4,5,8]. For this reason most of the proofs are omitted. Chapter 3 starts with results from weak convergence theory, especially for sequences of (semi)martingales. As in chapter 2, only proofs of new results are given. The second part of this chapter is devoted to the study of likelihood ratios for counting process. The proofs that we present differ slightly from those in the references [16,24]. Also an alternative characterization of local asymptotic normality is derived. This LAN property turns out to be very helpful in designing recursive parameter estimation algorithms. This is explained in chapter 4 where we also study convergence properties of such algorithms.

The technique we use to prove almost sure convergence involves stochastic Lyapunov functions and a convergence result for nonnegative semimartingales. In chapter 5 we discuss counting process systems. In detail we treat minimality for conditionally Poisson systems and a realization problem for selfexciting systems.

Chapter 2

Background in Stochastic Processes

In this chapter we briefly summarize the theory of stochastic processes as far as we need it. Basic references for this chapter are the books by DELLACHERIE and MEYER [4,5], JACOD [8], or specialized to point processes the one by BRÉMAUD [3] and [14,41]. We do not give definitions and results of most basic notions in probability theory.

A probability space is a triple $(\Omega,\mathcal{F},P)$, where $(\Omega,\mathcal{F})$ is a measurable space and P a probability measure on it. If $(E,\mathcal{E})$ is another measurable space, then an E-valued random variable X is a measurable mapping $X:(\Omega,\mathcal{F})\to(E,\mathcal{E})$. In the sequel E will usually be a subset of some $\mathbb{R}^n$, and $\mathcal{E}$ its Borel σ-algebra. Stochastic processes to be defined below will have $[0,\infty)$ as continuous "time set". However occassionally it will be replaced by $\mathbb{N}$, as a discrete "time set".

2.1 BASIC CONCEPTS

DEFINITION 2.1.1: A *stochastic process* is a mapping $X:\Omega\times[0,\infty)\to E$ such that for all $t\geq 0$ the mapping $X(\cdot,t):\Omega\to E$ is a random variable. The map $X(\omega,\cdot):[0,\infty)\to E$ is said to be a trajectory or path of X. For $X(\cdot,t)$ we will often write X_t.

Given a stochastic process one can form its associated family of finite dimensional distributions by computing $P(x_{t_1}\in A_1,\ldots,x_{t_n}\in A_n)$ for $t_i\in[0,\infty)$ and $A_i\in\mathcal{E}(i=1,\ldots,n)$. Denote these quantities by $\Phi(t_1,\ldots,t_n;A_1,\ldots,A_n)$. Then obviously one has $\Phi(t_{\pi(1)},\ldots,t_{\pi(n)};A_{\pi(1)},\ldots,A_{\pi(n)})$ for every permutation π of $\{1,\ldots,n\}$ and

$$\Phi(t_1,\ldots,t_n;A_1,\ldots,A_{n-1},E)=\Phi(t_1,\ldots,t_{n-1};A_1,\ldots,A_{n-1}).$$

A natural question to ask is then whether these two properties plus probably some extra completely characterize in some sense a stochastic process. This question is answered by the following theorem, essentially due to Kolmogorov.

THEOREM 2.1.2: *Let the family* $\Phi=\{\Phi(t_1,\ldots,t_n;A_1,\ldots,A_n)\mid n\in\mathbb{N},t_i\in[0,\infty),A_i\in\mathcal{E}\}$ *satisfy the above permutation and consistency property. Then there exist a probability space* $(\Omega,\mathcal{F},P)$ *and a stochastic process* $X:\Omega\times[0,\infty)\to E$ *such that the set of finite dimensional distributions of* X *coincides with* Φ.

The space Ω in this theorem has to be rich enough in order to carry the desired

property. In Kolmogorov's proof of the theorem $\Omega = E^{[0,\infty)}$, and $\mathcal{F}$ is the smallest σ-algebra that makes all cylinder sets in Ω measurable. For most purposes however this space is not suitable to answer questions like: What is the probability of the set $\{\omega: X(\omega,\cdot):[0,\infty)\to E \text{ is continuous}\}$? Also it is not clear how one should define double integrals like $E\int_0^t X(s,\omega)dsP(d\omega)$. These problems can historically be accounted for by noticing that the definition of stochastic process has first been given for a countable time set like $\mathbb{N}$ or $\mathbb{Z}$. Then all the relevant measurability properties are a simple consequence of the given definition of a stochastic process. For problems like the above ones more sophisticated measurability concepts are needed.
The assertion of the theorem admits more than one choice of a probability space. In fact one should always include $(\Omega,\mathcal{F},P)$ in the definition of a stochastic process. The fact that one can choose more than one probability space motivates the following definition:

DEFINITION 2.1.3: Let $(\Omega,\mathcal{F},P,X)$ and $(\Omega',\mathcal{F}',P',X')$ be two stochastic processes, both with values in $(E,\mathcal{E})$. They are said to be *equivalent* if their families of finite dimensional distributions coincide.

Even if one has a fixed suitable probability space $(\Omega,\mathcal{F},P)$, the stochastic process X is not completely determined by its family of finite dimensional distributions. We need more precise notions that tell us in what sense two stochastic process are the same. These are given in the next two definitions.

DEFINITION 2.1.4: Let X and X' be two stochastic processes on $(\Omega,\mathcal{F},P)$
1) X' is said to be a *modification* of X if $X_t = X'_t$ a.s. for every $t \geq 0$.
2) X' is said to be *indistinguishable* from X if

$$P(\omega: X_t(\omega) = X'_t(\omega) \text{ for all } t \geq 0) = 1.$$

One clearly has that two indistinguishable processes X and X' are modifications of each other and assuming only the latter, they are also equivalent. In the sequel statements that two processes X and Y are equal $(X = Y)$ will always mean that X and Y are indistinguishable. Similarly if a process X satisfying some property is unique, it will mean that it is indistinguishable from any other process that satisfies the same property. The next two examples of a process play an important role in the theory of stochastic processes. These are the Brownian Motion and the Poisson process.

DEFINITION 2.1.5: $W:\Omega\times[0,\infty)\to\mathbb{R}$ is a *Brownian* motion if
(i) $W_0 = 0$ a.s.
(ii) W has independent increments if for $t > s > u$ $W_t - W_s$ is independent of W_u.
(iii) $W_t - W_s$ has a normal distribution with mean 0 and variance $|t-s|$.

DEFINITION 2.1.6: $N:\Omega\times[0,\infty)\to\mathbb{N}_0$ is a *Poisson* process with intensity function $\lambda:\mathbb{R}_+\to\mathbb{R}_+$ which is locally integrable, if

i) $N_0=0$ a.s.
ii) N has independent increments
iii) N_t-N_s has a Poisson distribution with parameter $\int_s^t \lambda(u)du$

If $\lambda(t)=\lambda,\forall t$, then N is called a homogeneous Poisson process with intensity parameter λ. Later on we will define a generalization of a Poisson process. Although there are striking similarities between the Wiener process and the Poisson process, the differences are at least as important.

PROPOSITION 2.1.7:
1) *There exists a modification of the Brownian motion such that its paths are continuous and of unbounded variation. It is called the Wiener process.*
2) *There exists a modification of the Poisson process such that its paths are increasing, constant between the jumps, that are all of magnitude* $+1$ *and on each finite time interval there are only finitely many jumps.*

The current framework of the theory of stochastic processes has essentially been introduced by Doob. The important concept is that of a filtration, with the interpretation of a growing information pattern.

DEFINITION 2.1.8.: Let $(\Omega,\mathcal{F})$ be a measurable space. A *filtration* $\mathbb{F}=\{\mathcal{F}_t\}_{t\geq 0}$ on Ω is a family of sub σ-algebras of $\mathcal{F}$ such that $\mathcal{F}_t\supset\mathcal{F}_s$ if $t\geq s$. A filtered measurable space is then a triple $(\Omega,\mathcal{F},\mathbb{F})$, where $\mathbb{F}$ is a filtration on $(\Omega,\mathcal{F})$.

DEFINITION 2.1.9: A stochastic process X on $(\Omega,\mathcal{F})$ is said to be $\mathbb{F}$-*adapted* if X_t is $\mathcal{F}_t$-measurable for all $t\geq 0$.
Given a stochastic process X there always exist a filtration to which it is adapted. Take $\mathcal{F}_t=\bigvee_{s\leq t}\sigma(X_s)$.

DEFINITION 2.1.10: Let X be a stochastic process on $(\Omega,\mathcal{F})$ and let $\mathbb{F}$ be a filtration on $(\Omega,\mathcal{F})$. X is said to be *progressively measurable* (relative to $\mathbb{F}$) if for all $t\geq 0$ the mapping $(\omega,s)\to X(\omega,s)$ is a measurable mapping from $(\Omega\times[0,t),\mathcal{F}_t\times\mathcal{B}[0,t))$ in $(E,\mathcal{E})$.

The usefulness of this definition is motivated by the following. Let X be a progressively measurable process on $(\Omega,\mathcal{F},P)$. Assume that X is (Lebesgue)integrable on $\Omega\times[0,t]$ w.r.t. the product of P and the Lebesgue measure on $[0,t]$. Then from Fubini's theorem

$$\int_0^t X(\omega,s)ds$$

is $\mathcal{F}_t$-measurable. Hence Y defined by $Y(\omega,t)=\int_0^t X(\omega,s)ds$ is again an adapted process.

A filtration $\mathbb{F}$ on a (necessarily complete) probability space $(\Omega,\mathcal{F},P)$ is said to be *complete* if $\mathcal{F}_0$ contains all P-null sets. $\mathbb{F}$ is called right continuous if

$\mathcal{F}_{t+} = \cap_{s>t} \mathcal{F}_s$ equals $\mathcal{F}_t$. If $\mathbb{F}$ is both complete and right continuous than it is said to satisfy the *usual conditions.*
We need two other concepts of measurability.

DEFINITION 2.1.11: A stochastic process X is said to be *cadlag* if all its trajectories are right continuous and have left limits at all points.

Note: cadlag is a french abbreviation and stands for continue à droite et pourvue de limites à gauche finies.

DEFINITION 2.1.12: The *optional* σ-algebra $\mathcal{O}$ on $\Omega \times [0,\infty)$ is defined as the σ-algebra generated by all $\mathbb{F}$-adapted processes that are cadlag. A stochastic process X is said to be *optional* if it is measurable w.r.t. $\mathcal{O}$ as a function of both t and ω.

The *predictable* σ-algebra $\mathcal{P}$ on $\Omega \times [0,\infty)$ is defined as the σ-algebra generated by all $\mathbb{F}$-adapted processes that have left continuous trajectories. A stochastic process X is *predictable* if it is measurable as a function on $(0,\infty) \times \Omega$ w.r.t. $\mathcal{P}$.

The following sequence of implications holds: X is predictable $\Rightarrow X$ is optional. If X is moreover right continuous or left continuous, then it is also progressively measurable.

EXAMPLE: Let X be an adapted cadlag process. Define the process X_- by $(X_-)_t = X_{t-} = \lim_{s \uparrow t} X_s$. Then X_- is a predictable process.

Any cadlag process X is optional and by ΔX we mean the process with $\Delta X_t = X_t - X_{t-}$. Hence also ΔX is optional.

REMARK: All measurability concepts introduced above, from predictability to adaptedness hold with respect to the given filtration $\mathbb{F}$. Mostly it will be clear with what filtration we work. However if confusion may arise we will speak of $\mathbb{F}$-predictable process etc.

PROPOSITION 2.1.13: *Let X be a stochastic process such that almost all its paths admit left and right hand limits. Then almost all paths have only countable many points of discontinuity.*

The next concept that will be introduced is that of a stopping time. We will need it only occasionally.

DEFINITION 2.1.14: A *stopping time* is a random variable $T: \Omega \to [0,\infty]$ such that $\{T \leq t\} \in \mathcal{F}_t$ for all t.

PROPOSITION 2.1.15:
1) *If T is a stopping time, then $\{T < t\} \in \mathcal{F}_t$.*
2) *If T is a random variable such that $\{T < t\} \in \mathcal{F}_t$ and if the filtration is right continuous, then T is a stopping time.*

DEFINITION 2.1.16: The *pre-T σ-algebra* $\mathcal{F}_T$ is defined as $\mathcal{F}_T=\{A\in\mathcal{F}_\infty : A\cap\{T\leqslant t\}\in\mathcal{F}_t,\ \forall t\}$

PROPOSITION 2.1.17: *If S and T are two stopping times such that* $S\leqslant T$, *then* $\mathcal{F}_S\subset\mathcal{F}_T$.

We close this section with

DEFINITION 2.1.18: A stochastic process X is *uniformly integrable* if

$$\lim_{k\to\infty}\sup_{t\geqslant 0} E|X_t|1_{\{|X_t|\geqslant k\}}=0$$

EXAMPLE:

1) If X is bounded in $L^{1+\alpha}$ ($\alpha>0$), that is $\sup_{t\geqslant 0}E|X_t|^{1+\alpha}<\infty$. then X is uniformly integrable.
2) Let l be a random variable, $E|l|<\infty$. Define $X_t=E[l|\mathcal{F}_t]$, where $\{\mathcal{F}_t\}_{t\geqslant 0}$ is a filtration on $(\Omega,\mathcal{F},P)$. Then $\{X_t\}_{t\geqslant 0}$ is uniformly integrable.

2.2 MARTINGALE THEORY

In this section we summarize results on some important classes of stochastic processes, namely the martingales, supermartingales and submartingales. Throughout this section we assume that a complete probability space is given, together with a filtration $\mathbb{F}$ on it. All processes are adapted to this filtration and defined on the given probability space.

DEFINITION 2.2.1:

1) A stochastic process X (adapted to $\mathbb{F}$) is a *supermartingale* if
   i) $E|X_t|<\infty, \forall t\geqslant 0$
   ii) $E[X_t|\mathcal{F}_s]\leqslant X_s$ a.s. $\forall t\geqslant s$.
2) X is a *submartingale* if $-X$ is a supermartingale
3) X is *martingale* if X is both a supermartingale and a submartingale.

EXAMPLE:

1) Let $B:\Omega\times[0,\infty]\to\mathbb{R}$ be a Brownian Motion. Let for all t $\mathcal{F}_t^B=\sigma\{B_s, s\leqslant t\}$. Then B is a martingale. Indeed since B_t-B_s is independent of $\mathcal{F}_s$ we have $E[B_t-B_s|\mathcal{F}_s]=E[B_t-B_s]=0$
2) Let $N:\Omega\times[0,\infty)\to\mathbb{N}_0$ be a homogeneous Poisson process, with intensity λ. Let $\mathcal{F}_t^N=\sigma\{N_s, s\leqslant t\}$. Then X defined by $X_t=N_t-\lambda t$ is a martingale. Indeed since N_t-N_s is independent of $\mathcal{F}_s$ we have $E[N_t-N_s|\mathcal{F}_s]=E[N_t-N_s]=\lambda(t-s)$.
3) More generally let X be a process with independent increments such that $E|X_t|<\infty$. Then as in the two previous examples the process M defined by $M_t=X_t-EX_t$ is a martingale relative to the filtration generated by X.
4) Let $X:(\Omega,\mathcal{F},P)\to\mathbb{R}$ be a random variable, $E|X|<\infty$. Let $\mathbb{F}$ be a filtration

on $(\Omega,\mathcal{F},P)$ Then M defined by $M_t=E[X|\mathcal{F}_t]$ is a martingale. Observe that M is uniformly integrable.

It turns out that the trajectories of supermartingales enjoy desirable properties. Specifically one may almost always assume that they have cadlag paths. This statement will be made precise in the next sequence of theorems.

THEOREM 2.2.2: *Let $(\Omega,\mathcal{F},\mathbb{F},P)$ be a filtered probability space. Let D be a countable dense subset of $[0,\infty)$. Let X be a supermartingale on $(\Omega,\mathcal{F},\mathbb{F},P)$*

1) *Then for almost all ω $\lim_{\substack{s\downarrow\downarrow t\\ s\in D}} X_s(\omega)$ exists and is finite for all t. This limit will be denoted by $X_{t+}(\omega)$.*
2) *For almost all ω $\lim_{\substack{s\uparrow\uparrow t\\ s\in D}} X_s(\omega)$ exists and is finite for all $t\geq 0$. This limit is denoted by $X_{t-}(\omega)$.*
3) *The process $X_+=\{X_{t+}\}_{t\geq 0}$ is a supermartingale with respect to the filtration $\mathbb{F}^+=\{\mathcal{F}_{t+}\}_{t\geq 0}$.*
4) *The process $X_-=\{X_{t-}\}_{t>0}$, is a supermartingale with respect to the filtration $\mathbb{F}^-=\{\mathcal{F}_{t-}\}_{t>0}$. Here $\mathcal{F}_{t-}=\bigvee_{s<t}\mathcal{F}_s$.*

THEOREM 2.2.3: *Let X be a supermartingale on $(\Omega,\mathcal{F},\mathbb{F},P)$. Assume that almost all its paths are right continuous. Then X is also a supermartingale with respect to the smallest filtration $\mathbb{G}$ that contains $\mathbb{F}$ and that satisfies the usual conditions.*

THEOREM 2.2.4: *Assume that the filtration $\mathbb{F}$ satisfies the usual conditions and let X be a supermartingale on $(\Omega,\mathcal{F},\mathbb{F},P)$ such that $t\mapsto EX_t$ is a right continuous function. Then X admits a modification such that all its paths are cadlag. Hence every path has only countably many discontinuities.*

Although this is usually not stated, one can also prove that there exists a modification which has left continuous paths on $(0,\infty)$. In fact one then takes X^-.

In view of these results we will from now on, unless stated otherwise, assume that the filtrations satisfy the usual conditions and that supermartingales have cadlag paths. Martingales and supermartingales are especially interesting processes because under suitable conditions they converge almost surely as $t\to\infty$.

THEOREM 2.2.5: *Let X be a right continuous supermartingale. Suppose that $\sup\{E|X_t|,\ t\geq 0\}<\infty$ (or equivalently that $\lim_{t\to\infty} EX_t 1_{\{X_t<0\}}>-\infty$). Then $l=\lim_{t\to\infty} X_t$ exists and is an integrable random variable. If X is uniformly integrable, then X is right closed by its limit l, which means that $X_t\geq E[l|\mathcal{F}_t]$.*

REMARK: The same statement holds for left continuous supermartingales.

REMARK: As we have seen before X defined by $X_t = E[Y|\mathcal{F}_t]$ is a uniformly integrable martingale if $E|Y| < \infty$. The theorem tells us that the converse also holds. If X is a uniformly integrable martingale, then necessarily X is of the form $X_t = E[l|\mathcal{F}_t]$.

DEFINITION 2.2.6: A martingale X is called *square integrable* if $\sup_{t>0} EX_t^2 < \infty$.
One can also say that such a martingale is bounded in L^2. So in particular square integrable martingales are uniformly integrable and their limit for $t \to \infty$ exists.

In a number of cases the requirement that a process is a uniformly integrable martingale is too restictive. Therefore we introduce a wider class of processes, that of local martingales.

DEFINITION 2.2.7: A right continuous stochastic process X on $(\Omega, \mathcal{F}, \mathbf{F}, P)$ is said to be a *local martingale* if there exists an increasing sequence of stopping times $\{T_n\}$ with $\lim_{n\to\infty} T_n = \infty$ *a.s.* such that the process $\{X_{t\wedge T_n} 1_{\{T_n>0\}}\}_{t\geq 0}$ is a uniformly integrable martingale for each n. $\{T_n\}$ is called a *fundamental sequence* for X.

In order to see that local martingales are extension of martingales we mention the following result.

THEOREM 2.2.8: *Let X be a (super)martingale, right continuous. Let T be a stopping time. Then the stopped process $X^T (= \{X_{t\wedge T}\}_{t\geq 0})$ is again a (super)martingale.*

REMARK:

1) Let X be a right continuous martingale. Let $T_n \equiv n$. Then $X_t^{T_n} = X_{t\wedge n} = E[X_n|\mathcal{F}_t]$ for all t. Hence X^{T_n} is a uniformly integrable martingale and X is a local martingale.
2) X is a local martingale iff for the fundamental sequence $\{T_n\}$ for X we have that
   i) $X_0 1_{\{T_n>0\}}$ is integrable
   ii) $\{X_{t\wedge T_n} - X_0\}$ is a uniformly integrable martingale.
3) In fact the additional requirement that the stopped processes in this definition are uniformly integrable is superfluous. Indeed if we merely demand that $\{X_{t\wedge T_n} - X_0\}_{t\geq 0}$ is a martingale for each n, $\{X_{t\wedge T_n\wedge n} - X_o\}_{t\geq 0}$ is a uniformly integrable martingale for each n. Hence taking $\{n \wedge T_n\}$ as a fundamental sequence, then we see that indeed X is a local martingale.

DEFINITION 2.2.9: X is called a *locally square integrable* martingale if there exists an increasing sequence of stopping times $\{T_n\}$, $T_n \uparrow \infty$ such that $\{X_{t\wedge T_n} - X_0\}_{t\geq 0}$ is a square integrable martingale.

EXAMPLE: Every continuous martingale X is locally square integrable. Take for $T_n = \inf\{t>0:|X_t| \geq n\}$.

PROPOSITION 2.2.10: *Let X be a nonnegative local martingale and assume that $EX_0 < \infty$. Then X is a supermartingale, as follows from Fatou's lemma.*

DEFINITION 2.2.11: A stochastic process A on $(\Omega, \mathcal{F}, \mathbf{F}, P)$ is called *increasing* if its trajectories are right continuous, finite, increasing functions on $[0,\infty)$. It is called *integrable* if $\lim_{t\to\infty} EA_t < \infty$. The difference of two increasing processes is called a process of *finite variation.* If a process is the difference of two integrable increasing processes, then it is called a process of *integrable variation.*

PROPOSITION 2.2.12: *Let A be an increasing process. Then there exists a unique increasing process A^c, a sequence of stopping times $\{T_n\}$ and a sequence op positive constants $\{\lambda_n\}$ such that*

$$A_t = A_t^c + \sum_n \lambda_n 1_{\{T_n \leq t\}}$$

A^c is called the continuous parts of A, $A^d = A - A^c$ is the jump part of A. For processes of finite variation, being the difference of two increasing processes, a similar result holds.

DEFINITION 2.2.13: An increasing process is called locally integrable if there exists a sequence of stopping times $\{T_n\}$ such that $E[A_{T_n} - A_0] < \infty$ for each n and if A_0 is finite a.s.

Observe that for a locally integrable increasing process A and a nonnegative measurable process X, one can define $E\int_0^\infty X_s dA_s$ as a Lebesgue integral, which also yields $\int_0^t X_s dA_s$ adaptive.

PROPOSITION 2.2.14: *Let A be an increasing process, locally integrable. Then there exists a unique predictable increasing process B such that A-B is a local martingale. B is called the dual predictable projection of A, or the compensator of A.*

PROPOSITION 2.2.15: *Let A be an increasing locally integrable process. Then a predictable process B is its compensator if and only if $E\int_0^\infty X_s dA_s = E\int_0^\infty X_s dB_s$ for all nonnegative predictable processes X.*

EXAMPLE: Let N be a standard Poisson process. N is certainly locally integrable, which is immediately seen by taking $T_k = \inf\{t>0:N_t \geq k\}$. Since we have already seen that $N_t - t$ is a martingale, the process B with $B_t \equiv t$ is the compensator of N.
The dual predictable projection plays an important role in proving the decomposition theorem for supermartingales.

DECOMPOSITION OF SUPERMARTINGALES

Let A be an increasing process and let $EA_t<\infty$, for all t. Let M be a martingale. Then X defined by $X_t=M_t-A_t$ is a supermartingale. Conversely, given a supermartingale X can one always decompose it in the form $X_t=M_t-A_t$ above, and if this is the case, for a predictable process A this decomposition is unique. Clearly the latter is not the case if one considers the supermartingale $X_t=-n_t$, where n is a Poisson process with $En_t=\lambda t$. Now one can take $A_t=-n_t$ and $M_t=0$, but also one can take $A_t^1=-\lambda t$ and $M_t^1=\lambda t-n_t$, if the underlying filtration is given by $\mathcal{F}_t=\sigma\{n_s,s\leqslant t\}$. The difference between the two decompositions is that A^1 is a predictable process, whereas A is not. The surprising result is that one can always uniquely decompose a supermartingale as a difference of a martingale and a predictable increasing process. In the discrete time case this is almost trivial. Take A defined by

$$A_n-A_{n-1}=X_{n-1}-E(X_n|\mathcal{F}_{n-1}),\ A_0=0, \text{ and define } M_n=X_n+A_n$$

This is called the Doob-decomposition of a supermartingale. In continuous time it requires much more sophisticated techniques to prove a similar result, which is due to Meyer and is known as the Doob-Meyer decomposition of supermartingales.

THEOREM 2.2.16: *Let X be a supermartingale. Then X admits a unique decomposition $X=M-A$, where M is a local martingale and A a predictable increasing process, $A_0=0$. Moreover A is an integrable process if* $\lim_{t\to\infty}EX_t>-\infty$.

It should be noted that the process M and A depend on the given filtration. For instance in the Poisson process case $A_t=-\lambda t$ if we take $\mathcal{F}_t=\sigma\{n_s,s\leqslant t\}$. But $A_t=-n_t$ if we would take the (deterministic) filtration $\mathcal{F}_t=\mathcal{F}$ for all t. Now we are in the position to define a class of stochastic processes, which is closed under many operations.

DEFINITION 2.2.17: A stochastic process X on $(\Omega,\mathcal{F},\mathbf{F},P)$ is called a *semimartingale* if there exists a decomposition $X_t=X_0+A_t+M_t$, where M is a local martingale, $M_0=0$, A is a process of finite variation, $A_0=0$. A semimartingale is called *special* if there exists a decomposition such that A is predictable.

REMARK: The decomposition of a special semimartingale is unique. Indeed if $X_t-X_0=A_t+M_t=A_t^1+M_t^1$, then $M_t-M_t^1$ is predictable local martingale of finite variation and zero for $t=0$. Hence $M_t=M_t^1\equiv 0$.

An example of a special semi martingale is a supermartingale, according to the Doob-Meyer decomposition (theorem 2.2.16).

DEFINITION 2.2.18: A semimartingale X is said to be *locally square integrable* if the process Y defined by $Y_t=\sup\{|X_s-X_0|^2,s\leqslant t\}$ is locally integrable.

LEMMA 2.2.19:

i) *A locally square integrable semimartingale is special.*

ii) *A semimartingale is locally square integrable iff M is a locally square integrable martingale, where M is the local martingale of the canonical decomposition $X_t = X_0 + A_t + M_t$.*

Next we state and prove a convergence theorem for semimartingales, which is new and that will be used in subsequent chapters.

THEOREM 2.2.20: *Let X be a stochastic process such that $X_t = X_0 + A_t - B_t + M_t$. Here A and B are increasing processes with $A_0 = B_0 = 0$ and $\lim_{t\to\infty} A_t < \infty$ a.s. and M is a local martingale with $M_0 = 0$. Assume that $\inf\{X_t : t \geq 0\} > -\infty$ a.s. Then both $\lim_{t\to\infty} X_t$ and $\lim_{t\to\infty} B_t$ exist and are finite.*

PROOF: Assume without loss of generality that $X_0 = 0$. Observe that $\lim_{t\to\infty} X_t$ exists and is finite if and only if the same holds for $\lim_{t\to\infty} X_{t-}$. We will prove the latter. Let $\{T_n\}$ be a fundamental sequence for M. Define also another sequence of stopping times $R_k = \inf\{t > 0 : X_t^- + A_t > k\}$. Observe that $\{R_k\}$ is an increasing sequence with $\lim R_k = \infty$. Observe also that $\{R_k = \infty\} \uparrow \Omega$. Now for each n $\{M_{t\wedge T_n}\}_{t\geq 0}$ is a uniformly integrable martingale. But then also $\{M_{t\wedge T_n -}\}_{t>0}$ is a (uniformly integrable) martingale with respect to $\{\mathcal{F}_{t-}\}_{t>0}$, and the same is true for $\{M_{t\wedge R_k \wedge T_n -}\}_{t>0}$ for each k,n. Observe also that

$$\begin{aligned} M_{t\wedge R_k \wedge T_n -} &\geq X_{t\wedge R_k \wedge T_n -} - A_{t\wedge R_k \wedge T_n -} \\ &\geq -(X_{t\wedge R_k \wedge T_n -}^- + A_{t\wedge R_k \wedge T_n -}). \end{aligned}$$

Hence $M_{t\wedge R_k \wedge T_n -}^- \leq k$. So for fixed t and k $\{M_{t\wedge R_k \wedge T_n -}^-\}_{n\geq 0}$ is uniformly integrable. Then

$$\begin{aligned} E[M_{t\wedge R_k -} | \mathcal{F}_{s-}] &= E[\lim_{n\to\infty} M_{t\wedge R_k \wedge T_n -} | \mathcal{F}_{s-}] \\ &\leq \liminf_{n\to\infty} E[M_{t\wedge R_k \wedge T_n -} | \mathcal{F}_{s-}] \\ &= \liminf_{n\to\infty} M_{s\wedge R_k \wedge T_n -} = M_{s\wedge R_k -} \end{aligned}$$

The last inequality follows from Fatou's lemma for uniformly integrable families of random variables. So $\{M_{t\wedge R_k -}\}_{t\geq 0}$ is a supermartingale and moreover $\{M_{t\wedge R_k -}^-\}_{t\geq 0}$ is uniformly integrable since also $M_{t\wedge R_k -}^- \leq k$. Hence from the convergence theorem for supermartingales $\lim_{t\to\infty} M_{t\wedge R_k -}$ exists and is finite. But then also $\lim_{t\to\infty}(X_{t\wedge R_k -} + B_{t\wedge R_k -}) = \lim_{t\to\infty}(A_{t\wedge R_k -} + M_{t\wedge R_k -})$ exists and is finite.

Hence $\lim_{t\to\infty} X_{t\wedge R_k -}$ and $\lim_{t\to\infty} B_{t\wedge R_k -}$ exist and are finite because the latter limit always exists and cannot be infinity because of $\inf_{t\geq 0} X_t > -\infty$. Consequently $\lim_{t\to\infty} X_{t\wedge R_k -}$ also exists and is finite a.s. on $\{R_k = \infty\}$ and on

this set it equals $\lim_{t\to\infty} X_{t-}$. But as noticed before $\{R_k=\infty\}$ increases to Ω, which finishes the proof. □

REMARK: This theorem generalizes a result in [39], in the sense that we do not require X to be nonnegative, nor do we require that the jumps of A ae bounded. A similar result in discrete time can be found in [35]. We will apply this theorem in chapter 4 (cf. lemma 4.1.1.1), in order to prove almost sure convergence of a family of parameter estimators. However, there X_t will be nonnegative for all t.

DEFINITION 2.2.21: Let M be a square integrable martingale. Then $X=M^2$ is a submartingale in view of Jensen's inequality. Applying the Doob-Meyer decomposition theorem, we see that there exists a unique predictable increasing process, which we denote by $<M,M>$, such that $M^2-<M,M>$ is again a martingale. $<M,M>$ is called the *predictable variation process* of M. This definition easily extends to locally square integrable martingales M.

EXAMPLE: Let N be a Poisson process with parameter λ. Then one easily calculates that $<M,M>_t=\lambda t$.

DEFINITION 2.2.22: For two locally square integrable martingales X and Y one defines the predictable covariation process $<X,Y>$ via the polarization formula:

$$<X,Y> = \frac{1}{2}[<X+Y,X+Y>-<X,X>-<Y,Y>].$$

Then $XY-<X,Y>$ is a martingale.

REMARK: $<X,Y>=0$ iff XY is martingale. $<X,Y>$ is a process of bounded variation.

Clearly if one drops in the last two definitions the requirement that X and Y are locally square integrable, then there is no reason why $<X,X>$ or $<X,Y>$ should exist. However one can define another "variation"-process in this case.

DEFINITION 2.2.23: Let X be a local martingale. Then there exist a unique optional increasing process, denoted by $[X,X]$, such that
i) $X^2-[X,X]$ is a local martingale
ii) $\Delta[X,X]=(\Delta X)^2$.

For two local martingales X and Y, $[X,Y]$ is defined as $[X,Y]=$ $\frac{1}{2}\{[X+Y,X+Y]-[X,X]-[Y,Y]\}$. Hence $\Delta[X,Y]=\Delta[Y,X]=\Delta X\cdot\Delta Y$ and $XY-[X,Y]$ is a local martingale.

The processes $[X,Y],[X,X]$ are called the *optional covariation* and *optional*

variation processes. It is obvious (because of the possible jumps) that they are in general not predictable.

DEFINITION 2.2.24: Let X be a local martingale.
1) If $[X,X]$ is a purely discontinuous increasing process, we say that X is a *purely discontinuous local martingale*
2) If X has continuous paths, then we say that X is a *continuous local* martingale.

One can prove the following theorem.

THEOREM 2.2.25: *Let X be a local martingale. Then there exists a unique decomposition $X=X^c+X^d$, where X^c is a continuous local martingale and X^d is a purely discontinuous local martingale.*

REMARK: Different from the case where we considered increasing processes, the paths of X^d are not necessarily piecewise constant. With the aid of this theorem one can prove the following relation for any local martingale X:

$$[X,X]_t = <X^c,X^c>_t + \sum_{s\leq t}(\Delta X_s)^2.$$

Notice that $<X^c,X^c>$ is well defined since any continuous local martingale is locally square integrable. Notice also that $[X,X]^c=<X^c,X^c>$.

Furthermore if X is locally square integrable martingale then both $[X,X]$ and $<X,X>$ exist and $[X,X]-<X,X>=([X,X]-X^2)+(X^2-<X,X>)$ is a local martingale. Stated otherwise, in this case $<X,X>$ is the dual predictable projection of $[X,X]$.

EXAMPLE: Let N be a Poisson process with $EN_t=\lambda t$ and $M_t=N_t-\lambda t$. Then $\Delta[M,M]_t=(\Delta M_t)^2=\Delta N_t$. Hence $[M,M]=N$.

The next thing that we want to do is to define the process $[X,X]$ for a semimartingale X.

DEFINITION 2.2.26: Let X be a semimartingale and let $X_t=X_0+A_t+M_t$ be a decomposition. In this case the process $[X,X]$ is defined as

$$[X,X]_t = <M^c,M^c>_t + \sum_{s\leq t}(\Delta X_s)^2$$

REMARK: Although the decomposition of X in definition 2.2.26 is not unique, the continuous part M^c of M is. This can be shown in the same way as proving the uniqueness of the decomposition of a special semimartingale.

EXAMPLE: Let N be a Poisson process, then $[N,N]=N$.

We close this section by giving some rules that enable us to compute $[X,Y]$ for arbitrary semimartingales X and Y. If X is a local martingale and Y is a predictable process of finite variation, then $[X,Y]_t=\Sigma_{s\leq t}\Delta X_s\Delta Y_s$. In particular if Y is continuous, then $[X,Y]=0$. The same is true if X is a continuous local martingale. If both X and Y are semimartingales of finite variation, then again $[X,Y]_t=\Sigma_{s\leq t}\Delta X_s\Delta Y_s$.

2.3 Stochastic integration

Stochastic integration theory originated by considering integrals of the form $\int_0^t f_s dW_s$, where W is a Brownian motion. Then clearly a path by path definition as a Stieltjes integral is not possible, since the paths of Brownian motion are not of bounded variation. Itô has resolved this problem by defining a stochastic integral as an isometry between two Hilbert spaces, which is, although rather hidden in the definition that we will give in the sequel, still the core of stochastic integration theory.

Our goal will be to construct stochastic integrals of the form $\int_0^t H_s dX_s$, where H is a predictable process and X a semimartingale. Let $\mathcal{E}$ be the set of simple predictable processes, that is $H\in\mathcal{E}$ if and only if there exists an increasing sequence of stopping times $\{T_n\}, T_n\uparrow\infty$ such that $H_t=\Sigma_n H_n 1_{\{T_n<t\leq T_{n+1}\}}$, where H_n is $\mathcal{F}_{T_n}$-measurable. Let X be a semimartingale. Define the stochastic integral of H with respect to X to be the stochastic process $H\cdot X$ with

$$(H\cdot X)_t=\sum_n H_{n-1}(X_{T_n\wedge t}-X_{T_{n-1}\wedge t}).$$

Then the following theorem holds:

THEOREM 2.3.1: *Let X be a semimartingale. the map $H\mapsto H\cdot X$ on $\mathcal{E}$ admits a unique extension as a linear map on the space of all predictable locally bounded processes H. This extended map is also called the stochastic integral of H with respect to X. $H\cdot X$ is a stochastic process with cadlag paths and we will write $\int_0^t H_s dX_s$ for $(H\cdot X)_t$.*

THEOREM 2.3.2: *The stochastic integral is a stochastic processes satisfying the following properties*

a) *$H\cdot X$ is a semimartingale*
b) *For all locally bounded predictable processes H and K we have $K\cdot(H\cdot X)=(KH)\cdot X$.*
c) *The jumps of $H\cdot X$ are given by $\Delta(H\cdot X)=H\Delta X$*
d) *If X is a local martingale, then $H\cdot X$ is one too.*
e) *If X is of finite variation, then $H\cdot X$ is also of finite variation.*
f) *If H is left continuous and of finite variation and X is of finite variation then $H\cdot X$ is indistinguishable form an ordinary Stieltjes integral, that can be calculated path by path.*

REMARK: The last property (f) is of interest in the following chapters, where all relevant processes are of finite variation. Observe however that this

property is not evident from the definition of the stochastic integral as a linear map. This definition a priori gives no recipe for calculating $(H\cdot X)_t(\omega)$ for each $\omega\in\Omega$.

The next results will be often used.

THEOREM 2.3.3: *Let X and Y be two semimartingales, H and K locally bounded processes. Then $[H\cdot X, K\cdot Y]=HK\cdot[X,Y]$.*

If moreover the relevant predictable processes exist then also $<H\cdot X, K\cdot Y>=HK\cdot<X,Y>$.

THEOREM 2.3.4: *Let X and Y be two semimartingales. Then the stochastic integrals $X_-\cdot Y$ and $Y_-\cdot X$ are well defined and*

$$X_tY_t = X_0Y_0 + \int_0^t X_{s-}dY_s + \int_0^t Y_{s-}dX_s + [X,Y]_t.$$

This formula will often be used in differential form:

$$d(X_tY_t) = X_{t-}dY_t + Y_{t-}dX_t + d[X,Y]_t$$

and is known as the product formula of stochastic calculus. In theorem 2.3.5 we will encounter a more general chain rule for "differentiation" of stochastic processes. This theorem is the most important rule of stochastic calculus.

THEOREM 2.3.5: *(stochastic differentiation rule): Let $X=(X^1,\dots,X^n)$ where $X^1,\dots,X^n$ are real valued semi martingales. Let $F\in C^2(\mathbb{R}^n,\mathbb{R})$. Denote by D_i differentiation with respect to the i-th component. Then $F(X)$ is a semimartingale and*

$$F(X_t) = F(X_0) + \int_0^t \sum_{i=1}^n D_iF(X_{s-})dX_s^i + \frac{1}{2}\int_0^t \sum_{i,j} D_iD_jF(X_{s-})d[X^i,X^j]_s^c$$

$$+ \sum_{s\le t}[F(X_s)-F(X_{s-}) - \sum D_iF(X_{s-})\Delta X_s^i]$$

EXAMPLE: If one defines $F(x,y)=xy$, then theorem 2.3.5 yields the statements of theorem 2.3.4.

As an application of the stochastic calculus rule we have the following proposition, originally due to Lévy.

PROPOSITION 2.3.6: *Let X be a martingale with values in $\mathbb{R}$. X is assumed to have continuous paths $X_0=0$ and $[X,X]_t=t$. Then X is a Wiener process.*

PROOF: Let $u\in\mathbb{R}$. Consider e^{iuX_t}. From theorem 2.2.4, applied to complex

valued processes:

$$e^{iuX_t} = 1 + \int_0^t e^{iuX_s} iudX_s - \frac{1}{2}u^2 \int_0^t e^{iuX_s} ds$$

Hence for $v>0$:

$$e^{iu(X_{t+v}-X)} = 1 + \int_t^{t+v} e^{iu(X_s-X_t)} iudX_s - \frac{1}{2}u^2 \int_t^{t+v} e^{iu(X_s-X_t)} ds.$$

Take conditional expectations, conditional on $\mathcal{F}_t$. Then

$$E[e^{iu(X_{t+v}-X_t)}|\mathcal{F}_t] = 1 + E[-\frac{1}{2}u^2 \int_t^{t+v} e^{iu(X_s-X_t)} ds|\mathcal{F}_t].$$

Let

$$\Phi(s,t) = E[e^{iu(X_s-X_t)}|\mathcal{F}_t] \text{ for } s \geq t, \quad \Phi(t,t) = 1.$$

Then the above equation yields

$$\Phi(t+v,t) = 1 - \frac{1}{2}u^2 \int_t^{t+v} \Phi(s,t) ds$$

or $\Phi(t+v,t) = \exp(-\frac{1}{2}u^2 v)$, not depending on $\mathcal{F}_t$. This shows that $X_{t+v} - X_t$ is independent of $\mathcal{F}_t$ and that it has a normal distribution with zero mean and variance v. □

2.4 Point Processes

Definition 2.4.1: A *point process* is a sequence of random variables $T_n : \Omega \to \mathbb{R}_+$ such that $T_{n+1} > T_n$ on $\{T_n < \infty\}$. The point process is called non explosive if $T_\infty = \lim_{n\to\infty} T_n = \infty$. With a point process one can associate a counting process $N : \Omega \times [0,\infty) \to \mathbb{N}_0$ by $N_t = \Sigma_{n\in\mathbb{N}} 1_{\{T_n \leq t\}}$. The point process is then non explosive iff $N_t < \infty$ a.s. for all t. It is called *integrable* if $EN_t < \infty, \forall t$. The T_n are the jump times of N. We always assume that N is non explosive.

In order to fit this general definition into the framework that we have used before we assume to be given a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$ and we assume that N is adapted to $\mathbb{F}$, or equivalently that all T_n are stopping times. Since evidently N is an increasing locally integrable process its compensator A exists, hence $m = N - A$ is a local martingale. Observe that this result also follows from the Doob-Meyer decomposition theorem if N is integrable. From now on we will assume that the compensator A of N is absolutely continuous with respect to Lebesgue-measure. That is: there exists a nonnegative progressively measurable process λ such that $A_t = \int_0^t \lambda_s ds$. This process is called the *intensity process.* Of course from the definition of a compensator we

have that a progessively measurable process $\lambda:\Omega\times[0,\infty)\to\mathbb{R}_+$ is an intensity of N if and only if $E\int_0^\infty X_s dN_s = E\int_0^\infty X_s\lambda_s ds$, for all nonnegative predictable processes X. Observe that we wrote an intensity λ in the preceding sentence. Clearly if one changes $\lambda(\omega)$ on a set of Lebesgue measure zero, the integral $E\int_0^\infty X_s\lambda_s ds$ is left unchanged. However we have uniqueness of the intensity in the sense of the next proposition.

PROPOSITION 2.4.2: *Let N be a counting process, and assume that it admits an intensity process. Then there also exists a predictable intensity process λ. Moreover this λ is unique in the following sense. If $\tilde{\lambda}$ is another predictable intensity process, then λ and $\tilde{\lambda}$ coincide both $P(d\omega)dN_t(\omega)$ a.e. and $\lambda_t(\omega)dt$ a.e. Furthermore $\lambda_{T_n}>0$ a.s. on $\{T_n<\infty\}$.*

EXAMPLE: Let $\{S_k\}$ be an i.i.d. sequence of positive random variables with $P(S_k\geq t)=e^{-\lambda t}$ for some $\lambda>0$. Define $T_n=\Sigma_1^n S_k$. Then $\{T_n\}$ is a point process and its associated counting process is the Poisson process with parameter λ. As we have seen before $N_t-\lambda t$ is a martingale. Therefore λ is the intensity of N as just defined. This example more or less indicates that there is a connection between the distributions of the T_n (or the S_k) and the intensity of the counting process. We have the following precise statement that relates the compensator of N to the conditional distributions of the T_n.

PROPOSITION 2.4.3: *Let N be a counting process with jump times $\{T_n\}$ and compensator A. Then on $\{T_n\leq t<T_{n+1}\}$*

$$A_t = A_{T_n} + \int_0^{t-T_n} \frac{F_n(dx)}{F_n[x,\infty)}$$

where F_n is the conditional distribution function of $T_{n+1}-T_n$ given $\mathcal{F}_{T_n}$ and $F_n[x,\infty)=1-F_n(x)$.

But now it is immediately clear when a counting process admits an intensity. This is the case if and only if the conditional distributions F_n are absolutely continuous with respect to the Lebesgue measure. And in that case we have for the predictable intensity λ of N.

$$\lambda_t 1_{\{T_n<t\leq T_{n+1}\}} = \frac{f_n(t-T_n)}{F_n[t-T_n,\infty)}\, 1_{\{T_n<t\leq T_{n+1}\}}.$$

where $F_n(a,b]=\int_a^b f_n(x)dx$. Although it is in general difficult to characterize pre-T σ-algebras $\mathcal{F}_T$, it is relatively easy in the counting process case for $T=T_n$. One can prove that $\mathcal{F}_{T_n}=\sigma\{T_1,\ldots,T_n\}$.

Similarly to Levy's characterization of Brownian motion we have Watanabe's characterization of the (inhomogenous) Poisson process.

PROPOSITION 2.4.4: *Let N be a counting process and $\lambda:[0,\infty)\to\mathbb{R}^+$ a nonnegative locally integrable function such that $N_t-\int_0^t\lambda(s)ds=m_t$ is a martingale. Then*

n is a Poisson process.

PROOF: We also prove this using theorem 2.3.4. Application of this theorem gives

$$\begin{aligned} e^{iuN_t} &= 1 + iu\int_0^t e^{iuN_{s-}}\,dN_s + \sum_{s\le t}[e^{iuN_s} - e^{iuN_{s-}} - iue^{iuN_{s-}}] \\ &= 1 + \int_0^t e^{iuN_{s-}}(e^{iu}-1)dN_s \\ &= 1 + \int_0^t e^{iuN_s}(e^{iu}-1)\lambda(s)ds + \int_0^t e^{iuN_{s-}}(e^{iu}-1)dm_s \end{aligned}$$

Hence for $v \geqslant 0$:

$$E[e^{iu(N_{t+v}-N_t)}|\mathcal{F}_t] = 1 + (e^{iu}-1)E[\int_t^{t+v} e^{iu(N_s-N_t)}\lambda(s)ds|\mathcal{F}_t].$$

As in the proof of proposition 2.3.6 we deduce

$$E[e^{iu(N_{t+v}-N_t)}|\mathcal{F}_t] = \exp((e^{iu}-1)\int_t^{t+v}\lambda(s)ds),$$

which is the characteristic function of a Poisson random variable with parameter $\int_t^{t+v}\lambda(s)ds$ and which also shows that N is a process of independent increments. □

2.5 RANDOM MEASURES

The results of this section have been taken from JACOD & SHIRYAYEV [9]. Let E be an auxiliary space, with separable σ-algebra $\mathcal{E}$. In fact E will always be a subset of some $\mathbf{R}^d$ and $\mathcal{E}$ its Borel σ-algebra.

DEFINITION 2.5.1: A *random measure* on $[0,\infty)\times E$ is an indexed family $\mu=\{\mu(\omega,dt\times dx):\omega\in\Omega\}$ of positive measures on $([0,\infty)\times E, Bor[0,\infty)\times\mathcal{E})$ satisfying $\mu(\omega,\{0\}\times E)=0$ for all $\omega\in\Omega$.

Introduce $\tilde{\Omega}=\Omega\times[0,\infty)\times E$ and $\tilde{\mathcal{O}}=\mathcal{O}\times\mathcal{E}$, $\tilde{\mathcal{P}}=\mathcal{P}\times\mathcal{E}$, which are called the optional respectively predictable σ-algebra on $\tilde{\Omega}$. Functions on $\tilde{\Omega}$ are called optional, respectively predictable if they are $\tilde{\mathcal{O}}$-, respectively $\tilde{\mathcal{P}}$-, measurable. If W is an optional function on $\tilde{\Omega}$ such that for all t and all ω

$$\int_{[0,t]\times E} |W(\omega,s,x)|\mu(\omega,ds\times dx)$$

is finite, then we can define the proces $W\star\mu$ by

$$W\star\mu_t = \int_{[0,t]\times E} W(\omega,s,x)\mu(\omega,ds\times dx).$$

If $W\star\mu$ is an optional process for all optional W, then μ is called optional.

A random measure μ is called $\tilde{\mathcal{P}}-\sigma$-finite if there is a partition $\{A_n\}$ of $\tilde{\Omega}$, $A_n\in\tilde{\mathcal{P}}$ for all n such that $E(1_{A_n}\star\mu)_\infty<\infty$. We have the following result (extending proposition 2.2.4).

THEOREM 2.5.2: *Let μ be an optional $\tilde{\mathcal{P}}$- σ-finite random measure. There exists a random measure ν, called the dual predictable projection of μ, which is unique up to a P-null set with the following property:*
For all $\tilde{\mathcal{P}}$-measurable functions W with $|W|\star\mu\in A^+_{loc}$ we have $|W|\star\nu\in A^+_{loc}$ and $W\star\nu$ is the dual predictable projection of $W\star\mu$, so $W\star\mu-W\star\nu$ is a local martingale.

EXAMPLE: A rather trivial case is the following. Let A be an increasing process. Let $E=\{1\}$. Define $\mu(\omega,dt\times dx)=dA_t(\omega)\epsilon_{(1)}(dx)=dA_t(\omega)$. (Here $\epsilon_{(z)}$ is the Dirac measure at point z). In this case ν is given by the dual predictable projection $\tilde{A}$ of A: $\nu(\omega,dt\times dx)=d\tilde{A}_t(\omega)$.

DEFINITION 2.5.3: A random measure μ is called *integer valued* if
i) $\mu(\omega,\{t\}\times E)\leq 1$, $\forall\omega\in\Omega$
ii) $\mu(\cdot,A)\in\mathbb{N}_0$, for all $A\in Bor[0,\infty)\times\mathcal{E}$
iii) μ is optional and $\tilde{\mathcal{P}}$- σ-finite.

PROPOSITION 2.5.4: *Let μ be an integer valued random measure. Then there exists a sequence of stopping times $\{T_n\}$ and an E-valued optional process β such that $[\![T_n]\!]\wedge[\![T_m]\!]=\varnothing$, where $[\![T_n]\!]$ is the graph of T_n, for $m\neq n$ and with $D=\cup_n[\![T_n]\!]$*

$$\mu(\omega,dt\times dx)=\sum_{s\geq 0}1_D(\omega,s)\epsilon_{(s,\beta_s(\omega))}(ds,dx)$$

Consequently, if W is a nonnegative optional function, then

$$W\star\mu_t=\sum_{s\geq 0}W(s,\beta_s)1_D(s)=\sum_n W(T_n,\beta_{T_n})1_{\{T_n\leq t\}},$$

where we have suppressed the dependence on ω.

The important example of an integer valued random measure is the jump measure associated with a cadlag adapted process $X:[0,\infty)\times\Omega\to\mathbb{R}^d$. Here $E=\mathbb{R}^d$, $\mathcal{E}=Bor(\mathbb{R}^d)$ and $\mu=\mu^X$ is defined as

$$\mu^X(\omega,dt\times dx)=\sum_{s\geq 0}1_{\{\Delta X_s\neq 0\}}\epsilon_{(s,\Delta X_s)}(dt\times dx).$$

In the terminology of proposition 2.5.4 above $D=\{(t,\omega):\Delta X_t(\omega)\neq 0\}$ and $\beta=\Delta X$. We are mainly concerned with quasi-left continuous processes X, that is a process such that $\Delta X_T=0$, for all predictable stopping times T. In that case we have

PROPOSITION 2.5.5: *Let X be an adapted cadlag process, μ^X its jump measure. Then X is quasi-left continuous if and only if there is a version ν of the dual predictable projection of μ^X such that $\nu(\omega,\{t\}\times E)=0$, $\forall\omega,t$.*

EXAMPLE: Let N be a counting process with compensator A. Then we may take $E=\{1\}$, and

$$\mu^N(\omega,dt\times dx) = \sum_s 1_{\{\Delta N_s=1\}}\epsilon_{(s,\Delta N_s)}(dt\times dx) =$$

$$= \sum_n\sum_s 1_{\{T_n(\omega)=s\}}\epsilon_{\{T_n(\omega),1\}}(dt\times dx),$$

where $\{T_n\}$ is the sequence of jump times of N. So $\mu^N(\omega,[0,t]\times\{1\})=N_t$. The dual predictable projection ν of μ^N is given by $\nu(\omega,dt\times\{1\})=dA_t(\omega)$. It obviously follows that N is quasi left continuous iff A is continuous.

2.6 LOCAL CHARACTERISTICS OF SEMI MARTINGALES

2.6.1 Processes with Independent Increments

The local characteristics of a semi martingale, to be defined in 2.6.2 can be considered as an extension of the characteristics of a process with independent increments. We will see in section 3.1.5, that they play an important role in for instance studying weak convergence. In order to develop some feeling for what these local characteristics are, we will first briefly discuss the case of a process of independent increments. See JACOD [8] for details.

Let X be a a process of independent increments and assume that X has no fixed discontinuities, meaning that for all t $P(\Delta X_t=0)=1$, $X:\Omega\times[0,\infty)\to\mathbb{R}^d$. Let h be a truncation function, that is, $h(x)=x$ in a neighbourhood of 0, h is bounded and h has compact support. Transposed matrices are denoted by superscript T. We have the following result (Lévy-Khintchine formula)

PROPOSITION 2.6.1.1. *There exists a unique triplet (B^h,C,ν) where $B^h:[0,\infty)\to\mathbb{R}^d$ is a continuous function, $C:[0,\infty)\to\mathbb{R}^{d\times d}$ is continuous and $C_t-C_s\geq 0$ for $t\geq s$, ν is a positive measure on $[0,\infty)\times\mathbb{R}^d$ (Lévy-measure) satisfying $\nu([0,\infty)\times\{0\})=0$, $\nu(\{t\}\times\mathbb{R}^d)=0$ and $\int|x|^2\wedge 1\,\nu([0,t]\times dx)<\infty$, $\forall t\geq 0$ such that*

$$E\exp(iu^T(X_t-X_s)) = \exp\{iu^T(B^h_t-B^h_s)-\tfrac{1}{2}(u^T(C_t-C_s)u$$

$$-\int_s^t\int_{\mathbb{R}^d}(e^{iu^Tx}-1-iu^Th(x))\nu(dr\times dx)\}$$

REMARK 1: Observe that we get as a corollary that

$$\exp(iu^T X_t - (iu^T B_t^T - \frac{1}{2}u^T C_t u - \int_0^t \int_{\mathbf{R}^d} (e^{iu^T x} - 1 - iu^T h(x))\nu(dr \times dx))$$

is a martingale. A similar result will turn out to hold in the case where X is a semimartingale.

REMARK 2: If we had taken another truncation function, h' say, then the relation between $B^{h'}$ and B^h is given by $B^{h'} = B^h + (h' - h) \star \nu$.
One can show that there exists an equivalent description of the triplet (B^h, C, ν). First we need some definitions:

$$\begin{aligned} X_t^h &= \sum_{s \leq t} (\Delta X_s - h(\Delta X_s)), \\ \tilde{X}_t^h &= X_t^h - B_t^h \\ \tilde{C}_t^h &= C_t + hh^T \star \nu_t \end{aligned}$$

PROPOSITION 2.6.1.2. *Let X be a process of independent increments, h a truncation function. (B^h, C, ν) as defined above is the unique triplet satisfying*
i) $\tilde{X}^h$ *is a martingale*
ii) $\tilde{X}^h (\tilde{X}^h)^T - \tilde{C}^h$ *is a martingale*
iii) $\Sigma_{s \leq t} f(\Delta X_s) - f \star \nu_t$ *is a martingale for all bounded measurable f that are zero in a neighbourhood of zero.*

This last proposition, opens the way to defining the local characteristics of a general semimartingale. However we will *not* discuss this notion in its full generality, but we will restrict ourselves to semimartingales that are quasi left continuous. Later on we will confine ourselves to the case that the semimartingales are also locally square integrable. In case $E|X_t|^2$ is finite one can show that $|x|^2 \star \nu_t < \infty$.
Therefore $(x - h(x)) \star \nu_t$ is well defined as well as $B_t = B_t^h + (x - h(x)) \star \nu_t$ and $C_t = C_t + xx^T \star \nu_t$.

In this case proposition 2.6.1.2 reads

PROPOSITION 2.6.1.3: *Let X be a process with independent increments such that $E|X_t|^2$ is finite. With the above notation (B, C, ν) is the unique triplet satisfying*
i) $X - B$ *is a martingale*
ii) $(X - B)(X - B)^T - C$ *is a martingale*
iii) $\Sigma_{s \leq t} f(\Delta X_s) - f \star \nu_t$ *is a martingale for all measurable f such that $|f(x)| \leq c|x|^2$ for some $c \in (0, \infty)$.*

2.6.2 Semimartingales

We will follow the notation previously employed for processes with independent increments. Let X be a semimartingale and let h be a truncation function. We make the following assumption

A1: X is a quasi left-continuous process.

This means that for any increasing sequence of stopping times $\{T_n\}$ with limit T we have on $\{T<\infty\}$ $\lim_{n\to\infty} X_{T_n}=X_T$ a.s. Then the following holds

PROPOSITION 2.6.2.1: *There exist a continuous process of finite variation* $B^h:\Omega\times[0,\infty)\to\mathbb{R}^d$, $B_0^h=0$, *a continuous process* $C:\Omega\times[0,\infty)\to\mathbb{R}^{d\times d}$ *such that* $C_t(\omega)-C_s(\omega)\geq 0$, *meaning that* $C_t(\omega)-C_s(\omega)$ *is nonnegative definite if* $t\geq s$ *and a predictable random measure* ν *on* $[0,\infty)\times\mathbb{R}^d$ *such that*

i) $\nu(\omega,\{t\}\times\mathbb{R}^d)=0$, $\forall t\geq 0$, $\nu(\omega,[0,\infty)\times\{0\})=0$

ii) $\int_0^t\int_{\mathbb{R}^d}|x|^2\wedge 1\nu(\omega,dt\times dx)<\infty$ *satisfying the following properties*

i) X^h-B^h *is a local martingale*

ii) $\tilde{X}^h(\tilde{X}^h)^T-\tilde{C}^h$ *is a local martingale, where* $\tilde{C}_t^h=C_t+hh^T\star\nu_t$

iii) $\Sigma_{s\leq t}g(\Delta X_s)-g\star\nu_t$ *is a local martingale for all measurable and bounded functions* g *on* $\mathbb{R}^d$ *which are zero in a neighbourhood of zero.*

An alternative way of getting the triplet (B^h,C,ν) is the following. Observe that X^h is a special semimartingale since it has bounded jumps. Let $X_t^h=X_0+B_t^h+M_t^h$ be its unique special semimartingale decomposition. Split $M_t^h=M_t^c+M_{1t}$, where M^c is a continuous local martingale and M_{1t} a compensated sum of jumps martingale. Hence we get

$$X_t = X_0+B_t^h+M_t^c+M_{1t}+\sum_{s\leq t}(\Delta X_s-h(\Delta X_s))$$

Denote as usual by μ^X the jump measure of X and by ν its dual predictable projection. Observe that $\Delta M_{1t}=h(\Delta X_t)$. Then we can write

$$X_t = X_0+B_t^h+M_t^c+\int_{\mathbb{R}^d}h(x)(\mu^X([0,t]\times dx)-\nu([0,t]\times dx))$$
$$+\int_{\mathbb{R}^d}(x-h(x))\mu^X([0,t]\times dx)$$

or

$$X_t = X_0+B_t^h+M_t^c+h\star(\mu^X-\nu)_t+(x-h(x))\star\nu_t$$

Define also $C=<M^c,(M^c)^T>=[X,X^T]^c$. Then one can check that the triplet (B^h,C,ν) is the same as that in proposition 2.6.2.1; similarly the process $\tilde{C}^h$ in

proposition 2.6.2.1 is given by $\tilde{C}^h = <M^h, (M^h)^T>$.

REMARK: Notice that C and ν, in contrast to $\tilde{C}^h$ and B^h, don't depend on the specific choice of the truncation function h.

There is a third way of defining the local characteristics of a semimartingale. Let $g_u(x) = e^{iu^T x} - 1 - iu^T h(x)$. Since there exists a constant c_u such that $|g_u(x)| \leq c_u(|x|^2 \wedge 1)$, $g_u * \nu$ is well defined and we can then define the process $A(u)$ by $A(u)_t = iu^T B_t^h - \frac{1}{2}u^T C_t u + g_u * \nu_t$. Observe that $A(u)$ does not depend on the function h, and that $A(u)$ (in general being predictable and of finite variation) is a continuous process under the assumption that X is quasi left continuous.

PROPOSITION 2.6.2.2: *Let X be a cadlag process, $X: \Omega \times [0,\infty) \to \mathbb{R}^d$. Let h be a truncation function and let (B^h, C, ν) be defined as in 2.6.2.1 and $A(u)$ as defined above. Then the following statements are equivalent.*

a) *X is a semimartingale with local characteristics (B^h, C, ν).*

b) *For all $u \in \mathbb{R}^d$, the process $M(u): \Omega \times [0,\infty) \to \mathbb{C}$ defined by*

$$M(u)_t = \exp(iuX_t) - \int_0^t \exp(iuX_{s-}) dA(u)_s \text{ is a local martingale.}$$

COROLLARY 2.6.2.3: *If X is a semimartingale with local characteristics (B^h, C, ν) and (A1) holds, then*

$$\exp(iu^T X - A(u)) \text{ is a local martingale.}$$

Observe the resemblance of corollary 2.6.2.3, with the remark following proposition 2.6.1.1. Parallel to the case where processes with independent increments were considered we have the following

PROPOSITION 2.6.2.3: *Assume that the semimartingale X satisfies*
(A2): X is locally square integrable.

i) *Then $|x|^2 * \nu_t < \infty$ a.s. $\forall t \geq 0$, and X is a special semimartingale with canonical decomposition $X = X_0 + B + M$ where $B = B^h + (x - h(x)) * \nu$*

ii) $<M^T, M> = \tilde{C} = C + (xx^T) * \nu$

iii) $\Sigma_{s \leq t} g(\Delta X_s) - g * \nu_t$ *is a local martingale for all measurable g such that* $|g(x)| \leq c|x|^2$.

EXAMPLE: Let N be a counting process with intensity process λ. Then the above applies with $B_t = \int_0^t \lambda_s ds$, $C \equiv 0$, $\nu(dt \times dx) = \lambda_t dt \otimes \epsilon_{\{1\}}(dx)$. Or if we would have taken $E = \{1\}$ as the auxiliary space, then we can take $\tilde{\Omega} = \Omega \times [0,\infty)$ and we simply get $\nu(dt) = \lambda_t dt$.

Chapter 3

Weak Convergence and Local Asymptotic Normality

3.1 Weak Convergence

Weak convergence of stochastic processes is a subject in probability theory, that equally well can be viewed as a subject belonging to functional analysis. To explain this clearly, we have to introduce the relevant vector spaces with suitable topologies and the relevant weakly compact subsets. After having done this we have to show how we can treat stochastic processes within this framework and we have to present verifiable criteria that are sufficient to ensure weak convergence.

Historically, weak convergence was studied first for sequences of independent random variables, later on for stochastic procesess with independent increments.

As announced in section 2.6.1 the local characteristics of a semimartingale play an important role in studying weak convergence of a sequence of semimartingales. In order to obtain some insight in this statement, consider first the case where the semimartingales are processes of independent increments. In this case the characteristics are related to the process via the Lévy-Khintchine formula, which is in fact a formula for the characteristic function. Clearly it is sufficient for finite dimensional convergence of a sequence of semimartingales with independent increments that the corresponding characteristic functions converge, which condition can then be formulated in terms of convergence of the corresponding characteristics. Hence one may anticipate similar conditions (although of course more restrictive) on the local characteristics of the semimartingales that ensure weak convergence of the sequence under consideration. This is partly attributable to the analogy of the results of propositions 2.6.1 and 2.6.2.3.

Before we arrive at these results, we will briefly review some basic facts on weak convergence of general stochastic processes. Relevant references for this chapter are unless others are mentioned, Liptser and Shiryayev [29], Jacod and Shiryayev [8] or, Jacod [18] and Billingsley [2].

3.1.1 General Concepts

Let B be a real Banach space, with norm $\|\cdot\|$. The dual space B' is the vector space of all continuous linear functionals on B with norm

$$\|A\| = \sup\{|Ax| : x \in B,\ \|x\| \leqslant 1\}, \quad A \in B'.$$

The weak topology on B is defined by the neighbourhoods of zero of the form

$$V_{\epsilon,n} = \{x \in B : |A_i x| < \epsilon,\ A_i \in B',\ i = 1,\ldots,n\},$$

or equivalently by saying that a sequence $\{x_n\} \subset B$ converges to $x \in B$ iff $Ax_n \to Ax$, for all $A \in B'$.

The weak $\star$-topology on B' is defined by the neighbourhoods of zero of the form

$$U_{\epsilon,n} = \{A \in B' : |Ax_i| < \epsilon,\ x_i \in B,\ i = 1,\ldots,n\}$$

REMARK: If B is separable, the weak $\star$-topology on B' is metrizable.

DEFINITION 3.1.1.1: A subset E is called *weakly compact* if it is compact for the weak topology on B.

A subset E in B is called *weakly sequentially compact* if every sequence in E has a weakly convergent subsequence with limit in E.

PROPOSITION 3.1.1.2: *If K is a weakly compact set of a separable Banach space, then it is also weakly sequentially compact.*

The usual setup for weak convergence in probability theory is the following. Consider a metric space S, with its Borel σ-algebra $\mathcal{S}$. Let P, P_n, $n = 1,2,\ldots$ be a family of probability measures or $(S, \mathcal{S})$. Clearly each P, P_n can be considered as a continuous linear functional on $C_b(S)$, the space of bounded continuous functions on S, via $Pf = \int_S f dP$. Then $\|f\| = \sup\{|f(x)| : x \in S\}$ makes $C_b(S)$ a Banach space that plays the role of B in the preceding paragraphs and $|Pf| \leq \|f\|$, so indeed $P \in C_b(S)'$. Observe that $\|P\| = 1$, for all probabilities P. The situation in which we are interested is that of stochastic processes. A stochastic process $X : \Omega \times [0,\infty) \to \mathbb{R}^d$ can be viewed as a random variable $X : \Omega \to (\mathbb{R}^d)^{[0,\infty)}$. In order to apply the preceding definitions we have to turn $(\mathbb{R}^d)^{[0,\infty)}$ into a metric space. Of course the space $(\mathbb{R}^d)^{[0,\infty)}$ is too large for this. Since we assumed that all of the stochastic processes have cadlag paths, we replace $(\mathbb{R}^d)^{[0,\infty)}$ with $D^d[0,\infty)$, by definition the space of cadlag functions from $[0,\infty)$ into $\mathbb{R}^d$. The next point is to metrize $D[0,\infty)$. This will be done by means of the so called Skorokhod metric to be defined below. It should be noticed here that weak convergence problems considerably simplify if the processes under consideration have continuous paths. We will give some results for the case where indeed the paths $X.(\omega)$ belong to $C^d[0,\infty)$, the space of continuous functions from $[0,\infty)$ into $\mathbb{R}^d$. The results for $D^d[0,\infty)$-valued random variables will then be seen to be close in formulation to this case. For the particular applications we are interested in, we can even restrict ourselves to studying weak convergence of stochastic processes with paths in $D^d[0,1]$ or $C^d[0,1]$, the space of cadlag (respectively continuous) functions from $[0,1]$ into $\mathbb{R}^d$.

DEFINITION 3.1.1.3: Let P, P_n, $n = 1,2,\ldots$ be probability measures on $(S, \mathcal{S})$. The

sequence $\{P_n\}$ is said to converge weakly to P if $P_n f \to Pf$, for all $f \in C_b(S)$. Notation: $P_n \Rightarrow P$.

Obviously, this fits in the general framework outlined above by taking $B = C_b(S)$.

DEFINITION 3.1.1.4: Let for each n $X_n:(\Omega,\mathcal{F},P)\to(S,\mathcal{S})$ be a random variable. Then we say that $\{X_n\}$ converges weakly to a random variable $X:(\Omega,\mathcal{F},P)\to(S,\mathcal{S})$ if the induced probability measures $P^n = PX_n^{-1}$ on $(S,\mathcal{S})$ converge weakly to PX^{-1}. Notation $X_n \xrightarrow{\mathcal{D}} X$.

The plan of attack in proving weak convergence in usually as follows. Firstly one shows that a sequence of probability measures is weakly compact. Secondly it is shown that the set of limit points is a singleton. This second point can be checked by considering finite dimensional distributions, and it turns out that the first step is the difficult part. A theorem by Prokhorov gives an equivalent formulation for relative compactness. First we need a definition.

DEFINITION 3.1.1.5:

1) A family Π of probability measures on $(S,\mathcal{S})$ is said to be tight if $\forall \epsilon > 0$ there exists a compact subset K of S such that $P(K) \geq 1-\epsilon$, $\forall P \in \Pi$.
2) A sequence of random variables $\{X_n\}$, $X_n:\Omega\to S$ will be called tight iff the induced sequence of probability measures $\{P_n\}$, $P_n = PX_n^{-1}$ on $(S,\mathcal{S})$ is tight.

THEOREM 3.1.1.6: *A set* Π *of probability measures on* $(S,\mathcal{S})$ *is weakly compact if and only if it is tight.*

3.1.2 Weak convergence in $C^d[0,1]$

We first make $C^d[0,1]$ a Banach space by defining for $f \in C^d[0,1]$ the norm $\|f\| = \sup\{|f(x)|: x \in [0,1]\}$. Next we define for $f \in C^d[0,1]$ its modulus of continuity $W_f(\delta)$ by

$$W_f(\delta) = \sup\{|f(t)-f(s)|: t,s \in [0,1], \quad |t-s| \leq \delta\}.$$

and for any interval $I \subseteq [0,1]$

$$W_f(I) = \sup\{|f(t)-f(s)|: t,s \in I\}$$

In order to check the tightness condition of theorem 3.1.1.6 we need a characterization of the compact sets in $C^d[0,1]$. This is given by the Arzela-Ascoli theorem.

PROPOSITION 3.1.2.1: *A subset* K *of* $C^d[0,1]$ *is relatively compact iff*

i) $\sup\{|f(0)|: f \in K\} < \infty$
ii) $\lim_{\delta \to 0} \sup\{W_f(\delta): f \in K\} = 0$.

We are now in the position to state what tightness in $C^d[0,1]$ means.

PROPOSITION 3.1.2.2: *A sequence* $\{P_n\}$ *of probability measures on* $C^d[0,1]$ *is tight iff*

i) $\forall\eta>0$: $\exists a>0$ *such that* $\sup_n P_n\{x\in C^d[0,1]:|x(0)|>a\}\leqslant\eta$

ii) $\forall\epsilon,\eta>0$: $\exists\delta\in(0,1), n_0\in\mathbb{N}$: $\sup_{n\geqslant n_0} P_n\{x\in C^d[0,1]:w_x(\delta)\geqslant\epsilon\}\leqslant\eta$

3.1.3 Weak convergence in $D^d[0,1]$

In order to apply Prokhorov's characterization of weak compactness, we first have to make $D^d[0,1]$ a separable metric space. This will be done via Skorokhod's topology. One of the requirements of a topology on $D^d[0,1]$ is that the relative topology on $C^d[0,1]$ induced by it, should coincide with the norm-topology. Indeed Skorokhod's topology fulfills this requirement. Let us introduce some terminology. A time change on $[0,1]$ is a strictly increasing continuous function $\lambda:[0,1]\to[0,1]$ with $\lambda(0)=0$, $\lambda(1)=1$. The set of time changes is denoted by Λ.

DEFINITION 3.1.3.1: The *Skorokhod topology* is the topology defined by the following convergence concept. Let α, $\alpha_n\in D^d[0,1], \forall n$. Then $\alpha_n\to\alpha$ if there exist a sequence $\{\lambda_n\}$ of time changes on $[0,1]$ such that

i) $\sup\{|\lambda_n(t)-t|:t\in[0,1]\}\to 0$ for $n\to\infty$

ii) $\sup\{|\alpha_n(\lambda_n(t))-\alpha(t)|:t\in[0,1]\}\to 0$, for $n\to\infty$

PROPOSITION 3.1.3.2: *Let* $\alpha,\beta\in D^d[0,1]$. *The function* $d:D^d[0,1]\times D^d[0,1]\to\mathbb{R}^+$ *given by*

$$d(\alpha,\beta) = \inf\{\epsilon:\exists\lambda\in\Lambda:\ \sup\{\log\frac{\lambda(t)-\lambda(s)}{t-s}:t\neq s\}\leqslant\epsilon,$$

$$\sup\{|\alpha(\lambda(t))-\beta(t)|:t\in[0,1]\}\leqslant\epsilon\}$$

defines a metric on $D^d[0,1]$ *that induces the Skorokhod topology, and for which* $D^d[0,1]$ *is a separable metric space.*

Checking whether a sequence $\{\alpha_n\}\subset D^d[0,1]$ converges to a function $\alpha\in D$ may be a complicated task. However in some special situations it is easy to verify whether this is the case.

PROPOSITION 3.1.3.3: *Let* α,α_n *be increasing functions in* $D^d[0,1]$ *with* $\alpha_n(0)=\alpha(0)=0$ $\forall n$ *and assume that* α *is a continuous function. There is equivalence between the following three statements.*

i) $\alpha_n\to\alpha$ *for the Skorokhod topology*

ii) $\alpha_n\to\alpha$ *uniformly on* $[0,1]$

iii) $\alpha_n(t)\to\alpha(t)$ *for all* $t\in[0,1]$.

It turns out that there exists an Arzela-Ascoli type theorem for $(D^d[0,1])$. Therefore we need an extension of the modulus of continuity as defined for continuous functions. Define $w_x^1(\delta)=\inf\{\max_{0<i\leqslant r} w_x[t_{i-1},t_i)\}$, where the infimum

is taken over all partitions $\{t_i\}$ satisfying $0=t_0<\cdots<t_r=1$, and $\inf|t_i-t_{i-1}|\geqslant\delta$.

PROPOSITION 3.1.3.4: *A subset K of* $D^d[0,1]$ *is relatively compact iff*

i) $\sup_{x\in K}\sup_{t\in[0,1]}|x(t)|<\infty$

ii) $\lim_{\delta\to 0}\sup_{x\in K} w_x^1(\delta)=0.$

Via proposition 3.1.3.4 we arrive, similarly to the situation in $C^d[0,1]$, at a criterion for tightness in $D^d[0,1]$.

PROPOSITION 3.1.3.5: *A sequence* $\{P_n\}$ *is tight iff*

i) *for all* $\epsilon>0,\exists a>0$ *such that* $\sup_n P_n\{\sup_t|x(t)|>a\}\leqslant\epsilon$

ii) $\forall\epsilon,\eta>0:\exists\delta>0,\exists n_0\in\mathbb{N}$ *such that* $\sup_{n\geqslant n_0}P_n(w_x^1(\delta)\geqslant\eta)\leqslant\epsilon.$

The following result, which will be convenient later on, is now intuitively clear. (In fact one only has to check tightness of the Z^n below)

PROPOSITION 3.1.3.6: *Assume that* $X^n\xrightarrow{\mathcal{L}}X$ *and that* $\sup_{s\leqslant t}|Y_s^n|\xrightarrow{\mathcal{L}}0$ *(hence in probability) then* $Z^n=X^n+Y^n\xrightarrow{\mathcal{L}}X$.

The cases that will be of interest to us are the ones in which the limiting measures are concentrated on $C^d[0,1]$. A sequence satisfying this property is called C-tight. In this case we can replace proposition 3.1.3.5 with

PROPOSITION 3.1.3.7: *The following statements are equivalent.*

i) *the sequence* $\{X^n\}$ *is C-tight*

ii) *condition 3.1.3.5 (i) holds and* $\forall\epsilon,\eta>0:\exists n_0\in\mathbb{N},\delta>0$ *such that*

$$P^n(w_{X^n}(\delta)\geqslant\eta)\leqslant\epsilon \quad \text{for} \quad n\geqslant n_0$$

iii) *the sequence* $\{X^n\}$ *is tight and* $\forall\epsilon>0$ *we have*

$$\lim_{n\to\infty}P^n(\sup_t|\Delta X_t^n|>\epsilon)=0.$$

The verification of the second criterion of proposition 3.1.3.5 may be difficult since it involves the calculation of $(w_{X^n}^1(\delta))$. The following lemma gives a sufficient condition in terms of stopping times. Assume that each X^n is adapted to a filtration $\mathbb{F}^n(=\{\mathcal{F}_t^N\}_{t\in[0,1]})$ and let $\mathbb{T}^n$ be the set of $\mathbb{F}^n$-stopping times.

LEMMA 3.1.3.8: *If* $\forall N\in\mathbb{N},\forall\epsilon>0,\forall\eta>0:\exists n_0\in\mathbb{N},\delta>0$ *such that*

$$n\geqslant n_0\Rightarrow\sup\{P^n(|X_T^n-X_S^n|>\eta):S,T\in\mathbb{T}^n,|S-T|\leqslant\delta\}\leqslant\epsilon$$

Then $\forall N\in\mathbb{N},\epsilon>0,\eta>0:\exists n_0\in\mathbb{N},\delta>0$ *such that*

$$n\geqslant n_0\Rightarrow P^n(w_{X^n}^1(\delta)>\eta)\leqslant\epsilon.$$

This lemma is particularly useful if we consider the case where the X^n's are $\mathbb{F}^n$-martingales, that are locally square integrable. First we state an auxiliary result.

LEMMA 3.1.3.9: *(Lenglart's inequality): Let X be a stochastic process adapted to a filtration* $\mathbb{F}=\{\mathcal{F}_t\}_{t\in[0,1]}$. *Let A be a* $\mathbb{F}$*-predictable increasing process, such that for all* $\mathbb{F}$*-stopping times T one has* $E|X_T|\leq EA_T$. *Then for all* $\epsilon,\eta>0$ *and all stopping times T*

$$P(\sup_{s\leq T}|X_s|\geq\epsilon)\leq\frac{\eta}{\epsilon}+P(A_T\geq\eta).$$

COROLLARY 3.1.3.10: *Let X be a locally square integrable martingale. Let* $A=\Sigma_j<X^j,X^j>$. *Then* $E|X_T|^2\leq EA_T$ *by Fatou's lemma and*

$$P(\sup_{S\leq T}|X_s|\geq\epsilon)\leq\frac{\eta}{\epsilon^2}+P(A_T\geq\eta).$$

It is now very easy to derive a tightness condition for locally square integrable martingales:

THEOREM 3.1.3.11: *Let* $\{X^n\}$ *be a sequence of locally square integrable martingales and assume that* $X_0^n=0$ *for all n. Then* $\{X^n\}$ *is tight if* $\{A^n\}$ *is C-tight, where* $A^n=\Sigma_j<X^{nj},X^{nj}>$.

It is also easy to provide a sufficient condition for C-tightness of the sequence $\{A^n\}$ as defined in corollary 3.1.3.10.

PROPOSITION 3.1.3.12: *Let* $\{A^n\}$ *be a sequence of increasing processes. Suppose that* $A_t^n\xrightarrow{P}f(t)$, *where f is a continuous deterministic function. Then* $\{A^n\}$ *is C-tight.*

In general it is difficult to check the hypotheses for tightness as given in propositions 3.1.3.5,7, since they require knowledge of the distributions of the involved stochastic processes. Our aim is therefore to formulate sufficient conditions that ensure the necessary and sufficient condition for tightness. It turns out that the local characteristics are useful when we consider the case where the sequence of processes $\{X_n\}$ is a sequence of semimartingales. The idea behind this is motivated by the case where in addition $\{X_n\}$ is a sequence of processes with independent increments, since in this case the distribution of each of the X_n is completely characterized by the corresponding characteristics. However let us remark that in the case where we deal with weak convergence of stochastic processes there is no equivalent description in terms of characteristic functions as in the case where all X_n's take their values in some finite dimensional Euclidean space. See BILLINGSLEY [2] for examples. It is indeed this lack of analogy that leads to studying tightness of a family of distributions

on $C^d[0,1]$ or $D^d[0,1]$. On the other hand if one is only interested in convergence of the finite dimensional distributions, such an analogy indeed exists. Motivated by this brief explanation above we first give some results for processes with independent increments.

3.1.4 Weak convergence of processes with independent increments

In view of proposition 2.6.1 it is easy to give conditions on the characteristics of the processes under consideration that are equivalent to converging of X_t^n to X_t in distribution for each t. Since weak convergence of the sequence $\{X^n\}$ deals with the whole paths one anticipates conditions on the characteristics that also involve the whole trajectories. Here is the result.

THEOREM 3.1.4.1: *Let $\{X^n\},X$ be processes with independent increments without fixed discontinuities, and h a truncation function. Let $(B^{h,n},C^n,\nu^n)$ and (B^h,C,ν) be the corresponding characteristics. Then $X^n \xrightarrow{\mathcal{L}} X$ is equivalent to the following set of conditions*

i) $\sup_{[0,t]}|B_t^{h,n}-B_t^h|\to 0$

ii) $\sup_{[0,t]}|\tilde{C}_t^{h,n}-\tilde{C}_t^h|\to 0$

iii) $\sup_{[0,t]}|f\star(\nu^n-\nu)_t|\to 0$, *for every bounded continuous function $f:\mathbf{R}^d\to\mathbf{R}^+$, which is zero in a neighbourhood of zero.*

REMARKS: One can show that these conditions may be reformulated in terms of convergence for the Skorokhod topology.

In the proof of theorem 3.1.4.1 its conditions are mainly used to prove tightness. The fact that the law of X is completely determined by (B^h,C,ν) completes the proof.

PROPOSITION 3.1.4.2: *Condition ii) and iii) of theorem 3.1.3.1 are equivalent with respectively*

ii) $\tilde{C}_t^{h,n}\to\tilde{C}_t^h$, $\forall t\geqslant 0$

iii) $f\star\nu_t^n\to f\star\nu_t$, $\forall t\geqslant 0$ *and for all $f:\mathbf{R}^d\to\mathbf{R}^+$, which are continuous, bounded and zero in a neighbourhood of 0.*

This proposition immediately follows from proposition 3.1.3.3 since the functions that we consider are increasing and have continuous limits. We also have as an immediate consequence from the continuity theorem for characteristic functions

COROLLARY 3.1.4.3: *Suppose that the processes with independent increments $\{X^n\},X$ have no fixed discontinuities and suppose that $X_t^n \xrightarrow{\mathcal{L}} X_t$, for all $t\geqslant 0$. Then $X^n \xrightarrow{\mathcal{L}} X$ if and only if $\sup_{[0,t]}|B_t^{h,n}-B_t^h|\to 0$.*

3.1.5 Weak convergence of semimartingales to a process with independent increments

Let $\{X^n\}$ be a sequence of semimartingales. Each X^n is defined on a complete filtered probability space $(\Omega^n,\mathcal{F}^n,\mathbb{F}^n,P^n)$. In addition we assume that each X^n satisfies condition A 1 of section 2.6.2. Denote by (B^n,C^n,ν^n) the corresponding local characteristics. X is supposed to be a process with independent increments without fixed discontinuities that has characteristics (B,C,ν).

The next theorem 3.1.5.1 parallels to a certain extent theorem 3.1.3.1 in that the given conditions are similar but only sufficient, and that the convergence takes place in distribution and thus in measure since the limiting processes are deterministic.

THEOREM 3.1.5.1: *Under the assumptions presented above we have* $X^n \xrightarrow{\mathcal{L}} X$ *if each of the next conditions is satisfied.*

i) $\sup_{t\le 1}|B_t^{h,n}-B_t^n| \xrightarrow{P} 0$

ii) $\tilde{C}_t^{h,n} \xrightarrow{\mathcal{L}} \tilde{C}_t^h, \quad \forall t\in[0,1]$

iii) $f\star\nu_t^n \xrightarrow{\mathcal{L}} f\star\nu_t$, *for all continuous bounded functions* f *which are zero in a neighbourhood of zero.*

By specialization to the case where the semimartingales and the limiting process are locally square integrable we obtain

PROPOSITION 3.1.5.2: $X^n \xrightarrow{\mathcal{L}} X$ *if we have (iii) of theorem 3.1.5.1 and*

i) $\sup_{t\le 1}|B_t^n-B_t| \xrightarrow{P} 0$

ii) $\tilde{C}_t^n \xrightarrow{\mathcal{L}} \tilde{C}_t, \quad \forall t\in[0,1]$

iii) $\lim_{b\uparrow\infty}\limsup_n P^n(|x|^2 1_{\{|x|>b\}}\star\nu_t^n>\epsilon)=0, \quad \forall\epsilon>0, \quad \forall t\in[0,1]$

where as usual $B^n=B^{h,n}+(x-h(x))\star\nu^n$ *and* $\tilde{C}^n=C^n+xx^T\star\nu_t^n$ *and likewise for* B *and* $\tilde{C}$.

Further specializing to the case where X is a continuous process with independent increments and characteristics $(B,C,0)$, we get

PROPOSITION 3.1.5.3: $X^n \xrightarrow{\mathcal{L}} X$ *if*

i) $|x|^2 1_{\{|x|>\epsilon\}}\star\nu_t^n \xrightarrow{P} 0, \quad \forall t,\epsilon>0$

ii) $\sup_{t\le 1}|B_t^n-B_t| \xrightarrow{P} 0.$

iii) $\tilde{C}_t^n \xrightarrow{\mathcal{L}} \tilde{C}_t, \quad \forall t>0.$

In the particular case that we deal with locally square integrable martingales proposition 3.1.5.3 reads

PROPOSITION 3.1.5.4: *Let* $\{X^n\}$ *be a sequence of locally square integrable martingales,* $\mathbb{R}$*-valued, and* X *a continuous Gaussian martingale with characteristics* $(0,C,0)$. *Then* $X^n \xrightarrow{\mathcal{L}} X$ *if*

i) $|x|^2 1_{\{|x|>\epsilon\}} \star \nu_t^n \xrightarrow{\mathcal{L}} 0$

ii) $\tilde{C}_t^n \xrightarrow{\mathcal{L}} \tilde{C}_t, \ \forall t>0.$

REMARK 1: Condition i) is also known as "strong asymptotic rarefaction of the jumps of the second type", see [34]. Both proposition 3.1.5.3 and proposition 3.1.5.4 are known under the name "functional central limit theorem" for (semi)martingales.

REMARK 2: The sufficient conditions in the above propositions and theorems that involve the third characteristic are easily verified in the case that $\sup_t |\Delta X_t^n| \leq c_n, \ \forall n$, where the deterministic sequence $\{c_n\}$ tends to zero.

REMARK 3: Of course the limiting process X is a Brownian motion if $C_t = t$.

REMARK 4: condition 3.1.5.4 i) implies that $1_{\{|x|^2>\epsilon\}} \star \nu_t^n \xrightarrow{\mathcal{L}} 0$.

In LIPTSTER & SHIRYAYEV [29] it has been proved that this last property is equivalent to $\sup_{s \leq t} |\Delta X_s^n| \xrightarrow{P_n} 0$. Therefore in view of propositions 3.1.3.7 and 3.1.3.11 one sees that conditions 3.1.5.4 i) and ii) imply that the sequence $\{X^n\}$ is C-tight.

The convergence in condition 3.1.5.4 (i) takes place in probability. If we replace it with the stronger condition that convergence holds in L_1, then we get Lindeberg's condition:

$$E|x|^2 1_{\{|x|>\epsilon\}} \star \nu_t^n \to 0.$$

Observe that

$$E|x|^2 1_{\{|x|>\epsilon\}} \star \nu_t^n = E|x|^2 1_{\{|x|>\epsilon\}} \star \mu_t^n = E \sum_{s \leq t} |\Delta X_s^n|^2 1_{\{|\Delta X_s^n|>\epsilon\}},$$

which enables us to formulate sufficient conditions for weak convergence directly in terms of the jumps of X^n.

It may turn out to be convenient to work with other sufficient conditions. Some of these are listed below. The statement in the next proposition is obvious. Nevertheless we give the proof, since this proposition cannot be found in the standard references.

PROPOSITION 3.1.5.5: *Condition 3.1.5.3(i) is implied by* $|x|^{2+\alpha} \star \nu_t^n \xrightarrow{\mathcal{L}} 0$, *for some* $\alpha>0$.

PROOF: Observe first that $|x|^{2+\alpha} 1_{\{|x|>\epsilon\}} \star \nu_t^n \geq \epsilon^{2+\alpha} 1_{\{|x|>\epsilon\}} \star \nu_t^n$. Application of this and Hölder's inequality (with $p = \frac{2}{2+\alpha}, q = \frac{\alpha}{2+\alpha}$) yields

$$|x|^2 1_{\{|x|>\epsilon\}} \star \nu_t^n \leq (|x|^{2+\alpha} \star \nu_t^n)^{\frac{2}{2+\alpha}} (1_{\{|x|>\epsilon\}} \star \nu_t^n)^{\frac{\alpha}{2+\alpha}} \leq$$

$$\leqslant \epsilon^{-\alpha}(|x|^{2+\alpha}\star\nu_t^n)^{\frac{2}{2+\alpha}}(|x|^{2+\alpha}\star\nu_t^n)^{\frac{\alpha}{2+\alpha}}$$

$$= \epsilon^{-\alpha}|x|^{2+\alpha}\star\nu_t^n. \qquad \square$$

If we replace the convergence in proposition 3.1.5.5 by L_1-convergence, we get Lyapunov's condition:

$$E|x|^{2+\alpha}\star\nu_t^n \to 0$$

As above this condition can be reformulated in terms of the jumps of the X^n as follows:

$$E\sum_{s\leqslant t}|\Delta X_s^n|^{2+\alpha}\to 0.$$

Parallel to proposition 3.1.5.5, one can show (see [46] for a different proof) that Lyapunov's condition implies Lindeberg's condition. The advantage of both these conditions is the fact that it is sometimes comparatively simple to compute expectations.

3.1.6.

In this subsection we discuss some other forms of the central limit theorem. The difference with previous sections is that we work with a single semimartingale and study the asymptotic behaviour as t tends to infinity. In the literature, see for example [26,28] self contained proofs have been given for the results below. Here we will embed these into the framework that has been used throughout this chapter, thus giving alternative proofs. Results of this section will be used in determining the asymptotic distribution of certain recursive estimators that play a role in sections 4.2 and 4.3. First we present after some introductory notation and definitions the result of [26]. Let μ be an integer valued random measure and ν its dual predictable projection. Let f be a measureable function on $\tilde{\Omega}$ (see the terminology of section 2.5) such that $\forall T\geqslant 0$:

$$\int_0^T\int_{\mathbb{R}^d}|f(t,x)|^2\nu(dt\times dx)<0 \text{ a.s.}$$

Then the following process ζ is well defined:

$$\zeta_t = \int_0^t\int_{\mathbb{R}^d} f(s,x)(\mu-\nu)(ds\times dx).$$

In [26] the following result can be found.

PROPOSITION 3.1.6.1: *Assume that there exists a function* $g:[0,\infty)\to[0,\infty)$ *such that* $g(t)\to\infty$ *for* $t\to\infty$ *with*

i) $\lim_{T\to\infty} g(T)^{-2}\int_0^T\int_{\mathbb{R}} f^2(t,x)\nu(dt\times dx)\xrightarrow{P}1$

ii) $\lim_{T\to\infty} g(T)^{-(2+\alpha)}E\int_0^T\int_{\mathbb{R}}|f(t,x)|^{2+\alpha}\nu(dt\times dx) = 0$, *for some* $\alpha>0$

Then $g(t)^{-1}\zeta_t \xrightarrow{\mathcal{L}} N(0,1)$ *as* $t\to\infty$, *where* $N(0,1)$ *denotes a standard Gaussian random variable.*

PROOF: We have to show that for all sequences b_n with $b_n\to\infty$ the random variable ζ_{b_n} has a distribution which is asymptotically normal with parameter $(0,g(b_n)^2)$. Define $a_n=g(b_n)^2$. Without loss of generality we can assume that g is strictly increasing. Hence its inverse h is well defined. Let $t\in[0,1]$ and define

$$M_t^n = \frac{1}{\sqrt{a_n}} \int_0^{h(\sqrt{a_n}t)} \int_z f(s,x)(\mu-\nu)(ds\times dx).$$

Let $\mathcal{F}_t^n=\mathcal{F}_{h(\sqrt{a_n}t)}$, then M^n is $\mathbb{F}^n$-adapted. We will now show that M^n satisfies the Lyapunov condition.

$$\Delta M_t^n = \frac{1}{\sqrt{a_n}} \int_z f(h(\sqrt{a_n}t),x)\mu(\{h(\sqrt{a_n}t)\}\times dx)$$

Because $\mu(\{(\sqrt{a_n}t)\})$ is in fact a Dirac measure for each ω on $\mathbb{R}^d$, concentrated on some point $z=z(a_nt,\omega)$ [8], we have

$$|\Delta M_t^n|^{2+\alpha} = a_n^{-1-\alpha/2} \int_z |f(h(\sqrt{a_n}t),x)|^{2+\alpha}\mu(\{h(\sqrt{a_n}t)\}\times dx)$$

Hence

$$\sum_{s\le t}|\Delta M_s^n|^{2+\alpha} = a_n^{-1-\alpha/2} \int_0^{h(\sqrt{a_n}t)} \int_z |f(s,x)|^{2+\alpha}\mu\ (ds\times dx)$$

Since $\mu-\nu$ is a local martingale measure

$$E\sum_{s\le t}|\Delta M_s^n|^{2+\alpha} = a_n^{-1-\alpha/2}E \int_0^{h(\sqrt{a_n}t)} \int_z |f(s,x)^{2+\alpha}\nu(ds\times dx)$$

$$\le a_n^{-1-\alpha/2}E \int_0^{h(\sqrt{a_n})} \int_z |f(s,x)|^{2+\alpha}\nu(ds\times dx)$$

$$= (g(h(\sqrt{a_n})))^{-2-\alpha}\ E \int_0^{h(\sqrt{a_n})} \int_z |f(s,x)|^{2+\alpha}\nu(ds\times dx)\to 0$$

by assumption (ii). So a fortiori the sequence $\{M^n\}$ satisfies assumption 3.1.5.4 (i). We proceed to investigate the process $<M^n>$. A simple calculation gives

$$<M^n>_t = \frac{1}{a_n} \int_0^{h(\sqrt{a_n}t)} \int_z |f(s,x)|^2\nu(ds\times dx)$$

$$= t\frac{1}{(g(h\sqrt{a_n}t)))^2} \int_0^{h(\sqrt{a_n}t)} \int_z |f(s,x)|^2\nu(ds\times dx)\to t$$

in probability by assumption i).
We are now in the position to apply proposition 3.1.5.4 and we conclude that $M^n \xrightarrow{\mathcal{L}} W$, where W is a standard brownian motion. In particular

$$M_1^n \xrightarrow{\mathcal{L}} N(0,1)$$

or

$$\frac{1}{\sqrt{a_n}} \int_0^{h(\sqrt{a_n})_z} \int f(s,x)(\mu-\nu)(ds\times dx) \xrightarrow{\mathcal{L}} N(0,1)$$

which gives us the desired result by definition of a_n.

Corresponding to proposition 3.1.5.4 we have in this context the following

PROPOSITION 3.1.6.2: *Assume that there exists a function* $g:[0,\infty)\to g[0,\infty)$ *such that* $g(t)\to\infty$ *as* $t\to\infty$ *and*

i) $g(T)^{-2}\int_0^T\int_{\mathbb{R}^d}|f(s,x)|^2\nu(ds\times dx) \xrightarrow{P} 1$

ii) $g(T)^{-2}\int_0^T\int_{\mathbb{R}^d}|f(s,x)|^2 1_{\{|f(s,x)|>\epsilon g(T)\}}\nu(ds\times dx) \xrightarrow{P} 0$

Then $g(T)^{-1}\zeta_T \xrightarrow{\mathcal{L}} N(0,1)$.

PROOF: As in the proof of proposition 3.1.6.1, we define a sequence of local martingales M^n by

$$M_t^n = \frac{1}{\sqrt{a_n}} \int_0^{h(\sqrt{a_n}t)} \int_{\mathbb{R}^d} f(s,x)(\mu-\nu)(ds\times dx).$$

Again we see that $<M^n>_t \xrightarrow{\mathcal{L}} t$, for fixed t. Next we need to verify that 3.1.5.4i holds. We get

$$\sum_{s\le t}|\Delta M_s^n|^2 1_{\{|\Delta M_s^n|\ge\epsilon\}} =$$

$$= \sum_{s\le t}\frac{1}{a_n}\int_{\mathbb{R}^d}|f(h(\sqrt{a_n}t),x)|^2 1_{\{|f(h(\sqrt{a_n}s),x)|>\epsilon\sqrt{a_n}\}}\cdot\mu(\{h(\sqrt{a_n}s)\}\times dx)$$

$$= \int_0^{h(\sqrt{a_n}t)}\int_{\mathbb{R}^d}\frac{1}{a_n}|f(s,x)|^2 1_{\{|f(s,x)|>\epsilon\sqrt{a_n}\}}\mu(ds\times dx),$$

Hence the strong asymptotic rarefaction of jumps property in this case becomes:

$$\frac{1}{a_n}\int_0^{h(\sqrt{a_n}t)}\int_{\mathbb{R}^d}|f(s,x)|^2 1_{\{|f(s,x)|>\epsilon\sqrt{a_n}\}}\nu\,(ds\times dx)\xrightarrow{P}0$$

Now this integral is less than

$$\frac{1}{g(b_n)^2}\int_0^{b_n}\int |f(s,x)|^2 1_{\{|f(s,x)|>\varepsilon g(b_n)\}}\nu(ds\times dx),$$

by definition of a_n. Indeed this quantity tends to zero in probability by assumption.

REMARK: Observe that proposition 3.1.6.1 is a special case of proposition 3.1.6.2 in view of the relation between the conditions (ii) in both propositions.

Of particular interest for us is the case where the random measure μ is in fact a counting process n, and its compensator ν is of the form $\nu(dt)=\lambda_t dt$. In this case we have $\zeta_t=\int_0^t f_s(dn_s-\lambda_s ds)$ and proposition 3.1.6.2 reads as follows:

PROPOSITION 3.1.6.3: *Assume that there exists a function* $g:[0,\infty)\to[0,\infty)$ *such that* $g(t)\to\infty$ *as* $t\to\infty$ *and*

i) $g(t)^{-2}\int_0^t f_s^2\lambda_s ds \xrightarrow{P} 1.$

ii) $g(t)^{-2}\int_0^t f_s^2 1_{\{|f_s|>\varepsilon g(t)\}}\lambda_s ds \xrightarrow{P} 0.$

Then $g(t)^{-1}\int_0^t f_s(dn_s-\lambda_s ds)\xrightarrow{\mathcal{L}} N(0,1).$

It is indeed this type of central limit theorem that will be used in sections 4.2 and 4.3 to obtain the asymptotic distribution of certain recursive estimators.

3.2 LIKELIHOOD RATIOS AND LOCAL ASYMPTOTIC NORMALITY

In this section we discuss what is called local asymptotic normality (LAN) for counting processes. This LAN property is a special case of what is known as contiguity for two sequences of probability measures. First we spend a few words on the situation where the probabilities are concentrated on a finite dimensional space. This discussion facilitates the understanding of the relevant definition for the case when we deal with a sequence of stochastic processes. The raison d'être of this section partly lies in the fact that we use some of the results in chapter 4.

Consider a sequence of binary experiments $(\Omega^n,\mathcal{F}^n,P_1^n,P_0^n)$, each P_i^n being a probability measure on $(\Omega^n,\mathcal{F}^n)$. the sequences $\{P_1^n\}$ and $\{P_0^n\}$ are said to be *contiguous* if for all $A_n\in\mathcal{F}^n$,

$$\lim_{n\to\infty}P_1^n(A_n)=0\Leftrightarrow\lim_{n\to\infty}P_0^n(A_n)=0$$

Think now of P_1^n and P_0^n as two alternative distributions of a random vector $X^n:\Omega^n\to\mathbb{R}^k$. So the P_i^n are now defined on $(\mathbb{R}^k,Bor(\mathbb{R}^k))$. Let p_i^n be the corresponding densities with respect to some dominating σ-finite measure and assume that for each n $P_1^n\sim P_0^n$. In this case one has the following result, known as Le Cam's first lemma [7].

PROPOSITION 3.2.1: *Denote by* $\mathcal{L}^n$ *the law of* $\log\frac{p_1^n}{p_0^n}$ *under* P_0^n. $\{P_0^n\}$ *and* $\{P_1^n\}$ *are contiguous if and only if the sequence* $\{\mathcal{L}^n\}$ *is weakly compact and each limit point* $\mathcal{L}$ *satisfies*

$$\int_{\mathbf{R}} \exp(y)\mathcal{L}(dy) = 1.$$

COROLLARY 3.2.2: *Suppose that* $\mathcal{L}^n$ *converges to a normal* $N(\mu,\sigma^2)$ *distribution. Then* $\{P_1^n\}$ *and* $\{P_0^n\}$ *are contiguous iff* $\mu = -\frac{1}{2}\sigma^2$.

This corollary makes the next definition of a specific case of contiguity understandable. Let $\{P_\theta^n\}$ be a sequence of probability measures indexed by a parameter $\theta\in\Theta\subset\mathbf{R}^d$. Think of P_θ^n being the law of a random vector $X^n:\Omega^n\to\mathbf{R}^k$. Let θ_0 play the role of the "true" parameter and let $u\in\mathbf{R}^d$. Let $\{M^n\}$ be a sequence of matrices in $\mathbf{R}^{d\times d}$ such that $M^n\to 0$ and define $\theta^n=\theta_0+M^n u$. We will assume that $\theta_0\in Int\Theta$ and therefore that $\theta^n\in\Theta$ for all n. Denote by P_0^n the probability $P_{\theta_0}^n$ and by P_1^n the probability P_θ^n for $\theta=\theta^n$. Define Z^n to be the sequence of likelihood ratios

$$Z^n = Z^n(u) = \frac{dP_1^n}{dP_0^n}.$$

DEFINITION 3.2.3: The family of measures $\{P_\theta^n\}$ is called *locally asymptotically normal* (LAN) at θ_0, with normalizing sequence $\{M^n\}$ if Z^n admits the representation

$$Z^n = \exp(u^T\Delta^n - \tfrac{1}{2}u^T u + r^n),$$

where the Δ^n are random variables that converge weakly to a standard Gaussian random variable Δ under the sequence $\{P_0^n\}$ and $\lim P_0^n(|r^n|>\delta) = 0$, $\forall\delta>0$.

REMARK:

1 Observe that $E(u^T\Delta - \frac{1}{2}u^T u) = -\frac{1}{2}u^T u$ and $Var(u^T\Delta - \frac{1}{2}u^T u) = u^T u$, which corresponds to the situation described in corollary 3.2.2.
2 Of course we may replace the discrete parameter n, above by a continuous one.

Now we will turn to the case where we are dealing with stochastic processes instead of finite dimensional random variables. Suppose that we have a sequence of stochastic processes X^n, each of which is defined on a space $(\Omega^n,\mathcal{F}^n,\mathbf{F}^N)$ and suppose that the time set is $[0,1]$, and assume that $X^n\in D[0,1]^k$. Denote by P_θ^n the law of X^n on $D[0,1]^k$, where as before the parameter θ is taken from a set $\Theta\subset\mathbf{R}^d$. With the same notation as before,

define

$$\theta^n = \theta_0 + M^n u,\ P_0^n = P_{\theta_0}^n \text{ and } P_1^n \text{ to be } P_\theta^n \text{ for } \theta=\theta^n.$$

Assume that $P_1^n \ll P_0^n$ and define the likelihood ratio process Z_t^n to be the Radon-Nikodym derivative of P_1^n with respect to P_0^n, restricted to $\mathcal{F}_t^n$.

DEFINITION 3.2.4: The family of measures $\{P_\theta^n\}$ is called locally asymptotically normal at θ_0 with normalizing sequence $\{M^n\}$ if Z^n admits the representation

$$Z_t^n = \exp(u^T\Delta_t^n - \tfrac{1}{2}u^T < W >_t u + r_t^n).$$

Here $\{\Delta^n\}$ is a sequence of $\mathbf{F}^n$-adapted stochastic processes such that $\Delta^n \xrightarrow{\mathcal{L}(P_0^n)} W$, where the weak convergence to the Gaussian martingale W with continuous quadratic variation $< W >$ takes place relative to the sequence $\{P_0^n\}$ and

$$P_0^n(\sup_{s\le t}|r_s^n| > \epsilon) \to 0.$$

REMARK: Observe that in definition 3.2.4

$$\log Z^n \xrightarrow{\mathcal{L}(P_0^n)} u^T W - \tfrac{1}{2}u^T < W > u.$$

Furthermore by taking $t=1$ we get $Z_1^n = \exp(u^T\Delta_1^n - \frac{1}{2}u^T u + r_1^n)$ and

$$\Delta_1^n \xrightarrow{\mathcal{L}(P_0^n)} N(0,I),$$

thus we are back in the situation of definition 3.2.3.

The point is now to identify for a given sequence $\{X^n, P_\theta^n\}$ the processes Δ^n and the normalizing sequence M^n as well as the likelihood ratio process Z^n itself. This problem will be treated in subsequent sections for the case where the X^n are counting processes. Although we will not fully exploit the LAN property in the sense that we will not discuss asymptotic properties of (off-line) maximum likelihood estimators, we remark that establishing LAN is important for proving consistency and asymptotic normality. In a general context the reader is referred to IBRAGIMOV and KHASMINSKII [7] and for counting processes for instance to KUTOYANTS [23,24] or LIN'KOV [27]. The heuristic idea is however simple enough to present. Suppose that one works with a single counting process N defined on $[0,\infty)$, and suppose that an unknown parameter θ, entering in the intensity process, is to be estimated. In this case one usually establishes LAN as an asymptotic property for $t\to\infty$. If LAN holds for the likelihood ratio Z_t with a normalizing sequence $M(t)$ we have with $\theta^t = \theta_0 + M(t)u$ for $Z_t = \dfrac{dP_{\theta^t}}{dP_{\theta_0}}$ the representation

$$Z_t = \exp(u^T\Delta_t - \tfrac{1}{2}u^T u + r_t).$$

By ignoring the remainder term r we have for the maximum likelihood estimator of θ_0 in terms of u: $\hat{u}_t = \Delta_t$ and hence the "real" maximum likelihood estimator $\hat{\theta}_t$ should then be approximately equal to

$$\theta_0 + M(t)\hat{u}_t = \theta_0 + M(t)\Delta_t.$$

Hence $\hat{\theta}_t - \theta_0 \approx M(t)\Delta_t$, which tends to zero in P_{θ_0} probability and $\hat{\theta}_t - \theta_0$ is asymptotically distributed as a normal $N(0, M(t)^2)$ random variable. Similar properties are desired for recursive estimators to be treated in chapter 4. The analysis of their asymptotic behaviour as presented later on differs considerably from what can be done for "off-line" estimators, since it is often not clear what the minimizing criterion is, in contrast with eg. maximum likelihood estimation. However a clever interpretation of the form Z_t at least offers a way to guess a possible recursive algorithm that generates estimators with good asymptotic properties.

In studying LAN for counting processes, we use proposition 3.2.5 below. Suppose that a counting process N on $(\Omega, \mathcal{F}, \mathbb{F}, P_0)$ admits an intensity process λ. Let $m_t = N_t - \int_0^t \lambda_s ds$, m is a local martingale. Let ρ be another nonnegative predictable process such that

$$\int_0^t (\rho_s - 1)\lambda_s ds \text{ is a.s. finite for all } t.$$

Then X defined by $X_t = \int_0^t (\rho_s - 1) dm_s$ is again a local martingale. Let Z be the solution of the Doléans equation $dZ_t = Z_{t-} dX_t$, $Z_0 = 1$. Then $Z_t = \exp(\int_0^t \log \rho_s dN_s - \int_0^t (\rho_s - 1)\lambda_s ds)$, and Z is a nonnegative local martingale. Assume that $EZ_1 = 1$. It is known [37], that we can define another measure P_1 on $(\Omega, \mathcal{F})$ such that for each t the restriction of P_1 to $\mathcal{F}_t$ is absolutely continuous with respect to the restriction of P_0 to the same $\mathcal{F}_t$, and the Radon-Nikodym derivative on $\mathcal{F}_t$ equals Z_t. The next proposition is a special case of a result in [37].

PROPOSITION 3.2.5: *Under P_1, N has the intensity $\rho\lambda$.*

PROOF: Define m^1 by $m_t^1 = m_t - \int_0^t (\rho_s - 1)\lambda_s ds = N_t - \int_0^t \rho_s \lambda_s ds$. We will prove that m_1 is a local martingale under P_1. Consider first the process $m^1 Z$. By the stochastic calculus rule

$$\begin{aligned} d(m^1 Z)_t &= m_{t-}^1 dZ_{t-} + Z_{t-} dm_t^1 + d[m^1, Z]_t \\ &= m_{t-}^1 Z_{t-} dX_t + Z_{t-} dm_t - Z_t(\rho_t - 1)\lambda_t dt + Z_{t-} d[m, X]_t \\ &= m_{t-}^1 Z_{t-} dX_t + Z_{t-} dm_t + Z_{t-}(\rho_t - 1)[dN_t - \lambda_t dt] \\ &= m_{t-}^1 Z_{t-} dX_t + Z_{t-} \rho_t dm_t. \end{aligned}$$

Hence we see that $m^1 Z$ is local martingale under P_0. Let $\{T_n\}$ be a to ∞ increasing sequence of stopping times such that both $\{m_{t \wedge T_n}^1 Z_{t \wedge T_n}\}$ and $\{Z_{t \wedge T_n}\}$ are martingales under P_0. Write E_i for the expectation under the

measure P_i. Then

$$E_1[m^1_{t\wedge T_n}|\mathcal{F}_s] = \frac{E_0[m^1_{t\wedge T_n} Z_{t\wedge T_n}|\mathcal{F}_s]}{E_0[Z_{t\wedge T_n}|\mathcal{F}_s]} = m_{s\wedge T_n}. \quad \square$$

In the sequel we always assume that, whenever there are two probabilities P_0 and P_1 involved, the latter is an absolutely continuous transformation of the first. This is of course a restriction, but not too serious, if one accepts the idea behind the following example. Let $(\Omega,\mathcal{F},P)$ be a probability space with a standard Poisson process N defined on it, and $\mathcal{F}=\sigma\{N_t, t<\infty\}$. Let $(\tilde{\Omega},\tilde{\mathcal{F}})$ be another measurable space with two measures $\tilde{P}^0$ and $\tilde{P}^1$ which are mutually singular, for example $\tilde{\Omega}=\{0,1\}$, $\tilde{P}^i$ is the Dirac measure at i. Form the product space $(\Omega\times\tilde{\Omega},\mathcal{F}\times\tilde{\mathcal{F}})$ with the two product measures $P\times\tilde{P}^i$. If we define $\bar{N}$ on the product space by $\bar{N}_t(\omega,i)=N_t(\omega)$ then clearly $\bar{N}$ has intensity 1 under each of the two product measures, although $P\times\tilde{P}^0 \perp P\times\tilde{P}^1$. Suppose now that we are only interested in what happens on $\mathcal{F}_\infty$. We see that the projections of $P\times\tilde{P}^1$ and $P\times\tilde{P}^0$ on this σ-algebra coincide. So under this assumption the fact that $P\times\tilde{P}^1 \perp P\times\tilde{P}^0$ is innocuous.

In order to study LAN for counting processes in a proper way, we adopt the following approach. Suppose that we are given a sequence of binary experiments $(\Omega^n,\mathcal{F}^n,\mathbb{F}^n,P^n_1,P^n_0)$. Here the filtrations $\mathbb{F}^n$ are $\mathbb{F}^n=\{\mathcal{F}^n_t\}_{t\in[0,1]}$. Let $\{N^n\}$ be a sequence of counting processes, each N^n defined on $(\Omega,^n,\mathcal{F}^n,\mathbb{F}^n)$ and assume that N^n admits the intensity λ^n under P^n_0 and that P^n_1 is such that N^n has the intensity $\rho^n\lambda^n$ under this measure, for some nonnegative predictable process ρ^n. It follows that that $P^n_1 \ll P^n_0$. The Hellinger process H^n is in this situation defined as

$$H^n_t = \int_0^t (\sqrt{\rho^n_s}-1)^2\lambda^n_s\, ds. \tag{3.1}$$

For a definition of the Hellinger process in a more general situation see [6]. In general the behaviour of the Hellinger processes H^n characterizes the behaviour of the likelihood ratio $\dfrac{dP^n_1}{dP^n_0}$ [6]. For multivariate counting processes this has been explained in a fashion, tailored for nonparametric applications in [16]. In this section we will follow an approach that is close to [16], thus avoiding explicit technical conditions as given by other authors as LIN'KOV [27] and KUTOYANTS [23,24]. In agreement with the previous notation we have for

$$Z^n_t = \frac{dP^n_1}{dP^n_0}|\mathcal{F}^n_t$$

the expression

$$Z^n_t = \exp\Big(\int_0^t \log\rho^n_s\, dN^n_s - \int_0^t (\rho^n_s-1)\lambda^n_s\, ds\Big). \tag{3.2}$$

So Z^n is the exponential of the local martingale $\overline{M}^n$, defined by $\overline{M}^n_t = \int_0^t (\rho^n_s - 1) dm^n_s$, where m^n is the local martingale part of N^n under P^n_0. Let W be a continuous Gaussian martingale with (deterministic) quadratic variation process $<W>$. The result that we want to get is weak convergence of the sequence Z^n to the exponential of W, so to $\exp(W - \frac{1}{2}<W>)$, which is the content of the next theorem.

THEOREM 3.2.6: *Assume that H^n_t converges to $\frac{1}{4}<W>_t$ in P^n_0-probability for each t and that for all $\epsilon>0$*

$$H^{n,\epsilon}_t = \int_0^t 1_{\{|\sqrt{\rho^n_s}-1|>\epsilon\}} (\sqrt{\rho^n_s} - 1)^2 \lambda^n_s ds$$

tends to zero in P^n_0-probability. Then

$$Z^n \xrightarrow{\mathcal{L}(P^n_0)} \exp(W - \tfrac{1}{2}<W>),$$

where Z^n is as defined in (3.2) and the weak convergence takes place with respect to the sequence $\{P^n_0\}$.

PROOF: Write

$$\begin{aligned} \log Z^n_t &= 2M^n_t - 2(H^n_t - \tfrac{1}{4}<W>_t) - \tfrac{1}{2}<W>_t - \\ &\qquad \int_0^t (\sqrt{\rho^n_s} - 1)^2 dm^n_s + 2\int_0^t \psi(\sqrt{\rho^n_s}) dN^n_s, \\ &= 2M^n_t - \tfrac{1}{2}<W>_t + Y^n_t \end{aligned}$$

where $M^n_t = \int_0^t (\sqrt{\rho^n_s} - 1) dm^n_s$, $\psi(x) = \log x - x + 1 + \frac{1}{2}(x-1)^2$, and Y^n_t is simply defined as $\log Z^n_t - 2M^n_t + \frac{1}{2}<W>_t$. Our plan of attack is to prove the following steps.

1 $2M^n - \frac{1}{2}<W> \xrightarrow{\mathcal{L}(P^n_0)} W - \frac{1}{2}<W>$.

2 $\sup_{t\in[0,1]} |Y^n_t| \xrightarrow{(P^n_0)} 0$

As soon as we accomplished doing this, then from proposition 3.1.3.6 we get $\log Z^n \xrightarrow{\mathcal{L}(P^n_0)} W - \frac{1}{2}<W>$.

Step 1: Since $<W>$ is deterministic it is clearly sufficient to prove that $M^n \xrightarrow{\mathcal{L}(P^n_0)} \frac{1}{2}W$. But this is immediate from proposition 3.1.5.4 under the assumptions that we have made.

Step 2: In order to prove this we will split Y^n into parts $Y^n = Y^{n,1} + Y^{n,2} + Y^{n,3}$ and we will prove that $\sup_t |Y_t^{n,i}| \xrightarrow{P_0^n} 0$ for $i = 1,2,3$.

Step 2.1: Let $Y_t^{n,1} = 2(H_t^n - \frac{1}{4} < W >_t)$. Observe that H^n and $< W >$ are increasing processes, that $< W >$ is continuous, and that t is an element of a compact set. Hence proving that $\sup |Y_t^{n,1}| \xrightarrow{P_0^n} 0$ is now equivalent to proving that $|Y_t^{n,1}| \xrightarrow{P_0^n} 0$ for all t, which follows by assumption.

Step 2.2.: Let $\epsilon > 0$. Write

$$Y_t^{n,2} = \int_0^t 1_{\{|\sqrt{\rho_s^n} - 1| < \epsilon\}} (\sqrt{\rho_s^n} - 1)^2 dm_s^n + \int_0^t 1_{\{|\sqrt{\rho_s^n} - 1| \geq \epsilon\}} (\sqrt{\rho_s^n} - 1)^2 dm_s^n$$

$$= Y_t^{n,2,\epsilon} + \overline{Y}_t^{n,2,\epsilon}$$

First we will prove step 2.2.1: $\sup |Y_t^{n,2,\epsilon}| \xrightarrow{P_0^n} 0$. Observe that $Y^{n,2,\epsilon}$ is a locally square integrable martingale, so we can apply Lenglart's inequality in the following way:

$$P_0^n(\sup_{s \leq t} |Y_s^{n,2,\epsilon}| \geq \delta) \leq \frac{\eta}{\delta^2} + P_0^n(< Y^{n,2,\epsilon} >_t \geq \eta)$$

Here

$$< Y^{n,2,\epsilon} >_t = \int_0^t 1_{\{|\sqrt{\rho_s^n} - 1| < \epsilon\}} (\sqrt{\rho_s^n} - 1)^4 \lambda_s^n ds$$

$$\leq \epsilon^2 H_t^n$$

Therefore

$$P_0^n(< Y^{n,2,\epsilon} >_t \geq \eta) \leq P(H_t^n \geq \frac{\eta}{\epsilon^2}) \leq$$

$$\leq P_0^n(|H_t^n - \frac{1}{4} < W >_t| \geq \frac{\eta}{2\epsilon^2}) + 1_{\{< W >_t \geq \frac{\eta}{8\epsilon^2}\}}.$$

By taking ϵ sufficiently small, the indicator disappears. Then by letting n tend to infinity, the last probability goes to zero. Since η is arbitrary step 2.2.1 has been proved.

Step 2.2.2: We have to prove that $\sup_t |\overline{Y}^{n,2,\epsilon}| \xrightarrow{(P_0^n)} 0$. Here we have the inequality

$$\sup_{t \leq T} |\overline{Y}_t^{n,2,\epsilon}| \leq \int_0^T 1_{\{\sqrt{\rho_s^n} - 1| \geq \epsilon\}} (\sqrt{\rho_s^n} - 1)^2 dN_s^n +$$

$$+ \int_0^T 1_{\{|\sqrt{\rho_s^n}-1|\geq\epsilon\}}(\sqrt{\rho_s^n}-1)^2\lambda_s^n ds. \tag{3.3}$$

The last term in (3.3) tends to zero in $\{P_0^n\}$ probability by assumption. For the first one we have for any $\delta<\epsilon^2$.

$$P_0^n(\int_0^T 1_{\{|\sqrt{\rho_s^n}-1|\geq\epsilon\}}(\sqrt{\rho_s^n}-1)^2 dN_s^n \geq\delta)\leq$$

$$P_0^n(\exists s\leq T:\ 1_{\{|\sqrt{\rho_s^n}-1|\geq\epsilon\}}\Delta N_s^n = 1) =$$

$$P_0^n(\exists s\leq T:|\Delta M_s^n|\geq\epsilon) = P_0^n(\sup_{s\leq T}|\Delta M_s^n|\geq\epsilon).$$

However this last probability tends to zero since M_n converges weakly to a continuous martingale (step 1), hence $\{M_n\}$ is C-tight and the claim follows from proposition 3.1.3.7. Step 2.2.2 has been proved.

Step 2.3: This proof is similar to the one of step 2.2. First we split

$$\int_0^t \psi(\sqrt{\rho_s^n})dN_s^n = \tag{3.4}$$

$$= \int_0^t 1_{\{|\sqrt{\rho_s^n}-1|\geq\epsilon\}}\psi(\sqrt{\rho_s^n})dN_s^n + \int_0^t 1_{\{|\sqrt{\rho_s^n}-1|<\epsilon\}}\psi(\sqrt{\rho_s^n})dN_s^n$$

For the first term on the right hand side of (3.4) we can apply the procedure of step 2.2.2. Notice that by a Taylor expansion

$$|\psi(x)|\leq\frac{2}{(1-\epsilon)^3}|x-1|^3 \text{ for } |x-1|\leq\epsilon.$$

Hence

$$\sup_{t\leq T}|\int_0^t 1_{\{|\sqrt{\rho_s^n}-1|<\epsilon\}}\psi(\sqrt{\rho_s^n})dN_s^n|\leq$$

$$\frac{2\epsilon}{(1-\epsilon)^3}\int_0^T 1_{\{|\sqrt{\rho_s^n}-1|<\epsilon\}}(\sqrt{\rho_s^n}-1)^2\lambda_s^n ds + \frac{2\epsilon}{(1-\epsilon)^3}\sup_{t\leq T}|Y_t^{n,2,\epsilon}| \tag{3.5}$$

where we have followed the notation in step 2.2. We have already proved that

$$\sup_{t\leq T}|Y_t^{n,2,\epsilon}|\xrightarrow{P_0^n}0.$$

The remaining term in (3.5) can be treated as in step 2.2.1. The theorem has been proved. □

PROPOSITION 3.2.7: *Under the assumption of theorem 3.2.6 we have*

$$2\int_0^t(\sqrt{\rho_s^n}-1)dm_s^n\xrightarrow{\mathcal{L}(P_0^n)}W \text{ iff } \int_0^t(\rho_s^n-1)dm_s^n\xrightarrow{\mathcal{L}(P_0^n)}W.$$

PROOF: Let $M_t^n = 2\int_0^t (\sqrt{\rho_s^n} - 1) dm_s^n$ and $\overline{M}_t^n = \int_0^t (\rho_s^n - 1) dm_s^n$ Then $\overline{M}_t^n - M_t^n = \int_0^t (\sqrt{\rho_s^n} - 1)^2 dm_s^n$. So the conclusion follows as soon as we have proved that

$$\sup_{s \leq t} |\overline{M}_s^n - M_s^n| \xrightarrow{P_0^n} 0,$$

but this is exactly step 2.2 in the proof of theorem 3.2.6 □ .

The next corollary is now immediate.

COROLLARY 3.2.8: *Let the assumptions of theorem 3.2.6 hold. Then*

$$Z_t^n = \exp\left(\int_0^t (\rho_s^n - 1) dm_s^n - \frac{1}{2} < W >_t + r_t^n\right),$$

where

$$\int_0^t (\rho_s^n - 1) dm_s^n \xrightarrow{\mathcal{L}(P_0^n)} W \text{ and } \sup_{s \leq t} |r_s^n| \xrightarrow{(P_0^n)} 0$$

REMARK: Observe that in corollary 3.2.8 we have given conditions such that $\overline{M}^n \xrightarrow{\mathcal{L}(P_0^n)} W$ and at the same time $Z^n = \mathcal{E}(\overline{M}^n) \xrightarrow{\mathcal{L}(P_0^n)} \mathcal{E}(W)$. Of course this procedure can be generalized to the situation where the processes Z^n loose their interpretation of being likelihood ratios. Furthermore it is noticed that corollary 3.2.8 can be proved directly by imposing the sufficient conditions for weak convergence of the sequence $\{\overline{M}^n\}$ as given in proposition 3.1.5.4. These are

$$\overline{H}_t^n = \int_0^t (\rho_s^n - 1)^2 \lambda_s^n ds \xrightarrow{P_0^n} < W >_t, \ \forall t > 0.$$

and

$$\overline{H}_t^{n,\epsilon} = \int_0^t 1_{\{|\rho_s^n - 1| \geq \epsilon\}} (\rho_s^n - 1)^2 \lambda_s^n ds \xrightarrow{P_0^n} 0, \ \forall t, \epsilon > 0.$$

A natural question is then to ask whether these conditions and those of theorem 3.2.6 coincide. The answer turns out to be affirmative. See theorem 3.2.12 below. This means that there is an alternative but equivalent sufficient condition for LAN available. In order to prove theorem 3.2.12 we need some auxiliary results.

LEMMA 3.2.9: *Let X be a locally integrable increasing process, $X_0 = 0$, and $\tilde{X}$ its dual predictable projection. Then for all bounded stopping times T one has*

$EX_T = E\tilde{X}_T \leq \infty$.

PROOF: Let $M = X - \tilde{X}$, then M is a local martingale. Observe that by by convention $X_0 = 0 \Rightarrow \tilde{X}_0 = 0$, so $M_0 = 0$. Let $\{T_n\}$ be a fundamental sequence for M. Then $M^n = \{M_{T_n \wedge t}\}_{t \geq 0}$ is a uniformly integrable martingale. Hence $EX_{T \wedge T_n} = E\tilde{X}_{T \wedge T_n} \leq EX_T \leq \infty$. Because $X_T = \lim X_{T \wedge T_n}$ we have from Fatou's lemma.

$$EX_T \leq \liminf EX_{T \wedge T_n} \leq E\tilde{X}_T.$$

Similarly $E\tilde{X}_T \leq EX_T$. □

PROPOSITION 3.2.10: *Let X and $\tilde{X}$ be as in lemma 3.2.9 and assume moreover that $\tilde{X}$ is continuous and that T is a bounded predictable stopping time. Then* $\forall \epsilon, \eta > 0$:

$$P(\tilde{X}_T \geq \epsilon) \leq \frac{\eta}{\epsilon} + P(X_T \geq \eta).$$

PROOF: From lemma 3.2.9 we obtain $E\tilde{X}_T = EX_T$.
Now $0 = \Delta\tilde{X}_T = E[\Delta X_T | \mathcal{F}_{T-}]$. Hence $E\Delta X_T = 0$ and we get $E\tilde{X}_T = EX_{T-}$. Because the process $\{X^-\}$ is predictable, application of Lenglart's inequelity (3.1.3.9) yields

$$P(\tilde{X}_t \geq \epsilon) \leq \frac{\eta}{\epsilon} + P(X_{T-} \geq \eta) \leq \frac{\eta}{\epsilon} + P(X_T \geq \eta)$$

because of the fact that X is increasing. □

REMARK: Trivially proposition 3.2.10 holds for a deterministic time t

PROPOSITION 3.2.11: *Let*

$$\tilde{H}_t^{n,\epsilon} = \int_0^t 1_{\{|\rho_s^n - 1| > \epsilon\}} (\sqrt{\rho_s^n} - 1)^2 \lambda_s^n ds$$

$$\hat{\tilde{H}}_t^{n,\epsilon} = \int_0^t 1_{\{|\rho_s^n - 1| \leq \epsilon\}} (\sqrt{\rho_s^n} - 1)^2 \lambda_s^n ds$$

$$\overline{H}_t^{n,\epsilon} = \int_0^t 1_{\{|\rho_s^n - 1| > \epsilon\}} (\rho_s^n - 1)^2 \lambda_s^n ds$$

$$\hat{\overline{H}}_t^{n,\epsilon} = \int_0^t 1_{\{|\rho_s^n - 1| \leq \epsilon\}} (\rho_s^n - 1)^2 \lambda_s^n ds$$

The statements (i) and (ii) are equivalent:

(i) $\tilde{H}_t^{n,\epsilon} \xrightarrow{P_0^n} 0$, $\hat{\tilde{H}}_t^{n,\epsilon} \xrightarrow{P_0^n} \frac{1}{4} < W >_t$, $\forall \epsilon, t > 0$.

(ii) $\overline{H}_t^{n,\epsilon} \xrightarrow{P_0^n} 0$, $\hat{\overline{H}}_t^{n,\epsilon} \xrightarrow{P_0^n} < W >_t$, $\forall \epsilon, t > 0$.

PROOF: The first thing that we prove is

$$\forall \epsilon>0,\ \forall t\geqslant 0:\ \tilde{H}_t^{n,\epsilon} \xrightarrow{P_0^n} 0 \Leftrightarrow \forall \epsilon>0,\ \forall t\geqslant 0:\ \bar{H}_t^{n,\epsilon} \xrightarrow{P_0^n} 0.$$

Since $(x-1)^2 = (\sqrt{x}-1)^2(\sqrt{x}+1)^2 \geqslant (\sqrt{x}-1)^2$ we have $\bar{H}_t^{n,\epsilon} \geqslant \tilde{H}_t^{n,\epsilon}$.
So we only have to prove that $\forall \epsilon>0, \forall t\geqslant 0$: $\tilde{H}_t^{n,\epsilon} \xrightarrow{P_0^n} 0$ implies

$\forall \epsilon>0,\ \forall t\geqslant 0$: $\bar{H}_t^{n,\epsilon} \xrightarrow{P_0^n} 0$. Define $X_t^{n,\epsilon} = \int_0^t 1_{\{|\rho_s^n-1|\geqslant\epsilon\}}(\sqrt{\rho_s^n}-1)^2 dN_s^n$ and

$\bar{X}_t^{n,\epsilon} = \int_0^t 1_{\{|\rho_s^n-1|\geqslant\epsilon\}}(\rho_s^n-1)^2 dN_s^n$.

Observe that $X^{n,\epsilon}$ has compensator $\tilde{H}^{n,\epsilon}$ and $\bar{X}^{n,\epsilon}$ has compensator $\bar{H}_t^{n,\epsilon}$. From Lenglart's inequelity we get

$$P_0^n(X_t^{n,\epsilon}\geqslant\delta) < \frac{\eta}{\delta} + P_0^n(\tilde{H}_t^{n,\epsilon}\geqslant\eta)$$

and from proposition 3.2.7

$$P_0^n(\tilde{H}_t^{n,\epsilon}\geqslant\delta) \leqslant \frac{\eta}{\delta} + P_0^n(X_t^{n,\epsilon}\geqslant\eta).$$

So $\tilde{H}_t^{n,\epsilon} \xrightarrow{P_0^n} 0$ if and only if $X_t^{n,\epsilon} \xrightarrow{P_0^n} 0$ and of course the same holds for $\bar{H}_t^{n,\epsilon}$ and $\bar{X}_t^{n,\epsilon}$. Therefore we have to prove the implication

$$\forall \epsilon>0,\ \forall t\geqslant 0:\ X_t^{n,\epsilon} \xrightarrow{P_0^n} 0 \Rightarrow \forall \epsilon>0,\ \forall t\geqslant 0: \bar{X}_t^{n,\epsilon} \xrightarrow{P_0^n} 0.$$

Let $\delta<\epsilon^2\wedge\epsilon_1^2$, where $\epsilon_1^2 = (1-\sqrt{1+\epsilon})^2$.

$$P_0^n(\bar{X}_t^{n,\epsilon}\geqslant\delta) = P_0^n(\sup_{s\leqslant t}(\rho_s^n-1)^2\Delta N_s^n>\epsilon^2)$$

$$\leqslant P_0^n(\sup_{s\leqslant t}(\sqrt{\rho_s^n}-1)^2\Delta N_s^n>\epsilon_1^2) = P_0^n(X_t^{n,\epsilon}\geqslant\epsilon_1^2)\leqslant P_0^n(X_t^{n,\epsilon}\geqslant\delta)$$

which tends to zero by assumption. One remark about the choice of δ. We can obviously restrict ourselves to small values of δ, because $P_0^n(\bar{X}_t^{n,\epsilon}\geqslant\delta)$ is decreasing if δ increases. The next point is showing that (i) implies

$$\forall \epsilon>0, t\geqslant 0: \hat{\bar{H}}_t^{n,\epsilon} \xrightarrow{P_0^n} <W>_t$$

Let

$$\delta^+(\epsilon) = \tfrac{1}{4}(\sqrt{1+\epsilon}+1)^2-1 \text{ and } \delta^-(\epsilon) = 1-\tfrac{1}{4}(\sqrt{1-\epsilon}+1)^2$$

Then

$$4(1-\delta^-(\epsilon))\hat{\tilde{H}}_t^{n,\epsilon} \leqslant \hat{\bar{H}}_t^{n,\epsilon} \leqslant 4((1+\delta^+(\epsilon))\hat{\tilde{H}}_t^{n,\epsilon}$$

Let $t,\eta,\epsilon>0$. Observe that $\lim_{\epsilon\downarrow 0}\delta^+(\epsilon)=\lim_{\epsilon\downarrow 0}\delta^-(\epsilon)=0$. Choose ϵ' such that $\epsilon'\leqslant\epsilon$ and

$$\delta^+(\epsilon')<W>_1 \leqslant \frac{\eta}{2} \text{ and } \delta^-(\epsilon')<W>_1 \leqslant \frac{\eta}{2}$$

Then

$$P_0^n(|\hat{\overline{H}}_t^{n,\epsilon} - <W>_t| \geq 2\eta)$$
$$\leq P_0^n(|\hat{\overline{H}}_t^{n,\epsilon} - \hat{\overline{H}}_t^{n,\epsilon'}| \geq \eta) + P_0^n(|\hat{\overline{H}}_t^{n,\epsilon'} - <W>_t| \geq \eta).$$

Now

$$P_0^n(|\hat{\overline{H}}_t^{n,\epsilon} - \hat{\overline{H}}_t^{n,\epsilon'}| \leq \eta) \leq P_0^n(\overline{H}_t^{n,\epsilon} \geq \eta),$$

which tends to zero from the first part of the proof.

$$P_0^n(|\hat{\overline{H}}_t^{n,\epsilon'} - <W>_t| \geq \eta) \leq$$
$$P_0^n(4(1+\delta^+(\epsilon'))(\hat{\overline{H}}_t^{n,\epsilon'} - \frac{1}{4}<W>_t) \geq \eta - \delta^+(\epsilon')<W>_1) \leq$$
$$P_0^n(4(1+\delta^+(\epsilon'))(\hat{\overline{H}}_t^{n,\epsilon'} - \frac{1}{4}<W>_t) \geq \frac{\eta}{2}),$$

which tends to zero because of (i). A similar inequality holds for $P_0^n(\hat{\overline{H}}_t^{n,\epsilon'} - <W>_t \leq -\eta)$ which completes this part of the proof. By a same way of reasoning one obtains that (ii) implies that
$\forall \epsilon > 0, t \geq 0$: $\hat{\overline{H}}_t^{n,\epsilon} \xrightarrow{P_0^n} \frac{1}{4}<W>_t$. □

THEOREM 3.2.12: *Let W be a Gaussian martingale with deterministic quadratic variation $<W>$. The conditions of theorem 3.2.6 are equivalent to*

$$\overline{H}_t^n = \int_0^t (\rho_s^n - 1)^2 \lambda_s^n ds \xrightarrow{P_0^n} <W>_t, \quad \forall t.$$

and

$$\overline{H}_t^{n,\epsilon} = \int_0^t 1_{\{|\rho_s^n - 1| \geq \epsilon\}} (\rho_s^n - 1)^2 \lambda_s^n ds \xrightarrow{P_0^n} 0, \quad \forall t, \epsilon > 0.$$

In either case we have the representation

$$Z_t^n = \exp\left(\int_0^t (\rho_s^n - 1) dm_s^n - \frac{1}{2}<W>_t + r_t^n\right),$$

where

$$\int_0^{\cdot} (\rho_s^n - 1) dm_s^n \xrightarrow{\mathcal{L}(P_0^n)} W \quad \text{and} \quad \sup_{s \leq t} |r_s^n| \xrightarrow{P_0^n} 0.$$

PROOF: The equivalence in the theorem easily follows from the fact that for $\epsilon < \frac{1}{2}$:

$$|\sqrt{x} - 1| < \epsilon \Rightarrow |x - 1| \leq 3\epsilon$$

and conversely that $|x-1|\leqslant\epsilon \Rightarrow |\sqrt{x}-1|\leqslant\epsilon$ and by applying proposition 3.2.11. The representation for Z^n now follows from corollary 3.2.8. □

Thus far we have established LAN for a sequence of counting processes $\{N^n\}$. It is also relevant to study LAN for a single counting process N where the time parameter t tends to infinity. It is possible to give sufficient conditions on the intensity process λ of N that ensures LAN, see for instance [24,27] or the slightly different conditions of proposition 3.2.18 below. However it may also be useful to see in the specific situation at hand whether LAN can be proved directly by inspecting whether the sufficient conditions of theorem 3.2.12 hold after a suitable transformation of the given process N to a sequence of counting processes $\{N^n\}$. We will carry out this last procedure for counting processes with a particular form of the intensity process. This will be done with an eye to the recursive estimators that will be discussed in section 4.2.

Assume that we are given a counting process N on a filtered probability space $(\Omega,\mathcal{F},\mathbb{F},P_\theta)$ such that it admits under the probability P_θ the Doob-Meyer decomposition

$$dN_t = \phi_t^T\theta dt + dm_t \quad N_0 = 0. \tag{3.6}$$

Here ϕ is a predictable process, $\phi:\Omega\times[0,\infty)\to[0,\infty)^d$ and $\theta\in[\epsilon,\infty)^d$ for some $\epsilon>0$. Let θ_0 be the "true" parameter. We will prove LAN in proposition 3.2.14 under the following assumption.

ASSUMPTION 3.2.13: Let

$$Q_t^{-1} = \int_0^t \frac{\phi_s\phi_s^T}{\phi_s^T\theta_0}ds.$$

Suppose that there exists an increasing function $g:[0,\infty)\to[0,\infty)$ with $g(t)\to\infty$ as $t\to\infty$ and a positive definite matrix $P=P^T\in\mathbb{R}^{d\times d}$ such that

$$P_{\theta_0} - \lim_{t\to\infty}\frac{1}{g(t)^2}P^{1/2}Q_t^{-1}P^{1/2} = I.$$

As said before we wish to prove LAN by applying theorem 3.2.12. Therefore we have to transform the above model to a different one that fits into the context of this theorem. This will be done as follows. Let $\{b_n\}\subset\mathbb{R}^+$, with $b_n\uparrow\infty$. Define $a_n=g(b_n)^2$. Without loss of generality we may assume that g is strictly increasing and $g\in C^1(\mathbb{R}^+)$. Then the same holds for its inverse function h. Define $h_n(t)=h(\sqrt{a_n}t)$. Our sequence of counting processes $\{N^n\}$ will now be defined via $N_t^n=N_{h_n(t)}, t\in[0,1]$. The relevant probability spaces $(\Omega^n,\mathcal{F}^n,\mathbb{F}^n,P_\theta^n)$ are now defined as $\Omega^n=\Omega,\mathcal{F}^n=\mathcal{F}$ for all n and $\mathcal{F}_t^n=\mathcal{F}_{h_n(t)}$, $P_\theta^n=P_\theta|\mathcal{F}_1^n=P_\theta|\mathcal{F}_{b_n}$. Denote by θ_0 the true parameter and write P_0^n for $P_{\theta_0}^n$. For $u\in\mathbb{R}^d$ we define $\theta^n=\theta_0+M_nu$, where $M_n=g(b_n)^{-1}P^{1/2}=a_n^{-1/2}P^{1/2}$. Write P_1^n for $P_{\theta^n}^n$. Under the measure P_0^n we have

$$N_t^n = \int_0^{h_n(t)}\phi_s^T\theta_0 ds + m_{h_n(t)}$$

$$= \int_0^t \lambda_s^n ds + m_t^n$$

Here $\lambda_t^n = \phi_{h_n(t)}^T \theta_0 h'_n(t)$ and m^n is a local martingale under the measure P_0^n. A similar expression holds for the decomposition of N_n under the measure P_1^n. In the notation that we have used before

$$\rho_t^n = \frac{\phi_{h_n(t)}^T \theta^n}{\phi_{h_n(t)}^T \theta_0} = \frac{(\phi_t^n)^T \theta^n}{(\phi_t^n)^T \theta_0}$$

where the definition of the process ϕ^n is obvious. Having introduced the relevant notation, we can write the Hellinger process H^n as

$$H_t^n = \int_0^{h_n(t)} (\sqrt{\tilde{\rho}_s^n} - 1)^2 \phi_s^T \theta_0 ds$$

where

$$\tilde{\rho}_t^n = \frac{\phi_t^T \theta^n}{\phi_t^T \theta_0}.$$

A similar expression holds for the process $H_t^{n,\delta}$ (which is $H_t^{n,\epsilon}$ with ϵ replaced by δ).

PROPOSITION 3.2.14: *Under assumption 3.2.13 the following hold:*

(i) $H_t^n \xrightarrow{P_0^n} \frac{1}{4} u^T u t$ and $H_t^{n,\delta} \xrightarrow{P_0^n} 0$, $\forall t \in [0,1], \forall \delta > 0$.

(ii) $\int_0^t (\sqrt{\tilde{\rho}_s^n} - 1)^2 \phi_s^T \theta_0 ds \xrightarrow{P_0} \frac{1}{4} u^T u$ as $t \to \infty$ and

$\int_0^t 1_{\{|\sqrt{\tilde{\rho}_s^n} - 1| > \delta\}} (\sqrt{\tilde{\rho}_s^n} - 1)^2 \phi_s^T \theta_0 ds \xrightarrow{P_0}$ as $t \to \infty$.

PROOF: Let $\delta > 0$ and choose N such that $|M_n u| < \delta\epsilon$ for $n \geq N$. Clearly such N exists by assumption 3.2.13. Then

$$|\sqrt{\tilde{\rho}_s^n} - 1| = \frac{|\tilde{\rho}_s^n - 1|}{\sqrt{\tilde{\rho}_s^n} + 1} \leq |\tilde{\rho}_s^n - 1| = \frac{|\phi_s^T M_n u|}{\phi_s^T \theta_0} \leq$$

$$\leq \frac{(\phi_s^T \phi_s)^{1/2} |M_n u|}{\phi_s^T \theta_0} \leq \frac{\phi_s^T \mathbf{1} |M_n u|}{\phi_s^T \mathbf{1} \cdot \epsilon} < \delta$$

where $\mathbf{1}$ is the column vector with all its elements equal to $+1$. So trivially we have that $H_t^{n,\delta} \xrightarrow{P_0^n} 0$. In order to prove the convergence of H_t^n, take δ and N as above. Observe that we can write

$$H_t^n = u^T M_n \int_0^{h_n(t)} \frac{\phi_s \phi_s^T}{\phi_s^T \theta_0} (\sqrt{\tilde{\rho}_s^n} + 1)^{-2} ds \; M_n u \tag{3.7}$$

Since we have for $n \geq N$ $|\tilde{\rho}_s^n - 1| < \delta$ (see above) we get the inequalities $1-\delta \leq \sqrt{\tilde{\rho}_s^n} \leq 1+\frac{1}{2}\delta$, which we use to obtain from (3.7)

$$(2+\tfrac{1}{2}\delta)^{-2} u^T M_n Q_{h_n(t)}^{-1} M_n u \leq H_t^n \leq (2-\delta)^{-2} u^T M_n Q_{h_n(t)}^{-1} M_n u \tag{3.8}$$

Now we can write

$$\begin{aligned} M_n Q_{h_n(t)}^{-1} M_n &= a_n^{-1} g^2(h_n(t)) g^{-2}(h_n(t)) P^{1/2} Q_{h_n(t)}^{-1} P^{1/2} \\ &= t g^{-2}(h_n(t)) P^{1/2} Q_{h_n(t)} P^{1/2} \end{aligned} \tag{3.9}$$

which by assumption 3.2.13 tends to tI in P_0^n-probability. Since δ is arbitrary the assertion follows from (3.8). The second part follows from (i) by taking $t=1$ and by taking any sequence $\{b_n\}$ with $b_n \to \infty$. □

REMARK: The second part of the above proposition can be proved under a considerably weaker assumption than 3.2.13. If one has instead:

ASSUMPTION 3.2.15: There exists a map $M:[0,\infty) \to \mathbb{R}^{d \times d}$ such that $M(t)=M(t)^T>0$, $M(t) \to 0$ for $t \to \infty$ and

$$P_{\theta_0} - \lim_{t \to \infty} M(t) Q_t^{-1} M(t) = I.$$

The proof is then as follows. In the notation we have used above we replace $h_n(t)$ by t and arrive at $M(t)Q_t^{-1}M(t)$ instead of (3.9).

Now we are in the position to prove LAN for the model (3.6) (see definition 3.2.3).

COROLLARY 3.2.16: *Consider the model (3.6). Let assumption 3.2.15 hold. Define* $\theta^t = \theta_0 + M(t)u$ *and*

$$\rho_s^t = \frac{\phi_s^T \theta^t}{\phi_s^T \theta_0}.$$

Define

$$Z_t = \exp\left(\int_0^t \log \rho_s^t \, dN_s - \int_0^t (\rho_s^t - 1) \phi_s^T \theta_0 \, ds\right)$$

Then

$$Z_t = \exp(u^T \Delta_t - \tfrac{1}{2} u^T u + r_t),$$

where $\Delta_t = M(t) \int_0^t \frac{\phi_s}{\phi_s^T \theta_0} \, dm_s$, $\Delta_t \xrightarrow{\mathcal{L}(P_0^n)} N(0,I)$ *and* $r_t \xrightarrow{P_0^n} 0$.

PROOF: This is now a direct consequence of corollary 3.2.8, proposition 3.2.14

and by noticing that

$$\rho_s^t - 1 = \frac{\phi_s^T M(t) u}{\phi_s^T \theta_0} . \quad \square$$

It is also possible to state LAN for a counting process that admits an intensity process $\{\lambda_t(\theta)\}$ under the probability P_θ by giving conditions on $\lambda_t(\theta)$ directly. We will do this for a scalar parameter θ. These conditions (assumption 3.2.17 below) are close to those given in [24].

ASSUMPTION 3.2.17: There exists a function $M:[0,\infty)\to\mathbb{R}$ such that $M(t)\to 0$ and

(i) $P_{\theta_0} - \lim_{t\to\infty} M(t)^2 \int_0^t \frac{\lambda'_s(\theta_0)^2}{\lambda_s(\theta_0)}\, ds = 1$

(ii) $P_{\theta_0} - \lim_{t\to\infty} M(t)^2 \int_0^t 1_{\{|\frac{\lambda'_s(\theta_0)}{\lambda_s(\theta_0)}| \geq \epsilon M(t)^{-1}\}} \frac{\lambda'_s(\theta_0)^2}{\lambda'_s(\theta_0)} ds = 0$

(iii) $P_{\theta_0} - \lim_{t\to\infty} M(t)^\alpha \sup_{s\leq t} \sup_{|\theta-\theta_0|\leq CM(t)} |\frac{\lambda''_s(\theta)}{\lambda'_s(\theta_0)}| = 0$ for some $\alpha\in(0,1)$ and $\forall C>0$.

PROPOSITION 3.2.18: *Let assumption 3.2.17 hold. Then the family $\{P_\theta\}$ is LAN at the point θ_0 with normalizing sequence $\{M(t)\}$.*

PROOF: We have to check the conditions of theorem 3.2.6, or those concerning $\overline{H}_t$ and $\overline{H}_t^\epsilon$ in view of theorem 3.2.12

$$\overline{H}_t = \int_0^t (\rho_s^t - 1)^2 \lambda_s(\theta_0) ds, \text{ where } \rho_s^t = \frac{\lambda_s(\theta_t)}{\lambda_s(\theta_0)}, \ \theta_t = \theta_0 + M(t)u.$$

By a Taylor expansion

$$\lambda_s(\theta_t) = \lambda_s(\theta_0) + (\theta_t - \theta_0)\lambda'_s(\theta_0) + \tfrac{1}{2}(\theta_t - \theta_0)^2 \lambda''_s(\tilde{\theta}_t)$$

where $\tilde{\theta}_t$ is between θ_0 and $\tilde{\theta}_t$. Hence

$$\lambda_s(\theta_t) - \lambda_s(\theta_0) = (\theta_t - \theta_0)\lambda'_s(\theta_0)(1 + \tfrac{1}{2}(\theta_t - \theta_0)\frac{\lambda''_s(\tilde{\theta}_t)}{\lambda'_s(\theta_0)})$$

$$= (\theta_t - \theta_0)\lambda'_s(\theta_0)(1 + M(t)^{1-\alpha}\epsilon_s^t)$$

where

$$P_{\theta_0} - \lim_{t\to\infty} \sup_{s\leq t} |\epsilon_s^t| = 0$$

by assumption 3.2.17 (iii). Hence

$$\overline{H}_t = \int_0^t (\theta_t - \theta_0)^2 \frac{\lambda'_s(\theta_0)^2}{\lambda_s(\theta_0)} (1 + M(t)^{1-\alpha}\epsilon_s^t)^2 ds$$

$$= u^2M(t)^2[\int_0^t \frac{\lambda'_s(\theta_0)^2}{\lambda_s(\theta_0)}(1+2M(t)^{1-\alpha}\epsilon_s^t + (M(t)^{1-\alpha}\epsilon_s^t)^2)ds$$

$$= u^2M^2(t)\int_0^t \frac{\lambda'_s(\theta_0)^2}{\lambda_s(\theta_0)}ds + R_{1t} + R_{2t}.$$

From assumption 3.2.17 (i) we see that we are left to verify that both R_{1t} and R_{2t} tend to zero in P_{θ_0}-probability. We will prove this for R_{1t} only, since for R_{2t} the procedure is the same. For R_{1t} this immediately follows from the next inequality

$$|R_{1t}| \leqslant 2u^2M(t)^2\int_0^t \frac{\lambda'_s(\theta_0)^2}{\lambda_s(\theta_0)}ds \cdot M(t)^{1-\alpha} \sup_{s\leqslant t}|\epsilon_s^t|.$$

The last thing that we have to show is that $\overline{H}_t^\epsilon \xrightarrow{P_{\theta_0}} 0$, where

$$\overline{H}_t^\epsilon = \int_0^t 1_{\{|\rho_s^t - 1| \geqslant \epsilon\}}(\rho_s^t - 1)^2\lambda_s(\theta_0)ds.$$

Let $\epsilon_t = \sup_{s\leqslant t}|\epsilon_s^t|$. Then

$$|\rho_s^t - 1| \leqslant |uM(t)\frac{\lambda'_s(\theta_0)}{\lambda_s(\theta_0)}|(1 + M(t)^{1-\alpha}\epsilon_t)$$

Hence

$$\overline{H}_t^\epsilon \leqslant 4u^2M(t)^2\int_0^t 1_{\{|\frac{\lambda'_s(\theta_0)}{\lambda_s(\theta_0)}| \geqslant \frac{\epsilon}{2u}M(t)^{-1}, M(t)^{1-\alpha}\epsilon_t \leqslant 1\}}\frac{\lambda'_s(\theta_0)^2}{\lambda_s(\theta_0)}ds$$

$$+ u^2M(t)^2\int_0^t 1_{\{M(t)^{1-\alpha}\epsilon_t > 1\}}\frac{\lambda'_s(\theta_0)^2}{\lambda_s(\theta_0)}(1 + M(t)^{1-\alpha}\epsilon_t)^2ds \qquad (3.10)$$

The first term in the right hand side of (3.10) tends to zero because of assumption 3.2.17 (ii), whereas the last term tends to zero since $\overline{H}_t \to u^2$ and $P_{\theta_0}(M(t)^{1-\alpha}\epsilon_t > 1) \to 0$. □

Chapter 4

Recursive Estimation

In this chapter we will treat the problem of recursively estimating an unknown parameter that occurs in the intensity process associated with a given counting process. Contrary to the case where off-line procedures are studied, there are hardly any results for the recursive estimation problem except for a first attempt by VERE-JONES [47] and NIKITIN and SNEGOVOY [31]. Maximum likelihood estimation (off-line) has been treated by various authors such as LIN'KOV [27] KUTOYANTS [23], OGATA [32], and KONECNY [21,22]. They proved that under conditions that differ from paper to paper, the maximum likelihood estimator has desirable properties like consistency, asymptotic normality and efficiency. These properties still have to be investigated for recursive estimators. In sections 4.1, 4.2 and 4.3 we will do this for a rather specific case, viz. the case where the intensity process exhibits a linear structure. The first basic problem one encounters is that of designing a recursive procedure for parameter estimation. One of the justifications that we give for a certain choice of such an algorithm is based on the asymptotic expression of the likelihood ratio process for counting processes, which has been studied in section 3.2. We present a number of recursive parameter estimation algorithms for counting processes that admit an intensity process that is linear in the parameter. This is the model that we have encountered in section 3.2 (see equation (3.6)). Recall that this model is given by

$$dN_t = \phi_t^T \theta_0 dt + dm_t \tag{4.1}$$

where N is our counting process, ϕ is a predictable process, $\phi:[0,\infty)\times\Omega\to\mathbb{R}^d$ and $\theta_0 \in \mathbb{R}^d$. Occasionally we will need some additional requirements for ϕ or θ_0.

4.1 RECURSIVE LEAST SQUARES ESTIMATION

In this section we will study the model (4.1). So $dN_t = \phi_t^T \theta_0 dt + dm_t$. The least squares estimator $\hat{\theta}_t$ by definition minimizes the quadratic form in θ

$$V_t(\theta) = \int_0^t (\phi_s^T \theta)^2 ds - 2\int_0^t \phi_s^T \theta dN_s \tag{4.2}$$

For a heuristic justification of minimizing the criterion $V_t(\theta)$, observe that it is formally equivalent to minimizing the undefined expression

$$\int_0^t (\phi_s^T \theta - \frac{dN_s}{ds})^2 ds.$$

Assume for a moment that the matrix $\int_0^t \phi_s \phi_s^T ds$ is invertible, and denote by R_t its inverse. Put $\bar{\Phi}_t = \int_0^t \phi_s dN_s$. Then eq. (4.2) reads

$$V_t(\theta) = (\theta - R_t\bar{\Phi}_t)^T R_t^{-1}(\theta - R_t\bar{\Phi}_t) - \bar{\Phi}_t R_t \bar{\Phi}_t \tag{4.3}$$

From (4.3) we see that $\hat{\theta}_t = R_t\bar{\Phi}_t$ minimizes $V_t(\theta)$. Because $\frac{d}{dt}R_t = -R_t\phi_t\phi_t^T R_t$ we get by applying the stochastic calculus rule to the product $R_t\bar{\Phi}_t$ for $\hat{\theta}_t$ the equation

$$d\hat{\theta}_t = R_t\phi_t(dN_t - \phi_t^T\hat{\theta}_t dt) \tag{4.4}$$

However it is a priori not clear that the matrix R_t as introduced above is well defined. Therefore we will modify its definition slightly. Let R_0^{-1} be a strictly positive definite matrix. Define now R_t to be the inverse of $R_0^{-1} + \int_0^t \phi_s \phi_s^T ds$. One immediately sees that now R_t is well defined. Usually one will wish R_0^{-1} to be small in norm. Observe also that $R_0 = (R_0^{-1})^{-1}$, which makes the notation consistent. As before we still have

$$\frac{d}{dt}R_t = -R_t\phi_t\phi_t R_t dt.$$

Furthermore let $\hat{\theta}_0$ be any vector in $\mathbb{R}_+^d$. Now we are in the position to define a recursive (least squares) parameter estimation algorithm as the following couple of stochastic differential equations together with their initial values

$$d\hat{\theta}_t = R_t\phi_t(dN_t - \phi_t^T\hat{\theta}_t dt), \quad \hat{\theta}_0 \tag{4.5a}$$

$$dR_t = -R_t\phi_t\phi_t R_t dt, \quad R_0 \tag{4.5b}$$

The equations (4.5a,b) will be referred to as least squares algorithm. Observe that this system of equations has a unique global solution since N has no explosions.

REMARK: The algorithm (4.5) is invariant under non-singular linear transformations in the following sense. Let $S \in \mathbb{R}^{d\times d}$ be a non-singular matrix. Write $\eta = S\theta, \hat{\eta}_t = S\hat{\theta}_t, \xi_t = S^{-T}\phi_t$ and $T_t = SR_tS^T$. Then (4.5) transforms into

$$d\hat{\eta}_t = T_t\xi_t(dN_t - \xi_t^T\hat{\eta}_t dt), \quad \hat{\eta}_0$$

$$dT_t = -T_t\xi_t\xi_t^T T_t dt, \quad T_0$$

which is exactly the least squares algorithm that corresponds to $dN_t = \xi_t^T\eta dt + dm_t$, but this is nothing else but (4.1) because $\xi_t^T\eta = \phi_t^T\theta$.

4.1.1 Convergence of the least squares algorithm

In proving almost sure convergence of the estimators $\{\hat{\theta}_t\}$ defined by (4.5a, b) we will use the following lemma, which is nothing else but 2.2.20 for nonnegative x. We restate it for convenience. Compare also to [39,42].

LEMMA 4.1.1.1: *Let x,a,b be nonnegative stochastic processes and m a local martingale such that $x = a - b + m$. Assume that a and b are increasing processes,*

$a_0=b_0=0$ *and that* $\lim_{t\to\infty} a_t<\infty$ *a.s. Then*
(i) $\lim_{t\to\infty} x_t$ *exists and is finite a.s.*
(ii) $\lim_{t\to\infty} b_t$ *is finite a.s.*

THEOREM 4.1.1.2 [45]: *Consider the algorithm (4.5). Let* θ_0 *be the true parameter value. Let* $\tilde{\theta}_t=\hat{\theta}_t-\theta_0$ *and let* $\psi_t=\phi_t^T\phi_t, \Psi_t=\int_0^t \psi_s ds + tr(R_0^{-1})$.

Assume:
(i) $\lim_{t\to\infty} \Psi_t=\infty$ *a.s.*
(ii) $\int_0^\infty \Psi_t^{-2}\psi_t\phi_t dt<\infty$ *a.s.*
(iii) $\lim_{t\to\infty} \Psi_t^{-1}\int_0^t \phi_s\phi_s^T ds=C$, *where* $C\in\mathbb{R}^{m\times m}$ *is positive definite a.s.*
Then
(a) $\lim_{t\to\infty} \hat{\theta}_t=\theta_0$ *a.s.*
(b) $\lim_{t\to\infty} \Psi_t^{-1}\int_0^t (\phi_s^T\tilde{\theta}_s)^2 ds=0$ *a.s.*

PROOF: From (4.5) it follows that

$$d\tilde{\theta}_t=R_t\phi_{t-}(dN_t-\phi_t^T\hat{\theta}_t dt)=R_t\phi_{t-}(dm_t-\phi_t^T\tilde{\theta}dt)$$
$$dR_t^{-1}=\phi_t\phi_t^T dt$$

Define the Lyapunov like process $u_t=\tilde{\theta}_t^T R_t^{-1}\tilde{\theta}_t+\int_0^t(\tilde{\theta}_s^T\phi_s)^2 ds$. Applying the stochastic calculus rule to u_t, we obtain

$$du_t=2\tilde{\theta}_{t-}^T\phi_t dm_t+\phi_t^T R_t\phi_t dN_t.$$

Observe that $\Psi_t=tr(R_t^{-1})$. Define $w_t=u_t\Psi_t^{-1}$, then

$$dw_t=-\Psi_t^{-1}w_t\psi_t dt+\phi_t^T R_t\phi_t\Psi_t^{-1}\theta_0^T\phi_t dt+dm_{1t}, \tag{4.6}$$

where m_1 is local martingale. We want now to apply lemma 4.1.1.1 to equation (4.6). Because u,w,Ψ are positive, we then see that the only thing we have to check is

$$\int_0^\infty \phi_t^T R_t\phi_t\Psi_t^{-1}\theta_0^T\phi_t dt<\infty.$$

To that end, let $\rho_t=trR_t$. Let γ_{it} be one of the eigenvalues of R_t^{-1}, then $\lim_{t\to\infty}\Psi_t^{-1}\gamma_{it}=c_i>0$ by assumption (iii) of the theorem. Hence $\gamma_{it}=c_i\Psi_t(1+o(1))$ $(t\to\infty)$. Now γ_{it}^{-1} is an eigenvalue of R_t, $\gamma_{it}^{-1}=c_i^{-1}\Psi_t(1+o(1)),(t\to\infty)$. Hence

$$\rho_t = \Psi_t^{-1}(\Sigma c_i^{-1} + o(1))(t \to \infty), \quad \text{or} \quad \rho_t = O(\Psi_t^{-1}), (t \to \infty).$$

Recall that for a positive definite matrix A, $x^T A x \leq x^T x . tr(A)$ and $x^T A^2 x \leq x^T x (tr(A))^2$. Then

$$\int_0^\infty \phi_t^T R_t \Psi_t^{-1} \phi_t \theta_0^T \phi_t dt = \int_0^\infty \phi_t^T R_t R_t^{-1} R_t \phi_t \Psi_t^{-1} \theta_0^T \phi_t dt \leq$$

$$\leq \int_0^\infty \phi_t R_t^2 \phi_t \theta_0^T \phi_t dt \leq \int_0^\infty \phi_t^T \phi_t \rho_t^2 \theta_0^T \phi_t dt =$$

$$\theta_0^T \int_0^\infty \psi_t \rho_t^2 \phi_t dt = \theta_0^T \int_0^\infty \phi_t \psi_t O(\Psi_t^{-2}) dt < \infty, \text{ by assumption (ii).}$$

Then from lemma 4.1.1.1 we conclude that w and $\int_0^\infty w_s \Psi_s^{-1} \psi_s ds$ almost surely converge. We claim that $\lim_{t\to\infty} w_t = 0$ a.s. If not, there exists a subset of Ω with positive probability and an $\epsilon > 0$, such that $\lim_{t\to\infty} w_t \geq 2\epsilon$ on this subset. But then we also have on the same subset

$$\int_0^\infty \Psi_t^{-1} w_t \psi_t dt \geq \epsilon \int_0^\infty \Psi_t^{-1} \psi_t dt = \left[\log(\Psi_t)\right]_0^\infty = \infty, \text{ by assumption (i).}$$

But this contradicts the second assertion of lemma 4.1.1.1. Since w is the sum of two positive quantities we have both

$$\lim_{t\to\infty} \Psi_t^{-1} \int_0^t (\tilde{\theta}_s^T \phi_s)^2 ds = 0 \text{ a.s. and}$$

$$\lim_{t\to\infty} \tilde{\theta}_t^T \frac{R_t^{-1}}{\Psi_t} \tilde{\theta}_t = 0 \text{ a.s.}$$

Because of assumption (iii) we know that $\liminf_{t\to\infty} \Psi_t^{-1} R_t^{-1} = C > 0$, hence $\lim_{t\to\infty} \tilde{\theta}_t = 0$ a.s.

REMARK: It is possible the relax the third assumption of theorem 4.1.1.2 to the one in [15]. The analysis of the algorithm then becomes a bit more complicated. We will not discuss this. However we will follow a similar procedure in section 4.3 for a different algorithm. We will give some examples for which the assumption of theorem 4.1.1.2 hold.

EXAMPLE 1: Let $\phi:[0,\infty) \to \mathbb{R}_+^2$, $\phi_t = [1, 1+\sin t]$, $\theta = [a,b]^T$. Then

$$\Psi_t = \frac{5}{2}t - 2\cos t - \frac{1}{4}\sin 2t + tr(R_0^{-1}).$$

Clearly assumptions 4.1.1.1 i) and ii) are satisfied and

$$\lim_{t\to\infty} \Psi_t^{-1} \int_0^t \phi_s \phi_s^T ds = \lim_{t\to\infty} \frac{2}{5t} \int_0^t \begin{bmatrix} 1 & 1+\sin s \\ 1+\sin s & \frac{3}{2} + 2\sin s - \frac{1}{2}\cos 2s \end{bmatrix} ds$$

$$= \frac{1}{5}\begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$$

EXAMPLE 2. Let $\phi: T \times \Omega \to \mathbb{R}^2$, $\phi_t = (1, 1+(-1)^{N_t})$ and $\theta = (a,b) \in \mathbb{R}^2_+$. As in [11, p. 59] the second component of ϕ jumps like a random telegraph process. Conditions (i) and (ii) of theorem 4.1.1.2 are easily verified. To check conditions (iii), let us first define

$$X_t = t^{-1} \int_0^t (-1)^{N_s} ds$$

Then

$$\Psi_t^{-2} \int_0^t \phi_s \phi_s^T ds = (3 + t^{-1} tr(R_0^{-1}) + 2X_t)^{-1} \begin{bmatrix} 1 & 1+X_t \\ 1+X_t & 2+2X_t \end{bmatrix}.$$

We now proceed to compute $as - \lim_{t \to \infty} X_t$. Since $N_t = (a+b)t + btX_t + m_t$, we find that

$$X_t = b^{-1}(t^{-1}N_t - t^{-1}m_t - a - b)$$

The quadratic variation process $<m>_t = (a+b)t + btX_t \leqslant (a+2b)t$. It then follows from the strong law of large numbers for martingales that $t^{-1}m_t \to 0$ a.s. Finally we have to evaluate the asymptotic behaviour of $t^{-1}N_t$. Define $T_k = \inf\{t \geqslant 0 : N_t = k\}$. Then

$$\sum_{k=0}^{\infty} \frac{k}{T_{k+1}} 1_{\{T_k \leqslant t < T_{k+1}\}} \leqslant t^{-1} N_t \leqslant \sum_{k=0}^{\infty} \frac{k}{T_k} 1_{\{T_k \leqslant t < T_{k+1}\}}.$$

Consequently

$$as - \lim_{t \to \infty} t^{-1} N_t = as - \lim_{t \to \infty} \frac{k}{T_k}.$$

Let $\tau_j = T_j - T_{j-1}, j = 1,2,\ldots$. then $\{\tau_j\}$ is a sequence of independent random variables, and $E\tau_{2l} = a^{-1}, E\tau_{2l+1} = (a+2b)^{-1}$. Now the strong law of large numbers for independent random variables applies and we get:

$$as - \lim_{t \to \infty} \frac{T_k}{k} = as - \lim_{t \to \infty} \frac{1}{k} \sum_{j=1}^{k} \tau_j = \frac{1}{2}\left[\frac{1}{a} + \frac{1}{a+2b}\right] = \frac{a+b}{a(a+2b)}.$$

Collecting the above results we find

$$as - \lim_{t \to \infty} X_t = \frac{1}{b}\left[\frac{a(a+2b)}{a+b} - a - b\right] = -\frac{b}{a+b}.$$

Conclusion:

$$as - \lim_{t \to \infty} \Psi_t^{-1} \int_0^t \phi_s \phi_s^T ds = \frac{1}{3a+b} \begin{bmatrix} a+b & a \\ a & 2a \end{bmatrix} > 0.$$

EXAMPLE 3: Let X be a Markov process which takes its values in $\{0,1\}$.

Assume that the holding times in 0 and 1 are exponentially distributed with means μ_0 and μ_1 respectively. Assume that N_t has intensity $\theta_1 X_{t-} + \theta_0(1-X_{t-})$ which is left continuous, thus predictable. So $\phi_t = [X_{t-}, 1-X_{t-}]^T$. Now $\Psi_t = t + tr(R_0^{-1})$. Again assumptions 4.1.1.2 i, ii are easy to verify and

$$\lim_{t\to\infty} \frac{1}{\Psi_t} \int_0^t \phi_s \phi_s^T ds = \lim_{t\to\infty} \frac{1}{t} \int_0^t \begin{bmatrix} X_s & 0 \\ 0 & 1-X_s \end{bmatrix} ds = \frac{1}{\mu_1+\mu_0} \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_0 \end{bmatrix}$$

4.1.2 Asymptotic distribution of the least squares algorithm

In this section we will show that the algorithm (4.5) provides us with estimators $\hat{\theta}_t$ that are asymptotically normally distributed if we impose some additional requirements on the process ϕ. We use some of the central limit theorems of section 3.2. It immediately follows from (4.5) that

$$\hat{\theta}_t = R_t[R_0^{-1}\hat{\theta}_0 + \int_0^t \phi_s dN_s]$$

and

$$\tilde{\theta}_t = \hat{\theta}_t - \theta_0 = R_t[R_0^{-1}(\hat{\theta}_0 - \theta_0) + \int_0^t \phi_s dm_s]$$

Introduce the vector valued martingale

$$M_t = \int_0^t \phi_s dm_s \tag{4.7}$$

then

$$<M>_t = \int_0^t \phi_s \phi_s^T \phi_s^T \theta_0 ds$$

Clearly the distributions of $\hat{\theta}_t$ and $\tilde{\theta}_t$ are governed by the ones of R_t and M_t. For the latter we have the following result.

THEOREM 4.1.2.1: *Let M be as defined in (4.7). Assume that there exists a function μ:$[0,\infty)\to[0,\infty)$ with $\mu(t)\to\infty$ as $t\to\infty$ such that*

(i) $P-\lim_{t\to\infty} \mu(t)^{-1} <M>_t = D$, *where $D \in \mathbb{R}^{d\times d}$ is a positive definite non random matrix*

(ii) $P-\lim_{t\to\infty} \mu(t)^{-1} \int_0^t \phi_s^T \phi_s I_{\{\phi_s^T \phi_s > \epsilon\mu(t)\}} \phi_s^T \theta_0 ds = 0, \forall \epsilon > 0.$

Then

$$<M>_t^{-\frac{1}{2}} M_t \xrightarrow{\mathcal{L}} N(0,I).$$

PROOF: i) Let $\lambda \in \mathbb{R}^d$ and define $M_t^\lambda = \lambda^T D^{-\frac{1}{2}} M_t = \int_0^t \lambda^T D^{-\frac{1}{2}} \phi_s dm_s$. Then

$$<M^\lambda>_t = \lambda^T D^{-\frac{1}{2}} <M>_t D^{-\frac{1}{2}} \lambda$$

Hence

$$(\lambda^T \lambda \mu(t))^{-1} <M^\lambda>_t = (\lambda^T \lambda)^{-1} \lambda^T D^{-\frac{1}{2}} \mu(t)^{-1} <M>_t D^{-\frac{1}{2}} \lambda \to 1$$

in probability. Hence condition *i* in proposition 3.1.6.3 is satisfied with $g(t)^2 = \lambda^T \lambda \mu(t)$. In order to establish condition 3.1.6.3 ii) for $f_t = \lambda^T D^{-\frac{1}{2}} \phi_t$ we compute

$$(\lambda^T \lambda \mu(t))^{-1} \int_0^t \lambda^T D^{-\frac{1}{2}} \phi_s \phi_s^T D^{-\frac{1}{2}} \lambda I_{\{|\lambda^T D^{-\frac{1}{2}} \phi_s| > \epsilon (\lambda^T \lambda \mu(t))^{\frac{1}{2}}\}} \phi_s^T \theta_0 ds =$$

$$(\lambda^T \lambda \mu(t))^{-1} \lambda^T D^{-\frac{1}{2}} \int_0^t \phi_s^T \phi_s I_{\{\lambda^T D^{-\frac{1}{2}} \phi_s \phi_s^T D^{-\frac{1}{2}} \lambda > \epsilon^2 \lambda^T \lambda \mu(t)\}} \phi_s^T \theta_0 ds D^{-\frac{1}{2}} \lambda \leqslant$$

$$(\lambda^T \lambda \mu(t))^{-1} \lambda^T D^{-1} \lambda \int_0^t \phi_s^T \phi_s I_{\{\lambda^T D^{-1} \lambda \phi_s^T \phi_s > \epsilon^2 \lambda \mu(t)\}} \phi_s^T \theta_0 ds$$

which tends to zero in probability according to assumption ii since we can replace ϵ by $\epsilon^2 \lambda^T \lambda (\lambda^T D^{-1} \lambda)^{-1}$. Now we have proved

$$(\lambda^T \lambda \mu(t))^{-\frac{1}{2}} M_t^\lambda \xrightarrow{\mathcal{L}} N(0,1)$$

(ii) According to the Cramer-Wold device

$$\forall \lambda \in \mathbb{R}^d : (\lambda^T \lambda \mu(t))^{-\frac{1}{2}} M_t^\lambda \xrightarrow{\mathcal{L}} N(0,1)$$

if and only if

$$\mu(t) D^{-\frac{1}{2}} M_t \xrightarrow{\mathcal{L}} N(0, I).$$

Since

$$\mu(t)^{-\frac{1}{2}} D^{-\frac{1}{2}} M_t = (D^{-\frac{1}{2}} \mu(t)^{-\frac{1}{2}} <M>_t^{\frac{1}{2}}) <M>_t^{-\frac{1}{2}} M_t$$

and

$$D^{-\frac{1}{2}} \mu(t)^{-\frac{1}{2}} <M>_t^{\frac{1}{2}} \to I$$

in probability, we have finished the proof.

REMARK: Stronger conditions than 4.1.2.1 (ii) are the corresponding Lindeberg or Lyapunov conditions

$$\forall \epsilon > 0 : \mu(t)^{-1} E \int_0^t \phi_s^T \phi_s I_{\{\phi_s^T \phi_s > \epsilon \mu(t)\}} \phi_s^T \theta_0 ds \to 0$$

$$\exists \delta > 0 : \mu(t)^{-1-\delta/2} E \int_0^t \|\phi_s\|^{3+\delta} ds \to 0,$$

where $\|\cdot\|$ denotes the (Euclidean) norm on $\mathbb{R}^d$.

COROLLARY 4.1.2.2: *Under the conditions of theorem 4.1.2.1 we have*

$$<M>_t^{-\frac{1}{2}} R_t^{-1}(\hat{\theta}_t - \theta_0) \xrightarrow{\mathcal{L}} N(0,I)$$

PROOF:

$$<M>_t^{-\frac{1}{2}} R_t^{-1}(\hat{\theta}_t - \theta_0) = <M>_t^{-\frac{1}{2}} [R_0^{-1}(\hat{\theta}_0 - \theta_0) + M_t].$$

The fact that $<M>_t^{-\frac{1}{2}} R_0^{-1}(\hat{\theta}_0 - \theta_0) \to 0$ in probability (this follows from 4.1.2.1 i) gives us the desired result.

REMARK: $<M>_t$ depends on the unknown parameter θ_0. As usual we can estimate $<M>_t$ by substituting $\hat{\theta}_t$, which is strongly consistent, for θ_0.

The examples given below are continuations of examples 1-3 of section 4.1.1.

EXAMPLE 1: $\phi(t) = [1, 1+\sin t]^T, \theta_0 = (\theta_1, \theta_2)$. Take $\mu(t) = t$. Then we can calculate

$$\lim_{t\to\infty} \mu(t)^{-1} <M>_t = \begin{bmatrix} \theta_1 + \theta_2 & \theta_1 + \frac{3}{2}\theta_2 \\ \theta_1 + \frac{3}{2}\theta_2 & \frac{3}{2}\theta_1 + \frac{5}{2}\theta_2 \end{bmatrix}$$

which is a positive definite matrix. So assumption 4.1.2.1 (i) is satisfied. To establish that assumption 4.1.2.1 (ii) holds it is sufficient to remark that $\phi_s^T \phi_s \leq 5$. Hence for $t > \frac{5}{\epsilon}$ we have

$$I_{\{\phi_s^T \phi_s > \epsilon t\}} = 0.$$

Another calculation shows that we have asymptotically

$$(\hat{\theta}_t - \theta_0) \approx N(0, \frac{1}{t} \begin{bmatrix} 3\theta_1 + \theta_2 & -2\theta_1 - \theta_2 \\ -2\theta_1 - \theta_2 & 2\theta_1 + 2\theta_2 \end{bmatrix})$$

EXAMPLE 2: $\phi_t = [1, 1+(-1)^{N_t}]$, $\theta_0 = (\theta_1, \theta_2)$. Take $\mu(t) = t$. Then a simple calculation yields:

$$as - \lim_{t\to\infty} \mu(t)^{-1} <M>_t = \frac{\theta_1^2 + 2\theta_1\theta_2}{\theta_1 + \theta_2} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

which is positive definite. As in example 1 $\phi_s^T \phi_s$ is bounded, so again assumption 4.1.2.1 (ii) trivially holds. Combined with an expression for R_t we can calculate that

$$(\hat{\theta}_t - \theta_0) \approx N(0, \frac{1}{t} \frac{\theta_1 + \theta_2}{\theta_1^2 + 2\theta_1\theta_2} \begin{bmatrix} 2\theta_1^2 & -\theta_1^2 \\ -\theta_1^2 & (\theta_1 + \theta_2)^2 + \theta_2^2 \end{bmatrix}).$$

EXAMPLE 3: $\phi_t = [X_{t-}, 1 - X_{t-}]$. Again take $\mu(t) = t$. Then

$$as - \lim_{t\to\infty} \mu(t)^{-1} <M>_t = \frac{1}{\mu_1 + \mu_0} \begin{bmatrix} \theta_1\mu_1 & 0 \\ 0 & \theta_0\mu_0 \end{bmatrix}.$$

Since $\phi_t^T\phi_t=1$, again assumption 4.2.2.1 (ii) is trivially satisfied. Asymptotically we have

$$(\hat{\theta}_t-\theta_0)\approx N(0,\frac{\mu_1+\mu_0}{t}\begin{bmatrix}\theta_1/\mu_1 & 0\\ 0 & \theta_0/\mu_0\end{bmatrix}).$$

REMARK: Observe that implicitly corollary 4.1.2.2 provides us with an estimate if the asymptotic speed of convergence of $\hat{\theta}_t-\theta_0$. If we consider for instance $V_t(\hat{\theta}_t-\theta_0)$ where V_t tends to infinity, but $V_tR_t<M>_t^{\frac{1}{2}}$ tends to zero, then certainly $V_t(\hat{\theta}_t-\theta_0)\to 0$ in probability.

The advantage of the least squares estimator $\hat{\theta}_t$ of (4.5) is that we are able to obtain an explicit expression (see the first paragraphs of section 4.1). One of the drawbacks however is that they are in general not asymptotically efficient. (Compare with section 4.2.3 below). This is one of the reasons why we present in the next section another type of a recursive estimation algorithm.

4.2 RECURSIVE MAXIMUM LIKELIHOOD ESTIMATION

In this section we study another parameter estimation algorithm for the model (4.1). In contrast with the least squares algorithm there seems to be no explicit expression for a criterion that is minimized by the estimators $\{\hat{\theta}_t\}$ to be defined below, so we have to find another way to obtain an algorithm. Before stating the estimation algorithm, we prefer to formulate a preliminary version of it and provide a heuristic derivation. The preliminary algorithm is

$$d\hat{\theta}_t = \frac{Q_t\phi_t}{\phi_t^T\hat{\theta}_{t-}}(dN_t-\phi_t^T\hat{\theta}_t dt),\quad \hat{\theta}_0 \tag{4.8a}$$

$$dQ_t = -\frac{Q_t\phi_t\phi_t^TQ_t}{\phi_t^T\hat{\theta}_t}dt,\quad Q_0 \tag{4.8b}$$

We will give three approaches that justify, at least heuristically, the form of this preliminary algorithm. The first one is based on a "implicit-function theorem" type argument (4.2.1). The second approach is based on an associated filtering problem (4.2.2) while the last one uses an asymptotic expression of the likelihood functional (4.2.3). Assume that ϕ_t is $\mathbb{F}^N$-predictable. Let P_t be the measure on the trajectory space of counting processes defined on $[0,t]$ that is induced by (4.1) and let Q_t be the measure on the same space induced by a standard Poisson process. In order to express the dependence of dP_t/dQ_t on θ we write $L_t(\theta)=dP_t/dQ_t$. Then the following expression holds.

$$L_t(\theta) = \exp[\int_0^t \log\phi_s^T\theta dN_s-\int_0^t(\phi_s^T\theta-1)ds]$$

If ϕ is not $\mathbb{F}^N$-adapted, we can always interpret $L_t(\theta)$ as a partial likelihood.

The maximum likelihood estimator $\hat{\theta}_t$ by definition maximizes $L_t(\theta)$. Equivalently, $\hat{\theta}_t$ minimizes

$$J_t(\theta) = \int_0^t \phi_s^T \theta ds - \int_0^t \log \phi_s^T \theta dN_s$$

If differentiation with respect to θ under the integral sign is allowed we look for zeros of

$$l_t(\theta) = \nabla_\theta J_t(\theta) = \int_0^t \phi_s ds - \int_0^t \frac{\phi_s}{\phi_s^T \theta} dN_s$$

If $J_t(\theta) = J(t, \theta)$ happens to be a smooth function of both θ and t, it follows from the implicit function theorem that $\hat{\theta}_t$ satisfies the equation

$$\frac{d}{dt}\hat{\theta}_t = -[\nabla_\theta l_t(\hat{\theta}_{t-})]^{-1} \frac{\partial}{\partial_t} l_t(\hat{\theta}_t)$$

A similar expression in the present situation where $l_t(\theta)$ is not smooth, but has jumps, is

$$d\hat{\theta}_t = -[\nabla l_t(\hat{\theta}_{t-})]^{-1} \partial_t l_t(\hat{\theta}_{t-}) \tag{4.9}$$

where ∂_t is the forward partial differential operator with respect to t. Since we have

$$\partial_t l_t(\hat{\theta}_{t-}) = \phi_t dt - \frac{\phi_t}{\phi_t^T \hat{\theta}_{t-}} dN_t$$

and

$$\nabla_\theta l_t(\theta) = \int_0^t \frac{\phi_s \phi_s^T}{(\phi_s^T \theta)^2} dN_s$$

equation (4.9) becomes after writing $Q_t = [\nabla_\theta l_t(\hat{\theta}_t)]^{-1}$

$$d\hat{\theta}_t = \frac{Q_{t-} \phi_t}{(\phi_t^T \hat{\theta}_{t-})} (dN_t - \phi_t^T \hat{\theta}_t dt) \tag{4.10}$$

The next problem is to find an evolution equation for Q. One of the objectives is that the algorithm gives us strongly consistent estimators. Therefore we should have for large t, $\hat{\theta}_t \approx \theta_0$. Hence for large t

$$Q_t^{-1} \approx \int_0^t \frac{\phi_s \phi_s^T}{(\phi_s^T \theta_0)^2} dN_s = \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \theta_0} ds + \int_0^t \frac{\phi_s \phi_s^T}{(\phi_s^T \theta^0)^2} dm_s \tag{4.11}$$

The last term of the right hand side of (4.11) is a zero mean martingale. We get a new approximation of Q_t^{-1} by deleting this last term.

$$Q_t^{-1} \approx \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \theta_0} ds$$

Finally we replace θ_0 by $\hat{\theta}_s$ and we arrive at

$$dQ_t = -\frac{Q_t \phi_t \phi_t^T Q_t}{\phi_t^T \hat{\theta}_t} dt \tag{4.12}$$

Observe from (4.12) that Q_t is continuous. Consequently (4.10) is indeed (4.8a).

Another way of justifying (4.8) is the following. Consider the following filtering problem. We have an observation equation

$$dN_t = \phi_t^T \theta dt + dm_t, n_0 = 0$$

Here ϕ is a $\mathcal{F}_t^N$-predictable random process where $\mathcal{F}_t^N = \sigma\{N_s, 0 \leqslant s \leqslant t\}$ and θ is an unobserved random parameter, that is $\sigma(\theta) \not\subset \mathcal{F}_t^N$ for all t. It is known [3] that the optimal (in mean squared error sense) estimator of θ given the observations $\mathcal{F}_t^N$ is $\hat{\theta}_t := E[\theta | \mathcal{F}_t^N]$, and that satisfies the following equation

$$d\hat{\theta}_t = \frac{P_{t-}\phi_t}{\phi_t^T \hat{\theta}_{t-}}(dN_t - \phi_t^T \hat{\theta}_t dt), \ \hat{\theta}_0 = E\theta.$$

Here P_t is the conditional covariance matrix $E[(\theta - \hat{\theta}_t)(\theta - \hat{\theta}_t)^T | \mathcal{F}_t^n]$ and satisfies

$$dP_t = -\frac{P_t \phi_t \phi_t^T P_t}{\phi_t^T \hat{\theta}_t} dt + [E[(\theta - \hat{\theta}_t)(\theta - \hat{\theta}_t)^T (\theta - \hat{\theta}_t)^T \phi_t | \mathcal{F}_t^n] - \frac{P_t \phi_t \phi_t^T P_t}{\phi_t^T \hat{\theta}_t}]_{t=t-} - \frac{1}{\phi_t^T \hat{\theta}_{t-}}(dN_t - \phi_t^T \hat{\theta}_t dt)$$

In this setting the innovations process $N_t - \int_0^t \phi_s^T \hat{\theta}_s ds$ is a martingale with zero mean. We can approximate this equation by setting the martingale term equal to zero. Denoting the approximation of P_t by Q_t we find as a truncated second order filter

$$d\hat{\theta}_t = \frac{Q_t \phi_t}{\phi_t^T \hat{\theta}_{t-}}(dN - \phi_t^T \hat{\theta}_t dt) \tag{4.13a}$$

$$dQ_t = -Q_t \frac{\phi_t \phi_t^T}{\phi_t^T \hat{\theta}_t} Q_t dt \tag{4.13b}$$

It can be argued that the effect of the prior distribution of θ decays with time. Hence we will eventually get estimators $\hat{\theta}_t$ of θ that are hardly depending on the prior distribution. Consequently the $\hat{\theta}_t$'s for large t will not change much if we would take θ as a deterministic parameter. This suggests the use of the same formulas (4.13) for our original estimation problem.

A third way to obtain the recursive scheme (4.8) is to make use of an asymptotic expression of the logarithm of the likelihood functional. See section 3.2 for the relevant results and conditions. To illustrate what our aim is, consider the case where the process ϕ is deterministic. Define

$$\bar{Q}_t = [\int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \theta_0} ds]^{-1}$$

Then we have in a notation similar to that of section 3.2:

$$\log Z_t(u) = u^T \bar{Q}^{-\frac{1}{2}} \int_0^t \frac{\phi_s}{\phi_s^T \theta_0} dm_s - \frac{1}{2} u^T u + p_t$$

where p is a stochastic process that converges to zero in probability for $t \to \infty$ and

$$\bar{Q}^{-1/2} \int_0^t \frac{\phi_s}{\phi_s^T \theta_0} dm_s$$

converges in law to a gaussian $(0,I)$ random variable. If we set $p_t = 0$, then the value of u that maximizes $\log Z_t(u)$ is

$$\hat{u}_t = \bar{Q}_t^{-\frac{1}{2}} \int_0^t \frac{\phi_s}{\phi_s^T \theta_0} dm_s$$

Hence an approximate maximum likelihood estimator of θ_0 is

$$\bar{\theta}_t = \theta_0 + \bar{Q}_t \int_0^t \frac{\phi_s}{\phi_s^T \theta_0} dN_s$$

Or

$$\bar{\theta}_t = \bar{Q}_t \int_0^t \frac{\phi_s}{\phi_s^T \theta_0} dN_s$$

Observe that $\bar{Q}_t^{-\frac{1}{2}} (\bar{\theta}_t - \theta_0)$ converges in law to a gaussian $(0,I)$ random variable. Of course $\bar{\theta}_t$ is useless as an estimator of θ_0, since it depends on θ_0. We just use it at an intermediate step in obtaining our algorithm (4.8). A simple calculation shows that $\bar{\theta}_t$ and $\bar{Q}_t$ satisfy

$$d\bar{\theta}_t = \frac{\bar{Q}_t \phi_t}{\phi_t^T \theta_0} (dN_t - \phi_t^T \bar{\theta}_t dt) \tag{4.14a}$$

$$d\bar{Q}_t = -\frac{\bar{Q}_t \phi_t \phi_t^T \bar{Q}_t}{\phi_t^T \theta_0} dt \tag{4.14b}$$

As before since one is looking for $\hat{\theta}_t$'s that are close to θ_0 (and thus $\bar{\theta}_t$) we replace θ_0 and $\bar{\theta}_t$ in (4.14) by $\hat{\theta}_t$, thus arriving again at (4.8). Other justifications of (4.14) can be given by using one step improvement techniques or suitable weighted least squares and some additional approximations.

Having finished the explanation of the preliminary version of our algorithm, we will now present it in its final form. The change that has been made prevents estimators to escape to infinity. The reasons for the change will be apparent from the proof of theorem 4.2.1.1. We give a little discussion that tells us that this change does not affect the eventual performance of the algorithm. Suppose that θ_t given by (4.8), converges almost surely to θ_0. Then eventually θ_t will be in any neighbourhood of θ_0. Hence if $\epsilon \in \mathbb{R}_+^d$ is such that

all its components are smaller than the corresponding components of θ_0 we have $\phi_t^T\hat{\theta}_t > \phi_t^T\epsilon$ eventually. This is exactly the property that we need in the analysis. However (4.8) do not guarantee us, that this inequality holds. Obviously the modification below has the desired property. Define the indicator process I_t as follows

$$I_t = I_{\{\phi_t^T x_t > \phi_t^T \epsilon\}}$$

where $\epsilon \in \mathbb{R}_+^d$ is such that $0 < \epsilon_i < \theta_{0i}, i = 1,...,d$. We are now in the position to state our

APPROXIMATE MAXIMUM LIKELIHOOD (AML) ALGORITHM

$$dx_t = \frac{Q_t\phi_t}{\phi_t^T\hat{\theta}_{t-}}(dN_t - \phi_t^T x_t dt), \quad x_0 \tag{4.15a}$$

$$dQ_t = -\frac{Q_t\phi_t\phi_t^T Q_t}{\phi_t^T\hat{\theta}_t}dt, \quad Q_0 \tag{4.15b}$$

$$\hat{\theta}_t = x_t I_t + \epsilon(1 - I_t) \tag{4.15c}$$

Here x_0 is taken such that $\hat{\theta}_0 = x_0$, and Q_0 is a symmetric positive definite matrix

Apparently one should be able to establish lower bounds for the components θ_0 in order to compute $\hat{\theta}_t$ according to (4.15). In practical situations there are often physical considerations that enable us to do so. As for the least squares algorithm we can also prove invariance of (4.15) under non singular linear transformations. Contrary to (4.5) we even have invariance of (4.15) under time transformations. Let $\tau = f(t)$ be a (possibly random) time transformation with inverse $t = g(\tau)$. Assume that g has a derivative g' almost everywhere and $g' \geqslant 0$. Write $\bar{y}_t = y_{g(\tau)}$ for the time transformed process y. Then we have

$$d\bar{N}_t = \bar{\phi}_t^T\theta_0 g'(\tau)d\tau + d\bar{m}_t \tag{4.16}$$

The algorithm corresponding to (4.16)

$$d\bar{x}_t = \frac{\bar{Q}_\tau\bar{\phi}_{\tau-}}{\bar{\phi}_{\tau-}\bar{\theta}_{\tau-}}(d\bar{N}_\tau - \bar{\phi}_\tau^T\bar{x}_\tau\, g'(\tau)d\tau)$$

$$d\bar{Q}_\tau = \frac{\bar{Q}_\tau\bar{\phi}_\tau\bar{\phi}_\tau^T\bar{Q}_\tau}{\bar{\phi}_\tau^T\bar{\theta}_\tau}g'(\tau)d\tau$$

$$\bar{\theta}_\tau = \bar{x}_\tau\bar{I}_\tau + \epsilon(1 - \bar{I}_\tau)$$

which is indeed the same as the time transformed version of (4.15).

4.2.1 Convergence of the AML algorithm

The procedure that we follow is similar to the one of section 4.1.1. Again lemma 4.1.1.1 plays a key role.

THEOREM 4.2.1.1: *Let* $\theta_0 \in \mathbb{R}_+^d$ *and let* $\epsilon \in \mathbb{R}_+^d$ *be such that* $\theta_0 - \epsilon \in \mathbb{R}_+^d$. *Let* $\Phi_t = \int_0^t \phi_s ds$ *and assume*

i) $\Phi_t^T \theta_0 \to \infty$ *a.s.* $(t \to \infty)$

ii) $\liminf_{t\to\infty} \frac{1}{\Phi^T \theta_0} \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \theta_0} ds = C > 0.$

Then

i) $a.s. \lim_{t\to\infty} \hat{\theta}_t = \theta_0$

ii) $a.s. \lim_{t\to\infty} \frac{1}{\theta_0^T \Phi_t} \int_0^t \frac{(\phi_s^T(\hat{\theta}_s - \theta_0))^2}{\phi_s^T \theta_0} ds = 0$

Before proving the theorem we notice that conditions 4.2.1.1 i, ii are equivalent with $(\mathbf{1} = (1,...,1)^T)$

i') $\Phi_t^T \mathbf{1} \to \infty a.s.$

ii') $\liminf \frac{1}{\Phi_t^T \mathbf{1}} \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \mathbf{1}} ds > 0$

The equivalence of i) and i') can easily be seen by noting that $\underline{\theta} \Phi_t^T \mathbf{1} \leq \theta_0^T \Phi_t \leq \bar{\theta} \Phi_t^T \mathbf{1}$, where $\underline{\theta} = \min\{\theta_{0i}, i = 1,...,d\}$, $\bar{\theta} = \max\{\theta_{0i}, i = 1,...,\}$. The equivalence of ii) and ii') follows similary.

PROOF OF THEOREM 4.2.1.1: i) Let $\tilde{x}_t = x_t - \theta_0$, Then

$$d\tilde{x}_t = \frac{Q_t \phi_t}{\phi_t^T \hat{\theta}_{t-}} (dm_t - \phi_t^T \tilde{x}_t dt),$$

Applying the stochastic calculus rule to $\tilde{x}_t^T Q_t^{-1} \tilde{x}_t$ we obtain

$$d(\tilde{x}_t^T Q_t^{-1} \tilde{x}_t) = \frac{2\phi_t^T \tilde{x}_{t-}}{\phi_t^T \hat{\theta}_{t-}} dm_t - \frac{(\tilde{x}_t^T \phi_t)^2}{\phi_t^T \hat{\theta}_t} dt + \frac{\phi_t^T Q_t \phi_t}{(\phi_t^T \hat{\theta}_{t-})^2} dN_t \qquad (4.17)$$

$$= dm_{1t} - \frac{(\tilde{x}_t^T \phi_t)^2}{(\phi_t^T \hat{\theta}_t)^2} dt + \frac{\phi_t^T Q_t \phi_t}{(\phi_t^T \hat{\theta}_t)^2} \phi_t^T \theta_0 dt$$

where we have summarized the martingale term of (4.17) as dm_{1t}. Define

$$r_t = tr(Q_0^{-1}) + \int_0^t \frac{\phi_s^T \phi_s}{\phi_s^T \epsilon} ds$$

Then

$$r_t \geq tr(Q_0^{-1}) + \int_0^t \frac{\phi_s^T \phi_s}{\phi_s^T \hat{\theta}_s} ds = tr(Q_t^{-1}).$$

Define the Lyapunov function

$$u_t = r_t^{-1}[\tilde{x}_t^T Q_t^{-1} \tilde{x}_t + \int_0^t \frac{(\phi_s^T \tilde{x}_s)^2}{\phi_s^T \hat{\theta}_s} ds],$$

then

$$du_t = -r_t^{-1} \frac{\phi_t^T \phi_t}{\phi_t^T \epsilon} u_t dt + r_t^{-1} dm_{1t} + r_t^{-1} \frac{\phi_t^T Q_t \phi_t}{(\phi_t^T \hat{\theta}_t)^2} \phi_t^T \theta_0 dt \tag{4.18}$$

We are able to apply lemma 4.1.1.1 as soon we have verified assumption ii) which leads us to the calculation of

$$\begin{aligned} 0 \leqslant \int_0^\infty r_t^{-1} \frac{\phi_t^T Q_t \phi_t}{(\phi_t^T \hat{\theta}_t)^2} \phi_t^T \theta_0 dt &\leqslant tr \int_0^\infty r_t^{-1} (Q_t^{-1}) \frac{Q_t \phi_t \phi_t^T Q_t}{\phi_t^T \hat{\theta}_t} \frac{\phi_t^T \theta_0}{\phi_t^T \hat{\theta}_t} dt \\ &\leqslant tr \int_0^\infty \frac{Q_t \phi_t \phi_t^T Q_t}{\phi_t^T \hat{\theta}_t} \frac{\phi_t^T \theta_0}{\phi_t^T \hat{\theta}_t} dt \leqslant tr \int_0^\infty (-dQ_t) \frac{\bar{\theta} \phi_t^T \mathbb{1}}{\underline{\epsilon} \phi_t^T \mathbb{1}} dt \\ &= \frac{\bar{\theta}}{\underline{\epsilon}} tr \int_0^\infty (-dQ_t) \leqslant \frac{\bar{\theta}}{\underline{\epsilon}} tr(Q_0) < \infty. \end{aligned}$$

Having verified assumption (ii) we conclude that as-lim u_t exists and is finite a.s. We also get from the same lemma and eq (4.18)

$$\text{a.s.} \lim_{t \to \infty} \int_0^\infty r_t^{-1} \frac{\phi_t^T \phi_t}{\phi_t^T \epsilon} u_t dt < \infty. \tag{4.19}$$

Now

$$r_t \geqslant \frac{1}{\bar{\epsilon}} \int_0^t \frac{\phi_s^T \phi_s}{\phi_s^T \mathbb{1}} ds + tr(Q_0^{-1}) \geqslant \frac{1}{d\bar{\epsilon}} \int_0^t \phi_s^T \mathbb{1} ds + tr(Q_0^{-1}).$$

where we used in the last inequality that $\phi_t^T \phi_t \geqslant \frac{1}{d}(\phi_t^T \mathbb{1})^2$. Hence from assumption 4.3.1.1 (i) $r_t \to \infty$ a.s. Suppose now that on a set $\Omega_1 \subset \Omega$ of positive probability we have $\lim u_t \geqslant \delta$ for some $\delta > 0$. Then there is τ such that $t \geqslant \tau$ implies $u_t \geqslant \frac{1}{2}\delta$. But then

$$\int_\tau^\infty r_t^{-1} \frac{\phi_t^T \phi_t}{\phi_t^T \epsilon} u_t dt \geqslant \frac{1}{2}\delta \int_\tau^\infty d \log r_t = \infty$$

which contradicts (4.19). Hence as-$\lim_{t \to \infty} u_t = 0$. Since u is the sum of two positive processes we have in particular

$$\text{as} - \lim_{t \to \infty} \tilde{x}_t^T \frac{Q_t^{-1}}{r_t} \tilde{x}_t = 0. \tag{4.20}$$

Define now $\bar{\theta}_t = \sup\{\hat{\theta}_{is}, s \in [0,t],\ i = 1,\ldots,d\}$ and write $\lambda_{\min}(A) = \min \sigma(A)$ for the minimal eigenvalue of a matrix A. Then

$$Q_t^{-1} - Q_0^{-1} = \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \hat{\theta}_s} ds \geqslant \frac{1}{\bar{\theta}_t} \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \mathbf{1}} ds.$$

Hence

$$0 \leqslant \frac{\tilde{x}_t^T \tilde{x}_t}{\bar{\theta}_t} \lambda_{\min} \left[\frac{1}{\Phi_t^T \mathbf{1} + tr(Q_0^{-1})} \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \mathbf{1}} ds \right] \leqslant$$

$$\leqslant \tilde{x}_t^T \tilde{x}_t \lambda_{\min} \left[\frac{1}{\Phi_t^T \mathbf{1} + tr(Q_0^{-1})} \left[Q_t^{-1} - Q_0^{-1} \right] \right]$$

$$\leqslant \frac{1}{\Phi_t^T \mathbf{1} + tr(Q_0^{-1})} \tilde{x}_t^T (Q_t^{-1} - Q_0^{-1}) \tilde{x}_t \leqslant \frac{1}{\underline{\epsilon}} r_t^{-1} \tilde{x}_t^T (Q_t^{-1} - Q_0^{-1}) \tilde{x}_t$$

which tends to zero by (4.20). Consequently form assumption 4.2.1.1 (ii)

$$\frac{\tilde{x}_t^T \tilde{x}_t}{\bar{\theta}_t} \to 0 \text{ a.s.} \tag{4.21}$$

Now it is easy to prove that $\bar{\theta}$ is bounded. For suppose not, then there is $\hat{\theta}_{it}$ such that limsup $\hat{\theta}_{it} = \infty$. But then also limsup $\tilde{x}_{it} = \infty$ and we get immediately from (4.21) that this cannot happen. Hence $\theta_t \leqslant K$ for some $K > 0$. But then from

$$\tilde{x}_t^T \tilde{x}_t \leqslant K \frac{\tilde{x}_t^T \tilde{x}_t}{\bar{\theta}_t}.$$

we see that $\tilde{x}_t \to 0$ and so eventually

$$\phi_t^T x_t = \phi_t^T \tilde{x}_t^T + \phi_t^T \theta_0 > \phi_t^T \epsilon.$$

Then $I_t \to 1$ and consequently

$$\hat{\theta}_t = (\tilde{x}_t + \theta_0) I_t + \epsilon(1 - I_t) \to \theta_0$$

ii) $\tilde{\theta}_t = \tilde{x}_t I_t + (1 - I_t)(\epsilon - \theta_0)$. Let τ be such that $t \geqslant \tau$ implies $I_t = 1$. Then for $t \geqslant \tau$ $\tilde{\theta}_t = \tilde{x}_t$. Hence

$$\int_0^t \frac{(\phi_s^T \tilde{\theta}_s)^2}{\phi_s^T \hat{\theta}_s} ds = \int_0^\tau \frac{(\phi_s^T \tilde{\theta}_s)^2}{\phi_s^T \hat{\theta}_s} ds + \int_\tau^t \frac{(\phi_s^T \tilde{x}_s)^2}{\phi_s^T \hat{\theta}_s} ds$$

From the fact that $u_t \to 0$ we have

$$\frac{1}{r_t} \int_0^\tau \frac{(\phi_s^T \tilde{\theta}_s)^2}{\phi_s^T \hat{\theta}_s} ds \to 0.$$

But then it is easy to deduce from the fact that $\hat{\theta}_s \to \theta_0$ a.s. that we also have

$$\frac{1}{r_t} \int_\tau \frac{(\phi_s^T \tilde{\theta}_s)^2}{\phi_s^T \hat{\theta}_0} ds \to \text{ a.s.}$$

and

$$\frac{1}{\theta_0^T \Phi_s} \int_0^t \frac{(\phi_s^T \tilde{\theta}_s)^2}{\phi_s^T \theta_0} ds \to 0 \text{ a.s.} \qquad \text{Q.E.D.}$$

Before giving a few examples to which the theorem can be applied let us remark that a necessary condition for assumption 4.2.1.1 is

$$\liminf_{t\to\infty} \frac{1}{\Phi^T \mathbf{1}} \int_0^t \phi_{is} ds > 0.$$

Clearly this condition is not sufficient. $\phi_t = \mathbf{1}$ is a counterexample.

REMARK: It is possible to relax condition 4.2.1.1 (ii) in such a way that we still have the conclusion of theorem 4.2.2.1. However we have a small price to pay for this, which is a slight modification of the algorithm (4.15). See section 4.3.

We give some examples for which the assumption in theorem 4.2.1.1 hold.

EXAMPLE 1: Let $\phi:[0,\infty)\to\mathbb{R}_+^2$, $\phi_t=[1,1+\sin t]^T$. The following result will be used. For $a>b\geqslant 0$.

$$\int_0^{2\pi} \frac{1}{a+b\sin x} dx = \frac{2\pi}{\sqrt{a^2-b^2}},$$

hence

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \frac{1}{a+b\sin x} dx = \frac{1}{\sqrt{a^2-b^2}}.$$

Then

$$\lim_{t\to\infty} \frac{1}{\Phi_t^T \mathbf{1}} \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \mathbf{1}} ds = \lim_{t\to\infty} \frac{1}{2t-\cos t+1} \int_0^t \begin{bmatrix} 1 & 1+\sin x \\ 1+\sin x & 1+2\sin x+\sin^2 x \end{bmatrix} \frac{dx}{2+\sin x}$$

$$= \frac{1}{2} \begin{bmatrix} \frac{1}{3}\sqrt{3} & 1-\frac{1}{3}\sqrt{3} \\ 1-\frac{1}{3}\sqrt{3} & \frac{1}{3}\sqrt{3} \end{bmatrix}, \text{ which is positive definite.}$$

EXAMPLE 2: Let $\phi:[0,\infty)\times\Omega\to\mathbb{R}_+^2$, $\phi_t=[1,1+(-1)^{N_t}]^T, \theta=[\theta_1\ \theta_2]^T$ Introduce

$$X_t = \frac{1}{t} \int_0^t (-1)^{N_s} ds.$$

Then

$$\frac{1}{\Phi_t^T \mathbf{1}} \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \mathbf{1}} ds = \frac{1}{t(2+X_t)} \int_0^t \begin{bmatrix} 1 & 1+(-1)^{N_s} \\ 1+(-1)^{N_s} & 2+2(-1)^{N_s} \end{bmatrix} \frac{ds}{2+(-1)^{N_s}}$$

$$= \frac{1}{3t(2+X_t)} \int_0^t \begin{bmatrix} 2-(-1)^{N_s} & 1+(-1)^{N_s} \\ 1+(-1)^{N_s} & 2+2(-1)^{N_s} \end{bmatrix} ds$$

$$= \frac{1}{3(2+X_t)} \begin{bmatrix} 2-X_t & 1+X_t \\ 1+X_t & 2+2X_t \end{bmatrix}.$$

In section 4.1.1. we have found that

$$as - \lim_{t\to\infty} X_t = -\frac{\theta_2}{\theta_1+\theta_2}.$$

So

$$\lim_{t\to\infty} \frac{1}{\Phi_t^T \mathbf{1}} \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \mathbf{1}} ds = \frac{1}{3(2\theta_1+\theta_2)} \begin{bmatrix} 2\theta_1+3\theta_2 & \theta_1 \\ \theta_1 & 2\theta_1 \end{bmatrix} > 0.$$

EXAMPLE 3: Let X be a Markov process that takes its values in $\{0,1\}$. Assume that the holding times in 0 and 1 are exponentially distributed with means μ_0 and μ_1 respectively. Assume that N_t has the intensity $\theta_1 X_{t-} + \theta_0(1-X_{t-})$, which corresponds to $\phi_s = [X_{t-}, 1-X_{t-}]^T$ and $\theta = [\theta_1 \ \theta_0]^T$. Then

$$\lim_{Pt\to\infty} \frac{1}{\Phi_t^T \mathbf{1}} \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \mathbf{1}} ds = \lim_{t\to\infty} \frac{1}{t} \int_0^t \begin{bmatrix} X_s & 0 \\ 0 & 1-X_s \end{bmatrix} ds = \frac{1}{\mu_1+\mu_0} \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_0 \end{bmatrix}.$$

4.2.2 Asymptotic distribution of the AML algorithm

The purpose of this section is to show that the $\hat{\theta}_t$'s generated by (4.15) have a limit distribution which is approximately normal. After some definitions we state a useful lemma. Define the following matrix valued stochastic processes

$$\bar{Q}_t^{-1} = Q_0^{-1} + \int_0^t \frac{\phi_s^T \theta_0}{\phi_s^T \theta_0} ds$$

$$V_t = \int_0^t \frac{\phi_s \phi_s^T}{(\phi_s^T \hat{\theta}_s)^2} \phi_s^T \theta_0 ds$$

LEMMA 4.2.2.1: *Let $\hat{\theta}_t, Q_t$ as defined by (4.15) and let the assumptions of theorem 4.2.1.1 be in force. Then*

i) $$as - \lim_{t\to\infty} \bar{Q}_t^{1/2} Q_t^{-1} \bar{Q}_t^{1/2} = I \tag{4.22}$$

ii) $$as - \lim_{t\to\infty} \bar{Q}_t^{1/2} V_t \bar{Q}_t^{1/2} = I. \tag{4.23}$$

PROOF:

i) Let $\delta > 0$ and fix ω, taken from the set with probability one where $\hat{\theta}_t(\omega) \to \theta_0$. Then there is $\tau = \tau(\omega)$ such that $\forall t \geq \tau$ we have $|\hat{\theta}_{it} - \theta_{0i}| \leq \delta$ for all components i. Consequently $(1-\delta)\phi_t^T\theta_0 \leq \phi_t^T\hat{\theta}_t \leq (1+\delta)\phi_t^T\theta_0$ for $t \geq \tau$. In the ordering of positive definite matrices we then have

$$\frac{1}{1+\delta}\int_\tau^t \frac{\phi_s\phi_s^T}{\phi_s^T\theta_0} \leqslant Q_t^{-1} - Q_\tau^{-1} \leqslant \frac{1}{1-\delta}\int_\tau^t \frac{\phi_s\phi_s^T}{\phi_s^T\theta_0} ds$$

or

$$\frac{1}{1+\delta}(\bar{Q}_t^{-1} - \bar{Q}_\tau^{-1}) \leqslant Q_t^{-1} - Q_\tau^{-1} \leqslant \frac{1}{1-\delta}(\bar{Q}_t^{-1} - \bar{Q}_\tau^{-1})$$

which yields

$$\frac{1}{1+\delta}(I - \bar{Q}_t^{\frac{1}{2}}\bar{Q}_\tau^{-1}\bar{Q}_t^{\frac{1}{2}}) \leqslant \bar{Q}_t^{\frac{1}{2}}(Q_t^{-1} - Q_\tau^{-1})\bar{Q}_t^{\frac{1}{2}} \leqslant \frac{1}{1-\delta}(I - \bar{Q}_t^{\frac{1}{2}}\bar{Q}_\tau^{-1}\bar{Q}_t^{\frac{1}{2}}).$$

Now take limits for $t\to\infty$ and use that $\bar{Q}_t\to 0$ to get

$$\frac{1}{1+\delta} I \leqslant \liminf_{t\to\infty} \bar{Q}_t^{\frac{1}{2}} Q_t^{-1} \bar{Q}_t^{\frac{1}{2}} \leqslant \limsup_{t\to\infty} \bar{Q}_t^{\frac{1}{2}} Q_t^{-1} \bar{Q}_t^{\frac{1}{2}} \leqslant \frac{1}{1-\delta} I \tag{4.24}$$

Since (4.24) holds for all $\delta>0$ the proof of (4.22) is complete. □

ii) The proof of (4.23) is analogous.

The following vector valued martingale is important. Define

$$M_t = \int_0^t \frac{\phi_s}{\phi_s^T\hat{\theta}_{s-}} dm_s$$

Notice that we have $<M>_t = V_t$.

THEOREM 4.2.2.2: *Assume that there exists a function* $\mu:[0,\infty)\to[0,\infty)$ *such that*

$$P - \lim_{t\to\infty} \mu(t)^{-1}\theta_0^T\Phi_t = 1 \tag{4.25}$$

Then

$$\bar{Q}_t^{\frac{1}{2}} M_t \xrightarrow{\mathcal{L}} N(0,I).$$

PROOF: Let C be as in assumption 4.2.1.2 (ii)

$$C = \text{as}-\lim_{t\to\infty} \frac{1}{\theta_0^T\Phi_t}\int_0^t \frac{\phi_s\phi_s^T}{\phi_s\theta_0} ds = \text{as}-\lim_{t\to\infty} \frac{1}{\theta_0^T\Phi_t}\bar{Q}_t^{-1}$$

Then we also have

$$P - \lim_{t\to\infty} \frac{1}{\mu(t)}\int_0^t \frac{\phi_s\phi_s^T}{\phi_s^T\theta_0} ds = C. \tag{4.26}$$

Define

$$M_t^\lambda = \lambda^T C^{-\frac{1}{2}} M_t = \int_0^t \frac{\lambda^T C^{-\frac{1}{2}}\phi_s}{\phi_s^T\hat{\theta}_{s-}} dm_s,$$

then

$$(\lambda^T\lambda\mu(t))^{-1}<M>_t=$$
$$=(\lambda^T\lambda)^{-1}\lambda^T C^{-\frac{1}{2}}\bar{Q}_t^{-\frac{1}{2}}\mu(t)^{-\frac{1}{2}}\bar{Q}_t^{\frac{1}{2}}V_t\bar{Q}_t^{\frac{1}{2}}\mu(t)^{-\frac{1}{2}}\bar{Q}_t^{-\frac{1}{2}}C^{-\frac{1}{2}}\lambda \xrightarrow{P} 1$$

by (4.25), (4.26). Hence assumption 3.1.6.3i is satisfied. As in the proof of lemma 4.2.2.1, let $\tau(\omega)$ be such that $t\geqslant\tau(\omega)$ implies

$$|\phi_t^T\hat{\theta}_t-\phi_t^T\theta_0|\leqslant\phi_t^T\theta_0\delta.$$

Consider

$$(\lambda^T\lambda\mu(t))^{-1}\int_0^t\frac{(\lambda^T C^{-\frac{1}{2}}\phi_s)^2}{(\phi_s^T\hat{\theta}_s)^2}I_{\{|\frac{\lambda^T C^{-\frac{1}{2}}\phi_s}{\phi_s^T\hat{\theta}_s}|\geqslant\delta(\lambda^T\lambda\mu(t))^{-\frac{1}{2}}\}}\phi_s^T\theta_0 ds \tag{4.27}$$

Let us split the integral in two pieces, one with integration bounds 0 and $t\wedge\tau$ and the second with bounds $t\wedge\tau$ and t. Then clearly $(\lambda^T\lambda\mu(t))^{-1}$ times the former integral tends to zero almost surely. Hence we continue our investigation of the second integral which is after multiplication by $(\lambda^T\lambda\mu(t))^{-1}$ less than

$$(1-\delta)^{-2}(\lambda^T\lambda\mu(t))^{-1}\int_{t\wedge\tau}^t\frac{(\lambda^T C^{-\frac{1}{2}}\phi_s)^2}{(\phi_s^T\theta_0)^2}I_{\{\frac{\lambda^T C^{-\frac{1}{2}}\phi_s}{\phi_s^T\theta_0(1-\delta)}\geqslant\delta(\lambda^T\lambda\mu(t))^{\frac{1}{2}}\}}\phi_s^T\theta_0 ds\leqslant$$

$$\leqslant(1-\delta)^{-2}(\lambda^T\lambda\mu(t))^{-1}\int_{t\wedge\tau}^t\frac{\lambda^T C^{-1}\lambda\phi_s^T\phi_s)^2}{(\phi_s^T\theta_0)^2}I_{\{\frac{\lambda^T C^{-1}\lambda\phi_s^T\phi_s}{(\phi_s^T\theta_0)^2}\geqslant\delta^2\lambda^T\lambda\mu(t))\}}\phi_s^T\theta_0 ds \tag{4.28}$$

Now let t be such that

$$\mu(t)\geqslant\frac{\lambda^T C^{-1}\lambda}{\delta^2(1-\delta)^2\lambda^T\lambda\underline{\theta}^2},\ \text{where}\ \underline{\theta}=\min\{\theta_{0i},i=1,\ldots,d\}$$

Then

$$\delta^2\lambda^T\lambda\mu(t)\geqslant\frac{\lambda^T C^{-1}\lambda\phi_s^T\phi_s}{\underline{\theta}^2(1-\delta)^2(\phi_s^T\mathbb{1})^2}\geqslant\frac{\lambda^T C^{-1}\lambda\phi_s^T\phi_s}{(1-\delta)^2(\phi_s^T\theta_0)^2}$$

Consequently for large t the indicator appearing in the integral in (4.28) will be zero. As a result (4.27) converges to zero almost surely and a fortiori in probability, which gives us condition 3.1.6.3ii. Conclusion

$$(\lambda^T\lambda\mu(t))^{-\frac{1}{2}}M_t^\lambda \xrightarrow{\mathcal{L}} N(0,I).$$

As in the proof of theorem 4.1.2.1 the Cramer - Wold device gives us

$$(\mu(t)C)^{-\frac{1}{2}}M_t \xrightarrow{\mathcal{L}} N(0,I),$$

if and only if

$$(\lambda^T\lambda\mu(t))^{-\frac{1}{2}}M_t^\lambda \xrightarrow{\mathcal{L}} N(0,I),$$

which has just been proved.

Finally

$$\bar{Q}_t^{\frac{1}{2}} M_t = \mu(t)^{\frac{1}{2}} C^{\frac{1}{2}} \mu(t)^{-\frac{1}{2}} C^{-\frac{1}{2}} M_t.$$

We know from (4.26) that $\mu(t)^{\frac{1}{2}} \bar{Q}^{\frac{1}{2}} C^{\frac{1}{2}} \to I$ in probability, which completes the proof.

COROLLARY 4.2.2.3: *Under the assumptions of theorem 4.2.2.2*
i) $Q_t^{-\frac{1}{2}} \tilde{\theta}_t \xrightarrow{\mathcal{L}} N(0,I)$
ii) $Q_t^{-\frac{1}{2}} \tilde{\theta}_t \xrightarrow{\mathcal{L}} N(0,I)$.

PROOF: i) By writing out the stochastic differential equation for $Q_t^{-1}\tilde{x}_t$ one can show that the following relation holds

$$\tilde{x}_t = Q_t[\int_0^t \frac{\phi_s}{\phi_s^T \hat{\theta}_{s-}} dm_s + Q_0^{-1}(x_0 - \theta_0)]$$

And consequently

$$Q_t^{-\frac{1}{2}} \tilde{\theta}_t = I_t Q_t^{-\frac{1}{2}} M_t + I_t Q_t^{-\frac{1}{2}} Q_0 (x_0 - \theta_0) + Q_t^{-\frac{1}{2}} (1 - I_t)(\epsilon - \theta_0) \quad (4.29)$$

Since $I_t \to 1$ a.s. and $Q_t^{\frac{1}{2}} \to 0$ a.s. as $t \to \infty$ we see from (4.29) that the asymptotic distribution of $Q_t^{-\frac{1}{2}} \theta_t$ will be same as that of $Q_t^{\frac{1}{2}} M_t$. From lemma 4.2.2.1 we know that we can replace Q_t by $\bar{Q}_t$ and the conclusion follows from theorem 4.2.2.2.
ii) This is an immediate consequence of i)

The examples below are examples 1-3 of section 4.2.1 continued.

EXAMPLE 1: $\phi(t) = [1, 1+\sin t]^T$. Take $\mu(t) = t$. Then one finally gets after some tedious calculations: Approximately

$$\tilde{\theta}_t \approx N(0, \frac{\theta_1 + \theta_2 + \sqrt{\theta_1^2 + 2\theta_1\theta_2}}{t} V)$$

with

$$V = \begin{bmatrix} (\frac{1}{\theta_2} - \frac{\theta_1}{\theta_2^2})\sqrt{\theta_1^2 + 2\theta_1\theta_2} + \frac{\theta_1^2}{\theta_2^2} & -\frac{1}{\theta_2}(\sqrt{\theta_1^2 + 2\theta_1\theta_2} - \theta_1) \\ -\frac{1}{\theta_2}(\sqrt{\theta_1^2 + 2\theta_1\theta_2} - \theta_1) & 1 \end{bmatrix}$$

EXAMPLE 2: $\phi_t = [1, 1+(-1)^{N_t-}]$. One gets

$$\text{as} - \lim_{t\to\infty} \frac{1}{t} \bar{Q}_t^{-1} = \frac{1}{\theta_1(\theta_1 + 2\theta_2)(\theta_1 + \theta_2)} \begin{bmatrix} (\theta_1 + \theta_2)^2 + \theta_2^2 & \theta_1^2 \\ \theta_1^2 & 2\theta_1^2 \end{bmatrix}$$

and the asymptotic distribution

$$\tilde{\theta}_t \approx N(0, \frac{\theta_1 + \theta_2}{(\theta_1^2 + 2\theta_1\theta_2)t} \begin{bmatrix} 2\theta_1^2 & -\theta_1^2 \\ -\theta_1^2 & (\theta_1 + \theta_2)^2 + \theta_2^2 \end{bmatrix})$$

EXAMPLE 3: $\phi_t = [X_{t-}, 1 - X_{t-}]$. Here

$$as - \lim_{t\to\infty} \frac{1}{t} \bar{Q}_t^{-1} = \frac{1}{\mu_1 + \mu_0} \begin{bmatrix} \mu_1/\theta_1 & 0 \\ 0 & \mu_0/\theta_0 \end{bmatrix}$$

and asymptotically

$$\tilde{\theta}_t \approx N(0, \frac{\mu_1 + \mu_0}{t} \begin{bmatrix} \theta_1/\mu_1 & 0 \\ 0 & \theta_0/\mu_0 \end{bmatrix}).$$

We see that in this case the asymptotic variance of $\tilde{\theta}_t$ is the same as in example 3 of section 4.1.2.

REMARK: The basic assumption in getting a limit distribution for $\hat{\theta}_t$ or $\tilde{\theta}_t$ which is Gaussian is 4.1.2.1(i) or (4.25) depending on the algorithm. This assumption more or less tells us that the quadratic variation process of the martingale M becomes deterministic as t grows. If this assumption is dropped one can still derive results for the asymptotic distribution of $\hat{\theta}_t$. The idea then is to perform some random time transformation $\tau = f(t)$ after which the transformed version of $<M>$ becomes deterministic. For the transformed algorithm (which looks the same in the AML case (4.16)) we can then infer asymptotic normality as τ tends to infinity [36]. In the AML case a useful transformation is $\tau = \Phi_t^T \theta_0$. This idea has also been carried out in [36] for the off-line maximum likelihood estimation problem. Another way of getting other limit distributions is to look at process ϕ such that Local Mixed Asymptotic Normality holds for the associated family of probability distributions. See [1] for a definition. In this situation one may anticipate asymptotic distributions for the $\hat{\theta}_t$'s which are convolutions of a normal distribution and some other distribution. Information about the asymptotic behaviour of the recursive estimators can also be obtained by comparing their properties with general results in [16]. These approaches will not be discussed here.

4.2.3 Asymptotic efficiency of the AML algorithm

From the given examples it becomes clear that the asymptotic distributions of $\tilde{\theta}_t$, generated by (4.5), or (4.15) will differ in general. Thus they cannot both give us efficient estimators. In general we have the following Cramer-Rao inequality. An unbiased estimator of θ based on the observations in $[0,t]$ has a covariance matrix which is at least

$$C_t(\theta) = \{E_\theta[\frac{\partial}{\partial\theta} \log L_t(\theta)][\frac{\partial}{\partial\theta} \log L_t(\theta)]^T\}^{-1} \tag{4.30}$$

where the likelihood ratio $L_t(\theta)$ is as before. Calculation of (4.30) gives us

$$C_t(\theta) = \left[E_\theta \int_0^t \frac{\phi_s \phi_s^T}{\phi_s^T \theta} ds\right]^{-1}$$

This means that $\hat{\theta}_t$ is an asymptotically efficient estimator if we have

$$C_t(\theta_0)^{1/2}(\hat{\theta}_t - \theta_0) \xrightarrow{\mathcal{L}} N(0,I) \tag{4.31}$$

Clearly by comparing corollary 4.1.2.2 and (4.31) we see that the LS estimator of θ will not be asymptotically efficient in general except for some specific choices of ϕ. Recall that this was one of the considerations for seeking another algorithm than (4.5). On the other hand the AML estimator given by (4.15) is a good candidate for being an asymptotically efficient estimator by corollary 4.2.2.3 (ii). We will indeed have this property as soon as $C_t(\theta_0)\underline{Q}_t^{-1} \to I$ in probability. However assumption (4.25) in theorem 4.2.2.2 does not seem to be sufficient for guaranteeing this. But if we impose as an additional requirement that $\mu(t)^{-1}C_t(\theta_0)^{-1} \to C$ then indeed from (4.26).

$$C_t(\theta_0)\bar{Q}_t^{-1} = C_t(\theta_0)\mu(t)\mu(t)^{-1}\bar{Q}_t^{-1} \xrightarrow{P} C^{-1}C = I.$$

In fact under the assumption (4.25) requiring $\mu(t)^{-1}\underline{C}_t(\theta_0)$ to converge to C is nothing else but demanding the collection $\{C_t(\theta_0)\underline{Q}_t^{-1}\}_{t\geq 0}$ to be uniformly integrable.
Let us summarize the discussion of the proceeding paragraph in

PROPOSITION 4.2.3.1: *Assume that there exists a function* $\mu:[0,\infty)\to[0,\infty)$ *such that*

$$P - \lim_{t\to\infty} \mu(t)^{-1}\Phi_t^T\theta_0 = 1$$

$$\lim_{t\to\infty} \mu(t)^{-1}C_t(\theta_0)^{-1} = C$$

where C is as in assumption 4.2.1.2 (ii). Then the AML estimator $\hat{\theta}_t$ *generated by (4.15) is asymptotically efficient.*

One easily checks that one can take in the proceeding examples $\mu(t)=t$.

4.3 RECURSIVE MAXIMUM LIKELIHOOD ESTIMATION II

In this section we present a slight modification of the AML algorithm (4.15) that enables us to prove almost sure convergence of the $\{\hat{\theta}_t\}$ under weaker conditions. These conditions are close to those in [15,25]. However we do not require all of the conditions of [15] to hold. As such this can be considered an improvement. On the other hand we clearly deal with a more specific model than the general semimartingale regression presented in [15]. Recall that our model is given by (4.1). Throughout this section the following assumption holds.

ASSUMPTION 4.3.1: θ_0 lies in a compact subset of $\mathbb{R}_+^d$. Hence there exists $\epsilon>0$ such that $\epsilon<\theta_{0i}<\frac{1}{\epsilon}$, $\forall i=1,\ldots,d$.

AML II ALGORITHM:

$$dX_t = \frac{Q_t\phi_t}{\phi_t^T\hat{\theta}_{t-}}(dN_t - \phi_t^T X_t dt), X_0 \tag{4.32a}$$

$$dQ_t = -\frac{Q_t\phi_t\phi_t Q_t}{\phi_t^T\hat{\theta}_t}dt, \quad Q_0 > 0 \tag{4.32b}$$

$$\hat{\theta}_t = I_{1t}I_{2t}X_t + \epsilon(1-I_{1t})\mathbf{1} + \epsilon^{-1}(1-I_{2t})\mathbf{1} \tag{4.32c}$$

$$I_{1t} = 1_{\{\phi_t^T X_t \geq \epsilon\phi_t^T\mathbf{1}\}} \tag{4.32d}$$

$$I_{2t} = 1_{\{\phi_t^T X_t \leq \epsilon^{-1}\phi_t\mathbf{1}\}} \tag{4.32e}$$

COMMENT: Introducing the ϵ above is done to establish a.s. convergence of $\{\hat{\theta}_t\}$ to θ_0. If we compare (4.32) to the AML algorithm (5.15) we see that we use the extra indicator process I_2. Clearly we require knowledge of ϵ to compute the $\hat{\theta}_t$. The proof of $\hat{\theta}_t \to \theta_0$ a.s. that we will give parallels to a certain extent the procedure in [15]. First we state an auxiliary result.

Define $\bar{Q}_t^{-1} = \int_0^t \frac{\phi_s\phi_s^T}{\phi_s^T\theta_0}ds$. Denote by $\underline{\lambda}_t$ the minimal eigenvalue of $\bar{Q}_t^{-1}$ and by $\bar{\lambda}_t$ its maximal eigenvalue.

LEMMA 4.3.2.: *There exist constants c and $\bar{c}$ such that*

i) $\bar{c} + \epsilon^2\bar{\lambda}_t \leq \lambda_{\max}(Q_t^{-1}) \leq \epsilon^{-2}\bar{\lambda}_t + \bar{c}$

ii) $\underline{c} + \epsilon^2\underline{\lambda}_t \leq \lambda_{\min}(Q_t^{-1}) \geq \epsilon^{-2}\underline{\lambda}_t + \underline{c}$

PROOF: Define $\underline{c} = \inf_{|x|=1} x^T Q_0^{-1} x$ and $\bar{c} = \sup_{|x|=1} x^T Q_0^{-1} x$.

Since $\epsilon\phi^T\mathbf{1} \leq \phi^T\hat{\theta}_t \leq \epsilon^{-1}\phi^T\mathbf{1}$ we have for all $x \in \mathbb{R}^d$:

$$x^T Q_0^{-1} x + \epsilon^2 x^T \bar{Q}_t^{-1} x \leq x^T Q_t^{-1} x \leq x^T Q_0^{-1} x + \epsilon^{-2} x^T \bar{Q}_t^{-1} x. \tag{4.33}$$

By taking infima in (4.33) in the right order we get (i). The second assertion follows by taking suprema.

THEOREM 4.3.3: *Consider the AML II algorithm (4.32). Assume that $\underline{\lambda}_t \to \infty$ a.s. and that there exists a function $f:[0,\infty) \to [0,\infty)$ such that $\lim_{x\to\infty} \frac{f(x)}{x} = \infty$ and such that*

$$\sup_{t\geq 0} \frac{f(\log\bar{\lambda}_t)}{\underline{\lambda}_t} < \infty \text{ a.s.}$$

Then $\hat{\theta}_t \to \theta_0$ a.s.

REMARKS

1. Observe that $\underline{\lambda}_t \to \infty$ a.s. implies that $N_t \to \infty$ a.s. because
$$\int_0^t \phi_s^T \theta_0 ds = \theta_0^T \bar{Q}_t^{-1} \theta_0 \geqslant \underline{\lambda}_t \theta_0^T \theta_0.$$
2. A possible choice of f that can be found in the literature [15,25] is $f(x) = x^{1+\alpha}$, with $\alpha > 0$.

The crucial step in the proof of theorem 4.3.3 is lemma 4.3.4 below. We will postpone the proof of this lemma and show first, after stating the lemma, how we use it in the proof of theorem 4.3.3.

LEMMA 4.3.4: *Consider (4.32). Let* $\tilde{X}_t = X_t - \theta_0$ *and* $P_t = \tilde{X}_t^T Q_t^{-1} \tilde{X}_t$. *Then* $P_t = O(\log \bar{\lambda}_t)$ *a.s.* $(t \to \infty)$.

PROOF OF THEOREM 4.3.3.:

$$\tilde{X}_t^T \tilde{X}_t = \tilde{X}_t^T Q_t^{-\frac{1}{2}} Q_t Q_t^{-\frac{1}{2}} \tilde{X}_t \leqslant \lambda_{\max}(Q_t) P_t =$$
$$= \frac{P_t}{\lambda_{\min}(Q_t^{-1})} = \frac{f(\log \bar{\lambda}_t)}{\underline{\lambda}_t} \cdot \frac{\underline{\lambda}_t}{\lambda_{\min}(Q_t^{-1})} \cdot \frac{\log \bar{\lambda}_t}{f(\log \bar{\lambda}_t)} \cdot \frac{P_t}{\log \bar{\lambda}_t} \qquad (4.34)$$

Consider the right hand side of (4.34). Its last factor is bounded in view of lemma 4.3.4. The first factor is bounded because of the assumption in the theorem. The second factor is bounded because of lemma 4.3.2 and the third factor tends to zero because of the assumption on f. We conclude that $\tilde{X}_t \to 0$ a.s. But now it is easy to show that $\hat{\theta}_t \to \theta_0$ a.s.

$$\tilde{\theta}_t = \hat{\theta}_t - \theta_0 = \tilde{X}_t I_{1t} I_{2t} + (1 - I_{1t})(\epsilon \mathbf{1} - \theta_0) + (1 - I_{2t})(\epsilon^{-1} \mathbf{1} - \theta_0).$$

Since $\phi_t^T \theta_0 > \phi_t^T \mathbf{1}\epsilon$ there is $\eta > 0$ such that $\phi_t^T \theta_0 \geqslant \phi_t^T \mathbf{1}(\epsilon + \eta)$. Because $\tilde{X}_t \to 0$ we eventually have $|\tilde{X}_{it}| < \eta, \forall i$. But then

$$\phi_t^T X_t = \phi_t^T \tilde{X}_t + \phi_t^T \theta_0 \geqslant -\phi_t^T \mathbf{1}\eta + \phi_t^T \mathbf{1}(\epsilon + \eta) = \phi^T \mathbf{1}\epsilon.$$

Therefore $I_{1t} \to 1$. In a similar way one can prove that $I_{2t} \to 1$, which implies that $\tilde{\theta}_t \to 0$ a.s. □

The proof of lemma 4.3.4 involves a series of other lemmas.

LEMMA 4.3.5: *Let* $P_0 > 0$, $P_0 \in \mathbb{R}^{k \times k}$ *and let* $P_t = P_0 + \int_0^t \xi(s)\xi(s)^T ds$ *for a left continuous function* $\xi : [0, \infty) \to \mathbb{R}^k$. *Then*

(i) $\int_0^t \xi(s)^T P_s^{-1} \xi(s) ds = \log \det(P_t) - \log \det(P_0)$

(ii) $\int_0^t \xi(s)^T P_s^{-1} \xi(s) ds = O(\log \lambda_{\max}(P_t))$.

PROOF: Let $B \in \mathbb{R}^{k\times k}, W \in \mathbb{R}^k$ and $A = B + WW^T$. Assume that A is invertible. Then

$$W^T A^{-1} W = 1 - \frac{\det B}{\det A} \tag{4.35}$$

This can be seen as follows. Observe that $WW^T A^{-1}$ has $k-1$ eigenvalues zero and that the other eigenvalue is $W^T A^{-1} W$. Hence the characteristic polynomial of $WW^T A^{-1}$ is $p(\lambda) = \lambda^{k-1}(\lambda - W^T A^{-1} W)$. Observe now that

$$\det(I - WW^T A^{-1}) = p(1) = 1 - W^T A^{-1} W,$$

whence (4.35).

For $\Delta t \downarrow 0$ we have by definition of P_t

$$P_t = P_{t-\Delta t} + \xi(t)\xi(t)^T \Delta t + o(\Delta t)$$

Notice that $P_t \geq P_0 > 0$. Hence P_t^{-1} exists and $\det(P_t) > 0$. Application of (4.35) and the continuity of $t \mapsto \det(P_t)$ yield

$$\Delta t \xi(t)^T P_t^{-1} \xi(t) = 1 - \frac{\det(P_{t-\Delta t})}{\det(P_t)} + o(\Delta t)$$

Hence

$$\xi(t)^T P_t^{-1} \xi(t) = \frac{\det(P_t) - \det(P_{t-\Delta t})}{\Delta t} \cdot \frac{1}{\det(P_t)} + o(1)$$

or

$$\xi(t)^T P_t^{-1} \xi_t = D^- \log \det(P_t)$$

where D^- denotes left derivative. So (i) has been proved. The second assertion of the lemma is a simple consequence of (i).

LEMMA 4.3.6: *Let m be a quasi left-continuous locally square integrable martingale with $<m> = A$. Let $f : [0,\infty) \to [0,\infty)$ be a differentiable increasing function with*

$$\lim_{x\to\infty} f(x) = \infty \quad \text{and} \quad \int_0^\infty \frac{dx}{(1+f(x))^2} < \infty.$$

Define $g_t = 1 + f(A_t)$. Then both $g_t^{-1} m_t$ and $g_t^{-2}[m,m]_t$ converge almost surely for $t \to \infty$. On $\{A_\infty = \infty\}$ both limits equal zero a.s.

PROOF: This is a simple application of lemma 4.1.1.1. Consider $g_t^{-1} m_t$. Define $X_t = g_t^{-2} m_t^2$. Then application of the stochastic calculus rule yields

$$\begin{aligned} dX_t &= -2g_t^{-3} f'(A_t) m_t^2 dA_t + g_t^{-2}(2m_{t-} dm_t + d[m,m]_t) \\ &= -2g_t^{-1} f'(A_t) X_t dA_t + g_t^{-2} dA_t + g_t^{-2}(2m_{t-} dm_t + d([m,m]_t - A_t)) \end{aligned}$$

Notice that $f'(A_t) \geqslant 0$. Application of lemma 4.1.1.1 immediately yields the desired result since

$$\int_0^\infty g_t^{-2} dA_t = \int_0^{A_\infty} \frac{dx}{(1+f(x))^2} < \infty.$$

On $\{A_\infty = \infty\}$ the second part of lemma 4.1.1.1 yields that $X_t \to 0$ because

$$\int_0^\infty g_t^{-1} f'(A_t) X_t dA_t = \int_0^\infty X_t d\log\, g_t.$$

The statement about $g_t^{-2}[m,m]_t$ can be proved similarly. □

REMARKS

1. The statements of the lemma can be summarized as
$$m_t = o(g_t) + \mathcal{O}(1) \text{ and } [m,m]_t = o(g_t^2) + \mathcal{O}(1).$$
2. Of course we may replace g_t in the lemma by $f(A_t)$ since we consider the behaviour for $t \to \infty$
3. Convenient choices of f in applications are $f(x) = x^{\frac{1}{2}(1+\alpha)}$, with $\alpha > 0$.

PROOF OF LEMMA 4.3.4: For $\tilde{X}$ we have the following equation

$$d\tilde{X}_t = \frac{Q_t \phi_t}{\phi^T \hat{\theta}_{t-}} (dm_t - \phi_t^T \tilde{X}_t dt)$$

Hence

$$dP_t = d(\tilde{X}_t^T Q_t^{-1} \tilde{X}_t) = 2\frac{\phi_t^T \tilde{X}_{t-}}{\phi_t^T \hat{\theta}_{t-}} (dm_t - \phi_t^T \tilde{X}_t dt) + \frac{(\tilde{X}_t^T \phi_t)^2}{\phi_t^T \hat{\theta}_t} dt + \frac{\phi_t^T Q_t \phi_t}{(\phi_t^T \hat{\theta}_{t-})^2} dN_t$$

or

$$P_t - P_0 + \int_0^t \frac{(\tilde{X}_s^T \phi_s)^2}{\phi_s^T \hat{\theta}_s} ds =$$

$$2\int_0^t \phi_s^T \hat{\theta}_{s-} dm_s + \int_0^t \frac{\phi_s^T Q_s \phi_s}{(\phi_s^T \hat{\theta}_s)^2} \phi_s^T \theta_0 ds + \int_0^t \frac{\phi_s^T Q_s \phi_s}{(\phi_s^T \hat{\theta}_{s-})^2} dm_s \tag{4.36}$$

Write (4.36) in obvious notation as

$$P_t - P_0 + L_t = 2M_{1t} + R_t + M_{2t} \tag{4.37}$$

Compute

$$<M_1>_t = \int_0^t \frac{(\phi_s^T \tilde{X}_s)^2}{\phi_s^T \hat{\theta}_s} \cdot \frac{\phi_s^T \theta_0}{\phi_s^T \hat{\theta}_s} ds.$$

Observe that

$$\epsilon^2 \leqslant \frac{\phi_s^T \theta_0}{\phi_s^T \hat{\theta}_s} \leqslant \epsilon^{-2}$$

Hence $\epsilon^2 L_t \leq <M_1>_t \leq \epsilon^{-2} L_t$. Hence $M_{1t} = o(L_t) + O(1)$ in view of lemma 4.3.6 (take $f(x)=x$), and remarks 1 and 2 that follow this lemma. Consider now R_t and notice that

$$\epsilon^2 \int_0^t \frac{\phi_s^T Q_s \phi_s}{\phi_s^T \hat{\theta}_s} ds \leq R_t \leq \epsilon^{-2} \int_0^t \frac{\phi_s^T Q_s \phi_s}{\phi_s^T \hat{\theta}_s} ds \tag{4.38}$$

The integrals in the extreme sides of (4.38) are of the form encountered in lemma 4.3.5. (Take $\xi(s) = \dfrac{\phi_s}{(\phi_s^T \hat{\theta}_s)^{1/2}}$, $Q_t^{-1} = P_t$).

Therefore $R_t = O(\log \lambda_{\max}(Q_t^{-1}))$. The last term to analyze in (4.37) is M_{2t}.

$$<M_2>_t = \int_0^t \frac{(\phi_s^T Q_s \phi_s)^2}{(\phi_s^T \hat{\theta}_s)^4} \phi_s^T \theta_0 ds = \int_0^t -\frac{\phi_s^T dQ_s \phi_s}{(\phi_s^T \hat{\theta}_s)^2} \frac{\phi_s^T \theta_0}{\phi_s^T \hat{\theta}_s} ds$$

$$\leq \int_0^t \frac{\phi_s^T \phi_s}{(\phi_s^T \hat{\theta}_s)^2} \cdot \frac{\phi_s^T \hat{\theta}_0}{\phi_s^T \hat{\theta}_s} d\, tr(-Q_s) \leq \epsilon^{-4} \int_0^t d\, tr(-Q_s)$$

$$\leq \epsilon^{-4} tr(Q_0) < \infty.$$

From lemma 4.3.6 we conclude that $\dfrac{M_2}{<M_2>}$ converges to a finite limit and since $<M_2>_t \leq \epsilon^{-4} tr(Q_0)$, M_2 is a.s. bounded. Collecting the above results we get from (4.37)

$$P_t - P_0 + L_t = o(L_t) + O(1) + O(\log \lambda_{\max}(Q_t^{-1})) + O(1)$$

or

$$P_t - P_0 + L_t(1+o(1)) = O(1) + O(\log \lambda_{\max}(Q_t^{-1}))$$

From lemma 4.3.2 we obtain after dividing by $\log \bar{\lambda}_t$

$$\frac{P_t}{\log \bar{\lambda}_t} + (1+o(1)) \frac{L_t}{\log \bar{\lambda}_t} = O(1)$$

Since both P_t and $(1+o(1))L_t$ are (eventually) nonnegative we get $P_t = O(\log \bar{\lambda}_t)$, as was to be proven. □

We close this section by proving that the limit distribution of the AML II estimators defined by (4.32) is asymptotically normal. Since this proof is essentially the same as the end of 4.2.2.3 and related results we will only give the principal steps.

THEOREM 4.3.7: *Assume that* $\{\hat{\theta}_t\}$ *given by (4.32) is a.s. convergent. Assume that there exist* $P:[0,\infty) \to \mathbb{R}^{d \times d}$ *and* $h:[0,1] \times [0,\infty) \to \mathbb{R}_+$ *such that*

(i) *h is an increasing function of each of its arguments,* $h(t,T) \leq h(1,T) = T$, $\forall t,T \in [0,1] \times [0,\infty)$ *and* $\lim_{T \to \infty} h(t,T) = \infty$, $\forall t \in (0,1]$.

(ii) $R(t) = \lim_{T\to\infty} P(T)^{-\frac{1}{2}}\, P(h(t,T))P(T)^{-\frac{1}{2}}$ *exists and* $R(t)>0$ *for* $t\in(0,1]$

(iii) $P(t)^{+\frac{1}{2}}\bar{Q}_t\tilde{P}(t)^{+\frac{1}{2}}\to I$ *in probability for* $t\to\infty$.
Then $Q_t^{-\frac{1}{2}}\tilde{\theta}_t \xrightarrow{\mathcal{L}} N(0,I)$.

PROOF: We use the same techniques as in section 4.2.2. It is easy to see that the asymptotic distribution of $Q_t^{-\frac{1}{2}}\tilde{\theta}_t$ is the same as that of $Z_t = \bar{Q}_t^{-\frac{1}{2}}\bar{M}_t$, where $\bar{M}_t = \int_0^t \frac{\phi_s}{\phi_s^T\theta_0} dm_s$. Define

$$Z_t^T = \lambda^T P(T)^{-\frac{1}{2}} \int_0^{h(t,T)} \frac{\phi_s}{\phi_s^T\theta_0} dm_s \quad \text{for } \lambda\in\mathbb{R}^d,\ t\in[0,1],\ T\in[0,\infty).$$

Then

$$<Z^T>_t = \lambda^T P(T)^{\frac{1}{2}}\bar{Q}_{h(t,T)}^{-1}P(T)^{\frac{1}{2}}\lambda.$$

Hence $<Z^T>_t \to \lambda^T R(t)\lambda$ in probability as $T\to\infty$, because of the assumptions in the theorem. Now choose T_0 such that $\lambda^T P(T)^{\frac{1}{2}}\lambda \leqslant \frac{\delta^2}{\epsilon^2}$ for $T>T_0$. Then

$$\lambda^T P(T)^{\frac{1}{2}}\phi_s\phi_s^T P(T)^{\frac{1}{2}}\lambda \leqslant \lambda^T P(T)\lambda\phi_s^T\phi_s \leqslant \frac{\delta^2}{\epsilon^2}\phi_s^T\phi_s \leqslant \frac{\delta^2}{\epsilon^2}\phi_s^T\phi_s \leqslant \frac{\delta^2}{\epsilon^2}(\phi_s^T\mathbf{1})^2 \leqslant \delta^2(\phi_s^T\theta_0)^2.$$

Hence for such T

$$\lambda^T P(T)^{\frac{1}{2}} \int_0^{h(t,T)} \frac{\phi_s\phi_s^T}{\phi_s^T\theta_0} 1_{\{\frac{|\lambda^T P(T)^{\frac{1}{2}}\phi_s|}{\phi_s^T\theta_0} > \delta\}} ds P(T)^{\frac{1}{2}}\lambda = 0,\ \forall t\in[0,1].$$

The above implies that the assumptions of proposition 3.1.5.4 are satisfied. Hence $Z^T \xrightarrow{\mathcal{L}} W$, where W is a Gaussian martingale with $<W>_t = R(t)$. In particular $Z_1^T = \bar{Q}_1^T = \bar{Q}_T^{\frac{1}{2}}\bar{M}_t \xrightarrow{\mathcal{L}} N(0,I)$, because $R(1)=I$. □

Chapter 5

Counting Process Systems

In this chapter we treat some problems for counting process systems with a finite state space. Two specific classes of counting process systems will be treated viz. conditionally Poisson systems and self-exciting systems, which are the topics of sections 5.2 and 5.3. The main problem we adress is the characterization of minimality of a system, which means minimality of the state space. The reason why this topic is important lies partly in identification problems for such systems in the situation where the state process cannot directly be observed. It is known for instance in deterministic linear system theory that a state space, which is too large for explaining the behaviour of the output process, contains unobservable components. This implies among other things that if one wants to perform output-based parameter estimation one will not be able to identify the true parameter values that govern the behaviour of the state process in an unobservable part of the state space. For counting process systems to be treated in the next section a similar reasoning holds. If for instance one wants to identify transition rates of the state process (which turns out to be a Markov process) and if two different states yield the same behaviour of the observed counting process, then one is clearly not able to distinguish whether the state process assumes one of these two values, let alone that one is able to draw reliable conclusions about rates that govern a transition from one of these states to the other one. Therefore one can anticipate that minimality is also a prerequisite for consistency of estimators generated by recursive algorithms as in chapter 4, in the situation where the state process is not observed. The lesson of these considerations, as is well known, is that one should always work with minimal representation of a stochastic system.

5.1 Counting process systems

Counting process systems form a subclass of what is known as stochastic systems. Roughly speaking a stochastic system without input consists of two stochastic processes X and Y where X is called the state process and Y the output process. As in deterministic system theory the state process at time t should summarize all the relevant information about the past of the system in order to describe the future output. Contrary to what can be done in deterministic system theory the state process at time t cannot exactly predict the values of Y_s for $s \geqslant t$. It can only describe the probabilistic behaviour of the output process. These notions are made precise in definition 5.1.1 that in abstract terms describes what a stochastic system without inputs is. This definition is

followed by a more detailed treatment of stochastic systems where the output process is a counting process. First we have to introduce some notation. Let a complete probability space $(\Omega,\mathcal{F},P)$ be given together with a filtration $\mathbb{F}$. Let X and Y be $\mathbb{F}$-adapted stochastic processes. Then $\mathcal{F}_t^X=\sigma\{X_s,s\leq t\}$ and $\mathcal{F}_t^Y=\sigma\{Y_s,s\leq t\}$ are the σ-algebras generated by the past of the procesesses X and Y. Similarly $\mathcal{F}_t^{X+}=\sigma\{X_s,s\geq t\}$ contains the information of the future of X after t. We also use the σ-algebra that describes the future increments of the output process Y, $\mathcal{F}_t^{\Delta Y+}=\sigma\{Y_s-Y_t,s\geq t\}$.
If $\mathcal{F}_1,\mathcal{F}_2$ and $\mathcal{G}$ are sigma algebras contained in $\mathcal{F}$, then we say that $\mathcal{F}_1$ and $\mathcal{F}_2$ are conditionally independent given $\mathcal{G}$, if for all integrable $\mathcal{F}_1$-measurable functions X_1 the following relation holds

$$E[X_1|\mathcal{F}_2\vee\mathcal{G}]=E[X_1|\mathcal{G}].$$

We will use the notation $(\mathcal{F}_1,\mathcal{F}_2|\mathcal{G})\in CI$.

DEFINITION 5.1.1: [38]: A *continuous time stochastic system* is a multiple $(\Omega,\mathcal{F},P,T,\mathbb{F},X,Y,\mathcal{X},\mathcal{Y})$ such that
(i) $(\Omega,\mathcal{F},P)$ is a complete probability space
(ii) $T\subset\mathbb{R}$, T an interval.
(iii) $\mathbb{F}=\{\mathcal{F}_t\}_{t\in T}$ a filtration on $(\Omega,\mathcal{F},P)$
(iv) X and Y are $\mathbb{F}$-adapted processes with values in the measurable spaces $\mathcal{X}$ and $\mathcal{Y}$ respectively.
(v) $(\mathcal{F}_t^{X+}\vee\mathcal{F}_t^{\Delta Y+},\mathcal{F}_t|\sigma(X_t))\in CI$ for all $t\geq 0$.

Formally speaking each of the components of the multiple in definition 5.1.1 is part of the definition. However if no confusion can arise we will often write (X,Y) for a stochastic system. The crucial property in the definition of a stochastic system is (v), which says that given a whole past $\mathcal{F}_t$ it is sufficient to use only X_t for the prediction of the future values of X and the future increments of Y. Observe that 5.1.1 (v) implies that X is a Markov process with respect to the filtration $\mathbb{F}$. Finally it is noticed that usually $\mathcal{F}_t=\mathcal{F}_t^X\vee\mathcal{F}_t^Y$ and $T=(-\infty,\infty)$ or $T=[0,\infty)$.
Clearly the above definition is too abstract for practical purposes. In particular cases one has to specify the distribution of the state and output process. One way to do this is to pose stochastic differential equations that X and Y satisfy. In this chapter we will treat stochastic systems where the output process is a counting process and X a finite state process.

DEFINITION 5.1.2: A *counting process system* is a stochastic system where the output Y is a counting process. We write in this case N for the output process instead of Y. The shorthand notation is then (X,N) for a counting process system with state process X.

We will treat in more detail the class of conditionally Poisson systems.

DEFINITION 5.1.3: Let $N:\Omega\times[0,\infty)\to\mathbb{N}$ be a counting process, $\mathbb{F}$-adapted with

Doob-Meyer decomposition w.r.t. $\mathbb{F}$: $dN_t = \lambda_t dt + dm_t$. Let $\mathcal{F}_\infty^\lambda = \sigma\{\lambda_t, t \geq 0\}$. $\mathbb{N}$ is called a *conditionally Poisson process,* or a *doubly stochastic Poisson process,* iff for all $t,h \geq 0$, $u \in \mathbb{R}$

$$E[\exp(iu(N_{t+h} - N_t))|\mathcal{F}_t \vee \mathcal{F}_\infty^\lambda] = \exp((e^{iu} - 1) \int_t^{t+h} \lambda_s ds)$$

So conditioned upon $\mathcal{F}_t \vee \mathcal{F}_\infty^\lambda$ $N_{t+h} - N_t$ has a Poisson distribution with mean $\int_t^{t+h} \lambda_s ds$

PROPOSITION 5.1.4: *N is a conditionally Poisson process iff m as given in 5.1.3 is a martingale w.r.t. $\tilde{\mathbb{F}} = \{\tilde{\mathcal{F}}_t\}_{t \geq 0}$, where $\tilde{\mathcal{F}}_t = \mathcal{F}_t \vee \mathcal{F}_\infty^\lambda$.*

PROOF: If N is conditionally Poisson, then

$$E[m_{t+h} - m_t|\tilde{\mathcal{F}}_t] = E[N_{t+h} - N_t|\tilde{\mathcal{F}}_t] - E[\int_t^{t+h} \lambda_s ds|\tilde{\mathcal{F}}_t]$$

$$= \int_t^{t+h} \lambda_s ds - \int_t^{t+h} \lambda_s ds = 0.$$

Conversely assume that m is a martingale w.r.t. $\tilde{\mathbb{F}}$. Apply the stochastic calculus rule to $\exp(iuN_t)$ to obtain

$$\exp(iuN_{t+h}) = \exp(iuN_t) + (e^{iu} - 1) \int_t^{t+h} \exp(iuN_{s-}) dN_s$$

$$= \exp(iuN_t) + (e^{iu} - 1) \int_t^{t+h} \exp(iuN_{s-})(\lambda_s ds + dm_s)$$

Take conditional expectation w.r.t. $\tilde{\mathcal{F}}_t$ and get

$$E[\exp(iuN_{t+h})|\tilde{\mathcal{F}}_t] = \exp(iuN_t) + (e^{iu} - 1) \int_t^{t+h} E[\exp(iuN_s)|\tilde{\mathcal{F}}_t] \lambda_s ds$$

Define $g(t+h,t) = E[\exp(iu(N_{t+h} - N_t))|\tilde{\mathcal{F}}_t]$. Then we get

$$g(t+h,t) = 1 + (e^{iu} - 1) \int_t^{t+h} g(s,t) \lambda_s ds,$$

from which we find $g(t+h,t) = \exp((e^{iu} - 1) \int_t^{t+h} \lambda_s ds)$. □

Next we present a method for the construction of a counting process system. Let a probability space $(\Omega, \mathcal{F}, P_0)$ be given together with a standard Poisson process N and a Markov process X (with state space $\mathcal{X}$) defined on it such that N and X are independent processes. Notice that such a probability space always exists. We assume that X has cadlag paths. Consider the following filtrations: $\mathbb{F}^N$, $\mathbb{F}^X$, $\mathbb{F} = \{\mathcal{F}_t^N \vee \mathcal{F}_t^X\}_{t \geq 0}$, $\tilde{\mathbb{F}} = \{\mathcal{F}_t^N \vee \mathcal{F}_\infty^X\}_{t \geq 0}$. The following observation is important. Let $\tilde{m}_t = N_t - t$. By definition $\tilde{m}$ is an $\mathbb{F}^N$-martingale.

However because of the independence assumption $\tilde{m}$ is also an $\mathbb{F}$- and $\tilde{\mathbb{F}}$-martingale. Similarly X is also Markov with respect to the filtration $\mathbb{F}$. Let $\lambda:[0,\infty)\times\mathcal{X}\to(0,\infty)$ be a measurable function such that $E_0\int_0^t\lambda(s,X_s)ds<\infty$, $\forall t$. Write $\lambda_t=\lambda(t,X_{t-})$. Then $\{\lambda_t\}$ is clearly both $\mathbb{F}$ and $\tilde{\mathbb{F}}$-predictable. Then M defined by $M_t=\int_0^t(\lambda_s-1)d\tilde{m}_s$ is an $\tilde{\mathbb{F}}$-martingale and let $\Lambda_t=\mathcal{E}(M_t)$. Then

$$\Lambda_t=\exp(\int_0^t\log\lambda_s dN_s-\int_0^t(\lambda_s-1)ds)$$

and Λ is an $\mathbb{F}$- and $\tilde{\mathbb{F}}$-local martingale. We make the following assumption: $E_0\Lambda_t=1$, $\forall t\geq 0$. We can now define a new measure P on $(\Omega,\tilde{\mathcal{F}}_\infty)=(\Omega,\mathcal{F}_\infty)$ as follows. If $A\in\tilde{\mathcal{F}}_t$ then by definition $P(A)=E_0[1_A\Lambda_t]$. The extension to $\tilde{\mathcal{F}}_\infty$ follows by Caratheodory's theorem. Observe that the restriction of P to $\tilde{\mathcal{F}}_t$ is absolutely continuous with respect to the restriction of P to $\tilde{\mathcal{F}}_t$ with Λ_t as Radon-Nikodym derivative and that $\Lambda_t>0$ P_0 a.s. Observe also that the restrictions of P and P_0 to $\mathcal{F}_\infty^X$ coincide.

PROPOSITION 5.1.5: *Under the new measure P*

(i) $m_t=N_t-\int_0^t\lambda_s ds$ *defines a martingale with respect to* $\mathbb{F}$ *and* $\tilde{\mathbb{F}}$.

(ii) *X is a Markov process with respect to* $\mathbb{F}$.

PROOF: The first assertion has already been proved in proposition 3.2.5. So here we prove only (ii). Let f be a bounded measurable function on $\mathcal{X}$ and $h>0$. Then because Λ_t is the Radon-Nikodym derivative $\frac{dP|\tilde{\mathcal{F}}_t}{dP_0|\tilde{\mathcal{F}}_t}$

$$E[f(X_{t+h})|\tilde{\mathcal{F}}_t]=\frac{E_0[f(X_{t+h})\Lambda_t|\tilde{\mathcal{F}}_t]}{E_0[\Lambda_t|\tilde{\mathcal{F}}_t]}=E_0[f(X_{t+h})|\tilde{\mathcal{F}}_t]=E_0[f(X_{t+h})|\sigma(X_t)]$$

In the second equality we have used the fact Λ_t is $\mathcal{F}_t$-measurable and in the third one that X is $\mathbb{F}$-Markov under P_0. □.

THEOREM 5.1.6: *Under the new measure P the pair (X,N) forms a stochastic system.*

PROOF: From part (i) of proposition 5.1.5 and proposition 5.1.4 we obtain that N is conditionally Poisson. Notice that we even have

$$E[\exp(iu(N_{t+h}-N_t))|\tilde{\mathcal{F}}_t]=\exp((e^{iu}-1)\int_t^{t+h}\lambda_s ds).$$

Hence

$$E[\exp(iu(N_{t+h}-N_t))|\tilde{\mathcal{F}}_t\vee\mathcal{F}_t^{X+}]=E[\exp(iu(N_{t+h}-N_t))|\mathcal{F}_t^{X+}]$$

which shows that

$$(\mathcal{G}_t^{\Delta N+}, \mathcal{F}_t | \mathcal{G}_t^{X+}) \in CI, \quad \forall t \geqslant 0.$$

The fact that X is $\mathbb{F}$-Markov yields

$$(\mathcal{G}_t^{X+}, \mathcal{F}_t | \sigma(X_t)) \in CI, \quad \forall t \geqslant 0.$$

Now we can use the following result which is obvious. Let F_1, F_2, G be σ-algebras. Then $(F_1, F_2 | G) \in CI$ and $(F_1, F_3 | G \vee F_2) \in CI$ is equivalent with $(F_1, F_2 \vee F_3 | G) \in CI$.
In our case we take $G = \sigma(X_t)$, $F_2 = \mathcal{G}_t^{X+}$, $F_1 = \mathcal{F}_t$ and $F_3 = \mathcal{G}_t^{\Delta N+}$ and we obtain $(\mathcal{F}_t, \mathcal{G}_t^{X+} \vee \mathcal{G}_t^{\Delta N+} | \sigma(X_t)) \in CI$. □

Thus we have constructed a stochastic system where (as always) X is a Markov process and the output process is a conditional Poisson process. Notice that so far we have used an evolution equation for N whereas for X we only have the Markov property. The next objective is to describe the evolution of X in terms of a stochastic differential equation. Throughout the rest of this chapter the following assumption will be in force.

ASSUMPTION: The state process X takes its values in the finite set $\mathcal{X} = \{x_1, \ldots, x_n\}$, where the x_i are different. Moreover for all i and $t > 0$: $P(X_t = x_i) > 0$.

Define $Y: \Omega \times [0, \infty) \to \{0,1\}^n$ by its components $Y_{it} := 1_{\{X_t = x_i\}}$ Denote by $\Phi(t,s)$ the matrix of transition probabilities of X. That is for $t \geqslant s$, with the notation $z^+ = z^{-1} 1_{\{z \neq 0\}}$ and the understanding $\frac{0}{0} = 0$

$$\Phi_{ij}(t,s) = P(X_t = x_i | X_s = x_j) = (EY_{js})^+ E(Y_{js} Y_{it}).$$

Then we have the following well known facts. Semigroup property: $\Phi(t,s) = \Phi(t,u)\Phi(u,s)$ for $t \geqslant u \geqslant s$. Assume that for all $t \geqslant 0$ the following limit exists

$$A(t) := \lim_{h \downarrow 0} \frac{1}{h} [\Phi(t+h, t) - I]$$

$A(t)$ will be called the generator of X at time t. So $A(t)$ has nonpositive diagonal elements, the other entries are nonnegative and the column sums are zero. Such a matrix will be called a Markov matrix. Then $\frac{\partial}{\partial t} \Phi(t,s) = A(t)\Phi(t,s)$ In particular $\frac{\partial}{\partial t} \Phi(t, 0) = A(t)\Phi(t, 0)$. From this equation we get $\det \Phi(t, 0) = \exp(\int_0^t tr A(s) ds)$. Hence, by definition of $A(t)$, we see that $\Phi(t, 0)$ is invertible for all $t \geqslant 0$.

PROPOSITION 5.1.7: *Define* $Z: \Omega \times [0, \infty) \to \mathbb{R}^n$ *by* $Z_t = \Phi(t, 0)^{-1} Y_t$. *Then* Z *is an* $\mathbb{F}$*-martingale and* Y *satisfies the stochastic differential equation*

$$dY_t = A(t)Y_t dt + \Phi(t,0)dZ_t \tag{5.1}$$

PROOF: Using a representation of a conditional expectation when the conditioning σ-algebra is generated by a finite number of disjoint sets we get

$$\begin{aligned} E[Z_t|\mathcal{F}_s] &= \Phi(t,0)^{-1}E[Y_t|\mathcal{F}_s] = \Phi(t,0)^{-1}E[Y_t|\sigma(X_s)] = \\ &= \Phi(t,0)^{-1}E[Y_t|\sigma(Y_s)] = \Phi(t,0)^{-1}\sum_j E[Y_{js}]^+ E[Y_t Y_{js}]Y_{js} = \\ &= \Phi(t,0)^{-1}\Phi(t,s)Y_s = \Phi(s,0)^{-1}Y_s = Z_s. \end{aligned}$$

The second assertion can easily be proved by applying the stochastic differentiation rule to the product $Y_t = \Phi(t,0)Z_t$. □

Notice that $\int_0^t \Phi(s,0)dZ_s$ appearing in (5.1) is again a $\mathbb{F}$-martingale since $\Phi(\cdot,0)$ is trivially predictable.
Proposition 5.1.7 thus gives us a representation of Markov processes in terms of a linear stochastic differential equation driven by a martingale. The next result gives a converse statement.

PROPOSITION 5.1.8: *Let $X:\Omega\times[0,\infty)\to\mathcal{X}$ be a stochastic process, $\mathbb{F}$-adapted, and let Y be associated with X as before. Assume that Y satisfies*

$$dY_t = A(t)Y_t dt + dm_t^Y \tag{5.2}$$

Here $A:[0,\infty)\to\mathbb{R}^{n\times n}$ is a Lebesgue measurable function (deterministic !) and m^Y an $\mathbb{F}$-adapted martingale. Then X and Y are $\mathbb{F}$-Markov processes.

PROOF: We have to prove that $E[f(X_t)|\mathcal{F}_s] = E[f(X_t)|\sigma(X_s)]$ for all $f:\{x_1,\ldots,x_n\}\to\mathbb{R}$. Since $f(X_t) = \Sigma_j f(c_j)Y_{jt}$ we only have to prove $E[Y_t|\mathcal{F}_s] = E[Y_t|\sigma(X_s)]$. Let: $B(t) = \Phi(t,0)$. Then

$$\dot{B}(t) = A(t)B(t),\ B(0) = I$$

Now we can write the solution Y_t of (5.2) as

$$Y_t = B(t)Y_0 + B(t)\int_0^t B^{-1}(s)dm_s^Y.$$

Notice again that $\int_0^t B^{-1}(s)dm_s^Y$ is an $\mathbb{F}$-martingale and $B(t)$ deterministic. Hence

$$\begin{aligned} E[Y_t|\mathcal{F}_s] &= B(t)Y_0 + B(t)\int_0^s B^{-1}(u)dm_u^Y = \\ &= B(t)Y_0 + B(t)[B^{-1}(s)Y_s - Y_0] = B(t)B^{-1}(s)Y_s \end{aligned}$$

Since we have $\sigma(X_s) = \sigma(Y_s) \subset \mathcal{F}_s$ we get

$$\begin{aligned} E[Y_t|\sigma(X_s)] &= E[E[Y_t|\mathcal{F}_s]|\sigma(Y_s)] = E[B(t)B^{-1}(s)Y_s|\sigma(Y_s)] \\ &= B(t)B^{-1}(s)Y_s = E[Y_t|\mathcal{F}_s]. \end{aligned}$$ □

Concluding we see that the statement X and Y are $\mathbb{F}$-Markov is equivalent with saying that the indicator process Y satisfies equation (5.2).

Next we give a result on Markov solutions of stochastic differential equations (see also [33] for related problems).

PROPOSITION 5.1.9: *Let X be the solution of the stochastic differential equation*

$$dX_t = g(t,X_t)dt + dm_t^X, \quad X_0 \tag{5.3}$$

where m^x is an $\mathbb{F}$-martingale and $g:[0,\infty)\times\mathcal{X}\to\mathbb{R}$. Assume that the jump measure μ of X admits a compensator ν (with respect to $\mathbb{F}$ and P) such that $\nu(dt,dy,\omega)=p(t,X_t(\omega),dy)dt$. Then X is an $\mathbb{F}$-Markov process.

PROOF: We show that for the indicator process Y the representation of proposition 5.1.8 holds. From (5.3) we get from the stochastic calculus rule for all $k\geqslant 0$:

$$\begin{aligned} dX_t^k &= kX_{t-}^{k-1}dX_t + \int_{\mathcal{X}}[(X_{t-}+y)^k - X_{t-}^k - kX_{t-}^{k-1}y]\mu(dt,dy) \\ &= kX_t^{k-1}(g(t,X_t) + \int_{\mathcal{X}}[(X_t+y)^k - X_t^k - kX_t^{k-1}y]p(t,X_t,dy))dt \\ &\quad + d\tilde{m}_t^{(k)} \end{aligned} \tag{5.4}$$

Here $d\tilde{m}_t^{(k)}$ summarizes all the martingale terms in (5.4). In a more compact notation we can write (5.4) as

$$dX_t^k = g^{(k)}(t,X_t)dt + d\tilde{m}_t^{(k)} \tag{5.5}$$

where $g^{(k)}:[0,\infty)\times\mathcal{X}\to\mathbb{R}$. Now we can write X_t^k as $[x_1^k,\ldots,x_n^k]Y_t$ and $g^{(k)}(t,X_t)$ as $G^{(k)}(t)Y_t$ where $G^{(k)}(t)=[g^{(k)}(t,x_1),\ldots,g^{(k)}(t,x_n)]$. Introduce the following notation. V is the $(n\times n)$ matrix with k-th row equal to $[x_1^{k-1},\ldots,x_n^{k-1}]$ $(k=1,\ldots,n)$. $G(t)$ is the $(n\times n)$ matrix with k-th row $G^{(k-1)}(t)$ $(k=1,\ldots,n)$. $\tilde{M}_t$ is the martingale with components $\tilde{m}_t^{(k)}$. If we consider (5.5) as a system of equations for $k=0,\ldots,n-1$ we can summarize it (with $G(t)$ and V as defined above) as

$$VdY_t = G(t)Y_t dt + d\tilde{M}_t \tag{5.6}$$

Observe that V is a Vandermonde matrix, that is nonsingular because all the x_i are different. Let $A(t)=V^{-1}G(t)$ and $M_t^Y=V^{-1}\tilde{M}_t$ then (5.6) becomes

$$dY_t = A(t)Y_t dt + dM_t^Y \tag{5.7}$$

Because M is an $\mathbb{F}$-martingale and $A(t)$ is nonrandom, we obtain from (5.7) by applying proposition 5.1.8 that X is $\mathbb{F}$-Markov, with generator $A(t)$. □

If we collect the above results we get the following

THEOREM 5.1.10: *Let the process X and the counting process N satisfy the following equation*

$$dX_t = g(t,X_t)dt + dm_t^X, \quad X_0$$
$$dN_t = \lambda(t,X_t)dt + dm_t, \quad N_0=0$$

Here λ and g are measurable functions from $[0,\infty)\times\mathcal{X}$ to $\mathbb{R}$ and $\mathbb{R}^+$ respectively and m^X and m are $\mathbf{F}$-martingales. Assume moreover that m is a martingale with respect to $\tilde{\mathbf{F}}=\{\mathcal{F}_t^N\vee\mathcal{F}_\infty^X\}$ and that the jump measure μ of X admits a compensator ν of the form $\nu(dt,dy,\omega)=p(t,X_t(\omega),dy)dt$. Then the pair (X,N) is a counting process system.

5.2 MINIMALITY OF CONDITIONALLY POISSON SYSTEMS

In this section we will confine ourselves to stationary systems. This means that the functions A,g and λ in theorem 5.1.10 are not explicitly dependent on t. So we use the representations

$$dY_t = AY_tdt + dM_t^Y, \quad Y_0 \tag{5.8a}$$
$$dN = CY_tdt + dm_t, \quad N_0=0 \tag{5.8b}$$

Here C is a row vector in $\mathbb{R}^n$ with elements $c_i=\lambda(x_i)$.

Equation (5.8) is called the *forward representation* of the system (X,N). It is also possible to give a *backward representation.* The starting point of this section is the system of equation (5.8). The word minimality in the title refers to the minimality of size of the state space $\mathcal{X}$ in a way to be made precise below. The external behaviour of the system (X,N) is the process N. We call (X,N) minimal if we cannot find a system $(\bar{X},N)$ where $\bar{X}$ has a smaller state space than X. Observe that the external behaviours $(\bar{X},N)$ and (X,N) are both given by the same process N. For $(\bar{X},N)$ we use the equation (5.8) with Y,A and C replaced with $\bar{Y},\bar{C},\bar{A}$.

DEFINITION 5.2.1: The forward representation (5.8) of the system (X,N) is called *strongly reducible* if there exists a set $\bar{\mathcal{X}}$ of lower cardinality than $\mathcal{X}$ and a function $f:\mathcal{X}\to\bar{\mathcal{X}}$ such that with $\bar{X}_t=f(X_t)$, the pair $(\bar{X},N)$ is a stochastic system with a forward representation of the form (5.8) and such that $CY_t=\bar{C}\bar{Y}_t$. In this case (X,N) is called *strongly forwardly reducible.* If (X,N) is not strongly forwardly reducible, it is called *strongly forwardly minimal.*

Some remarks are appropriate.

1. If (X,N) is strongly reducible then the "new" state process $\bar{X}$ is again Markov.

2. The adverb strongly in definition 5.2.1 can be thought of as opposed to weakly. One may call a system weakly reducible if there exists a counting process system $(\bar{X},\bar{N})$ on some possibly different probability space $(\bar{\Omega},\bar{\mathcal{F}},\bar{P})$ such

that the state space of $\bar{X}$ has strictly smaller cardinality than that of $\bar{X}$ and such that $\bar{N}$ is equal to N in distribution. One can also define strong reducibility for the backward representation of (X,N). We will not treat weak problems and problems for the backward representation. For this reason we will speak of mimimality and reducibility throughout this section when we mean strong forward minimality and strong forward reducibility.

The problem that we want to treat is the characterization of minimal counting process systems. In view of remark 1 above we first focus our attention on functions of a Markov process.

From the equivalence of $\mathbb{F}$-Markov processes and solutions of certain linear stochastic differential equations (Propositions 5.1.7 and 5.1.8) it is easy to see when functions of a Markov chain again yield a Markov chain. We give a result that also holds for non stationary chains.

To be specific let as before X be a $\mathbb{F}$-Markov chain with state space $\mathcal{X}$. Let H be another set and $f:\mathcal{X}\to H$ a function. Clearly $f(X)$ is again Markov if f is injective. To avoid trivialities let us assume that $H=\{h_1,\ldots,h_m\}$, $m<n$ and that f is onto. Write $Z_t=f(X_t)$. Associate with Z the indicator process W as usual:

$$W:\Omega\times[0,\infty)\to\{0,1\}^m,\ W_{it}=1_{\{Z_t=h_i\}}.$$

Define $F\in\mathbb{R}^{m\times n}$ by $F_{ij}=1_{\{f(x_j)=h_i\}}$. Notice that $\mathbf{1}_m^T F=\mathbf{1}_n^T$, where $\mathbf{1}_m$ is a column vector with as elements $+1$. Then $W_t=FY_t$. Notice that because f is onto, F has rank m, i.e. it has full row rank. Let $K\in\mathbb{R}^{n\times(n-m)}$ be a fixed matrix such that it columns span *Ker F*. Let as before $A(t)$ be the matrix of transition intensities of X. We have the following.

THEOREM 5.2.2: *Let X be $\mathbb{F}$-Markov with finite state space $\mathcal{X}$. Let $f:\mathcal{X}\to H$. Then $f(X)$ is again $\mathbb{F}$-Markov iff $FA(t)K=0$ where the columns of K span Ker F and F is related to f as indicated above. If this condition is satisfied, then the matrix $B(t)$ of transition intensities of $f(X)$ is given by $B(t)=FA(t)\hat{F}$, where $\hat{F}$ is any right inverse of F.*

PROOF: We have $dY_t=A(t)Y_t dt+dm_t^Y$. Hence

$$dW_t=FA(t)Y_t dt+F dm_t^Y$$

Now Z is $\mathbb{F}$-Markov iff $dW_t=B(t)W_t dt+dm_t^W$ for some matrix-valued function B and a $\mathbb{F}$-martingale m^W. By the uniqueness of the special semimartingale decomposition Z is $\mathbb{F}$-Markov if and only if there is a $B(\cdot)$ such that $FA(t)=B(t)F$. Let $\hat{F}$ be a fixed right inverse of F. It exists, since F has full row rank. Then the last equation implies $B(t)=FA(t)\hat{F}$. Of course for B to be well defined it should not depend on the particular choice of $\hat{F}$.

Starting from $\hat{F}$ all other right inverses G of F are given by $G=\hat{F}+KX$, where $X\in\mathbb{R}^{(n-m)\times m}$ is an arbitrary matrix. Hence $B(t)$ is well defined iff $FA(t)\hat{F}=FA(t)(\hat{F}+KX)$ or iff $FA(t)K\equiv 0$. □

REMARK: The result as such is not new but can be found in a slightly different form in KEMENY and SNELL [10, p.126] where Markov chains in discrete time are considered. However the proof given here is shorter.
We will work with a special right inverse of F, the Moore-Penrose inverse which is defined as $F^+ = F^T(FF^T)^{-1}$ Because of the prominent role that matrices F as defined before play, we will refer to these as reduction matrices. Observe that the only invertible transformations of the state space $\mathcal{X}$ are permutations, which correspond to special reduction matrices F, the permutation matrices, that also have the property $F\mathbf{1}=\mathbf{1}$.

PROPOSITION 5.2.3: *The pair (X,N) is reducible iff there exists a reduction matrix F such that with $\bar{A}=FAF^+, \bar{C}=CF^+$ the equalities $FA=\bar{A}F$ and $C=\bar{C}F$ hold. Moreover for the reduced system $(\bar{X},N)$ the generator of $\bar{X}$ is $\bar{A}$ and the intensity of N is given by $\bar{C}\bar{Y}_t$. In this case one says that F reduces (X,N).*

PROOF: Obvious in view of remark 1 after definition 5.2.1.

REMARK: Observe that from purely algebraic considerations $FA=\bar{A}F$ implies that indeed $\bar{A}$ is a generating matrix of some Markov process. Indeed, let $k=k(j)$ be the unique integer such that $F_{kj}=1$. Then

$$\bar{A}_{ik} = \sum_l F_{il}A_{lj} = \sum_{l\neq j} F_{il}A_{lj} + F_{ij}A_{jj}.$$

Now if $i=k$, then $\bar{A}_{ik} = \Sigma_{l\neq j}F_{il}A_{lj}+A_{jj} \leq \Sigma_{l\neq j}A_{lj}+A_{jj}=0$. And if $i\neq k$, then $\bar{A}_{ik}=\Sigma_{l\neq j}F_{il}A_{lj}\geq 0$. Furthermore $\mathbf{1}^T\bar{A}=0$. Observe also that $FA=\bar{A}F$ is equivalent with $F\Phi(t)=\bar{\Phi}(t)F$, where $\Phi(t)=\exp(At)$ and $\bar{\Phi}(t)=\exp(\bar{A}t)$.

Since the stochastic nature of the pair (X,N) is determined by the pair (A,C) in view of eq. (5.8), we will often speak of minimality or reducibility of (A,C) instead of (X,N).

Observe that the reduction procedure is transitive, which means the following. Suppose F_1 reduces (X,N) into a new system (X_1,N) and suppose that F_2 reduces (X_1,N). Then F_2F_1 reduces the original system (X,N). Indeed if F_1 reduces (X,N) then $F_1A=A_1F_1$ for $A_1=F_1AF_1^+$ and $C=C_1F_1$ for $C_1=CF_1^+$. If then also F_2 reduces (X_1,N), then we can write $F_2A_1=A_2F_2$ and $C_1=C_2F_2$. But then $F_2F_1A=F_2A_1F_1=A_2F_2F_1$ and $C=C_1F_1=C_2F_2F_1$ which is what we have to prove. Notice however that given a reduction matrix F that reduces (X,N) one cannot always decompose F as $F=F_2F_1$, where F_1 reduces (X,N) and F_2 reduces (X_1,N). A simple example is the following. Suppose that X has generator

$$A=\begin{bmatrix} -2 & 1 & 2 \\ 1 & -3 & 2 \\ 1 & 2 & -4 \end{bmatrix}$$

and N has constant intensity $\lambda=\lambda \mathbf{1}^T Y_t$. Then clearly $F=[1\ 1\ 1]$ reduces (X,N) but no reduction matrix $F\in\mathbb{R}^{2\times 3}$ reduces (X,N) as can easily be checked.

DEFINITION 5.2.4: Let the row vector $C\in\mathbb{R}^n$ be given. Then D is defined to be the diagonal matrix *diag*(C) which has as the j-th diagonal element c_j. For $u\in\mathbb{R}, D(u)=(e^{iu}-1)D$.

LEMMA 5.2.5: *Let F be a reduction matrix, with right inverse F^+ and let K be a matrix whose columns span Ker F. Let $\bar{C}=CF^+$ and $\bar{D}=FDF^+$. The following statements are equivalent*
(i) $C=\bar{C}F$
(ii) $FDK=0$
(iii) $FD=\bar{D}F$

PROOF: (i) $\Rightarrow$ (ii):

$$(FDK)_{ij}=\Sigma_k F_{ik}c_k K_{kj}=\Sigma_{k,l}F_{ik}\bar{c}_l F_{lk}K_{kj}.$$

Because of the special form of the matrix F, there is only one nonzero element in each column. Hence a product $F_{ik}F_{lk}$ equals zero if $i\neq l$. Therefore the last summation can be written as

$$\sum_k F_{ik}\bar{c}_i F_{ik}K_{kj}=\bar{c}_i\sum_k F_{ik}^2 K_{kj}=\bar{c}_i\sum_k F_{ik}K_{kj}=0.$$

(ii)$\Rightarrow$(iii): $FDK=0$ means that FD is contained in the left kernel of K which is F. Hence there is a matrix L such that $FD=LF$. But then by postmultiplying with F^+ we obtain $L=FDF^+=\bar{D}$.
(iii)$\Rightarrow$(i): $FD=\bar{D}F$ implies that $\mathbf{1}^T FD=\mathbf{1}^T\bar{D}F$ or $\mathbf{1}^T D=\mathbf{1}^T\bar{D}F$. However $\mathbf{1}^T D=C$ and $\mathbf{1}^T\bar{D}=\bar{C}$. □

REMARK: Assume that $FD=\bar{D}F$ for some reduction matrix F. Then $\bar{D}$ is necessarily diagonal. Indeed we have from this assumption: $F_{ij}c_j=\bar{D}_{ik}F_{kj}+\Sigma_{l\neq j}\bar{D}_{il}F_{lj}$. Assume that $i\neq k$ and multiply this equation with F_{kj}. Then, since $F_{ij}F_{kj}=0$ for $i\neq k$ we have $0=\bar{D}_{ik}F_{kj}$, and hence $\bar{D}_{ik}\Sigma_j F_{kj}=0$. Since the summation $\Sigma_j F_{kj}\geq 1$ for all k, we have $\bar{D}_{ik}=0$.

LEMMA 5.2.6: *Let F and K be as in lemma 5.2.5 and let e_i be the i-th basis vector of $\mathbb{R}^n$. Let (X,N) be a stochastic system as in (5.8). Assume that $FAK=0$. Then F reduces (X,N) if C is such that $Fe_k=Fe_j$ for some k and j implies $c_k=c_j$.*

PROOF: We only have to prove that we can write $C=\bar{C}F$, where $\bar{C}=CF^+$. Observe first that

$$(FF^T)_{ij}=\sum_k F_{ik}F_{jk}=\sum_k F_{ik}\delta_{ij},$$

where δ_{ij} is the Kronecker symbol. In particular $(FF)^T_{ii}=\Sigma_k F_{ik}$ Observe furthermore that for all i,j,k $c_k F_{ik}F_{ij}=c_j F_{ik}F_{ij}$ because of the assumption on C.

Now we calculate $(\bar{C}F)_j=(CF^T(FF^T)^{-1}F)_j=$

$$= \sum_{i,k,l} c_k F_{ik}(FF^T)^{-1}_{il} F_{lj} = \sum_{i,k} c_k F_{ik} F_{ij}(FF^T)^{-1}_{ii} =$$

$$c_j \sum_i F_{ij}(FF^T)^{-1}_{ii} \sum_k F_{ik} = c_j \sum_i F_{ij} = c_j.$$

So $\bar{C}F=C$. □

REMARK: From proposition 5.2.3 it follows that a necessary condition for reduction of (X,N) (or (A,C)) is that some of the c_i are identical. However this condition is not sufficient, since also the transformed process $f(X)$ has to be Markov. See the example that follows after proposition 5.2.3.
However if F reduces a pair (A,C), then, as follows from lemma 5.2.6, at the same time it reduces any other pair $(A,\tilde{C})$, where $\tilde{C}=\tilde{C}F$ for some $\tilde{C}$. Observe that here all the $\tilde{c}_i$ may be different, which is not necessarily the case for the $\bar{c}_i$. This means that if F reduces (A,C), it also reduces any other pair $(A,\hat{C})=(A,\tilde{C}F)$, if there exists a map g such that $g(\tilde{c}_j)=\bar{c}_i$. Or, equivalently, if there exists a map g such that $g(\hat{c}_i)=c_i$. Indeed this equivalence holds, because $g(\hat{c}_i)=g(\Sigma_j \tilde{c}_j F_{ji})=\Sigma_j g(\tilde{c}_j)F_{ji}=\Sigma_j \bar{c}_j F_{ji}=c_i$.

To see whether a system (X,N) is reducible one may check whether the criteria of proposition 5.2.3 hold for a reduction matrix F. If the state space $\mathcal{X}$ is very large this is of course quite a task. So we are looking for more easily verifiable criteria. It turns out, as can be expected, that a definition of stochastic observability offers an alternative approach to find a possible reduction. Before defining this concept, we have to introduce some notation and we also need some properties that are satisfied by the objects that play a role in the following definition.

DEFINITION 5.2.7: Let for each integer $k\geqslant 1$, U^k be the set of bounded left continuous functions from $\mathbb{R}^+$ to $\mathbb{R}^k$. Write $U=U^1$ and if $u\in U^k$, then $u(t)$ will be written as a row vector. Define for $u\in U$, $v\in U^m$, a reduction matrix $F\in\mathbb{R}^{m\times n}$ and $T\geqslant t\geqslant 0$

$$g_F^{u,v}(t,T) = E[\exp(i\int_t^T u(s)dN_s + i\int_t^T v(s)FY_s ds)|\mathcal{F}_t]. \tag{5.9}$$

Because (X,N) is a stochastic system, we may replace the conditioning σ-algebra in (5.9) by $\sigma(X_t)$. Hence there exists a deterministic $h_F^{u,v}(t,T)\in\mathbb{C}^{1\times n}$, such that $g_F^{u,v}(t,T) = h_F^{u,v}(t,T)Y_t$.

The following proposition gives a representation for $h_F^{u,v}(t,T)$ as defined above. We use the following notation throughout the rest of this section. Let x be a row or column vector in $\mathbb{R}^n$. Then diag (x) is the $n\times n$ diagonal matrix with i-th diagonal element equal to x_i. Note that $g_F^{u,v}(t,T)$ can be interpreted as a

conditional characteristic function of part of the future behaviour of the system, given its entire past.

PROPOSITION 5.2.8: *Let $h_F^{u,v}(t,T)$ be as in definition 5.2.7. Then it satisfies the integral equation:*

$$h_F^{u,v}(t,T) = \mathbf{1}^T + \int_t^T h_F^{u,v}(s,T)(i\ diag(v(s)F) + D(u(s)))\Phi(s-t)ds \quad (5.10)$$

In the points where $h_F^{u,v}(\cdot,T)$ is differentiable, we have

$$\frac{\partial}{\partial t}h_F^{u,v}(t,T) = -\ h_F^{u,v}(t,T)(i\ diag(v(t)F) + D(u(t)) + A) \quad (5.11)$$

PROOF: We suppress in this proof the dependence on u,v and F.

Let $\tilde{g}(T)=\exp(i\int_0^T u(s)dN_s + i\int_0^T v(s)FY_s ds)$. Then from the stochastic calculus rule (theorem 2.3.5), we obtain

$$\tilde{g}(T) = 1+\int_0^T \tilde{g}(s\text{-})(e^{iu(s)}-1)dN_s + \int_0^T \tilde{g}(s)iv(s)FY_s ds. \quad (5.12)$$

Now we take conditional expectations in (5.12) w.r.t. $\tilde{\mathcal{F}}_t = \mathcal{F}_t \vee \mathcal{F}_\infty^X$. Because CY is also the $\tilde{\mathbf{F}}$-intensity of N, we get

$$\begin{aligned} E[\tilde{g}(T)|\tilde{\mathcal{F}}_t] &= \tilde{g}(t) + E[\int_t^T \tilde{g}(s)(e^{iu(s)}-1)CY_s ds|\tilde{\mathcal{F}}_t] + \\ &+ E[\int_t^T \tilde{g}(s)iv(s)FY_s ds|\tilde{\mathcal{F}}_t] \\ &= \tilde{g}(t) + \int_t^T E[\tilde{g}(s)|\mathcal{F}_t](e^{iu(s)}-1)CY_s ds + \\ &+ \int_t^T E[\tilde{g}(s)|\tilde{\mathcal{F}}_t]iv(s)FY_s ds. \end{aligned} \quad (5.13)$$

Define $\tilde{g}(t,T)=E[\tilde{g}(t)^{-1}\tilde{g}(T)|\tilde{\mathcal{F}}_t]=\tilde{g}(t)^{-1}E[\tilde{g}(T)|\tilde{\mathcal{F}}_t]$. Then (5.13) yields

$$\tilde{g}(t,T) = 1 + \int_t^T \tilde{g}(t,s)((e^{iu(s)}-1)CY_s + iv(s)FY_s)ds.$$

So, $\tilde{g}(t,T)=\exp(\int_t^T((e^{iu(s)}-1)C+iv(s)F)Y_s ds)$. From this expression we also obtain a "backward" integral equation:

$$\tilde{g}(t,T) = 1 + \int_t^T \tilde{g}(s,T)((e^{iu(s)}-1)C + iv(s)F)Y_s ds. \quad (5.14)$$

Define now $g(t,T)=E[\tilde{g}(t,T)|\mathcal{F}_t]$, and observe that this is indeed the quantity in definition 5.2.7. So we can write $g(t,T)=h(t,T)Y_t$. Then from (5.14)

$$
\begin{aligned}
g(t,T) &= 1 + E[\int_t^T E[\tilde{g}(s,T)|\mathcal{F}_s]((e^{iu(s)}-1)C + iv(s)F)Y_s ds|\mathcal{F}_t] \\
&= 1 + E[\int_t^T g(s,T)((e^{iu(s)}-1)C + iv(s)F)Y_s ds|\mathcal{F}_t] \\
&= 1 + E[\int_t^T h(s,T)Y_s((e^{iu(s)}-1)C + iv(s)F)Y_s ds|\mathcal{F}_t] \\
&= 1 + E[\int_t^T h(s,T)(D(u(s)) + i\,diag(v(s)F))Y_s ds|\mathcal{F}_t] \\
&= 1 + \int_t^T h(s,T)(D(u(s)) + i\,diag(v(s)F))\Phi(s-t)ds\,Y_t \qquad (5.15)
\end{aligned}
$$

Or, since $\mathbb{1}^T Y_t = 1$, $g(t,T) = h(t,T)Y_t$ and (5.15) has to hold for all possible outcomes of Y_t, we get $h(t,T) = \mathbb{1}^T + \int_t^T h(s,T)(D(u(s)) + i\,diag(v(s)F)\,\Phi(s-t)ds$. This proves (5.10).
Furthermore, if differentiation w.r.t. t is allowed, (5.10) yields

$$
\begin{aligned}
\frac{\partial h}{\partial t}(t,T) &= -h(t,T)(D(u(t)) + i\,diag(v(t)F) \\
&\quad - \int_t^T h(s,T)(D(u(s)) + i\,diag(v(s)F)\Phi(s-t)A ds \\
&= -h(t,T)(D(u(t)) + i\,diag(v(t)F)) + (\mathbb{1}^T - h(t,T))A \\
&= -h(t,T)(D(u(t)) + i\,diag(v(t)F) + A),
\end{aligned}
$$

because $\mathbb{1}^T A = 0$. Hence (5.11) holds . □

In several cases an explicit expression for $h_F^{u,v}(t,T)$ is available. We need the following notation. Let $M_1,...,M_k$ be square matrices of the same order. Then we denote by $\Pi_{i=1}^k \star M_i$ the ordered product $M_k M_{k-1}....M_1$.

COROLLARY 5.2.9: *Let $t = t_0 < t_1 < .... < t_k = T$. Let for $j = 1,...,k$, $u_j \in \mathbb{R}$ and $v_j \in \mathbb{R}^m$ and define $u(s) = \Sigma_{j=1}^k u_j 1_{(t_{j-1},t_j]}(s)$, $v(s) = \Sigma_{j=1}^k v_j 1_{(t_{j-1},t_j]}(s)$. Then with this choice of the functions u and v we have*

$$
h_F^{u,v}(t,T) = \mathbb{1}^T \prod_{j=1}^k \star \exp((i\,diag(v_j F) + D(u_j) + A)(t_j - t_{j-1})). \qquad (5.16)
$$

PROOF: Follows directly from equation (5.11).

The usefulness of the $h_F^{u,v}(t,T)$ is partly the content of the next lemma.

LEMMA 5.2.10: *Assume that F reduces (X,N). Let $\mathcal{K} = Ker\,F$ and K be a matrix whose columns span $\mathcal{K}$. Then $h_F^{u,v}(t,T)K \equiv 0$, and hence there exists a factorization $h_F^{u,v}(t,T) = \bar{h}^{u,v}(t,T)F$.*

PROOF: Observe first that, always, $F\ diag(v(s)F) = diag(v(s))F$. From the fact that F reduces (A,C), we have $FAK = FDK = 0$. Hence there exist matrices $N_1, N_1(t)$ and N_2 such that $AK = KN_1$, $\Phi(t)K = KN_1(t)$, $DK = KN_2$. Therefore, with suppression of the dependence on u,v and F:

$$h(t,T)K = \mathbf{1}^T K + \int_t^T h(s,T)(i\ diag(v(s)F) + D(u(s)))\Phi(s-t)Kds =$$

$$= 0 + \int_t^T h(s,T)K(i\,N(s) + (e^{iu(s)} - 1)N_2)N_1(s-t)ds. \quad (5.17)$$

Since $h(t,T)K \equiv 0$ is a solution of (5.17), and since solutions are unique, the proof is finished. □

Apparently, for stepfunctions u and v as in corollary 5.2.9, only the differences $t_j - t_{j-1}$ are important. Therefore we introduce functions $h_{F,k}^{u,v}(t)$ as follows. Let $\{t_j\}_{j=1}^{\infty} \subset \mathbb{R}^+, \{u_j\}_{j=1}^{\infty} \subset \mathbb{R}$, $\{v_j\}_{j=1}^{\infty} \subset \mathbb{R}^m$ and define $h_{F,k}^{u,v}(t)$ as in (5.16) with the differences $t_j - t_{j-1}$ replaced by t_j. Let H be the cone $\{(t,T) \in \mathbb{R}^2 : T \geqslant t \geqslant 0\}$. Clearly for all $u \in U$, $v \in U^m$ and $(t,T) \in H$, $h_F^{u,v}(t,T)$ induces a linear map from $\mathbb{R}^n$ into $\mathbb{C}$. So we can introduce $h_F \in \mathfrak{L}(\mathbb{R}^n, \mathbb{C}^{U \times U^m \times H})$ by $h_F(u,v,(t,T)) = h_F^{u,v}(t,T) \in \mathfrak{L}(\mathbb{R}^n, \mathbb{C})$. Denote by $\mathcal{K}_F$ the kernel of h_F.

In a similar way we can introduce operators $h_{F,k}$, by considering the functions $h_{F,k}^{u,v}(t)$, and their kernels $\mathcal{K}_{F,k}$. Now we can prove the following.

THEOREM 5.2.11: *Let $\mathcal{K}_F$ be Ker h_F and $\mathcal{K}_{F,k} = Ker\ h_{F,k}$. Then*

(i) $\mathcal{K}_{F,1} \supset \mathcal{K}_{F,2} \supset \ldots$ *and* $\bigcap_{j=1}^{\infty} \mathcal{K}_{F,j} = \mathcal{K}_F$

(ii) *If for some j $\mathcal{K}_{F,j} = \mathcal{K}_{F,j+1}$, then $\mathcal{K}_{F,j}$ is D,A and $diag(vF)$ invariant, for all $v \in \mathbb{R}^m$ and $\mathcal{K}_F = \mathcal{K}_{F,j}$*

(iii) $\mathcal{K}_F \subset Ker\ F$

(iv) *If moreover h_F factorizes as $h_F = \bar{h}F$, then $\mathcal{K}_F = Ker\ F$*

PROOF: Since we work with fixed F, we suppress the dependence on F. For notational convenience we also suppress dependence on u and v.

(i) Let $\bar{t}_j = t_{j+1}, j \geqslant 1$. Let $v_1 = v, u_1 = u$.

$$\text{Then } h_{j+1}(t) = h_j(\bar{t})\ \exp((i\ diag(vF) + D(u) + A)t_1). \quad (5.18)$$

Now $h_{j+1}(t)\mathcal{K}_{j+1} \equiv 0$. So in particular for $t_1 = 0$, we get $h_j(\bar{t})\mathcal{K}_{j+1} \equiv 0$, which shows that $\mathcal{K}_{j+1} \subset \mathcal{K}_j$.

Certainly $\mathcal{K} \subset \mathcal{K}_j$, for all j, so $\mathcal{K} \subset \cap_{j=1}^{\infty} \mathcal{K}_j$. But since any $u \in U$ and $v \in U^m$ are pointwise limits of stepfunctions, also the reversed inequality holds.

(ii) Assume $\mathcal{K}_j = \mathcal{K}_{j+1}$. Differentiation of (5.18) with respect to t_1 gives

$$0 \equiv \frac{\partial}{\partial t_1} h_{j+1}(t)\mathcal{K}_j = h_{j+1}(t)\ (i\ diag(vF) + D(u) + A)\mathcal{K}_j. \quad (5.19)$$

Now take in (5.19) $t_1=0$, $u=0$ and $v=0$. Then

$$0\equiv h_j(\bar{t})A\mathcal{K}_j$$

which yields $\mathcal{K}_j$ to be A-invariant. With this information we take in (5.19) $t_1=0$ and $u=0$ but we allow v to be free. This yields $\mathcal{K}_j$ is also $diag(vF)$ invariant for all v. Similarly $\mathcal{K}_j$ is also $D(u)$ invariant for all u, hence D invariant. Hence $\mathcal{K}=\cap_{l=1}^{\infty}\mathcal{K}_l=\mathcal{K}_j$.

(iii) From (ii) we know that $\mathcal{K}$ is $diag(vF)$ invariant (for all v). Hence $Fdiag(vF)\mathcal{K}=0$ or $diag(v)F\mathcal{K}=0 \Rightarrow F\mathcal{K}=0$.

(iv) Obvious in view of (iii).

PROPOSITION 5.2.12: *The following statements are equivalent.*

(i) $h_F^{u,v}(t,T) = \bar{h}^{u,v}(t,T)F$

(ii) $\bar{h}^{u,v}$ *satisfies the integral equation*

$$\bar{h}^{u,v}(t,T) = \mathbb{1}^T + \int_t^T \bar{h}^{u,v}(s,T)(i\,diag(v(s)) + \bar{D}(u(s)))\bar{\Phi}(s-t)ds \quad (5.20)$$

where $FD = \bar{D}F$, $F\Phi(t) = \bar{\Phi}(t)F$

PROOF: (i) $\Rightarrow$ (ii): From theorem 5.2.11, we know that $Ker\,F=Ker\,h_F$ is a D,A and $diag(v(s)F)$ invariant subspace of $\mathbb{R}^n$. So there exist matrices $\bar{D}$ and $\bar{A}$ such that $FD=\bar{D}F$, $FA=\bar{A}F$, and as always we have $F\,diag(v(s)F)=diag(v(s))F$. Hence

$$\begin{aligned}\bar{h}_F^{u,v}(t,T)F &= \mathbb{1}^T + \int_t^T \bar{h}^{u,v}(s,T)F(i\,diag(v(s)F) + D(u(s)))\Phi(s-t)ds \\ &= \mathbb{1}^T + \int_t^T \bar{h}^{u,v}(s,T)\,(i\,diag(v(s)) + \bar{D}(u(s)))\bar{\Phi}(s-t)dsF\end{aligned}$$

After postmultiplication with F^+, the claim follows.

(ii) $\Rightarrow$ (i): Postmultiply (5.20) by F, then we see that $\bar{h}^{u,v}(t,T)F$ satisfies the same integral equation as $h_F^{u,v}(t,T)$. Because $\bar{h}^{u,v}(T,T)F=\mathbb{1}^TF=\mathbb{1}^T=h_F^{u,v}(T,T)$, the claim follows. □

The following proposition, that summarizes some of the preceding results forms the basis of definition 5.2.14 below and makes it understandable if one keeps the interpretation of $g_F^{u,v}(t,T)$ as a conditional characteristic function in mind.

PROPOSITION 5.2.13: *Let* $\bar{X}_t=f(X_t)$, $\bar{Y}_t=FY_t$, *where the reduction matrix* F *is associated with* f *as usual. There is equivalence between*

(i) $E[g_F^{u,v}(t,T)|\sigma(\bar{X}_t)]=g_F^{u,v}(t,T)$ *for all* u,v *and* $t\leq T$.

(ii) (X,N) *is a stochastic system and* $g_F^{u,v}(t,T)=E[g_F^{u,v}(t,T)|\mathcal{F}_t^{\bar{X}}\vee\mathcal{F}_t^N]$ *for all* u,v *and* $t\leq T$.

(iii) *There exists a factorization* $h_F^{u,v}(t,T) \equiv \bar{h}^{u,v}(t,T)F$.

PROOF: (i) $\Rightarrow$ (iii): There exists a matrix Q_t such that $E[Y_t|\sigma(\bar{X}_t)] = Q_t \bar{Y}_t$. (i) then implies $h_F^{u,v}(t,T)Y_t = h_F^{u,v}(t,T)Q_t\bar{Y}_t = h_F^{u,v}(t,T)Q_tFY_t$. So take $\bar{h}^{u,v}(t,T) = h_F^{u,v}(t,T)Q_t$

(iii) $\Rightarrow$ (i): $E[g_F^{u,v}(t,T)|\sigma(\bar{X}_t)] = E[\bar{h}^{u,v}(t,T)FY_t|\sigma(\bar{X}_t)] = E[\bar{h}^{u,v}(t,T)\bar{Y}_t|\sigma(\bar{X}_t)]$

$$= \bar{h}^{u,v}(t,T)\bar{Y}_t = h_F^{u,v}(t,T)Y_t.$$

(iii) $\Rightarrow$ (ii): $E[g_F^{u,v}(t,T)|\mathcal{F}_t^{\bar{X}} \vee \mathcal{F}_t^N] = E[E[g_F^{u,v}(t,T)|\mathcal{F}_t^X \vee \mathcal{F}_t^N]|\mathcal{F}_t^{\bar{X}} \vee \mathcal{F}_t^N]$

$$= E[h_F^{u,v}(t,T)Y_t|\mathcal{F}_t^{\bar{X}} \vee \mathcal{F}_t^N] = E[\bar{h}_F^{u,v}(t,T)\bar{Y}_t|\mathcal{F}_t^{\bar{X}} \vee \mathcal{F}_t^N] = g_F^{u,v}(t,T).$$

This, together with proposition 5.2.12 also shows that $(\bar{X},N)$ is a stochastic system.

(ii) $\Rightarrow$ (iii): $h_F^{u,v}(t,T)Y_t = E[g_F^{u,v}(t,T)|\mathcal{F}_t^{\bar{X}} \vee \mathcal{F}_t^N] = E[g_F^{u,v}(t,T)|\sigma(\bar{X}_t)]$
$= \bar{h}_F^{u,v}(t,T)\bar{Y}_t$ for some deterministic $\bar{h}_F^{u,v}(t,T)$ since the last conditional expectation is a function of $\bar{X}_t$. Because $\bar{Y}_t = FY_t$ the result now follows. □

DEFINITION 5.2.14: The n-th order system (X,N) is said to be *strongly stochastically observation equivalent* with some m-th order system $(m \leqslant n)$ if there exists a reduction matrix $F \in \mathbb{R}^{m \times n}$ such that a factorization $h_F = \bar{h}F$ holds. If any such factorization implies that F is a permutation matrix, then (X,N) will be called *strongly stochastically observable.*

Some comments are appropriate. Let (X,N) be described by equation (5.8). If (X,N) is strongly stochastically observation equivalent with some m-th order system, then from propositions 5.2.12 and 5.2.13 it follows that this one is described via matrices $\bar{A}$ and $\bar{C}$ by an equation like (5.8). Therefore we will also say that (A,C) is strongly stochastically observation equivalent with $(\bar{A},\bar{C})$.

The interpretation is as follows. If we condition the distribution of the future of the bivariate stochastic process $(f(X),N)$ on the entire past of (X,N), or equivalently just on the current state X_t, then this determines $f(X_t)$ only, instead of X_t itself. We also know from proposition 5.2.13 that $(f(X),N)$ is again a stochastic system.

Suppose now that (X,N) is strongly stochastically observable and that F is a $m \times n$ reduction matrix $(m < n)$. Then a factorization $h_F = \bar{h}G$ always exists for another reduction matrix G, which may be the identity (or a permutation matrix). However, because then $Ker\, G \subset Ker\, h_F \subset Ker\, F$ (see theorem 5.2.11 (iii)), where the last inclusion is strict, it follows that there exists yet another reduction matrix H such that $HG = F$. Hence the conditioning of the distribution of $(f(X),N)$ on X_t determines strictly more then $f(X_t)$. Stated otherwise, $f(X_t)$ is not sufficient to predict the future distribution of $(f(X),N)$. Note also that in this case (X,N) can only be strongly stochastically observation equivalent with another n-th order system.

We also mention that this definition differs from the current definition of stochastic observability in the literature for linear Gaussian systems, where the

future evolution of the state processes is disregarded. However the Gaussian analogue of our definition is equivalent with what can be found in the literature. The reason behind our alternative is that we now force the transformed process $f(X)$ to be Markov, which is automatically the case in the linear Gaussian situation. Therefore a slightly different terminology appears to be advisable. The idea behind strong stochastical observation equivalence, is that it should provide us with a link to (strong forward) reducibility. Moreover it should give us information about what reductions of the original systems are possible. This is the content of the next result which, although obvious, brings the concepts mimimality and observability together.

THEOREM 5.2.15: *Let (X,N) be given by equation (5.8). Let F be a reductioin matrix and define $\bar{A}=FAF^+$ and $\bar{C}=CF^+$.*

(i) *F reduces (A,C) if and only if (A,C) is strongly stochastically observation equivalent with (A,C).*

(ii) *(A,C) is (strongly forwardly) minimal if and only if (A,C) strongly stochastically observable.*

PROOF: Direct consequence of definition 5.2.14, theorem 5.2.11 and proposition 5.2.12. □

At first glance this theorem seems to be not very helpful, if one is looking for possible reduction of (A,C), since also definition 5.2.14 involves the unknown F that describes the reduction. But it turns out that it is a useful step to the finding of the F (if any) that reduces (A,C). We first introduce some new notation and an auxiliary result. Take in the definition of $h^{u,v}_F(t,T)$ the function v to be identically zero and write instead $h^u(t,T)$. Observe that this quantity does not depend on the specific F anymore. By taking u to be a stepfunction we can again, parallel to what we did after lemma 5.2.10, construct $h^u_k(t)\in\mathbb{C}^n$ and from these the operators h and h_k. Some of the properties of the $h^{u,v}_F(t,T)$ and $h^{u,v}_{F,k}(t)$ carry over to $h^u(t,T)$ and the $h^u_k(t)$. There are however some differences. The precise result is the following.

PROPOSITION 5.2.16: *Let $\mathcal{K}=Ker\,h$, $\mathcal{K}_j=Ker\,h_j$. Then*

(i) *$\mathcal{K}_1\supset\mathcal{K}_2\supset\cdots$, and $\bigcap_{j=1}^{\infty}\mathcal{K}_j=\mathcal{K}$*

(ii) *If for some j $\mathcal{K}_j=\mathcal{K}_{j+1}$, then $\mathcal{K}=\mathcal{K}_j$ and $\mathcal{K}=\mathcal{K}_j$ is D and A invariant.*

(iii) *If F reduces (A,C), then $Ker\,F\subset\mathcal{K}$, hence there exists a factorization $h=\bar{h}F$.*

(iv) *If a factorization $h=\bar{h}F$ exists such that $Ker\,h=Ker\,F$ for a reduction matrix F, then F reduces (A,C).*

(v) *There is equivalence between*

  (a) *$h^u(t,T)\equiv\bar{h}^u(t,T)G$ and $Ker\,h=Ker\,G$ for some matrix G and*

  (b) *$\bar{h}^u(t,T)$ satisfies the integral equation*
  $\bar{h}^u(t,T)=\mathbb{1}^T+\int\bar{h}^u(s,T)\bar{D}(u(s))\bar{\Phi}(s-t)ds$
  where $\bar{D}G=G\bar{D}$, $\bar{D}(u(s))=(e^{iu(s)}-1)\bar{D}$ and $\bar{\Phi}(t)G=G\Phi(t)$.

PROOF: (i) and (ii) are proved in the same way as (i), (ii) of theorem 5.2.11, (iii) follows in the same way as lemma 5.2.10, (iv) follows from (ii) and (v) can be proved as proposition 5.2.12. □

REMARK: The most striking difference between h and h_F is the following. Suppose that $\mathcal{K}=Ker\, h\neq\{0\}$. Then we have a factorization $h=\bar{h}G$ for some G which is such that $Ker\, h=Ker\, G$. It may happen that it is impossible to choose G to be a reduction matrix. See examples 5.2.23 and 5.2.24. Notice also that we imposed in (v) of proposition 5.2.16 that $Ker\, h=Ker\, G$, whereas for the analogous statement of proposition 5.2.12 the equality $Ker\, h_F=Ker\, F$ automatically holds. The next proposition implicitely offers a way to compute the $\mathcal{K}_j$ and $\mathcal{K}$.

PROPOSITION 5.2.17: *There exist a sequence of matrices W_j, as indicated in the proof, such that $Ker\, h_j=Ker\, W_j$ for all $j\geq 1$.*

PROOF: Let $z=e^{iu}-1$ and let (with a little abuse of notation) $h_1^z(t)=h_1^u(t)$. Let $W_1(z)$ be the $n\times n$ matrix with j-th row equal to $(\frac{\partial}{\partial t})^j h_1^z(0)=\mathbb{1}^T(zD+A)^{j-1}$ (use equation (5.11) with $v=0$). By the Caley-Hamilton theorem for $k\geq n$ one has $(\frac{\partial}{\partial t})^k h_1^z(0)=\Sigma_{j=0}^{n-1}\alpha_{kj}\mathbb{1}^T(zD+A)^j$ for some real numbers α_{kj}. Hence $Ker\, h_1=\mathcal{K}$ if and only if $W_1(z)\mathcal{K}=\{0\}$ for all z. Next we form the matrix W_1 in the following way. Each row $\mathbb{1}^T(zD+A)^{j-1}$ of $W_1(z)$ can be written as $\Sigma_{k=0}^{j-1}z^k\beta_{kj}$, where the β_{kj} are row vectors in $\mathbb{R}^n$. W_1 is now the matrix obtained by stacking all the β_{kj} in a large matrix with n columns. It is evident that $Ker\, h_1=Ker\, W_1$. In an analogous way we can also construct matrices $W_j(z)$ via the partial derivatives of $h_j^z(t)$ with respect to the vector t evaluated at $t=0$. And as above by grouping the equal powers of z that appear in the rows of $W_j(z)$, we obtain a matrix W_j. Hence the $\mathcal{K}_j$ appearing in proposition 5.2.16 are the same as the kernels of the matrices W_j. □

Some additional properties of h are described by the following lemma.

LEMMA 5.2.18:

(i) *For all $u\in U$ the function $h^u(\cdot,T)$ is left differentiable at $t=T$ and $\frac{\partial}{\partial t}h^u(T,T)=-(e^{iu(T)}-1)C$.*

(ii) *Let V be the $n\times n$ Vandermonde matrix with j-th row equal to $\mathbb{1}^T D^{j-1}$. Then $Ker\, h\subset Ker\, V$.*

(iii) *Assume that there exists a reduction matrix F such $h^u(t,T)\equiv\bar{h}^u(t,T)F$. Then C can be written as $\bar{C}F$. If moreover all the elements of $\bar{C}$ are different from each other, then $\mathcal{K}=Ker\, h=Ker\, F$.*

(iv) *If all the c_i are different then $Ker\, h=\{0\}$.*

PROOF: (i). Immediately follows from eq. (5.11) since $h(T,T)=\mathbb{1}^T$ and $\mathbb{1}^T A=0$.

(ii) From proposition 5.2.16 we know that $Ker\,h$ is D invariant and since $\mathbb{1}^T Ker\,h = h(T,T)$, $Ker\,h = \{0\}$ we have $\mathbb{1}^T D^{j-1} Ker\,h = \mathbb{1}^T Ker\,h = \{0\}$.

(iii) We have to prove that $\bar{\mathcal{K}} = Ker\,\bar{h} = \{0\}$. Because of (i) and lemma 5.2.5 there exists a diagonal matrix $\bar{D}$ such that $FD = \bar{D}F$. Now $\bar{\mathcal{K}} = F\bar{\mathcal{K}}$. Hence $\bar{D}\bar{\mathcal{K}} = \bar{D}F\bar{\mathcal{K}} = FD\bar{\mathcal{K}} \subset F\bar{\mathcal{K}} = \bar{\mathcal{K}}$. So $\bar{\mathcal{K}}$ is $\bar{D}$ invariant. If $\bar{V}$ is the Vandermonde matrix with j-th row equal $\mathbb{1}^T \bar{D}^{j-1}$ then we have as in (ii) $\bar{\mathcal{K}} \subset Ker\,\bar{V}$. The latter is zero, since all the elements of $\bar{C}$ are assumed to be different.

(iv) Follows from (iii).

The role that the $h^u(t,T)$ play in the finding of a matrix F that reduces (X,N) is revealed by the following theorem.

THEOREM 5.2.19: *There is equivalence between*

(i) (A,C) *is strongly stochastically observation equivalent with* $(\bar{A},\bar{C})$.

(ii) *There exists a reduction matrix* F *such that* $h^u(t,T) = \bar{h}^u(t,T)F$ *for all* $t \leq T$ *and all* $u \in U$ *and a similar factorization holds for any other pair* $(A,\hat{C})$ *where* $\hat{C} = \tilde{C}F$. *So if* $\hat{h}$ *is related to* $(A,\hat{C})$ *as* h *is to* (A,C), *then:* $\hat{h}^u(t,T) = \tilde{h}^u(t,T)F$ *for all* $t \leq T$ *and all* $u \in U$.

PROOF:

(i) $\Rightarrow$ (ii). From theorem 5.2.15 we know that there exists a reduction matrix F such that $FA = \bar{A}F$ and $C = \bar{C}F$. But then in view of the remark after lemma 5.2.6 F also reduces any $(A,\hat{C})$ where $\hat{C}$ can be written as $\tilde{C}F$. Hence from proposition 5.2.16 we have both the factorization $h^u(t,T) = \bar{h}^u(t,T)F$ and $\hat{h}^u(t,T) = \tilde{h}^u(t,T)F$.

(ii) $\Rightarrow$ (i). Since the assumption holds for any $\hat{C} = \tilde{C}F$, we may take all the elements of $\tilde{C}$ to be different. Then from lemma 5.2.18 (iii) $Ker\,\hat{h} = Ker\,F$ and from proposition 5.2.16 $Ker\,F$ is A invariant, so $FA = \bar{A}F$, with $\bar{A} = FAF^+$. By assumption and from lemma 5.2.5 $FD = \bar{D}F$. Hence F is a matrix that reduces (A,C). The result now follows from theorem 5.2.15 (i).

The following result is closely related to theorem 5.2.19.

PROPOSITION 5.2.20: *Let* (X,N) *satisfy equation (5.8) and let* $(\hat{X},\hat{N})$ *be another stochastic system that satisfies an equation like (5.8) with* $C = [c_1,\ldots,c_n]$ *replaced with* $\hat{C} = [\hat{c}_1,\ldots,\hat{c}_n]$ *and let* $\mathcal{K} = Ker\,h$ *and* $\hat{\mathcal{K}} = Ker\,\hat{h}$. *Assume that there exists a map* g *such that* $c_i = g(\hat{c}_i)$. *Then* $\hat{\mathcal{K}} \subset \mathcal{K}$

PROOF: $\hat{\mathcal{K}}$ is a $\hat{D}$ and A invariant subspace of $\mathbb{R}^n$, where $\hat{D} = diag(\hat{C})$ (proposition 5.2.16). We claim that $\hat{\mathcal{K}}$ is also D invariant. If the claim holds, then it immediately follows from equation (5.10) with $v \equiv 0$, that $h^u(t,T)\hat{\mathcal{K}} = \{0\}$, since $\mathbb{1}^T\hat{\mathcal{K}} = \{0\}$. Since $\hat{\mathcal{K}}$ is $\hat{D}$ invariant, it is spanned by some eigenvectors of $\hat{D}$. So let $k \in \hat{\mathcal{K}}$ be such that $\hat{D}k = \hat{c}_i k$, for one of the eigenvalues $\hat{c}_i$ of $\hat{D}$. Hence for all j we have $\hat{c}_j k_j = \hat{c}_i k_j$ if $k = [k_1,\ldots,k_n]^T$. If $k_j = 0$ then certainly $c_j k_j = c_i k_j$. If

$k_j \neq 0$ then $\hat{c}_j = \hat{c}_i$, but then also $c_j = c_i$. So again we have $c_j k_j = c_i k_j$. Hence k is an eigenvector of D with eigenvalue $c_i = g(\hat{c}_i)$, which shows that $\mathcal{K}$ is also D invariant.

All results so far obtained form the basis of algorithm 5.2.21 below, that yields for a stochastic system (X,N) a minimal representation.

ALGORITHM 5.2.21:

1. Compute *Ker h* and find a reduction matrix F such that $h^u(t,T) \equiv \bar{h}^u(t,T)F$ and such that $\bar{h}^u(t,T)$ cannot be factorized further by means of some other reduction matrix.
2. If *Ker h* = *Ker F*, then the algorithm produces F as its outcome. Else we go to step 3.
3. Let $\hat{C} = CF$, where all the $\tilde{c}_i$ are different from each other. Form $\hat{h}^u(t,T)$ (which is related to $(A,\hat{C})$ as was $h^u(t,T)$ to (A,C)).
4. Apply step 1 to $\hat{h}^u(t,T)$ in lieu of $h^u(t,T)$.

Before proving that the matrix F produced by the algorithm, induces a minimal pair $(\bar{A},\bar{C})$ defined by $\bar{A} = FAF^+, \bar{C} = CF^+$, we discuss the way it works. The finding of F in step 1 is relatively simple. Compute *Ker h* by using the matrices W_k of proposition 5.2.9 as far as needed. This results in a factorization $h^u(t,T) = \tilde{h}^u(t,T)G$, where G is such that *Ker G* = *Ker h*. Next one inspects the columns of G. If any two of them are identical, then the same holds for the corresponding columns of F, which determines F up to a permutation of its columns. An alternative way is to inspect the elements of the $h_k^u(t)$ for all k as far as needed. If two columns of F are identical then the same holds for the corresponding elements of all the $h_k^u(t)$ and vice versa. If the algorithm stops at step 2, then it follows from proposition 5.2.16 that F reduces (A,C). If instead *Ker h* $\neq$ *Ker F* and step 3 is performed then we know from lemma 5.2.18 that some of the elements of $\hat{C}$ (which is such that $C = \bar{C}F$) are identical. Hence it makes sense to construct $\bar{C}$ as prescribed. Then from proposition 5.2.20 we obtain that $\mathit{Ker}\,\hat{h} \subset \mathit{Ker}\,h$ and moreover that this inclusion in strict, since also $\mathit{Ker}\,\hat{h} \subset \mathit{Ker}\,F$ in view of lemma 5.2.18 (ii), applied to the Vandermonde matrix with rows $1^T \hat{D}^{j-1}$, which has kernel equal to *Ker F*. Hence the algorithm constructs a strictly decreasing sequence of kernels, until it terminates which happens after finitely many iterations.

THEOREM 5.2.22: *Let F be the final result of algorithm 5.2.21. Then F reduces (A,C). Hence there exist $\bar{A},\bar{C}$ with $FA = \bar{A}F, C = \bar{C}F$. Moreover $(\bar{A},\bar{C})$ is minimal.*

PROOF: The resulting F has the property that in the final iteration a factorization of the form $\hat{h}^u(t,T) = \tilde{h}^u(t,T)F$ holds, where $\hat{h}^u(t,T)$ corresponds to some pair $(A,\hat{C})$ and where $\mathit{Ker}\,\hat{h} = \mathit{Ker}\,F$. So *Ker F* is $\hat{D}$ and A invariant (proposition 5.2.16) and also D invariant (see the proof of proposition 5.2.20). Hence F reduces (A,C). Now let F_1 be a matrix that gives a maximal reduction of

(A,C). So with $\bar{A}=F_1AF_1^+$ and $\bar{C}=CF_1^+$ we have that $(\bar{A},\bar{C})$ is a minimal pair. F_1 is determined up to a permutation of its columns. Then in step 1 of the algorithm we have a factorization (as follows from proposition 5.2.16) $h^u(t,T)=\bar{h}^u(t,T)F_2F_1$, where possibly another reduction matrix is involved. Suppose that step 2 is skipped, otherwise the proof is complete. So we construct $\tilde{C}=CF_2F_1$. Then of course (lemma 5.2.5) $Ker\,F_1$ is $\hat{D}$ invariant and therefore $\hat{h}^u(t,T)$ factorizes as $\tilde{h}^u(t,T)F_3F_1$, with possibly again another reduction matrix F_3, which has the property that $Ker\,F_3 \subset Ker\,F_2$, because $Ker\,(F_3F_1)\subset Ker\,\hat{h}\subset Ker\,\hat{V}=Ker\,(F_2F_1)$, where $\hat{V}$ is the Vandermonde matrix with j-th row equal to $1^T\hat{D}^{j-1}$. (Use also lemma 5.2.18). Hence in each iteration of the algorithm a factorization of functions like $h^u(t,T)$ holds, where the matrix F_1 is always part of the factorization, and where the kernels of the F_2, F_3 etc. are shrinking. Therefore in the final step of the algorithm we have a factorization of the form $\hat{h}=\tilde{h}F_*F_1$. From the first part of the proof we know that F_*F_1 is a matrix that reduces (A,C), but since F_1 gives the minimal reduction F_* has to be a permutation matrix. □

In the next two examples, we apply algorithm 5.2.21.

EXAMPLE 5.2.23: *Let X takes its values in $\{1,2,3,4,5\}$ and let*

$$A=\begin{bmatrix} -14 & 1 & 1 & 1 & 1 \\ 1 & -14 & 2 & 3 & 1 \\ 9 & 9 & -7 & 7 & 1 \\ 1 & 1 & 2 & -12 & 1 \\ 3 & 3 & 2 & 1 & -4 \end{bmatrix}$$

Assume that N has the intensity CY, where $C=[1\,1\,1\,1\,2]$.
The matrix W_1 of proposition 5.2.17 now becomes (use $1^TA=0$):

$$W_1=\begin{bmatrix} 1^T \\ 1^TD \\ 1^TD^2 \\ 1^TDA \\ 1^TD^3 \\ 1^T(D^2A+DAD) \\ 1^TDA^2 \\ 1^TD^4 \\ 1^T(D^3A+D^2AD+DAD^2) \\ 1^T(D^2A^2+DADA+DA^2D) \\ 1^TDA^3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 & 4 \\ 3 & 3 & 2 & 1 & -4 \\ 1 & 1 & 1 & 1 & 8 \\ 12 & 12 & 8 & 4 & -20 \\ -32 & -32 & -11 & 10 & 25 \\ 1 & 1 & 1 & 1 & 16 \\ 33 & 33 & 22 & 11 & -68 \\ -172 & -172 & -63 & 46 & 166 \\ 402 & 402 & 51 & -300 & 165 \end{bmatrix}$$

Now $Ker\,h_1=Ker\,W_1$ is spanned by $[1\,-1\,0\,0\,0]^T$ and $[0\,1\,-2\,1\,0]^T$. Observe that these two vectors are eigenvectors of both D and A. So $Ker\,h_1=Ker\,h$.

The reduction matrix F in step 1 of algorithm 5.2.21 is easily seen to be

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

since the first two columns of W_1 are identical. Clearly $Ker\ F \neq Ker\ h$. So step 3 of the algorithm applies. Let $C=[1\,3\,4\,2]$, $\hat{C}=[1\,1\,3\,4\,2]$. Of course one can now construct a matrix $\hat{W}_1$. Then $Ker\ \hat{W}_1 \subset Ker\ F$ (lemma 5.2.18 (ii)), which is spanned by $[1\ -1\ 0\ 0\ 0]^T$. Since, as observed above, $[1\ -1\ 0\ 0\ 0]^T$ is $\hat{A}$ and D invariant, we see that $Ker \hat{W}_1 = Ker\ F$ and also, as above, $Ker\ \hat{h} = Ker\ \hat{W}_1$. Hence the outcome of the algorithm is

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The next (partially worked) example is apart from an illustration of algorithm 5.2.21 also interesting in the light of the remark that followed proposition 5.2.16.

EXAMPLE 5.2.24: Change the matrix A in the preceding example into

$$A = \begin{bmatrix} -4 & 1 & 1 & 2 & 1 \\ 1 & -5 & 4 & 2 & 2 \\ 1 & 3 & -8 & 1 & 3 \\ 1 & 0 & 1 & -8 & 4 \\ 1 & 1 & 2 & 3 & -10 \end{bmatrix}$$

but let C be the same. If one again computes the matrix W_1 than it turns out that its kernel $\mathcal{K}_1$ is again spanned by $k_1^T=[1\ -1\ 0\ 0\ 0]^T$ and $k_2^T=[0\,1\ -2\,1\,0]^T$. Let $K=[k_1 k_2]$. A calculation shows that

$$AK = K \begin{bmatrix} -5 & 1 \\ 1 & -10 \end{bmatrix} \text{ and } DK=K$$

Hence $Ker\ h$ is spanned by k_1 and k_2, since already $\mathcal{K}_1$ is a D and A invariant subspace. The matrix F in step 1 of algorithm 5.2.21 is the same as in the preceding example. Take again $\hat{C}=[1\,1\,3\,4\,2]$. The matrix $\hat{W}_1$ contains one row equal to $\mathbf{1}^T\hat{D}\hat{A}=[6\,7\ -11\ -19\,8]$. Since $Ker\ \hat{h}$ belongs to both $Ker\ \hat{W}_1$ and $Ker\ F$ as explained in the discussion after the description of the algorithm, we see that $Ker\ \hat{h}=\{0\}$. The F resulting from the algorithm is therefore the identity matrix (or another permuation matrix). The intriguing feature of this example is, that in spite of the fact that most of the c_i are equal, no reduction is possible.

We close this section with some considerations that indicate ways of future research. Let first $\bar{X}_t = f(X_t)$ and let F the reduction matrix associated with f. Assume $f:\{1,...,n\}\to\{1,...,m\}$. Trivially each entry F_{ij} of F has the following interpretation: $F_{ij} = P(X_t = i | \bar{X}_t = j)$. In both the two examples above we can factorize h as $\bar{h}G$, where

$$G = \begin{bmatrix} 1 & 1 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Observe that each column of G can be considered as a probability vector. The idea is now to extend the interpretation of the F_{ij} as a conditional probability to the entries of G. This idea allows us to consider so called probabilistic reductions of the system (X,N) be looking at suitably defined random functions of X_t. This new approach seems to be connected with the behaviour of the solutions of the filtering problem that is defined by the finding of $E[Y_t|\mathcal{P}_t^N]$. Results in this direction will be reported in another publication. We only mention that in the last example the following identity hold: $GA = \bar{A}G$, where

$$\bar{A} = \begin{bmatrix} -2\frac{1}{2} & 4\frac{1}{2} & 4\frac{1}{2} \\ 1\frac{1}{2} & -7\frac{1}{2} & 5\frac{1}{2} \\ 1 & 3 & -10 \end{bmatrix},$$

which is indeed the rate matrix of some Markov process, that lives on a state space with 3 elements. This already indicates that some reduction, of another type than described in this section, should be possible.

5.3 Self exciting counting process systems

In this section we study what are called self-exciting counting process systems. These can be considered as being on the opposite side of the whole spectrum of counting processes if they are compared to the conditionally Poisson systems. As shown in section 5.1, conditionally Poisson systems can be constructed by a measure transformation, where under the original measure the state process and the counting process were independent. As a consequence the state process and the counting process never have jumps at the same time with probability 1. In this section we will see that for self-exciting systems (the previously made assumption that the state space $\mathcal{X}$ is finite is maintained) the state process can only jump when N jumps. The motivation for studying this class of systems is roughly the following. From a practical point of view it is attractive when the state process evolves on a finite space. For instance finite dimensional filters for state estimation exist in this case. On the other hand one can argue, see also Boel [13], that in the situation where one cannot observe a state process and where there are no physical grounds that lead to an obvious choice of a state model, it is perhaps better to use self-exciting models for identification purposes.

Here we adopt both these points and the question arises whether this yields an interesting model. To put it a little bit more precise, we want to characterize

the class of counting processes that admit an intensity, which is a function of a finite state process which is Markov with respect to the flow of σ-algebras generated by such a counting process. Or, to formulate it in terms of a stochastic realization problem, given a counting process, under what conditions can it be represented as the output of a stochastic system, where the state process assumes finitely many values, and is Markov with respect to the filtration generated by the output.
The purpose of this section is to present a solution of the above stated problems. In particular a detailed investigation is made of finite state processes which are Markov with respect to a given counting process.

DEFINITION 5.3.1: A *self-exciting counting process system* is a stochastic system (X,N) such that $\mathcal{F}_t^X \subset \mathcal{F}_t^N, \forall t \geqslant 0$. So $(\mathcal{F}_t^{X+} \vee \mathcal{F}_t^{\Delta N+}, \mathcal{F}_t^N | \sigma(X_t)) \in CI$.

This definition implies that the state process X is Markov with respect to the filtration $\mathbb{F}^N$, whereas of course for N we still have the equation

$$dN_t = \lambda_t(X_t)dt + dm_t$$

where m is now an $\mathbb{F}^N$-adapted martingale. A good deal of this section is devoted to $\mathbb{F}^N$-Markov processes and we make again the restriction that the state space of X is finite, so assumption 5.1.7 holds.

Before we are treating these $\mathbb{F}^N$-Markov processes we present some preliminary results that will be used later on. Let $(\Omega,\mathcal{F},P)$ be a complete probability space. Let $N:\Omega\times[0,\infty)\to\mathbb{N}_0$ be a counting process and let $\mathcal{F}_t^N = \sigma\{N_s, s \leqslant t\}$ be the σ-algebra generated by the collection $\{N_s, s \leqslant t\}$. Write $\mathbb{F}^N = \{\mathcal{F}_t^N, t \geqslant 0\}$. Assume that N admits the minimal decomposition

$$dN_t = \lambda_t dt + dm_t$$

where $\lambda:\Omega\times[0,\infty)\to\mathbb{R}_+$ is the $\mathbb{F}^N$ -predictable intensity process of N and $m:\Omega\times[0,\infty)\to\mathbb{R}$ is a $\mathbb{F}^N$ -adapted martingale.
The following lemma, known as the martingale representation theorem, plays a crucial role.

LEMMA 5.3.2:

(1) *Let $M:\Omega\times[0,\infty)\to\mathbb{R}$ be an $\mathbb{F}^N$ -adapted martingale. Then there exists an $\mathbb{F}^N$ -predictable process $k:\Omega\times[0,\infty)\to\mathbb{R}$ such that for all $t \geqslant 0$*

$$M_t = M_0 + \int_0^t k_s(dN_s - \lambda_s ds)$$

The process k is $P(d\omega)\lambda_t(\omega)$ a.e. uniquely defined and for all $t \geqslant 0$

$$\int_0^t k_s \lambda_s ds < \infty \text{ a.s.}$$

(2) *Let $S:\Omega\times[0,\infty)\to\mathbb{R}$ be an $\mathbb{F}^N$ -adapted semi-martingale of the form*

$S_t = S_0 + V_t + M_t$. Here V is a process of bounded variation which is assumed to be continuous, $V_0 = 0$ and M is a $\mathbb{F}^N$-adapted martingale.

i) *S can jump only when N jumps i.e. $\Delta S_t \neq 0 \Rightarrow \Delta N_t = 1$*

ii) *If moreover S is a pure jump process (which is the case if it takes its values in a countable set), then*

$$S_t = S_0 + \int_0^t k_s dN_s$$

and V is absolutely continuous satisfying

$$V_t = \int_0^t k_s \lambda_s ds$$

where k is as in lemma 2.1.

PROOF: The proof of (1) can be found in [3, p. 76]. for (2) we have

i) From lemma 5.3.2 we know that $M_t = \int_0^t k_s (dn_s - \lambda_s ds)$ for some $\mathbb{F}^N$-predictable process k. But then from the assumption that V is continuous $\Delta S_t = \Delta M_t = k_t \Delta N_t$.

ii) Now $S_t - S_0 = \Sigma_{u \leq t} \Delta S_u = \Sigma_{u \leq t} k_u \Delta N_u = \int_0^t k_s dN_s$ and $V_t = S_t - S_0 - M_t = \int_0^t k_s dN_s - \int_0^t k_s (dN_s - \lambda_s ds) = \int_0^t k_s \lambda_s ds$. a.s.

REMARK: The assumption that V is continuous implies that the given decomposition of S is $\mathbb{F}^N$ unique, since S is now a fortiori a special semi-martingale.

PROPOSITION 5.3.3: *Let N and $\tilde{N}$ be two counting processes and let λ and $\tilde{\lambda}$ be their $\mathbb{F}^N$-, respectively $\mathbb{F}^{\tilde{N}}$-predictable intensities. Equivalent are*

(i) *$\mathcal{F}_t^N \subset \mathcal{F}_t^{\tilde{N}}$, and $\mathcal{F}_\infty^N$ and $\mathcal{F}_t^{\tilde{N}}$ are conditionally independent given $\mathcal{F}_t^N$.*

(ii) $\tilde{N}_t = \int_0^t 1_{\{\tilde{\lambda}_s > 0\}} dN_s$ and $\tilde{\lambda}_t = 1_{\{\tilde{\lambda}_t > 0\}} \lambda_t$.

PROOF OF PROPOSITION 5.3.3: We will use the following result which is obvious. Consider two filtrations $\mathbb{F}$ and $\mathbb{G}$, such that for all $t \geq 0$: $\mathcal{F}_t \subset \mathcal{G}_t$. Then there is equivalence between

(i) Any $\mathbb{F}$-martingale is a $\mathbb{G}$-martingale

(ii) $\mathcal{F}_\infty$ and $\mathcal{G}_t$ are conditionally independent given $\mathcal{F}_t$.

(i)$\Rightarrow$(ii): Write $d\tilde{N}_t = \tilde{\lambda}_t dt + d\tilde{m}_t$, the Doob-Meyer decomposition of $\tilde{N}$ with respect to $\mathbb{F}^N$. From the above equivalence $\tilde{m}$ is also an $\mathbb{F}^N$-martingale. Hence $\tilde{m}_t = \int_0^t h_s dm_s$ for a $P(d\omega) \otimes dN_s(\omega)$ a.e. unique process h. Then $d\tilde{N}_t = (\tilde{\lambda}_t - \lambda_t h_t) dt + h_t dN_t$, which gives $dN_t = h_t dN_t$ and $\lambda_t = h_t \lambda_t$. Therefore on the jump times T_k of N we have $h_{T_k}^2 = h_{T_k}$. Hence we can also write $d\tilde{N}_t = h_t d\tilde{N}_t = h_t \tilde{\lambda}_t + h_t d\tilde{m}_t$. From the fact that predictable intensities are unique,

we find $\tilde{\lambda}_t = h_t\tilde{\lambda}_t$ a.s., which implies that $h_t 1_{\{\tilde{\lambda}_t>0\}} = 1_{\{\tilde{\lambda}_t>0\}}$. An obvious choice of h that satisfies this relation is $h'_t = 1_{\{\tilde{\lambda}_t>0\}}$. It is certainly $\mathbb{F}^N$-predictable and

$$E\int_0^\infty 1_{\{h_t\neq 1_{\{\tilde{\lambda}_t>0\}}\}} dN_t = E\sum_{n\geq 1} 1_{\{h_{T_n}\neq 1_{\{\tilde{\lambda}_{T_n}>0\}}\}} =$$

$$= E\sum_{n\geq 1}[1_{\{h_{T_n}=1,\tilde{\lambda}_{T_n}=0\}} + 1_{\{h_{T_n}=0,\tilde{\lambda}_{T_n}>0\}}] = 0,$$

which can be seen as follows. It $h_{T_n}=1$, then $\tilde{N}$ jumps at T_n, so that $\tilde{\lambda}_{T_n}>0$, and if $h_{T_n}=0$, then $\tilde{\lambda}_{T_n}=0$ from $\tilde{\lambda}_t = h_t\tilde{\lambda}_t$. The uniqueness of the process h now gives the result.

(ii)⇒(i): Notice first that $\mathcal{F}_t^{\tilde{N}} \subset \mathcal{F}_t^N$, since by the assumption $\tilde{N}_t = \int_0^t 1_{\{\tilde{\lambda}_s>0\}} dN_s$, the sequence $\{\tilde{T}_k\}$ of jump times of $\tilde{N}$ is contained in the sequence $\{T_k\}$. It is now sufficient to prove that any $\mathbb{F}^N$-martingale is a $\mathbb{F}^N$-martingale. So let M be a $\mathbb{F}^N$-martingale. Then there is a $\mathbb{F}^N$-predictable process h such that $M_t = M_0 + \int_0^t h_s d\tilde{m}_s$. Now

$$\tilde{\lambda}_t dt + d\tilde{m}_t = d\tilde{N}_t = 1_{\{\tilde{\lambda}_t>0\}} dN_t = 1_{\{\tilde{\lambda}_t>0\}}\lambda_t dt + 1_{\{\tilde{\lambda}_t>0\}} dm_t = \tilde{\lambda}_t dt + 1_{\{\tilde{\lambda}_t>0\}} dm_t$$

by assumption. Because of $\mathcal{F}_t^{\tilde{N}} \subset \mathcal{F}_t^N$ the process $1_{\{\tilde{\lambda}_t>0\}}$ is $\mathbb{F}^N$-predictable, hence $\tilde{m}$ is also a $\mathbb{F}^N$-martingale. But then the same conclusions holds for M. □

REMARK: The formulation of condition (ii) of proposition 5.3.4 can be replaced by

(ii)′ There exists a $\mathbb{F}^N$-predictable process u such that

$$\tilde{N}_t = \int_0^t u_s dN_s \text{ and } \tilde{\lambda}_t = u_t\lambda_t.$$

Later on one can identify u as $u_t = 1_{\{\tilde{\lambda}_t>0\}}$, showing that it even becomes $\mathbb{F}^{\tilde{N}}$-predictable.

The next object that we want to study is the class of $\mathbb{F}^N$-Markov processes. We will combine the results of corollary 5.3.3 and propositions 5.1.7, 5.1.8 applied to the situation where $\mathbb{F} = \mathbb{F}^N$ in order to find an integral representation of a finite state $\mathbb{F}^N$-Markov process in terms of its infinitesimal characteristics and the intensity of the counting process. Let as before $\lambda_t^+ = \frac{1}{\lambda_t} 1_{\{\lambda_t>0\}}$, with the understanding that $\frac{0}{0}=0$.

THEOREM 5.3.4: *Let X be an $\mathbb{F}^N$-Markov process with state space $\{x_1, \ldots, x_n\}$ and let Y be the indicator process associated to X as before. Then*

$$Y_t = Y_0 + \int_0^t \lambda_s^+ A(s) Y_{s-} dN_s \qquad (5.15)$$

PROOF: Y is a pure jump process satisfying $Y_t = Y_0 + \int_0^t A(s)Y_s ds + m_t^Y$ where m^Y is a $\mathbb{F}^N$-martingale. Hence a multivariate extension of 5.3.2-(2) applies: $Y_t = Y_0 + \int_0^t k_s dN_s$, where $k:\Omega\times[0,\infty)\to\mathbb{R}^n$ is $\mathbb{F}^N$-predictable. In the notation of this corollary we have $V_t = \int_0^t A(s)Y_s ds$. So k satisfies for all $t \geqslant 0$

$$\int_0^t A(s)Y_s ds = \int_0^t k_s \lambda_s ds.$$

Hence, in order to ensure $\mathbb{F}^N$-predictability of k we have

$$A(t)Y_{t-} = k_t \lambda_t \tag{5.16}$$

Now define $\tilde{k}$ by $\tilde{k}_t = k_t 1_{\{\lambda_t > 0\}}$. Then $\tilde{k}$ is $\mathbb{F}^N$-predictable and

$$0 \leqslant E\int_0^\infty 1_{\{k_t \neq \tilde{k}_t\}} \lambda_t dt \leqslant E\int_0^\infty 1_{\{\lambda_t = 0\}} \lambda_t dt = 0.$$

Hence k and $\tilde{k}$ are $P(d\omega)\lambda_t(\omega)$ a.e. the same. From the uniqueness result of 5.3.2 we may use $\tilde{k}$ as well as k. So we have

$$A(t)Y_{t-} = \tilde{k}_t \lambda_t.$$

Hence

$$A(t)Y_{t-}\lambda_t^+ = \tilde{k}_t \lambda_t \lambda_t^+ = \tilde{k}_t.$$

Now drop the tilde on $\tilde{k}$ and the proof is complete. □

COROLLARY 5.3.5: *We have the following explicit expression for Y:*

$$Y_t 1_{\{T_k \leqslant t < T_{k+1}\}} = \prod_{l=1}^{k} (\lambda_{T_l}^+ A(T_l) + I) Y_0 1_{\{T_k \leqslant t < T_{k+1}\}}$$

PROOF: Immediate from theorem 5.3.4 by noting that $Y_{T_l-} = Y_{T_{l-1}}$ and the fact that Y is right continuous. □

EXAMPLE: Assume that the intensity process λ does not depend on t. Then $\lambda_t(\omega) = \lambda$ for some non random constant λ since $\lambda_0(\cdot)$ is $\mathcal{F}_0^N$-measurable. Assume $\lambda > 0$. Assume further that X is a homogeneous Markov process. Then

$$Y_t 1_{\{T_k \leqslant t < T_{k+1}\}} = (\lambda^{-1}A + I)^k Y_0 1_{\{T_k \leqslant t < T_{k+1}\}}$$

or

$$Y_t = (\lambda^{-1}A + I)^{N_t} Y_0$$

Since Y_t is a unit vector for all t, $\lambda^{-1}A + I$ is a semi-permutation matrix in the sense that each of its columns has exactly one $+1$ entry and the other entries are zero. Of course two $+1$ entries may occur in the same row. Consequently

all the diagonal elements A_{ii} of A are either zero or equal to $-\lambda$. If some $A_{ii}=-\lambda$ then there is in the i-th column A_i of A exactly one A_{ji} equal to $+\lambda$. All the other entries of A_i are zero. If $A_{ii}=0$ for some i then the whole column $A_i=0$.

A similar remark applies to the general expression in corollary 5.3.5. We have for all i $A_{ii}(T_l)\leqslant 0$. Then if $A_{ii}(T_l)<0$ there is exactly one $j=j(i,T_l)$ such that $A_{ji}(T_l)=-A_{ii}(T_l)$. Since T_l can assume any value >0, we have that for each i and t there is exactly one $j=j(i,t)$ such that $A_{ji}(t)=-A_{ii}(t)$, all the other entries in the column $A_i(t)$ being zero.

We will now investigate how λ and A are related. Equation (5.16) relates the intensity λ_t of the counting process with the matrix $A(t)$ of transition intensities of X by means of the intermediate process k. In this subsection we will study this relation a little further.
Multiply (5.16) by Y_{t-}^T to obtain

$$\lambda_t Y_{t-}^T k_t = Y_{t-}^T A(t) Y_{t-} \tag{5.17}$$

At a jump time T_n of the counting process there are two possibilities. If X also jumps then $Y_{T_n}\neq Y_{T_n-}=Y_{T_{n-1}}$ and $Y_{T_{n-1}}^T k_{T_n}=Y_{T_{n-1}}^T(Y_{T_n}-Y_{T_{n-1}})=-1$. If X does not jump then $Y_{T_{n-1}}^T k_{T_n}=0$. So assuming that X jumps we get from (5.17)

$$\lambda_{T_n} = -Y_{T_{n-1}}^T A(T_n) Y_{T_{n-1}} \tag{5.18}$$

This last equation (5.18) suggests the following connection between λ and A:

$$\lambda_t = -Y_{t-}^T A(t) Y_{t-}$$

This connection will be studied in the sequel. First we need a definition. Define $\tilde{N}:\Omega\times[0,\infty)\to\mathbb{R}$ by $\tilde{N}_t=\frac{1}{2}[Y^T,Y]_t$. Here $[Y^T,Y]$ is the optional quadratic variation process of Y. It satisfies

$$Y_t^T Y_t = Y_0^T Y_0 + 2\int_0^t Y_{s-}^T dY_s + [Y^T,Y]_t \tag{5.19}$$

Observe that $\tilde{N}$ counts the transitions of the Markov chain. We now have the following proposition.

PROPOSITION 5.3.6:

i) *$\tilde{N}$ is an $\mathbb{F}^N$ and $\mathbb{F}^Y$-adapted counting process with $\mathbb{F}^N$ and $\mathbb{F}^Y$-predictable intensity $\tilde{\lambda}_t=-Y_{t-}^T A(t) Y_{t-}$*

ii) *$N-\tilde{N}$ is also a counting process. It is only $\mathbb{F}^N$-adapted and has $\mathbb{F}^N$-predictable intensity $\lambda_t+Y_{t-}^T A(t) Y_{t-}$*

iii) *$\tilde{N}_t=\int_0^t 1_{\{\tilde{\lambda}_s>0\}} dN_s$ and $\tilde{\lambda}_t=1_{\{\tilde{\lambda}_t>0\}}\lambda_t$*

iv) *$\mathcal{F}_t^N$ and $\mathcal{F}_\infty^N$ are conditionally independent given $\mathcal{F}_t^{\tilde{N}}$*

PROOF:

i) In view of eq (5.19) we have

$$0 = Y_t^T Y_t - Y_s^T Y_s = 2\int_s^t Y_{u-}^T dY_u + [Y^T, Y]_t - [Y^T, Y]_s =$$
$$= 2\int_s^t Y_u^T A(u) Y_u du + 2\int_s^t Y_{u-}^T dm_u^Y + 2(\tilde{N}_t - \tilde{N}_s).$$

By observing that $\int_0 Y_{u-}^T dm_u^Y$ is again a $\mathbb{F}^N$ and $\mathbb{F}^Y$ martingale we get the desired result according to the definition of intensity.

ii) From known results in stochastic calculus we get $2\tilde{N}_t = [Y^T, Y]_t = \int_0^t k_s^T k_s dN_s = 2\int_0^t 1_{\{Y_s \neq Y_{s-}\}} dN_s$ because we only need to know k at the jump times T_n. If X does not jump when N does then $k_{T_n} = 0$, and if it jumps then $k_{T_n}^T k_{T_n} = 2$. Hence $N_t \leqslant N_t$ for all t and $N_t - \tilde{N}_t = \int_0^t 1_{\{Y_s = Y_{s-}\}} dN_s$ which yields in view of (i) that $N - \tilde{N}$ has the described intensity.

iii) Notice that $1_{\{Y_{T_k} \neq Y_{T_k - 1}\}} = 1_{\{\lambda_{T_k} \geq 0, \tilde{\lambda}_{T_k} > 0\}} = 1_{\{\tilde{\lambda}_{T_k} > 0\}}$, since $\tilde{\lambda}_{T_k} \leqslant \lambda_{T_k}$. Hence $d\tilde{N}_t = 1_{\{\tilde{\lambda}_t > 0\}} dN_t$. But then $d\tilde{N}_t = 1_{\{\tilde{\lambda}_t > 0\}} \lambda_t dt + 1_{\{\tilde{\lambda}_t > 0\}} dm_t$, which shows that $1_{\{\tilde{\lambda}_t > 0\}} \lambda_t$ is the $\mathbb{F}^N$-intensity of N which is then also equal to $\tilde{\lambda}_t$ by part (i).

iv) This is an alternative formulation of (iii) in view of proposition 5.3.3.

An important corollary of this proposition occurs when all the $A_{ii}(t)$ are strictly negative. It is stated as the next

THEOREM 5.3.7:

i) *Let all the $A_{ii}(t)$ be strictly negative. Then $N = \tilde{N}$, $\mathcal{F}_t^Y = \mathcal{F}_t^N$ for all $t > 0$ and* $\lambda_t = -Y_{t-}^T A(t) Y_{t-}$

ii) $k_t = -(Y_{t-}^T A(t) Y_{t-})^{-1} A(t) Y_{t-} = -\Sigma \frac{A_i(t)}{A_{ii}(t)} Y_{it-}$, where k satisfies $d\tilde{N}_t = k_t dN_t$

PROOF:

i) From eq. (5.17) we have

$$\lambda_{T_n} Y_{T_{n-1}}^T k_{T_n} = Y_{T_{n-1}}^T A(T_n) Y_{T_{n-1}} = \Sigma_i A_{ii}(T_n) 1_{\{X_{T_{n-1}} = c_i\}} < 0.$$

Hence $\lambda_{T_n} > 0$ and $k_{T_n} \neq 0$, which means that X always jumps as soon as N jumps. Hence $N = \tilde{N}$. Since always $\mathcal{F}_t^{\tilde{N}} \subset \mathcal{F}_t^Y \subset \mathcal{F}_t^N$ we now also have $\mathcal{F}_t^N = \mathcal{F}_t^Y$. Finally $N = \tilde{N}$ implies $\lambda_t = \tilde{\lambda}_t = -Y_{t-}^T A(t) Y_{t-}$.

ii) This is a simple consequence of formula (5.15) and part i) of the theorem.

It is appropriate to inspect the results of proposition 5.3.6 and theorem 5.3.7 a little closer. In general we have for all $t \geqslant 0$ $\mathcal{F}_t^N \subset \mathcal{F}_t^Y \subset \mathcal{F}_t^N$. In the case described in theorem 5.3.7, we get equality of those σ-algebra's. Since now N is also the total number of jumps (or transitions) of the Markov chain and $\mathcal{F}_t^Y = \mathcal{F}_t^N$ it seems logical to expect that we have in the general situation (where

$\tilde{N}$ counts the transitions of the chain) $\mathcal{F}_t^{\tilde{N}} = \mathcal{F}_t^Y$, which means in words that if we have a Markov chain adapted to a counting process then it is also adapted to the counting process that describes the total number of transitions of the the chain. One could say that $\tilde{N}$ is sort of "minimal" counting processes to which X is adapted.
Next we show that the claim $\mathcal{F}_t^{\tilde{N}} = \mathcal{F}_t^Y$ holds true. It is a consequence of

THEOREM 5.3.8: *Let X be finite state $\mathbb{F}^N$-Markov, then Y_t is $\mathcal{F}_t^{\tilde{N}}$-measurable.*

PROOF: Let $\tilde{T}_1, \tilde{T}_2, \ldots$ be the possibly finite sequence of jump times of $\tilde{N}$. From the discussion leading to (5.18) we see that $\lambda_{\tilde{T}_j} = -Y_{\tilde{T}_j-1}^T A(\tilde{T}_j) Y_{\tilde{T}_{j-1}} > 0$. Consider first $\tilde{T}_1$. Then $\lambda_{\tilde{T}_1}$ is a (measurable) function of $\tilde{T}_1$ only. Hence from $Y_{\tilde{T}_1} = (\lambda_{\tilde{T}_1}^{-1} A(\tilde{T}_1) + I) Y_0$ the random variable $Y_{\tilde{T}_1}$ is also a measurable function of $\tilde{T}_1$ only. But then by induction we find that $Y_{\tilde{T}_n} = (\lambda_{\tilde{T}_n}^{-1} A(\tilde{T}_n) + I) Y_{\tilde{T}_{n-1}}$ is a measurable function of $\tilde{T}_1, \ldots, \tilde{T}_n$, say $Y_{\tilde{T}_n} = y_n(\tilde{T}_1, \ldots, \tilde{T}_n)$.
Consequently, by right continuity of Y, we get with $y_0 = Y_0$

$$Y_t = Y_0 + \sum_{n=1}^{\infty} y_n(\tilde{T}_1, \ldots, \tilde{T}_n) 1_{\{\tilde{T}_n \leq t < \tilde{T}_{n+1}\}}.$$

Notice that y_n is $\mathcal{F}_{\tilde{T}_n}^{\tilde{N}}$-measurable since $\mathcal{F}_{\tilde{T}_n}^{\tilde{N}} = \sigma(\tilde{T}_1, \ldots, \tilde{T}_n)$. Now we invoke the fact that $\mathcal{F}_{\tilde{T}_n}^{\tilde{N}} \cap \{\tilde{T}_n \leq t < \tilde{T}_{n+1}\} = \mathcal{F}_t^{\tilde{N}} \cap \{\tilde{T}_n \leq t < \tilde{T}_{n+1}\}$ (see BRÉMAUD [3, p.308]) to see that indeed Y_t is $\mathcal{F}_t^{\tilde{N}}$ measurable. □

The statement of the theorem is sometimes immediately seen in specific cases. Consider for example the case where $\lambda_t \equiv \lambda > 0$ and A is a constant matrix. Then we have in fact $Y_t = (\lambda^{-1} A + I)^{N_t} Y_0$.

COROLLARY 5.3.9: *If X is a finite state $\mathbb{F}^N$-Markov process, then it is also $\mathbb{F}^{\tilde{N}}$ Markov.*

PROOF: Since a process that is Markov with respect to some filtration is also Markov with respect to any other smaller filtration to which is adapted, this is an immediate consequence of theorem 5.3.8.

So far we have seen the following results. Given the fact that we have a $\mathbb{F}^N$-Markov process X, X is also $\mathbb{F}^N$-Markov and $\tilde{N}$ has intensity $\tilde{\lambda}_t = -Y_{t-}^T A(t) Y_{t-}$, where $\tilde{N}$ is as before the process that counts all the transitions of X. As such these results form necessary conditions that follow from the existence of such processes. One might raise the question how to formulate sufficient conditions on a given Markov matrix function $A(\cdot)$ such that there exists an associated $\mathbb{F}^N$-Markov chain X.
Secondly, given that a process X is $\mathbb{F}^N$-Markov, what other counting processes N do exist such that X is also $\mathbb{F}^N$-Markov.
Answering the first question will be postponed until the end of this section.

Concerning the second one we have - as a converse of previous results -

PROPOSITION 5.3.10: *Let X be $\mathbb{F}^{\tilde{N}}$-Markov. Let N be another counting process with $\mathbb{F}^N$-predictable intensity λ such that*

(i) $\tilde{N}_t = \int_0^t 1_{\{\tilde{\lambda}_s > 0\}} dN_s$

(ii) $\tilde{\lambda}_t = 1_{\{\tilde{\lambda}_t > 0\}} \lambda_t$

Then X is also $\mathbb{F}^N$-Markov.

PROOF: From proposition 5.3.3, we see that $\mathcal{F}_t^{\tilde{N}} \subset \mathcal{F}_t^N$ and that $\mathcal{F}_\infty^{\tilde{N}}$ and $\mathcal{F}_t^N$ are conditionally independent given $\mathcal{F}_t^{\tilde{N}}$. Hence X is certainly $\mathbb{F}^N$-adapted.
Observe first that $\tilde{\lambda}_t = 0 \Leftrightarrow Y_t^T{}_- A(t) Y_{t-} = 0$ implies $A(t) Y_{t-} = 0$ as a result of the fact that $A(t)$ is a Markov-matrix. Since X is $\mathbb{F}^N$-Markov: $dY_t = \tilde{\lambda}_t^+ A(t) Y_{t-} d\tilde{N}_t$ (theorem 5.3.4). Hence

$$\begin{aligned} dY_t &= \tilde{\lambda}_t^+ \lambda_t A(t) Y_t dt + \tilde{\lambda}_t^+ A(t) Y_{t-} d\tilde{m}_t \\ &= A(t) Y_t dt + \tilde{\lambda}_t^+ A(t) Y_{t-} d\tilde{m}_t. \end{aligned}$$

From the conditional independence relation (proposition 5.3.4) the last term is an $\mathbb{F}^N$-martingale. Therefore application of proposition 5.1.8 completes the proof.

REMARK: In view of the remark following the proof of proposition 5.3.3 one can replace conditions (i) and (ii) in proposition 5.3.10 by $\tilde{N}_t = \int_0^t u_s dN_s$ and $\tilde{\lambda}_t = u_t \lambda_t$ for some $\mathbb{F}^N$-predictable process u.

Until now we have studied processes X that are $\mathbb{F}^N$-Markov and thus $\mathbb{F}^{\tilde{N}}$-adapted. As mentioned before, one of the results is then, that X is also $\mathbb{F}^{\tilde{N}}$-Markov (corollary 5.3.9). Knowing this, one can prove all the results mentioned in the foregoing, such as $\tilde{\lambda}_t = Y_t^T{}_- A(t) Y_{t-}$ etc.
An interesting question is to see whether a process which is Markov with respect to its own flow of σ-algebras and which is $\mathbb{F}^N$-adapted, is also $\mathbb{F}^{\tilde{N}}$-Markov. In general this is not true. For instance if N is standard Poisson process and X is defined by $X_t = N_{\frac{1}{2}t}$, then X is $\mathbb{F}^X$-Markov, but not $\mathbb{F}^N$-Markov. Theorem 5.3.11 gives a sufficient condition for an affirmative answer. Let us first remark that any bounded process that is a semi-martingale with respect to some filtration is special. See DELLACHERIE & MEYER [5, VII.25]

THEOREM 5.3.11: *Let X be a finite state $\mathbb{F}^X$-Markov chain and assume that X is adapted to $\mathbb{F}^N$ - for some counting process N. Assume moreover that the indicator process Y, being a $\mathbb{F}^N$ - special semi martingale, admits a decomposition such that the predictable process of finite variation is continuous. Then $\mathcal{F}_t^X = \mathcal{F}_t^N$ $\forall t \geq 0$ and X is $\mathbb{F}^N$-adapted and thus $\mathbb{F}^N$-Markov.*

PROOF: From corollary 5.3.2 we know that $dY_t = k_t dN_t$ for some $\mathbb{F}^N$-

predictable process k. By definition of $\tilde{N}$ we have $d\tilde{N}_t = \frac{1}{2} d[Y^T, Y]_t = [Y^T, Y]_t = \frac{1}{2} k_t^T k_t dN_t$. So $\Delta\tilde{N}_t = 0$ iff $k_t = 0$. Therefore we can write $dY_t = k_t d\tilde{N}_t$. Observe that $\tilde{N}$ is $\mathbb{F}^Y$-adapted. As in BRÉMAUD [3, p. 2, 13], we can interpret k_t as a Radon-Nikodym derivative $\frac{dY_t}{d\tilde{N}_t}$ on the $\mathbb{F}^Y$-predictable sets. Therefore we may take k to be $\mathbb{F}^Y$-predictable. For $\tilde{N}$ we have by its definition

$$d\tilde{N}_t = -Y_{t-}^T dY_t = -Y_{t-}^T A(t) Y_{t-} dt - Y_{t-} dm_t^Y$$

so

$$dY_t = k_t d\tilde{N}_t = -k_t Y_{t-}^T A(t) Y_{t-} dt - k_t Y_{t-}^T dm_t^Y \tag{5.20}$$

on the other hand

$$dY_t = A(t) Y_t dt + dm_t^Y \tag{5.21}$$

Since all processes in (5.20) and (5.21) are $\mathbb{F}^Y$-adapted, we have from the uniqueness of the decomposition of a special semi martingale that $-k_t Y_{t-}^T A(t) Y_{t-} = A(t) Y_{t-}$ a.s., which then leads to $k_t = -(Y_{t-}^T A(t) Y_{t-})^+ A(t) Y_{t-}$. As in the proof of the theorem 5.3.8 we can conclude that Y is $\mathbb{F}^N$-measurable. Therefore $\mathcal{F}_t^Y \subset \mathcal{F}_t^N \subset \mathcal{F}_t^Y$. Hence X is $\mathbb{F}^X$-Markov is now equivalent to X is $\mathbb{F}^Y = \mathbb{F}^N$-Markov. □

REMARK: The statement of theorem 5.3.11 indicates why $N_{\frac{1}{2}t}$ cannot be $\mathbb{F}^N$-Markov. This is immediately seen by noting that $N_{\frac{1}{2}t}$ is $\mathbb{F}^N$-predictable. Hence its dual predictable projection with respect to $\mathbb{F}^N$ is the process itself, which is discontinuous.

We proceed with some consequences of the foregoing for the case where X is a homogeneous chain.

COROLLARY 5.3.12: *Assume that X is a homogeneous chain*

i) *If $A_{ii} < 0$, then in the corresponding column A_i of A there is exactly one $j = j(i)$ such that $A_{ji} = -A_{ii}$ and all other A_{ki}'s are zero. If $A_{ii} = 0$ then the whole column $A_i = 0$.*

ii) *k is now a left continuous piecewise constant process and satisfies*

$$k_t 1_{\{T_n < t \leq T_{n+1}\}} = -\sum_i A_{ii}^+ A_i 1_{\{X_{T_n} = x_i\}} 1_{\{T_n < t \leq T_{n+1}\}}$$

iii) *The sampled chain $\hat{X}_n := X_{T_n}$ is now a deterministic process and completely known given the initial state $\hat{X}_0 = X_0$.*

iv) *If there are no absorbing states, then the process λ assumes only a finite number of values. Specifically $\lambda_t \in \{-A_{11}, \ldots, -A_{nn}\}$.*

PROOF: i), iii) iv) follow immediate from the explicit expression in corollary

5.3.5 ii) requires a little work. Recall that we have $k_t = \lambda_t^+ A Y_{t-}$. Let T be the absorption time of the chain. Then $A Y_{t-} 1_{\{t>T\}} \equiv 0$. Hence $\tilde{\lambda}_t > 0 \Leftrightarrow t \leqslant T$. Therefore

$$\lambda_t 1_{\{t \leqslant T\}} = \tilde{\lambda}_t 1_{\{t \leqslant T\}} = -Y_{t-}^T A Y_{t-} 1_{\{t \leqslant T\}}.$$

Hence $k_t = -\Sigma_i A_{ii}^+ A_i Y_{it-} 1_{\{t \leqslant T\}} = -\Sigma_i A_{ii}^+ A_i Y_{it-}$, because $A_i Y_{it-} 1_{\{t>T\}} = 0$.

At this point one might raise the question in virtue of corollary 5.3.12 iv) whether λ is also a Markov process. Clearly this is the case if all the A_{ii} are different or when they are all the same. Interesting is the case when there exists at least one pair (i,j) such that $A_{ii} = A_{jj}$. We will answer this question by means of theorem 5.2.2. Assume that there are $2 \leqslant m \leqslant n-1$ distinct values among the A_{ii}. Call these $a_1, \ldots, a_m$ and denote for all $j = 1, \ldots, m$ by E_i the set of of all j such that $A_{jj} = a_i$. Define $F \in \mathbb{R}^{m \times n}$ by $F_{ij} = 1_{\{j \in E_i\}}$ We have the following result in the terminology of theorem 5.2.2.

PROPOSITION 5.3.13: *In the terminology of theorem 5.2.2: λ is an $\mathbb{F}^N$-Markov chain iff $FAK = 0$. If the last condition is satisfied then the matrix B of transition intensities of λ is given by $FA\hat{F}$.*

EXAMPLE

i) If

$$A = \begin{bmatrix} -a & 0 & 0 & b \\ a & -b & 0 & 0 \\ 0 & b & -a & 0 \\ 0 & 0 & a & -b \end{bmatrix}$$

then λ is Markov with $B = \begin{bmatrix} -a & b \\ a & -b \end{bmatrix}$ and state space $\{-a, -b\}$. Here we should take $F = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$

ii) If

$$A = \begin{bmatrix} -a & 0 & 0 & b \\ a & -a & 0 & 0 \\ 0 & a & -b & 0 \\ 0 & 0 & b & -b \end{bmatrix}$$

then λ is not Markov, which is seen by calculating $FAK = \begin{bmatrix} a-b \\ -a \quad b \end{bmatrix}$ with, $F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$, $K^T = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$

REMARKS:

i) Although it might happen that λ is not Markov of course (λ, X_-) is jointly Markov.

ii) Since it follows from proposition 5.3.8 iv that the number of values that λ can assume is always at most the number of states that X can assume, we see that a necessary condition for a process X to be Markov is, that it takes values in a set which is at least as big as the set of values of λ: So $n \geq \#\{\lambda_t : t \geq 0\}$. Hence a homogeneous chain X cannot have a finite state space if λ has a continuously varying component. In the same way as checking, whether λ is $\mathbb{F}^N$-Markov one can investigate whether there exist Markov processes X^1 with a smaller state space than X by considering all possible choises of F. Thus obtaining a description of a "minimal" Markov process. This is of some relevance in connection with the stochastic realization problem to be posed at the end of this section.

iii) The case where λ is $\mathbb{F}^N$-Markov itself implies here that it changes value as soon as N jumps. Thus we can immediately see from the A-matrix whether λ is $\mathbb{F}^N$-Markov or not. In the previous example i we see that at jump times λ switches from a to b or conversely, which is in agreement with the fact that it is Markov. In second part of the example we see that it is possible that λ stays in a even when N jumps.

We have seen that the existence of a homogeneous $\mathbb{F}^N$-Markov chain X does not necessarily imply that λ is also $\mathbb{F}^N$-Markov. Hereafter we describe some consequences of the situation where indeed λ is an $\mathbb{F}^N$-Markov process with finite state space. Since in this case λ assumes only a finite number of values it follows that λ (being predictable) may be taken as a left continuous process. Write $X_t = \lambda_{t+}$, the right continuous version of λ. We will apply the previous results to this particular choice of X.

Denote by $\{\lambda_1, \ldots, \lambda_n\}$ the state space of X. If there are no absorbing states then $A_{ii} < 0$ and we have that $\lambda_i = -A_{ii}$ for all i in view of corollary 5.3.12 iv. So all $\lambda_i > 0$.

For reasons of completeness we will show what happens if some of the A_{ii} are equal to zero or if one of the λ_i equals zero. The latter case clearly implies that the corresponding $A_{ii} = 0$. Hence this case is covered by the first one. Define $B \subset \{1, \ldots, n\}$ to be the set of integers i such that λ_i is an absorbing state. Define also $T = inf\{t \geq 0 : X_t \in \{\lambda_i, i \in B\}\}$.

Notice that $T < \infty$ a.s. if and only if $B \neq \varnothing$, and for $i \in B$ we have $A_{ii}(t) \equiv 0$, and hence the whole column $A_i(t) \equiv 0$. The principal result of this subsection is the next proposition which tells that for $t \leq T$ we can more or less identify the intensity λ_t as one of the $A_{ii}(t)$'s, and that $A_{ii}(t)$ only assumes the values $-\lambda_i$ or 0.

PROPOSITION 5.3.14: *Assume that λ is $\mathbb{F}^N$-Markov with state space $\{\lambda_1, \ldots, \lambda_n\}$ and transition intensity matrix $A(t)$. Let T be the absorption time as defined above and B the set of integers corresponding to the absorbing states. Then*

$$\lambda_t = \lambda_T 1_{\{t>T\}} + \sum_{i\in B^c} \lambda_i 1_{\{\lambda_t=\lambda_i\}} 1_{\{A_{ii}(t)=0\}} - \sum_{i\in B^c} A_{ii}(t) 1_{\{\lambda_t=\lambda_i\}}$$

and for $i\in B^c$: $A_{ii}(t) = -\lambda_i$ *if* $A_{ii}(t)<0$.

PROOF: Let $X_t=\lambda_{t+}$, then $Y_{it}=1_{\{X_t=\lambda_i\}}$ and $Y_{it-}=1_{\{\lambda_t=\lambda_i\}}$. In the notation that we have used previously, $\tilde{N}$ has rate

$$\tilde{\lambda}_t = -Y_{t-}^T A(t) Y_{t-} = -\Sigma_{i\in B^c} A_{ii}(t) 1_{\{\lambda_t=\lambda_i\}} 1_{\{t\leqslant T\}}$$

Since $\tilde{\lambda}_t = {}_{\{\tilde{\lambda}_t>0\}}\lambda_t$ (proposition 5.3.6 iii) we have

$$\begin{aligned}\tilde{\lambda}_t 1_{\{t\leqslant T\}} &= 1_{\{\tilde{\lambda}_t>0\}} 1_{\{t\leqslant T\}} \lambda_t + 1_{\{\tilde{\lambda}_t>0\}} 1_{\{t>T\}} \lambda_t \\ &= 1_{\{\tilde{\lambda}_t>0\}} \lambda_t,\end{aligned}$$

since $\tilde{\lambda}_t>0$ implies $t\leqslant T$ and conversely $t>T$ implies $\tilde{\lambda}_t=0$. Hence

$$-\sum_i A_{ii}(t) 1_{\{\lambda_t=\lambda_i\}} 1_{\{t\leqslant T\}} = 1_{\{\tilde{\lambda}_t>0\}} \sum_i \lambda_i 1_{\{\lambda_t=\lambda_i\}}$$

Now let $i\in B^c$. Then

$$-A_{ii}(t) 1_{\{\lambda_t=\lambda_i\}} 1_{\{t\leqslant T\}} = 1_{\{\tilde{\lambda}_t>0\}} 1_{\{\lambda_t=\lambda_i\}} \lambda_i.$$

Observe that

$$1_{\{\lambda_t>0\}} 1_{\{\lambda_t=\lambda_i\}} = 1_{\{A_{ii}(t)<0\}} 1_{\{\lambda_t=\lambda_i\}}$$

and for $i\in B^c$ $\lambda_t=\lambda_i$ implies $t\leqslant T$. Hence we get

$$-A_{ii}(t) 1_{\{\lambda_t=\lambda_i\}} = 1_{\{A_{ii}(t)<0\}} 1_{\{\lambda_t=\lambda_i\}} \lambda_i.$$

Since we may assume that $P(\lambda_t=\lambda_i)>0$ we now get by taking expectations

$$-A_{ii}(t) = 1_{\{A_{ii}(t)<0\}} \lambda_i$$

which proves the second assertion of the proposition. Furthermore

$$\begin{aligned}\lambda_t &= \lambda_T 1_{\{t>T\}} + \lambda_t 1_{\{t\leqslant T\}} = \\ &= \lambda_T 1_{\{t>T\}} + 1_{\{t\leqslant T\}} \sum_{i\in B^c} 1_{\{\lambda_t=\lambda_i\}} \lambda_i \\ &= \lambda_T 1_{\{t\geqslant T\}} + \sum_{i\in B^c} 1_{\{\lambda_t=\lambda_i, A_{ii}(t)=0\}} \lambda_i + \sum_{i\in B^c} 1_{\{\lambda_t=\lambda_i, A_{ii}(t)<0\}} \lambda_i \\ &= \lambda_T 1_{\{t>T\}} + \sum_{i\in B^c} 1_{\{\lambda_t=\lambda_i, A_{ii}(t)=0\}} \lambda_i - \sum_{i\in B^c} A_{ii}(t) 1_{\{\lambda_t=\lambda_i\}}\end{aligned}$$

which proves the first assertion. □

REMARKS

1. If λ is a homogeneous $\mathbb{F}^N$-Markov chain, then A is a constant matrix and we have for $i\in B^c$ the identity $A_{ii}(t)\equiv -\lambda_i$. Hence

$$\lambda_t = \lambda_T 1_{\{t>T\}} - \sum_{i\in B^c} A_{ii} 1_{\{\lambda_t=\lambda_i\}}.$$

And of course if there are no absorbing states or if the value zero is the only one, then $A_{ii}(t) \equiv -\lambda_i$ for all i and $\lambda_t = -\Sigma_{i=1}^{n} A_{ii} 1_{\{\lambda_t = \lambda_i\}}$

2. Now it is easy to see that for any function f which is not injective or constant $f(\lambda)$ cannot be a $\mathbb{F}^N$-Markov chain, since we have tacitly assumed that all the λ_i are different. Hence the number of states of λ is now the minimal number of elements that a set should have in order that it can serve as a state space for some $\mathbb{F}^N$-Markov process. In this sense one can say that λ, if it is $\mathbb{F}^N$-Markov, is the minimal $\mathbb{F}^N$-Markov chain.

We conclude this section by solving a certain stochastic realization problem. The solution involves a technical result on the existence of $\mathbb{F}^N$-Markov processes which is formulated in lemma 5.3.15.

It is known that given a Markov-matrix function $A:[0,\infty) \to \mathbb{R}^{N \times n}$, one can always construct a probability space $(\Omega, \mathcal{F}, P)$ and a Markov process $X:\Omega \times [0,\infty) \to \{1, \ldots, n\}$, such that its transition probabilities are generated by A. This is a consequence of Kolomogorov theorem (2.12)
In this section we are concerned with a version of this problem under a restrictive condition, namely given a complete probability space $(\Omega, \mathcal{F}, P)$ a counting process $n:\Omega \times [0,\infty) \to \mathbb{N}_0$ and a Markov matrix function $A:[0,\infty) \to \mathbb{R}^{n \times n}$, does there exist an $\mathbb{F}^N$-Markov process $X:\Omega \times [0,\infty) \to \{1, \ldots, n\}$ such that A generates its transition probabilities. We know from previous results that given such a process we have the identities $\tilde{\lambda}_t = -Y_{t-}^T A(t) Y_{t-}$ and $\tilde{\lambda}_t = \lambda_t 1_{\{\tilde{\lambda}_t > 0\}}$ and that for each (i,t) such that $A_{ii}(t) < 0$, there exists only one j such that $A_{ji}(t) = -A_{ii}(t)$. Hence for the existence of such a process X this imposes some necessary conditions on the matrix $A(t)$. In lemma 5.3.15 we present a set of sufficient conditions that implies the existence of such a desired process X, and we also give a construction for X. Before stating the theorem let us emphasize that one should not overestimate its content, since in a sense it looks like a tautology. On the other hand it shows how one can extract a $\mathbb{F}^N$-Markov process that is hidden in a suitable matrix function A. After having proved the theorem we give an example, how to use the construction of X.

LEMMA 5.3.15: *Given a counting process N with $\mathbb{F}^N$-predictable intensity λ and a Markov matrix function $A:[0,\infty) \to \mathbb{R}^{n \times n}$. There exists a $\mathbb{F}^N$-Markov process $X:\Omega \times [0,\infty) \to \{1, \ldots, n\}$ with A as its infinitesimal generator if there is a unique sequence of random variables $\{x_m\}_{m \geq 0}, x_m:\Omega \to \{1, \ldots, n\}$ such that the following two conditions hold*

a) $A_{x_m x_m}(T_m)(A_{x_m x_m}(T_m) + \lambda_{T_m} I) = 0, \forall_m$.

b) *If $A_{x_m x_m}(T_m) < 0$ then x_{m+1} is such that $A_{x_{m+1} x_m}(T_m) = -A_{x_m x_m}(T_m)$ and if $A_{x_m x_m}(T_m) = 0$, then $x_{m+1} = x_m$.*

PROOF: Let us define a process $Y^- : \Omega \times [0,\infty) \to \{0,1\}^n$ by requiring that $Y_t^- 1_{\{T_{m-1} < t \leq T_m\}} = Y_{T_m}^- 1_{\{T_{m-1} < t \leq T_m\}}$ and $Y_{iT_m}^- = 1_{\{x_m = i\}}$. Then

$$\begin{aligned}
&\lambda_{T_m}^+ \sum_j A_{ij}(T_m) Y_{jT_m}^- = \lambda_{T_m}^+ \sum_j A_{ij}(T_m) 1_{\{x_m = j\}} = \\
&= \lambda_{T_m}^+ A_{ix_m}(T_m) = \lambda_{T_m}^+ A_{x_{m+1}x_m}(T_m) 1_{\{x_{m+1} = i\}} + \\
&\lambda_{T_m}^+ A_{x_m x_m}(T_m) 1_{\{x_m = i\}} + \lambda_{T_m}^+ A_{ix_m}(T_m) 1_{\{x_m \neq i, x_{m+1} \neq i\}} \\
&= -\lambda_{T_m}^+ A_{x_m x_m}(T_m) 1_{\{x_{m+1} = i\}} 1_{\{A_{x_m x_m}(T_m) < 0\}} + \\
&\lambda_{T_m}^+ A_{x_m x_m}(T_m) 1_{\{x_m = i\}} + 0 \\
&= -\lambda_{T_m}^+ A_{x_m x_m}(T_m) [1_{\{x_{m+1} = i\}} - 1_{\{x_m = i\}}] \\
&= 1_{\{x_{m+1} = i\}} - 1_{\{x_m = i\}} = Y_{iT_{m+1}}^- - Y_{iT_m}^-.
\end{aligned}$$

So in vector notation we have

$$Y_{T_{m+1}}^- - Y_{T_m}^- = \lambda_{T_m}^+ A(T_m) Y_{T_m}^- \tag{5.22}$$

Notice that $\lambda_{T_m} = 0$ implies $A(T_m) Y_{T_m}^- = 0$. Therefore with the usual convention that $\frac{0}{0} = 0$ we have from (5.22)

$$Y_{T_{m+1}}^- - Y_{T_m}^- = \lambda_{T_m}^{-1} A(T_m) Y_{T_m}^-.$$

Define now $Y : \Omega \times [0,\infty) \to \{0,1\}^n$ by $Y_t = Y_{t+}^-$. Then $Y_{T_{m+1}}^- = Y_{T_m}$. Hence (5.22) reads

$$Y_{T_m} - Y_{T_{m-1}} = \lambda_{T_m}^{-1} A(T_m) Y_{T_{m-1}}$$

which can be rephrased as

$$dY_t = \lambda_t^{-1} A(t) Y_{t-} dN_t$$

or

$$dY_t = A(t) Y_{t-} dt + \lambda_t^+ A(t) Y_{t-} dm_t$$

We now want to apply proposition 5.1.8. Therefore we have to verify that Y_{t-} is $\mathbb{F}^N$-predictable. Observe that

$$Y_{t-} 1_{\{T_m < t \leq T_{m+1}\}} = Y_{T_m} 1_{\{T_m < t \leq T_{m+1}\}} \tag{5.23}$$

Now the sequence $\{x_m\}_{m \geq 0}$ is such that x_{m+1} is selected on the basis of knowing x_m and T_m, or iteratively is selected on the knowledge of $\{T_1, \ldots, T_m\}$. Therefore $Y_{iT_m}^- = Y_{iT_{m+1}}^- = 1_{\{x_{m+1} = i\}}$ only depends on $\{T_1, \ldots, T_m\}$. From (5.23) and [3, p.307] we now find the desired result. □

EXAMPLE: Let λ be constant between the jump times T_i and evolve according to $\lambda_1, \lambda_2, \lambda_3, \lambda_1, \lambda_2, \lambda_3 \cdots$ etc. Let

$$A_1 = \begin{bmatrix} -\lambda_1 & 0 & \lambda_2 \\ \lambda_1 & -\lambda_3 & 0 \\ 0 & \lambda_3 & -\lambda_2 \end{bmatrix}.$$

Then we see that A_1 cannot be a transition matrix of a $\mathbb{F}^N$-Markov chain $X:\Omega\times[0,\infty)\to\{1,2,3\}$. Because from condition a) of the theorem we see that $X_t=1$ iff $\lambda_t=\lambda_1$, $X_t=3$ iff $\lambda_t=\lambda_1$ and $X_t=2$ iff $\lambda_t=\lambda_3$. From $X_t=1$ it can only jump to 2 according to A_1. But from the given sequence of λ's it should jump from 1 to 3. However

$$A_2=\begin{bmatrix}-\lambda_1 & 0 & \lambda_3\\ \lambda_1 & -\lambda_2 & 0\\ 0 & \lambda_2 & -\lambda_3\end{bmatrix}$$

is compatible with the sequence of λ's as one can easily verify and thus A_2 can act as the transition matrix of a $\mathbb{F}^N$-Markov chain $X:\Omega\times[0,\infty)\to\{1,2,3\}$.

Finally we will adress a certain stochastic realization problem, and see how we can solve it by means of lemma 5.3.15. Let us state the problem precisely.
We are given a complete filtered probability space $(\Omega,\mathcal{F},\mathbb{F}^N,P)$, where the filtration $\mathbb{F}^N$ is generated by a counting process satisfying $dN_t=\lambda_t dt+dm_t$, where λ is the $\mathbb{F}^N$-predictable intensity process and m a $\mathbb{F}^N$-martingale.
We pose the following question. Does there exist a homogeneous $\mathbb{F}^N$-Markov process X with finite state space $\mathcal{X}$ and a (measurable) function $f:\mathcal{X}\to\mathbb{R}^+$ such that $\lambda_t=f(X_{t-})$?
One can reformulate this question in terms that are used in stochastic realization theory as follows. Given a counting process N on $(\Omega,\mathcal{F},P)$ can we find a stochastic system on $(\Omega,\mathcal{F},\mathbb{F}^N,P)$ such that its state process X is homogeneous and has finite state space $\mathcal{X}$ and such that the output processes is N with $\mathbb{F}^N$-predictable intensity $f(X_{t-})$ for some $f:\mathcal{X}\to\mathbb{R}_+$.
Let us suppose that we can affirmatively answer this question. From corollary 5.3.12 we see that the sequence $\{\lambda_{T_n}\}$ is eventually constant or periodic. This observation also gives us a sufficient condition for solving the problem, which is the content of the next theorem.

THEOREM 5.3.16: *There exists on $(\Omega,\mathcal{F},\mathbb{F}^N,P)$ a finite state $\mathbb{F}^N$-Markov process X with state space $\mathcal{X}$ and a function $f:\mathcal{X}\to\mathbb{R}^+$ such that $\lambda_t=f(X_{t-})$ if and only if there exist a jump time T_k of N such that the sequence $\{\lambda_{T_n}\}$ for $T_n\geq T_k$ is either constant or periodic.*

PROOF: We only have to prove that this condition on λ is sufficient for the existence of X.

(i) Consider first the case where $\{\lambda_{T_m}\}$ is eventually constant. Let k be the smallest integer such that $\lambda_{T_m}=\lambda_{T_N}$ for all $m\geq k$. Now we can construct a $\mathbb{F}^N$-Markov process X with state space $\{1,\ldots,k+1\}$ as follows. Define $A\in\mathbb{R}^{(k+1)\times(k+1)}$ as follows $A_{ii}=-\lambda_{T_{i-1}}$, $A_{i+1,i}=-A_{ii}=\lambda_{T_i-1}$, for $i=1,\ldots,k$ and all other A_{ij}'s equal to zero.

$$A=\begin{bmatrix}-\lambda_{T_0} & & & \\ +\lambda_{T_0} & & & \\ & -\lambda_{T_{k-1}} & & \\ & +\lambda_{T_{k-1}} & 0 & \end{bmatrix}$$

This matrix clearly satisfies the conditions of lemma 5.3.15, which yields the existence of the desired X. The function $f:\{1,\ldots,k+1\}\to\mathbb{R}^+$ we are looking for is of course defined by $f(i)=\lambda_{T_{i-1}}$, $i=1,\ldots,k+1$.

(ii) Consider now the case where $\{\lambda_{T_n}\}$ is eventually cyclic, which means that there exist integers k' and p' such that $\lambda_{T_{i+p'}}=\lambda_{T_i}$ for $i\geq k'$. Let k and p be the smallest of such integers. Now we can construct an $\mathbb{F}^N$-Markov process X with state space $\{1,\ldots,k+p\}$ as follows. Define $A\in\mathbb{R}^{(k+p)\times(k+p)}$ by $A_{ii}=-\lambda_{T_{i-1}}$ for $i=1,\ldots,k+p$,

$$A_{i+1,i}=-A_{ii}=\lambda_{T_{i-1}} \quad \text{for } i=1,\ldots,k+p-1 \text{ and } A_{k+1,k+p}=\lambda_{T_{k+p-1}}.$$

All other A_{ij} are zero.

$$A=\begin{bmatrix} -\lambda_0 & & & & \\ \lambda_0 & & & & \\ & -\lambda_{T_k} & & & +\lambda_{T_{k+p-1}} \\ & +\lambda_{T_k} & & & \\ & & -\lambda_{T_{k+p-2}} & & \\ & & +\lambda_{T_{k+p-2}} & & -\lambda_{T_{k+p-1}} \end{bmatrix}$$

As in the first case the existence of the X we are looking for is guaranteed by theorem 5.1 and f is defined by $f(i)=\lambda_{T_{i-1}}, i=1,\ldots,k+p$ □

REMARK: The behaviour of the system for $t\leq T_k$ (T_k as defined in the proof of theorem 5.3.16) can be considered as the transient behaviour of the system. If one would assume that time runs from minus infinity, instead from zero, then the necessary and sufficient condition in theorem 5.3.16 would read: The sequence $\{\lambda_{T_n}\}$ is either periodic or constant.

One other problem that remains to be solved is that of minimality of the solution of the realization problem. In our context minimality means minimality of the number of elements of the state space E. We have the following result.

COROLLARY 5.3.17: *The solution of the stochastic realization problem as presented in the proof of theorem 5.3.16 is minimal.*

PROOF: In principle one can prove the corollary by applying the $FAK=0$ criterion. Here we give an alternative proof. Consider first case (i). Assume that there exists a function g such that $g(X)$ is Markov and a function h such that $h(g(X_t))=f(X_t)=\lambda_t$. Consider a state j of $X, j\leq k$. Then there is no $i\leq j-1$ such that $g(i)=g(j)$, otherwise the sequence $\{\lambda_{T_n}\}$ would reach a loop, which is forbidden by assumption. Similarly there is no $i\leq k$ such that $g(i)=g(k+1)$,

otherwise the absorption time would be smaller than T_k, which is minimal by construction. This shows that g is injective, so that E is minimal. A similar argument applies to case (ii). Assume again that there is a function g such that $g(X)$ is Markov. For the transient states we have the same argument as in case (i). For the cyclic part of the chain we have for each recurrent state j that there is by definition no transient state $i<j$ such that $g(i)=g(j)$, but also no recurrent state $i<j$ such that $g(i)=g(j)$, because that would contradict the minimality of the number (period) p. Again g is injective. □

The object that we have studied in this section was a stochastic process X that is $\mathbb{F}^N$-Markov, where $\mathbb{F}^N$ denotes the filtration that was generated by some given counting process N, and has finite state space. The additional requirement that X is homogeneous resulted in the fact that then X has to be eventually either cyclic or constant. Consequently the idea of viewing N as the output of a stochastic system, with such a process X as state process, leads to a rather restricted class of counting processes that satisfy this requirement. This partly negative result answers a question posed in the beginning of this section, namely whether we get an interesting class of counting processes that obeys the afore mentioned conditions.

Chapter 6

Conclusions

In this monograph we have considered recursive parameter estimation algorithms and realization problems for counting process systems. As stated earlier one of the problems that show up in recursive estimation is the design of an algorithm. It turned out that exploiting the asymptotic structure of the likelihood ratio process offered a way to find a possible form of a recursive algorithm. The use of the likelihood ratio process for this purpose motivated a detailed study of this process, which has been presented in chapter 3. In chapter 4 we have studied asymptotic properties of various recursive algorithms. The underlying model was a counting process with an intensity of the form $\lambda_t = \theta^T \phi_t$. Because of this linear structure quadratic Lyapunov functions appeared to be a useful tool in establishing almost sure convergence of the recursive estimators. It is of course a serious restriction to confine oneself to intensity processes that exhibit this linear structure. The results that we obtained should be understood as a first step towards an analysis of recursive algorithms for the general case where λ_t depends on θ in a nonlinear way, which problem is of course a real challenge. This kind of problems occur for instance in adaptive filtering. Even if $\lambda_t = \theta^T \phi_t$ where ϕ is a process that is not observed, nonlinear problems arise, because to compute estimates we have to use the conditional expectation of ϕ_t given the past observations of the counting process, which is in general a nonlinear function of θ. We feel that the procedure that we have followed to find a recursive algorithm for the situation, where the intensity has a linear structure, also yields useful algorithms in the nonlinear case. However proving that the resulting estimators converge is not as easy. The difficult point is to find a suitable Lyapunov function. It is not clear whether quadratic forms, which were helpful in the linear case, are again a good choice. Much research remains to be done. In chapter 5 we considered minimality questions for counting process systems. These questions arise in the context of stochastic realization theory. We presented a criterion that enables one to judge whether a given realization of a conditionally Poisson system is minimal. For selfexciting counting processes we have also shown under what conditions such a process can be seen as the output of a stochastic system with a finite state space. For conditionally Poisson processes this is still an open problem, which is interesting enough to merit further research.

REFERENCES

1. I.V. BASAWA and D.J. SCOTT (1983), *Asymptotic optimal inference for nonergodic models,* LNS 17, Springer.
2. P. BILLINGSLEY (1968), *Convergence of probability measures,* Wiley.
3. P. BRÉMAUD (1981), *Point processes and queues,* Springer.
4. C. DELLACHERIE and P.A. MEYER (1975), *Probabilités et potentiel,* Ch I-IV, Hermann.
5. C. DELLACHERIE and P.A. MEYER (1980), *Probabilités et potentiel,* Ch V-VIII, Hermann.
6. P. GREENWOOD and A.N. SHIRYAYEV (1985), *Contiguity and the statistical invariance principle,* Stochastic monographs Vol. 1, Gordon & Breach.
7. I.A. IBRAGIMOV and R.Z. HAS'MINISKII (1985), *Statistical Estimation,* Applications of mathematics 16, Springer.
8. J. JACOD (1979), *Calcul stochastique et problèmes de martingales,* LNM 714, Springer.
9. J. JACOD and A.N. SHIRYAYEV (1987), *Limit theorems for stochastic processes,* to appear
10. J.G. KEMENY and J.L. SNELL (1960), *Finite Markov chains,* Van Nostrand.
11. E. WONG (1971), *Stochastic processes in information and dynamical systems.* Mc Graw-Hill.
12. O.O. AALEN (1978), Non parametric inference for a family of counting processes, *Ann. Stat 6,* pp. 701-726.
13. R.K. BOEL (1985), Modelling, estimation and prediction for jump processes, In: Advances in statistical signal processing, *Vol. 1, JAI Press.*
14. P. BRÉMAUD and J. JACOD (1977), Processus ponctuels et martingales: Résultats récents sur la modélisation et le filtrage, *Adv. Appl. Prob. 9,* pp. 362-416.
15. N. CHRISTOPEIT (1986). Quasi-least squares estimation in semimartingale regression models, *Stochastics 16,* pp. 255-278.
16. K. DZHAPARIDZE (1986), On LAN for counting processes, *Report MS-R8606,* Centre for Mathematics and Computer Science, Amsterdam.
17. M. JACOBSEN (1984), Maximum likelihood estimation in the multiplicative intensity model, *Int. Stat. Rev. 52,* pp. 193-207.
18. J. JACOD (1983), Théorèmes limite pour les processes, In: Ecole d'été de Probabilités de St. Flour XIII, *LNM 1117,* Springer.
19. N. KEIDING (1974), Estimation in the birth process, *Biometrika 61,* pp. 71-80.
20. B. KEMP (1986), An optimal monitor of the rapid-eye-movement brain state, *Biol. Cybern. 54,* pp. 133-139.
21. F. KONECNY (1984), Parameter estimation for point processes with partial observations: A filtering approach, *System & Control letters 4,* pp. 281-286.
22. F. KONECNY (1986), Maximum likelihood estimation for doubly stochastic Poisson process with partial observations, *Stochastics 16,* pp. 51-63.

23. YU. A. KUTOYANTS (1979), Intensity parameter estimation of an inhomogeneous Poisson process, *Probl. Control. Inf. Theory 8, no. 2,* pp. 137-149.
24. YU. A. KUTOYANTS (1979), Locally asymptotic normality for processes of Poisson type, *Izv. Akad. Nauk. Armjan. SSR Ser. Mat. 14, no 1,* pp. 3-20.
25. T.L. LAI and C.Z. WEI (1982), Least squares estimates in Stochastic Regression models with applications to identification and control of dynamic systems, *Ann. Statist. 10,* 154-166.
26. YU. N. LIN'KOV (1981), The asymptotic normality of stochastic integrals and the statistics of discontinuous processes, *Th. Prob. Math. Stat. 23,* pp. 97-107.
27. YU. N. LIN'KOV (1982), Estimation of parameters of counting processes, *Probl. Inf. Transm. 18,* pp. 63-76.
28. YU. N. LIN'KOV (1983), Asymptotic normality of locally square integrable martingales in a series scheme, *Th. Prob. Math. Stat. 27,* pp. 95-103.
29. R.S. LIPTSER and A.N. SHIRYAYEV (1980), A functional central limit theorem for semimartingales, *Th. Prob. Applic. 25,* pp. 667-688.
30. M.F.A.M. VAN MAARSEVEEN (1982), Application of martingales in stochastic systems theory-Surveillance and control of freeway traffic flow, *Ph.D. Thesis,* University of Twente.
31. N.N. NIKITIN and A.A. SNEGOVOY (1984), Recurrent parameter estimation of a nonstationary Poisson flow, *Automatika & Telemechanika,* pp. 86-90.
32. Y. OGATA (1978), The asymptotic behaviour of maximum likelihood estimators for stationary point processes, *Ann. Inst. Statist. Math. 30, Part A,* pp. 243-261.
33. PH. PROTTER (1977), Markov solutions of stochastic differential equations, *Zfw 41,* pp. 39-58.
34. R. REBOLLEDO (1980), Central limit theorems for local martingales, *Zfw 51,* pp. 269-286.
35. H. ROBBINS and D. SIEGMUND (1971), A convergence theorem for nonnegative almost supermartingales and some applications. In: Optimizing methods in statistics, J.S. Rustagi Ed., *Academic Press,* pp. 233-256.
36. B. SAGALOVSKY (1983), Maximum likelihood and related estimation methods in point processes and point process systems, *Ph.D. Thesis,* University of California, Berkeley.
37. J.H. VAN SCHUPPEN and E. WONG (1974), Transformation of local martingales under a change of law, *Ann. Prob. 2,* pp. 879-888.
38. J.H. VAN SCHUPPEN (1979), Stochastic filtering theory: a discussion of concepts, methods and results. In: *"Stochastic control theory and stochastic differential systems",* M. Kohlmann and W. Vogel eds. pp. 209-226.
39. J.H. VAN SCHUPPEN (1983), Convergence results for continuous time adaptive stochastic filtering algorithms, *J. Math. An. Applic. 96,* pp. 209-225.
40. J.H. VAN SCHUPPEN (1984), The weak stochastic realization problem for discrete-time counting processes. Analysis & Optimization of Systems Part 1, A. Bensoussan, J.L. Lions Eds. *LNCIS 62,* Springer.

41. A.N. SHIRYAYEV (1981), Martingales; Recent developments, Results and Applications, *Int. Statist. Rev. 49,* pp. 199-233.
42. V. SOLO: (1979), The convergence of AML, *IEEE Trans. Aut. Control 24,* 958-962
43. P.J.C. SPREIJ and G. KOCH (1983), Software reliability as an application of martingale and filtering theory, *IEEE Trans. Reliability 32,* pp. 342-345.
44. P.J.C. SPREIJ (1985), Parameter estimation for a specific software reliability model. *IEEE Trans. Reliability 34,* pp. 323-328.
45. P.J.C. SPREIJ (1986), An on-line parameter estimation algorithm for counting process observations, *IEEE Trans. Inf. Th. 32,* pp. 300-303.
46. P.J.C. SPREIJ (1986), Recursive parameter estimation for counting processes with linear intensity, *Stochastics 18,* pp. 277-312.
47. D. VERE-JONES (1975), On updating algorithms and inference for stochastic point processes. In: *"Perspectives of in probability and statistics"* J. Gani, Ed.
48. D. SNYDER (1975), *Random point processes,* Wiley.

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print; replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservation of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statement.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Emperical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallman spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *Go-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewal theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequences.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.

2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.

3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.

4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.

5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.

6 A.P.W. Böhm. *Dataflow computation.* 1984.

7 A. Blokhuis. *Few-distance sets.* 1984.

8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.

9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.

10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.

11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.

12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.

13 D. Grune. *On the design of ALEPH.* 1985.

14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.

15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.

16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.

17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.

18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.

19 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 1: Philosophy, framework, computer science.* 1986.

20 B.F. Schriever. *Order dependence.* 1986.

21 D.P. van der Vecht. *Inequalities for stopped Brownian motion.* 1986.

22 J.C.S.P. van der Woude. *Topological dynamix.* 1986.

23 A.F. Monna. *Methods, concepts and ideas in mathematics: aspects of an evolution.* 1986.

24 J.C.M. Baeten. *Filters and ultrafilters over definable subsets of admissible ordinals.* 1986.

25 A.W.J. Kolen. *Tree network and planar rectilinear location theory.* 1986.

26 A.H. Veen. *The misconstrued semicolon: Reconciling imperative languages and dataflow machines.* 1986.

27 A.J.M. van Engelen. *Homogeneous zero-dimensional absolute Borel sets.* 1986.

28 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 2: Applications to natural language.* 1986.

29 H.L. Trentelman. *Almost invariant subspaces and high gain feedback.* 1986.

30 A.G. de Kok. *Production-inventory control models: approximations and algorithms.* 1987.

31 E.E.M. van Berkum. *Optimal paired comparison designs for factorial experiments.* 1987.

32 J.H.J. Einmahl. *Multivariate empirical processes.* 1987.

33 O.J. Vrieze. *Stochastic games with finite state and action spaces.* 1987.

34 P.H.M. Kersten. *Infinitesimal symmetries: a computational approach.* 1987.

35 M.L. Eaton. *Lectures on topics in probability inequalities.* 1987.

36 A.H.P. van der Burgh, R.M.M. Mattheij (eds.). *Proceedings of the first international conference on industrial and applied mathematics (ICIAM 87).* 1987.

37 L. Stougie. *Design and analysis of algorithms for stochastic integer programming.* 1987.

38 J.B.G. Frenk. *On Banach algebras, renewal measures and regenerative processes.* 1987.

39 H.J.M. Peters, O.J. Vrieze (eds.). *Surveys in game theory and related topics.* 1987.

40 J.L. Geluk, L. de Haan. *Regular variation, extensions and Tauberian theorems.* 1987.

41 Sape J. Mullender (ed.). *The Amoeba distributed operating system: Selected papers 1984-1987.* 1987.

42 P.R.J. Asveld, A. Nijholt (eds.). *Essays on concepts, formalisms, and tools.* 1987.

43 H.L. Bodlaender. *Distributed computing: structure and complexity.* 1987.

44 A.W. van der Vaart. *Statistical estimation in large parameter spaces.* 1988.

45 S.A. van de Geer. *Regression analysis and empirical processes.* 1988.

46 S.P. Spekreijse. *Multigrid solution of the steady Euler equations.* 1988.

47 J.B. Dijkstra. *Analysis of means in some non-standard situations.* 1988.

48 F.C. Drost. *Asymptotics for generalized chi-square goodness-of-fit tests.* 1988.

49 F.W. Wubs. *Numerical solution of the shallow-water equations.* 1988.

50 F. de Kerf. *Asymptotic analysis of a class of perturbed Korteweg-de Vries initial value problems.* 1988.

51 P.J.M. van Laarhoven. *Theoretical and computational aspects of simulated annealing.* 1988.

52 P.M. van Loon. *Continuous decoupling transformations for linear boundary value problems.* 1988.

53 K.C.P. Machielsen. *Numerical solution of optimal control problems with state constraints by sequential quadratic programming in function space.* 1988.

54 L.C.R.J. Willenborg. *Computational aspects of survey data processing.* 1988.

55 G.J. van der Steen. *A program generator for recognition, parsing and transduction with syntactic patterns.* 1988.

56 J.C. Ebergen. *Translating programs into delay-insensitive circuits.* 1989.

57 S.M. Verduyn Lunel. *Exponential type calculus for linear delay equations.* 1989.

58 M.C.M. de Gunst. *A random model for plant cell population growth.* 1989.

59 D. van Dulst. *Characterizations of Banach spaces not containing l^1.* 1989.

60 H.E. de Swart. *Vacillation and predictability properties of low-order atmospheric spectral models.* 1989.

61 P. de Jong. *Central limit theorems for generalized multilinear forms.* 1989.

62 V.J. de Jong. *A specification system for statistical software.* 1989.

63 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part I.* 1989.

64 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part II.* 1989.

65 B.M.M. de Weger. *Algorithms for diophantine equations.* 1989.

66 A. Jung. *Cartesian closed categories of domains.* 1989.

67 J.W. Polderman. *Adaptive control & identification: Conflict of conflux?.* 1989.

68 H.J. Woerdeman. *Matrix and operator extensions.* 1989.

69 B.G. Hansen. *Monotonicity properties of infinitely divisible distributions.* 1989.

70 J.K. Lenstra, H.C. Tijms, A. Volgenant (eds.). *Twenty-five years of operations research in the Netherlands: Papers dedicated to Gijs de Leve.* 1990.

71 P.J.C. Spreij. *Counting process systems. Identification and stochastic realization.* 1990.

72 J.F. Kaashoek. *Modeling one dimensional pattern formation by anti-diffusion.* 1990.

73 A.M.H. Gerards. *Graphs and polyhedra. Binary spaces and cutting planes.* 1990.

74 B. Koren. *Multigrid and defect correction for the steady Navier-Stokes equations. Application to aerodynamics.* 1990.

75 M.W.P. Savelsbergh. *Computer aided routing.* 1990.