Entropy Methods for Martingales

Yoichi Nishiyama

H

Entropy Methods for Martingales

Entropie Methodes voor Martingalen
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de Rector Magnificus, Prof. dr. H.O. Voorma ingevolge het besluit van het College voor Promoties in het openbaar te verdedigen op maandag 25 mei 1998 des namiddags te 12.45 uur

door

Yoichi Nishiyama

geboren op 14 januari 1969 te Himeji, Japan

Promotor:
Prof. dr. Richard D. Gill
Mathematisch Instituut, Universiteit Utrecht

Author:
Yoichi Nishiyama
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku
Tokyo 106-8569
Japan
nisiyama@ism.ac.jp

1991 Mathematics Subject Classification:
Primary: 60F05, 60G44, 62G20, 62M09
Secondary: 60F17, 60G42, 60G55, 60J60, 62E20, 62F10, 62F12, 62G05, 62G07, 62M02, 62M05, 62M40

Nishiyama, Yoichi
Entropy Methods for Martingales
Proefschrift Universiteit Utrecht – Met samenvatting in het Nederlands

ISBN 90-393-2015-2
Printed in the Netherlands

Preface

Let me begin with stating my personal history; readers who are not interested may skip to the next paragraph. In the spring of 1992, I was looking for a subject for my master's thesis. I took an interest in censoring problems, and learned that O.O. Aalen's paper in 1978 is the origin of the martingale approach to those problems. The approach has been one of the most active areas in statistics since the early 80's. (Needless to say, R.D. Gill's pioneering monograph in 1980 is also important.) Having read Aalen's paper, I made a conjecture, my first as a statistician, concerning the weak convergence of Nelson-Aalen's estimator in the multiplicative intensity model of point processes with general marks, where the estimator is considered as a set-indexed stochastic process. Unfortunately (or fortunately?), I was not able to solve it in my master's thesis, which consequently dealt with another problem. (At that time, I didn't know the prominent paper by R.M. Dudley in 1978; this means that I didn't know anything about the modern theory of empirical processes.) However, the conjecture brought me the motivation of my current research subject —how to manage entropy methods, which have been developed mainly for i.i.d. empirical processes, in the framework of martingales. The result up to the present is this thesis. An answer to the conjecture is presented in Section 4.1.

I would like to express my greatest gratitude to Prof. R.D. Gill for his advice, comments, kindness, patience and encouragement. He has always been the first reader of my drafts during the last two years, and gave me useful advice every time. His enthusiasm really accelerated my study. Also, although my stay in Utrecht was not originally intended to end up in a Doctor's degree, he has kindly given me this opportunity. It would be a great honor for me to succeed in obtaining a degree at the prestigious University of Utrecht.

I am really grateful to my supervisor in Osaka, Prof. N. Inagaki, for his general statistical advice and constant encouragement; without him, I might not be a statistician. I would like to thank Prof. N. Yoshida for stimulating discussion at many stages of my work since I moved to Tokyo; without him, my statistics might be much weaker. My thanks also go to: Dr. S. Aki for his lectures on empirical processes in the winter of 1993–1994, which inspired me to do this project; Prof. A.W. van der Vaart for a discussion in February 1997 which taught me the importance of maximal inequalities; Prof. S.A. van de Geer for a discussion in November 1997 which improved my understanding of M-estimation. I express my gratitude to Profs. N. Ikeda, K. Isii, S. Shirahata, M. Fukushima, R. Shimizu, K. Hirano, T. Matsunawa, Y. Ogata, S. Eguchi, J. Jacod, Yu.A. Kutoyants and B.Y. Levit for their advice and encouragement.

My work in Utrecht from June 1996 to May 1998 has been supported by a JSPS Fellowship for Research Abroad from the Japan Society for the Promotion of Science. These two years living here have made a strong positive impact on my study, and I hope that my scientific fruits may have achieved the Society's high expectations. I also owe a great deal to the Institute of Statistical Mathematics for allowing such a long period of leave, and for support from the Tokyo side; and to the kind hospitality of the Mathematical Institute at the University of Utrecht, which has made my period of work here exciting beyond my expectations. My special thanks go to the secretarial staff and computer managers both in Tokyo and Utrecht, and to the stochastics group in Utrecht, for their help and kindness; in particular, to Damien White for correcting grammatical erros and awkward sentenses contained in a draft of the preface, and to Erik van Zwet for translating the summary into Dutch.

April 1998, in Utrecht — Yoichi Nishiyama

Contents

Chapter 1

Introduction

1.1 Overview

The purpose of this study is to develop entropy methods, which were first introduced for empirical processes of i.i.d. data, in order to handle some martingales with applications to statistical inference for stochastic processes. Among a lot of directions of statistical applications of the methods, we are concerned with two main themes in this monograph. The following brief description of them is intended to illustrate also the motivation of our work.

Theme 1: Asymptotic normality and efficiency in ℓ^∞-spaces. Let $(E, \mathcal{E})$ be a measurable space. Let $\{Z_i\}_{i\in\mathbb{N}}$ be a sequence of E-valued i.i.d. random variables with the common law P, and let Ψ be a subset of $\mathcal{L}^2(P) = \mathcal{L}^2(E, \mathcal{E}, P)$ with an envelope function belonging to $\mathcal{L}^2(P)$. We are interested in estimating the $\ell^\infty(\Psi)$-valued unknown parameter $P = (P(\psi)|\psi \in \Psi)$ given by $P(\psi) = \int_E \psi(z)P(dz)$; a natural estimator is the empirical process $\mathbb{P}^n = (\mathbb{P}^n(\psi)|\psi \in \Psi)$ given by $\mathbb{P}^n(\psi) = n^{-1}\sum_{i=1}^n \psi(Z_i)$. The Donsker property is then nothing else than the asymptotic normality of $\mathbb{P}^n$ about P; if the class Ψ is P-Donsker then the residual process $\sqrt{n}(\mathbb{P}^n - P)$ converges weakly in $\ell^\infty(\Psi)$ to a Brownian bridge indexed by Ψ. A sufficient condition for the class Ψ to be a Donsker class is that the class satisfies an integrability condition for metric entropy with $\mathcal{L}^2(P)$-bracketing, given by Ossiander (1987), which we will recall later in this chapter.

When we have the asymptotic normality of an estimator, the next interest should be to show its asymptotic efficiency. For this purpose, a general procedure based on the Le Cam theory for finite- or infinite-dimensional parameters goes as follows (see e.g. Chapter 3.11 of van der Vaart and Wellner (1996)):

(i) show the *local asymptotic normality* of a model;

(ii) show the *differentiability* of an unknown parameter;

(iii) show that the sequence of proposed estimators converges weakly to the distribution of the efficient bound specified in terms of some factors appearing in the steps (i) and (ii).

Then, the asymptotic efficiency in the sense of the local asymptotic minimax theorem with certain loss functions follows from the weak convergence shown at the step (iii) and the continuous mapping theorem. Furthermore if the proposed estimator can be shown to be *regular*, then the asymptotic efficiency in the sense of the convolution theorem is also fulfilled.

Van der Vaart and Wellner (1996) illustrated the usefulness of this approach by a discussion about the asymptotic efficiency of the estimator $\mathbb{P}^n$ for the unknown parameter P (see their Section 3.11.1 for the details); the Donsker theorems for empirical processes are applied at the step (iii) above. On the other hand, a merit of the Le Cam theory is that, as seen in the step (i), the i.i.d. setup has been generalized up to local asymptotic normality. It is thus meaningful to present some new limit theorems, which should be useful at step (iii), in order to make full use of the general approach. In particular, such theorems in ℓ^∞-spaces were given mostly for row-independent arrays in the 80's, and have been recently established also for stationary sequences (see Notes to Chapter 3). We consider this subject in some martingale contexts. $\diamondsuit$

Theme 2: Rate of convergence of M-estimators. Let $\theta \rightsquigarrow \gamma(\theta)$ be a deterministic process with parameter θ in a set Θ. Suppose that we are interested in estimating a maximum point θ_0 of the function $\theta \rightsquigarrow \gamma(\theta)$. If $\theta \rightsquigarrow \gamma(\theta)$ is well approximated by a stochastic process $\theta \rightsquigarrow \Gamma^n(\theta)$, a natural estimator would be a maximum point $\widehat{\theta}^n$ of the latter, that is, an M-estimator with respect to the criterion process $\theta \rightsquigarrow \Gamma^n(\theta)$.

In the case of i.i.d. data, those processes are typically given by $\gamma(\theta) = P(\psi_\theta)$ and $\Gamma^n(\theta) = \mathbb{P}^n(\psi_\theta)$, where $\{\psi_\theta : \theta \in \Theta\}$ is a given class of elements of $\mathcal{L}^1(P)$ indexed by Θ. When the data is a sample from a density p_θ with respect to a measure on $(E, \mathcal{E})$, the maximum likelihood estimator is an M-estimator for $\psi_\theta = \log p_\theta$. On the other hand, when $E = \Theta = \mathbb{R}$, if we set $\psi_\theta = 1_{[\theta-a,\theta+a]}$ for a constant $a > 0$, then the maximum point θ_0 of $\theta \rightsquigarrow \gamma(\theta) = P([\theta - a, \theta + a])$ is something like a mode of the unknown distribution P.

Such M-estimation procedures allowing (Θ, d) to be a general metric space have been studied in recent years. A general approach to derive the rate of convergence requires the following (see Theorem 3.2.5 of van der Vaart and Wellner (1996) for the details; a version of the theorem, with some modifications, is given also in this monograph, namely Theorem 5.1.2):

(1.1.1) $$\gamma(\theta) - \gamma(\theta_0) \leq -d(\theta, \theta_0)^2 \quad \text{in a neighborhood of } \theta_0;$$

$$(1.1.2) \qquad E^* \sup_{d(\theta,\theta_0)\le\delta} |(\Gamma^n - \gamma)(\theta) - (\Gamma^n - \gamma)(\theta_0)| \le \phi^n(\delta) \quad \text{for small } \delta > 0.$$

Here, $\delta \leadsto \phi^n(\delta)$ is an appropriate non-decreasing function. When we have checked those conditions, by choosing some constants $r_n > 0$ which satisfy $\phi^n(r_n^{-1}) \le r_n^{-2}$, we can conclude that $r_n d(\widehat{\theta}^n, \theta_0) = O_{P^*}(1)$ for M-estimators $\widehat{\theta}^n = \text{argmax}_{\theta\in\Theta}\Gamma^n(\theta)$. The crucial point of this approach is how to get a moment inequality for the residual processes $\theta \leadsto (\Gamma^n - \gamma)(\theta)$ as in (1.1.2). In the case of i.i.d. data mentioned above, the residual $(\Gamma^n - \gamma)(\theta)$ equals $(\mathbb{P}^n - P)(\psi_\theta)$, and the function $\phi^n(\delta)$ is typically of the form $\phi^n(\delta) = n^{-1/2}\varphi(\delta)$ for a function $\delta \leadsto \varphi(\delta)$ not depending on n; the function $\varphi(\delta) = \delta$ leads to the standard rate $r_n = n^{1/2}$, while $\varphi(\delta) = \sqrt{\delta}$ does to the "cube root asymptotics" $r_n = n^{1/3}$. It should be noted, however, that this method possesses a good potential to be applied in much broader situations. As a matter of fact, some authors have already taken this kind of approach in non-i.i.d. settings, for instance regression models, but most of them are based on some maximal inequalities for i.i.d. empirical processes. With this aim in mind, we develop moment inequalities to obtain a bound (1.1.2) when the residual $(\Gamma^n - \gamma)(\theta)$ is the terminal variable of a martingale. $\diamondsuit$

To handle martingales, we introduce a quantity called "quadratic modulus" in Chapter 2, which plays a key role in this work. For the sake of intuitive explanation, let us recall Ossiander's central limit theorem for i.i.d. sequences under the entropy condition for L^2-bracketing, and next see how to generalize it to a dependent case; the idea of the quantity naturally appears there.

Let $(E, \mathcal{E})$ be a measurable space. Let $\{Z_i\}_{i\in\mathbb{N}}$ be a sequence of E-valued i.i.d. random variables with the common law P. Let Ψ be a subset of $\mathcal{L}^2(P)$ with an envelope function belonging to $\mathcal{L}^2(P)$. For every $\varepsilon \in (0, 1]$ choose $N(\varepsilon)$ pairs of elements of $\mathcal{L}^2(P)$, namely, $[l^{\varepsilon,k}, u^{\varepsilon,k}]$, $k = 1, ..., N(\varepsilon)$, such that for every $\psi \in \Psi$ the relation $l^{\varepsilon,k} \le \psi \le u^{\varepsilon,k}$ holds for some k and that

$$(1.1.3) \qquad \sqrt{\int_E |u^{\varepsilon,k}(z) - l^{\varepsilon,k}(z)|^2 P(dz)} \le \varepsilon.$$

Ossiander's theorem says that if this bracketing procedure can be accomplished with

$$(1.1.4) \qquad \int_0^1 \sqrt{\log N(\varepsilon)}d\varepsilon < \infty,$$

then the sequence of stochastic processes $\psi \leadsto X^n(\psi)$ defined by

$$\begin{aligned} X^n(\psi) &= \sqrt{n}(\mathbb{P}^n - P)(\psi) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{\psi(Z_i) - \int_E \psi(z)P(dz)\right\} \end{aligned}$$

converges weakly in $\ell^\infty(\Psi)$ to a Brownian bridge indexed by Ψ.

Now, let $\{Z_i\}_{i\in\mathbb{N}}$ be an arbitrary sequence of E-valued random variables, and denote by P_i the conditional law of Z_i given $\mathcal{F}_{i-1} = \sigma\{Z_1, ..., Z_{i-1}\}$ ($\mathcal{F}_0$ is the null σ-field). We are interested in the sequence of stochastic processes $\psi \rightsquigarrow X^n(\psi)$ defined by

$$X^n(\psi) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{\psi(Z_i) - \int_E \psi(z)P_i(dz)\right\}.$$

Consider the bracketing procedure as above with (1.1.3) replaced by

$$\sqrt{\int_E |u^{\varepsilon,k}(z) - l^{\varepsilon,k}(z)|^2 P_i(dz)} \le K_i\varepsilon \qquad \text{almost surely,}$$

where K_i is a random variable, not depending on ε and k, that is $\mathcal{F}_{i-1}$-measurable; since the left hand side is random in the present case, we have allowed the random coefficient K_i in the right hand side. Then, a result given in Chapter 3 (Theorem 3.2.2 or 3.3.1) says that if the entropy condition (1.1.4) is satisfied and if the sequence of random variables $\overline{K}^n$ defined by

$$\overline{K}^n = \sqrt{\frac{1}{n}\sum_{i=1}^n |K_i|^2}$$

is bounded in probability, then the asymptotic tightness of the processes $\psi \rightsquigarrow X^n(\psi)$ follows from the finite-dimensional convergence and a Lindeberg condition on an envelope function of Ψ. Ossiander's theorem can be thought as the case of $P_i \equiv P$ and $K_i \equiv 1$. Some quantities "quadratic modulus", which we will define for three kinds of martingales in Chapter 2, have the same spirit as the random variables $\overline{K}^n$; a closer explanation might be that

$$\text{"quadratic modulus"} = \sup_{\varepsilon\in(0,1]} \max_{1\le k\le N(\varepsilon)} \frac{\sqrt{n^{-1}\sum_{i=1}^n \int_E |u^{\varepsilon,k}(z) - l^{\varepsilon,k}(z)|^2 P_i(dz)}}{\varepsilon}.$$

Since this random variable depends awkwardly on the choice of the brackets, and moreover since we will treat also random weight functions, we will actually define the quantities in a slightly different way based on a series of finite partitions of Ψ, avoiding the explicit construction of pairs $[l^{\varepsilon,k}, u^{\varepsilon,k}]$ in the definition of brackets.

The entropy methods were at first recognized to be useful to statisticians chiefly through efforts to seek for sharper and/or more general versions of uniform laws of large numbers and central limit theorems for empirical processes especially in multi-dimensional cases. However, some recent works have shown that a core part of them, namely, chaining and bracketing techniques controlled in terms of entropies, can be applied also to other problems in statistics which are not directly connected to those limit theorems; a good example is M-estimation (recall Theme 2). From our point of view, an important advantage of the methods is that some of the techniques work well also for the dependent case above on the set $\{\overline{K}^n \leq L\}$ for a given constant $L > 0$. Hence, some problems of statistical inference for stochastic processes can be solved by handling some truncations such that the complements like $\{\overline{K}^n > L\}$ are asymptotically negligible for a fixed, large constant L. This is the basic outline of the approach which we frequently take in this work.

Chapters 2 and 3 are concerned with some abstract martingales, while the remaining chapters deal with concrete models in statistics. To be more precise, the stochastic processes treated in the former chapters are some classes $(X^\psi|\psi \in \Psi)$ of martingales, indexed by an arbitrary set Ψ, in the sense that each coordinate process $t \rightsquigarrow X_t^\psi$ for every ψ is an $\mathbb{R}$-valued martingale. We consider the following three situations: (i) each coordinate process $t \rightsquigarrow X_t^\psi$ is represented as a stochastic integral, namely,

$$\begin{aligned} X_t^\psi &= W^\psi * (\mu - \nu)_t \\ &= \int_{[0,t]\times E} W^\psi(\omega, s, z)(\mu(\omega; ds, dz) - \nu(\omega; ds, dz)), \end{aligned}$$

where $W^\psi = W^\psi(\omega, t, z)$ is a predictable function on $\Omega \times \mathbb{R}_+ \times E$, μ is an E-valued multivariate point process, and ν is the predictable compensator of μ; (ii) each process $t \rightsquigarrow X_t^\psi$ is a partial sum process of a discrete time martingale, namely,

$$X_t^\psi = \sum_{i=1}^{\sigma_t} \xi_i^\psi,$$

where $\{\xi_i^\psi\}_{i\in\mathbb{N}}$ is a discrete time martingale and $(\sigma_t)_{t\in\mathbb{R}_+}$ is an increasing family of finite stopping times; (iii) each process $t \rightsquigarrow X_t^\psi$ is a continuous local martingale. There are three reasons why we choose the martingales as the objects of our study. First, the Bernstein inequality, which is a basic tool in the i.i.d. case, is already provided also in the framework of martingales with the modification that a truncation based on the predictable quadratic variation is introduced (Lemma 2.1.1). The second reason is that we can take advantage of the well-developed martingale central limit theorems to establish the finite-dimensional convergence in our situation. Last, but not least, the martingale is a vital concept in analyzing a rich class of statistical models, including the multiplicative intensity models for survival data, Markov chains, the Gaussian white noise model, and diffusion processes derived from stochastic differential equations.

The organization of the monograph is as follows. In Chapter 2, we introduce the quantities "quadratic modulus" and "exponential modulus" for the three kinds of martingales above, and establish maximal inequalities, namely, some bounds for

$$E \sup_t \sup_{\psi,\phi\in\Psi} |X_t^\psi - X_t^\phi| 1_B$$

with the truncation by the set B in terms of the modulus. Those inequalities are not asymptotic estimates, and are applied not only for the proofs of weak convergence theorems in Chapter 3 but also as a crucial tool to derive the rate of convergence of M-estimators in Chapters 5 and 6. As for the case (iii) of continuous local martingales, we study also the continuity of the sample paths along the direction of parameter $\psi \rightsquigarrow X_t^{\psi}$.

Chapter 3 is devoted to weak convergence theorems for the three kinds of martingales. An essential part of the proofs is the asymptotic tightness, which is established by using the maximal inequalities in Chapter 2. As we have mentioned above, the sufficient condition that we present is that the quadratic modulus is bounded in probability and that an entropy condition of the type (1.1.4) is satisfied. In particular, natural generalizations of Jain-Marcus' and Ossiander's central limit theorems are presented. The entropy condition for the cases (i) and (ii) above is analogous to that for L^2-bracketing, while the case (iii) is based on the metric entropy condition without bracketing. The results of this chapter are repeatedly applied to derive the asymptotic distribution of estimators in the subsequent chapters.

Some results concerning Theme 1 are given in Chapter 4. We consider the multiplicative intensity model for point processes with general marks, and derive the asymptotic normality and efficiency in ℓ^∞-spaces of a generalized Nelson-Aalen estimator. An interesting difference from the i.i.d. case, where the L^2-bracketing condition is optimal, is that an L^p-bracketing condition with $p \in [2, \infty]$ is sometimes preferable; this fact is valid also for other problems in this monograph, and the multiplicative intensity model provides a good illustration. We also study two non-linear models, of continuous semimartingales and of counting processes, both with time-dependent covariates.

Theme 2, the M-estimation procedure, is studied in Chapters 5 and Chapter 6 stressing non-standard rates of convergence. First, in Section 5.1, we present a general criterion for rate of convergence. A difference from known results in this area is that a kind of "twice differentiability" of criterion functions is generalized to a "p-times differentiability", that is, "$d(\theta, \theta_0)^2$" appearing in (1.1.1) is replaced by "$d(\theta, \theta_0)^p$". Sections 5.2 and 5.3 are concerned with some estimation problems of Euclidean parameters in the Gaussian white noise model and the multiplicative intensity model, respectively. Jump point estimation, among other things, is considered in those models.

Chapter 6 is devoted to the study of rate of convergence of non-parametric maximum likelihood estimators. The models considered there are the Gaussian white noise model, the multiplicative intensity model, a counting process model with non-linear covariates, and the diffusion type processes. The third model above contains a discussion about the Lexis diagram, which is important in the context of survival analysis.

The last chapter contains three independent topics. Since the setups of these problems are simple, this chapter, as well as Section 4.1, perhaps gives a guideline of the usage of the weak convergence theorems in Chapter 3. Except for Subsection 7.1.2, we do not use any results presented in Chapters 4, 5 and 6. In Section 7.1, we study the asymptotics of local random fields of kernel type estimators. The results are applied to the problem of estimating the mode of a density function; we derive the asymptotic behavior of an estimator defined as the argmax of kernel density estimator by using also the general theorem for M-estimators presented in Section 5.1. Section 7.2 is devoted to deriving the asymptotic behavior of log-likelihood ratio random fields in a general discrete-time statistical experiment with abstract parameters. An application to Markov chains is also discussed. In Section 7.3, we study a testing problem for a non-parametric regression model with dependent noise.

1.2 General Notations and Remarks

(1) $\mathbb{R} = (-\infty, \infty)$; $\mathbb{R}_+ = [0, \infty)$; $\mathbb{N} = \{1, 2, ...\}$; $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$; $\mathbb{Z} = \{\text{integers}\}$; $\mathbb{Q} = \{\text{rational numbers}\}$.

(2) We denote by $|\cdot|$ the Euclidean norm, even in the multi-dimensional case. We denote by $\text{Leb}(B)$ the Lebesgue measure of a Borel measurable subset B of a Euclidean space.

(3) The inequality "$x \lesssim y$" ($x, y \in [0, \infty)$) means that there exists a universal constant $C > 0$ such that $x \leq Cy$.

(4) $\ell^\infty(\Psi)$ denotes the set of all bounded functions defined on a set Ψ. We denote by $\|\cdot\|_\infty$ the supremum-norm on $\ell^\infty(\Psi)$.

(5) For every $p \in [1, \infty]$, we denote by $\mathcal{L}^p(E, \mathcal{E}, \lambda)$ the set of all p-integrable functions defined on a measure space $(E, \mathcal{E}, \lambda)$ (when $p = \infty$, it means the set of all λ-essentially bounded functions), and by $L^p(E, \mathcal{E}, \lambda)$ the equivalent classes of elements of $\mathcal{L}^p(E, \mathcal{E}, \lambda)$. These are often abbreviated to $\mathcal{L}^p$ and L^p, respectively. This kind of notational abbreviations of spaces are given section-wisely.

(6) $\text{Card}(\Psi)$ denotes the number of the elements of a set Ψ, allowing ∞.

(7) $\text{Diam}(\Psi, \rho)$ denotes the diameter of a set Ψ with respect to a semimetric ρ.

(8) When a semimetric space $(\mathcal{S}, \rho)$ is given, we denote by $B_{(\mathcal{S},\rho)}(x; \varepsilon)$ the closed ball with center $x \in \mathcal{S}$ and ρ-radius $\varepsilon > 0$; when there is no danger of confusion, it is

sometimes denoted by $B_\rho(x;\varepsilon)$ or even by $B(x;\varepsilon)$. This notational abbreviation is also given section-wisely.

(9) When a semimetric space $(\mathcal{S},\rho)$ and a subset Ψ of $\mathcal{S}$ are given, we denote by $N(\Psi,\rho;\varepsilon)$ the minimum number of closed balls with ρ-radius $\varepsilon>0$ which cover Ψ, allowing ∞. The centers of the balls need not belong to Ψ.

(10) Let $\mathcal{S}$ be a linear space of $\mathbb{R}$-valued functions ψ defined on a set, and let a seminorm $\|\cdot\|$ on $\mathcal{S}$ be given. For a given pair $l,u\in\mathcal{S}$, we denote $[l,u]=\{\psi\in\mathcal{S}:l\leq\psi\leq u\}$. Such $[l,u]$ is called a $(\|\cdot\|,\varepsilon)$-bracket in $\mathcal{S}$ if $\|u-l\|\leq\varepsilon$. For a given class $\Psi\subset\mathcal{S}$, the bracketing number $N_{[]}(\Psi,\rho;\varepsilon)$ is the minimum number of $(\|\cdot\|,\varepsilon)$-brackets which cover Ψ; that is, the smallest $N\in\mathbb{N}\cup\{\infty\}$ such that: there exists $l^k,u^k\in\mathcal{S}$, $k=1,...,N$, such that $\Psi\subset\bigcup_{k=1}^N[l^k,u^k]$ and that $\|u^k-l^k\|\leq\varepsilon$ for all k.

(11) Let $\alpha,H>0$ be given, and denote by $\underline{\alpha}$ the greatest integer strictly smaller than α. Let a bounded, convex subset E in $\mathbb{R}^d$ with nonempty interior E^i be given. We denote by $C_H^\alpha(E)$ the set of all continuous functions $f:E\to\mathbb{R}$ such that $\|f\|_\alpha\leq H$, where

$$\|f\|_\alpha=\max_{k.\leq\underline{\alpha}}\sup_{x\in E^i}|D^kf(x)|+\max_{k.=\underline{\alpha}}\sup_{\substack{x,y\in E^i\\x\neq y}}\frac{|D^kf(x)-D^kf(y)|}{|x-y|^{\alpha-\underline{\alpha}}}$$

with the notations $k.=\sum_{i=1}^d k_i$ and

$$D^k=\frac{\partial^{k.}}{\partial x_1^{k_1}\cdots\partial x_d^{k_d}}$$

for every vector $k=(k_1,...,k_d)$ of d non-negative integers. It is well-known that there exists a constant $K>0$ depending only on α and d such that

$$\log N(C_H^\alpha(E),\|\cdot\|_\infty;\varepsilon)\leq K\cdot\mathrm{Leb}(E_+)\left(\frac{H}{\varepsilon}\right)^{d/\alpha}\qquad\forall\varepsilon>0,$$

where $E_+=\{x:|x-E|<1\}$ (see, e.g., Theorem 2.7.1 of van der Vaart and Wellner (1996)).

(12) A *random semimetric* ϱ on a set T is a collection $\{\varrho(\omega;\cdot,\cdot):\omega\in\Omega\}$ of semimetric on T indexed by a probability space $(\Omega,\mathcal{F},P)$, although we do not require any measurability. We often denote a random semimetric by ϱ and a (non-random) semimetric by ρ.

(13) The words *increasing* and *decreasing* mean "non-decreasing" and "non-increasing", respectively (the situation where we should use the words *strictly increasing* or *strictly decreasing* does not appear in the monograph).

(14) We follow the standard definitions and notations of the martingale theory, which can be found in the book by Jacod and Shiryaev (1987).

(15) We refer to Part 1 of van der Vaart and Wellner (1996) for the weak convergence theory which does not require the measurability of random sequences. In particular, see their Chapter 1.2 for the definitions of the notations E^*, E_*, P^* and P_* that mean the outer integral, inner integral, outer probability and inner probability, respectively. Let $(\mathcal{X}, d)$ be a metric space, and for every $n \in \mathbb{N}$ let X^n be a mapping from a probability space $(\Omega^n, \mathcal{F}^n, P^n)$ to $\mathcal{X}$. We denote by "$X^n \overset{P^n}{\Longrightarrow} X$ in $\mathcal{X}$" the weak convergence of X^n to a tight, Borel measurable random element X taking values in $\mathcal{X}$; by "$X^n \overset{P^{n*}}{\longrightarrow} c$" the convergence in P^{n*}-probability to a non-random element c of $\mathcal{X}$; by "$X^n \overset{P^n}{\longrightarrow} c$" the convergence in P^n-probability (in this case X^n is assumed to be Borel measurable).

Chapter 2

Maximal Inequalities

2.1 Preliminaries

This chapter is devoted to getting some bounds for expectation of supremum of martingales up to a universal constant; throughout we use the notation "$\lesssim$" given in (3) of Section 1.2. The present section prepares two things, namely, quotation of two known inequalities which are used in Sections 2.2, 2.3 and 2.4, and introduction of two definitions which are necessary to formulate a quantity "quadratic modulus" in Sections 2.2 and 2.3. Thus, readers who are interested only in continuous local martingales studied in Section 2.4 may skip the latter.

First, let us state two lemmas which are well-known. The first one is the Bernstein inequality for martingales with bounded jumps; see e.g. Section 4.13 of Liptser and Shiryaev (1989) for the proof. The second one, which is used in connection with the former, is an adaptation of Lemma 2.2.10 of van der Vaart and Wellner (1996).

Lemma 2.1.1 *Let $t \rightsquigarrow X_t$ be an $\mathbb{R}$-valued, locally square-integrable martingale such that $X_0 = 0$ and that $|\Delta X| \leq a$ for a constant $a \geq 0$, and τ a bounded stopping time. Then, it holds that for every $\Gamma > 0$*

$$P\left(\sup_{t\in[0,\tau]} |X_t| > \varepsilon,\ \langle X, X\rangle_\tau \leq \Gamma\right) \leq 2\exp\left(-\frac{\varepsilon^2}{2(a\varepsilon + \Gamma)}\right) \qquad \forall \varepsilon > 0.$$

Lemma 2.1.2 *Let $N \in \mathbb{N}$ and let $X_1, ..., X_N$ be arbitrary $\mathbb{R}$-valued random variables. Assume that for a measurable set B and some constants $a \geq 0$ and $\Gamma > 0$*

$$P(|X_i| > \varepsilon,\ B) \leq 2\exp\left(-\frac{\varepsilon^2}{2(a\varepsilon + \Gamma)}\right) \qquad \forall \varepsilon > 0,\ \forall i = 1, ..., N.$$

Then, it holds that

$$E \max_{1\leq i\leq N} |X_i| 1_B \lesssim a\log(1+N) + \sqrt{\Gamma\log(1+N)}.$$

Combining these inequalities, we can easily get the following.

Corollary 2.1.3 *Let $N \in \mathbb{N}$. Let $t \rightsquigarrow X_t = (X_t^1, ..., X_t^N)$ be an $\mathbb{R}^N$-valued, locally square-integrable martingale such that $X_0^i = 0$ and that $|\Delta X^i| \leq a$ for a constant $a \geq 0$, and let τ be a finite stopping time. Then, for any constant K satisfying*

$$a\sqrt{\log(1+N)} \leq K,$$

it holds that

$$E \sup_{t\in[0,\tau]} \max_{1\leq i,j\leq N} |X_t^i - X_t^j| 1_B \lesssim K\sqrt{\log(1+N)}$$

where

$$B = \left\{ \max_{1\leq i,j\leq N} \sqrt{\langle X^i - X^j, X^i - X^j \rangle_\tau} \leq K \right\}.$$

The purpose of this chapter is to study what happens in the case of "$N = \infty$". We consider this problem for three kinds of martingales in Sections 2.2, 2.3 and 2.4, respectively.

Next, let us give two definitions for Sections 2.2 and 2.3.

Definition 2.1.4 *Let $(\mathcal{X}, \mathcal{A}, \lambda)$ be a σ-finite measure space. For a given mapping $Z : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$, we denote by $[Z]_{\mathcal{A},\lambda}$ any $\mathcal{A}$-measurable function $U : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ such that:*
(i) *$U \geq Z$ holds identically;*
(ii) *$\tilde{U} \geq U$ holds λ-almost everywhere, for every $\mathcal{A}$-measurable function $\tilde{U}$ such that $\tilde{U} \geq Z$ holds λ-almost everywhere.*

The existence of such a random variable $[Z]_{\mathcal{A},\lambda}$ and its uniqueness up to a λ-negligible set can be shown by using Lemma 1.2.1 of van der Vaart and Wellner (1996).

Definition 2.1.5 *Let Ψ be an arbitrary set such that* Card$(\Psi) = \infty$.

$\Pi = \{\Pi(\varepsilon)\}_{\varepsilon\in(0,\Delta_\Pi]}$, *where $\Delta_\Pi \in (0,\infty) \cap \mathbb{Q}$, is called a Decreasing series of Finite Partitions (DFP) of Ψ if it satisfies the following* (i), (ii) *and* (iii)*:*
(i) *each $\Pi(\varepsilon) = \{\Psi(\varepsilon; k) : 1 \leq k \leq N_\Pi(\varepsilon)\}$ is a finite partition of Ψ, that is, $\Psi = \bigcup_{k=1}^{N_\Pi(\varepsilon)} \Psi(\varepsilon; k)$ and $\Psi(\varepsilon; k_1) \cap \Psi(\varepsilon; k_2) = \emptyset$ whenever $k_1 \neq k_2$;*
(ii) *$N_\Pi(\Delta_\Pi) = 1$ and $\lim_{\varepsilon\downarrow 0} N_\Pi(\varepsilon) = \infty$;*
(iii) *$N_\Pi(\varepsilon) \geq N_\Pi(\varepsilon')$ whenever $\varepsilon \leq \varepsilon'$.*

$\Pi = \{\Pi(\varepsilon)\}_{\varepsilon\in(0,\Delta_\Pi]}$, *where $\Delta_\Pi \in (0,\infty) \cap \mathbb{Q}$, is called a Nested series of Finite Partitions (NFP) of Ψ if it satisfies the above* (i) *and* (ii) *and the following* (iii′)*:*
(iii′) *$\Pi(\varepsilon)$ is a refinement of $\Pi(\varepsilon')$ whenever $\varepsilon \leq \varepsilon'$.*

The ε-entropy $H_\Pi(\varepsilon)$ and the modified ε-entropy $\tilde{H}_\Pi(\varepsilon)$ of a DFP Π are defined by:

$$\begin{aligned} H_\Pi(\varepsilon) &= \sqrt{\log N_\Pi(\varepsilon)}; \\ \tilde{H}_\Pi(\varepsilon) &= \sqrt{\log(1+N_\Pi(\varepsilon))}. \end{aligned}$$

Notice that any NFP is a DFP. Although the converse is not true, we can sometimes construct a new NFP from a given DFP, due to Lemma 2.2.2 given later, without loss of generality for our purpose. Notice also that for any DFP Π

$$\int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon \le \delta\sqrt{\log 2} + \int_0^\delta H_\Pi(\varepsilon)d\varepsilon \qquad \forall \delta \in (0, \Delta_\Pi]$$

if the integral on the right hand side is finite; in fact, it holds that

$$\sqrt{\log(1+N)} \le \sqrt{\log 2N} \le \sqrt{\log 2} + \sqrt{\log N} \qquad \forall N \ge 1. \tag{2.1.1}$$

2.2 Multivariate Point Processes

Let $(E, \mathcal{E})$ be a Blackwell space, and let $\mathbf{B} = (\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}, P)$ be a stochastic basis. We put $\tilde{\Omega} = \Omega \times \mathbb{R}_+ \times E$ and $\tilde{\mathcal{P}} = \mathcal{P} \otimes \mathcal{E}$, where $\mathcal{P}$ is the predictable σ-field on $\Omega \times \mathbb{R}_+$. For a given predictable function W on $\tilde{\Omega}$ and a given random measure μ on $\mathbb{R}_+ \times E$, we denote by $W * \mu$ the integral process defined as the path-wise Lebesgue-Stieltjes integral: for every $t \in \mathbb{R}_+$,

$$W * \mu_t(\omega) = \begin{cases} \int_{[0,t]\times E} W(\omega, s, x)\mu(\omega; ds, dx) & \text{if } \int_{[0,t]\times E} |W(\omega, s, x)|\mu(\omega; ds, dx) < \infty, \\ \infty & \text{otherwise.} \end{cases}$$

(See Section II.1a of Jacod and Shiryaev (1987) for the detail).

Let μ be an E-valued multivariate point process. Let ν be a "good" version of the predictable compensator of μ (thus $\nu(\omega; \{t\} \times E) \le 1$). We introduce the Doléans measure M_ν^P on $(\tilde{\Omega}, \tilde{\mathcal{P}})$, which is $\tilde{\mathcal{P}}$-σ-finite, given by

$$M_\nu^P(d\omega, dt, dx) = P(d\omega)\nu(\omega; dt, dx)$$

(See Section II.1b, Definition III.1.23 and III.3.15 of Jacod and Shiryaev (1987)).

Let $\mathcal{W} = \{W^\psi : \psi \in \Psi\}$ be a family of predictable functions on $\tilde{\Omega}$ indexed by an arbitrary set Ψ. We give a definition, using the notation of Definition 2.1.4, which plays the key role in our context.

Definition 2.2.1 *The predictable envelope $\overline{W}$ of $\mathcal{W} = \{W^\psi : \psi \in \Psi\}$ is defined by*

$$\overline{W} = \left[\sup_{\psi\in\Psi} |W^\psi|\right]_{\tilde{\mathcal{P}}, M_\nu^P}.$$

For a given DFP Π of Ψ, the quadratic Π-modulus $||\mathcal{W}||_\Pi$ of $\mathcal{W} = \{W^\psi : \psi \in \Psi\}$ is defined as the $[0,\infty]$-valued increasing process $t \rightsquigarrow ||\mathcal{W}||_{\Pi,t}$ given by

$$||\mathcal{W}||_{\Pi,t} = \sup_{\varepsilon\in(0,\Delta_\Pi]\cap\mathbb{Q}} \max_{1\le k\le N_\Pi(\varepsilon)} \frac{\sqrt{|W(\Psi(\varepsilon;k))|^2 * \nu_t}}{\varepsilon} \qquad \forall t \in \mathbb{R}_+,$$

where

$$W(\Psi') = \left[\sup_{\psi,\phi\in\Psi'} |W^\psi - W^\phi|\right]_{\tilde{\mathcal{P}},M_\nu^P} \qquad \forall \Psi' \subset \Psi. \tag{2.2.1}$$

For a given DFP Π of Ψ and a given constant $a > 0$, the exponential (Π, a)-modulus $||\mathcal{W}||_\Pi^{\mathcal{E}_a}$ of $\mathcal{W} = \{W^\psi : \psi \in \Psi\}$ is defined as the $[0,\infty]$-valued increasing process $t \rightsquigarrow ||\mathcal{W}||_{\Pi,t}^{\mathcal{E}_a}$ given by

$$||\mathcal{W}||_{\Pi,t}^{\mathcal{E}_a} = \sup_{\varepsilon\in(0,\Delta_\Pi]\cap\mathbb{Q}} \max_{1\le k\le N_\Pi(\varepsilon)} \frac{\sqrt{(\mathcal{E}_a(W(\Psi(\varepsilon;k)))) * \nu_t}}{\varepsilon} \qquad \forall t \in \mathbb{R}_+,$$

where

$$\mathcal{E}_a(x) = \begin{cases} 2a^2\left(\exp(a^{-1}x) - 1 - a^{-1}x\right) & \forall x \in [0,\infty), \\ \infty & x = \infty, \end{cases} \tag{2.2.2}$$

and $W(\Psi')$ is defined by (2.2.1) for every $\Psi' \subset \Psi$.

It is clear that there exist some increasing versions of $t \rightsquigarrow ||\mathcal{W}||_{\Pi,t}$ and $t \rightsquigarrow ||\mathcal{W}||_{\Pi,t}^{\mathcal{E}_a}$ uniquely up to a P-negligible set, respectively. It holds that $||\mathcal{W}||_{\Pi,t} \le ||\mathcal{W}||_{\Pi,t}^{\mathcal{E}_a}$ almost surely, since $|x|^2 \le \mathcal{E}_a(x)$. Notice also that all of $\overline{W}$, $||\mathcal{W}||_\Pi$ and $||\mathcal{W}||_\Pi^{\mathcal{E}_a}$ depend on $\mathbf{F}$, P and ν, through $\tilde{\mathcal{P}}$ and M_ν^P. One may find that the exponential modulus above is based on the "Bernstein norm" (see 324 page of van der Vaart and Wellner (1996) for the i.i.d. case; a discrete martingale version is contained in van de Geer (1997)).

Lemma 2.2.2 *For any DFP Π of Ψ such that $\int_0^{\Delta_\Pi} H_\Pi(\varepsilon)d\varepsilon < \infty$, there exists a NFP Π' such that:*

$$\begin{aligned} \Delta_{\Pi'} &= \Delta_\Pi; \\ \int_0^{\Delta_{\Pi'}} H_{\Pi'}(\varepsilon)d\varepsilon &\le 4\int_0^{\Delta_\Pi} H_\Pi(\varepsilon)d\varepsilon; \\ \int_0^{\Delta_{\Pi'}} \tilde{H}_{\Pi'}(\varepsilon)d\varepsilon &\le 4\int_0^{\Delta_\Pi} \tilde{H}_\Pi(\varepsilon)d\varepsilon; \\ ||\mathcal{W}||_{\Pi',t} &\le ||\mathcal{W}||_{\Pi,t} \qquad \forall t \in \mathbb{R}_+. \end{aligned}$$

Proof. For every $\varepsilon \in (0, \Delta_\Pi]$, let us define

$$\Pi'(\varepsilon) = \bigvee_{i_0\le j\le i} \Pi(2^{-j}) \qquad \text{if} \quad \varepsilon \in [2^{-i}, 2^{-i+1}) \cap (0, \Delta_\Pi], \quad i \ge i_0,$$

where $i_0 = \min\{i \in \mathbb{Z} : 2^{-i} \leq \Delta_\Pi\}$. Then, the constructed $\Pi' = \{\Pi'(\varepsilon)\}_{\varepsilon\in(0,\Delta_{\Pi'}]}$ is a NFP such that $\Delta_{\Pi'} = \Delta_\Pi$. Further, since $N_{\Pi'}(2^{-i}) \leq \prod_{i_0\leq j\leq i} N_\Pi(2^{-j})$ we have

$$\begin{aligned}
\int_0^{\Delta_{\Pi'}} H_{\Pi'}(\varepsilon)d\varepsilon &\leq \sum_{i=i_0}^{\infty} 2^{-i} H_{\Pi'}(2^{-i}) \\
&\leq \sum_{i=i_0}^{\infty} 2^{-i} \sum_{j=i_0}^{i} H_\Pi(2^{-j}) \\
&= \sum_{j=i_0}^{\infty} H_\Pi(2^{-j}) \sum_{i=j}^{\infty} 2^{-i} \\
&= 4\sum_{j=i_0}^{\infty} 2^{-j-1} H_\Pi(2^{-j}) \\
&\leq 4\int_0^{\Delta_\Pi} H_\Pi(\varepsilon)d\varepsilon.
\end{aligned}$$

The same argument is valid also for the modified entropies. The last inequality is trivial from the construction of Π'. □

Supposing that there exists a version of $\overline{W}$ such that $\overline{W} * \nu_t(\omega) < \infty$ for all $\omega \in \Omega$ and $t \in \mathbb{R}_+$, let us define the random variables X_t^ψ and $X_t^{a,\psi}$ by

$$X_t^\psi = W^\psi * (\mu - \nu)_t \qquad \forall t \in \mathbb{R}_+ \ \forall \psi \in \Psi \tag{2.2.3}$$

and

$$X_t^{a,\psi} = W^\psi 1_{\{\overline{W}\leq a\}} * (\mu - \nu)_t \qquad \forall t \in \mathbb{R}_+ \ \forall \psi \in \Psi \ \forall a > 0, \tag{2.2.4}$$

respectively. Then, the process $t \rightsquigarrow X_t^\psi$ and the process $t \rightsquigarrow X_t^{a,\psi}$ is a locally square-integrable martingale on $\mathbf{B}$, both of which have finite variation. (see Lemma I.3.10 and Proposition II.1.28 of Jacod and Shiryaev (1987)). The following theorem gives some maximal inequalities for these processes in terms of $||\mathcal{W}||_\Pi$.

Theorem 2.2.3 *Let μ be an E-valued multivariate point process defined on a stochastic basis $\mathbf{B}$, and ν a "good" version of the predictable compensator of μ. Let $\mathcal{W} = \{W^\psi : \psi \in \Psi\}$ be a family of predictable functions on $\tilde{\Omega}$, indexed by an arbitrary set Ψ, such that $\overline{W} * \nu_t(\omega) < \infty$ for all $\omega \in \Omega$ and $t \in \mathbb{R}_+$ for a version of predictable envelope $\overline{W}$ of $\mathcal{W}$. Let τ be a finite stopping time. Then, we have the following* (i) *and* (ii).

(i) *It holds for any NFP Π of Ψ and any constants $\delta \in (0, \Delta_\Pi]$ and $K > 0$ that*

$$E^* \sup_{t\in[0,\tau]} \sup_{\substack{1\leq k\leq N_\Pi(\delta) \\ \psi,\phi\in\Psi(\delta;k)}} |X_t^{a,\psi} - X_t^{a,\phi}| 1_{\{||\mathcal{W}||_{\Pi,\tau}\leq K\}} \lesssim K \int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon,$$

where the random variables $X_t^{a,\psi}$ are defined by (2.2.4) with $a = a(\delta, K) = \delta K/\tilde{H}_\Pi(\delta/2)$.

(ii) *It holds for any DFP* Π *of* Ψ *and any constants* $K, L > 0$ *that*

$$E^* \sup_{t\in[0,\tau]} \sup_{\psi,\phi\in\Psi} |X_t^\psi - X_t^\phi| 1_{\{\|\mathcal{W}\|_{\Pi,\tau}\le K,\ |\overline{W}|^2 * \nu_\tau \le L\}} \lesssim K \int_0^{\Delta_\Pi} \tilde{H}_\Pi(\varepsilon)d\varepsilon + \frac{L}{\Delta_\Pi K},$$

where the random variables X_t^ψ *are defined by (2.2.3).*

Theorem 2.2.4 *Consider the same situation as Theorem 2.2.3. It holds for any random semimetric* ϱ *on* Ψ, *any NFP* Π *of* Ψ *and any constants* $\delta \in (0, \Delta_\Pi]$ *and* $K > 0$ *that*

$$E^* \sup_{t\in[0,\tau]} \sup_{\substack{\psi,\phi\in\Psi \\ \varrho(\psi,\phi)\le\delta}} |X_t^{a,\psi} - X_t^{a,\phi}| 1_{\{\|\mathcal{W}\|_{\Pi,\tau}\le K\}\cap B} \lesssim K \int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon,$$

where

$$B = \left\{ \sup_{\substack{\psi,\phi\in\Psi \\ \varrho(\psi,\phi)\le\delta}} \frac{\sqrt{\langle X^{a,\psi} - X^{a,\phi}, X^{a,\psi} - X^{a,\phi}\rangle_\tau}}{\delta} \le K \right\} \tag{2.2.5}$$

and the random variables $X_t^{a,\psi}$ *are defined by (2.2.4) with* $a = a(\delta, K) = \delta K/\tilde{H}_\Pi(\delta/2)$.

Proof of Theorem 2.2.3 (i). Fix any $\delta, K > 0$; we may assume $\delta \in \mathbb{Q}$ without loss of generality. For every integer $p \ge 0$, we set

$$a_p = 2^{-p+1}\delta K/\tilde{H}_\Pi(2^{-p-1}\delta).$$

Next, choosing an element $\psi_{p,k}$ from each partitioning set $\Psi(2^{-p}\delta; k)$ such that

$$\{\psi_{p,k} : 1 \le k \le N_\Pi(2^{-p}\delta)\} \subset \{\psi_{p+1,k} : 1 \le k \le N_\Pi(2^{-(p+1)}\delta)\},$$

we define for every $\psi \in \Psi$

$$\begin{cases} \pi_p\psi &= \psi_{p,k}, \\ \Pi_p\psi &= \Psi(2^{-p}\delta; k), \end{cases} \qquad \text{if} \quad \psi \in \Psi(2^{-p}\delta; k).$$

For every integer $q \ge 1$, we introduce the stopping time

$$\tau_q = \inf\left\{ t \in \mathbb{R}_+ : \nu([0,t]\times E) > \frac{|\tilde{H}_\Pi(2^{-q-2}\delta)|^2}{16} - 1 \right\} \wedge \tau.$$

Since $\nu([0,\tau]\times E) < \infty$ almost surely and $\lim_{\varepsilon\downarrow 0} N_\Pi(\varepsilon) = \infty$, it holds that $\tau_q \uparrow \tau$ as $q \to \infty$ almost surely. Hence it is enough to show that

$$E^* \sup_{t\in[0,\tau_q]} \sup_{\psi\in\Psi} |X_t^{a,\psi} - X_t^{a,\pi_0\psi}| 1_{\{\|\mathcal{W}\|_{\Pi,\tau}\le K\}} \lesssim K \int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon \qquad \forall q \ge 1, \tag{2.2.6}$$

where $a = a(\delta, K)$.

Let us now fix any integer $q \geq 1$, and denote $\tau = \tau_q$ [there should be no danger of confusion]. For every $p = 0, 1, ..., q$, we consider the predictable functions $W(\Pi_p\psi)$ on $\tilde{\Omega}$ defined by (2.2.1). Since $\Pi = \{\Pi(\varepsilon)\}_{\varepsilon \in (0,\Delta_\Pi]}$ is nested, it follows from Definition 2.1.4 that

$$2\overline{W} \geq W(\Pi_0\psi) \geq W(\Pi_1\psi) \geq \cdots \geq W(\Pi_q\psi), \tag{2.2.7}$$

M_ν^P-almost everywhere. Defining the values on the exceptional sets as zero, we can choose some versions such that the above inequality holds *identically*. Notice also that $W(\Pi_p\psi) = W(\Pi_p\phi)$ holds identically, whenever $\psi, \phi \in \Psi(2^{-q}\delta; k)$ for some k. Next, let us introduce the following predictable functions on $\tilde{\Omega}$:

$$\begin{aligned} A_p(\psi) &= 1_{\{W(\Pi_0\psi)\leq a_0,...,W(\Pi_{p-1}\psi)\leq a_{p-1},W(\Pi_p\psi)\leq a_p\}}, \quad p = 0, 1, ..., q; \\ B_p(\psi) &= 1_{\{W(\Pi_0\psi)\leq a_0,...,W(\Pi_{p-1}\psi)\leq a_{p-1},W(\Pi_p\psi)> a_p\}}, \quad p = 1, ..., q; \\ B_0(\psi) &= 1_{\{W(\Pi_0\psi)>a_0\}}. \end{aligned}$$

It is important that $A_p(\psi)$ and $B_p(\psi)$ depend on ψ only through the subsets $\Pi_0\psi, ..., \Pi_p\psi$ of Ψ. Next observe the identity

$$\begin{aligned} W^\psi - W^{\pi_0\psi} &= (W^\psi - W^{\pi_0\psi})B_0(\psi) \\ &\quad + \sum_{p=1}^{q}(W^\psi - W^{\pi_p\psi})B_p(\psi) \\ &\quad + (W^\psi - W^{\pi_q\psi})A_q(\psi) \\ &\quad + \sum_{p=1}^{q}(W^{\pi_p\psi} - W^{\pi_{p-1}\psi})A_{p-1}(\psi). \end{aligned} \tag{2.2.8}$$

Since $a_0 = 2a(\delta, K)$, we have $B_0(\psi) \leq 1_{\{\overline{W}>a(\delta,K)\}}$. Hence we obtain

$$\sup_{t\in[0,\tau]}\sup_{\psi\in\Psi}|X_t^{a(\delta,K),\psi} - X_t^{a(\delta,K),\pi_0\psi}| \leq (I_1) + (I_2) + (II_1) + (II_2) + (III),$$

where

$$\begin{aligned} (I_1) &= \sup_{\psi\in\Psi}\sum_{p=1}^{q} W(\Pi_p\psi)B_p(\psi) * \mu_\tau, \\ (I_2) &= \sup_{\psi\in\Psi}\sum_{p=1}^{q} W(\Pi_p\psi)B_p(\psi) * \nu_\tau, \\ (II_1) &= \sup_{\psi\in\Psi} W(\Pi_q\psi)A_q(\psi) * \mu_\tau, \\ (II_2) &= \sup_{\psi\in\Psi} W(\Pi_q\psi)A_q(\psi) * \nu_\tau, \\ (III) &= \sup_{t\in[0,\tau]}\sup_{\psi\in\Psi}\sum_{p=1}^{q}\left|(W^{\pi_p\psi} - W^{\pi_{p-1}\psi})A_{p-1}(\psi) * (\mu-\nu)_t\right|. \end{aligned}$$

Further, it holds that $(I_1) \le (I_1') + (I_2)$ where

$$(I_1') = \sup_{\psi\in\Psi} \sum_{p=1}^{q} |W(\Pi_p\psi)B_p(\psi) * (\mu-\nu)_\tau| \,,$$

and that $(II_1) \le (II_1') + (II_2)$ where

$$(II_1') = \sup_{\psi\in\Psi} |W(\Pi_q\psi)A_q(\psi) * (\mu-\nu)_\tau| \,.$$

Hereafter we will obtain bounds for the terms $(I_1'), (I_2), (II_1'), (II_2)$ and (III).

Estimation of (I_2) *and* (II_2). We can easily see that

$$\begin{aligned}(I_2) &\le \sup_{\psi\in\Psi} \sum_{p=1}^{q} \frac{1}{a_p} |W(\Pi_p\psi)|^2 B_p(\psi) * \nu_\tau \\ &\le \max_{1\le p\le q} \sup_{\psi\in\Psi} \frac{|W(\Pi_p\psi)|^2 B_p(\psi) * \nu_\tau}{|2^{-p}\delta|^2} \cdot \sum_{p=1}^{q} \frac{|2^{-p}\delta|^2}{a_p} \\ &\le K \sum_{p=1}^{q} 2^{-p-1}\delta \tilde{H}_\Pi(2^{-p-1}\delta) \qquad \text{on the set} \quad \{\|\mathcal{W}\|_{\Pi,\tau} \le K\}.\end{aligned}$$

On the other hand, it follows from Schwarz's inequality that

$$\begin{aligned}(II_2) &\le \sup_{\psi\in\Psi} \sqrt{|W(\Pi_q\psi)|^2 * \nu_\tau} \cdot \sqrt{\nu([0,\tau]\times E)} \\ &\le 2^{-q}\delta K \cdot \frac{\tilde{H}_\Pi(2^{-q-2}\delta)}{4}.\end{aligned}$$

Hence we have

$$\begin{aligned}E|(I_2) + (II_2)|1_{\{\|\mathcal{W}\|_{\Pi,\tau}\le K\}} &\le K \sum_{p=1}^{q+1} 2^{-p-1}\delta \tilde{H}_\Pi(2^{-p-1}\delta) \\ &\le 2K \int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon.\end{aligned}$$

Estimation of (I_1'), (II_1') *and* (III). Let us consider the term (I_1'). We will apply the Bernstein inequality (Lemma 2.1.1) to the processes

$$t \rightsquigarrow M_t = W(\Pi_p\psi)B_p(\psi) * (\mu-\nu)_t.$$

It follows from

$$0 \le W(\Pi_p\psi)B_p(\psi) \le W(\Pi_{p-1}\psi)B_p(\psi) \le a_{p-1}$$

that $|\Delta M| \le a_{p-1}$; it is also clear that

$$\begin{aligned}\langle M, M\rangle_\tau &\le |W(\Pi_p\psi)|^2 B_p(\psi) * \nu_\tau \\ &\le |2^{-p}\delta K|^2 \qquad \text{on the set} \quad \{\|\mathcal{W}\|_{\Pi,\tau} \le K\}.\end{aligned}$$

Thus we have

$$P\left(\sup_{t\in[0,\tau]}|W(\Pi_p\psi)B_p(\psi)*(\mu-\nu)_t|>\varepsilon,\ \|\mathcal{W}\|_{\Pi,\tau}\le K\right)$$
$$\le\ 2\exp\left(-\frac{\varepsilon^2}{2(a_{p-1}\varepsilon+|2^{-p}\delta K|^2)}\right)\qquad\forall\varepsilon>0.$$

Hence it follows from Lemma 2.1.2 that

$$\begin{aligned} &E\sup_{\psi\in\Psi}\sup_{t\in[0,\tau]}|W(\Pi_p\psi)B_p(\psi)*(\mu-\nu)_t|\,1_{\{\|\mathcal{W}\|_{\Pi,\tau}\le K\}}\\ &\lesssim\ a_{p-1}|\tilde{H}_\Pi(2^{-p}\delta)|^2+2^{-p}\delta K\tilde{H}_\Pi(2^{-p}\delta)\\ &\le\ 5K\cdot 2^{-p}\delta\tilde{H}_\Pi(2^{-p}\delta), \end{aligned}$$

where it should be noted that "$\sup_{\psi\in\Psi}$" appearing on the left hand side is actually "$\max_{1\le k\le N_\Pi(2^{-p})}$". We therefore obtain

$$\begin{aligned} E|(I_1')|1_{\{\|\mathcal{W}\|_{\Pi,\tau}\le K\}}\ &\lesssim\ 5K\sum_{p=1}^{q}2^{-p}\delta\tilde{H}_\Pi(2^{-p}\delta)\\ &\le\ 5K\int_0^\delta\tilde{H}_\Pi(\varepsilon)d\varepsilon. \end{aligned}$$

Exactly the same calculation as for (I_1') yields some bounds for (II_1') and (III), which lead to the inequality (2.2.6). □

Proof of Theorem 2.2.3 (ii). Due to Lemma 2.2.2, it suffices to show the assertion in the case of Π being a NFP. We extend the given NFP $\Pi=\{\Pi(\varepsilon)\}_{\varepsilon\in(0,\Delta_\Pi]}$ to $\Pi=\{\Pi(\varepsilon)\}_{\varepsilon\in(0,2\Delta_\Pi]}$ where $N_\Pi(\varepsilon)=1$ for all $\varepsilon\in[\Delta_\Pi,2\Delta_\Pi]$. In order to apply the assertion (i) with $\delta=2\Delta_\Pi$, we consider the truncated processes $X_t^{a,\psi}$ with $a=a(2\Delta_\Pi,K)=2\Delta_\Pi K/\sqrt{\log 2}$; notice that

$$\begin{aligned} \sup_{t\in[0,\tau]}\sup_{\psi,\phi\in\Psi}|X_t^\psi-X_t^\phi|\ &\le\ \sup_{t\in[0,\tau]}\sup_{\psi,\phi\in\Psi}|X_t^{a,\psi}-X_t^{a,\phi}|\\ &\quad+2\overline{W}1_{\{\overline{W}>a\}}*\mu_\tau+2\overline{W}1_{\{\overline{W}>a\}}*\nu_\tau. \end{aligned}$$

First we have

$$\begin{aligned} \overline{W}1_{\{\overline{W}>a\}}*\nu_\tau\ &\le\ \frac{|\overline{W}|^2*\nu_\tau}{a}\\ &\le\ \frac{L}{a}\qquad\text{on the set}\quad\{|\overline{W}|^2*\nu_\tau\le L\}. \end{aligned}$$

Next, let us introduce the predictable time

$$S=\inf\{t\in\mathbb{R}_+:|\overline{W}|^2*\nu_t>L\}.$$

Take an announcing sequence $\{S_n\}$ for S (see I.2.16 of Jacod and Shiryaev (1987)). Since $0 \le S_n < S$ almost surely on the set $\{S > 0\}$, it holds that $|\overline{W}|^2 * \nu_{S_n} \le L$ almost surely. Thus it follows also from Doob's stopping theorem that

$$\begin{aligned} E\overline{W}1_{\{\overline{W}>a\}} * \mu_{S_n \wedge T_m} &= E\overline{W}1_{\{\overline{W}>a\}} * \nu_{S_n \wedge T_m} \\ &\le \frac{E|\overline{W}|^2 * \nu_{S_n \wedge T_m}}{a} \le \frac{L}{a}, \end{aligned}$$

where $\{T_m\}$ is a localizing sequence for the local martingale $t \rightsquigarrow \overline{W}1_{\{\overline{W}>a\}} * (\mu - \nu)_t$. By letting $n, m \to \infty$, we obtain $E\overline{W}1_{\{\overline{W}>a\}} * \mu_S \le L/a$. The predictable time S appearing in this inequality can be replaced by τ on the set $\{|\overline{W}|^2 * \nu_\tau \le L\}$.

Hence it follows from the assertion (i) with $\delta = 2\Delta_\Pi$ that

$$\begin{aligned} &E^* \sup_{t\in[0,\tau]} \sup_{\psi,\phi\in\Psi} |X_t^\psi - X_t^\phi| 1_{\{\|\mathcal{W}\|_{\Pi,\tau}\le K,\ |\overline{W}|^2*\nu_\tau\le L\}} \\ &\lesssim K\int_0^{2\Delta_\Pi} \tilde{H}_\Pi(\varepsilon)d\varepsilon + 4\cdot\frac{L}{2\Delta_\Pi K/\sqrt{\log 2}} \\ &\le 2\left\{K\int_0^{\Delta_\Pi} \tilde{H}_\Pi(\varepsilon)d\varepsilon + \frac{L}{\Delta_\Pi K}\right\}. \end{aligned}$$

□

Proof of Theorem 2.2.4. We use the notations introduced in the first paragraph of the proof of Theorem 2.2.3 ($p = 0$ only). Notice that

$$|X_t^{a,\psi} - X_t^{a,\phi}| \le |X_t^{a,\psi} - X_t^{a,\pi_0\psi}| + |X_t^{a,\phi} - X_t^{a,\pi_0\phi}| + |X_t^{a,\pi_0\psi} - X_t^{a,\pi_0\phi}|$$

and thus

$$\sup_{\varrho(\psi,\phi)\le\delta} |X_t^{a,\psi} - X_t^{a,\phi}| \le 2 \sup_{\substack{1\le k\le N_\Pi(\delta) \\ \psi,\phi\in\Psi(\delta;k)}} |X_t^{a,\psi} - X_t^{a,\phi}| + \sup_{\varrho(\psi,\phi)\le\delta} |X_t^{a,\pi_0\psi} - X_t^{a,\pi_0\phi}|.$$

The second term on the right hand side equals

$$\max_{1\le k_1,k_2\le N_\Pi(\delta)} |X_t^{a,\pi_{0,k_1}} - X_t^{a,\pi_{0,k_2}}| 1_{A_{k_1,k_2}},$$

where

$$A_{k_1,k_2} = \left\{\omega\in\Omega: \quad \exists\psi,\phi\in\Psi \text{ s.t. } \begin{Bmatrix} \pi_0\psi = \pi_{0,k_1} \\ \pi_0\phi = \pi_{0,k_2} \end{Bmatrix} \text{ and } \varrho(\psi,\phi)(\omega)\le\delta\right\}.$$

Here, notice that for every $\psi, \phi \in \Psi$

$$\begin{aligned} \langle X^{a,\pi_0\psi} - X^{a,\pi_0\phi}, X^{a,\pi_0\psi} - X^{a,\pi_0\phi}\rangle_\tau \le\ & 3\langle X^{a,\pi_0\psi} - X^{a,\psi}, X^{a,\pi_0\psi} - X^{a,\psi}\rangle_\tau \\ &+3\langle X^{a,\pi_0\phi} - X^{a,\phi}, X^{a,\pi_0\phi} - X^{a,\phi}\rangle_\tau \\ &+3\langle X^{a,\psi} - X^{a,\phi}, X^{a,\psi} - X^{a,\phi}\rangle_\tau \end{aligned}$$

on the set $\{\|\mathcal{W}\|_{\Pi,\tau} \leq K\} \cap B$, where B is given by (2.2.5). Thus, for every $k_1, k_2 = 1, ..., N_\Pi(\delta)$ and every $\omega \in \{\|\mathcal{W}\|_{\Pi,\tau} \leq K\} \cap B \cap A_{k_1,k_2}$, by choosing some appropriate $\psi = \psi(\omega, k_1, k_2)$ and $\phi = \phi(\omega, k_1, k_2)$, we get

$$\langle X^{a,\pi_{0,k_1}} - X^{a,\pi_{0,k_2}}, X^{a,\pi_{0,k_1}} - X^{a,\pi_{0,k_2}} \rangle_\tau \leq 9\delta^2 K^2.$$

Thus it follows from Lemma 2.1.1 that

$$P\left(\sup_{t\in[0,\tau]} |X_t^{a,\pi_{0,k_1}} - X_t^{a,\pi_{0,k_2}}| 1_{A_{k_1,k_2}} > \varepsilon,\ \|\varrho\|_\Pi \leq K \right) \leq 2\exp\left(-\frac{\varepsilon^2}{2(2a\varepsilon + 9K^2\delta^2)} \right)$$

for every $\varepsilon > 0$. Hence, we obtain from Lemma 2.1.2 that

$$\begin{aligned} & E \sup_{t\in[0,\tau]} \sup_{\varrho(\psi,\phi)\leq\delta} |X_t^{a,\pi_0\psi} - X_t^{a,\pi_0\phi}| 1_{\{\|\mathcal{W}\|_{\Pi,\tau}\leq K\}\cap B} \\ \lesssim\ & 2a\log(1 + N_\Pi(\delta)^2) + 3\delta K\sqrt{\log(1 + N_\Pi(\delta)^2)} \\ \leq\ & 4K\delta\frac{\log(1 + N_\Pi(\delta))}{\sqrt{\log(1 + N_\Pi(\delta/2))}} + 3\sqrt{2}\delta K\sqrt{\log(1 + N_\Pi(\delta))}. \end{aligned}$$

This, together with (i) of Theorem 2.2.3, yields the assertion. □

So far we have been concerned with the truncated processes $X^{a,\psi} = W^\psi 1_{\{\overline{W}\leq a\}} * (\mu - \nu)$ (except for (ii) of Theorem 2.2.3). This means that the predictable functions of integrands should be uniformly bounded, and this assumption is sometimes too strong. However, as is explained in Chapter 3.4 of van der Vaart and Wellner, it can be replaced by a moment assumption of exponential order (see, in particular, their Lemmas 3.4.2 and 3.4.3 which are concerned with the i.i.d. case); the key tool for this purpose is the extended Bernstein inequality (e.g., their Lemma 2.2.11). In our situation, we can make use of a martingale version of the inequality given by van de Geer (1995b, Lemma 2.2); we will quote it as Lemma 2.2.6 below. Let us begin with giving a version of Theorem 2.2.3.

Lemma 2.2.5 *Consider the same situation as Theorem 2.2.3. It holds for any NFP* Π *of* Ψ *and any constants* $\delta \in (0, \Delta_\Pi]$ *and* $K > 0$ *that*

$$\begin{aligned} & E^* \sup_{t\in[0,\tau]} \sup_{\substack{1\leq k\leq N_\Pi(\delta) \\ \psi,\phi\in\Psi(\delta;k)}} |(W^\psi - W^\phi) * (\mu - \nu)_t| 1_{\{\|\mathcal{W}\|_{\Pi,\tau}\leq K\}} \\ \lesssim\ & K\int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon + E^* \sup_{t\in[0,\tau]} \max_{1\leq k\leq N_\Pi(\delta)} |W(\Psi(\delta;k)) * (\mu - \nu)_t| 1_{\{\|\mathcal{W}\|_{\Pi,\tau}\leq K\}} \end{aligned}$$

(recall the notation $W(\Psi')$ *defined by (2.2.1) for every* $\Psi' \subset \Psi$*).*

Proof. Recall the proof of (i) of Theorem 2.2.3; it suffices to consider the first term on the right hand side of (2.2.8). Notice that

$$(W^{\psi} - W^{\pi_0\psi})B_0(\psi) * (\mu - \nu)_t \leq W(\Pi_0\psi)B_0(\psi) * (\mu - \nu)_t + 2W(\Pi_0\psi)B_0(\psi) * \nu_t.$$

The second term on the right hand side is bounded by

$$\begin{aligned}\frac{2}{a_0}|W(\Pi_0\psi)|^2 * \nu_\tau &\leq \frac{\tilde{H}_\Pi(\delta/2)}{\delta K}(\delta K)^2 \qquad \text{on the set } \{\|\mathcal{W}\|_{\Pi,\tau} \leq K\} \\ &\leq 2K\int_0^{\delta/2} \tilde{H}_\Pi(\varepsilon)d\varepsilon.\end{aligned}$$

On the other hand, the first term is bounded by

$$|W(\Pi_0\psi) * (\mu - \nu)_t| + |W(\Pi_0\psi)1_{\{W(\Pi_0\psi)\leq a_0\}} * (\mu - \nu)_t|,$$

and it follows from Corollary 2.1.3 that

$$E^* \sup_{t\in[0,\tau]} \sup_{\psi\in\Psi} |W(\Pi_0\psi)1_{\{W(\Pi_0\psi)\leq a_0\}} * (\mu - \nu)_t|1_{\{\|\mathcal{W}\|_{\Pi,\tau}\leq K\}} \lesssim \delta K\tilde{H}_\Pi(\delta).$$

The assertion follows from these inequalities. □

Thus the problem is how to manage the second term on the right hand side of Lemma 2.2.5. As we announced above, this can be solved by means of the following lemma that is an easy consequence of Lemma 2.2 of van de Geer (1995b). From now on, we will assume that $\nu(\{t\} \times E) = 0$ for every $t \in \mathbb{R}_+$ for simplicity.

Lemma 2.2.6 *Let μ be an E-valued multivariate point process which has the predictable compensator ν such that $\nu(\omega; \{t\} \times E) = 0$ for all $\omega \in \Omega$ and $t \in \mathbb{R}_+$. Let W be a predictable function, and suppose that for a given constant $a > 0$ it holds that* $\exp(a^{-1}|W|) * \nu_t(\omega) < \infty$ *for all $\omega \in \Omega$ and $t \in \mathbb{R}_+$. Let τ be a finite stopping time. Then, it holds for every $\Gamma > 0$*

$$P\left(\sup_{t\in[0,\tau]} |W * (\mu - \nu)_t| > \varepsilon,\ (\mathcal{E}_a(|W|)) * \nu_\tau \leq \Gamma\right) \leq 2\exp\left(-\frac{\varepsilon^2}{2(a\varepsilon + \Gamma)}\right) \qquad \forall\varepsilon > 0.$$

Lemma 2.2.7 *Consider the same situation as Theorem 2.2.3. For a given constant $a > 0$, suppose also that for every $\omega \in \Omega$ it holds that $\nu(\omega; \{t\} \times E) = 0$ and that* $\exp(a^{-1}\overline{W}) * \nu_t(\omega) < \infty$ *for all $t \in [0, \tau(\omega)]$. Then, it holds that for any NFP Π of Ψ and any constants $\delta \in (0, \Delta_\Pi]$ and $K > 0$ that*

$$E^* \sup_{t\in[0,\tau]} \sup_{\substack{1\leq k\leq N_\Pi(\delta) \\ \psi,\phi\in\Psi(\delta;k)}} |(W^{\psi} - W^{\phi}) * (\mu - \nu)_t|1_{\{\|\mathcal{W}\|_{\Pi,\tau}\leq K\}\cap B}$$

$$\lesssim K \int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon + a\log(1 + N_\Pi(\delta)),$$

where

$$B = \left\{ \max_{1\le k\le N_\Pi(\delta)} \frac{\sqrt{(\mathcal{E}_{2a}(W(\Psi(\delta;k)))) * \nu_\tau}}{\delta} \le K \right\}.$$

Proof. Lemmas 2.2.6 and 2.1.2 yield that

$$E^* \sup_{t\in[0,\tau]} \sup_{\psi\in\Psi} |W(\Psi(\delta;k)) * (\mu - \nu)_t| 1_B \lesssim \delta K \tilde{H}_\Pi(\delta) + a\log(1 + N_\Pi(\delta)),$$

which implies the assertion. □

While the argument δ in the above lemma is arbitrary, it sometimes suffices to consider a specific range of δ. In the context of M-estimation studied in Chapter 6, we will use it in the following form that is reasonably weak and simple. It should be noted that one may get different versions by going back to Lemmas 2.2.6 and 2.2.7; even removing the assumption that $\nu(\omega;\{t\}\times E) = 0$ is also possible.

Theorem 2.2.8 *Consider the same situation as Theorem 2.2.3. For a given constant $a > 0$, suppose also that for every $\omega \in \Omega$ it holds that $\nu(\omega;\{t\}\times E) = 0$ and that* $\exp(a^{-1}\overline{W}) * \nu_t(\omega) < \infty$ *for all $t \in [0,\tau(\omega)]$. Let Π be an arbitrary NFP of Ψ.*

(i) *For any constants $\delta \in (0, \Delta_\Pi]$ and $K > 0$ satisfying*

$$a\int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon \le K\delta^2, \qquad (2.2.9)$$

it holds that

$$E^* \sup_{t\in[0,\tau]} \sup_{\substack{1\le k\le N_\Pi(\delta)\\ \psi,\phi\in\Psi(\delta;k)}} |X_t^\psi - X_t^\phi| 1_{\{\|\mathcal{W}\|_{\Pi,\tau}^{\mathcal{E}_{2a}} \le K\}} \lesssim K\int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon.$$

(ii) *If a given random semimetric ϱ on Ψ satisfies that*

$$\sqrt{(\mathcal{E}_{2a}(|W^\psi - W^\phi|)) * \nu_\tau} \le \varrho(\psi,\phi) \qquad \forall\psi,\phi\in\Psi \qquad P_*\text{-almost surely},$$

then, for any constants $\delta \in (0,\Delta_\Pi]$ and $K > 0$ satisfying (2.2.9), it holds that

$$E^* \sup_{t\in[0,\tau]} \sup_{\substack{\psi,\phi\in\Psi\\ \varrho(\psi,\phi)\le K\delta}} |X_t^\psi - X_t^\phi| 1_{\{\|\mathcal{W}\|_{\Pi,\tau}^{\mathcal{E}_{2a}} \le K\}\cap\{\|\varrho\|_\Pi \le K\}} \lesssim K\int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon,$$

where

$$\|\varrho\|_\Pi = \sup_{\varepsilon\in(0,\Delta_\Pi]} \max_{1\le k\le N_\Pi(\varepsilon)} \sup_{\psi,\phi\in\Psi(\varepsilon;k)} \frac{\varrho(\psi,\phi)}{\varepsilon}.$$

Proof. The second term of Lemma 2.2.7 is bounded by $a\delta^{-2}\left|\int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon\right|^2$, which is bounded by $K\int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon$ whenever $a\delta^{-2}\int_0^\delta \tilde{H}_\Pi(\varepsilon)d\varepsilon \leq K$. Thus we have obtained the first assertion.

For the proof of the second assertion, we use the notations introduced in the first paragraph of the proof of Theorem 2.2.3 ($p = 0$ only). The method of the proof is quite similar to that of Theorem 2.2.4. Notice that

$$|X_t^\psi - X_t^\phi| \leq |X_t^\psi - X_t^{\pi_0\psi}| + |X_t^\phi - X_t^{\pi_0\phi}| + |X_t^{\pi_0\psi} - X_t^{\pi_0\phi}|$$

and thus

$$\sup_{\varrho(\psi,\phi)\leq K\delta} |X_t^\psi - X_t^\phi| \leq 2 \sup_{\substack{1\leq k\leq N_\Pi(\delta)\\ \psi,\phi\in\Psi(\delta;k)}} |X_t^\psi - X_t^\phi| + \sup_{\varrho(\psi,\phi)\leq K\delta} |X_t^{\pi_0\psi} - X_t^{\pi_0\phi}|.$$

The second term on the right hand side equals

$$\max_{1\leq k_1,k_2\leq N_\Pi(\delta)} |X_t^{\pi_{0,k_1}} - X_t^{\pi_{0,k_2}}| 1_{A_{k_1,k_2}},$$

where

$$A_{k_1,k_2} = \left\{\omega\in\Omega: \quad \exists\psi,\phi\in\Psi \text{ s.t. } \left\{\begin{array}{c}\pi_0\psi = \pi_{0,k_1}\\ \pi_0\phi = \pi_{0,k_2}\end{array}\right\} \text{ and } \varrho(\psi,\phi)(\omega)\leq K\delta\right\}.$$

Here notice that for every $\psi,\phi\in\Psi$

$$\begin{aligned}\sqrt{(\mathcal{E}_{2a}(|W^{\pi_0\psi} - W^{\pi_0\phi}|)) * \nu_\tau} &\leq \varrho(\pi_0\psi,\pi_0\phi)\\ &\leq \varrho(\pi_0\psi,\psi) + \varrho(\pi_0\phi,\phi) + \varrho(\psi,\phi)\\ &\leq 2K\delta + \varrho(\psi,\phi)\end{aligned}$$

on the set $\{\|\varrho\|_\Pi \leq K\}$. Thus, for every $k_1, k_2 = 1, ..., N_\Pi(\delta)$, and every $\omega \in \{\|\varrho\|_\Pi \leq K\}\cap A_{k_1,k_2}$, by choosing some appropriate $\psi = \psi(k_1,k_2,\omega)$ and $\phi = \phi(k_1,k_2,\omega)$, we can get

$$\sqrt{(\mathcal{E}_{2a}(|W^{\pi_{0,k_1}} - W^{\pi_{0,k_2}}|)) * \nu_\tau} \leq 3K\delta.$$

Thus Lemma 2.2.6 yields that

$$P\left(\sup_{t\in[0,\tau]} |X_t^{\pi_{0,k_1}} - X_t^{\pi_{0,k_2}}| 1_{A_{k_1,k_2}} > \varepsilon,\ \|\varrho\|_\Pi \leq K\right) \leq 2\exp\left(-\frac{\varepsilon^2}{2(2a\varepsilon + 9K^2\delta^2)}\right)$$

for every $\varepsilon > 0$. Hence it follows from Corollary 2.1.2 that

$$\begin{aligned}&E^* \sup_{t\in[0,\tau]} \sup_{\varrho(\psi,\phi)\leq K\delta} |X_t^{\pi_0\psi} - X_t^{\pi_0\phi}| 1_{\{\|\varrho\|_\Pi\leq K\}}\\ &\lesssim 2a\log(1+N_\Pi(\delta)^2) + 3K\delta\sqrt{\log(1+N_\Pi(\delta)^2)}\\ &\leq 4a\log(1+N_\Pi(\delta)) + 3\sqrt{2}K\delta\sqrt{\log(1+N_\Pi(\delta))}.\end{aligned}$$

This, together with the first assertion of the theorem, yields the second. □

2.3 Martingale Difference Arrays

Let a discrete-time stochastic basis $\mathbf{B} = (\Omega, \mathcal{F}, \mathbf{F}, P)$ be given, where $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathbf{F} = \{\mathcal{F}_i\}_{i \in \mathbb{N}_0}$ is an increasing sequence of sub-σ-fields of $\mathcal{F}$ indexed by $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$. Let Ψ be an arbitrary set.

Definition 2.3.1 $\{\xi_i\}_{i \in \mathbb{N}} = \{(\xi_i^\psi | \psi \in \Psi)\}_{i \in \mathbb{N}}$ *is called an* $\ell^\infty(\Psi)$*-valued martingale difference array on* $\mathbf{B}$ *if:*
(i) ξ_i *is a mapping from* Ω *to* $\ell^\infty(\Psi)$ *for every* $i \in \mathbb{N}$*;*
(ii) $\{\xi_i^\psi\}_{i \in \mathbb{N}}$ *is an* $\mathbb{R}$*-valued martingale difference array on* $\mathbf{B}$ *for every* $\psi \in \Psi$*.*

It is required in (ii) that ξ_i^ψ is $\mathcal{F}_i$-measurable and $E_{i-1}\xi_i^\psi = 0$ almost surely, for every $\psi \in \Psi$, where E_{i-1} denotes the $\mathcal{F}_{i-1}$-conditional expectation; the exceptional sets may depend on ψ. Notice also that we do not require any measurability of the $\ell^\infty(\Psi)$-valued random element ξ_i.

Based on the notation of Definition 2.1.4, we make the following definition.

Definition 2.3.2 *The adapted envelope* $\{\overline{\xi}_i\}_{i \in \mathbb{N}}$ *of* $\{\xi_i\}_{i \in \mathbb{N}}$ *is defined by*

$$\overline{\xi}_i = \left[\sup_{\psi \in \Psi} |\xi_i^\psi|\right]_{\mathcal{F}_i, P} \qquad \forall i \in \mathbb{N}.$$

For a given DFP Π *of* Ψ*, the quadratic* Π*-modulus* $\|\xi\|_\Pi$ *of* $\{\xi_i\}_{i \in \mathbb{N}}$ *is defined as the* $\mathbb{R}_+ \cup \{\infty\}$*-valued increasing process* $\{\|\xi\|_{\Pi,i}\}_{i \in \mathbb{N}}$ *given by*

$$\|\xi\|_{\Pi,i} = \sup_{\varepsilon \in (0, \Delta_\Pi] \cap \mathbb{Q}} \max_{1 \le k \le N_\Pi(\varepsilon)} \frac{\sqrt{\sum_{j=1}^i E_{j-1}|\xi_j(\Psi(\varepsilon; k))|^2}}{\varepsilon} \qquad \forall i \in \mathbb{N},$$

where

$$\xi_i(\Psi') = \left[\sup_{\psi, \phi \in \Psi'} |\xi_i^\psi - \xi_i^\phi|\right]_{\mathcal{F}_i, P} \qquad \forall i \in \mathbb{N} \quad \forall \Psi' \subset \Psi. \tag{2.3.1}$$

For a given DFP Π *of* Ψ *and a given constant* $a > 0$*, the exponential* (Π, a)*-modulus* $\|\xi\|_\Pi^{\mathcal{E}_a}$ *of* $\{\xi_i\}_{i \in \mathbb{N}}$ *is defined as the* $[0, \infty]$*-valued increasing process* $\{\|\xi\|_{\Pi,i}^{\mathcal{E}_a}\}_{i \in \mathbb{N}}$ *given by*

$$\|\xi\|_{\Pi,i}^{\mathcal{E}_a} = \sup_{\varepsilon \in (0, \Delta_\Pi] \cap \mathbb{Q}} \max_{1 \le k \le N_\Pi(\varepsilon)} \frac{\sqrt{\sum_{j=1}^i E_{j-1}\mathcal{E}_a(\xi_j(\Psi(\varepsilon; k)))}}{\varepsilon} \qquad \forall i \in \mathbb{N},$$

where $\mathcal{E}_a(x)$ *is defined by (2.2.2) and* $\xi_i(\Psi')$ *is defined by (2.3.1) for every* $\Psi' \subset \Psi$*.*

Theorem 2.3.3 *Let* $\{\xi_i\}_{i \in \mathbb{N}}$ *be an* $\ell^\infty(\Psi)$*-valued martingale difference array, and let* σ *be a finite stopping time, both of which are defined on a discrete-time stochastic basis* $\mathbf{B}$*. Then, we have the following* (i) *and* (ii)*.*

(i) *For any NFP* Π *of* Ψ *and any constants* $\delta \in (0, \Delta_\Pi]$ *and* $K > 0$,

$$E^* \max_{1 \le m \le \sigma} \sup_{\substack{1 \le k \le N_\Pi(\delta) \\ \psi, \phi \in \Psi(\delta;k)}} \left| \sum_{i=1}^m (\xi_i^{a,\psi} - \xi_i^{a,\phi}) \right| 1_{\{\|\xi\|_{\Pi,\sigma} \le K\}} \lesssim K \int_0^\delta \tilde{H}_\Pi(\varepsilon) d\varepsilon,$$

where $\xi_i^{a,\psi} = \xi_i^\psi 1_{\{\bar{\xi}_i \le a\}}$ *with* $a = a(\delta, K) = \delta K / \tilde{H}_\Pi(\delta/2)$.

(ii) *For any DFP* Π *of* Ψ *and any constants* $K, L > 0$,

$$E^* \max_{1 \le m \le \sigma} \sup_{\psi, \phi \in \Psi} \left| \sum_{i=1}^m (\xi_i^\psi - \xi_i^\phi) \right| 1_{\{\|\xi\|_{\Pi,\sigma} \le K,\ \sum_1^\sigma E_{i-1} |\bar{\xi}_i|^2 \le L\}} \lesssim K \int_0^{\Delta_\Pi} \tilde{H}_\Pi(\varepsilon) d\varepsilon + \frac{L}{\Delta_\Pi K}.$$

The result above is similar to Theorem 2.2.3, although the proof needs a careful discussion about the choice of versions of conditional expectations. It gives us the analogue of Theorem 2.2.4.

Theorem 2.3.4 *Consider the same situation as Theorem 2.3.3. It holds for any random semimetric* ρ *on* Ψ, *any NFP* Π *of* Ψ *and any constants* $\delta \in (0, \Delta_\Pi]$ *and* $K > 0$ *that*

$$E^* \max_{1 \le m \le \sigma} \sup_{\substack{1 \le k \le N_\Pi(\delta) \\ \psi, \phi \in \Psi(\delta;k)}} \left| \sum_{i=1}^m (\xi_i^{a,\psi} - \xi_i^{a,\phi}) \right| 1_{\{\|\xi\|_{\Pi,\sigma} \le K\} \cap B} \lesssim K \int_0^\delta \tilde{H}_\Pi(\varepsilon) d\varepsilon,$$

where

$$B = \left\{ \sup_{\substack{\psi, \phi \in \Psi \\ \varrho(\psi,\phi) \le \delta}} \frac{\sqrt{\sum_{i=1}^\sigma E_{i-1} |\xi_i^{a,\psi} - \xi_i^{a,\phi}|^2}}{\delta} \le K \right\} \qquad (2.3.2)$$

and where $\xi_i^{a,\psi} = \xi_i^\psi 1_{\{\bar{\xi}_i \le a\}}$ *with* $a = a(\delta, K) = \delta K / \tilde{H}_\Pi(\delta/2)$.

Proof of Theorem 2.3.3. Fix any $\delta, K > 0$, and define a_p, π_p and Π_p for every integer $p \ge 0$ in the same way as the first paragraph of the proof of Theorem 2.2.3 (i). For every integer $q \ge 1$ we introduce the finite stopping time

$$\sigma_q = \inf \left\{ i \in \mathbb{N} : i > \frac{|\tilde{H}_\Pi(2^{-q-2}\delta)|^2}{16} - 1 \right\} \wedge \sigma.$$

Then, we have $\sigma_q \uparrow \sigma$ as $q \to \infty$ almost surely. Hence it is enough for getting the assertion (i) to show that

$$E^* \max_{1 \le m \le \sigma_q} \sup_{\psi \in \Psi} \left| \sum_{i=1}^m (\xi_i^{a,\psi} - \xi_i^{a,\pi_0\psi}) \right| 1_{\{\|\xi\|_{\Pi,\sigma} \le K\}} \lesssim K \int_0^\delta \tilde{H}_\Pi(\varepsilon) d\varepsilon \qquad \forall q \ge 1, \qquad (2.3.3)$$

where $a = a(\delta, K)$.

Let us now fix any integer $q \geq 1$, and choose some appropriate versions of $\overline{\xi}_i$ and $\xi_i(\Pi_p\psi)$, $p = 0, 1, ..., q$ (recall the argument about (2.2.7)). We define:

$$\begin{aligned} A_{i,p}(\psi) &= 1_{\{\xi_i(\Pi_0\psi)\leq a_0,...,\xi_i(\Pi_{p-1}\psi)\leq a_{p-1},\xi_i(\Pi_p\psi)\leq a_p\}}, \quad p = 0, 1, ..., q; \\ B_{i,p}(\psi) &= 1_{\{\xi_i(\Pi_0\psi)\leq a_0,...,\xi_i(\Pi_{p-1}\psi)\leq a_{p-1},\xi_i(\Pi_p\psi)> a_p\}}, \quad p = 1, ..., q; \\ B_{i,0}(\psi) &= 1_{\{\xi_i(\Pi_0\psi)>a_0\}}. \end{aligned}$$

Next observe the identity

$$\begin{aligned} \xi_i^\psi - \xi_i^{\pi_0\psi} &= (\xi_i^\psi - \xi_i^{\pi_0\psi})B_{i,0}(\psi) \\ &\quad + \sum_{p=1}^q (\xi_i^\psi - \xi_i^{\pi_p\psi})B_{i,p}(\psi) \\ &\quad + (\xi_i^\psi - \xi_i^{\pi_q\psi})A_{i,q}(\psi) \\ &\quad + \sum_{p=1}^q (\xi_i^{\pi_p\psi} - \xi_i^{\pi_{p-1}\psi})A_{i,p-1}(\psi). \end{aligned}$$

Taking the $\mathcal{F}_{i-1}$-conditional expectations of all terms, we obtain

$$\begin{aligned} (2.3.4) \qquad (0 =)\ E_{i-1}\xi_i^\psi - E_{i-1}\xi_i^{\pi_0\psi} &= E_{i-1}(\xi_i^\psi - \xi_i^{\pi_0\psi})B_{i,0}(\psi) \\ &\quad + \sum_{p=1}^q E_{i-1}(\xi_i^\psi - \xi_i^{\pi_p\psi})B_{i,p}(\psi) \\ &\quad + E_{i-1}(\xi_i^\psi - \xi_i^{\pi_q\psi})A_{i,q}(\psi) \\ &\quad + \sum_{p=1}^q E_{i-1}(\xi_i^{\pi_p\psi} - \xi_i^{\pi_{p-1}\psi})A_{i,p-1}(\psi), \end{aligned}$$

almost surely. Further, it holds that

$$(2.3.5) \qquad |E_{i-1}(\xi_i^\psi - \xi_i^{\pi_0\psi})B_{i,0}(\psi)| \leq E_{i-1}\xi_i(\Pi_0\psi)B_{i,0}(\psi)$$

almost surely, and that

$$(2.3.6) \quad |E_{i-1}(\xi_i^\psi - \xi_i^{\pi_p\psi})B_{i,p}(\psi)| \leq E_{i-1}\xi_i(\Pi_p\psi)B_{i,p}(\psi) \leq a_{p-1}, \qquad p = 1, ..., q,$$

almost surely, and that

$$(2.3.7) \qquad |E_{i-1}(\xi_i^{\pi_p\psi} - \xi_i^{\pi_{p-1}\psi})A_{i,p-1}(\psi)| \leq a_{p-1}$$

almost surely. Here we choose versions of conditional expectations as follows: first choose some versions of the terms $E_{i-1}\xi_i(\Pi_p(\psi))B_{i,p}(\psi)$ of (2.3.5) and (2.3.6), which are non-negative and fulfill the second inequalities of (2.3.6), identically; next, on the exceptional sets of (2.3.4), (2.3.5), (2.3.6) and (2.3.7), we define the values of all other conditional expectations as zero. Then, the values of $E_{i-1}\xi_i(\Pi_p\psi)B_{i,p}(\psi)$ and $E_{i-1}(\xi_i^{\pi_p\psi} - \xi_i^{\pi_{p-1}\psi})A_{i,p-1}(\psi)$ depend on ψ only through $\Pi_0\psi, ..., \Pi_p\psi$ and $\pi_{p-1}\psi, \pi_p\psi$, respectively, while (2.3.4), (2.3.5), (2.3.6) and (2.3.7) hold *identically* for *all* $\psi \in \Psi$.

Since $a_0 = 2\delta K\tilde{H}_\Pi(\delta/2)$ we have $B_{i,0}(\psi)1_{\{\bar{\xi}_i \leq a(\delta,K)\}} = 0$. This implies that

$$\begin{aligned}
& E^*\left(\sup_{\psi\in\Psi}\left|E_{i-1}(\xi_i^\psi - \xi_i^{\pi_0\psi})B_{i,0}(\psi)\right| 1_{\{\bar{\xi}_i \leq a(\delta,K)\}}\right) \\
\leq\ & E\left(\sup_{\psi\in\Psi}|E_{i-1}\xi_i(\Pi_0\psi)B_{i,0}(\psi)|\, 1_{\{\bar{\xi}_i\leq a(\delta,K)\}}\right) \qquad \text{since (2.3.5) holds identically} \\
\leq\ & 2E\left(\bar{\xi}_i B_{i,0}(\psi)1_{\{\bar{\xi}_i\leq a(\delta,K)\}}\right) = 0,
\end{aligned}$$

and thus

$$\sup_{\psi\in\Psi}\left|E_{i-1}(\xi_i^\psi - \xi_i^{\pi_0\psi})B_{i,0}(\psi)\right| 1_{\{\bar{\xi}_i\leq a(\delta,K)\}} = 0$$

almost surely. Hence we obtain

$$\max_{1\leq m\leq\sigma_q}\sup_{\psi\in\Psi}\left|\sum_{i=1}^m(\xi_i^{a,\psi} - \xi_i^{a,\pi_0\psi})\right| \leq (I_1') + 2(I_2) + (II_1') + 2(II_2) + (III),$$

where

$$\begin{aligned}
(I_1') &= \sup_{\psi\in\Psi}\sum_{p=1}^q\left|\sum_{i=1}^{\sigma_q}\{\xi_i(\Pi_p\psi)B_{i,p}(\psi) - E_{i-1}\xi_i(\Pi_p\psi)B_{i,p}(\psi)\}\right|, \\
(I_2) &= \sup_{\psi\in\Psi}\sum_{p=1}^q\sum_{i=1}^{\sigma_q}E_{i-1}\xi_i(\Pi_p\psi)B_{i,p}(\psi), \\
(II_1') &= \sup_{\psi\in\Psi}\left|\sum_{i=1}^{\sigma_q}\{\xi_i(\Pi_q\psi)A_{i,q}(\psi) - E_{i-1}\xi_i(\Pi_q\psi)A_{i,q}(\psi)\}\right|, \\
(II_2) &= \sup_{\psi\in\Psi}\sum_{i=1}^{\sigma_p}E_{i-1}\xi_i(\Pi_q\psi)A_{i,q}(\psi), \\
(III) &= \max_{1\leq m\leq\sigma_q}\sup_{\psi\in\Psi}\sum_{p=1}^q\left|\sum_{i=1}^m\Big\{(\xi_i^{\pi_p\psi} - \xi_i^{\pi_{p-1}\psi})A_{i,p-1}(\psi)\right. \\
&\qquad \left. - \sum_{i=1}^m E_{i-1}(\xi_i^{\pi_p\psi} - \xi_i^{\pi_{p-1}\psi})A_{i,p-1}(\psi)\Big\}\right|,
\end{aligned}$$

almost surely. To get (2.3.3), we can deal with terms $(I_1'), (I_2), (II_1'), (II_2)$ and (III) exactly in the same way as those of the proof of Theorem 2.2.3 (i).

The assertion (ii) can be proved in the same way as that of Theorem 2.2.3, paying attention to the choice of conditional expectations; introduce a continuous-time stochastic basis and repeat the argument with an announcing sequence (see page 14 and I.2.43 of Jacod and Shiryaev (1987)). □

Proof of Theorem 2.3.4. The result follows from the same argument as Theorem 2.2.4. □

When a maximal inequality not for the truncated $\xi_i^{a,\psi}$'s but for the original ξ_i^{ψ}'s is needed, one may follow exactly the same discussion as Lemmas 2.2.5, 2.2.7 and Theorem 2.2.8, replacing Lemma 2.2.6 by the following version of the extended Bernstein inequality due to van de Geer (1995b).

Lemma 2.3.5 *Let $\xi = \{\xi_i\}_{i \in \mathbb{N}_0}$ be an $\mathbb{R}$-valued martingale difference array on a discrete stochastic basis* $\mathbf{B}$. *Suppose that for a given constant $a > 0$ it holds that $E\exp(a^{-1}|\xi_i|) < \infty$ for every $i \in \mathbb{N}$. Let σ be a finite stopping time on* $\mathbf{B}$. *Then, it holds for every $\Gamma > 0$*

$$P\left(\max_{1 \le m \le \sigma} \left| \sum_{i=1}^{m} \xi_i \right| > \varepsilon, \ \sum_{i=1}^{\sigma} E_{i-1}\mathcal{E}_a(|\xi_i|)) \le \Gamma \right) \le 2\exp\left(-\frac{\varepsilon^2}{2(a\varepsilon + \Gamma)} \right) \qquad \forall \varepsilon > 0.$$

We state here the analogue of Theorem 2.2.8 only.

Theorem 2.3.6 *Let $\{\xi_i\}_{i \in \mathbb{N}}$ be an $\ell^\infty(\Psi)$-valued martingale difference array and let σ be a finite stopping time both of which defined on a discrete-time stochastic basis* $\mathbf{B}$. *Suppose also that for a given constant $a > 0$ it holds that $E\exp(a^{-1}\overline{\xi}_i) < \infty$ for every $i \in \mathbb{N}$. Let Π be an arbitrary NFP of Ψ.*

(i) *For any constants $\delta \in (0, \Delta_\Pi]$ and $K > 0$ satisfying*

$$a \int_0^\delta \tilde{H}_\Pi(\varepsilon) d\varepsilon \le K\delta^2, \tag{2.3.8}$$

it holds that

$$E^* \max_{1 \le m \le \sigma} \sup_{\substack{\psi,\phi \in \Psi \\ \rho(\psi,\phi) \le \delta}} \left| \sum_{i=1}^{m} (\xi_i^\psi - \xi_i^\phi) \right| 1_{\{\|\xi\|_{\Pi,\sigma}^{\mathcal{E}_{2a}} \le K\}} \lesssim K \int_0^\delta \tilde{H}_\Pi(\varepsilon) d\varepsilon,$$

(ii) *If a given semimetric ϱ on Ψ satisfies that*

$$\sqrt{\sum_{i=1}^{\sigma} E_{i-1}\mathcal{E}_{2a}(|\xi_i^\psi - \xi_i^\phi|)} \le \varrho(\psi, \phi) \qquad \forall \psi, \phi \in \Psi \qquad P_*\text{-almost surely,}$$

then, for any constant $\delta \in (0, \Delta_\Pi]$ and $K > 0$ satisfying (2.3.8), it holds that

$$E^* \max_{1 \le m \le \sigma} \sup_{\substack{\psi,\phi \in \Psi \\ \varrho(\psi,\phi) \le K\delta}} \left| \sum_{i=1}^{m} (\xi_i^\psi - \xi_i^\phi) \right| 1_{\{\|\xi\|_{\Pi,\sigma}^{\mathcal{E}_{2a}} \le K\} \cap \{\|\varrho\|_\Pi \le K\}} \lesssim K \int_0^\delta \tilde{H}_\Pi(\varepsilon) d\varepsilon,$$

where

$$\|\varrho\|_\Pi = \sup_{\varepsilon \in (0, \Delta_\Pi]} \max_{1 \le k \le N_\Pi(\varepsilon)} \sup_{\psi,\phi \in \Psi(\varepsilon;k)} \frac{\varrho(\psi, \phi)}{\varepsilon}.$$

2.4 Continuous Local Martingales

Let $\mathbf{B} = (\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}, P)$ be a stochastic basis and (Ψ, ρ) a *proper* metric space. Let $X = (X^\psi | \psi \in \Psi)$ be a family of continuous local martingales defined on $\mathbf{B}$ indexed by Ψ. We introduce a kind of "quadratic modulus" which fits in this situation.

Definition 2.4.1 *A quadratic ρ-modulus $\|X\|_\rho$ of a family $X = (X^\psi|\psi \in \Psi)$ of continuous local martingales is defined as a $[0,\infty]$-valued stochastic process $t \rightsquigarrow \|X\|_{\rho,t}$ given by*

$$\|X\|_{\rho,t} = \sup_{\substack{\psi,\phi\in\Psi \\ \psi\neq\phi}} \frac{\sqrt{\langle X^\psi - X^\phi, X^\psi - X^\phi\rangle_t}}{\rho(\psi,\phi)} \qquad \forall t \in \mathbb{R}_+.$$

Since the set Ψ is not necessarily countable, the random element $\|X\|_{\rho,t}$ may not have any measurability. Moreover, although the predictable covariation $\langle X^\psi, X^\phi\rangle$ is uniquely determined up to a negligible set for every pair $\psi, \phi \in \Psi$, due to the same reason the quadratic ρ-modulus of X may not be unique even in the almost sure sense. However, we do not require its uniqueness because the assertion of the following theorem is valid for *any* choice of quadratic ρ-modulus of X.

Theorem 2.4.2 *Let (Ψ, ρ) be a totally bounded metric space. Let $X = (X^\psi|\psi \in \Psi)$ be a family of continuous local martingales indexed by Ψ such that $X_0^\psi = 0$, and τ a finite stopping time, both of which are defined on a stochastic basis $\mathbf{B}$.*

Then, for any choice of quadratic ρ-modulus $\|X\|_\rho$ of X, it holds that for every $\delta, K > 0$

$$\sup_{\substack{\Psi^*\subset\Psi \\ \text{countable}}} E^* \sup_{t\in[0,\tau]} \sup_{\substack{\psi,\phi\in\Psi^* \\ \rho(\psi,\phi)\leq\delta}} |X_t^\psi - X_t^\phi| 1_{\{\|X\|_{\rho,\tau}\leq K\}} \lesssim K \int_0^\delta \sqrt{\log(1 + N(\Psi, \rho; \varepsilon))} d\varepsilon,$$

provided the integral on the right hand side is finite (the first supremum is taken over all countable subsets Ψ^* of Ψ).

Proof. Fix any countable subset Ψ^* of Ψ. Let $\{\Psi^m\}_{m\in\mathbb{N}}$ be a sequence of finite subsets of Ψ^* such that $\Psi^m \uparrow \Psi^*$ as $m \to \infty$. For every $m \in \mathbb{N}$ and $p \in \mathbb{Z}$, let us denote by $q(m,p)$ the smallest integer such that $q(m,p) > p$ and that each of closed balls with centers in Ψ^m and ρ-radius $2 \cdot 2^{-q(m,p)}$ contains exactly one point in Ψ^m. Then it is clear that $\text{Card}(\Psi^m) \leq N(\Psi, \rho; 2^{-q(m,p)})$.

Next let us introduce some mappings $\pi_r^{m,p} : \Psi^m \to \Psi_r^{m,p}$, $p \leq r \leq q(m,p)$, defined by

$$\pi_r^{m,p} = \lambda_r^{m,p} \circ \lambda_{r+1}^{m,p} \circ \cdots \circ \lambda_{q(m,p)}^{m,p},$$

where the sets $\Psi_r^{m,p} \subset \Psi^m$ and the mappings $\lambda_r^{m,p} : \Psi^m \to \Psi_r^{m,p}$ should be specified in the following way. For $p \le r < q(m,p)$, choose $\Psi_r^{m,p}$ and define $\lambda_r^{m,p}$ which satisfy the following two conditions: (i) $\mathrm{Card}(\Psi_r^{m,p}) \le N(\Psi, \rho; 2^{-r})$; (ii) $\rho(\psi, \lambda_r^{m,p}(\psi)) \le 2 \cdot 2^{-r}$ for every $\psi \in \Psi^m$. For $r = q(m,p)$, put $\Psi_{q(m,p)}^{m,p} = \Psi^m$ and denote by $\lambda_{q(m,p)}^{m,p}$ the identity mapping on Ψ^m.

In term of the mappings $\pi_r^{m,p}$ which have been introduced, we consider the *chaining* given as follows: for every $t \in \mathbb{R}_+$ and $\psi \in \Psi^*$

$$|X_t^\psi - X_t^\phi| \le (I) + (II)$$

where the terms on the right hand side are given by:

$$\begin{aligned}(I) &= \sum_{r=p+1}^{q(m,p)} |X_t^{\pi_r^{m,p}(\psi)} - X_t^{\pi_{r-1}^{m,p}(\psi)}| + \sum_{r=p+1}^{q(m,p)} |X_t^{\pi_r^{m,p}(\phi)} - X_t^{\pi_{r-1}^{m,p}(\phi)}|;\\ (II) &= |X_t^{\pi_p^{m,p}(\psi)} - X_t^{\pi_p^{m,p}(\phi)}|.\end{aligned}$$

First let us consider the term (*I*). It follows from Lemma 2.1.1 that for every $\varepsilon, T > 0$

$$P\left(\sup_{t\in[0,\tau\wedge T]} |X_t^{\pi_r^{m,p}(\psi)} - X_t^{\pi_{r-1}^{m,p}(\psi)}| > \varepsilon,\ \|X\|_{\rho,\tau} \le K\right) \le 2\exp\left(-\frac{\varepsilon^2}{2\cdot 2^{-2r}K^2}\right),$$

and by letting $T \to \infty$ we can replace "$\tau \wedge T$" by "τ" on the left hand side. Thus we obtain from Lemma 2.1.2 that

$$E \sup_{\psi\in\Psi^m} \sup_{t\in[0,\tau]} |X_t^{\pi_r^{m,p}(\psi)} - X_t^{\pi_{r-1}^{m,p}(\psi)}| 1_{\{\|X\|_{\rho,\tau}\le K\}} \lesssim 2^{-r}K\sqrt{\log(1 + N(\Psi, \rho; 2^{-r}))}.$$

Next let us consider the term (*II*). Notice that

$$\begin{aligned}&\rho(\pi_p^{m,p}(\psi), \pi_p^{m,p}(\phi))\\ &\le \sum_{r=p+1}^{q(m,p)} \rho(\pi_r^{m,p}(\psi), \pi_{r-1}^{m,p}(\psi)) + \sum_{r=p+1}^{q(m,p)} \rho(\pi_r^{m,p}(\phi), \pi_{r-1}^{m,p}(\phi)) + \rho(\psi, \phi)\end{aligned}$$

and the right hand side is not bigger than $9 \cdot 2^{-p}$ whenever $\rho(\psi, \phi) \le 2^{-p}$. Hence it follows from Lemmas 2.1.1 and 2.1.2 that

$$\begin{aligned}&E \sup_{\substack{\psi,\phi\in\Psi^m\\ \rho(\psi,\phi)\le 2^{-p}}} \sup_{t\in[0,\tau]} |X_t^{\pi_p^{m,p}(\psi)} - X_t^{\pi_p^{m,p}(\phi)}| 1_{\{\|X\|_{\rho,\tau}\le K\}}\\ &\lesssim 9\cdot 2^{-p}K\sqrt{\log(1 + N(\Psi, \rho; 2^{-p})^2)} \le 9\sqrt{2}\cdot 2^{-p}K\sqrt{\log(1 + N(\Psi, \rho; 2^{-p}))}.\end{aligned}$$

To show the assertion of the theorem, for a given $\delta > 0$ choose $p \in \mathbb{Z}$ such that $2^{-p-1} < \delta \le 2^{-p}$. Then, the estimates for the terms (*I*) and (*II*) yield that

$$\begin{aligned}&E \sup_{\substack{\psi,\phi\in\Psi^m\\ \rho(\psi,\phi)\le\delta}} \sup_{t\in[0,\tau]} |X_t^\psi - X_t^\phi| 1_{\{\|X\|_{\rho,\tau}\le K\}}\\ &\lesssim \sum_{r=p}^{q(m,p)} 2^{-r}K\sqrt{\log(1 + N(\Psi, \rho; 2^{-r}))} \le 2K\int_0^{2\delta} \sqrt{\log(1 + N(\Psi, \rho; \varepsilon))}d\varepsilon.\end{aligned}$$

The proof is accomplished by letting $m \to \infty$. □

One may sometimes encounter the question whether the paths $\psi \rightsquigarrow X_\tau^\psi$ and $(t,\psi) \rightsquigarrow X_t^\psi$ are continuous and/or bounded. Applying the result above, we can get two kinds of answers to this problem. The first one is concerned with the case where Ψ is countable.

Theorem 2.4.3 *Consider the same situation as Theorem 2.4.2. Suppose also that Ψ is countable and that*

$$P(\|X\|_{\rho,\tau} < \infty) = 1 \quad \textit{and} \quad \int_0^1 \sqrt{\log N(\Psi,\rho;\varepsilon)}d\varepsilon < \infty.$$

Then, almost all paths of $\psi \rightsquigarrow X_\tau^\psi$ are uniformly ρ-continuous on Ψ; moreover, they belong to $\ell^\infty(\Psi)$. Furthermore, when $\tau > 0$ is a constant, almost all paths of $(t,\psi) \rightsquigarrow X_t^\psi$ are uniformly $\widetilde{\rho}$-continuous on $[0,\tau]\times\Psi$, where $\widetilde{\rho}((t,\psi),(s,\phi)) = |t-s| \vee \rho(\psi,\phi)$; moreover, they belong to $\ell^\infty([0,\tau]\times\Psi)$.

Proof. It follows from Theorem 2.4.2 that for every $i \in \mathbb{N}$ there exists $\delta_i > 0$ such that

$$E \sup_{t\in[0,\tau]} \sup_{\substack{\psi,\phi\in\Psi\\ \rho(\psi,\phi)\le\delta_i}} |X_t^\psi - X_t^\phi| 1_{\{\|X\|_{\rho,\tau}\le K\}} \lesssim K\cdot 4^{-i}.$$

Here, we set

$$A_i = \{\sup_{\rho(\psi,\phi)\le\delta_i} \sup_{t\in[0,\tau]} |X_t^\psi - X_t^\phi| > 2^{-i}\} \qquad \forall i \in \mathbb{N}.$$

Then, since $\sum_i P(A_i \cap \{\|X\|_{\rho,\tau} \le K\}) < \infty$, it follows from the Borel-Cantelli lemma that $P(\limsup_i A_i \cap \{\|X\|_{\rho,\tau}\le K\}) = 0$ for every $K > 0$. Noting $\bigcup_{K\in\mathbb{N}}\{\|X\|_{\rho,\tau}\le K\} = \{\|X\|_{\rho,\tau} < \infty\}$, we obtain that $P(\limsup_i A_i) = 0$, which implies the uniform continuity. Since (Ψ,ρ) is totally bounded, almost all pathes are bounded. □

When Ψ is uncountable, the following gives a sufficient condition for the existence of a continuous version of $\psi \rightsquigarrow X_\tau^\psi$.

Theorem 2.4.4 *Consider the same situation as Theorem 2.4.2. Suppose also that it holds for a choice of quadratic ρ-modulus $\|X\|_{\rho,\tau}$ of X that*

$$P\left([\|X\|_{\rho,\tau}]_{\mathcal{F}_\tau,P} < \infty\right) = 1 \quad \textit{and} \quad \int_0^1 \sqrt{\log N(\Psi,\rho;\varepsilon)}d\varepsilon < \infty.$$

Then, there exists a family $\{\widetilde{X}(\psi) : \psi\in\Psi\}$ of $\mathcal{F}_\tau$-measurable random variables such that $\widetilde{X}(\psi) = X_\tau^\psi$ almost surely for every $\psi\in\Psi$ and that almost all paths of $\psi \rightsquigarrow \widetilde{X}(\psi)$ are uniformly ρ-continuous; moreover, they belong to $\ell^\infty(\Psi)$. (Such a process $\psi \rightsquigarrow \widetilde{X}(\psi)$ is called a ρ-continuous version of $\psi \rightsquigarrow X_\tau^\psi$.)

Proof. Consider the $\mathcal{F}_\tau$-measurable partition $\Omega = \bigcup_{K\in\mathbb{N}\cup\{\infty\}} \Omega(K)$ given by

$$\begin{aligned}\Omega(K) &= \left\{[\|X\|_{\rho,\tau}]_{\mathcal{F}_\tau,P} \in [K-1,K)\right\} \in \mathcal{F}_\tau \qquad \forall K\in\mathbb{N},\\ \Omega(\infty) &= \left\{[\|X\|_{\rho,\tau}]_{\mathcal{F}_\tau,P} = \infty\right\} \in \mathcal{F}_\tau,\end{aligned}$$

and define the process $\psi \rightsquigarrow Y_K(\psi)$ by $Y_K(\psi) = X_\tau^\psi 1_{\Omega(K)}$ for every $K \in \mathbb{N}$. Notice that (Ψ,ρ) is separable. For every $K\in\mathbb{N}$, since $\psi \rightsquigarrow Y_K(\psi)$ is continuous in probability by Theorem 2.4.2, it admits a separable version $\psi \rightsquigarrow \widetilde{Y}_K(\psi)$; here, we may define $Y_K \equiv 0$ on the set $\Omega\setminus\Omega(K)$. In the same way as Theorem 2.4.3, we can show that almost all paths of $\psi \rightsquigarrow \widetilde{Y}_K(\psi)$ are uniformly ρ-continuous. Thus the process $\widetilde{X} = \sum_{K\in\mathbb{N}} \widetilde{Y}_K$ satisfies the required properties. □

Notice that the constructed $\widetilde{X}(\psi)$ is not the terminal variable of a continuous local martingale any more. However, it is conjectured that such a construction, including also the parameter t, would be possible.

In Theorem 2.4.2, the requirement that ρ should be a *proper* metric on Ψ is strong for some applications. The following theorem is concerned with an adaptation to a (random) semimetric ϱ which is "weaker" than the metric ρ; the entropy number should be still computed with respect to the metric ρ. The proof is similar to (and easier than) that for (ii) of Theorem 2.2.8, hence is omitted.

Theorem 2.4.5 *Let (Ψ,ρ) be a totally bounded metric space. Let $X = (X^\psi|\psi\in\Psi)$ be a family of continuous local martingales indexed by Ψ such that $X_0^\psi = 0$, and let τ be a finite stopping time, both of which are defined on a stochastic basis* $\mathbf{B}$. *If a given random semimetric ϱ on Ψ satisfies that*

$$\sqrt{\langle X^\psi - X^\phi, X^\psi - X^\phi\rangle_\tau} \le \varrho(\psi,\phi) \qquad \forall \psi,\phi\in\Psi \qquad P_*\text{-almost surely,}$$

then it holds that for every $\delta, K > 0$

$$\sup_{\substack{\Psi^*\subset\Psi\\ \text{countable}}} E^* \sup_{t\in[0,\tau]} \sup_{\substack{\psi,\phi\in\Psi^*\\ \varrho(\psi,\phi)\le K\delta}} |X_t^\psi - X_t^\phi| 1_{\{\|\varrho\|_\rho \le K\}} \lesssim K\int_0^\delta \sqrt{\log(1+N(\Psi,\rho;\varepsilon))}d\varepsilon,$$

where

$$\|\varrho\|_\rho = \sup_{\substack{\psi,\phi\in\Psi\\ \psi\ne\phi}} \frac{\varrho(\psi,\phi)}{\rho(\psi,\phi)},$$

provided the integral on the right hand side is finite (the first supremum is taken over all countable subsets Ψ^* of Ψ).

2.A Notes

The martingale version of the Bernstein inequality (Lemma 2.1.1) is due to Freedman (1975) who dealt with the discrete-time case. The inequality requires that the jumps of a martingale are bounded, but this assumption has been replaced by a kind of higher order moment condition by van de Geer (1995b), which we quoted as Lemmas 2.2.6 and 2.3.5.

The usefulness of bounds for expectation of supremum was shown by Pollard (1989). See also Pollard (1990), Kim and Pollard (1990) and, for more details, van der Vaart and Wellner (1996). The inequalities given in Section 2.2 have the same nature as that of van de Geer (1995b, 1997) who derived a probability inequality with a different definition of brackets.

Related to Theorems 2.4.3 and 2.4.4, one can find a general theory of the regularity of sample paths in Chapter 11 of Ledoux and Talagrand (1991).

Chapter 3

Weak Convergence Theorems

3.1 Preliminaries

Let us quote here a tightness criterion for sequences of random elements taking values in ℓ^∞-spaces. The proof can be found in Chapter 1.5 of van der Vaart and Wellner (1996).

Theorem 3.1.1 *Let T be an arbitrary set. For every $n \in \mathbb{N}$, let $(\Omega^n, \mathcal{F}^n, P^n)$ be a probability space and X^n a mapping from Ω^n to $\ell^\infty(T)$. Consider the following statements:*
(i) *X^n converges weakly in $\ell^\infty(T)$ to a tight, Borel law;*
(ii) *every finite-dimensional marginal of $t \rightsquigarrow X^n(t)$ converges weakly to a (tight,) Borel law;*
(iii) *for every $\varepsilon, \eta > 0$ there exists a finite partition $\{T_k : 1 \leq k \leq N\}$ of T such that*

$$\limsup_{n\to\infty} P^{n*}\left(\max_{1\leq k\leq N}\sup_{t,s\in T_k}|X^n(t)-X^n(s)|>\varepsilon\right)\leq\eta;$$

(iv) *there exist a semimetric ρ on T such that (T,ρ) is totally bounded and that for every $\varepsilon, \eta > 0$ there exists $\delta > 0$ such that*

$$\limsup_{n\to\infty} P^{n*}\left(\sup_{\substack{t,s\in T\\ \rho(t,s)\leq\delta}}|X^n(t)-X^n(s)|>\varepsilon\right)\leq\eta.$$

Then, there is the equivalence (i) $\iff$ (ii) + (iii) $\iff$ (ii) + (iv). *Furthermore, if the finite-dimensional marginals of a process $t \rightsquigarrow X(t)$ have the same laws as those of the limits in* (ii), *then there exists a version $\widetilde{X}$ of X such that $X^n \overset{P^n}{\Longrightarrow} \widetilde{X}$ in $\ell^\infty(T)$ and that almost all paths $t \rightsquigarrow \widetilde{X}(t)$ are uniformly ρ-continuous, where ρ is the semimetric appearing in* (iv). *Furthermore, if the finite-dimensional marginals of the process $t \rightsquigarrow X(t)$ are Gaussian, the semimetric ρ_2 defined by*

$$\rho_2(t,s)=\sqrt{E|X(t)-X(s)|^2}\qquad\forall t,s\in T$$

satisfies the same properties as ρ.

Remark. No measurability of X^n has been assumed. On the other hand, in the latter part of the conclusion, all finite-dimensional marginals of the processes $t \rightsquigarrow X(t)$ and $t \rightsquigarrow \widetilde{X}(t)$ are implicitly assumed to be Borel measurable. Moreover, the assertion means that it is possible to choose a version $\widetilde{X}$ that is Borel measurable in $(\ell^\infty(T), \|\cdot\|_\infty)$.

3.2 Multivariate Point Processes

Let $(E, \mathcal{E})$ be a Blackwell space and Ψ an arbitrary set. For every $n \in \mathbb{N}$, let μ^n be an E-valued multivariate point process defined on a stochastic basis $\mathbf{B}^n = (\Omega^n, \mathcal{F}^n, \mathbf{F}^n = (\mathcal{F}_t^n)_{t \in \mathbb{R}_+}, P^n)$, and ν^n a "good" version of the predictable compensator of μ^n. Let $\mathcal{W}^n = \{W^{n,\psi} : \psi \in \Psi\}$ be a family of predictable functions on $\tilde{\Omega}^n = \Omega^n \times \mathbb{R}_+ \times E$ indexed by Ψ. Let a DFP Π of Ψ be given. Notice that $(E, \mathcal{E})$, Ψ and Π do not depend on n, while all other objects are indexed by $n \in \mathbb{N}$ (we will discuss the case of DFP's Π^n varying with $n \in \mathbb{N}$ at the end of this section). In the same way as Section 2.2, we introduce the following notations:

- the predictable envelope $\overline{W}^n$ of $\mathcal{W}^n$;
- the quadratic Π-modulus $\|\mathcal{W}^n\|_\Pi$ of $\mathcal{W}^n$.

Further, let a finite stopping time τ^n on $\mathbf{B}^n$ be given. Throughout this section, we shall assume:

(3.2.1) $\qquad$ the process $\;t \rightsquigarrow \overline{W}^n * \nu_t^n\;$ takes values in $[0, \infty)$.

As in Section 2.2, we define the local martingales $t \rightsquigarrow X_t^{n,\psi}$ and the locally square-integrable martingales $t \rightsquigarrow X^{n,a,\psi}$ on $\mathbf{B}^n$ by

$$X_t^{n,\psi} = W^{n,\psi} * (\mu^n - \nu^n)_t \qquad \forall t \in \mathbb{R}_+\ \forall \psi \in \Psi$$

and

$$X_t^{n,a,\psi} = W^{n,\psi} 1_{\{\overline{W} \le a\}} * (\mu^n - \nu^n)_t \qquad \forall t \in \mathbb{R}_+\ \forall \psi \in \Psi\ \forall a > 0,$$

respectively. We will derive the asymptotic behavior of the processes $\psi \rightsquigarrow X_{\tau^n}^{n,\psi}$ and $(t, \psi) \rightsquigarrow X_t^{n,\psi}$, as $n \to \infty$.

Let us now introduce several conditions. The first one is the *Partitioning Entropy condition*, which is a natural generalization of the metric entropy condition for L^2-bracketing in the I.I.D. case:

[PE] there exists a DFP Π of Ψ such that

$$\|\mathcal{W}^n\|_{\Pi, \tau^n} = O_{P^n}(1) \quad \text{and} \quad \int_0^{\Delta_\Pi} H_\Pi(\varepsilon) d\varepsilon < \infty.$$

Notice that, due to Lemma 2.2.2, under **[PE]** we can always construct a new NFP Π which satisfies the displayed conditions. Next, we shall also consider two kinds of *Lindeberg conditions*:
[L1] $\overline{W}^n 1_{\{\overline{W}^n > \varepsilon\}} * \nu^n_{\tau^n} \xrightarrow{P^n} 0$ for every $\varepsilon > 0$;
[L2] $|\overline{W}^n|^2 1_{\{\overline{W}^n > \varepsilon\}} * \nu^n_{\tau^n} \xrightarrow{P^n} 0$ for every $\varepsilon > 0$.
When we mention **[L2]**, the assumption that

$$\text{(3.2.2)} \qquad \text{the process} \quad t \rightsquigarrow |\overline{W}^n|^2 * \nu^n_t \quad \text{takes values in } [0, \infty)$$

is also implicitly imposed in addition to (3.2.1), and in this case the process $t \rightsquigarrow X^{n,\psi}_t$ is a locally square-integrable martingale on $\mathbf{B}^n$. It is trivial that **[L2]** implies **[L1]**.

Next let us introduce some conditions prescribing the asymptotic behavior of the quadratic covariations. Let S be a subset of $\mathbb{R}_+$, and suppose that the family $\{C^{(\psi,\phi)}_t : t \in \mathbb{R}_+, (\psi, \phi) \in \Psi^2\}$ of constants in the following satisfies that

$$\text{(3.2.3)} \qquad t \rightsquigarrow C^{(\psi,\phi)}_t \text{ is continuous for every } (\psi, \phi) \in \Psi^2 :$$

[C1] $[X^{n,\psi}, X^{n,\phi}]_t \xrightarrow{P^n} C^{(\psi,\phi)}_t$ for every $t \in S$ and $(\psi, \phi) \in \Psi^2$;
[C2] $\langle X^{n,\psi}, X^{n,\phi}\rangle_t \xrightarrow{P^n} C^{(\psi,\phi)}_t$ for every $t \in S$ and $(\psi, \phi) \in \Psi^2$;
[C1$_a$] $[X^{n,a,\psi}, X^{n,a,\phi}]_t \xrightarrow{P^n} C^{(\psi,\phi)}_t$ for every $t \in S$ and $(\psi, \phi) \in \Psi^2$, for every $a > 0$;
[C2$_a$] $\langle X^{n,a,\psi}, X^{n,a,\phi}\rangle_t \xrightarrow{P^n} C^{(\psi,\phi)}_t$ for every $t \in S$ and $(\psi, \phi) \in \Psi^2$, for every $a > 0$.
Similarly to the remark following **[L2]**, the assumption (3.2.2) is implicitly imposed when we mention **[C2]**. It is well-known that the quadratic covariations are given by

$$[X^{n,\psi}, X^{n,\phi}]_t = \sum_{s \le t} \Delta X^{n,\psi}_s \Delta X^{n,\phi}_s$$

and (under the assumption (3.2.2))

$$\langle X^{n,\psi}, X^{n,\phi}\rangle_t = (W^{n,\psi} W^{n,\phi}) * \nu^n_t - \sum_{s \le t} \widehat{W}^{n,\psi}_s \widehat{W}^{n,\phi}_s,$$

where $\widehat{W}^{n,\psi}_t(\omega) = \int_E W^{n,\psi}(\omega, t, x)\nu(\omega; \{t\} \times dx)$, respectively.

Using the constants $\{C^{\psi,\phi}_t\}$ appeared in the conditions above, we set (formally)

$$\text{(3.2.4)} \qquad \rho_2((t, \psi), (s, \phi)) = \sqrt{C^{(\psi,\psi)}_t + C^{(\phi,\phi)}_s - 2C^{(\psi,\phi)}_{t \wedge s}}$$

for every $(t, \psi), (s, \phi) \in \mathbb{R}_+ \times \Psi$. Any of **[C1]**, **[C2]**, **[C1$_a$]** or **[C2$_a$]** implies that the value under the square-root is non-negative for every $(t, \psi), (s, \phi) \in S \times \Psi$, hence the $\mathbb{R}_+$-valued function ρ_2 is well-defined by the formula (3.2.4) at least on $(S \times \Psi)^2$. Further, by virtue of (3.2.3), this is true also on $([0, \tau] \times \Psi)^2$ if S is a dense subset of the finite interval $[0, \tau]$ with τ being a constant.

The assertions in the following lemma are clear or rather well-known (see e.g. Theorem VIII.3.6 of Jacod and Shiryaev (1987) for the part (ii) below), but we state the proofs with minor modification to our situation.

Lemma 3.2.1 (i) *The condition* **[L1]** *implies the following:*
(i_1) $\overline{W}^n 1_{\{\overline{W}^n>\varepsilon\}} * \mu^n_{\tau^n} \xrightarrow{P^n} 0$ *for every* $\varepsilon > 0$*;*
(i_2) $\sup_{t\in[0,\tau^n]}\sup_{\psi\in\Psi}|X^{n,\psi}_t - X^{n,a,\psi}_t| \xrightarrow{P^{n*}} 0$ *for every* $a > 0$*;*
(i_3) $\sup_{t\in[0,\tau^n]}\Delta(\overline{W}^n * \mu^n_t) \xrightarrow{P^n} 0$ *and* $\sup_{t\in[0,\tau^n]}\Delta(\overline{W}^n * \nu^n_t) \xrightarrow{P^n} 0$*;*
(i_4) $\sup_{t\in[0,\tau^n]}\sup_{\psi\in\Psi}|\Delta X^{n,a,\psi}_t| \xrightarrow{P^{n*}} 0$ *for every* $a > 0$*.*

(ii) *Let* $\tau^n \equiv \tau$ *be a fixed constant, and suppose that* S *is a subset of the finite interval* $[0,\tau]$*. Then, under* **[L1]** *it holds that* **[C1]** $\Leftrightarrow$ **[C1$_a$]** $\Leftrightarrow$ **[C2$_a$]**. *Under* **[L2]**, *the condition* **[C2]** *is also equivalent to any of them.*

Proof. It follows from Lenglart's inequality that

$$P^n\left(\overline{W}^n 1_{\{\overline{W}^n>\varepsilon\}} * \mu^n_{\tau^n} \geq \eta\right) \leq \eta + P^n\left(\overline{W}^n 1_{\{\overline{W}^n>\varepsilon\}} * \nu^n_{\tau^n} \geq \eta^2\right) \qquad \forall \eta > 0,$$

hence the condition **[L1]** implies (i_1). The assertions (i_2), (i_3) and (i_4) are immediate from (i_1).

Next we show the part (ii) of the lemma. By polarization it is enough to consider the case $\phi = \psi$. Observe that

$$\begin{aligned}
&\left|[X^{n,\psi}, X^{n,\psi}]_t - [X^{n,a,\psi}, X^{n,a,\psi}]_t\right| \\
&= \left|\sum_{s\leq t}(\Delta X^{n,\psi}_s + \Delta X^{n,a,\psi}_s)(\Delta X^{n,\psi}_s - \Delta X^{n,a,\psi}_s)\right| \\
&\leq 2\sum_{s\leq t}\left|\Delta W^{n,\psi} 1_{\{\overline{W}>a\}} * (\mu^n - \nu^n)_s\right| \qquad \text{on the set } \Omega^n_1 \\
&\leq 2\left|\overline{W}^n 1_{\{\overline{W}>a\}} * \mu^n_\tau + \overline{W}^n 1_{\{\overline{W}>a\}} * \nu^n_\tau\right|,
\end{aligned}$$

where $\Omega^n_1 = \{\sup_{t\in[0,\tau]}|\Delta X^{n,\psi}_t| \leq 1\} \cup \{\sup_{t\in[0,\tau]}|\Delta X^{n,a,\psi}_t| \leq 1\}$. The assertion that **[C1]** $\Leftrightarrow$ **[C1$_a$]** under **[L1]** is now derived from (i_1), (i_3) and (i_4).

To show the equivalence **[C1$_a$]** $\Leftrightarrow$ **[C2$_a$]** under **[L1]**, fix any $a > 0$, and we set $Y^n = [X^{n,a,\psi}, X^{n,a,\psi}] - \langle X^{n,a,\psi}, X^{n,a,\psi}\rangle$. We will prove that $\sup_{s\in[0,t]}|\Delta Y^n_s| \xrightarrow{P^n} 0$ for every $t \in S$ under either **[L1]** + **[C1$_a$]** or **[L1]** + **[C2$_a$]**. Since $X^{n,a,\psi}$ is a locally square-integrable martingale, we have that Y^n is a local martingale and that so is $|Y^n|^2 - [Y^n, Y^n]$ (see Proposition I.4.50 of Jacod and Shiryaev (1987)). Hence Lenglart's inequality yields that for every $\varepsilon, \eta > 0$

$$P^n\left(\sup_{s\in[0,t]}|Y^n_s|^2 \geq \varepsilon\right) \leq \frac{1}{\varepsilon}\left(\eta + E^n \sup_{s\in[0,t]}\Delta[Y^n, Y^n]_s\right) + P^n([Y^n, Y^n]_t \geq \eta)$$

$$\leq \frac{2\eta}{\varepsilon} + \left(\frac{16a^4}{\varepsilon} + 1\right) P^n([Y^n, Y^n]_t \geq \eta),$$

because $\Delta[Y^n, Y^n] = |\Delta Y^n|^2 \leq (|\Delta[X^{n,a,\psi}, X^{n,a,\psi}]|^2 \vee |\Delta\langle X^{n,a,\psi}, X^{n,a,\psi}\rangle|^2) \leq 16a^4$. Thus it suffices to show that

$$[Y^n, Y^n]_t \xrightarrow{P^n} 0 \qquad \forall t \in S \tag{3.2.5}$$

under either **[L1]** + **[C1$_a$]** or **[L1]** + **[C2$_a$]**.

Since the local martingale Y^n has finite variation, we have

$$\begin{aligned}
[Y^n, Y^n]_t &= \sum_{s \leq t} |\Delta Y_s^n|^2 \\
&\leq \sum_{s \leq t} |\Delta[X^{n,a,\psi}, X^{n,a,\psi}]_s|^2 + \sum_{s \leq t} |\Delta\langle X^{n,a,\psi}, X^{n,a,\psi}\rangle_s|^2 \\
&\leq \alpha_t^n A_t^n + \beta_t^n B_t^n,
\end{aligned}$$

where

$$\begin{aligned}
\alpha_t^n &= \sup_{s \in [0,t]} \Delta[X^{n,a,\psi}, X^{n,a,\psi}]_s, & A_t^n &= [X^{n,a,\psi}, X^{n,a,\psi}]_t, \\
\beta_t^n &= \sup_{s \in [0,t]} \Delta\langle X^{n,a,\psi}, X^{n,a,\psi}\rangle_s, & B_t^n &= \langle X^{n,a,\psi}, X^{n,a,\psi}\rangle_t.
\end{aligned}$$

Using (i$_3$), we obtain that $\alpha_t^n \xrightarrow{P^n} 0$ and $\beta_t^n \xrightarrow{P^n} 0$ for every $t \in S$, under **[L1]**. On the other hand, Lenglart's inequality yields that

$$P^n(A_t^n \geq \varepsilon) \leq \frac{\eta + 2a^2}{\varepsilon} + P^n(B_t^n \geq \eta) \qquad \forall \varepsilon, \eta > 0$$

and that

$$P^n(B_t^n \geq \varepsilon) \leq \frac{\eta + 4a^2}{\varepsilon} + P^n(A_t^n \geq \eta) \qquad \forall \varepsilon, \eta > 0.$$

Hence **[C2$_a$]** implies that $A_t^n = O_{P^n}(1)$, and **[C1$_a$]** does that $B_t^n = O_{P^n}(1)$. The claim (3.2.5) has been established.

The equivalence that **[C2]** $\Leftrightarrow$ **[C2$_a$]** under **[L2]** follows from the inequality

$$\begin{aligned}
&\left|\langle X^{n,\psi}, X^{n,\psi}\rangle_t - \langle X^{n,a,\psi}, X^{n,a,\psi}\rangle_t\right| \\
&\leq |\overline{W}^n|^2 1_{\{\overline{W}^n > a\}} * \nu_\tau^n \\
&\quad + \sum_{t \leq \tau} \int_E 2\overline{W}^n(t,x)\nu(\{t\} \times dx) \int_E \overline{W}^n(t,x) 1_{\{\overline{W}^n(t,x) > a\}} \nu(\{t\} \times dx) \\
&\leq |\overline{W}^n|^2 1_{\{\overline{W}^n > a\}} * \nu_\tau^n + 2\overline{W}^n 1_{\{\overline{W}^n > a\}} * \nu_\tau^n \qquad \text{on the set } \Omega_2^n,
\end{aligned}$$

where $\Omega_2^n = \{\sup_{t \in [0,\tau]} \Delta(\overline{W}^n * \nu_t^n) \leq 1\}$. □

The first result of this section is concerned with the processes $\psi \rightsquigarrow X_{\tau^n}^{n,\psi}$.

Theorem 3.2.2 *Consider the above situation with (3.2.1). Suppose that every finite-dimensional marginal of* $X^n_{\tau^n} = (X^{n,\psi}_{\tau^n} | \psi \in \Psi)$ *converges weakly to a (tight,) Borel law, and also that the conditions* **[PE]** *and* **[L1]** *are satisfied. Then* $X^n_{\tau^n}$ *converges weakly in* $\ell^\infty(\Psi)$ *to a tight, Borel law.*

The result above is a direct consequence of the next lemma, applying Theorem 3.1.1

Lemma 3.2.3 *The conditions* **[PE]** *and* **[L1]** *imply that for every* $\varepsilon, \eta > 0$ *there exists a finite partition* $\{\Psi_k : 1 \leq k \leq N\}$ *of* Ψ *such that*

$$\limsup_{n\to\infty} P^{n*}\left(\sup_{t\in[0,\tau^n]} \sup_{\substack{1\leq k\leq N \\ \psi,\phi\in\Psi_k}} |X^{n,\psi}_t - X^{n,\phi}_t| > \varepsilon \right) \leq \eta.$$

Proof. Take a NFP Π which satisfies the requirements of **[PE]**. Fix any $\varepsilon, \eta > 0$. First notice that for any $\delta \in (0, \Delta_\Pi]$ and $K > 0$

$$P^{n*}\left(\sup_{t\in[0,\tau^n]} \sup_{\substack{1\leq k\leq N_\Pi(\delta) \\ \psi,\phi\in\Psi(\delta;k)}} |X^{n,a(\delta,K),\psi}_t - X^{n,a(\delta,K),\phi}_t| > \varepsilon \right) \leq (I) + (II), \tag{3.2.6}$$

where the terms on the right hand side are given by:

$$\begin{aligned} (I) &= P^n(\|\mathcal{W}^n\|_{\Pi,\tau^n} > K); \\ (II) &= \frac{1}{\varepsilon} E^{n*} \sup_{t\in[0,\tau^n]} \sup_{\substack{1\leq k\leq N_\Pi(\delta) \\ \psi,\phi\in\Psi(\delta;k)}} |X^{n,a(\delta,K),\psi}_t - X^{n,a(\delta,K),\phi}_t| 1_{\{\|\mathcal{W}^n\|_{\Pi,\tau^n}\leq K\}}, \end{aligned}$$

where $a(\delta, K) = \delta K / \tilde{H}_\Pi(\delta/2)$. It follows from (i) of Theorem 2.2.3 that there exists a universal constant $C > 0$ such that

$$(II) \leq C \cdot \frac{K}{\varepsilon} \int_0^\delta \tilde{H}_\Pi(\epsilon) d\epsilon. \tag{3.2.7}$$

Now, the first condition of **[PE]** yields that there exists a constant $K = K_\eta > 0$ such that $\limsup_{n\to\infty}(I) \leq \eta/2$. Next, since $\tilde{H}_\Pi(\epsilon) \leq 1 + H_\Pi(\epsilon)$, the second condition of **[PE]** implies that we can choose a sufficiently small constant $\delta = \delta_{\varepsilon,\eta} > 0$ such that the right hand side of (3.2.7) is not bigger than $\eta/2$. Consequently, (i_2) of Lemma 3.2.1 with $a = a(\delta_{\varepsilon,\eta}, K_\eta)$ yields the assertion. □

The next result deals with the processes $(t, \psi) \rightsquigarrow X^{n,\psi}_t$.

Theorem 3.2.4 *Consider the above situation with (3.2.1) where* $\tau^n \equiv \tau$ *is a fixed positive constant, and let* S *be a dense subset of the finite interval* $[0, \tau]$ *containing* τ*. Suppose that either* **[PE]** + **[L1]** + **[C1]** *or* **[PE]** + **[L2]** + **[C2]** *is satisfied. Then, it* *holds that* $X^n \overset{P^n}{\Longrightarrow} X$ *in* $\ell^\infty([0,\tau]\times\Psi)$, *where each finite-dimensional marginal of the process* $(t,\psi) \rightsquigarrow X_t^\psi$ *has the Gaussian distribution* $N(0,\Sigma)$ *with* $\Sigma = \{\Sigma_{ij}\}$ *given by* $\Sigma_{ij} = C_{t_i \wedge t_j}^{(\psi_i,\psi_j)}$. *Furthermore, the formula (3.2.4) defines a semimetric* ρ_2 *on* $[0,\tau]\times\Psi$ *such that* $[0,\tau]\times\Psi$ *is totally bounded with respect to* ρ_2, *and that almost all paths of* X *are uniformly* ρ_2*-continuous.*

The following lemma, which is rather well-known, is used to show the result above.

Lemma 3.2.5 *Under* **[L1]** + **[C1]**, *for every* $\psi \in \Psi$ *and every* $\varepsilon, \eta > 0$ *there exists* $\delta > 0$ *such that*

$$\limsup_{n\to\infty} P^n\left(\sup_{\substack{t,s\in[0,\tau]\\ |t-s|\le\delta}} |X_t^{n,\psi} - X_s^{n,\psi}| > \varepsilon\right) \le \eta.$$

Proof. Fix any $N \in \mathbb{N}$ for a while, and put $a = N^{-1}$. By (ii) of Lemma 3.2.1 we may assume **[L1]** + **[C2$_a$]**. It always holds that $C_0^{(\psi,\psi)} = 0$ and that $t \rightsquigarrow C_t^{(\psi,\psi)}$ is increasing, because so does $t \rightsquigarrow \langle X^{n,a,\psi}, X^{n,a,\psi}\rangle_t$. We may assume $C_\tau^{(\psi,\psi)} > 0$ without loss of generality. Since $t \rightsquigarrow C_t^{(\psi,\psi)}$ is continuous and S is dense in $[0,\tau]$, we can choose some points $\tau_i \in S$ $(i = 1, ..., N)$ such that $C_{\tau_i}^{(\psi,\psi)} - C_{\tau_{i-1}}^{(\psi,\psi)} = C_\tau^{(\psi,\psi)} N^{-1}$, where $\tau_0 = 0$ and $\tau_N = \tau$. It follows from Lemma 2.1.1 that for every $\varepsilon > 0$

$$P^n\left(\sup_{t\in[\tau_{i-1},\tau_i]} |X_t^{n,a,\psi} - X_{\tau_{i-1}}^{n,a,\psi}| > \varepsilon,\ \Omega_N^n\right) \le 2\exp\left(-\frac{\varepsilon^2}{2[\varepsilon a + 2C_\tau^{(\psi,\psi)} N^{-1}]}\right),$$

where

$$\Omega_N^n = \bigcap_{i=1}^N \left\{\langle X^{n,a,\psi}, X^{n,a,\psi}\rangle_{\tau_i} - \langle X^{n,a,\psi}, X^{n,a,\psi}\rangle_{\tau_{i-1}} \le 2C_\tau^{(\psi,\psi)} N^{-1}\right\}.$$

Hence we have

$$P^n\left(\max_{1\le i\le N} \sup_{t\in[\tau_{i-1},\tau_i]} |X_t^{n,a,\psi} - X_{\tau_{i-1}}^{n,a,\psi}| > \varepsilon,\ \Omega_N^n\right) \le 2N\exp\left(-\frac{\varepsilon^2 N}{2[\varepsilon + 2C_\tau^{(\psi,\psi)}]}\right).$$

Here notice that $\lim_{n\to\infty} P^n(\Omega_N^n) = 1$. Choosing a large number N, and then letting $n \to \infty$, we can easily deduce the assertion from (i$_2$) of Lemma 3.2.1. □

Proof of Theorem 3.2.4. Let us check the conditions of Theorem 3.1.1. First, Theorem VIII.3.11 of Jacod and Shiryaev (1987) says that either of **[L1]** + **[C1$_a$]** or **[L2]** + **[C2$_a$]** implies the finite-dimensional convergence of $X^{n,a}$ for any $a > 0$ (recall also (i$_4$) of Lemma 3.2.1). Thus the finite-dimensional convergence of X^n follows from (i$_2$) and (ii) of Lemma 3.2.1. The condition (iii) of Theorem 3.1.1 can be shown by means of Lemmas 3.2.3 and 3.2.5. □

Let us close this section with discussing the case of DFP's Π^n varying with $n \in \mathbb{N}$. In this case, we shall check the condition (iv) of Theorem 3.1.1 instead of (iii). We thus introduce the following condition.

[PE^n] there exists a semimetric ρ on Ψ such that (Ψ, ρ) is totally bounded, and for every $n \in \mathbb{N}$ there exists a DFP Π^n of Ψ such that:

$$||\mathcal{W}^n||_{\Pi^n,\tau^n} = O_{P^n}(1); \qquad \lim_{\delta\downarrow 0}\limsup_{n\to\infty}\int_0^\delta H_{\Pi^n}(\varepsilon)d\varepsilon = 0;$$

$$\lim_{K\to\infty}\limsup_{\delta\downarrow 0}\limsup_{n\to\infty} P^{n*}\left(\sup_{\substack{\psi,\phi\in\Psi \\ \rho(\psi,\phi)\leq\delta}} \frac{\sqrt{|W^{n,\psi} - W^{n,\phi}|^2 1_{\{\overline{W}^n \leq 1\}} * \nu_{\tau^n}}}{\delta} > K\right) = 0.$$

We then have an analogue of Lemma 3.2.3.

Lemma 3.2.6 *The conditions* **[PE^n]** *and* **[L1]** *imply that for every* $\varepsilon, \eta > 0$ *there exists* $\delta > 0$ *such that*

$$\limsup_{n\to\infty} P^{n*}\left(\sup_{t\in[0,\tau^n]}\sup_{\substack{\psi,\phi\in\Psi \\ \rho(\psi,\phi)\leq\delta}} |X_t^{n,\psi} - X_t^{n,\phi}| > \varepsilon\right) \leq \eta.$$

Proof. Repeat the same argument as Lemma 3.2.3 using Theorem 2.2.4 instead of (i) of Theorem 2.2.3. □

Consequently, Theorems 3.2.2 and 3.2.4 hold also with **[PE^n]** instead of **[PE]**. In particular, Theorem 3.2.4 under **[PE^n]** + **[L2]** + **[C2]** refines Theorem 2.2 of Nishiyama (1997); the condition (B_3) there has been removed.

3.3 Martingale Difference Arrays

We give some analogues of Theorems 3.2.2 and 3.2.4; those can be shown using Theorem 2.3.3 (i) instead of Theorem 2.2.3 (i) (or using Theorem 2.3.4 instead of Theorem 2.2.4), thus the proofs are omitted. Let Ψ be an arbitrary set. For every $n \in \mathbb{N}$, let $\{\xi_i^n\}_{i\in\mathbb{N}}$ be an $\ell^\infty(\Psi)$-valued martingale difference array on a discrete-time stochastic basis $\mathbf{B}^n = (\Omega^n, \mathcal{F}^n, \mathbf{F}^n = \{\mathcal{F}_i^n\}_{i\in\mathbb{N}_0}, P^n)$.

Let Π a DFP of Ψ, and let Π^n be a sequence of DPF's of Ψ. In the same way as Section 2.3, we introduce the following notations:

- the adapted envelope $\{\overline{\xi}_i^n\}_{i\in\mathbb{N}}$ of $\{\xi_i^n\}_{i\in\mathbb{N}}$;
- the quadratic Π-modulus $\{||\xi^n||_{\Pi,i}\}_{i\in\mathbb{N}}$ of $\{\xi_i^n\}_{i\in\mathbb{N}}$;

- the quadratic Π^n-modulus $\{\|\xi^n\|_{\Pi^n,i}\}_{i\in\mathbb{N}}$ of $\{\xi_i^n\}_{i\in\mathbb{N}}$.

We shall always assume:

$$E^n\overline{\xi}_i^n < \infty \qquad \forall i \in \mathbb{N}. \tag{3.3.1}$$

For a given finite stopping time σ^n, we make the following conditions:

[PE′] there exists a DFP Π of Ψ such that

$$\|\xi^n\|_{\Pi,\sigma^n} = O_{P^n}(1) \quad \text{and} \quad \int_0^{\Delta_\Pi} H_\Pi(\varepsilon)d\varepsilon < \infty;$$

[PE$^{n\prime}$] there exists a semimetric ρ on Ψ such that (Ψ,ρ) is totally bounded, and for every $n\in\mathbb{N}$ there exists a DFP Π^n of Ψ such that

$$\|\xi^n\|_{\Pi^n,\tau^n} = O_{P^n}(1), \qquad \lim_{\delta\downarrow 0}\limsup_{n\to\infty}\int_0^\delta H_{\Pi^n}(\varepsilon)d\varepsilon = 0,$$

and

$$\lim_{K\to\infty}\limsup_{\delta\downarrow 0}\limsup_{n\to\infty} P^{n*}\left(\sup_{\substack{\psi,\phi\in\Psi\\ \rho(\psi,\phi)\le\delta}} \frac{\sqrt{\sum_{i=1}^{\sigma^n} E_{i-1}^n(|\xi_i^{n,\psi}-\xi_i^{n,\phi}|^2\wedge 1)}}{\delta} > K\right) = 0;$$

[L1′] $\sum_{i=1}^{\sigma^n} E_{i-1}^n\overline{\xi}_i^n 1_{\{\overline{\xi}_i^n>\varepsilon\}} \xrightarrow{P^n} 0$ for every $\varepsilon>0$;

[L2′] $\sum_{i=1}^{\sigma^n} E_{i-1}^n|\overline{\xi}_i^n|^2 1_{\{\overline{\xi}_i^n>\varepsilon\}} \xrightarrow{P^n} 0$ for every $\varepsilon>0$.

When we mention **[L2′]**, the assumption that

$$E^n|\overline{\xi}_i^n|^2 < \infty \qquad \forall i\in\mathbb{N}, \tag{3.3.2}$$

which is stronger than (3.3.1), is implicitly imposed. It is clear that **[L2′]** implies **[L1′]**.

Theorem 3.3.1 *Consider the above situation with (3.3.1). Suppose that every finite-dimensional marginal of $X^n = (X^{n,\psi}|\psi\in\Psi)$ given by $X^{n,\psi} = \sum_{i=1}^{\sigma^n}\xi_i^{n,\psi}$ converges weakly to a (tight,) Borel law, and also that either of* **[PE′]** + **[L1′]** *or* **[PE$^{n\prime}$]** + **[L1′]** *is satisfied. Then X^n converges weakly in $\ell^\infty(\Psi)$ to a tight, Borel law.*

Next, let us consider the process $(t,\psi)\rightsquigarrow X_t^{n,\psi}$ given by

$$X_t^{n,\psi} = \sum_{i=1}^{\sigma_t^n}\xi_i^{n,\psi} \qquad \forall t\in[0,\tau]\ \forall\psi\in\Psi, \tag{3.3.3}$$

where $\tau>0$ is a constant, and $(\sigma_t^n)_{t\in[0,\tau]}$ is a family of finite stopping times on $\mathbf{B}^n$ such that $\sigma_0^n = 0$ and that each path $t\rightsquigarrow\sigma_t^n$ is increasing, càd, with jumps equal to 1. We introduce two kinds of conditions, in which the family $\{C_t^{(\psi,\phi)} : t\in\mathbb{R}_+, (\psi,\phi)\in\Psi^2\}$ of constants should satisfy (3.2.3):

[C1′] $\sum_{i=1}^{\sigma_t^n} \xi_i^{n,\psi}\xi_i^{n,\phi} \xrightarrow{P^n} C_t^{(\psi,\phi)}$ for every $t \in S$ and $(\psi,\phi) \in \Psi^2$;
[C2′] $\sum_{i=1}^{\sigma_t^n} E_{i-1}^n \xi_i^{n,\psi}\xi_i^{n,\phi} \xrightarrow{P^n} C_t^{(\psi,\phi)}$ for every $t \in S$ and $(\psi,\phi) \in \Psi^2$.
Similarly to the remark following **[L2′]**, the assumption (3.3.2) is implicitly imposed when we mention **[C2′]**.

Theorem 3.3.2 *Let S be a dense subset of the finite interval $[0,\tau]$ containing τ. Consider the above situation with (3.3.1), and assume* **[PE′]** *or* **[PE$^{n\prime}$]** *with $\sigma^n = \sigma_\tau^n$. Suppose also that either* **[L1′]** + **[C1′]** *or* **[L2′]** + **[C2′]** *is satisfied. Then, the same conclusion as Theorem 3.2.4 holds for the sequence of processes $X^n = (X_t^{n,\psi} | (t,\psi) \in [0,\tau] \times \Psi)$ defined by (3.3.3).*

Let us close this section with stating a generalization of Jain-Marcus' (1975) central limit theorem to the case of martingale difference arrays. We denote by $N(\Psi,\rho;\varepsilon)$ the ε-covering number of a semimetric space (Ψ,ρ).

Proposition 3.3.3 *Let (Ψ,ρ) be a totally bounded semimetric space. For every $n \in \mathbb{N}$, let $\{\xi_i^n\}_{i\in\mathbb{N}}$ be an $\ell^\infty(\Psi)$-valued martingale difference array on a discrete-time stochastic basis $\mathbf{B}^n$ such that*

$$|\xi_i^{n,\psi} - \xi_i^{n,\phi}| \le K_i^n \rho(\psi,\phi) \qquad \forall \psi,\phi \in \Psi,$$

where $\{K_i^n\}_{i\in\mathbb{N}}$ is an $\mathbb{R}_+$-valued adapted process. For given finite stopping time σ^n, a sufficient condition for **[PE′]** *is*

$$\sum_{i=1}^{\sigma^n} E_{i-1}^n |K_i^n|^2 = O_{P^n}(1) \quad \text{and} \quad \int_0^1 \sqrt{\log N(\Psi,\rho;\varepsilon)}\,d\varepsilon < \infty.$$

3.4 Continuous Local Martingales

Let us begin with giving a definition.

Definition 3.4.1 *A family $X = (X^\psi | \psi \in \Psi)$ of continuous local martingales indexed by a metric space (Ψ,ρ) is said to be ρ-separable if there exist a countable subset Ψ^* of Ψ and a negligible set $N \in \mathcal{F}$ such that for every $\varepsilon > 0$ and $\omega \in \Omega \setminus N$*

$$X_t^\psi(\omega) \in \overline{\{X_t^\phi(\omega) : \phi \in \Psi^*,\ \rho(\psi,\phi) < \varepsilon\}} \qquad \forall t \in \mathbb{R}_+,\ \forall \psi \in \Psi,$$

where the closure is taken in $\mathbb{R} \cup \{-\infty, +\infty\}$.

When (Ψ,ρ) is separable, a sufficient condition for the ρ-separability is that almost all paths $\psi \rightsquigarrow X_t^\psi$ are ρ-continuous, but it is not always easy to check the continuity in general. On the other hand, it is clear that any family of continuous local martingales indexed by a countable set Ψ is ρ-separable (for any metric ρ on Ψ).

Let us now turn to the context of weak convergence. Let (Ψ, ρ) be a totally bounded *proper* metric space. For every $n \in \mathbb{N}$, let $X^n = (X^{n,\psi} | \psi \in \Psi)$ be a (not necessarily ρ-separable) family of continuous local martingales indexed by Ψ defined on a stochastic basis $\mathbf{B}^n = (\Omega^n, \mathcal{F}^n, \mathbf{F}^n = (\mathcal{F}_t^n)_{t \in \mathbb{R}_+}, P^n)$. Let τ^n be a finite stopping time on $\mathbf{B}^n$. We introduce the *Metric Entropy condition.*

[ME] Given finite stopping time τ^n on $\mathbf{B}^n$,

$$\|X^n\|_{\rho,\tau^n} = O_{P^{n*}}(1) \qquad \text{and} \qquad \int_0^1 \sqrt{\log N(\Psi, \rho; \varepsilon)} d\varepsilon < \infty.$$

Theorem 3.4.2 *In the above situation, suppose that the family $X^n = (X^{n,\psi} | \psi \in \Psi)$ is ρ-separable and that $X^n_{\tau^n} = (X^{n,\psi}_{\tau^n} | \psi \in \Psi)$ takes values in $\ell^\infty(\Psi)$ almost surely. Suppose also that every finite-dimensional marginal of $X^n_{\tau^n}$ converges to a (tight,) Borel law, and that the condition* **[ME]** *is satisfied. Then $X^n_{\tau^n}$ converges weakly in $\ell^\infty(\Psi)$ to a tight, Borel law.*

Proof. The assertion is immediate from Theorems 2.4.2 and 3.1.1. □

The result above generalizes Theorem 1 of Bae and Levental (1995b) who assumed the continuity of $(t, \psi) \rightsquigarrow X_t^{n,\psi}$. According to Theorem 2.4.3, when Ψ is countable and **[ME]** is assumed, a sufficient condition for $X^n_{\tau^n} = (X^{n,\psi}_{\tau^n} | \psi \in \Psi)$ to take values in $\ell^\infty(\Psi)$ almost surely is that $P^n(\|X^n\|_{\rho,\tau^n} < \infty) = 1$. If one encounters the situation where Ψ is uncountable and the ρ-separability of the family $X^n = (X^{n,\psi} | \psi \in \Psi)$ itself is difficult to show, the following result concerning a uniformly ρ-continuous version $\psi \rightsquigarrow \widetilde{X}^n(\psi)$ of $\psi \rightsquigarrow X^{n,\psi}_{\tau^n}$ will be helpful. Recall that a sufficient condition for the existence such a version is given by Theorem 2.4.4, and that $\widetilde{X}^n(\psi)$ is not the terminal variable of a continuous local martingale.

Corollary 3.4.3 *In the above situation, suppose that every finite-dimensional marginal of $X^n_{\tau^n}$ converges to a (tight,) Borel law, and that the condition* **[ME]** *is satisfied. Suppose also that it holds for a choice of the quadratic ρ-modulus $\|X^n\|_\rho$ that*

$$P^n\left([\|X^n\|_{\rho,\tau^n}]_{\mathcal{F}^n_{\tau^n}, P^n} < \infty \right) = 1 \qquad \forall n \in \mathbb{N},$$

and choose a uniformly ρ-continuous version $\psi \rightsquigarrow \widetilde{X}^n(\psi)$ of the process $\psi \rightsquigarrow X^{n,\psi}_{\tau^n}$. Then $\widetilde{X}^n$ converges weakly in $\ell^\infty(\Psi)$ to a tight, Borel law. If the limit of each finite-dimensional marginal of $\psi \rightsquigarrow X^{n,\psi}_{\tau^n}$ coincides with that of a process $\psi \rightsquigarrow X(\psi)$, then there exists a uniformly ρ-continuous version $\widetilde{X}$ of X such that $\widetilde{X}^n \xrightarrow{P^n} \widetilde{X}$ in $\ell^\infty(\Psi)$.

Proof. The finite-dimensional convergence of $\psi \rightsquigarrow \widetilde{X}^n(\psi)$ follows from that of $\psi \rightsquigarrow X^{n,\psi}_{\tau^n}$. Next choose a countable dense subset Ψ^* of Ψ. Then, by Theorem 3.4.2, the statement (iv) of Theorem 3.1.1 is satisfied for $(X^{n,\psi}_{\tau^n}|\psi \in \Psi^*)$, thus also for $(\widetilde{X}^n(\psi)|\psi \in \Psi)$. □

Next we consider the process $(t,\psi) \rightsquigarrow X^{n,\psi}_t$. Given subset S of $\mathbb{R}_+$, we make also the following condition, in which the family $\{C^{(\psi,\phi)}_t : t \in \mathbb{R}_+,\ (\psi,\phi) \in \Psi^2\}$ of constants should satisfy (3.2.3):
[C2] $\langle X^{n,\psi}, X^{n,\phi}\rangle_t \xrightarrow{P^n} C^{(\psi,\phi)}_t$ for every $t \in S$ and $(\psi,\phi) \in \Psi^2$.

Theorem 3.4.4 *Let S be a dense subset of the a finite interval $[0,\tau]$ containing τ. In the above situation, suppose that the family $X^n = (X^{n,\psi}|\psi \in \Psi)$ is ρ-separable and that $(X^{n,\psi}_t|(t,\psi) \in [0,\tau]\times\Psi)$ takes values in $\ell^\infty([0,\tau]\times\Psi)$ almost surely. Assume* **[ME]** *with $\tau^n \equiv \tau$ and* **[C2]**. *Then, the same conclusion as Theorem 3.2.4 holds for X^n.*

Proof. Repeat the same assertion as Lemma 3.2.5 to obtain under **[C2]** that for every $\psi \in \Psi$ and every $\varepsilon, \eta > 0$ there exists $\delta > 0$ such that

$$\limsup_{n\to\infty} P^n\left(\sup_{\substack{t,s\in[0,\tau]\\ |t-s|\le\delta}} |X^{n,\psi}_t - X^{n,\psi}_s| > \varepsilon\right) \le \eta.$$

This fact, together with Theorem 2.4.2, implies the asymptotic tightness. Thus, the assertion follows from Theorem 3.1.1 and the martingale central limit theorem. □

The result above refines Theorem 2.3 of Nishiyama (1997); the condition (C_3) there has been removed and the condition (C_5) has been weakened.

3.A Notes

The study of the Donsker theorems for i.i.d. empirical processes indexed by classes of sets or functions was initiated by the landmark paper by Dudley (1978), and was developed in the 80's. There are two types of sufficient conditions for the Donsker property, namely, the *uniform entropy condition* (Pollard (1982)) and the *L^2-bracketing condition* (Ossiander (1987)); see van der Vaart and Wellner (1996) and references therein for refinement and generalizations up to row-independent arrays. In particular, it should be noted that Andersen et al. (1988) contains a result based on a weaker condition than L^2-bracketing one in a row-independent case.

In the recent years, some authors have considered to remove the assumption of independence: Arcone and Yu (1994) and Doukhan et al. (1995) for stationary sequences based on mixing conditions; Levental (1989), Bae (1995) and Bae and Levental (1995a) for stationary martingale sequences; Bae and Levental (1995b) for continuous martingales; Nishiyama (1997) for some continuous-time semimartingales. One can find the roots of the quantity "quadratic modulus" in the works by Bae and Levental cited above. A major part of the results in this chapter was originally presented by Nishiyama (1997), although some of the conditions have been refined as mentioned in the main text. The refinement is partly due to the use of the tightness criterion in terms of partitioning (i.e., (iii) of Theorem 3.1.1) rather than the well-known stochastic ρ-equicontinuity criterion. Van der Vaart and Wellner (1996) is apparently the first to present the partitioning criterion.

For the weak convergence of (finite-dimensional) semimartingales, one should consult the excellent book by Jacod and Shiryaev (1987) which we have referred many times.

Chapter 4

Integral Estimators

4.1 Multiplicative Intensity Model

Let $(E, \mathcal{E})$ be a Blackwell space on which a measure λ is defined. For every $n \in \mathbb{N}$, let μ^n be an E-valued multivariate point process defined on a filtered measurable space $(\Omega^n, \mathcal{F}^n, \mathbf{F}^n = (\mathcal{F}_t^n)_{t \in \mathbb{R}_+})$. Notice that μ^n can be identified with an E-marked point process $\{(T_i^n, Z_i^n);\ i \in \mathbb{N}\}$ through the equality

$$\mu^n(\omega; dt, dz) = \sum_i \varepsilon_{(T_i^n(\omega), Z_i^n(\omega))}(dt, dz),$$

where $0 < T_1^n < T_2^n < \cdots$ and each Z_i^n is an E-valued random variable. We suppose that the predictable compensator ν^n of μ^n under the probability measure P^n on $(\Omega^n, \mathcal{F}^n)$ is given by

$$\nu^n(\omega; dt, dz) = \alpha(t, z) Y^n(\omega, t, z) dt \lambda(dz),$$

where $\alpha(t, z)$ is a $[0, \infty)$-valued measurable function on $\mathbb{R}_+ \times E$, and $Y^n(\omega, t, z)$ is a $[0, \infty)$-valued predictable function on $\Omega^n \times \mathbb{R}_+ \times E$.

Let a constant $\tau > 0$ be given. Throughout this section, we always assume

$$\int_{[0,\tau] \times E} \alpha(t, z) dt \lambda(dz) < \infty;$$

then $\alpha(t, z) dt \lambda(dz)$ defines a finite measure on $[0, \tau] \times E$. We will use the following notation

$$\mathcal{L}^p(\alpha) = \mathcal{L}^p([0, \tau] \times E, \mathfrak{B}[0, \tau] \otimes \mathcal{E}, \alpha(t, z) dt \lambda(dz)) \qquad \forall p \in [1, \infty].$$

We denote by $\|\cdot\|_{\mathcal{L}^p(\alpha)}$ the L^p-seminorm on $\mathcal{L}^p(\alpha)$.

4.1.1 Asymptotic Normality

Let Ψ be a subset of $\mathcal{L}^{2p}(\alpha)$ with an envelop function φ belonging to $\mathcal{L}^{2p}(\alpha)$, where $p \in [1, \infty]$ should be specified in connection with another assumption in Condition 4.1.1

below. Our goal is to estimate the functional $(t,\psi) \rightsquigarrow A(\psi)_t$ where

$$A(\psi)_t = \int_{[0,t]\times E} \psi(s,z)\alpha(s,z)ds\lambda(dz) \qquad \forall (t,\psi) \in [0,\tau]\times\Psi.$$

To do it, we introduce the *generalized inverse* Y^{n-} of Y^n defined by

$$Y^{n-}(\omega,t,z) = \frac{1_{\{Y^n(\omega,t,z)\geq 1\}}}{Y^n(\omega,t,z)},$$

with the convention $0/0 = 0$; we define also

$$I^n(\omega,t,z) = 1_{\{Y^n(\omega,t,z)\geq 1\}}.$$

It then holds that $Y^{n-}Y^n = I^n$. We propose the estimator $(t,\psi) \rightsquigarrow \widehat{A}^n(\psi)_t$ defined by

$$\widehat{A}^n(\psi)_t(\omega) = (\psi Y^{n-}) * \mu_t^n(\omega) \qquad \forall (t,\psi) \in [0,\tau]\times\Psi.$$

The main step is to derive the weak convergence of the residual $\sqrt{n}(\widehat{A}^n - \widetilde{A}^n)$, where $(t,\psi) \rightsquigarrow \widetilde{A}^n(\psi)_t$ is given by

$$\widetilde{A}^n(\psi)_t(\omega) = \int_{[0,t]\times E} \psi(s,z)I^n(\omega,s,z)\alpha(s,z)dt\lambda(dz) \qquad \forall (t,\psi) \in [0,\tau]\times\Psi.$$

We make the following condition.

Condition 4.1.1 *For some $p,q \in [1,\infty]$ such that $(1/p)+(1/q) = 1$, and a measurable function $y = y(t,z)$ on $[0,\tau]\times E$, which is bounded away from zero, it holds that:*

(4.1.1) $$\varphi \in \mathcal{L}^{2p}(\alpha) \quad \text{and} \quad \int_0^1 \sqrt{\log N_{[]}(\Psi, \|\cdot\|_{\mathcal{L}^{2p}(\alpha)};\varepsilon)}d\varepsilon < \infty;$$

(4.1.2) $$\|nY^{n-}\|_{\mathcal{L}^q(\alpha)} = O_{P^n}(1);$$

(4.1.3) $$\Big\|\varphi^2 \cdot |nY^{n-} - (1/y)|\Big\|_{\mathcal{L}^1(\alpha)} = o_{P^n}(1).$$

This condition generalizes (8.4.1) of Andersen et al. (1993). Although the framework of this section does not contain empirical processes of i.i.d. data, one may find an interesting "difference" between the i.i.d. and the present cases. In the i.i.d. case, since the random elements nY^{n-} do not appear, the entropy condition for L^2-bracketing is always optimal. In the present case, however, it is sometimes wise to seek for the entropy condition (4.1.1) with respect to a stronger semimetric when the requirement (4.1.2) can be checked only for $q < \infty$. For instance, when $E = \mathbb{R}$ and $\psi = 1_{(-\infty,z]}$, the bracketing entropy condition is satisfied with the L^{2p}-semimetric for any $p < \infty$, and thus (4.1.2) for an arbitrary $q > 1$ is sufficient. Compare the present Condition 4.1.1 with Conditions 4.1 and 4.2 of Nishiyama (1997) which were concerned only with the case of $p = 1$ and $q = \infty$.

Theorem 4.1.2 *Suppose that a given class Ψ satisfies Condition 4.1.1. Then, it holds that $\sqrt{n}(\widehat{A}^n - \widetilde{A}^n) \overset{P^n}{\Longrightarrow} G$ in $\ell^\infty([0,\tau]\times\Psi)$, where $(t,\psi) \rightsquigarrow G_t^\psi$ is a zero-mean Gaussian process such that*

$$EG_t^\psi G_s^\phi = \int_{[0,t\wedge s]\times E} \psi(u,z)\phi(u,z)\frac{\alpha(u,z)}{y(u,z)}du\lambda(dz) \qquad \forall (t,\psi),(s,\phi) \in [0,\tau]\times\Psi$$

and that almost all paths are uniformly ρ_2-continuous on $[0,\tau]\times\Psi$, where

$$\rho_2((t,\psi),(s,\phi)) = \sqrt{E|G_t^\psi - G_s^\phi|^2} \qquad \forall (t,\psi),(s,\phi)\in[0,\tau]\times\Psi.$$

Further if

(4.1.4) $$\|\varphi\cdot(1-I^n)\|_{\mathcal{L}^1(\alpha)} = o_{P^n}(n^{-1/2}),$$

then it also holds that $\sqrt{n}(\widehat{A}^n - A) \overset{P^n}{\Longrightarrow} G$ in $\ell^\infty([0,\tau]\times\Psi)$.

Proof. We will apply Theorem 3.2.4 to $\mathcal{W}^n = \{W^{n,\psi} : \psi \in \Psi\}$ given by $W^{n,\psi} = \sqrt{n}\psi Y^{n-}$. It indeed holds that $\sqrt{n}(\widehat{A}^n(\psi)_t - \widetilde{A}^n(\psi)_t) = W^{n,\psi} * (\mu^n - \nu^n)_t$.

First notice that

$$\begin{aligned}&\langle W^{n,\psi} * (\mu^n - \nu^n), W^{n,\phi} * (\mu^n - \nu^n)\rangle_t \\ &= n\int_{[0,t]\times E} \psi(s,z)\phi(s,z)Y^{n-}(s,z)\alpha(s,z)ds\lambda(dz).\end{aligned}$$

Since

$$\begin{aligned}&\left|\int_{[0,t]\times E}\psi(s,z)\phi(s,z)\left\{nY^{n-}(s,z) - \frac{1}{y(s,z)}\right\}\alpha(s,z)ds\lambda(dz)\right| \\ &\le \int_{[0,\tau]\times E}|\varphi(s,z)|^2\left|nY^{n-}(s,z) - \frac{1}{y(s,z)}\right|\alpha(s,z)ds\lambda(dz) \\ &= \left\|\varphi^2\cdot|nY^{n-} - (1/y)|\right\|_{\mathcal{L}^1(\alpha)},\end{aligned}$$

the condition **[C2]** follows from (4.1.3). To show **[L2]**, notice that $\overline{W}^n \le \sqrt{n}\varphi Y^{n-}$ and that, when $p\in[1,\infty)$,

(4.1.5) $$\begin{aligned}&\left(|\sqrt{n}\varphi Y^{n-}|^2 1_{\{\sqrt{n}\varphi Y^{n-}>\varepsilon\}}\right) * \nu_\tau^n \\ &= \left\|\varphi^2\cdot nY^{n-}1_{\{\sqrt{n}\varphi Y^{n-}>\varepsilon\}}\right\|_{\mathcal{L}^1(\alpha)} \\ &\le \|\varphi\|_{\mathcal{L}^{2p}(\alpha)}\cdot\left\|\varphi\cdot nY^{n-}1_{\{\sqrt{n}\varphi Y^{n-}>\varepsilon\}}\right\|_{\mathcal{L}^{2p/(2p-1)}(\alpha)} \\ &\le \|\varphi\|_{\mathcal{L}^{2p}(\alpha)}\cdot\frac{1}{\sqrt{n}\varepsilon}\|\varphi\cdot nY^{n-}\|_{\mathcal{L}^{2p/(2p-1)}(\alpha)} \\ &\le \|\varphi\|_{\mathcal{L}^{2p}(\alpha)}\cdot\frac{1}{\sqrt{n}\varepsilon}\cdot\|\varphi\|_{\mathcal{L}^{2p}(\alpha)}\cdot\|nY^{n-}\|_{\mathcal{L}^q(\alpha)},\end{aligned}$$

which converges in P^n-probability to zero by $\varphi \in \mathcal{L}^{2p}(\alpha)$ and (4.1.2). The case of $p = \infty$ and $q = 1$ is easier.

To check the condition **[PE]**, for every $\varepsilon \in (0,1]$, choose $(\|\cdot\|_{\mathcal{L}^{2p}}, \varepsilon)$-brackets $[l^{\varepsilon,k}, u^{\varepsilon,k}]$ in $\mathcal{L}^{2p}$ which cover Ψ. Introduce a DFP Π of Ψ induced from these brackets, that is, $\Pi(\varepsilon) = \{\Psi(\varepsilon;k) : 1 \le k \le N_\Pi(\varepsilon)\}$ is given by $\Psi(\varepsilon;k) = \{\psi \in \Psi : l^{\varepsilon,k} \le \psi \le u^{\varepsilon,k}\}$ with modification to make the partition disjoint. This can be done with $N_\Pi(\varepsilon) = N_{[]}(\Psi, \|\cdot\|_{\mathcal{L}^{2p}}; \varepsilon)$. Since

$$
\begin{aligned}
|\sqrt{n}(u^{\varepsilon,k} - l^{\varepsilon,k})Y^{n-}|^2 * \nu_\tau^n &= \left\| |u^{\varepsilon,k} - l^{\varepsilon,k}|^2 nY^{n-} \right\|_{\mathcal{L}^1(\alpha)} \\
&\le \left\| |u^{\varepsilon,k} - l^{\varepsilon,k}|^2 \right\|_{\mathcal{L}^p(\alpha)} \cdot \|nY^{n-}\|_{\mathcal{L}^q(\alpha)} \\
&= \|u^{\varepsilon,k} - l^{\varepsilon,k}\|_{\mathcal{L}^{2p}(\alpha)}^2 \cdot \|nY^{n-}\|_{\mathcal{L}^q(\alpha)} \\
&\le \varepsilon^2 \|nY^{n-}\|_{\mathcal{L}^q(\alpha)},
\end{aligned}
$$

the quadratic Π-modulus $\|\mathcal{W}^n\|_{\Pi,\tau}$ is bounded by $\sqrt{\|nY^{n-}\|_{\mathcal{L}^q(\alpha)}}$, which is bounded in P^n-probability. This completes the proof. □

4.1.2 Asymptotic Efficiency

Let us discuss the asymptotic efficiency of the estimator $\widehat{A}^n$ following the general theory developed in Chapter 3.11 of van der Vaart and Wellner (1996). We set:

$$
\begin{aligned}
\mathbb{H} &= L^2([0,\tau] \times E, \mathfrak{B}[0,\tau] \otimes \mathcal{E}, \tfrac{\alpha(t,z)}{y(t,z)} dt\lambda(dz)); \\
H &= L^\infty([0,\tau] \times E, \mathfrak{B}[0,\tau] \otimes \mathcal{E}, dt\lambda(dz)).
\end{aligned} \tag{4.1.6}
$$

Here, the function y is the one appearing in Condition 4.1.1. We equip $\mathbb{H}$ with the usual L^2-inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$. Since $1/y$ is assumed to be bounded, H is a linear subspace of the Hilbert space $\mathbb{H}$. Let $\mathbf{P}^n = \{P_h^n : h \in H\}$ be a family of probability measures on $(\Omega^n, \mathcal{F}^n)$ indexed by H.

Suppose that the predictable compensator of $N^{n,i}$ under the probability measure P_h^n is given by

$$\alpha_h^n(t,z) Y^n(t,z) dt,$$

where $\alpha_h^n = \alpha_h^n(t,z)$ is defined by

$$\alpha_h^n = \left(1 + \frac{h}{2\sqrt{ny}}\right)^2 \alpha.$$

Notice that $\alpha_0^n = \alpha$.

STEP I: LOCAL ASYMPTOTIC NORMALITY. Assume $P_h^n \ll P_0^n$. It is well-known that, under some conditions, a version of the log-likelihood ratio is given by

$$\log \frac{dP_h^n|\mathcal{F}_\tau^n}{dP_0^n|\mathcal{F}_\tau^n} = 2\left(\log\left|1+\frac{h}{2\sqrt{ny}}\right|\right) * \mu_\tau^n - \left(\left|1+\frac{h}{2\sqrt{ny}}\right|^2 - 1\right) * \nu_\tau^{n,0} \tag{4.1.7}$$

(see, e.g., Theorem III.5.43 of Jacod and Shiryaev (1987)).

Proposition 4.1.3 *Suppose that*

$$\left\|(1/y^2)\cdot|n^{-1}Y^n - y|\right\|_{\mathcal{L}^1(\alpha)} = o_{P_0^n}(1)$$

is satisfied, where y is the function appearing in Condition 4.1.1, and introduce the Hilbert space $\mathbb{H}$ given by (4.1.6). Let C be a given subset of H, and suppose that $P_h^n \ll P_0^n$ and (4.1.7) hold for every $h \in C$. Then, it holds that for every $h \in C$

$$\log \frac{dP_h^n|\mathcal{F}_\tau^n}{dP_0^n|\mathcal{F}_\tau^n} = \Delta^n(h) - \frac{1}{2}\|h\|_{\mathbb{H}}^2 + \epsilon_n(h),$$

where

$$\Delta^n(h) = \left(\frac{h}{\sqrt{ny}}\right) * (\mu^n - \nu^{n,0})_\tau$$

and $\epsilon_n(h) = o_{P_0^n}(1)$. Furthermore, it also holds that $(\Delta^n(h_1), ..., \Delta^n(h_d)) \overset{P_0^n}{\Longrightarrow} N(0,\Sigma)$ where $\Sigma_{ij} = \langle h_i, h_j\rangle_{\mathbb{H}}$.

Proof. Since $|\log(1+x) - x + \frac{x^2}{2}| \le \frac{8}{3}x^3$ for all $x \in [-\frac{1}{2}, \frac{1}{2}]$, we have that for any $\varepsilon \in (0,1]$,

$$\left|\log \frac{dP_h^n|\mathcal{F}_\tau^n}{dP_0^n|\mathcal{F}_\tau^n} - \Delta^n(h) + D^n(h) + \tilde{D}^n(h)\right| \le \frac{16}{3}\varepsilon D^n(h)$$

whenever $\sup_{t,z}\frac{|h(t,z)|}{\sqrt{ny(t,z)}} \le \varepsilon$, where:

$$D^n(h) = \left|\frac{h}{2\sqrt{ny}}\right|^2 * \mu_\tau^n;$$
$$\tilde{D}^n(h) = \left|\frac{h}{2\sqrt{ny}}\right|^2 * \nu_\tau^{n,0}.$$

Notice that $\tilde{D}^n(h) \xrightarrow{P_0^n} \frac{1}{4}\|h\|_{\mathbb{H}}^2$. Also, using Lenglart's inequality, we have $D^n(h) - \tilde{D}^n(h) \xrightarrow{P_0^n} 0$. These facts imply the first assertion. The finite-dimensional convergence of $h \rightsquigarrow \Delta^n(h)$ follows easily from the martingale central limit theorem. □

STEP II: DIFFERENTIABILITY OF UNKNOWN PARAMETER. We consider the functional $(t,\psi) \rightsquigarrow A(\psi)_t$ as an unknown parameter taking values in the Banach space $(\ell^\infty([0,\tau]\times\Psi), \|\cdot\|_\infty)$. We denote by $\ell^\infty([0,\tau]\times\Psi)^*$ the dual space of $\ell^\infty([0,\tau]\times\Psi)$. Introducing a sequence of operators $A^n : H \to \ell^\infty([0,\tau]\times\Psi)$, which should be regarded as a local sequence of A, we aim to get a derivative operator $\dot{A} : H \to \ell^\infty([0,\tau]\times\Psi)$ with rate r_n ($=\sqrt{n}$ in the present case) which should satisfy

$$r_n\|A^n(h) - A^n(0) - \dot{A}(h)\|_\infty \to 0 \qquad \forall h \in H,$$

and its adjoint operator $\dot{A}^* : \ell^\infty([0,\tau]\times\Psi)^* \to \bar{H}$, where $\bar{H}$ is the completion of H in $\mathbb{H}$ (thus $\bar{H} = \mathbb{H}$ in the present case), which is determined by

$$\langle \dot{A}^* b^*, h\rangle_{\mathbb{H}} = b^*\dot{A}(h) \qquad \forall h \in H$$

for every $b^* \in \ell^\infty([0,\tau]\times\Psi)^*$.

Now, we define the sequence of operators $A^n : H \to \ell^\infty([0,\tau]\times\Psi)$ by

$$A^n(h)(\psi)_t = \int_{[0,t]\times E} \psi(s,z)\alpha_h^n(s,z)dt\lambda(dz).$$

Then the sequence A^n is differentiable with rate $\sqrt{n}$ and its derivative $\dot{A} : H \to \ell^\infty([0,\tau]\times \Psi)$ is given by

$$\dot{A}(h)(t,\psi) = \langle 1_{[0,t]}\psi, h\rangle_{\mathbb{H}}.$$

We denote by $\pi_{t,\psi} : \ell^\infty([0,\tau]\times\Psi) \to \mathbb{R}$ the projection on the (t,ψ)-coordinate, which belongs to $\ell^\infty([0,\tau]\times\Psi)^*$. The above formula shows that $\dot{A}^*\pi_{t,\psi} = 1_{[0,t]}\psi$ for every $(t,\psi) \in [0,\tau]\times\Psi$, and this means that the process $(t,\psi) \rightsquigarrow G_t^\psi$ appearing in the limit of Theorem 4.1.2 satisfies that

$$EG_t^\psi G_s^\phi = \langle \dot{A}^*\pi_{t,\psi}, \dot{A}^*\pi_{s,\phi}\rangle_{\mathbb{H}} \qquad \forall (t,\psi),(s,\phi) \in [0,\tau]\times\Psi.$$

Since the law of the process $(t,\psi) \rightsquigarrow G_t^\psi$ is characterized by its finite-dimensional distributions, we can conclude that it coincides with the bound of asymptotic efficiency (see Theorem 3.11.2 of van der Vaart and Wellner (1996)).

STEP III: ASYMPTOTIC EFFICIENCY. In order to discuss the asymptotic efficiency in the sense of the convolution theorem, it remains to show that the estimator $(t,\psi) \rightsquigarrow \widehat{A}^n(\psi)_t$ for $(t,\psi) \rightsquigarrow A^n(h)(\psi)_t$ is regular. The following is an easy consequence of Theorem 4.1.2 and Proposition 4.1.3 which implies the contiguity.

Proposition 4.1.4 *Suppose that a given class Ψ satisfies Condition 4.1.1 and the assumption (4.1.4) with $P^n = P_0^n$, and introduce the Hilbert space $\mathbb{H}$ given by (4.1.6). Let*

C be an arbitrary subset of H, and suppose that all assumptions of Proposition 4.1.3 are satisfied. Then, it holds that

$$\sqrt{n}(\widehat{A}^n - A^n(h)) \overset{P_h^n}{\Longrightarrow} G \qquad \text{in } \ell^\infty([0,\infty] \times \Psi) \qquad \forall h \in C,$$

where $(t,\psi) \rightsquigarrow G_t^\psi$ is the process appearing in the limit of Theorem 4.1.2.

Proof. The local asymptotic normality established in Proposition 4.1.3 implies that P_h^n and P_0^n are two-sided contiguous (see, e.g., Definition 3.10.1 and Example 3.10.6 of van der Vaart and Wellner (1996)), although $P_h^n \lhd P_0^n$ suffices for us. Hence we have that Condition 4.1.1 and the assumption (4.1.4) are satisfied with $P^n = P_h^n$ for all $h \in C$. Noting also that $\|A^n(h) - A\|_\infty \to \infty$ for every $h \in C$, we can obtain the assertion in the same way as Theorem 4.1.2. □

Consequently, the estimator $\widehat{A}^n$ has shown to be regular. It also holds for any bounded, continuous function $\ell : \ell^\infty([0,\tau] \times \Psi) \to [0,\infty)$ that

$$\sup_{I \subset C} \limsup_{n\to\infty} \sup_{h \in I} E_{h*}^n \ell\left(\sqrt{n}(\widehat{A}^n - A^n(h))\right) = E\ell(G), \tag{4.1.8}$$

where the supremum with respect to $I \subset C$ is taken over all finite subsets I of C. Summarizing the above discussion, we can conclude that:

Corollary 4.1.5 *Suppose that all assumptions of Proposition 4.1.4 are satisfied. If C is a convex cone in $\mathbb{H}$ such that its closed linear span coincides with $\mathbb{H}$, then the estimator $\widehat{A}^n$ for A^n is asymptotically efficient in the sense of the convolution and the locally asymptotic minimax theorems with respect to bounded, continuous, subconvex loss functions.*

See Theorems 3.11.2 and 3.11.5 of van der Vaart and Wellner (1996) for the convolution and the locally asymptotic minimax theorems. See also their Example 3.11.8 for the choice of loss functions in the latter theorem. In particular, when we choose a loss function $\ell : \ell^\infty([0,\tau] \times \Psi) \to [0,\infty)$ of the type $\ell(z) = \ell_0(\|z\|_\infty)$, where $\ell_0 : [0,\infty) \to [0,\infty)$ is a bounded, continuous, increasing function, their Theorem 3.11.5 says that: for any $T^n : \Omega^n \to \ell^\infty([0,\tau] \times \Psi)$ such that $T^n(t,\psi)$ is $\mathcal{F}_\tau^n$-measurable for every $(t,\psi) \in [0,\tau] \times \Psi$, it holds that

$$\sup_{I \subset C} \liminf_{n\to\infty} \sup_{h \in I} E_{h*}^n \ell\left(\sqrt{n}(T^n - A^n(h))\right) \geq E\ell(G),$$

where the supremum with respect to $I \subset C$ is taken over all finite subsets I of C. Recalling (4.1.8), we can see that the estimator $\widehat{A}^n$ achieves this bound.

4.2 Continuous Semimartingales with Non-linear Covariates

Let $(E, \mathcal{E}, \lambda)$ be a finite measure space; this is the state space of covariate processes in the following model. In the n-th statistical experiment, we consider k_n continuous, adapted processes $X^{n,i}$ defined on a filtered measurable space $(\Omega^n, \mathcal{F}^n, \mathbf{F}^n = (\mathcal{F}_t^n)_{t\in\mathbb{R}_+})$. Suppose that $X^{n,i}$'s are special semimartingales under the probability measure P^n on $(\Omega^n, \mathcal{F}^n)$, and that their canonical decompositions are given by

$$dX_t^{n,i} = \alpha(t, Z_t^{n,i})Y_t^{n,i}dt + Y_t^{n,i}dB_t^{n,i} \qquad \forall i = 1, ..., k_n,$$

where $\alpha(t,z)$ is an $\mathbb{R}$-valued $\mathfrak{B}(\mathbb{R}_+) \otimes \mathcal{E}$-measurable function, $t \rightsquigarrow Z_t^{n,i}$'s are E-valued predictable processes, $t \rightsquigarrow Y_t^{n,i}$'s are $\{0,1\}$-valued predictable processes, and $t \rightsquigarrow B_t^{n,i}$'s are orthogonal Brownian motions. It has implicitly been assumed that

$$\int_0^t |\alpha(s, Z_s^{n,i}(\omega))| Y_s^{n,i}(\omega)ds < \infty \qquad \forall\omega \in \Omega^n, \ \forall t \in \mathbb{R}_+$$

for every i. Notice that we do not assume the independence of $Z^{n,i}$'s and $Y^{n,i}$'s (thus $X^{n,i}$'s, too). We are interested in estimating the functional $(t, \psi) \rightsquigarrow A(\psi)_t$, which is given by

$$A(\psi)_t = \int_{[0,t]\times E} \psi(s,z)\alpha(s,z)ds\lambda(dz), \tag{4.2.1}$$

where ψ belongs to an appropriate class of measurable functions on $\mathbb{R}_+ \times E$.

4.2.1 Preliminaries

Let $E = \bigcup_m E_m^n$ be a partition of $\mathcal{E}$-measurable sets, that is at most countable. Set $\mathcal{E}^n = \sigma\{E_m^n : m = 1, 2, ...\}$. We denote $\tilde{\mathcal{P}}^n = \mathcal{P}^n \otimes \mathcal{E}$ and $\tilde{\mathcal{Q}}^n = \mathcal{P}^n \otimes \mathcal{E}^n$, where $\mathcal{P}^n$ is the predictable σ-field on $\Omega^n \times \mathbb{R}_+$. We introduce a kind of "generalized inverse" Y^{n-} defined by

$$Y^{n-}(\omega, t, z) = \sum_m \frac{1_{\{\overline{Y}_m^n(\omega,t)>0\}}\lambda(E_m^n)}{\overline{Y}_m^n(\omega,t)} 1_{\{z\in E_m^n\}}$$

with the convention $0/0 = 0$, where

$$\overline{Y}_m^n(\omega, t) = \sum_{i=1}^{k_n} Y_t^{n,i}(\omega)1_{\{Z_t^{n,i}(\omega)\in E_m^n\}}.$$

Here notice that

$$0 \le Y^{n-} \le \sup_m \lambda(E_m^n) < \infty. \tag{4.2.2}$$

Define also

$$I^n(\omega,t,z) = \sum_m 1_{\{\overline{Y}^n_m(\omega,t)>0\}} 1_{\{z\in E^n_m\}}.$$

We denote by $\overline{\alpha}^n$ the $\mathfrak{B}(\mathbb{R}_+)\otimes\mathcal{E}^n$-measurable majorant of α, and by $\underline{\alpha}^n$ the $\mathfrak{B}(\mathbb{R}_+)\otimes\mathcal{E}^n$-measurable minorant of α. By using the notation of Definition 2.1.4, they are given by:

$$\begin{aligned}\overline{\alpha}^n &= [\alpha]_{\mathfrak{B}(\mathbb{R}_+)\otimes\mathcal{E}^n,dt\lambda(dz)};\\ \underline{\alpha}^n &= -[-\alpha]_{\mathfrak{B}(\mathbb{R}_+)\otimes\mathcal{E}^n,dt\lambda(dz)}.\end{aligned}$$

The "generalized inverse" Y^{n-} will be useful through the following lemma.

Lemma 4.2.1 *For any $\mathbb{R}$-valued, $\tilde{\mathcal{Q}}^n$-measurable function W^n on $\Omega^n\times\mathbb{R}_+\times E$, it holds that*

$$\sum_{i=1}^{k_n}\int_0^t W^n(s,Z^{n,i}_s)Y^{n-}(s,Z^{n,i}_s)Y^{n,i}_s ds = \int_{[0,t]\times E} W^n(s,z)I^n(s,z)ds\lambda(dz) \qquad \forall t\in\mathbb{R}_+,$$

identically, provided the integrals are finite. Furthermore, it holds that

$$\left|\sum_{i=1}^{k_n}\int_0^t W^n(s,Z^{n,i}_s)\alpha(s,Z^{n,i}_s)Y^{n-}(s,Z^{n,i}_s)Y^{n,i}_s ds - \int_{[0,t]\times E} W^n(s,z)\alpha(s,z)I^n(s,z)ds\lambda(dz)\right|$$

$$\le \int_{[0,t]\times E}|W^n(s,z)|(\overline{\alpha}^n-\underline{\alpha}^n)(s,z)I^n(s,z)ds\lambda(dz) \qquad \forall t\in\mathbb{R}_+,$$

identically, provided the integrals are finite.

Proof. The first equality indeed holds since

$$\begin{aligned}&\sum_{i=1}^{k_n}\int_{[0,t]\times E} W^n(\omega,s,Z^{n,i}_s(\omega))Y^{n-}(\omega,s,Z^{n,i}_s(\omega))Y^{n,i}_s(\omega)ds\\ &= \int_{[0,t]\times E}\sum_m W^n(\omega,s,z^n_m)1_{\{\overline{Y}^n_m(\omega,s)>0\}}\lambda(E^n_m)ds\\ &= \int_{[0,t]\times E} W^n(\omega,s,z)I^n(\omega,s,z)ds\lambda(dz),\end{aligned}$$

where z^n_m is any point of E^n_m. The second inequality is an easy consequence of the first. □

Let a constant $\tau>0$ be given, and denote:

(4.2.3) $$\mathcal{L}^p = \mathcal{L}^p([0,\tau]\times E,\mathfrak{B}[0,\tau]\otimes\mathcal{E},dt\lambda(dz)) \qquad \forall p\in[1,\infty];$$

(4.2.4) $$\mathcal{L}^{p,n} = \mathcal{L}^p([0,\tau]\times E,\mathfrak{B}[0,\tau]\otimes\mathcal{E}^n,dt\lambda(dz)) \qquad \forall p\in[1,\infty].$$

We denote by $\|\cdot\|_{\mathcal{L}^p}$ the L^p-seminorm on $\mathcal{L}^p$. For every $n \in \mathbb{N}$, we introduce the mapping $\pi^n : \mathcal{L}^1 \to \mathcal{L}^{1,n}$ by

$$\pi^n(\psi)(t,z) = \sum_m \frac{1_{\{\lambda(E_m^n)>0\}}}{\lambda(E_m^n)} \int_{E_m^n} \psi(t,w)\lambda(dw) 1_{\{z\in E_m^n\}} \tag{4.2.5}$$

with the convention $0/0 = 0$.

Lemma 4.2.2 (i) *For every $p \in [1,\infty]$, it holds that $\pi^n \mathcal{L}^p \subset \mathcal{L}^{p,n}$, and the mapping $\pi^n : \mathcal{L}^p \to \mathcal{L}^{p,n}$ is linear.*
(ii) *If $\psi, \phi \in \mathcal{L}^1$ and $\psi \leq \phi$, then $\pi^n(\psi) \leq \pi^n(\phi)$.*
(iii) *For every $p \in [1,\infty)$, it holds for any $\psi \in \mathcal{L}^1$ and any function $f : [0,\tau]\times E \to [0,\infty)$ which is $\mathfrak{B}[0,\tau] \otimes \mathcal{E}^n$-measurable that*

$$\int_{[0,\tau]\times E} |\pi^n(\psi)(t,z)|^p f(t,z) dt\lambda(dz) \leq \int_{[0,\tau]\times E} |\psi(t,z)|^p f(t,z) dt \lambda(dz),$$

provided the right hand side is finite.
(iv) *It holds for any $\psi \in \mathcal{L}^\infty$ that $\|\pi^n(\psi)\|_{\mathcal{L}^\infty} \leq \|\psi\|_{\mathcal{L}^\infty}$.*

Proof. The assertions (i), (ii) and (iv) are trivial, and (iii) follows from that

$$\begin{aligned} |\pi^n(\psi)(t,z)|^p &= \sum_m \frac{1_{\{\lambda(E_m^n)>0\}}}{\lambda(E_m^n)^p} \left| \int_{E_m^n} \psi(t,w)\lambda(dw)\right|^p 1_{\{z\in E_m^n\}} \\ &\leq \sum_m \frac{1_{\{\lambda(E_m^n)>0\}}}{\lambda(E_m^n)} \int_{E_m^n} |\psi(t,w)|^p \lambda(dw) 1_{\{z\in E_m^n\}}. \end{aligned}$$

□

4.2.2 Asymptotic Normality

Let $\tau > 0$ be a constant, and let Ψ be a countable class of $\mathfrak{B}[0,\tau] \otimes \mathcal{E}$-measurable functions on $[0,\tau] \times E$. We will always assume

$$\sup_{\psi\in\Psi} \int_{[0,\tau]\times E} |\psi(t,z)|(|\overline{\alpha}^n| \vee |\underline{\alpha}^n|)(t,z) dt\lambda(dz) < \infty; \tag{4.2.6}$$

then, the functional $(t,\psi) \rightsquigarrow A(\psi)_t$ defined by (4.2.1) can be thought as an unknown parameter taking values in the space $\ell^\infty([0,\tau] \times \Psi)$. The main step in our approach is to estimate the modified unknown parameter $(t,\psi) \rightsquigarrow \widetilde{A}^n(\psi)_t$ given by

$$\widetilde{A}^n(\psi)_t(\omega) = \int_{[0,t]\times E} \pi^n(\psi)(s,z) I^n(\omega,s,z) ds\lambda(dz) \qquad \forall (t,\psi) \in [0,\tau] \times \Psi.$$

It follows from Lemma 4.2.2 that this also takes values in $\ell^\infty([0,\tau]\times\Psi)$ under the assumption (4.2.6). We will discuss later under which conditions this $\widetilde{A}^n$ is indeed "close" to A. We propose the estimator $(t,\psi)\rightsquigarrow \widehat{A}^n(\psi)_t$ given by

$$\widehat{A}^n(\psi)_t = \sum_{i=1}^{k_n}\int_0^t \pi^n(\psi)(s,Z_s^{n,i})Y^{n-}(s,Z_s^{n,i})dX_s^{n,i} \qquad \forall(t,\psi)\in[0,\tau]\times\Psi.$$

The stochastic integral appearing above is well-defined if

$$\text{(4.2.7)}\quad \sup_{t\in[0,\tau]}|\pi^n(\psi)(t,Z_t^{n,i}(\omega))|Y^{n-}(\omega,t,Z_t^{n,i}(\omega))<\infty \qquad \forall\omega\in\Omega^n,\ \forall i=1,...,k_n$$

(see Theorem I.4.31 of Jacod and Shiryaev (1987)). We will always assume (4.2.7), and define $t\rightsquigarrow \widehat{A}^n(\psi)_t$ by any version of the stochastic integral, which is unique up to a P^n-negligible set. It is immediate from the definitions of π^n and Y^{n-} that a sufficient condition for (4.2.7) is that

$$\sup_{t\in[0,\tau]}\sup_m\int_{E_m^n}|\psi(t,z)|\lambda(dz)<\infty.$$

However, it is still unclear even under this condition that $(t,\psi)\rightsquigarrow\widehat{A}^n(\psi)_t$ takes values in $\ell^\infty([0,\tau]\times\Psi)$. This requirement will be shown to be fulfilled under (4.2.2) and (4.2.8) given below, by means of Theorem 2.4.3.

Condition 4.2.3 *For some $p,q\in[1,\infty]$ such that $(1/p)+(1/q)=1$, and a $\mathfrak{B}[0,\tau]\otimes\mathcal{E}$-measurable function $y=y(t,z)$, which is bounded away from zero, it holds that $\Psi\subset\mathcal{L}^{2p}$, and that $\|\cdot\|_{\mathcal{L}^{2p}}$ defines a proper metric on Ψ, and that:*

$$\text{(4.2.8)}\qquad \int_0^1\sqrt{\log N(\Psi,\|\cdot\|_{\mathcal{L}^{2p}};\varepsilon)}d\varepsilon<\infty;$$

$$\text{(4.2.9)}\qquad \|nY^{n-}\|_{\mathcal{L}^q}=O_{P^n}(1);$$

$$\text{(4.2.10)}\qquad \Big\||\pi^n(\psi)|^2\cdot|nY^{n-}-(1/y)|\Big\|_{\mathcal{L}^1}=o_{P^n}(1)\qquad\forall\psi\in\Psi;$$

$$\text{(4.2.11)}\qquad \|\{\pi^n(\psi)\pi^n(\phi)-\psi\phi\}\cdot(1/y)\|_{\mathcal{L}^1}\to0\qquad\forall\psi,\phi\in\Psi;$$

$$\text{(4.2.12)}\qquad \sup_{\psi\in\Psi}\|\pi^n(\psi)\cdot(\overline{\alpha}^n-\underline{\alpha}^n)\cdot I^n\|_{\mathcal{L}^1}=o_{P^n}(n^{-1/2}).$$

Compare this with Condition 4.1.1; the conditions (4.2.8), (4.2.9) and (4.2.10) correspond to (4.1.1) (4.1.2), and (4.1.3), respectively, while (4.2.11) and (4.2.12) are needed for the present situation.

Theorem 4.2.4 *Suppose that a given countable class Ψ satisfies the assumptions (4.2.6) and (4.2.7) and Condition 4.2.3. Then, the random elements $(t,\psi) \rightsquigarrow \widehat{A}^n(\psi)_t$ and $(t,\psi) \rightsquigarrow \widetilde{A}^n(\psi)_t$ take values in $\ell^\infty([0,\tau]\times\Psi)$, P^n-almost surely, and it holds that $\sqrt{n}(\widehat{A}^n - \widetilde{A}^n) \overset{P^n}{\Longrightarrow} G$ in $\ell^\infty([0,\tau]\times\Psi)$, where $(t,\psi) \rightsquigarrow G_t^\psi$ is a zero-mean Gaussian process such that*

$$EG_t^\psi G_s^\phi = \int_{[0,t\wedge s]\times E} \frac{\psi(u,z)\phi(u,z)}{y(u,z)} du\lambda(dz) \qquad \forall(t,\psi),(s,\phi)\in[0,\tau]\times\Psi$$

and that almost all paths are uniformly ρ_2-continuous on $[0,\tau]\times\Psi$, where

$$\rho_2((t,\psi),(s,\phi)) = \sqrt{E|G_t^\psi - G_s^\phi|^2} \qquad \forall(t,\psi),(s,\phi)\in[0,\tau]\times\Psi.$$

Proof. Observe that

$$\sqrt{n}(\widehat{A}^n(\psi)_t - \widetilde{A}^n(\psi)_t) = M_t^{n,\psi} + N_t^{n,\psi},$$

where:

$$\begin{aligned} M_t^{n,\psi} &= \sqrt{n}\sum_{i=1}^{k_n}\int_0^t \pi^n(\psi)(s,Z_s^{n,i})Y^{n-}(s,Z_s^{n,i})Y_s^{n,i}dB_s^{n,i}; \\ N_t^{n,\psi} &= \sqrt{n}\sum_{i=1}^{k_n}\int_0^t \pi^n(\psi)(s,Z_s^{n,i})\alpha(s,Z_s^{n,i})Y^{n-}(s,Z_s^{n,i})Y_s^{n,i}ds. \end{aligned}$$

It follows from Lemmas 4.2.1 and 4.2.2 that the term $N^{n,\psi}$ is well-defined under the assumption (4.2.6), and that

$$\left|N_t^{n,\psi} - \sqrt{n}\widetilde{A}^n(\psi)_t\right| \le \sqrt{n}\int_{[0,t]\times E} |\pi^n(\psi)(s,z)|(\overline{\alpha}^n - \underline{\alpha}^n)(s,z)I^n(s,z)ds\lambda(dz),$$

which converges in P^n-probability to zero uniformly in $(t,\psi)\in[0,\tau]\times\Psi$ by (4.2.12).

On the other hand, $t \rightsquigarrow M_t^{n,\psi}$ is a continuous local martingale under the assumption (4.2.7). Since $B^{n,i}$'s are orthogonal Brownian motions, and since $|Y^{n,i}|^2 = Y^{n,i}$, we obtain from Lemma 4.2.1 that

$$\langle M^{n,\psi}, M^{n,\phi}\rangle_t = n\int_{[0,t]\times E} \pi^n(\psi)(s,z)\pi^n(\phi)(s,z)Y^{n-}(s,z)ds\lambda(dz).$$

It follows from Lemma 4.2.2 that

$$\begin{aligned} &\langle M^{n,\psi} - M^{n,\phi}, M^{n,\psi} - M^{n,\phi}\rangle_\tau \\ &= n\left\||\pi^n(\psi) - \pi^n(\phi)|^2 Y^{n-}\right\|_{\mathcal{L}^1} \\ &\le n\left\||\psi - \phi|^2 Y^{n-}\right\|_{\mathcal{L}^1} \\ &\le n\||\psi-\phi|^2\|_{\mathcal{L}^p}\cdot\|Y^{n-}\|_{\mathcal{L}^q} \\ &= n\|\psi-\phi\|_{\mathcal{L}^{2p}}^2\cdot\|Y^{n-}\|_{\mathcal{L}^q}. \end{aligned}$$

Thus we have that the quadratic $\|\cdot\|_{\mathcal{L}^{2p}}$-modulus of $M^n = (M^{n,\psi}|\psi \in \Psi)$ is bounded by $\sqrt{n\|Y^{n-}\|_{\mathcal{L}^q}}$. Hence, Theorem 2.4.3 implies that almost all paths of $(t,\psi) \rightsquigarrow M_t^{n,\psi}$ take values in $\ell^\infty([0,\tau] \times \Psi)$, and thus so do those of $(t,\psi) \rightsquigarrow \widehat{A}^n(\psi)_t$.

In order to derive the weak convergence of the processes $(t,\psi) \rightsquigarrow M_t^{n,\psi}$, we will apply Theorem 3.4.4. The above computation of the quadratic $\|\cdot\|_{\mathcal{L}^{2p}}$-modulus together with (4.2.9), yields **[ME]**. Next, notice that

$$\begin{aligned}
&\left|\int_{[0,t]\times E} \pi^n(\psi)(s,z)\pi^n(\phi)(s,z)\left\{nY^{n-}(s,z) - \frac{1}{y(s,z)}\right\} ds\lambda(dz)\right| \\
&\quad\leq \sqrt{\int_{[0,\tau]\times E} |\pi^n(\psi)(s,z)|^2 \left|nY^{n-}(s,z) - \frac{1}{y(s,z)}\right| ds\lambda(dz)} \\
&\quad\quad\times \sqrt{\int_{[0,\tau]\times E} |\pi^n(\psi)(s,z)|^2 \left|nY^{n-}(s,z) - \frac{1}{y(s,z)}\right| ds\lambda(dz)}
\end{aligned}$$

and that

$$\begin{aligned}
&\left|\int_{[0,t]\times E} \{\pi^n(\psi)(s,z)\pi^n(\phi)(s,z) - \psi(s,z)\phi(s,z)\} \frac{1}{y(s,z)} ds\lambda(dz)\right| \\
&\quad\leq \|\{\pi^n(\psi)\pi^n(\phi) - \psi\phi\}\cdot(1/y)\|_{\mathcal{L}^1}.
\end{aligned}$$

Thus (4.2.10) and (4.2.11) imply **[C2]**. This finishes the proof. □

In order to derive the weak convergence of $\sqrt{n}(\widehat{A}^n - A)$ rather than $\sqrt{n}(\widehat{A}^n - \widetilde{A}^n)$, we have to show that $\sqrt{n}(\widetilde{A}^n - A) \stackrel{P^n}{\Longrightarrow} 0$ in $\ell^\infty([0,\tau] \times \Psi)$. For this purpose, it suffices to check the following:

(4.2.13) $$\sup_{\psi\in\Psi} \|(\pi^n(\psi) - \psi)\cdot\alpha\|_{\mathcal{L}^1} = o(n^{-1/2});$$

(4.2.14) $$\sup_{\psi\in\Psi} \|\psi\cdot(1 - I^n)\cdot\alpha\|_{\mathcal{L}^1} = o_{P^n}(n^{-1/2}).$$

Example: Euclidean covariates

Set $E = [0,1]^d$, and equip it with the Lebesgue measure. Suppose that $(t,z) \rightsquigarrow \alpha(t,z)$ is continuous on $[0,\tau] \times [0,1]^d$ with respect to the $(d+1)$-dimensional Euclidean metric. Consider the class of indicator functions $\Psi = \{1_{[0,u_1]\times\cdots\times[0,u_d]}(z) : u \in [0,1]^d\}$. This problem in the case of $d = 1$ was studied by McKeague and Utikal (1990), based on a classical criterion of tightness of sequences of stochastic processes with finite-dimensional parameters. Among the assumptions appearing above, the entropy condition (4.2.8) is satisfied for any $p < \infty$. If we choose a partition $[0,1]^d = \bigcup_m E_m^n$ of d-dimensional rectangles with side length at most b_n, where $b_n n^{1/2} \to 0$, then the weak convergence of $\sqrt{n}(\widehat{A}^n - A)$ in $\ell^\infty([0,\tau] \times [0,1]^d)$ follows from that:

$$\|nY^{n-}\|_{\mathcal{L}^q} = O_{P^n}(1) \qquad \text{for some } q > 1;$$
$$\|nY^{n-} - (1/y)\|_{\mathcal{L}^1} = o_{P^n}(1);$$
$$\|(1 - I^n) \cdot \alpha\|_{\mathcal{L}^1} = o_{P^n}(n^{-1/2}).$$

The first condition above refines (A3) of McKeague and Utikal (1990), which is corresponding to the case of $q = 3$.

4.2.3 Asymptotic Efficiency

Let us discuss the asymptotic efficiency of the estimator $(t, \psi) \rightsquigarrow \widehat{A}^n(\psi)_t$ along the general theory exposited in Chapter 3.11 of van der Vaart and Wellner (1996), again. We set:

$$(4.2.15) \qquad \begin{aligned} \mathbb{H} &= L^2([0,\tau] \times E, \mathfrak{B}[0,\tau] \otimes \mathcal{E}, \tfrac{1}{y(t,z)} dt\lambda(dz)); \\ H &= L^\infty([0,\tau] \times E, \mathfrak{B}[0,\tau] \otimes \mathcal{E}, dt\lambda(dz)); \\ H^n &= L^\infty([0,\tau] \times E, \mathfrak{B}[0,\tau] \otimes \mathcal{E}^n, dt\lambda(dz)). \end{aligned}$$

Here, the function y is the one which appears in Condition 4.2.3. We equip $\mathbb{H}$ with the usual L^2-inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$. Notice that H is a linear subspace of the Hilbert space $\mathbb{H}$ since $1/y$ is assumed to be bounded. Let $\mathbf{P}^n = \{P_h^n : h \in H\}$ be a family of probability measures on $(\Omega^n, \mathcal{F}^n)$ indexed by H. Define the mapping $\pi^n : H \to H^n$ by (4.2.5). Hereafter, we denote by $\pi^n(h)$ any function of the equivalent elements in H^n.

Suppose that the canonical decompositions of special semimartingales $X^{n,i}$ under the probability measure P_h^n are given by

$$(4.2.16) \qquad dX_t^{n,i} = \alpha_h^n(t, Z_t^{n,i}) Y_t^{n,i} dt + Y_t^{n,i} dB_t^{n,h,i} \qquad \forall i = 1, ..., k_n,$$

where $\alpha_h^n = \alpha_h^n(t,z)$ is defined by

$$(4.2.17) \qquad \alpha_h^n = \alpha + n^{-1/2} \pi^n(h/y),$$

and where $t \rightsquigarrow B_t^{n,h,i}$'s are orthogonal Brownian motions on $[0,\tau]$ under P_h^n. We should first see that the local model (4.2.16) is "well-defined" in the sense that it does not depend on the choice of a version of $\pi^n(h/y) \in H^n$. To see this, notice that, if $f, g \in \mathcal{L}^\infty([0,\tau] \times E, \mathfrak{B}[0,\tau] \otimes \mathcal{E}^n, dt\lambda(dz))$ such that $f(t,z) = g(t,z)$ for $dt\lambda(dz)$-almost all (t,z), then it holds that for every $\omega \in \Omega^n$

$$f(t, Z_t^{n,i}(\omega)) = g(t, Z_t^{n,i}(\omega)) \qquad dt\text{-almost all } t.$$

Thus, for every $\omega \in \Omega^n$, the function $t \rightsquigarrow \alpha_h^n(t, Z_t^{n,i}(\omega))$ is uniquely determined by (4.2.17) up to a negligible set with respect to the Lebesgue measure on $[0, \tau]$, not depending on the choice of a version of $\pi^n(h/y) \in H^n$. Hence the model (4.2.16) is well-defined, and in particular, it holds that $\alpha_0^n(t, Z_t^{n,i}(\omega)) = \alpha(t, Z_t^{n,i}(\omega))$ for almost all t.

STEP I: LOCAL ASYMPTOTIC NORMALITY. It is well-known that, under some conditions, a version of the log-likelihood ratio is given by

$$\text{(4.2.18)} \qquad \log \frac{dP_h^n|\mathcal{F}_\tau^n}{dP_0^n|\mathcal{F}_\tau^n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{k_n} \int_0^\tau \pi^n(h/y)(t, Z_t^{n,i}) Y_t^{n,i} dB_t^{n,0,i} - \frac{1}{2n} \sum_{i=1}^{k_n} \int_0^\tau |\pi^n(h/y)(t, Z_t^{n,i})|^2 Y_t^{n,i} dt$$

(see, e.g., Theorem III.5.32 of Jacod and Shiryaev (1987)). Again, this representation does not depend on the choice of a version of $\pi^n(h/y) \in H^n$.

Proposition 4.2.5 *Let C be an arbitrary subset of H. Suppose that the function y appearing in Condition 4.2.3 satisfies that*

$$\text{(4.2.19)} \qquad y \in \mathcal{L}^1 \quad \text{and} \quad \left\| \frac{I^n}{nY^{n-}} - y \right\|_{\mathcal{L}^1} = o_{P_0^n}(1)$$

with the convention $0/0 = 0$, and that

$$\text{(4.2.20)} \qquad \|\pi^n(h_1/y)\pi^n(h_2/y)y - (h_1h_2/y)\|_{\mathcal{L}^1} \to 0 \qquad \forall h_1, h_2 \in C,$$

and introduce the Hilbert space $\mathbb{H}$ given by (4.2.15). Suppose also that $P_h^n \ll P_0^n$ and (4.2.18) hold for every $h \in C$. Then, it holds that for every $h \in C$

$$\log \frac{dP_h^n|\mathcal{F}_\tau^n}{dP_0^n|\mathcal{F}_\tau^n} = \Delta^n(h) - \frac{1}{2}\|h\|_{\mathbb{H}}^2 + \epsilon_n(h),$$

where

$$\Delta^n(h) = \frac{1}{\sqrt{n}} \sum_{i=1}^{k_n} \int_0^\tau \pi^n(h/y)(t, Z_t^{n,i}) Y_t^{n,i} dB_t^{n,0,i}$$

and $\epsilon_n(h) = o_{P_0^n}(1)$. Furthermore, it also holds that $(\Delta^n(h_1), ..., \Delta^n(h_d)) \stackrel{P_0^n}{\Longrightarrow} N(0, \Sigma)$ where $\Sigma_{ij} = \langle h_i, h_j \rangle_{\mathbb{H}}$.

Proof. Notice that $\Delta^n(h)$ is the terminal variable of the continuous local martingale $M^{n,h}$ given by

$$\begin{aligned} M_t^{n,h} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{k_n} \int_0^t \pi^n(h/y)(s, Z_s^{n,i}) I^n(s, Z_s^{n,i}) Y_s^{n,i} dB_s^{n,0,i} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{k_n} \int_0^t \pi^n(h/y)(s, Z_s^{n,i}) \frac{I^n(s, Z_s^{n,i})}{Y^{n-}(s, Z_s^{n,i})} Y^{n-}(s, Z_s^{n,i}) Y_s^{n,i} dB_s^{n,0,i}. \end{aligned}$$

It follows from Lemma 4.2.1 that

$$\langle M^{n,h_1}, M^{n,h_2}\rangle_\tau = \int_{[0,\tau]\times E} \pi^n(h_1/y)(t,z)\pi^n(h_2/y)(t,z)\frac{I^n(t,z)}{nY^{n-}(t,z)}dt\lambda(dz).$$

Since

$$\left|\int_{[0,\tau]\times E} \pi^n(h_1/y)(t,z)\pi^n(h_2/y)(t,z)\left\{\frac{I^n(t,z)}{nY^{n-}(t,z)} - y(t,z)\right\}dt\lambda(dz)\right|$$
$$\leq \|\pi^n(h_1/y)\|_{\mathcal{L}^\infty}\cdot\|\pi^n(h_2/y)\|_{\mathcal{L}^\infty}\cdot\left\|\frac{I^n}{nY^{n-}} - y\right\|_{\mathcal{L}^1},$$

the finite-dimensional convergence of $h \rightsquigarrow \Delta^n(h)$ follows from (4.2.19) and (4.2.20) by means of the martingale central limit theorem.

On the other hand, it follows again from Lemma 4.2.1 that

$$\frac{1}{n}\sum_{i=1}^{k_n}\int_0^\tau |\pi^n(h/y)(t,Z_t^{n,i})|^2 Y_t^{n,i}dt = \int_{[0,\tau]\times E}|\pi^n(h/y)(t,z)|^2\frac{I^n(t,z)}{nY^{n-}(t,z)}dt\lambda(dz),$$

which converges in P^n-probability to $\|h\|_{\mathbb{H}}^2$ by (4.2.19) and (4.2.20). □

STEP II: DIFFERENTIABILITY OF UNKNOWN PARAMETER. The discussion here is similar to that at STEP II of Subsection 4.1.2. Recall the first paragraph there (we use exactly the same notation).

Now, we define the sequence of operators $A^n : H \to \ell^\infty([0,\tau]\times\Psi)$ by

$$A^n(h)(\psi)_t = \int_{[0,t]\times E}\psi(s,z)\left(\alpha(s,z) + n^{-1/2}\pi^n(h/y)(s,z)\right)ds\lambda(dz).$$

Under the assumption that

$$\sup_{(t,\psi)\in[0,\tau]\times\Psi}\left|\langle 1_{[0,t]}\psi, \pi^n(h/y)y - h\rangle_{\mathbb{H}}\right| \to 0 \qquad \forall h\in H, \tag{4.2.21}$$

the sequence A^n is differentiable with rate $\sqrt{n}$ and its derivative $\dot{A} : H \to \ell^\infty([0,\tau]\times\Psi)$ is given by

$$\dot{A}(h)(t,\psi) = \langle 1_{[0,t]}\psi, h\rangle_{\mathbb{H}}.$$

We denote by $\pi_{t,\psi} : \ell^\infty([0,\tau]\times\Psi) \to \mathbb{R}$ the projection to the (t,ψ)-coordinate, which belongs to $\ell^\infty([0,\tau]\times\Psi)^*$ (do not confuse this with the mapping π^n given by (4.2.5)). The above formula shows that $\dot{A}^*\pi_{t,\psi} = 1_{[0,t]}\psi$ for every $(t,\psi)\in[0,\tau]\times\Psi$. By the same reason as STEP II of Subsection 4.1.2, the law of the limit process $(t,\psi)\rightsquigarrow G_t^\psi$ appearing in Theorem 4.2.4 coincides with the bound of asymptotic efficiency.

STEP III: ASYMPTOTIC EFFICIENCY. Let us show the regularity of the estimator $(t,\psi)\rightsquigarrow \widehat{A}^n(\psi)_t$ for $(t,\psi)\rightsquigarrow A^n(h)(\psi)_t$.

Proposition 4.2.6 *Let C be an arbitrary subset of H. Suppose that a given countable class Ψ satisfies the assumptions (4.2.6), (4.2.7), (4.2.13) and (4.2.14), and Condition 4.2.3 for $P^n = P_0^n$, and introduce the Hilbert space $\mathbb{H}$ given by (4.2.15). Suppose also that all assumptions of Proposition 4.2.5 are satisfied. Then, the random element $(t,\psi) \rightsquigarrow \widehat{A}^n(\psi)_t$ takes values in $\ell^\infty([0,\tau] \times \Psi)$, P_h^n-almost surely, for every $h \in C$, and it holds that*

$$\sqrt{n}(\widehat{A}^n - A^n(h)) \overset{P_h^n}{\Longrightarrow} G \qquad \text{in } \ell^\infty([0,\tau]\times\Psi) \qquad \forall h \in C,$$

where $(t,\psi) \rightsquigarrow G_t^\psi$ is the process appearing in the limit of Theorem 4.2.4.

Proof. In view of the contiguity, a consequence of Proposition 4.2.5, all assumptions concerning convergence in P_0^n-probability hold also in P_h^n-probability for every $h \in C$. Thus the assertion can be proved in the same way as Theorem 4.2.4. □

Notice that the assumptions (4.2.20) and (4.2.21) follow from that

$$||\pi^n(h/y)y - h||_{\mathbb{H}} \to 0 \qquad \forall h \in H.$$

Summarizing the above discussion, we can get the asymptotic efficiency of the estimator $\widehat{A}^n$ in the same fashion as Corollary 4.1.5, under the assumption that C is a convex cone in $\mathbb{H}$ such that its closed linear span coincides with $\mathbb{H}$.

4.3 Counting Processes with Non-linear Covariates

Let $(E,\mathcal{E})$ be a Blackwell space on which a σ-finite measure λ is defined; this is the state space of covariate processes in the following model. In the n-th statistical experiment, we consider k_n adapted point processes on $\mathbb{R}_+$, namely $N^{n,i}$, $i = 1,...,k_n$, defined on a filtered measurable space $(\Omega^n, \mathcal{F}^n, \mathbf{F}^n = (\mathcal{F}_t^n)_{t\in\mathbb{R}_+})$; we then denote $T_j^{n,i} = \inf\{t \in \mathbb{R}_+ : N_t^{n,i} = j\}$ for every $j \in \mathbb{N}$ (see page 34 of Jacod and Shiryaev (1987)). Suppose that the predictable compensator of $N^{n,i}$ under the probability measure P^n on $(\Omega^n,\mathcal{F}^n)$ is given by

$$\alpha(t, Z_t^{n,i})Y_t^{n,i}dt,$$

where $\alpha(t,z)$ is a $[0,\infty)$-valued $\mathfrak{B}(\mathbb{R}_+)\otimes\mathcal{E}$-measurable function, $t \rightsquigarrow Z_t^{n,i}$ is an E-valued predictable process, and $t \rightsquigarrow Y_t^{n,i}$ is a $\{0,1\}$-valued predictable process. Let $\tau > 0$ be a constant, and suppose that we can observe the point processes, the processes $t \rightsquigarrow Y_t^{n,i}$, and the covariate processes $t \rightsquigarrow Z_t^{n,i}$ on the random sets $\{t \in [0,\tau] : Y_t^{n,i}(\omega) = 1\}$. The goal of this section is to estimate the unknown parameter $(t,\psi) \rightsquigarrow A(\psi)_t = \int_{[0,t]\times E} \psi(s,z)\alpha(s,z)ds\lambda(dz)$ where ψ's are appropriate functions.

We analyze this problem by using the E-valued multivariate point processes

$$\mu^n(dt,dz) = \sum_{i=1}^{k_n}\sum_j \varepsilon_{(T_j^{n,i}, Z^{n,i}_{T_j^{n,i}})}(dt,dz)$$

which has the predictable compensator

$$\nu^n(dt,dz) = \sum_{i=1}^{k_n} \alpha(t,z)Y_t^{n,i}\varepsilon_{Z_t^{n,i}}(dz)dt.$$

Here, we suppose that $T_j^{n,i} \neq T_{j'}^{n,i'}$ whenever $i \neq i'$; then the basic requirement that $\mu^n(\{t\}\times E) \leq 1$ is indeed satisfied. The approach which we take here is quite close to that in the preceding section.

4.3.1 Preliminaries

In the same way as Section 4.2.1, we consider a partition $E = \bigcup_m E_m^n$ of $\mathcal{E}$-measurable sets, which is at most countable, such that $\sup_m \lambda(E_m^n) < \infty$ (recall (4.2.2)). Introduce $\mathcal{E}^n$, $\tilde{\mathcal{P}}^n$, $\tilde{\mathcal{Q}}^n$, Y^{n-}, I^n, $\overline{\alpha}^n$ and $\underline{\alpha}^n$ given there. We then have Lemmas 4.2.1 and 4.2.2. For a given $\tilde{\mathcal{P}}^n$-measurable function W^n, we denote:

$$A(W^n)_t = \int_{[0,t]\times E} W^n(s,z,x)\alpha(s,z)ds\lambda(dz);$$

$$\overline{A}^n(W^n)_t = \int_{[0,t]\times E} W^n(s,z,x)\overline{\alpha}^n(s,z)ds\lambda(dz).$$

It follows from Lemma 4.2.1 that for any $\tilde{\mathcal{Q}}^n$-measurable function H^n the following two inequalities hold provided the integrals are finite:

$$\begin{aligned}(4.3.1)\qquad & \left|H^nY^{n-} * \nu_t^n - A(H^nI^n)_t\right| \\ & \leq \int_{[0,t]\times E} |H^n(s,z)|(\overline{\alpha}^n - \underline{\alpha}^n)(s,z)I^n(s,z)ds\lambda(dz);\end{aligned}$$

$$(4.3.2)\qquad \left|H^nY^{n-} * \nu_t^n\right| \leq \overline{A}^n(|H^n|I^n)_t.$$

Let Ψ be an arbitrary set. We will deal with a family $\{K^{n,\psi} : \psi \in \Psi\}$ of $\tilde{\mathcal{Q}}^n$-measurable functions which satisfies that

(4.3.3) the process $t \rightsquigarrow (|\overline{K}^nY^{n-}| \vee |\overline{K}^nY^{n-}|^2) * \nu_t^n$ is locally integrable

where

$$\overline{K}^n = \left[\sup_{\psi\in\Psi}|K^{n,\psi}|\right]_{\tilde{\mathcal{Q}}^n, M_{\nu^n}^{P^n}}$$

(cf., Definition 2.2.1). Then, for every $\psi \in \Psi$, the process $t \rightsquigarrow X_t^{n,\psi}$ defined by

$$X_t^{n,\psi} = K^{n,\psi} Y^{n-} * (\mu^n - \nu^n)_t \tag{4.3.4}$$

is an $\mathbb{R}$-valued locally square-integrable martingale with bounded variation. As an easy consequence of (4.3.2), we have that a sufficient condition for (4.3.3) is

$$\text{the process} \quad t \rightsquigarrow \overline{A}^n(|\overline{K}^n| \vee |\overline{K}^n|^2 \cdot I^n)_t \quad \text{is locally integrable} \tag{4.3.5}$$

(recall also (4.2.2)). The following lemma gives some tractable conditions to ensure the weak convergence of the $\ell^\infty([0,\tau] \times \Psi)$-valued random element $(t, \psi) \rightsquigarrow X_t^{n,\psi}$ by using Theorem 3.2.4.

Lemma 4.3.1 *Let $\tau > 0$ be a constant, and let S be a finite or dense subset of $[0, \tau]$ such that $\tau \in S$. Let $\{K^{n,\psi} : \psi \in \Psi\}$ be a family of $\tilde{\mathcal{Q}}^n$-measurable functions satisfying (4.3.3), and consider $X^n = (X_t^{n,\psi} | (t, \psi) \in [0, \tau] \times \Psi)$ defined by (4.3.4). When* Card$(\Psi) < \infty$, *suppose that the following* (i), (ii) *and* (iii) *and satisfied: when* Card$(\Psi) = \infty$, *suppose that the following* (i), (ii), (iii) *and* (iv) *are satisfied:*
(i) $\int_{[0,\tau]\times E} |\overline{K}^n(t,z)|^2 Y^{n-}(t,z)(\overline{\alpha}^n - \underline{\alpha}^n)(t,z) dt \lambda(dz) \xrightarrow{P^n} 0$;
(ii) $A\left(|\overline{K}^n|^2 Y^{n-} 1_{\{\overline{K}^n Y^{n-} > \varepsilon\}}\right)_\tau \xrightarrow{P^n} 0 \quad \forall \varepsilon > 0$;
(iii) $A(K^{n,\psi} K^{n,\phi} Y^{n-})_t \xrightarrow{P^n} C_t^{(\psi,\phi)} \quad \forall t \in S \ \forall (\psi, \phi) \in \Psi^2$, *where $\{C_t^{(\psi,\phi)}\}$ is a family of constants satisfying (3.2.3);*
(iv) *there exists a DFP $\Pi = \{\Psi(\varepsilon; k) : 1 \le k \le N_\Pi(\varepsilon)\}_{\varepsilon \in (0,1]}$ of Ψ such that*

$$\sup_{\varepsilon \in (0,1] \cap \mathbb{Q}} \max_{1 \le k \le N_\Pi(\varepsilon)} \frac{\overline{A}^n(|K^n(\Psi(\varepsilon; k))|^2 Y^{n-})_\tau}{\varepsilon^2} = O_{P^n}(1) \quad \textit{and} \quad \int_0^1 \sqrt{\log N_\Pi(\varepsilon)} d\varepsilon < \infty,$$

where

$$K^n(\Psi') = \left[\sup_{\psi,\phi \in \Psi'} |K^{n,\psi} - K^{n,\phi}| \right]_{\tilde{\mathcal{Q}}^n, \overline{A}^n(dt,dz)} \qquad \forall \Psi' \subset \Psi.$$

Then, it holds that $X^n \stackrel{P^n}{\Longrightarrow} X$ in $\ell^\infty(S \times \Psi)$, where $(t, \psi) \rightsquigarrow X_t^\psi$ is a zero-mean Gaussian process such that $E X_t^\psi X_s^\phi = C_{t \wedge s}^{(\psi,\phi)}$. Furthermore, the formula (3.2.4) defines a semi-metric ρ_2 on $S \times \Psi$ such that $S \times \Psi$ is totally bounded with respect to ρ_2, and that almost all paths of X are uniformly ρ_2-continuous. When S is dense in $[0, \tau]$, the space $S \times \Psi$ appearing in the conclusion can be replaced by $[0, \tau] \times \Psi$.

Proof. By using the inequality (4.3.1) and the assumption (i), we get the conditions **[L2]** and **[C2]** of Theorem 3.2.4 from (ii) and (iii), respectively. The condition **[PE]** is immediate from the inequality (4.3.2) and (iv). □

4.3.2 Asymptotic Normality

Let a constant $\tau > 0$ be given, and let us consider the estimation problem of $(t, \psi) \rightsquigarrow A(\psi)_t$. Recall the notations $\mathcal{L}^p$ and $\mathcal{L}^{p,n}$ given by (4.2.3) and (4.2.4), and the definition of π^n by (4.2.5), respectively. In the following, we will always assume:

$$\alpha_N^\star = \bigvee_{n \geq N} \overline{\alpha}^n \in \mathcal{L}^1 \text{ for some } N \in \mathbb{N}. \tag{4.3.6}$$

Thus $\alpha_N^\star(t, z)dt\lambda(dz)$ defines a finite measure on $[0, \tau] \times E$. We then denote

$$\mathcal{L}^p(\alpha_N^\star) = \mathcal{L}^p([0, \tau] \times E, \mathfrak{B}[0, \tau] \otimes \mathcal{E}, \alpha_N^\star(t, z)dt\lambda(dz)) \qquad \forall p \in [1, \infty].$$

Furthermore, we denote by $\|\cdot\|_{\mathcal{L}^p(\alpha_N^\star)}$ the L^p-seminorm on $\mathcal{L}^p(\alpha_N^\star)$. These should not be confused with the notations $\mathcal{L}^p$ and $\|\cdot\|_{\mathcal{L}^p}$ given by (4.2.3).

Let Ψ be a subset of $\mathcal{L}^{2p}(\alpha_N^\star) \cap \mathcal{L}^1$ with an envelope function $\varphi \in \mathcal{L}^{2p}(\alpha_N^\star) \cap \mathcal{L}^1$, where $p \in [1, \infty]$ should be specified in connection with another assumption in Condition 4.3.2. We propose the estimator $(t, \psi) \rightsquigarrow \widehat{A}^n(\psi)_t$ defined by

$$\widehat{A}^n(\psi)_t(\omega) = (\pi^n(\psi)Y^{n-}) * \mu_t^n(\omega) \qquad \forall (t, \psi) \in [0, \tau] \times \Psi,$$

where π^n is the mapping defined by (4.2.5). We shall first consider the residual $\sqrt{n}(\widehat{A}^n - \widetilde{A}^n)$, where

$$\widetilde{A}^n(\psi)_t(\omega) = \int_{[0,t] \times E} \pi^n(\psi)(s, z)I(\omega, s, z)\alpha(s, z)ds\lambda(dz) \qquad \forall (t, \psi) \in [0, \tau] \times \Psi,$$

Condition 4.3.2 *For some $p, q \in [1, \infty]$ such that $(1/p) + (1/q) = 1$, and a $\mathfrak{B}[0, \tau] \otimes \mathcal{E}$-measurable function $y = y(t, z)$ on $[0, \tau] \times E$, which is bounded away from zero, it holds that:*

$$\varphi \in \mathcal{L}^{2p}(\alpha_N^\star) \cap \mathcal{L}^1 \quad \text{and} \quad \int_0^1 \sqrt{\log N_{[]}(\Psi, \|\cdot\|_{\mathcal{L}^{2p}(\alpha_N^\star)}; \varepsilon)}d\varepsilon < \infty, \tag{4.3.7}$$
where the brackets should be constructed in $\mathcal{L}^{2p}(\alpha_N^\star) \cap \mathcal{L}^1$;

$$\|nY^{n-}\|_{\mathcal{L}^q(\alpha_N^\star)} = O_{P^n}(1); \tag{4.3.8}$$

$$\Big\| |\pi^n(\varphi)|^2 \cdot |nY^{n-} - (1/y)| \Big\|_{\mathcal{L}^1(\alpha_N^\star)} = o_{P^n}(1); \tag{4.3.9}$$

$$\|\{\pi^n(\psi)\pi^n(\phi) - \psi\phi\} \cdot (\alpha/y)\|_{\mathcal{L}^1} \to 0 \qquad \forall \psi, \phi \in \Psi; \tag{4.3.10}$$

$$\|\varphi \cdot (\overline{\alpha}^n - \underline{\alpha}^n)\|_{\mathcal{L}^1} = o(n^{-1/2}). \tag{4.3.11}$$

Theorem 4.3.3 *Assume (4.3.6). Suppose that a given class Ψ satisfies Conditions 4.3.2. Then, it holds that $\sqrt{n}(\widehat{A}^n - \widetilde{A}^n) \overset{P^n}{\Longrightarrow} G$ in $\ell^\infty([0, \tau] \times \Psi)$, where $(t, \psi) \rightsquigarrow G_t^\psi$ is a zero-mean Gaussian process such that*

$$EG_t^\psi G_s^\phi = \int_{[0,t\wedge s] \times E} \psi(u, z)\phi(u, z)\frac{\alpha(u, z)}{y(u, z)}dt\lambda(dz) \qquad \forall (t, \psi), (s, \phi) \in [0, \tau] \times \Psi$$

and that almost all paths are uniformly ρ_2-continuous on Ψ, where

$$\rho_2((t,\psi),(s,\phi)) = \sqrt{E|G_t^\psi - G_s^\phi|^2} \qquad \forall (t,\psi),(s,\phi) \in [0,\tau]\times\Psi.$$

Proof. It follows from (4.3.1) and (4.3.11) that

$$\sup_{(t,\psi)\in[0,\tau]\times\Psi} \left|(\pi^n(\psi)Y^{n-}) * \nu_t^n - A(\pi^n(\psi)I^n)_t\right| = o_{P^{n*}}(n^{-1/2}).$$

Thus it suffices to derive the weak convergence of the sequence of processes $(t,\psi) \rightsquigarrow X_t^{n,\psi}$ defined by

$$X_t^{n,\psi} = \sqrt{n}(\pi^n(\psi)Y^{n-}) * (\mu^n - \nu^n)_t.$$

We will apply Lemma 4.3.1 with $K^{n,\psi} = \sqrt{n}\pi^n(\psi)$. The condition (4.3.5) for $n \geq N$ follows (4.3.6) and the fact $\varphi \in \mathcal{L}^2(\alpha_N^\star)$. To show the condition (i), observe that

$$\begin{aligned}
&\left| n\int_{[0,\tau]\times E} |\pi^n(\varphi)(t,z)|^2 Y^{n-}(t,z)(\overline{\alpha}^n - \underline{\alpha}^n)(t,z)dt\lambda(dz)\right| \\
&\leq \left\||\pi^n(\varphi)|^2\cdot|nY^{n-} - (1/y)|\cdot(\overline{\alpha}^n - \underline{\alpha}^n)\right\|_{\mathcal{L}^1} + \left\||\pi^n(\varphi)|^2\cdot(1/y)\cdot(\overline{\alpha}^n - \underline{\alpha}^n)\right\|_{\mathcal{L}^1} \\
&\leq \left\||\pi^n(\varphi)|^2\cdot|nY^{n-} - (1/y)|\right\|_{\mathcal{L}^1(\alpha_N^\star)} + \|1/y\|_{\mathcal{L}^\infty}\cdot\left\|\varphi^2\cdot(\overline{\alpha}^n - \underline{\alpha}^n)\right\|_{\mathcal{L}^1}.
\end{aligned}$$

The first term on the right hand side converges in P^n-probability to zero by (4.3.9). On the other hand, since it follows from (4.3.11) that $(\overline{\alpha}^n - \underline{\alpha}^n)(t,z) \to 0$ for $\varphi(t,z)dt\lambda(dz)$-almost all (t,z), the dominated convergence theorem yields that the second term also vanishes. Thus the condition (i) has been proved. Next, it follows from the same computation as (4.1.5) that

$$\begin{aligned}
&nA\left(|\pi^n(\varphi)|^2 Y^{n-} 1_{\{\sqrt{n}\pi^n(\varphi)Y^{n-}>\varepsilon\}}\right) \\
&\leq \left\||\pi^n(\varphi)|^2\cdot nY^{n-}\cdot 1_{\{\sqrt{n}\pi^n(\varphi)Y^{n-}>\varepsilon\}}\right\|_{\mathcal{L}^1(\alpha_N^\star)} \\
&\leq \frac{1}{\sqrt{n}\varepsilon}\cdot\|\pi^n(\varphi)\|_{\mathcal{L}^{2p}(\alpha_N^\star)}^2\cdot\|nY^{n-}\|_{\mathcal{L}^q(\alpha_N^\star)},
\end{aligned}$$

which converges in P^n-probability to zero by (4.3.7) and (4.3.8). Thus the condition (ii) has been proved. The condition (iii) can be shown by using (4.3.9) and (4.3.10). This finishes the proof of the case where the set Ψ is finite.

On the other hand, it follows from Lemma 4.2.2 that, if $[l^k, u^k]$'s are $(\|\cdot\|_{\mathcal{L}^{2p}(\alpha_N^\star)},\varepsilon)$-brackets in $\mathcal{L}^{2p}(\alpha_N^\star)\cap\mathcal{L}^1$ which cover the class Ψ, then $[\pi^n(l^k),\pi^n(u^k)]$'s are $(\|\cdot\|_{\mathcal{L}^{2p}(\alpha_N^\star)},\varepsilon)$-brackets in $\mathcal{L}^{2p}(\alpha_N^\star)\cap\mathcal{L}^1$ which cover the class $\pi^n\Psi$. Hence, by using the Hölder inequality, it is shown that the DFP of Ψ induced from the minimum brackets satisfies the requirements of (iv) of Lemma 4.3.1. □

In order to derive the weak convergence of $\sqrt{n}(\widehat{A}^n - A)$ rather than $\sqrt{n}(\widehat{A}^n - \widetilde{A}^n)$, we have to show that $\sqrt{n}(\widetilde{A}^n - A) \overset{P^n}{\Longrightarrow} 0$ in $\ell^\infty([0,\tau] \times \Psi)$. For this purpose, it suffices to check the following:

$$\sup_{\psi \in \Psi} \|(\pi^n(\psi) - \psi) \cdot \alpha\|_{\mathcal{L}^1} = o(n^{-1/2}); \tag{4.3.12}$$

$$\|\varphi \cdot (1 - I^n) \cdot \alpha\|_{\mathcal{L}^1} = o_{P^n}(n^{-1/2}). \tag{4.3.13}$$

See the discussion after Theorem 4.2.4 for getting simple sufficient conditions for all assumptions appearing above in the case where $E = [0,1]^d$.

4.3.3 Asymptotic Efficiency

In order to discuss the asymptotic efficiency, we set:

$$\begin{aligned} \mathbb{H} &= L^2([0,\tau] \times E, \tfrac{\alpha(t,z)}{y(t,z)} dt\lambda(dz)); \\ H &= L^\infty([0,\tau] \times E, \mathfrak{B}[0,\tau] \otimes \mathcal{E}, dt\lambda(dz)); \\ H^n &= L^\infty([0,\tau] \times E, \mathfrak{B}[0,\tau] \otimes \mathcal{E}^n, dt\lambda(dz)). \end{aligned} \tag{4.3.14}$$

Here, the function y is the one which appears in Condition 4.3.2. We equip $\mathbb{H}$ with the usual L^2-inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$. Since we always assume (4.3.6), and since $1/y$ is assumed to be bounded, H is a linear subspace of the Hilbert space $\mathbb{H}$. Let $\mathbf{P}^n = \{P_h^n : h \in H\}$ be a family of probability measures on $(\Omega^n, \mathcal{F}^n)$ indexed by H. Define the mapping $\pi^n : H \to H^n$ by (4.2.5). Hereafter, we denote by $\pi^n(h)$ any function of the equivalent elements in H^n.

Suppose that the predictable compensator of $N^{n,i}$ under the probability measure P_h^n is given by

$$\alpha_h^n(t, Z_t^{n,i}) Y_t^{n,i} dt, \tag{4.3.15}$$

where $\alpha_h^n = \alpha_h^n(t,z)$ is defined by

$$\alpha_h^n = \left\{1 + n^{-1/2} \pi^n(h/2y)\right\}^2 \alpha. \tag{4.3.16}$$

To see that the compensator (4.3.15) is well-defined, recall the discussion after (4.2.16) and (4.2.17). In particular, it holds that $\alpha_0^n(t, Z_t^{n,i}(\omega)) = \alpha(t, Z_t^{n,i}(\omega))$ for almost all t. The predictable compensator $\nu^{n,h}$ of μ^n under the probability measure P_h^n is then given by

$$\nu^{n,h}(dt, dz) = \alpha_h^n(t,z) \overline{\nu}^n(dt, dz)$$

where

$$\overline{\nu}^n(dt, dz) = \sum_{i=1}^{k_n} Y_t^{n,i} \varepsilon_{Z_t^{n,i}}(dz) dt.$$

STEP I: LOCAL ASYMPTOTIC NORMALITY. Assume $P_h^n \ll P_0^n$. It is well-known that, under some conditions, a version of the log-likelihood ratio is given by

$$\log \frac{dP_h^n|\mathcal{F}_\tau^n}{dP_0^n|\mathcal{F}_\tau^n} = 2\left(\log\left|1 + n^{-1/2}\pi^n(h/2y)\right|\right) * \mu_\tau^n - \left(\left|1 + n^{-1/2}\pi^n(h/2y)\right|^2 - 1\right) * \nu_\tau^{n,0} \tag{4.3.17}$$

(see, e.g., Theorem III.5.13 of Jacod and Shiryaev (1987)). This representation does not depend on the choice of a version of $\pi^n(h/2y) \in H^n$, because it holds that $\nu^{n,0}(\omega; B) = 0$ identically and that $P_0^n(\mu^n(B) = 0) = 1$ for any $B \in \mathfrak{B}[0,\tau] \otimes \mathcal{E}^n$ such that $\mathrm{Leb} \otimes \lambda(B) = 0$.

Proposition 4.3.4 *Let C be an arbitrary subset of H. Assume (4.3.6). Suppose that the function y appearing in Condition 4.3.2 satisfies the following:*

$$y \in \mathcal{L}^1(\alpha_N^\star) \quad \textit{and} \quad \left\| \frac{I^n}{nY^{n-}} - y \right\|_{\mathcal{L}^1(\alpha_N^\star)} = o_{P_0^n}(1);$$

$$\left\| \left\{ \pi^n(h_1/2y)\pi^n(h_2/2y) - (h_1h_2)/(4y^2) \right\} \cdot y \cdot \alpha \right\|_{\mathcal{L}^1} \to 0 \qquad \forall h_1, h_2 \in C;$$

$$\|y \cdot (\overline{\alpha}^n - \underline{\alpha}^n)\|_{\mathcal{L}^1} \to 0.$$

Introduce the Hilbert space $\mathbb{H}$ given by (4.3.14). Suppose also that $P_h^n \ll P_0^n$ and (4.3.17) hold for every $h \in C$. Then, it holds that for every $h \in C$

$$\log \frac{dP_h^n|\mathcal{F}_\tau^n}{dP_0^n|\mathcal{F}_\tau^n} = \Delta^n(h) - \frac{1}{2}\|h\|_{\mathbb{H}}^2 + \epsilon_n(h),$$

where

$$\Delta^n(h) = n^{-1/2}(2\pi^n(h/2y)) * (\mu^n - \nu^{n,0})_\tau$$

and $\epsilon_n(h) = o_{P_0^n}(1)$. Furthermore, it also holds that $(\Delta^n(h_1), ..., \Delta^n(h_d)) \overset{P_0^n}{\Longrightarrow} N(0, \Sigma)$ where $\Sigma_{ij} = \langle h_i, h_j \rangle_{\mathbb{H}}$.

Proof. Since $|\log(1+x) - x + \frac{x^2}{2}| \le \frac{8}{3}x^3$ for all $x \in [-\frac{1}{2}, \frac{1}{2}]$, we have that for any $\varepsilon \in (0,1]$,

$$\left| \log \frac{dP_h^n|\mathcal{F}_\tau^n}{dP_0^n|\mathcal{F}_\tau^n} - \Delta^n(h) + D^n(h) + \tilde{D}^n(h) \right| \le \frac{16}{3}\varepsilon D^n(h)$$

whenever $\sup_{t,z} |n^{-1/2}\pi^n(h/2y)(t,z)| \le \varepsilon$, where:

$$D^n(h) = |n^{-1/2}\pi^n(h/2y)|^2 * \mu_\tau^n;$$

$$\tilde{D}^n(h) = |n^{-1/2}\pi^n(h/2y)|^2 * \nu_\tau^{n,0}.$$

Since

$$\begin{aligned}
&\left|\left(|\pi^n(h/2y)|^2\frac{I^n}{nY^{n-}}Y^{n-}\right)*\nu_\tau^{n,0}-A\left(|\pi^n(h/2y)|^2\frac{I^n}{nY^{n-}}\right)_\tau\right| \\
\leq\ & \left\||\pi^n(h/2y)|^2\cdot\frac{I^n}{nY^{n-}}\cdot(\overline{\alpha}^n-\underline{\alpha}^n)\right\|_{\mathcal{L}^1} \\
\leq\ & \left\||\pi^n(h/2y)|^2\cdot\left|\frac{I^n}{nY^{n-}}-y\right|\cdot(\overline{\alpha}^n-\underline{\alpha}^n)\right\|_{\mathcal{L}^1}+\left\||\pi^n(h/2y)|^2\cdot y\cdot(\overline{\alpha}^n-\underline{\alpha}^n)\right\|_{\mathcal{L}^1} \\
\leq\ & \|\pi^n(h/2y)\|^2_{\mathcal{L}^\infty}\left\{\left\|\frac{I^n}{nY^{n-}}-y\right\|_{\mathcal{L}^1(\alpha_N^*)}+\|y\cdot(\overline{\alpha}^n-\underline{\alpha}^n)\|_{\mathcal{L}^1}\right\},
\end{aligned}$$

and since

$$A\left(|\pi^n(h/2y)|^2\left|\frac{I^n}{nY^{n-}}-y\right|\right)_\tau\leq\|\pi^n(h/2y)\|^2_{\mathcal{L}^\infty}\cdot\left\|\frac{I^n}{nY^{n-}}-y\right\|_{\mathcal{L}^1(\alpha_N^*)},$$

it holds that $\tilde{D}^n(h)\xrightarrow{P_0^n}\frac{1}{4}\|h\|^2_{\mathbb{H}}$. Also, using Lenglart's inequality, we have $D^n(h)-\tilde{D}^n(h)\xrightarrow{P_0^n}0$. These facts imply the first assertion. Applying Lemma 4.3.1 to the family of $\widetilde{Q}^n$-measurable functions $\{K^{n,h}:h\in C\}$ given by

$$K^{n,h}=n^{1/2}\cdot 2\pi^n(h/2y)\cdot\frac{I^n}{nY^{n-}},$$

we can show the finite-dimensional convergence of $h\rightsquigarrow\Delta^n(h)$. □

Discussion about Step II and Step III is similar to the preceding sections, hence is omitted.

4.A Notes

A part of the results in Section 4.1 was presented by Nishiyama (1997). As mentioned in the main text, a progress from the preceding work is the introduction of L^{2p}-bracketing entropies rather than the L^2-bracketing one. This would be important also for other applications in non-i.i.d. settings; see, e.g., Chapter 6. The problems considered in Sections 4.2 and 4.3 were posed by McKeague and Utikal (1990) who treated the case where the state space of covariates is $[0,1]$. Our way of constructing the estimator is motivated by their work. A moment assumption of theirs has been weakened.

Chapter 5

M-Estimators: General Criterion and Euclidean Parameters

5.1 General Criterion

The common structure of the models treated in this and next chapters is as follows.

Formulation 5.1.1 *For every $n \in \mathbb{N}$, let $(\Omega^n, \mathcal{F}^n)$ be a measurable space and $\mathbf{P}^n = \{P_u^n : u \in U^n\}$ a family of probability measures on $(\Omega^n, \mathcal{F}^n)$ indexed by an arbitrary set U^n. For every $n \in \mathbb{N}$ and $u \in U^n$, let the following be given:*

(i) *a space Θ^n, a random point $\theta_u^n \in \Theta^n$, and a $[0, \infty)$-valued stochastic process $\theta \rightsquigarrow d_u^n(\theta, \theta_u^n)$ with parameter in Θ^n;*

(ii) *some stochastic processes $\theta \rightsquigarrow \Gamma_u^n(\theta)$ and $\theta \rightsquigarrow \gamma_u^n(\theta)$, with parameters in Θ^n.*

We then denote $R_u^n(\delta) = \{\theta \in \Theta^n : (\delta/2) < d_u^n(\theta, \theta_u^n) \leq \delta\}$ for every $\delta \in (0, \infty)$.

The $[0, \infty)$-valued stochastic process $\theta \rightsquigarrow d_u^n(\theta, \theta_u^n)$ above is usually given by a (random) semimetric $d_u^n(\theta, \vartheta)$ on Θ^n and a (random) point θ_u^n which should be regarded as an (approximate) true point of unknown parameter. We refer the processes $\theta \rightsquigarrow \Gamma_u^n(\theta)$ and $\theta \rightsquigarrow \gamma_u^n(\theta)$ as the "criterion process" and the "contrast process", respectively; the latter is sometimes taken to be deterministic, and in that case it is referred as the "contrast function". The following result is a version of Theorem 3.4.1 of van der Vaart and Wellner (1996), into which contributions by some other authors in this area are condensed (see Notes at the end of this chapter).

Theorem 5.1.2 *Consider Formulation 5.1.1 above. Suppose that the following M-*CRITERION *is satisfied for some $\delta_0 \in (0, \infty]$, $p > 0$, $a \in (0, p)$, not depending on n and u, some functions $\phi_u^n : (0, \delta_0) \to (0, \infty)$ such that $\delta \rightsquigarrow \delta^{-a}\phi_u^n(\delta)$ is decreasing, and some positive constants $r_{n,u}$ such that $r_{n,u}^{-1} \in (0, \delta_0)$ and that $\phi_u^n(r_{n,u}^{-1}) \leq r_{n,u}^{-p}$.*

M-CRITERION. For every $\varepsilon > 0$ there exists some $c_\varepsilon, C_\varepsilon, K_\varepsilon > 0$ and $n_\varepsilon \in \mathbb{N}$ such that: for every $n \geq n_\varepsilon$ and $u \in U^n$ there exists a set $B_u^n(\varepsilon) \subset \Omega^n$ such that

$$\gamma_u^n(\theta) - \gamma_u^n(\theta_u^n) \leq -c_\varepsilon \delta^p \qquad \forall \theta \in R_u^n(\delta) \qquad \text{on the set } B_u^n(\varepsilon)$$

and

$$E_u^{n*} \sup_{\theta \in R_u^n(\delta)} |(\Gamma_u^n - \gamma_u^n)(\theta) - (\Gamma_u^n - \gamma_u^n)(\theta_u^n)| 1_{B_u^n(\varepsilon)} \leq C_\varepsilon \phi_u^n(\delta)$$

whenever $\delta \in [K_\varepsilon r_{n,u}^{-1}, \delta_0)$, and that $P_u^{n*}(\Omega^n \setminus B_u^n(\varepsilon)) \leq \varepsilon$.
Then, for any mappings $\widehat{\theta}_u^n : \Omega^n \to \Theta^n$ such that

$$\lim_{L\to\infty} \limsup_{n\to\infty} \sup_{u \in U^n} P_u^{n*}\left(\Gamma_u^n(\widehat{\theta}_u^n) < \Gamma_u^n(\theta_u^n) - L r_{n,u}^{-p}\right) = 0 \tag{5.1.1}$$

and that

$$\lim_{n\to\infty} \sup_{u \in U^n} P_u^{n*}\left(d_u^n(\widehat{\theta}_u^n, \theta_u^n) > \delta_0/2\right) = 0, \tag{5.1.2}$$

it holds that

$$\lim_{L\to\infty} \limsup_{n\to\infty} \sup_{u \in U^n} P_u^{n*}\left(r_{n,u} d_u^n(\widehat{\theta}_u^n, \theta_u^n) > L\right) = 0.$$

When M-CRITERION is satisfied for $\delta_0 = \infty$, the assumption (5.1.2) is unnecessary.

Remark. In the sequel, we refer the first and the second displayed inequalities of M-CRITERION as the "FIRST INEQUALITY" and the "SECOND INEQUALITY", respectively.

Keeping a two-term Taylor expansion of the function $\theta \rightsquigarrow \gamma_u^n(\theta)$ in their mind, van der Vaart and Wellner (1996) presented some results of this fashion for the case of $p = 2$ as their Theorems 3.2.5 and 3.4.1. The adaptation to the case of general p will be useful in Sections 5.2, 5.3 and 7.2. The truncation introduced in the SECOND INEQUALITY fits in our maximal inequalities based on the quadratic modulus. The last difference is the uniformity in the underlying probability measures (this is clear if the conclusion is given in the form of a probability inequality by using universal constants; see, e.g., Birgé and Massart (1993) and van de Geer (1995b)). Although the change of the proof is minor, we state the whole proof following exactly the same line as that of van der Vaart and Wellner (1996).

Proof of Theorem 5.1.2. Fix any $\varepsilon > 0$, and choose some constants $c_\varepsilon, C_\varepsilon, K_\varepsilon > 0$ and $n_\varepsilon \in \mathbb{N}$ according to M-CRITERION. In the following we will consider $n \geq n_\varepsilon$ only.

Now, fix any $L > 0$ for a while, and choose any $J \in \mathbb{N}$ such that $c_\varepsilon - 2^{-p(J-1)}L > 0$ and that $2^J \geq K_\varepsilon$. Put $J_u^n = \max\{j \in \mathbb{N} : 2^j r_{n,u}^{-1} < \delta_0\}$ (we have implicitly assumed $\delta_0 < \infty$, but the case of $\delta_0 = \infty$ is easier; read the following argument replacing "$J \leq \forall j \leq J_u^n$"

by "$\forall j \geq J$"). We denote:

$$
\begin{aligned}
S_u^n(j) &= \left\{\omega \in \Omega^n : \widehat{\theta}_u^n(\omega) \in R_u^n(2^j r_{n,u}^{-1})\right\} \qquad J \leq \forall j \leq J_u^n; \\
\Omega_u^n(\varepsilon, L) &= \left\{\omega \in \Omega^n : \Gamma_u^n(\widehat{\theta}_u^n(\omega))(\omega) - \Gamma_u^n(\theta_u^n)(\omega) \geq -Lr_{n,u}^{-p}\right\} \cap B_u^n(\varepsilon).
\end{aligned}
$$

Then it holds on the set $S_u^n(j) \cap \Omega_u^n(\varepsilon, L)$ that

$$
\sup_{\theta \in R_u^n(2^j r_{n,u}^{-1})} \Gamma_u^n(\theta) - \Gamma_u^n(\theta_u^n) \geq -Lr_{n,u}^{-p}
$$

and that

$$
\inf_{\theta \in R_u^n(2^j r_{n,u}^{-1})} \gamma_u^n(\theta) - \gamma_u^n(\theta_u^n) \leq -c_\varepsilon 2^{p(j-1)} r_{n,u}^{-p};
$$

thus we have

$$
\begin{aligned}
\text{(5.1.3)} \quad \sup_{\theta \in R_u^n(2^j r_{n,u}^{-1})} \{(\Gamma_u^n - \gamma_u^n)(\theta) - (\Gamma_u^n - \gamma_u^n)(\theta_u^n)\} &\geq (c_\varepsilon 2^{p(j-1)} - L) r_{n,u}^{-p} \\
&\geq (c_\varepsilon - 2^{-p(J-1)} L) 2^{p(j-1)} r_{n,u}^{-p}.
\end{aligned}
$$

Since $\{d_u^n(\widehat{\theta}_u^n, \theta_u^n) \leq \delta_0/2\} \subset \{d_u^n(\widehat{\theta}_u^n, \theta_u^n) \leq 2^{J_u^n} r_{n,u}^{-1}\}$ it holds that

$$
\begin{aligned}
& P_u^{n*}\left(2^{J-1} r_{n,u}^{-1} < d_u^n(\widehat{\theta}_u^n, \theta_u^n) \leq \delta_0/2,\ \Omega_u^n(\varepsilon, L)\right) \\
& \leq \sum_{J \leq j \leq J_u^n} P_u^{n*}\left(S_u^n(j) \cap \Omega_u^n(\varepsilon, L)\right),
\end{aligned}
$$

where the summation with respect to j can be read as zero when $J > J_u^n$. If $J \leq J_u^n$, recalling (5.1.3), we obtain from the Markov inequality and the SECOND INEQUALITY of M-CRITERION that

$$
\begin{aligned}
\sum_{J \leq j \leq J_u^n} P_u^{n*}\left(S_u^n(j) \cap \Omega_u^n(\varepsilon, L)\right) &\leq \frac{C_\varepsilon}{c_\varepsilon - 2^{-p(J-1)} L} \sum_{J \leq j \leq J_u^n} \frac{\phi_u^n(2^j r_{n,u}^{-1})}{2^{p(j-1)} r_{n,u}^{-p}} \\
&\leq \frac{C_\varepsilon}{c_\varepsilon - 2^{-p(J-1)} L} \sum_{J \leq j \leq J_u^n} \frac{2^{aj} \phi_u^n(r_{n,u}^{-1})}{2^{p(j-1)} r_{n,u}^{-p}} \\
&\leq \frac{2^p C_\varepsilon}{c_\varepsilon - 2^{-p(J-1)} L} \sum_{j \geq J} 2^{(a-p)j}.
\end{aligned}
$$

Here we have also used the fact that $\phi_u^n(c\delta) \leq c^a \phi_u^n(\delta)$ for every $c > 1$.

Consequently we have

$$
\begin{aligned}
& P_u^{n*}\left(r_{n,u} d_u^n(\widehat{\theta}_u^n, \theta_u^n) > 2^{J-1}\right) \\
& \leq P_u^{n*}(\Omega^n \setminus \Omega_u^n(\varepsilon, L)) + P_u^{n*}\left(d_u^n(\widehat{\theta}_u^n, \theta_u^n) > \delta_0/2\right) + \frac{2^p C_\varepsilon}{1 - 2^{(a-p)}} \cdot \frac{2^{(a-p)J}}{c_\varepsilon - 2^{-p(J-1)} L}.
\end{aligned}
$$

This inequality holds also in the case of $\delta_0 = \infty$ by regarding the second term on the right hand side as zero. Notice that the last term on the right hand side does not depend on $n \in \mathbb{N}$ and $u \in U^n$ and converges to zero as $J \to \infty$ since $a < p$. To get the assertion, first choose large $L > 0$ according to the assumption (5.1.1), and next let $J \to \infty$. □

In the remaining sections of this chapter, we are concerned with some problems of estimating finite-dimensional parameters. Here, we sketch a procedure for deriving the asymptotic distribution of M-estimators based on a continuous mapping theorem for argmax functionals (Theorem 3.2.2 of van der Vaart and Wellner (1996)). In any case, we shall consider some rescaled criterion processes $h \rightsquigarrow \mathbb{M}^n(h)$ of the form

$$\mathbb{M}^n(h) = a_n\{\Gamma^n(\theta_0 + r_n^{-1}h) - \Gamma^n(\theta_0)\},$$

where r_n and a_n are some appropriate constants. Thus the first problem should be to find the "rate of convergence" r_n, and Theorem 5.1.2 is useful at this step. The constant a_n should be determined in connection with r_n. Next, according to Theorem 3.2.2 of van der Vaart and Wellner (1996), we shall show the following.

(i) The uniform tightness of the local sequence $\widehat{h}^n = r_n(\widehat{\theta}^n - \theta_0)$.

(ii) The weak convergence of the process $h \rightsquigarrow \mathbb{M}^n(h)$ to a continuous process $h \rightsquigarrow \mathbb{M}(h)$ in $\ell^\infty(K)$, for every compact subset K of the space of local parameters.

(iii) The existence of a unique maximum point $\widehat{h}$ of the path $h \rightsquigarrow \mathbb{M}(h)$.

Any Borel random variable on a Polish space is tight, hence so is $\widehat{h}$. Thus a result of the form "$r_n(\widehat{\theta}^n - \theta_0) \stackrel{P}{\Longrightarrow} \widehat{h}$" follows from the argmax continuous mapping theorem.

The reason why we are content with the case of finite-dimensional parameters in this approach is that the uniform tightness of the local sequence $\widehat{h}^n$ (the step (i) above) is equivalent to "$r_n|\widehat{\theta}^n - \theta_0| = O_P(1)$", which is actually the consequence of Theorem 5.1.2. This is not always true when the parameter space is general, but Theorem 5.1.2 is still useful at least for deriving the rate of convergence as we see in Chapter 6. We will make use of the results given in Chapter 3 at step (ii). For simplicity, we will not discuss the uniformity in the underlying probability measures in Sections 5.2 and 5.3.

5.2 Gaussian White Noise Model

5.2.1 Criterion for Rate of Convergence

For every $n \in \mathbb{N}$, let $X^n = (X_t^n)_{t \in [0,1]}$ be a continuous stochastic process given by

(5.2.1) $$dX_t^n = f(t)dt + n^{-1/2}dB_t, \qquad X_0^n = x_0 \in \mathbb{R},$$

where $f \in \mathcal{L}^2[0,1]$, and $B = (B_t)_{t\in[0,1]}$ is a standard Brownian motion on a stochastic basis $\mathbf{B} = (\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t\in[0,1]}, P)$. Let (Θ, d) be a separable metric space; we will take it to be a Euclidean space later. Let $w = \{w_\theta : \theta \in \Theta\}$ be a class of elements of $\mathcal{L}^2[0,1]$. Equip Θ with the semimetric ρ_w given by

$$\rho_w(\theta, \vartheta) = \|w_\theta - w_\vartheta\|_{\mathcal{L}^2[0,1]} \qquad \forall \theta, \vartheta \in \Theta. \tag{5.2.2}$$

We consider the criterion process $\theta \rightsquigarrow \Gamma^n(\theta)$ defined by

$$\Gamma^n(\theta) = \int_0^1 w_\theta(t) dX_t^n \qquad \forall \theta \in \Theta, \tag{5.2.3}$$

and the contrast function $\theta \rightsquigarrow \gamma(\theta)$ defined by

$$\gamma(\theta) = \int_0^1 w_\theta(t) f(t) dt \qquad \forall \theta \in \Theta. \tag{5.2.4}$$

The former is indeed well-defined as the stochastic integral with respect to the semimartingale $t \rightsquigarrow X_t^n$ (see, e.g., Theorem I.4.31 of Jacod and Shiryaev (1987)). When Θ is not countable, the process $\theta \rightsquigarrow \Gamma^n(\theta)$ is not unique even in the almost sure sense. However, the following argument holds for *any version* of the process, because we shall always consider a countable subset Θ^* of Θ when we apply Theorem 2.4.2 to the terminal variables of the continuous martingales

$$\Gamma^n(\theta) - \gamma(\theta) = n^{-1/2} \int_0^1 w_\theta(t) dB_t \qquad \forall \theta \in \Theta^* \subset \Theta.$$

We denote by $B_{(\Theta,d)}(\theta; \delta)$ the closed ball in Θ with center $\theta \in \Theta$ and d-radius $\delta > 0$.

Theorem 5.2.1 *Let (Θ, d) be a separable metric space. For a given class $w = \{w_\theta : \theta \in \Theta\} \subset \mathcal{L}^2[0,1]$, introduce ρ_w, Γ^n and γ given by (5.2.2), (5.2.3) and (5.2.4). Suppose that there exists a countable, d-dense subset Θ^* of Θ such that ρ_w defines a proper metric on Θ^*. For a given point $\theta_0 \in \Theta$, suppose also that $\theta \rightsquigarrow \gamma(\theta)$ is d-continuous at θ_0 and that the following conditions are satisfied for some $\delta_0 \in (0, \infty]$, $p > 0$, $a \in (0, p)$, $c > 0$ and a function $\varphi : (0, \delta_0) \to (0, \infty)$ such that $\delta \rightsquigarrow \delta^{-a}\varphi(\delta)$ is decreasing:*

$$\gamma(\theta) - \gamma(\theta_0) \le -c d(\theta, \theta_0)^p \qquad \forall \theta \in B_{(\Theta,d)}(\theta_0; \delta_0); \tag{5.2.5}$$

$$\sup_{\theta \in \Theta^*} \int_0^\infty \sqrt{\log N(B_{(\Theta,d)}(\theta; \delta), \rho_w; \varepsilon)} d\varepsilon \le \varphi(\delta) \qquad \forall \delta \in (0, \delta_0); \tag{5.2.6}$$

$$\sup_{\theta \in \Theta^*} \mathrm{Diam}(B_{(\Theta,d)}(\theta; \delta), \rho_w) \le \varphi(\delta) \qquad \forall \delta \in (0, \delta_0). \tag{5.2.7}$$

Choose any constants $r_n > 0$ such that $r_n^{-1} \in (0, \delta_0)$ and that $n^{-1/2}\varphi(r_n^{-1}) \le r_n^{-p}$. Then, for any Θ^*-valued random sequence $\widehat{\theta}^n$ such that $d(\widehat{\theta}^n, \theta_0) = o_{P^*}(1)$ and that

$$\Gamma^n(\widehat{\theta}^n) \ge \sup_{\theta \in \Theta^*} \Gamma^n(\theta) - \epsilon_n$$

for some $\epsilon_n = O_{P^*}(r_n^{-p})$, *it holds that* $d(\widehat{\theta}^n, \theta_0) = O_{P^*}(r_n^{-1})$. *When* $\delta_0 = \infty$, *the assumption* "$d(\widehat{\theta}^n, \theta_0) = o_{P^*}(1)$" *is unnecessary. When* $\theta_0 \in \Theta^*$, *the assumption that* $\theta \rightsquigarrow \gamma(\theta)$ *is* d*-continuous at* θ_0 *is unnecessary.*

Proof. We will apply Theorem 5.1.2. Since $\theta \rightsquigarrow \gamma(\theta)$ is d-continuous at θ_0, we can choose a point $\theta_{\theta_0}^n \in \Theta^*$ such that $|\gamma(\theta_{\theta_0}^n) - \gamma(\theta_0)| \leq (c/2^{p+1}) \cdot r_n^{-p}$ and that $d(\theta_{\theta_0}^n, \theta_0) \leq r_n^{-1}$ (when $\theta_0 \in \Theta^*$, the choice $\theta_{\theta_0}^n = \theta_0$ satisfies these requirements, thus the assumption that $\theta \rightsquigarrow \gamma(\theta)$ is d-continuous is unnecessary). We then denote $R_{\theta_0}^n(\delta) = \{\theta \in \Theta^* : (\delta/2) < d(\theta, \theta_{\theta_0}^n) \leq \delta\}$ for every $\delta \in (0, \infty)$.

For any $\delta \in [r_n^{-1}, \delta_0)$, it holds that

$$\begin{aligned} \gamma(\theta) - \gamma(\theta_{\theta_0}^n) &= \gamma(\theta) - \gamma(\theta_0) + \gamma(\theta_0) - \gamma(\theta_{\theta_0}^n) \\ &\leq -cd(\theta, \theta_0)^p + \frac{c}{2^{p+1}} \cdot r_n^{-p} \\ &\leq -\frac{c}{2^p} \cdot \delta^p + \frac{c}{2^{p+1}} \cdot \delta^{-p} \qquad \forall \theta \in R_{\theta_0}^n(\delta) \\ &\leq -\frac{c}{2^{p+1}} \cdot \delta^p. \end{aligned}$$

Thus the FIRST INEQUALITY of M-CRITERION is fulfilled. Next, to show the SECOND INEQUALITY, notice that for every $\delta \in (0, \delta_0)$

$$\begin{aligned} & E \sup_{\theta, \vartheta \in R_{\theta_0}^n(\delta)} |(\Gamma^n - \gamma)(\theta) - (\Gamma^n - \gamma)(\vartheta)| \\ &\leq n^{-1/2} E \sup_{\theta, \vartheta \in B_{(\Theta,d)}(\theta_{\theta_0}^n; \delta) \cap \Theta^*} \left| \int_0^1 w_\theta(t) dB_t - \int_0^1 w_\vartheta(t) dB_t \right|. \end{aligned}$$

Since the quadratic ρ_w-modulus of the family of continuous martingales

$$\left\{ \int_0^{\cdot} w_\theta(t) dB_t : \theta \in \Theta^* \right\}$$

is bounded by 1, and since $\mathrm{Diam}(B_{(\Theta,d)}(\theta_{\theta_0}^n; \delta), \rho_w) \leq \varphi(\delta)$, it follows from Theorem 2.4.2 and (2.1.1) that the right hand side is bounded by (up to a multiplicative universal constant)

$$\begin{aligned} & n^{-1/2} \int_0^{\varphi(\delta)} \sqrt{\log(1 + N(B_{(\Theta,d)}(\theta_{\theta_0}^n; \delta), \rho_w; \varepsilon))} d\varepsilon \\ &\leq n^{-1/2} \left\{ \varphi(\delta) \sqrt{\log 2} + \int_0^{\varphi(\delta)} \sqrt{\log N(B_{(\Theta,d)}(\theta_{\theta_0}^n; \delta), \rho_w; \varepsilon)} d\varepsilon \right\} \\ &\leq \left\{ \sqrt{\log 2} + 1 \right\} \cdot n^{-1/2} \varphi(\delta). \end{aligned}$$

Thus the SECOND INEQUALITY is satisfied with $\phi_{\theta_0}^n = n^{-1/2}\varphi$. □

The assumptions (5.2.6) and (5.2.7) are analogous to an assumption of Theorem 3.2.10 of van der Vaart and Wellner (1996). Although the supremum with respect to δ comes out of the integral, this condition may still look awkward at first sight. Indeed, it requires a calculation of certain covering numbers of the sets $B_{(\Theta,d)}(\theta;\delta)$ for all sufficiently small $\delta > 0$. However, when the parameter space (Θ, d) is Euclidean, this condition can be replaced by a simple relationship between the two metrics d and ρ_w, as is given in the next theorem.

Theorem 5.2.2 *Let Θ be a subset of a finite-dimensional Euclidean space with the usual metric d. For a given class $w = \{w_\theta : \theta \in \Theta\} \subset \mathcal{L}^2[0,1]$, introduce ρ_w, Γ^n and γ given by (5.2.2), (5.2.3) and (5.2.4). Suppose that there exists a countable, d-dense subset Θ^* of Θ such that ρ_w defines a proper metric on Θ^*. For a given point $\theta_0 \in \Theta$, suppose also that $\theta \rightsquigarrow \gamma(\theta)$ is d-continuous at θ_0 and that there exist some $\delta_0 \in (0,\infty]$ and some constants $p > q > 0$ and $c, C > 0$ such that:*

$$(5.2.8) \qquad \left\{ \begin{array}{rcll} \gamma(\theta) - \gamma(\theta_0) & \leq & -cd(\theta,\theta_0)^p & \forall \theta \in B_{(\Theta,d)}(\theta_0;\delta_0); \\ \rho_w(\theta,\vartheta) & \leq & Cd(\theta,\vartheta)^q & \forall \theta,\vartheta \in \Theta. \end{array} \right.$$

Then, the same conclusion as Theorem 5.2.1 holds for $r_n = n^{1/2(p-q)}$.

Proof. It suffices to show that the conditions (5.2.6) and (5.2.7) of Theorem 5.2.1 are satisfied with $\varphi(\delta) = \text{const.}\delta^q$. First notice that for every $\delta > 0$

$$(5.2.9) \qquad d(\theta,\vartheta) \leq \varepsilon^{1/q}\delta \quad \Longrightarrow \quad \rho_w(\theta,\vartheta) \leq C\varepsilon\delta^q.$$

Thus we have for every $\theta \in \Theta^*$

$$N(B_{(\Theta,d)}(\theta;\delta), \rho_w; C\varepsilon\delta^q) \leq N(B_{(\Theta,d)}(\theta;\delta), d; \varepsilon^{1/q}\delta).$$

The right hand side is bounded by $\{(2\delta)/(\varepsilon^{1/q}\delta) + 1\}^r$ for every $\varepsilon \in (0,1]$, where r is the dimension of Θ. Hence, noting also $N(B_{(\Theta,d)}(\theta;\delta), \rho_w; C\delta^q) = 1$, we obtain

$$\begin{aligned} & C^{-1}\delta^{-q} \int_0^\infty \sqrt{\log N(B_{(\Theta,d)}(\theta;\delta), \rho_w; \varepsilon)}d\varepsilon \\ = & \int_0^1 \sqrt{\log N(B_{(\Theta,d)}(\theta;\delta), \rho_w; C\varepsilon\delta^q)}d\varepsilon \\ \leq & \int_0^1 \sqrt{r\log\{2\varepsilon^{-1/q} + 1\}}d\varepsilon \ (= K_0 \text{ say}) \ < \infty. \end{aligned}$$

On the other hand, by putting $\varepsilon = 1$ in (5.2.9) we obtain $\text{Diam}(B_{(\Theta,d)}(\theta;\delta), \rho_w) \leq 2C\delta^q$. Hence, (5.2.6) and (5.2.7) are satisfied with $\varphi(\delta) = C \cdot (K_0 \vee 2) \cdot \delta^q$. □

In so-called "regular" parametric models, the condition (5.2.8) is satisfied with $p = 2$ and $q = 1$, which leads to the "square root asymptotics". The "cube root asymptotics" investigated by Kim and Pollard (1990), whose origin goes back at least to Chernoff (1964), corresponds to the case of $p = 2$ and $q = 1/2$.

In both theorems, we have to show the consistency of estimators somehow. Thus let us state here a sufficient condition based on Corollary 3.2.3 of van der Vaart and Wellner (1996).

Proposition 5.2.3 *Let (Θ, d) be a separable metric space. For a given class $w = \{w_\theta : \theta \in \Theta\} \subset \mathcal{L}^2[0,1]$, introduce ρ_w, Γ^n and γ given by (5.2.2), (5.2.3) and (5.2.4). Suppose also that there exists a countable, d-dense subset Θ^* of Θ such that ρ_w defines a proper metric on Θ^*. Suppose that it holds for a given point $\theta_0 \in \Theta$ that*

$$\gamma(\theta_0) > \sup_{\theta \notin G} \gamma(\theta) \qquad (5.2.10)$$

for every d-open set $G \subset \Theta$ that contains θ_0, and that

$$\int_0^\infty \sqrt{\log N(\Theta, \rho_w; \varepsilon)}\, d\varepsilon < \infty.$$

Then, for any Θ^*-valued random sequence $\widehat{\theta}^n$ such that

$$\Gamma^n(\widehat{\theta}^n) \geq \sup_{\theta \in \Theta^* \cup \{\theta_0\}} \Gamma^n(\theta) - \epsilon_n$$

for some $\epsilon_n = o_{P^*}(1)$, it holds that $d(\widehat{\theta}^n, \theta_0) = o_{P^*}(1)$.

Proof. We will apply (i) of Corollary 3.2.3 of van der Vaart and Wellner (1996) to $\Theta^\star = \Theta^* \cup \{\theta_0\}$ (if $\rho_w(\theta_0^*, \theta_0) = 0$ for some $\theta_0^* \in \Theta^*$, set $\Theta^\star = (\Theta^* \setminus \{\theta_0^*\}) \cup \{\theta_0\}$ to make $(\Theta^\star, \rho_w)$ a proper metric space). Notice that the inequality (5.2.10) still holds for any $G \subset \Theta^\star$ containing θ_0 which is open in the relative topology. Theorem 2.4.2 yields that

$$\begin{aligned} E \sup_{\theta \in \Theta^\star} |\Gamma^n(\theta) - \gamma(\theta)| &\leq C \cdot n^{-1/2} \int_0^D \sqrt{\log(1 + N(\Theta, \rho_w; \varepsilon))}\, d\varepsilon \\ &\leq C \cdot n^{-1/2} \left\{ D\sqrt{\log 2} + \int_0^D \sqrt{\log N(\Theta, \rho_w; \varepsilon)}\, d\varepsilon \right\}, \end{aligned}$$

where $D = \mathrm{Diam}(\Theta, \rho_w) < \infty$ by our entropy condition, and $C > 0$ is a universal constant. This implies that

$$\sup_{\theta \in \Theta^\star} |\Gamma^n(\theta) - \gamma(\theta)| = o_P(1),$$

which completes the proof. □

5.2.2 Examples

Example 1: Peak point of F

Let us consider estimating the value of

$$\theta_0 = \operatorname*{argmax}_{\theta\in[0,1]} F(\theta),$$

where $t \rightsquigarrow F(t)$ is the cumulative function of f defined by $F(t) = \int_0^t f(s)ds$. This problem can be treated in our general framework by setting

$$w_\theta(t) = 1_{[0,\theta]}(t) \qquad \forall\theta \in [0,1].$$

The criterion process and the contrast function, defined by (5.2.3) and (5.2.4), turn out to be $\Gamma^n(\theta) = X^n_\theta$ and $\gamma(\theta) = F(\theta)$, respectively.

We equip $\Theta = [0,1]$ with the usual metric $d(\theta,\vartheta) = |\theta-\vartheta|$ to apply Theorem 5.2.2. It holds that $\rho_w(\theta,\vartheta) = \sqrt{|\theta-\vartheta|}$, and thus ρ_w is defines a proper metric on $[0,1]$. The function $\theta \rightsquigarrow \gamma(\theta)$ is indeed continuous. Hence, if θ_0 is an inner point of $[0,1]$ and if there exist some constants $\delta_0, c > 0$ and $p > 1/2$ such that

$$F(\theta) - F(\theta_0) \le -c|\theta-\theta_0|^p \qquad \forall\theta \in [\theta_0-\delta_0, \theta_0+\delta_0] \tag{5.2.11}$$

then the same conclusion as Theorem 5.2.1 holds for $r_n = n^{1/(2p-1)}$.

To derive the asymptotic behavior of the rescaled residual $n^{1/(2p-1)}(\widehat{\theta}^n - \theta_0)$, let us introduce an assumption on the function $t \rightsquigarrow F(t)$.

Condition 5.2.4 *Let $p \in \mathbb{N}$ be given. For a given point $\theta_0 \in (0,1)$, the function $t \rightsquigarrow F(t)$ is $(p-1)$-times continuously differentiable in a neighborhood of θ_0 with derivatives $F^{(m)}, m = 1,...,p-1$, and has p-th left- and right-derivatives $F^{(p)}_-$ and $F^{(p)}_+$ at θ_0, respectively, which satisfy:*

- *when $p \ge 2$: $F^{(m)}(\theta_0) = 0$ for every $m = 1,...,p-1$;*
- *when p is odd: $F^{(p)}_-(\theta_0) > 0 > F^{(p)}_+(\theta_0)$;*
- *when p is even: $F^{(p)}_-(\theta_0) \vee F^{(p)}_+(\theta_0) < 0$.*

The condition (5.2.11) follows from this assumption by a Taylor expansion. Moreover, we obtain the following result.

Proposition 5.2.5 *Under Condition 5.2.4, for any $[0,1]\cap\mathbb{Q}$-valued random sequence $\widehat{\theta}^n$ such that $\widehat{\theta}^n \xrightarrow{P^*} \theta_0$ and that*

$$X^n_{\widehat{\theta}^n} \ge \sup_{\theta\in[0,1]\cap\mathbb{Q}} X^n_\theta - \epsilon_n$$

for some $\epsilon_n = o_{P^*}(n^{-p/(2p-1)})$, *it holds that* $n^{1/(2p-1)}(\widehat{\theta}^n - \theta_0) \stackrel{P}{\Longrightarrow} \text{argmax}_{h\in\mathbb{R}}\{\mathbb{A}(h)+\mathbb{B}(h)\}$ *in* $\mathbb{R}$, *where* $h \rightsquigarrow \mathbb{A}(h)$ *is the deterministic process given by*

$$\mathbb{A}(h) = \begin{cases} h^p F_+^{(p)}(\theta_0)/p!, & \forall h \geq 0, \\ h^p F_-^{(p)}(\theta_0)/p!, & \forall h < 0, \end{cases}$$

and where $h \rightsquigarrow \mathbb{B}(h)$ *is the two-sided Brownian motion, that is, a zero-mean, continuous Gaussian process such that* $E|\mathbb{B}(h_1) - \mathbb{B}(h_2)|^2 = |h_1 - h_2|$ *for every* $h_1, h_2 \in \mathbb{R}$.

Proof. It has already been shown by Theorem 5.2.2 that the sequence $n^{1/(2p-1)}(\widehat{\theta}^n - \theta_0)$ is uniformly tight. Let us set:

$$\begin{aligned} H^n &= \{h \in \mathbb{R} : \theta_0 + n^{-1/(2p-1)}h \in [0,1]\}; \\ H^{n*} &= \{h \in \mathbb{R} : \theta_0 + n^{-1/(2p-1)}h \in [0,1] \cap \mathbb{Q}\}. \end{aligned}$$

We consider the stochastic process $h \rightsquigarrow \mathbb{M}^n(h)$, with parameter in H^n, defined by

$$\begin{aligned} \mathbb{M}^n(h) &= n^{p/(2p-1)}\left\{\Gamma^n(\theta_0 + n^{-1/(2p-1)}h) - \Gamma^n(\theta_0)\right\} \\ &= \mathbb{A}^n(h) + \mathbb{B}^n(h), \end{aligned}$$

where:

$$\begin{aligned} \mathbb{A}^n(h) &= n^{p/(2p-1)} \int_0^1 \left\{w_{\theta_0 + n^{-1/(2p-1)}h}(t) - w_{\theta_0}(t)\right\} f(t)dt; \\ \mathbb{B}^n(h) &= n^{1/(4p-2)} \int_0^1 \left\{w_{\theta_0 + n^{-1/(2p-1)}h}(t) - w_{\theta_0}(t)\right\} dB_t. \end{aligned}$$

By Theorem 2.4.4, there exists a continuous version $h \rightsquigarrow \widetilde{\mathbb{B}}^n(h)$ of $h \rightsquigarrow \mathbb{B}^n(h)$. Thus we first consider $\widetilde{\mathbb{M}}^n = \mathbb{A}^n + \widetilde{\mathbb{B}}^n$ instead of $\mathbb{M}^n$. An easy computation gives that $\lim_{n\to\infty} \mathbb{A}^n(h) = \mathbb{A}(h)$ for every $h \in \mathbb{R}$. Furthermore, since $h \rightsquigarrow \mathbb{A}^n(h)$ and $h \rightsquigarrow \mathbb{A}(h)$ are continuous, this convergence is uniform on every compact set $K \subset \mathbb{R}$. On the other hand, it follows from Corollary 3.4.3 that $\widetilde{\mathbb{B}}^n \stackrel{P}{\Longrightarrow} \mathbb{B}$ in $\ell^\infty(K)$ for every compact set $K \subset \mathbb{R}$. Thus we have $\widetilde{\mathbb{M}}^n \stackrel{P}{\Longrightarrow} \mathbb{M} = \mathbb{A} + \mathbb{B}$ in $\ell^\infty(K)$ for every compact set $K \subset \mathbb{R}$. The existence and the uniqueness of the maximum point of $h \rightsquigarrow \mathbb{M}(h)$ follow from Khinchin's law of the iterated logarithm (see, e.g., page 61 of Hida (1980)) and Lemma 2.6 of Kim and Pollard (1990), respectively. Hence Theorem 3.2.2 of van der Vaart and Wellner (1996) yields the following CLAIM: for any uniformly tight sequence $\widehat{h}^n$ satisfying

$$\widetilde{\mathbb{M}}^n(\widehat{h}^n) \geq \sup_{h\in K} \widetilde{\mathbb{M}}^n(h) - \epsilon_{n,K}$$

for some $\epsilon_{n,K} = o_{P^*}(1)$ for each compact set $K \subset \mathbb{R}$, it holds that $\widehat{h}^n \stackrel{P}{\Longrightarrow} \text{argmax}_{h\in\mathbb{R}} \mathbb{M}(h)$ in $\mathbb{R}$.

Now, for every $n \in \mathbb{N}$ such that $K \subset H^n$, using the continuity of $h \rightsquigarrow \widetilde{\mathbb{M}}^n(h)$, we have

$$\begin{aligned} \sup_{h\in K} \widetilde{\mathbb{M}}^n(h) &= \sup_{h\in K\cap H^{n*}} \widetilde{\mathbb{M}}^n(h) & P\text{-almost surely} \\ &= \sup_{h\in K\cap H^{n*}} \mathbb{M}^n(h) & P\text{-almost surely} \\ &\leq \sup_{h\in H^{n*}} \mathbb{M}^n(h). \end{aligned}$$

Hence, we can apply the above CLAIM to $\widehat{h}^n = n^{1/(2p-1)}(\widehat{\theta}^n - \theta_0)$, which takes values in H^{n*} identically, in order to obtain that $n^{1/(2p-1)}(\widehat{\theta}^n - \theta_0) \stackrel{P}{\Longrightarrow} \mathrm{argmax}_{h\in\mathbb{R}} \mathbb{M}(h)$ in $\mathbb{R}$. □

Since $\Gamma^n(\theta) = X^n_\theta$, and since $\theta \rightsquigarrow X^n_\theta$ is continuous, it is possible to apply Theorem 3.4.4 to $\mathbb{B}^n$, without introducing the continuous version $\widetilde{\mathbb{B}}^n$ by Theorem 2.4.4. However, since it is not always easy to show the ρ-separability of the original family of continuous local martingales $\{\int_0^\cdot w_\theta(t)dB_t\}$, we presented the above approach using $\widetilde{\mathbb{B}}^n$. This argument is indeed necessary for Example 3 given later.

Example 2: Steepest interval of F

Fix a constant $b \in (0, 1/2)$. We aim to estimate the value of

$$\theta_0 = \operatorname*{argmax}_{\theta\in\Theta} \int_{\theta-b}^{\theta+b} f(t)dt,$$

which is the center of the interval with length $2b$ where the function $t \rightsquigarrow F(t)$ increases most rapidly. This problem fits in our general framework by setting

$$w_\theta(t) = 1_{[\theta-b,\theta+b]}(t) \qquad \forall\theta \in [b, 1-b].$$

The criterion process and the contrast function, defined by (5.2.3) and (5.2.4), turn out to be $\Gamma^n(\theta) = X^n_{\theta+b} - X^n_{\theta-b}$ and $\gamma(\theta) = F(\theta + b) - F(\theta - b)$, respectively.

Here we make an assumption which is similar to Condition 5.2.4 in the preceding example.

Condition 5.2.6 *Let an even integer $p \geq 2$ be given. For given $\theta_0 \in (b, 1-b)$, the function $t \rightsquigarrow f(t)$ is $(p-1)$-times continuously differentiable on an open set containing $\theta_0 - b$ and $\theta_0 + b$ with derivatives $f^{(m)}, m = 1, ..., p-1$, satisfying:*

- *$f^{(m)}(\theta_0 - b) = f^{(m)}(\theta_0 + b)$ for every $m = 0, ..., p-2$;*
- *$f^{(p-1)}(\theta_0 - b) > f^{(p-1)}(\theta_0 + b)$.*

Proposition 5.2.7 *Under Condition 5.2.6, for any $[b,1-b]\cap\mathbb{Q}$-valued random sequence $\widehat{\theta}^n$ such that $\widehat{\theta}^n \xrightarrow{P^*} \theta_0$ and that*

$$X^n_{\widehat{\theta}^n+b} - X^n_{\widehat{\theta}^n-b} \geq \sup_{\theta\in[b,1-b]\cap\mathbb{Q}} \left\{ X^n_{\theta+b} - X^n_{\theta-b} \right\} - \epsilon_n$$

for some $\epsilon_n = o_{P^*}(n^{-p/(2p-1)})$, it holds that $n^{1/(2p-1)}(\widehat{\theta}^n - \theta_0) \xrightarrow{P} \operatorname{argmax}_{h\in\mathbb{R}}\{\mathbb{A}(h)+\mathbb{B}(h)\}$ in $\mathbb{R}$, where $h \rightsquigarrow \mathbb{A}(h)$ is the deterministic process given by

$$\mathbb{A}(h) = 2^{-1/2}h^p\{f^{(p-1)}(\theta_0+b) - f^{(p-1)}(\theta_0-b)\}/p! \qquad \forall h\in\mathbb{R},$$

and where $h \rightsquigarrow \mathbb{B}(h)$ is the two-sided Brownian motion.

Proof. It follows from Condition 5.2.6 and a Taylor expansion that

$$\gamma(\theta) - \gamma(\theta_0) = \frac{f^{(p-1)}(\tilde{\theta}_+) - f^{(p-1)}(\tilde{\theta}_-)}{p!}(\theta-\theta_0)^p,$$

where $\tilde{\theta}_+$ (resp. $\tilde{\theta}_-$) is a point on the segment connecting $\theta+b$ and θ_0+b (resp. $\theta-b$ and θ_0-b). Thus, since p is even, it holds that $\gamma(\theta)-\gamma(\theta_0) \leq -c|\theta-\theta_0|^p$ in a neighborhood of θ_0 for a constant $c>0$. On the other hand, it is clear that $\rho_w(\theta,\vartheta) = \sqrt{2|\theta-\vartheta|}$, and thus ρ_w defines a proper metric on $[b,1-b]$. Hence Theorem 5.2.2 implies that $n^{1/(2p-1)}(\widehat{\theta}^n-\theta_0)$ is uniformly tight. Repeating the same argument as that in the proof of Proposition 5.2.5 to the stochastic process $h \rightsquigarrow \mathbb{M}^n(h)$ defined by

$$\mathbb{M}^n(h) = 2^{-1/2}n^{p/(2p-1)}\left\{ (X^n_{\theta_0+b+n^{-1/(2p-1)}h} - X^n_{\theta_0+b}) - (X^n_{\theta_0-b+n^{-1/(2p-1)}h} - X^n_{\theta_0-b}) \right\},$$

we can obtain the assertion. □

Example 3: Jump point of f

Let us introduce a model for the estimation problem of jump point of f.

Condition 5.2.8 *For an inner point θ_0 of $[0,1]$, there exists a constant $a\in(0,1/2)$ such that the function $t \rightsquigarrow f(t)$ is càdlàg on the interval $[\theta_0-a,\theta_0+a]$ and that*

$$D = (R_\star - L^\star) - (L^\star - L_\star) \vee (R^\star - R_\star) > 0$$

where

$$\begin{aligned} L^\star &= \sup_{t\in[\theta_0-a,\theta_0)} f(t), & R^\star &= \sup_{t\in[\theta_0,\theta_0+a]} f(t), \\ L_\star &= \inf_{t\in[\theta_0-a,\theta_0)} f(t), & R_\star &= \inf_{t\in[\theta_0,\theta_0+a]} f(t). \end{aligned}$$

The constant $a > 0$ in the above assumption should be known to construct the estimator given later, but we do not specify any concrete shape of the function $t \rightsquigarrow f(t)$, even the value of the constant $D > 0$. Condition 5.2.8 means that the function $t \rightsquigarrow f(t)$ has a positive jump at θ_0, namely $f(\theta_0) - f(\theta_0-) \geq R_\star - L^\star$, which is the biggest one in the interval $[\theta_0 - a, \theta_0 + a]$. This interpretation shows how natural this assumption is in the present context.

Let the parameter space $\Theta = [a, 1-a]$ be equipped with the Euclidean metric $d(\theta, \vartheta) = |\theta - \vartheta|$. Fixing a constant $b \in (0, a)$ we define

$$w_\theta(t) = k_b(t-\theta) \qquad \forall \theta \in [a, 1-a], \tag{5.2.12}$$

where

$$k_b(x) = \begin{cases} -x-b, & x \in [-b, 0), \\ -x+b, & x \in [0, b], \\ 0, & \text{otherwise.} \end{cases}$$

Proposition 5.2.9 *Under Condition 5.2.8, consider the criterion process $\theta \rightsquigarrow \Gamma^n(\theta) = \int_0^1 w_\theta(t)dX_t^n$ with w_θ given by (5.2.12). For any $[a, 1-a] \cap \mathbb{Q}$-valued random sequence $\widehat{\theta}^n$ such that $\widehat{\theta}^n \xrightarrow{P^*} \theta_0$ and that*

$$\Gamma^n(\widehat{\theta}^n) \geq \sup_{\theta \in [a, 1-a] \cap \mathbb{Q}} \Gamma^n(\theta) - \epsilon_n$$

for some $\epsilon_n = o_{P^*}(n^{-1})$, it holds that $n(\widehat{\theta}^n - \theta_0) \xrightarrow{P} \text{argmax}_{h \in \mathbb{R}}\{\mathbb{A}(h) + \mathbb{B}(h)\}$ in $\mathbb{R}$, where $h \rightsquigarrow \mathbb{A}(h)$ is the deterministic process given by

$$\mathbb{A}(h) = \begin{cases} h\left\{(2b)^{-1}\int_{\theta_0-b}^{\theta_0+b} f(t)dt - f(\theta_0)\right\}, & \forall h \geq 0, \\ h\left\{(2b)^{-1}\int_{\theta_0-b}^{\theta_0+b} f(t)dt - f(\theta_0-)\right\}, & \forall h < 0, \end{cases}$$

and where $h \rightsquigarrow \mathbb{B}(h)$ is the two-sided Brownian motion.

Proof. It holds that for any $\theta \in [\theta_0, \theta_0 + a - b]$

$$\begin{aligned} \gamma(\theta) - \gamma(\theta_0) &\leq -(2b - |\theta - \theta_0|)R_\star|\theta - \theta_0| + |\theta - \theta_0|(R^\star + L^\star)b \\ &\leq -|\theta - \theta_0|\{b[(R_\star - L^\star) - (R^\star - R_\star)] - |\theta - \theta_0|R_\star\} \\ &\leq -|\theta - \theta_0|\{bD - |\theta - \theta_0|R_\star\} \end{aligned}$$

and that, in the same way, for any $\theta \in [\theta_0 - a + b, \theta_0)$

$$\gamma(\theta) - \gamma(\theta_0) \leq -|\theta - \theta_0|\{bD - |\theta - \theta_0|L^\star\}.$$

Thus, choosing sufficiently small constants $\delta_0, c > 0$ we have $\gamma(\theta) - \gamma(\theta_0) \leq -c|\theta - \theta_0|$ for every $\theta \in [\theta_0 - \delta_0, \theta_0 + \delta_0]$. On the other hand, an easy computation implies that

$\rho_w(\theta, \vartheta) \le C\sqrt{|\theta - \vartheta|}$ with $C = \sqrt{4b^2 + 6b}$. Hence Theorem 5.2.2 yields that the rate of convergence in this model is $r_n = n$. Repeat the same argument as Proposition 5.2.5 to the stochastic process $h \rightsquigarrow \mathbb{M}^n(h)$ defined by $\mathbb{M}^n(h) = (2b)^{-1}n\{\Gamma^n(\theta_0 + n^{-1}h) - \Gamma^n(\theta_0)\}$ to get the assertion. □

5.2.3 Remarks for Non-Gaussian Cases

Instead of the Gaussian white noise model (5.2.1), let us consider the following:

$$dX_t^n = f(t)dt + dM_t^n, \qquad X_0^n = x_0 \in \mathbb{R},$$

where $f \in \mathcal{L}^2[0,1]$ is as before, and $M^n = (M_t^n)_{t\in[0,1]}$ is a continuous local martingale, defined on a stochastic basis $\mathbf{B}$, with the quadratic covariation given by $\langle M^n, M^n \rangle_t = \int_0^t g^n(s)ds$. Then, all results in Subsection 5.2.1 remain true whenever

$$\sup_{t\in[0,1]} g^n(t) = O_P(n^{-1}).$$

Furthermore, if

$$\sup_{t\in[0,1]} |ng^n(t) - 1| = o_P(1),$$

then all results in Subsection 5.2.2 also hold without change of limit distributions. More generally, if there exists $g \in \mathcal{L}^\infty[0,1]$ such that

$$\sup_{t\in[0,1]} |ng^n(t) - g(t)| = o_P(1),$$

then one can get some results similar to those in Subsection 5.2.2, under some smoothness assumptions on g, with modification of limit distributions.

5.3 Multiplicative Intensity Model

Let μ^n be a 1-dimensional point process defined on a stochastic basis $(\Omega^n, \mathcal{F}^n, \mathbf{F}^n = (\mathcal{F}_t^n)_{t\in[0,1]}, P^n)$ with the predictable compensator ν^n given by

$$\nu^n(\omega; dt) = \alpha(t)Y_t^n(\omega)dt,$$

where α is a $[0,\infty)$-valued measurable function on $[0,1]$, and $t \rightsquigarrow Y_t^n$ is a $[0,\infty)$-valued predictable process. Let a class $\{w_\theta : \theta \in \Theta\}$ of bounded, measurable functions on $[0,1]$ be given. Assuming $\int_0^1 \alpha(t)dt < \infty$, we consider the contrast function given by

$$\gamma(\theta) = \int_0^1 w_\theta(t)\alpha(t)dt \qquad \forall\theta \in \Theta, \tag{5.3.1}$$

and the criterion process given by

$$\Gamma^n(\theta) = (w_\theta Y^{n-}) * \mu_1^n \qquad \forall \theta \in \Theta, \tag{5.3.2}$$

where the generalized inverse process Y^{n-} of Y^n is defined by

$$Y_t^{n-}(\omega) = \begin{cases} \frac{1}{Y_t^n(\omega)} & \text{if } Y_t^n(\omega) \geq 1, \\ 0 & \text{otherwise.} \end{cases}$$

In this section, specializing the class w, we consider two kinds of estimation problems. In both problems, we will assume the following.

Condition 5.3.1 *There exists a measurable function y on $[0,1]$, which is bounded away from zero, such that*

$$\sup_{t\in[0,1]} \left| n^{-1}Y_t^n - y(t) \right| \xrightarrow{P^n} 0.$$

Now, setting

$$\Omega^n(L) = \left\{ \sup_{t\in[0,1]} \left| n^{-1}Y_t^n - y(t) \right| \leq \frac{1}{L} \right\} \qquad L = \sup_{t\in[0,1]} \frac{2}{y(t)}, \tag{5.3.3}$$

we have for all $n \geq L$ that

$$\Gamma^n(\theta) - \gamma(\theta) = (w_\theta Y^{n-}) * (\mu^n - \nu^n)_1 \qquad \text{on the set } \Omega^n(L)$$

and that

$$\sup_{t\in[0,1]} nY_t^{n-} \leq L \qquad \text{on the set } \Omega^n(L).$$

Noting also that $\lim_{n\to\infty} P^n(\Omega^n \setminus \Omega^n(L)) = 0$, we will use these facts to establish the SECOND INEQUALITY of Theorem 5.1.2.

Problem 1: Peak point of α

Let us consider estimating the location of (approximate) peak of the function $t \rightsquigarrow \alpha(t)$. This problem can be treated by setting

$$\gamma(\theta) = \int_0^1 1_{[\theta-b,\theta+b]}(t)\alpha(t)dt \qquad \forall \theta \in (b, 1-b),$$

where $b \in (0, 1/2)$ is a given constant.

Proposition 5.3.2 *Suppose that Conditions 5.3.1 and 5.2.6 with f replaced by α are satisfied, and that the function $t \rightsquigarrow y(t)$ appearing in the former is continuous at points $\theta_0 - b$ and $\theta_0 + b$ appearing in the latter. Define $\Gamma^n(\theta)$ by (5.3.2) for the class $w =$*

$\{1_{[\theta-b,\theta+b]} : \theta \in [b, 1-b]\}$. *Then, for any* $[b, 1-b]$*-valued random sequence* $\widehat{\theta}^n$ *such that* $\widehat{\theta}^n \xrightarrow{P^{n*}} \theta_0$ *and that*

$$\Gamma^n(\widehat{\theta}^n) \geq \sup_{\theta \in [b,1-b]} \Gamma^n(\theta) - \epsilon_n$$

for some $\epsilon_n = o_{P^{n*}}(n^{-p/(2p-1)})$, *it holds that* $n^{1/(2p-1)}(\widehat{\theta}^n - \theta_0) \xrightarrow{P^n} \mathrm{argmax}_{h\in\mathbb{R}}\{\mathbb{A}(h) + \mathbb{B}(h)\}$ *in* $\mathbb{R}$, *where* $h \rightsquigarrow \mathbb{A}(h)$ *is the deterministic function given by*

$$\mathbb{A}(h) = h^p \cdot \frac{\alpha^{(p-1)}(\theta_0+b) - \alpha^{(p-1)}(\theta_0-b)}{p!} \left| \frac{\alpha(\theta_0+b)}{y(\theta_0+b)} + \frac{\alpha(\theta_0-b)}{y(\theta_0-b)} \right|^{-1/2},$$

and where $h \rightsquigarrow \mathbb{B}(h)$ *is the two-sided Brownian motion.*

Proof. First, let us apply Theorem 5.1.2 to get $|\widehat{\theta}^n - \theta_0| = O_{P^{n*}}(n^{-1/(2p-1)})$. The FIRST INEQUALITY of M-CRITERION can be shown in the same way as Proposition 5.2.7. To show the SECOND INEQUALITY, we apply (ii) of Theorem 2.2.3 to the class

$$\mathcal{W}_\delta^n = \{1_{[\theta-b,\theta+b]} Y^{n-} : \theta \in \Theta_\delta\}$$

where $\Theta_\delta = [\theta_0 - \delta, \theta_0 + \delta] \cap [b, 1-b]$. Notice that there exist some constants $\delta_0, M > 0$ such that $|\alpha(\theta+b) + \alpha(\theta-b)| \leq M$ for all $\theta \in \Theta_{\delta_0}$. For every $\delta \in (0, \delta_0)$, we construct a DFP $\Pi_\delta = \{\Pi_\delta(\varepsilon)\}_{\varepsilon\in(0,1]}$ of Θ_δ as follows: for every $\varepsilon \in (0,1]$ we divide the interval Θ_δ, with length less than 2δ, into $N_{\Pi_\delta}(\varepsilon)$ sub-intervals with length at most $2\delta\varepsilon^2$; this can be done with $N_{\Pi_\delta}(\varepsilon) \leq \frac{1}{\varepsilon^2} + 1$. Then it holds that

$$\|\mathcal{W}_\delta^n\|_{\Pi_\delta,1} \leq \sqrt{2\delta \cdot M \cdot n^{-1}L} \quad \text{on the set } \Omega^n(L)$$

and that

$$|\overline{W}_\delta^n|^2 * \nu_1^n \leq 2\delta \cdot M \cdot n^{-1}L \qquad \text{on the set } \Omega^n(L),$$

where the constant $L > 0$ and the set $\Omega^n(L)$ is given by (5.3.3). Thus we have for all $n \geq L$ that

$$E^{n*} \sup_{\theta\in\Theta_\delta} |(\Gamma^n - \gamma)(\theta) - (\Gamma^n - \gamma)(\theta_0)| 1_{\Omega^n(L)}$$
$$\lesssim \; n^{-1/2}\delta^{1/2} \left\{ \sqrt{2ML} \int_0^1 \sqrt{\log\left(\frac{1}{\varepsilon^2} + 2\right)} d\varepsilon + 1 \right\}.$$

Since Condition 5.3.1 implies that $\lim_{n\to\infty} P^n(\Omega^n \setminus \Omega^n(L)) = 0$, we can deduce from Theorem 5.1.2 with $\phi_u^n(\delta) = n^{-1/2}\delta^{1/2}$ that $|\widehat{\theta}^n - \theta_0| = O_{P^{n*}}(n^{-1/(2p-1)})$.

Next, let us derive the asymptotic behavior of the sequence of rescaled processes

$$\mathbb{M}^n(h) = n^{p/(2p-1)} C_0 \left\{ \Gamma^n(\theta_0 + n^{-1/(2p-1)} h) - \Gamma^n(\theta_0) \right\}$$

where

$$C_0 = \left| \frac{\alpha(\theta_0 + b)}{y(\theta_0 + b)} + \frac{\alpha(\theta_0 - b)}{y(\theta_0 - b)} \right|^{-1/2}.$$

Let us decompose $\mathbb{M}^n(h) = \mathbb{A}^n(h) + \mathbb{B}^n(h)$, where

$$\begin{aligned} \mathbb{A}^n(h) &= W^{n,h} * \nu_1^n, \\ \mathbb{B}^n(h) &= W^{n,h} * (\mu^n - \nu^n)_1, \end{aligned}$$

with

$$W^{n,h} = n^{p/(2p-1)} C_0 (1_{[\theta_0 + n^{-1/(2p-1)}h - b, \theta_0 + n^{-1/(2p-1)}h + b]} - 1_{[\theta_0 - b, \theta_0 + b]}) Y^{n-}.$$

An easy calculation implies that $\mathbb{A}^n(h) \xrightarrow{P^n} \mathbb{A}(h)$ for every $h \in \mathbb{R}$. Furthermore, since $h \rightsquigarrow \mathbb{A}^n(h)$ and $h \rightsquigarrow \mathbb{A}(h)$ are continuous, this convergence is uniform on every compact set $K \subset \mathbb{R}$. On the other hand, Theorem 3.2.4 yields that $\mathbb{B}^n \overset{P^n}{\Longrightarrow} \mathbb{B}$ in $\ell^\infty(K)$ for every compact set $K \subset \mathbb{R}$. Hence, the same argument as Proposition 5.2.5 implies the conclusion. □

Problem 2: Jump point of α

The second problem is concerned with a jump point of the hazard function $t \rightsquigarrow \alpha(t)$.

Proposition 5.3.3 *Assume Conditions 5.3.1 and 5.2.8 with f replaced by α. Fixing a constant $b \in (0, a)$, define $\Gamma^n(\theta) = (w_\theta Y^{n-}) * \mu_1^n$, where $w_\theta = k_b(t - \theta)$ is given by (5.2.12). Then, for any $[a, 1-a]$-valued random sequence $\widehat{\theta}^n$ such that $\widehat{\theta}^n \xrightarrow{P^{n*}} \theta_0$ and that*

$$\Gamma^n(\widehat{\theta}^n) \geq \sup_{\theta \in [a, 1-a]} \Gamma^n(\theta) - \epsilon_n$$

$\epsilon_n = o_{P^{n*}}(n^{-1})$, it holds that $|\widehat{\theta}^n - \theta_0| = O_{P^{n*}}(n^{-1})$.

Proof. The FIRST INEQUALITY of M-CRITERION of Theorem 5.1.2 with $p = 1$ follows from the same computation as Proposition 5.2.9. Using Theorem 2.2.3 again, we can show that the SECOND INEQUALITY holds with the truncation by the set $\Omega^n(L)$ defined by (5.3.3) and with $\phi_u^n(\delta) = n^{-1/2}\delta^{1/2}$. □

We do not derive the asymptotic behavior of a sequence of "rescaled processes" as in the proof of Proposition 5.3.2; the Lindeberg condition is not satisfied in the present situation, thus no result in Section 3.2 works well. It is conjectured that the limit of $n(\widehat{\theta}^n - \theta_0)$ would be the argmax of a process with jumps.

5.A Notes

The general criterion Theorem 5.1.2 is an adaptation of Theorem 3.4.1 of van der Vaart and Wellner (1996) which is a fruit of recent works by some other authors, including Birgé and Massart (1993), van de Geer (1995a) and Wong and Shen (1995). The two inequalities of M-CRITERION are from, e.g., (1.2) and (1.3) of van de Geer (1995a), respectively.

The problem of estimating the mode of a density function, which is the motivation of Propositions 5.2.7 and 5.3.2, was studied by Chernoff (1964). It is treated also by Kim and Pollard (1990) in their systematic study of the cube root asymptotics. A progress of our results is that the smoothness around the mode has come into our scope; this is possible also in the i.i.d. case although we did not state in the main text. We continue this study further in Section 7.2.

Related to Propositions 5.2.9 and 5.3.3, let us briefly review some known results of jump point estimation. The asymptotic distribution of the maximum likelihood estimator $\widehat{\theta}^n$ of a jump point θ_0 of $t \rightsquigarrow f(t)$ in the Gaussian white noise model (5.2.1) can be found in Ibragimov and Has'minskii (1981, Section VII.2). More precisely, they derived the asymptotic behavior of $n(\widehat{\theta}^n - \theta_0)$ when the function f is of the form $f_\theta(t) = S(t-\theta)$ with S being a known function, along the approach of finite-dimensional parametric estimation. Korostelev (1987) showed the rate of convergence is still order n in a certain non-parametric model. Wang (1995) considered a broader model, including not only jumps but also cusps, and derived that the rate of convergence of a jump point estimator is $n|\log n|^{-\eta}$ with any constant $\eta > 0$, which is quite close to the best rate. Our setting is more general than that of Korostelev (1987), but does not contain that of Wang (1995). The point of Proposition 5.2.7 is that we have gotten the asymptotic distribution result of the rate n. By contrary, Proposition 5.3.3 gives the rate n only. See also Müller and Wang (1990) who considered estimating the point where a hazard function changes most rapidly.

Chapter 6

Non-parametric Maximum Likelihood Estimators

6.1 Gaussian White Noise Model

Let $\tau > 0$ be a fixed constant, and Θ a subset of $\mathcal{L}^2[0,\tau]$. For every $n \in \mathbb{N}$, let $t \rightsquigarrow X_t^n$ be a continuous, adapted process on a filtered measurable space $(\Omega^n, \mathcal{F}^n, \mathbf{F}^n = (\mathcal{F}_t^n)_{t\in[0,\tau]})$, and let $\mathbf{P}^n = \{P_\theta^n : \theta \in \Theta\}$ be a family of probability measures on $(\Omega^n, \mathcal{F}^n)$ indexed by Θ. Suppose that X^n is a special semimartingale under P_θ^n with the canonical decomposition

$$dX_t^n = \theta(t)dt + n^{-1/2}dB_t^{n,\theta}, \qquad X_0^n = x_0 \in \mathbb{R},$$

where $t \rightsquigarrow B^{n,\theta}$ is a standard Brownian motion defined on $(\Omega^n, \mathcal{F}^n, \mathbf{F}^n, P_\theta^n)$. It is well-known that, under some conditions, the log-likelihood ratio is given by

$$\text{(6.1.1)} \quad \log \frac{P_\theta^n|\mathcal{F}_\tau^n}{P_\vartheta^n|\mathcal{F}_\tau^n} = \int_0^\tau \{\theta(t) - \vartheta(t)\}dX_t^n - \frac{1}{2}\left\{||\theta||^2_{\mathcal{L}^2[0,1]} - ||\vartheta||^2_{\mathcal{L}^2[0,1]}\right\} \quad \forall \theta, \vartheta \in \Theta$$

(see e.g. Theorem III.5.34 of Jacod and Shiryaev (1987)). Thus the maximum likelihood estimator is the maximizer of the criterion process $\theta \rightsquigarrow \Gamma^n(\theta)$ defined by

$$\text{(6.1.2)} \qquad \Gamma^n(\theta) = \int_0^\tau \theta(t)dX_t^n - \frac{1}{2}||\theta||^2_{\mathcal{L}^2[0,\tau]} \qquad \forall \theta \in \mathcal{L}^2[0,\tau].$$

For every $\theta_0 \in \Theta$, the corresponding contrast function $\theta \rightsquigarrow \gamma_{\theta_0}(\theta)$ under $P_{\theta_0}^n$ turns out to be

$$\text{(6.1.3)} \qquad \begin{aligned} \gamma_{\theta_0}(\theta) &= \int_0^\tau \theta(t)\theta_0(t)dt - \frac{1}{2}||\theta||^2_{\mathcal{L}^2[0,\tau]} \qquad \forall \theta \in \mathcal{L}^2[0,\tau] \\ &= -\frac{1}{2}||\theta - \theta_0||^2_{\mathcal{L}^2[0,\tau]} + \frac{1}{2}||\theta_0||^2_{\mathcal{L}^2[0,\tau]}. \end{aligned}$$

We shall *not* use the fact that (6.1.1) gives the log-likelihood ratio, and thus $\Gamma^n(\theta)$ can be thought just as a random variable defined by (6.1.2). Moreover, we have defined $\Gamma^n(\theta)$

and $\gamma_{\theta_0}(\theta)$ for all elements θ of $\mathcal{L}^2[0,\tau]$ in order to consider "sieved" maximum likelihood estimators. In view of (6.1.3) and the FIRST INEQUALITY of M-CRITERION of Theorem 5.1.2, it is natural to adopt the L^2-semimetric as the canonical semimetric d_u^n on the parameter space Θ, that is, $d_u^n(\theta,\vartheta) = \|\theta - \vartheta\|_{\mathcal{L}^2[0,\tau]}$.

Let U be a subset of Θ (one may take $U = \Theta$ or $\{\theta_0\}$ for instance). We now introduce a *local* entropy condition on a "sieve" $\Theta^n \subset \mathcal{L}^2[0,\tau]$, which need not be contained in Θ, *uniformly* over U. Throughout this section, we denote by $B(\theta;\delta)$ the closed ball in $\mathcal{L}^2[0,\tau]$ with center θ and $\|\cdot\|_{\mathcal{L}^2[0,\tau]}$-radius δ.

Condition 6.1.1 *Let a set $U \subset \Theta \subset \mathcal{L}^2[0,\tau]$ and some countable sets $\Theta^n \subset \mathcal{L}^2[0,\tau]$ be given. For every $n \in \mathbb{N}$ and $\theta_0 \in U$, there exist a proper metric $\widehat{\rho}_{\theta_0}^n$ on Θ^n such that $\|\cdot\|_{\mathcal{L}^2[0,\tau]} \leq \widehat{\rho}_{\theta_0}^n$, and a function $\varphi_{\theta_0}^n : (0,\infty) \to (0,\infty)$ such that $\delta \rightsquigarrow \delta^{-1}\varphi_{\theta_0}^n(\delta)$ is decreasing and that*

$$\text{(6.1.4)} \qquad \int_0^\delta \sqrt{\log(1 + N(\Theta^n \cap B(\theta_0;\delta), \widehat{\rho}_{\theta_0}^n;\varepsilon))}\,d\varepsilon \leq \varphi_{\theta_0}^n(\delta) \qquad \forall \delta \in (0,\infty).$$

Then, choose some positive constants r_{n,θ_0} such that $n^{-1/2}\varphi_{\theta_0}^n(r_{n,\theta_0}^{-1}) \leq r_{n,\theta_0}^{-2}$.

Theorem 6.1.2 *Let a set $U \subset \Theta \subset \mathcal{L}^2[0,\tau]$ and some countable sets $\Theta^n \subset \mathcal{L}^2[0,\tau]$ be given. Suppose that Condition 6.1.1 is satisfied, and choose some constants r_{n,θ_0} described there. Suppose also that there exists a constant $M > 0$ such that: for every $n \in \mathbb{N}$ and $\theta_0 \in U$ there exists $\theta_{\theta_0}^n \in \Theta^n$ such that*

$$\text{(6.1.5)} \qquad \|\theta_0 - \theta_{\theta_0}^n\|_{\mathcal{L}^2[0,\tau]} \leq M r_{n,\theta_0}^{-1}.$$

Then, for any mapping $\widehat{\theta}^n : \Omega^n \to \Theta^n$ such that

$$\text{(6.1.6)} \qquad \Gamma^n(\widehat{\theta}^n) \geq \sup_{\theta\in\Theta^n} \Gamma^n(\theta) - r_n^{-2} \quad \text{with} \quad r_n = \sup_{\theta_0\in U} r_{n,\theta_0},$$

it holds that

$$\lim_{L\to\infty} \limsup_{n\to\infty} \sup_{\theta_0\in U} P_{\theta_0}^{n*}\left(r_{n,\theta_0}\|\widehat{\theta}^n - \theta_0\|_{\mathcal{L}^2[0,\tau]} > L\right) = 0.$$

It is trivial that the assumption (6.1.5) is not a real restriction when $\Theta^n = \Theta$; it is satisfied with $\theta_{\theta_0}^n = \theta_0$. On the other hand, it should be noted that the positive constant $M = M_U$ appearing there may depend on U; the case of $U = \Theta$ leads to the rate of convergence uniformly in the true parameter θ_0, while the case of $U = \{\theta_0\}$ implies the point-wise assertion only. However, from the practical point of view we should choose a sieve Θ^n satisfying $\Theta \subset \bigcup_{\theta\in\Theta^n} B(\theta; M^* r_n^{-1})$ with a positive constant M^* not depending on U even in the case of $U = \{\theta_0\}$, because statisticians do not know which is the true point θ_0.

Proof. We will apply Theorem 5.1.2. Formulation 5.1.1 should be read as follows: for every $\theta_0 \in U(=U^n)$,

(i) the semimetric space $(\Theta^n, \|\cdot\|_{\mathcal{L}^2[0,\tau]})$ and the point $\theta^n_{\theta_0} \in \Theta^n$;

(ii) the criterion processes $\theta \rightsquigarrow \Gamma^n(\theta)$ and the contrast functions $\theta \rightsquigarrow \gamma_{\theta_0}(\theta)$, with parameters in Θ^n, defined by (6.1.2) and (6.1.3), respectively.

We then denote $R^n_{\theta_0}(\delta) = \left\{\theta \in \Theta^n : (\delta/2) < \|\theta - \theta^n_{\theta_0}\|_{\mathcal{L}^2[0,\tau]} \leq \delta\right\}$ for every $\delta \in (0,\infty)$.

To show the FIRST INEQUALITY of M-CRITERION with $p = 2$, first observe that

$$\begin{aligned}
\|\theta^n_{\theta_0} - \theta_0\|_{\mathcal{L}^2[0,\tau]} &\leq Mr^{-1}_{n,\theta_0} \\
&\leq \frac{1}{4}\cdot\frac{\delta}{2} \qquad \text{whenever } \delta \geq 8Mr^{-1}_{n,\theta_0} \\
&\leq \frac{1}{4}\|\theta - \theta^n_{\theta_0}\|_{\mathcal{L}^2[0,\tau]} \qquad \text{whenever } \theta \in R^n_{\theta_0}(\delta).
\end{aligned}$$

Thus we have for every $\delta \geq 8Mr^{-1}_{n,\theta_0}$ and every $\theta \in R^n_{\theta_0}(\delta)$

$$\begin{aligned}
\gamma_{\theta_0}(\theta) - \gamma_{\theta_0}(\theta^n_{\theta_0}) &= \frac{1}{2}\left\{-\|\theta - \theta_0\|^2_{\mathcal{L}^2[0,\tau]} + \|\theta^n_{\theta_0} - \theta_0\|^2_{\mathcal{L}^2[0,\tau]}\right\} \\
&\leq \frac{1}{2}\left\{-\|\theta - \theta^n_{\theta_0}\|^2_{\mathcal{L}^2[0,\tau]} + 2\|\theta - \theta^n_{\theta_0}\|_{\mathcal{L}^2[0,\tau]}\cdot\|\theta^n_{\theta_0} - \theta_0\|_{\mathcal{L}^2[0,\tau]}\right\} \\
&\leq \frac{1}{2}\left\{-\|\theta - \theta^n_{\theta_0}\|^2_{\mathcal{L}^2[0,\tau]} + 2\|\theta - \theta^n_{\theta_0}\|_{\mathcal{L}^2[0,\tau]}\cdot\frac{\|\theta - \theta^n_{\theta_0}\|_{\mathcal{L}^2[0,\tau]}}{4}\right\} \\
&= -\frac{1}{4}\|\theta - \theta^n_{\theta_0}\|^2_{\mathcal{L}^2[0,\tau]} \\
&\leq -\frac{1}{16}\delta^2.
\end{aligned}$$

This means that the FIRST INEQUALITY holds for $c_\varepsilon = \frac{1}{16}$ and $K_\varepsilon = 8M$ (with $\delta_0 = \infty$). On the other hand, by (6.1.5) we have $R^n_{\theta_0}(\delta) \subset \Theta^n \cap B(\theta_0; \frac{9}{8}\delta)$ whenever $\delta \in [8Mr^{-1}_{n,\theta_0}, \infty)$. Thus Theorem 2.4.2 implies that

$$\begin{aligned}
&E^{n*}_{\theta_0} \sup_{\theta \in R^n_{\theta_0}(\delta)} |(\Gamma^n - \gamma)(\theta) - (\Gamma^n - \gamma)(\theta^n_{\theta_{\theta_0}})| \\
&\lesssim n^{-1/2}\int_0^\delta \sqrt{\log(1 + N(\Theta^n \cap B(\theta_0; \tfrac{9}{8}\delta), \widetilde{\rho}^n_{\theta_0}; \varepsilon))}d\varepsilon \\
&\leq n^{-1/2}\varphi^n_{\theta_0}(\tfrac{9}{8}\delta) \leq \tfrac{9}{8}\cdot n^{-1/2}\varphi^n_{\theta_0}(\delta),
\end{aligned}$$

which means that the SECOND INEQUALITY holds with $\phi^n_{\theta_0} = n^{-1/2}\varphi^n_{\theta_0}$. Hence Theorem 5.1.2 yields that

$$\lim_{L\to\infty}\limsup_{n\to\infty}\sup_{\theta_0\in U} P^{n*}_{\theta_0}\left(r_{n,\theta_0}\|\widehat{\theta}^n - \theta^n_{\theta_0}\|_{L^2[0,1]} > L\right) = 0.$$

Recalling (6.1.5), we obtain the assertion. □

Example 1: Monotone functions

Let us set Θ to be the class of monotone functions $\theta : [0, \tau] \to [0, 1]$. Then it follows from Theorem 2.7.5 of van der Vaart and Wellner (1996) that

$$\int_0^\delta \sqrt{\log N(\Theta, \|\cdot\|_{\mathcal{L}^2[0,\tau]}; \varepsilon)} d\varepsilon \leq \text{const.}\delta^{1/2} \qquad \forall \delta > 0.$$

This suggests that, by choosing a sieve Θ^n appropriately, Condition 6.1.1 should be fulfilled with $\varphi^n_{\theta_0}(\delta) = \text{const.}(\delta^{1/2} \vee \delta)$, which leads to the rate $r_n = n^{1/3}$, not depending on $\theta_0 \in U = \Theta$.

Proposition 6.1.3 *Choosing any grids* $0 = t^n_0 < t^n_1 < \cdots < t^n_{k_n} = \tau$ *such that* $t^n_i - t^n_{i-1} \leq n^{-2/3}$, *define* Θ^n *as the class of monotone functions* $\theta : [0, \tau] \to V^n$ *which are piecewise constant on each interval* $[t^n_{i-1}, t^n_i)$, *where* $V^n = \{j \cdot n^{-2/3} : j \in \mathbb{Z}\} \cap [0, 1]$. *Then, it holds that* $\Theta \subset \bigcup_{\theta \in \Theta^n} B(\theta; \sqrt{\tau + 1} n^{-1/3})$.

Proof. Fix any $\theta \in \Theta$. Let us choose $\theta^u, \theta^l \in \Theta^n$ given by

$$\begin{cases} \theta^u(t) &= u_i \\ \theta^l(t) &= l_i \end{cases} \qquad \text{for } t \in [t^n_{i-1}, t^n_i), \ i = 1, ..., k_n,$$

where

$$\begin{aligned} u_i &= \min\left\{ y \in V^n : \sup_{s \in [t^n_{i-1}, t^n_i)} \theta(s) \leq y \right\}, \\ l_i &= \max\left\{ y \in V^n : \inf_{s \in [t^n_{i-1}, t^n_i)} \theta(s) \geq y \right\}, \end{aligned}$$

and $\theta^u(\tau) = \theta^l(\tau) = 0$. If the function $t \rightsquigarrow \theta(t)$ is increasing, then $u_i = l_{i+1} + n^{-2/3}$. Thus we have $\|\theta - \theta^l\|^2_{\mathcal{L}^2[0,\tau]} \leq \|\theta^u - \theta^l\|^2_{\mathcal{L}^2[0,\tau]} \leq \|\theta^u - \theta^l\|_{\mathcal{L}^1[0,\tau]} \leq (\tau + 1) n^{-2/3}$. The case of $t \rightsquigarrow \theta(t)$ being decreasing is also shown in the same way. □

Since $\Theta^n \subset \Theta$, Condition 6.1.1 is indeed satisfied for $\widehat{\rho}^n_{\theta_0} = \|\cdot\|_{\mathcal{L}^2[0,\tau]}$ and $\varphi^n_{\theta_0}(\delta) = \text{const.}(\delta^{1/2} \vee \delta)$. The above proposition says that the assumption (6.1.5) is also fulfilled. Consequently, it holds for any $\widehat{\theta}^n$ satisfying (6.1.6) with Θ^n given in Proposition 6.1.3 that

$$\lim_{L \to \infty} \limsup_{n \to \infty} \sup_{\theta_0 \in \Theta} P^{n*}_{\theta_0} \left(n^{1/3} \|\widehat{\theta}^n - \theta_0\|_{\mathcal{L}^2[0,\tau]} > L \right) = 0.$$

It should be noted that grids of order $n^{-2/3}$ is sufficient to get this rate, and the discrete observation of the process $t \rightsquigarrow X^n_t$ only on the grids is enough to compute the estimator.

Example 2: Smooth functions

Let some constants $\alpha > 1/2$ and $H > 0$ be given. Let us consider the class $\Theta = C_H^\alpha([0,\tau])$ defined in (11) of Section 1.2. Recall that

$$\int_0^\delta \sqrt{\log N(C_H^\alpha([0,\tau]), \|\cdot\|_\infty; \varepsilon)}d\varepsilon \leq \text{const.}\delta^{1-(1/2\alpha)}.$$

This suggests that Condition 6.1.1 should be fulfilled with $\varphi_{\theta_0}^n(\delta) = \text{const.}(\delta^{1-(1/2\alpha)} \vee \delta)$, which leads to the rate $r_n = n^{\alpha/(2\alpha+1)}$.

Choosing any grids $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = \tau$ such that $t_i^n - t_{i-1}^n \leq n^{-\alpha/(2\alpha+1)}$, define the mapping $\pi^n : \Theta \to \ell^\infty[0,\tau]$ by

$$\pi^n\theta(t) = \sum_{i=1}^{k_n} \frac{1_{[t_{i-1}^n, t_i^n)}(t)}{t_i^n - t_{i-1}^n} \int_{t_{i-1}}^{t_i} \theta(s)ds \qquad \forall t \in [0,\tau].$$

This mapping π^n is a special case of (4.2.5). Using also Lemma 4.2.2, we have:

$$\|\pi^n\theta - \pi^n\vartheta\|_{\mathcal{L}^2[0,\tau]} \leq \|\theta - \vartheta\|_{\mathcal{L}^2[0,\tau]} \leq \tau\|\theta - \vartheta\|_\infty \qquad \forall \theta, \vartheta \in \Theta;$$

$$\|\theta - \pi^n\theta\|_{\mathcal{L}^2[0,\tau]} \leq \tau\|\theta - \pi^n\theta\|_\infty \leq \tau H n^{-\alpha/(2\alpha+1)} \qquad \forall \theta \in \Theta.$$

Hence, if we choose a sequence of countable subsets Θ^n of $\pi^n\Theta$ such that $\pi^n\Theta \subset \bigcup_{\theta\in\Theta^n} B(\theta; Mn^{-\alpha/(2\alpha+1)})$ for a constant $M > 0$ not depending on n, then Condition 6.1.1 and the assumption (6.1.5) are satisfied for $U = \Theta$, $\widehat{\rho}_{\theta_0}^n = \|\cdot\|_\infty$ and $\varphi_{\theta_0}^n(\delta) = \text{const.}(\delta^{1-(1/2\alpha)} \vee \delta)$. Thus, the assertion of Theorem 6.1.2 holds for such a sieve Θ^n, with $U = \Theta$.

Similarly to the preceding example, this result says that taking some grids of order $n^{-\alpha/(2\alpha+1)}$ is enough to get the convergence rate $r_n = n^{\alpha/(2\alpha+1)}$.

6.2 Multiplicative Intensity Model

Let $(E, \mathcal{E})$ be a Blackwell space on which a measure λ is defined, and let $\tau > 0$ be a fixed constant. Let us denote:

$$\mathcal{L}_+^p = \{f \in \mathcal{L}^p([0,\tau] \times E, \mathfrak{B}[0,\tau] \otimes \mathcal{E}, dt\lambda(dz)) : \; f(t,z) \geq 0\} \qquad \forall p \in [1,\infty].$$

We also denote by $\|\cdot\|_{\mathcal{L}^p}$ the L^p-seminorm on $\mathcal{L}^p([0,\tau] \times E, \mathfrak{B}[0,\tau] \otimes \mathcal{E}, dt\lambda(dz))$.

For every $n \in \mathbb{N}$, let μ^n be an E-valued multivariate point process on a filtered measurable space $(\Omega^n, \mathcal{F}^n, \mathbf{F}^n = (\mathcal{F}_t^n)_{t\in[0,\tau]})$. Let $\mathbf{P}^n = \{P_\theta^n : \theta \in \Theta\}$ be a family of probability measures on $(\Omega^n, \mathcal{F}^n)$ indexed by a subset Θ of $\mathcal{L}_+^p$ for some $p \in [1,\infty]$

specified later. We suppose that the predictable compensator of μ^n with respect to the probability measure P_θ^n is given by

$$\nu^{n,\theta}(\omega; dt, dz) = \theta(t, z)Y^n(\omega, t, z)dt\lambda(dz),$$

where Y^n is a $[0, \infty)$-valued predictable function. It is well-known that, under some conditions, the log-likelihood ratio is given by

$$\log \frac{dP_\theta^n|\mathcal{F}_\tau^n}{dP_\vartheta^n|\mathcal{F}_\tau^n} = l^n(\theta) - l^n(\vartheta) \qquad \forall \theta, \vartheta \in \Theta,$$

where

(6.2.1) $$l^n(\theta) = (\log \theta) * \mu_\tau^n - \theta * \overline{\nu}_\tau^n$$

with

(6.2.2) $$\overline{\nu}^n(\omega; dt, dz) = Y^n(\omega, t, z)dt\lambda(dz).$$

However, as in the preceding section, we shall not use the fact that $l^n(\theta)$ above is a component of the log-likelihood ratio; it may be regarded just as a $[-\infty, \infty)$-valued random variable defined by the right hand side of (6.2.1). As a matter of fact, in the following we will define $l^n(\theta)$ for all elements θ of $\mathcal{L}_+^p$ in order to discuss some "sieved" maximum likelihood estimators, again.

Condition 6.2.1 *For some $p, q \in [1, \infty]$ such that $(1/p) + (1/q) = 1$, it holds that*

$$\Theta \subset \mathcal{L}_+^p \qquad \textit{and} \qquad Y^n(\omega, \cdot, \cdot) \in \mathcal{L}_+^q \quad \forall \omega \in \Omega^n.$$

Furthermore, for a given subset U of Θ, it holds that

$$\lim_{L\to\infty} \limsup_{n\to\infty} \sup_{\theta_0 \in U} P_{\theta_0}^{n*}(\Omega^n \setminus \Omega_q^n(L)) = 0,$$

where

(6.2.3) $$\Omega_q^n(L) = \left\{\omega \in \Omega^n : \sqrt{n^{-1}\|Y^n\|_{\mathcal{L}^q(\omega)}} \le L\right\} \qquad \forall L > 0.$$

A typical case, considered by van de Geer (1995b), is $p = 1$ and $q = \infty$. This choice is optimal in the context of censoring models, where it indeed holds that $Y^n \le n$. For our discussion of rate of convergence, we adopt the random Hellinger semimetric ϱ^n defined by

$$\begin{aligned}\varrho^n(\theta, \vartheta) &= \sqrt{\frac{1}{n}\left|\sqrt{\theta} - \sqrt{\vartheta}\right|^2 * \overline{\nu}_\tau^n} \\ &= \sqrt{\frac{1}{n}\int_{[0,\tau]\times E}\left|\sqrt{\theta(t,z)} - \sqrt{\vartheta(t,z)}\right|^2 Y^n(t,z)dt\lambda(dz)} \qquad \forall \theta, \vartheta \in \mathcal{L}_+^p.\end{aligned}$$

Due to the first requirement of Condition 6.2.1, the above formula indeed defines a random semimetric. On the other hand, an entropy condition on the sieve should be given in terms of the L^{2p}-Hellinger semimetric ρ_{2p} defined by

$$\begin{aligned}\rho_{2p}(\theta,\vartheta) &= \left\|\sqrt{\theta}-\sqrt{\vartheta}\right\|_{\mathcal{L}^{2p}} \\ &= \left(\int_{[0,\tau]\times E}\left|\sqrt{\theta(t,z)}-\sqrt{\vartheta(t,z)}\right|^{2p} dt\lambda(dz)\right)^{1/2p} \qquad \forall\theta,\vartheta\in\mathcal{L}^p_+.\end{aligned}$$

Then it follows from the Hölder inequality that $\varrho^n \leq L\rho_{2p}$ on the set $\Omega^n_q(L)$ given by (6.2.3). Throughout this section, we denote by $B_{\rho_{2p}}(\theta;\varepsilon)$ the closed ball in $\mathcal{L}^p_+$ with center θ and ρ_{2p}-radius ε, and by $B_{\varrho^n}(\theta;\varepsilon)$ the (random) closed ball in $\mathcal{L}^p_+$ with center θ and ϱ^n-radius ε. We consider a sequence Θ^n of subsets of $\mathcal{L}^p_+$ which satisfies the following condition.

Condition 6.2.2 *Let $U \subset \Theta \subset \mathcal{L}^p_+$ be given, where $p\in[1,\infty]$. For every $n\in\mathbb{N}$ and $\theta_0\in U$, there exist a function $\varphi^n_{\theta_0}:(0,\infty)\to(0,\infty)$ and some sets $\Theta^n(\theta_0;\delta)\subset\Theta^n\subset\mathcal{L}^p_+$ for $\delta\in(0,\infty)$ such that $\delta\rightsquigarrow\delta^{-1}\varphi^n_{\theta_0}(\delta)$ is decreasing and that*

$$(6.2.4)\qquad \int_0^\delta \sqrt{\log(1+N_{[]}(\Theta^n(\theta_0;\delta),\rho;\varepsilon))}d\varepsilon \leq \varphi^n_{\theta_0}(\delta) \qquad \forall\delta\in(0,\infty).$$

Then, choose some positive constants r_{n,θ_0} such that $n^{-1/2}\varphi^n_{\theta_0}(r^{-1}_{n,\theta_0})\leq r^{-2}_{n,\theta_0}$.

The subsets $\Theta^n(\theta_0;\delta)$ of Θ^n have to be chosen to satisfy not only (6.2.4) but also (6.2.6) below. It can be taken to be $\Theta^n\cap B_{\rho_{2p}}(\theta_0;\delta)$ if the random semimetric ϱ^n is "asymptotically equivalent" to the semimetric ρ_{2p} (i.e., the assumption (6.2.8) below). Generally speaking, a smaller choice of $\Theta^n(\theta_0;\delta)$'s makes it easy to check the entropy condition (6.2.4), but does it difficult to check the condition (6.2.6). If we choose $\Theta^n(\theta_0;\delta)=\Theta^n$, the condition (6.2.6) is always satisfied; thus this choice is wise when it does not affect the inequality (6.2.4).

Theorem 6.2.3 *Let $U\subset\Theta\subset\mathcal{L}^p_+$ be given, where $p\in[1,\infty]$. Suppose that Conditions 6.2.1 and 6.2.2 are satisfied for some $\Theta^n(\theta_0;\delta)\subset\Theta^n\subset\mathcal{L}^p_+$, and choose some constants r_{n,θ_0} described there. Suppose also that there exists a constant $M>0$ such that: for every $n\in\mathbb{N}$ and $\theta_0\in U$ there exists $\theta^n_{\theta_0}\in\Theta^n$ such that*

$$(6.2.5)\qquad \theta_0\leq M\theta^n_{\theta_0} \quad \text{and} \quad \rho_{2p}(\theta_0,\theta^n_{\theta_0})\leq Mr^{-1}_{n,\theta_0},$$

and that

$$(6.2.6)\qquad \begin{aligned}&\lim_{K\to\infty}\limsup_{n\to\infty}\sup_{\theta_0\in U}P^{n*}_{\theta_0}(\Omega^n\setminus\Omega^n_{\theta_0}(K))=0\\ &\text{where}\quad \Omega^n_{\theta_0}(K)=\left\{R^n_{\theta_0}(\delta)\cup\{\theta^n_{\theta_0}\}\subset\Theta^n(\theta_0;K\delta)\quad\forall\delta\in[Kr^{-1}_{n,\theta_0},\infty)\right\}\\ &\text{with}\quad R^n_{\theta_0}(\delta)=\left\{\theta\in\Theta^n:(\delta/2)<\varrho^n(\theta,\theta^n_{\theta_0})\leq\delta\right\}.\end{aligned}$$

Then, for any mapping $\widehat{\theta}^n : \Omega^n \to \Theta^n$ *such that*

$$(6.2.7) \qquad l^n(\widehat{\theta}^n) \geq \sup_{\theta\in\Theta^n} l^n(\theta) - nr_n^{-2} \qquad \textit{with} \qquad r_n = \sup_{\theta_0\in U} r_{n,\theta_0},$$

it holds that

$$\lim_{L\to\infty} \limsup_{n\to\infty} \sup_{\theta_0\in U} P^n_{\theta_0}\left(r_{n,\theta_0}\varrho^n(\widehat{\theta}^n,\theta_0) > L\right) = 0.$$

When $\Theta^n(\theta_0;\delta) = \Theta^n$, *the assumption (6.2.6) is unnecessary. When* $\Theta^n(\theta_0;\delta) = \Theta^n \cap B_{\rho_{2p}}(\theta_0;\delta)$, *the assumption (6.2.6) is satisfied if*

$$(6.2.8) \qquad \lim_{K\to\infty} \limsup_{n\to\infty} \sup_{\theta_0\in U} P^{n*}_{\theta_0}\left(\sup_{\substack{\theta,\vartheta\in\Theta^n \\ \varrho^n(\theta,\vartheta)>Kr^{-1}_{n,\theta_0}}} \frac{\rho_{2p}(\theta,\vartheta)}{\varrho^n(\theta,\vartheta)} > K \right) = 0.$$

It is trivial that the assumption (6.2.5) is not a real restriction when $\Theta^n = \Theta$; it is satisfied with $M = 1$ and $\theta^n_{\theta_0} = \theta_0$. Recall also the remark following Theorem 6.1.2.

To prove the above result, we shall apply Theorem 5.1.2 not to the naive criterion process $\theta \rightsquigarrow \Gamma^n(\theta) = n^{-1}l^n(\theta)$ but to the process $\theta \rightsquigarrow \Gamma^n_{\theta_0}(\theta)$ given by

$$(6.2.9) \qquad \Gamma^n_{\theta_0}(\theta) = \frac{1}{n}\left\{\left(\log\frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}}\right) * \mu^n_\tau - \left(\left(\frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}} - 1\right)\theta^n_{\theta_0}\right) * \overline{\nu}^n_\tau\right\}$$

where $\theta^n_{\theta_0}$ is an element of Θ^n satisfying (6.2.5). Then it is natural to introduce the process $\theta \rightsquigarrow \gamma^n_{\theta_0}(\theta)$ given by

$$\begin{aligned}(6.2.10) \qquad \gamma^n_{\theta_0}(\theta) &= \frac{1}{n}\left\{\left(\log\frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}}\right) * \nu^{n,\theta_0}_\tau - \left(\left(\frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}} - 1\right)\theta^n_{\theta_0}\right) * \overline{\nu}^n_\tau\right\} \\ &= \frac{1}{n}\left(\left(\log\frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}}\right)\theta_0 - \left(\frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}} - 1\right)\theta^n_{\theta_0}\right) * \overline{\nu}^n_\tau,\end{aligned}$$

which can be approximated by

$$\frac{1}{n}\left(\left(\log\frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}} - \frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}} + 1\right)\theta^n_{\theta_0}\right) * \overline{\nu}^n_\tau$$

if $\theta^n_{\theta_0}$ is "close" to θ_0. We should have noticed in advance that

$$(6.2.11) \qquad \frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}} \quad \text{should be read as} \quad \frac{(\theta+\theta^n_{\theta_0})1_{\{\theta^n_{\theta_0}>0\}}}{2\theta^n_{\theta_0}} \quad (\text{where } 0/0 = 1).$$

However, it follows from the assumption (6.2.5) that $\{\theta^n_{\theta_0} = 0\} \subset \{\theta_0 = 0\}$, and we also have $P^n_{\theta_0}(\mu^n(\{\theta_0 = 0\}) = 0) = 1$. These facts allow us to adopt the notational convention (6.2.11). The merit of these "modified" processes is that

$$(6.2.12) \qquad \frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}} \geq \frac{1}{2}.$$

To justify the "modification", we should first see the following.

Lemma 6.2.4 *Under the first requirement of (6.2.5), for any mapping $\widehat{\theta}^n : \Omega^n \to \Theta^n$ satisfying (6.2.7), it holds that*

$$\Gamma^n_{\theta_0}(\widehat{\theta}^n) \geq \Gamma^n_{\theta_0}(\theta^n_{\theta_0}) - \frac{1}{2} r_n^{-2} \qquad P^n_{\theta_0}\text{-almost surely.}$$

Proof. Since the function $x \rightsquigarrow \log(x)$ is concave, it holds that

$$\begin{aligned} n\Gamma^n_{\theta_0}(\widehat{\theta}^n) &= \left(\log \frac{\widehat{\theta}^n + \theta^n_{\theta_0}}{2\theta^n_{\theta_0}}\right) * \mu^n_\tau - \frac{1}{2}(\widehat{\theta}^n - \theta^n_{\theta_0}) * \overline{\nu}^n_\tau \\ &\geq \frac{1}{2}\left(\log \frac{\widehat{\theta}^n}{\theta^n_{\theta_0}}\right) * \mu^n_\tau - \frac{1}{2}(\widehat{\theta}^n - \theta^n_{\theta_0}) * \overline{\nu}^n_\tau \qquad \text{on the set } \{\mu^n(\{\theta^n_{\theta_0} = 0\}) = 0\} \\ &\geq \frac{1}{2}(l^n(\widehat{\theta}^n) - l^n(\theta^n_{\theta_0})) \\ &\geq -\frac{1}{2} n r_n^{-2}. \end{aligned}$$

Since $\Gamma^n_{\theta_0}(\theta^n_{\theta_0}) = 0$, and since $P^n_{\theta_0}(\mu^n(\{\theta^n_{\theta_0} = 0\}) = 0) \geq P^n_{\theta_0}(\mu^n(\{\theta_0 = 0\}) = 0) = 1$, we obtain the assertion. □

For computation of the Hellinger semimetric, we will use the following inequalities:

$$\frac{1}{2}\left|\sqrt{x} - \sqrt{y}\right| \leq \left|\sqrt{\frac{x+y}{2}} - \sqrt{y}\right| \leq \left|\sqrt{x} - \sqrt{y}\right| \qquad \forall x, y \in [0, \infty); \tag{6.2.13}$$

$$\left|\sqrt{x+a} - \sqrt{y+a}\right| \leq \left|\sqrt{x} - \sqrt{y}\right| \qquad \forall x, y, a \in [0, \infty). \tag{6.2.14}$$

Proof of Theorem 6.2.3. We will apply Theorem 5.1.2. Formulation 5.1.1 should be read as follows: for every $\theta_0 \in U(= U^n)$:

(i) the random semimetric space (Θ^n, ϱ^n) and the point $\theta^n_{\theta_0}$ satisfying (6.2.5) and (6.2.6);

(ii) the stochastic processes $\theta \rightsquigarrow \Gamma^n_{\theta_0}(\theta)$ and $\theta \rightsquigarrow \gamma^n_{\theta_0}(\theta)$, with parameters in Θ^n, defined by (6.2.9) and (6.2.10), respectively.

As in (6.2.6), we denote $R^n_{\theta_0}(\delta) = \left\{\theta \in \Theta^n : (\delta/2) < \varrho^n(\theta, \theta^n_{\theta_0}) \leq \delta\right\}$ for every $\delta \in (0, \infty)$. Notice that $\Gamma^n_{\theta_0}(\theta^n_{\theta_0}) = \gamma^n_{\theta_0}(\theta^n_{\theta_0}) = 0$.

To show the FIRST INEQUALITY of M-CRITERION, let us write $\gamma^n_{\theta_0}(\theta) = (I) + (II)$, where:

$$\begin{aligned} (I) &= \frac{1}{n}\left(\left(\log \frac{\theta + \theta^n_{\theta_0}}{2\theta^n_{\theta_0}} - \frac{\theta + \theta^n_{\theta_0}}{2\theta^n_{\theta_0}} + 1\right)\theta^n_{\theta_0}\right) * \overline{\nu}^n_\tau; \\ (II) &= \frac{1}{n}\left(\left(\log \frac{\theta + \theta^n_{\theta_0}}{2\theta^n_{\theta_0}}\right)(\theta_0 - \theta^n_{\theta_0})\right) * \overline{\nu}^n_\tau. \end{aligned}$$

Since $\log x - x + 1 \leq -|\sqrt{x} - 1|^2$ for all $x > 0$, we have

$$\begin{aligned} (I) &\leq -\frac{1}{n}\left(\left|\sqrt{\frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}}} - 1\right|^2 \theta^n_{\theta_0}\right) * \overline{\nu}^n_\tau \\ &= -\left|\varrho^n\left(\frac{\theta+\theta^n_{\theta_0}}{2}, \theta^n_{\theta_0}\right)\right|^2 \\ &\leq -\frac{1}{4}\left|\varrho^n(\theta, \theta^n_{\theta_0})\right|^2 \qquad \text{by (6.2.13)} \\ &\leq -\frac{1}{16}\delta^2 \qquad \text{whenever } \theta \in R^n_{\theta_0}(\delta). \end{aligned}$$

On the other hand, since $|\log x| \leq c_2|\sqrt{x} - 1|$ for all $x \geq 1/2$ where $c_2 = (2+\sqrt{2})\log 2$, it follows from (6.2.12) that

$$\begin{aligned} |(II)| &\leq \frac{c_2}{n}\left(\left|\sqrt{\frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}}} - 1\right| \left|\sqrt{\theta^n_{\theta_0}} + \sqrt{\theta_0}\right| \left|\sqrt{\theta^n_{\theta_0}} - \sqrt{\theta_0}\right|\right) * \overline{\nu}^n_\tau \\ &\leq \frac{c_2(1+\sqrt{M})}{n}\left(\left|\sqrt{\frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}}} - 1\right| \sqrt{\theta^n_{\theta_0}} \cdot \left|\sqrt{\theta^n_{\theta_0}} - \sqrt{\theta_0}\right|\right) * \overline{\nu}^n_\tau \\ &\leq c_2(1+\sqrt{M})\varrho^n\left(\frac{\theta+\theta^n_{\theta_0}}{2}, \theta^n_{\theta_0}\right)\varrho^n(\theta^n_{\theta_0}, \theta_0) \\ &\leq c_2(1+\sqrt{M})\varrho^n(\theta, \theta^n_{\theta_0})\varrho^n(\theta^n_{\theta_0}, \theta_0) \qquad \text{by (6.2.13)} \\ &\leq c_2(1+\sqrt{M}) \cdot \delta \cdot M r^{-1}_{n,\theta_0} \qquad \text{whenever } \theta \in R^n_{\theta_0}(\delta). \end{aligned}$$

Thus, we have

$$\gamma^n_{\theta_0}(\theta) - \gamma^n_{\theta_0}(\theta^n_{\theta_0}) \leq -\frac{1}{32}\delta^2 \qquad \forall \theta \in R^n_{\theta_0}(\delta)$$

whenever $\delta \geq K r^{-1}_{n,\theta_0}$ with any $K \geq 32 \cdot c_2(1+\sqrt{M})M$.

To show the SECOND INEQUALITY, observe that

$$(\Gamma^n_{\theta_0} - \gamma^n_{\theta_0})(\theta) - (\Gamma^n_{\theta_0} - \gamma^n_{\theta_0})(\theta^n_{\theta_0}) = \frac{1}{n} \cdot W^{n,\theta} * (\mu^n - \nu^{n,\theta_0})_\tau \qquad \forall \theta \in \Theta^n,$$

where $W^{n,\theta}$'s are defined by

$$W^{n,\theta} = \log\left(\frac{\theta+\theta^n_{\theta_0}}{2\theta^n_{\theta_0}}\right) \qquad \forall \theta \in \mathcal{L}^p_+$$

(we have extended the parameter space Θ^n to $\mathcal{L}^p_+$ in the latter). For a given $\epsilon > 0$, choose some $L_\epsilon, K_\epsilon \geq 1$ and $n_\epsilon \in \mathbb{N}$ such that $P^n_{\theta_0}(\Omega^n \setminus \Omega^n_q(L_\epsilon) \cap \Omega^n_{\theta_0}(K_\epsilon)) \leq \epsilon$ holds for all $n \geq n_\epsilon$, where the sets $\Omega^n_q(L_\epsilon)$ and $\Omega^n_{\theta_0}(K_\epsilon)$ are given by (6.2.3) and (6.2.6), respectively. For every $n \geq n_\epsilon$ and $\theta_0 \in U$, fix any $\delta \in [r^{-1}_{n,\theta_0}, \infty)$, and we will apply

Theorem 2.2.8 to the class $\mathcal{W} = \{W^{n,\theta} : \theta \in \Theta^n(\theta_0; K_\epsilon\delta)\}$ of predictable functions with $a = 1$. Indeed, since $\nu^n(\omega; [0,\tau] \times E) < \infty$ for all $\omega \in \Omega^n$, we have that $\exp(|\overline{W}|) * \nu^n_\tau(\omega) < \infty$ by using also (6.2.5). For every $\varepsilon > 0$, choose (ρ_{2p}, ε)-brackets $[l^{\varepsilon,k}, u^{\varepsilon,k}]$, $k = 1, ..., N_{[]}(\Theta^n(\theta_0; K_\epsilon\delta), \rho_{2p}; \varepsilon)$, in $\mathcal{L}^p_+$, which cover the class $\Theta^n(\theta_0; K_\epsilon\delta)$. Construct a NFP Π of $\Theta^n(\theta_0; K_\epsilon\delta)$ from this series of brackets. We have that

$$|W^{n,\theta} - W^{n,\vartheta}| \leq W^{n,u^{\varepsilon,k}} - W^{n,l^{\varepsilon,k}} \qquad \text{if } l^{\varepsilon,k} \leq \theta, \vartheta \leq u^{\varepsilon,k}$$

and that, since $0 \leq 2(x - 1 - \log x) \leq |x-1|^2$ for all $x \geq 1$,

$$\begin{aligned}
&\frac{1}{4}\left(\mathcal{E}_2(|W^{n,u^{\varepsilon,k}} - W^{n,l^{\varepsilon,k}}|)\right) * \nu^{n,\theta_0}_\tau \\
&= 2\left(\exp\left(\frac{1}{2}|W^{n,u^{\varepsilon,k}} - W^{n,l^{\varepsilon,k}}|\right) - 1 - \frac{1}{2}|W^{n,u^{\varepsilon,k}} - W^{n,l^{\varepsilon,k}}|\right) * \nu^{n,\theta_0}_\tau \\
&= 2\left(\left(\sqrt{\frac{u^{\varepsilon,k} + \theta^n_{\theta_0}}{l^{\varepsilon,k} + \theta^n_{\theta_0}}} - 1 - \log\sqrt{\frac{u^{\varepsilon,k} + \theta^n_{\theta_0}}{l^{\varepsilon,k} + \theta^n_{\theta_0}}}\right)\theta_0\right) * \overline{\nu}^n_\tau \\
&\leq 2M\left(\left(\sqrt{\frac{u^{\varepsilon,k} + \theta^n_{\theta_0}}{l^{\varepsilon,k} + \theta^n_{\theta_0}}} - 1 - \log\sqrt{\frac{u^{\varepsilon,k} + \theta^n_{\theta_0}}{l^{\varepsilon,k} + \theta^n_{\theta_0}}}\right)\theta^n_{\theta_0}\right) * \overline{\nu}^n_\tau \\
&\leq M\left(\left|\sqrt{\frac{u^{\varepsilon,k} + \theta^n_{\theta_0}}{l^{\varepsilon,k} + \theta^n_{\theta_0}}} - 1\right|^2 \theta^n_{\theta_0}\right) * \overline{\nu}^n_\tau \\
&\leq M\left(\left|\sqrt{u^{\varepsilon,k} + \theta^n_{\theta_0}} - \sqrt{l^{\varepsilon,k} + \theta^n_{\theta_0}}\right|^2\right) * \overline{\nu}^n_\tau \\
&= Mn|\varrho^n(u^{\varepsilon,k} + \theta^n_{\theta_0}, l^{\varepsilon,k} + \theta^n_{\theta_0})|^2 \\
&\leq Mn|\varrho^n(u^{\varepsilon,k}, l^{\varepsilon,k})|^2 \qquad \text{by (6.2.14)} \\
&\leq L^2_\epsilon Mn|\rho_{2p}(u^{\varepsilon,k}, l^{\varepsilon,k})|^2 \qquad \text{on the set } \Omega^n_q(L_\epsilon) \\
&\leq L^2_\epsilon Mn\varepsilon^2,
\end{aligned}$$

where the set $\Omega^n_q(L_\epsilon)$ is given by (6.2.3). Thus we obtain $\|\mathcal{W}\|^{\mathcal{E}_2}_{\Pi,\tau} \leq \sqrt{4Mn}L_\epsilon$ on the set $\Omega^n_q(L_\epsilon)$. Likewise, it holds that

$$\sqrt{\mathcal{E}_2(|W^{n,\theta} - W^{n,\vartheta}|) * \nu^{\theta_0}_\tau} \leq \sqrt{4Mn}\varrho^n(\theta, \vartheta) \qquad \forall \theta, \vartheta \in \Theta^n;$$

this can be shown by computing on the sets $\{(t,z) : \theta(t,z) \leq \vartheta(t,z)\}$ and $\{(t,z) : \theta(t,z) > \vartheta(t,z)\}$ separately. This suggests that we should apply Theorem 2.2.8 to the random metric $\varrho = \sqrt{4Mn}\varrho^n$. Notice that $\|\sqrt{4Mn}\varrho^n\|_\Pi \leq \sqrt{4Mn}L_\epsilon$ on the set $\Omega^n_q(L_\epsilon)$. Hence, applying the theorem to $K = \sqrt{4Mn}L_\epsilon K_\epsilon$ and $\delta > 0$ we obtain

$$E^n_{\theta_0} \sup_{\substack{\theta,\vartheta \in \Theta^n(\theta_0; K_\epsilon\delta) \\ \varrho^n(\theta,\vartheta) \leq \delta}} |(W^{n,\theta} - W^{n,\vartheta}) * (\mu^n - \nu^{n,\theta_0})_\tau| 1_{\Omega^n_q(L_\epsilon)} \tag{6.2.15}$$

$$\begin{aligned}&\lesssim \sqrt{4MnL_\epsilon K_\epsilon}\int_0^\delta \sqrt{\log(1+N_{[]}(\Theta^n(\theta_0;K_\epsilon\delta),\rho;\varepsilon))}d\varepsilon\\ &\leq \sqrt{4MnL_\epsilon K_\epsilon}\varphi^n_{\theta_0}(K_\epsilon\delta) \leq \sqrt{4MnL_\epsilon K_\epsilon^2}\varphi^n_{\theta_0}(\delta)\end{aligned}$$

whenever δ satisfies the restriction that $\varphi^n_{\theta_0}(K_\epsilon\delta) \leq \sqrt{4MnL_\epsilon K_\epsilon}\cdot\delta^2$; assuming $M \geq 1/4$ without loss of generality, this restriction is weaker than $n^{-1/2}\varphi^n_{\theta_0}(\delta) \leq \delta^2$. Thus the above inequality holds for any $\delta \in [r_n^{-1},\infty)$. Now, recall that $R^n_{\theta_0}(\delta)\cup\{\theta^n_{\theta_0}\} \subset \Theta^n(\theta_0;K_\epsilon\delta)$ for all $\delta \in [K_\epsilon r^{-1}_{n,\theta_0},\infty)$ on the set $\Omega^n_{\theta_0}(K_\epsilon)$ given by (6.2.6). Multiplying the both sides of (6.2.15) by n^{-1}, we have that

$$E^n_{\theta_0}\sup_{\theta\in R^n_{\theta_0}(\delta)}|(\Gamma^n_{\theta_0}-\gamma^n_{\theta_0})(\theta)-(\Gamma^n_{\theta_0}-\gamma^n_{\theta_0})(\theta^n_{\theta_0})|1_{\Omega^n(L_\epsilon)\cap\Omega^n_{\theta_0}(K_\epsilon)} \lesssim \sqrt{4ML_\epsilon K_\epsilon^2}n^{-1/2}\varphi^n_{\theta_0}(\delta)$$

for any $\delta \in [K_\epsilon r_n^{-1},\infty)$. The SECOND INEQUALITY has been established. □

6.3 Counting Processes with Non-linear Covariates

Let $(E,\mathcal{E})$ be a Blackwell space on which a σ-finite measure λ is defined; this is the state space of covariate processes in the following model. Let $\tau > 0$ be a constant. We define the notations $\mathcal{L}^p_+$ and $\|\cdot\|_{\mathcal{L}^p}$ in the same way as the first paragraph of Section 6.2.

In the n-th statistical experiment, we consider k_n adapted point processes on $[0,\tau]$, namely $N^{n,i}$, $(i = 1,\ldots,k_n)$, defined on a filtered measurable space $(\Omega^n,\mathcal{F}^n,\mathbf{F}^n = (\mathcal{F}^n_t)_{t\in[0,\tau]})$; we then denote $T^{n,i}_j = \inf\{t \in [0,\tau] : N^{n,i}_t = j\}$ for every $j \in \mathbb{N}$ (see page 34 of Jacod and Shiryaev (1987)). Let $\mathbf{P}^n = \{P^n_\theta : \theta \in \Theta\}$ be a family of probability measures on $(\Omega^n,\mathcal{F}^n)$ indexed by a subset Θ of $\mathcal{L}^p_+$ for some $p \in [1,\infty]$ specified later. Suppose that the predictable compensator of $N^{n,i}$ with respect to the probability measure P^n_θ is given by

$$\theta(t,Z^{n,i}_t)Y^{n,i}_t dt,$$

where $t \rightsquigarrow Z^{n,i}_t$ is an E-valued predictable process and $t \rightsquigarrow Y^{n,i}_t$ is a $[0,\infty)$-valued predictable process. It is implicitly assumed that $\int_0^\tau \theta(t,Z^{n,i}_t(\omega))Y^{n,i}_t(\omega)dt < \infty$ for every $\omega \in \Omega^n$. Suppose that we can observe the processes $N^{n,i}$, $Y^{n,i}$ and $Z^{n,i}$ on the random sets $\{t \in [0,\tau] : Y^{n,i}_t(\omega) > 0\}$. The goal of this section is to derive the rate of convergence of sieved maximum likelihood estimators for θ.

We analyze this problem by using the E-valued multivariate point processes

$$\mu^n(dt,dz) = \sum_{i=1}^{k_n}\sum_j \varepsilon_{(T^{n,i}_j, Z^{n,i}_{T^{n,i}_j})}(dt,dz).$$

Here, we suppose that $T_j^{n,i} \neq T_{j'}^{n,i'}$ whenever $i \neq i'$; then all requirement for μ^n to be an E-valued multivariate point process, including that $\mu^n(\{t\} \times E) \leq 1$, are indeed satisfied (see Definition III.1.23 of Jacod and Shiryaev (1987)). The predictable compensator of μ^n with respect to the probability measure P_θ^n is given by

$$\nu^{n,\theta}(dt, dz) = \theta(t, z)\overline{\nu}^n(dt, dz)$$

where

$$\overline{\nu}^n(dt, dz) = \sum_{i=1}^{k_n} Y_t^{n,i} \varepsilon_{Z_t^{n,i}}(dz)dt. \tag{6.3.1}$$

Under some conditions, the log-likelihood ratio is given by

$$\log \frac{dP_\theta^n | \mathcal{F}_\tau^n}{dP_\vartheta^n | \mathcal{F}_\tau^n} = l^n(\theta) - l^n(\vartheta) \qquad \forall \theta, \vartheta \in \Theta,$$

where

$$l^n(\theta) = (\log \theta) * \mu_\tau^n - \theta * \overline{\nu}_\tau^n,$$

although we shall not require any property of the log-likelihood ratio.

For our discussion of rate of convergence, we adopt the random Hellinger "semimetric" ϱ^n which is "formally" defined by

$$\begin{aligned} \varrho^n(\theta, \vartheta) &= \sqrt{\frac{1}{n}\left|\sqrt{\theta} - \sqrt{\vartheta}\right|^2 * \overline{\nu}_\tau^n} \\ &= \sqrt{\frac{1}{n}\sum_{i=1}^{k_n}\int_0^\tau \left|\sqrt{\theta(t, Z_t^{n,i})} - \sqrt{\vartheta(t, Z_t^{n,i})}\right|^2 Y_t^{n,i} dt} \end{aligned} \tag{6.3.2}$$

for every $\theta, \vartheta \in \mathcal{L}_+^p$. The meaning of the quotation marks is that $\varrho^n(\theta, \vartheta) < \infty$ may *not* hold, although it has been implicitly assumed at least for $\theta, \vartheta \in \Theta$. On the other hand, an entropy condition on the sieve should be given in terms of the L^{2p}-Hellinger semimetric ρ_{2p} defined by

$$\begin{aligned} \rho_{2p}(\theta, \vartheta) &= \left\|\sqrt{\theta} - \sqrt{\vartheta}\right\|_{\mathcal{L}^{2p}} \\ &= \left(\int_{[0,\tau]\times E} \left|\sqrt{\theta(t, z)} - \sqrt{\vartheta(t, z)}\right|^{2p} dt\lambda(dz)\right)^{1/2p} \qquad \forall \theta, \vartheta \in \mathcal{L}_+^{2p}. \end{aligned}$$

A main difficulty in this model is that the random measure $\overline{\nu}^n(dt, dz)$ on $[0, \tau] \times E$ defined by (6.3.1) is not dominated by the measure $dt\lambda(dz)$; compare (6.2.2) and (6.3.1). Hence, in the present situation, an entropy condition in terms of ρ_{2p} is not directly translated into that in terms of ϱ^n as in the multiplicative intensity model where the relation $\varrho^n \leq L\rho_{2p}$ holds on the set $\Omega_q^n(L)$ given by (6.2.3). To solve this problem, we will take the same approach as Sections 4.2 and 4.3. Let $E = \bigcup_m E_m^n$ be a partition of $\mathcal{E}$-measurable sets, which is at most countable, such that $\lambda(E_m^n) \in (0, \infty)$. Set $\mathcal{E}^n = \sigma\{E_m^n : m = 1, 2, ...\}$, and denote by $\mathcal{L}_+^{p,n}$ the space of elements f of $\mathcal{L}_+^p$ that are $\mathfrak{B}[0, \tau] \otimes \mathcal{E}^n$-measurable; it is trivial that $\mathcal{L}_+^{p,n} \subset \mathcal{L}_+^p$. We define the predictable function Y^n on $\Omega^n \times [0, \tau] \times E$ by

$$Y^n(\omega, t, z) = \sum_m \sum_{i=1}^{k_n} Y_t^{n,i}(\omega) 1_{\{Z_t^{n,i}(\omega) \in E_m^n\}} \frac{1_{\{z \in E_m^n\}}}{\lambda(E_m^n)} \tag{6.3.3}$$

(do not confuse this $Y^n(\omega, t, z)$ and the original $Y_t^{n,i}(\omega)$'s). Then, it holds for any $f \in \mathcal{L}_+^{p,n}$ that

$$f * \overline{\nu}_\tau^n = \int_{[0,\tau] \times E} f(t, z) Y^n(\cdot, t, z) dt \lambda(dz) \tag{6.3.4}$$

if the integral is finite. We thus introduce the following condition.

Condition 6.3.1 *For some $p, q \in [1, \infty]$ such that $(1/p) + (1/q) = 1$, it holds that*

$$\Theta \subset \mathcal{L}_+^p \quad \textit{and} \quad Y^n(\omega, \cdot, \cdot) \in \mathcal{L}_+^{q,n} \quad \forall \omega \in \Omega^n. \tag{6.3.5}$$

Furthermore, for a given subset U of Θ, it holds that

$$\lim_{L \to \infty} \limsup_{n \to \infty} \sup_{\theta_0 \in U} P_{\theta_0}^{n*}(\Omega^n \setminus \Omega_q^n(L)) = 0,$$

where

$$\Omega_q^n(L) = \left\{\omega \in \Omega^n : \sqrt{n^{-1} \|Y^n\|_{\mathcal{L}^q}} \leq L\right\} \qquad \forall L > 0 \tag{6.3.6}$$

and where Y^n is defined by (6.3.3).

Under (6.3.5), the equation (6.3.4) does hold for all $f \in \mathcal{L}_+^{p,n}$ by the Hölder inequality, and thus the same relation as (6.2.2) is fulfilled on the smaller σ-field $\mathfrak{B}[0, \tau] \otimes \mathcal{E}^n$. In this case, the formula (6.3.2) indeed defines a random semimetric on $\mathcal{L}_+^{p,n} \cup \Theta$. In particular, we have that

$$\varrho^n(\theta, \vartheta) = \sqrt{\frac{1}{n} \int_{[0,\tau] \times E} \left|\sqrt{\theta(t, z)} - \sqrt{\vartheta(t, z)}\right|^2 Y^n(t, z) dt \lambda(dz)} \qquad \forall \theta, \vartheta \in \mathcal{L}_+^{p,n}$$

and that

$$\varrho^n(\theta, \vartheta) \leq L \rho_{2p}(\theta, \vartheta) \qquad \forall \theta, \vartheta \in \mathcal{L}_+^{p,n} \qquad \text{on the set } \Omega_q^n(L).$$

Hence it would be convenient for computation of entropy to construct a sieve $\{\Theta^n\}$ of subsets of $\mathcal{L}_+^{p,n}$'s rather than $\mathcal{L}_+^p$. To do it, we introduce the mapping $\pi^n : \mathcal{L}_+^p \to \mathcal{L}_+^{p,n}$ defined by

$$\pi^n \theta(t, z) = \sum_m \frac{1}{\lambda(E_m^n)^2} \left|\int_{E_m^n} \sqrt{\theta(t, w)} \lambda(dw)\right|^2 1_{\{z \in E_m^n\}}$$

(notice that this is different from that of Section 4.2). Then we have the following:

Lemma 6.3.2 (i) *If $\theta \leq \vartheta$ where $\theta, \vartheta \in \mathcal{L}_+^p$, then $\pi^n\theta \leq \pi^n\vartheta$.*
(ii) *If f is a $[0,\infty)$-valued $\mathfrak{B}[0,\tau] \otimes \mathcal{E}^n$-measurable function, then for every $\theta, \vartheta \in \mathcal{L}_+^p$*

$$\int_{[0,\tau]\times E} \left| \sqrt{\pi^n\theta(t,z)} - \sqrt{\pi^n\vartheta(t,z)} \right|^{2p} f(t,z)dt\lambda(dz)$$
$$\leq \int_{[0,\tau]\times E} \left| \sqrt{\theta(t,z)} - \sqrt{\vartheta(t,z)} \right|^{2p} f(t,z)dt\lambda(dz),$$

provided the integrals are finite.

Proof. The assertion (i) is trivial, and (ii) follows from that

$$\left| \sqrt{\pi^n\theta(t,z)} - \sqrt{\pi^n\vartheta(t,z)} \right|^{2p}$$
$$= \sum_m \frac{1}{\lambda(E_m^n)^{2p}} \left| \int_{E_m^n} \left(\sqrt{\theta(t,w)} - \sqrt{\vartheta(t,w)} \right) \lambda(dw) \right|^{2p} 1_{\{z \in E_m^n\}}$$
$$\leq \sum_m \frac{1}{\lambda(E_m^n)} \int_{E_m^n} \left| \sqrt{\theta(t,w)} - \sqrt{\vartheta(t,w)} \right|^{2p} \lambda(dw) 1_{\{z \in E_m^n\}}.$$

□

Consequently, we obtain that: if we choose (ρ_{2p}, ε)-brackets in $\mathcal{L}_+^p$, namely $[l^{\varepsilon,k}, u^{\varepsilon,k}]$'s, which cover the class Θ, then it holds on the set $\Omega_q^n(L)$ that $[\pi^n l^{\varepsilon,k}, \pi^n u^{\varepsilon,k}]$'s form an $(\varrho^n, L\varepsilon)$-brackets in $\mathcal{L}_+^{p,n}$ which cover the class $\pi^n\Theta$. This allows us to make an entropy condition with respect to the non-random semimetric ρ_{2p} rather than ϱ^n. Hereafter, we denote by $B_{\rho_{2p}}(\theta;\varepsilon)$ the closed ball in $\mathcal{L}_+^p$ with center θ and ρ_{2p}-radius ε.

Condition 6.3.3 *There exists a function $\varphi : (0,\infty) \to (0,\infty)$ such that $\delta \rightsquigarrow \delta^{-1}\varphi(\delta)$ is decreasing and that*

$$\text{(6.3.7)} \qquad \int_0^\delta \sqrt{\log(1 + N_{[]}(\Theta, \rho_{2p}; \varepsilon))} d\varepsilon \leq \varphi(\delta) \qquad \forall \delta \in (0,\infty).$$

Then, choose some positive constants r_n such that $n^{-1/2}\varphi(r_n^{-1}) \leq r_n^{-2}$.

Although we have stated the version corresponding to the case $\Theta^n(\theta_0;\delta) = \Theta^n$ of Condition 6.2.2 only, it is also possible to replace Θ of the entropy assumption (6.3.7) by the local ball $\Theta \cap B_{\rho_{2p}}(\theta_0;\delta)$ when we can show an "asymptotic equivalence" of ϱ^n and ρ_{2p} (i.e., (6.2.8)).

Theorem 6.3.4 *Let $U \subset \Theta \subset \mathcal{L}_+^p$ for some $p \in [1,\infty]$. Assume Conditions 6.3.1 and 6.3.3, and choose a sequence r_n described there. For a given sequence of subsets*

$\Theta^n \subset \pi^n\Theta$, suppose that there exists a constant $M > 0$ such that: for every $n \in \mathbb{N}$ and $\theta_0 \in U$ there exists $\theta^n_{\theta_0} \in \Theta^n$ such that

$$\text{(6.3.8)} \qquad \theta_0 \le M\theta^n_{\theta_0} \qquad \text{and} \qquad \rho_{2p}(\theta_0, \theta^n_{\theta_0}) \le Mr_n^{-1}.$$

Then, for any mapping $\widehat{\theta}^n : \Omega^n \to \Theta^n$ such that

$$l^n(\widehat{\theta}^n) \ge \sup_{\theta \in \Theta^n} l^n(\theta) - nr_n^{-2},$$

it holds that

$$\lim_{L\to\infty} \limsup_{n\to\infty} \sup_{\theta_0 \in U} P^n_{\theta_0}\left(r_n\varrho^n(\widehat{\theta}^n, \theta_0) > L\right) = 0.$$

This result can be proved exactly in the same way as Theorem 6.2.3 by means of Lemma 6.2.4; those proofs can be read with little change of notation (notice that "$\theta \in \Theta^n$" there should be read as "$\pi^n\theta \in \Theta^n$" here).

The next problem we should consider is how to check the assumption (6.3.8) for the sieve Θ^n given as a subset of $\pi^n\Theta$. When Θ is a class of "smooth" functions, the assumption is always satisfied if we use a slightly different sieve; a part of the idea has already appeared in Example 2 of Section 6.1. Define the mapping $\pi^n_+ : \mathcal{L}^p_+ \to \mathcal{L}^{p,n}_+$ by

$$\text{(6.3.9)} \qquad \pi^n_+\theta(t,z) = \sum_m \frac{1}{\lambda(E^n_m)^2}\left|\int_{E^n_m} \sqrt{\theta(t,w) + r_n^{-1}}\lambda(dw)\right|^2 1_{\{z\in E^n_m\}}.$$

It is easy to show the same facts as Lemma 6.3.2 with π^n replaced by π^n_+. Thus we have:

Corollary 6.3.5 *Let $U \subset \Theta \subset \mathcal{L}^p_+$. Assume Conditions 6.3.1 and 6.3.3, and choose a sequence r_n described there. Define $\Theta^n = \pi^n_+\Theta$. If there exists a constant $M > 0$ such that*

$$\sup_{t\in[0,\tau]} \sup_m \left| \sup_{z\in E^n_m} \sqrt{\theta_0(t,z)} - \inf_{z\in E^n_m} \sqrt{\theta_0(t,z)} \right| \le Mr_n^{-1} \qquad \forall \theta_0 \in U,\ \forall n \in \mathbb{N},$$

and if $\lambda(E) < \infty$, then the same conclusion as Theorem 6.3.4 holds.

Proof. Notice that: if we choose (ρ_{2p}, ε)-brackets in $\mathcal{L}^p_+$ which cover the class Θ, namely $[l^{\varepsilon,k}, u^{\varepsilon,k}]$'s, then it holds on the set $\Omega^n_q(L)$ that $[\pi^n_+ l^{\varepsilon,k} - r_n^{-1}, \pi^n_+ u^{\varepsilon,k} - r_n^{-1}]$'s form $(\varrho^n, L\varepsilon)$-brackets in $\mathcal{L}^{p,n}_+$ which cover $\Theta^n = \pi^n_+\Theta$. Hence it suffices to prove that (6.3.8) is satisfied for the present sieve $\Theta^n = \pi^n_+\Theta$; we will show it for $\theta^n_{\theta_0} = \pi^n_+\theta_0$. First observe that

$$\sqrt{\frac{\theta_0(t,z)}{\pi^n_+\theta_0(t,z)}} \le \frac{\inf_{z\in E^n_m}\sqrt{\theta_0(t,z)} + Mr_n^{-1}}{\inf_{z\in E^n_m}\sqrt{\theta_0(t,z)} + r_n^{-1}} \le 1 + M.$$

Next, since

$$\inf_{z\in E_m^n}\sqrt{\theta_0(t,z)} \le \sqrt{\pi_+^n\theta_0(t,z)} - r_n^{-1} \le \sup_{z\in E_m^n}\sqrt{\theta_0(t,z)}$$

we have

$$\sup_{(t,z)\in[0,\tau]\times E}\left|\sqrt{\theta_0(t,z)} - \sqrt{\pi_+^n\theta_0(t,z)}\right| \le (1+M)r_n^{-1}.$$

Thus it holds that $\rho_{2p}(\theta_0, \pi_+^n\theta_0) \le (1+M)\tau\lambda(E)\cdot r_n^{-1}$. □

Example 1: Smooth functions

Let us take $E = [0,1]^d$ endowed with the Lebesgue measure. Let $\alpha > (d+1)/2$ and $H > 0$, and we consider the class $C_H^\alpha = C_H^\alpha([0,\tau]\times[0,1]^d)$ given in (11) of Section 1.2. We set $\Theta = \{\theta \in \mathcal{L}_+^\infty : \sqrt{\theta} \in C_H^\alpha\}$. Then, Condition 6.3.3 is satisfied with $U = \Theta$ and $\varphi(\delta) = \text{const.}(\delta^{1-(d+1)/2\alpha} \vee \delta)$; thus we set $r_n = n^{\alpha/(2\alpha+d+1)}$. Let $[0,1]^d = \bigcup_m E_m^n$ be a partition of Borel measurable subsets of $[0,1]^d$ such that $\mathrm{Diam}(E_m^n) \le r_n^{-1}$. Then, the displayed assumption of Corollary 6.3.5 is satisfied also with $U = \Theta$. Hence, we can get the conclusion of Theorem 6.3.4 if Condition 6.3.1 for $p = \infty$ and $q = 1$ is satisfied. A sufficient condition for this is that $k_n = n$ and the processes $Y^{n,i}$ take values only in a bounded set $[0,K]$. As in the next example, the state space of the process $Y^{n,i}$ is typically $\{0,1\}$ in the context of survival analysis, hence the result above is always valid.

Example 2: Lexis diagram

Let us discuss the Lexis diagram which is an important method describing models in survival analysis (see Keiding (1990) or Chapter X of Andersen et al. (1993) for the details). Let $E = [0,\tau]$ and $k_n = n$. We suppose that the covariate process $Z^{n,i}$ is given by

$$Z_t^{n,i} = (t - e^{n,i}) \vee 0,$$

where $e^{n,i}$ is a $[0,\tau]$-valued random variable representing the *entry time* of individual i; then $Z_t^{n,i}$ is considered to represent the *age* or *duration* of the individual i at *calendar time* $t \in [0,\tau]$. We also suppose that the process $Y^{n,i}$ is given by

$$Y_t^{n,i} = 1_{\{e^{n,i}\le t,\ Z_t^{n,i} < U^{n,i}\}},$$

where $U^{n,i}$ is a $[0,\infty]$-valued random variable; it is typically of the form $U^{n,i} = (T_1^{n,i} - e^{n,i}) \wedge C^{n,i}$ where $C^{n,i}$ is a $[0,\infty]$-valued random variable representing the *censoring time* of the individual i. To guarantee the predictability of the processes $Y^{n,i}$ and $Z^{n,i}$, the random variables $e^{n,i}$ and $e^{n,i} + U^{n,i}$ are assumed to be stopping times.

We consider the parameter space Θ given in the preceding example with $d = 1$ (and $E = [0, 1]$ is replaced by $[0, \tau]$), and thus we set $r_n = n^{\alpha/(2\alpha+2)}$. If we take a sequence of partitions $[0, \tau] = \bigcup_m E_m^n$ of Borel measurable sets E_m^n such that $\mathrm{Diam}(E_m^n) \leq r_n^{-1}$, the conclusion of Theorem 6.3.4 for the sieve $\Theta^n = \pi_+^n\Theta$ given by (6.3.9).

6.4 Diffusion-type Processes

Let us consider the stochastic differential equation

$$dX_t = \theta(t, X)dt + n^{-1/2}dB_t, \qquad X_0 = x_0 \in \mathbb{R}, \tag{6.4.1}$$

where $t \rightsquigarrow B_t$ is a standard Brownian motion defined on a stochastic basis $(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t\in[0,\tau]}, P)$ and $\tau > 0$ is a fixed constant. The functional θ appearing above should satisfy some appropriate properties described as follows. We equip $C[0, \tau]$, the canonical space of sample paths, with the σ-field $\mathcal{H}_t = \sigma\{x_s : s \leq t\}$ for every $t \in [0, \tau]$.

Definition 6.4.1 *We denote by $\mathcal{A}$ the set of functionals $\theta : [0, \tau] \times C[0, \tau] \to \mathbb{R}$ such that:*
(i) *$x \rightsquigarrow \theta(t, x)$ is $\mathcal{H}_t$-measurable for every $t \in [0, \tau]$;*
(ii) $\sup_{t\in[0,\tau]} \sup_{x\in C[0,\tau]} |\theta(t, x)| < \infty$.

Definition 6.4.2 *For a given constant $H > 0$, we denote by $\mathcal{L}_H$ the set of functionals $\theta \in \mathcal{A}$ such that*

$$|\theta(t, x) - \theta(t, y)| \leq H \sup_{s\in[0,t]} |x_s - y_s| \qquad \forall x, y \in C[0, \tau], \ \forall t \in [0, \tau].$$

It is well-known that the stochastic differential equation (6.4.1) has a unique strong solution $X^{n,\theta} = (X_t^{n,\theta})_{t\in[0,\tau]}$ whenever $\theta \in \mathcal{L}_H$ for some $H > 0$ (see e.g. Theorem 13.1 of Rogers and Williams (1987)). We denote by $x^\theta = (x_t^\theta)_{t\in[0,\tau]}$ the solution of the ordinary differential equation

$$dx_t = \theta(t, x)dt, \qquad x_0 \in \mathbb{R}. \tag{6.4.2}$$

Fix any $H > 0$. Let us introduce three notations. For every $n \in \mathbb{N}$ and $\theta_0 \in \mathcal{L}_H$ we define the random semimetric $\varrho_{\theta_0}^n$ on $\mathcal{A}$ by

$$\varrho_{\theta_0}^n(\theta, \vartheta) = \sqrt{\frac{1}{\tau}\int_0^\tau |\theta(t, X^{n,\theta_0}) - \vartheta(t, X^{n,\theta_0})|^2 dt} \qquad \forall \theta, \vartheta \in \mathcal{A}.$$

For $\theta_0 \in \mathcal{L}_H$ we define the semimetric ρ_{θ_0} on $\mathcal{A}$ by

$$\rho_{\theta_0}(\theta, \vartheta) = \sqrt{\frac{1}{\tau}\int_0^\tau |\theta(t, x^{\theta_0}) - \vartheta(t, x^{\theta_0})|^2 dt} \qquad \forall \theta, \vartheta \in \mathcal{A}.$$

For every $\theta_0 \in \mathcal{L}_H$ and $\varepsilon > 0$, we define

$$\mathcal{L}_H^\star(\rho_{\theta_0};\varepsilon) = \{\theta \in \mathcal{A} : \quad \exists \theta_1, \theta_2 \in \mathcal{L}_H \text{ such that } \theta_1 \le \theta \le \theta_2 \text{ and } \rho_{\theta_0}(\theta_1,\theta_2) \le \varepsilon\}.$$

It is clear that $\mathcal{L}_H \subset \mathcal{L}_H^\star(\rho_{\theta_0};\varepsilon) \subset \mathcal{A}$.

We denote by $\mathbb{P}^{n,\theta}$ the induced probability distribution of $X^{n,\theta}$ on the canonical space $C[0,\tau]$. Then, the family $\{\mathbb{P}^{n,\theta} : \theta \in \mathcal{L}_H\}$ is equivalent, and the log-likelihood ratio is given by

$$\log \frac{\mathbb{P}_\theta^n}{\mathbb{P}_\vartheta^n}(X) = n\{l(\theta;X) - l(\vartheta;X)\} \qquad \forall \theta, \vartheta \in \mathcal{L}_H$$

where

$$l(\theta;X) = \int_0^\tau \theta(t,X)dX_t - \frac{1}{2}\int_0^\tau |\theta(t,X)|^2 dt \tag{6.4.3}$$

(see, e.g., page 29 of Kutoyants (1994)). Although the representation of log-likelihood ratio relies on the existence of unique strong solution of the stochastic differential equation (6.4.1), the formula (6.4.3) itself is well-defined for all $\theta \in \mathcal{A}$. We then consider a maximum likelihood estimator (MLE) on the sieve $\Theta^n \subset \mathcal{A}$

$$\widehat{\theta}^n(X) = \operatorname*{argmax}_{\theta \in \Theta^n} l(\theta;X).$$

The precise description will be given in the main theorem below.

The condition which we shall assume is as follows; we denote by $B_{\rho_{\theta_0}}(\theta;\delta)$ the closed ball in $\mathcal{A}$ with center θ and ρ_{θ_0}-radius δ.

Condition 6.4.3 *Let $U \subset \Theta \subset \mathcal{L}_H$ and $\Theta^n \subset \mathcal{A}$ be given. For every $n \in \mathbb{N}$ and $\theta_0 \in U$, there exist a proper metric $\widehat{\rho}_{\theta_0}^n$ on Θ^n such that $\rho_{\theta_0} \le \widehat{\rho}_{\theta_0}^n$, and a function $\varphi_{\theta_0}^n : (0,\infty) \to (0,\infty)$ such that $\delta \rightsquigarrow \delta^{-1}\varphi_{\theta_0}^n(\delta)$ is decreasing and that*

$$\int_0^\delta \sqrt{\log(1 + N(\Theta^n \cap B_{\rho_{\theta_0}}(\theta_0;\delta), \widehat{\rho}_{\theta_0}^n;\varepsilon))}d\varepsilon \le \varphi_{\theta_0}^n(\delta) \qquad \forall \delta \in (0,\infty). \tag{6.4.4}$$

Then, choose some constants $r_{n,\theta_0} \in (0, n^{1/2}]$ such that $n^{-1/2}\varphi_{\theta_0}^n(r_{n,\theta_0}^{-1}) \le r_{n,\theta_0}^{-2}$.

Theorem 6.4.4 *Let $U \subset \Theta \subset \mathcal{L}_H$ for a constant $H > 0$. Suppose that Condition 6.4.3 is satisfied for some countable sets $\Theta^n \subset \mathcal{A}$, and choose some constants r_{n,θ_0} described there. Suppose also that there exists a constant $M > 0$ such that: for every $n \in \mathbb{N}$ and $\theta_0 \in U$*

$$\Theta^n \subset \mathcal{L}_H^\star(\rho_{\theta_0}; M r_{n,\theta_0}^{-1}) \qquad \text{and} \qquad \rho_{\theta_0}(\theta_0,\theta_{\theta_0}^n) \le M r_{n,\theta_0}^{-1} \quad \text{for some } \theta_{\theta_0}^n \in \Theta^n.$$

Then, for any mapping $\widehat{\theta}^n : C[0,\tau] \to \Theta^n$ such that

$$l(\widehat{\theta}^n(X);X) \ge \sup_{\theta \in \Theta^n} l(\theta;X) - r_n^{-2} \quad \text{with} \quad r_n = \sup_{\theta_0 \in U} r_{n,\theta_0}, \tag{6.4.5}$$

it holds that

$$\lim_{L\to\infty}\limsup_{n\to\infty}\sup_{\theta_0\in U} P^*\left(r_{n,\theta_0}\rho_{\theta_0}(\widehat{\theta}^n(X^{n,\theta_0}),\theta_0) > L\right) = 0;$$
$$\lim_{L\to\infty}\limsup_{n\to\infty}\sup_{\theta_0\in U} P^*\left(r_{n,\theta_0}\varrho^n_{\theta_0}(\widehat{\theta}^n(X^{n,\theta_0}),\theta_0) > L\right) = 0.$$

In order to prove the result above, we will apply Theorem 5.1.2 to the processes $\theta \rightsquigarrow \Gamma^n_{\theta_0}(\theta)$ and $\theta \rightsquigarrow \gamma^n_{\theta_0}(\theta)$ given by:

$$\text{(6.4.6)} \qquad \begin{aligned}\Gamma^n_{\theta_0}(\theta) &= l(\theta; X^{n,\theta_0})\\ &= \int_0^\tau \theta(t,X^{n,\theta_0})dX^{n,\theta_0} - \frac{1}{2}\int_0^\tau |\theta(t,X^{n,\theta_0})|^2dt;\end{aligned}$$

$$\text{(6.4.7)} \qquad \begin{aligned}\gamma^n_{\theta_0}(\theta) &= \int_0^\tau \theta(t,X^{n,\theta_0})\theta_0(t,X^{n,\theta_0})dt - \frac{1}{2}\int_0^\tau |\theta(t,X^{n,\theta_0})|^2dt\\ &= -\frac{1}{2}\varrho^n_{\theta_0}(\theta,\theta_0)^2 + \frac{1}{2}\int_0^\tau |\theta_0(t,X^{n,\theta_0})|^2dt.\end{aligned}$$

The key point of the proof is that

$$\text{(6.4.8)} \qquad \Gamma^n_{\theta_0}(\theta) - \gamma^n_{\theta_0}(\theta) = n^{-1/2}\int_0^\tau \theta(t,X^{n,\theta_0})dB_t.$$

First, let us investigate the relationship between ρ^n_θ and $\varrho^n_{\theta_0}$.

Lemma 6.4.5 *Let $H > 0$ and $\varepsilon > 0$ be arbitrary constants. For every $\theta_0 \in \mathcal{L}_H$ and every $\theta,\vartheta \in \mathcal{L}^\star_H(\rho_{\theta_0},\varepsilon)$ it holds that*

$$|\varrho^n_{\theta_0}(\theta,\vartheta) - \rho_{\theta_0}(\theta,\vartheta)|^2 \le n^{-1}\cdot 24H^2\tau e^{2H\tau}\sup_{t\in[0,\tau]}|B_t|^2 + 48\varepsilon^2$$

P-almost surely.

Proof. Observe that

$$\begin{aligned}&|\varrho^n_{\theta_0}(\theta,\vartheta) - \rho_{\theta_0}(\theta,\vartheta)|^2\\ &\le \int_0^\tau \left|\{\theta(X^{n,\theta_0}) - \vartheta(X^{n,\theta_0})\} - \{\theta(x^{\theta_0}) - \vartheta(x^{\theta_0})\}\right|^2 dt\\ &\le 2\int_0^\tau \left|\theta(X^{n,\theta_0}) - \theta(x^{\theta_0})\right|^2 dt + 2\int_0^\tau \left|\vartheta(X^{n,\theta_0}) - \vartheta(x^{\theta_0})\right|^2 dt.\end{aligned}$$

Here, for given $\theta \in \mathcal{L}^\star_H(\rho_{\theta_0},\varepsilon)$ choose some $\theta_1,\theta_2 \in \mathcal{L}_H$ such that $\theta_1 \le \theta \le \theta_2$ and that $\rho_{\theta_0}(\theta_1,\theta_2) \le \varepsilon$. Then we have

$$\begin{aligned}|\theta(t,X^{n,\theta_0}) - \theta(t,x^{\theta_0})| &\le |\theta_1(t,X^{n,\theta_0}) - \theta_2(t,x^{\theta_0})| + |\theta_2(t,X^{n,\theta_0}) - \theta_1(t,x^{\theta_0})|\\ &\le |\theta_1(t,X^{n,\theta_0}) - \theta_1(t,x^{\theta_0})| + |\theta_2(t,X^{n,\theta_0}) - \theta_2(t,x^{\theta_0})|\\ &\quad +2|\theta_1(t,x^{\theta_0}) - \theta_2(t,x^{\theta_0})|,\end{aligned}$$

thus

$$|\theta(t, X^{n,\theta_0}) - \theta(t, x^{\theta_0})|^2 \le 6H^2 \sup_{s\in[0,t]} |X_s^{n,\theta_0} - x_s^{\theta_0}|^2 + 12|\theta_1(t, x^{\theta_0}) - \theta_2(t, x^{\theta_0})|^2.$$

It follows from the Grownwall inequality that

$$\sup_{t\in[0,\tau]} |X_t^{n,\theta_0} - x_t^{\theta_0}| \le \mathrm{e}^{H\tau} \sup_{t\in[0,\tau]} |n^{-1/2} B_t|,$$

hence we obtain

$$\int_0^\tau |\theta(t, X^{n,\theta_0}) - \theta(t, x^{\theta_0})|^2 dt \le 6H^2 \tau \mathrm{e}^{2H\tau} \sup_{t\in[0,\tau]} |n^{-1/2} B_t|^2 + 12\varepsilon^2.$$

Those inequalities imply the assertion. □

We will use the above lemma in the following form.

Lemma 6.4.6 *Let U be an arbitrary subset of $\mathcal{L}_H$ for some $H > 0$, and let $M > 0$ be an arbitrary constant. For every $n \in \mathbb{N}$ and $\theta_0 \in U$, let r_{n,θ_0} be some positive constants such that $r_{n,\theta_0} \le n^{1/2}$, and let Θ^n a countable subset of the set $\bigcap_{\theta_0\in U} \mathcal{L}_H^\star(\rho_{\theta_0}, M r_{n,\theta_0}^{-1})$. Then, it holds that:*

$$\lim_{L\to\infty} \sup_{n\in\mathbb{N}} \sup_{\theta_0\in U} P\left(\sup_{\theta,\vartheta\in\Theta^n} |\varrho_{\theta_0}^n(\theta,\vartheta) - \rho_{\theta_0}(\theta,\vartheta)| > L r_{n,\theta_0}^{-1} \right) = 0;$$

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \sup_{\theta_0\in U} P\left(\sup_{\substack{\theta,\vartheta\in\Theta^n \\ \rho_{\theta_0}(\theta,\vartheta) > K r_{n,\theta_0}^{-1}}} \frac{\rho_{\theta_0}(\theta,\vartheta)}{\varrho_{\theta_0}^n(\theta,\vartheta)} > 2 \right) = 0.$$

Proof. It follows from Lemma 6.4.5 that for every $\varepsilon > 0$ there exists a constant $L_\varepsilon > 0$ such that $\sup_{n\in\mathbb{N}} \sup_{\theta_0\in\Theta} P(\Omega \setminus \Omega_{\theta_0}^n(\varepsilon)) \le \varepsilon$, where

$$\Omega_{\theta_0}^n(\varepsilon) = \left\{ \sup_{\theta,\vartheta\in\Theta^n} |\varrho_{\theta_0}^n(\theta,\vartheta) - \rho_{\theta_0}(\theta,\vartheta)| \le L_\varepsilon r_{n,\theta_0}^{-1} \right\}.$$

The first assertion is nothing else than this fact. On the other hand, it holds on the set $\Omega_{\theta_0}^n(\varepsilon)$ that

$$\begin{aligned}
\sup_{\substack{\theta,\vartheta\in\Theta^n \\ \rho_{\theta_0}(\theta,\vartheta) > K r_{n,\theta_0}^{-1}}} \frac{\rho_{\theta_0}(\theta,\vartheta)}{\varrho_{\theta_0}^n(\theta,\vartheta)} &\le \sup_{\substack{\theta,\vartheta\in\Theta^n \\ \rho_{\theta_0}(\theta,\vartheta) > K r_{n,\theta_0}^{-1}}} \frac{\rho_{\theta_0}(\theta,\vartheta)}{\rho_{\theta_0}(\theta,\vartheta) - L_\varepsilon r_{n,\theta_0}^{-1}} \\
&\le \sup_{\substack{\theta,\vartheta\in\Theta^n \\ \rho_{\theta_0}(\theta,\vartheta) > K r_{n,\theta_0}^{-1}}} \frac{\rho_{\theta_0}(\theta,\vartheta)}{\rho_{\theta_0}(\theta,\vartheta) - \frac{1}{2}\rho_{\theta_0}(\theta,\vartheta)} \qquad \text{whenever } K \ge 2L_\varepsilon \\
&= 2.
\end{aligned}$$

This implies the second assertion. □

Proof of Theorem 6.4.4. We will apply Theorem 5.1.2. Formulation 5.1.1 should be read as follows: for every $\theta_0 \in U(= U^n)$,

(i) the random semimetric space $(\Theta^n, \varrho^n_{\theta_0})$ and the point $\theta^n_{\theta_0} \in \Theta^n$;

(ii) the stochastic processes $\theta \rightsquigarrow \Gamma^n_{\theta_0}(\theta)$ and $\theta \rightsquigarrow \gamma^n_{\theta_0}(\theta)$ with parameters in Θ^n, given by (6.4.6) and (6.4.7), respectively.

We then denote $R^n_{\theta_0}(\delta) = \{\theta \in \Theta^n : (\delta/2) < \varrho^n_{\theta_0}(\theta, \theta^n_{\theta_0}) \leq \delta\}$ for every $\delta \in (0, \infty)$.

First of all, it follows from Lemma 6.4.6 that for every $\varepsilon > 0$ there exists a constant $K_\varepsilon > 0$ such that

$$\sup_{n\in\mathbb{N}} \sup_{\theta_0\in U} P(\Omega \setminus \Omega^n_{\theta_0}(K_\varepsilon)) \leq \varepsilon,$$

where

$$\Omega^n_{\theta_0}(K_\varepsilon) = \left\{\varrho^n_{\theta_0}(\theta^n_{\theta_0}, \theta_0) \leq K_\varepsilon r^{-1}_{n,\theta_0}\right\} \cap \bigcap_{\substack{\theta,\vartheta\in\Theta^n \\ \rho_{\theta_0}(\theta,\vartheta) > K_\varepsilon r^{-1}_{n,\theta_0}}} \left\{\frac{1}{2} \leq \frac{\varrho^n_{\theta_0}(\theta,\vartheta)}{\rho_{\theta_0}(\theta,\vartheta)} \leq 2\right\}.$$

To show the FIRST INEQUALITY of M-CRITERION, observe that for any $L \geq 16L_\varepsilon$

$$\begin{aligned}
\varrho^n_{\theta_0}(\theta^n_{\theta_0}, \theta_0) &\leq K_\varepsilon r^{-1}_{n,\theta_0} && \text{on the set } \Omega^n_{\theta_0}(K_\varepsilon) \\
&\leq \frac{1}{4}\cdot\frac{\delta}{2} && \text{whenever } \delta \geq 8K_\varepsilon r^{-1}_{n,\theta_0} \\
&\leq \frac{1}{4}\varrho^n_{\theta_0}(\theta, \theta^n_{\theta_0}) && \text{whenever } \theta \in R^n_{\theta_0}(\delta).
\end{aligned}$$

Hence, we have for every $\delta \geq 8K_\varepsilon r^{-1}_{n,\theta_0}$ and $\theta \in R^n_{\theta_0}(\delta)$ that

$$\begin{aligned}
\gamma^n_{\theta_0}(\theta) - \gamma^n_{\theta_0}(\theta^n_{\theta_0}) &= \frac{1}{2}\left\{-\varrho^n_{\theta_0}(\theta,\theta_0)^2 + \varrho^n_{\theta_0}(\theta^n_{\theta_0},\theta_0)^2\right\} \\
&\leq \frac{1}{2}\left\{-\varrho^n_{\theta_0}(\theta,\theta^n_{\theta_0})^2 + 2\varrho^n_{\theta_0}(\theta,\theta^n_{\theta_0})\varrho^n_{\theta_0}(\theta^n_{\theta_0},\theta_0)\right\} \\
&\leq \frac{1}{2}\left\{-\varrho^n_{\theta_0}(\theta,\theta^n_{\theta_0})^2 + 2\varrho^n_{\theta_0}(\theta,\theta^n_{\theta_0})\frac{\varrho^n_{\theta_0}(\theta,\theta^n_{\theta_0})}{4}\right\} \\
&= -\frac{1}{4}\varrho^n_{\theta_0}(\theta,\theta^n_{\theta_0})^2 \\
&\leq -\frac{1}{4}\delta^2
\end{aligned}$$

on the set $\Omega^n_{\theta_0}(K_\varepsilon)$. This means that the FIRST INEQUALITY holds for $p = 2$.

Next, notice that whenever $K_\varepsilon \geq 2$

$$\Omega^n_{\theta_0}(K_\varepsilon) \subset \left\{R^n_{\theta_0}(\delta) \cup \{\theta^n_{\theta_0}\} \subset \Theta^n \cap B_{\rho_{\theta_0}}(\theta^n_{\theta_0}; K_\varepsilon\delta) \quad \forall\delta \in [K_\varepsilon r^{-1}_{n,\theta_0}, \infty)\right\}.$$

Since $\Theta^n \cup B_{\rho_{\theta_0}}(\theta^n_{\theta_0}; K_\varepsilon\delta)$ is contained in $\Theta^n \cup B_{\rho_{\theta_0}}(\theta_0; (K_\varepsilon + M)\delta)$, we can deduce from Theorem 2.4.5 that for every $\delta \geq K_\varepsilon r^{-1}_{n,\theta_0}$

$$\begin{aligned} & E^* \sup_{\theta \in R^n_{\theta_0}(\delta)} |(\Gamma^n_{\theta_0} - \gamma^n_{\theta_0})(\theta) - (\Gamma^n_{\theta_0} - \gamma^n_{\theta_0})(\theta^n_{\theta_0})| 1_{\Omega^n_{\theta_0}(K_\varepsilon)} \\ \lesssim \ & K_\varepsilon \cdot n^{-1/2} \int_0^\delta \sqrt{\log(1 + N(\Theta^n \cap B_{\rho_{\theta_0}}(\theta_0; (K_\varepsilon + M)\delta), \widehat{\rho}^n_{\theta_0}; \varepsilon))} d\varepsilon \\ \leq \ & K_\varepsilon \cdot n^{-1/2} \varphi^n_{\theta_0}((K_\varepsilon + M)\delta) \\ \leq \ & K_\varepsilon (K_\varepsilon + M) \cdot n^{-1/2} \varphi^n_{\theta_0}(\delta). \end{aligned}$$

Thus the SECOND INEQUALITY is fulfilled with $\phi^n_{\theta_0} = n^{-1/2}\varphi^n_{\theta_0}$.

Hence it follows from Theorem 5.1.2 that

$$\lim_{L\to\infty} \limsup_{n\to\infty} \sup_{\theta_0 \in U} P^* \left(r_{n,\theta_0} \rho_{\theta_0}(\widehat{\theta}^n(X^{n,\theta_0}), \theta^n_{\theta_0}) > L \right) = 0.$$

Noting $\rho_{\theta_0}(\theta^n_{\theta_0}, \theta_0) \leq M r^{-1}_{n,\theta_0}$, we get the first conclusion of the theorem. The second one follows from the first and Lemma 6.4.6. □

Example: Markovian case

Consider the stochastic differential equation

$$dX_t = \theta(X_t)dt + n^{-1/2} dB_t, \qquad X_0 = x_0 \in \mathbb{R},$$

where $\theta : \mathbb{R} \to \mathbb{R}$ is a bounded, Lipschitz continuous function. Then, the unique strong solution $X^{n,\theta} = (X^{n,\theta}_t)_{t\in[0,\tau]}$ is a time-homogeneous Markov process. In the same way as the general case, we denote by $x^\theta = (x^\theta_t)_{t\in[0,\tau]}$ the solution of the ordinary differential equation

$$dx_t = \theta(x_t)dt, \qquad x_0 \in \mathbb{R}.$$

For a given bounded, Lipschitz continuous function $\theta_0 : \mathbb{R} \to \mathbb{R}$, we define

$$\varrho^n_{\theta_0}(\theta, \vartheta) = \sqrt{\frac{1}{\tau} \int_0^\tau |\theta(X^{n,\theta_0}_t) - \vartheta(X^{n,\theta_0}_t)|^2 dt} \qquad \forall \theta, \vartheta \in \mathcal{A}$$

and

$$\rho_{\theta_0}(\theta, \vartheta) = \sqrt{\frac{1}{\tau} \int_0^\tau |\theta(x^{\theta_0}_t) - \vartheta(x^{\theta_0}_t)|^2 dt} \qquad \forall \theta, \vartheta \in \mathcal{A},$$

where $\mathcal{A}$ denotes the space of bounded measurable functions on $\mathbb{R}$.

Let some constants $\alpha \geq 1$ and $H > 0$ be given. We consider the class $C^\alpha_H(\mathbb{R})$ given in (11) of Section 1.2. Due to the fact that $t \rightsquigarrow x^\theta_t$ takes values only in $[x_0 - H\tau, x_0 + H\tau]$ whenever $\theta \in C^\alpha_H(\mathbb{R})$, and recalling also Lemma 6.4.5, we define the parameter space

$$\Theta = \{\theta \in C^\alpha_H : x \rightsquigarrow \theta(x) \text{ is constant on } E^c\},$$

where $E = [x_0 - H\tau - 1, x_0 + H\tau + 1]$. We denote by $\|\cdot\|_\infty$ the supremum norm on $\ell^\infty(\mathbb{R})$; notice that $\rho_{\theta_0} \leq \|\cdot\|_\infty$. Since we have

$$\log N(\Theta, \|\cdot\|_\infty; \varepsilon) \leq \text{const.}\varepsilon^{-1/\alpha},$$

Condition 6.4.3 is satisfied with $\varphi(\delta) = \text{const.}(\delta^{1-(1/2\alpha)} \vee \delta)$ whenever $\Theta^n \subset \Theta$. In this case, we get the rate $r_n = n^{\alpha/(2\alpha+1)}$. Consequently, if we choose a countable subset Θ^n of Θ such that $\Theta \subset \bigcup_{\theta \in \Theta^n} B_{(\Theta, \|\cdot\|_\infty)}(\theta; Mn^{-\alpha/(2\alpha+1)})$ for some $M > 0$ not depending on n, then it holds for any Θ^n-sieved MLE $\widehat{\theta}^n$ that:

$$\lim_{L\to\infty} \limsup_{n\to\infty} \sup_{\theta_0 \in \Theta} P^*\left(n^{\alpha/(2\alpha+1)} \rho_{\theta_0}(\widehat{\theta}^n(X^{n,\theta_0}), \theta_0) > L\right) = 0;$$

$$\lim_{L\to\infty} \limsup_{n\to\infty} \sup_{\theta_0 \in \Theta} P^*\left(n^{\alpha/(2\alpha+1)} \varrho^n_{\theta_0}(\widehat{\theta}^n(X^{n,\theta_0}), \theta_0) > L\right) = 0.$$

6.A Notes

The rate of convergence of infinite-dimensional M-estimators has been studied vigorously by Birgé and Massart (1993), van de Geer (1990, 1993, 1995a, 1995b), Wong and Severini (1991) and Wong and Shen (1995); see also Chapters 3.2 and 3.4 of van der Vaart and Wellner (1996) and the bibliographical Notes there. Among the preceding works, van de Geer (1995b) is a unique paper that deals with dependent data. Based on her general result for counting processes, she considered non-parametric maximum likelihood estimators in the multiplicative intensity model (without marks); it should be emphasized that, although there are some differences, a major part of Section 6.2 has been already known through her work. However, the marks and the discussion about sieves that have been newly added there are important for analyzing the non-linear covariate model in Section 6.3.

Some M-estimation problems in finite-dimensional parametric models of diffusion-type processes have been studied by Lánska (1979), Genon-Catalot (1990), Yoshida (1990, 1992) and Kutoyants (1994, Chapter 7); see also the references therein. The results in Section 6.4 seem to be the first attempt in the infinite-dimensional model.

Chapter 7

Miscellanies

7.1 Local Random Fields of Kernel Estimators

It is well-known that kernel density estimators for i.i.d. data have point-wise asymptotic normality. However, since the density f is originally defined as a Radon-Nikodym derivative with respect to Lebesgue measure, the value $f(x)$ at each point x does not intrinsically make sense. Thus, an assertion in some functional sense is preferable in order for, e.g., the construction of confidence intervals.

The purpose of this section is to extend the asymptotic normality of kernel density estimators to the functional sense with respect to a local parameter. The localizing constants should be chosen to be the same as the bandwidth. Further, in Subsection 7.1.2, we apply it to the estimation problem of the mode of f using also Theorem 5.1.2. The generalizations of those results to some dependent cases are discussed in Subsection 7.1.3.

7.1.1 I.I.D. Case

Let $\{X_i\}_{i\in\mathbb{N}}$ be an i.i.d. sequence of $\mathbb{R}^d$-value random variables with Lebesgue density f. Let $x_0 \in \mathbb{R}^d$ be a fixed point, and let $\{b_n\}_{n\in\mathbb{N}}$ be a sequence of positive constants such that $b_n \downarrow 0$ as $n \to \infty$. We are interested in estimating the local function $u \rightsquigarrow f(x_0+b_n u)$, where the parameter u runs through a subset U of $\mathbb{R}^d$. We consider the kernel estimator

$$\widehat{f}_n(x_0 + b_n u) = \frac{1}{nb_n^d} \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{b_n} - u\right) \qquad \forall u \in U,$$

where $K(x)$ is a kernel function on $\mathbb{R}^d$. Throughout this section, the notation $x^{(p)}$ means the p-th component of a vector $x \in \mathbb{R}^d$. We make two kinds of conditions either of which the kernel function should satisfy.

Condition 7.1.1 (smooth kernel) *The function $K : \mathbb{R}^d \to \mathbb{R}$ satisfies that:*

(i) $\int_{\mathbb{R}^d} K(x)dx = 1$, $K(x) = K(-x)$ *for every* $x \in \mathbb{R}^d$, *and* K *has a compact support;*
(ii) *there exist* $\alpha \in (0,1]$ *and* $L > 0$ *such that* $|K(x) - K(y)| \leq L|x-y|^\alpha$ *for every* $x, y \in \mathbb{R}^d$.

Condition 7.1.2 (monotone kernel) *The function* $K : \mathbb{R}^d \to \mathbb{R}$ *is of the product form* $K(x) = \prod_{p=1}^d K_p(x^{(p)})$ *of some functions* $K_p : \mathbb{R} \to \mathbb{R}$, $p = 1, \dots, d$. *The functions* K_p *need not be the same, but each of them satisfies:*
(i) $\int_{\mathbb{R}} K_p(x)dx = 1$, $K_p(x) = K_p(-x)$ *for every* $x \in \mathbb{R}$, *and* K_p *has a compact support;*
(ii) *the function* $x \rightsquigarrow K_p(x)$ *is decreasing on* $[0, \infty)$.

We aim to derive the asymptotic behavior of the sequence of (normalized) residual processes $R^n = (R^n(u)|u \in U)$ defined by

$$R^n(u) = \sqrt{nb_n^d}\left\{\widehat{f}_n(x_0 + b_n u) - f(x_0 + b_n u)\right\} \qquad \forall u \in U.$$

The key point is to investigate the processes $Z^n = (Z^n(u)|u \in U)$ given by

$$Z^n(u) = \sqrt{nb_n^d}\left\{\widehat{f}_n(x_0 + b_n u) - \widetilde{f}_n(x_0 + b_n u)\right\} \qquad \forall u \in U,$$

where

$$\begin{aligned}\widetilde{f}_n(x_0 + b_n u) &= \frac{1}{b_n^d}\int_{\mathbb{R}^d} K\left(\frac{x - x_0}{b_n} - u\right) f(x)dx \\ &= \int_{\mathbb{R}^d} K(y) f(x_0 + b_n(u+y))dy \qquad \forall u \in U.\end{aligned}$$

Notice that the processes R^n and Z^n are not necessarily continuous in the case of a monotone kernel, and thus we treat them as $\ell^\infty(U)$-valued random elements. This is natural especially in the multi-dimensional case.

Proposition 7.1.3 *Choose a kernel function* $K : \mathbb{R}^d \to \mathbb{R}$ *satisfying either Condition 7.1.1 or 7.1.2, and let* $\{b_n\}$ *be a sequence of positive constants such that* $b_n \downarrow 0$ *and that* $nb_n^d \uparrow \infty$ *as* $n \to \infty$. *If* f *is continuous at* x_0, *and if* U *is bounded, then it holds that* $Z^n \Longrightarrow Z$ *in* $\ell^\infty(U)$, *where* $u \rightsquigarrow Z(u)$ *a zero-mean, continuous Gaussian process such that*

$$E(Z(u_1)Z(u_2)) = f(x_0)\int_{\mathbb{R}^d} K(x-u_1)K(x-u_2)dx \qquad \forall u_1, u_2 \in U. \tag{7.1.1}$$

Remark. The continuity of the limit process $u \rightsquigarrow Z(u)$ is considered with respect to the Euclidean metric.

Let $U_1 \subset U_2 \subset \cdots$ be a sequence of bounded subsets of $\mathbb{R}^d$ such that $\bigcup_{i=1}^\infty U_i = \mathbb{R}^d$. We denote by $\ell^\infty_{\text{loc}}(\mathbb{R}^d)$ the set of all functions $z : \mathbb{R}^d \to \mathbb{R}$ that are bounded on every U_i, and equip it with the local uniform metric d defined by

$$d(z_1, z_2) = \sum_{i=1}^\infty \left(\sup_{u \in U_i} |z_1(u) - z_2(u)| \wedge 1\right) 2^{-i}.$$

Using Theorem 1.6.1 of van der Vaart and Wellner (1996), we obtain the following.

Theorem 7.1.4 *Choose a kernel function $K : \mathbb{R}^d \to \mathbb{R}$ and a sequence of constants $\{b_n\}$ as in Proposition 7.1.3.*

(i) *If f is continuous at x_0, then it holds that $Z^n \Longrightarrow Z$ in $\ell^\infty_{\text{loc}}(\mathbb{R}^d)$, where $u \rightsquigarrow Z(u)$ a zero-mean, continuous Gaussian process whose covariance $E(Z(u_1)Z(u_2))$ is given by (7.1.1) for every $u_1, u_2 \in \mathbb{R}^d$.*

(ii) *If f is twice continuously differentiable in a neighborhood of x_0, and if*

$$\lim_{n\to\infty} nb_n^{5d} = h < \infty,$$

then it holds that $R^n \Longrightarrow z_0 + Z$ in $\ell^\infty_{\text{loc}}(\mathbb{R}^d)$, where

$$z_0 = \frac{\sqrt{h}}{2} \sum_{p=1}^{d}\sum_{q=1}^{d} \int_{\mathbb{R}^d} y^{(p)} y^{(q)} K(y)dy \left.\frac{\partial^2 f(x)}{\partial x^{(p)} \partial x^{(q)}}\right|_{x=x_0}. \qquad (7.1.2)$$

This result can be applied to construct a confidence band, substituting estimators for $f(x_0)$ in the covariance of the limit process Z and for the second derivatives of f at x_0 in the constant z_0. Another application is given in Subsection 7.1.2. Notice that the assumptions appearing above are exactly the same as those in the context of point-wise asymptotic normality, and thus are quite reasonable. Our conclusion is that the local smoothness of the density f implies not only the point-wise asymptotic normality but also the weak convergence of local residual processes R^n.

Proof of Proposition 7.1.3. We can write $Z^n(u) = \sum_{i=1}^n \xi_i^n(u)$ where

$$\xi_i^n(u) = \frac{1}{\sqrt{nb_n^d}} \left\{ K\left(\frac{X_i - x_0}{b_n} - u\right) - \int_{\mathbb{R}^d} K\left(\frac{x - x_0}{b_n} - u\right) f(x)dx \right\}.$$

We will check the conditions of Theorem 3.3.2. For every $u_1, u_2 \in U$, since

$$\begin{aligned} E\xi_i^n(u_1)\xi_i^n(u_2) &= \frac{1}{nb_n^d}\left\{ \int_{\mathbb{R}^d} K\left(\frac{x-x_0}{b_n} - u_1\right) K\left(\frac{x-x_0}{b_n} - u_2\right) f(x)dx \right. \\ &\qquad \left. - \int_{\mathbb{R}^d} K\left(\frac{x-x_0}{b_n} - u_1\right) f(x)dx \int_{\mathbb{R}^d} K\left(\frac{x-x_0}{b_n} - u_2\right) f(x)dx \right\} \\ &= \frac{1}{n}\left\{ \int_{\mathbb{R}^d} K(y-u_1)K(y-u_2)f(x_0+b_n y)dy \right. \\ &\qquad \left. - b_n^d \int_{\mathbb{R}^d} K(y-u_1)f(x_0+b_n y)dy \int_{\mathbb{R}^d} K(y-u_2) f(x_0 + b_n y)dy \right\}, \end{aligned}$$

we easily obtain

$$\lim_{n\to\infty} \sum_{i=1}^n E\xi_i^n(u_1)\xi_i^n(u_2) = f(x_0)\int_{\mathbb{R}^d} K(y-u_1)K(y-u_2)dy.$$

The Lindeberg condition **[L2′]** follows from the assumption $nb_n^d \uparrow \infty$. In the following, we will show that **[PE′]** of Theorem 3.3.2 is satisfied under either Condition 7.1.1 or

7.1.2, and that the limit process $u \rightsquigarrow Z(u)$ is continuous with respect to the Euclidean metric.

[The case of smooth kernel.] Assume Condition 7.1.1. First notice that for any $u_1, u_2 \in U$

$$|K(y-u_1) - K(y-u_2)| \leq L|u_1 - u_2|^\alpha \qquad \forall y \in \mathbb{R}^d. \tag{7.1.3}$$

We can take a compact set S which is a common support of the functions $y \rightsquigarrow K(y-u)$ for all $u \in U$. Now, for every $\varepsilon > 0$, choose a finite partition $\Pi(\varepsilon) = \{U(\varepsilon;k) : 1 \leq k \leq N_\Pi(\varepsilon)\}$ of U such that the diameter of each partitioning set is not bigger than $\varepsilon^{1/\alpha}$. This can be done with $N_\Pi(\varepsilon) \leq$ const. $\varepsilon^{-d/\alpha}$; thus it holds that $\int_0^1 \sqrt{\log N_\Pi(\varepsilon)} d\varepsilon < \infty$. On the other hand, it follows from (7.1.3) that if $|u_1 - u_2| \leq \varepsilon^{1/\alpha}$ then

$$|\xi_i^n(u_1) - \xi_i^n(u_2)| \leq \frac{L\varepsilon}{\sqrt{nb_n^d}} \left\{ 1_S\left(\frac{X_i - x_0}{b_n}\right) + \int_{\mathbb{R}^d} 1_S\left(\frac{x - x_0}{b_n}\right) f(x)dx \right\}.$$

We thus have

$$\begin{aligned}
\|\xi^n\|_\Pi^2 &\leq \frac{L^2}{nb_n^d} \sum_{i=1}^n E\left| 1_S\left(\frac{X_i - x_0}{b_n}\right) + \int_{\mathbb{R}^d} 1_S\left(\frac{x - x_0}{b_n}\right) f(x)dx \right|^2 \\
&\leq \frac{4L^2}{b_n^d} \int_{\mathbb{R}^d} 1_S\left(\frac{x - x_0}{b_n}\right) f(x)dx \\
&= 4L^2 \int_{\mathbb{R}^d} 1_S(y) f(x_0 + b_n y)dy \\
&\leq 4L^2 \cdot \mathrm{Leb}(S) \cdot \sup_{x \in N} f(x) \qquad \text{for all sufficiently large } n \in \mathbb{N},
\end{aligned}$$

where N is a neighborhood of x_0. The condition **[PE′]** of Theorem 3.3.2 has been established.

[The case of monotone kernel.] Assume Condition 7.1.2. For every $p = 1, ..., d$, choose a constant $c_p > 0$ such that $[-c_p, c_p]$ is a support of K_p and that $U \subset \prod_{p=1}^d (-c_p, c_p]$.

For every $\varepsilon > 0$ and every $p = 1, ..., d$, we introduce a finite partition $(-c_p, c_p] = \bigcup_{k_p=1}^{N_p(\varepsilon)} I_p(\varepsilon; k_p)$ where $I_p(\varepsilon; k_p) = (\gamma_p(\varepsilon; k_p - 1), \gamma_p(\varepsilon; k_p)]$, such that

$$0 < \gamma_p(\varepsilon; k_p) - \gamma_p(\varepsilon; k_p - 1) \leq \varepsilon^2, \qquad k_p = 1, ..., N_p(\varepsilon).$$

This can be done with $N_p(\varepsilon) \leq [2c_p\varepsilon^{-2}] + 1$. Now, to check the condition **[PE′]** of Theorem 3.3.2, we consider the DFP $\Pi = \{\Pi(\varepsilon)\}_{\varepsilon \in (0,1]}$ of U given by

$$\Pi(\varepsilon) = \left\{ U \cap \prod_{p=1}^d I_p(\varepsilon; k_p) : \quad 1 \leq k_p \leq N_p(\varepsilon),\ 1 \leq p \leq d \right\}.$$

Then, since

$$N_\Pi(\varepsilon) = \#\Pi(\varepsilon) \leq \prod_{p=1}^d N_p(\varepsilon) \leq \prod_{p=1}^d \left(\left[\frac{2c_p}{\varepsilon^2}\right] + 1 \right)$$

we have $\int_0^1 \sqrt{\log N_\Pi(\varepsilon)}d\varepsilon < \infty$.

Next, for every $u^{(p)} \in (-c_p, c_p]$ we define

$$K_p^{n,u^{(p)}}(x^{(p)}) = \frac{1}{\sqrt{b_n}} K_p\left(\frac{x^{(p)} - x_0^{(p)}}{b_n} - u^{(p)}\right), \qquad \forall x^{(p)} \in \mathbb{R}.$$

Then it holds that

$$(7.1.4) \qquad |K_p^{n,u_1^{(p)}} - K_p^{n,u_2^{(p)}}| \le \overline{K}_p^{n,\varepsilon,k_p} \qquad \text{whenever } u_1^{(p)}, u_2^{(p)} \in I_p(\varepsilon; k_p),$$

where

$$\overline{K}_p^{n,\varepsilon,k_p}(x^{(p)}) = \begin{cases} \dfrac{K_p(0)}{\sqrt{b_n}}, & \text{if } \dfrac{x^{(p)} - x_0^{(p)}}{b_n} \in I_p(\varepsilon; k_p), \\ |K_p^{n,\gamma_p(\varepsilon;k_p)} - K_p^{n,\gamma_p(\varepsilon;k_p-1)}|(x^{(p)}), & \text{otherwise.} \end{cases}$$

The key points are the following:

$$(7.1.5) \qquad \int_{\mathbb{R}} |\overline{K}_p^{n,\varepsilon,k_p}(x_0^{(p)} + b_n y^{(p)})|^2 dy^{(p)} \le \frac{3|K_p(0)|^2\varepsilon^2}{b_n};$$

$$(7.1.6) \qquad \text{Support}(\overline{K}_p^{n,\varepsilon,k_p}) \subset [x_0^{(p)} - 2b_n c_p, x_0^{(p)} + 2b_n c_p];$$

$$(7.1.7) \qquad f_0 = \limsup_{n\to\infty} \sup_{y\in S} f(x_0 + b_n y) < \infty, \quad \text{where } S = \prod_{p=1}^d [-2c_p, 2c_p].$$

The fact (7.1.5) will be proved later, while (7.1.6) and (7.1.7) are trivial.

Let us proceed with the main part of the proof. It follows from (7.1.4) that

$$\begin{aligned}
&\frac{1}{\sqrt{b_n^d}}\left|K\left(\frac{x - x_0}{b_n} - u_1\right) - K\left(\frac{x - x_0}{b_n} - u_2\right)\right| \\
&= \left|\prod_{p=1}^d K_p^{n,u_1^{(p)}}(x^{(p)}) - \prod_{p=1}^d K_p^{n,u_2^{(p)}}(x^{(p)})\right| \\
&= \left|\sum_{p=1}^d \left(\prod_{q=1}^{p-1} K_q^{n,u_2^{(q)}}(x^{(q)})\right)\left(\prod_{q=p+1}^d K_q^{n,u_1^{(q)}}(x^{(q)})\right)\left\{K_p^{n,u_1^{(p)}}(x^{(p)}) - K_p^{n,u_2^{(p)}}(x^{(p)})\right\}\right| \\
&\le \sum_{p=1}^d \left(\prod_{q\ne p} \frac{K_q(0)}{\sqrt{b_n}}\right)\left|K_p^{n,u_1^{(p)}}(x^{(p)}) - K_p^{n,u_2^{(p)}}(x^{(p)})\right| \\
&\le \sum_{p=1}^d \left(\prod_{q\ne p} \frac{K_q(0)}{\sqrt{b_n}}\right)\overline{K}_p^{n,\varepsilon,k_p}(x^{(p)}), \qquad \text{if } u_1, u_2 \in \prod_{p=1}^d I_p(\varepsilon; k_p).
\end{aligned}$$

Here, for every $p = 1, ..., d$, we obtain from (7.1.5), (7.1.6) and (7.1.7) that for all sufficiently large $n \in \mathbb{N}$

$$\left(\prod_{q\ne p} \frac{K_q(0)}{\sqrt{b_n}}\right)^2 \int_{\mathbb{R}^d} |\overline{K}_p^{n,\varepsilon,k_p}(x^{(p)})|^2 f(x)dx$$

$$\begin{aligned}
&= \left(\prod_{q\neq p}|K_q(0)|^2\right) b_n \int_{\mathbb{R}^d} |\overline{K}_p^{n,\varepsilon,k_p}(x_0^{(p)}+b_n y^{(p)})|^2 f(x_0+b_n y)dy \\
&\le \left(\prod_{q\neq p}|K_q(0)|^2\right) b_n(f_0+1)\cdot \int_S |\overline{K}_p^{n,\varepsilon,k_p}(x_0^{(p)}+b_n y^{(p)})|^2 dy \\
&\le \left(\prod_{q\neq p}|K_q(0)|^2\right) b_n(f_0+1)\cdot \left(\prod_{q\neq p}^{d} 4c_q\right)\frac{3|K_p(0)|^2\varepsilon^2}{b_n} \\
&= \varepsilon^2 D_p, \qquad \text{where} \qquad D_p = \left(\prod_{q=1}^{d} 4c_q|K_q(0)|^2\right)\frac{3(f_0+1)}{4c_p},
\end{aligned}$$

which implies that

$$\int_{\mathbb{R}^d}\left|\sum_{p=1}^{d}\left(\prod_{q\neq p}\frac{K_q(0)}{\sqrt{b_n}}\right)\overline{K}_p^{n,\varepsilon,k_p}(x^{(p)})\right|^2 f(x)dx \le \varepsilon^2 d\sum_{p=1}^{d} D_p.$$

We therefore have

$$\limsup_{n\to\infty}\|\xi^n\|_{\Pi} \le \sqrt{4d\sum_{p=1}^{d} D_p}.$$

It remains to prove (7.1.5). Observe that

$$\int_{\mathbb{R}} |\overline{K}_p^{n,\varepsilon,k_p}(x_0^{(p)}+b_n y^{(p)})|^2 dy^{(p)} = (I) + \frac{|K_p(0)|^2}{b_n}\varepsilon^2 + (II),$$

where:

$$\begin{aligned}
(I) &= \frac{1}{b_n}\int_{-\infty}^{\gamma_p(\varepsilon;k_p-1)} \left|K_p(y^{(p)}-\gamma_p(\varepsilon;k_p)) - K_p(y^{(p)}-\gamma_p(\varepsilon;k_p-1))\right|^2 dy^{(p)}; \\
(II) &= \frac{1}{b_n}\int_{\gamma_p(\varepsilon;k_p)}^{\infty} \left|K_p(y^{(p)}-\gamma_p(\varepsilon;k_p)) - K_p(y^{(p)}-\gamma_p(\varepsilon;k_p-1))\right|^2 dy^{(p)}.
\end{aligned}$$

Further, it holds that

$$\begin{aligned}
(II) &= \frac{1}{b_n}\int_0^{\infty} \left|K_p(y^{(p)}) - K_p(y^{(p)}+\gamma_p(\varepsilon,k_p)-\gamma_p(\varepsilon,k_p-1))\right|^2 dy^{(p)} \\
&\le \frac{1}{b_n}\int_0^{\infty} \left|K_p(y^{(p)}) - K_p(y^{(p)}+\varepsilon^2)\right|^2 dy^{(p)} \\
&\le \frac{K_p(0)}{b_n}\int_0^{\infty} \left|K_p(y^{(p)}) - K_p(y^{(p)}+\varepsilon^2)\right| dy^{(p)} \\
&= \frac{K_p(0)}{b_n}\int_{-\varepsilon^2}^{0} K_p(y^{(p)}+\varepsilon^2)dy^{(p)} \\
&\le \frac{K_p(0)}{b_n}\cdot K_p(0)\varepsilon^2.
\end{aligned}$$

Since the same bound holds also for (I), we get (7.1.5).

[Continuity of the limit process.] Theorem 3.3.2 says that the process $u \rightsquigarrow Z(u)$ is continuous with respect to the pseudo-metric ρ on U defined by

$$\rho(u_1, u_2) = \sqrt{\int_{\mathbb{R}^d} |K(x - u_1) - K(x - u_2)|^2 dx} \qquad \forall u_1, u_2 \in U.$$

Hence it suffices to show that $u_1 \rightsquigarrow \rho(u_1, u_2)$ is continuous at u_2 with respect to the Euclidean metric for every $u_2 \in U$. This is immediate from (7.1.3) in the case of a smooth kernel. On the other hand, in the case of a monotone kernel, the claim follows from the inequality

$$|K(x - u_1) - K(x - u_2)| \leq \sum_{p=1}^{d} \left(\prod_{q \neq p} K_q(0) \right) \left| K_p(x^{(p)} - u_1^{(p)}) - K_p(x^{(p)} - u_2^{(p)}) \right|$$

which can be easily shown by the same argument as above. □

Proof of Theorem 7.1.4. The assertion (i) is immediate from Proposition 7.1.3 and Theorem 1.6.1 of van der Vaart and Wellner (1996). Next, observe that

$$\widetilde{f}_n(x_0 + b_n u) - f(x_0 + b_n u) = \int_{\mathbb{R}^d} K(y) \{ f(x_0 + b_n(u + y)) - f(x_0 + b_n u) \} dy$$

and that

$$\begin{aligned} f(x_0 + b_n(u + y)) - f(x_0 + b_n u) &= b_n^d \sum_{p=1}^{d} y^{(p)} \left. \frac{\partial f(x)}{\partial x^{(p)}} \right|_{x = x_0 + b_n u} \\ &\quad + \frac{b_n^{2d}}{2} \sum_{p=1}^{d} \sum_{q=1}^{d} y^{(p)} y^{(q)} \left. \frac{\partial^2 f(x)}{\partial x^{(p)} \partial x^{(q)}} \right|_{x = \tilde{x}_n}, \end{aligned}$$

where $\tilde{x}_n$ is a point on the segment connecting $x_0 + b_n u$ and $x_0 + b_n(u + y)$. We can obtain the assertion (ii) using the assumption that the kernel function K is symmetric. □

7.1.2 Estimation of Mode

In this subsection, we consider the 1-dimensional case only. We are interested in estimating the mode of a density f of an i.i.d. data, namely, $x_0 = \mathrm{argmax}_{x \in \mathbb{R}} f(x)$. A natural estimator would be $\widehat{\theta}^n = \mathrm{argmax}_{x \in \mathbb{R}} \widehat{f}_n(x)$, where

$$\widehat{f}_n(x) = \frac{1}{n b_n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{b_n} \right) \qquad \forall x \in \mathbb{R}.$$

Here, b_n is a vanishing sequence of positive constants, and K is a kernel function on $\mathbb{R}$. We now introduce a condition on f in a neighborhood of $x_0 \in \mathbb{R}$ (we do not assume that

x_0 is the maximum point over the whole line $\mathbb{R}$; the point x_0 should be regarded as a local mode of f).

Condition 7.1.5 *For an even integer $p \geq 2$, the function $x \rightsquigarrow f(x)$ is p-times continuously differentiable in a neighborhood N of x_0 with derivatives $f^{(m)}$, $m = 1, ..., p$ such that:*

- $f^{(m)}(x_0) = 0$ *for every* $m = 1, ..., p-1$;
- $\sup_{x \in N} f^{(p)}(x) < 0$.

Theorem 7.1.6 *For a given point $x_0 \in \mathbb{R}$, suppose that Condition 7.1.5 is satisfied for an even integer $p \geq 2$. Put the bandwidth $b_n = n^{-1/(2p+1)}$, and choose a kernel function K on $\mathbb{R}$ following either of Condition 7.1.1 with $\alpha = 1$ or Condition 7.1.2. Then, for any $\mathbb{R}$-valued random sequence $\widehat{\theta}^n$ such that $\widehat{\theta}^n \xrightarrow{P} x_0$ and that*

$$\widehat{f}_n(\widehat{\theta}^n) \geq \widehat{f}_n(x_0) + \epsilon_n$$

for some $\epsilon_n = O_{P^*}(n^{-p/(2p+1)})$, it holds that $|\widehat{\theta}^n - x_0| = O_{P^*}(n^{-1/(2p+1)})$.

Proof. We will check the conditions of Theorem 5.1.2 for $r_{n,u}^{-1} = b_n = n^{-1/(2p+1)}$. Formulation 5.1.1 should be as follows: (Θ^n, d_u^n) is the Euclidean space $\mathbb{R}$, and $\theta_u^n = x_0$; let $\Gamma^n(x) = \widehat{f}_n(x)$ and $\gamma^n(x) = f_n(x)$, where

$$\begin{aligned} f_n(x) &= \frac{1}{b_n}\int_{\mathbb{R}} K\left(\frac{y-x}{b_n}\right) f(y)dy \\ &= \int_{\mathbb{R}} K(y) f(x + b_n y)dy. \end{aligned}$$

We then denote $R(\delta) = \{x \in \mathbb{R} : (\delta/2) < |x - x_0| \leq \delta\}$ for every $\delta \in (0, \infty)$.

To show the FIRST INEQUALITY of M-CRITERION, we denote

$$c_p = \inf_{x \in N} |f^{(p)}(x)| \quad \text{and} \quad C_p = \sup_{x \in N} |f^{(p)}(x)|.$$

It follows from a p-term Taylor expansion of f around $x_0 + b_n y$ that

$$\begin{aligned} \gamma^n(x) - \gamma^n(x_0) &= \int_{\mathbb{R}} K(y)\{f(x + b_n y) - f(x_0 + b_n y)\}\,dy \\ &= (I) + (II), \end{aligned}$$

where

$$(I) = \sum_{m=1}^{p-1} \frac{(x - x_0)^m}{m!} \int_{\mathbb{R}} K(y) f^{(m)}(x_0 + b_n y)dy$$

and

$$\begin{aligned}(II) &\leq -\frac{|x-x_0|^p}{p!}\int_{\mathbb{R}} K(y)c_p dy \\ &\leq -\frac{c_p}{p!2^p}\delta^p \qquad \text{whenever } x \in R(\delta).\end{aligned}$$

Furthermore, since $f^{(k)}(x_0) = 0$ for $k = 1, ..., p-1$, it follows from a $(p-m)$-term Taylor expansion of $f^{(m)}$ around x_0 that

$$f^{(m)}(x_0 + b_n y) = \frac{(b_n y)^{p-m}}{(p-m)!} f^{(p)}(\tilde{x}_n) \quad \text{for some } \tilde{x}_n \in N,$$

for every $m = 1, ..., p-1$. Thus it holds that if $\delta \in [Lb_n, \infty)$ then

$$\begin{aligned}|(I)| &\leq \sum_{m=1}^{p-1} \frac{|x-x_0|^m}{m!} \cdot b_n^{p-m} \cdot \frac{C_p}{(p-m)!} \int_{\mathbb{R}} |y|^{p-m} K(y) dy \\ &\leq \delta^p \sum_{m=1}^{p-1} \frac{C_p}{m!(p-m)!L^{p-m}} \int_{\mathbb{R}} |y|^{p-m} K(y) dy \qquad \text{whenever } x \in R(\delta).\end{aligned}$$

Thus, choosing a sufficiently large $L > 0$ we can conclude that the FIRST INEQUALITY is satisfied.

To check the SECOND INEQUALITY, we will apply Theorem 2.3.3 (ii) to $\{\xi_i\}_{i\in\mathbb{N}} = \{(\xi_i^{n,x} | x \in \Psi)\}_{i\in\mathbb{N}}$, where Ψ is a subset of $\mathbb{R}$ and

$$\xi_i^{n,x} = \frac{1}{n}\left\{\frac{1}{b_n} K\left(\frac{X_i - x}{b_n}\right) - f_n(x)\right\}.$$

Notice that

$$\begin{aligned}|\xi_i^{n,x} - \xi_i^{n,y}| &\leq \frac{1}{nb_n}\left|K\left(\frac{X_i - x}{b_n}\right) - K\left(\frac{X_i - y}{b_n}\right)\right| \\ &\quad + \frac{1}{nb_n}\int_{\mathbb{R}}\left|K\left(\frac{z-x}{b_n}\right) - K\left(\frac{z-y}{b_n}\right)\right| f(z) dz.\end{aligned}$$

We may choose a sufficiently small $\delta_0 > 0$ so that $[x_0 - 2\delta_0, x_0 + 2\delta_0] \subset N$; then we have $f^* = \sup_{z\in[x_0-2\delta_0, x_0+2\delta_0]} f(z) < \infty$. We discuss the cases of smooth and monotone kernels, separately; in both cases, let $L > 0$ be a constant such that $[-L, L]$ is a support of K. In each case, for every $\delta \in (0, \delta_0)$ we will construct a DFP $\Pi_\delta = \{\Pi_\delta(\varepsilon)\}_{\varepsilon\in(0,1]}$ of $\Psi = \Psi_\delta = [x_0 - \delta, x_0 + \delta]$ such that

$$\sup_{\delta\in(0,\delta_0)} \int_0^1 \sqrt{\log(1 + N_{\Pi_\delta}(\varepsilon))} d\varepsilon < \infty \tag{7.1.8}$$

and that

$$\|\xi^n\|_{\Pi_\delta} \leq \text{const.}\varphi_1^n(\delta) \qquad \forall \delta \in [Lb_n, \delta_0),\ \forall n \in \mathbb{N}$$

for some appropriate functions $\delta \rightsquigarrow \varphi_1^n(\delta)$ indexed by $n \in \mathbb{N}$. Notice that we have in both cases that with $\Psi_\delta = [x_0 - \delta, x_0 + \delta]$

$$\begin{aligned}\sum_{i=1}^n E|\overline{\xi}_i^n|^2 &\leq n \cdot \left|\frac{\sup_x K(x)}{nb_n}\right|^2 \cdot (2\delta + 2Lb_n) \cdot f^* \\ &\leq D \cdot \varphi_2^n(\delta) \qquad \text{whenever } \delta \in [Lb_n, \delta_0)\end{aligned}$$

where $D = 4\sup_x |K(x)|^2 f^*$ and $\varphi_2^n(\delta) = n^{-1}b_n^{-2}\delta$. Then, Theorem 2.3.3 (ii) yields that

$$E \sup_{x \in R(\delta)} |(\Gamma^n - \gamma^n)(x) - (\Gamma^n - \gamma^n)(x_0)| \leq \text{const.}\phi^n(\delta) \qquad \forall \delta \in [Lb_n, \delta_0)$$

where $\phi^n(\delta) = \varphi_1^n(\delta) \vee (\varphi_2^n(\delta)/\varphi_1^n(\delta))$.

[The case of smooth kernel.] For every $\varepsilon \in (0,1]$, we make a finite partition $(x_0 - \delta, x_0 + \delta] = \bigcup_{k=1}^{N_{\Pi_\delta}(\varepsilon)} (u_{k-1}, u_k]$ such that $u_k - u_{k-1} \leq \varepsilon\delta$. This can be done with $N_{\Pi_\delta}(\varepsilon) \leq [2\varepsilon^{-1}] + 1$, thus (7.1.8) is satisfied. On the other hand, if $x, y \in [u_{k-1}, u_k]$ with $u_k - u_{k-1} \leq \varepsilon\delta$, then

$$\left|K\left(\frac{z-x}{b_n}\right) - K\left(\frac{z-y}{b_n}\right)\right| \leq \frac{L\varepsilon\delta}{b_n} \cdot 1_{[u_{k-1}-Lb_n, u_k+Lb_n]}(z) \qquad \forall z \in \mathbb{R},$$

where $L > 0$ is a constant appearing Condition 7.1.1. Thus it holds that if $Lb_n \leq \delta < \delta_0$ then

$$\begin{aligned}||\xi^n||_{\Pi_\delta} &\leq \sqrt{4n \cdot \frac{1}{n^2 b_n^2} \cdot \left|\frac{L\delta}{b_n}\right|^2 \cdot (\delta + 2Lb_n) \cdot f^*} \\ &\leq \sqrt{12L^2 f^*} \cdot \varphi_1^n(\delta)\end{aligned}$$

where $\varphi_1^n(\delta) = n^{-1/2}b_n^{-2}\delta^{3/2}$. We thus have $\phi^n(\delta) = (n^{-1/2}b_n^{-2}\delta^{3/2}) \vee (n^{-1/2}\delta^{1/2})$. The relation $\phi^n(b_n) = b_n^p$ holds since we put $b_n = n^{-1/(2p+1)}$.

[The case of uniform kernel.] For every $\varepsilon \in (0,1]$, we make a finite partition $(x_0 - \delta, x_0 + \delta] = \bigcup_{k=1}^{N_{\Pi_\delta}(\varepsilon)} (u_{k-1}, u_k]$ such that $u_k - u_{k-1} \leq \varepsilon^2\delta$. This can be done with $N_{\Pi_\delta}(\varepsilon) \leq [2\varepsilon^{-2}] + 1$, thus (7.1.8) is satisfied. On the other hand, if $x, y \in [u_{k-1}, u_k]$ then

$$\left|K\left(\frac{z-x}{b_n}\right) - K\left(\frac{z-y}{b_n}\right)\right| \leq \left|K\left(\frac{z-u_{k-1}}{b_n}\right) - K\left(\frac{z-u_k}{b_n}\right)\right| \qquad \forall z \in \mathbb{R}.$$

Here notice that if $u_k - u_{k-1} \leq \varepsilon^2\delta$ then

$$\begin{aligned}\int_{\mathbb{R}} \left|K\left(\frac{z-u_{k-1}}{b_n}\right) - K\left(\frac{z-u_k}{b_n}\right)\right|^2 dz &= \int_{\mathbb{R}} \left|K\left(\frac{z}{b_n}\right) - K\left(\frac{z-\varepsilon^2\delta}{b_n}\right)\right|^2 dz \\ &\leq K(0) \int_{\mathbb{R}} \left|K\left(\frac{z}{b_n}\right) - K\left(\frac{z-\varepsilon^2\delta}{b_n}\right)\right| dz \\ &\leq 3|K(0)|^2\varepsilon^2\delta.\end{aligned}$$

Thus it holds that if $Lb_n \leq \delta < \delta_0$ then

$$\begin{aligned} \|\xi^n\|_{\Pi_\delta} &\leq \sqrt{4n \cdot \frac{1}{n^2 b_n^2} \cdot 3|K(0)|^2 \delta \cdot f^*} \\ &\leq \sqrt{12|K(0)|^2 f^*} \cdot \varphi_1^n(\delta) \end{aligned}$$

where $\varphi_1^n(\delta) = n^{-1/2} b_n^{-1} \delta^{1/2}$. We thus have $\phi^n(\delta) = n^{-1/2} b_n^{-1} \delta^{1/2}$. The relation $\phi^n(b_n) = b_n^p$ holds since we put $b_n = n^{-1/(2p+1)}$. □

Corollary 7.1.7 *For a given point $x_0 \in \mathbb{R}$, suppose that Condition 7.1.5 is satisfied for an even integer $p \geq 2$. Put the bandwidth $b_n = n^{-1/(2p+1)}$, and set $K(x) = \frac{1}{2} \cdot 1_{[-1,1]}(x)$. Then, for any $\mathbb{R}$-valued random sequence $\widehat{\theta}^n$ such that $\widehat{\theta}^n \xrightarrow{P^*} x_0$ and that*

$$\widehat{f}_n(\widehat{\theta}^n) \geq \sup_{x \in \mathbb{R}} \widehat{f}_n(x) + \epsilon_n$$

for some $\epsilon_n = o_{P^*}(n^{-p/(2p+1)})$, it holds that $n^{1/(2p+1)}(\widehat{\theta}^n - x_0) \stackrel{P}{\Longrightarrow} \mathrm{argmax}_{h \in \mathbb{R}}\{\mathbb{A}(h) + \mathbb{B}(h+1) - \mathbb{B}(h-1)\}$, where the deterministic process $h \rightsquigarrow \mathbb{A}(h)$ is given by

$$\mathbb{A}(h) = \frac{f^{(p)}(x_0)}{\sqrt{f(x_0)}p!} \int_{-1}^{1} (h+y)^p dy$$

and $h \rightsquigarrow \mathbb{B}(h)$ is the two-sided Brownian motion.

Proof. Theorem 7.1.6 asserts that the sequence $b_n^{-1}(\widehat{\theta}^n - x_0)$ is uniformly tight. Let us consider the stochastic process $h \rightsquigarrow \mathbb{M}^n(h)$ defined by

$$\begin{aligned} \mathbb{M}^n(h) &= b_n^{-p}\left\{\widehat{f}_n(x_0 + b_n h) - \widehat{f}_n(x_0)\right\} \\ &= Y^n(h) + Z^n(h), \end{aligned}$$

where:

$$\begin{aligned} Y^n(h) &= b_n^{-p}\left\{f_n(x_0 + b_n h) - f(x_0)\right\}; \\ Z^n(h) &= b_n^{-p}\left\{\widehat{f}_n(x_0 + b_n h) - f_n(x_0 + b_n h)\right\}. \end{aligned}$$

Noting that $\sqrt{nb_n} = b_n^{-p}$, we obtain from Proposition 7.1.3 that $Z^n \stackrel{P}{\Longrightarrow} Z$ in $\ell^\infty(K)$ for any compact set $K \subset \mathbb{R}$, where $Z(h) = \sqrt{f(x_0)}\{\mathbb{B}(h+1) - \mathbb{B}(h-1)\}/2$. On the other hand, an easy computation shows that $\lim_{n\to\infty} Y^n(h) = Y(h) = \sqrt{f(x_0)}\mathbb{A}(h)/2$ for every $h \in \mathbb{R}$. Since $h \rightsquigarrow Y^n(h)$ and $h \rightsquigarrow Y(h)$ are continuous, this convergence is uniform on every compact set $K \subset \mathbb{R}$. Hence, by the same argument as the last part of the proof of Proposition 5.2.5, we can obtain the assertion. □

7.1.3 Remarks for Non-I.I.D. Cases

Gaussian White Noise Model

For every $n \in \mathbb{N}$, let $X^n = (X_t^n)_{t\in[0,1]}$ be a continuous stochastic process given by

$$dX_t^n = f(t)dt + n^{-1/2}dB_t,$$

where $f \in \mathcal{L}^2[0,1]$, and $B = (B_t)_{t\in[0,1]}$ is a standard Brownian motion. Let $t_0 \in (0,1)$ be a fixed point, and let b_n be a vanishing sequence of positive constants. In order to estimate the local function $u \rightsquigarrow f(t_0 + b_n u)$, a natural estimator would be

$$\widehat{f}_n(t_0 + b_n u) = \frac{1}{nb_n}\int_0^1 K\left(\frac{t-t_0}{b_n}\right) dX_t^n$$

where K is a kernel function on $\mathbb{R}$ satisfying either of Condition 7.1.1 or 7.1.2 with $d = 1$. Then, we can get the same conclusions as Theorem 7.1.4 by using Theorem 3.4.4.

Multiplicative Intensity Model

For every $n \in \mathbb{N}$, let μ^n be an $\mathbb{R}^d$-valued multivariate point process on a stochastic basis $\mathbf{B}^n = (\Omega^n, \mathcal{F}^n, \mathbf{F}^n, P^n)$; μ^n can be identified with an $\mathbb{R}^d$-marked point process $\{(T_i^n, Z_i^n);\ i \in \mathbb{N}\}$ through the equality

$$\mu^n(\omega; dt, dz) = \sum_i \varepsilon_{(T_i^n(\omega), Z_i^n(\omega))}(dt, dz),$$

where $0 < T_1^n < T_2^n < \cdots$ and each Z_i^n is an $\mathbb{R}^d$-valued random variable.

We assume that the predictable compensator ν^n of μ^n is given by

$$\nu^n(\omega; dt, dz) = \alpha(t,z)Y^n(\omega,t,z)dtdz,$$

where $\alpha(t,z)$ is a $[0,\infty)$-valued measurable function on $\mathbb{R}_+ \times \mathbb{R}^d$, and $Y^n(\omega,t,z)$ is a $[0,\infty)$-valued predictable function on $\Omega^n \times \mathbb{R}_+ \times \mathbb{R}^d$.

Let $(t_0, z_0) \in (0,\infty) \times \mathbb{R}^d$ be a fixed point, and let b_n and c_n be vanishing sequences of positive constants. In order to estimate the local function $(u,v) \rightsquigarrow \alpha(t_0 + b_n u, z_0 + c_n v)$, a natural estimator would be

$$\widehat{\alpha}_n(t_0 + b_n u, z_0 + c_n v) = \frac{1}{nb_n c_n^d}\int_{\mathbb{R}_+\times\mathbb{R}^d} K\left(\frac{t-t_0}{b_n}, \frac{z-z_0}{c_n}\right) Y^{n-}(\cdot,t,z)\mu^n(\cdot; dt, dz),$$

where K is a kernel function on $\mathbb{R}^{d+1}$ satisfying either of Condition 7.1.1 or 7.1.2 with "d" replaced by "$d+1$", and where Y^{n-} is the generalized inverse of Y^n.

Assume the "local" version of Condition 5.3.1, that is:

Condition 7.1.8 *There exists a measurable function $y = y(t, z)$ on a neighborhood N of (t_0, z_0) such that*

$$\sup_{(t,z)\in N} \left| n^{-1}Y^n(\cdot, t, z) - y(t, z) \right| \xrightarrow{P^{n*}} 0.$$

Then, under some conditions of smoothness of the functions $(t, z) \rightsquigarrow \alpha(t, z)$ and $(t, z) \rightsquigarrow y(t, z)$, we can derive some conclusions about the residual processes $(u, v) \rightsquigarrow R^n(u, v)$ given by

$$R^n(u, v) = \sqrt{nb_n c_n^d}\,\{\widehat{\alpha}_n(t_0 + b_n u, z_0 + c_n v) - \alpha(t_0 + b_n u, z_0 + c_n v)\}$$

similarly to those of Theorem 7.1.4 by using Theorem 3.2.4; the term "$f(x_0)$" in (7.1.1) is replaced by "$\alpha(t_0, z_0)/y(t_0, z_0)$", while the change of (7.1.2) is clear.

7.2 Log-likelihood Ratio Random Fields

7.2.1 Results

For every $n \in \mathbb{N}$, let $\mathbf{B}^n = (\Omega^n, \mathcal{F}^n, \mathbf{F}^n = \{\mathcal{F}_i^n\}_{i\in\mathbb{N}_0}, P^n)$ be a discrete-time stochastic basis. Let $\mathbf{P}^n = \{P^{n,\psi} : \psi \in \Psi\}$ be a family of probability measures on $(\Omega^n, \mathcal{F}^n)$, indexed by an arbitrary set Ψ, such that $P^{n,\psi} \ll P^n$ for every $\psi \in \Psi$. We denote

$$Z_i^{n,\psi} = \frac{dP_i^{n,\psi}}{dP_i^n},$$

where $P_i^{n,\psi}$ (resp. P_i^n) is the restriction of $P^{n,\psi}$ (resp. P^n) on the σ-field $\mathcal{F}_i^n$. We assume $P_0^{n,\psi} = P_0^n$ for every $\psi \in \Psi$, hence we can set $Z_0^{n,\psi} = 1$. For a given finite stopping time σ^n on $\mathbf{B}^n$, and we suppose also that the random element $\log Z_{\sigma^n}^n = (\log Z_{\sigma^n}^{n,\psi} | \psi \in \Psi)$ takes values in $\ell^\infty(\Psi)$. Here we set

$$\zeta_i^{n,\psi} = \sqrt{\frac{Z_{i\wedge\sigma^n}^{n,\psi}}{Z_{(i-1)\wedge\sigma^n}^{n,\psi}}} - 1 \qquad \forall i \in \mathbb{N}\ \forall \psi \in \Psi.$$

Theorem 7.2.1 *In the above situation, suppose that the following conditions hold:*
(a$_1$) $\sum_{i=1}^{\sigma^n} 4E_{i-1}^n \zeta_i^{n,\psi}\zeta_i^{n,\phi} \xrightarrow{P^n} C(\psi, \phi)$ *(some constant) for every $\psi, \phi \in \Psi$;*
(a$_2$) $\sup_{\psi\in\Psi} \left| \sum_{i=1}^{\sigma^n} 4E_{i-1}^n |\zeta_i^{n,\psi}|^2 - C(\psi, \psi) \right| \xrightarrow{P^{n*}} 0$;
(b) $\sum_{i=1}^{\sigma^n} E_{i-1}^n |\overline{\zeta_i^n}|^2 1_{\{|\overline{\zeta_i^n}|>\varepsilon\}} \xrightarrow{P^n} 0$ *for every $\varepsilon > 0$;*
(c) *there exists a DFP Π of Ψ such that*

$$||\zeta^n||_{\Pi,\sigma^n} = O_{P^n}(1) \quad \text{and} \quad \int_0^{\Delta_\Pi} H_\Pi(\varepsilon)d\varepsilon < \infty.$$

Then, it holds that $\log Z^n_{\sigma^n} \overset{P^n}{\Longrightarrow} X$ *in* $\ell^\infty(\Psi)$, *where* $X(\psi) = -\frac{1}{2}C(\psi,\psi) + G(\psi)$ *and* $\psi \rightsquigarrow G(\psi)$ *is a zero-mean Gaussian process such that* $EG(\psi)G(\phi) = C(\psi,\phi)$. *Furthermore, the formula*

$$\rho(\psi,\phi) = \sqrt{C(\psi,\psi) + C(\phi,\phi) - 2C(\psi,\phi)} \qquad \forall \psi,\phi \in \Psi$$

defines a semimetric on Ψ *such that* (Ψ,ρ) *is totally bounded and that almost all paths of* X *are uniformly* ρ*-continuous.*

Remark. If a version of the conditional expectation $E^n_{i-1}\zeta_i^{n,\psi}\zeta_i^{n,\phi}$ satisfies the assumption (a_1), then so does any version. However, this is not true in (a_2); the assumption means that there exist *some* versions of $E^n_{i-1}|\zeta_i^{n,\psi}|^2$'s which satisfy the requirement.

Example: Ergodic Markov chains

Let $\{X_i\}_{i\in\mathbb{N}_0}$ be an ergodic Markov chain, defined on a probability space $(\Omega,\mathcal{F},P)$, with values in an arbitrary state space $(E,\mathcal{E})$. Let $\mu(dx)$ denote the initial distribution, $p(x,dy)$ the transition distribution, and $\pi(dx)$ the invariant distribution. Let us equip the space $\mathcal{L}^2 = \mathcal{L}^2(E\times E, \pi(dx)p(x,dy))$ with the "inner product" given by

$$\langle h_1,h_2\rangle_{\mathcal{L}^2} = \int_{E\times E} h_1(x,y)h_2(x,y)\pi(dx)p(x,dy) \qquad \forall h_1,h_2 \in \mathcal{L}^2.$$

The meaning of the quotation marks is that $\|h\|_{\mathcal{L}^2} = \sqrt{\langle h,h\rangle_{\mathcal{L}^2}}$ is merely a "semi-"norm. Next we define the subset $\mathcal{L}^2_0$ of $\mathcal{L}^2$ by

$$\mathcal{L}^2_0 = \left\{ h \in \mathcal{L}^2 : \quad \int_E h(x,y)p(x,dy) = 0 \quad \forall x \in E \quad \text{and} \quad h > -1 \right\}.$$

Fix a subset $\mathcal{H} \subset \mathcal{L}^2_0$. For every $n \in \mathbb{N}$, let us consider a family of probability measures $\mathbf{P}^n = \{P^{n,h} : h \in \mathcal{H}\}$ on $(\Omega,\mathcal{F})$ such that: under $P^{n,h}$, the process $\{X_i\}_{i\in\mathbb{N}_0}$ is the Markov chain with initial distribution μ and transition distribution $p^{n,h}$ given by

$$p^{n,h}(x,dy) = \left(1 + \frac{h(x,y)}{\sqrt{n}}\right) p(x,dy).$$

Here we set $\mathcal{F}_i = \sigma\{X_0,...,X_i\}$. Then it holds that

$$Z_i^{n,h} = \frac{dP_i^{n,h}}{dP_i} = \prod_{j=1}^{i}\left(1 + \frac{h(X_{j-1},X_j)}{\sqrt{n}}\right).$$

We need some more notations to state the following result, which concerns the asymptotic behavior of the process $\log Z^n_n = (\log Z^{n,h}_n | h \in \mathcal{H})$. For a given $K \in \mathcal{L}^2(E,\pi(dx))$ we define the semimetric ρ_K on $\mathcal{L}^2$ by

$$\rho_K(h_1,h_2) = \sup_{x\in E}\frac{\rho_x(h_1,h_2)}{|K(x)| \vee 1} \qquad \forall h_1,h_2 \in \mathcal{L}^2,$$

where

$$\rho_x(h_1,h_2) = \sqrt{\int_E |h_1(x,y) - h_2(x,y)|^2 p(x,dy)} \qquad \forall x \in E.$$

Proposition 7.2.2 *Let $\{X_i\}_{i\in\mathbb{N}_0}$, $(\Omega, \mathcal{F}, \mathbf{F} = \{\mathcal{F}_i\}_{i\in\mathbb{N}_0}, P)$ and $\mathbf{P}^n = \{P^{n,h} : h \in \mathcal{H}\}$ as above be given. Suppose that there exists $h^* \in \mathcal{L}^4(E \times E, \pi(dx)p(x,dy))$ such that $\sup_{h\in\mathcal{H}} |h| \leq h^*$, and also that there exists $K \in \mathcal{L}^2(E, \pi(dx))$ such that*

$$\int_0^1 \sqrt{\log N_{[]}(\mathcal{H}, \rho_K; \varepsilon)}d\varepsilon < \infty.$$

Then, it holds that $\log Z_n^n \stackrel{P}{\Longrightarrow} X$ in $\ell^\infty(\mathcal{H})$, where $X(h) = -\frac{1}{2}\|h\|_{\mathcal{L}^2}^2 + G(h)$ and $h \rightsquigarrow G(h)$ is a zero-mean Gaussian process such that $EG(h_1)G(h_2) = \langle h_1, h_2\rangle_{\mathcal{L}^2}$. Furthermore, almost all paths of X are uniformly $\|\cdot\|_{\mathcal{L}^2}$-continuous.

This result is easily derived from the ergodic theorem and Theorem 7.2.1, hence the proof is omitted. Here we give a statistical application. Fix a subset $\mathcal{H} \subset \mathcal{L}_0^2$ such that $\|h\|_{\mathcal{L}^2} > 0$ for every $h \in \mathcal{H}$. Let us consider the testing problem:

$$\begin{array}{rll} \text{hypothesis} & H_0: & p \\ \text{against} & H_1^n: & p^{n,h} \quad \text{for some } h \in \mathcal{H}. \end{array}$$

We propose the test statistics

$$S^n = \sup_{h\in\mathcal{H}} \left| \frac{1}{2}\|h\|_{\mathcal{L}^2}^2 + \log Z_n^{n,h} \right|.$$

Assume the same conditions as in Proposition 7.2.2. Then, it holds that

$$S^n \stackrel{P^{n,u}}{\Longrightarrow} \sup_{h\in\mathcal{H}} |\langle h, u\rangle_{\mathcal{L}^2} + G(h)| \qquad \text{in } \mathbb{R} \qquad \forall u \in \{0\} \cup \mathcal{H},$$

where the process $h \rightsquigarrow G(h)$ is as above. This fact follows from Proposition 7.2.2 that implies local asymptotic normality and contiguity, together with Le Cam's third lemma and the continuous mapping theorem. In view of Anderson's lemma (e.g., Lemma 3.11.4 of van der Vaart and Wellner (1996)), the statistics S^n seems reasonable.

7.2.2 Proofs

Let us denote:

$$\begin{aligned} \tilde{Z}_i^{n,\psi} &= \frac{Z_{i\wedge\sigma^n}^{n,\psi}}{Z_{(i-1)\wedge\sigma^n}^{n,\psi}} && \forall i \in \mathbb{N}\ \forall\psi \in \Psi; \\ \lambda_i^{n,a,\psi} &= \log \tilde{Z}_i^{n,\psi} 1_{\{\bar{\zeta}_i^n \leq a\}} && \forall i \in \mathbb{N}\ \forall\psi \in \Psi\ \forall a > 0. \end{aligned}$$

The process $\psi \rightsquigarrow \log Z^{n,\psi}_{\sigma^n} = \sum_{i=1}^{\sigma^n} \log \tilde{Z}^{n,\psi}_i$ can be well-approximated by the process $\psi \rightsquigarrow \Lambda^{n,a,\psi} = \sum_{i=1}^{\sigma^n} \lambda^{n,a,\psi}_i$. As a matter of fact, it holds that

$$\begin{aligned} \sup_{\psi\in\Psi} 1_{\{\log Z^{n,\psi}_{\sigma^n} \neq \Lambda^{n,a,\psi}\}} &\leq \sum_{i=1}^{\sigma^n} 1_{\{\overline{\zeta}^n_i > a\}} \\ &\leq \frac{1}{a^2}\sum_{i=1}^{\sigma^n} |\overline{\zeta}^n_i|^2 1_{\{\overline{\zeta}^n_i > a\}}, \end{aligned}$$

hence using also Lenglart's inequality we obtain

$$\sup_{\psi\in\Psi} |\log Z^{n,\psi}_{\sigma^n} - \Lambda^{n,a,\psi}| \xrightarrow{P^{n*}} 0.$$

We consider the decomposition

$$\Lambda^{n,a,\psi} = \sum_{i=1}^{\sigma^n} E^n_{i-1}\lambda^{n,a,\psi}_i + \sum_{i=1}^{\sigma^n}\left\{\lambda^{n,a,\psi}_i - E^n_{i-1}\lambda^{n,a,\psi}_i\right\}. \tag{7.2.1}$$

We will derive the uniform convergence of the first term in (outer) probability, and apply Theorem 3.3.2 to the martingale difference array $\{\xi^n_i\}_{i\in\mathbb{N}_0}$ of the second term, that is, $\xi^{n,\psi}_i = \lambda^{n,a,\psi}_i - E^n_{i-1}\lambda^{n,a,\psi}_i$. We use the following lemma which will be proved later.

Lemma 7.2.3 *For every $a \in (0,1)$, there exist some versions of the conditional expectations $E^n_{i-1}\lambda^{n,a,\psi}_i$ such that:*
(i) *if $\sup_{\psi\in\Psi} C(\psi,\psi) < \infty$ then $\sup_{\psi\in\Psi}\left|\sum_{i=1}^{\sigma^n} E^n_{i-1}\lambda^{n,a,\psi}_i + \frac{1}{2}C(\psi,\psi)\right| \xrightarrow{P^{n*}} 0$;*
(ii) *$\sum_{i=1}^{\sigma^n} E^n_{i-1}\lambda^{n,a,\psi}_i\lambda^{n,a,\phi}_i \xrightarrow{P^n} C(\psi,\phi)$ for every $\psi,\phi\in\Psi$;*
(iii) *$\sum_{i=1}^{\sigma^n} |E^n_{i-1}\lambda^{n,a,\psi}_i|^2 \xrightarrow{P^n} 0$ for every $\psi\in\Psi$.*

Remark. (i) We will see later that the conditions of the theorem actually implies that $\sup_{\psi\in\Psi} C(\psi,\psi) < \infty$. (ii) The choice of versions of the conditional expectations $E^n_{i-1}\lambda^{n,a,\psi}_i\lambda^{n,a,\phi}_i$ is not important.

Let us proceed with the main part of the proof. It is clear that there exists a constant $\delta \in (0,1)$ such that $|\log x - \log y| \leq 2|\sqrt{x}-\sqrt{y}|$ whenever $x, y \in [1-\delta, 1+\delta]$. We consider the decomposition (7.2.1) for $a = \sqrt{1+\delta} - 1$; then it holds that $\{x : |\sqrt{x}-1| \leq a\} \subset \{x : |x-1| \leq \delta\}$.

First we show the weak convergence of the second term of the decomposition (7.2.1). The condition **[L2′]** is direct from (ii) and (iii) of Lemma 7.2.3. It is also easy to see that the assumption (b) implies the Lindeberg condition **[L2′]**. Finally, recalling the choice of δ and the relationship between a and δ, we have for any subset $\Psi' \subset \Psi$

$$E^n_{i-1}\left[\sup_{\psi,\phi\in\Psi'} |\lambda^{n,a,\psi}_i - \lambda^{n,a,\phi}_i|\right]^2_{\mathcal{F}^n_i, P^n}$$

$$\begin{aligned} &= E_{i-1}^n \left[\sup_{\psi,\phi\in\Psi} |\log \tilde{Z}_i^{n,\psi} - \log \tilde{Z}_i^{n,\phi}| 1_{\{\overline{\zeta}_i^n \le a\}} \right]^2_{\mathcal{F}_i^n, P^n} \\ &\le E_{i-1}^n \left[\sup_{\psi,\phi\in\Psi'} 2|\zeta_i^{n,\psi} - \zeta_i^{n,\phi}| \right]^2_{\mathcal{F}_i^n, P^n}. \end{aligned}$$

Thus the assumption (c) implies the condition **[PE′]**. Consequently, Theorem 3.3.2 yields $\sum_i^n \xi_i^n \overset{P^n}{\Longrightarrow} G$ in $\ell^\infty(\Psi)$.

Next we consider the first term of the decomposition. Observe that

$$\begin{aligned} \sqrt{C(\psi,\psi)} &= \sqrt{E|G(\psi)|^2} \\ &\le \sqrt{E|G(\psi)-G(\phi)|^2} + \sqrt{E|G(\phi)|^2} = \rho(\psi,\phi) + \sqrt{C(\psi,\phi)}. \end{aligned}$$

The inequality above and the total boundedness of (Ψ,ρ), a consequence of Theorem 3.3.2, imply that $\sup_{\psi\in\Psi} C(\psi,\psi) < \infty$. Hence (i) of Lemma 7.2.3 works to show the uniform convergence of the first term of (7.2.1). Also, it is trivial from the above inequality that $\psi \rightsquigarrow \sqrt{C(\psi,\psi)}$ is uniformly ρ-continuous, thus so is $\psi \rightsquigarrow C(\psi,\psi)$.

Proof of Lemma 7.2.3. For every $\varepsilon > 0$ we denote:

$$\begin{aligned} B^{n,\varepsilon}(\psi) &= \sum_{i=1}^{\sigma^n} E_{i-1}^n \lambda_i^{n,\varepsilon,\psi}; \\ C^{n,\varepsilon}(\psi,\phi) &= \sum_{i=1}^{\sigma^n} E_{i-1}^n \lambda_i^{n,\varepsilon,\psi} \lambda_i^{n,\varepsilon,\phi}. \end{aligned}$$

[STEP 1] First we prove the following facts: for given $a \in (0,1)$ there exist constants $K_1, K_2, K_3 > 0$ such that for every $\varepsilon \in (0,a]$

$$\sup_{\psi\in\Psi} \left| B^{n,\varepsilon}(\psi) + \frac{1}{2} C(\psi,\psi) \right| \le \varepsilon K_1 + o_{P^n}(1), \tag{7.2.2}$$

$$|C^{n,\varepsilon}(\psi,\phi) - C(\psi,\phi)| \le \varepsilon K_2 + o_{P^n}(1) \qquad \forall \psi,\phi \in \Psi, \tag{7.2.3}$$

$$\sum_{i=1}^{\sigma^n} |E_{i-1}^n \lambda_i^{n,\varepsilon,\psi}| \le \frac{1}{2} C(\psi,\psi) + \varepsilon K_3 + o_{P^n}(1) \qquad \forall \psi \in \Psi. \tag{7.2.4}$$

In order to show (7.2.2), first notice that there exists a constant $K > 0$ such that $|\log x - (x-1) + 2(\sqrt{x}-1)^2| \le K|\sqrt{x}-1|^3$ whenever $|\sqrt{x}-1| \le a$. Hence, for fixed $\varepsilon \in (0,a]$ we obtain

$$\begin{aligned} &\left| B^{n,\varepsilon}(\psi) + 2\sum_{i=1}^{\sigma^n} E_{i-1}^n |\zeta_i^{n,\psi}|^2 1_{\{\overline{\zeta}_i^n \le \varepsilon\}} \right| \\ &\le \varepsilon K \sum_{i=1}^{\sigma^n} E_{i-1}^n |\zeta_i^{n,\psi}|^2 1_{\{\overline{\zeta}_i^n \le \varepsilon\}} + \left| \sum_{i=1}^{\sigma^n} E_{i-1}^n (\tilde{Z}_i^{n,\psi} - 1) 1_{\{\overline{\zeta}_i^n \le \varepsilon\}} \right| \end{aligned}$$

almost surely. Since $E^n_{i-1}(\tilde{Z}^{n,\psi}_i - 1) = 0$ almost surely, the last term on the right hand side equals to

$$\begin{aligned}&\left|\sum_{i=1}^{\sigma^n} E^n_{i-1}(\tilde{Z}^{n,\psi}_i - 1)1_{\{\overline{\zeta}^n_i > \varepsilon\}}\right| \\ &\leq \sum_{i=1}^{\sigma^n} E^n_{i-1}|\tilde{Z}^{n,\psi}_i - 1|1_{\{\overline{\zeta}^n_i > \varepsilon\}} \\ &\leq \frac{\varepsilon+2}{\varepsilon}\sum_{i=1}^{\sigma^n} E^n_{i-1}|\zeta^{n,\psi}_i|^2 1_{\{\overline{\zeta}^n_i > \varepsilon\}}\end{aligned}$$

almost surely. Thus we obtain

$$\begin{aligned}&\left|B^{n,\varepsilon}(\psi) + \frac{1}{2}C(\psi,\psi)\right| \\ &\leq (2+\varepsilon K)\left|\sum_{i=1}^{\sigma^n} E^n_{i-1}|\zeta^{n,\psi}_i|^2 - \frac{1}{4}C(\psi,\psi)\right| + \varepsilon K \cdot \frac{1}{4}C(\psi,\psi) \\ &\quad + \left(2 + \frac{\varepsilon+2}{\varepsilon}\right)\sum_{i=1}^{\sigma^n} E^n_{i-1}|\overline{\zeta}^n_i|^2 1_{\{\overline{\zeta}^n_i > \varepsilon\}}\end{aligned} \tag{7.2.5}$$

almost surely. In order to get the estimate for all $\omega \in \Omega^n$, we can choose the versions of conditional expectations as follows: first, we may without loss of generality choose a version of $E^n_{i-1}|\overline{\zeta}^n_i|^2 1_{\{\overline{\zeta}^n_i > \varepsilon\}}$ which is non-negative identically; next, on the union of all exceptional sets for the estimates appeared above, we define the values of all other conditional expectations as zero. Then, the inequality (7.2.5) holds *identically* for *all* $\psi \in \Psi$. By taking the supremum of (7.2.5) with respect to $\psi \in \Psi$, and letting $n \to \infty$, we obtain the assertion (7.2.2).

A similar argument yields (7.2.4). In fact, it is much easier than (7.2.2), because the assertion of (7.2.4) is ψ-wise, for which we do not need any argument about versions of conditional expectations. Also, it is easy to show (7.2.3) if we notice the following fact: for given $a \in (0,1)$ there exists a constant $K > 0$ such that $|\log x \cdot \log y - 4(\sqrt{x}-1)(\sqrt{y}-1)| \leq K \max\{|\sqrt{x}-1|^3, |\sqrt{y}-1|^3\}$ whenever $\max\{|\sqrt{x}-1|, |\sqrt{y}-1|\} \leq a$.

[STEP 2] Next we prove the following facts:

$$\sup_{\psi\in\Psi}|B^{n,a}(\psi) - B^{n,\varepsilon}(\psi)| \xrightarrow{P^{n*}} 0 \qquad \forall \varepsilon \in (0,a); \tag{7.2.6}$$

$$|C^{n,a}(\psi,\phi) - C^{n,\varepsilon}(\psi,\phi)| \xrightarrow{P^n} 0 \qquad \forall \psi,\phi \in \Psi\ \forall \varepsilon \in (0,a). \tag{7.2.7}$$

In order to show (7.2.6), notice that for given $a \in (0,1)$ there exists a constant $K > 0$ such that $|\log x| \leq K|\sqrt{x}-1|^2$ whenever $|\sqrt{x}-1| \leq a$. For every $\varepsilon \in (0,a)$ it holds that

$$\left|\sum_{i=1}^{\sigma^n} E^n_{i-1}\lambda^{n,a,\psi}_i - \sum_{i=1}^{\sigma^n} E^n_{i-1}\lambda^{n,\varepsilon,\psi}_i\right| \tag{7.2.8}$$

$$
\begin{aligned}
&= \left| \sum_{i=1}^{\sigma^n} E_{i-1}^n \log \tilde{Z}_i^{n,\psi} 1_{\{\varepsilon < \overline{\zeta}_i^n \le a\}} \right| \\
&\le K \sum_{i=1}^{\sigma^n} E_{i-1}^n |\zeta_i^{n,\psi}|^2 1_{\{\varepsilon < |\zeta_i^n| \le a\}} \\
&\le K \sum_{i=1}^{\sigma^n} E_{i-1}^n |\overline{\zeta}_i^n|^2 1_{\{\overline{\zeta}_i^n > \varepsilon\}},
\end{aligned}
$$

almost surely. We can choose some versions of conditional expectations such that the estimate above holds identically for all $\psi \in \Psi$, in the same way as in the proof of (7.2.2). Take the supremum of (7.2.8) with respect to $\psi \in \Psi$, and let $n \to \infty$, then we get (7.2.6). A similar computation yields (7.2.7).

[STEP 3] Now it is easy to see that (7.2.2) and (7.2.6) imply the assertion (i), and that (7.2.3) and (7.2.7) do the assertion (ii); first choose $\varepsilon > 0$ small enough, and then, let $n \to \infty$. In order to show the assertion (iii), notice that for any $\varepsilon \in (0, a)$

$$
\begin{aligned}
&\left| \sum_{i=1}^{\sigma^n} |E_{i-1}^n \lambda_i^{n,a,\psi}|^2 - \sum_{i=1}^{\sigma^n} |E_{i-1}^n \lambda_i^{n,\varepsilon,\psi}|^2 \right| \\
&= \left| \sum_{i=1}^{\sigma^n} E_{i-1}^n (\lambda_i^{n,a,\psi} + \lambda_i^{n,\varepsilon,\psi}) E_{i-1}^n (\lambda_i^{n,a,\psi} - \lambda_i^{n,\varepsilon,\psi}) \right| \\
&\le 2|\log(1-a^2)| \sum_{i=1}^{\sigma^n} \left| E_{i-1}^n \log \tilde{Z}_i^{n,\psi} 1_{\{\varepsilon < \overline{\zeta}_i^n \le a\}} \right| = o_{P^n}(1),
\end{aligned}
$$

hence

$$
\begin{aligned}
\sum_{i=1}^{\sigma^n} |E_{i-1}^n \lambda_i^{n,a,\psi}|^2 &= \sum_{i=1}^{\sigma^n} |E_{i-1}^n \lambda_i^{n,\varepsilon,\psi}|^2 + o_{P^n}(1) \\
&\le |\log(1-\varepsilon^2)| \sum_{i=1}^{\sigma^n} |E_{i-1}^n \lambda_i^{n,\varepsilon,\psi}| + o_{P^n}(1).
\end{aligned}
$$

We therefore obtain (iii) by virtue of (7.2.4); first choose $\varepsilon > 0$ small enough, and then let $n \to \infty$. □

7.3 Model Checking for a Non-linear Times Series

Let us consider the $\mathbb{R}$-valued time series $\{X_i\}_{i \in \mathbb{Z}}$ given by

$$X_i = \psi(X_{i-1}) + \varepsilon_i,$$

where ψ is an $\mathbb{R}$-valued function on $\mathbb{R}$ and ε_i is an $\mathbb{R}$-valued random variable such that

$$P(\varepsilon_i \le 0 | \mathcal{F}_{i-1}) = \frac{1}{2} \qquad \text{almost surely,}$$

where $\mathcal{F}_i = \sigma\{X_j : j \leq i\}$.

Let $K : \mathbb{R} \to [0, \infty)$ be a kernel function with a compact support, and let $\{b_n\}_{n\in\mathbb{N}}$ be a sequence of positive constants such that $b_n \downarrow 0$ as $n \to \infty$. We introduce the stochastic process $V^n = (V^n(x)|x \in \mathbb{R})$ given by

$$V^n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i^n(x) Z_i,$$

where

$$Y_i^n(x) = \frac{1}{b_n} \int_{-\infty}^{x} K\left(\frac{X_{i-1} - u}{b_n}\right) du$$

and

$$Z_i = \mathrm{sign}\left(X_i - \psi(X_{i-1})\right),$$

with

$$\mathrm{sign}(x) = \begin{cases} -1 & \text{for } x \leq 0, \\ 1 & \text{for } x > 0. \end{cases}$$

The process V^n is a "smoothed" version of that of Erlenmaier (1997) who considered a process V^n with $Y_i^n(x)$ replaced by $1_{(-\infty,x]}(X_{i-1})$.

In order to state some conditions which we shall assume, we denote by $\widehat{F}_n$ and $\widehat{f}_n$ the empirical distribution function and the empirical density function associated to the kernel function K and the bandwidth b_n, of the data $X_0, ..., X_{n-1}$: that is,

$$\begin{aligned} \widehat{F}_n(x) &= \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,x]}(X_{i-1}); \\ \widehat{f}_n(x) &= \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{X_{i-1} - x}{b_n}\right). \end{aligned}$$

We are ready to present our result.

Theorem 7.3.1 *Assume the following conditions* (i) *and* (ii)*:*
(i) $\widehat{F}_n(x) \xrightarrow{P} F(x)$ *for every* $x \in \mathbb{R}$, *where* F *is a continuous distribution function on* $\mathbb{R}$*;*
(ii) *there exists a function* $g \in \mathcal{L}^p$ *such that* $\|\widehat{f}_n/g\|_{\mathcal{L}^q(\mathbb{R})} = O_P(1)$ *for some* $p, q \in (1, \infty)$ *such that* $(1/p) + (1/q) = 1$.
Then, it holds that $V^n \xRightarrow{P} V$ *in* $\ell^\infty(\mathbb{R})$ *where* $V(x) = B_{F(x)}$ *and* $t \rightsquigarrow B_t$ *is a standard Brownian motion on* $[0, 1]$.

Notice that, when the time series takes values only in a bounded subset of $\mathbb{R}$, a sufficient condition for (ii) is that $\|\widehat{f}_n\|_{\mathcal{L}^q(\mathbb{R})} = O_P(1)$ for an arbitrary $q > 1$. The result above, together with the continuous mapping theorem, yields the following.

Corollary 7.3.2 *Define* $S^n = \sup_{x\in\mathbb{R}} |V^n(x)|$. *Under the conditions* (i) *and* (ii) *of Theorem 7.3.1, it holds that* $S^n \stackrel{P}{\Longrightarrow} \sup_{t\in[0,1]} |B_t|$ *in* $\mathbb{R}$, *where* $t \rightsquigarrow B_t$ *is a standard Brownian motion on* $[0,1]$.

We have thus obtained the asymptotically distribution-free test statistics S^n. Notice that the time series $\{X_i\}_{i\in\mathbb{Z}}$ need not be Markovian (the noise ε_i may depend on the whole past).

Proof of Theorem 7.3.1. We can write $V^n(x) = \sum_{i=1}^n \xi_i^{n,x}$, where

$$\xi_i^{n,x} = \frac{1}{\sqrt{n}} Y_i^n(x) Z_i.$$

It is clear that $\{\xi_i^n\}_{i\in\mathbb{N}} = \{(\xi_i^{n,x}|x \in \mathbb{R})\}_{i\in\mathbb{N}}$ is an $\ell^\infty(\mathbb{R})$-valued martingale difference array on the discrete-time stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_i\}_{i\in\mathbb{N}_0}, P)$, where $\mathcal{F}_i = \sigma\{X_j : j \le i\}$. We will check the conditions of Theorem 3.3.2.

To check **[PE′]**, for every $\varepsilon > 0$, choose some finite points $\{x_{\varepsilon,k} : 1 \le k \le N(\varepsilon) - 1\}$ of $\mathbb{R}$ such that

$$x_{\varepsilon,0} = -\infty < x_{\varepsilon,1} < x_{\varepsilon,2} < \cdots < x_{\varepsilon,N(\varepsilon)-1} < \infty = x_{\varepsilon,N(\varepsilon)}$$

and that

$$\int_{x_{\varepsilon,k-1}}^{x_{\varepsilon,k}} |g(x)|^p dx \le \varepsilon^{2p} \qquad \forall k = 1, ..., N(\varepsilon).$$

This can be done with $N(\varepsilon) \le \text{const.}\varepsilon^{-2p}$, thus it holds that $\int_0^1 \sqrt{\log N(\varepsilon)} d\varepsilon < \infty$. On the other hand, it holds that

$$\begin{aligned} |\xi_i^{n,x} - \xi_i^{n,y}| &\le \frac{1}{\sqrt{n}b_n} \int_{x\wedge y}^{x\vee y} K\left(\frac{X_{i-1} - u}{b_n}\right) du \\ &\le \frac{1}{\sqrt{n}b_n} \int_{x_{\varepsilon,k-1}}^{x_{\varepsilon,k}} K\left(\frac{X_{i-1} - u}{b_n}\right) du \qquad \text{if } x, y \in (x_{\varepsilon,k-1}, x_{\varepsilon,k}]. \end{aligned}$$

So we have

$$\begin{aligned} &\sum_{i=1}^n E_{i-1}\left[\sup_{x,y\in(x_{\varepsilon,k-1},x_{\varepsilon,k}]} |\xi_i^{n,x} - \xi_i^{n,y}|\right]^2_{\mathcal{F}_i} \\ &\le \frac{1}{n}\sum_{i=1}^n \left|\frac{1}{b_n}\int_{x_{\varepsilon,k-1}}^{x_{\varepsilon,k}} K\left(\frac{X_{i-1} - u}{b_n}\right) du\right|^2 \\ &\le \frac{1}{n}\sum_{i=1}^n \frac{1}{b_n}\int_{x_{\varepsilon,k-1}}^{x_{\varepsilon,k}} K\left(\frac{X_{i-1} - u}{b_n}\right) du \\ &= \int_{\mathbb{R}} 1_{(x_{\varepsilon,k-1},x_{\varepsilon,k}]}(u)\widehat{f}_n(u)du \\ &= \left(\int_{\mathbb{R}} |1_{(x_{\varepsilon,k-1},x_{\varepsilon,k}]}(u)g(u)|^p du\right)^{1/p} \cdot \left(\int_{\mathbb{R}} |\widehat{f}_n(u)/g(u)|^q du\right)^{1/q} \\ &\le \varepsilon^2 \cdot \|\widehat{f}_n/g\|^q. \end{aligned}$$

Hence the condition (ii) implies **[PE′]**.

To check **[C2′]**, first observe that

$$\int_{-\infty}^{x} \widehat{f}_n(u)du \xrightarrow{P} F(x) \qquad \forall x \in \mathbb{R}. \tag{7.3.1}$$

This fact is proved as follows. Since K has a compact support and since $b_n \downarrow 0$, it holds that for any $\varepsilon > 0$

$$\widehat{F}_n(x-\varepsilon) \leq \frac{1}{nb_n}\sum_{i=1}^{n}\int_{-\infty}^{x} K\left(\frac{X_{i-1}-u}{b_n}\right)du \leq \widehat{F}_n(x+\varepsilon)$$

for all sufficiently large $n \in \mathbb{N}$. Due to (i), the left and the right hand side converge in probability to $F(x-\varepsilon)$ and $F(x+\varepsilon)$, respectively. The claim (7.3.1) follows from the assumption that $x \rightsquigarrow F(x)$ is continuous.

Let us now turn to the convergence of $\sum_{i=1}^{n} E_{i-1}\xi_i^{n,x}\xi_i^{n,y}$'s. In case of $x \neq y$, we have that

$$\begin{aligned}\sum_{i=1}^{n} E_{i-1}\xi_i^{n,x}\xi_i^{n,y} &= \frac{1}{n}\sum_{i=1}^{n} Y_i^n(x)Y_i^n(y) \\ &= \frac{1}{n}\sum_{i=1}^{n} Y_i^n(x\wedge y) \qquad \text{for all sufficiently large } n \in \mathbb{N} \\ &= \int_{-\infty}^{x\wedge y} \widehat{f}_n(u)du,\end{aligned}$$

which converges in probability to $F(x\wedge y)$ due to (7.3.1). In case of $x = y$, observe that

$$\sum_{i=1}^{n} E_{i-1}\xi_i^{n,x-\varepsilon}\xi_i^{n,x} \leq \sum_{i=1}^{n} E_{i-1}|\xi_i^{n,x}|^2 \leq \sum_{i=1}^{n} E_{i-1}\xi_i^{n,x}\xi_i^{n,x+\varepsilon} \qquad \forall \varepsilon > 0,$$

and that the left and the right hand side converge in probability to $F(x-\varepsilon)$ and $F(x)$, respectively. Thus it follows from the continuity of $x \rightsquigarrow F(x)$ that the middle of the above inequalities converges in probability to $F(x)$. We therefore have shown that

$$\sum_{i=1}^{n} E_{i-1}\xi_i^{n,x}\xi_i^{n,y} \xrightarrow{P} F(x\wedge y) \qquad \forall x, y \in \mathbb{R}.$$

It is trivial that the Lindeberg condition **[L2′]** is satisfied, and all conditions of Theorem 3.3.2 have been established. □

7.A Notes

The bandwidth processes or/and deviation processes of kernel density estimators as random elements taking values in the space C were studied by Krieger and Pickands, III

(1981), Müller and Prewitt (1992, 1993), and Müller and Wang (1990). Theorem 7.1.4 could be obtained from a general study of "local empirical processes" by Einmahl and Mason (1997) combined with a uniform Donsker theorem by Sheehy and Wellner (1992). Although the notion of "local empirical process" is more general than the local kernel estimators, their approach is essentially based on the i.i.d. setup.

The asymptotic behavior of the log-likelihood ratio random fields in finite-dimensional parametric models has been studied by many authors including Le Cam (1970), Inagaki and Ogata (1975), Ogata and Inagaki (1977), Ibragimov and Has'minskii (1981), Kutoyants (1984) and Vostrikova (1987). Although no result for infinite-dimensional cases seems to have been presented in the literature so far, some results in i.i.d. cases are immediate from the Donsker theorems for empirical processes. Theorem 7.2.1 seems the first to consider the general statistical experiment with abstract parameters.

The problem considered in Section 7.2 and the basic idea of the test statistics were posed by Erlenmaier (1997), who obtained the same conclusion as Theorem 7.3.1 for a slightly different statistics in a Markovian case under an explicit assumption on the transition kernel.

Bibliography

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

[2] Andersen, N.T., Giné, E., Ossiander, M. and Zinn, J. (1988). The central limit theorem and the law of iterated logarithm for empirical processes under local conditions. *Probab. Theory Relat. Fields* **77** 271-305.

[3] Arcones, M.A. and Yu, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *J. Theoret. Probab.* **7** 47-71.

[4] Bae, J. (1995). An empirical CLT for stationary martingale differences. *J. Korean Math. Soc.* **32** 427-446.

[5] Bae, J. and Levental, S. (1995a). Uniform CLT for Markov chains and its invariance principle: a martingale approach. *J. Theoret. Probab.* **8** 549-570.

[6] Bae, J. and Levental, S. (1995b). A uniform CLT for continuous martingales. *J. Korean Statist. Soc.* **24** 225-231.

[7] Birgé, L. and Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Relat. Fields* **97** 113-150.

[8] Chernoff, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31-41.

[9] Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. Henri Poincaré Probab. Statist.* **31** 393-427.

[10] Dudley, R.M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899-929. Correction, **7** 909-911.

[11] Einmahl, U. and Mason, D.M. (1997). Gaussian approximation of local empirical processes indexed by functions. *Probab. Theory Relat. Fields* **107** 283-311.

[12] Erlenmaier, U. (1997). A new criterion for tightness of stochastic processes and an application to Markov processes. *Manuscript.*

[13] Freedman, D.A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100-118.

[14] Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations. *Statistics* **21** 99-116.

[15] Gill, R.D. (1980). *Consoring and Stochastic Integrals.* Mathematical Centre Tracts 124. Mathematisch Centrum, Amsterdam.

[16] Hida, T. (1980). *Brownian Motion.* Springer-Verlag, New York.

[17] Ibragimov, I.A. and Has'minskii, R.Z. (1981). *Statistical Estimation.* Springer-Verlag, New York.

[18] Inagaki, N. and Ogata, Y. (1975). The weak convergence of likelihood ratio random fields and its applications. *Ann. Inst. Statist. Math.* **27** 391-419.

[19] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin Heidelberg.

[20] Jain, N. and Marcus, M. (1975). Central limit theorems for $C(S)$-valued random variables. *J. Funct. Anal.* **19** 216-231.

[21] Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191-219.

[22] Korostelev, A. (1987). On minimax estimation of discontinuous signal. *Theory Probab. Appl.* **32** 727-730.

[23] Krieger, A.M. and Pickands, J., III (1981). Weak convergence and efficient density estimation at a point. *Ann. Statist.* **9** 1066-1078.

[24] Kutoyants, Yu. (1984). *Parameter Estimation for Stochastic Processes.* Heldermann Verlag, Berlin.

[25] Kutoyants, Yu. (1994). *Identification of Dynamical Systems with Small Noise.* Kluwer Academic Publishers, Dordrecht.

[26] Lánska, V. (1979). Minimum contrast estimation in diffusion processes. *J. Appl. Probab.* **16** 65-75.

[27] Le Cam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Statist.* **41** 802-828.

[28] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces.* Springer-Verlag, Berlin Heidelberg.

[29] Levental, S. (1989). A uniform CLT for uniformly bounded families of martingale differences. *J. Theoret. Probab.* **2** 271-287.

[30] Liptser, R.S. and Shiryayev A.N. (1989). *Theory of Martingales.* Kluwer Academic Publishers, Dordrecht.

[31] McKeague, I.W. and Utikal, K.J. (1990). Identifying nonlinear covariate effects in semimartingale regression models. *Probab. Theory Relat. Fields* **87** 1-25.

[32] Müller, H.G. and Prewitt, K.A. (1992). Weak convergence and adaptive peak estimation for spectral densities. *Ann. Statist.* **20** 1329-1349.

[33] Müller, H.G. and Prewitt, K.A. (1993). Multiparameter bandwidth processes and adaptive surface smoothing. *J. Multivariate Anal.* **47** 1-21.

[34] Müller, H.G. and Wang, J.L. (1990). Locally adaptive hazard smoothing. *Probab. Theory Relat. Fields* **85** 523-538.

[35] Nishiyama, Y. (1997). Some central limit theorems for ℓ^∞-valued semimartingales and their applications. *Probab. Theory Relat. Fields* **108** 459-494.

[36] Ogata, Y. and Inagaki, N. (1977). The weak convergence of the likelihood ratio random fields for Markov observations. *Ann. Inst. Statist. Math.* **29** 165-187.

[37] Ossiander, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15** 897-919.

[38] Pollard, D. (1982). A central limit theorem for empirical processes. *J. Austral. Math. Soc. Ser. A* **33** 235-248.

[39] Pollard, D. (1989). Asymptotics via empirical processes (with discussion). *Statistical Science* **4** 341-366.

[40] Pollard, D. (1990). *Empirical Processes: Theory and Applications.* Inst. Math. Statist., Hayward.

[41] Rogers, L.C.G. and Williams, D. (1987). *Diffusions, Markov Processes, and Martingales. Volume 2: Itô Calculus.* John Wiley & Sons, Chichester New York Brisbane Tronto Singapore.

[42] Sheehy, A. and Wellner, J.A. (1992). Uniform Donsker classes of functions. *Ann. Probab.* **20** 1983-2030.

[43] van de Geer, S. (1990). Estimating a regression function. *Ann. Statist.* **18** 907-924.

[44] van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21** 14-44.

[45] van de Geer, S. (1995a). The method of sieves and minimum contrast estimators. *Math. Methods Statist.* **1** 20-38.

[46] van de Geer, S. (1995b). Exponential inequalities for martingales, with application to maximum likelihood likelihood estimation for counting processes. *Ann. Statist.* **23** 1779-1801.

[47] van de Geer, S. (1997). *Empirical Processes and the Quality of Estimators in Complex Parameter Space. Manuscript for forthcoming Lecture Notes.*

[48] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag, New York.

[49] Vostrikova, L. (1987). On the weak convergence of likelihood ratio processes of general statistical parametric models. *Stochastics* **23** 277-298.

[50] Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82** 385-397.

[51] Wong, W.H. and Severini, T.A. (1991). On maximum likelihood estimation in infinite dimensional parameter spaces. *Ann. Statist.* **19** 603-632.

[52] Wong, W.H. and Shen, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* **23** 339-362.

[53] Yoshida, N. (1990). Asymptotic behavior of M-estimator and related random field for diffusion process. *Ann. Inst. Statist. Math.* **42** 221-251.

[54] Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.* **41** 220-242.

Samenvatting/Summary

Samenvatting

Het doel van deze studie is entropie methoden te ontwikkelen, welke voor het eerst werden geïntroduceerd voor empirische processen van onderling onafhankelijke, identiek verdeelde (Engels: i.i.d.) data, ten einde bepaalde martingalen te behandelen met toepassingen bij de statistische analyse van stochastische processen. Gebaseerd op maximale ongelijkheden, afgeleid in Hoofdstuk 2, vormt Hoofdstuk 3 een hoogtepunt waarin zwakke convergentie resultaten voor l^∞-waardige martingalen worden gegeven. De resterende hoofdstukken zijn gewijd aan statistische toepassingen van deze resultaten; wij houden ons bezig met twee hoofd thema's, namelijk *asymptotische normaliteit en efficiëntie in* l^∞*-ruimtes* (Hoofdstuk 4) en *M-schatten* (Hoofdstuk 5 en 6), terwijl Hoofdstuk 7 drie onafhankelijke onderwerpen bevat.

De motivatie is als volgt. Sinds het prominente werk van Dudley in 1978, werden in de tachtiger jaren de entropie methoden bestudeerd om het bewijs te leveren voor wetten van grote aantallen en centrale limiet stellingen voor empirische processen geïndiceerd door klassen van verzamelingen of functies. Voorts, hebben enkele recente werken aangetoond dat de methoden niet alleen bruikbaar zijn voor deze limiet stellingen, maar ook voor andere problemen in de statistiek. Het boek van Van der Vaart en Wellner uit 1996 geeft een fraaie uiteenzetting van zowel de methoden als een groot aantal toepassingen, met de nadruk op i.i.d. data. Echter, hoewel bepaalde aspecten van de methoden een groot potentieel hebben om ook toegepast te worden op niet-i.i.d. data, is er geen systematisch onderzoek gedaan in verband met martingalen, waarvan het bekend is dat ze belangrijk zijn by de analyse van een rijke klasse statistische modellen. We zijn van plan om een stap te doen, zij het een kleine, om deze leemte in de literatuur op te vullen.

Hoofdstuk 2 en 3 gaan over drie soorten abstracte martingalen, te weten: (i) stochastische integralen ten opzichte van multivariate punt processen; (ii) martingalen in discrete tijd; (iii) continue martingalen. In Hoofdstuk 2 introduceren wij eerst bepaalde grootheden geheten kwadratische modulus en exponentiele modulus, welke een belangrijke rol spelen in ons onderzoek. In termen van deze grootheden en van entropie getallen, verkrijgen we enige moment-ongelijkheden voor de martingalen. Onze aanpak in gevallen (i) en (ii) is in essentie hetzelfde als de L^2-'bracketing' entropie, terwijl geval (iii) gebaseerd is op metrische entropie zonder 'bracketing'.

In Hoofdstuk 3 bewijzen we zwakke convergentie stellingen in l^∞-ruimtes door gebruik te maken van de ongelijkheden uit Hoofdstuk 2. Onder deze stellingen bevinden zich natuurlijke generalisaties van Jain-Marcus' en Ossiander's centrale limiet stellingen voor het geval van martingalen in discrete tijd. Deze resultaten worden herhaaldelijk gebruikt in de resterende hoofdstukken.

Hoofdstuk 4 is gewijd aan het eerste hoofd thema dat we eerder noemden. In sectie 4.1, beschouwen we het multiplicatieve intensiteit model voor punt processen met algemene labels, en introduceren we een veralgemeniseerde Nelson-Aalen schatter die een empirisch proces is, geïndiceerd door klassen van functies op de tijd-label ruimte. We leiden de asymptotische normaliteit van de schatter af, onder een L^p voorwaarde, $p \in [2, \infty]$, waarbij we een resultaat toepassen uit Hoofdstuk 3. Een verschil met het geval van i.i.d. empirische processen is dat de L^2-'bracketing' conditie niet altijd optimaal is in deze context. We tonen ook de asymptotische efficiëntie aan in de zin van de convolutie- en de lokaal asymptotische minimax stellingen. In de secties 4.2 en 4.3, geven we een aanpak van niet-lineaire modellen volgens een idee van McKeage en Utikal uit 1990. We beschouwen enkele statistische modellen van continue semi-martingalen en tel-processen met tijdsafhankelijke covariaten die waarden aannemen een algemene toestandsruimtes, en leiden de asymptotische normaliteit en efficiëntie af van enkele schatters van het integraal-type. Als we ons specialiseren tot het geval van Euclidische covariaten, verzwakt ons resultaat een aanname van McKeage en Utikal.

Sectie 5.1 bevat een algemeen resultaat voor het verkrijgen van de convergentie snelheid van M-schatters, die waarden aannemen in een abstracte ruimte, dat gebruikt wordt door heel Hoofdstuk 5 en 6. De overgebleven secties van Hoofdstuk 5 gaan over M-schatten van Euclidische parameters met niet-standaard convergentie snelheid. We beschouwen het Gaussische witte ruis model en het multiplicatieve intensiteit model in secties 5.2 en 5.3, respectievelijk, en leiden het asymptotisch gedrag af van sommige schatters. We geven, onder andere, een asymptotisch verdelings resultaat met wortel-n snelheid voor een sprong-punt schatter in het eerstgenoemde model. Zowel de maximale ongelijkheden van Hoofdstuk 2, als de resultaten van Hoofdstuk 3, zijn hierbij nodig.

Hoofdstuk 6 gaat over de convergentie snelheid van niet-parametrische meest aannemelijke schatters. We beschouwen het Gaussische witte ruis model, het multiplicatieve intensiteit model, tel-processen met niet-lineaire covariaten en processen van het diffusie type, afgeleid van stochastische differentiaal vergelijkingen. Een discussie over het Lexis diagram is in de beschouwing bevat als een toepassing van het derde genoemde model.

In sectie 7.1 breiden we de puntsgewijze asymptotische normaliteit van kern dichtheidsschatters uit in de lokaal functionale zin. De lokalisatie constante dient gelijk gekozen te worden aan de bandbreedte. Bij wijze van toepassing, bespreken we het schatten van de modus van onbekende dichtheidsfuncties. Sectie 7.2 is gewijd aan het afleiden van het asymptotisch gedrag van 'log-likelihood ratio random fields' in een algemeen statistisch experiment met abstracte parameters. Een toepassing voor een ergodische Markov keten wordt ook gepresenteerd. In sectie 7.3 beschouwen we een niet-parametrische test voor het beoordelen van een niet-lineaire tijdreeks. Een asymptotisch verdelings-vrije test wordt verkregen.

Summary

The purpose of this study is to develop entropy methods, which were first introduced for empirical processes of i.i.d. data, in order to handle some martingales with applications to statistical inference for stochastic processes. Based on maximal inequalities derived in Chapter 2, a highlight is Chapter 3 that gives weak convergence theorems for ℓ^∞-valued martingales. The remaining chapters are devoted to statistical applications of those results; we are concerned with two main themes, namely, *asymptotic normality and efficiency in ℓ^∞-spaces* (Chapter 4) and *M-estimation* (Chapters 5 and 6), while Chapter 7 contains three independent topics.

The motivation is as follows. Since the prominent work of Dudley in 1978, the entropy methods were studied to establish laws of large numbers and central limit theorems for empirical processes indexed by classes of sets or functions in the 80's. Furthermore, some recent works have shown that the methods are useful not only for those limit theorems but also for other problems in statistics. The book by van der Vaart and Wellner in 1996 gives a nice exposition of the methods as well as a lot of applications, with emphasis on i.i.d. data. However, although some parts of the methods have a good potential to be applied also for non-i.i.d. data, no systematic study has been done in the framework of martingales, which are known to be important for analyzing a rich class of statistical models. We intend to make a step, which is still small though, to fill this gap in the literature.

Chapters 2 and 3 deal with three kinds of abstract martingales, that is: (i) stochastic integrals with respect to multivariate point processes; (ii) discrete time martingales; (iii) continuous martingales. In Chapter 2, we first introduce some quantities called quadratic modulus and exponential modulus, which play a key role in our study. In terms of the quantities and entropy numbers, we obtain some moment inequalities for the martingales.

Our approach to the cases (i) and (ii) is essentially the same as the L^2-bracketing entropy, while the case (iii) is based on the metric entropy without bracketing.

In Chapter 3, we establish weak convergence theorems in ℓ^∞-spaces by using the inequalities given in Chapter 2. Among them, natural generalizations of Jain-Marcus' and Ossiander's central limit theorems to discrete time martingales are presented. The results obtained there are repeatedly applied in the remaining chapters.

Chapter 4 is devoted to the first main theme mentioned above. In Section 4.1, we consider the multiplicative intensity model of point processes with general marks, and introduce a generalized Nelson-Aalen estimator, which is an empirical process indexed by classes of functions on the time-mark space. We derive the asymptotic normality of the estimator, under an L^p-bracketing condition with $p \in [2, \infty]$, applying a result given in Chapter 3. A difference from the case of i.i.d. empirical processes is that the L^2-bracketing condition is not always optimal in this context. We also show its asymptotic efficiency in the sense of the convolution and the locally asymptotic minimax theorems. In Sections 4.2 and 4.3, we present an approach to non-linear models along the idea of McKeague and Utikal in 1990. We consider some statistical models of continuous semimartingales and counting processes with time-dependent covariates taking values in general state spaces, and derive the asymptotic normality and efficiency of some integral-type estimators. Specialized to the case of Euclidean covariates, our result weakens an assumption of McKeague and Utikal.

Section 5.1 contains a general result for obtaining the rate of convergence of M-estimators taking values in an abstract space, and it is used throughout Chapters 5 and 6. The remaining sections of Chapter 5 deal with M-estimation of Euclidean parameters with non-standard rate of convergence. We consider the Gaussian white noise model and the multiplicative intensity model in Sections 5.2 and 5.3, respectively, and derive the asymptotic behavior of some estimators. We present an asymptotic distribution result with rate n for a jump point estimator in the former model, among other things. The maximal inequalities in Chapter 2, as well as the results in Chapter 3, are needed for the discussion there.

Chapter 6 is concerned with the rate of convergence of non-parametric maximum likelihood estimators. We consider the Gaussian white noise model, the multiplicative intensity model, counting processes with non-linear covariates, and diffusion-type processes derived from stochastic differential equations. A discussion about the Lexis diagram is contained as an application of the third model above.

In Section 7.1, we extend the point-wise asymptotic normality of kernel density estimators to the local functional sense. The localizing constant should be chosen to be

the same as the bandwidth. As an application, we discuss the estimation of the mode of unknown density functions. Section 7.2 is devoted to deriving the asymptotic behavior of log-likelihood ratio random fields in a general statistical experiment with abstract parameters. An application to an ergodic Markov chain is also presented. In Section 7.3, we consider a non-parametric test for checking a non-linear time series. An asymptotically distribution-free test is obtained.

Curriculum Vitae

- 14 January 1969: Yoichi Nishiyama was born in Himeji, Japan.
- April 1984 – March 1987: Student at Tatsuno High School.
- April 1987 – March 1991: Undergraduate course at the Department of Mathematics, Osaka University (B.A. degree in March 1991).
- April 1991 – March 1993: Master's course at the Department of Mathematical Science, Osaka University (Master's degree in March 1993).
- April 1993 – March 1994: Doctor's course at the Department of Mathematical Science, Osaka University.
- April 1994 – the present: Assistant Professor at the Institute of Statistical Mathematics, Tokyo.
- June 1996 – (May 1998): JSPS Fellow for Research Abroad at the Mathematical Institute, Utrecht University (supported by the Japan Society for the Promotion of Science).