

## DETECTING ABRUPT CHANGES IN RANDOM FIELDS

ANTOINE CHAMBAZ<sup>1, 2</sup>

**Abstract.** This paper is devoted to the study of some asymptotic properties of a  $M$ -estimator in a framework of detection of abrupt changes in random field's distribution. This class of problems includes *e.g.* recovery of sets. It involves various techniques, including  $M$ -estimation method, concentration inequalities, maximal inequalities for dependent random variables and  $\phi$ -mixing. Penalization of the criterion function when the size of the true model is unknown is performed. All the results apply under mild, discussed assumptions. Simple examples are provided.

**Mathematics Subject Classification.** 60E15, 62C99, 62F12, 62G20, 62M40.

### 1. INTRODUCTION

The problem of abrupt changes detecting includes a wide range of subjects unified by a common basic framework: observation of a random process whose distribution is long-scale heterogeneous but short-scale homogeneous on some regions. Comprehensive presentations can be found in the three books [5–7].

The mathematical methods include  $M$ -estimation as in the present paper or [14, 15] and also nonparametric or Bayesian techniques, see *e.g.* [2, 13].

Handling estimation or test in a multiple changes case with an unknown number of changes is crucial and intricate. Akaike's and Schwarz's papers [1, 23] are most of the time invoked as milestones, as Yao's [29], who proved consistency of Schwarz's criterion based estimator in case of independent Gaussian observations. Penalization methods are widely used, for instance in context of the estimation of the order of a process (see [1]), of the order of a mixture (see [8]), or generally in a context of statistical learning theory (see for instance the lectures notes [16]). Barron *et al.* obtained in [4] some precise bounds in a framework of regression and density estimation. Penalization in view of estimating a number of change-points is widely used, for instance among the previous citations in [14, 15].

#### *Examples*

One of the simplest models for change-points can be summarized by the following model: one observes responses  $Y(X_i)$  at  $X_i = i$  with  $Y(X_i) = \vartheta^*(X_i) + \varepsilon(X_i)$  for centered possibly dependent  $\varepsilon(X_i)$  and some piecewise constant function  $\vartheta^*$ . Here,  $X_1, \dots, X_n$  should be understood as *regular times of observation*. A first natural extension consists of observing at random points  $X_i$  on a *d-dimensional lattice*. Then to allow observation throughout some *general d-dimensional space*  $\mathcal{X}$ . And finally to observe some process  $Y$  indexed by  $x \in \mathcal{X}$  at *randomly chosen points*  $X_i$  of  $\mathcal{X}$ .

---

*Keywords and phrases:* Detection of change-points,  $M$ -estimation, penalized  $M$ -estimation, concentration inequalities, maximal inequalities, mixing.

<sup>1</sup> UMR C 8628 du CNRS, Équipe de Probabilités, Statistique et Modélisation, Université Paris-Sud, France;  
e-mail: [Antoine.Chambaz@math.u-psud.fr](mailto:Antoine.Chambaz@math.u-psud.fr)

<sup>2</sup> FTR&D, 38 rue du Général Leclerc, 92130 Issy-les-Moulineaux, France.

Recovery of sets obviously enters in this framework, too: one observes an image  $\mathcal{X}$  composed of an object  $\tau_0^*$  and a background through noisy observations  $(X_i, Y_i)$  ( $i = 1, \dots, n$ ), with independent  $X_i \in \mathcal{X}$  and responses  $Y_i = f(X_i)\mathbb{1}\{X_i \in \tau_0^*\} + \xi_i$ , for some function  $f$  bounded away from 0 and random centered noise  $\xi_i$ . Here  $(X_1, \dots, X_n)$  are supposed independent of the mutually independent  $n$ -tuple  $(\xi_1, \dots, \xi_n)$ . The aim is to estimate  $\tau_0^*$ , or equivalently the partition  $\tau^* = (\tau_0^*, \mathcal{X} - \tau_0^*)$ .

#### *Aim of this paper*

We address in this paper the estimation of a partition  $\tau^*$  of  $\mathcal{X}$  from possibly dependent random observations  $Y_i$  at independent and identically  $P$ -distributed points  $X_i$  in  $\mathcal{X}$ . The proofs are based on Lavielle's paper [12]. The model actually consists of a couple  $(\tau^*, \theta^*)$ :  $\tau^* = (\tau_j^*)_{1 \leq j \leq K^*}$  is a partition with  $K^*$  subsets, where  $K^*$  (the *cardinality* of  $\tau^*$ ) is possibly unknown, and  $\theta^*$  is a collection of  $K^*$  finite-dimensional parameters  $\theta_j^*$ . We define for convenience  $\vartheta^* = \sum_{j=1}^{K^*} \theta_j^* \mathbb{1}\{\tau_j^*\}$ . We consider that changes affect the marginal distribution of  $Y_i$ 's: conditionally on  $X_i$ ,  $Y_i$  has a distribution which depends on  $\vartheta^*(X_i)$ . We assume that there exists an *ad hoc* contrast  $J_n$  associated to the problem.

Indeed, we estimate  $\vartheta^*$  by minimum contrast estimation and related techniques. Suppose first that we choose *a priori* the cardinality  $K$  of the estimator. By definition, the value of the contrast computed at the estimator  $\hat{\vartheta}_n = (\hat{\tau}_n, \hat{\theta}_n)$  bounds below the contrast  $J_n(\tau, \theta)$  computed at any model  $(\tau, \theta)$  of cardinality  $K$ .

#### *Involved techniques*

$J_n$  is naturally decomposed into the sum of a first term that depends only on  $X_1, \dots, X_n$  and a term of random centered fluctuations. Fluctuations take the form  $\Sigma_n(G) = \sum_{i=1}^n Z_i \mathbb{1}\{X_i \in G\}$  for  $Z_i = Y_i - \mathbb{E}(Y_i|X_i)$  and any  $G$  in a set  $\mathcal{G}$ . Section 4 is devoted to the control of those fluctuations *via* maximal inequalities.

A maximal inequality consists of an upper bound of the probability for  $\sup\{\|\Sigma_n(G)\|_\infty : G \in \mathcal{G}\}$  to be greater than some  $\delta > 0$ . In the simple case where partitions are constructed with elementary rectangles, one can derive easily such maximal inequalities from mild control of the second order moment of the fluctuations (see [19, 20]). This problem is more difficult in a general framework where partitions are constructed with elements of a larger class of sets (see [9]). In comparison with the previous simple case, control of moments of order any  $p > 2$  is needed here.

Denote  $P_n$  the empirical measure of  $(X_1, \dots, X_n)$ . Another theoretical complication arises from the need to derive lower bounds for  $(P(G) - P_n(G))/P(G)$  from bounds of  $P(G)$  for a large class of sets  $G$ . Actually, this is possible with large probability for sets  $G$  satisfying  $P(G) \geq r_n$  for some sequence  $\{r_n\} \downarrow 0$  carefully chosen. We cope with this difficulty thanks to concentration inequalities (refer to [18, 25]), see Section 3.1.

#### *Results for a priori known cardinality $K^*$ . Penalization*

We finally obtain under mild assumptions and for *a priori* known  $K^*$  that estimation is asymptotically consistent and we bound below rates of convergences. Quite surprisingly, but according to Lavielle's former results, the rate of convergence of the estimate  $\hat{\tau}_n$  of  $\tau^*$  does not seem to depend on the dependence structure of  $Y_i$ 's. It is strongly related to the rate  $\{r_n\}$  mentioned in the previous section.

One can generalize those results for known  $K^*$ . We can indeed construct an estimator  $\hat{\vartheta}_{n,K}$  of  $\vartheta^*$  for any *a priori* choice of the cardinality  $K$  of the estimator. The point is then to select the best estimator among them. This is roughly speaking the aim of the penalization method: replace the contrast  $J_n(\tau, \theta)$  by its penalized version  $J_n(\tau, \theta) + \beta_n K$ , with  $\beta_n > 0$ . The added term  $\beta_n K$  penalizes the models with large cardinality whereas those models are favoured when minimizing  $J_n(\tau, \theta)$  alone.

We prove that, for sequences  $\{\beta_n\} \downarrow 0$  slowly enough, penalized estimation yields a consistent estimated triplet  $(\hat{K}_n, \hat{\tau}_n, \hat{\theta}_n)$ . Naturally, the dependence structure of  $Y_i$ 's affects the maximum rate of convergence for  $\{\beta_n\}$ .

#### *Comparison with previous works*

We noticed earlier that the field of recovery of sets is part of the general problem of abrupt changes detection. Thus, we may wish to compare our results to classical ones in that field. Choose Mammem and Tsybakov's [17]

paper where the authors derive some optimal convergence rates. Recall the previous crude description of the recovery of sets problem. Here, the partition to estimate has cardinality 2, so the penalization procedure is not needed. The point is to estimate  $\tau_0^*$ . Roughly speaking, the authors prove that the risk for the maximum likelihood estimator (which is also a  $M$ -estimator) achieves the best possible rate of convergence in the minimax approach. Nevertheless, those results rely on independence of responses  $Y_i$ . On the contrary, our results apply in a framework of  $M$ -estimation of abrupt changes from dependent observations and are satisfying in this context, see again the former citations.

### Asymptotics

This paper is concerned with asymptotic results. In the whole text, the expression “as  $n, \delta \uparrow \infty$ ” will correspond to limits  $\lim_{\delta \rightarrow \infty} \lim_{n \rightarrow \infty}$ , and *idem* for “as  $n, \eta \uparrow \infty$ ”.

The practical interest of detecting abrupt changes in the general setting described above is certain, though our asymptotic results are mainly of theoretical value. They ensure confidence in a reasonable idealistic framework and encourage to find practical recipes to apply. Indeed, rigorous minimum contrast estimation is here computationally intractable and the penalization coefficient  $\beta_n$  would have to take a fixed value for real observed data. The choice of such a value would be justified by practical considerations as presented *e.g.* in [3, 11]. An automatic choice would require non asymptotic theory, see for example [4], but this is beyond the scope of this paper.

### Notation

In the whole paper, different positive constants might be denoted by the same letter  $C$ .

The organization of the paper is as follows: we introduce in Section 2 the partitions and the associated parameters to be studied and we define a pseudo-distance between them with useful properties. Section 3 is dedicated to the description of both the observations and the contrast to be minimized. Further assumptions are presented in Section 4. They deal with some crucial maximal inequalities. We consider estimation for known cardinality  $K^*$  of  $\tau^*$  in Section 5 and use those results to address the unknown cardinality case in Section 6. The Appendix 7 consists of three parts: the first one devoted to the postponed proof of a proposition; the second one to an exploration of the assumptions presented in Section 4; the third one to a sketch of proof of a technical lemma.

## 2. THE PARTITIONS AND THE ASSOCIATED PARAMETERS

### 2.1. Introducing partitions and associated parameters

Set a probability space  $(\Omega, \mathcal{A}, P)$  upon which random variables will be defined.

Consider some probability space  $(\mathcal{X}, \mathcal{G}, P)$  where  $P$  has support  $\mathcal{X}$ , *i.e.*  $\{x \in \mathcal{X} : \mathcal{O} \ni x \implies P(\mathcal{O}) > 0\} = \mathcal{X}$  ( $\mathcal{O}$  denotes an open set).  $\mathcal{X}$  is typically included in  $\mathbb{R}^d$ . We will define partitions of  $\mathcal{X}$  in the next paragraphs.

First, choose some set  $\mathcal{F}_0 \subset \mathcal{G}$  of measurable sets. Roughly speaking, a partition  $\tau$  of  $\mathcal{X}$  will be constructed as a collection  $(\tau_k)$  satisfying  $\cup_k \tau_k = \mathcal{X}$  and where any  $\tau_k$  is a finite union of elements of  $\mathcal{F}_0$ .

Then, define  $\mathcal{F}$  which contains all finite unions of elements of  $\mathcal{F}_0$  and pairwise intersection of such sets. Moreover, we suppose for sake of simplicity (that is to overcome measurability difficulties) that all the mathematical expressions in this paper involving suprema over subsets of  $\mathcal{F}$  are measurable (it suffices that for each of them, suprema are  $P$ -almost surely equal to suprema over some countable subsets).

Examples of  $\mathcal{F}_0$  when  $\mathcal{X} \subset \mathbb{R}^d$  include the set of all rectangles of the form  $\prod_{i=1}^d (a_i, b_i]$  (simply called *rectangles* in the whole paper); the set of all the polygons whose edges have lengths bounded below by some positive constant (*polygons* for short); or more generally (including rectangles and polygons), some Vapnik–Červonenkis class whose Vapnik–Červonenkis dimension is finite (for references, see *e.g.* [26–28]). In the sequel,  $VC$  will stand for Vapnik–Červonenkis.

Other assumptions will concern  $\mathcal{F}_0$  and  $\mathcal{F}$ : we will state them in Section 3.

**Definition 2.1.** We will consider  $\mathcal{F}$ -partitions (or shortly *partitions*) of  $\mathcal{X}$ . The set of all the partitions is denoted  $\mathcal{T}$ . Any  $\tau \in \mathcal{T}$ ,  $\tau = (\tau_k)_{1 \leq k \leq K}$ , is a collection of subsets of  $\mathcal{X}$ .  $K$  is called *cardinality* of  $\tau$ , also denoted  $\text{card}(\tau)$ . Any  $\tau_k$  can be written as an union  $\cup_l \tau_k(l)$  of non intersecting elements  $\tau_k(l)$  of  $\mathcal{F}_0$  whose  $P$ -probabilities must be bounded below by some fixed  $\Delta^* > 0$ .

$\mathcal{T}_K$  denotes the set of partitions with cardinality  $K$ .

**Remark 2.2.** Condition of minimal  $P$ -probability for the pieces  $\tau_k(l)$  of  $\tau_k = \cup_l \tau_k(l)$  stands for technical reasons. We will actually suppose that we know some lower bound of  $\Delta^*$ . Besides, this condition yields that there exists a finite maximal partition cardinality  $\overline{K}$  and that any  $\tau_k$  is a finite union of  $\tau_k(l)$ .

Some parameters are to be associated to a partition in the following way: a partition  $\tau \in \mathcal{T}$  with cardinality  $K$  may go with a collection  $\theta$  of  $K$   $\Theta$ -valued vectors. Here,  $\Theta$  is an open and precompact subset of  $\mathbb{R}^p$ . Thus, for  $\tau = (\tau_k)_{1 \leq k \leq K}$  and  $\theta = (\theta_k)_{1 \leq k \leq K}$ , the parameter  $\theta_k$  goes with  $\tau_k$ . We will denote  $\Theta_K = \Theta^K$ .

## 2.2. Pseudo-distances for partitions and parameters

To start with, let us recall some notations. For two sets  $A$  and  $B$ ,  $A \nabla B$  denotes their asymmetrical difference and  $A \Delta B$  their symmetrical difference, that is

$$A \nabla B = A \setminus A \cap B \quad \text{and} \quad A \Delta B = (A \nabla B) \cup (B \nabla A).$$

We wish to define a pseudo-distance between two  $\mathcal{F}$ -partitions of the set  $\mathcal{X}$  that generalizes the natural definition in the usual one-dimensional case, see [12]: for  $t$  and  $t^*$  two increasing vectors (respectively of length  $K$  and  $K^*$ ), the pseudo-distance is taken to be  $\max_{1 \leq j \leq K^*} \min_{1 \leq k \leq K} |t_k - t_j^*|$ . Thus, that pseudo-distance is the largest distance between points of  $t^*$  and their respective closest point in  $t$ . Observe that it is zero if and only if each point of  $t^*$  appears in  $t$ . These considerations lead to the following:

**Definition 2.3.** Let  $\tau$  and  $\tau^*$  be two  $\mathcal{F}$ -partitions of the set  $\mathcal{X}$ . Denote  $K$  and  $K^*$  their respective cardinality. The gap  $g(\tau, \tau^*)$  between them is defined as follows

$$g(\tau, \tau^*) = \max_{1 \leq j \leq K^*} \min_{\mathcal{K}} P \left( \left( \bigcup_{k \in \mathcal{K}} \tau_k \right) \Delta \tau_j^* \right).$$

The index  $\mathcal{K}$  in the infimum ranges over all subsets of  $\{1, \dots, K\}$ .

For  $j = 1, \dots, K^*$ , we denote  $\mathcal{K}_j$  a smallest subset of  $\{1, \dots, K\}$  achieving the minimum in the definition for fixed  $j$ . Consequently, we have

$$g(\tau, \tau^*) = \max_{1 \leq j \leq K^*} P \left( \left( \bigcup_{k \in \mathcal{K}_j} \tau_k \right) \Delta \tau_j^* \right).$$

Let us present a few interesting properties of the gap  $g$ .

**Proposition 2.4.** Consider two  $\mathcal{F}$ -partitions  $\tau^* = (\tau_j^*)_{1 \leq j \leq K^*}$  and  $\tau = (\tau_k)_{1 \leq k \leq K}$ .

- (i) Let  $j$  be in  $\{1, \dots, K^*\}$  and  $k$  in  $\{1, \dots, K\}$ . Observe that if  $\tau_k \subset \tau_j^*$ , then  $k \in \mathcal{K}_j$  whereas  $\tau_k \cap \tau_j^* = \emptyset$  implies  $k \notin \mathcal{K}_j$ . Not surprisingly, if  $k \notin \mathcal{K}_j$ , then  $P(\tau_k \nabla \tau_j^*) \geq P(\tau_k \cap \tau_j^*)$ . On the contrary, if  $k \in \mathcal{K}_j$  and  $\text{card}(\mathcal{K}_j) > 1$ , then  $P(\tau_k \nabla \tau_j^*) \leq P(\tau_k \cap \tau_j^*)$ . When  $\mathcal{K}_j = \{k\}$ , the former inequality holds as soon as  $g(\tau, \tau^*) \leq \Delta^*/2$ .
- (ii) Set  $j_0 \neq j_1$  and  $k_0 \in \mathcal{K}_{j_0}$ . Then  $P(\tau_{k_0} \cap \tau_{j_1}^*) \leq g(\tau, \tau^*)$ .

- (iii) If  $g(\tau, \tau^*) = 0$ , then for all  $j$ , there exists  $\mathcal{K}_j$  such that  $\tau_j^* = \bigcup_{k \in \mathcal{K}_j} \tau_k$  (equalities hold up to  $P$ -null sets, as the following conclusions). We derive from this that  $\mathcal{K}_j$ 's are mutually disjoint and  $K \geq K^*$ :  $\tau$  is a sub-partition of  $\tau^*$ . Thus, when  $g(\tau, \tau^*) = 0$  with  $K = K^*$ , we do have  $\tau = \tau^*$ . Suppose now that  $g(\tau, \tau^*) < \Delta^*/2$ . We still have mutually disjoint  $\mathcal{K}_j$ 's, and therefore again,  $K \geq K^*$ . In particular, when  $K = K^*$ , one can assume that  $\mathcal{K}_j = \{j\}$  for each  $j$ . Observe that  $g(\tau, \tau^*) \geq \Delta^*$  as soon as  $K < K^*$ .
- (iv) To conclude with, note that  $\bigcup_{1 \leq j \leq K^*} \mathcal{K}_j$  does not necessarily cover  $\{1, \dots, K\}$ . Nevertheless, it does when  $g(\tau, \tau^*) < \Delta^*/\bar{K}$ .

*Proof.* (i) Denoting  $\tilde{\mathcal{K}} = \mathcal{K} \cup \{k\}$  with  $k \notin \mathcal{K}$ , the following equalities hold true:

$$\begin{aligned} P\left(\tau_j^* \Delta \left(\bigcup_{k \in \tilde{\mathcal{K}}} \tau_k\right)\right) &= P\left(\tau_j^* \nabla \left(\bigcup_{k \in \tilde{\mathcal{K}}} \tau_k\right)\right) + \sum_{k \in \tilde{\mathcal{K}}} P(\tau_k \nabla \tau_j^*) \\ &= P\left(\tau_j^* \nabla \left(\bigcup_{k \in \mathcal{K}} \tau_k\right)\right) - P(\tau_k \cap \tau_j^*) + \sum_{k \in \mathcal{K}} P(\tau_k \nabla \tau_j^*) + P(\tau_k \nabla \tau_j^*) \\ &= P\left(\tau_j^* \Delta \left(\bigcup_{k \in \mathcal{K}} \tau_k\right)\right) + P(\tau_k \nabla \tau_j^*) - P(\tau_k \cap \tau_j^*). \end{aligned}$$

We conclude taking on the one hand  $\mathcal{K} = \mathcal{K}_j$  and  $\mathcal{K} = \mathcal{K}_j - \{j\}$  on the other hand. If  $\mathcal{K}_j = \{k\}$  and  $g(\tau, \tau^*) \leq \Delta^*/2$ , then use  $P(\tau_k \nabla \tau_j^*) \leq \Delta^*/2$  and  $\Delta^* \leq P(\tau_k) = P(\tau_k \nabla \tau_j^*) + P(\tau_k \cap \tau_j^*)$ .

(ii) We have indeed

$$P(\tau_{k_0} \cap \tau_{j_1}^*) \leq P(\tau_{k_0} \nabla \tau_{j_0}^*) \leq P\left(\left(\bigcup_{k \in \mathcal{K}_{j_0}} \tau_k\right) \nabla \tau_{j_0}^*\right) \leq P\left(\left(\bigcup_{k \in \mathcal{K}_{j_0}} \tau_k\right) \Delta \tau_{j_0}^*\right) \leq g(\tau, \tau^*).$$

(iii) Let  $g(\tau, \tau^*) < \Delta^*/2$ . Suppose  $k \in \mathcal{K}_{j_0}$  and  $k \in \mathcal{K}_{j_1}$  with  $j_0 \neq j_1$ , too. Then

$$P(\tau_k) = P((\tau_k \nabla \tau_{j_0}^*) \cup (\tau_k \nabla \tau_{j_1}^*)) \leq P(\tau_k \nabla \tau_{j_0}^*) + P(\tau_k \nabla \tau_{j_1}^*) < \Delta^*,$$

which is excluded.

(iv) To see that, suppose we can take  $k_0 \notin \bigcup_{1 \leq j \leq K^*} \mathcal{K}_j$ :

$$P(\tau_{k_0}) = \sum_{j=1}^{K^*} P(\tau_{k_0} \cap \tau_j^*) \leq \sum_{j=1}^{K^*} P\left(\tau_j^* \nabla \left(\bigcup_{k \in \mathcal{K}_j} \tau_k\right)\right) < K^* \cdot \Delta^*/\bar{K},$$

and that is impossible. □

We will use two pseudo-distances between two parameters  $\theta$  and  $\theta^*$  respectively associated to  $\tau$  and  $\tau^*$  (as explained in 2.1) that are compatible with the definition of the gap  $g(\tau, \tau^*)$ . Set some parameter  $\theta^* = (\theta_j^*)_{1 \leq j \leq K^*}$  in  $\bar{\Theta}_{K^*}$  having no equal coordinates. Let  $v$  be a nonnegative continuous function on the set  $\{\theta_1^*, \dots, \theta_{K^*}^*\} \times \bar{\Theta}$ , continuously differentiable with respect to its second variable, whose derivative has continuous extension on  $\{\theta_1^*, \dots, \theta_{K^*}^*\} \times \bar{\Theta}$ . Furthermore,  $v$  must verify the condition  $v(\theta_j^*, \theta) = 0$  ( $1 \leq j \leq K^*$ ,  $\theta \in \bar{\Theta}$ ) if and only if  $\theta = \theta_j^*$ .

**Definition 2.5.** Let  $\tau$  and  $\tau^*$  be two  $\mathcal{F}$ -partitions of the set  $\mathcal{X}$ ,  $K$  and  $K^*$  their respective cardinality. Let  $\theta$  and  $\theta^*$  be two parameters, respectively taken in  $\Theta_K$  and  $\Theta_{K^*}$ ,  $\theta^*$  having no equal coordinates. We define the two following pseudo-distances between them

$$\begin{aligned} d_2(\theta, \theta^*) &= \max_{1 \leq j \leq K^*} \max_{k \in \mathcal{K}_j} \|\theta_j^* - \theta_k\|_2, \\ d_v(\theta, \theta^*) &= \max_{1 \leq j \leq K^*} \max_{k \in \mathcal{K}_j} v(\theta_j^*, \theta_k). \end{aligned}$$

Here, indexes  $\mathcal{K}_j$  arise from definition of  $g(\tau, \tau^*)$ .

**Remark 2.6.**

- Naturally, thanks to properties of  $v$ ,  $d_v \leq Cd_2$  holds (recall that in the whole paper, different positive constants might be denoted  $C$ ).
- If  $d_v(\theta, \theta^*) = 0$ , then  $\mathcal{K}_j$ 's do not intersect each other and  $K \geq K^*$ . Moreover, if  $g(\tau, \tau^*) = d_v(\theta, \theta^*) = 0$  and  $K = K^*$ , then  $\tau = \tau^*$  and  $\theta = \theta^*$ .

### 3. MODELIZATION, OBSERVATIONS, CONTRAST

#### Modelization

We assume that the set  $\mathcal{X}$  is covered by a  $\mathcal{F}$ -partition  $\tau^*$  consisting of  $K^*$  subsets  $\tau_j^*$  ( $j = 1, \dots, K^*$ ). In the whole paper, index  $j$  will generally be devoted to the description of objects related to  $(\tau^*, \theta^*)$ . Index  $k$  will correspond to other  $(\tau, \theta)$ . A parameter  $\theta^*$  with no equal coordinates of  $\Theta_{K^*}$  is associated with  $\tau^*$  and  $\vartheta^*$  is defined by  $\vartheta^* = \sum_j \theta_j^* \mathbb{1}\{\tau_j^*\}$ . We define respectively  $\mathcal{T}_{K, \delta}$  and  $\Theta_{K, \delta}$  as the sets of all partitions  $\tau$  of cardinality  $K$  such that  $g(\tau, \tau^*) > \delta$  and all parameters  $\theta$  of length  $K$  such that  $d_2(\theta, \theta^*) > \delta$ .

We suppose the existence of a random field indexed by  $x \in \mathcal{X}$  of possibly dependent random variables (rv): for any  $x \in \mathcal{X}$ , a rv  $Y_x$  taking its values in  $\mathbb{R}^q$  is generated according to a law which depends on  $\vartheta(x)$ .

Our aim is to estimate  $K^*$ ,  $\tau^*$  and  $\theta^*$  from random observations under as mild conditions as possible.

#### Two classical examples

- *Detection in the mean*

Here,  $Y_x = \vartheta^*(x) + Y'_x$  for some strictly stationary field of centered rv  $(Y'_x)_{x \in \mathcal{X}}$ . Consequently, the vector of true parameters  $\theta^*$  is understood as the vector of the true possible means. Thus,  $Y_x$  has mean  $\theta_j^*$  if and only if  $x \in \tau_j^*$ .

- *Detection in both mean and variance*

Denoting  $\theta^* = (\mu^*, s^{2*})$  and  $\vartheta_1^* = \sum_j \mu_j^* \mathbb{1}\{\tau_j^*\}$ ,  $\vartheta_2^* = \sum_j s_j^{2*} \mathbb{1}\{\tau_j^*\}$ , we define  $Y_x = \vartheta_1^* + \vartheta_2^{*1/2} Y'_x$ . Here,  $(Y'_x)_{x \in \mathcal{X}}$  is a strictly stationary field of centered rv with variance 1. In this example,  $Y_x$  has mean  $\mu_j^*$  and variance  $s_j^{2*}$  if and only if  $x \in \tau_j^*$ .

#### 3.1. Observations and first assumptions

One observes  $n$  rv  $(X_i, Y_i)$  ( $i = 1, \dots, n$ ).  $X_1^n$  denotes the vector of mutually independent variables  $(X_1, \dots, X_n)$  generated independently of  $(Y_x)_{x \in \mathcal{X}}$ .  $X_i$  takes its values in  $\mathcal{X}$  and  $Y_i := Y_{X_i}$  takes its in  $\mathbb{R}^q$ .  $P$  is the common distribution of  $X_1, \dots, X_n$ .  $P_n$  is the empirical law.

**Remark 3.1.** All the proofs still hold up to minor changes when  $X_1, \dots, X_n$  are not identically distributed but still independent and  $P$  denotes the arithmetic mean of their respective distribution  $P_{X_i}$ . Furthermore, we emphasize that the rv  $Y_1, \dots, Y_n$  are possibly dependent.

Consider now the first three assumptions: they concern  $\mathcal{F}$  and  $P$ .

**A1:** The random variable  $\sup\{|P_n(F) - P(F)| : F \in \mathcal{F}\}$  converges  $P$ -almost surely ( $P$ -as) to zero. In other words,  $\mathcal{F}$  is  $P$ -Glivenko–Cantelli.

**A2:** There exists a sequence  $\{r_n\} \downarrow 0$  such that  $\liminf_n nr_n > 0$  and

$$\lim_{\eta \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P} \left( \sup \left\{ \frac{P(F) - P_n(F)}{P(F)} : F \in \mathcal{F}, P(F) \geq \eta r_n \right\} \geq \frac{1}{2} \right) = 0.$$

**Remark 3.2** (on Assumptions **A1** and **A2**). Assumption **A1** is fulfilled whenever  $\mathcal{F}$  is a finite VC-dimension VC class. In the sequel, the case of  $\mathcal{F}$  finite VC-dimension VC class will be the more general example for  $\mathcal{F}$ . For a wide family of examples, see for instance [26]. On the other hand, Proposition 3.3 below (whose proof, postponed in Appendix 7.1, require *independence* of  $X_i$ 's) casts some light on the Assumption **A2**.

**Proposition 3.3.** *Assumption **A2** holds whenever  $\mathcal{F}$  is a finite VC-dimension VC class and the sequence  $\left\{ \frac{\log r_n}{nr_n} \right\}$  is bounded.*

**Remark 3.4** (on Prop. 3.3). Choices of  $r_n = (\log^\alpha n)^\beta/n$  with integer  $\alpha \geq 1$  and positive  $\beta$  are obviously included (with notation  $\log^{\alpha+1} = \log \circ \log^\alpha$ ,  $\log^1 = \log$ ).

The last assumption of this Section concerns the control of the moment of order  $h$  of  $P_n(G)$  for  $G \in \mathcal{G}$ :

**A3:** For any  $h \in (1, 2)$  and  $G \in \mathcal{G}$ , for some constant  $A > 0$  depending on  $h$  only,

$$\mathbb{E} (P_n(G)^h) \leq A (\mathbb{E} P_n(G))^h = A P(G)^h.$$

**Remark 3.5** (on Assumption **A3**). Note that Jensen's inequality yields straightforwardly to the reversed lower bound  $P(G)^h \leq \mathbb{E}(P_n(G)^h)$ . Assumption **A3** is always satisfied for independent, non necessarily identically distributed, rv: it is a simple consequence of Rosenthal's inequality, see *e.g.* [21]. We will use this inequality to derive useful maximal inequalities in Section 4.

### 3.2. Further assumptions: on the contrast

The following assumption ensures the existence of a contrast  $J_n$  adapted to our model.  $J_n(\tau, \theta)$  is obtained as a sum of local contrasts  $W_n(\tau_k, \theta_k)$  computed at  $(\tau_k, \theta_k)$ .

**A4:** Let  $\varphi : \bar{\Theta} \rightarrow \mathbb{R}$  and  $\psi : \bar{\Theta} \rightarrow \mathbb{R}^r$  be two continuously differentiable functions with continuous extensions of the derivatives on  $\bar{\Theta}$ . Let  $\xi : \mathbb{R}^q \rightarrow \mathbb{R}^r$  be such that  $\xi(Y_x) \in L^1(\mathbb{P})$  for any  $x \in \mathcal{X}$  and  $\xi(Y_X) \in L^1(\mathbb{P})$  for  $X \sim P_{X_i}$ -distributed. Define the local contrasts for  $(\tau_k, \theta_k)$  ( $k = 1, \dots, K$ ) by

$$W_n(\tau_k, \theta_k) = n^{-1} \sum_{i=1}^n \left\{ \varphi(\theta_k) + \psi(\theta_k)^T \xi(Y_i) \right\} \mathbb{1}\{X_i \in \tau_k\}$$

and introduce the corresponding limit contrast  $w : \{\theta_1^*, \dots, \theta_{K^*}^*\} \times \bar{\Theta} \rightarrow \mathbb{R}$ , which is supposed to satisfy:

- $P$ -as for all  $i$  such that  $X_i \in \tau_j^*$  and any  $\theta \in \bar{\Theta}$ ,

$$w(\theta_j^*, \theta) = \varphi(\theta) + \langle \psi(\theta), \mathbb{E}(\xi(Y_i) | X_i) \rangle; \tag{1}$$

- $w(\theta_j^*, \theta) \geq w(\theta_j^*, \theta_j^*)$  for any  $(\theta_j^*, \theta) \in \{\theta_1^*, \dots, \theta_{K^*}^*\} \times \bar{\Theta}$ , equality if and only if  $\theta = \theta_j^*$ .

Denote  $v$  the centered limit contrast, that is  $v(\theta_j^*, \theta) = w(\theta_j^*, \theta) - w(\theta_j^*, \theta_j^*)$ , any  $(\theta_j^*, \theta)$ . Then  $v$  is nonnegative, continuous on  $\{\theta_1^*, \dots, \theta_{K^*}^*\} \times \bar{\Theta}$ , continuously differentiable on  $\{\theta_1^*, \dots, \theta_{K^*}^*\} \times \Theta$  with respect to its second variable. Its derivative has continuous extension on  $\{\theta_1^*, \dots, \theta_{K^*}^*\} \times \bar{\Theta}$ . Finally,  $v$  is zero only on the diagonal. Thus, following Definition 2.5 in Section 2.2, we can define a pseudo-distance  $d_v$  from  $v$ . Furthermore, since  $\{v(\theta_j^*, \cdot), j = 1, \dots, K^*\}$  are continuous, there exist  $\rho^*, v^* > 0$  such that, for any  $j_0 \neq j_1$ ,

$$\inf \left\{ v(\theta_{j_0}^*, \theta) : \|\theta - \theta_{j_1}^*\|_2 \leq \rho^* \right\} - \sup \left\{ v(\theta_{j_0}^*, \theta) : \|\theta - \theta_{j_0}^*\|_2 \leq \rho^* \right\} \geq v^*.$$

**Remark 3.6** (on Assumption **A4**).

- Condition (1) in the former assumption controls the way the rv  $Y_i$  depends on  $X_i$  through  $\tau^*$ . In particular,  $P$ -as for  $i, i'$  such that  $X_i, X_{i'} \in \tau_j^*$ ,

$$E(\xi(Y_i) | X_i) - E(\xi(Y_{i'}) | X_{i'}) \in \text{Vect}(\psi(\Theta))^\perp$$

and they are equal as soon as  $\text{Vect}(\psi(\Theta)) = \mathbb{R}^r$ , which is clearly the case for  $r = 1$  and  $\xi \neq 0$ .

- Note that  $E(W_n(\tau_j^*, \theta) | X_1^n) = P_n(\tau_j^*) w(\theta_j^*, \theta)$ , where  $P_n(\tau_j^*)$  tends to  $P(\tau_j^*)$   $P$ -as. Thus,  $w(\theta_j^*, \cdot)$  can be understood as a rescaled limit conditional expectation of the local contrast computed at  $\tau_j^*$ .

The next assumption concerns  $v$ :

**A5:** There exist  $B > 0$ ,  $\sigma > 0$  such that (up to change of  $\rho^*$ )

$$\text{if } \|\theta - \theta_j^*\|_2 \leq \rho^*, \text{ then } v(\theta_j^*, \theta) \geq B \|\theta - \theta_j^*\|_2^\sigma \quad (j = 1, \dots, K^*).$$

*Back to the classical examples*

- *Detection in the mean*

We choose the following local criterion function

$$W_n(\tau_k, \theta_k) = n^{-1} \sum_{i=1}^n (Y_i - \theta_k)^2 \mathbb{1}\{X_i \in \tau_k\} - n^{-1} \sum_{i=1}^n Y_i^2 \mathbb{1}\{X_i \in \tau_k\}.$$

Here,  $\varphi(\theta) = \theta^2$ ,  $\psi(\theta) = -2\theta$  and  $\xi(y) = y$ . For this particular criterion,  $v(\theta_j^*, \theta) = (\theta - \theta_j^*)^2$  and Assumption **A5** above is satisfied.

- *Detection in both mean and variance*

This time, we choose

$$W_n(\tau_k, \theta_k) = n^{-1} \sum_{i=1}^n \left\{ \frac{(Y_i - \mu_k)^2}{s_k^2} + \log s_k^2 \right\} \mathbb{1}\{X_i \in \tau_k\}.$$

Here,  $\varphi(\mu, s^2) = \mu^2/s^2 + \log s^2$ ,  $\psi(\mu, s^2) = (-2\mu, 1)/s^2$  and  $\xi(y) = (y, y^2)$ . Moreover, we have

$$v(\theta_j^*, \theta) = \frac{(\mu_j^* - \mu)^2}{s^2} + \log \frac{s^2}{s_j^{*2}} + \frac{s_j^{*2}}{s^2} - 1.$$

Thus,  $v(\theta_j^*, \theta)$  is twice the Kullback–Leibler divergence  $H(\mathcal{N}_{\theta_j^*} | \mathcal{N}_\theta)$  for Gaussian rv  $\mathcal{N}_{\theta_j^*}$  (resp.  $\mathcal{N}_\theta$ ) of mean and variance  $\theta_j^*$  (resp.  $\theta$ ). Besides, Assumption **A5** above does hold for  $\Theta = ]a, b[ \times ]c, d[$  with  $d > c > 0$ .

Note that in both cases, minimization of  $\theta \mapsto W_n(\tau_j^*, \theta)$  leads to the natural least squares estimators of the parameter  $\theta_j^*$ .

#### 4. CONTROLLING RANDOM FLUCTUATIONS VIA MAXIMAL INEQUALITIES

Let us define the centered random field of fluctuations  $(Z_x)_{x \in \mathcal{X}}$  and the rv  $Z_i$  ( $i = 1, \dots, n$ ) by

$$Z_x = \xi(Y_x) - E(\xi(Y_x)) \quad \text{and} \quad Z_i = \xi(Y_i) - E(\xi(Y_i) | X_i)$$



and for any  $x_1^n \in \mathcal{X}^n$ , define the corresponding sums over any set  $G \in \mathcal{G}$

$$\Sigma_{x_1^n}(G) = \sum_{i=1}^n Z_{x_i} \mathbb{1}\{x_i \in G\} \quad \text{and} \quad \Sigma_{X_1^n}(G) = \sum_{i=1}^n Z_i \mathbb{1}\{X_i \in G\}.$$

Denote finally  $S_n(G; \theta) = \psi(\theta)^T \Sigma_{X_1^n}(G)$  and, for any  $\mathcal{F}$ -partition  $\tau$ ,  $n_{kj} = nP_n(\tau_k \cap \tau_j^*)$ . Then

$$W_n(\tau_k, \theta_k) = n^{-1} \sum_{j=1}^{K^*} \{n_{kj} w(\theta_j^*, \theta_k) + S_n(\tau_k \cap \tau_j^*; \theta_k)\}.$$

We require

**A6:** There exist  $C_1 > 0$  and  $h \in (1, 2)$  such that, for any  $\delta > 0$ ,  $G \in \mathcal{G}$ ,

$$P\left(\sup\left\{\|\Sigma_{X_1^n}(F)\|_\infty : F \in \mathcal{F}(G)\right\} \geq \delta \mid X_1^n\right) \leq \frac{C_1}{\delta^2} \left(\sum_{i=1}^n \mathbb{1}\{X_i \in G\}\right)^h \quad P\text{-as.}$$

Here,  $\mathcal{F}(G)$  denotes the set  $\{F \cap G : F \in \mathcal{F}\}$ .

Observe that Assumptions **A3** and **A6** yield (uncondition with respect to  $X_1^n$  and then bound above) the following maximal inequality:

**Lemma 4.1.** *Under Assumptions **A3** and **A6**, there exists  $C_2 > 0$  such that, for any  $\delta > 0$  and  $G \in \mathcal{G}$ ,*

$$P\left(\sup\left\{\|\Sigma_{X_1^n}(F)\|_\infty : F \in \mathcal{F}(G)\right\} \geq \delta\right) \leq \frac{C_2 n^h P(G)^h}{\delta^2}.$$

**Remark 4.2** (on Assumption **A6** and Lem. 4.1). Aim of Assumption **A6** consists of ensuring result of Lemma 4.1. The linearity of  $S_n$  in  $\Sigma_{X_1^n}$  is needed to derive uniform control of  $S_n(G; \theta)$  in  $(G, \theta)$  from uniform control of  $\Sigma_{X_1^n}(G)$  in  $G$ . Besides, we can propose some mild alternative condition to Assumption **A6**, see Section 7.2.

*Back to the classical examples*

Applying Lemma 4.1 and Assumption **A1**, we get that

$$(nP_n(\tau_j^*))^{-1} \|\Sigma_{X_1^n}(\tau_j^*)\|_\infty = o_P(1).$$

The previous result yields that the least squares estimators obtained by minimization of the respective contrasts at  $\tau_j^*$  are consistent.

Finally, state the last assumption: it concerns  $\mathcal{F}_0$ .

**A7:** For some constant  $\gamma > 0$  depending on  $\mathcal{F}_0$  only, for any  $G \in \mathcal{G}$  and  $r > 0$ , there exists  $\tilde{G} \in \mathcal{G}$  subset of  $G$  with  $P(\tilde{G}) \leq \gamma r$  such that

$$\left\{F \in \tilde{\mathcal{F}}_0(G) : P(F) \leq r\right\} \subset \mathcal{F}(\tilde{G}). \tag{2}$$

Here,  $\tilde{\mathcal{F}}_0(G)$  denotes the set  $\{F \cap G : F \in \mathcal{F}_0, F \nabla G \neq \emptyset\}$ .

**Remark 4.3** (on Assumption **A7**). Simplest examples are again when  $\mathcal{F}_0$  is composed of rectangles or polygons. Besides, this result still holds when  $\tilde{\mathcal{F}}_0(G)$  is replaced in (2) by the set  $\tilde{\mathcal{F}}(G) = \{F \cap G : F \nabla G \neq \emptyset\}$  where  $F$  is an union of at most  $\bar{K}$  elements of  $\mathcal{F}_0$ .

We can now state a key-result that completes Lemma 4.1:

**Lemma 4.4.** *Under Assumptions **A3**, **A6**, **A7**, there exists  $C_3 > 0$  such that, for any  $G \in \mathcal{G}$ , any  $v > 0$ ,*

$$\mathbb{P} \left( \sup \left\{ \frac{\|\Sigma_{X_1^n}(F)\|_\infty}{nP(F)} : F \in \tilde{\mathcal{F}}(G), P(F) \geq v \right\} \geq \delta \right) \leq \frac{C_3(nv)^{h-2}}{\delta^2}.$$

*Proof.* The event whose  $P$ -probability we want to bound above is included in the union over  $j \geq 0$  of the events

$$\sup \left\{ \|\Sigma_{X_1^n}(F)\|_\infty : F \in \tilde{\mathcal{F}}(G), P(F) \leq 2^{j+1}v \right\} \geq 2^j nv \cdot \delta.$$

Assumption **A7** yields that the former event indexed by  $j$  is itself included in the following one

$$\sup \left\{ \|\Sigma_{X_1^n}(F)\|_\infty : F \in \mathcal{F}(\tilde{G}_j) \right\} \geq 2^j nv \cdot \delta$$

for some subset  $\tilde{G}_j$  of  $G$  satisfying  $P(\tilde{G}_j) \leq \gamma 2^{j+1}v$ . Lemma 4.1 allows to conclude.  $\square$

## 5. THE CASE OF KNOWN CARDINALITY OF THE TRUE PARTITION

### 5.1. Definition of the estimator

We address in this section the consistency of our estimator when the cardinal  $K$  of the estimator of  $\tau^*$  is *a priori* fixed.

The estimator  $(\hat{\tau}_n, \hat{\theta}_n)$  of  $(\tau^*, \theta^*)$  is constructed by minimization over  $\mathcal{T}_K \times \Theta_K$  of the contrast  $J_n$ , or equivalently of the centered contrast  $U_n$ , with

$$\begin{aligned} J_n(\tau, \theta) &= \sum_{k=1}^K W_n(\tau_k, \theta_k), \\ U_n(\tau, \theta) &= J_n(\tau, \theta) - J_n(\tau^*, \theta^*) = u_n(\tau, \theta) + e_n(\tau, \theta) \end{aligned}$$

where

$$\begin{aligned} u_n(\tau, \theta) &= n^{-1} \sum_{j=1}^{K^*} \sum_{k=1}^K n_{kj} v(\theta_j^*, \theta_k) \quad \text{and} \\ e_n(\tau, \theta) &= n^{-1} \sum_{j=1}^{K^*} \sum_{k=1}^K \{S_n(\tau_k \cap \tau_j^*; \theta_k) - S_n(\tau_k \cap \tau_j^*; \theta_j^*)\}. \end{aligned}$$

In the sequel, we will denote  $\hat{\theta}_n(\tau_k) = \arg \min\{W_n(\tau_k, \theta) : \theta \in \bar{\Theta}\}$  for any  $\tau \in \mathcal{T}_K$  and  $1 \leq k \leq K$ . Observe then that  $(\hat{\theta}_n(\tau_k))_k = \arg \min\{J_n(\tau, \theta) : \theta \in \bar{\Theta}_K\}$ . Moreover, we will denote  $\hat{\theta}_n^* = \hat{\theta}_n(\tau^*)$ . We will write  $\hat{\theta}_{nj}^*$  for the  $j^{\text{th}}$  coordinate of  $\hat{\theta}_n^*$  and  $\hat{\theta}_{nj}$  for the  $j^{\text{th}}$  coordinate of  $\hat{\theta}_n = \hat{\theta}_n(\hat{\tau}_n)$ .

The next proposition casts some light on the behaviour of the total fluctuation term  $e_n$ . It is a direct consequence of Lemma 4.1 since  $S_n(\tau_k \cap \tau_j^*; \theta) = \psi(\theta)^T \Sigma_{X_1^n}(\tau_k \cap \tau_j^*)$  and  $\psi$  is bounded.

**Proposition 5.1.** *Under the assumptions of Lemma 4.1,  $e_n$  is uniformly  $o_{\mathbb{P}}(1)$  over  $\mathcal{T}_K \times \Theta_K$ .*

5.2. Consistency

Consistency is of course hopeless for  $K < K^*$ , since then  $g(\tau, \tau^*) \geq \Delta^*/2$ . We prove that our estimator is consistent as soon as  $K \geq K^*$ :

**Theorem 5.2.** *Set  $K \geq K^*$  and let  $(\hat{\tau}_n, \hat{\theta}_n)$  be the estimator defined in Section 5.1. Under Assumptions **A1**, **A3**, **A4** and **A6**,  $(\hat{\tau}_n, \hat{\theta}_n)$  is consistent, i.e. that both  $g(\hat{\tau}_n, \tau^*)$  and  $d_2(\hat{\theta}_n, \theta^*)$  converge to 0 in  $P$ -probability.*

The proof of Theorem 5.2 is based on a technical lemma of great importance throughout this paper, namely Lemma 5.3, and on application of Proposition 5.1.

**Lemma 5.3.** *Under Assumption **A1**, there exists  $C^* > 0$  such that, for any  $K$  and all  $\delta > 0$ ,*

$$\lim_{n \rightarrow \infty} P \left( \bigcap_{l \geq n} [\forall (\tau, \theta) \in \mathcal{T}_K \times \Theta_K, u_l(\tau, \theta) \geq C^* d_v(\theta, \theta^*)] \right) = 1,$$

$$\lim_{n \rightarrow \infty} P \left( \bigcap_{l \geq n} [\forall (\tau, \theta) \in \mathcal{T}_{K, \delta} \times \Theta_K, u_l(\tau, \theta) \geq C^* g(\tau, \tau^*)] \right) = 1.$$

A sketch of proof of Lemma 5.3 can be found in Appendix 7.3. We are able now to demonstrate Theorem 5.2.

*Proof.* (Th. 5.2). Let us prove that  $g(\hat{\tau}_n, \tau^*) = o_P(1)$ . We would show that  $d_2(\hat{\theta}_n, \theta) = o_P(1)$  along the same lines.

If  $g(\hat{\tau}_n, \tau^*) > \delta$ , then  $\inf\{U_n(\tau, \theta) : (\tau, \theta) \in \mathcal{T}_{K, \delta} \times \Theta_K\}$  is non positive and consequently, thanks to Lemma 5.3, for some constant  $c > 0$ , for any  $\varepsilon > 0$  and for  $n$  large enough,

$$\sup \{|e_n(\tau, \theta)| : (\tau, \theta) \in \mathcal{T}_K \times \Theta_K\} > c\delta$$

with probability at least  $1 - \varepsilon$ . Thus, the proof is complete, since  $e_n$  is uniformly  $o_P(1)$ . □

5.3. Rate of convergence

The following theorem gives lower bounds for rates of convergence of  $\hat{\tau}_n$  and  $\hat{\theta}_n$  for  $K = K^*$ . Observe that the case  $K > K^*$  would be dealt with the same proof, up to minor changes.

**Theorem 5.4.** *Set  $K = K^*$  and let  $(\hat{\tau}_n, \hat{\theta}_n)$  be the estimator defined in Section 5.1. Under Assumptions **A1** to **A7**, assuming moreover that the coefficient  $\sigma$  that appears in Assumption **A5** satisfies  $\sigma \geq 2/h$ , the sequences*

$$\{r_n^{-1} g(\hat{\tau}_n, \tau^*)\} \quad \text{and} \quad \{n^{(2-h)/2(\sigma-1)} d_2(\hat{\theta}_n, \theta^*)\}$$

are uniformly bounded in  $P$ -probability.

**Remark 5.5.** This lower bound of the rate of convergence of  $\hat{\tau}_n$  does not depend on the dependence structure of the sequence  $(Y_i)$  since the coefficient  $h$  of Assumption **A6** does not appear in the bound and  $\{r_n\}$  depends solely on  $X_1^n$ :  $g(\hat{\tau}_n, \tau^*) = O_P(r_n)$ . On the contrary, the rate of convergence of  $\hat{\theta}_n$  does depend on the dependence structure of  $(Y_i)$ . Moreover, this rate is the same than when the true partition  $\tau^*$  is known.

*Proof.* Denote in the sequel  $\zeta = 1/2(\sigma - 1)$ . We shall prove that both probabilities

$$P \left( \delta_n \geq g(\hat{\tau}_n, \tau^*) \geq \delta r_n, d_2(\hat{\theta}_n, \theta^*) < \delta_n \right), \tag{3}$$

$$P \left( \delta_n \geq d_2(\hat{\theta}_n, \theta^*) \geq \delta n^{\zeta(h-2)}, g(\hat{\tau}_n, \tau^*) < \delta_n \right) \tag{4}$$

tend to zero as  $n, \delta \uparrow \infty$ , where the sequence  $\{\delta_n\} \downarrow 0$  slowly enough, precisely with  $n^{\zeta(2-h)}\delta_n \rightarrow \infty$ , and is driven from consistency. Let us deal first with the first of them: we restrict ourselves to the event whose probability is written in (3).

By definition,  $\hat{\tau}_n$  minimizes  $\tau \mapsto U'_n(\tau) = \sum_{k=1}^{K^*} W_n(\tau_k; \hat{\theta}_n(\tau_k)) - \sum_{j=1}^{K^*} W_n(\tau_j^*; \hat{\theta}_{nj}^*)$ . Furthermore, the definition of  $\hat{\theta}_{nj}^*$  and simple decomposition of  $U'_n(\tau)$  yield the next first inequality

$$\begin{aligned} U'_n(\tau) &= \sum_{j=1}^{K^*} W_n(\tau_j^*; \hat{\theta}_{nj}) - \sum_{j=1}^{K^*} W_n(\tau_j^*; \hat{\theta}_{nj}^*) + \sum_{j=1}^{K^*} \left\{ W_n(\tau_j \nabla \tau_j^*; \hat{\theta}_{nj}) - W_n(\tau_j^* \nabla \tau_j; \hat{\theta}_{nj}) \right\} \\ &\geq \sum_{j=1}^{K^*} \left\{ W_n(\tau_j \nabla \tau_j^*; \hat{\theta}_{nj}) - W_n(\tau_j^* \nabla \tau_j; \hat{\theta}_{nj}) \right\} \\ &\geq n^{-1} \sum_{j=1}^{K^*} \sum_{k \neq j} \left\{ n_{kj} v^* + S_n(\tau_k \cap \tau_j^*; \hat{\theta}_{nk}) - S_n(\tau_k \cap \tau_j^*; \hat{\theta}_{nj}) \right\}, \end{aligned}$$

where the previous one holds as soon as  $\hat{\theta}_n$  is close enough to  $\theta^*$ , that is for  $n$  large.

The point is now to separate the fluctuations in terms of  $X_i$  and  $Y_i$ .

Since  $g(\tau, \tau^*) \geq \delta r_n$ , there exist  $k_0 \neq j_0$  such that  $P(\tau_{k_0} \cap \tau_{j_0}^*) \geq \delta r_n$ , up to substitution of  $\delta$ . Furthermore, Assumption **A2** yields that, for any  $\varepsilon > 0$  and for  $n, \delta$  large enough, for any  $F \in \mathcal{F}$  such that  $P(F) \geq \delta r_n$ , we have

$$P_n(F) \geq P(F)/2 \geq \delta r_n/2, \quad (5)$$

with probability  $1 - \varepsilon$ . Forwardly, up to other substitutions on  $\delta$  and  $v^*$ , applying Cauchy-Schwarz's inequality and invoking boundedness of  $\psi$ , we get the next first inequality

$$\begin{aligned} U'_n(\tau) &\geq n^{-1} \sum_{j=1}^{K^*} \sum_{k \neq j} \left\{ (n_{kj} \vee \delta n r_n) v^* + S_n(\tau_k \cap \tau_j^*; \hat{\theta}_{nk}) - S_n(\tau_k \cap \tau_j^*; \hat{\theta}_{nj}) \right\} \\ &\geq C n^{-1} \sum_{j=1}^{K^*} \sum_{k \neq j} \left\{ (n P(\tau_k \cap \tau_j^*) \vee \delta n r_n) - c \|\Sigma_{X_1^n}(\tau_k \cap \tau_j^*)\|_\infty \right\} \end{aligned}$$

with probability at least  $1 - \varepsilon$  ( $c$  is a positive constant, independent of  $\tau$ ).

The delicate point in the previous display takes place in the second inequality. We have to verify that  $n_{kj} \vee \delta n r_n \geq C n P(\tau_k \cap \tau_j^*) \vee \delta n r_n$ . Carefully considering cases where, on the one hand,  $n_{kj}$  is greater than  $\delta n r_n$ , and on the other hand, where it is less than  $\delta n r_n$  (with subcases  $P(\tau_k \cap \tau_j^*)$  greater or less than  $\delta r_n$ ) yields the expected result. Note that this inequality does not hold anymore if we replace  $\{r_n\}$  by some sequence  $\{r'_n\}$  that decreases faster to 0.

Hence, the proof will be complete if we show that the convergences to 0 as  $n, \delta \uparrow \infty$  of the probabilities of the following events hold for any  $c > 0$  and  $j_0 \neq j_1$  (where  $F$  denotes any set of the form  $\tau \cap \tau_{j_1}^*$  with  $\tau \in \mathcal{F}$  such that  $P(\tau \cap \tau_{j_0}^*) \leq \delta_n$ ):

$$\begin{aligned} \sup \left\{ \|\Sigma_{X_1^n}(F)\|_\infty : P(F) \leq \delta r_n \right\} &\geq \delta n r_n, \\ \sup \left\{ \frac{\|\Sigma_{X_1^n}(F)\|_\infty}{n P(F)} : P(F) \geq \delta r_n \right\} &\geq c. \end{aligned}$$

This is a direct consequence of Lemmas 4.1 and 4.4.

Let us show now that the expression in (4) goes to 0 when  $n, \delta \uparrow \infty$ , too. Observe that for  $n$  large enough and on the events whose probabilities are given by (4), we have (with a view to application of Lem. 5.3) the lower bounding  $d_v(\hat{\theta}_n, \theta^*) \geq B\delta^\sigma n^{\sigma\zeta(h-2)}$ . Moreover, Assumption **A1** together with  $g(\tau, \tau^*) < \min_j P(\tau_j^*)/2$  imply that, for any  $\varepsilon > 0$  and for  $n$  large enough,  $P_n(\tau_j \cap \tau_j^*) \geq P(\tau_j^*)/4$  with probability at least  $1 - \varepsilon$ . Thus, for any  $(\tau, \theta)$  of  $\mathcal{T}_{K^*} \times \Theta_{K^*}$  satisfying the same conditions than  $(\hat{\tau}_n, \hat{\theta}_n)$  on the events whose probabilities are written in (4), we have with probability at least  $1 - \varepsilon$ , for  $n$  large enough and any  $1 \leq j \leq K^*$ ,

$$u_n(\tau, \theta) \geq C^* \delta^\sigma n^{\sigma\zeta(h-2)} \vee a_1 \|\theta_j - \theta_j^*\|_2^\sigma \tag{6}$$

for some  $a_1 > 0$  independent of  $(\tau, \theta)$ . Note that the first term in the maximum comes from Lemma 5.3. To conclude this first step, remark that the preceding inequality together with the following one

$$x/y^\sigma \vee y^{\sigma/(\sigma-1)}z \geq x^{1/\sigma}z^{(\sigma-1)/\sigma} \quad (x, y, z > 0),$$

yield (for some constant  $a_2 > 0$  independent of  $(\tau, \theta)$ )

$$u_n(\tau, \theta) \geq a_2 \|\theta_j - \theta_j^*\|_2 \delta^{\sigma-1} n^{h/2-1}. \tag{7}$$

The second step consists of the same kind of arguments than in the first part of the proof: we will bound from below  $(\tau, \theta) \mapsto U_n(\tau, \theta) = \sum_{k=1}^{K^*} W_n(\tau_k; \theta_k) - \sum_{j=1}^{K^*} W_n(\tau_j^*; \theta_j^*)$  taking care of separating cases  $k = j$  and  $k \neq j$  and distributing weight we know we can count on. Precisely, consider  $(\tau, \theta)$  such as above: on events of probability at least  $1 - \varepsilon$ , for  $n$  large enough,

$$\begin{aligned} U_n(\tau, \theta) &\geq n^{-1} \sum_{j=1}^{K^*} \{a_3 n u_n(\tau, \theta) + S_n(\tau_j \cap \tau_j^*; \theta_j) - S_n(\tau_j \cap \tau_j^*; \theta_j^*)\} \\ &\quad + n^{-1} \sum_{j=1}^{K^*} \sum_{k \neq j} \{(n_{kj} \vee n u_n(\tau, \theta)) a_4 + S_n(\tau_k \cap \tau_j^*; \theta_k) - S_n(\tau_k \cap \tau_j^*; \theta_j^*)\}, \end{aligned}$$

for some constants  $a_3, a_4 > 0$ . Hence, thanks to Assumption **A2** (as before, see (5)), Taylor-Lagrange's inequality and (6, 7), with probability  $1 - 2\varepsilon$  for  $n, \delta$  large enough,

$$\begin{aligned} U_n(\tau, \theta) &\geq C n^{-1} \sum_{j=1}^{K^*} \left\{ \|\theta_j - \theta_j^*\|_2 \delta^{\sigma-1} n^{h/2} - a \|\Sigma_{X_1^n}(\tau_j \cap \tau_j^*)\|_\infty \|\theta_j - \theta_j^*\|_2 \right\} \\ &\quad + C n^{-1} \sum_{j=1}^{K^*} \sum_{k \neq j} \left\{ (nP(\tau_k \cap \tau_j^*) \vee \delta^\sigma n^{\sigma\zeta(h-2)+1}) b - c \|\Sigma_{X_1^n}(\tau_j \cap \tau_j^*)\|_\infty \right\}, \end{aligned}$$

for constants  $a, b, c > 0$ .

Consequently, the proof will be complete if we show that the convergence to 0 as  $n, \delta \uparrow \infty$  of the probabilities of the following events hold for any  $c > 0$  and  $j_0 \neq j_1$  (where  $F$  denotes any set of the form  $\tau \cap \tau_{j_1}^*$  with  $\tau \in \mathcal{F}$  such that  $P(\tau \Delta \tau_{j_0}^*) \leq \delta_n$ ):

$$\begin{aligned} &\sup \left\{ \|\Sigma_{X_1^n}(F)\|_\infty : F \in \mathcal{F} \right\} \geq \delta n^{h/2}, \\ &\sup \left\{ \|\Sigma_{X_1^n}(F)\|_\infty : P(F) \leq \delta n^{\sigma\zeta(h-2)} \right\} \geq \delta n^{\sigma\zeta(h-2)+1}, \\ &\sup \left\{ \frac{\|\Sigma_{X_1^n}(F)\|_\infty}{nP(F)} : P(F) \geq \delta n^{\sigma\zeta(h-2)} \right\} \geq c. \end{aligned}$$

This is a direct consequence of Lemmas 4.1 and 4.4 since  $\sigma \geq 2/h$ .  $\square$

#### 5.4. Number of misclassified observations

The standard scheme of proof applied to show that the number of misclassified observations is  $O_P(1)$  requires to bound below (up to a multiplicative constant) a generic term of the form  $P_n(F)$  by  $P(F)$  for  $P(F)$  possibly less than  $\delta r_n$ . Thus, Assumption **A2** is useless and we can not conclude. This difficulty is overcome independently of the dimension  $d$  when proving that the number of misclassified observations is  $O_P(nr_n)$ . We can obtain the boundedness in probability in the 1-dimensional case. The proof relies then on the natural ordering over  $\mathbb{R}$ .

**Proposition 5.6.** *Let  $N_n(\hat{\tau}_n) = \sum_{j=1}^{K^*} \sum_{k \neq j} n_{kj}$  denote the number of misclassified observations for  $\hat{\tau}_n$  with respect to  $\tau^*$ . Under assumptions of Theorem 5.4,  $N_n(\hat{\tau}_n) = O_P(1)$  for  $d = 1$  and  $N_n(\hat{\tau}_n) = O_P(nr_n)$  for higher dimensions.*

*Sketch of the proof.* Let  $\eta \in \{0, 1\}$ . If  $N_n(\hat{\tau}_n) \geq \delta(nr_n)^\eta$  and  $g(\hat{\tau}_n, \tau^*) \vee d_2(\hat{\theta}_n, \theta^*) \leq \delta r_n$  (study of that configuration suffices thanks to Th. 5.4), then  $\hat{\tau}_n$  minimizes  $U'_n$  which is lower bounded (up to the usual substitutions and to some multiplicative constant) for large enough  $n, \delta$  by

$$n^{-1} \sum_{j=1}^{K^*} \sum_{k \neq j} \left\{ (n_{kj} \vee \delta(nr_n)^\eta) - c \|\Sigma_{X_1^n}(\tau_k \cap \tau_j^*)\|_\infty \right\}.$$

Set  $d = 1$  and  $\eta = 0$ . We can follow the strategy of proof in [12]. Application of triangle's inequality shifts the problem to the control of the  $P$ -probabilities of the following events (and their left-symmetric)

$$\begin{aligned} & \sup \left\{ \left\| \sum_{l=0}^t Z_{(s^*+t)} \right\|_\infty : t^*, 0 \leq t \leq \delta \right\} \geq \delta, \\ & \sup \left\{ \frac{\left\| \sum_{l=0}^t Z_{(s^*+t)} \right\|_\infty}{t} : t^*, t \geq \delta \right\} \geq c. \end{aligned}$$

Here,  $Z_{(s)} = Z_i$  for  $X_{(s)} = X_i$  ( $\{X_{(s)}\}_s$  denotes the increasing ordered vector  $X_1^n$ ). Index  $t^*$  in the supremum ranges over all right extremities of intervals constituting subsets of  $\tau^*$ . For some  $t^*$ ,  $X_{(s^*)}$  corresponds to the nearest  $X_i$  greater than  $t^*$ . Such probabilities do go to zero, as provided by the simplest M6ricz's one dimensional inequalities that we apply here in place of Lemmas 4.1 and 4.4, and proof is complete for this case. Heuristically, the conclusion holds because the union of all intervals containing at most  $\delta$  points  $X_i$  of observation contains itself a  $O(\delta)$  number of such points.

For  $d \geq 2$ , we can not proceed as above. Actually, the union of all subsets  $\tau_k \cap \tau_j^*$  that contain at most  $\delta$  points  $X_i$  of observation may contain much more than  $O(\delta)$  points itself, *i.e.* generally a  $O(n)$ . Thus, we must conclude as in proof of Theorem 5.4, *i.e.* choose  $\eta = 1$  and conclude that  $N_n(\hat{\tau}_n) = O_P(nr_n)$ .  $\square$

**Remark 5.7.** Note that in a very specific multidimensional case usually called *pixel case*, we get  $N_n(\hat{\tau}_n) = O_P(1)$ . Indeed, suppose that  $\mathcal{F}_0$  is composed of rectangles and that  $[0, 1]^d$  is decomposed into the union of  $n^d$  mutually disjoint rectangular boxes, the *pixels*. Suppose that  $X_1, \dots, X_{n^d}$  are chosen uniformly in each box. Then,  $N_n(\hat{\tau}_n) = O_P(1)$ . The scheme of proof applied in the one dimensional case above applies here, because the union of all subsets  $\tau_k \cap \tau_j^*$  that contain at most  $\delta$  points  $X_i$  of observation contains a  $O(\delta(\log \delta)^{d-1})$  points.

## 6. THE CASE OF UNKNOWN CARDINALITY OF THE TRUE PARTITION

### 6.1. Definition of the estimator

We address in this section the case of an unknown cardinality  $K^*$  of  $\tau^*$ . According to the former section, we can construct an estimator  $(\hat{\tau}_{n,K}, \hat{\theta}_{n,K})$  of any cardinality  $K$ , *i.e.* for any *a priori* fixed cardinality of the

estimator. The point is to select the best estimator among the family  $(\hat{\tau}_{n,K}, \hat{\theta}_{n,K})_K$ . Naturally, models with large cardinality  $K$  are favoured, hence the idea of penalizing the contrast  $J_n$  by adding a penalization term  $\beta_n K$ . Its role is to balance out this effect.

Thus, estimation of the triplet  $(K^*, \tau^*, \theta^*)$  is performed by minimizing a penalized contrast constructed with  $J_n$  as defined in Section 5. The estimator of  $(K^*, \tau^*, \theta^*)$  is taken to achieve the minimization of the penalized contrast  $\tilde{J}_n$  given by

$$\tilde{J}_n(K, \tau, \theta) = J_n(\tau, \theta) + \beta_n K$$

for  $K \in \{1, \dots, \overline{K}\}$  and  $(\tau, \theta)$  ranging over  $\mathcal{T}_K \times \Theta_K$  (recall that  $K^*$  is bounded above by  $\overline{K}$ , as a consequence of the definition of a  $\mathcal{F}$ -partition).

The sequence  $\{\beta_n\}$  is positive and tends to zero. The difficulty relies on the choice of its rate of convergence: since large  $\beta_n$  (precisely slow rate of convergence) favours simple models (that is models with small cardinality  $K$ ) and *vice versa*, the point is to calibrate its rate of convergence. Indeed,  $\beta_n$  appears as a trade-off between fitting the observations and avoiding too big models to be selected. Actually, the calibration will follow from the rate of convergence of the estimator  $(\hat{\tau}_n, \hat{\theta}_n)$  studied in the previous section, for *a priori* known  $K^*$ .

### 6.2. Consistency

**Theorem 6.1.** *Let  $\{\beta_n\}$  be a sequence of positive numbers satisfying both*

$$\beta_n \rightarrow 0 \quad \text{and} \quad n^{(2-h)/2(\sigma-1)} \beta_n \rightarrow \infty.$$

*Under the assumptions of Theorem 5.4,  $\hat{K}_n = K^*$  with P-probability tending to one. Hence, the consistency of  $(\hat{\tau}_n, \hat{\theta}_n)$  as defined in Theorem 5.2 still holds.*

*Proof.* Proof that  $P(\hat{K}_n < K^*)$  tends to zero is straightforward with Lemma 5.3. Indeed, if  $\hat{K}_n < K^*$ , then  $g(\hat{\tau}_n, \tau^*) \geq \Delta^*$  (see Property (iii)), hence  $u_n(\hat{\tau}_n, \hat{\theta}_n) \geq C^* \Delta^*$  with probability tending to one. Moreover,  $\hat{K}_n < K^*$  yields  $\beta_n(K^* - \hat{K}_n) \geq u_n(\hat{\tau}_n, \hat{\theta}_n) + e_n(\hat{\tau}_n, \hat{\theta}_n)$ . Thus (since  $\beta_n \rightarrow 0$ ) we can conclude if we control for  $n$  large enough the sum over  $K$  from 1 to  $K^* - 1$  of the probabilities that  $\sup\{|e_n(\tau, \theta)|\} > c$  for some  $c > 0$  (where  $(\tau, \theta)$  in the supremum ranges over  $\mathcal{T}_K \times \Theta_K$ ). This part is then complete, because  $e_n$  is uniformly  $o_P(1)$ .

Let us consider now  $P(\hat{K}_n > K^*)$ . It is bounded above by the sum over  $K$  from  $K^* + 1$  up to  $\overline{K}$  of probabilities that

$$\inf \{U_n(\tau, \theta) : (\tau, \theta) \in \mathcal{T}_K \times \Theta_K\} + \beta_n \leq 0.$$

For  $(\tau, \theta)$  ranging over  $\mathcal{T}_{K, \beta_n} \times \Theta_K$ , we can replace the previous events by

$$\inf \{U'_n(\tau) : \tau \in \mathcal{T}_{K, \beta_n}\} \leq 0$$

and proceed as in the first part of the proof of Theorem 5.4; and when it ranges over  $\mathcal{T}_K \times \Theta_{K, \beta_n}$ , proof is similar to its second part (that is why we impose  $n^{(2-h)/2(\sigma-1)} \beta_n \rightarrow \infty$ ). Thus, we have to focus on the  $P$ -probabilities of those events for  $(\tau, \theta)$  in  $\mathcal{T}_K \times \Theta_K$  satisfying  $g(\tau, \tau^*) \vee d_2(\theta, \theta^*) \leq \beta_n$ .

For  $(\tau, \theta)$  as described above, we get, applying Taylor–Lagrange’s inequality ( $b, c$  are some positive constants):

$$\begin{aligned} U_n(\tau, \theta) + \beta_n \geq e_n(\tau, \theta) + \beta_n &\geq \pi n^{-1} \sum_{j=1}^{K^*} \sum_{k \in K_j} \{n\beta_n - b\beta_n \|\Sigma_{X_1^n}(\tau_k \cap \tau_j^*)\|_\infty\} \\ &\quad + \pi n^{-1} \sum_{j=1}^{K^*} \sum_{k \notin K_j} \{n\beta_n - c \|\Sigma_{X_1^n}(\tau_k \cap \tau_j^*)\|_\infty\}. \end{aligned}$$

Finally, the proof will be complete if we show that the convergence to 0 as  $n \uparrow \infty$  of the probabilities of the following events hold for any  $c > 0$  and  $j_0 \neq j_1$  (where  $F$  denotes any set of the form  $\tau \cap \tau_{j_1}^*$  with  $\tau \in \mathcal{F}$  such that  $P(\tau \Delta \tau_{j_0}^*) \leq \delta_n$ , for  $\{\delta_n\} \downarrow 0$  driven from consistency):

$$\begin{aligned} \sup \{ \|\Sigma_{X_1^n}(F) : F \in \mathcal{F} \} &\geq cn, \\ \sup \{ \|\Sigma_{X_1^n}(F) : P(F) \leq \beta_n \} &\geq cn\beta_n. \end{aligned}$$

Once again, this is a direct consequence of Lemmas 4.1 and 4.4.  $\square$

## 7. APPENDIX

### 7.1. Proof of Proposition 3.3

*Proof.* Obviously, it suffices to prove that

$$\lim_{\eta \rightarrow \infty} \lim_{n \rightarrow \infty} P \left( \sup \left\{ \left| \frac{P(F) - P_n(F)}{P(F)} \right| : F \in \mathcal{F}, P(F) \geq \eta r_n \right\} \geq \frac{1}{2} \right) = 0.$$

Thanks to Talagrand's concentration inequalities for the supremum of empirical processes (see [18], Th. 2.4, p. 266), basic analysis yields an upper bound  $\exp\{-f(n, \eta)\}$  where  $f(n, \eta) > 0$  tends to infinity as  $n, \eta \uparrow \infty$ , as soon as the expectation of the supremum in the former equation goes to zero.

Let us first study the following expectation for fixed integers  $n, p$  and some  $\eta > 0$ :

$$E \left( \sup \left\{ \left| \frac{P(F) - P_n(F)}{P(F)} \right| : F \in \mathcal{F}_n^p \right\} \right),$$

for  $\mathcal{F}_n^p = \{F : F \in \mathcal{F}, 2^p \eta r_n \leq P(F) < 2^{p+1} \eta r_n\}$ . It is bounded above by

$$2^{-p} (\eta r_n)^{-1} E(\sup\{|P(F) - P_n(F)| : F \in \mathcal{F}_n^p\}).$$

Symmetrization arguments (refer to [18]) yield that the previous expression is bounded above by

$$2^{-p+1} (\eta n r_n)^{-1} E \left( \sup \left\{ \left| \sum_{i=1}^n \varepsilon_i \mathbb{1}\{X_i \in F\} \right| : F \in \mathcal{F}_n^p \right\} \right) \quad (8)$$

for independent identically distributed Rademacher rv  $\varepsilon_i$  (*i.e.*  $P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = 1/2$ ) independent of  $X_1^n$ . Furthermore, for any  $\mathcal{C} \subset \mathcal{F}$ , the next result holds (as a consequence of Hoeffding's inequality, see Problem 2.14.8 of [27]): for  $a = \sup_{F \in \mathcal{C}} P(F)$ ,  $V$  the VC-dimension of  $\mathcal{F}$  and some constant  $A(\mathcal{F})$  depending on  $\mathcal{F}$  only,

$$E \left( \sup \left\{ \left| \sum_{i=1}^n \varepsilon_i \mathbb{1}\{X_i \in F\} \right| : F \in \mathcal{C} \right\} \right) \leq C n^{1/2} \left[ \left( a + \frac{V}{n} \log \frac{V}{a} \right) \log \frac{A(\mathcal{F})}{a} \right]^{1/2}.$$

We apply this result together with the inequality  $(x + y)^{1/2} \leq x^{1/2} + y^{1/2}$  ( $x, y > 0$ ) to (8) and get that, for  $A' = A(\mathcal{F}) \vee V$ ,

$$E \left( \sup \left\{ \left| \frac{P(F) - P_n(F)}{P(F)} \right| : F \in \mathcal{F}_n^p \right\} \right) \leq C \sqrt{2}^{-p} \left( \frac{\log(A' (\eta r_n)^{-1})}{\eta n r_n} \right)^{1/2} + C 2^{-p} (2A')^{1/2} \frac{\log(A' (\eta r_n)^{-1})}{\eta n r_n}.$$



Finally, the expectation of the supremum of interest over  $\mathcal{F}_n$  is bounded above by the next sum that can be controlled as shown

$$\sum_{p \geq 0} E \left( \sup \left\{ \left| \frac{P(F) - P_n(F)}{P(F)} \right| : F \in \mathcal{F}_n^p \right\} \right) \leq C \left( \frac{\log \left( A' (\eta r_n)^{-1} \right)}{\eta n r_n} \right)^{1/2} + C A'^{1/2} \frac{\log \left( A' (\eta r_n)^{-1} \right)}{\eta n r_n}.$$

The result follows immediately. □

## 7.2. Exploring Assumption A6

*Two alternative assumptions*

We propose in this section to study Assumption **A6**. In order to make things clearer, we will suppose through this section that  $r = 1$ , that is that  $\xi$  is real valued. Thus,  $\|\cdot\|_\infty$  is systematically replaced by absolute values. The results are easy to adapt to the general case. Let us start with a very special case where slight control of the conditional second order moments suffices to ensure Assumption **A6**. State

**A6a:** There exist  $C_0 > 0$  and  $h \in (1, 2)$  such that, for any  $G \in \mathcal{G}$ ,

$$E \left( \Sigma_{X_1^n}(G)^2 \mid X_1^n \right) \leq C_0 \left( \sum_{i=1}^n \mathbb{1}\{X_i \in G\} \right)^h \quad P\text{-as.}$$

Note that **A6a** would obviously be satisfied for  $h = 1$  if the rv  $Z_i$  were independent. Moreover, Assumption **A6a** implies Assumption **A6** in the basic but fundamental case of rectangles:

**Proposition 7.1.** *Assumption A6 is satisfied as soon as A6a holds when  $\mathcal{F}_0$  is composed of rectangles.*

The proof of Proposition 7.1 relies on an adaptation of the method proposed in [19,20] to show such a result on the real line and its extension to the multidimensional case. It can be done by induction and uses then thoroughly basic properties of decomposition of rectangles in union of rectangles.

The sequel draws its inspiration from the theory of dependent variables and random fields, see [10,22] and especially [9]. Actually, a natural loosened conditional Marcinkiewicz–Zygmund inequality yields Assumption **A6** for VC class  $\mathcal{F}$ . Indeed, let Assumption **A6b** consists of the following:

**A6b:** There exist  $C_0 > 0$  and  $h \in [1, 2)$  such that, for any  $p > 2$  and  $G \in \mathcal{G}$ ,

$$E \left( |\Sigma_{X_1^n}(G)|^p \mid X_1^n \right) \leq C_0^p p^{\frac{h}{2}} \left( \sum_{i=1}^n \mathbb{1}\{X_i \in G\} \right)^{h \frac{p}{2}} \quad P\text{-as.}$$

**Remark 7.2** (on Assumption **A6b**). The previous inequality is said to be “loosened” because of the power  $h$  in the right hand term, where  $h = 1$  is usually expected. It is sharp an inequality thanks to the particular form of the factor  $C_0^p p^{\frac{h}{2}}$ : it allows some efficient optimization in  $p$  producing the expected final result *via* Pisier’s method for some rich class  $\mathcal{F}$  – namely finite VC-dimension VC class. Precise statement is given by Proposition 7.3. It underlines how exceptional seems Proposition 7.1, where only control of second order moments is needed, to compare with Assumption **A6b** and control of moments of order any  $p > 2$ . It is known that such a simple condition of control of moments can not be sufficient in the simple case of polygons.

**Proposition 7.3.** *Assumption A6 is satisfied as soon as Assumption A6b holds when  $\mathcal{F}$  is a finite VC-dimension VC class.*

*Proof.* All the inequalities to come hold  $P$ -as. We set  $\Sigma_n(G)$  for  $\Sigma_{X_1^n}(G)$  and  $E_n$  for  $E(\cdot \mid X_1^n)$ .  $\|\cdot\|_p$  denotes the  $L^p$  norm with respect to the conditional probability  $P(\cdot \mid X_1^n)$ .

Pisier’s method (see [9]) consists of writing

$$\begin{aligned} E_n \left( \sup_{F \in \mathcal{F}(G)} |\Sigma_n(F)|^2 \right) &\leq \left\| \sup_{F \in \mathcal{F}(G)} \|\Sigma_n(F)\|_p^2 \right\| \leq \left( \sum_{F \in \Gamma_n} E_n (|\Sigma_n(F)|^p) \right)^{\frac{2}{p}} \\ &\leq N_n^{\frac{2}{p}} \max_{F \in \Gamma_n} \|\Sigma_n(F)\|_p^2 \leq N_n^{\frac{2}{p}} p C_0^2 \left( \sum_{i=1}^n \mathbb{1}\{X_i \in G\} \right)^h, \end{aligned}$$

where  $\Gamma_n$  denotes a family of sets of minimal cardinality  $N_n$  such that any separation of  $\{X_1, \dots, X_n\} \cap G$  by elements of  $\mathcal{F}(G)$  can be achieved with an element of  $\Gamma_n$ . Since  $\mathcal{F}$  has finite VC-dimension, so does  $\mathcal{F}(G)$  and  $N_n$  is finite. Set  $H_n = \log N_n$ : VC theory (see for instance [28]) ensures that  $H_n$  is bounded above by  $V(1 + \log \sum \mathbb{1}\{X_i \in G\})$  ( $V$  denotes the VC-dimension of  $\mathcal{F}$ ).

Optimization in  $p$  yields

$$\begin{aligned} E_n \left( \sup_{F \in \mathcal{F}(G)} |\Sigma_n(F)|^2 \right) &\leq 2C_0^2 H_n \left( \sum_{i=1}^n \mathbb{1}\{X_i \in G\} \right)^h \\ &\leq 2C_0^2 V \left( \sum_{i=1}^n \mathbb{1}\{X_i \in G\} \right)^h \left( 1 + \log \sum_{i=1}^n \mathbb{1}\{X_i \in G\} \right). \end{aligned}$$

Set  $\varepsilon > 0$  such that  $h + \varepsilon < 2$ : there exists  $C_1$  depending on  $\varepsilon$  but neither on  $G$  nor on  $n$  such that

$$E_n \left( \sup_{F \in \mathcal{F}(G)} |\Sigma_n(F)|^2 \right) \leq C_1 \left( \sum_{i=1}^n \mathbb{1}\{X_i \in G\} \right)^{h+\varepsilon}$$

and Markov’s inequality concludes the proof. □

*Assumptions **A6a** and **A6b** for regular lattices*

We propose here to illustrate both Assumption **A6a** and Assumption **A6b** in a simple but nevertheless natural and interesting case where  $\mathcal{X} = \mathbb{Z}^d$  is a regular lattice and the field  $(Z_x)_{x \in \mathcal{X}}$  is bounded and strictly stationary. The sequel is widely inspired from [9] again. Our aim is to determine some ultimate assumptions that imply both **A6a** and **A6b**.

First, let us recall the definition of the  $\phi$ -mixing coefficient  $\phi(\mathcal{U}, \mathcal{V})$  of two  $\sigma$ -algebras  $\mathcal{U}$  and  $\mathcal{V}$  of  $\mathcal{A}$ : it is given by

$$\phi(\mathcal{U}, \mathcal{V}) = \sup \{ \|P(V|\mathcal{U}) - P(V)\|_\infty : V \in \mathcal{V} \}.$$

Note that  $\phi(\mathcal{U}, \mathcal{V}) \in [0, 1]$ , with value 0 for independent  $\sigma$ -fields only. Let us introduce the nonincreasing sequence  $\{\phi(t)\}_{t \geq 1}$  with definition

$$\phi(t) = \sup \{ \phi(\sigma(Z_y, y \in \mathcal{Y}), \sigma(Z_x)) : x \in \mathcal{X}, \mathcal{Y} \subset \mathcal{X}, d(x, \mathcal{Y}) \geq t \},$$

where  $\mathcal{Y}$  denotes any finite subset of  $\mathcal{X}$  and  $d(x, \mathcal{Y})$  is the infimum for  $y$  ranging throughout  $\mathcal{Y}$  of  $d(x, y) = \min_{1 \leq i \leq d} |x_i - y_i|$ . Its role is to resume the whole system of dependency of the field  $(Z_x)_{x \in \mathcal{X}}$ .

We can state now the final result of this section and infer from it a sufficient rate of convergence to 0 for  $\{\phi(t)\}$  to ensure that Assumptions **A6a** and **A6b** hold true:

**Proposition 7.4.** *Suppose that  $\mathcal{X} = \mathbb{Z}^d$  and that  $(Z_x)_{x \in \mathcal{X}}$  is bounded and strictly stationary. Assumptions **A6a** and **A6b** are satisfied as soon as, for some  $1 \leq h < 2$ ,  $C_0 > 0$  depending on  $h$  and for any  $n \geq 1$ ,*

$$\sum_{t=1}^n t^{d-1} \phi(t) \leq C_0 n^{h-1}. \tag{9}$$

The previous inequality is satisfied e.g. when  $\phi(t) = O(t^{-(d+1-h)})$ .

Proposition 7.4 is a corollary of Proposition 7.6 below. The strategy of the proof relies on a Burkholder-like inequality as shown by Dedecker in [9], namely in Proposition 1(a) of this paper. Following his method, we get that, for any  $G \in \mathcal{G}$  and  $\mathcal{Y}_n \subset \mathcal{X}$  of cardinality  $n$ ,

$$\begin{aligned} E(|\Sigma_{\mathcal{Y}_n}(G)|^p) &\leq \left( 2p \sum_{x \in \mathcal{Y}_n} \left\{ \|Z_x^2\|_{p/2} + \sum_{x' \in \mathcal{Y}_n} \|Z_{x'} E_{d(x,x')}(Z_x)\|_{p/2} \mathbb{1}\{x' \in G\} \right\} \mathbb{1}\{x \in G\} \right)^{p/2} \\ &\leq (2p)^{p/2} \left( \|Z_0\|_\infty + \|Z_0\|_\infty^2 \right) \\ &\quad \left( \sum_{x \in \mathcal{Y}_n} \mathbb{1}\{x \in G\} + \sum_{x,x' \in \mathcal{Y}_n} \|E_{d(x,x')}(Z_x)\|_{p/2} \mathbb{1}\{x,x' \in G\} \right)^{p/2}. \end{aligned} \tag{10}$$

In the previous display,  $E_{d(x,x')}(Z_x)$  denotes the conditional expectation of  $Z_x$  with respect to the  $\sigma$ -field  $\sigma(Z_y : y \in \mathcal{Y}_n, d(x,y) \geq d(x,x'))$ .

We can derive from (10) some first ultimate condition on the field  $(Z_x)_{x \in \mathcal{X}}$ :

**Proposition 7.5.** *Suppose that  $\mathcal{X} = \mathbb{Z}^d$  and that  $(Z_x)_{x \in \mathcal{X}}$  is centered, bounded and strictly stationary. Then Assumptions **A6a** and **A6b** hold as soon as, for some  $1 \leq h < 2$ ,  $C_0 > 0$  depending on  $h$  and for any  $p \geq 2$ ,  $G \in \mathcal{G}$ ,  $n \geq 1$  and  $\mathcal{Y}_n \subset \mathcal{X}$  of cardinality  $n$ , for any  $x \in \mathcal{Y}_n$ ,*

$$\sum_{x' \in \mathcal{Y}_n} \|E_{d(x,x')}(Z_x)\|_{p/2} \leq C_0 n^{h-1}.$$

The condition above holds with  $h = 1$  for  $m$ -conditionally centered fields (i.e. fields such that  $E_{d(x,x')}(Z_x) = 0$  for  $d(x,x') \geq m$ ). This includes  $m$ -dependence and consequently, independence.

Furthermore, combining the next upper bound for the right hand term of (10)

$$(2p)^{p/2} \left( \|Z_0\|_\infty + \|Z_0\|_\infty^2 \right) \left( \sum_{x \in \mathcal{Y}_n} \mathbb{1}\{x \in G\} + \sum_{x,x' \in \mathcal{Y}_n} \|E_{d(x,x')}(Z_x)\|_\infty \mathbb{1}\{x,x' \in G\} \right)^{p/2}$$

with Serfling’s inequality (see [24])

$$\|E_{d(x,x')}(Z_x)\|_\infty \leq 2\phi(d(x,x'))\|Z_x\|_\infty,$$

yields:

**Proposition 7.6.** *Suppose that  $\mathcal{X} = \mathbb{Z}^d$  and that  $(Z_x)_{x \in \mathcal{X}}$  is centered, bounded and strictly stationary. Then Assumptions **A6a** and **A6b** hold as soon as, for some  $1 \leq h < 2$ ,  $C_0 > 0$  depending on  $h$  and for any  $G \in \mathcal{G}$ , any  $n \geq 1$  and any  $\mathcal{Y}_n \subset \mathcal{X}$  of cardinality  $n$ , for any  $x \in \mathcal{Y}_n$ ,*

$$\sum_{x' \in \mathcal{Y}_n} \|E_{d(x,x')}(Z_x)\|_\infty \leq C_0 n^{h-1}.$$

Condition (9) in Proposition 7.4 is sufficient.

### 7.3. Sketch of proof of Lemma 5.3

*Sketch of the proof.* First, define  $f_{ij}(\theta^*, \alpha) = \inf\{\alpha v(\theta_j^*, \theta) + (1 - \alpha)v(\theta_i^*, \theta) : \theta \in \Theta\}$  ( $1 \leq i \neq j \leq K^*$ ,  $0 < \alpha < 1$ ) and  $A_{ij}(\theta^*) = 2f_{ij}(\theta^*, 1/2)$ ,  $\underline{A} = \min A_{ij}(\theta^*)$ : then  $\underline{A} > 0$  and  $f_{ij}(\theta^*, \alpha) \geq \underline{A} \min(\alpha, 1 - \alpha)$ .

Set  $\varepsilon > 0$  and define the events  $\Omega_n(\delta) = \bigcap_{l \geq n} [\sup\{|P_l(F) - P(F)| : F \in \mathcal{F}\} \leq \delta]$  (any  $\delta > 0$ ,  $n \geq 1$ ). Since  $P(\Omega_n(\delta)) \uparrow 1$  as  $n \uparrow \infty$ , there exists  $n_1 \geq 1$  such that  $P(\Omega_{n_1}(\delta_1)) \geq 1 - \varepsilon$  with  $\delta_1 = \Delta^*/16\overline{K}^2$ . Let us restrict ourselves to the event  $\Omega_{n_1}(\delta_1)$  and consider  $(\tau, \theta)$  in  $\mathcal{T}_{K, \Delta^*/4\overline{K}} \times \Theta_K$ ,  $n \geq n_1$ . We can prove that  $u_n(\tau, \theta)$  is bounded from below by a positive constant independent of  $n$  and  $(\tau, \theta)$ . Indeed, we have (applying Prop. 2.4(i)) the existence of  $k, j_0, j_1$  such that  $n_{kj_0}/n$  and  $n_{kj_1}/n$  are both bounded from below by such a constant  $c$ , and then, for  $\alpha = n_{kj_0}/(n_{kj_0} + n_{kj_1})$ ,

$$u_n(\tau, \theta) \geq \frac{n_{kj_0} + n_{kj_1}}{n} (\alpha v(\theta_{j_0}^*, \theta_k) + (1 - \alpha)v(\theta_{j_1}^*, \theta_k)) \geq c \underline{A}.$$

Then, since  $g(\tau, \tau^*) \leq 1$  and  $v$  is bounded from above by its supremum over the compact set  $\overline{\Theta} \times \overline{\Theta}$ , the study of the case  $g(\tau, \tau^*) > \Delta^*/4\overline{K}$  is completed.

Let us deal now with the lower bounding in  $d_v(\theta, \theta^*)$  for the models  $(\tau, \theta)$  ranging over  $(\mathcal{T}_K - \mathcal{T}_{K, \Delta^*/4\overline{K}}) \times \Theta_K$ . Since

$$u_n(\tau, \theta) \geq \max_{1 \leq j \leq K^*} \max_{k \in \mathcal{K}_j} \frac{n_{kj}}{n} v(\theta_j^*, \theta_k),$$

it suffices to bound by below all  $P(\tau_k \cap \tau_j^*)$ 's by some positive constant independent of  $n$  and  $(\tau, \theta)$  and greater than  $\delta_1$ . We do so thanks to Proposition 2.4(i) again.

Set  $\delta > 0$ . We still have to show that  $u_n(\tau, \theta) \geq C^* g(\tau, \tau^*)$  for  $\tau$  verifying  $\delta < g(\tau, \tau^*) \leq \Delta^*/4\overline{K}$  and  $\theta \in \Theta_K$ , with probability larger than  $1 - \varepsilon$ . First, as above, there exists  $n_0 \geq n_1$  such that  $P(\Omega_{n_0}(\delta_0)) \geq 1 - \varepsilon$ , for  $\delta_0 = \min(\delta_1, \delta/2)$ . Let us restrict ourselves to the event  $\Omega_{n_0}(\delta_0)$  from now. Set  $(\tau, \theta) \in \mathcal{T}_{K, \delta} \times \Theta_K$  with  $g(\tau, \tau^*) \leq \Delta^*/4\overline{K}$ . Suppose that  $g(\tau, \tau^*)$  is achieved for  $j_0$  and  $\mathcal{K}_{j_0}$ . Since then

$$\sum_{k \notin \mathcal{K}_{j_0}} \frac{n_{kj_0}}{n} + \sum_{j \neq j_0} \sum_{k \in \mathcal{K}_j} \frac{n_{kj}}{n} \geq g(\tau, \tau^*)/2,$$

it is sufficient to prove that  $u_n(\tau, \theta)$  is greater than the left hand term of the preceding inequation, up to some positive multiplicative constant independent of both  $n$  and  $(\tau, \theta)$ . Careful writing and Proposition 2.4(ii) yields the result.  $\square$

The author would like to thank warmly his two supervisors Elisabeth Gassiat and Marc Lavielle. Many thanks to Jérôme Dedecker for his clear, helpful introduction to mixing. The author is grateful to one of the referees whose remarks helped to improve the presentation.

## REFERENCES

- [1] H. Akaike, A new look at the statistical model identification. *IEEE Trans. Automat. Control* **AC-19** (1974) 716-723. System identification and time-series analysis.
- [2] A. Antoniadis, I. Gijbels and B. MacGibbon, Non-parametric estimation for the location of a change-point in an otherwise smooth hazard function under random censoring. *Scand. J. Statist.* **27** (2000) 501-519.
- [3] Z.D. Bai, C.R. Rao and Y. Wu, Model selection with data-oriented penalty. *J. Statist. Plann. Inference* **77** (1999) 103-117.
- [4] A. Barron, L. Birgé and P. Massart, Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** (1999) 301-413.
- [5] M. Basseville and I.V. Nikiforov, *Detection of abrupt changes: Theory and application*. Prentice Hall Inc. (1993).
- [6] B.E. Brodsky and B.S. Darkhovsky, *Nonparametric methods in change-point problems*. Kluwer Academic Publishers Group (1993).
- [7] E. Carlstein, H.-G. Müller and D. Siegmund, *Change-point problems*. Institute of Mathematical Statistics, Hayward, CA (1994). Papers from the AMS-IMS-SIAM Summer Research Conference held at Mt. Holyoke College, South Hadley, MA July 11-16, 1992.
- [8] D. Dacunha-Castelle and E. Gassiat, The estimation of the order of a mixture model. *Bernoulli* **3** (1997) 279-299.
- [9] J. Dedecker, Exponential inequalities and functional central limit theorems for random fields. *ESAIM P&S* **5** (2001) 77.
- [10] P. Doukhan, *Mixing*. Springer-Verlag, New York (1994). Properties and examples.
- [11] M. Lavielle, *On the use of penalized contrasts for solving inverse problems. Application to the DDC (Detection of Divers Changes) problem* (submitted).
- [12] M. Lavielle, Detection of multiple changes in a sequence of dependent variables. *Stochastic Process. Appl.* **83** (1999) 79-102.
- [13] M. Lavielle and E. Lebarbier, An application of MCMC methods for the multiple change-points problem. *Signal Process.* **81** (2001) 39-53.
- [14] M. Lavielle and C. Ludeña, The multiple change-points problem for the spectral distribution. *Bernoulli* **6** (2000) 845-869.
- [15] M. Lavielle and E. Moulines, Least-squares estimation of an unknown number of shifts in a time series. *J. Time Ser. Anal.* **21** (2000) 33-59.
- [16] G. Lugosi, *Lectures on statistical learning theory*. Presented at the Garchy Seminar on Mathematical Statistics and Applications, available at <http://www.econ.upf.es/~lugosi> (2000).
- [17] E. Mammen and A.B. Tsybakov, Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** (1995) 502-524.
- [18] P. Massart, Some applications of concentration inequalities to statistics. *Ann. Fac. Sci. Toulouse Math. (6)* **9** (2000) 245-303.
- [19] F. Móricz, A general moment inequality for the maximum of the rectangular partial sums of multiple series. *Acta Math. Hungar.* **41** (1983) 337-346.
- [20] F.A. Móricz, R.J. Serfling and W.F. Stout, Moment and probability bounds with quasisuperadditive structure for the maximum partial sum. *Ann. Probab.* **10** (1982) 1032-1040.
- [21] V.V. Petrov, *Limit theorems of probability theory*. The Clarendon Press Oxford University Press, New York (1995). Sequences of independent random variables, Oxford Science Publications.
- [22] E. Rio, *Théorie asymptotique des processus aléatoires faiblement dépendants*. Springer (2000).
- [23] G. Schwarz, Estimating the dimension of a model. *Ann. Statist.* **6** (1978) 461-464.
- [24] R.J. Serfling, Contributions to central limit theory for dependent variables. *Ann. Math. Statist.* **39** (1968) 1158-1175.
- [25] M. Talagrand, New concentration inequalities in product spaces. *Invent. Math.* **126** (1996) 505-563.
- [26] A.W. van der Vaart, *Asymptotic statistics*. Cambridge University Press (1998).
- [27] A.W. van der Vaart and J.A. Wellner, *Weak convergence and empirical processes*. Springer-Verlag, New York (1996). With applications to statistics.
- [28] V.N. Vapnik, *Statistical learning theory*. John Wiley & Sons Inc., New York (1998).
- [29] Y.-C. Yao, Estimating the number of change-points via Schwarz's criterion. *Statist. Probab. Lett.* **6** (1988) 181-189.