A CHANGE-POINT PROBLEM AND INFERENCE FOR SEGMENT SIGNALS

VICTOR-EMMANUEL BRUNEL*

Abstract. We address the problem of detection and estimation of one or two change-points in the mean of a series of random variables. We use the formalism of set estimation in regression: to each point of a design is attached a binary label that indicates whether that point belongs to an unknown segment and this label is contaminated with noise. The endpoints of the unknown segment are the change-points. We study the minimal size of the segment which allows statistical detection in different scenarios, including when the endpoints are separated from the boundary of the domain of the design, or when they are separated from one another. We compare this minimal size with the minimax rates of convergence for estimation of the segment under the same scenarios. The aim of this extensive study of a simple yet fundamental version of the change-point problem is two-fold: understanding the impact of the location and the separation of the change points on detection and estimation and bringing insights about the estimation and detection of convex bodies in higher dimensions.

1. INTRODUCTION

1.1. The change-point problem

Change-point problems have been studied extensively, especially in time series analysis, where the goal is to detect or estimate breakpoints in the distribution of ordered observations. The breakpoints can occur in the mean [2, 14, 16, 20–22, 26] and/or in the variance [5, 23], in a location parameter [1], in the tail of the distribution [11], in a general parameter of the distribution [18], or the whole distribution can change at the breakpoints [3]. For more details on the account of change-point problems in time series analysis, we refer to (see [2, 24] and the references therein. Here, we focus on change-points in ordered data that do not necessarily come from time series. Then, if the change-points occur in the mean, this problem can be stated in terms of inference on breakpoints of a regression function. A general problem has been addressed in [8] and includes cases where breakpoints occur in the mean: A sample $Y_1, \ldots, Y_n$ is observed, where Y_i has a density $f(\cdot, \phi(\frac{i}{n}))$ with respect to a given measure, for $i = 1, \ldots, n$. The map f belongs to some parametric class of densities. The real valued function ϕ is assumed to be piecewise constant on $[0, 1]$, with a finite number K of jumps (called the change points, or breakpoints), where K is not necessarily known. It is shown that at least one change-point can be detected consistently. When $f(\cdot, \mu)$ is the Gaussian density with mean μ and given variance $\sigma^2 > 0$, the

problem was addressed in [17]. That problem can also be interpreted as estimating the jumps of a regression function. Its simplest form, with only one change-point, reads as follows:

$$Y_i = \mathbb{1}(X_i \leq \theta) + \xi_i, i = 1, \ldots, n, \tag{1.1}$$

where $X_1, \ldots, X_n$ are given numbers (possibly random) in $[0, 1]$, $\xi_1, \ldots, \xi_n$ are i.i.d. random variables independent of the X_i's and $\theta \in [0, 1]$ is the change-point (or breakpoint).

Model (1.1) was studied in Section 1.9 from [16] and a continuous-time version was addressed in [14], where the aim is to estimate the breakpoint θ. In the continuous-time version, Korostelev [14] focused on a more general framework, where the regression function has a jump but is not necessarily an indicator function. In [14, 16], the change-point θ is estimated with an expected accuracy of order $1/n$. Ibragimov and Khasminskii [12, 13] defined a consistent estimator of the discontinuity point of a regression function, with rate $1/n$ as well. However, a key assumption in all these works is that the change-point is separated from the boundaries of the domain: $h \leq \theta \leq 1 - h$, for some $h \in (0, 1/2)$.

The separation assumption from the boundaries of the domain is also made in higher dimensional problems. For instance, in [15] and Chapter 3 from [16], a boundary fragment (which plays the higher dimensional role of $[0, \theta]$ in (1.1)) is estimated, under the assumption that its edge function is uniformly separated from 0 and 1.

As part of this work, we prove that this separation assumption is only technical and that the estimation rate $1/n$ (with no extra logarithmic factor) is achieved without separation in model (1.1). Our focus is a one-dimensional model, where the number of breakpoints is known and is either one or two, which allows an interpretation of the model in terms of detection and estimation of segments.

More generally, we consider the following statistical model:

$$Y_i = \mathbb{1}(X_i \in G) + \xi_i, i = 1, \ldots, n, \tag{1.2}$$

where $\mathbb{1}(\cdot)$ is the indicator function. The collection $\mathcal{X} = \{X_1, \ldots, X_n\}$ is called the design and it is observed, as well as the labels $Y_1, \ldots, Y_n$. The unknown set G is a segment on $[0, 1]$ and the noise terms ξ_i are unobserved i.i.d. random variables, independent of the design.

Throughout the paper, we assume that the noise terms $\xi_1, \ldots, \xi_n$ from model (1.2) are subgaussian, *i.e.*, satisfy

$$\mathbb{E}\left[e^{u\xi_i}\right] \leq e^{\sigma^2 u^2/2}, \quad \forall u \in \mathbb{R}, \quad i = 1, \ldots, n, \tag{1.3}$$

for some positive constant $\sigma > 0$. This constant need not be known. Note that (1.3) implies that the noise terms have mean zero. If they are centered Gaussian random variables, then they satisfy (1.3) with $\sigma^2 = \mathsf{Var}(\xi_1)$.

Since the design and the noise are assumed to be independent, reordering the X_i's does not affect the model. Indeed, there exists a reordering $\{i_1, \ldots, i_n\}$ of $\{1, \ldots, n\}$, such that $X_{i_1} \leq \ldots \leq X_{i_n}$. The random indices $i_1, \ldots, i_n$ are independent of the noise, therefore, the new noise vector $(\xi_{i_1}, \ldots, \xi_{i_n})$ has the same distribution as $(\xi_1, \ldots, \xi_n)$. Hence, we assume from now on that $\mathcal{X}$ is the reordering of a preliminary design, *i.e.*, $X_1 \leq \ldots \leq X_n$ almost surely, without loss of generality. We distinguish two types of designs:

(DD) Deterministic, regular design: $X_i = i/n, i = 1, \ldots, n$;
(RD) Random, uniform design: the variables $X_i, i = 1, \ldots, n$, are the reordering of i.i.d. uniform random variables in $[0, 1]$.

Other designs are considered in the literature (*e.g.*, see [16], Sect. 1.9) but we prefer to restrict ourselves to the designs (DD) and (RD), which yield straightforward extensions to other types of designs. Note that model (1.2) can also be interpreted as a nonlinear regression, where we do inference on the support G of the regression function. Actually, it is the one-dimensional version of the model studied in [4], where G plays the role of a one-dimensional convex body. In [4], it is explained that hardness of estimation of G, in model (1.2), can be explained by two factors. The first factor is the complexity of the class of possible candidates G and the

second one is that G might be too small to be detected by any statistical procedure. In the present case, G is a segment, hence the corresponding class is parametric and one can hope to estimate G at the fast speed $1/n$, up to a positive multiplicative constant. However, we show that surprisingly, if the class contains arbitrarily short segments, the speed of estimation may be deteriorated. We try to understand what type of scenarios allow estimation of G at the fast rate $1/n$ and under what other scenarios the best rate of estimation of G is significantly worse than $1/n$, in a minimax sense. In particular, we study the two following assumptions on G, where $\mu \in (0, 1)$ is fixed throughout the paper.

Assumption 1. $G = [0, \theta]$, for some unknown number $0 \leq \theta \leq 1$.

This assumption carries information about the location of G. In terms of the change-point problem, it implies that there is only one change-point and the problem becomes equivalent to (1.1), where no separation of θ from 0 and 1 is imposed.

The second assumption implies that G is not too short, *i.e.*, that the two change-points are separated from one another:

Assumption 2. $|G| \geq \mu$, where $\mu \in (0, 1)$ is a known positive number.

Here, $|G|$ is the length of the segment G.

Model (1.2) with random design has been addressed in higher dimensions in [4], where G is assumed to be a convex polytope with fixed number of vertices. In that work, there were no assumptions of the type of 1 or 2 and the rate of estimation was $(\ln n)/n$, which, we believe, would become faster under similar assumptions as Assumptions 1 or 2. Further details on this account are given in the discussion in Section 4.

The detection problem consists of testing whether $G = \emptyset$ in model (1.2), *i.e.*, whether there is no change-point. It is addressed, for example, in [6], where, unlike here, the authors do not assume the strength of the signal to be known:

$$Y_i = \delta \mathbb{1}(X_i \in G) + \xi_i, i = 1, \ldots, n, \tag{1.4}$$

where δ is an unknown positive number. For the signal to be detectable, there should be a tradeoff between its length $|G|$ and its strength δ. Intuitively, if δ is small, then the set G should be big enough and conversely, if δ is large, the set G can be short and the signal still be detected. In that framework, testing the presence of a signal reduces to decide whether $\delta = 0$, which makes the problem different from ours, where we know δ and impose $\delta = 1$. In [6], the authors mainly study the power of two tests under design (DD): the scan – or maximum – likelihood ratio and the average likelihood ratio. These are compared in two regimes: signals of small scales, *i.e.* $|G| \longrightarrow 0$, when $n \to \infty$ and signals of large scales, *i.e.* $\liminf_{n \to \infty} |G| > 0$. It is proved that if $\delta \sqrt{n|G|} \geq \sqrt{2 \ln \frac{1}{|G|} + b_n}$, for some sequence b_n such that $b_n \longrightarrow \infty$, then there is a test with asymptotic power 1. For fixed δ, Chan and Walther's condition implies that $|G|$ must be of order at least $(\ln n)/n$. We prove a similar condition when $\delta = 1$ is known. This means that knowing δ does not make the detection problem easier and the logarithmic factor in Chan and Walther's condition is not due to adaptation to δ.

Note that $\delta \sqrt{|G|}$ is exactly the L^2-norm of the signal. In [19], signals of unknown shape but known smoothness were considered. The authors test whether the observations are pure noise and give exact minimax separation rates in terms of the L^2-norm of the signal. Detection is harder in that framework, because unlike in model (1.2) or (1.4), the shape of the signal is unknown and only its smoothness is known. This is why the separation rates are larger than those corresponding to models (1.2) and (1.4), in the sense that they allow less freedom for the size of the signal to be detected. A similar detection problem has also been studied in [7], where three cases are considered: either the shape of the signal (up to an affine transform), or its smoothness is known, or nothing is known. However, this problem is different from ours, since we are only concerned with the location of the signal, not the signal itself. This is why, and also for the sake of simplicity, we only deal with signals of known shape and amplitude in the present work.

1.2. Definitions and notation

If G is a segment in $[0,1]$, we denote by $\mathbb{P}_G$ the joint probability measure of the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ that satisfy model (1.2) and by $\mathbb{E}_G$ and $\mathbb{V}_G$ the corresponding expectation and variance operators. We may omit the subscript G and write only $\mathbb{P}$, $\mathbb{E}$ or $\mathbb{V}$ if there is no ambiguity.

If G_1 and G_2 are two segments in $[0,1]$, we denote by $G_1 \Delta G_2$ their symmetric difference. Its Lebesgue measure $|G_1 \Delta G_2|$ is also called the Nykodim distance between G_1 and G_2.

An estimator of G is a segment (possibly empty) $\hat{G}_n$ of $[0,1]$, whose construction depends on the observations. We measure the accuracy of an estimator $\hat{G}_n$ in a minimax framework. The risk of $\hat{G}_n$ on a class $\mathcal{C}$ of segments is defined as

$$\mathcal{R}_n(\hat{G}_n; \mathcal{C}) = \sup_{G \in \mathcal{C}} \mathbb{E}_G[|G \Delta \hat{G}_n|]. \tag{$*$}$$

The rate (a sequence which depends on n) of an estimator on a class $\mathcal{C}$ is the speed at which its risk converges to zero when the number n of available observations tends to infinity. The minimax risk on a class $\mathcal{C}$, when n observations are available, is defined as

$$\mathcal{R}_n(\mathcal{C}) = \inf_{\tilde{G}_n} \mathcal{R}_n(\hat{G}_n; \mathcal{C}), \tag{$**$}$$

where the infimum is taken over all estimators $\tilde{G}_n$ depending on n observations. If $\mathcal{R}_n(\mathcal{C})$ converges to zero, we call the minimax rate of convergence on the class $\mathcal{C}$ the speed at which $\mathcal{R}_n(\mathcal{C})$ tends to zero.

For the detection problem, let h be a positive number, that may depend on n. We consider the following hypotheses:

$$\begin{cases} H_0 : G = \emptyset \text{ (the null hypothesis)}, \\ H_1 : |G| \geq h \text{ (the alternative hypothesis)}. \end{cases}$$

The performance of a test $\tau_n \in \{0,1\}$ on a class $\mathcal{C}$ is measured by the sum of its type one and two errors, *i.e.*,

$$\gamma_n(\tau_n, \mathcal{C}) = \mathbb{P}_\emptyset[\tau_n = 1] + \sup_{G \in \mathcal{C}, |G| \geq h} \mathbb{P}_G[\tau_n = 0].$$

We say that τ_n is consistent on the class $\mathcal{C}$ if and only if $\gamma_n(\tau_n, \mathcal{C}) \longrightarrow 0$ when $n \to \infty$. We call the separation rate on the class $\mathcal{C}$ any sequence of positive numbers r_n such that:

- if $\dfrac{h}{r_n} \underset{n \to \infty}{\longrightarrow} \infty$, then there exists a consistent test on $\mathcal{C}$ and
- if $\dfrac{h}{r_n} \underset{n \to \infty}{\longrightarrow} 0$, then no test is consistent test on $\mathcal{C}$.

We define three different classes of segments:

- $\mathcal{S} = \{[a,b] : 0 \leq a \leq b \leq 1\}$ is the class of all segments on $[0,1]$,
- $\mathcal{S}_0 = \{[0,\theta] : 0 \leq \theta \leq 1\}$ is the class of all segments on $[0,1]$, satisfying Assumption 1,
- $\mathcal{S}(\mu) = \{G \in \mathcal{S} : |G| \geq \mu\}$ is the class of all segments on $[0,1]$, satisfying Assumption 2.

For two real valued sequences A_n and B_n and a parameter ϑ, which may be multidimensional, we will write $A_n \asymp_\vartheta B_n$ when there exist positive constants $c(\vartheta)$ and $C(\vartheta)$, which depend on ϑ only, such that $c(\vartheta)B_n \leq A_n \leq C(\vartheta)B_n$, for n large enough. If we put no subscript under the sign $\asymp$, this means that the corresponding constants are universal, *i.e.*, do not depend on any parameter.

1.3. Outline and contributions

In Section 2, we tackle the detection problem. We prove that the separation rates are at least $1/n$ for the class $\mathcal{S}_0$ and $(\ln n)/n$ for $\mathcal{S}$ and that these are exactly the separation rates when the noise is Gaussian (Thm. 2.1). In Section 3, we estimate the unknown set and we prove minimax rates of on the classes $\mathcal{S}$, $\mathcal{S}_0$ and $\mathcal{S}(\mu)$. Specifically, we show that the breakpoint θ in (1.1) can be estimated at the speed $1/n$ without imposing separation from 0 and 1: we prove that the minimax rate of convergence on the class $\mathcal{S}_0$ is $1/n$ (Thm. 3.5). We also show that recovering a general segment (or, equivalently, two change-points) with a uniform risk of order $1/n$ is possible only if it is known *a priori* that the unknown set is not too small (Thm. 3.9). Otherwise, a logarithmic factor appears in the minimax rate (Thm. 3.14). For $G \in \mathcal{S}(\mu)$, we define a two step estimator inspired by [14]. We first locate G with high probability using the first half of the sample and we apply the same ideas as in the class $\mathcal{S}_0$ to estimate the endpoints of G separately, using the second half of the sample.

In Section 4, we draw conclusions, followed by a discussion about possible extensions and all the proofs are deferred to Section 5.

2. DETECTION OF A SET

In this section, we find separation rates for testing $H_0 : G = \emptyset$ against $H_1 : |G| \geq h$, when G belongs to some classes of segments.

The idea for the class $\mathcal{S}_0$ is the following. Under H_1, since $G \in \mathcal{S}_0$, we necessarily have that $[0, h] \subseteq G$. Therefore, we check among those pairs (X_i, Y_i) for which $X_i \leq h$ if there are sufficiently many Y_i's that are large, *e.g.* larger than $1/2$. Let $N = \max\{i = 1, \ldots, n : X_i \leq h\} = \#(\mathcal{X} \cap [0, h])$, where $\#$ stands for cardinality. Let S be the following test statistic:

$$ S = \# \left\{ i = 1, \ldots, N : Y_i \leq \frac{1}{2} \right\}. $$

If the alternative hypothesis is true, *i.e.* if $|G| \geq h$, then all the X_i's, $i \leq N$, fall inside G and the corresponding Y_i's should not be too small. We define the test $T_n^0 = \mathbb{1}(S \leq cN)$, where c is any number strictly between $\mathbb{P}[\xi_1 \leq -1/2])$ and $\mathbb{P}[\xi_1 \leq 1/2]$. Note that this definition requires some knowledge about the noise distribution. In many cases, it may be reasonable to assume that the noise is symmetric, hence to take $c = 1/2$. However, in general, it is not clear how to calibrate c if no information about the noise is available.

For the class $\mathcal{S}$, we consider a pseudo likelihood ratio test (which is a likelihood ratio test if the noise is Gaussian). For $G \in \mathcal{S}$, let $R(G) = \sum_{i=1}^{n} Y_i \mathbb{1}(X_i \in G) - \frac{\#(\mathcal{X} \cap G)}{2}$, and let $R = \sup_{|G| \geq h} R(G)$. Under the alternative hypothesis, R should be quite large, hence, we define the test $T_n^1 = \mathbb{1}(R \geq 0)$. Note that this test reduces to scanning the interval $[0, 1]$ and seeking for a large enough quantity of successive observations with large Y_i, which can be done in a quadratic number (in n) of steps.

Theorem 2.1. *Let model* (1.2) *hold.*

1. *Assume that the design is (DD) or (RD) and that the noise satisfies:*

$$ \mathbb{P}[\xi_1 \leq -1/2] < \mathbb{P}[\xi_1 \leq 1/2]. $$

   Then, if $nh \longrightarrow \infty$, the test T_n^0 is consistent, i.e. $\gamma_n(T_n^0, \mathcal{S}_0) \longrightarrow 0$. In addition, if the noise is Gaussian, then $r_n = 1/n$ is a separation rate on the class $\mathcal{S}_0$.
2. *Assume that the design is (DD) or (RD). Then, if $nh/\ln n \longrightarrow \infty$, the test T_n^1 is consistent, i.e. $\gamma_n(T_n^1, \mathcal{S}) \longrightarrow 0$. In addition, if the noise is Gaussian, then $r_n = (\ln n)/n$ is a separation rate on the class $\mathcal{S}$.*

In the next section, we show that the separation rates given in Theorem 2.1 are equal to the minimax rates of convergence on the corresponding classes.

3. Estimation of a segment

3.1. A least square estimator

Let model (1.2) hold. For $G' \in \mathcal{S}$, let $A_0(G') = \sum_{i=1}^{n} (Y_i - \mathbb{1}(X_i \in G'))^2$ be the sum of squared errors. In order to estimate the true and unknown set G, we find a random set $\hat{G}_n$ which minimizes $A_0(G')$, among all possible candidates G'. Note that minimizing $A_0(G')$ is equivalent to maximizing

$$A(G') = \sum_{i=1}^{n} (2Y_i - 1)\mathbb{1}(X_i \in G'). \tag{3.1}$$

Denote by $S = \{i = 1, \ldots, n : X_i \in G\}$ and by $S' = \{i = 1, \ldots, n : X_i \in G'\}$, for some $G' \in \mathcal{S}$. If we write $A(G')$ as a function of S', (3.1) becomes

$$
\begin{aligned}
A(S') &= \sum_{i \in S'} (2Y_i - 1) \\
&= \sum_{i \in S'} (2\mathbb{1}(X_i \in G) + 2\xi_i - 1) \\
&= 2\#(S \cap S') - \#S' + 2\sum_{i \in S'} \xi_i,
\end{aligned}
$$

so,

$$A(S') - A(S) = -\#(S \Delta S') + 2\left(\sum_{i \in S' \setminus S} \xi_i - \sum_{i \in S \setminus S'} \xi_i\right). \tag{3.2}$$

We call a set $S' \subseteq \{1, \ldots, n\}$ convex if and only if it is of the form $\{i, \ldots, j\}$, for some $1 \le i \le j \le n$. It is clear that if a convex subset S' of $\{1, \ldots, n\}$ maximizes $A(S') - A(S)$, then the segment $G' = [X_{\min S'}, X_{\max S'}]$ maximizes $A(G')$.

3.2. Estimation of one change-point

3.2.1. Under the deterministic design (DD)

Let model (1.2) hold, with design (DD). Assume that $G = [0, \theta] \in \mathcal{S}_0$, where $\theta \in [0, 1]$. Let us make one preliminary remark. For any estimator $\hat{G}_n$ of G, the random segment $\tilde{G}_n = [0, \sup \hat{G}_n]$ performs better than $\hat{G}_n$, since $|\tilde{G}_n \Delta G| \le |\hat{G}_n \Delta G|$ almost surely. Therefore, it is sufficient to only consider estimators of the form $\hat{G}_n = [0, \hat{\theta}_n]$, where $\hat{\theta}_n$ is a random variable. Then, $|\hat{G}_n \Delta G| = |\hat{\theta}_n - \theta|$, and the performance of the estimator $\hat{G}_n$ of G is that of the estimator $\hat{\theta}_n$ of the change-point θ. Let us define a least square estimator (LSE) of θ. For $M = 1, \ldots, n$, let

$$
\begin{aligned}
F(M) &= A(\{1, \ldots, M\}) \\
&= \sum_{i=1}^{M} (2Y_i - 1).
\end{aligned}
\tag{3.3}
$$

Let $\hat{M}_n \in \underset{M=1,\ldots,n}{\operatorname{ArgMax}} F(M)$, and $\hat{\theta}_n = X_{\hat{M}_n}$. We have the following theorem.

Theorem 3.1. *Let $n \geq 1$. Let model* (1.2) *hold, with design (DD). Let $\hat{G}_n = [0, \hat{\theta}_n]$. Then,*

$$\sup_{G \in \mathcal{S}_0} \mathbb{P}_G \left[|\hat{G}_n \Delta G| \geq \frac{x}{n} \right] \leq C_0 e^{-x/(8\sigma^2)}, \quad \forall x > 0,$$

where C_0 is a positive constant which depends on σ only.

A simple application of Fubini's theorem yields the following result.

Corollary 3.2. *Let the assumptions of Theorem* 3.1 *be satisfied. Then, for all $q > 0$, there exists a positive constant A_q which depends on q and σ only, such that*

$$\sup_{G \in \mathcal{S}_0} \mathbb{E}_G \left[|\hat{G}_n \Delta G|^q \right] \leq \frac{A_q}{n^q}.$$

This corollary shows that the minimax risk on the class $\mathcal{S}_0$ is bounded from above by $1/n$, up to multiplicative constants.

3.2.2. Under the random design (RD)

We now consider model (1.2) hold, with design (RD). Assume that G belongs to $\mathcal{S}_0$ and consider again the function $F(M)$ as defined in (3.3), with $\hat{M}_n \in \underset{M=1,\ldots,n}{\mathrm{ArgMax}} \, F(M)$, $\hat{\theta}_n = X_{\hat{M}_n}$ and $\hat{G}_n = [0, \hat{\theta}_n]$, as in the previous section. Then, the following theorem holds.

Theorem 3.3. *Let $n \geq 1$ and let model* (1.2) *hold, with random design (RD). Then, the estimator $\hat{G}_n$ satisfies the following moment inequalities:*

$$\sup_{G \in \mathcal{S}_0} \mathbb{E}_G \left[|\hat{G}_n \Delta G|^q \right] \leq \frac{2C(2q)!(16\sigma^2)^q}{n^q},$$

for all positive integer q, where $C = 2 + 16\sigma^2$. Moreover, there exist universal constants A_1, A_2 and A_3 such that the following deviation inequality holds for all $x \geq 0$:

$$\mathbb{P} \left[|\hat{G}_n \Delta G| \geq \frac{\sigma^2}{n} (x + A_1 \sigma^2) \right] \leq e^{-\frac{x}{A_2 \sqrt{x} + A_3 \sigma^2}}.$$

Note that the estimator that we define is the same as in Section 1.9 from [16], *i.e.*, the least square estimator. However, we do not make use of a separation assumption of θ from 0 and 1, unlike in their proof, which shows that this assumption, which is only technical, is not linked to their estimation method. Next theorem shows that up to constants, $1/n$ is also a lower bound on the minimax risk for both designs (DD) and (RD).

Theorem 3.4. *Consider model* (1.2), *with design (DD) or (RD). Then for all integer $n \geq 1$,*

$$\mathcal{R}_n(\mathcal{S}_0) \geq \frac{1}{8n}.$$

Note that in the case of the deterministic design, the lower bound is determined by the fact that the parameter θ in (1.1) is not identified by the model.

As a consequence, the minimax rate of convergence on the class $\mathcal{S}_0$ is $1/n$ under both designs (DD) and (RD).

Theorem 3.5. *Consider model* (1.2), *with either design (DD) or (RD). Then, the minimax risk on the class*

$\mathcal{S}_0$ *satisfies*

$$\mathcal{R}_n(\mathcal{S}_0) \asymp_\sigma \frac{1}{n}.$$

3.3. Estimation of two change-points

Let us now assume that the unknown segment G does not necessarily contain 0. We will prove that the classes $\mathcal{S}$ and $\mathcal{S}(\mu)$ yield two minimax rates that differ by a logarithmic factor.

When G is only assumed to belong to the largest class $\mathcal{S}$, we define the same LSE estimator as we did in [4] for convex polytopes and we prove an upper bound of the same order, *i.e.*, $(\ln n)/n$. This estimator is obtained by maximising $A(G')$ (see (3.1)) over all segments G' with edge points that are integer multiples of $1/n$.

If $|G|$ is *a priori* known to be greater or equal to μ, we split the available sample into two parts. With the first half, we define the same LSE estimator as in the general case seen above. This estimator may not be minimax optimal on $\mathcal{S}(\mu)$, but it is close to G with high probability as shown in Theorem 3.6. Thus, the middle point $\hat{m}_n$ of this estimator is inside G with high probability. On that event, we use the second half of the sample in order to estimate the two edge points of G, one on each side of $\hat{m}_n$, using the same technique as on the class S_0, where the base point 0 is now replaced with $\hat{m}_n$.

3.3.1. On the class $\mathcal{S}$

Let us first state the following theorem, which is, for the design (RD), a particular case of Theorem 1 in [4], for $d = 1$.

Theorem 3.6. *Let $n \geq 2$. Let model (1.2) hold, with design (DD) or (RD). Let $\hat{G}_n \in \underset{G' \in \mathcal{S}}{\mathrm{ArgMax}}\ A(G')$ be a LSE estimator of G. Then, there exist two positive constants C_1 and C_2 which depend on σ only, such that*

$$\sup_{G \in \mathcal{S}} \mathbb{P}_G \left[n \left(|\hat{G}_n \Delta G| - \frac{4 \ln n}{C_2 n} \right) \geq x \right] \leq C_1 e^{-C_2 x}, \forall x > 0.$$

The expressions of C_1 and C_2 are given in the proof of Theorem 1 from [4], for the design (RD). For the design (DD), we do not give a proof of this theorem here, but it can be easily adapted from that of the case of the design (RD). The next corollary comes as an immediate consequence.

Corollary 3.7. *Let the assumptions of Theorem 3.6 be satisfied. Then, for all $q > 0$, there exists a positive constant B_q which depends on q and σ only, such that*

$$\sup_{G \in \mathcal{S}} \mathbb{E}_G \left[|\hat{G}_n \Delta G|^q \right] \leq B_q \left(\frac{\ln n}{n} \right)^q.$$

This corollary shows that the minimax risk on the class $\mathcal{S}$ is bounded from above by $\ln(n)/n$, up to a multiplicative constant. The following theorem establishes a lower bound, if the noise is supposed to be Gaussian.

Theorem 3.8. *Consider model (1.2), with design (DD) or (RD). Assume that the noise terms ξ_i are i.i.d. Gaussian random variables, with variance $\sigma^2 > 0$. For any large enough n,*

$$\mathcal{R}_n(\mathcal{S}) \geq \frac{\alpha^2 \sigma^2 \ln n}{n},$$

where α is a universal positive constant.

This lower bound comes from Theorem 2 from [4] in the case of the design (RD), and the proof is easily adapted for the design (DD). As a consequence, the minimax risk on the class $\mathcal{S}$ is of the order $\ln(n)/n$:

Theorem 3.9. *Consider model* (1.2), *with design (DD) or (RD). Assume that the noise terms ξ_i are i.i.d. Gaussian random variables, with variance $\sigma^2 > 0$. The minimax risk on the class $\mathcal{S}$ satisfies, asymptotically:*

$$\mathcal{R}_n(\mathcal{S}) \asymp_\sigma \frac{\ln n}{n}.$$

3.3.2. On the class $\mathcal{S}(\mu)$

In this section we shall combine both Theorems 3.1 and 3.6 to find the minimax rate on the class $\mathcal{S}(\mu)$. Let model (1.2) hold and let $G \in \mathcal{S}(\mu)$. First, we split the sample into two equal parts. Let $\mathcal{D}_0$ be the set of sample points with even indices, and $\mathcal{D}_1$ the set of sample points with odd indices. Note that $\mathcal{D}_0 \cup \mathcal{D}_1$ is exactly the initial sample, that these two subsamples are independent, and that each of them has at least $(n-1)/2$ ordered design points. Let $\hat{G}_n$ be the LSE estimator of G given in Theorem 3.6, computed from the subsample $\mathcal{D}_0$, and let $\hat{m}_n$ be the middle point of $\hat{G}_n$. As it will be shown in the proof of the next theorem, $\hat{m}_n$ satisfies both following properties, with high probability:

1. $\hat{m}_n \in G$,
2. $\mu/4 \leq \hat{m}_n \leq 1 - \mu/4$.

Then, each endpoint of G is estimated separately, using the same technique as in Section 3.2.

Let I_1 be the set of odd integers between 1 and n, so $\mathcal{D}_1 = \{(X_i, Y_i) : i \in I_1\}$. Define $I_1^+ = \{i \in I_1 : X_i \geq \hat{m}_n\}$ and $I_1^- = \{i \in I_1 : X_i < \hat{m}_n\}$. Then, for all $i \in I_1^+$, $Y_i = \mathbb{1}(X_i \leq b) + \xi_i$ and for all $i \in I_1^-$, $Y_i = \mathbb{1}(X_i \geq a) + \xi_i = \mathbb{1}(1 - X_i \leq 1 - a) + \xi_i$, where a and b are the left and right boundaries of G, respectively. Let $F_+(M) = \sum\limits_{\substack{i \in I_1^+ \\ i \leq M}} (2Y_i - 1)$ and $F_-(M) = \sum\limits_{\substack{i \in I_1^- \\ i \geq M}} (2Y_i - 1)$. If $I_1^+ = \emptyset$, let $\tilde{M}^+ = n$, otherwise, let $\tilde{M}^+ \in \underset{M \in I_1^+}{\mathrm{ArgMax}} \, F_+(M)$. Similarly, if $I_1^+ = \emptyset$, let $\tilde{M}^- = 1$, otherwise, let $\hat{M}^- \in \underset{M \in I_1^-}{\mathrm{ArgMax}} \, F_-(M)$. Finally, set $\tilde{G}_n = [\tilde{a}, \tilde{b}]$, where $\tilde{a} = X_{\hat{M}^-}$ and $\tilde{b} = X_{\hat{M}^+}$.

3.3.3. Under the design (DD)

By adapting Theorem 3.1, we know that under the design (DD), both edge points of G can be estimated at the speed $1/n$, up to multiplicative constants, when the two properties above are satisfied by $\hat{m}_n$. We get the following theorem.

Theorem 3.10. *Consider model* (1.2), *with design (DD). There exists an estimator $\tilde{G}_n$ of G, such that*

$$\sup_{G \in \mathcal{S}(\mu)} \mathbb{P}_G \left[|\tilde{G}_n \Delta G| \geq \frac{x}{n} \right] \leq 2C_0 e^{-\mu x/(256\sigma^2)} + C_1 n^4 e^{-C_2 \mu n/2}, \quad \forall x > 0,$$

for n large enough. The positive constants C_0 and C_2 are the same as in Theorems 3.1 and 3.6.

Naturally, Theorem 3.10 yields next corollary.

Corollary 3.11. *Let the assumptions of Theorem 3.10 be satisfied. Then, for all $q > 0$, there exists a positive constant B_q' which depends on q, μ and σ only, such that*

$$\sup_{G \in \mathcal{S}(\mu)} \mathbb{E}_G \left[|\tilde{G}_n \Delta G|^q \right] \leq \frac{B_q'}{n^q}.$$

This corollary, for $q = 1$, shows that the minimax risk on the class $\mathcal{S}(\mu)$ is bounded from above by $1/n$, up to a multiplicative constant.

3.3.4. Under the design (RD)

When the random design (RD) is considered, we prove the following inequality, which is less strong than the deviation inequality obtained for the deterministic design, and yet enough for our purposes.

Theorem 3.12. *Let $n \geq 1$ and consider model* (1.2)*, with design (RD). There exist positive constants c_j, $j = 1, \ldots, 5$, that depend on σ^2 and μ only, such that the following holds.*

$$\sup_{G \in \mathcal{S}(\mu)} \mathbb{P}_G \left[|\tilde{G}_n \Delta G| \geq \frac{x + c_1}{n} \right] \leq c_2 n^4 e^{-c_3 n} + c_4 e^{-c_5 \sqrt{x}},$$

for all $x \geq 0$.

As a consequence, we have the following moment inequalities.

Corollary 3.13. *Let the assumptions of Theorem 3.12 hold. Then, for all $q > 0$, there exists a positive constant B_q' which depends on q, μ and σ only, such that*

$$\sup_{G \in \mathcal{S}(\mu)} \mathbb{E}_G \left[|\tilde{G}_n \Delta G|^q \right] \leq \frac{B_q'}{n^q}.$$

As in the case of the deterministic design, this corollary shows that the minimax risk on the class $\mathcal{S}(\mu)$ is bounded from above by $1/n$, up to a multiplicative constant.

A very similar proof to that of Theorem 3.4 yields a lower bound for this minimax risk, and we get the next theorem.

Theorem 3.14. *Consider model* (1.2)*, with design (DD) or (RD). The minimax risk on the class $\mathcal{S}(\mu)$ satisfies, asymptotically,*

$$\mathcal{R}_n(\mathcal{S}(\mu)) \asymp_{\mu, \sigma} \frac{1}{n}.$$

Remark 3.15. Note that, in Theorem 3.10, the upper bound contains one residual term which does not depend on x. This term, in order to be sufficiently small and to still yield Corollary 3.11 for $q = 1$, requires that μ – if allowed to depend on n – is of order larger than $(\ln n)/n$. This is consistent with Theorem 2.1 which shows that $(\ln n)/n$ is the order of magnitude of the length of the shortest segment that can be detected consistently, in a minimax sense. If μ becomes too small, *i.e.*, of order less or equal to $\ln(n)/n$, then the lower bound proved for the minimax risk on $\mathcal{S}(\mu)$ will break down to $(\ln n)/n$ and the proof of the lower bound in Theorem 3.8 can be applied to $\mathcal{S}(\mu)$, which would yield $\mathcal{R}_n(\mathcal{S}(\mu)) \asymp_\sigma \mathcal{R}_n(\mathcal{S}) \asymp_\sigma \frac{\ln n}{n}$ as expected.

Remark 3.16. We have only considered uniform or regular designs. However, the rates would be deteriorated if we allow densities that can get arbitrarily close to zero. For instance, as in [9], if the density f of the design points satisfies $f(x) \sim x^\beta$ as $x \to 0$, where $\beta > 0$, then adapting the proofs yields lower bounds of the order of $n^{-1/(\beta+1)}$ for the minimax estimation rates on the three classes $\mathcal{S}_0$, $\mathcal{S}$ and $\mathcal{S}(\mu)$, which is strictly slower than $(\ln n)/n$.

TABLE 1. Minimax risks and separation rates for the classes $\mathcal{S}$, $\mathcal{S}_0$ and $\mathcal{S}(\mu)$.

	$\mathcal{S}$	$\mathcal{S}_0$	$\mathcal{S}(\mu)$
Minimax rate for estimation	$\ln(n)/n$	$1/n$	$1/n$
Separation rate for detection	$\ln(n)/n$	$1/n$	$\cdot$

4. CONCLUSION AND DISCUSSION

We summarize our results in Table 1. The rates that are written in this table hold when the noise is Gaussian.

We have shown that asymptotically, a segment can be estimated infinitely faster when it is *a priori* supposed either to contain a given point (here, 0), or to be large enough. In terms of the change-point problem, the change-points can be detected and their location estimated at the fast rate $1/n$ either when one is known (*i.e.*, there is only one unknown change-point), or when they are separated from one another. In particular, in the case of a single change-point, we show that the assumption that it is separated away from the boundaries of the domain, which is always imposed in change-point estimation (*e.g.*, [12–14, 16]) is only technical, not fundamental. When there are two change points that are not separated away from one another, we showed the presence of an extra logarithmic factor in both the minimax rate for estimation and separation rate for detection. As we mentioned in the introduction [6], prove a similar result for detection in the case of two change-points that are not assumed to be separated from each other. However, in their work, the amplitude of the jump is not known. In our results, we have proved that even if the amplitude is known (it is 1 in our case), the logarithmic factor is unavoidable.

Detection is a simpler problem than estimation. Hence, studying detection gives a benchmark for the optimal rates in estimation. Here, we did inference on segments, which are parametric objects, and we showed that detection and estimation are asymptotically equivalent, in the sense that in each scenario that we studied, separation rates for detection and minimax rates for estimation coincide. Because of the sharp jumps of the indicator function $\mathbb{1}(x \in G)$ at the boundaries of G, one would expect the parametric object G to be estimable at the rate n^{-1}. However, we showed that in general, if G can be arbitrarily small, hence, hard to detect, the rate n^{-1} becomes negligible compared with the detection separated rate $(\ln n)/n$, which becomes the optimal estimation rate. If the detection threshold is at most of order n^{-1}, as it is the case in the class $\mathcal{S}_0$ or, obviously, in the class $\mathcal{S}(\mu)$, then n^{-1} is the optimal estimation rate, as expected.

In [4], a higher dimensional version of the model is treated, where G is a d-dimensional convex body ($d \geq 1$). In that case, G is no longer a parametric object, and the optimal rate of convergence is affected by the complexity of the class of convex sets. Morally, the optimal rates of convergence are subsequent of a competition between separation rates for detectability and a terms that relates to the complexity of the class of sets G. In dimension 1, the complexity of the parametric class of segments yields a term of order $1/n$, which is never larger than the separation rate for detectability. In higher dimensions [4], shows that the minimax rate of convergence for convex bodies is $n^{-2/(d+1)}$, although for the parametric class of convex polytopes with bounded number of vertices, it is exactly $(\ln n)/n$, which can be shown to be the separation rate. For the class of convex bodies, the term $n^{-2/(d+1)}$ dominates the separation rate, whereas for parametric classes of polytopes, the separation rate $(\ln n)/n$ dominates $1/n$ that is yielded by the complexity of the class.

In [4], there are no assumptions of the type of Assumptions 1 or 2 and G is estimated without any restriction on its size or location. We believe that if one restricts G to having volume at least μ, for some $\mu > 0$ (corresponding to the scenario 2 here), there would not be any term of order $(\ln n)/n$ in the minimax rates, and for parametric classes, the minimax rates for estimation of G would become $1/n$, *i.e.*, the logarithmic factor would disappear. In a similar fashion, we believe that under a similar scenario as 1, the separation rate would decrease to $1/n$ and the minimax rate for estimation would become $1/n$ as well for parametric classes (such as polytopes with bounded number of vertices). In higher dimensions, scenario 1 would have to be adapted. Indeed, one can easily modify the proof of the lower bound in Theorem 2 from [4] to show that the minimax rate is still at least $(\ln n)/n$, even under the extra assumption that $0 \in G$. In fact, Assumption 1 should take the following form in dimension $d \geq 2$.

Assumption 3. $[0,1]^{d-1} \times \{0\} \subseteq G$.

Particular sets satisfying this assumption are boundary fragment (see [15] and Chap. 3 from [16]). A boundary fragment in $\mathbb{R}^d$ $(d \geq 2)$ is a set of the form

$$G = \{(x, y) \in [0, 1]^{d-1} \times \mathbb{R} : 0 \leq y \leq g(x)\},$$

for some nonnegative function $g : [0, 1]^{d-1} \to \mathbb{R}$, called the *edge function* of G. As a side remark, in [15] and Chapter 3 from [16], the edge function g is assumed to satisfy $h \leq g(x) \leq 1 - h, \forall x \in [0, 1]^{d-1}$, for some $h > 0$ and we believe that, similarly to what we have proven here in dimension one, this assumption is only technical. A particular case of polytopes with bounded number of faces that satisfy Assumption 3 is boundary fragments with edge function of the form $g = \min_{1 \leq k \leq K} f_k$, where the f_k's are affine maps on $[0, 1]^{d-1}$ and K is some fixed positive integer. We believe that the correct minimax rate of convergence on such a parametric class is $1/n$, although the only known upper bound at the moment is of order $(\ln n)/n$, see [4].

5. Proofs

5.1. Proof of Theorem 2.1

5.1.1. On the class $\mathcal{S}_0$

Upper bound. Let us first prove the upper bound, *i.e.* assume that $nh \to \infty$, and prove that there exists a consistent test. Recall that $N = \max\{i = 1, \ldots, n : X_i \leq h\} = \#(\mathcal{X} \cap [0, h])$. If the design is (DD), then N is just equal to the integer part of nh. If the design (RD), then N is a binomial random variable, with parameters n and h. Let us show first that the error of the first kind of the test T_n^0 goes to zero, when $n \to \infty$.

$$\mathbb{P}_\emptyset\left[S \leq cN\right] = \mathbb{P}_\emptyset\left[\#\left\{i = 1, \ldots, N : Y_i > \frac{1}{2}\right\} \geq (1-c)N\right]$$

$$\leq \mathbb{E}\left[\mathbb{P}_\emptyset\left[\#\left\{i = 1, \ldots, N : \xi_i > \frac{1}{2}\right\} \geq (1-c)N | \mathcal{X}\right]\right].$$

Since the ξ_i's are independent of $\mathcal{X}$, the distribution of $\#\{i = 1, \ldots, N : \xi_i > \frac{1}{2}\}$ conditionally to $\mathcal{X}$ is binomial, with parameters N and β, where $\beta = \mathbb{P}\left[\xi_1 > 1/2\right] \in [0, 1)$. Thus, by Bernstein's inequality for binomial random variables, by defining $\gamma = \dfrac{(1 - c - \beta)^2}{2\beta(1 - \beta) + (1 - c - \beta)/3} > 0$,

$$\mathbb{P}_\emptyset\left[S \leq cN\right] \leq \mathbb{E}\left[\exp\left(-\gamma N\right)\right].$$

If $\mathcal{X}$ satisfies (DD), then $N \geq nh - 1$ and it is clear that $\mathbb{P}_\emptyset\left[S \leq cN\right] \longrightarrow 0$. If $\mathcal{X}$ satisfies (RD), then

$$\mathbb{E}\left[\exp\left(-\gamma N\right)\right] = \exp\left(-nh\left(1 - e^{-\gamma}\right)\right),$$

so $\mathbb{P}_\emptyset\left[S \leq cN\right] \longrightarrow 0$.

Let us show, now, that the error of the second kind goes to zero as well. Let $G \in \mathcal{S}_0$ satisfying the alternative hypothesis, *i.e.* $|G| \geq h$. Denote by $\beta' = \mathbb{P}[\xi_1 \leq -1/2]$, and by $\gamma' = \dfrac{(c - \beta')^2}{2\beta'(1 - \beta') + (c - \beta')/3} > 0$

$$\mathbb{P}_G\left[S > cN\right] = \mathbb{P}_G\left[\#\left\{i = 1, \ldots, N : Y_i \leq \frac{1}{2}\right\} > cN\right]$$

$$\leq \mathbb{E}\left[\mathbb{P}_\emptyset\left[\#\left\{i = 1, \ldots, N : \xi_i \leq -\frac{1}{2}\right\} > cN | \mathcal{X}\right]\right]$$

$$\leq \mathbb{E}\left[\exp\left(-\gamma' N\right)\right],$$

by a similar computation to that for the error of the first kind. Since the right-side of the last inequality does not depend on G,

$$\sup_{|G| \geq h} \mathbb{P}_G [S > cN] \leq \mathbb{E}\left[\exp\left(-\gamma' N\right)\right]$$

and therefore, by the same argument as for the error of the first kind, goes to zero when $n \to \infty$, for both designs (DD) and (RD).

Lower bound. Assume, now, that $nh \to 0$. Let τ_n be any test. Let $G_1 = [0, h]$. We denote by $\mathcal{H}$ the Hellinger distance between probability measures. The following computation uses properties of this distance, which can be found in [25].

$$\begin{aligned}
\gamma_n(\tau_n, \mathcal{C}) &\geq \mathbb{E}_\emptyset [\tau_n] + \mathbb{E}_{G_1} [1 - \tau_n] \\
&\geq \int \min\left(d\mathbb{P}_\emptyset, d\mathbb{P}_{G_1}\right) \\
&\geq \frac{1}{2}\left(1 - \frac{\mathcal{H}(\mathbb{P}_\emptyset, \mathbb{P}_{G_1})^2}{2}\right)^2.
\end{aligned} \tag{5.1}$$

Let $G, G' \in \mathcal{S}$. A simple computation shows that, for the design (DD),

$$1 - \frac{\mathcal{H}(\mathbb{P}_G, \mathbb{P}_{G'})^2}{2} = \exp\left(-\frac{\#\left(\mathcal{X} \cap (G\Delta G')\right)}{8\sigma^2}\right), \tag{5.2}$$

and for the design (RD),

$$1 - \frac{\mathcal{H}(\mathbb{P}_G, \mathbb{P}_{G'})^2}{2} = \left(1 - \left(1 - e^{-\frac{1}{8\sigma^2}}\right)|G\Delta G'|\right)^n. \tag{5.3}$$

In particular, for the design (DD),

$$1 - \frac{\mathcal{H}(\mathbb{P}_\emptyset, \mathbb{P}_{G_1})^2}{2} \geq \exp\left(-\frac{nh}{8\sigma^2}\right),$$

and for the design (RD),

$$1 - \frac{\mathcal{H}(\mathbb{P}_\emptyset, \mathbb{P}_{G_1})^2}{2} = \left(1 - \left(1 - e^{-\frac{1}{8\sigma^2}}\right)h\right)^n.$$

In both cases, we showed that the right side of (5.1) tends to $1/2$, when $n \to \infty$. Therefore the test τ_n is not consistent.

5.1.2. On the class $\mathcal{S}$

Upper bound. Assume that $\dfrac{nh}{\ln n} \longrightarrow \infty$. Let us first show that the error of the first kind of T_n^1 goes to zero, when $n \to \infty$. Recall that $T_n^1 = \mathbb{1}(R \geq 0)$, where $R = \sup_{|G| \geq h} R(G)$ and $R(G) = \sum_{i=1}^n Y_i \mathbb{1}(X_i \in G) - \frac{\#(\mathcal{X} \cap G)}{2}$, for all $G \in \mathcal{S}$. Note that $R(G)$ is piecewise constant, and can only take a finite number of values. It is clear that

$$\{R(G) : G \in \mathcal{S}, |G| \geq h\} = \{R([X_k, X_l]) : 1 \leq k < l \leq n, X_l - X_k > h\}.$$

Recall that for $1 \leq k < l \leq n$, $R([X_k, X_l)) = \dfrac{1}{2}\sum_{i=k}^{l-1}(2Y_i - 1)$. Therefore,

$$
\begin{aligned}
\mathbb{P}_{\emptyset}[R \geq 0] = \mathbb{P}_{\emptyset}&\left[\max_{\substack{1 \leq k < l \leq n \\ X_l - X_k > h}} R([X_k, X_l)) > 0 \right] \\
&\leq \mathbb{P}_{\emptyset}\left[\bigcup_{1 \leq k < l \leq n} \{R([X_k, X_l)) > 0\} \cap \{X_l - X_k > h\} \right] \\
&\leq \sum_{1 \leq k < l \leq n} \mathbb{P}_{\emptyset}\left[R([X_k, X_l)) > 0, X_l - X_k > h \right] \\
&\leq \sum_{1 \leq k < l \leq n} \mathbb{P}_{\emptyset}\left[\sum_{i=k}^{l-1}(2\xi_i - 1) > 0 \right] \mathbb{P}[X_l - X_k > h].
\end{aligned}
\tag{5.4}
$$

For $1 \leq k < l \leq n$,

$$
\mathbb{P}_{\emptyset}\left[\sum_{i=k}^{l-1}(2\xi_i - 1) > 0 \right] \leq \exp\left(-\frac{(l-k)\sigma^2}{8} \right),
\tag{5.5}
$$

using Markov's inequality and (1.3).

If the design is (DD), then $\mathbb{P}[X_l - X_k > h]$ is 1 if and only if $l - k > nh$, 0 otherwise, so from (5.4) and (5.5) we get that

$$
\begin{aligned}
\mathbb{P}_{\emptyset}[R \geq 0] &\leq \sum_{l-k>nh} \exp\left(-\frac{(l-k)\sigma^2}{8} \right) \\
&\leq \sum_{l-k>nh} \exp\left(\frac{(-nh)\sigma^2}{8} \right) \\
&\leq \frac{n^2}{2} \exp\left(\frac{(-nh)\sigma^2}{8} \right) \longrightarrow 0,
\end{aligned}
$$

when $n \to \infty$, which proves that the error of the first kind goes to zero.

If the design is (RD), let us use the following Lemma.

Lemma 5.1. *Let $X_1, \ldots, X_n$ be the (RD) design. Then, for any $1 \leq k < l \leq n$, and $h > 0$,*

$$
\mathbb{P}[X_l - X_k > h] \leq n \exp\left(-(n-1)h(1 - e^{-u}) + u(l-k) \right), \quad \forall u > 0.
$$

5.1.3. Proof of Lemma 5.1

Note that the event $\{X_l - X_k > h\}$ is equivalent to $\{\#(\mathcal{X} \cap (X_k, X_k + h)) < l - k\}$. Let us denote by $X_1', \ldots, X_n'$ the preliminary design, from which $X_1, \ldots, X_n$ is the reordered version. The random variables $X_1', \ldots, X_n'$ are then i.i.d., with uniform distribution on $[0, 1]$.

Hence,

$$\mathbb{P}[X_l - X_k > h] = \sum_{j=1}^{n} \mathbb{P}\left[\#(\mathcal{X} \cap (X_k, X_k + h)) < l - k, X_k = X_j'\right]$$

$$\leq \sum_{j=1}^{n} \mathbb{P}\left[\#(\mathcal{X} \cap (X_j', X_j' + h)) < l - k\right]$$

$$= \sum_{j=1}^{n} \mathbb{E}\left[\mathbb{P}\left[\#(\mathcal{X} \cap (X_j', X_j' + h)) < l - k | X_j'\right]\right]$$

$$= \sum_{j=1}^{n} \mathbb{E}\left[f(X_j')\right], \tag{5.6}$$

where $f(x) = \mathbb{P}\left[\#(\tilde{\mathcal{X}} \cap (x, x + h)) < l - k\right]$, for all $x \in [0,1]$, and $\tilde{\mathcal{X}} = \{X_1', \ldots, X_{n-1}'\}$. The random variable $\#(\tilde{\mathcal{X}} \cap (x, x + h))$ is binomial with parameters $n - 1$ and h, hence, $f(x) = \mathbb{P}[N \geq n - 1 - (l - k)]$, where N is a binomial random variable with parameters $n - 1$ and $1 - h$. By Markov's inequality, for all $u > 0$,

$$f(x) \leq e^{-u(n-1-l+k)}\mathbb{E}[e^{uN}]$$

$$= e^{-u(n-1-l+k)}\left((1 - h)e^u + h\right)^{n-1}$$

$$= e^{u(l-k)}\left(1 - h(1 - e^{-u})\right)^{n-1}$$

$$\leq e^{u(l-k)}e^{-(n-1)h(1-e^{-u})}, \tag{5.7}$$

where we used the inequality $1 - x \leq e^{-x}, \forall x \in \mathbb{R}$ in the last line. Then, (5.6) and (5.7) yield the lemma. □

Therefore, by (5.4)–(5.6), and Lemma 5.1 with $u = \sigma^2/8$,

$$\mathbb{P}_\emptyset[R \geq 0] \leq \sum_{1 \leq k < l \leq n} n \exp\left(-\frac{(l-k)\sigma^2}{8} - (n-1)h(1 - e^{-u}) + u(l - k)\right)$$

$$\leq \frac{n^3}{2} \exp\left(-(n-1)h(1 - e^{-\sigma^2/8})\right) \longrightarrow 0,$$

when $n \to \infty$, which proves that the error of the first kind goes to zero.

Let us bound, now, the error of the second kind. Let $G \in \mathcal{S}$ satisfying $|G| \geq h$. For this G, denote by $N_G = \#(\mathcal{X} \cap G)$. Then,

$$\mathbb{P}_G[R < 0] \leq \mathbb{P}_G\left[R(G) \leq N_G/2\right]$$

$$\leq \mathbb{P}\left[\sum_{i=1}^{n} \xi_i \mathbb{1}(X_i \in G) \leq -N_G/2\right]. \tag{5.8}$$

For the design (DD), N_G is the integer part of $n|G|$, so $N_G \geq nh$. Therefore, by Markov's inequality, and by (1.3), (5.8) becomes

$$\mathbb{P}_G[R < 0] \leq \exp\left(-\frac{\sigma^2 N_G}{8}\right) \leq \exp\left(-\frac{nh\sigma^2}{8}\right). \tag{5.9}$$

For the design (RD), N_G is a random binomial variable, with parameters n and $|G|$. By conditioning to the

design and using Markov's inequality, (5.8) becomes

$$
\begin{aligned}
\mathbb{P}_G[R < 0] &\leq \mathbb{P}\left[\sum_{i=1}^n -\xi_i \mathbb{1}(X_i \in G) \geq N_G/2\right] \\
&\leq \mathbb{E}\left[\exp\left(-\frac{\sigma^2 N_G}{8}\right)\right] \\
&\leq \exp\left(-Cn|G|\right) \leq \exp\left(-Cnh\right),
\end{aligned} \tag{5.10}
$$

where $C = 1 - e^{-\frac{\sigma^2}{8}}$.

In both cases (5.9) and (5.10), the right side does not depend on G, and goes to zero as $n \to \infty$. We conclude that, for both designs (DD) and (RD),

$$
\sup_{|G| \geq h} \mathbb{P}_G[R < 0] \longrightarrow 0,
$$

which ends the proof of the upper bound.

Lower bound. We more or less reproduce the proof of Theorem 3.1 in [10]. Here, the noise is supposed to be zero-mean Gaussian, with variance σ^2. Let us assume that $\dfrac{nh}{\ln n} \longrightarrow 0$. Let $M = 1/h$, assumed to be an integer, without loss of generality. For $q = 1, \ldots, M$, let $G_q = [(q-1)h, qh]$. For $q = 1, \ldots, M$, let $Z_q = \frac{d\mathbb{P}_{G_q}}{d\mathbb{P}_\emptyset}(X_1, Y_1, \ldots, X_n, Y_n)$, and denote by $\bar{Z} = \frac{1}{M}\sum_{q=1}^M Z_q$. Let τ_n be any test. Then,

$$
\begin{aligned}
\gamma_n(\tau_n, \mathcal{S}) &\geq \mathbb{P}_\emptyset[\tau_n = 1] + \frac{1}{M}\sum_{q=1}^M \mathbb{P}_{G_q}[\tau_n = 0] \\
&\geq \frac{1}{M}\sum_{q=1}^M \left(\mathbb{P}_\emptyset[\tau_n = 1] + \mathbb{P}_{G_q}[\tau_n = 0]\right) \\
&\geq \frac{1}{M}\sum_{q=1}^M \left(\mathbb{E}_\emptyset[\tau_n] + \mathbb{E}_{G_q}[1 - \tau_n]\right) \\
&\geq \frac{1}{M}\sum_{q=1}^M \mathbb{E}_\emptyset[\tau_n + (1 - \tau_n)Z_q] \\
&\geq \mathbb{E}_\emptyset\left[\tau_n + (1 - \tau_n)\bar{Z}\right] \\
&\geq \mathbb{E}_\emptyset\left[\left(\tau_n + (1 - \tau_n)\bar{Z}\right)\mathbb{1}(\bar{Z} \geq 1/2)\right] \\
&\geq \frac{1}{2}\mathbb{P}_\emptyset\left[\bar{Z} \geq 1/2\right].
\end{aligned} \tag{5.11}
$$

Let us prove that $\mathbb{E}_\emptyset[\bar{Z}] = 1$, and that $\mathbb{V}_\emptyset[\bar{Z}] \longrightarrow 0$. This will imply that the right side term of (5.11) goes to $1/2$, when $n \to \infty$.

For $q = 1, \ldots, M$, under the null hypothesis,

$$
\begin{aligned}
Z_q &= \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n \left((Y_i - \mathbb{1}(X_i \in G_q))^2 - Y_i^2\right)\right) \\
&= \exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^n (2\xi_i - 1)\mathbb{1}(X_i \in G_q)\right).
\end{aligned} \tag{5.12}
$$

By its definition, Z_q has expectation 1 under $\mathbb{P}_\emptyset$:

$$\mathbb{E}_\emptyset[\bar{Z}] = 1. \tag{5.13}$$

Since, almost surely, no design point falls in two G_q's at the time, a simple computation shows that the random variables $Z_q, q = 1, \ldots, M$, are not correlated. Thus,

$$\mathbb{V}_\emptyset[\bar{Z}] = \frac{1}{M^2} \sum_{q=1}^{M} \mathbb{V}_\emptyset[Z_q].$$

Let us bound from above $\mathbb{V}_\emptyset[Z_q]$, for $q = 1, \ldots, M$:

$$\begin{aligned}
\mathbb{V}_\emptyset[Z_q] &\leq \mathbb{E}_\emptyset[Z_q^2] \\
&= \mathbb{E}\left[\exp\left(-\frac{\#(\mathcal{X} \cap G_q)}{\sigma^2}\right) \mathbb{E}_\emptyset\left[\exp\left(\frac{2}{\sigma^2} \sum_{i=1}^{n} \xi_i \mathbb{1}(X_i \in G_q)\right) | \mathcal{X}\right]\right] \\
&= \mathbb{E}\left[\exp\left(\frac{\#(\mathcal{X} \cap G_q)}{\sigma^2}\right)\right].
\end{aligned} \tag{5.14}$$

If the design is (DD), then we get that

$$\mathbb{V}_\emptyset[Z_q] \leq \exp\left(\frac{nh+1}{\sigma^2}\right),$$

and the variance of $\bar{Z}$ is then bounded from above:

$$\mathbb{V}_\emptyset[\bar{Z}] \leq h \exp\left(\frac{nh+1}{\sigma^2}\right). \tag{5.15}$$

If the design is (RD), then $\#(\mathcal{X} \cap G_q)$ is a binomial random variable with parameters n and h, so from (5.14), we get that

$$\begin{aligned}
\mathbb{V}_\emptyset[Z_q] &\leq \left(1 + \left(e^{1/\sigma^2} - 1\right)h\right)^n \\
&\leq \exp\left(Cnh\right),
\end{aligned}$$

where $C = e^{1/\sigma^2} - 1$, and the variance of $\bar{Z}$ is then bounded from above:

$$\mathbb{V}_\emptyset[\bar{Z}] \leq h \exp\left(Cnh\right). \tag{5.16}$$

Since we assumed that $nh/\ln n \longrightarrow 0$, the right side terms of (5.15) and (5.16) go to zero, and therefore, for both designs (DD) and (RD),

$$\mathbb{V}_\emptyset[\bar{Z}] \longrightarrow 0. \tag{5.17}$$

Finally, we get from (5.11), (5.13) and (5.17), that

$$\liminf_{n \to \infty} \gamma_n(\tau_n, \mathcal{S}) \geq \frac{1}{2}.$$

This concludes the proof. □

5.2. Proof of Theorem 3.1

The beginning of this proof holds for any design $\{X_1, \ldots, X_n\}$, independent of the noise $\xi_i, i = 1, \ldots, n$. Let $G \in \mathcal{S}_0$. Let $M = \max\{i = 1, \ldots, n : X_i \in G\}$ – set $M = 0$ if the set is empty. Then, $\hat{M}_n \in \underset{M'=1,\ldots,n}{\text{ArgMax}}$ $(F(M') - F(M))$, and, by (3.2),

$$F(M') - F(M) = -|M' - M| + \begin{cases} 2\sum_{i=M'+1}^{M} \xi_i & \text{if } M > M', \\ 0 & \text{if } M' = M, \\ -2\sum_{i=M+1}^{M'} \xi_i & \text{if } M < M'. \end{cases}$$

Let us complete the i.i.d. sequence $\xi_1, \ldots, \xi_n$ to obtain an infinite double sided i.i.d. sequence $(\xi_i)_{i \in \mathbb{Z}}$, independent of the design. Let $k \in \mathbb{N}^*$ be any positive integer. Define, for $i \in \mathbb{Z}$, $\tilde{\xi}_i = \xi_{i+M}$. Since M depends on the design only, it is independent of the $\xi_i, i \in \mathbb{Z}$, and therefore, the $\tilde{\xi}_i, i \in \mathbb{Z}$ are i.i.d., with same distribution as ξ_1. Let E_k be the event $\{\hat{M}_n \geq M + k\}$. If E_k holds, then $F(j) - F(M) \geq 0$ for some $j \geq M + k$, yielding

$$0 \leq \max_{M+k \leq j \leq n} \left(M - j - 2 \sum_{i=M+1}^{j} \xi_i \right)$$

$$= \max_{M+k \leq j \leq n} \left(M - j - 2 \sum_{i=1}^{j-M} \tilde{\xi}_i \right)$$

$$= \max_{k \leq j \leq n-M} \left(-j - 2 \sum_{i=1}^{j} \tilde{\xi}_i \right)$$

$$\leq \max_{j \geq k} \left(-j - 2 \sum_{i=1}^{j} \tilde{\xi}_i \right).$$

Hence, for all $u > 0$,

$$\mathbb{P}_G[E_k] \leq \mathbb{P}_G \left[\max_{k \leq j} \left(-j - 2 \sum_{i=1}^{j} \hat{\xi}_i \right) \geq 0 \right]$$

$$\leq \mathbb{P} \left[\max_{k \leq j} \left(-j - 2 \sum_{i=1}^{j} \xi_i \right) \geq 0 \right]$$

$$\leq \sum_{j=k}^{\infty} \mathbb{P} \left[-2 \sum_{i=1}^{j} \xi_i \geq j \right]$$

$$\leq \sum_{j=k}^{\infty} \mathbb{E} \left[e^{-2u \sum_{i=1}^{j} \xi_i} \right] e^{-uj}, \text{ by Markov's inequality}$$

$$\leq \sum_{j=k}^{\infty} e^{(-u+2\sigma^2 u^2)j}, \text{ by (1.3)}$$

and, by choosing $u = 1/(4\sigma^2)$,

$$\mathbb{P}_G[E_k] \leq C e^{-k/(8\sigma^2)},$$

where $C = \left(1 - e^{-1/(8\sigma^2)}\right)^{-1}$ is a positive constant. By symmetry, we obtain that:

$$\mathbb{P}_G[|\hat{M}_n - M| \geq k] \leq 2Ce^{-k/(8\sigma^2)}. \tag{5.18}$$

If the design is (DD), the conclusion is straightforward, since for all $i, j = 1, \ldots, n, |X_i - X_j| = \frac{|i-j|}{n}$, and Theorem 3.1 is proved.

5.3. Proof of Theorem 3.3

First, note that inequality (5.18) holds for any design. Hence, under the random design (RD),

$$\mathbb{P}_G[|\hat{M}_n - M| \geq k] \leq 2Ce^{-k/(8\sigma^2)}, \tag{5.19}$$

where we recall that $M = \max\{i = 1, \ldots, n : X_i \in G\}$, $M = 0$ if the set is empty.

Let q be a positive integer. We aim to bound $\mathbb{E}_G\left[|X_{\hat{M}_n} - \theta|^q\right]$ from above, where $G = [0, \theta]$, $\theta \in [0, 1]$.

$$\mathbb{E}_G\left[|X_{\hat{M}_n} - \theta|^q\right] = \sum_{k=-\infty}^{\infty} \mathbb{E}_G\left[|X_{\hat{M}_n} - X_M|^q \mathbb{1}(\hat{M}_n = M + k)\right]$$

$$= \sum_{k=-\infty}^{0} \mathbb{E}_G\left[|X_{\hat{M}_n} - \theta|^q \mathbb{1}(\hat{M}_n = M + k)\right]$$

$$+ \sum_{k=1}^{\infty} \mathbb{E}_G\left[|X_{\hat{M}_n} - \theta|^q \mathbb{1}(\hat{M}_n = M + k)\right]. \tag{5.20}$$

Let us bound from above the second sum only. The first one requires exactly the same techniques.

Just for the sake of notation, we set X_j to zero whenever j is not in the range $\{1, \ldots, n\}$ and we identify X_M with θ in the next computation.

$$\sum_{k=1}^{\infty} \mathbb{E}_G\left[|X_{\hat{M}_n} - \theta|^q \mathbb{1}(\hat{M}_n = M + k)\right] = \sum_{k=1}^{\infty} \mathbb{E}_G\left[(X_{M+k} - \theta)^q \mathbb{1}(\hat{M}_n = M + k)\right]$$

$$\leq \sum_{k=1}^{\infty} k^{q-1} \sum_{j=0}^{k-1} \mathbb{E}_G\left[(X_{M+j+1} - X_{M+j})^q \mathbb{1}(\hat{M}_n = M + k)\right]$$

$$\leq \sum_{k=1}^{\infty} k^{q-1} \sum_{j=0}^{k-1} \mathbb{E}_G\left[(X_{M+j+1} - X_{M+j})^{2q}\right]^{1/2} \mathbb{P}_G\left[\hat{M}_n = M + k\right]^{1/2}, \tag{5.21}$$

where the last line follows from Cauchy–Schwarz inequality. By (5.19), for $k \geq 1$,

$$\mathbb{P}_G\left[\hat{M}_n = M + k\right] \leq \mathbb{P}_G\left[|\hat{M}_n - M| \geq k\right] \leq 2Ce^{-k/(8\sigma^2)}. \tag{5.22}$$

Let j be a nonnegative integer. Conditioning on the random variable M yields

$$\mathbb{E}_G\left[(X_{M+j+1} - X_{M+j})^{2q}\right] = \sum_{l=1}^{n} \mathbb{E}_G\left[(X_{M+j+1} - X_{M+j})^{2q} | M = l\right] \mathbb{P}[M = l]. \tag{5.23}$$

To compute the first term in the right hand side of (5.23), we use the following lemma. For all vectors $(Y_1, \ldots, Y_n)$ with pairwise disjoint entries, we denote by $(Y_{(1)}, \ldots, Y_{(n)})$ their increasing reordering, i.e., $Y_{(1)} < \ldots < Y_{(n)}$.

Lemma 5.2. *Let n be a positive integer, $\theta \in [0,1]$ and $Y_1, \ldots, Y_n$ be i.i.d. random variables with uniform distribution in $[0,1]$. Let $M = \max\{i = 1, \ldots, n : Y_{(i)} \leq \theta\}$ and $l \in \{1, \ldots, n-1\}$. Let $Z_1, \ldots, Z_{n-l}$ be i.i.d random variables with uniform distribution in $[\theta, 1]$. Then, the conditional distribution of $(Y_{(M+1)}, \ldots, Y_{(n)})$ conditional on the event $\{M = l\}$ is equal to the distribution of $Z_{(1)}, \ldots, Z_{(n-l)}$.*

Proof. Recall that the joint density of $(Y_{(1)}, \ldots, Y_{(n)})$ is given by

$$n!\,\mathbb{1}(0 \leq y_1 \leq \ldots \leq y_n), \qquad y_1, \ldots, y_n \in \mathbb{R}.$$

Let $f : \mathbb{R}^{n-l} \longrightarrow \mathbb{R}$ be a continuous and bounded function. Then,

$$
\begin{aligned}
\mathbb{E}\left[f(Y_{(M+1)}, \ldots, Y_{(n)})|M = l\right] &= \mathbb{E}\left[f(Y_{(l+1)}, \ldots, Y_{(n)})\mathbb{1}(M = l)\right] (\mathbb{P}[M = l])^{-1} \\
&= \int_{[0,1]^n} f(y_{l+1}, \ldots, y_n)\mathbb{1}(y_1 \leq \ldots \leq y_l \leq \theta \leq y_{l+1} \leq \ldots \leq y_n)\mathrm{d}y_1 \ldots \mathrm{d}y_n \\
&\quad \times n!\,(\mathbb{P}[M = l])^{-1} \\
&= \int_{[0,1]^{n-l}} f(y_{l+1}, \ldots, y_n)\mathbb{1}(\theta \leq y_{l+1} \leq \ldots \leq y_n)\mathrm{d}y_{l+1} \ldots \mathrm{d}y_n \\
&\quad \times n!\int_{[0,1]^l} \mathbb{1}(y_1 \leq \ldots \leq y_l \leq \theta) \,(\mathbb{P}[M = l])^{-1}, \qquad (5.24)
\end{aligned}
$$

which shows that the conditional joint density of $(Y_{(l+1)}, \ldots, Y_{(n)})$ is proportional to $\mathbb{1}(\theta \leq y_{l+1} \leq \ldots \leq y_n \leq 1)$ and therefore shows Lemma 5.2. $\qquad \square$

As a consequence of this lemma, for all $l \in \{1, \ldots, n\}$ and $j \in \{0, \ldots, n-l-1\}$,

$$
\begin{aligned}
\mathbb{E}_G\left[(X_{M+j+1} - X_{M+j})^{2q}|M = l\right] &= \mathbb{E}\left[(\min(Z_1, \ldots, Z_{n-l}) - \theta)^{2q}\right] \\
&= 2q(1-\theta)^{2q}\frac{\Gamma(2q)\Gamma(n-l+1)}{\Gamma(n+2q-l+1)},
\end{aligned}
$$

where $Z_1, \ldots, Z_{n-l}$ are i.i.d. unformly distributed in $[\theta, 1]$ and Γ is Euler's Gamma function.

If $j \geq n-l$, then the previous conditional expectation is zero.

In addition,

$$\mathbb{P}[M = l] = \binom{n}{l}\theta^l(1-\theta)^{n-l}.$$

Hence, following (5.23), for $j \in \{0, \ldots, k-1\}$

$$
\begin{aligned}
\mathbb{E}_G\left[(X_{M+j+1} - X_{M+j})^{2q}\right] &= \sum_{l=1}^{n} \frac{n!}{l!(n-l)!}\theta^l(1-\theta)^{n-l}2q(1-\theta)^{2q}\frac{\Gamma(2q)\Gamma(n-l+1)}{\Gamma(n+2q-l+1)} \\
&= \frac{n!(2q)!}{(n+2q)!}\sum_{l=1}^{n}\frac{(n+2q)!}{l!(n+2q-l)!}\theta^l(1-\theta)^{n+2q-l} \\
&\leq \frac{n!(2q)!}{(n+2q)!} \\
&\leq \frac{(2q)!}{n^{2q}}. \qquad (5.25)
\end{aligned}
$$

Finally, combining (5.21) with (5.19) and (5.25) yields:

$$\sum_{k=1}^{\infty} \mathbb{E}_G\left[|X_{\hat{M}_n} - \theta|^q \mathbb{1}(\hat{M}_n = M + k)\right] \leq \frac{\sqrt{(2q)!}}{n^q} \sum_{k=1}^{\infty} k^q e^{-k/(16\sigma^2)}. \tag{5.26}$$

In order to bound the sum in the right hand side of (5.26), we use the following lemma.

Lemma 5.3. *For all $\lambda > 0$ and all $q > 0$,*

$$\sum_{k=1}^{\infty} k^q e^{-\lambda k} \leq (2 + \lambda^{-1})\frac{q!}{\lambda^{q+1}}.$$

Proof. Denote by $\phi(x) = x^q e^{-\lambda x}$, for all $x \geq 0$. Then, ϕ is positive and it is increasing on $[0, q/\lambda]$ and decreasing on $[q/\lambda, \infty)$. Therefore, one can write:

$$\begin{aligned}
\sum_{k=1}^{\infty} \phi(k) &\leq \sum_{k \leq q/\lambda - 1} \phi(k) + 2\phi(q/\lambda) + \sum_{k \geq q/\lambda + 1} \phi(k) \\
&\leq \int_0^{q/\lambda} \phi(t)\, \mathrm{d}t + 2\phi(q/\lambda) + \int_{q/\lambda}^{\infty} \phi(t)\, \mathrm{d}t \\
&= \frac{q!}{\lambda^{q+1}} + 2\phi(q/\lambda).
\end{aligned} \tag{5.27}$$

Note that

$$\phi(q/\lambda) = q^q e^{-q}\lambda^{-q} \leq \frac{q!}{\lambda^q}.$$

Hence, (5.27) yields the conclusion of Lemma 5.3.

$\square$

Thus, for $\lambda = (16\sigma^2)^{-1}$, (5.26) continues as

$$\begin{aligned}
\sum_{k=1}^{\infty} \mathbb{E}_G\left[|X_{\hat{M}_n} - \theta|^q \mathbb{1}(\hat{M}_n = M + k)\right] &\leq C\frac{\sqrt{(2q)!q!}}{\lambda^q n^q} \\
&\leq C\frac{(2q)!}{\lambda^q n^q},
\end{aligned}$$

where we used $(q!)^2 \leq (2q)!$ in the last inequality and $C = 2 + \lambda^{-1}$.

Similar arguments would yield the same upper bound for the first sum in (5.20). Therefore,

$$\mathbb{E}_G\left[|X_{\hat{M}_n} - \theta|^q\right] \leq 2C\frac{(2q)!}{\lambda^q n^q}, \quad \forall q \geq 1.$$

Finally, we use the following lemma, that is similar to a Bernstein-type inequality, for non centered random variables.

Lemma 5.4. *Let $A \geq 1$ and $\alpha > 0$ be given numbers. Let Z be a random variable satisfying $\mathbb{E}[Z^q] \leq A(2q)!\alpha^{2q}$ for all positive integer q. Then, for all $x \geq 0$,*

$$\mathbb{P}\left[Z \geq \alpha^2(2x + 4A)\right] \leq e^{-\frac{x}{8\sqrt{x} + 192A}}.$$

Proof. Let $T = \sqrt{Z}$. Then, the second assumption yields

$$\mathbb{E}[T^{2q}] \leq A(2q)!\alpha^{2q}, \quad \forall q \geq 1. \tag{5.28}$$

As a consequence, for all integer $q \geq 2$, Hölder's inequality yields

$$\mathbb{E}[T^{2q-1}] \leq \mathbb{E}[T^{2q}]^{\frac{2q-1}{2q}} \leq A(2q)!\alpha^{2q-1}, \tag{5.29}$$

since $A \geq 1$. In particular, (5.28) and (5.29) yield

$$\mathbb{E}[T^q] \leq A(q+1)!\alpha^q, \quad \forall q \geq 1. \tag{5.30}$$

Let $t \geq 0$. By Markov's inequality, for all $\kappa > 0$, $\mathbb{P}[T - \mathbb{E}[T] \geq t] \leq e^{-\kappa t}\mathbb{E}\left[e^{\kappa(T-\mathbb{E}[T])}\right]$ and a Taylor expansion yields

$$\mathbb{P}[T - \mathbb{E}[T] \geq t] \leq e^{-\kappa t}\left(1 + \sum_{q=2}^{\infty} \frac{\kappa^q}{q!}\mathbb{E}\left[(T - \mathbb{E}[T])^q\right]\right)$$

$$\leq e^{-\kappa t}\left(1 + \sum_{q=2}^{\infty} \frac{\kappa^q}{q!}2^{q-1}\mathbb{E}\left[T^q + \mathbb{E}[T]^q\right]\right)$$

$$\leq e^{-\kappa t}\left(1 + \sum_{q=2}^{\infty} \frac{\kappa^q}{q!}2^q\mathbb{E}\left[T^q\right]\right),$$

where we used the fact that $\mathbb{E}[T]^q \leq \mathbb{E}[T^q]$ in the last inequality. Hence, by (5.30), if $\kappa \leq 1/(4\alpha)$,

$$\mathbb{P}[T - \mathbb{E}[T] \geq t] \leq e^{-\kappa t}\left(1 + A\sum_{q=2}^{\infty}(q+1)(2\kappa\alpha)^q\right)$$

$$\leq e^{-\kappa t}\left(1 + 4A(\kappa\alpha)^2\sum_{q=0}^{\infty}(q+3)2^{-q}\right)$$

$$= e^{-\kappa t}\left(1 + 48A(\kappa\alpha)^2\right)$$

$$\leq \exp\left(-\kappa t + 48A(\kappa\alpha)^2\right), \tag{5.31}$$

where we used the inequality $(1 + u) \leq e^u, \forall u \in \mathbb{R}$.

Set $\kappa = \dfrac{t}{4\alpha t + 96A\alpha^2}$. Then, $\kappa \leq 1/(4\alpha)$ and the expression inside the exponent in the right hand side of (5.31) is bounded from above by $-t^2/(4\alpha(2t + 48A\alpha))$.

Now, note that $\mathbb{E}[T] \leq \mathbb{E}[T^2]^{1/2} = \mathbb{E}[Z]^{1/2} \leq (2A)^{1/2}\alpha$, hence, (5.31) implies

$$\mathbb{P}\left[T \geq t + (2A)^{1/2}\alpha\right] \leq e^{-\frac{t^2}{4\alpha(2t+48A\alpha)}}. \tag{5.32}$$

Since $Z = T^2$, we get

$$\mathbb{P}\left[Z \geq 2t^2 + 4A\alpha^2\right] \leq e^{-\frac{t^2}{4\alpha(2t+48A\alpha)}},$$

which implies Lemma 5.4, by taking $x = t^2/\alpha^2$. $\square$

Applying Lemma 5.4 to the random variable $|X_{\hat{M}} - \theta|$ yields the second part of the theorem.

5.4. Proof of Theorem 3.4

5.4.1. Deterministic design

The proof is straightforward for the case of the deterministic design (DD). Let $G_1 = [0,0]$ and $G_2 = [0, 1/(2n)]$. Then $\mathbb{P}_{G_1} = \mathbb{P}_{G_2}$, since no point of the design falls in $G_1 \Delta G_2$, and for any estimator $\hat{G}_n$,

$$
\begin{aligned}
2 \sup_{G \in \mathcal{S}_0} \mathbb{E}_G \left[|\hat{G}_n \Delta G| \right] &\geq \mathbb{E}_{G_1} \left[|\hat{G}_n \Delta G_1| \right] + \mathbb{E}_{G_2} \left[|\hat{G}_n \Delta G_2| \right] \\
&\geq \mathbb{E}_{G_1} \left[|\hat{G}_n \Delta G_1| + |\hat{G}_n \Delta G_2| \right] \\
&\geq \mathbb{E}_{G_1} \left[|G_1 \Delta G_2| \right] \text{ by the triangle inequality} \\
&\geq \frac{1}{2n}.
\end{aligned}
$$

$\square$

5.4.2. Random design

The proof is only slightly different for the case of random design (RD). Let $G_1 = [0,0]$ and $G_2 = [0, 1/(2n)]$. The key is to note equality of the conditional joint distributions $\mathbb{P}_{G_1} \left[\ \cdot \ \middle| X_i > 1/(2n), i = 1, \ldots, n \right]$ and $\mathbb{P}_{G_2} \left[\ \cdot \ \middle| X_i > 1/(2n), i = 1, \ldots, n \right]$. Indeed, if $Y_i = \xi_i$ and $Y_i' = \mathbb{1}(X_i \leq 1/(2n)) + \xi_i$, for $i = 1, \ldots, n$, then conditionally to the event $\{X_i > 1/(2n), i = 1, \ldots, n\}$, $(Y_i)_{i=1,\ldots,n} = (Y_i')_{i=1,\ldots,n}$ and therefore have the same conditional law. Hence, one can write:

$$
\begin{aligned}
2 \sup_{G \in \mathcal{S}_0} \mathbb{E}_G \left[|\hat{G}_n \Delta G| \right] &\geq 2 \sup_{G \in \mathcal{S}_0} \mathbb{E}_G \left[|\hat{G}_n \Delta G| \mathbb{1}(X_i > 1/(2n), i = 1, \ldots, n) \right] \\
&\geq \mathbb{E}_{G_1} \left[|\hat{G}_n \Delta G_1| \mathbb{1}(X_i > 1/(2n), i = 1, \ldots, n) \right] \\
&\quad + \mathbb{E}_{G_2} \left[|\hat{G}_n \Delta G_2| \mathbb{1}(X_i > 1/(2n), i = 1, \ldots, n) \right] \\
&= \Big(\mathbb{E}_{G_1} \left[|\hat{G}_n \Delta G_1| \middle| X_i > 1/(2n), i = 1, \ldots, n \right] \\
&\quad + \mathbb{E}_{G_2} \left[|\hat{G}_n \Delta G_2| \middle| X_i > 1/(2n), i = 1, \ldots, n \right] \Big) \\
&\quad \times \mathbb{P}\left[X_1 > 1/(2n), i = 1, \ldots, n \right] \\
&= \Big(\mathbb{E}_{G_1} \left[|\hat{G}_n \Delta G_1| \middle| X_i > 1/(2n), i = 1, \ldots, n \right] \\
&\quad + \mathbb{E}_{G_1} \left[|\hat{G}_n \Delta G_2| \middle| X_i > 1/(2n), i = 1, \ldots, n \right] \Big) \mathbb{P}\left[X_1 > 1/(2n) \right]^n \\
&\geq \mathbb{E}_{G_1} \left[|G_1 \Delta G_2| \middle| X_i > 1/(2n), i = 1, \ldots, n \right] \\
&\quad \times \mathbb{P}\left[X_1 > 1/(2n) \right]^n \text{ by the triangle inequality} \\
&\geq \frac{1}{2n} \left(1 - \frac{1}{2n} \right)^n \geq \frac{1}{4n}.
\end{aligned}
$$

$\square$

5.5. Proof of Theorem 3.10

Consider the event $E = \{ |\hat{G}_n \Delta G| \leq \mu/2 \}$. Since $\{X_i : i \in I_0\}$ is a deterministic and regular design, with step $2/n$, Theorem 3.6 yields $\mathbb{P}_G[E] \geq 1 - C_1 n^4 e^{-C_2 \mu n/2}$. On the event E, $|\hat{G}_n \Delta G| < \mu \leq |G|$, yielding that $\hat{G}_n$ and

G must intersect. Thus, still on the event E,

$$|\hat{G}_n \Delta G| = |\hat{b}_n - b| + |\hat{a}_n - a|,$$

where we denoted by $G = [a, b]$ and $\hat{G}_n = [\hat{a}_n, \hat{b}_n]$. Let $m = \frac{a+b}{2}$ and $\hat{m}_n = \frac{\hat{a}_n + \hat{b}_n}{2}$ be, respectively, the middle points of G and $\hat{G}_n$. From now on, let us assume that E holds. Then,

$$\begin{aligned}
|\hat{m}_n - m| &\leq \frac{1}{2}(|\hat{b}_n - b| + |\hat{a}_n - a|) \\
&= \frac{1}{2}|\hat{G}_n \Delta G| \\
&\leq \frac{\mu}{4}.
\end{aligned} \tag{5.33}$$

Therefore, $\hat{m}_n \in G$ and, combining (5.33) with the fact that $|G| \geq \mu$,

$$\min(\hat{m}_n, 1 - \hat{m}_n) \geq \frac{\mu}{4}. \tag{5.34}$$

By (5.34), $\#I_1^\epsilon \geq \frac{\mu n}{8} - 1 \geq \frac{\mu n}{16}$ for n large enough, and for $\epsilon \in \{+, -\}$. Note that $\{X_i : i \in I_1^+\}$ (resp. $\{X_i : i \in I_1^-\}$) is a deterministic and regular design of the segment $[\hat{m}_n, 1]$ with $b \in [\hat{m}_n, 1]$ (resp. $[0, \hat{m}_n]$ with $a \in [0, \hat{m}_n]$), of cardinality greater or equal to $\frac{\mu n}{16}$, as we saw just before. Then, the change-points a and b are estimated as in Theorem 3.1, using the subsamples $\{(X_i, Y_i) : i \in I_1^+\}$ and $\{(X_i, Y_i) : i \in I_1^-\}$, respectively:

$$\mathbb{P}_G\left[|\tilde{a}_n - a| \geq \frac{16y}{\mu n}\Big|E\right] \leq C_0 e^{-y/(8\sigma^2)}$$

and

$$\mathbb{P}_G\left[|\tilde{b}_n - b| \geq \frac{16y}{\mu n}\Big|E\right] \leq C_0 e^{-y/(8\sigma^2)},$$

for all $y > 0$. Finally,

$$\begin{aligned}
\mathbb{P}_G\left[|\tilde{G}_n \Delta G| \geq \frac{y}{n}\right] &\leq \mathbb{P}_G\left[|\tilde{G}_n \Delta G| \geq \frac{y}{n}\Big|E\right] + \mathbb{P}_G[\bar{E}] \\
&\leq 2C_0 e^{-\mu y/(256\sigma^2)} + C_1 n^4 e^{-C_2 \mu n/2},
\end{aligned}$$

where $\bar{E}$ stands for the complement of the event E. $\qquad \square$

5.6. Proof of Theorem 3.12

Now, we consider the random design (RD). For simplicity's sake, we assume that n is even, without loss of generality. As in the proof of Theorem 3.10, let E be the event $\{|\hat{G}_n \Delta G| \leq \mu/2\}$, where $\hat{G}_n$ is the preliminary estimator of G based on the first subsample $\mathcal{D}_0$.

Let $k \in \{1, \ldots, n/2\}$. By a similar reasoning as in Lemma 5.2, conditional on the subsample $\mathcal{D}_0$ and on the event $\{\#I_1^+ = k\}$, the vectors $(X_i, i \in I_1^+)$ and $(\xi_i, i \in I_1^+)$ are independent, the conditional distribution of $(\xi_i, i \in I_1^+)$ is the same as the unconditional distribution of $(\xi_i, i = 1, \ldots, k)$ and the conditional distribution of $(X_i, i \in I_1^+)$ is equal to the distribution of the increasing reordering of k i.i.d. random variables uniformly

distributed in $[\hat{m}_n, 1]$. Let $t > 0$. We first write that

$$\mathbb{P}[|\tilde{G}_n \Delta G| \geq t|\mathcal{D}_0] \leq \mathbb{P}[|\tilde{G}_n \Delta G| \geq t|\mathcal{D}_0]\mathbb{1}(E) + \mathbb{1}(\bar{E}), \qquad (5.35)$$

where $\mathbb{1}(E)$ is one if E is satisfied, 0 otherwise. Note that if E is satisfied, then $|\tilde{G}_n \Delta G| = |\tilde{a}_n - a| + |\tilde{b}_n - b|$. Hence, (5.35) becomes

$$\mathbb{P}[|\tilde{G}_n \Delta G| \geq t|\mathcal{D}_0] \leq \left(\mathbb{P}[|\tilde{a}_n - a| \geq t/2|\mathcal{D}_0] + \mathbb{P}[|\tilde{b}_n - b| \geq t/2|\mathcal{D}_0]\right)\mathbb{1}(E) + \mathbb{1}(\bar{E}). \qquad (5.36)$$

Let us bound the first term.

$$\mathbb{P}[|\tilde{a}_n - a| \geq t/2|\mathcal{D}_0] = \sum_{k=0}^{n/2} \mathbb{P}[|\tilde{a}_n - a| \geq t/2|\mathcal{D}_0, \#I_1^- = k]\mathbb{P}[\#I_1^- = k|\mathcal{D}_0]. \qquad (5.37)$$

Assume that the event E is satisfied. Then, $\#I_1 \geq N := \#\{i \in I_1 : X_i \geq \mu/4\}$, which is a binomial random variable with parameters $n/2, 1 - \mu/4$. Denote by $\lambda = (1 - \mu/4)/4$. By Hausdorff inequality, for all integers $k \leq n\lambda$,

$$\mathbb{P}[\#I_1^- = k|\mathcal{D}_0] \leq \mathbb{P}[N \leq k]$$
$$\leq e^{-4n\lambda^2}.$$

By the second part of Theorem 3.3, for all $k \geq 1$,

$$\mathbb{P}[|\tilde{a}_n - a| \geq t/2|\mathcal{D}_0, \#I_1^- = k] \leq \begin{cases} 1 \text{ if } t < \frac{A_1\sigma^4}{k}, \\ \Psi\left(\frac{kt}{\sigma^2} - A_1\sigma^2\right) \text{ otherwise,} \end{cases} \qquad (5.38)$$

where we denoted by $\Psi(x) = e^{-\frac{x}{A_2\sqrt{x}+A_3\sigma^2}}$. In the sequel, we assume that $t \geq (1 + \eta)\frac{A_1\sigma^4}{\lambda n}$, where $\eta > 0$ is a constant that we will choose later. Then, for all $k > n\lambda$, it holds that $t \geq \frac{A_1\sigma^4}{k}$ and we can use the inequality of the second case in (5.38). Hence, it follows from (5.37) that

$$\mathbb{P}[|\tilde{a}_n - a| \geq t/2|\mathcal{D}_0] \leq \sum_{k=0}^{\lambda n} e^{-4n\lambda^2} + \sum_{k=\lambda n}^{n/2} \Psi\left(\frac{kt}{\sigma^2} - A_1\sigma^2\right)\mathbb{P}[\#I_1^- = k|\mathcal{D}_0]. \qquad (5.39)$$

Now, we choose the constant $\eta = \frac{A_3\sigma^2}{A_2A_1}$, so that, for $k \geq \lambda n$ and $x = \frac{kt}{\sigma^2} - A_1\sigma^2$, $\Psi(x) \leq e^{-\frac{\sqrt{x}}{2A_3\sigma^2}}$. In addition, still for $k \geq \lambda n$, $x \geq \frac{\eta}{1+\eta}\frac{kt}{\sigma^2}$. Hence, (5.39) becomes

$$\mathbb{P}[|\tilde{a}_n - a| \geq t/2|\mathcal{D}_0] \leq \lambda n e^{-4n\lambda^2} + \sum_{k=\lambda n}^{\infty} e^{-C_1\sqrt{kt}}\mathbb{P}[\#I_1^- = k|\mathcal{D}_0]$$
$$\leq \lambda n e^{-4n\lambda^2} + e^{-C_1\sqrt{\lambda nt}}, \qquad (5.40)$$

where C_1 is a positive constant that depends on σ^2 only.

The same reasoning yields, still under the assumption that the event E holds,

$$\mathbb{P}[|\tilde{b}_n - b| \geq t/2|\mathcal{D}_0] \leq \lambda_1 n e^{-4n\lambda_1^2} + e^{-C_1\sqrt{\lambda_1 n t}}, \tag{5.41}$$

for all $t \geq (1+\eta)A_1\sigma^4/(\lambda_1 n)$, where $\lambda_1 = \mu/16$. Finally, (5.36) becomes

$$\mathbb{P}[|\tilde{G}_n \Delta G| \geq t|\mathcal{D}_0] \leq B_1 e^{-B_2 n} + 2e^{-C_3\sqrt{nt}} + \mathbb{1}\left(\bar{E}\right), \tag{5.42}$$

for all $t \geq C_4/n$, where C_1, C_2, C_3 and C_4 are positive constants that depend on μ and σ^2 only. Since Theorem 3.6 yields $\mathbb{P}_G[E] \geq 1 - C_1 n^4 e^{-C_2\mu n/2}$, taking the expectation in (5.42) with respect to $\mathcal{D}_0$ gives the desired result. $\qquad\square$

REFERENCES

[1] J. Antoch and M. Hušková, Bayesian-type estimators of change points. Prague Workshop on Perspectives in Modern Statistical Inference: Parametrics, Semi-parametrics, Non-parametrics 1998. *J. Stat. Plan. Inference* **91** (2000) 195–208.

[2] A. Betken, Testing for change-points in long-range dependent time series by means of a self-normalized Wilcoxon test. *J. Time Ser. Anal.* **37** (2016) 785–809.

[3] A.A. Borovkov and Yu.Yu. Linke, Asymptotically optimal estimates in the smooth change-point problem. *Math. Method. Stat.* **13** (2004) 1–24.

[4] V.-E. Brunel, Adaptive estimation of convex polytopes and convex sets from noisy data. *Electron. J. Stat.* **7** (2013) 1301–1327.

[5] J. Chen and A.K. Gupta, Statistical inference of covariance change points in gaussian model. *Statistics* **38** (2004) 17–28.

[6] H.P. Chan and G. Walther, Detection with the scan and the average likelihood ratio. *Stat. Sin.* **23** (2013) 409–428.

[7] T.T. Cai and M. Yuan, *Rate-optimal detection of very short signal segments*. Technical report (2014).

[8] K. Frick, A. Munk and H. Sieling, Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** (2014) 495–580. With 32 discussions by 47 authors and a rejoinder by the authors.

[9] S. Gaïffas, Convergence rates for pointwise curve estimation with a degenerate design. *Math. Method. Stat.* **14** (2005) 1–27.

[10] G. Gayraud, Minimax hypothesis testing about the density support. *Bernoulli* **7** (2001) 507–525.

[11] K. Gadeikis and V. Paulauskas, On the estimation of a changepoint in a tail index. *Liet. Mat. Rink.* **45** (2005) 333–348.

[12] I.A. Ibragimov and R.Z. Khasminskii, Estimation of the parameter of a discontinuous signal in Gaussian white noise. *Prob. Peredači Inform.* **11** (1975) 31–43.

[13] I.A. Ibragimov and R.Z. Khasminskii, Statistical Estimation: Asymptotic Theory. Springer Science & Business, New York, (2013).

[14] A.P. Korostelev, On minimax estimation of a discontinuous signal (translation of a paper in russian, published in 1986). *Theory Probab. Appl.* **32** (2006) 727–730.

[15] A.P. Korostelëv and A.B. Tsybakov, Asymptotically minimax image reconstruction problems, in Topics in Nonparametric Estimation. Vol. 12 of *Advances in Soviet Mathematics*. American Mathematical Society, Providence, RI (1992) 45–86.

[16] A.P. Korostelëv and A.B. Tsybakov, Minimax theory of image reconstruction. Vol. 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York (1993).

[17] E. Lebarbier, Detecting multiple change-points in the mean of gaussian process by model selection. *Signal Process.* **85** (2005) 717–736.

[18] S. Lee, J. Ha, O. Na and S. Na, The cusum test for parameter change in time series models. *Scand. J. Stat.* **30** (2003) 781–796.

[19] O.V. Lepski and A.B. Tsybakov, Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Rel. Fields* **117** (2000) 17–48.

[20] Y. Ninomiya, Information criterion for Gaussian change-point model. *Stat. Probab. Lett.* **72** (2005) 237–247.

[21] C.-W. Park and W.-C. Kim, Estimation of a regression function with a sharp change point using boundary wavelets. *Stat. Probab. Lett.* **66** (2004) 435–448.

[22] M. Raimondo, Minimax estimation of sharp change points. *Ann. Stat.* **26** (1998) 1379–1397.

[23] Y.S. Son and S.W. Kim, Bayesian single change point detection in a sequence of multivariate normal observations. *Statistics* **39** (2005) 373–387.

[24] X. Shao and X. Zhang, Testing for change points in time series. *J. Am. Stat. Assoc.* **105** (2010) 1228–1240.

[25] A.B. Tsybakov, Introduction to nonparametric estimation. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. *Springer Series in Statistics*. Springer, New York (2009).

[26] Y. Wu, Inference for change-point and post-change mean with possible change in variance. *Seq. Anal.* **24** (2005) 279–302.