Regression level set estimation — the excess mass approach in the fixed design case

Zailong Wang (Ray)

Department of Mathematical Sciences

New Jersey Institute of Technology

Newark, NJ 07102, USA

CAMS Report 0607-03, Fall 2006

Center for Applied Mathematics and Statistics

NJIT

Regression level set estimation — the excess mass approach in the fixed design case

Zailong Wang (Ray)

Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102, USA

Abstract

Considering the d-dimensional nonparametric regression ($d \geqslant 1$) with fixed design, the estimation of regression level sets is studied based on the maximization of empirical excess masses. By applying modern marked empirical process theory to excess masses, consistency and rates of convergence for the resulting estimators are studied under the conditions on bracketing numbers and modulus of continuity, and a functional central limit theorem for standardized excess mass processes is derived. The consistency and rates of convergence for test statistic of multi-modality are also presented.

Key words: Excess mass; Empirical process; Bracketing number; Modulus of continuity; Image segmentation.
1991 MSC: 62G05, 62H35

Email address: zailong.wang@yahoo.com, Tel: 862-778-8855, Fax: 973-781-4663 (Zailong Wang (Ray)).

1 Introduction

For regression analysis, there are two scenarios on how the data $\{(X_i, Y_i)\}_{i=1}^n$ have been generated. The first setting is concerned with independent, identically distributed (iid) random variables $\{(X_i, Y_i)\}_{i=1}^n$ referred as the random design regression. [3,23] considered the estimation of regression level sets under this setting. The second is concerned with controlled, non-stochastic X-variables. By contrast, this is called the fixed design regression. In this setting, the design points $X_1, \cdots, X_n$ and the responses $Y_1, \cdots, Y_n$ are related as:

$$Y_i = m(X_i) + \sigma\varepsilon_i, \quad 1 \leqslant i \leqslant n, \tag{1.1}$$

where $\varepsilon_i, 1 \leqslant i \leqslant n$, is a sequence of iid random variables with mean 0 and variance 1, and m is a real-valued function defined on the real space R^d.

The fixed design approach is meaningful in many practical situations where there is a controlled input X_i (a fixed grid, a signal message, a period of time, controlled temperature and pressure, etc.) which produces a random output Y_i disturbed by random noise. In particular, chemometrics offers a huge potential application of functional regression under fixed design [9]. Another well-known example for fixed design model is the study of human growth curves where the X-variable has been determined well in advance by a team of pediatricians [10]. A number of consistent and robust estimators for linear regression models with fixed designs have been studied in the literature [25,6,28,24]. An overview of most nonparametric techniques is also given in [12].

Although the estimation of the regression function $m(x)$ is of particular interest in statistics [8], the estimation of location and size of a peak of a regression function has received increasing attention in recent years [18,1]. A more general statistical problem is the estimation of regression level sets of m at the

level λ, which is defined as

$$C(\lambda) := \overline{\{x : m(x) > \lambda\}}, \tag{1.2}$$

where for a set $A \subset \mathbf{R}^d$ we let $\bar{A}$ denote its closure. Since every level set (corresponding to $\lambda < \sup m(x)$) contains at least one of the modes, estimating level sets implicitly provides information on the modes.

The approach based on level sets can be employed to develop a nonparametric quality control tool with a set base approach similar to that proposed by Devroye and Wise [7]. Pattern Analysis and image reconstruction are often related with the support and level set estimation (see [4,16] and simulation results in section 5).

For estimation of regression level sets, we utilize the usual *excess mass* approach introduced by Müller and Sawitzki [17] and Hartigan [11] independently. In this paper it is shown that the excess mass approach as a whole can be translated to the fixed design regression setting, leading, among others, to consistent estimates of regression level sets. In addition to those parallel to the results in the random design case [23], we extend the consistency of the empirical generalized λ-cluster to the different conditions (Theorem 3.2). Furthermore, we develop the rates of convergence for the estimators of level sets (Theorem 3.3, 3.4) and for multi-modality test statistic (Theorem 4.1, 4.2). An algorithm based on rolling ball method [30] is developed for estimating regression level sets and is applied to image segmentation (section 5).

Other statistical theories and applications of level sets include Klemelä [14,15], who developed level set trees for visualization of multivariate density estimates and applied penalized excess mass functional to estimation of the support of a multivariate distribution (level set at $\lambda = 0$), and Jang [13], who extended an idea due to [27,2] to construct confidence sets using the asymptotic distribution of loss function, etc. See also [5] for constructing confidence sets in arbitrary

dimensions.

The paper is organized as follows. In the remaining of this section, basic concepts and technical assumptions required for the statement of the main results are given. Section 2 derives the asymptotic behavior of the empirical excess mass. Section 3 studies the asymptotic behavior of the empirical generalized λ-clusters including consistency and rates of convergence. Consistency and rates of convergence for multi-modality test statistic are shown in Section 4. Some simulation results for image segmentation are presented in Section 5. Most of the technical details are deferred to the appendix section.

We close the introduction with basic concepts for estimating the regression level sets with *excess mass* approach. We also state some basic technical assumptions which are required for the statement of the main results in following sections.

Excess mass: Let $\mathbb{C}$ be a class of measurable subsets on R^d. Let F denote the measure for fixed design X and G the distribution function of ε. Let

$$I(C) = \int_C m(x)\,\mathrm{d}F(x), \qquad C \in \mathbb{C} \tag{1.3}$$

denote the integrated regression functional. By a measure-theoretic argument, the function I uniquely determines m provided $\mathbb{C}$ is rich enough. Hence one is able to draw certain conclusions about m from an analysis of I. In the sequel we always assume that:

Assumption 1 $\emptyset \in \mathbb{C}$ *and* $\Omega = \cup_{C\in\mathbb{C}} C \in \mathbb{C}$ *with* $F(\Omega) = 1$. *Furthermore there exists a constant* $M > 0$, *such that*

$$\sup_{x\in\Omega} |m(x)| \leqslant M < \infty. \tag{1.4}$$

With this assumption, we have

$$\sup_{C\in\mathbb{C}} |I(C)| \leqslant M \cdot F(\Omega) \leqslant M < \infty. \tag{1.5}$$

The next two definitions are central.

Definition 1.1 *For $\lambda \in R$, let $K_\lambda(C) := I(C) - \lambda F(C)$, $C \in \mathbb{C}$.* The excess mass over $\mathbb{C}$ *(at level λ) is defined as:*

$$\mathbb{E}_{\mathbb{C}}(\lambda) := \sup_{C\in\mathbb{C}}\{K_\lambda(C)\}.$$

Every set $\Gamma_{\mathbb{C}}(\lambda) \in \mathbb{C}$ with $\mathbb{E}_{\mathbb{C}}(\lambda) = K_\lambda(\Gamma_{\mathbb{C}}(\lambda))$ is called generalized λ-cluster.

Definition 1.2 *The excess mass functional of $m(x)$ is defined as:*

$$\lambda \to \mathbb{E}(\lambda) := \int (m(x) - \lambda)^+ d\,F(x).$$

The connection between the objects just defined and the regression level sets is that, once $C(\lambda) \in \mathbb{C}$, then $\mathbb{E}_{\mathbb{C}}(\lambda) = \mathbb{E}(\lambda)$. In this case, $C(\lambda)$ is a generalized λ-cluster, which guarantees the existence of $\Gamma_{\mathbb{C}}(\lambda)$. If in addition uniqueness of generalized λ-clusters is assumed then the assumption $C(\lambda) \in \mathbb{C}$ implies that generalized λ-clusters only depend on m. Uniqueness will be assumed for the theoretical results.

Assumption 2 *For all λ there exists a generalized λ-cluster $\Gamma_{\mathbb{C}}(\lambda)$ which is unique up to F-nullsets.*

Assumption 2 is fulfilled if for instance $C(\lambda) \in \mathbb{C}$, and m has no flat part (see Lemma 3.2). In general, however, Assumption 2 does not hold. Consider, for instance, a one-dimensional bimodal regression function symmetric around zero. If we choose $\mathbb{C}$ to be the class of intervals, then, if λ is larger than the antimode (and smaller than the maximum of m), so that the corresponding level sets consist of unions of exactly two intervals, then each of these two intervals is a generalized λ-cluster in $\mathbb{C}$.

Empirical excess mass: For given $(X_i, Y_i), 1 \leqslant i \leqslant n$, the empirical version of $I(C)$ is:

$$I_n(C) = n^{-1} \sum_{i=1}^{n} Y_i \mathbf{1}_{\{X_i \in C\}}. \tag{1.6}$$

Then by replacing the unknown functions I by its empirical versions I_n, we obtain the *empirical* excess mass over $\mathbb{C}$:

Definition 1.3 *The empirical excess mass function over $\mathbb{C}$ is defined as*

$$\mathbb{E}_{n,\mathbb{C}}(\lambda) := \sup_{C \in \mathbb{C}} \{K_{n,\lambda}(C)\}, \tag{1.7}$$

where

$$K_{n,\lambda}(C) := I_n(C) - \lambda F(C), \qquad C \in \mathbb{C}.$$

Every set $\Gamma_{n,\mathbb{C}}(\lambda) \in \mathbb{C}$ with

$$\mathbb{E}_{n,\mathbb{C}}(\lambda) = K_{n,\lambda}(\Gamma_{n,\mathbb{C}}(\lambda)) \tag{1.8}$$

is called an empirical *generalized λ-cluster.*

The empirical generalized λ-cluster serves as an estimator of the generalized λ-cluster (see section 4). The existence of empirical generalized λ-clusters is guaranteed for many classes $\mathbb{C}$ which consist of closed sets such as all closed convex sets [23]. For the results formulated in section 4, we also need the following assumption.

Assumption 3 *For all λ there exists an empirical generalized λ-cluster $\Gamma_{n,\mathbb{C}}(\lambda)$.*

We would like to mention that in [3] an empirical excess mass is defined similar to (1.7) but with F being replaced by a kernel estimate of F. However in our case here F is known. Recall that if $C(\lambda) \in \mathbb{C}$ then $C(\lambda)$ is a generalized λ-cluster. In such a case, the existence of a generalized λ-cluster is guaranteed, and estimation of the generalized λ-clusters means estimation of the regression level sets.

Remark on measurability: If the random quantities considered in the sequel are

not measurable the results given below are still valid when using convergence in outer probability. The results given below are formulated as if everything were measurable.

2 Asymptotic Behavior of Excess Masses

In this section we first derive consistency of $\mathbb{E}_{n,\mathbb{C}}$ and then a functional central limit theorem for its standardization.

2.1 Consistency

Since $\mathbb{E}_{n,\mathbb{C}}$ involves I_n we first study properties of I_n. For this purpose we use the well-known concept of *bracketing number* of $\mathbb{C}$ in $L_1(F)$ which is defined as follows. For a general definition of bracketing [29].

$$\begin{aligned} N_I(\varepsilon, \mathbb{C}, L_1(F)) := & \, min\{n : \exists \text{ a set } \mathbb{D} := \{C_1, ..., C_n\} \subset \mathbb{C}, \text{ such that} \\ & \text{for every } C \in \mathbb{C}, \text{ there exist } C_l \in \mathbb{D}, C^u \in \mathbb{D} \\ & \text{with } C_l \subseteq C \subseteq C^u \text{ and } d_F(C^u,\ C_l) < \varepsilon\}. \end{aligned} \tag{2.1}$$

where $d_F(C,\ D) := F(C \,\Delta\, D)$ and $C \,\Delta\, D := (C \setminus D) \cup (D \setminus C)$ denotes set-theoretic symmetric difference. Then the logarithm of the bracketing number

$$H_I(\varepsilon, \mathbb{C}, L_1(F)) := \log N_I(\varepsilon, \mathbb{C}, L_1(F)) \tag{2.2}$$

is called the metric entropy with bracketing.

From the property of I_n, it is not difficult to induce the following results:

Lemma 2.1 (Consistency) *For any class $\mathbb{C}$ we have*

$$\sup_{\lambda \in R} |\mathbb{E}_{n,\mathbb{C}}(\lambda) - \mathbb{E}_{\mathbb{C}}(\lambda)| \leqslant \sup_{C \in \mathbb{C}} |I_n(C) - I(C)|.$$

Hence if $N_I(\varepsilon, \mathbb{C}, L_1(F)) < \infty$ for every $\varepsilon > 0$, then we have

$$\lim_{n\to\infty} \sup_{\lambda\in R} \mid \mathbb{E}_{n,\mathbb{C}}(\lambda) - \mathbb{E}_{\mathbb{C}}(\lambda) \mid \overset{a.s.}{=} 0.$$

2.2 Functional central limit theorem

In this section, we study the asymptotic behavior of the standardized empirical excess mass

$$n^{1/2}(\mathbb{E}_{n,\mathbb{C}}(\lambda) - \mathbb{E}_{\mathbb{C}}(\lambda)). \tag{2.3}$$

Notice that the error in estimating $\mathbb{E}_{\mathbb{C}}(\lambda) = K_\lambda(\Gamma_{\mathbb{C}}(\lambda))$ by $\mathbb{E}_{n,\mathbb{C}}(\lambda) = K_{n,\lambda}(\Gamma_{n,\mathbb{C}}(\lambda))$ comes from two sources: the replacement of I by its empirical version I_n, and the estimation of $\Gamma_{\mathbb{C}}(\lambda)$ through $\Gamma_{n,\mathbb{C}}(\lambda)$. Ignoring the estimation of $\Gamma_{\mathbb{C}}(\lambda)$, we consider $\mathbb{E}^*_{n,\mathbb{C}}(\lambda) = K_{n,\lambda}(\Gamma_{\mathbb{C}}(\lambda))$ at this moment. Then we have $\mathbb{E}^*_{n,\mathbb{C}}(\lambda) - \mathbb{E}_{\mathbb{C}}(\lambda) = (I_n - I)(\Gamma_{\mathbb{C}}(\lambda))$. Hence, if we define

$$S_n := \sqrt{n}(I_n - I), \tag{2.4}$$

then we have

$$\sqrt{n}\ (\mathbb{E}^*_{n,\mathbb{C}}(\lambda) - \mathbb{E}_{\mathbb{C}}(\lambda)) = S_n(\Gamma_{\mathbb{C}}(\lambda)), \tag{2.5}$$

and asymptotic normality of the left hand side follows from Lemma 2.2 which is directly from Theorem 3.3 of [23].

Lemma 2.2 *Suppose $\sigma^2 < \infty$. If the bracketing integral*

$$J_I(1, \mathbb{C}, L_1(F)) := \int_0^1 \sqrt{H_I(\varepsilon, \mathbb{C}, L_1(F))}\, d\varepsilon < \infty \tag{2.6}$$

then we have

$$S_n \xrightarrow{L} \mathbb{Z} \qquad as \quad n \to \infty,$$

where $\{\mathbb{Z}(C) :\ C \in \mathbb{C}\}$ is a mean zero Gaussian process with

$$E\mathbb{Z}(C_1)\mathbb{Z}(C_2) = E(Y^2 \mathbf{1}_{\{X\in C_1C_2\}}) - I(C_1)I(C_2).$$

The next result concerns the excess mass process. For a set Λ let $\mathbb{C}(\Lambda)$ denote the space of all real-valued continuous functions on Λ equipped with the sup-norm $\|.\|_\Lambda$.

Theorem 2.1 *Let $\Lambda \subset R$ be compact. Suppose that $(\mathbb{C}, d_F)$ is complete and that the assumptions of Lemma 2.2 hold. Then*

$$n^{1/2}(\mathbb{E}_{n,\mathbb{C}}(\lambda) - \mathbb{E}_{\mathbb{C}}(\lambda)) \xrightarrow{L} \mathbb{Z}_{\mathbb{C}}(\lambda) \quad as \quad n \to \infty, \tag{2.7}$$

in $\mathbb{C}(\Lambda)$, where $\mathbb{Z}_{\mathbb{C}}(\lambda)$ is a Gaussian process with mean zero and covariance function

$$\sigma(\lambda_i, \lambda_j) = E(Y^2 \mathbf{1}_{\{X \in \Gamma_{\mathbb{C}}(\lambda_i)\Gamma_{\mathbb{C}}(\lambda_j)\}}) - I(\Gamma_{\mathbb{C}}(\lambda_i))I(\Gamma_{\mathbb{C}}(\lambda_j)). \tag{2.8}$$

Remark: The covariance structure (2.8) is much less complex than that in stochastic design cases as in [23]. [22] showed that for the density case the covariance would be $F(\Gamma_{\mathbb{C}}(\lambda_i)\Gamma_{\mathbb{C}}(\lambda_j)) - F(\Gamma_{\mathbb{C}}(\lambda_i))F(\Gamma_{\mathbb{C}}(\lambda_j))$, and if in addition the generalized λ-clusters are nested, the limiting process is a Brownian bridge with transformed time axis.

3 Empirical Generalized λ-Cluster

In this section we mainly study the asymptotic behavior of $\Gamma_{n,\mathbb{C}}(\lambda)$ as an estimator of $\Gamma_{\mathbb{C}}(\lambda)$. The consistency results are given in subsection 4.1 and the rates of convergence are derived in subsection 4.2.

3.1 Consistency

Recall that if the regression level sets lie in the considered class $\mathbb{C}$, then they are generalized λ-clusters in $\mathbb{C}$. Hence, in such a case, the existence of the generalized λ-clusters is guaranteed, and estimation of the generalized λ-clusters

means estimation of the regression level sets. However, we do not necessarily want to assume the regression level sets to lie in $\mathbb{C}$. This case can then be interpreted as a situation where the chosen model (determined by the class $\mathbb{C}$) is not necessarily correct. Let us first consider the case where $C(\lambda)$ is not necessarily assumed to lie in $\mathbb{C}$.

Theorem 3.1 *Suppose that the space* $(\mathbb{C}, d_F)$ *is complete and* $N_I(\varepsilon, \mathbb{C}, L_1(F)) < \infty$ *for every* $\varepsilon > 0$*. Then we have with probability 1 that*

$$\sup_{\lambda \in R} d_F(\Gamma_{\mathbb{C}}(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) \to 0, \qquad as \;\; n \to \infty.$$

Remarks: In the density case, consistency of the empirical generalized λ-cluster was shown in [22]. [23] showed the consistency property for random design case. [11] considered the class of all closed convex sets in R^2 equipped with the Hausdorff metric and [19] used the class of ellipsoids in R^d in a more parametric setup.

Next we consider the case where $C(\lambda) \in \mathbb{C}$. This additional assumption allows us to derive explicit upper bounds for $d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda))$, which are the key for deriving rates of convergence in the case $C(\lambda) \in \mathbb{C}$. Following Lemma gives the explicit upper bounds for $d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda))$.

Lemma 3.1 *Suppose* $C(\lambda) \in \mathbb{C}$ *for a fixed* $\lambda \in R$*. Then the following inequality holds for every* $\eta > 0$*:*

$$\begin{aligned} d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) \leqslant & F\{x : |m(x) - \lambda| < \eta\} \\ & + \eta^{-1}[(I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda)) - (I_n - I)(C(\lambda))]. \end{aligned} \tag{3.1}$$

For the following conclusion, we need the assumption that "m has no flat part" at certain levels. We say that: *m has no flat part at the level λ if and only if* $F\{x : m(x) = \lambda\} = 0$. If m has no flat part for all $\lambda \in \Lambda$, then we say that *m has no flat part in* Λ. The following lemma shows the properties of m

with no flat part.

Lemma 3.2 *Let $\Lambda \subset R$ be a closed subset. Then the following statements are equivalent:*

(i) *m has no flat part in Λ.*

(ii) $\sup_{\lambda\in\Lambda} F\{x : |m(x) - \lambda| < \eta\} \to 0, \qquad$ *as* $\eta \to 0$.

(iii) *$\lambda \to C(\lambda)$ is uniformly continuous in Λ for the d_F-pseudometric.*

The following result is directly from (3.1) and Lemma 3.1 of [23].

Theorem 3.2 *Let $\Lambda \subset R$ be a closed set such that $C(\lambda) \in \mathbb{C}$ for all $\lambda \in \Lambda$. If $N_I(\varepsilon, \mathbb{C}, L_1(F)) < \infty$ for every $\varepsilon > 0$, and*

$$\sup_{\lambda\in\Lambda} F\{x : \ |m(x) - \lambda| < \eta\} \to 0, \qquad \text{as } \eta \to 0, \tag{3.2}$$

then we have, with probability 1, that

$$\sup_{\lambda\in\Lambda} d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) \to 0, \qquad \text{as } n \to \infty.$$

3.2 Rates of convergence

In order to derive the rates of convergence for $d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda))$, we define *the modulus of continuity of g* as

$$\omega_g(\delta, \mathbb{C}) := \sup\{|g(C) - g(D)| : C, D \in \mathbb{C}, d_F(C, D) < \delta\},$$

for a function $g : \mathbb{C} \to R$, where $\mathbb{C}$ is equipped with the d_F-pseudometric.

Theorem 3.3 *Let $\{\alpha_n\}$ be a sequence of positive real numbers with $\alpha_n \to \infty$ as $n \to \infty$ and let r be a positive real number. Assume $\Lambda \subset R$ be a closed set such that $C(\lambda) \in \mathbb{C}$ for all $\lambda \in \Lambda$. If $N_I(\varepsilon, \mathbb{C}, L_1(F)) < \infty$ for every $\varepsilon > 0$, and the following two conditions hold:*

$$\sup_{\lambda\in\Lambda} F\{x : |m(x) - \lambda| < \eta\} = O(\eta^r), \qquad \text{as } \eta \to 0, \tag{3.3}$$

$$\forall \eta > 0: \quad \lim_{\delta \to 0} \limsup_{n \to \infty} P\{\alpha_n \omega_{I_n - I}(\delta, \mathbb{C}) > \eta\} = 0; \tag{3.4}$$

then we have

$$\sup_{\lambda \in \Lambda} d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) = O_P(\alpha_n^{-r/(1+r)}), \qquad as\ n \to \infty.$$

It follows from Lemma 2.2 that condition (3.4) is fulfilled with $\alpha_n = n^{1/2}$ if conditions in Lemma 2.2 are satisfied. The reason for this is that Lemma 2.2 implies the asymptotic equicontinuity of $\{S_n(C), C \in \mathbb{C}\}$ (see the definition of S_n in (2.4)), which means that for $\forall\ \eta > 0$ we have

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P\{\omega_{S_n}(\delta, \mathbb{C}) > \eta\} = 0. \tag{3.5}$$

Therefore we immediately obtain the following corollary:

Corollary 3.1 *Suppose $E(Y^2 \mid X) \leqslant K^2$. Let r be a positive real number and $\Lambda \subset R$ be a closed set such that $C(\lambda) \in \mathbb{C}$ for all $\lambda \in \Lambda$. If the conditions* (3.3) *and* (2.6) *hold, then we have*

$$\sup_{\lambda \in \Lambda} d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) = O_P(n^{-r/2(1+r)}), \qquad as\ n \to \infty.$$

The following lemma shows that additional knowledge on the behavior of the modulus of continuity of the modern marked empirical process (see (3.7) below) can be used to accelerate the rates.

Lemma 3.3 *Assume that* (3.3) *holds with $r > 0$ and that $C(\lambda) \in \mathbb{C}$ for all $\lambda \in \Lambda$. Let $\{\delta_n\}$ be a sequence of positive real numbers converging to 0 as $n \to \infty$ and g be a real-valued function with $g(\delta) \to 0$ as $\delta \to 0$ such that the following two conditions hold:*

$$\sup_{\lambda \in \Lambda} d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) = O_P(\delta_n), \qquad as\ n \to \infty; \tag{3.6}$$

$$\omega_{S_n}(\delta_n, \mathbb{C}) = O_P(g(\delta_n)), \qquad as\ n \to \infty. \tag{3.7}$$

Then we have

$$\sup_{\lambda\in\Lambda} d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) = O_P(n^{-1/2} g(\delta_n))^{r/(1+r)}, \qquad as\ n \to \infty.$$

The Lemma 3.3 can be used iteratively. The following lemma shows the faster rate which could be almost obtained from an iteration of Lemma 3.3 is $O_P(n^{-r^*/(2(1-\beta r^*))})$, where $r^* = r/(1+r)$.

Lemma 3.4 *Suppose the conditions of Lemma 3.3 hold with* $\delta_n = n^{-\alpha}$, $\alpha > 0$, *and* $g(\delta) = \delta^\beta (\log \delta^{-1})^\eta$, $\beta > 0$, $\eta \geqslant 0$. *If* $\beta r^* < 1$, $\eta r^* < 1$ *where* $r^* = r/(1+r)$, *then we have for every* $\varepsilon > 0$ *that*

$$\sup_{\lambda\in\Lambda} d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) = O_P(n^{-r^*/(2(1-\beta r^*))+\varepsilon}), \qquad as\ n \to \infty.$$

Theorem 3.3 and Corollary 3.1 can be used to obtain a starting rate for the iteration. The following lemma shows that (3.7) is satisfied for *Vapnik-Cervonenkis (VC)-classes* [21] under some conditions on Y and δ_n.

Lemma 3.5 *If the class* $\mathbb{C}$ *is a VC-class and* $E(Y^2) = K^2$, $E(Y^6) < \infty$, *then for* $\delta_n (\to 0) \geqslant n^{-2/3} \ln n/(K+2M)^2$, *we have*

$$\omega_{S_n}(\delta_n, \mathbb{C}) = O_P(\delta_n^{1/4} (\ln \delta_n^{-1})^{1/2}); \qquad as \quad n \to \infty.$$

Now we can easily formulate the main theorem for the rate of convergence from results above:

Theorem 3.4 *Suppose conditions in Corollary 3.1 and Lemma 3.5 hold, then we have for every* $\varepsilon > 0$ *that*

$$\sup_{\lambda\in\Lambda} d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) = O_P(n^{-r^*/(2(1-r^*/4))+\varepsilon}), \qquad as\ n \to \infty.$$

Example: For levels λ where $\|\mathrm{grad}\, m(x)\|$ is bounded away from zero in a neighborhood of $\{x : m(x) = \lambda\}$ we have $r = 1$ and $r^* = 1/2$. Let m be a

smooth unimodal regression function. Then, for $d = 1$, the regression level sets are intervals. Therefore we choose $\mathbb{C} = I$ (the class of all intervals). For $d \geqslant 2$, we assume the regression level sets to be boxes, balls or ellipsoids. For these situations we obtain from Theorem 3.4 that

$$d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) = O_P(n^{-2/7+\varepsilon}).$$

4 Testing for Multimodality

In this section we study the testing for multimodality problem. Let $\mathbb{C}_0$ and $\mathbb{C}$ be two classes of measurable subsets of R^d with $\mathbb{C}_0 \subset \mathbb{C}$, and let $\Lambda \subset R$. One can consider the hypothesis:

$$H_0 : \Gamma_{\mathbb{C}}(\lambda) \in \mathbb{C}_0 \qquad \text{for all} \quad \lambda \in \Lambda.$$

That is the generalized λ-clusters in $\mathbb{C}$ already lie in the subset class $\mathbb{C}_0$. Let $\Delta_n(\mathbb{C}_0, \mathbb{C}, \lambda) = \mathbb{E}_{n,\mathbb{C}}(\lambda) - \mathbb{E}_{n,\mathbb{C}_0}(\lambda)$. Then the test statistic for this testing problem is ([22]):

$$T_n(\mathbb{C}_0, \mathbb{C}, \Lambda) = \sup_{\lambda \in \Lambda} \Delta_n(\mathbb{C}_0, \mathbb{C}, \lambda).$$

The quantity $\Delta_n(\mathbb{C}_0, \mathbb{C}, \lambda)$ is nonnegative for every $\lambda \in \Lambda$. The large values of Δ_n for some λ suggest a violation of the hypothesis H_0.

Let I_m be the class of unions of at most m intervals. If we consider the univariate case and choose $\mathbb{C}_0 = I_1$ and $\mathbb{C} = I_2$, then this testing problem can be regarded as looking for unimodality versus bimodality. For multivariate case, a similar choice can be made such as $\mathbb{C}_0 = \mathcal{B}_k^d$ and $\mathbb{C} = \mathcal{B}_m^d$ for $k < m$, where $\mathcal{B}_m^d$ stands for the class of unions of at most m closed convex sets in R^d. Then the problem is concerning about k-modes against the alternative of m-modes.

4.1 Consistency

Define $\Delta(\mathbb{C}_0, \mathbb{C}, \lambda) := \mathbb{E}_{\mathbb{C}}(\lambda) - \mathbb{E}_{\mathbb{C}_0}(\lambda)$ and $T(\mathbb{C}_0, \mathbb{C}, \Lambda) := \sup_{\lambda \in \Lambda} \Delta(\mathbb{C}_0, \mathbb{C}, \lambda)$. Then the following proposition follows from Lemma 2.1 immediately.

Proposition 4.1 *For every choice of $\mathbb{C}_0$ and $\mathbb{C}$ we have*

$$\sup_{\lambda \in R} |\Delta_n(\mathbb{C}_0, \mathbb{C}, \lambda) - \Delta(\mathbb{C}_0, \mathbb{C}, \lambda)| \leqslant \sup_{C \in \mathbb{C}} \|I_n(C) - I(C)\| + \sup_{C \in \mathbb{C}_0} \|I_n(C) - I(C)\|.$$

Hence if $N_I(\varepsilon, \mathbb{C}, L_1(F)) < \infty$ for every $\varepsilon > 0$, then we have for any $\Lambda \subset R$,

$$|T_n(\mathbb{C}_0, \mathbb{C}, \Lambda) - T(\mathbb{C}_0, \mathbb{C}, \Lambda)| \stackrel{a.s.}{\to} 0, \qquad as \quad n \to \infty.$$

If in addition H_0 holds, then we have

$$T_n(\mathbb{C}_0, \mathbb{C}, \Lambda) \stackrel{a.s.}{\to} 0, \qquad as \quad n \to \infty.$$

4.2 Rates of convergence

As in Proposition 4.1 we consider $\sup_{\lambda \in \Lambda} |\Delta_n(\mathbb{C}_0, \mathbb{C}, \lambda) - \Delta(\mathbb{C}_0, \mathbb{C}, \lambda)|$. Rates of convergence for this quantity immediately give upper bound rates for T_n centered at T.

Theorem 4.1 *Let $\Lambda \subset R$ be closed and $\mathbb{C}_0, \mathbb{C}$ be two classes of measurable subsets of R^d with $\mathbb{C}_0 \subset \mathbb{C}$. Assume that (2.6) holds and with probability 1 $\sup_{\lambda \in \Lambda} d_F(\Gamma_{\mathbb{C}_0}(\lambda), \Gamma_{n,\mathbb{C}_0}(\lambda))$ and $\sup_{\lambda \in \Lambda} d_F(\Gamma_{\mathbb{C}}(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) \to 0$ as $n \to \infty$.*

Then we have

$$\sup_{\lambda \in \Lambda} |\Delta_n(\mathbb{C}_0, \mathbb{C}, \lambda) - \Delta(\mathbb{C}_0, \mathbb{C}, \lambda)| = O_P(n^{-1/2}); \qquad as\ n \to \infty.$$

Under H_0 rates of convergence can be given under assumptions on the class $\mathbb{C}$ only:

Theorem 4.2 *Let $\Lambda \subset R$ be closed and $\mathbb{C}_0, \mathbb{C}$ be two classes of measurable subsets of R^d with $\mathbb{C}_0 \subset \mathbb{C}$. Assume that* (2.6) *holds and with probability 1 that* $\sup_{\lambda\in\Lambda} d_F(\Gamma_{\mathbb{C}}(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) \to 0$ *as* $n \to \infty$.

Then we have under H_0 that

$$\sup_{\lambda\in\Lambda} \Delta_n(\mathbb{C}_0, \mathbb{C}, \lambda) = o_P(n^{-1/2}); \qquad \text{as } n \to \infty.$$

One may obtain its asymptotic distribution and faster rates of convergence for fixed or even random design regression cases. A further study on these projects is still on the way. One thing for sure is that if $\mathbb{E}_{\mathbb{C}_0}(\lambda) - \mathbb{E}_{\mathbb{C}}(\lambda) > 0$ for some $\lambda \in \Lambda$, then it follows from Proposition 4.1 that the power of a test based on $T_n(\mathbb{C}_0, \mathbb{C}, \Lambda)$ converges to 1 as n tends to infinity.

5 Simulation Studies

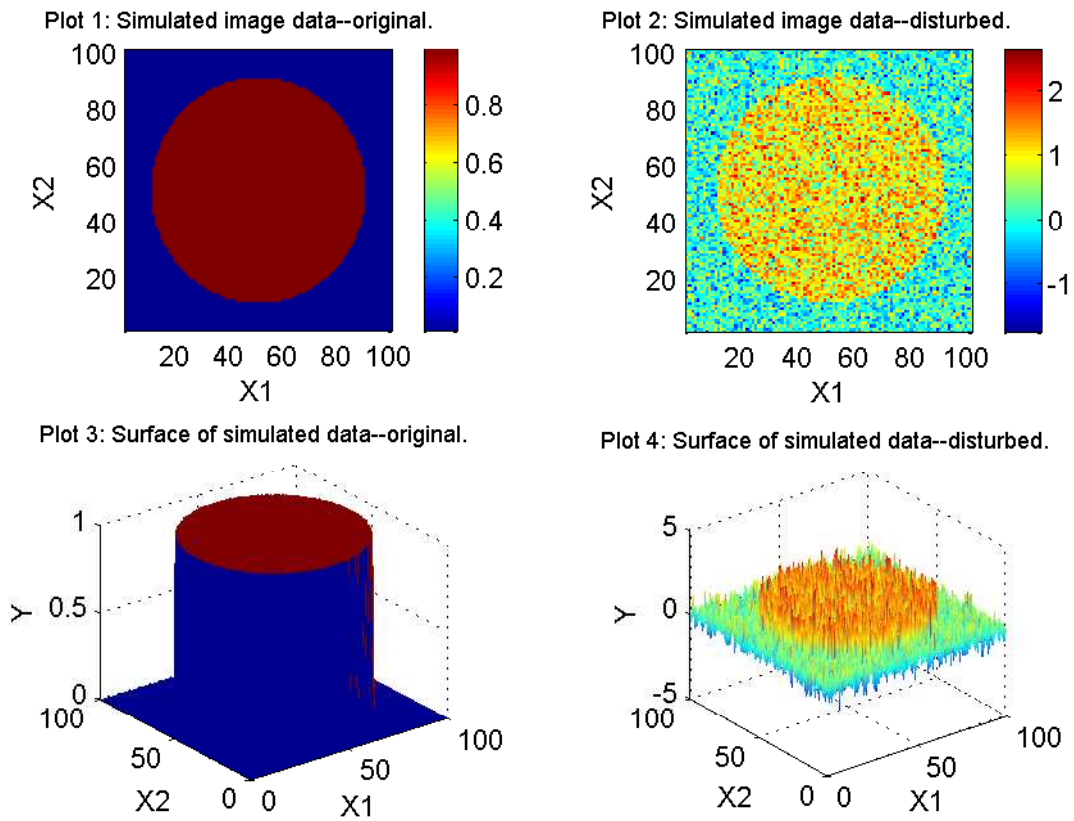
We now present some simulation results for image segmentation with excess mass approach. The problem discussed in this section is intended to give another possible application for fixed design case with excess mass approach.

Image segmentation is the process of subdividing a digit image such as a cDNA microarray scan raw data into homogeneous regions (e.g. background and foreground). This is generally as a prelude to further analysis [26]. For simplicity, we consider that a true image consists of a disc of intensity 1 (foreground) against a background of a lower intensity 0. The image was then disturbed with an additive Gaussian noise $\varepsilon \sim N(0, \sigma^2)$, independently at each pixel on a 100×100 grid. That is, the model considered here is:

$$y = \begin{cases} 1 + \varepsilon, \; if \; (x_1 - a_1)^2 + (x_2 - a_2)^2 < r^2; \\ \\ \varepsilon, \qquad otherwise; \end{cases} \tag{5.1}$$

where $\varepsilon \sim N(0, \sigma^2)$. Figure 1 shows the data plots from model (5.1) with x_1, x_2 from 0 to 100; $a_1 = a_2 = 50, r^2 = 5000/\pi$ and $\sigma = 0.5$. The purpose of image segmentation for this model is equivalent to the estimation of the regression level set at level λ. We consider $\lambda = 0.5$ for this data set.

Fig. 1. Simulated image data set: left plots show data from true mode and right the noise data; lower plots show perspective plots of the same surfaces displayed as image above.



For this purpose, we use the empirical generalized λ-cluster which will be calculated by the following algorithm. The basic idea is from rolling ball method discussed by Walther [30] for density level set estimation. From excess mass point of view, we want to find a set C such that $I_n(C) - \lambda F(C) = n^{-1} \sum_{i=1}^{n} (Y_i - \lambda) \mathbf{1}_{\{X_i \in C\}}$ attains maximum. Since F is uniform over the range of (x_1, x_2), we use $F(C) = n^{-1} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \in C\}}$ with fixed n for this simulation study. Then for given data set, we look for C such that within C, the value of

$\sum(Y_i - \lambda)$ is maximum. Therefore C could be regarded as the union of many tiny sets within which the value of $\sum(Y_i - \lambda)$ is positive. Hence we develop the following algorithm for C.

(1) Begin from $X_{(i_n)}$ corresponding to $Y_{(n)} = \max\{Y_i\}$, that is $C = \{X_{(i_n)}\}$ if $Y_{(n)} \geqslant \lambda$. Let $D = C$.

(2) For each point $x_0 \in D$, calculate the mean values of Y for all tiny sets that contain x_0. Let C_0 be the union of those tiny sets with means greater than λ, then $C' = C \bigcup C_0$.

(3) If $C' = C$ then stop. Otherwise, let $D = C' \backslash C$ and update C by $C = C'$, then go back to step 2.

Remarks: a) The size and shape of tiny sets serve as smoothing parameters for estimation. The larger and smoother the tiny sets is, the smoother the estimated set. That means the variance of the estimation is smaller. However the bias of the estimation will become larger. One needs to choose the appropriate smoothing parameter to balance between bias and variance. In our simulation for image segmentation, we consider the smoothing parameter $s = 2, 3, 4$, and 5 meaning that the tiny sets contain exactly s points including x_0. The remaining $s - 1$ points are the closest to x_0.

b) The algorithm above can give only one connected set C which is feasible for our model. There is no restriction on the shape of the estimated set except for connection. One only need to repeat the algorithm for the remained data set $C^c = \Omega \backslash C$ if there is any hint that the level set may be the union of disjoint sets. Hence the algorithm is quite useful for level set estimation in high-dimensional case.

This algorithm is used in a Monte Carlo simulation for three different cases ($\sigma = 0.2, 0.5, 1$) with 200 repetitions for each case. The simulated results are shown in table 1. Figure 2 gives the estimated image plots for $\sigma = 0.5$. The notations used in the table are defined as follows: $\overline{d_F}$ is the average dis-

tance between true set $\Gamma_{\mathbb{C}}(\lambda)$ and estimated set $\Gamma_{n,\mathbb{C}}(\lambda)$ using pseudo-metric $F(\Gamma_{\mathbb{C}}(\lambda)\Delta\Gamma_{n,\mathbb{C}}(\lambda))$; STD is the standard deviation for d_F; and s is smoothing parameter (see Remarks: a) below the algorithm).

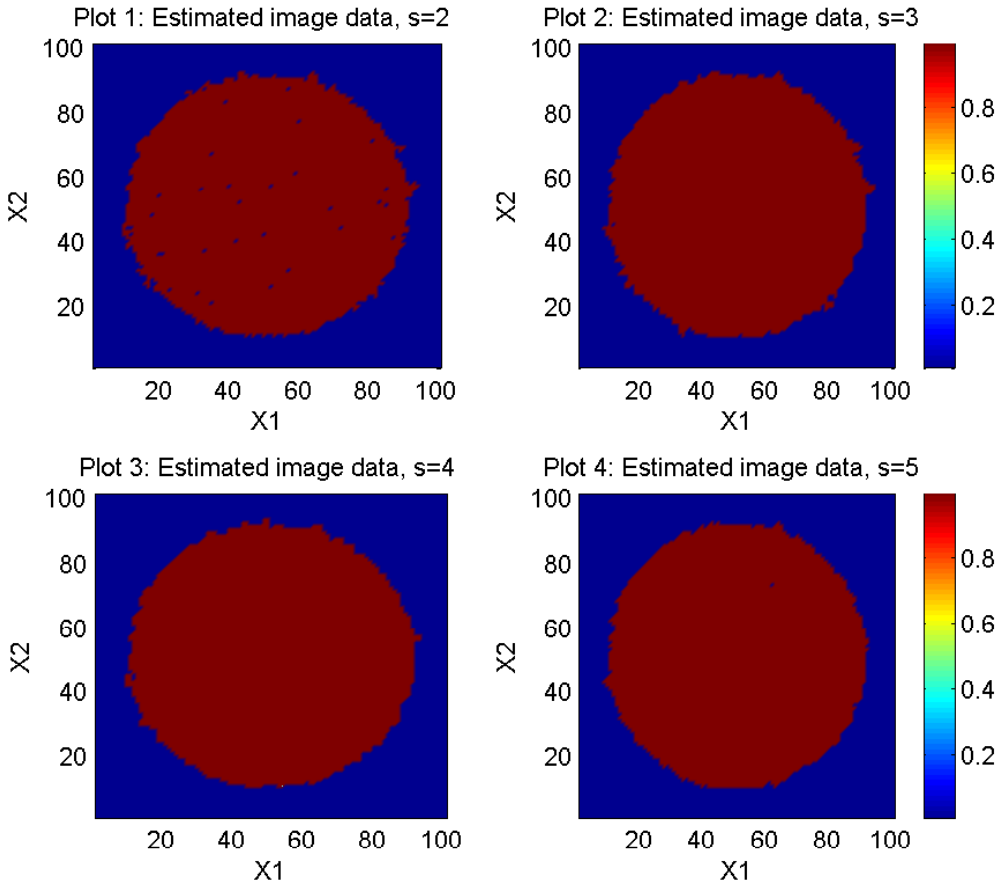
Table 1

Performances of the empirical generalized λ-clusters for model (5.1).

	$\overline{d_F}$				STD			
σ	$s=2$	$s=3$	$s=4$	$s=5$	$s=2$	$s=3$	$s=4$	$s=5$
0.2	0.0126	0.0226	0.0275	0.0230	0.0008	0.0004	0.0005	0.0006
0.5	0.0225	0.0259	0.0318	0.0265	0.0333	0.0018	0.0016	0.0330
0.8	0.0484	0.0458	0.0510	0.0319	0.0317	0.0325	0.0321	0.0329

Table 1 indicates that the smaller the variance of the error distribution, the more satisfactory level set estimate could be obtained as we have expected. That is the mean and standard deviation of $d_F(\Gamma_{\mathbb{C}}(\lambda), \Gamma_{n,\mathbb{C}}(\lambda))$ are smaller. On the other hand, the behavior of smoothing parameter s is little more complicated and varies from one σ to another σ. Generally speaking, smaller s leads to smaller mean and larger standard deviation of d_F especially for $s = 2, 3$, and 4. That means smaller bias and larger variance of estimated level sets. For $s = 5$, however, it even gets smallest mean and largest deviation for $\sigma = 0.8$. The reason may due to the larger σ and the different shapes of tiny sets for different values of parameter s. From Figure 2, we can also see the trend that the larger the value of parameter s, the smoother the estimated images.

Fig. 2. Estimated images (level sets) at $\lambda = 0.5$ with $s = 2, 3, 4, 5$ and $\sigma = 0.5$.



6 Appendix

Proof of Lemma 2.1: Notice that $K_{n,\lambda}(C) = K_\lambda(C) + (I_n - I)(C)$. Hence we have

$$\begin{aligned}|E_{n,\mathbb{C}}(\lambda) - E_{\mathbb{C}}(\lambda)| &= |\sup_{C\in\mathbb{C}} K_{n,\lambda}(C) - \sup_{C\in\mathbb{C}} K_\lambda(C)| \\ &\leqslant \sup_{C\in\mathbb{C}} |K_{n,\lambda}(C) - K_\lambda(C)| \\ &= \sup_{C\in\mathbb{C}} |I_n(C) - I(C)|.\end{aligned}$$

The result hence follows from Lemma 3.1 of [23].$\diamondsuit$

Proof of Theorem 2.1: From equality (2.5), we have

$$\sqrt{n}\,(\mathbb{E}_{n,\mathbb{C}}(\lambda) - \mathbb{E}_{\mathbb{C}}(\lambda)) = S_n(\Gamma_{\mathbb{C}}(\lambda)) + R_n(\lambda)$$

where

$$
\begin{aligned}
0 \leqslant R_n(\lambda) &= \sqrt{n}\,(\mathbb{E}_{n,\mathbb{C}}(\lambda) - \mathbb{E}^*_{n,\mathbb{C}}(\lambda)) \\
&= \sqrt{n}[K_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) - K_\lambda(\Gamma_{\mathbb{C}}(\lambda)) + (I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda)) - (I_n - I)(\Gamma_{\mathbb{C}}(\lambda))] \\
&\leqslant S_n(\Gamma_{n,\mathbb{C}}(\lambda)) - S_n(\Gamma_{\mathbb{C}}(\lambda)).
\end{aligned}
$$

We now argue that $\sup_{\lambda\in\Lambda} |R_n(\lambda)| = o_P(1)$. First notice that Lemma 2.2 implies asymptotic stochastic equicontinity of S_n, meaning that for all $\eta > 0$ we have

$$
\lim_{\delta\to 0} \limsup_{n\to\infty} P\left(\sup_{d_F(C,D)<\delta} \mid S_n(C) - S_n(D) \mid > \eta \right) = 0. \tag{6.1}
$$

Further, Theorem 3.1 provides us with uniform consistency of the generalized λ-cluster. Hence $\sup_{\lambda\in\Lambda} |R_n(\lambda)| = o_P(1)$ from (6.1).

It follows that the asymptotic behavior of the process $\sqrt{n}\,(\mathbb{E}_{n,\mathbb{C}}(\lambda) - \mathbb{E}_{\mathbb{C}}(\lambda))$ is the same as the one of $S_n(\Gamma_{\mathbb{C}}(\lambda))$. Convergence of the finite dimensional distributions of the latter follows directly from the CLT. It remains to prove tightness of $S_n(\Gamma_{\mathbb{C}}(\lambda)), \lambda \in \Lambda$. Since the map $\lambda \to \Gamma_{\mathbb{C}}(\lambda)$ is continuous in the metric d_F (Lemma 6.2), equicontinuity of $S_n(C)$ implies the one of $\sqrt{n}\,(\mathbb{E}_{n,\mathbb{C}}(\lambda) - \mathbb{E}_{\mathbb{C}}(\lambda))$. This completes the proof. $\diamondsuit$

In order to prove Theorem 3.1 we need the following two lemmas stating properties of $K_\lambda(C)$ and $\Gamma_{\mathbb{C}}(\lambda)$.

Lemma 6.1 *(Properties of K_λ)*

(1) If $N_I(\varepsilon, \mathbb{C}, L_1(F)) < \infty$ for every $\varepsilon > 0$, then

$$
\sup_{\lambda\in R} \mid K_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) - K_\lambda(\Gamma_{\mathbb{C}}(\lambda)) \mid \overset{a.s}{\to} 0, \qquad as \;\; n \to \infty.
$$

(2) The function $C \;\to\; K_\lambda(C),\; C \in (\mathbb{C}, d_F)$ is continuous.

Proof of Lemma 6.1:

(1) Notice that

$$
K_\lambda = K_{n,\lambda} - (I_n - I) + \lambda(F_n - F).
$$

From the definition of $\Gamma_{n,\mathbb{C}}(\lambda)$ and $\Gamma_{\mathbb{C}}(\lambda)$ it follows $K_{n,\lambda}(\Gamma_{n,\mathbb{C}}(\lambda)) \geqslant K_{n,\lambda}(\Gamma_{\mathbb{C}}(\lambda))$ and $K_\lambda(\Gamma_{\mathbb{C}}(\lambda)) \geqslant K_\lambda(\Gamma_{n,\mathbb{C}}(\lambda))$. Therefore

$$\begin{aligned}
0 &\leqslant K_\lambda(\Gamma_{\mathbb{C}}(\lambda)) - K_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) \\
&= K_{n,\lambda}(\Gamma_{\mathbb{C}}(\lambda)) - K_{n,\lambda}(\Gamma_{n,\mathbb{C}}(\lambda)) \\
&\quad +(I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda)) - (I_n - I)(\Gamma_{\mathbb{C}}(\lambda)) \\
&\leqslant (I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda)) - (I_n - I)(\Gamma_{\mathbb{C}}(\lambda))
\end{aligned} \tag{6.2}$$

The assertion follows from Lemma 3.1 of [23].

(2) The result follows from the fact that I is dominated by F. $\diamondsuit$

The following property of $\Gamma_{\mathbb{C}}(\lambda)$ is from Lemma 6.3 of [23].

Lemma 6.2 *Suppose that conditions of Theorem 3.1 hold. Then $\lambda \to \Gamma_{\mathbb{C}}(\lambda)$ is uniformly continuous in Λ for the d_F-pseudometric.*

Proof of Theorem 3.1: Using Lemma 6.1 and 6.2, the proof is analog to the proof of Theorem 3.2 and 3.3 of [22]. $\diamondsuit$

Proof of Lemma 3.1: Let $D_{n,\mathbb{C}}(\lambda) = C(\lambda) \triangle \Gamma_{n,\mathbb{C}}(\lambda)$, we have

$$\begin{aligned}
&d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) = F(D_{n,\mathbb{C}}(\lambda)) \\
&= F(D_{n,\mathbb{C}}(\lambda) \cap \{x : |m(x) - \lambda| < \eta\}) \\
&\quad +F(D_{n,\mathbb{C}}(\lambda) \cap \{x : |m(x) - \lambda| \geqslant \eta\}).
\end{aligned} \tag{6.3}$$

The first term is dominated by $F\{x : |m(x) - \lambda| < \eta\}$. As for the second term, notice that, on the one hand,

$$K_\lambda(C(\lambda)) - K_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) \leqslant (I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda)) - (I_n - I)(C(\lambda))$$

from (6.2). On the other hand,

$$\begin{aligned}
&K_\lambda(C(\lambda)) - K_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) = \int_{D_{n,\mathbb{C}}(\lambda)} |m(x) - \lambda| \mathrm{d}F(x) \\
&\geqslant \eta F(D_{n,\mathbb{C}}(\lambda) \cap \{x : |m(x) - \lambda| \geqslant \eta\}).
\end{aligned}$$

Hence

$$
\begin{aligned}
&F(D_{n,\mathbb{C}}(\lambda) \cap \{x : |m(x) - \lambda| \geqslant \eta\}) \\
\leqslant &\eta^{-1}[(I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda)) - (I_n - I)(C(\lambda))]
\end{aligned} \tag{6.4}
$$

and the assertion follows. $\diamondsuit$

Proof of Theorem 3.3: Let $\{\eta_n\}$ be a sequence of real numbers converging to 0 as $n \to \infty$. We have (see (6.3)):

$$
\begin{aligned}
&d_F(C(\lambda), \Gamma_{n,\mathbb{C}}(\lambda)) = F(D_{n,\mathbb{C}}(\lambda)) \\
=&F(D_{n,\mathbb{C}}(\lambda) \cap \{x : |m(x) - \lambda| < \eta_n\}) \\
&+F(D_{n,\mathbb{C}}(\lambda) \cap \{x : |m(x) - \lambda| \geqslant \eta_n\}).
\end{aligned}
$$

The first term is dominated by $F\{x : |m(x) - \lambda| < \eta_n\}$, thus it is of the order $O(\eta_n^r)$ from (3.3). The second term is dominated by $\eta_n^{-1}[(I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda)) - (I_n - I)(C(\lambda))]$ from (6.4), and so it is of the order $O_P((\alpha_n \eta_n)^{-1})$ from (3.4). The assertion hence follows by choosing $\eta_n = \alpha_n^{-1/(1+r)}$. $\diamondsuit$

Proof of Lemma 3.3: Similar to the proof of Theorem 3.3 with $\alpha_n = n^{1/2}(g(\delta_n))^{-1}$.$\diamondsuit$

Proof of Lemma 3.4: It is easy to show that an l-fold iteration of Lemma 3.3 gives the rate

$$
O_P(n^{-h(l,\beta)}(\log n)^{h(l,\eta)}),
$$

where $h(l, c) = (r^*/2)\sum_{j=0}^{l}(cr^*)^j + \alpha(cr^*)^l$, for $c = \beta, \eta$. This iteration can be done arbitrarily (but finitely) often. Since $cr^* < 1$, the quantity $h(l, c)$ converges to $r^*/(2(1 - cr^*))$. The assertion hence follows from ignoring log–term since $\varepsilon > 0$ appears in the exponent of the n-term. $\diamondsuit$

Proof of Lemma 3.5: For the proof of Lemma 3.5, we need a result from following lemma which is extended from Theorem 3.4 of [20]. Denote

$$
\|S_n\|_\delta := \sup_{C,D \in \mathbb{C}:\ \rho(C,D) < \delta} |S_n(C) - S_n(D)|,
$$

where the pseudo-metric $\rho(C, D) := (K + 2M)d_F^{1/2}(C, D)$.

Lemma 6.3 *If the class $\mathbb{C}$ is a VC-class and $E(Y^6) < \infty$, then there exist a*

constant c_0, depending on the VC-index, such that for all $\varepsilon > 0$, $\delta_n(\to 0) \geqslant n^{-1/3}\sqrt{\ln n}$, and $\eta_n^2 = 8c_0(2+\varepsilon)\delta_n \ln(\delta_n^{-1})$, we have

$$P(\|S_n\|_{\delta_n} > \eta_n) \to 0, \qquad \text{as } n \to \infty.$$

Notice that $\omega_{S_n}(\delta_n, \mathbb{C}) = \|S_n\|_{(K+2M)\delta_n^{1/2}}$. If $(K+2M)\delta_n^{1/2} \geqslant n^{-1/3}\sqrt{\ln n}$, then $\delta_n \geqslant n^{-2/3}\ln n/(K+2M)^2$. Substitute $(K+2M)\delta_n^{1/2}$ in the formula of η_n for δ_n, we have $\eta_n^2 = 8c_0(2+\varepsilon)(K+2M)\delta_n^{1/2}\ln\frac{1}{(K+2M)\delta_n^{1/2}}$. Therefore $\eta_n = O(\delta_n^{1/4}(\ln\delta_n^{-1})^{1/2})$. The assertion hence follows from Lemma 6.3. $\diamondsuit$

Proof of Theorem 4.1: Notice that

$$\begin{aligned}\mathbb{E}_{n,\mathbb{C}}(\lambda) - \mathbb{E}_{\mathbb{C}}(\lambda) &= (I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda) + K_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) - K_\lambda(\Gamma_{\mathbb{C}}(\lambda)) \\ &= (I_n - I)(\Gamma_{\mathbb{C}}(\lambda)) + (I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda)) - (I_n - I)(\Gamma_{\mathbb{C}}(\lambda)) \\ &\quad + K_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) - K_\lambda(\Gamma_{\mathbb{C}}(\lambda)).\end{aligned}$$

This also holds for $\mathbb{C}_0$. Therefore we have

$$\sup_{\lambda\in\Lambda} |\Delta_n(\mathbb{C}_0, \mathbb{C}, \lambda) - \Delta(\mathbb{C}_0, \mathbb{C}, \lambda)| \leqslant G_n(\Lambda) + R_n(\Lambda),$$

where

$$G_n(\Lambda) = \sup_{\lambda\in\Lambda} |(I_n - I)(\Gamma_{\mathbb{C}}(\lambda)) - (I_n - I)(\Gamma_{\mathbb{C}_0}(\lambda))|$$

and

$$\begin{aligned}R_n(\Lambda) = &\sup_{\lambda\in\Lambda} |(I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda)) - (I_n - I)(\Gamma_{\mathbb{C}}(\lambda))| \\ &+ \sup_{\lambda\in\Lambda} |(I_n - I)(\Gamma_{n,\mathbb{C}_0}(\lambda)) - (I_n - I)(\Gamma_{\mathbb{C}_0}(\lambda))| \\ &+ \sup_{\lambda\in\Lambda} |K_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) - K_\lambda(\Gamma_{\mathbb{C}}(\lambda))| \\ &+ \sup_{\lambda\in\Lambda} |K_\lambda(\Gamma_{n,\mathbb{C}_0}(\lambda)) - K_\lambda(\Gamma_{\mathbb{C}_0}(\lambda))|.\end{aligned}$$

Lemma 2.2 implies $G_n(\Lambda)$ is of the order $O_P(n^{-1/2})$ as $n \to \infty$. $R_n(\Lambda)$ is of the order $O_P(n^{-1/2})$ because of stochastical equicontinuity of S_n (3.5) and inequality (6.2). $\diamondsuit$

Proof of Theorem 4.2: Notice that under H_0 we have

$$\begin{aligned}\mathbb{E}_{n,\mathbb{C}}(\lambda) - K_\lambda(\Gamma_{\mathbb{C}}(\lambda)) &= (I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda) + K_\lambda(\Gamma_{n,\mathbb{C}}(\lambda)) - K_\lambda(\Gamma_{\mathbb{C}}(\lambda)) \\ &\leqslant (I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda))\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}_{n,\mathbb{C}_0}(\lambda) - K_\lambda(\Gamma_{\mathbb{C}}(\lambda)) &= K_{n,\lambda}(\Gamma_{n,\mathbb{C}_0}(\lambda)) - K_\lambda(\Gamma_{\mathbb{C}}(\lambda)) \\ &\geqslant K_{n,\lambda}(\Gamma_{\mathbb{C}}(\lambda)) - K_\lambda(\Gamma_{\mathbb{C}}(\lambda)) \\ &= (I_n - I)(\Gamma_{\mathbb{C}}(\lambda)).\end{aligned}$$

Therefore we have

$$0 \leqslant \Delta_n(\mathbb{C}_0, \mathbb{C}, \lambda) \leqslant (I_n - I)(\Gamma_{n,\mathbb{C}}(\lambda)) - (I_n - I)(\Gamma_{\mathbb{C}}(\lambda)).$$

The result hence follows from stochastical equicontinuity of S_n (3.5). $\diamondsuit$

7 Acknowledgments

The author would like to thank Professor Wolfgang Polonik for drawing my attention to the excess mass approach and for advice and support during author's dissertation research. Furthermore, I thank my dissertation committee members, Professor Hans-Georg Müller, Jane-Ling Wang, Jiming Jiang, and Fushing Hsieh, for valuable comments and suggestions concerning the subject.

References

[1] Bai, Z. D. and Huang, L. M. On consistency of the best-r-points-average estimator for the maximizer of a nonparametric regression function. The Indian J. of Stat. **61, A**, (1999) Pt. 2, 208-217.

[2] Beran, R. and Dümbgen, L. Modulation of estimators and confidence sets. Ann. Stat. **26,** (1998) 1826-1856.

[3] Cavalier, L. Nonparametric estimation of regression level sets. Statistics, **29,** (1997) 1131-1160.

[4] Cuevas, A. On pattern analysis in the non-convex case. Kybernetes, **19,** (1990) 26-33.

[5] DasGupta, A., Ghosh, J. K. and Zen, M. M. A new general method for constructing confidence sets in arbitrary dimensions: with applications. Ann. Stat. **23,** (1995) 1408-1432.

[6] Davies, P. L. The asymptotic of S-estimators in the linear regression model. Ann. Stat. **18,** (1990) 1651-1675.

[7] Devroye, L. and Wise, G. Detection of abnormal behavior via nonparametric estimation of the support. SIAM J. App. Math. **38,** (1980) 480-488.

[8] Eubank, R. L. Spline smoothing and nonparametric regression. Dekker, New York, 1988.

[9] Frank, I. E. and Friedman, J. H. A statistical view of some chemometrics regression tools. Technometrics, **35,** (1993) 109-135.

[10] Gasser, T., Müller, H. G., Kohler, W., Molinari, L., and Prader, A. Nonparametric regression analysis of growth curves. Ann. Stat. **12,** (1984) 210-229.

[11] Hartigan, J. A. Estimation of a convex density contour in two dimensions. J. Amer. Stat. Asso. **82,** (1987) 267-270.

[12] Hastie, T. J. and Tibshirani, R. J. Generalized additive models. London, Chapman-Hall, 1990.

[13] Jang, W. Nonparametric density estimation and clustering with application to cosmology. Manuscript on http://www-2.cs.cmu.edu/∼learning/talks-2003/2003-02-03.wjang.pdf.

[14] Klemelä, J. Visualization of multivariate density estimates with level set trees. J. Comp. and Grap. Stat. **13,** (2004a) 599-620.

[15] Klemelä, J. Complexity penalized support estimation. J. Multivariate Anal. **88,** (2004b) 274-294.

[16] Korostelev, A. P. and Tsybakov, A. B. Minimax theory of image reconstruction. Springer-Verlag, New York, 1993.

[17] Müller, D. W. and Sawitzki, G. Excess mass estimates and tests for multimodality. J. Amer. Stat. Asso. **86,** (1991) 738-746.

[18] Müller, H. G. Adaptive nonparametric peak estimation. Ann. Stat. **17,** (1989) 1053-1069.

[19] Nolan, D. The excess-mass ellipsoid. J. Multivariate Anal. **39,** (1991) 348-371.

[20] Ossiander, M. A central limit theorem under metric entropy with L_2 bracketing. Ann. Prob. **15,** (1987) 897-919.

[21] Pollard, D. Convergence of stochastic processes. Springer, New York, 1984.

[22] Polonik, W. Measuring mass concentrations and estimating density contour clusters–an excess mass approach. Ann. Stat. **23,** (1995) 855-881.

[23] Polonik, W. and Wang, Z. Estimation of regression contour clusters — An application of the excess mass approach to regression. J. Multivariate Anal. **94,** (2005) 227-249.

[24] Ren, J. Regression M-estimators with non-i.i.d. doubly censored data. Ann. Stat. **31,** (2003) 1186-1219.

[25] Silvapulle, M. J. Asymptotic behavior of robust estimators of regression and scale parameters with fixed carriers. Ann. Stat. **13,** (1985) 1490-1497.

[26] Sonka, M., Hlavac, V., and Boyle, R. Image processing, analysis and machine vision. London, Chapman and Hall, 1993.

[27] Stein, C. Estimation of the mean of multivariate normal distribution. Ann. Stat. **9,** (1981) 1135-1151.

[28] Tran, L. T.; Roussas, G.; Yakowitz, S. and Van, B. T. Fixed design regression for linear time series. Ann. Stat. **24,** (1996) 975-991.

[29] van der Vaart, A. W. Asymptotic statistics. Cambridge University Press, 1998.

[30] Walther, G. Granulometric smoothing. Ann. Stat. **25,** (1997) 2273-2299.