Universität Rostock
Traditio et Innovatio

Minimax Estimation of the Mode of Functional Data

Dissertation

zur
Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Rostock

vorgelegt von
Dennis Müller, geb. am 06. November 1993 in Güstrow

Rostock, 2021

https://doi.org/10.18453/rosdok_id00003428

Gutachter:

Prof. Dr. Alexander Meister, Universität Rostock, Institut für Mathematik

Prof. Dr. Alois Kneip, Universität Bonn, Institute of Finance and Statistics

Prof. Dr. Frédéric Ferraty, Universität Toulouse, Institut de Mathématiques

Jahr der Einreichung: 2021
Jahr der Verteidigung: 2021

Abstract

We investigate the mode of a probability distribution defined on a function space, e.g. the space of integrable functions or a class of smooth functions. Describing such distributions with the help of densities is often complicated, since they can only be defined with respect to some abstract reference measure. Therefore, we give a definition of the mode that does not rely on a density function, but instead uses small ball probabilities. We use entropy methods, e.g. finite covers, to define an estimator of the mode and to deduce its asymptotic behaviour. We show strong consistency and continue to derive the optimal rate of convergence over a class of distributions whose modes are contained in a totally bounded subset of the function space.

Zusammenfassung

Wir untersuchen den Modalwert einer Wahrscheinlichkeitsverteilung, die auf einem Funktionenraum wie etwa dem Raum integrierbarer Funktionen oder einer Klasse glatter Funktionen definiert ist. Die Beschreibung solcher Verteilungen mit Hilfe von Dichten ist oft kompliziert, da diese nur bezüglich eines abstrakten Referenzmaßes angegeben werden können. Daher definieren wir den Modalwert nicht unter Zuhilfenahme einer Dichtefunktion, sondern verwenden stattdessen Small-Ball-Wahrscheinlichkeiten.
Wir benutzen Entropiemethoden wie etwa endliche Überdeckungen für die Definition eines Modalwertschätzers und die Beschreibung seines asymptotischen Verhaltens. Wir zeigen die starke Konsistenz und ermitteln die optimale Konvergenzrate für eine Klasse von Verteilungen, deren Modalwerte in einer totalbeschränkten Teilmenge des Funktionenraums liegen.

Acknowledgements

First and foremost, I want to thank my advisor, Alexander Meister, for posing this challenging problem and for offering his continued support whenever my progress came to a halt.

It has always been pleasant fulfilling teaching duties and exploring mathematical problems among such inspiring colleagues as I am glad I have found at the Institute of Mathematics in Rostock and at our stochastic section in particular.

I owe thanks to Hartmut Milbrodt, who did not only propose I should start working on a Ph.D. but also offered me a post at his chair until his retirement in 2020.

My dear colleague and good friend Lucas Radloff deserves my special gratitude for proofreading this thesis and thereby giving helpful advice. Studying mathematics at Rostock would not have been the same experience without you.

Finally, I want to thank my grandparents Marianne and Udo, my mother Sandra, my sister Maria and my friends with whom I share an abundance of joyful memories and who were there to cheer me up when times were tough.

Contents

Chapter 1

Introduction

It is common that datasets emerging from observations in the fields of finance, chemometrics (especially spectrometrics), biometrics, econometrics or medicine consist of a collection of functions, e.g. smooth curves or discrete plots. Thus, there has been an extensive study of statistical models that are suitable for such a sample of functions and the developed techniques have been applied to a broad spectrum of scientific fields. To illustrate this, we can refer to Laukaitis (2008), who applies methods of functional data analysis to cash flow and transaction data in finance or to Hyndman and Shang (2010), who apply the techniques of functional principal component analysis to age-specific mortality rates in a demographic context. Chapados and Levitin (2008) use cubic splines in combination with methods of functional variance analysis to study the emotional response of music listeners. The list of functional data studies is plentiful, which is why we refer to Ullah and Finch (2013), who provide a thorough overview of the applications of functional methods to datasets from various scientific branches between the years 1995 and 2010. More recently, considerable attention has been paid to integrating functional data into the procedures of machine learning, e. g. neural networks or deep learning (e.g. Perdices et al. (2021) and Rao et. al. (2020)).

Ferraty and Vieu (2006) emphasise the advantages of functional data analysis methods over using traditional approaches. If, for instance, the dataset is a collection of n finely discretised random curves where each of them comes from measuring some phenomenon (e.g. the local temperature) over a discrete time grid $(t_1, \dots, t_m)$, then one usually faces some challenges when attempting to apply conventional statistical methods to the $n \cdot m$ single observations. First and foremost, difficulties can arise from the dependency structure between the single variables. Thus, the necessity to develop methods that take into account the functional structure of the data becomes obvious.

A random function (or functional random variable) takes on values in an infinite-dimensional space. In the usual statistical framework, the distribution of that variable is unknown and some functional of the respective probability measure can only be estimated from a sample of (independent and identically distributed) observations. These functionals of interest typically include measures of location, e.g. the mean, median or mode. Our thesis will be about the latter of these three. In finite-dimensional settings, the mode is a popular notion of centrality in classification tasks because of its usefulness in depicting groups. It is also less sensitive to outliers than the mean. It is our goal in this work to extend the concept of a mode to infinite-dimensional spaces and estimate the mode of a functional random variable (or of its probability distribution, which is an equivalent task). Therefore, we will suggest a mode estimator that converges almost surely to the actual (and unknown) mode as the sample size tends to infinity. Then we will continue to derive an asymptotic upper bound for the (squared) maximum risk and verify that the order of that bound coincides with the order of the asymptotic lower bound of the (squared) minimax risk over some class of probability measures. Thus, the proposed estimator will achieve the optimal rate of convergence.

The first challenge is to define what the mode of a probability distribution on an infinite-dimensional space is. In textbook literature, modes are typically considered as a feature of some Lebesgue-continuous probability measure defined on the (finite-dimensional) Euclidean space. Following these assumptions, the mode is then set equal to a point at which the density function attains some maximum (e.g. Hogg et al. (2012), Mood et al. (1974), Ross (2019), Kendall et al. (1987), Witting and Müller-Funk (1995) and Milbrodt (2010)). Usually, every local maximum point of the (uni- or multivariate) density is considered a mode, which means that the mode is not necessarily unique. Some authors refer to distributions which attain a unique mode as unimodal (e.g. Kendall et al. (1987)), whereas Feller (1971) and Milbrodt (2010) consider a (univariate) distribution unimodal if and only if the density function has exactly one change in its monotonicity behaviour: They impose that the density of a unimodal distribution is first increasing and then decreasing; intervals of constancy are not excluded. Furthermore, the definitions differ in requiring smoothness constraints for the density at and/or around the mode(s). For instance, whereas continuity at the mode is a typical requirement (e.g. Mood et al. (1974)), Meister (2011) and Milbrodt (2010) give definitions omitting any smoothness constraint at the mode.

Since there is no analogue of the Lebesgue-measure on an infinite-dimensional (Banach) space, it is usually difficult to describe probability measures defined on such spaces by the means of density functions. We will solve that issue by defining the mode in a way that only relies on the small ball probability functions of centre points in the space. We will prove that our approach is not only consistent with the common, density-based

definition but much rather an extension of it. The uniqueness of the mode will be a direct implication of the statement of its definition.
The history of mode estimation in a univariate setting goes back to Parzen (1962), who introduces the kernel density estimator (KDE). Therein, conditions are established which guarantee the uniqueness of the (global) maximum point of the KDE, which is then declared the mode estimator. Parzen continues to prove both consistency and asymptotic normality. The properties of that kernel mode estimator (KME) $\hat{\theta}_n$ of the actual mode θ have been thoroughly investigated and extended to a multivariate setting in the last decades. Some significant progress was achieved by the results of Vieu (1996), Abraham et al. (2003), Herrmann and Ziegler (2004) and Shi et al. (2008), who proved that the measurement error $\|\hat{\theta}_n - \theta\|$ (with respect to the Euclidean norm) attains an upper bound of the order $\ln(n)^{c_1} \cdot n^{-c_2}$ almost surely. The constants c_1 and c_2 typically depend on the dimension of the support of the density, the smoothness of the density at or around the mode and the steepness of it in a neighbourhood of the mode. The paper of Herrmann and Ziegler (2004) examines the KME in the absence of any smoothness conditions. Eddy (1980) has shown for a univariate setting that by selecting a specific kernel and appropriately adjusting the bandwidth parameter one achieves $\mathbb{E}\big(\hat{\theta}_n - \theta\big)^2 \in \mathcal{O}(n^{-c})$, where c is a smoothness parameter. Minimax optimality was first given in Tsybakov (1990), who deduces that the optimal rate of convergence of the KME over a Hölder-class of (multivariate) densities has the order n^{-c}, where c is once again a smoothness parameter that also depends on the dimension. Donoho and Liu (1991) and Klemelä (2005) continued the study of the rate optimality of the KME using smoothness restrictions for the density at and/or around the mode. Meister (2011) proposes a different mode estimator in a deconvolution context and derives the minimax rate in absence of any smoothness constraints. The rates established depend on the asymptotic behaviour of the Fourier transform of the error density, where smooth and supersmooth error densities are distinguished.
In fact, analogues of the KME in an infinite-dimensional setting have already been studied (compare Ferraty and Vieu (2006) and Dabo-Niang et al. (2010)), although minimax optimality has yet to be established. The authors assume that the distribution of the functional random variable X, which is supposed to take values in a separable, infinite-dimensional semi-metric space, can be described by the means of a density function f with respect to some abstract measure μ. Consequently, using kernel techniques, they define a functional version of the KME and prove its consistency for the estimation of the mode, which itself is set equal to the unique maximum point of the abstract density. The constraints imposed in these works include the uniform continuity of f as well as the regularity of the small ball probability functions.

The notion of density for a square-integrable, compactly supported random function X has been taken up by Delaigle and Hall (2010), who assume that the covariance operator K of X is positive definite and thus admits the spectral decomposition

$$K(s,t) = \sum_{j=1}^{\infty} \theta_j \psi_j(s) \psi_j(t),$$

where $(\theta_j)_j$ and $(\psi_j)_j$ are the eigenvalues and eigenfunctions of K and s and t are elements of the compact support interval I. The random function itself then admits the representation $X = \sum_{j=1}^{\infty} \theta_j^{1/2} X_j \psi_j$, where the real-valued random variables $(X_j)_j$, which are called the *scores of* X, are assumed to be independent. Let f_j be the density of X_j and let m_j be the mode of f_j (or of X_j equivalently). Then the mode (function) is defined by

$$x_{\text{mode}} = \sum_{j=1}^{\infty} \theta_j^{1/2} m_j \psi_j$$

and the corresponding (principal component) mode estimator is given by

$$\hat{x}_{\text{mode}} = \sum_{j=1}^{T} \hat{\theta}_j^{1/2} \hat{m}_j \widehat{\psi}_j.$$

Therein, $T = T(n)$ is a truncation point and $\hat{\theta}_j$ as well as $\widehat{\psi}_j$ are estimates of the eigenvalues and eigenfunctions that are computed using an empirical version of the covariance operator from a sample $X^{(1)}, \ldots, X^{(n)}$ of i.i.d. observations (compare Ramsay and Silverman (2005)). Delaigle and Hall (2010) proceed to derive an estimator $\widehat{f}_j$ of the score density f_j using kernel methods and set $\hat{m}_j$ equal to the mode of $\widehat{f}_j$. They deduce that the proposed estimator $\widehat{f}_j$ is equivalent to the respective kernel density estimator $\bar{f}_j$ one would use if the values $(\theta_j)_j$ and functions $(\psi_j)_j$ were known explicitly.
To our best knowledge as of today, there exist no results on minimax optimality of the mode estimation problem on an infinite-dimensional space. Our approach will combine the concepts of small ball probability functions and covering numbers. We will assume that the mode is contained within some set $\mathcal{Y}$ for which finite ε-covers exist, which means there is a finite collection of balls with radii smaller than or equal to ε such that $\mathcal{Y}$ is covered by the union of these balls. We will define two different mode estimators and analyse their asymptotic behaviour under additional requirements, e.g. bounds on small ball probabilities. In either scenario, the mode estimator will be set equal to the centre point of some ball for which the amount of data points that are located within it is maximised. We will show that the possible event of ties among different balls is

negligible. Rates of convergence are established for one of these two mode estimators under constraints for the small ball probability functions of the mode itself and of points in a neighbourhood of it. In particular, in order to derive the lower bound of the minimax risk, we will restrict our considerations to the setting in which $\mathcal{Y}$ is equal to a Sobolev class of functions.
The thesis is organised as follows:

- **Chapter 2** provides an overview of the main tools that we use thereafter, namely covering numbers and minimax theory. Asymptotic bounds for the covering numbers of relevant function spaces are given and the main ideas of Le Cam's minimax theory are collected.

- **Chapter 3** contains our definition of the mode of a probability distribution defined on a Polish metric space. That definition will then be applied to several exemplary probability measures on both finite and infinite-dimensional sets. The chapter concludes with a consistency theorem for our first mode estimator.

- **Chapter 4** will be about establishing the upper and lower rate of convergence for our second mode estimator. The lower bound will be obtained over a class of distributions for which the mode is located within some Sobolev ellipsoid.

Chapter 2

Preliminaries

The intention of this paragraph is to provide an overview of the topological and information-theoretic tools that will be put to use in our two main chapters. Before we can proceed we want to elaborate on certain aspects of our notation. Therefore, let $f, g : (0, \infty) \to (0, \infty)$ be two positive functions and let $L \in [0, \infty)$. Then we will write

$$\begin{aligned} \liminf_{x \to 0^+} f(x) = L :\Longleftrightarrow \forall \varepsilon > 0 \, \exists \delta = \delta(\varepsilon) > 0 : x \in (0, \delta) \Longrightarrow f(x) > L - \varepsilon \\ \wedge \, \forall \varepsilon > 0 \, \forall \delta > 0 : \exists x \in (0, \delta) : f(x) < L + \varepsilon. \end{aligned}$$

For brevity, we omit presenting the analogue definitions for the limit superior as well as for the cases where x tends to some other positive real value (from either side) or to infinity. Asymptotic equivalence will be denoted by

$$f(x) \sim g(x) \text{ as } x \to 0^+ :\Longleftrightarrow \lim_{x \to 0^+} \frac{f(x)}{g(x)} = 1.$$

We will use

$$f(x) \asymp g(x) \text{ as } x \to 0^+ :\Longleftrightarrow 0 < \liminf_{x \to 0^+} \frac{f(x)}{g(x)} \leq \limsup_{x \to 0^+} \frac{f(x)}{g(x)} < \infty$$

to abbreviate the statement of f and g behaving similarly as x approaches zero from the right side, which is equivalent to the existence of three constants $0 < c_1 < c_2$ and $s > 0$ such that

$$c_1 g(x) \leq f(x) \leq c_2 g(x), \qquad \forall x \in (0, s).$$

Once again, $x \to \infty$ can be considered analogously.

2.1 Polish metric spaces

This subparagraph serves as a collection of definitions of certain metric spaces which we will work with. Additionally, some topological properties of metric and normed spaces are repeated.

Definition 2.1. Let $\mathcal{F} \neq \emptyset$.

(a) A *metric* d on $\mathcal{F}$ is a function $d\colon \mathcal{F} \times \mathcal{F} \to [0, \infty)$ such that for every $x, y, z \in \mathcal{F}$ the following three properties hold:

  (i) $d(x, y) = 0 \iff x = y$

  (ii) $d(x, y) = d(y, x)$

  (iii) $d(x, z) \leq d(x, y) + d(y, z)$

(b) If d is a metric on $\mathcal{F}$, then the pair $(\mathcal{F}, d)$ is called a *metric space*.

Let $(\mathcal{F}, d)$ be a metric space.

(c) For each $x \in \mathcal{F}$ and $r > 0$ we define by $B_d^o(x, r) := \{y \in \mathcal{F} | d(x, y) < r\}$ the *open ball* and by $B_d(x, r) := \{y \in \mathcal{F} | d(x, y) \leq r\}$ the *closed ball* in $(\mathcal{F}, d)$ with centre x and radius r.

Definition 2.2. Let $\mathcal{F} \neq \emptyset$ be a vector space over the field $\mathbb{R}$.

(a) A *norm* $\|\cdot\|$ on $\mathcal{F}$ is a function $\|\cdot\| : \mathcal{F} \to [0, \infty)$ such that for every $x, y \in \mathcal{F}$ and $\alpha \in \mathbb{R}$ the following three properties hold:

  (i) $\|x\| = 0 \Longrightarrow x = 0$

  (ii) $\|\alpha x\| = |\alpha| \|x\|$

  (iii) $\|x + y\| \leq \|x\| + \|y\|$

(b) If $\|\cdot\|$ is a norm on $\mathcal{F}$, then the pair $(\mathcal{F}, \|\cdot\|)$ is called a *normed space*.

Remark 2.3. It is well known that if $(\mathcal{F}, \|\cdot\|)$ is a normed space, then $(\mathcal{F}, d_{\|\cdot\|})$, where $d_{\|\cdot\|}(x, y) := \|x - y\|$ for every $x, y \in \mathcal{F}$, is a metric space. This metric $d_{\|\cdot\|}$ is called the *induced metric*. Thus, we can consider any normed space a metric space with respect to the metric it induces.

Examples 2.4. (a) Let $k \in \mathbb{N}$, $x = (x_1, \dots, x_k)$, $y = (y_1, \dots, y_k) \in \mathbb{R}^k$ and define

$$d_{Euc}(x, y) := \left(\sum_{n=1}^{k} (x_n - y_n)^2 \right)^{1/2}.$$

Then $(\mathbb{R}^k, d_{Euc})$ is a metric space called the *(k-dimensional) Euclidean metric space.*

(b) Let $k \in \mathbb{N}, D \subset \mathbb{R}^k$ be a compact set and let

$$C_0(D) := \{f : D \longrightarrow \mathbb{R} | f \text{ is continuous}\}$$

denote the space of continuous functions defined on D. Since D is compact, every $f \in C_0(D)$ is bounded. Hence, by setting $\|f\|_\infty := \sup_{x \in D} |f(x)|$ for every $f \in C_0(D)$ we can define a norm on $C_0(D)$, called the *supremum norm*. The induced metric $d_\infty := d_{\|\cdot\|_\infty}$ is called the *supremum metric*. We will typically consider the metric space $(C_0(D), d_\infty)$ where $D = [0, 1]^k$ is the k-dimensional unit hypercube.

(c) Let $(\Omega, \mathcal{A}, \mu)$ denote a measure space, let $p \in (0, \infty)$ and define

$$\mathcal{L}^p(\Omega, \mathcal{A}, \mu) := \mathcal{L}^p_\mu(\Omega) := \left\{ f : \Omega \longrightarrow \mathbb{R}, f \text{ is measurable} \;\middle|\; \int_\Omega |f(x)|^p \, \mu(dx) < \infty \right\}$$

$$\text{and} \qquad \tilde{d}_p(f, g) := \left(\int_\Omega |f(x) - g(x)|^p \, \mu(dx) \right)^{1/p}, \quad \forall f, g \in \mathcal{L}^p_\mu(\Omega).$$

In general, the pair $(\mathcal{L}^p_\mu(\Omega), \tilde{d}_p)$ is not a metric space, e.g. if $\Omega = \mathbb{R}$ and $\mu \equiv \lambda\!\!\!\lambda$ is the Lebesgue measure, then $\tilde{d}_p(f, g) = 0$ does not necessarily imply $f \equiv g$. However, if we define

$$\begin{aligned} [f]_\mu &:= \{g \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu) | \tilde{d}_p(f, g) = 0\}, & \forall f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu), \\ L^p(\Omega, \mathcal{A}, \mu) := L^p_\mu(\Omega) &:= \{[f]_\mu | f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)\} & \text{and} \\ d_p([f]_\mu, [g]_\mu) &:= \tilde{d}_p(f, g), & \forall [f]_\mu, [g]_\mu \in L^p_\mu(\Omega), \end{aligned}$$

then $(L^p_\mu(\Omega), d_p)$ is always a metric space. By setting $\|[f]_\mu\|_p := d_p([f]_\mu, [0]_\mu)$ for every $[f]_\mu \in L^p_\mu(\Omega)$ we define a norm on $L^p_\mu(\Omega)$. From now on we will neglect the bracket notation for classes of μ-equivalent functions. Also, when the reference measure μ is stated explicitly, we will often write $L^p(\Omega)$ to abbreviate $L^p_\mu(\Omega)$.

Definition 2.5. Let $(\mathcal{F}, d)$ be a metric space.

(a) $(\mathcal{F}, d)$ is called *complete* if every Cauchy sequence $(x_n)_n$ in $\mathcal{F}$ has a limit point x that is in $\mathcal{F}$.

(b) $(\mathcal{F}, d)$ is called *separable* if a countable set $\mathcal{D} \subseteq \mathcal{F}$ exists that is dense in $\mathcal{F}$.

(c) $(\mathcal{F}, d)$ is called a *Polish space* if the space is complete and separable.

The notion of a Polish metric space is usually defined in a more general way where the condition is that the space $(\mathcal{F}, d)$ is separable and completely metrizable (or complete with respect to *some* metric d' on $(\mathcal{F}, d)$). However, as we will not be dealing with situations where this difference is relevant, it makes sense for us to refer to a metric space as Polish if it fulfills Definition 2.5 (c).
These definitions can also be applied to a normed space by referring to its induced metric. A complete normed space is also called a *Banach space*. Thus, a separable Banach space is always Polish. It happens that all the spaces considered in the Examples 2.4 are separable Banach spaces.
The Borel σ-algebra of a separable metric space can be derived from its system of open balls:

Definition 2.6. Let $(\mathcal{F}, d)$ be a separable metric space and let $\mathcal{O}(d)$ denote the set containing all open balls in $(\mathcal{F}, d)$. Then we set

$$\mathbb{B}(\mathcal{F}) := \bigcap_{\substack{\mathcal{A}\ \sigma\text{-algebra} \\ \mathcal{O}(d) \subseteq \mathcal{A}}} \mathcal{A}$$

and call $\mathbb{B}(\mathcal{F})$ the *Borel σ-algebra* associated to $\mathcal{F}$.

2.2 Covering numbers

Given some radius $\varepsilon > 0$, the covering number of a subset of a metric space $\mathcal{F}$ is equal to the minimum (finite oder infinite) amount of closed balls with radius ε such that the set is contained within the union of these balls. There are two reasons why we discuss them at this stage, the first one being that we will later stipulate entropy bounds for the metric spaces considered, which are directly related to bounds for their covering numbers (see Definitions 2.7(b) and 2.14). The second reason is that they provide a useful tool of approximation and discretisation that we will often exploit in the proofs of our main results.

This subsection will first include the definition and a collection of some fundamental properties of covering numbers. We refer to van der Vaart and Wellner (1996) for a comprehensive introduction into covering and entropy methods. In the second half we will give an overview of upper and lower bounds for the metric entropy functions of some exemplary (function) spaces.

Definition 2.7. Let $(\mathcal{F}, d)$ be a metric space, $A \subseteq \mathcal{F}$ be non-empty and $\varepsilon > 0$.

(a) A set $\mathcal{X} \subseteq \mathcal{F}$ is called an *ε-cover* of A, if $A \subseteq \bigcup_{x \in \mathcal{X}} B_d(x, \varepsilon)$.

(b) Let $\mathcal{C}$ denote the set containing all ε-covers of A, then the *ε-covering number* of A is defined by

$$N(A, d, \varepsilon) := N(A, \varepsilon) := \min_{\mathcal{X} \in \mathcal{C}} |\mathcal{X}| \in \mathbb{N} \cup \{+\infty\},$$

where $|\mathcal{X}|$ denotes the cardinality of $\mathcal{X}$.

(c) If $N(A, d, r) < \infty$ for every $r > 0$, then the set A is called *totally bounded*.

Note that ε-covers always exist for any non-empty set (e.g. the set itself is a trivial cover). We will write $N(\cdot, \cdot)$ instead of $N(\cdot, d, \cdot)$ whenever it is unambiguous what metric is considered. The quantity defined in (b) is sometimes referred to as the *extrinsic covering number*, because we merely impose that the elements of the covers of A lie in $\mathcal{F}$, which is, of course, less strict than requiring that they lie in A. The extrinsic covering number is always smaller than or equal to the respective *intrinsic covering number*. From now on, we will call an ε-cover of any subset of $\mathcal{F}$ a *minimum ε-cover* if its cardinality equals the covering number of the set.
The total boundedness of a metric space is a strong property that is linked to other topological concepts:

Proposition 2.8. *Let $(\mathcal{F}, d)$ be a metric space.*

(a) The following statements are equivalent:

(1) $(\mathcal{F}, d)$ is totally bounded and complete.

(2) $(\mathcal{F}, d)$ is compact.

(b) If $(\mathcal{F}, d)$ is totally bounded, then the space is separable.

Proof: This can be found in Dieudonné (1972), Theorems 3.16.1 and 3.16.2.

■

Proposition 2.8 implies that a totally bounded Polish metric space is necessarily compact. On the other hand, every complete and totally bounded metric space is Polish.

The concept of packing numbers is closely related to covering numbers and will now be formalised.

Definition 2.9. Let $(\mathcal{F}, d)$ be a metric space, $A \subseteq \mathcal{F}$ be non-empty and $\varepsilon > 0$.

(a) A set $\mathcal{X} \subseteq A$ is called an *ε-packing* of A, if $B_d(x, \varepsilon/2) \cap B_d(x', \varepsilon/2) = \emptyset$ for every $x, x' \in \mathcal{X}$ such that $x \neq x'$.

(b) Let $\mathcal{P}$ denote the set containing all ε-packings of A, then the *ε-packing number* of A is defined by

$$D(A, d, \varepsilon) := D(A, \varepsilon) := \max_{\mathcal{X} \in \mathcal{P}} |\mathcal{X}| \in \mathbb{N} \cup \{+\infty\}.$$

Analogously, a *maximum ε-packing* of some set denotes an ε-packing whose cardinality is equal to the packing number of the set. Packing and covering numbers are equivalent in the following sense:

Lemma 2.10. *Let $(\mathcal{F}, d)$ be a metric space, $A \subseteq \mathcal{F}$ be non-empty and $\varepsilon > 0$. Then we have*

$$D(A, d, 2\varepsilon) \leq N(A, d, \varepsilon) \leq D(A, d, \varepsilon).$$

Proof: This is Lemma 5.5 in Wainwright (2019).

■

The following Lemmata 2.11 and 2.12 summarise some basic properties of covering numbers. As all of these facts are well known, only sketches of their proofs will be given.

Lemma 2.11. *Let $(\mathcal{F}, d)$ be a metric space.*

(a) For every set $A \subseteq \mathcal{F}$, if $0 < \varepsilon' \leq \varepsilon$, then we have $N(A, \varepsilon') \geq N(A, \varepsilon)$.

(b) For every radius $\varepsilon > 0$, if $A, A' \subseteq \mathcal{F}$ such that $A' \subseteq A$, then we have $N(A', \varepsilon) \leq N(A, \varepsilon)$.

Proof: For (a) consider that any ε'-cover of A is an ε-cover of A and for (b) we observe that any ε-cover of A is an ε-cover of A'. ■

If $(\mathcal{F}, \|\cdot\|)$ is a normed space, then both covering and packing numbers can be applied with respect to the induced metric and we can write $N(\cdot, \|\cdot\|, \cdot) := N(\cdot, d_{\|\cdot\|}, \cdot)$ and $D(\cdot, \|\cdot\|, \cdot) := D(\cdot, d_{\|\cdot\|}, \cdot)$. If $A \subseteq \mathcal{F}, x \in \mathcal{F}$ and $\alpha \in \mathbb{R}\backslash\{0\}$, then we can set $A + x := \{a + x | a \in A\}$ and $\alpha A := \{\alpha a | a \in A\}$.

Lemma 2.12. *Let $(\mathcal{F}, \|\cdot\|)$ be a normed space and $A \subseteq \mathcal{F}$ be non-empty. Let $x \in \mathcal{F}, \alpha \in \mathbb{R}\backslash\{0\}$ and $\varepsilon > 0$. Then the following two relations hold:*

(a) $N(A + x, \|\cdot\|, \varepsilon) = N(A, \varepsilon)$

(b) $N(\alpha A, \|\cdot\|, |\alpha|\varepsilon) = N(A, \varepsilon)$

Proof: Let $U(\varepsilon)$ denote a minimum ε-cover of A.

(a) For every $x_1, x_2 \in \mathcal{F}$ and $\varepsilon > 0$ we can verify that $x_1 + B_{\|\cdot\|}(x_2, \varepsilon) = B_{\|\cdot\|}(x_1 + x_2, \varepsilon)$. Thus, $U(\varepsilon) + x$ is a minimum ε-cover of $A + x$.

(b) For every $y \in \mathcal{F}, \lambda \in \mathbb{R}\backslash\{0\}$ and $\varepsilon > 0$ we can verify that $\lambda B_{\|\cdot\|}(y, \varepsilon) = B_{\|\cdot\|}(\lambda y, |\lambda|\varepsilon)$. Thus, $\alpha U(\varepsilon)$ is a minimum $|\alpha|\varepsilon$-cover of αA.

■

For the remainder of this subsection we will cite results concerning the (asymptotic) behaviour of the covering numbers of certain exemplary spaces. We will start with finite-dimensional metric spaces and proceed towards (infinite-dimensional) function classes.

Lemma 2.13. *Let $(\mathcal{F}, d) = (\mathbb{R}^k, d_{Euc})$ be the k-dimensional Euclidean metric space.*

(a) For every $\varepsilon > 0$ we have

$$\left(\frac{1}{\varepsilon}\right)^k \leq N(B_{d_{Euc}}(0, 1), d_{Euc}, \varepsilon) \leq \left(1 + \frac{2}{\varepsilon}\right)^k.$$

(b) Let $C \subset \mathbb{R}^k$ such that $\delta := \operatorname{diam} C < \infty$. Then we have

$$N(C, d_{Euc}, \varepsilon) \leq \left(1 + \frac{\delta}{\varepsilon}\right)^k$$

for every $\varepsilon > 0$.

(c) If the set C from part (b) has an inner point, then we have $N(C, d_{Euc}, \varepsilon) \asymp \varepsilon^{-k}$ as $\varepsilon \to 0^+$.

Proof:

(a) See Wainwright (2019), Lemma 5.7.

(b) There exists some $x \in \mathbb{R}^k$ such that $C \subseteq B_{d_{Euc}}(x, \delta/2)$. Hence, we can apply part (b) of Lemma 2.11, Lemma 2.12 and part (a) to deduce

$$\begin{aligned} N(C, d_{Euc}, \varepsilon) &\leq N(B_{d_{Euc}}(x, \delta/2), d_{Euc}, \varepsilon) = N(B_{d_{Euc}}(0, \delta/2), d_{Euc}, \varepsilon) \\ &= N\Big(B_{d_{Euc}}(0,1), d_{Euc}, \frac{2\varepsilon}{\delta}\Big) \leq \Big(1 + \frac{\delta}{\varepsilon}\Big)^k. \end{aligned}$$

(c) It follows from part (b) that if $\varepsilon \in (0, \delta)$, then we have $N(C, d_{Euc}, \varepsilon) \leq (2\delta)^k \varepsilon^{-k}$. The lower bound can be derived from the fact that some k-dimensional Euclidean ball is contained within C, part (a) and our Lemmata 2.11 and 2.12.

■

If $(\mathcal{F}, d)$ is an infinite-dimensional metric space and $\mathcal{Y} \subseteq \mathcal{F}$ is a totally bounded, infinite-dimensional set, then it is typically cumbersome to deduce the exact asymptotic behaviour of $N(\mathcal{Y}, d, \varepsilon)$ as $\varepsilon \to 0^+$. Instead, the logarithm is considered.

Definition 2.14. Let $(\mathcal{F}, d)$ be a metric space and let $\mathcal{Y} \subseteq \mathcal{F}$ be a totally bounded subset. The function

$$m_{(\mathcal{Y},d)} : (0, \infty) \longrightarrow [0, \infty), \quad \varepsilon \longmapsto \ln N(\mathcal{Y}, d, \varepsilon)$$

is called the *metric entropy function* of $\mathcal{Y}$.

We will later stipulate that the mode lies in some totally bounded set $\mathcal{Y} \subseteq \mathcal{F}$ that fulfills $m_{(\mathcal{Y},d)}(\varepsilon) \asymp \varepsilon^{-q}$, where $q > 0$ is a constant (that would usually depend on certain parameters that appear in the definition of the space). This is an assumption that can be justified for a broad spectrum of relevant function classes. A collection of some relevant exemplary spaces for which it is fulfilled is given below.

Examples 2.15. (a) Let $k \in \mathbb{N}$, $n \in \mathbb{N}_0, \gamma \in (0, 1], L > 0$ and let $\mathcal{F}_{k,n,\gamma,L}$ denote the set of all functions

- which are defined on $[0, 1]^k$,
- whose partial derivatives up to the order n exist and are uniformly bounded by L and
- whose n-th partial derivatives are γ-Hölder-continuous with constant L.

Note that $f^{(0)} \equiv f$. Then we have

$$m_{(\mathcal{F}_{k,n,\gamma,L},d_\infty)}(\varepsilon) \asymp \varepsilon^{-\frac{k}{n+\gamma}}, \qquad \varepsilon \to 0^+.$$

The upper bound is proven by van der Vaart and Wellner (1996) in their Theorem 2.7.1, whereas the lower bound can be deduced from Theorem XIII in Kolmogorov and Tikhomirov (1961).
Note that $\mathcal{F}_{k,0,1,L}$ is a class of k-dimensional, Lipschitz-continuous functions defined on the unit hypercube.

(b) The covering numbers with respect to the L^2-metric of a class of square-integrable functions whose Fourier coefficients (with respect to the trigonometric L^2-basis) lie in a Sobolev ellipsoid (see Definition 4.8) also admit an exponential order. A detailed proof of the upper bound is given in Proposition 4.10.

(c) Let $k \in \mathbb{N}$ and define by

$$\mathcal{D}_{k,inc} := \{f : [0,1]^k \to [0,1] | f \textit{ is increasing in each variable}\}$$

a space of k-dimensional functions with explicit monotonicity properties. Then, if $p \geq 1$ such that $(k,p) \neq (2,2)$, we have

$$m_{(\mathcal{D}_{k,\mathrm{inc}},d_p)}(\varepsilon) \asymp \varepsilon^{-\alpha}, \qquad \varepsilon \to 0^+,$$

where $\alpha = \max(k,(k-1)p)$ (see Theorem 4.1 in Gao and Wellner (2007)). For the critical case $(k,p) = (2,2)$ they give an upper bound that has an additional logarithmic factor. The lower bound, however, remains the same.

We want to contrast the exponential dependency of the covering numbers of the preceding examples on their parameters (e.g. the dimension) as opposed to the polynomial dependency obtained in Lemma 2.13(c) for a finite-dimensional space, which Wainwright (2019) refers to as a 'dramatic manifestation of the curse of dimensionality'. It is an unfortunate consequence that optimal convergence rates achieved for nonparametric estimation problems on infinite-dimensional (function) spaces are typically logarithmic.

2.3 Minimax theory

Our first goal in this paragraph is to give a precise definition of the term 'optimal rate of convergence' with reference to an estimation problem. Later, we will present some information-theoretic tools that will help us to ascertain lower bounds. Therefore, let $(\mathcal{F}, d)$ be a Polish metric space, $n \in \mathbb{N}$ and let $X_1, \ldots, X_n$ be independent and identically distributed random variables on some probability space $(\Omega, \mathcal{A}, P)$ such that $X_j : (\Omega, \mathcal{A}) \to (\mathcal{F}, \mathbb{B}(\mathcal{F}))$ for every $j \in \{1, \ldots, n\}$. Let

$$\mathcal{W}_{(\mathcal{F},d)} := \Big\{ P \,\Big|\, P \text{ is a probability measure on } (\mathcal{F}, \mathbb{B}(\mathcal{F})) \Big\}$$

and assume that $X_1 \sim P_X \in \mathcal{W}_{(\mathcal{F},d)}$. Typically, P_X is unknown. Let $\mathcal{P} \subseteq \mathcal{W}_{(\mathcal{F},d)}$ be a class of probability distributions such that $|\mathcal{P}| \geq 2$ and assume $P_X \in \mathcal{P}$. Our objective is to define and discuss the properties of an estimator of $\theta(P_X)$, where $\theta : \mathcal{P} \to \mathcal{F}$ is some functional of interest. For our purposes, $\mathcal{P}$ will typically be a class of probability measures with a unique mode and θ will be the mapping that assigns the mode to some $Q \in \mathcal{P}$. Let $\widehat{\theta}_n := \widehat{\theta}_n(X_1, \ldots, X_n)$ be an estimator of $\theta(P_X)$ and let $(\widehat{\theta}_n)_{n \in \mathbb{N}}$ be the sequence of estimators depending on the sample size $n \in \mathbb{N}$. This can be done by formalising an infinite statistical (product) model and assuming there are infinitely many observations. Consistency is a typical criterion used to assess the quality of a sequence of estimators.

Definition 2.16. The following equations are given assuming that the distribution of a single variable is $Q \in \mathcal{P}$.

(a) $(\widehat{\theta}_n)_{n \in \mathbb{N}}$ is *strongly consistent* for $\theta :\Longleftrightarrow$

$$\forall Q \in \mathcal{P} \colon P\Big(\lim_{n \to \infty} d(\widehat{\theta}_n, \theta(Q)) = 0 \Big) = 1$$

Let $p \geq 1$.

(b) $(\widehat{\theta}_n)_{n \in \mathbb{N}}$ is *consistent in the p-th mean* for $\theta :\Longleftrightarrow$

$$\forall Q \in \mathcal{P} : \lim_{n \to \infty} \mathbb{E}_Q \, d^p(\widehat{\theta}_n, \theta(Q)) = 0$$

(c) $(\widehat{\theta}_n)_{n \in \mathbb{N}}$ is *uniformly consistent in the p-th mean* for $\theta :\Longleftrightarrow$

$$\lim_{n \to \infty} \sup_{Q \in \mathcal{P}} \mathbb{E}_Q \, d^p(\widehat{\theta}_n, \theta(Q)) = 0$$

The notion of uniform consistency in the p-th mean motivates the following two related risk concepts.

Definition 2.17. Let $p \geq 1$.

(a) The *maximum risk* of the estimator $\widehat{\theta}_n$ is defined by

$$r_n(\widehat{\theta}_n, \mathcal{P}) := \sup_{Q \in \mathcal{P}} \mathbb{E}_Q \, d^p(\widehat{\theta}_n, \theta(Q)).$$

(b) The *minimax risk* is defined by $\mathcal{R}_n(\mathcal{P}) := \inf_{\widehat{\theta}_n} r_n(\widehat{\theta}_n, \mathcal{P})$, where the infimum is taken over all estimators $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$.

The dependency on the weighting factor p is usually neglected in the notation. It is worth mentioning that both the maximum and the minimax risk depend on the class of distributions $\mathcal{P}$. If there exists a uniformly consistent sequence of estimators for which its maximum risk admits the same order as the minimax risk over $\mathcal{P}$ as the sample size increases, then that order is considered the optimal rate of convergence. It is unique up to a constant $c > 0$ and will typically depend on the parameters used in the definition of $\mathcal{P}$.

Definition 2.18. A positive, real-valued sequence $(\rho_n)_{n \in \mathbb{N}}$ converging to zero is called the *optimal rate of convergence* for the problem of estimating θ over $\mathcal{P}$ if the following two conditions are fulfilled:

(1) There is a sequence of estimators $(\widehat{\theta}_n)_{n \in \mathbb{N}}$ such that $\limsup_{n \to \infty} \dfrac{r_n(\widehat{\theta}_n, \mathcal{P})}{\rho_n} < \infty$.

(2) $\liminf_{n \to \infty} \dfrac{\mathcal{R}_n(\mathcal{P})}{\rho_n} > 0$

By presenting the main ideas of the methods developed by Le Cam we want to describe a general procedure to derive lower bounds for the minimax risk $\mathcal{R}_n(\mathcal{P})$. The boundaries involved are often expressed depending on some information-theoretic distance of two measures $P_0, P_1 \in \mathcal{P}$, e.g. the *Hellinger distance*, the *Kullback-Leibler distance* or the *total variation distance*. There exist relations between these distances, but for our purposes it will suffice to give lower bounds with respect to the total variation distance.

Definition 2.19. Let $P_0, P_1 \in \mathcal{W}_{(\mathcal{F},d)}$. Then the *total variation distance* of P_0 and P_1 is defined by $TV(P_0, P_1) := \sup_{A \in \mathbb{B}(\mathcal{F})} |P_0(A) - P_1(A)|$.

Remark 2.20. $(\mathcal{W}_{(\mathcal{F},d)}, TV)$ is a metric space. By Scheffé's theorem (see Lemma 2.1 in Tsybakov (2008)), if μ is some non-negative, σ-finite measure on $(\mathcal{F}, \mathbb{B}(\mathcal{F}))$ such that $P_i \ll \mu$ and $f_i \in dP_i/d\mu, i = 0, 1$, then the total variation distance admits the representations

$$TV(P_0, P_1) = \frac{1}{2} \int_{\mathcal{F}} |f_0(x) - f_1(x)| \, \mu(dx) = 1 - \int_{\mathcal{F}} \min(f_0(x), f_1(x)) \, \mu(dx).$$

Such a dominating measure always exists, e.g. we could choose $\mu = P_0 + P_1$.

An integral of the type $\int \min(f_0, f_1)$, where f_0 and f_1 are densities, will appear in the following inequality for the minimax risk. It is a consequence of the preceding remarks that such an integral can be expressed in terms of the total variation distance. To ease our notation we will write $\nu^{\otimes n} := \bigotimes_{i=1}^n \nu$ for the n-times product measure.

Proposition 2.21. *Let $p \geq 1, P_0, P_1 \in \mathcal{P}$ and let μ be a non-negative, σ-finite measure on $(\mathcal{F}, \mathbb{B}(\mathcal{F}))$ such that $P_i \ll \mu$ and $f_i \in dP_i/d\mu, i = 0, 1$. Let $f_{i,n} \in dP_i^{\otimes n}/d\mu^{\otimes n}, i = 0, 1$, be versions of the respective product densities. Then we have*

$$\mathcal{R}_n(\mathcal{P}) \geq \frac{d^p(\theta(P_0), \theta(P_1))}{2^{p+1}} \int_{\mathcal{F}^{\times n}} \min(f_{0,n}(x), f_{1,n}(x)) \, \mu^{\otimes n}(dx).$$

Proof: This follows from Theorem 2.2 in Tsybakov (2008) and the remarks given therein throughout Section 2.2.

■

It now seems desirable to be able to express the TV-distance of product measures in terms of the TV-distance of the (base) measures. Therefore, the following inequality will prove useful:

Lemma 2.22. *Let P_0, $P_1 \in \mathcal{P}$. Then we have*

$$TV(P_0^{\otimes n}, P_1^{\otimes n}) \leq n \, TV(P_0, P_1).$$

Proof: Set $\mu = P_0 + P_1$, let $f_i \in dP_i/d\mu$ and let $f_{i,n} \in dP_i^{\otimes n}/d\mu^{\otimes n}$ denote the n-times product density, $i = 0, 1$. Applying the triangle inequality and the Fubini theorem yields

$$
\begin{aligned}
2\,TV(P_0^{\otimes n}, P_1^{\otimes n}) &= \int\limits_{\mathcal{F}^{\times n}} |f_{0,n}(x) - f_{1,n}(x)|\,\mu^{\otimes n}(dx) = \int\limits_{\mathcal{F}^{\times n}} \Big|\prod_{k=1}^{n} f_0(x_k) - \prod_{k=1}^{n} f_1(x_k)\Big|\,\mu^{\otimes n}(dx) \\
&= \int\limits_{\mathcal{F}^{\times n}} \Big|\sum_{j=1}^{n}\Big(\prod_{k=1}^{j-1} f_1(x_k)\prod_{k=j}^{n} f_0(x_k) - \prod_{k=1}^{j} f_1(x_k)\prod_{k=j+1}^{n} f_0(x_k)\Big)\Big|\,\mu^{\otimes n}(dx) \\
&\leq \sum_{j=1}^{n}\int\limits_{\mathcal{F}^{\times n}} \Big(\prod_{k=1}^{j-1} f_1(x_k)\cdot|f_0(x_j) - f_1(x_j)|\cdot\prod_{k=j+1}^{n} f_0(x_k)\Big)\,\mu^{\otimes n}(dx) \\
&= \sum_{j=1}^{n}\int\limits_{\mathcal{F}} |f_0(x_j) - f_1(x_j)|\,\mu(dx_j) = 2n\,TV(P_0, P_1).
\end{aligned}
$$

Note that in the second line the difference of the two products in the previous expression is written as a telescoping sum.

■

Finally, we can combine the preceding results to express the lower bound for $\mathcal{R}_n(\mathcal{P})$ from Proposition 2.21 in terms of the total variation distance of two arbitrary probability measures $P_0, P_1 \in \mathcal{P}$.

Corollary 2.23. *(a) Let $p \geq 1$ and $P_0, P_1 \in \mathcal{P}$. Then we have*

$$
\mathcal{R}_n(\mathcal{P}) \geq \frac{d^p(\theta(P_0), \theta(P_1))}{2^{p+1}}\big(1 - n\,TV(P_0, P_1)\big).
$$

(b) If $(P_n^{(0)})_{n\in\mathbb{N}}$ and $(P_n^{(1)})_{n\in\mathbb{N}}$ are sequences of probability distributions in $\mathcal{P}$ and if both

$$
\limsup_{n\to\infty} n\,TV\big(P_n^{(0)}, P_n^{(1)}\big) < 1
$$

and $d^p\big(\theta(P_n^{(0)}), \theta(P_n^{(1)})\big) > 0, \forall n \in \mathbb{N}$, hold, then we have

$$
\liminf_{n\to\infty} \frac{\mathcal{R}_n(\mathcal{P})}{d^p\big(\theta\big(P_n^{(0)}\big), \theta\big(P_n^{(1)}\big)\big)} > 0.
$$

Proof:

(a) The integral in Proposition 2.21 can be expressed in terms of $TV(P_0^{\otimes n}, P_1^{\otimes n})$. The inequality follows after applying Lemma 2.22.

(b) Using part (a) yields

$$\mathcal{R}_n(\mathcal{P}) \geq \frac{d^p(\theta(P_n^{(0)}), \theta(P_n^{(1)}))}{2^{p+1}} \big(1 - n\,TV(P_n^{(0)}, P_n^{(1)})\big)$$

for every $n \in \mathbb{N}$.

■

Part (b) of the previous corollary unveils the strategy we are going to follow to deduce a lower bound for the minimax risk: We will define a class $\mathcal{P}$ of distributions with a mode and identify two sequences $(P_n^{(0)})_{n\in\mathbb{N}}$ and $(P_n^{(1)})_{n\in\mathbb{N}}$ in $\mathcal{P}$ such that $TV(P_n^{(0)}, P_n^{(1)})$ has the upper bound c/n for some $c \in (0, 1)$ and n sufficiently large, while the speed of convergence at which the (weighted) distances of their modes $d^p(\theta(P_n^{(0)}), \theta(P_n^{(1)}))$ converge is as slow as can be achieved. If we, in addition, succeed at finding a sequence of mode estimators such that the maximum risk admits the same order of convergence (up to a constant factor), then we have found the optimal rate.
The first step, however, is to define what the mode of a distribution on a Polish metric space is, which is what the following chapter is about.

Chapter 3

Distributions with a mode

In this chapter, the notion of the mode of a probability distribution defined on a Polish metric space is developed and analysed in-depth.

- **Section 3.1** contains a general definition of the mode.
- **Section 3.2** provides an analysis of conditions under which distributions on exemplary finite- or infinite-dimensional spaces have a unique mode. Therein, we will prove that if a probability measure is defined on the k-dimensional Euclidean space and is continuous with respect to the k-dimensional Lebesgue measure, then it has a unique mode under certain conditions on the corresponding density. We will then extend the study of examples to Gaussian measures defined on a function space (e.g. the space of continuous functions or the space of square-integrable functions).
- **Section 3.3** is about establishing a link between the small ball probability functions at points in a neighbourhood of the mode and the metric entropy function of a totally bounded subspace that contains the mode and attains a positive probability.
- **Section 3.4** contains the definition of an estimator of the mode. The section is concluded with a theorem about its consistency.

Throughout this chapter, if not stated otherwise, we denote by $(\mathcal{F}, d)$ a Polish metric space.

3.1 Definition of the mode

In the following, let $Q \in \mathcal{W}_{(\mathcal{F},d)}$.

Definition 3.1. (a) For every $x \in \mathcal{F}$ we define by

$$\varphi_x^Q : (0,\infty) \longrightarrow [0,1], \quad \varepsilon \longmapsto Q(B_d(x,\varepsilon))$$

the *small ball probability function* for balls around the centre x.

(b) The set $\text{supp}(Q) := \{x \in \mathcal{F} | \forall \varepsilon > 0 : \varphi_x^Q(\varepsilon) > 0\}$ is called the *support* of Q.

Lemma 3.2. *(a) The support of* Q *is a closed set.*

(b) We have $Q(\text{supp}(Q)) = 1$.

Proof:

(a) Recall that $\mathcal{O}(d)$ is the set containing all the open balls in $\mathcal{F}$ and define

$$U := \bigcup_{\substack{B \in \mathcal{O}(d) \\ Q(B)=0}} B,$$

(see Heinonen et al. (2015), p. 64). Since U is an open set, it suffices to show that $\text{supp}(Q) = \mathcal{F} \setminus U$. Let $x \in \text{supp}(Q)$, which implies $x \notin U$, because if $x \in U$, then it would follow from the openness of U that there is some radius $\varepsilon' > 0$ such that $\varphi_x^Q(\varepsilon') = 0$. If $x \in \mathcal{F} \setminus U$, then we clearly have $\varphi_x^Q(\varepsilon) > 0$ for every $\varepsilon > 0$.

(b) Let us first repeat the following known facts about separable metric spaces (e.g. see Theorem 16.11 in Willard (1970)). Assume that $\{f_n | n \in \mathbb{N}\} \subset \mathcal{F}$ is dense in $\mathcal{F}$, then

$$\mathcal{B} = \left\{B_d^o(f_n, 1/m) \middle| (n,m) \in \mathbb{N}^2\right\} \subset \mathcal{O}(d)$$

is a basis of $\mathcal{F}$, by which we mean that for every open set $O \subset \mathcal{F}$ there is a collection of at most countable open balls $(B_i)_{i \in I}$ taken from $\mathcal{B}$ such that $O = \bigcup_{i \in I} B_i$. In fact, if $x \in O$, then, for some $m \in \mathbb{N}$, we have $B_d^o(x, 1/m) \subset O$. Now there exists $j \in \mathbb{N}$ such that $d(x, f_j) < 1/2m$, which implies both $B_d^o(f_j, 1/2m) \subset O$ and $x \in B_d^o(f_j, 1/2m) \in \mathcal{B}$. This can be done for any $x \in O$. Note that $\mathcal{B}$ is countable. Now let U be defined as in part (a). Then there exists a sequence $(B_n)_{n \in \mathbb{N}}$ in $\mathcal{B}$ such that $U = \bigcup_{n \in \mathbb{N}} B_n$ and $Q(B_n) = 0$ for every $n \in \mathbb{N}$. This implies $Q(U) = 0$. Thus, $Q(\text{supp}(Q)) = Q(\mathcal{F}) - Q(U) = 1$. ■

The small ball probability functions are our main tool to define the mode as, in general, we do not assume that Q possesses some (abstract) density. The behaviour of the small ball probability functions for small radii will also play a crucial role in the analysis of the asymptotic properties of our mode estimators. Typically, additional constraints that we will later impose involve restrictions on the covering numbers of some subspace that contains the mode and for the values of the small ball probability functions of the mode and of points in its neighbourhood. Let us now set

$$M^Q(\delta) := \left\{ x \in \operatorname{supp}(Q) \,\middle|\, \liminf_{\varepsilon \to 0^+} \inf_{d(x,y) \geq \delta} \frac{\varphi_x^Q(\varepsilon)}{\varphi_y^Q(\varepsilon)} > 1 \right\}, \qquad \forall \delta > 0, \tag{3.1}$$

where we use the conventions $\inf \emptyset := \infty, \frac{1}{0} := \infty$ and $\liminf \infty := \infty$. For every $\delta > 0$, $x \in M^Q(\delta)$ implies that there is some radius $\varepsilon = \varepsilon(x) > 0$ such that for every $\varepsilon \in (0, \varepsilon(x))$ the value of the small ball probability function $\varphi_x^Q(\varepsilon)$ is greater than $\varphi_y^Q(\varepsilon)$ for every $y \in \mathcal{F}$ that has a distance towards x that is at least δ. We will now show that if the sets in (3.1) are non-empty for any $\delta > 0$, then there exists exactly one element in the intersection of the topological closures of the sets $M^Q(\delta)$. We will declare that element the mode of Q.

Proposition 3.3. *If $M^Q(\delta) \neq \emptyset$ for every $\delta > 0$, then there exists some $x \in \operatorname{supp}(Q)$ such that $\{x\} = \bigcap_{\delta > 0} \overline{M^Q(\delta)}$.*

Proof: We use arguments which are similar to the ones used to prove *Cantor's intersection theorem* (e.g. see section 7.8 in Lewin (2003)). At first, let us note that if $x, y \in M^Q(\delta)$, then $d(x,y) < \delta$. Thus, for every $\delta > 0$ we can find $x(\delta) \in \mathcal{F}$ such that $M^Q(\delta) \subseteq B_d(x(\delta), \frac{1}{2}\delta)$, which implies both

$$\operatorname{diam} M^Q(\delta) \overset{\delta \to 0^+}{\longrightarrow} 0 \qquad \text{and} \qquad \operatorname{diam} \bigcap_{\delta > 0} \overline{M^Q(\delta)} = 0.$$

Hence, $\bigcap_{\delta > 0} \overline{M^Q(\delta)}$ is either empty or consists of a single point. Let us now fix some $\delta > 0$. We consider that, for every $\delta' \in (0, \delta]$, we have $M^Q(\delta') \subseteq M^Q(\delta)$, which implies $\overline{M^Q(\delta')} \subseteq \overline{M^Q(\delta)}$. Let $(\delta_n)_{n \in \mathbb{N}}$ be a decreasing sequence of positive real numbers that converges to zero and let $(x_n)_{n \in \mathbb{N}}$ denote a sequence in $\mathcal{F}$ such that, for every $n \in \mathbb{N}$, we have $x_n \in M^Q(\delta_n)$. Since the sets $M^Q(\delta)$ are nested and the diameter of $M^Q(\delta)$ is bounded by δ, $(x_n)_n$ is a Cauchy sequence in $\mathcal{F}$. Since $\mathcal{F}$ is complete, the sequence converges and we can assume that $x \in \mathcal{F}$ exists such that $d(x_n, x) \overset{n \to \infty}{\longrightarrow} 0$. Since $x_n \in \overline{M^Q(\delta_m)}$ for every $m, n \in \mathbb{N}$ with $m \leq n$, we can deduce from the closedness of the sets $\overline{M^Q(\cdot)}$ that $x \in \overline{M^Q(\delta_n)}$ for every $n \in \mathbb{N}$. Now it follows from the monotonicity property of the sets $\overline{M^Q(\delta)}$ (w.r.t. to inclusion) that $x \in \overline{M^Q(\delta)}$ for all $\delta > 0$. This, however, is equivalent to $\{x\} = \bigcap_{\delta>0} \overline{M^Q(\delta)}$. Since $M^Q(\delta) \subseteq \text{supp}(Q), \forall \delta > 0$, and $\text{supp}(Q)$ is closed (see Lemma 3.2(a)), it follows that $x \in \text{supp}(Q)$.

■

We can now use Proposition 3.3 to state the definition of the mode.

Definition 3.4. (a) For every $\delta > 0$ assume that $M^Q(\delta) \neq \emptyset$ and let $x \in \mathcal{F}$ such that $\{x\} = \bigcap_{\delta>0} \overline{M^Q(\delta)}$. Then we set $\text{Mod}(Q) := x$ and call $\text{Mod}(Q)$ the *mode* of the probability distribution Q.

(b) By

$$\mathcal{L}_{(\mathcal{F},d)} := \{Q' \in \mathcal{W}_{(\mathcal{F},d)} | \forall \delta > 0 \colon M^{Q'}(\delta) \neq \emptyset\}$$

we denote the *set of all probability measures on* $(\mathcal{F}, \mathbb{B}(\mathcal{F}))$ *with a mode.*

It follows from Definition 3.4(a) that the mode of a probability distribution is always unique.

Remarks 3.5. (a) If $Q' \in \mathcal{L}_{(\mathcal{F},d)}$ and $(\delta_n)_{n\in\mathbb{N}}$ is a positive sequence that converges to zero, then any sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{F}$ such that $x_n \in M^{Q'}(\delta_n)$ for every $n \in \mathbb{N}$ fulfills $d(x_n, \text{Mod}(Q')) \overset{n\to\infty}{\longrightarrow} 0$.

(b) Let $(\delta_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of positive real numbers that converges to zero. To prove that $Q \in \mathcal{L}_{(\mathcal{F},d)}$ it suffices to find some $x_n = x_n(\delta_n) \in \mathcal{F}$ for every $n \in \mathbb{N}$ such that

$$\liminf_{\varepsilon\to 0^+} \inf_{d(x_n,y)\geq\delta_n} \frac{\varphi^Q_{x_n}(\varepsilon)}{\varphi^Q_y(\varepsilon)} > 1.$$

This follows from the circumstance that the sets $M^Q(\delta)$ are nested. Then we can use our result in (a) to ascertain the mode by taking the limit of the sequence $(x_n)_n$.

(c) According to Definition 3.4(a) and (3.1), the mode is equal to some point $x \in \mathcal{F}$ such that there exist points in its neighbourhood at which the small ball probabilities are large. The mode itself is not required to be such a point. For instance, if $(\mathcal{F}, d) = (\mathbb{R}, d_{Euc})$ and, for every $A \in \mathbb{B}(\mathbb{R})$, we have $Q(A) = \int_A f(x)\,dx$, where $f(x) := 2x \cdot 1_{[0,1]}(x), \forall x \in \mathbb{R}$, then by Proposition 3.9 we have $Q \in \mathcal{L}_{(\mathbb{R},d_{Euc})}$ and $\text{Mod}(Q) = 1$ (see Figure 3.1). However one can show that, for $\delta > 0$ small, we have $1 \notin M^Q(\delta)$. In the final section of this chapter we will show strong consistency for a mode estimator under additional requirements which are fulfilled in this case (see Corollary 3.26).

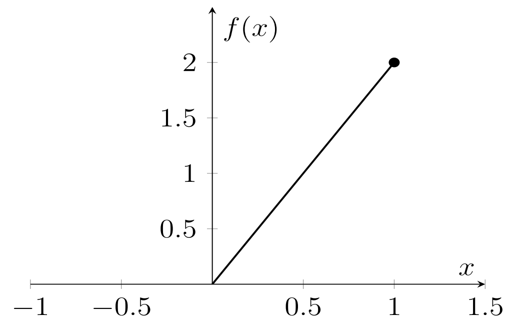

Figure 3.1: Depiction of the density f from Remark 3.5(c)

(d) The infimum in the definition of the sets in (3.1) eliminates some pathological cases. For instance, if we assume that $(\mathcal{F}, d) = (\mathbb{R}^k, d_{Euc})$ and $Q \ll \lambda\!\!\lambda^k$ such that $f \in dQ/d\lambda\!\!\lambda^k$ is continuous, attains a unique global maximum at $x \in \mathbb{R}^k$ and there is a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^k$ that is bounded away from x such that $f(x_n) \overset{n\to\infty}{\longrightarrow} f(x)$, then there exists some $\delta' > 0$ such that $M^Q(\delta) = \emptyset$ for every $\delta \in (0, \delta')$ (see Corollary 3.10(a)). A critical example is given in Example 3.11(a). In the literature, this case is typically avoided by imposing additional constraints for the steepness of the density around the mode (e.g. Abraham et al. (2003), Sager (1979) or Meister (2011)).

(e) If we set $\sup \emptyset := 0$, then we can deduce that

$$M^Q(\delta) = \left\{ x \in \operatorname{supp}(Q) \,\middle|\, \liminf_{\varepsilon\to 0^+} \frac{\varphi_x^Q(\varepsilon)}{\sup\limits_{d(x,y)\geq\delta} \varphi_y^Q(\varepsilon)} > 1 \right\}, \qquad \forall \delta > 0.$$

3.2 Examples

We will establish conditions under which some probability measure Q on $(\mathcal{F}, \mathbb{B}(\mathcal{F}))$ has a mode.

3.2.1 Distributions with densities

In this subsection, results will be given under the assumption that Q has a density. We will first turn our attention to the case where Q is a countable sum of Dirac measures and will later assume that there exists a density function with respect to the k-dimensional Lebesgue measure on $(\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k))$. Under that requirement, the small ball probability functions of Q can be expressed as an integral. Hence, we present a version of the

Lebesgue differentiation theorem that can be applied to evaluate the limit of a quotient of two small ball probability functions as the radius tends to zero.

Proposition 3.6. *For every* $x \in \mathcal{F}$*, let* $m_x \in \mathcal{W}_{(\mathcal{F},d)}$ *denote the Dirac measure at* x*.*

(a) Let $n \in \mathbb{N}, \{x_1, \dots, x_n\} \subseteq \mathcal{F}, q_1, \dots, q_n > 0$ *such that* $\sum_{i=1}^n q_i = 1$ *and assume that* $Q = \sum_{i=1}^n q_i m_{x_i}$*. If* $j \in \{1, \dots, n\}$*, then we have*

$$q_j > \max_{i \neq j} q_i \iff \big(Q \in \mathcal{L}_{(\mathcal{F},d)} \wedge \mathrm{Mod}(Q) = x_j\big),$$

where we set $\max \emptyset := 0$.

(b) Let $\mathcal{X} = \{x_1, x_2, \dots\} \subseteq \mathcal{F}$ *be countably infinite,* $q_1, q_2, \dots > 0$ *such that* $\sum_{i=1}^\infty q_i = 1$ *and assume that* $Q = \sum_{i=1}^\infty q_i m_{x_i}$*. If* $j \in \mathbb{N}$*, then we have*

$$q_j > \max_{i \neq j} q_i \iff \big(Q \in \mathcal{L}_{(\mathcal{F},d)} \wedge \mathrm{Mod}(Q) = x_j\big).$$

Proof: Let us first realise that if $x \in \mathcal{F}$, then

$$\lim_{\varepsilon \to 0^+} \varphi_x^Q(\varepsilon) = \lim_{n \to \infty} Q(B_d(x, 1/n)) = Q(\{x\}).$$

(a) If $n = 1$, then $Q = m_{x_1}$ and we obviously have $x_1 \in M^Q(\delta)$ for every $\delta > 0$, so let us assume that $n \geq 2$.
($\Longrightarrow$). Set $\rho := \frac{1}{2} \min_{i \neq i'} d(x_i, x_{i'}) > 0$. Then if $A \subseteq \{x_1, \dots, x_n\}$ is non-empty and $\varepsilon \leq \rho$, we have $\sup\limits_{x \in A} \varphi_x^Q(\varepsilon) = \max\limits_{x_i \in A} Q(\{x_i\})$. Hence, we can conclude that if $\delta \leq \rho$, then

$$\liminf_{\varepsilon \to 0^+} \inf_{d(x_j,y) \geq \delta} \frac{\varphi_{x_j}^Q(\varepsilon)}{\varphi_y^Q(\varepsilon)} = \frac{q_j}{\max\limits_{i \neq j} q_i} > 1$$

and thus $x_j \in M^Q(\delta)$ for every $\delta > 0$.

($\Longleftarrow$). Since $x_j \in \overline{M^Q(\delta)}$ for every $\delta > 0$ and $B_d(x_j, \varepsilon) \cap \mathrm{supp}(Q) = \{x_j\}$ for every $\varepsilon \in (0, \rho]$, we can conclude that $M^Q(\delta) = \{x_j\}$ for all $\delta \leq \rho$. If we assume that there was some $i \neq j$ such that $q_i \geq q_j$, then for every $\delta \in (0, \rho]$ we would have

$$\inf_{d(x_j,y) \geq \delta} \frac{\varphi_{x_j}^Q(\varepsilon)}{\varphi_y^Q(\varepsilon)} \leq \frac{\varphi_{x_j}^Q(\varepsilon)}{\varphi_{x_i}^Q(\varepsilon)} \overset{\varepsilon \to 0^+}{\longrightarrow} \frac{q_j}{q_i} \leq 1,$$

which is equivalent to $x_j \notin M^Q(\delta)$ for $\delta \leq \rho$.

(b) $(\Longrightarrow)$. Since $q_n \overset{n\to\infty}{\longrightarrow} 0$, we can fix some $x_{j'} \in \mathcal{X} \setminus \{x_j\}$ for which $Q(\{x_{j'}\}) = q_{j'} = \max_{i\neq j} q_i < q_j$ holds. We can choose a sufficiently small $\alpha > 0$ and find a set $\mathcal{X}' \subseteq \mathcal{X}$ such that the following conditions are fulfilled:

(i) $2 \leq |\mathcal{X}'| < \infty$

(ii) $Q(\mathcal{X}') \geq 1 - \alpha$

(iii) $q_{j'} + \alpha < q_j$

Define

$$\rho := \frac{1}{2} \min_{\substack{x_i, x_{i'} \in \mathcal{X}' \\ x_i \neq x_{i'}}} d(x_i, x_{i'}) > 0.$$

Then, for all $x \in \mathcal{X}'$ and $\varepsilon \in (0, \rho]$, we have $Q(B_d(x, \varepsilon) \cap \mathcal{X}') = Q(\{x\})$. Since $Q(\{x_j\}) = q_j > \alpha$, it follows by condition (ii) that $x_j \in \mathcal{X}'$ and we can deduce

$$\sup_{d(x_j,y)\geq\delta} \varphi_y^Q(\varepsilon) \leq q_{j'} + \alpha < q_j \leq \varphi_{x_j}^Q(\varepsilon), \qquad \forall\delta > 0, \forall\varepsilon \in (0, \rho],$$

which is equivalent to

$$\frac{\varphi_{x_j}^Q(\varepsilon)}{\sup\limits_{d(x_j,y)\geq\delta} \varphi_y^Q(\varepsilon)} \geq \frac{q_j}{q_{j'} + \alpha} > 1, \qquad \forall\delta > 0, \forall\varepsilon \in (0, \rho].$$

Thus, we have proven that $x_j \in M^Q(\delta)$ for every $\delta > 0$.

$(\Longleftarrow)$. We will first show that $x_j \in M^Q(\delta)$ for every $\delta > 0$. Let us assume there was some $\delta' > 0$ such that $x_j \notin M^Q(\delta')$, which implies that $x_j \notin M^Q(\delta)$ for every $\delta \in (0, \delta']$. Fix some $\delta \in (0, \delta']$ and recall that $x_j \in \overline{M^Q(\delta)}$. Then there is a sequence $(y_n)_{n\in\mathbb{N}}$ in $M^Q(\delta)$ that converges to x_j as n tends to infinity and fulfills $y_n \neq y_m$ for every $n, m \in \mathbb{N}$ such that $n \neq m$. It follows from $\mathrm{diam}(M^Q(\delta)) \leq \delta$ that we can fix some $z(\delta) \in \mathcal{F}$ such that $M^Q(\delta) \subset B_d(z(\delta), \delta/2)$. Since $\mathrm{diam}(\mathcal{X}) > 0$, if $\delta > 0$ is sufficiently small, then there exists some $x \in \mathcal{X} \setminus B_d(z(\delta), 3\delta/2)$. Applying the reverse triangle inequality yields

$$d(x, y_n) \geq |d(x, z(\delta)) - d(z(\delta), y_n)| \geq \delta$$

for every $n \in \mathbb{N}$.

Since $y_n \in M^Q(\delta)$, this leads to

$$\frac{Q(\{y_n\})}{Q(\{x\})} = \liminf_{\varepsilon \to 0^+} \frac{\varphi^Q_{y_n}(\varepsilon)}{\varphi^Q_x(\varepsilon)} \geq \liminf_{\varepsilon \to 0^+} \inf_{y: d(y_n, y) \geq \delta} \frac{\varphi^Q_{y_n}(\varepsilon)}{\varphi^Q_y(\varepsilon)} > 1$$

for every $n \in \mathbb{N}$, which would yield $Q\big(\bigcup_{n \in \mathbb{N}} \{y_n\}\big) = \infty$, because the values of the sequence $(y_n)_n$ are pairwise different. We can conclude that $x_j \in M^Q(\delta)$ for every $\delta > 0$ and we can deduce that $q_j > 0$. If we assume there was some $i \neq j$ such that $q_i \geq q_j$, then we would have

$$\frac{\varphi^Q_{x_j}(\varepsilon)}{\varphi^Q_{x_i}(\varepsilon)} \overset{\varepsilon \to 0^+}{\longrightarrow} \frac{q_j}{q_i} \leq 1,$$

which contradicts that $x_j \in M^Q(\delta)$ for every $\delta \leq d(x_i, x_j)$.

∎

We will now assume that $(\mathcal{F}, d) = (\mathbb{R}^k, d_{Euc})$ is the k-dimensional Euclidean space and $Q \ll \lambda\!\!\!\lambda^k$, where $\lambda\!\!\!\lambda^k$ denotes the k-dimensional Lebesgue measure for some $k \in \mathbb{N}$. We will set $\lambda\!\!\!\lambda := \lambda\!\!\!\lambda^1$. For every $\varepsilon > 0$ we set $v_k(\varepsilon) := \lambda\!\!\!\lambda^k(B_{d_{Euc}}(0, \varepsilon))$. Let $dQ/d\lambda\!\!\!\lambda^k$ denote the set containing all non-negative versions of the Lebesgue density of Q. In that case if $f \in dQ/d\lambda\!\!\!\lambda^k$, then for every $x \in \mathbb{R}^k$ and $\varepsilon > 0$ we have

$$\varphi^Q_x(\varepsilon) = \int_{B_{d_{Euc}}(x,\varepsilon)} f(y)\, \lambda\!\!\!\lambda^k(dy).$$

The following version of the Lebesgue differentiation theorem (e.g. Rudin (1987), Theorem 7.7) can be used to evaluate the limit of

$$\big(v_k(\varepsilon)\big)^{-1} \int_{B_{d_{Euc}}(x,\varepsilon)} f(y) \lambda\!\!\!\lambda^k(dy)$$

as $\varepsilon \to 0^+$ if x is a so called *Lebesgue point* of f.

Theorem 3.7. *Let $f \in L^1(\mathbb{R}^k, \mathbb{B}(\mathbb{R}^k), \lambda\!\!\lambda^k)$ and define*

$$N_f := N := \left\{ x \in \mathbb{R}^k \,\middle|\, \lim_{\varepsilon \to 0^+} \frac{1}{v_k(\varepsilon)} \int_{B_{d_{Euc}}(x,\varepsilon)} f(y)\, \lambda\!\!\lambda^k(dy) = f(x) \right\}.$$

Then the following statements hold:

(a) $\lambda\!\!\lambda^k(\mathbb{R}^k \setminus N) = 0$

(b) $C_f := C := \{x \in \mathbb{R}^k | f \text{ is continuous at } x\} \subseteq N$

Proof:

(a) This is the well known *Lebesgue differentiation theorem* (e.g. Rudin (1987), Theorem 7.7).

(b) We can assume that $C \neq \emptyset$ and fix some $x \in C$. For every $r > 0$ we can find $\delta = \delta(r) > 0$ such that $d_{Euc}(x, y) \leq \delta$ implies $|f(x) - f(y)| \leq r$. For every $\varepsilon \in (0, \delta)$ this yields

$$\begin{aligned} \left| \frac{1}{v_k(\varepsilon)} \int\limits_{B_{d_{Euc}}(x,\varepsilon)} f(y)\, \lambda\!\!\lambda^k(dy) - f(x) \right| &= \left| \frac{1}{v_k(\varepsilon)} \int\limits_{B_{d_{Euc}}(x,\varepsilon)} \big(f(y) - f(x)\big)\, \lambda\!\!\lambda^k(dy) \right| \\ &\leq \frac{1}{v_k(\varepsilon)} \int\limits_{B_{d_{Euc}}(x,\varepsilon)} |f(x) - f(y)|\, \lambda\!\!\lambda^k(dy) \\ &\leq r, \end{aligned}$$

which implies $x \in N$.

■

Remark 3.8. Heinonen et al. (2015) extend this version of the Lebesgue differentiation theorem to the more general setting where f is a locally integrable function defined on a separable metric space $(\mathcal{S}, \rho)$ that takes values in a Banach space (e.g., see (3.4.8) and (3.4.11) therein). The results they give are stated for an abstract reference measure ν that fulfills some technical requirements, which are satisfied if ν is a non-trivial Borel-measure where every open ball has a finite and positive measure such that there exists a constant $C \geq 1$ so that

$$\nu(B^o_\rho(x, 2\varepsilon)) \leq C\nu(B^o_\rho(x, \varepsilon))$$

holds for every $x \in \mathcal{S}$ and $\varepsilon > 0$. Therefore, it is possible to consider densities on more abstract spaces and still exploit the statement of the Lebesgue differentiation theorem, as we will now do in the Euclidean setting.

The set N_f as defined in Theorem 3.7 is also called the set of *Lebesgue points* of f. It will be used in the two following results.

Proposition 3.9. *We assume there exists a version $f \in dQ/d\lambda\!\!\!\lambda^k$ of the Lebesgue density of Q and some $x^M \in \mathbb{R}^k$ such that*

$$\sup_{d_{Euc}(x,x^M)\geq r} f(x) < \infty, \qquad \forall r > 0. \tag{3.2}$$

We define $\tilde{f}(x) := f(x) \cdot 1_{N_f}(x)$ for every $x \in \mathbb{R}^k$. Then we have

$$\Big(\forall r > 0: \sup_{d_{Euc}(x,x^M)\geq r} \tilde{f}(x) < \sup_{\mathbb{R}^k} \tilde{f}(x)\Big) \iff \Big(Q \in \mathcal{L}_{(\mathbb{R}^k,d_{Euc})} \wedge \operatorname{Mod}(Q) = x^M\Big).$$

Proof: ($\Longrightarrow$). It follows from Theorem 3.7(a) that $\tilde{f} \in dQ/d\lambda\!\!\!\lambda^k$. Additionally, we have $f(x) \geq \tilde{f}(x)$ for every $x \in \mathbb{R}^k$.
There exists a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^k$ such that $d_{Euc}(x_n, x^M) \overset{n\to\infty}{\longrightarrow} 0$ and $\tilde{f}(x_n) \overset{n\to\infty}{\longrightarrow} \sup_{\mathbb{R}^k} \tilde{f}(x)$, because if there was no such sequence, then there would be some $\rho > 0$ and another sequence $(x'_n)_{n\in\mathbb{N}}$ of points in $\{x \in \mathbb{R}^k | d_{Euc}(x, x^M) \geq \rho\}$ that fulfills $\tilde{f}(x'_n) \overset{n\to\infty}{\longrightarrow} \sup_{\mathbb{R}^k} \tilde{f}(x)$, which contradicts that

$$\sup_{\mathbb{R}^k} \tilde{f}(x) > \sup_{d_{Euc}(x,x^M)\geq\rho} \tilde{f}(x).$$

Fix $\delta > 0$. Then there exists some $n(\delta) \in \mathbb{N}$ such that both $d_{Euc}(x^M, x_{n(\delta)}) < \delta$ and $\tilde{f}(x_{n(\delta)}) > \sup_{d_{Euc}(x,x^M)\geq\delta} \tilde{f}(x)$ hold. This implies that $\tilde{f}(x_{n(\delta)}) > \sup_{d_{Euc}(x,x_{n(\delta)})\geq 2\delta} \tilde{f}(x)$. Since $\tilde{f}(x_{n(\delta)}) > 0$, we have $x_{n(\delta)} \in N_f$, which implies $x_{n(\delta)} \in \operatorname{supp}(Q)$. Thus, we can conclude that

$$\liminf_{\varepsilon\to 0^+} \frac{\varphi^Q_{x_{n(\delta)}}(\varepsilon)}{\sup\limits_{d_{Euc}(x,x_{n(\delta)})\geq 3\delta} \varphi^Q_x(\varepsilon)} \geq \liminf_{\varepsilon\to 0^+} \frac{\varphi^Q_{x_{n(\delta)}}(\varepsilon)}{v_k(\varepsilon) \sup\limits_{d_{Euc}(x,x_{n(\delta)})\geq 2\delta} \tilde{f}(x)} = \frac{\tilde{f}(x_{n(\delta)})}{\sup\limits_{d_{Euc}(x,x_{n(\delta)})\geq 2\delta} \tilde{f}(x)} > 1,$$

where the first inequality holds for ε sufficiently small (e.g. smaller than or equal to δ) and the final step is due to $f(x_{n(\delta)}) = \tilde{f}(x_{n(\delta)})$. This, however, yields $x_{n(\delta)} \in M^Q(3\delta)$ and shows that $M^Q(\delta) \neq \emptyset$ for every $\delta > 0$, which implies $Q \in \mathcal{L}_{(\mathbb{R}^k,d_{Euc})}$. Since $d_{Euc}(x_n, x^M) \overset{n\to\infty}{\longrightarrow} 0$, we have $\operatorname{Mod}(Q) = x^M$ (see Remark 3.5(a)).

($\Longleftarrow$). We can assume that $\sup_{\mathbb{R}^k} \tilde{f}(x) < \infty$ because by (3.2), if $\sup_{\mathbb{R}^k} \tilde{f}(x) = \infty$, then

$$\sup_{d_{Euc}(x,x^M)\geq r} \tilde{f}(x) \leq \sup_{d_{Euc}(x,x^M)\geq r} f(x) < \infty = \sup_{\mathbb{R}^k} \tilde{f}(x)$$

for every $r > 0$ so that the desired condition holds. If there was some $r' > 0$ such that

$$\sup_{d_{Euc}(x^M,x)\geq r'} \tilde{f}(x) = \sup_{\mathbb{R}^k} \tilde{f}(x),$$

then there would be a sequence $(x_n)_{n\in\mathbb{N}}$ of points located in the set $\{x \in \mathbb{R}^k | d_{Euc}(x, x^M) \geq r'\}$ such that $\lim_{n\to\infty} \tilde{f}(x_n) = \sup_{\mathbb{R}^k} \tilde{f}(x)$. We certainly have $\sup_{\mathbb{R}^k} \tilde{f}(x) > 0$, because otherwise we would have $\tilde{f} \equiv 0$ which contradicts $\tilde{f} \in dQ/d\lambda^k$. Hence, if $n' \in \mathbb{N}$ is chosen to be sufficiently large, then for every $n \geq n'$ we have that $\tilde{f}(x_n) > 0$, which implies $x_n \in N_f$ and, thus, $x_n \in \mathrm{supp}(Q)$.
Now let $\delta \in (0, r'/2)$ and fix some $y \in M^Q(\delta)$. This implies $d_{Euc}(x^M, y) \leq \delta < r'/2$. Now we can derive for $n \geq n'$ that

$$\liminf_{\varepsilon\to 0^+} \frac{\varphi_y^Q(\varepsilon)}{\sup\limits_{x:d_{Euc}(x,y)\geq\delta} \varphi_x^Q(\varepsilon)} \leq \liminf_{\varepsilon\to 0^+} \frac{v_k(\varepsilon) \cdot \sup_{\mathbb{R}^k} \tilde{f}(x)}{\varphi_{x_n}^Q(\varepsilon)} = \frac{\sup_{\mathbb{R}^k} \tilde{f}(x)}{\tilde{f}(x_n)} \overset{n\to\infty}{\longrightarrow} 1,$$

which contradicts the fact that $y \in M^Q(\delta)$. In consequence, such an $r' > 0$ does not exist.

■

Proposition 3.9 can be applied to certain Lebesgue-continuous distributions to identify their mode at a location where none of the versions of the density is continuous. In fact, we do not even necessarily require that there exists a version that is bounded in a neighbourhood of the mode.
When there is no continuous density, characterising modes via density functions is usually done by referring to a specific version. We can pick any version f that fulfills (3.2) and then check whether $f \cdot 1_{N_f}$ fulfills the left side of the equivalence statement of our proposition.
The following corollary, which is an application of Proposition 3.9, states that if there indeed exists a continuous version of the density f (which implies $N_f = \mathbb{R}^k$), then (3.2) is fulfilled if either of the conditions of the equivalence relation in Proposition 3.9 is fulfilled.

Corollary 3.10. *Assume there exists $f \in dQ/d\lambda\!\!\!\lambda^k$ such that f is continuous at every $x \in \mathbb{R}^k$. Then the following statements hold:*

(a) Let $x^M \in \mathbb{R}^k$. Then we have

$$\Big(\forall r > 0 : \sup_{d_{Euc}(x,x^M)\geq r} f(x) < f(x^M)\Big) \Longleftrightarrow \Big(Q \in \mathcal{L}_{(\mathbb{R}^k,d_{Euc})} \wedge \mathrm{Mod}(Q) = x^M\Big).$$

(b) If, for some $x^M \in \mathbb{R}^k$, we have $Q \in \mathcal{L}_{(\mathbb{R}^k,d_{Euc})}$ and $\mathrm{Mod}(Q) = x^M$, then $f(x^M) > f(x)$ holds for every $x \in \mathbb{R}^k \setminus \{x^M\}$.

Proof:

(a) We will first show that both the left and the right side of the equivalence statement imply the boundedness requirement (3.2) of Proposition 3.9. It is clear that if the left side can be assumed, then (3.2) is obviously fulfilled, so let us assume that $Q \in \mathcal{L}_{(\mathbb{R}^k,d_{Euc})}$ and $\mathrm{Mod}(Q) = x^M \in \mathbb{R}^k$.
Let $r > 0$. It follows from the continuity of f that the set of Lebesgue points N_f is equal to the space $\mathbb{R}^k$ (see Theorem 3.7(b)) and that $\max_{x\in B_{d_{Euc}}(x^M,r)} f(x) < \infty$. If $z \in M^Q(r/2)$ and $y \in \mathbb{R}^k$ exists such that $d_{Euc}(y, x^M) \geq r$ and $f(y) > 0$, then we have

$$\frac{\varphi_z^Q(\varepsilon)}{\varphi_y^Q(\varepsilon)} = \frac{\varphi_z^Q(\varepsilon)/v_k(\varepsilon)}{\varphi_y^Q(\varepsilon)/v_k(\varepsilon)} \overset{\varepsilon\to 0^+}{\longrightarrow} \frac{f(z)}{f(y)} > 1.$$

If such a y does not exist, then it is an immediate consequence that $f|_{\mathbb{R}^k\setminus B^o_{d_{Euc}}(x^M,r)} \equiv 0$. We can conclude that $\sup_{d_{Euc}(x,x^M)\geq r} f(x) \leq f(z) < \infty$.

$(\Longrightarrow)$. Let $\delta > 0$ and $\varepsilon \in (0, \delta/2)$. Since $N_f = \mathbb{R}^k$, we have

$$\frac{\varphi_{x^M}^Q(\varepsilon)}{\sup\limits_{d_{Euc}(x,x^M)\geq\delta} \varphi_x^Q(\varepsilon)} \geq \frac{\varphi_{x^M}^Q(\varepsilon)}{v_k(\varepsilon) \sup\limits_{d_{Euc}(x,x^M)\geq\delta/2} f(x)} \overset{\varepsilon\to 0^+}{\longrightarrow} \frac{f(x^M)}{\sup\limits_{d_{Euc}(x,x^M)\geq\delta/2} f(x)} > 1,$$

which shows that $x^M \in M^Q(\delta)$ for any $\delta > 0$.
$(\Longleftarrow)$. If we can prove that $\sup_{\mathbb{R}^k} f(x) = f(x^M)$ then this is the result of the application of Proposition 3.9. However, if there was some $x' \in \mathbb{R}^k$ such that $f(x') > f(x^M)$ then it is due to the continuity of f that for some sufficiently small $\delta > 0$ we have $f(x') > \max_{x\in B_d(x^M,2\delta)} f(x)$. Since $N_f = \mathbb{R}^k, x^M \in \overline{M^Q(\delta)}$ and $\mathrm{diam}\, M^Q(\delta) \leq \delta$, this is impossible.

(b) This follows immediately from (a), because the left side of the equivalence statement of (a) is indeed stronger than the inequality to prove.

■

For a Lebesgue-continuous distribution with a continuous density and a mode, that mode is always equal to the unique global maximum point of the density. As mentioned before, there are cases of distributions with continuous densities with a unique global maximum point that do not have a mode according to Definition 3.4(a). An example is given below.

Examples 3.11. (a) For every $\alpha, \beta > 0$ and $\gamma \in \mathbb{R}$ define

$$h_{\alpha,\beta,\gamma}(x) := \alpha\left(1 - \frac{|x-\gamma|}{\beta}\right) \cdot \mathbf{1}_{[\gamma-\beta,\gamma+\beta]}(x), \qquad \forall x \in \mathbb{R}$$

and let $f \in dQ/d\lambda\!\!\lambda^k$ such that

$$c \cdot f(x) = h_{1,1,0}(x) + \sum_{n=2}^{\infty} h_{1-\frac{1}{n^2},\frac{2}{n^2},n}(x), \qquad \forall x \in \mathbb{R},$$

where $c > 0$ is the normalization constant (see Figure 3.2). Since

$$\int h_{1-\frac{1}{n^2},\frac{2}{n^2},n}(x)\,dx \leq \frac{2}{n^2},$$

such a constant exists indeed and f is a density function.

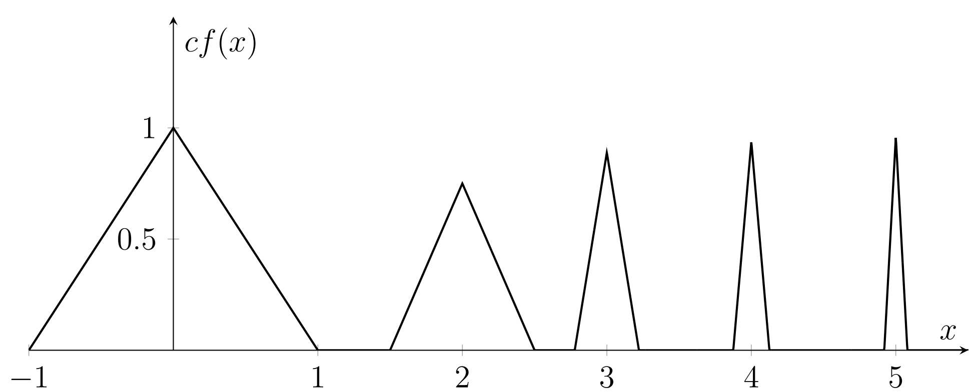

Figure 3.2: Depiction of the function $c \cdot f$ of Example 3.11(a) on the interval $[-1, 5.5]$

Then, f is continuous, $f(0) > f(x)$ holds for every $x \neq 0$ and we have $f(n) \overset{n\to\infty}{\longrightarrow} f(0)$. We can use Corollary 3.10(a) to deduce that $Q \notin \mathcal{L}_{(\mathbb{R},d_{Euc})}$, because the condition on the left side of the equivalence statement therein is not fulfilled. Consequently, we will not consider this or comparable cases in our further analyses, as is often done in the literature on mode estimation. Authors typically impose further conditions on the density to eliminate pathological cases as the one given above. For instance, Sager (1978) imposes that $\theta \in \mathbb{R}^k$ is the mode of a density g if and only if for every $\varepsilon > 0$ there is $\delta > 0$ such that $d_{Euc}(x, \theta) > \varepsilon$ implies $g(\theta) > g(x) + \delta$, which is similar to the condition (9) imposed by Meister (2011), who considers the case $k = 1$ and demands that $g(\theta) - g(x) \geq C \cdot \min(1, d_{Euc}(\theta, x)^2)$ is valid for every $x \in \mathbb{R}$. Herrmann and Ziegler (2004) also consider a univariate setting and propose a condition which is very similar to the one given in our Corollary 3.10(a): $g(\theta) > \sup_{d_{Euc}(x,\theta)>\varepsilon} g(x)$ for every $\varepsilon > 0$. Another common constraint is about the level sets of the density g (e.g. Abraham et al. (2003)). If $g(\theta) < \infty$ is assumed, then all of these different constraints imply (3.2).

(b) The boundedness requirement (3.2) is not a necessary condition for the existence of a mode of the (k-variate and Lebesgue-continuous) distribution Q. For instance, if the density has more than one singularity point, then there may still be a mode. E.g., if $k \geq 3$ and $x_1, x_2 \in \mathbb{R}^k$ such that $d_{Euc}(x_1, x_2) > 2$ and $f \in dQ/d\boldsymbol{\lambda}^k$ exists such that

$$f(x) = c_k \cdot \Big(\frac{1}{\|x - x_1\|_{Euc}^2} \cdot \mathbf{1}_{B_{d_{Euc}}(x_1,1)}(x) + \frac{1}{\|x - x_2\|_{Euc}} \cdot \mathbf{1}_{B_{d_{Euc}}(x_2,1)}(x) \Big), \quad \forall x \neq x_1, x_2,$$

where $c_k > 0$, then $Q \in \mathcal{L}_{(\mathbb{R}^k, d_{Euc})}$ and $\text{Mod}(Q) = x_1$. If $\delta \in (0, 2]$ then there exists $\varepsilon(\delta) > 0$ such that $\sup_{y: d_{Euc}(x_1,y) \geq \delta} \varphi_y^Q(\varepsilon) = \varphi_{x_2}^Q(\varepsilon)$ holds for every $\varepsilon \in (0, \varepsilon(\delta))$. It follows from $\varphi_{x_1}^Q(\varepsilon) \asymp \varepsilon^{k-2}$ and $\varphi_{x_2}^Q(\varepsilon) \asymp \varepsilon^{k-1}$ as $\varepsilon \to 0^+$ that

$$\lim_{\varepsilon \to 0^+} \frac{\varphi_{x_1}^Q(\varepsilon)}{\sup\limits_{d_{Euc}(x_1,y) \geq \delta} \varphi_y^Q(\varepsilon)} = \lim_{\varepsilon \to 0^+} \frac{\varphi_{x_1}^Q(\varepsilon)}{\varphi_{x_2}^Q(\varepsilon)} = \infty,$$

which yields $x_1 \in M^Q(\delta)$.

3.2.2 Gaussian measures

In order to apply our Definition 3.4(a) for the purpose of deciding whether a specific distribution Q has a mode or not, it is necessary to have some knowledge on the asymptotic behaviour of the small ball probability functions $\varphi_\cdot^Q$. In the preceding paragraph, densities were available, which made these probabilities accessible (e.g. through the Lebesgue differentiation theorem). One can generalise and extend our Proposition 3.9

to abstract ν-densities on separable metric spaces for which the conditions collected in our Remark 3.8 hold, which would allow us to apply the differentiation theorem. However, if these conditions are not met in an arbitrary infinite-dimensional setting, then computing the small ball probabilities explicitly or deriving their (logarithmic) asymptotic behaviour are usually very complicated problems and results exist only for a few special cases, e.g. for Gaussian measures defined on a Hilbert space or on the space of continuous functions. The small ball problem for the special case where Q is Gaussian has been thoroughly studied in the last decades (e.g. Dunker et al. (1991), Li and Shao (2001), Gao et al. (2003a), Gao et al. (2003b) and Bogachev (2015)) and we want to use these results to extend our analysis of examples to the setting where Q is a Gaussian measure.

While the research on small ball probabilities is usually concerned with describing the asymptotics where the centre point $x \in \mathcal{F}$ is the mean, we can apply the Cameron-Martin theorem (see Proposition 3.15(a)) to Gaussian measures to extend the existing results to certain other centre points $x' \in \mathcal{F}$ which lie in the Cameron-Martin space of Q.

Firstly, we will collect some of the main ideas of Gaussian measures and refer to Bogachev (2015) for a detailed report. We assume that $(\mathcal{F}, \|\cdot\|)$ is a separable Banach space and that Q is a probability measure on $(\mathcal{F}, \mathbb{B}(\mathcal{F}))$. By

$$\mathcal{F}^* := \{f : \mathcal{F} \to \mathbb{R} | f \text{ is linear and continuous}\}$$

we denote the topological dual space of $\mathcal{F}$. Furthermore, for every $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, let $\mathcal{N}_{\mu,\sigma^2}$ denote the univariate normal distribution (e.g.

$$\frac{d\mathcal{N}_{\mu,\sigma^2}}{d\lambda}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

for every $x \in \mathbb{R}$) and let us write $\mathcal{N}_{\mu,0} := m_\mu$ for every $\mu \in \mathbb{R}$.

Definition 3.12. (a) Q is called *Gaussian* if for every $f \in \mathcal{F}^*$ there exist $\mu \in \mathbb{R}$ and $\sigma^2 \geq 0$ such that $Q \circ f^{-1} = \mathcal{N}_{\mu,\sigma^2}$.

Let Q be Gaussian.

(b) The mapping

$$a_Q : \mathcal{F}^* \longrightarrow \mathbb{R}, \quad f \longmapsto \int_{\mathcal{F}} f(x)\, Q(dx)$$

is called the *mean* of Q.

(c) Q is called *centred* if, for every $f \in \mathcal{F}^*$, we have $a_Q(f) = 0$.

(d) The mapping

$$b_Q : \mathcal{F}^* \times \mathcal{F}^* \longrightarrow \mathbb{R}, \quad (f,g) \longmapsto \int_{\mathcal{F}} (f(x) - a_Q(f))(g(x) - a_Q(g))\, Q(dx)$$

is called the *covariance* of Q.

(e) For every $h \in \mathcal{F}$ we define $\|h\|_{H(Q)} := \sup_{\substack{f \in \mathcal{F}^* \\ b_Q(f,f) \leq 1}} f(h)$. The space

$$H(Q) := \{h \in \mathcal{F} : \|h\|_{H(Q)} < \infty\}$$

is called the *Cameron-Martin space* of Q.

It is elaborated in Bogachev (2015) that for a Gaussian measure Q on a separable Banach space $(\mathcal{F}, \mathbb{B}(\mathcal{F}))$ the mean a_Q can be represented by an element $h \in \mathcal{F}$ (see Remark 2.2.9 and Theorem 3.2.3 therein). From now on, we will identify the mean a_Q with that element.
A stochastic process $(X_t)_{t\in[0,1]}$ on a probability space $(\Omega, \mathcal{A}, P)$ is called a *Gaussian process* if for every $k \in \mathbb{N}, \lambda_1, \ldots, \lambda_k \in \mathbb{R}$ and $t_1, \ldots, t_k \in [0,1]$ there exist $\mu \in \mathbb{R}$ and $\sigma^2 \geq 0$ such that $\sum_{j=1}^k \lambda_j X_{t_j} \sim \mathcal{N}_{\mu,\sigma^2}$. The notions of Gaussian processes and Gaussian measures are closely related. Let $(X_t)_{t\in[0,1]}$ be a Gaussian process with paths in $C_0([0,1])$ and let P_X denote the probability measure induced by the random variable

$$X : (\Omega, \mathcal{A}, P) \longrightarrow (C_0([0,1]), \mathbb{B}(C_0([0,1]))), \quad \omega \longmapsto \big(t \mapsto X_t(\omega)\big),$$

e.g. if $A \in \mathbb{B}(C_0([0,1]))$, then $P_X(A) = P(X \in A)$. Then P_X is a Gaussian measure on $(C_0([0,1]), \mathbb{B}(C_0([0,1])))$ which we will refer to as the *induced Gaussian measure*. On the other hand if Q is Gaussian on $(C_0([0,1]), \mathbb{B}(C_0([0,1])))$, then there exists a corresponding Gaussian process with paths in $C_0([0,1])$ that induces Q. The analogue result holds if $C_0([0,1])$ is replaced by $L^2([0,1])$ (and the topology is the one induced by d_2) under an additional measurability requirement (see Theorems 1 and 2 in Rajput and Cambanis (1972)).
The Cameron-Martin space of a Gaussian measure Q is the space of all shifts $h \in \mathcal{F}$ for which Q and the shifted measure $Q_h := Q(\cdot - h)$ are equivalent (see Lemma 3.14(a)), which will be denoted by $Q \sim Q_h$ (and must not be confused with our notation $X \sim Q$ that indicates the law of the random variable X is Q). The pair $(H(Q), \|\cdot\|_{H(Q)})$ is a Hilbert space that is sometimes called the *reproducing kernel Hilbert space* (see Bogachev

(2015), Proposition 2.4.6). It is usually difficult to give an expression for the Cameron-Martin space of a Gaussian measure in terms of other (function) spaces, but we can do so in the special case of the Wiener measure.

Example 3.13. Let P^B denote the Wiener measure, e.g. the Gaussian measure on the space of continuous functions that is induced by the Brownian motion $(B_t)_{t\in[0,1]}$. Recall that the Brownian motion is a Lévy-process with continuous paths such that $B_t - B_s \sim \mathcal{N}_{0,t-s}$ for every $s, t \in [0,1]$ such that $t \geq s$. The Cameron-Martin space $H(P^B)$ is equal to the class of functions $f \in C_0([0,1])$ such that $f(0) = 0$, f is absolutely continuous (which means that the function has a derivative f' almost everywhere, see the text before Proposition 4.7) and f' is square-integrable (see Bogachev (2015), Lemma 2.3.14). The corresponding norm on the space $H(P^B)$ is $\|f\|_{H(P^B)} := \|f'\|_2$ for every $f \in H(P^B)$.

The following lemma contains three properties of the Cameron-Martin space of a Gaussian measure on a separable Banach space.

Lemma 3.14. *Assume that Q is a Gaussian measure and $(H(Q), \|\cdot\|_{H(Q)})$ is the Cameron-Martin space of Q. Then the following statements hold, where $\overline{H(Q)}$ denotes the closure in $\mathcal{F}$:*

(a) $H(Q) = \{h \in \mathcal{F} | Q \sim Q_h\}$

(b) $\exists C > 0 : \forall f \in H(Q) : \|f\| \leq C\|f\|_{H(Q)}$

(c) $\mathrm{supp}(Q) = a_Q + \overline{H(Q)}$.

Proof: These are Theorem 2.4.5, Proposition 2.4.6 and Theorem 3.6.1 in Bogachev (2015).

■

It follows from part (b) of the preceding lemma that the Cameron-Martin space is continuously embedded into the space $\mathcal{F}$.
If $h \in H(Q)$, then $Q \ll Q_h$ and we can use the Radon-Nikodym density of Q with respect to Q_h to make a statement about the limit of the quotient $\varphi_0^Q(\varepsilon)/\varphi_h^Q(\varepsilon)$ as $\varepsilon \to 0^+$. If $h \in \overline{H(Q)}$, then we use the following inequality provided by Li and Shao (2001) in order to give a lower bound.

Proposition 3.15. *Assume that Q is a centred Gaussian measure and let $(H(Q), \|\cdot\|_{H(Q)})$ be the Cameron-Martin space of Q.*

(a) For every $f \in H(Q)$ we have

$$1 \leq \frac{\varphi_0^Q(\varepsilon)}{\varphi_f^Q(\varepsilon)} \leq \exp\left(\frac{1}{2}\|f\|_{H(Q)}^2\right)$$

for every $\varepsilon > 0$. In particular, we have $\lim_{\varepsilon \to 0^+} \frac{\varphi_0^Q(\varepsilon)}{\varphi_f^Q(\varepsilon)} = \exp\left(\frac{1}{2}\|f\|_{H(Q)}^2\right)$ for every $f \in H(Q)$.

(b) Let $\varepsilon > 0$ and $f \in \overline{H(Q)}$ such that $\|f\| \geq \varepsilon$. Then we have

$$\frac{\varphi_0^Q(\varepsilon)}{\varphi_f^Q(\varepsilon)} \geq \exp\left(\frac{1}{2C^2}(\|f\| - \varepsilon)^2\right),$$

where $C > 0$ is the embedding constant taken from Lemma 3.14(b)

Proof: These results are deduced from the Theorems 3.1 and 3.2 in Li and Shao (2001) and the remarks they give in between the theorems. The upper bound in (a) can be proven by making use of the *Cameron-Martin formula* (see, for instance, Corollay 2.4.3 in Bogachev (2015)).

■

It may not be surprising that the mode of a Gaussian measure coincides with its mean. However, we are now in the position to apply our Definition 3.4(a) to Gaussian measures on separable Banach spaces.

Corollary 3.16. *Assume that Q is a Gaussian measure with mean $a_Q \in \mathcal{F}$ and full support* $\text{supp}(Q) = \mathcal{F}$*, then $Q \in \mathcal{L}_{(\mathcal{F},\|\cdot\|)}$ and* $\text{Mod}(Q) = a_Q$.

Proof: It suffices to consider the shifted measure $Q' := Q + a_Q$ and show that $0 \in M^{Q'}(\delta)$ for every $\delta > 0$. Since $\varphi_x^{Q'}(\varepsilon) = \varphi_{x+a_Q}^Q(\varepsilon)$ for every $x \in \mathcal{F}$ and $\varepsilon > 0$, we have $M^{Q'}(\delta) + a_Q = M^Q(\delta)$, which means that if $0 \in M^{Q'}(\delta)$ holds for every $\delta > 0$ then so does $a_Q \in M^Q(\delta)$. One can verify that Q' is a Gaussian measure with mean zero (e.g. via Fourier transformation, see Lemma 2.2.2 and Theorem 2.2.4 in Bogachev (2015)). It follows through the fact that the support of Q is the whole space and Lemma 3.14(c) that $\text{supp}(Q') = \text{supp}(Q)$ and thus $\overline{H(Q')} = \text{supp}(Q') = \mathcal{F}$. Now let $\delta > 0$ and assume that $\varepsilon \leq \delta$. Then, applying Proposition 3.15(b) yields

$$\liminf_{\varepsilon\to 0^+} \inf_{\|f\|\geq\delta} \frac{\varphi_0^{Q'}(\varepsilon)}{\varphi_f^{Q'}(\varepsilon)} \geq \liminf_{\varepsilon\to 0^+} \inf_{\|f\|\geq\delta} \exp\left(\frac{1}{2C^2}(\|f\|-\varepsilon)^2\right)$$
$$\geq \liminf_{\varepsilon\to 0^+} \exp\left(\frac{1}{2C^2}(\delta-\varepsilon)^2\right) = \exp\left(\frac{1}{2C^2}\delta^2\right) > 1.$$

It now follows that $0 \in M^{Q'}(\delta)$ for every $\delta > 0$, which shows that $a_Q \in M^Q(\delta)$ for every $\delta > 0$.

■

Proposition 3.15 provides bounds for the ratio of two small ball probabilities, which is more informative if the asymptotics of $\varphi_0^Q(\varepsilon)$ as $\varepsilon \to 0^+$ are known. Thus, we conclude this paragraph by collecting examples of certain Gaussian measures defined on a Hilbert or Banach space for which these asymptotics are known. One of the stipulations that we will later make is that the small ball probability function of the mode behaves exponentially as the radius tends to zero, as is the case in many of the examples we will discuss below.

Examples 3.17. We will first consider the Hilbert space setting. Therefore, let Q be a centred Gaussian measure on $\big(L^2([0,1]), \mathbb{B}(L^2([0,1]))\big)$ and let $(X_t)_{t\in[0,1]}$ be a square-integrable stochastic process that induces Q. It follows from Theorem 2 in Rajput and Cambanis (1972) that $(X_t)_t$ also has mean zero. Let $K_X(s,t) := \mathbb{E}(X_s X_t)$, $s,t \in [0,1]$, denote the covariance function of the process, which we assume to be continuous. Then, by the Karhunen-Loeve decomposition (e.g. Deheuvels and Martynov (2008)) we know that there is a sequence of independent, standard normal random variables $(Z_j)_{j\in\mathbb{N}}$ such that

$$X_t = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\nu_j(t) Z_j,$$

where the convergence of the series is with respect to the L^2-topology and uniform in t and where $(\nu_j)_{j\in\mathbb{N}}$ is an orthonormal basis of $L^2([0,1])$ formed by the eigenfunctions of the covariance operator of the process and $(\lambda_j)_{j\in\mathbb{N}}$ is the sequence of its eigenvalues. It follows that

$$\varphi_0^Q(\varepsilon) = P\Big(\|X_t\|_2 \leq \varepsilon\Big) = P\left(\sum_{j=1}^{\infty} \lambda_j Z_j^2 \leq \varepsilon^2\right).$$

We can give the exact asymptotic behaviour of $\varphi_0^Q(\varepsilon)$ if the distribution of the sum $\sum_{j=1}^{\infty} \lambda_j Z_j^2$ is known, which requires information on the sequence $(\lambda_j)_j$ of eigenvalues.

The following two results are well known examples (see Berghin et al. (2005)):

(a) Let $(X_t)_{t\in[0,1]}$ be the Brownian motion, then

$$P(\|X_t\|_2 \le \varepsilon) \sim \frac{4}{\pi^{1/2}}\,\varepsilon\,\exp\Big(-\frac{1}{8\varepsilon^2}\Big), \qquad \varepsilon \to 0^+.$$

(b) Let $(X_t)_{t\in[0,1]}$ be the Brownian bridge, i.e. $K_X(s,t) = \min(s,t) - st$, then

$$P(\|X_t\|_2 \le \varepsilon) \sim \frac{2\sqrt{2}}{\pi^{1/2}}\,\varepsilon\,\exp\Big(-\frac{1}{8\varepsilon^2}\Big), \qquad \varepsilon \to 0^+.$$

The list of types of Gaussian processes $(X_t)_{t\in[0,1]}$ for which constants $c_0, c_1 > 0, s \ge 0$ and $s' > 0$ exist such that

$$P(\|X_t\|_2 \le \varepsilon) \sim c_0\,\varepsilon^s\,\exp\Big(-\frac{c_1}{\varepsilon^{s'}}\Big), \qquad \varepsilon \to 0^+, \tag{3.3}$$

holds can be extended (e.g. Li and Shao (2001), Gao et al. (2003a) and Berghin et al. (2005)) and includes the integrated versions of the Brownian motion and Brownian bridge and the Ornstein-Uhlenbeck process. Dunker et al. (1998, Corollary 4.3) have shown that (3.3) holds if the eigenvalues of the covariance operator admit the polynomial form $\lambda_j = j^{-l}$ for some $l > 1$ and they computed the exact constants in (3.3) in that case.

If $(Y_t)_{t\in[0,1]}$ is another mean-zero, square-integrable Gaussian process with a continuous covariance function and eigenvalues $(\mu_j)_{j\in\mathbb{N}}$, then

$$P(\|Y_t\|_2 \le \varepsilon) \sim \Bigg(\prod_{j=1}^{\infty}\frac{\lambda_j}{\mu_j}\Bigg)^{1/2} P(\|X_t\|_2 \le \varepsilon), \qquad \varepsilon \to 0^+,$$

holds under the conditions that the eigenvalues $(\lambda_j)_j$ and $(\mu_j)_j$ are summable and that the infinite product $\prod_j \lambda_j/\mu_j$ converges (see Corollary 1 in Gao et al. (2003b)). Therefore, even if $(\mu_j)_j$ is not explicitly known one can often make a statement about the asymptotics of $P(\|Y_t\|_2 \le \varepsilon)$ by comparing it to the small deviation probabilities of the process (X_t), e.g. if the order at which the eigenvalues tend to zero is known.

If we now assume that Q is a centred Gaussian measure on the Banach space of continuous functions $\big(C_0([0,1]), \mathbb{B}(C_0([0,1]))\big)$ and $(X_t)_{t\in[0,1]}$ is the corresponding process, then deriving the exact asymptotic behaviour of $\varphi_0^Q(\varepsilon)$ is usually more complicated and there exist only few results where the explicit order is known.

(c) If $(X_t)_{t\in[0,1]}$ is the Brownian motion, then

$$P\Big(\sup_{t\in[0,1]}|X_t|\le\varepsilon\Big)\sim\frac{4}{\pi}\exp\left(-\frac{\pi^2}{8\varepsilon^2}\right),\qquad\varepsilon\to 0^+,$$

e.g. see Bogachev (2015), p. 187.

Li and Shao (2001) have verified the asymptotic relation

$$\ln P\Big(\sup_{t\in[0,1]}|X_t|\le\varepsilon\Big)\asymp-\varepsilon^{-r}\big(-\ln(\varepsilon)\big)^q,\qquad\varepsilon\to 0^+,$$

for some constants $r>0$ and $q\in\mathbb{R}$ for various types of Gaussian processes $(X_t)_{t\in[0,1]}$, including the fractional Brownian motion and its integrated version.

If Q is a centred Gaussian measure and the exact asymptotic behaviour of $\varphi_0^Q(\varepsilon)$ is known, then we can use Proposition 3.15(a) to derive the asymptotics of $\varphi_f^Q(\varepsilon)$ for any $f\in H(Q)$.

3.3 Metric entropy and small ball probabilities

The mode is defined as an element of the space $\mathcal{F}$ such that there exist points in its neighbourhood whose small ball probabilities are asymptotically larger than the small ball probabilities around other, distant centre points. This raises the question whether there is a link between the small ball probabilities of these points close to the mode and the covering numbers of the support of the distribution if the latter are assumed to be finite for any positive radius. Under certain conditions, we can give a positive answer to that question. Therefore, we will assume throughout this short section that $Q\in\mathcal{L}_{(\mathcal{F},d)}$ is supported on the totally bounded set $\mathcal{Y}\in\mathbb{B}(\mathcal{F})$, which means that $Q(\mathcal{Y})=1,\mathrm{Mod}(Q)\in\mathcal{Y}$ and $N(\mathcal{Y},d,\varepsilon)<\infty$ for every $\varepsilon>0$.

Lemma 3.18. *(a) There is some $\delta'>0$ such that, for every $\delta\in(0,\delta']$ and for every $x\in M^Q(\delta)$, we have*

$$\liminf_{\varepsilon\to0^+}\varphi_x^Q(\varepsilon)\cdot N(\mathcal{Y},d,\varepsilon)>0.$$

(b) Let $Q'\in\mathcal{L}_{(\mathcal{F},d)}$ such that $\mathrm{Mod}(Q')\in\mathcal{Y}$ and $\mathrm{Mod}(Q')\in M^{Q'}(\delta)$ for every $\delta>0$. If $Q'(\mathcal{Y})>0$, then we have

$$\liminf_{\varepsilon\to0^+}\varphi_{\mathrm{Mod}(Q')}^{Q'}(\varepsilon)\cdot N(\mathcal{Y},d,\varepsilon)>0.$$

Proof:

(a) If Q is a Dirac measure, e.g. if $Q = m_y$ for some $y \in \mathcal{Y}$, then $\{y\} = M^Q(\delta)$ for every $\delta > 0$ and the desired relation holds. In any other case, there exists some $y' \in \operatorname{supp}(Q)$ such that $\rho := d(\operatorname{Mod}(Q), y') > 0$. We can set $\delta' := \rho/4$ and fix both $\delta \in (0, \delta']$ and an arbitrary $x \in M^Q(\delta)$. For every $\varepsilon > 0$, let $C(\varepsilon) \subseteq \mathcal{F}$ denote a minimum ε-cover of $\mathcal{Y} \setminus B_d(x, 2\delta)$, which implies that for every $f \in C(\varepsilon)$ we have $d(x, f) \geq 2\delta - \varepsilon$, because otherwise there would be a contradiction to the minimality property of $C(\varepsilon)$. If ε is sufficiently small, then

$$\varphi_x^Q(\varepsilon) > \max_{f \in C(\varepsilon)} \varphi_f^Q(\varepsilon)$$

holds, which implies

$$|C(\varepsilon)| \varphi_x^Q(\varepsilon) > \sum_{f \in C(\varepsilon)} \varphi_f^Q(\varepsilon) \geq Q(\mathcal{Y} \setminus B_d(x, 2\delta)) > 0.$$

Since $N(\mathcal{Y}, d, \varepsilon) \geq |C(\varepsilon)|$ holds for every $\varepsilon > 0$, the result has been proven.

(b) Consider the probability measure

$$Q'_{\mathcal{Y}}(M) := \frac{Q'(M \cap \mathcal{Y})}{Q'(\mathcal{Y})}, \qquad \forall M \in \mathbb{B}(\mathcal{F}).$$

Since $Q'(\mathcal{Y}) > 0$, it indeed holds that $Q'_{\mathcal{Y}} \in \mathcal{W}_{(\mathcal{F},d)}$. By Lemma 3.2(b), we have $Q'_{\mathcal{Y}}\big(\operatorname{supp}(Q'_{\mathcal{Y}})\big) = 1$. Let $y \in \operatorname{supp}\big(Q'_{\mathcal{Y}}\big)$. If $Q'_{\mathcal{Y}}(\{y\}) > 0$, then $Q'(\{y\}) > 0$. Either we have $y = \operatorname{Mod}(Q')$ and the claim follows immediately or we have $y \neq \operatorname{Mod}(Q')$. In that case, since $\operatorname{Mod}(Q') \in M^{Q'}(\delta)$ for every $\delta > 0$, we can deduce that $Q'\big(\{\operatorname{Mod}(Q')\}\big) > 0$ must hold. Hence, the desired limit relation holds.
If $Q'_{\mathcal{Y}}(\{y\}) = 0$, then there must exist some $y' \in \operatorname{supp}(Q'_{\mathcal{Y}})$ such that $y \neq y'$, which means that $Q'\big(B_d(y, \varepsilon) \cap \mathcal{Y}\big) > 0$ and $Q'\big(B_d(y', \varepsilon) \cap \mathcal{Y}\big) > 0$ for every $\varepsilon > 0$. Hence, if $\delta > 0$ is sufficiently small, then we have $Q'\big(\mathcal{Y} \setminus B_d\big(\operatorname{Mod}(Q'), 2\delta\big)\big) > 0$. Fix some small $\delta > 0$, let $C(\varepsilon)$ denote a minimum cover of $\mathcal{Y} \setminus B_d\big(\operatorname{Mod}(Q'), 2\delta\big)$ for every $\varepsilon > 0$ and deduce that for sufficiently small ε one has

$$\varphi_{\operatorname{Mod}(Q')}^{Q'}(\varepsilon) > \max_{f \in C(\varepsilon)} \varphi_f^{Q'}(\varepsilon).$$

Now, the claim follows by considerations which are analogous to (a).

■

If δ is sufficiently small, then for every point $x \in M^Q(\delta)$ we can derive an asymptotic lower bound on its small ball probabilities that is a multiple of the inverse of the covering number of the support of Q. We will later see that the faster the order at which these small ball probabilities tend to zero, the slower the rate at which the risk of our mode estimator vanishes. Thus, it will prove helpful knowing that in the case of a probability distribution with a mode that is supported on a totally bounded set, these important probabilities admit a lower bound that merely depends on the metric entropy of the support.
If the mode itself is contained in every set $M^Q(\delta)$, then we can apply Lemma 3.18(a) to its small ball probabilities.

Corollary 3.19. *If* $\mathrm{Mod}(Q) \in M^Q(\delta)$ *for every* $\delta > 0$, *then there are constants* $c, \varepsilon' > 0$ *such that*

$$\varphi^Q_{\mathrm{Mod}(Q)}(\varepsilon) \geq \frac{c}{N(\mathcal{Y}, d, \varepsilon)}, \qquad \forall \varepsilon \in (0, \varepsilon').$$

Proof: This is a direct application of Lemma 3.18(a) to $x = \mathrm{Mod}(Q) \in M^Q(\delta)$. ■

We will conclude this section by applying our Lemma 3.18(a) to two cases where in the first example the support is finite-dimensional and in the second example its dimension is infinite.

Examples 3.20. (a) Assume that $(\mathcal{F}, d) = (\mathbb{R}^k, d_{Euc})$ and that $\mathcal{Y} \subset \mathbb{R}^k$ is compact. Then if $\delta' > 0$ is sufficiently small and $x \in M^Q(\delta)$, where $\delta \in (0, \delta']$, is arbitrary, we have

$$\liminf_{\varepsilon \to 0^+} \frac{\varphi^Q_x(\varepsilon)}{\varepsilon^k} > 0,$$

as follows from Lemma 2.13(b), because $\mathcal{Y}$ is contained in a closed ball with finite radius.

(b) Assume that there exists an $\alpha > 0$ such that $\limsup_{\varepsilon \to 0^+} \varepsilon^\alpha \ln N(\mathcal{Y}, d, \varepsilon) < \infty$, which holds true if $\mathcal{Y}$ is equal to any of the spaces considered in our Examples 2.15. If $\delta' > 0$ is sufficiently small and $x \in M^Q(\delta)$, where $\delta \in (0, \delta']$, is arbitrary, then there exists a constant $c > 0$ that does not depend on x such that

$$\liminf_{\varepsilon \to 0^+} \varphi^Q_x(\varepsilon) \exp\left(\frac{c}{\varepsilon^\alpha}\right) > 0.$$

In general it is not possible to derive an upper bound for the small ball probabilities based on the entropy of the support, even if the covering numbers are given explicitly. For instance, the mode could be an atom regardless of the entropy behaviour.

3.4 Consistent mode estimation

The second part of this thesis deals with the estimation of the mode and the asymptotic properties of two mode estimators. It is divided into this Section 3.4 and the following Chapter 4. We propose an estimator in each of these two parts and formalise further stipulations required for the convergence analysis.
While there are differences in these two approaches, we will assume in either scenario that the probability distribution has a mode according to our Definition 3.4(a) that is contained in some totally bounded set $\mathcal{Y} \in \mathbb{B}(\mathcal{F})$ that admits a positive probability. We will require bounds on the respective covering numbers and will use finite covers to find bounds on certain probabilities, e.g. we will use them to discretise the event that the distance between the mode and its estimator is larger than some $\varepsilon > 0$ and derive an upper bound for its probability. A challenge arises from the fact that if $C(r)$ is a (minimum) r-cover of $\mathcal{Y}$, then there may be elements in $\mathcal{Y}$ for which the distance between them and their nearest point in $C(r)$ is close to r. Thus, if we want to compare the small ball probabilities of two given points by approximation through the probabilities of some elements of $C(r)$, then we will need additional conditions for the behaviour of the small ball probability functions for small radii.
Our two settings in Section 3.4 and Chapter 4 differ in the requirements we impose on these quantities and in the definition of the estimator itself:

- The estimator in **Section 3.4** will be set equal to one data point from the sample for which, for a specified radius $r > 0$, the amount of data that fall into a ball around it with given radius r is maximised. Additionally, we will stipulate that, for certain points in a neighbourhood of the mode, the exact order at which their small ball probabilities tend to zero is known.

- In **Chapter 4** we will specifiy a finite r-cover $C'(r)$ of the set $\mathcal{Y}$. For every $x \in C'(r)$ we will count these data points that fall into the ball $B_d(x, r)$ and set the estimator equal to one point in the cover that maximises this amount. We will stipulate a lower bound on the small ball probability of the mode. Additionally, we require both upper and lower bounds for the quotient of the small ball probabilities of the mode and other distant points.

To summarise, in this section we will impose constraints on the small ball probabilities at single points in a neighbourhood of the mode and prove strong consistency. It is out of scope of this thesis to derive the minimax rate in the general setting we are working with in this section. The requirements in the following chapter are more restrictive and will additionally include a uniform bound for the ratio of certain probabilities. Therein, we will deduce the minimax rate of the mode estimation problem over a specified class of distributions.

3.4.1 A mode estimator

Let $n \in \mathbb{N}$. For the remainder of this chapter, let $X_1, \ldots, X_n$ be independent and identically distributed random variables defined on a probability space $(\Omega, \mathcal{A}, P)$ which take on values in $(\mathcal{F}, \mathbb{B}(\mathcal{F}))$. Set $P_X := P_{X_1}$ and assume that $P_X \in \mathcal{L}_{(\mathcal{F},d)}$. Our goals in this section are to find an estimator $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$ of $\operatorname{Mod}(P_X)$ and to show that the sequence $(\widehat{\theta}_n)_{n\in\mathbb{N}}$ converges to $\operatorname{Mod}(P_X)$ almost surely.

Therefore, let $n \in \mathbb{N}$ and define the random variables

$$Z_n(x, r) := \sum_{i=1}^{n} \mathbf{1}_{B_d(x,r)}(X_i)$$

for every $r > 0$ and $x \in \mathcal{F}$, which count the data points from the sample that fall into $B_d(x, r)$. We now fix some radius $r > 0$. The estimator $\widehat{\theta}_{n,r}$ will be set equal to some element of $\{X_1, \ldots, X_n\}$ for which the corresponding variable $Z_n(X_k, r)$ is maximised over all $k \in \{1, \ldots, n\}$. Note that since the data $X_1, \ldots, X_n$ take values in a separable space, the mappings $\omega \mapsto d\big(X_i(\omega), X_j(\omega)\big)$ are measurable for every $i \neq j$. To avoid problems arising with the possible occurrence of ties among the variables $Z_n(X_1, r), \ldots, Z_n(X_n, r)$, we will define independent and identically distributed random variables $U_{1,n}, \ldots, U_{n,n}$ such that $U_{1,n} \sim \mathcal{U}\big((0,1)\big)$ (which is the uniform distribution defined on the open interval $(0,1) \subset \mathbb{R}$) and also impose that the random vectors $\big(Z_n(X_1, r), \ldots, Z_n(X_n, r)\big)$ and $\big(U_{1,n}, \ldots, U_{n,n}\big)$ are stochastically independent.
Let us then set $\tilde{Z}_n(X_k, r) := Z_n(X_k, r) + U_{k,n}$ for every $k \in \{1, \ldots, n\}$ and define

$$\Big(\operatorname*{arg\,max}_{k\in\{1,\ldots,n\}} \tilde{Z}_n(X_k, r)\Big)(\omega) := \big\{k \in \{1, \ldots, n\} | \forall k' \in \{1, \ldots, n\} : \tilde{Z}_n(X_k, r) \geq \tilde{Z}_n(X_{k'}, r)\big\}$$

for every $\omega \in \Omega$. Observe that $\operatorname*{arg\,max}_{k\in\{1,\ldots,n\}} \tilde{Z}_n(X_k, r)$ is a random set, for which

$$P\Big(\Big|\operatorname*{arg\,max}_{k\in\{1,\ldots,n\}} \tilde{Z}_n(X_k, r)\Big| > 1\Big) \leq P(\exists i, j \in \{1, \ldots, n\}, i \neq j : U_{i,n} = U_{j,n}) = 0$$

and thus $P(|\operatorname{arg\,max}_{k\in\{1,\ldots,n\}} \tilde{Z}_n(X_k, r)| = 1) = 1$ holds. Hence, there is a set $A \in \mathcal{A}$ such that $P(A) = 1$ and $\big|\big(\operatorname{arg\,max}_{k\in\{1,\ldots,n\}} \tilde{Z}_n(X_k, r)\big)(\omega)\big| = 1$ for every $\omega \in A$. We can define a random variable $M_n(r)\colon(\Omega,\mathcal{A},P)\to(\{1,\dots,n\},2^{\{1,\dots,n\}})$ by

$$M_n(r)(\omega)=i \Longleftrightarrow \Big(\operatorname*{arg\,max}_{k\in\{1,\dots,n\}} \tilde{Z}_n(X_k,r)\Big)(\omega)=\{i\}, \quad \forall\omega\in A, \forall i\in\{1,\dots,n\} \tag{3.4}$$

and $M_n(r)(\omega)=1$ for every $\omega\in\Omega\setminus A$. Then

$$\widehat{\theta}_{n,r} := X_{M_n(r)}$$

is the proposed estimator we are working with in this section. The radius $r>0$ serves as a parameter and will depend on n. We will analyse the convergence properties of the sequence $(\widehat{\theta}_{n,r_n})_{n\in\mathbb{N}}$, where $(r_n)_{n\in\mathbb{N}}$ are positive real numbers that converge to zero. These radii must fulfill certain conditions, e.g. they must not converge too fast in order to receive strong asymptotic results.

3.4.2 Strong consistency

Two additional constraints are required in addition to the existence of a unique mode $\mathrm{Mod}(P_X)$, which is guaranteed by our assumption $P_X\in\mathcal{L}_{(\mathcal{F},d)}$. The first one is the following entropy inequality:

(A1) If $\mathcal{Y}:=\mathrm{supp}(P_X)\subseteq\mathcal{F}$, then there exist constants $\alpha>0, K>0$ and $\varepsilon'>0$ such that

$$\ln N(\mathcal{Y},d,\varepsilon)\leq\frac{K}{\varepsilon^\alpha}$$

holds for every $\varepsilon\in(0,\varepsilon']$.

Now let $c_0,c_1>0$, $s\geq0$ and $t>0$ be constants and define

$$g_{c_0,c_1,s,t}(\varepsilon)\colon(0,\infty)\longrightarrow(0,\infty), \quad \varepsilon\longmapsto c_0\,\varepsilon^s\,\exp\Big(-\frac{c_1}{\varepsilon^t}\Big).$$

As previously stated, a lower bound on certain small ball probability functions can be derived from Lemma 3.18(a) in terms of covering numbers. Hence, when we impose constraints on the order of small ball functions of elements taken from $M^{P_X}(\delta)$, then we must make sure that there is no contradiction to Lemma 3.18(a).

Lemma 3.21. *Let $c_0, c_1 > 0$, $s \geq 0$ and $t > 0$. Then the relation*

$$\liminf_{\varepsilon\to 0^+} g_{c_0,c_1,s,t}(\varepsilon)\ \exp(K\varepsilon^{-\alpha}) > 0$$

holds if and only if the constants c_1, s and t fulfill one of the following three conditions:

(i) $t < \alpha$,

(ii) $t = \alpha, c_1 < K$

(iii) $t = \alpha, c_1 = K, s = 0$

Proof: First realise that if $s > 0$, then for every $t^* \in \mathbb{R}$ and $c' > 0$ we have

$$\varepsilon^s \exp(-c'\varepsilon^{-t^*}) \overset{\varepsilon\to 0^+}{\longrightarrow} 0,$$

because $\ln\left(\varepsilon^s \exp(-c'\varepsilon^{-t^*})\right) = s\ln(\varepsilon) - c'/\varepsilon^{t^*} \overset{\varepsilon\to 0^+}{\longrightarrow} -\infty$. If $s = 0$ and $t^* > 0$, then the limit is also zero. We will now consider several cases. If $t > \alpha$, then we have $K/\varepsilon^\alpha - c_1/\varepsilon^t = \varepsilon^{-\alpha}(K - c_1/\varepsilon^{t-\alpha}) < -1/\varepsilon^\alpha$ for sufficiently small ε. This yields

$$\varepsilon^s \exp(-c_1/\varepsilon^t)\exp(K/\varepsilon^\alpha) < \varepsilon^s \exp(-\varepsilon^{-\alpha}) \overset{\varepsilon\to 0^+}{\longrightarrow} 0.$$

Hence, the limit relation in the lemma does not hold. If $t < \alpha$, then $K/\varepsilon^\alpha - c_1/\varepsilon^t = \varepsilon^{-t}(K/\varepsilon^{\alpha-t} - c_1) > \varepsilon^{-t}$ for small ε, which yields

$$\varepsilon^s \exp(-c_1/\varepsilon^t)\exp(K/\varepsilon^\alpha) > \varepsilon^s \exp(\varepsilon^{-t}) \overset{\varepsilon\to 0^+}{\longrightarrow} \infty$$

and, hence, the claim is verified if condition (i) holds. The last limit follows from $\ln(\varepsilon)\varepsilon^t \overset{\varepsilon\to 0^+}{\longrightarrow} 0$ (e.g. by l'Hopital's rule) and $\ln(\varepsilon)\varepsilon^t < 0$ for small ε, which yields $\ln(\varepsilon^s \exp(\varepsilon^{-t})) = \ln(\varepsilon)(s + (\ln(\varepsilon)\varepsilon^t)^{-1}) \overset{\varepsilon\to 0^+}{\longrightarrow} \infty$. Now assume that $t = \alpha$, then

$$\varepsilon^s \exp(-c_1/\varepsilon^t)\exp(K/\varepsilon^\alpha) = \varepsilon^s \exp\left(\frac{K - c_1}{\varepsilon^\alpha}\right).$$

The case of $c_1 < K$ can be treated analogously to the limit we have just considered. Hence, the function also tends to $+\infty$ and the claim holds if (ii) is fulfilled. If $c_1 > K$, then the limit is zero (see the first limit considered). Finally, if $c_1 = K$, then the desired relation obviously holds if and only if $s = 0$, which is condition (iii). There are no more cases to consider. ∎

Let $\mathcal{I}$ denote the set of all triplets $(c_1, s, t) \in \mathbb{R}^3$ such that $c_1, t > 0, s \geq 0$ and for which one of the conditions (i), (ii) or (iii) introduced in Lemma 3.21 holds.

(A2) There is $\delta' > 0$ such that for every $\delta \in (0, \delta']$ there exists $x(\delta) \in M^{P_X}(\delta)$ such that

$$\varphi_{x(\delta)}^{P_X}(\varepsilon) \sim g_{c_0,c_1,s,t}(\varepsilon), \quad \varepsilon \to 0^+,$$

where $c_0 > 0$ and $(c_1, s, t) \in \mathcal{I}$ depend on $x(\delta)$.

If the constants in (A2) fulfill the imposed conditions, then by Lemma 3.21 we have

$$\liminf_{\varepsilon \to 0^+} \varphi_{x(\delta)}^{P_X}(\varepsilon) \exp\left(K\varepsilon^{-\alpha}\right) = \liminf_{\varepsilon \to 0^+} \frac{\varphi_{x(\delta)}^{P_X}(\varepsilon)}{g_{c_0,c_1,s,t}(\varepsilon)} \, g_{c_0,c_1,s,t}(\varepsilon) \, \exp\left(K\varepsilon^{-\alpha}\right) > 0.$$

Recall that if δ is sufficiently small, then Lemma 3.18(a) provides a lower bound for $\varphi_{x(\delta)}^{P_X}$ in terms of the covering number of the support of P_X. Since (A1) holds, the lower bound given for $\varphi_{x(\delta)}^{P_X}$ in (A2) is not asymptotically smaller than the lower bound that Lemma 3.18(a) yields.

The first condition (A1) is an upper bound on the entropy of the support of P_X. Again we refer to our Examples 2.15 for a list of spaces for which (A1) can be justified. It coincides with the upper bound in constraint (2.3) imposed by Meister (2016), who also uses entropy and covering arguments to derive asymptotic properties of estimators.
Our second constraint (A2) deals with the small ball probabilities of certain points taken from $M^{P_X}(\delta)$, which is a set that is always contained in a neighbourhood of the mode and may even contain the mode itself. Recall that $\mathrm{Mod}(P_X) \in \overline{M^{P_X}(\delta)}$ for every $\delta > 0$. The order we propose is exponential with an additional rational factor ε^s, that may also be equal to 1, because $s = 0$ is allowed.
Small ball probability functions with an exponential behaviour for small radii are not uncommon for distributions defined on (infinite-dimensional) function spaces. We have seen in Section 3.2.2 (see Examples 3.17) that the small ball probabilities of certain Gaussian measures admit such a rate. Ferraty and Vieu (2006) also give asymptotic results for their functional kernel estimators assuming the small ball probabilities belong to an exponential type (see Definition 13.4 and Proposition 13.5 therein).

The following two lemmata contain useful inequalities, which we will later exploit to derive an upper bound for a probability of a set that involves the variables $Z_n(\cdot, \cdot)$. Therefore, let $\mathrm{Bin}(n, p)$ denote the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$.

Lemma 3.22. *If* $Y \sim Bin(n,p)$, *where* $p \in (0, \frac{1}{2})$, *then*

$$P\Big(Y \geq \frac{n}{2}\Big) \leq \Big(1 - \big(\sqrt{p} - \sqrt{1-p}\big)^2\Big)^n.$$

Proof: This follows from Theorem 1 in Hoeffding (1963). ■

Lemma 3.23. *Let* $A, B \in \mathbb{B}(\mathcal{F})$ *be disjoint, set* $p_1 := P_X(A)$ *and* $p_2 := P_X(B)$ *and assume that* $p_1 > p_2$. *Further define the random variables*

$$S_n^{(1)} := \sum_{i=1}^n \boldsymbol{1}_A(X_i) \quad \text{and} \quad S_n^{(2)} := \sum_{i=1}^n \boldsymbol{1}_B(X_i).$$

Then we have

$$P(S_n^{(2)} \geq S_n^{(1)}) \leq \left(1 - \left(\sqrt{p_1} - \sqrt{p_2}\right)^2\right)^n.$$

Proof:

In order to show the desired inequality we need to prove the following auxiliary statement:

$$(\bullet) \qquad S_n^{(2)}|_{S_n^{(1)}+S_n^{(2)}=k} \sim \text{Bin}\left(k, \frac{p_2}{p_1+p_2}\right), \quad \forall k \in \{1, \dots, n\}$$

We consider that, since A and B are disjoint, we have $S_n^{(1)} + S_n^{(2)} \sim \text{Bin}\big(n, p_1 + p_2\big)$. Let $k \in \{1, \dots, n\}, j \in \{0, \dots, k\}$, set $S_n^{(3)} = n - S_n^{(1)} - S_n^{(2)}$, consider that the vector $(S_n^{(1)}, S_n^{(2)}, S_n^{(3)})$ has a multinomial distribution and deduce

$$\begin{aligned}P(S_n^{(2)} = j | S_n^{(1)} + S_n^{(2)} = k) &= \frac{P(S_n^{(1)} = k-j, S_n^{(2)} = j, S_n^{(3)} = n-k)}{P(S_n^{(1)} + S_n^{(2)} = k)} \\ = \frac{\binom{n}{k}\binom{k}{j} p_1^{k-j} p_2^j (1-p_1-p_2)^{n-k}}{\binom{n}{k}(p_1+p_2)^k (1-p_1-p_2)^{n-k}} &= \binom{k}{j}\Big(\frac{p_1}{p_1+p_2}\Big)^{k-j}\Big(\frac{p_2}{p_1+p_2}\Big)^j.\end{aligned}$$

It follows from $p_1 > 0$ that $P(S_n^{(1)} + S_n^{(2)} = k) > 0$ for every $k \in \{1, \dots, n\}$, which concludes the proof of $(\bullet)$.

We can now show the claim. Let $Z, Z_1, \dots, Z_n$ denote random variables such that $Z \sim \text{Bin}(n, p_1+p_2)$ and $Z_k \sim \text{Bin}\Big(k, \frac{p_2}{p_1+p_2}\Big)$ for every $k \in \{1, \dots, n\}$. We have

$$P(S_n^{(2)} \geq S_n^{(1)} | S_n^{(1)} + S_n^{(2)} = 0) = 1$$

and deduce

$$
\begin{aligned}
P(S_n^{(2)} \geq S_n^{(1)}) &= \sum_{k=0}^{n} P(S_n^{(2)} \geq S_n^{(1)} | S_n^{(1)} + S_n^{(2)} = k) P(S_n^{(1)} + S_n^{(2)} = k) \\
&= \sum_{k=0}^{n} P(S_n^{(2)} \geq k/2 | S_n^{(1)} + S_n^{(2)} = k) P(S_n^{(1)} + S_n^{(2)} = k) \\
&= \sum_{k=0}^{n} P\Big(Z_k \geq k/2\Big) P(Z = k).
\end{aligned}
$$

We have used $S_n^{(1)} + S_n^{(2)} \sim \mathrm{Bin}\big(n, p_1 + p_2\big)$ and $(\bullet)$ in the last step. Since $p_1 > p_2$ (which implies that $\frac{p_2}{p_1+p_2} < \frac{1}{2}$), we can now use Lemma 3.22 to derive an upper bound of the probability $P\big(Z_k \geq k/2\big)$. By using the binomial theorem, we conclude that

$$
\begin{aligned}
&P(S_n^{(2)} \geq S_n^{(1)}) \\
&\leq \sum_{k=0}^{n} \left(1 - \left(\sqrt{\frac{p_1}{p_1 + p_2}} - \sqrt{\frac{p_2}{p_1 + p_2}}\right)^2\right)^k (p_1 + p_2)^k (1 - p_1 - p_2)^{n-k} \binom{n}{k} \\
&= \left(1 - (p_1 + p_2)\left(\frac{\sqrt{p_1} - \sqrt{p_2}}{\sqrt{p_1 + p_2}}\right)^2\right)^n = \Big(1 - \big(\sqrt{p_1} - \sqrt{p_2}\big)^2\Big)^n.
\end{aligned}
$$

■

We are now in the position to prove strong consistency for the mode estimator introduced at the beginning of this section. Another short lemma will be used in our main theorem.

Lemma 3.24. *Let $c_0, c_1 > 0$, $s \geq 0$ and $t > 0$ be constants. Let $h\colon (0,\infty) \to (0,\infty)$ be a positive functions that fulfills both $h(x) < x$ for every $x > 0$ and $\lim_{x \to 0^+} h(x)x^{-(1+t)} = 0$. Then we have*

$$
g_{c_0,c_1,s,t}(x - h(x)) \sim g_{c_0,c_1,s,t}(x + h(x)), \quad x \to 0^+.
$$

Proof: It suffices to prove

$$
\frac{g_{c_0,c_1,s,t}(x \pm h(x))}{g_{c_0,c_1,s,t}(x)} \sim 1, \qquad x \to 0^+.
$$

We observe that

$$
\frac{g_{c_0,c_1,s,t}(x \pm h(x))}{g_{c_0,c_1,s,t}(x)} = \left(\frac{x \pm h(x)}{x}\right)^s \exp\Big(c_1\big(x^{-t} - (x \pm h(x))^{-t}\big)\Big),
$$

which converges to 1 as $x \to 0^+$ if

$$\lim_{x\to 0^+} \frac{1}{x^t} - \frac{1}{(x \pm h(x))^t} = 0$$

holds, since the rational factor clearly converges to 1 because $\lim_{x\to 0^+} h(x)x^{-1} = 0$. We can consider

$$\frac{1}{x^t} - \frac{1}{(x+h(x))^t} = \frac{\left(1+\frac{h(x)}{x}\right)^t - 1}{(x+h(x))^t} \leq \frac{\left(1+\frac{h(x)}{x}\right)^{\lceil t\rceil} - 1}{(x+h(x))^t} = \frac{\sum_{k=0}^{\lceil t\rceil}\binom{\lceil t\rceil}{k}\left(\frac{h(x)}{x}\right)^k - 1}{(x+h(x))^t}$$
$$= \frac{\sum_{k=1}^{\lceil t\rceil}\binom{\lceil t\rceil}{k}\left(\frac{h(x)}{x^{t/k+1}}\right)^k}{(1+\frac{h(x)}{x})^t} = \frac{\sum_{k=1}^{\lceil t\rceil}\binom{\lceil t\rceil}{k}\left(x^{t(1-\frac{1}{k})}\frac{h(x)}{x^{t+1}}\right)^k}{(1+\frac{h(x)}{x})^t} \xrightarrow{x\to 0^+} 0,$$

because $t(1-\frac{1}{k}) \geq 0$ for every $k = 1, ..., \lceil t\rceil$. Recall that $\lceil t\rceil = \min\{k \in \mathbb{N} | k \geq t\}$. The second limit (where '+' is changed to '-') can be dealt with analogously.

■

Theorem 3.25. *We assume that P_X fulfills both (A1) and (A2). Let $(r_n)_{n\in\mathbb{N}}$ be a sequence of positive real numbers that converges to zero and satisfies*

$$\lim_{n\to\infty} r_n \ln(n)^{\frac{1}{\alpha(\alpha+1)}} = \infty,$$

where $\alpha > 0$ is the parameter taken from (A1). Then we have

$$d(\widehat{\theta}_{n,r_n}, \mathrm{Mod}(P_X)) \xrightarrow{n\to\infty} 0$$

almost surely.

Proof:
The proof is organised as follows: We show that, for every $\varepsilon > 0$ there exists $n(\varepsilon) \in \mathbb{N}$ such that for every $n \geq n(\varepsilon)$ we have $P(d(\mathrm{Mod}(P_X), \widehat{\theta}_{n,r_n}) > \varepsilon) \leq q(n, \varepsilon)$, where $q(n, \varepsilon) > 0$ is some upper bound depending on n and ε. If $\sum_n q(n, \varepsilon) < \infty$ holds for every $\varepsilon > 0$, then almost sure convergence follows from the Borel-Cantelli lemma.
We set $\mathcal{Y} := \mathrm{supp}(P_X) \subset \mathcal{F}$ and fix some $\varepsilon > 0$. Using (A2), if $\delta > 0$ is sufficiently small, then there exists $x(\delta) \in M^{P_X}(\delta)$ such that the following property holds:

$$\exists c_0 > 0, (c_1, s, t) \in \mathcal{I} : \varphi^{P_X}_{x(\delta)}(r) \sim g_{c_0,c_1,s,t}(r), \qquad r \to 0^+ \tag{3.5}$$

Recall that $\mathcal{I}$ is the set of all triplets that fulfill one of the conditions imposed in Lemma

3.21. We can assume that $\delta \leq \varepsilon/4$, which implies $d\big(\operatorname{Mod}(P_X), x(\delta)\big) \leq \varepsilon/4$. This yields

$$\begin{aligned} P\big(d\big(\operatorname{Mod}(P_X), \widehat{\theta}_{n,r_n}\big) > \varepsilon\big) &\leq P\big(d\big(\operatorname{Mod}(P_X), x(\delta)\big) + d\big(x(\delta), \widehat{\theta}_{n,r_n}\big) > \varepsilon\big) \\ &\leq P\big(d\big(x(\delta), \widehat{\theta}_{n,r_n}\big) > \varepsilon/2\big). \end{aligned}$$

Set $h_n := \Big(\dfrac{2K}{\ln(n)}\Big)^{1/\alpha}$, where $K > 0$ is the constant from (A1) and define the sets

$$A_n := \big\{d\big(x(\delta), \widehat{\theta}_{n,r_n}\big) > \varepsilon/2\big\} \text{ and } B_n := \big\{\exists i \in \{1, \dots, n\} : d\big(X_i, x(\delta)\big) \leq h_n\big\}$$

and we can conclude that

$$P\big(d\big(\operatorname{Mod}(P_X), \widehat{\theta}_{n,r_n}\big) > \varepsilon\big) \leq P(A_n) \leq P(A_n \cap B_n) + P(\Omega \setminus B_n).$$

In the following it is our goal to derive upper bounds for both $P(A_n \cap B_n)$ and $P(\Omega \setminus B_n)$ and we will start with the latter. Since $h_n \overset{n\to\infty}{\longrightarrow} 0$, we can use (3.5) and Lemma 3.21 to deduce that there exists a constant $c' > 0$ such that

$$\varphi_{x(\delta)}^{P_X}(h_n) \geq c' \exp\Big(-\frac{K}{h_n^\alpha}\Big) \tag{3.6}$$

holds for every sufficiently large n. Hence, we can deduce

$$\begin{aligned} P(\Omega \setminus B_n) &= P\Big(\forall j \in \{1, \dots, n\} : d(X_j, x(\delta)) > h_n\Big) = \Big(1 - \varphi_{x(\delta)}^{P_X}(h_n)\Big)^n \\ &\leq \exp\big(-n\varphi_{x(\delta)}^{P_X}(h_n)\big) \leq \exp\Big(-c'\, n \exp\big(-K h_n^{-\alpha}\big)\Big) \\ &= \exp\Bigg(-c'\, n \exp\Bigg(-\frac{1}{2}\ln(n)\Bigg)\Bigg) = \exp\Big(-c'\sqrt{n}\Big). \end{aligned}$$

The second equation is due to the fact that the variables $X_1, \dots, X_n$ are independent and identically distributed. The reasoning for the first inequality goes as follows: One can verify that $\ln(1-z) \leq -z$ holds for every $z < 1$. Multiplying each side by n and applying 'exp' yields $(1-z)^n \leq \exp(-nz)$, which also holds for $z = 1$. Finally, we have used (3.6) in the second inequality.
In order to find a suitable upper bound for $P(A_n \cap B_n)$, we will apply Lemma 3.23. Therefore, it is necessary to prove the following inequality: There is $m(\varepsilon) \in \mathbb{N}$ such that for every $n \geq m(\varepsilon)$ and every $y \in \mathcal{F}$ such that $d(x(\delta), y) \geq \delta$ we have

$$\frac{\varphi_{x(\delta)}^{P_X}(r_n - h_n)}{\varphi_y^{P_X}(r_n + h_n)} \geq 1 + c, \tag{3.7}$$

where $c > 0$ is a constant depending on ε.
Firstly, we have

$$\frac{\varphi_{x(\delta)}^{P_X}(r_n - h_n)}{\varphi_y^{P_X}(r_n + h_n)} = \frac{\varphi_{x(\delta)}^{P_X}(r_n + h_n)}{\varphi_y^{P_X}(r_n + h_n)} \cdot \frac{\varphi_{x(\delta)}^{P_X}(r_n - h_n)}{\varphi_{x(\delta)}^{P_X}(r_n + h_n)} =: T_n^{(1)} \cdot T_n^{(2)}.$$

Since $x(\delta) \in M^{P_X}(\delta)$ and $d\big(x(\delta), y\big) \geq \delta$, we can deduce that

$$T_n^{(1)} = \frac{\varphi_{x(\delta)}^{P_X}(r_n + h_n)}{\varphi_y^{P_X}(r_n + h_n)} \geq 1 + c^*$$

holds for every sufficiently large n and a constant $c^* > 0$. Both the minimum $n \in \mathbb{N}$ such that the inequality above holds and $c^* > 0$ depend on $x(\delta)$ and thus depend on ε, but they are both independent of y, because the supremum is taken in the definition of the set $M^{P_X}(\delta)$. Let us now realise that, since $t \leq \alpha$ (see Lemma 3.21), we have $\ln(n)^{\frac{1}{\alpha(1+t)}} \geq \ln(n)^{\frac{1}{\alpha(1+\alpha)}}$ and can deduce that

$$\lim_{n\to\infty} \frac{h_n}{r_n^{1+t}} = \lim_{n\to\infty} \frac{(2K)^{1/\alpha}}{r_n^{1+t}\ln(n)^{1/\alpha}} = \lim_{n\to\infty} \frac{(2K)^{1/\alpha}}{(r_n \ln(n)^{\frac{1}{\alpha(1+t)}})^{1+t}} \leq \lim_{n\to\infty} \frac{(2K)^{1/\alpha}}{(r_n \ln(n)^{\frac{1}{\alpha(1+\alpha)}})^{1+t}} = 0$$

Hence, we can use Lemma 3.24 and (3.5) to derive

$$\lim_{n\to\infty} T_n^{(2)} = \lim_{n\to\infty} \frac{\varphi_{x(\delta)}^{P_X}(r_n - h_n)}{\varphi_{x(\delta)}^{P_X}(r_n + h_n)} = \lim_{n\to\infty} \frac{g_{c_0,c_1,s,t}(r_n - h_n)}{g_{c_0,c_1,s,t}(r_n + h_n)} = 1.$$

It follows that if n is sufficiently large, then

$$\frac{\varphi_{x(\delta)}^{P_X}(r_n - h_n)}{\varphi_y^{P_X}(r_n + h_n)} \geq 1 + \frac{c^*}{2}$$

and (3.7) is shown. The bound given in (3.7) holds uniformly over all $y \in \mathcal{F}$ such that $d(x(\delta, y)) \geq \delta$, e.g. c is independent of y.
Now let $C(h_n) \subseteq \mathcal{F}$ denote a minimum h_n-cover of the set $\mathcal{Y} \setminus B_d(x(\delta), \varepsilon/2)$, which we will use to derive

$$\begin{aligned} P(A_n \cap B_n) &\leq P\Big(\exists j \in \{1, \ldots, n\} \colon \Big(d(X_j, x(\delta)) > \frac{\varepsilon}{2} \wedge Z_n\big(X_j, r_n\big) \geq Z_n\big(x(\delta), r_n - h_n\big)\Big)\Big) \\ &\leq P\left(\bigcup_{f \in C(h_n)} \Big\{Z_n(f, r_n + h_n) \geq Z_n\big(x(\delta), r_n - h_n\big)\Big\}\right) \end{aligned}$$

$$\le \sum_{f\in C(h_n)} P\Big(Z_n(f, r_n+h_n) \ge Z_n\big(x(\delta), r_n-h_n\big)\Big)$$
$$\le |C(h_n)| \cdot \sup_{f\in C(h_n)} P\Big(Z_n(f, r_n+h_n) \ge Z_n\big(x(\delta), r_n-h_n\big)\Big).$$

The first inequality holds due to the fact that if $d(x(\delta), \widehat{\theta}_{n,r_n}) > \varepsilon/2$, then there exists some $j \in \{1, \dots, n\}$ such that both $d\big(X_j, x(\delta)\big) > \varepsilon/2$ and $Z_n(X_j, r_n) \ge Z_n(X_{j'}, r_n)$ hold for every $j' \in \{1, \dots, n\}$. On the other hand, since $r_n > h_n$ for large n, if there is some $i \in \{1, \dots, n\}$ such that $d(x(\delta), X_i) \le h_n$, then an application of the triangle inequality yields

$$B_d\big(x(\delta), r_n - h_n\big) \subseteq B_d(X_i, r_n),$$

which implies

$$Z_n\big(x(\delta), r_n - h_n\big) \le Z_n(X_i, r_n).$$

For large n it holds that $i \neq j$.
The second inequality can be justified as follows: For every $y \in \mathcal{Y} \setminus B_d\big(x(\delta), \varepsilon/2\big)$ there exists some $f' \in C(h_n)$ such that $d(y, f') \le h_n$. We can again use the triangle inequality to deduce that if $d\big(X_j, x(\delta)\big) > \varepsilon/2$, then there exists $f' \in C(h_n)$ such that

$$Z_n(f', r_n + h_n) \ge Z_n(X_j, r_n).$$

Since $C(h_n)$ is a minimum h_n-cover of $\mathcal{Y} \setminus B_d\big(x(\delta), \varepsilon/2\big)$, we have for every $f \in C(h_n)$ that $d\big(x(\delta), f\big) \ge \varepsilon/2 - h_n$, because if $d\big(x(\delta), f\big) < \varepsilon/2 - h_n$ was true, then

$$B_d(f, h_n) \cap \Big(\mathcal{Y} \setminus B_d\big(x(\delta), \varepsilon/2\big)\Big) = \emptyset$$

and f could be erased from $C(h_n)$ without losing the covering property. Hence, $C(h_n)$ would not be a minimum h_n-cover. That is why if n is sufficiently large, then every $f \in C(h_n)$ fulfills $d\big(f, x(\delta)\big) \ge \varepsilon/4 \ge \delta$. Due to (3.7), if n is sufficiently large, then all the requirements to apply Lemma 3.23 to $A := B_d\big(x(\delta), r_n - h_n\big)$ and $B := B_d(f, r_n + h_n)$, where $f \in C(h_n)$, are fulfilled.
Thus, we can derive an upper bound on $P\big(Z_n(f, r_n + h_n) \ge Z_n\big(x(\delta), r_n - h_n\big)\big)$. We receive

$$\sup_{f\in C(h_n)} P\Big(Z_n(f, r_n+h_n) \ge Z_n(x(\delta), r_n-h_n)\Big)$$
$$\le \sup_{f\in C(h_n)} \left(1 - \left(\sqrt{\varphi^{P_X}_{x(\delta)}(r_n-h_n)} - \sqrt{\varphi^{P_X}_{f}(r_n+h_n)}\right)^2\right)^n$$

$$
\begin{aligned}
&\le \left(1 - \left(\sqrt{\varphi_{x(\delta)}^{P_X}(r_n - h_n)} - \sqrt{\frac{\varphi_{x(\delta)}^{P_X}(r_n - h_n)}{1+c}}\right)^2\right)^n \\
&\le \Big(1 - \tilde{c}\varphi_{x(\delta)}^{P_X}(r_n)\Big)^n \le \exp\Big(-\tilde{c}\cdot n\cdot \varphi_{x(\delta)}^{P_X}(r_n)\Big).
\end{aligned}
$$

The first inequality uses Lemma 3.23 and the second one uses (3.7). The third one holds due to the fact that $\varphi_{x(\delta)}^{P_X}(r_n - h_n) \sim \varphi_{x(\delta)}^{P_X}(r_n)$ as $n \to \infty$. Hence, there exists a positive constant $\tilde{c} < (1 - \sqrt{1/(1+c)})^2$ such that

$$
\Big(1 - \sqrt{\frac{1}{1+c}}\Big)^2 \varphi_{x(\delta)}^{P_X}(r_n - h_n) \ge \tilde{c}\varphi_{x(\delta)}^{P_X}(r_n)
$$

holds for every sufficiently large n. Before we will combine our previous efforts we want to state that, since $r_n \ln(n)^{\frac{1}{\alpha(1+\alpha)}} \ge (2K)^{1/\alpha}$ holds for large n, we can conclude that

$$
\exp\big(-K(r_n)^{-\alpha}\big) \ge \exp\Big(-\frac{1}{2}\ln(n)^{\frac{1}{1+\alpha}}\Big) \ge \exp\Big(-\frac{1}{2}\ln(n)\Big) = \frac{1}{\sqrt{n}}.
$$

Finally, we can summarise that

$$
\begin{aligned}
P(A_n \cap B_n) &\le |C(h_n)| \cdot \sup_{f \in C(h_n)} P\Big(Z_n(f, r_n + h_n) \ge Z_n(x(\delta), r_n - h_n)\Big) \\
&\le N(\mathcal{Y}, d, h_n) \exp\Big(-\tilde{c}\cdot n \cdot \varphi_{x(\delta)}^{P_X}(r_n)\Big) \\
&\le \exp\Big(\frac{K}{h_n^\alpha}\Big) \exp\Big(-\tilde{c}\cdot c' \cdot n \cdot \exp\Big(-\frac{K}{r_n^\alpha}\Big)\Big) \\
&\le \sqrt{n} \exp\Big(-\tilde{c}\cdot c' \cdot \sqrt{n}\Big).
\end{aligned}
$$

We have used in the second inequality that (3.6) also holds if h_n is replaced by r_n. Additionally, (A1) is used in that step. Thus, we have shown that for every $\varepsilon > 0$ there exists $n(\varepsilon) \in \mathbb{N}$ such that for every $n \ge n(\varepsilon)$ we have

$$
\begin{aligned}
P\Big(d\big(\widehat{\theta}_{n,r_n}, \mathrm{Mod}(P_X)\big) > \varepsilon\Big) &\le P(A_n \cap B_n) + P(\Omega \setminus B_n) \\
&\le \sqrt{n}\exp\Big(-\tilde{c}\cdot c' \cdot \sqrt{n}\Big) + \exp\Big(-c'\sqrt{n}\Big) \\
&\le 2\sqrt{n}\exp\big(-\tilde{\tilde{c}}\sqrt{n}\big),
\end{aligned}
$$

where $\tilde{\tilde{c}} > 0$ is a constant. The almost sure convergence of our estimator towards the mode follows from

$$\sum_{n=1}^{\infty} P\big(d(\widehat{\theta}_{n,r_n}, \mathrm{Mod}(P_X)) > \varepsilon\big)$$
$$\leq \sum_{n=1}^{n(\varepsilon)-1} P\big(d(\widehat{\theta}_{n,r_n}, \mathrm{Mod}(P_X)) > \varepsilon\big) + \sum_{n=n(\varepsilon)}^{\infty} \frac{2\sqrt{n}}{\exp(\tilde{\tilde{c}}\sqrt{n})} < \infty,$$

which is true for every $\varepsilon > 0$.

■

Theorem 3.25 is equivalent to the strong consistency of $\big(\widehat{\theta}_{n,r_n}\big)_{n\in\mathbb{N}}$ for $\mathrm{Mod}(P_X)$, if the rate at which the radii $(r_n)_n$ converge to zero is slower than some logarithmic term whose exponent depends on the parameter $\alpha > 0$. It can be concluded from the final part of the preceding proof that we can even achieve almost complete consistency.
Ferraty and Vieu (2006) prove almost complete consistency for their functional kernel mode estimator by assuming that the mode of the functional random variable Y lies in some set $\mathcal{C}$ that is contained within a finite union of certain balls. They further assume that there exists a density function f of P_Y with respect to some measure μ and a function ψ such that

$$\lim_{\varepsilon\to 0^+} \sup_{x\in\mathcal{C}} \left| \frac{\varphi_x^{P_Y}(\varepsilon)}{\psi(\varepsilon)} - f(x) \right| = 0, \tag{3.8}$$

where ψ is independent of x and fulfills additional regularity conditions. Due to the uniformity aspect and the fact that ψ is independent of x, (3.8) can be more difficult to verify than our assumption (A2), which we impose on the small ball functions of certain centre points in a neighbourhood of the mode. Furthermore, the existence of a function ψ and a μ-density f for which (3.8) holds is not guaranteed (see Remark 3.8 for conditions under which a generalised version of the Lebesgue differentiation theorem holds). The condition (H5) imposed in the work of Dabo-Niang et al. (2010), where another consistency result for a mode estimator is given, is identical to (3.8).
Delaigle and Hall (2010) consider a functional random variable that takes on values in the space of square-integrable functions. Hence, the small ball probability problem becomes equivalent to describing the distribution of an infinite sum of certain random variables (see Examples 3.17, where this analogy is formalised for a Gaussian measure). Then, they give approximation results for the small ball probabilities and use them to develop the notion of a density for functional data based on principal component decomposition.

We presented a consistency result on an arbitrary Polish metric space that requires an entropy bound for the support of the measure as well as exponentially decreasing small ball probability functions at certain points in a neighbourhood of the mode. If the mode is contained in $M^{P_X}(\delta)$ for every $\delta > 0$, then it suffices to assume that its small ball function admits an exponential order, e.g. $\varphi^{P_X}_{\mathrm{Mod}(P_X)}(\varepsilon) \sim g_{c_0,c_1,s,t}(\varepsilon)$ as $\varepsilon \to 0^+$, where the constants fulfill the usual conditions (see Lemma 3.21). This would already imply that (A2) holds.

We want to conclude this section with a corollary to our Theorem 3.25, for which we substitute our assumptions (A1) and (A2) for corresponding versions that are suitable for the setting where the space $(\mathcal{F}, d)$ is equal to the k-dimensional Euclidean space $(\mathbb{R}^k, d_{Euc})$, which we will assume for the remainder of this section. We will again prove consistency, but the technicalities in the proof can be carried out analogously to Theorem 3.25.

(a1) If $\mathcal{Y} := \mathrm{supp}(P_X) \subset \mathcal{F}$, then there exist constants $K > 0$ and $\varepsilon' > 0$ such that

$$N(\mathcal{Y}, d_{Euc}, \varepsilon) \leq \frac{K}{\varepsilon^k}$$

holds for every $\varepsilon \in (0, \varepsilon']$.

(a2) There exists some $\delta' > 0$ such that, for every $\delta \in (0, \delta']$, there exists some $x(\delta) \in M^{P_X}(\delta)$ and a constant $c_0 > 0$ depending on $x(\delta)$ such that

$$\varphi^{P_X}_{x(\delta)}(\varepsilon) \sim c_0 \varepsilon^k, \qquad \varepsilon \to 0^+.$$

By Lemma 2.13(b), if $\mathcal{Y}$ is the subset of a Euclidean ball with finite radius, then (a1) is fulfilled. Through Lemma 3.18(a) and condition (a1), we realise that $\varphi^{P_X}_{x(\delta)}(\varepsilon)$ is bounded from below by a multiple of ε^k and our condition (a2) additionally requires an upper bound of the same order. If P_X is Lebesgue-continuous with a density function f that is continuous at every point in a neighbourhood of the mode, then (a2) can be justified by the means of Theorem 3.7(b). Continuity of f at the mode is not a necessary condition to verify (a2). If, for instance, for every $\delta > 0$ there exists $x(\delta) \in M^{P_X}(\delta)$ such that $x(\delta) \neq \mathrm{Mod}(P_X)$ and f is continuous at $x(\delta)$, then (a2) holds.

Corollary 3.26. *We assume that $P_X \in \mathcal{L}_{(\mathbb{R}^k, d_{Euc})}$ fulfills both (a1) and (a2). Let $(r_n)_{n\in\mathbb{N}}$ be a sequence of positive real numbers that converges to zero and satisfies*

$$\lim_{n\to\infty} r_n \Big(\frac{n}{\ln(n)}\Big)^{1/k} = \infty.$$

Then we have

$$d(\widehat{\theta}_{n,r_n}, \mathrm{Mod}(P_X)) \xrightarrow{n\to\infty} 0$$

almost surely.

Proof: Proving this claim will be analogous to proving Theorem 3.25. We intend to avoid unnecessary repetition and therefore rely on the notation and definitions established therein if not explicitly specified otherwise in this proof. For brevity, we will only elaborate on the parts of this proof that differ from the one given for the theorem.

By (a2), we can assume that $\varphi^{P_X}_{x(\delta)}(r) \sim c_0 r^k, r \to 0^+$, where $c_0 > 0$ is a constant depending on $x(\delta)$. Set $h_n := \big(\frac{4}{c_0}\frac{\ln(n)}{n}\big)^{1/k}$. If r is sufficiently small, then we have

$$\varphi^{P_X}_{x(\delta)}(r) \geq \frac{c_0}{2} r^k. \tag{3.9}$$

Firstly, let us deduce that

$$P(\Omega \setminus B_n) \leq \exp\big(-n\varphi^{P_X}_{x(\delta)}(h_n)\big) \leq \exp\Big(-n\frac{c_0}{2}(h_n)^k\Big) = \exp(-2\ln(n)) = \frac{1}{n^2}.$$

Since

$$\frac{h_n}{r_n} = \frac{(4/c_0)^{1/k}}{r_n\big(\frac{n}{\ln(n)}\big)^{1/k}} \xrightarrow{n\to\infty} 0,$$

the relation (3.7) from the proof of Theorem 3.25 remains valid in this setting as

$$T_n^{(2)} = \frac{\varphi^{P_X}_{x(\delta)}(r_n - h_n)}{\varphi^{P_X}_{x(\delta)}(r_n + h_n)} \sim \left(\frac{r_n - h_n}{r_n + h_n}\right)^k = \left(\frac{1 - h_n/r_n}{1 + h_n/r_n}\right)^k \xrightarrow{n\to\infty} 1.$$

Let $q > 0$ be a constant. Then, by analogous argumentation, we receive for sufficiently large n that

$$P(A_n \cap B_n) \leq N(\mathcal{Y}, d_{Euc}, h_n) \exp\big(-q\cdot n\cdot \varphi^{P_X}_{x(\delta)}(r_n)\big) \leq \frac{K}{h_n^k}\exp\Big(-\frac{q\cdot c_0}{2} n r_n^k\Big),$$

where we used (a1) and (3.9). Since, for instance, $r_n > \Big(\frac{6}{q\cdot c_0}\frac{\ln(n)}{n}\Big)^{1/k}$ holds for sufficiently large n, we can conclude that

$$P(A_n \cap B_n) \leq K \cdot \frac{c_0}{4} \cdot \frac{n}{\ln(n)} \exp\big(-3\ln(n)\big) = \frac{q'}{\ln(n)n^2}$$

holds for large n and another constant $q' > 0$. The upper bounds given for both $P(\Omega \backslash B_n)$ and $P(A_n \cap B_n)$ hold for every $n \geq n(\varepsilon)$. Since $\sum_n n^{-2} < \infty$ and $\sum_n (\ln(n)n^2)^{-1} < \infty$, we can again conclude from the Borel-Cantelli lemma that $\widehat{\theta}_{n,r_n}$ converges to the mode $\text{Mod}(P_X)$ almost surely.

■

Our theory presented in the preceding corollary is also suitable for the situation where P_X is supported on a k'-dimensional subset of $\mathbb{R}^k$ and $1 \leq k' < k$, e.g. if $k \geq 2$ and the distribution of P_X is singular. If, in that case, (a1) and (a2) can be verified for k' instead of k, then strong consistency is achieved if k is substituted by k' in the requirement for the radii r_n in Corollary 3.26 and in its proof.
The statements in Theorem 3.25 and Corollary 3.26 are both given under the additional requirement that the radii $(r_n)_n$ do not converge too fast. It is not surprising that in the corollary, where the support is a subset of a finite-dimensional space, the requirement is less restrictive. The condition in the corollary coincides with the restriction (9.4) that Ferraty and Vieu (2006) impose on the bandwidth parameter of their kernel mode estimator if $\psi(t)$ in (9.1) therein is set equal to the Lebesgue measure of a ball with radius t. Recall that their condition (9.1) is equal to (3.8) in this work. In the following chapter we will impose further constraints on the small ball probability functions, which will allow us to explicitly state the optimal order of convergence of the radii. We will show that this particular order is identical to the optimal rate of convergence over a specified class of distributions.

Chapter 4

Rate-optimal mode estimation

Let $(\mathcal{F}, d)$ be a Polish metric space and let $\mathcal{Y} \in \mathbb{B}(\mathcal{F})$ be a non-empty, totally bounded set.

In this chapter, we will propose a second mode estimator and deduce its maximum risk over a class of distributions with a mode, which will be denoted by $\mathcal{P}$. Therefore, we will assume that every $Q \in \mathcal{P}$ assigns a positive probability to $\mathcal{Y}$ and has a mode that is contained in $\mathcal{Y}$. We will define an estimator based on finite covers of $\mathcal{Y}$. Contrary to the previous considerations, the estimator will be set equal to an element x of some finite cover of $\mathcal{Y}$ such that the amount of data points that fall into $B_d(x, r)$ is maximised over all elements of the cover. Therein, $r > 0$ is once again a parameter of estimation that depends on n.

It is necessary to require upper and lower bounds on the entropy of $\mathcal{Y}$. We will apply a system $\mathcal{C} := \big(C(r)\big)_{r>0}$ of finite covers of $\mathcal{Y}$. These covers do not need to be minimum covers, but we will impose that their logarithmic cardinality admits the same rational order as the one we impose on $\ln N(\mathcal{Y}, d, r)$.

As for the small ball probabilities, we will relax our previous requirements (A1) and (a1) in the sense that we merely demand a lower bound for the small ball probability function of the mode. The exact asymptotic order is not specified in our stipulations. However, for each $Q \in \mathcal{P}$, we will additionally impose both upper and lower bounds on the quotient

$$\frac{\varphi^Q_{\mathrm{Mod}(Q)}(\varepsilon)}{\varphi^Q_y(\varepsilon)}, \tag{4.1}$$

where $y \in \mathcal{F}$ and $\varepsilon > 0$. Upper bounds on (4.1) are usually more difficult to justify and will only be required for certain points y that are taken from the covers of the system $\mathcal{C}$.

The chapter is divided into the following two sections:

- **Section 4.1** contains the definition of both a class of distributions with a mode and a mode estimator. The maximum rate of that estimator over the class $\mathcal{P}$ will be established.

- In **Section 4.2** we will set $\mathcal{Y}$ equal to a Sobolev ellipsoid, which will be defined as a class of univariate, square-integrable functions on a compact interval such that the L^2-norms of the (weak) derivatives are bounded. For every radius $r > 0$, we will give an explicit definition of a finite r-cover $C(r)$ of $\mathcal{Y}$ that obeys the cardinality stipulations introduced in Section 4.1. The minimax rate will be deduced for that setting.

4.1 The maximum rate

We will start by formalising the necessary entropy inequalities for $\mathcal{Y}$ and the restrictions imposed on the system of finite covers $\mathcal{C}$.

Definition 4.1. Let $\mathcal{C} := \big(C(r)\big)_{r>0}$ be a family of subsets of $\mathcal{F}$ such that for every $r > 0$ the set $C(r)$ is a finite r-cover of $\mathcal{Y}$. Then the pair $(\mathcal{Y}, \mathcal{C})$ is said to fulfill *property (E)* for parameters $\alpha > 0$ and $\rho > 0$ and two constants $0 < L' < L$ if, additionally, the following three conditions hold:

(C1) $\forall r \in (0, \rho] : \dfrac{L'}{r^\alpha} \leq \ln N(\mathcal{Y}, d, r) \leq \dfrac{L}{r^\alpha}$

(C2) $\forall r \in (0, \rho] : \ln |C(r)| \leq \dfrac{L}{r^\alpha}$

(C3) $\forall r > 0, \forall x \in C(r) : B_d(x, r) \cap \mathcal{Y} \neq \emptyset$

The covers $C(r)$ from the system $\mathcal{C}$ in Definition 4.1 are not required to be minimum covers. By imposing condition (C3) we make sure that, for every $r > 0$, the distance between any $x \in C(r)$ and the set $\mathcal{Y}$ is less than or equal to r.

Now we specify a class of distributions with a mode that will be based on a pair $(\mathcal{Y}, \mathcal{C})$ that fulfills property (E).

Definition 4.2. Assume that the pair $(\mathcal{Y},\mathcal{C})$ fulfills property (E) for parameters $\alpha > 0, \rho > 0$ and two constants $0 < L' < L$. Let $\beta > \alpha, \gamma > 0$ and $\eta \in \big(0, \min(\rho,1)\big)$ be other parameters and let $K \in (0, L]$ be a constant. Then $\mathcal{P} := \mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})$ shall denote the class of probability distributions defined on the space $\big(\mathcal{F}, \mathbb{B}(\mathcal{F})\big)$ such that for every $Q \in \mathcal{P}$ the following four properties hold, where $C_1, C_2 > 0$ and $C_3 \in (0,1]$ are constants that do not depend on Q:

(P0) $Q(\mathcal{Y}) > 0$

(P1) There is some $m := m(Q) \in \mathcal{Y} \cap \operatorname{supp}(Q)$ such that

$$\forall r > 0 \colon \big(x \in \mathcal{F} \wedge d(m,x) \geq 4r\big) \Longrightarrow \frac{\varphi_m^Q(r)}{\varphi_x^Q(r)} \geq 1 + C_1(d(m,x) - r)^\gamma,$$

where we again set $1/0 := \infty$.

(P2) For every $r \in (0, \eta]$ there exists some $y := y(r) \in C(r^{1+\beta})$ such that $d(m,y) \leq r$, $\varphi_y^Q(r) > 0$ and

$$\frac{\varphi_m^Q(r)}{\varphi_y^Q(r)} \leq 1 + C_2 r^\gamma.$$

(P3) For every $r \in (0, \eta]$ we have $\varphi_m^Q(r) \geq C_3 \exp(-K r^{-\alpha})$.

We want to discuss our conditions (P0)-(P3) in greater detail.

Remarks 4.3. Let $Q \in \mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})$.

(a) Fix some $\delta > 0$, let $r \leq \delta/4$ and $x \in \mathcal{F}$ such that $d(m,x) \geq \delta$. Then, by (P1) we have

$$\frac{\varphi_m^Q(r)}{\varphi_x^Q(r)} \geq 1 + C_1\big(d(m,x) - r\big)^\gamma \Longrightarrow \liminf_{r \to 0^+} \inf_{d(m,x) \geq \delta} \frac{\varphi_m^Q(r)}{\varphi_x^Q(r)} \geq 1 + C_1 \delta^\gamma > 1,$$

from which it follows that $m \in M^Q(\delta)$ for every $\delta > 0$ and thus $m = \operatorname{Mod}(Q) \in \mathcal{Y}$. Hence, condition (P1) implies that $\mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C}) \subset \mathcal{L}_{(\mathcal{F},d)}$. Note that $\operatorname{supp}(Q)$ is not necessarily contained in $\mathcal{Y}$. But by imposing (P1) we restrict the set in which we have to search for the mode to $\mathcal{Y}$.
(P1) is stricter than simply requiring that $Q \in \mathcal{L}_{(\mathcal{F},d)}$ and $\operatorname{Mod}(Q) \in M^Q(\delta)$ for every $\delta > 0$, because it is not just a limit relation but an inequality for a quotient of the type (4.1) that involves the additional steepness summand $C_1\big(d(m,x) - \delta\big)^\gamma$. The condition can be motivated by Proposition 3.15(b), from which it follows that such an inequality holds for a centred Gaussian measure, e.g. for $m = 0, \gamma = 2$ and elements x of the topological closure of the Cameron-Martin space. Part (a) of that

proposition contains an upper bound on the quotient of two Gaussian small ball probabilities that is stated in terms of the Cameron-Martin norm.

(b) Since $\eta < 1$, we have that $r^{1+\beta} < r$ for every $r \in (0, \eta]$. If the relation

$$\frac{\varphi_m^Q(r)}{\varphi_m^Q(r - r^{1+\beta})} \leq 1 + C_2 r^\gamma, \qquad \forall r \in (0, \eta], \tag{4.2}$$

holds, then we can show that (P2) also holds, because there exists $y \in C(r^{1+\beta})$ such that $d(m, y) \leq r^{1+\beta} < r$, which implies $\varphi_y^Q(r) \geq \varphi_m^Q(r - r^{1+\beta})$. In order for (4.2) to hold, $\varphi_m^Q(r)/\varphi_m^Q(r - r^{1+\beta})$ must converge to one as $r \to 0^+$. For instance, if φ_m^Q is asymptotically equivalent to the exponential-type function $g_{c_0,c_1,s,t}$ introduced in Section 3.4.2, then we can deduce from Lemma 3.24 that if $t \leq \alpha$, since $\beta > \alpha$, we have that $\varphi_m^Q(r)/\varphi_m^Q(r - r^{1+\beta})$ converges to one as r tends to zero. The same convergence relation holds if the order of $\varphi_m^Q(r)$ is polynomial. Hence, (4.2) (and thus (P2)) can be verified in the case where tight bounds on $\varphi_m^Q(r)$ exist. (4.2) is not necessary for (P2) to hold, as we only impose that for every $r \in (0, \eta]$ an inequality can be checked at a single point $y \in C(r^{1+\beta})$ that is close to m.

(c) (P3) is a uniform lower bound for the small ball probability function of the mode. It follows by (P0), (P1), (C1) and Lemma 3.18(b), that we have

$$\liminf_{r \to 0^+} \varphi_m^Q(r) \exp\left(Lr^{-\alpha}\right) > 0,$$

since $m(Q) \in M^Q(\delta)$ for every $\delta > 0$ (see (a)). Since $\exp\big(-Lr^{-\alpha}\big) \leq \exp\big(-Kr^{-\alpha}\big)$, the lower bound proposed in (P3) does not contradict Lemma 3.18(b). Since $\ln N(\mathcal{Y}, d, r) \asymp r^{-\alpha}, r \to 0^+$, it would not be justified to propose an exponential lower bound with a parameter that is smaller than α.

We will now continue our considerations of the mode estimation problem over a class $\mathcal{P}$. Therefore, assume that $(\mathcal{Y}, \mathcal{C})$ is a pair that fulfills property (E) for parameters $\alpha > 0$ and $\rho > 0$ and two constants $0 < L' < L$. Let $\beta > \alpha, \gamma > 0, \eta \in \big(0, \min(\rho, 1)\big)$ and $K \in (0, L]$ and set $\mathcal{P}$ equal to the class $\mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y}, \mathcal{C})$ specified in Definition 4.2.

In order to define an estimator we can reuse much of the notation and definitions from Section 3.4.1. Let $n \in \mathbb{N}$ and let $X_1, \ldots, X_n$ be independent and identically distributed observations on a probability space $(\Omega, \mathcal{A}, P)$ that take on values in $(\mathcal{F}, \mathbb{B}(\mathcal{F}))$ such that $P_{X_1} =: P_X \in \mathcal{P}$. Let $r \in (0,1)$, set $C'(r) := C(r^{1+\beta})$ and assume that $C'(r) = \{x_1, \ldots, x_{|C'(r)|}\}$. Recall that $Z_n(x,r) = \sum_{i=1}^n \mathbf{1}_{B_d(x,r)}(X_i)$ for every $x \in \mathcal{F}$ and set

$$\tilde{Z}_n(x_j, r) := Z_n(x_j, r) + U_{j,n}, \qquad \forall j \in \{1, \ldots, |C'(r)|\},$$

where $U_{1,n}, \ldots, U_{|C'(r)|,n}$ are independent random variables with a uniform distribution on $(0,1) \subset \mathbb{R}$. We assume that $\big(Z_n(x_1, r), \ldots, Z_n(x_{|C'(r)|}, r)\big)$ and $\big(U_{1,n}, \ldots, U_{|C'(r)|,n}\big)$ are independent. The set $\arg\max_{j \in \{1,\ldots,|C'(r)|\}} \tilde{Z}_n(x_j, r)$ is a random set and almost surely contains exactly one element. Let $A \in \mathcal{A}$ be a set such that $P(A) = 1$ and

$$\Big|\Big(\operatorname*{arg\,max}_{j \in \{1,\ldots,|C'(r)|\}} \tilde{Z}_n(x_j, r)\Big)(\omega)\Big| = 1$$

for every $\omega \in A$. Define the random variable

$$R_n(C'(r)) \colon (\Omega, \mathcal{A}, P) \longrightarrow \big(\{1, \ldots, |C'(r)|\}, 2^{\{1,\ldots,|C'(r)|\}}\big)$$

by

$$R_n(C'(r))(\omega) = i \Longleftrightarrow \Big(\operatorname*{arg\,max}_{j \in \{1,\ldots,|C'(r)|\}} \tilde{Z}_n(x_j, r)\Big)(\omega) = \{i\},\ \forall \omega \in A, \forall i \in \{1, \ldots, |C'(r)|\}$$

and $R_n(C'(r))(\omega) = 1$ for every $\omega \in \Omega \setminus A$ and set

$$\widehat{m}_n(C'(r)) := x_{R_n(C'(r))}.$$

This is the mode estimator whose properties we will analyse in this section. It is important to note that it depends on the cover $C'(r) = C(r^{1+\beta})$ of $\mathcal{Y}$ and thus on the system $\mathcal{C}$. The radius r will usually depend on n and we will give a convergence result on the sequence $\big(\widehat{m}_n(C'(r_n))\big)_{n \in \mathbb{N}}$ where $(r_n)_{n \in \mathbb{N}}$ are positive real numbers that converge to zero.

It is our goal to derive the order of convergence of the uniform squared risk

$$\sup_{Q \in \mathcal{P}} \mathbb{E}_Q\, d^2\big(\widehat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big),$$

which, of course, coincides with the squared maximum risk of $\widehat{m}_n(C'(r_n))$ over $\mathcal{P}$. The notation $\mathbb{E}_\bullet$ will be used to indicate the distribution of a single observation. Asymptotic

relations will be given under additional conditions on the (order of) the radii $(r_n)_n$, which will once again serve as an important estimation parameter.
Fixing some $Q \in \mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})$, the following lemma contributes a useful inequality for a probability involving the estimation error $d\big(\hat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big)$.

Lemma 4.4. *Let $(r_n)_{n\in\mathbb{N}}$ be a sequence of positive real numbers that converges to zero. Fix $Q \in \mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})$ and set $m := \mathrm{Mod}(Q)$. If $t := \max\big(4, \big(\frac{3C_2}{C_1}\big)^{1/\gamma} + 1\big)$, where $C_1, C_2 > 0$ are the constants taken from the conditions (P1) and (P2), then there exist some $n' \in \mathbb{N}$ and a constant $q > 0$ that are both independent of Q and m such that*

$$P\big(d(\hat{m}_n(C'(r_n)), m) \geq tr_n\big) \leq |C(r_n^{1+\beta})| \exp\Big(-qnr_n^{2\gamma}\varphi_m^Q(r_n)\Big)$$

holds for every $n \geq n'$.

Proof: Let $n \in \mathbb{N}$ be sufficiently large such that $r_i \in (0,1)$ holds for every $i \geq n$. Recall that $C'(r_n) = C(r_n^{1+\beta})$ and again assume that $C'(r_n) = \big\{x_1^{(n)}, \ldots, x_{|C'(r_n)|}^{(n)}\big\}$. Let us begin by deducing that

$$\begin{aligned}
&P\big(d\big(\hat{m}_n(C'(r_n)), m\big) \geq tr_n\big) = \\
&\qquad P\Big(\exists j \in \big\{1, \ldots, |C'(r_n)|\big\} : d\big(x_j^{(n)}, m\big) \geq tr_n, \{j\} = \operatorname*{arg\,max}_{l\in 1,\ldots,|C'(r_n)|} \tilde{Z}_n(x_l, r_n)\Big).
\end{aligned}$$

We have $r_n \leq \eta$ for n large. Thus, using (P2), for n large, we can select some $x^* \in C'(r_n)$ such that

- $d(x^*, m) \leq r_n$,
- $\varphi_{x^*}^Q(r_n) > 0$ and
- $(1 + C_2 r_n^\gamma)\varphi_{x^*}^Q(r_n) \geq \varphi_m^Q(r_n)$.

Without loss of generality, we can assume that $x^* = x_1^{(n)}$. We are now prepared to deduce

$$\begin{aligned}
&P\big(d(\hat{m}_n(C'(r_n)), m) \geq tr_n\big) \\
&\qquad \leq P\Big(\exists j \in \big\{1, \ldots, |C'(r_n)|\big\} : d\big(x_j^{(n)}, m\big) \geq tr_n \wedge Z_n(x_j^{(n)}, r_n) \geq Z_n(x_1^{(n)}, r_n)\Big) \\
&\qquad = P\Bigg(\bigcup_{\{j\in\{1,\ldots,|C'(r_n)|\} | d(x_j^{(n)}, m) \geq tr_n\}} \Big\{Z_n(x_j^{(n)}, r_n) \geq Z_n(x_1^{(n)}, r_n)\Big\}\Bigg)
\end{aligned}$$

$$\le \sum_{\{j\in\{1,\ldots,|C'(r_n)|\}|d(x_j^{(n)},m)\ge tr_n\}} P\big(Z_n\big(x_j^{(n)},r_n\big)\ge Z_n\big(x_1^{(n)},r_n\big)\big)$$
$$\le |C'(r_n)|\cdot \sup_{\{j\in\{1,\ldots,|C'(r_n)|\}|d(x_j^{(n)},m)\ge tr_n\}} P\Big(Z_n(x_j^{(n)},r_n)\ge Z_n(x_1^{(n)},r_n)\Big)$$
$$=: |C'(r_n)|\cdot \sup_{\{j\in\{1,\ldots,|C'(r_n)|\}|d(x_j^{(n)},m)\ge tr_n\}} p_{j,n}.$$

We will now derive a uniform upper bound for the probability $p_{j,n}$ over all j such that $d(x_j^{(n)},m)\ge tr_n$. The following auxiliary result is required: Assuming that n is sufficiently large, we can show that

$$\frac{\varphi^Q_{x_1^{(n)}}(r_n)}{\varphi^Q_x(r_n)} \ge 1+C_2r_n^\gamma \tag{4.3}$$

is valid for any $x\in\mathcal{F}$ such that $d(x,m)\ge tr_n$. In fact, since $t=\max\big(4,\big(\frac{3C_2}{C_1}\big)^{1/\gamma}+1\big)$ and $d(x_1^{(n)},m)\le r_n$, we can use (P1) and (P2) to derive that

$$\frac{\varphi^Q_{x_1^{(n)}}(r_n)}{\varphi^Q_x(r_n)} = \frac{\varphi^Q_{x_1^{(n)}}(r_n)}{\varphi^Q_m(r_n)}\cdot\frac{\varphi^Q_m(r_n)}{\varphi^Q_x(r_n)} \ge \frac{1+C_1\big((t-1)r_n\big)^\gamma}{1+C_2r_n^\gamma} \ge \frac{1+3C_2r_n^\gamma}{1+C_2r_n^\gamma} = 1+\frac{2C_2}{1+C_2r_n^\gamma}r_n^\gamma$$
$$> 1+C_2r_n^\gamma.$$

The last step uses that $r_n\to 0$ as $n\to\infty$. Hence, (4.3) is proven. Now we can realise that, for n large, the requirements to apply Lemma 3.23 to $A:=B_d\big(x_1^{(n)},r_n\big)$ and $B:=B_d\big(x_j^{(n)},r_n\big)$, where $d(x_j^{(n)},m)\ge tr_n$, are all satisfied. In particular, since $t\ge 4$, we have $A\cap B=\emptyset$. It yields

$$p_{j,n} \le \bigg(1-\Big(\Big(\varphi^Q_{x_j^{(n)}}(r_n)\Big)^{1/2}-\Big(\varphi^Q_{x_1^{(n)}}(r_n)\Big)^{1/2}\Big)^2\bigg)^n$$
$$\le \exp\bigg(-n\Big(\Big(\varphi^Q_{x_1^{(n)}}(r_n)\Big)^{1/2}-\Big(\varphi^Q_{x_j^{(n)}}(r_n)\Big)^{1/2}\Big)^2\bigg)$$
$$\le \exp\left(-n\left(\left(\varphi^Q_{x_1^{(n)}}(r_n)\right)^{1/2}-\left(\frac{\varphi^Q_{x_1^{(n)}}(r_n)}{1+C_2r_n^\gamma}\right)^{1/2}\right)^2\right)$$
$$= \exp\left(-n\varphi^Q_{x_1^{(n)}}(r_n)\left(1-\frac{1}{(1+C_2r_n^\gamma)^{1/2}}\right)^2\right)$$
$$\le \exp\left(-n\frac{\varphi^Q_m(r_n)}{1+C_2r_n^\gamma}\left(1-\frac{1}{(1+C_2r_n^\gamma)^{1/2}}\right)^2\right) \le \exp\Big(-qnr_n^{2\gamma}\varphi^Q_m(r_n)\Big).$$

In the third inequality we made use of (4.3), from which we derive

$$\varphi^Q_{x_j^{(n)}}(r_n) \leq \frac{\varphi^Q_{x_1^{(n)}}(r_n)}{1+C_2 r_n^\gamma}$$

and in the first inequality on the fifth line we used (P2). We can find a constant $q > 0$ such that the last inequality holds for n large as we have $1 + C_2 r_n^\gamma \overset{n\to\infty}{\longrightarrow} 1$ and

$$1 - \frac{1}{(1+C_2 r_n^\gamma)^{1/2}} = \frac{(1+C_2 r_n^\gamma)^{1/2}-1}{(1+C_2 r_n^\gamma)^{1/2}} = \frac{C_2 r_n^\gamma}{(1+C_2 r_n^\gamma)^{1/2}\cdot\big((1+C_2 r_n^\gamma)^{1/2}+1\big)} \sim \frac{C_2}{2} r_n^\gamma$$

for $n \to \infty$.

Summarising all of our previous efforts we end up with

$$\begin{aligned} P\big(d(\hat{m}_{n,}(C'(r_n)), m) \geq t r_n\big) &\leq |C'(r_n)| \cdot \sup_{\{j\in\{1,\dots,|C'(r_n)|\} | d(x_j^{(n)}, m) \geq t r_n\}} p_{j,n} \\ &\leq |C(r_n^{1+\beta})| \exp\Big(-q n r_n^{2\gamma} \varphi^Q_m(r_n)\Big), \end{aligned}$$

for n large, which is the bound we wanted to attain.

■

The dependence on Q of the right side of the inequality in Lemma 4.4 can be eliminated by using the uniform lower bound for φ^Q_m that is provided in condition (P3). Hence, we can conclude that there exists $n' \in \mathbb{N}$ and a constant $q' > 0$ depending on the sequence $(r_n)_n$ and the parameters of $\mathcal{P}$ such that

$$\sup_{Q\in\mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})} P\big(d(\hat{m}_n(C'(r_n)), \mathrm{Mod}(Q)) \geq t r_n\big) \leq |C(r_n^{1+\beta})| \exp\left(-q' n r_n^{2\gamma} \exp\left(-K r_n^{-\alpha}\right)\right) \tag{4.4}$$

holds for every $n \geq n'$. The probability in the expression on the left side of (4.4) is evaluated under the assumption that the distribution of a single observation is Q and the supremum is taken over all $Q \in \mathcal{P}$.
The lemma is now used to derive conditions for $(r_n)_n$ under which $\big(\widehat{m}_n(C'(r_n))\big)_{n\in\mathbb{N}}$ has the desirable asymptotic properties, e.g. strong consistency (see Theorem 4.5) or uniform consistency in the second mean (see Theorem 4.6).

Theorem 4.5. *Let $Q \in \mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})$ and let $(r_n)_{n\in\mathbb{N}}$ be a sequence of positive real numbers that converges to zero. If*

$$\liminf_{n\to\infty} \big(r_n\big)^{\alpha} \ln(n) > K,$$

then we have

$$d\big(\widehat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) \overset{n\to\infty}{\longrightarrow} 0$$

almost surely.

Proof: If $\sum_{n=1}^{\infty} P\big(d\big(\widehat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) \le tr_n\big) < \infty$, where $t > 0$ is defined as in Lemma 4.4, then $\sum_{n=1}^{\infty} P\big(d\big(\widehat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) < \varepsilon\big) < \infty$ holds for every $\varepsilon > 0$ and the claim follows by the Borel-Cantelli lemma. Hence, if we can show that

$$\limsup_{n\to\infty} n^2 P\big(d\big(\widehat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) \le tr_n\big) < \infty,$$

then the proof is complete. There exists a constant $l \in (0,1)$ such that if n is sufficiently large, we have

$$\big(r_n\big)^{\alpha} \ln(n) \ge \frac{K}{l} \iff K\big(r_n\big)^{-\alpha} \le l \cdot \ln(n),$$

which implies $\exp\big(-K(r_n)^{-\alpha}\big) \ge \exp\big(-l\cdot\ln(n)\big) = n^{-l}$. Hence, there exists a constant $s_1 > 0$ such that, for n large, we have

$$\exp\big(-q' n r_n^{2\gamma} \exp\big(-K(r_n)^{-\alpha}\big)\big) \le \exp\Big(-s_1 n^{1-l}\cdot \ln(n)^{-2\gamma/\alpha}\Big),$$

where $q' > 0$ is the constant introduced in (4.4). On the other hand, we can use condition (C2) to deduce that there exists another constant $s_2 > 0$ such that

$$\big|C(r_n^{1+\beta})\big| \le \exp\big(L r_n^{-\alpha(1+\beta)}\big) \le \exp\big(s_2 \ln(n)^{1+\beta}\big) = n^{s_2\left(\ln(n)\right)^{\beta}}.$$

Finally, we can use (4.4) and our previous efforts to deduce that

$$n^2 P\big(d\big(\widehat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) \le tr_n\big) \le n^{s_2\left(\ln(n)\right)^{\beta}+2} \cdot \exp\Big(-s_1 n^{1-l}\cdot \ln(n)^{-2\gamma/\alpha}\Big) =: T_n.$$

Now we have $\lim_{n\to\infty} T_n = 0$, because for every $z_1, z_2 > 0$ it holds that $\lim_{n\to\infty} n^{z_1} \ln(n)^{-z_2} = \infty$ and thus

$$\ln T_n = \Big(s_2\big(\ln(n)\big)^{\beta} + 2\Big)\ln(n) - s_1 \frac{n^{1-l}}{\ln(n)^{2\gamma/\alpha}}$$
$$= -\frac{n^{1-l}}{\ln(n)^{2\gamma/\alpha}}\left(\frac{-\ln(n)^{2\gamma/\alpha}\Big(\big(s_2\ln(n)\big)^{\beta}+2\Big)\ln(n)}{n^{1-l}} + s_1\right) \overset{n\to\infty}{\longrightarrow} -\infty.$$

■

Strong consistency is achieved if the rate at which the radii $(r_n)_n$ tend to zero is sufficiently slow, which is similar to the condition under which Theorem 3.25 holds, where strong consistency was proven for our first mode estimator. The constraint we imposed on the radii in Theorem 4.5 is less restrictive than the one that was used in Theorem 3.25. This is due to the dependence of $\hat{m}_n(\cdot)$ on a cover and our conditions (P1) and (P2), which both make statements on the small ball probabilities at points that are not necessarily contained in the sets $M^Q(\cdot)$. It is worth noting that if, for instance, $r_n = \big(K/\ln(n)\big)^{1/\alpha}$, then the expression on the right side of (4.4) diverges to $+\infty$ as $n \to \infty$. Hence, in order to answer the question whether consistency still holds if the constraint on $(r_n)_n$ is relaxed, there is need for a different approach. We will soon realise that $\ln(n)^{-2/\alpha}$ is the order at which the squared maximum risk of $\widehat{m}_n$ over $\mathcal{P}$ converges to zero. In the following section, a result will be given that states that the squared minimax risk admits a lower bound of the same order if the pair $(\mathcal{Y}, \mathcal{C})$ is specified further. We have mentioned earlier that, in our understanding, the notion of the optimal rate of convergence is unique up to a constant and it is outside the scope of this thesis to give the exact constants that would lead to the sharpest (upper or lower) bounds.

Theorem 4.6. *Let $(r_n)_{n\in\mathbb{N}}$ be a sequence of positive real numbers that satisfies*

$$K < \liminf_{n\to\infty} \ln(n)\big(r_n\big)^{\alpha} \leq \limsup_{n\to\infty} \ln(n)\big(r_n\big)^{\alpha} < \infty.$$

Then we have

$$\limsup_{n\to\infty}\ \ln(n)^{2/\alpha} \sup_{Q\in\mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})} \mathbb{E}_Q\, d^2\big(\hat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) < \infty.$$

Proof: Let $n \in \mathbb{N}$ and $Q \in \mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})$. Since $\mathcal{Y}$ is totally bounded, it is contained within a ball of finite radius, because there is a finite union of balls with bounded radii that covers $\mathcal{Y}$. We have that $\widehat{m}_n(C'(r_n)) \in C\big(r_n^{1+\beta}\big)$ and thus, by our condition (C3),

$\inf_{y\in\mathcal{Y}} d\big(\widehat{m}_n(C'(r_n)), y\big) \leq r_n^{1+\beta}$. Hence, for n large, there exists a constant $D > 0$ such that

$$d^2\big(\hat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) \leq D^2.$$

We have

$$\begin{aligned}
\mathbb{E}_Q\, d^2\big(\hat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) &= \int_{[0,D^2]} P\Big(d^2\big(\hat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) \geq r\Big)\, dr \\
&= \int_{[0,(tr_n)^2)} + \int_{[(tr_n)^2,D^2]} P\Big(d^2\big(\hat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) \geq r\Big) dr \\
&\leq (tr_n)^2 + D^2 P\Big(d\big(\hat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) \geq tr_n\Big)
\end{aligned}$$

where $t > 0$ is defined as in Lemma 4.4. We can deduce that

$$\begin{aligned}
&\sup_{Q\in\mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})} \mathbb{E}_Q\, d^2\big(\hat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) \\
&\qquad\leq (r_n)^2\bigg(t^2 + \Big(\frac{D}{r_n}\Big)^2 \sup_{Q\in\mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})} P\Big(d(\hat{m}_n(C'(r_n)), \mathrm{Mod}(Q)) \geq tr_n\Big)\bigg).
\end{aligned}$$

Since $\limsup_{n\to\infty} \ln(n)^{2/\alpha}(r_n)^2 < \infty$, this means that the claim is proven if we succeed at verifying

$$\limsup_{n\to\infty} (r_n)^{-2} \sup_{Q\in\mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})} P\Big(d\big(\hat{m}_n(C'(r_n)), \mathrm{Mod}(Q)\big) \geq tr_n\Big) < \infty,$$

which we will now accomplish. We can revisit the considerations in the proof of Theorem 4.5, which were based on (4.4), to derive that there exist constants $s_1, s_2, s_3 > 0$ and $l \in (0, 1)$ such that

$$\begin{aligned}
&(r_n)^{-2} \sup_{Q\in\mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})} P\Big(d(\hat{m}_{n,r_n}, \mathrm{Mod}(Q)) \geq tr_n\Big) \\
&\qquad\leq s_3\big(\ln(n)^{2/\alpha}\big) n^{s_2\left(\ln(n)\right)^{\beta}} \cdot \exp\Big(-s_1 n^{1-l} \cdot \ln(n)^{-2\gamma/\alpha}\Big) := T_n'.
\end{aligned}$$

Since, for n large, we have $0 \leq T_n' \leq T_n$, where T_n is defined as in the proof of Theorem 4.5, we have $\lim_{n\to\infty} T_n' = 0$ and the proof is complete.

■

Theorem 4.6 implies that, for n large, the squared maximum risk of the estimator $\widehat{m}_n(C'(r_n))$ over $\mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})$ attains an upper bound of the order $\big(\ln(n)\big)^{-2/\alpha}$, if the radii $(r_n)_n$ are bounded and admit the order $\big(\ln(n)\big)^{-1/\alpha}$. The bigger the parameter

α, the slower the rate converges. A greater value of α in the definition of our class $\mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y},\mathcal{C})$ leads to greater (upper and lower) bounds on the covering numbers of $\mathcal{Y}$ (see condition (C1)) and, consequently, in a lower uniform bound on the small ball probability function of the mode (see condition (P3)). The lower the small ball probabilities at the mode are, the less likely it is for data points to fall into some (small) ball around it.
Furthermore, we realise that our rate is independent of the parameters β and γ. However, the conditions (P1) and (P2), where these parameters appear, are crucial to verify Lemma 4.4, which has been the essential tool in the proofs of the preceding theorems. Under (P3) and the condition $K < \liminf_{n\to\infty} \ln(n)\big(r_n\big)^{\alpha}$, which was imposed in both theorems, the expression $\exp\big(-q'nr_n^{2\gamma}\varphi_m^Q(r_n)\big)$ admits an upper bound of the type $\exp(-n^{1-l})$, where $l \in (0,1)$ is a constant. That exponential term dominates the convergence behaviour of both T_n and T_n' in our Theorems 4.5 and 4.6. To summarise, the rate of convergence critically depends on the quantifications we impose on the covering numbers of $\mathcal{Y}$ and the small ball probabilities of the mode.
The remainder of this chapter is dedicated towards deriving the lower rate of convergence, or the minimax risk. Therefore, we will consider the mode estimation problem for a specific pair $(\mathcal{Y},\mathcal{C})$ that fulfills property (E) and where both $\mathcal{Y}$ and $\mathcal{C}$ are described in detail.

4.2 The minimax rate

Set $(\mathcal{F},d) = \big(L^2\big([0,1]\big), d_2\big)$. For every $j \in \mathbb{N}$ and $x \in [0,1]$ we define

$$\nu_{2j}(x) := \sqrt{2}\cos(2\pi jx), \qquad \nu_{2j+1}(x) := \sqrt{2}\sin(2\pi jx)$$

and set $\nu_1(x) := 1$. Then we have $\nu_j \in \mathcal{F}$ for every $j \in \mathbb{N}$ and the sequence $(\nu_j)_{j\in\mathbb{N}}$ is an orthonormal basis of the Hilbert space $\mathcal{F}$. Hence, for every $f \in \mathcal{F}$ there exists a sequence $(\theta_j)_{j\in\mathbb{N}}$ of real numbers such that

$$f = \sum_{j=1}^{\infty} \theta_j \nu_j,$$

where the series converges with respect to d_2. It follows from Parseval's identity that

$$\|f\|_2^2 = \sum_{j=1}^{\infty} \theta_j^2.$$

The values of the sequence $(\theta_j)_j$ are called the *Fourier coefficients* of f. Let

$$\ell^2(\mathbb{N}) := \left\{ (x_j)_{j\in\mathbb{N}} \middle| \sum_{j=1}^{\infty} x_j^2 < \infty \right\}$$

denote the space of square-summable sequences. Then, the mapping

$$\Phi : \ell^2(\mathbb{N}) \longrightarrow L^2\big([0,1]\big), \quad (\psi_j)_{j\in\mathbb{N}} \longmapsto \sum_{j=1}^{\infty} \psi_j \nu_j, \tag{4.5}$$

is an isometric isomorphism between two Hilbert spaces. Hence, every $f \in L^2\big([0,1]\big)$ can be uniquely identified with the square-summable sequence of its Fourier coefficients $(\theta_j)_j$ (with respect to the L^2-basis $(\nu_j)_j$) and vice versa. Note that, by definition, $L^2\big([0,1]\big)$ is a space of equivalence classes of functions, which is usually neglected in the notation. Let $\Phi^{-1} : L^2\big([0,1]\big) \to \ell^2(\mathbb{N})$ denote the inverse mapping of Φ. Consequently, the ball around $f \in L^2\big([0,1]\big)$ with respect to d_2 and the ball around the sequence of its Fourier coefficients $(\theta_j)_j$ with respect to the ℓ^2-metric d_{ℓ^2} can be mutually identified by the following identity, where $\varepsilon > 0$:

$$\begin{aligned} B_{d_2}(f,\varepsilon) &= \left\{ \tilde{f} \in L^2\big([0,1]\big) \middle| \int_{[0,1]} \big(f - \tilde{f}\big)^2 \, d\lambda \le \varepsilon^2 \right\} \\ &= \left\{ \tilde{f} \in L^2\big([0,1]\big) \middle| \exists (\tilde{\theta}_j)_{j\in\mathbb{N}} \in \ell^2(\mathbb{N}) \colon \tilde{f} = \sum_{j=1}^{\infty} \tilde{\theta}_j \nu_j \wedge \sum_{j=1}^{\infty} \big(\theta_j - \tilde{\theta}_j\big)^2 \le \varepsilon^2 \right\} \\ &= \Phi\Big(B_{d_{\ell^2}}\big((\theta_j)_j, \varepsilon\big) \Big) \end{aligned}$$

We will use that relation between the Hilbert spaces $L^2\big([0,1]\big)$ and $\ell^2(\mathbb{N})$ to find a cover of some set $A \subseteq L^2\big([0,1]\big)$ by giving a cover of $\Phi^{-1}(A) \subseteq \ell^2(\mathbb{N})$: If $\mathcal{X} \subseteq \ell^2(\mathbb{N})$ is an ε-cover of $\Phi^{-1}(A)$ with respect to d_{ℓ^2}, then $\Phi(\mathcal{X})$ is an ε-cover of A with respect to d_2. Obviously, for any $\varepsilon > 0$, there do not exist finite ε-covers of the whole space $\mathcal{F} = L^2\big([0,1]\big)$, but if additional constraints are imposed on the elements of $\mathcal{F}$, e.g. smoothness or boundedness conditions, then such covers may exist. Generally speaking, the faster the sequence of Fourier coefficients $(\theta_j)_j$ of some $f \in L^2\big([0,1]\big)$ converges to zero, the smoother is f, e.g. the higher is the order up to which derivatives (or weak derivatives) exist. By

$$\Theta'(s,L) := \left\{ (\theta_j)_{j\in\mathbb{N}} \in \ell^2(\mathbb{N}) \middle| \sum_{j=1}^{\infty} a_j^2 \theta_j^2 \le \frac{L^2}{\pi^{2s}} \right\},$$

where $s \in \mathbb{N}, L > 0$ and $a_j = \begin{cases} j^s, & j \text{ is even} \\ (j-1)^s, & j \text{ is odd} \end{cases}$, we define some ellipsoid in $\ell^2(\mathbb{N})$.

Now recall that $f : [0,1] \to \mathbb{R}$ is called *absolutely continuous* if for every $\varepsilon > 0$ there exists some $\delta = \delta(\varepsilon) > 0$ such that, for any countable collection of pairwise disjoint, open intervals $I_k \subset [0,1]$, the implication

$$\sum_k \lambda(I_k) < \delta \Longrightarrow \sum_k |f(\sup I_k) - f(\inf I_k)| < \varepsilon$$

holds. Since f is defined on a compact interval, this implies the existence of the derivative f' of f at almost every $x \in (0,1)$. We will refer to f' (where the other values are chosen arbitrarily) as the *weak derivative* of f. We have $f' \in L^1([0,1])$ and consider f' as an equivalence class of functions. Hence, f' is unique and coincides with the conventional derivative if it exists (see Definition 7.17 and Theorem 7.20 in Rudin (1987)).
The following correspondence holds:

Proposition 4.7. *Let $s \in \mathbb{N}, L > 0$ and set $W'(s,L) := \Phi\big(\Theta'(s,L)\big)$. Then $W'(s,L)$ is equal to the set of all (equivalence classes of) functions $f : [0,1] \to \mathbb{R}$ such that*

(a) f is $s-1$-times differentiable,

(b) $f^{(s-1)}$ is absolutely continuous,

(c) $\|f^{(s)}\|_2^2 \leq L^2$ and

(d) $f^{(j)}(0) = f^{(j)}(1)$ holds for every $j \in \{0, \dots, s-1\}$.

Therein, for every $j \in \{0, \dots, s-1\}$, $f^{(j)}$ denotes the j-th derivative and $f^{(s)}$ is the weak derivative of $f^{(s-1)}$. We have $f^{(0)} = f$.

Proof: See Lemma A.3 in Tsybakov (2008). For $j = 0, \dots, s-1$, the values $f^{(j)}(0)$ and $f^{(j)}(1)$ are defined to be one-sided limits.

■

Since $a_1 = 0$, we have that, for every $c \in \mathbb{R}$, the constant function $f = c$ lies in $W'(s,L)$ for every $s \in \mathbb{N}$ and $L > 0$. For instance, let $c > 0$ and set $f_n = 2cn \in W'(s,L)$ for every $n \in \mathbb{N}$. Then $\{f_n | n \in \mathbb{N}\}$ is a c-packing of any space $W'(s,L)$. It follows by Lemma 2.10 that $N(W'(s,L), d_2, r) = \infty$ for every $r > 0$ and, thus, $W'(s,L)$ is not totally bounded. Hence, we will introduce the following notion of a Sobolev ellipsoid:

Definition 4.8. Let $s \in \mathbb{N}$ and $C > 0$. Then the set

$$\Theta(s, C) := \left\{ (\theta_j)_{j\in\mathbb{N}} \in \ell^2(\mathbb{N}) \,\middle|\, \sum_{j=1}^{\infty} j^{2s}\theta_j^2 \le C^2 \right\},$$

is called a *Sobolev ellipsoid of order* s. The function class

$$W(s, C) := \Phi\big(\Theta(s, C)\big) = \left\{ \sum_{j=1}^{\infty} \psi_j \nu_j \,\middle|\, (\psi_j)_{j\in\mathbb{N}} \in \Theta(s, C) \right\}$$

will also be referred to as a Sobolev ellipsoid.

In the literature, authors sometimes use the larger set $\Theta'(s, L)$ to define the term 'Sobolev ellipsoid'. Since $a_j \le j^s$ for every $j \in \mathbb{N}$, it is easy to see that $\Theta(s, C) \subset \Theta'(s, \pi^s C)$, because

$$\sum_{j=1}^{\infty} j^{2s}\theta_j^2 \le C^2 \Longrightarrow \sum_{j=1}^{\infty} a_j^2\theta_j^2 \le C^2 = \frac{(\pi^s C)^2}{\pi^{2s}}.$$

It follows that $W(s, C) \subseteq W'(s, \pi^s C) \subset L^2\big([0, 1]\big)$. In general, if $k \in \mathbb{N}$ and $p \in [1, \infty)$, then a Sobolev space is a set of k-times (weak) differentiable functions in L^p whose derivatives are contained in L^p, which serves as a motivation behind the term 'Sobolev ellipsoid' (see Proposition 4.7). We will see that the space $W(s, C)$ is totally bounded and the complexity of minimum r-covers can be deduced (see Proposition 4.10). Before we discuss the entropy of $W(s, C)$, we introduce an additional condition for a pair $(\mathcal{Y}, \mathcal{C})$ of a set $\mathcal{Y} \in \mathbb{B}(\mathcal{F})$ and a system of covers $\mathcal{C} = \big(C(r)\big)_{r>0}$, which we will require in order to define the sequences of probability measures in the proof of our minimax theorem.

Definition 4.9. We say that the pair $(\mathcal{Y}, \mathcal{C})$ satisfies *property (E')* for parameters $\alpha, \rho > 0$ and constants $L, L' > 0$ if it fulfills property (E) for these parameters and constants and, additionally, the following condition holds:

(C4) $\forall\, 0 < r' \le r\colon\ x \in C(r) \cap \mathcal{Y} \Longrightarrow x \in C(r')$.

By (C4) we impose that the family of sets $\big(C(r) \cap \mathcal{Y}\big)_{r>0}$ is decreasing with respect to inclusion. If we set $\mathcal{Y} = W(s, C)$, then it will follow from the next proposition that there exists a system of covers such that the properties (C1)-(C4) are fulfilled.

Proposition 4.10. *Let $s \in \mathbb{N}, C > 0$ and set $\mathcal{Y} = W(s, C)$. Then there exist a system $\mathcal{E} = \big(E(r)\big)_{r>0}$ of finite r-covers of $\mathcal{Y}$, some parameter $\rho > 0$ and constants $0 < L' < L$ such that the pair $\big(W(s, C), \mathcal{E}\big)$ satisfies property (E') for the parameters $\alpha = 1/s, \rho > 0$ and constants $L, L' > 0$.*

Proof: We want to refer to Kolmogorov and Tikhomirov (1961), who prove that

$$\ln N\big(W(s,C), \|\cdot\|_2, r\big) \asymp r^{-1/s}, \qquad r \to 0^+,$$

(see Theorem XVI therein). The lower bound is deduced by the construction of a sufficiently large $2r$-packing for r small, which provides a bound for the r-covering number (see our Lemma 2.10). This proves the lower bound in condition (C1) if ρ is chosen sufficiently small. Wainwright (2019) also gives a proof of both bounds (see Example 5.12 therein). We present a sketch of their proofs of the upper bound, because it involves the explicit construction of an economical cover that we will later work with. The adjustments we make to the construction principles of the covers in the two cited works will be elaborated in-depth and will not alter the results given by the authors. We have $N(W(s,C), d_2, r) = N(\Theta(s,C), d_{\ell^2}, r)$. Thus, we can construct an r-cover of $\Theta(s,C)$ and transform it into an r-cover of $W(s,C)$. Let $n \in \mathbb{N}$ and set $r_n = C \cdot 2^{-(n-1)}$. We can fix some integer $m_n \in \mathbb{N}$ such that

$$\frac{C}{m_n^s} \leq \frac{r_n}{\sqrt{2}} < \frac{C}{(m_n-1)^s}, \tag{4.6}$$

for instance, we take $m_n = \lceil 2^{(n-\frac{1}{2})\frac{1}{s}} \rceil$ and once again set $1/0 = \infty$. Now, for every $k \in \mathbb{N}$, define

$$\ell^2_{(k)}(\mathbb{N}) := \big\{(\theta_j)_j \in \ell^2(\mathbb{N}) \big| \forall j > k \colon \theta_j = 0\big\}$$

and by

$$\Theta_k(s,C) := \Theta(s,C) \cap \ell^2_{(k)}(\mathbb{N})$$

the truncated ellipsoid that is contained in a k-dimensional subspace of $\ell^2(\mathbb{N})$. We now claim that any $r_n/\sqrt{2}$-cover $\mathcal{X} \subseteq \ell^2_{(m_n)}(\mathbb{N})$ of $\Theta_{m_n}(s,C)$ is an r_n-cover of $\Theta(s,C)$. Let $(\theta_j)_{j\in\mathbb{N}} \in \Theta(s,C)$, then our claim follows from

$$\sum_{j=m_n+1}^{\infty} \theta_j^2 = \sum_{m_n+1}^{\infty} \frac{1}{j^{2s}} j^{2s}\theta_j^2 \leq (m_n)^{-2s} \sum_{j=m_n+1}^{\infty} j^{2s}\theta_j^2 \leq \Big(\frac{C}{m_n^s}\Big)^2 \leq \frac{r_n^2}{2},$$

as we deduce that

$$\min_{(\theta'_j)_{j\in\mathbb{N}}\in\mathcal{X}} \Big\| \big(\theta_j\big)_{j\in\mathbb{N}} - \big(\theta'_j\big)_{j\in\mathbb{N}} \Big\|_{\ell^2}^2 = \min_{(\theta'_j)_{j\in\mathbb{N}}\in\mathcal{X}} \sum_{j=1}^{m_n} \big(\theta_j - \theta'_j\big)^2 + \sum_{j=m_n+1}^{\infty} \theta_j^2 \leq \frac{r_n^2}{2} + \frac{r_n^2}{2} = r_n^2.$$

So the task of constructing an r_n-cover of $\Theta(s,C)$ comes down to finding an $r_n/\sqrt{2}$-cover of $\Theta_{m_n}(s, C)$ that is a subset of $\ell^2_{(m_n)}(\mathbb{N})$, which we can identify with $\mathbb{R}^{m_n}$. Now define

$$\mathcal{Z}_k := \left\{ (\theta_j)_{j\in\mathbb{N}} \in \ell^2(\mathbb{N}) \middle| (\theta_1, \ldots, \theta_{m_k}) \in \frac{r_k}{\sqrt{2m_k}} \mathbb{Z}^{m_k} \right\} \cap \ell^2_{(m_k)}(\mathbb{N})$$

for every $k \in \mathbb{N}$. In the following, for every $k \in \mathbb{N}$, we want to treat $\mathcal{Z}_k$ and $\Theta_{m_k}(s, C)$ as subsets of $\mathbb{R}^{m_k}$ and define by $Z'(r_k) \subset \mathbb{R}^{m_k}$ the set containing every $g \in \mathcal{Z}_k$ such that the m_k-dimensional hypercube with side length $r_k/\sqrt{2m_k}$ and centre point g has a non-empty intersection with $\Theta_{m_k}(s, C)$. Then, the hypercube around g is contained in $B_{d_{Euc}}(g, r_k/\sqrt{2}) \subset \mathbb{R}^{m_k}$. Hence, $Z'(r_k)$ is an $r_k/\sqrt{2}$-cover of $\Theta_{m_k}(s, C)$. By a volume argument, Kolmogorov and Tikhomirov (1961) show that there exists an $n' \in \mathbb{N}$ such that if $k \geq n'$, we have

$$\ln |Z'(r_k)| \leq Q(s) \left(\frac{\sqrt{2}C}{r_k} \right)^{1/s},$$

where the constant $Q(s) > 0$ depends on s. Now treat $\mathcal{Z}_1, \ldots, \mathcal{Z}_n$ as subsets of $\mathbb{R}^{m_n}$ and define $\mathcal{G}_n := \bigcup_{k=1}^n \mathcal{Z}_k \subset \mathbb{R}^{m_n}$. Define, by $E'(r_n)$, the set containing every $g \in \mathcal{G}_n$ such that the m_n-dimensional hypercube with side length $r_n/\sqrt{2m_n}$ and centre point g has a non-empty intersection with $\Theta_{m_n}(s, C)$. We now come back to consider $Z'(r_1), \ldots, Z'(r_n), \Theta_{m_1}(s, C), \ldots, \Theta_{m_n}(s, C)$ and $E'(r_n)$ as subsets of $\ell^2(\mathbb{N})$ by adding zeros to the elements of these sets. Since $Z'(r_n)$ is an $r_n/\sqrt{2}$-cover of $\Theta_{m_n}(s, C)$ and $Z'(r_n) \subset \ell^2_{(m_n)}(\mathbb{N})$, it is an r_n-cover of $\Theta(s, C)$ and so is $E'(r_n)$. Due to the convexity of the truncated ellipsoids and the fact that $\left(\frac{r_k}{\sqrt{2m_k}}\right)_k$ is decreasing, we can deduce that $\ln |E'(r_n)| \leq \ln \Big(\sum_{k=1}^n |Z'(r_k)| \Big)$. Let $n > n'$ be sufficiently large and set $t := \sum_{k=1}^{n'-1} |Z'(r_k)|$, then we can deduce

$$\begin{aligned}
\ln |E'(r_n)| &\leq \ln \left(t + \sum_{k=n'}^{n} |Z'(r_k)| \right) \leq \ln \left(t + n \cdot \max_{k=n', \ldots, n} |Z'(r_k)| \right) \\
&\leq \ln \left(t'n \cdot \max_{k=n', \ldots, n} |Z'(r_k)| \right) \leq \ln(t'n) + Q(s) \left(\frac{\sqrt{2}C}{r_n} \right)^{1/s} \\
&\leq 2Q(s) \cdot \left(\frac{\sqrt{2}C}{r_n} \right)^{1/s},
\end{aligned}$$

where $t' > 1$ is another constant. The last step uses that $\left(C \cdot (r_n)^{-1}\right)^{1/s} = 2^{(n-1)/s}$, which tends to infinity at a rate faster than $\ln(t'n)$.

Let $r > 0$ be arbitrary. If $r \geq C$, then we have $\Theta(s, C) \subseteq B_{d_{\ell^2}}(0, r)$. Thus, we simply set

$E'(r) := \{0\}$ in that case. If $r \in (0, C)$, then there is a unique n such that $r \in [r_{n+1}, r_n)$. We can set $E'(r) := E'(r_{n+1})$ in that case and deduce that there exists some $\rho \in (0, C)$ such that for every $r \in (0, \rho]$ we have

$$\ln|E'(r)| = \ln|E'(r_{n+1})| \leq 2Q(s)\left(\frac{\sqrt{2}C}{r_{n+1}}\right)^{1/s} = 2Q(s)\left(\frac{\sqrt{2}C}{\frac{1}{2}r_n}\right)^{1/s} < 2Q(s)\left(\frac{2\sqrt{2}C}{r}\right)^{1/s}.$$

We now define $\mathcal{E} := \big(E(r)\big)_{r>0}$ by

$$E(r) := \Phi\big(E'(r)\big), \qquad \forall r > 0, \tag{4.7}$$

and claim that the pair $\big(W(s, C), \mathcal{E}\big)$ satisfies property (E') for the parameters $\alpha = 1/s, \rho > 0$ and two constants. For the lower bound in (C1) we again refer to Kolmogorov and Tikhomirov (1961) or Wainwright (2019). Since for every $r \in (0, \rho]$ we have that $E(r)$ is an r-cover of $W(s, C)$ such that $\ln|E(r)| \leq 2Q(s)\big(2\sqrt{2}C\big)^{1/s} r^{-1/s} =: Lr^{-1/s}$, the upper bounds in (C1) and (C2) hold as well. In order to prove (C3), we can realise that, for every $r > 0$ and $x \in E(r)$, the distance between x and $W(s, C)$ is zero if $r \geq C$, and always less than or equal to $r/(2\sqrt{2})$, if $r \in (0, C)$, which is an upper bound for half the length of the diagonal of the m_{n+1}-dimensional hypercubes with side length $r_{n+1}/\sqrt{2m_{n+1}}$ that we used to define $E'(r_{n+1})$. Recall that $\Theta_k(s, C) \subset \Theta(s, C)$ for every $k \in \mathbb{N}$. We can now show (C4). If $r \geq C$, then $E(r) \cap W(s, C) = \{0\} \subset E(r')$ for every $r' \leq r$. Let $r < C$ and $x \in E(r) \cap W(s, C)$, which means that, for some $j \in \mathbb{N}$, we have $x \in \Phi\big(Z'(r_j)\big) \cap W(s, C) \subset E(r_{j'})$ for every $j' \geq j$, because the relevant intersections between hypercubes and truncated ellipsoids are never empty, as follows from the definition of $\mathcal{G}_n$ and the fact that the truncated ellipsoids are increasing with respect to set inclusion. This completes the proof.

■

Hence, we can set $(\mathcal{Y}, \mathcal{C}) := \big(W(s, C), \mathcal{E}\big)$ where $\mathcal{E}$ is defined in (4.7) and fix some $\rho > 0$ and constants $0 < L' < L$ such that $(\mathcal{Y}, \mathcal{C})$ satisfies property (E') for parameters $\alpha = 1/s, \rho > 0$ and constants $L', L > 0$. Fix three parameters $\beta > 1/s, \gamma > 0$ and $0 < \eta < \min(1, \rho)$ and a constant $0 < K \leq L$. Then

$$\mathcal{P} := \mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y}, \mathcal{C}) = \mathcal{P}_{1/s,\beta,\gamma,\eta,K}\big(W(s, C), \mathcal{E}\big) \tag{4.8}$$

denotes the class of distributions on $\big(L^2\big([0,1]\big), \mathbb{B}\big(L^2\big([0,1]\big)\big)\big)$ that fulfill the properties (P0)-(P3) specified in Definition 4.2. Every $Q \in \mathcal{P}$ has a mode in $\mathcal{Y} = W(s, C)$, which means it satisfies the smoothness properties of the class $W'(s, C)$ (see Proposition 4.7),

which depend on the explicit choice of the Hilbert basis $(\nu_j)_{j\in\mathbb{N}}$.
It is possible to show that $\mathcal{P}$ contains certain probability distributions that are supported on a finite-dimensional subspace of $L^2\big([0,1]\big)$, some of which can be described by Lebesgue density functions. This is the main idea behind the following theorem which provides a lower bound for the minimax risk of the mode estimation problem over $\mathcal{P}$.

Theorem 4.11. *Let $\mathcal{P}$ be the class of probability distributions defined in (4.8). Then we have*

$$\liminf_{n\to\infty} \ln(n)^{2s} \inf_{\hat\theta_n} \sup_{Q\in\mathcal{P}} \mathbb{E}_Q\, d_2^2\big(\hat\theta_n, \mathrm{Mod}(Q)\big) > 0,$$

where the infimum is taken over all estimators $\hat\theta_n = \hat\theta_n(X_1,\ldots,X_n)$.

Proof: Recall that $\mathcal{F} = L^2\big([0,1]\big)$. We know from our considerations in Chapter 2.3 (see Corollary 2.23(b)) that, in order to prove the claim, we need to find two sequences $\big(P_k^{(1)}\big)_{k\in\mathbb{N}}$ and $\big(P_k^{(2)}\big)_{k\in\mathbb{N}}$ of probability measures on $(\mathcal{F},\mathbb{B}(\mathcal{F}))$ such that

- there exists $k'\in\mathbb{N}$ such that $P_k^{(i)}\in\mathcal{P}$ for every $k\geq k'\in\mathbb{N}$ and $i=1,2$,
- $\limsup_{k\to\infty} k\, TV\big(P_k^{(1)},P_k^{(2)}\big) < 1$ and
- $\liminf_{k\to\infty} \ln(k)^s d_2\big(\mathrm{Mod}\,\big(P_k^{(1)}\big), \mathrm{Mod}\,\big(P_k^{(2)}\big)\big) > 0.$

Both $P_k^{(1)}$ and $P_k^{(2)}$ will be defined in a sense that they are supported on a k-dimensional subspace of $\mathcal{F}$ (with the dimension k tending to infinity) and can be described by using Lebesgue densities.
Let $k\in\mathbb{N}$ and set

$$q_k' := \frac{1}{3}\left(\frac{K}{\ln\big(8C_3k\big)}\right)^s,$$

where $C_3>0$ is the constant from our condition (P3) (see Definition 4.2). Let us begin by considering that $x\nu_1\in W(s,C)$ if and only if $x\in[-C,C]$. The definition of our probability measures will be given with the help of the covers of the system $\mathcal{E}$. Before we can define them, we need to verify the following claim. Set $r_k = C\cdot 2^{-(k-1)}$. Then if k is sufficiently large, there exist both an integer $n_k\in\mathbb{N}$ and some $q_k>0$ such that the following three properties hold:

(1) $q_k\in[q_k'/2, q_k']$

(2) $q_k\nu_1\in E(r_{n_k})\cap W(s,C)$

(3) $2q_k\leq r_{n_k}$

By construction of the covers in $\mathcal{E}$, the cover $E(r_{n_k})$ contains the set

$$\left\{ l \cdot \frac{r_{n_k}}{\sqrt{2m_{n_k}}}\nu_1 \middle| l \in \mathbb{Z}, l \cdot \frac{r_{n_k}}{\sqrt{2m_{n_k}}} \in [-C, C] \right\},$$

where m_{n_k} is the integer taken from the proof of Proposition 4.10 that fulfills (4.6). Hence, if k and n_k are sufficiently large and $r_{n_k}/\sqrt{2m_{n_k}} \leq \frac{1}{2}q_k'$ holds, then there exists $q_k > 0$ such that (1) and (2) are verified. Property (3) holds if $r_{n_k} \geq 2q_k'$. Since, by (4.6), $C/m_n^s \leq r_n/\sqrt{2}$, we have $r_{n_k}/\sqrt{2m_{n_k}} \leq \tilde{C} r_{n_k}^{1+\frac{1}{2s}}$, where $\tilde{C} > 0$ is a constant. We can conclude that (1)-(3) hold if k is sufficiently large and the inequalities

$$2q_k' \leq r_{n_k} \leq \left(\frac{q_k'}{2\tilde{C}}\right)^{\frac{2s}{2s+1}}$$

hold, which, by appropriate calculations, are equivalent to

$$\frac{\ln(C) - \ln(2) - \ln(q_k')}{\ln(2)} + 1 \geq n_k \geq \frac{\ln(C) + \frac{2s}{2s+1}\ln(2\tilde{C}) - \frac{2s}{2s+1}\ln(q_k')}{\ln(2)} + 1.$$

Since $-\ln(q_k') \stackrel{k\to\infty}{\longrightarrow} \infty$ and $-\ln(q_k') + \frac{2s}{2s+1}\ln(q_k') = -\frac{1}{2s+1}\ln(q_k') \stackrel{k\to\infty}{\longrightarrow} \infty$, from which it follows that the difference between the left and the right side of the inequalities above diverges to infinity, there exists, for k large, an integer n_k and some $q_k > 0$ such that the properties (1)-(3) hold. Henceforth we will assume that k is sufficiently large such that the pair $(n_k, q_k) \in \mathbb{N} \times (0, \infty)$ satisfies (1), (2) and (3).
We can now define both $P_k^{(1)}$ and $P_k^{(2)}$. Recall that for every $r > 0$ we write $v_k(r) = \lambda\!\!\lambda^k(B_{d_{Euc}}(0, r))$. We set $l_k = \left(\frac{K}{sk}\right)^s$ and assume that k is sufficiently large such that $l_k < q_k$. Let $m^{(k)} := (q_k, 0, \dots, 0) \in \mathbb{R}^k$ and define the following functions for every $x \in \mathbb{R}^k$:

$$\begin{aligned}
f_k^{(1)}(x) &:= \frac{1}{4\gamma \cdot v_k(l_k)} \left(1 - \left\|\frac{x - m^{(k)}}{l_k}\right\|_{\text{Euc}}^{\gamma}\right) \cdot 1_{B_{d_{Euc}}(m^{(k)}, l_k)}(x) \\
g_k^{(1)}(x) &:= \frac{1}{4\gamma \cdot v_k(l_k)} \left(1 - \left\|\frac{x + m^{(k)}}{l_k}\right\|_{\text{Euc}}^{\gamma}\right) \cdot 1_{B_{d_{Euc}}(-m^{(k)}, l_k)}(x) \\
f_k^{(2)}(x) &:= g_k^{(2)}(x) := \frac{1}{v_k(q_k)} \cdot \left(1 - \frac{1}{4(k+\gamma)}\right) \cdot \frac{1}{2^k} \cdot 1_{B_{d_{Euc}}(0, 2q_k)}(x) \\
f_k(x) &:= f_k^{(1)}(x) + f_k^{(2)}(x) \\
g_k(x) &:= g_k^{(1)}(x) + g_k^{(2)}(x)
\end{aligned}$$

It will later be shown that both f_k and g_k are Lebesgue density functions. They are continuous at every x for which $\|x\|_{\text{Euc}} < 2q_k$ and attain their unique global maximum at $m^{(k)}$ for f_k or at $-m^{(k)}$ for g_k, respectively. Set

$$\mathcal{F}^{(k)} := \left\{ \sum_{j=1}^{k} \psi_j \nu_j \in L^2\big([0,1]\big) \,\middle|\, (\psi_1, \ldots, \psi_k) \in \mathbb{R}^k \right\} \in \mathbb{B}(\mathcal{F})$$

and define the mappings

$$f : \mathcal{F}^{(k)} \longrightarrow \mathbb{R}^k, \qquad \sum_{j=1}^{k} \psi_j \nu_j \longmapsto (\psi_1, \ldots, \psi_k)$$

and

$$p_k : \mathbb{B}(\mathcal{F}) \longrightarrow \mathbb{B}(\mathbb{R}^k), \qquad A \longmapsto f\big(A \cap \mathcal{F}^{(k)}\big).$$

Note that, since $d_2\big(\sum_{j=1}^{k} \psi_j \nu_j, \sum_{j=1}^{k} \psi_j' \nu_j\big) = d_{Euc}\big((\psi_1, \ldots, \psi_k), (\psi_1', \ldots, \psi_k')\big)$, f is continuous. Furthermore, f is injective and, thus, $f(A \cap \mathcal{F}^{(k)}) \in \mathbb{B}(\mathbb{R}^k)$ for every $A \in \mathbb{B}(\mathcal{F})$. We can derive that, for pairwise disjoint sets $F_1, F_2, \ldots \in \mathbb{B}(\mathcal{F})$, we have $p_k\big(\bigcup_{j \in \mathbb{N}} F_j\big) = \bigcup_{j \in \mathbb{N}} p_k(F_j)$ and $p_k(F_1,), p_k(F_2), \ldots$ are also pairwise disjoint.

We can now define the measures

$$P_k^{(1)}(A) := \int\limits_{p_k(A)} f_k(x)\, \lambda^k(dx) \quad \text{and} \quad P_k^{(2)}(A) := \int\limits_{p_k(A)} g_k(x)\, \lambda^k(dx), \quad \forall A \in \mathbb{B}(\mathcal{F}).$$

Since f_k and g_k are non-negative, $P_k^{(1)}$ and $P_k^{(2)}$ are measures. Due to the above mentioned property of p_k, both measures are σ-additive. It remains to show that $P_k^{(i)}(\mathcal{F}) = 1$ for $i = 1, 2$. Therefore, note that $p_k(\mathcal{F}) = \mathbb{R}^k$. We have $f_k(x) = g_k(-x)$ for every $x \in \mathbb{R}^k$, which is why almost any of following computations involving one measure can be carried out analogously for the other one. For instance, it suffices to verify that $P_k^{(1)}$ is indeed a probability measure. We deduce that

$$\int\limits_{\mathbb{R}^k} f_k^{(1)}(x)\, \lambda^k(dx) = \frac{1}{4\gamma \cdot v_k(l_k)} \int\limits_{B_{d_{Euc}}(m^{(k)}, l_k)} \left(1 - \left\| \frac{x - m^{(k)}}{l_k} \right\|_{\text{Euc}}^{\gamma}\right) \lambda^k(dx)$$

$$= \frac{1}{4\gamma \cdot v_k(l_k)} \int\limits_{B_{d_{Euc}}(0, l_k)} \left(1 - \left\| \frac{x}{l_k} \right\|_{\text{Euc}}^{\gamma}\right) \lambda^k(dx) = \frac{s_{k-1}}{4\gamma \cdot v_k(l_k)} \cdot \int_0^{l_k} y^{k-1} \Big(1 - \Big(\frac{y}{l_k}\Big)^{\gamma}\Big)\, dy$$

$$= \frac{s_{k-1}}{4\gamma \cdot v_k(l_k)} \cdot \Big[\frac{y^k}{k} - \frac{y^{k+\gamma}}{(k+\gamma)(l_k)^\gamma}\Big]_0^{l_k} = \frac{1}{4\gamma \cdot v_k(l_k)} \cdot s_{k-1} \cdot (l_k)^k \cdot \frac{1}{k} \cdot \Big(1 - \frac{k}{k+\gamma}\Big)$$
$$= \frac{1}{4\gamma} \cdot \frac{\gamma}{k+\gamma} = \frac{1}{4(k+\gamma)},$$

where $s_{k-1} = 2\pi^{k/2}\Gamma(k/2)^{-1}$ is the area of the surface of the k-dimensional (Euclidean) sphere with radius one. In the third equation we used the k-dimensional substitution rule and applied it to an integral over a centred ball of a rotation invariant function. For details on k-dimensional spherical coordinates we refer to chapter 6 in Stromberg (1981).
The penultimate identity is due to $v_k(r) = \pi^{k/2}\Gamma(k/2+1)^{-1} \cdot r^k$, from which we derive

$$v_k(l_k)^{-1} \cdot s_{k-1} \cdot (l_k)^k \cdot \frac{1}{k} = \frac{2\Gamma(k/2+1)}{\Gamma(k/2)k} = 1.$$

Furthermore, we deduce

$$\int_{\mathbb{R}^k} f_k^{(2)}(x)\, \lambda^k(dx) = \frac{v_k(2q_k)}{v_k(q_k)} \cdot \frac{1}{2^k} \cdot \Big(1 - \frac{1}{4(k+\gamma)}\Big) = 1 - \frac{1}{4(k+\gamma)}$$

and conclude that

$$P_k^{(1)}(\mathcal{F}) = \int_{\mathbb{R}^k} f_k(x)\, \lambda^k(dx) = \int_{\mathbb{R}^k} f_k^{(1)}(x)\, \lambda^k(dx) + \int_{\mathbb{R}^k} f_k^{(2)}(x)\, \lambda^k(dx) = 1.$$

Hence, $P_k^{(1)}$ is a probability measure on $\big(\mathcal{F}, \mathbb{B}(\mathcal{F})\big)$ and so is $P_k^{(2)}$ by analogous reasoning. It is due to the fact that the supports of the functions $f_k^{(1)}$ and $g_k^{(1)}$ are disjoint as well as $f_k(x) - g_k(x) = f_k^{(1)}(x) - g_k^{(1)}(x)$ that we have $|f_k(x) - g_k(x)| = f_k^{(1)} + g_k^{(1)}$. Thus, we can conclude that

$$TV\big(P_k^{(1)}, P_k^{(2)}\big) = \sup_{A \in \mathbb{B}(\mathcal{F})} \big|P_k^{(1)}(A) - P_k^{(2)}(A)\big| = \sup_{A \in \mathbb{B}(\mathcal{F})} \Bigg| \int_{p_k(A)} \big(f_k(x) - g_k(x)\big)\, \lambda^k(dx) \Bigg|$$
$$\leq \sup_{B \in \mathbb{B}(\mathbb{R}^k)} \Bigg| \int_B \big(f_k(x) - g_k(x)\big) \lambda^k(dx) \Bigg| = \frac{1}{2} \int_{\mathbb{R}^k} \Big| f_k(x) - g_k(x) \Big|\, \lambda^k(dx)$$
$$= \frac{1}{2} \int_{\mathbb{R}^k} \Big(f_k^{(1)}(x) + g_k^{(1)}(x) \Big)\, \lambda^k(dx) = \frac{1}{4(k+\gamma)},$$

which leads to $\limsup_{k\to\infty} k \cdot TV\big(P_k^{(1)}, P_k^{(2)}\big) \leq 1/4$. The identity in the second line follows from Scheffé's theorem (see Lemma 2.1 in Tsybakov (2008)). Soon we will prove that $P_k^{(1)}$ and $P_k^{(2)}$ satisfy our condition (P1) and that we have $\mathrm{Mod}(P_k^{(1)}) = q_k\nu_1$ and $\mathrm{Mod}(P_k^{(2)}) = -q_k\nu_1$. For k large (such that $q_k \leq C$), this would mean that both modes lie in $\mathcal{Y} = W(s, C)$ and we deduce that

$$d_2^2\big(\mathrm{Mod}\left(P_k^{(1)}\right), \mathrm{Mod}\left(P_k^{(2)}\right)\big) = d_2^2(q_k\nu_1, -q_k\nu_1) = 4q_k^2 \geq (q_k')^2 \sim \frac{\text{const.}}{\ln(k)^{2s}}, \quad k \to \infty,$$

where we used that q_k satisfies property (1). By Corollary 2.23(b), this means that the proof is complete once we have finished to verify that, for k large, we have $P_k^{(1)}, P_k^{(2)} \in \mathcal{P}$. For the remainder of the proof, we will show that our distributions satisfy the conditions (P0)-(P3) established in Definition 4.2. As we have already mentioned, it suffices just to consider the measure $P_k^{(1)}$ instead of both measures, which is due to the symmetry relations between the densities f_k and g_k.

In order to prove that $P_k^{(1)}$ satisfies (P0), we first realise that, for every $k \in \mathbb{N}$, there exists some $\tau_k > 0$ such that

$$\left\{ \sum_{j=1}^{k} \psi_j\nu_j \in L^2\big([0,1]\big) \,\middle|\, \|(\psi_1, \dots, \psi_k)\|_{\mathrm{Euc}} \leq \tau_k \right\} \subseteq W(s, C)$$

and, hence, $B_{d_{Euc}}(0, \tau_k) \subseteq p_k(W(s, C))$, from which it follows that

$$P_k^{(1)}(W(s, C)) = \int\limits_{p_k(W(s,C))} f_k(x)\, \lambda^k(dx) \geq \int\limits_{B_{d_{Euc}}(0,\tau_k)} f_k(x)\, \lambda^k(dx) > 0.$$

We now claim that the probability measure $P_k^{(1)}$ satisfies (P1) for $m := q_k\nu_1$. Once we have succeeded at proving that claim, we know from our considerations in Remark 4.3(a) that $\mathrm{Mod}(P_k^{(1)}) = q_k\nu_1$. Verifying the condition involves deriving a lower bound for the quotient of two small ball probabilities. At first, we will provide an argument according to which it is sufficient to check (P1) at points located in $\mathcal{F}^{(k)}$. We want to show that, for every $r > 0$ and $x \in \mathcal{F}$ that satisfies $d_2(m, x) \geq 4r$, there exists $x' \in \mathcal{F}^{(k)}$ that also satisfies $d_2(m, x') \geq 4r$ such that

$$\varphi_{x'}^{P_k^{(1)}}(r) \geq \varphi_{x}^{P_k^{(1)}}(r) \tag{4.9}$$

holds.

Therefore, let $r > 0$ and $x = \sum_{j=1}^{\infty} \theta_j\nu_j \in \mathcal{F} \setminus \mathcal{F}^{(k)}$ such that $d_2(x, m) \geq 4r$. Observe that $m \in \mathcal{F}^{(k)}$ for every $k \in \mathbb{N}$. Define by $\tilde{x} = \sum_{j=1}^{k} \theta_j\nu_j$ the orthogonal projection of x

onto the subspace $\mathcal{F}^{(k)}$, such that

$$d_2^2(x,\tilde{x}) = \sum_{j=k+1}^{\infty} \theta_j^2 = \inf_{w\in\mathcal{F}^{(k)}} d_2^2(x,w) =: z^2$$

holds. If $z > r$, then $B_{d_2}(x,r) \cap \mathcal{F}^{(k)} = \emptyset$ and, thus, $p_k\big(B_{d_2}(x,r)\big) = \emptyset$, from which it follows that $\varphi_x^{P_k^{(1)}}(r) = 0$, so let us assume that $0 < z \le r$. Define

$$x' := m + \frac{\|x-m\|_2}{\|\tilde{x}-m\|_2}(\tilde{x}-m) \in \mathcal{F}^{(k)}.$$

One can see that $\|x'-m\|_2 = \|x-m\|_2 \ge 4r$. Additionally, since $x \notin \mathcal{F}^{(k)}$, it follows that $\|x-m\|_2 > \|\tilde{x}-m\|_2$. We can derive that

$$\begin{aligned}\|\tilde{x}-x'\|_2 &= \left(\frac{\|x-m\|_2}{\|\tilde{x}-m\|_2} - 1\right)\|\tilde{x}-m\|_2 = \|x-m\|_2 - \|\tilde{x}-m\|_2 \\ &= \|x-m\|_2 - \sqrt{\|x-m\|_2^2 - z^2} \le 4r - \sqrt{(4r)^2 - z^2},\end{aligned}$$

where, in the third identity, we applied the Pythagorean theorem (which is valid in any Hilbert space), which gives

$$\|\tilde{x}-m\|_2^2 = \|x-m\|_2^2 - \|x-\tilde{x}\|_2^2 = \|x-m\|_2^2 - z^2.$$

The final inequality uses that $y \mapsto y - \sqrt{y^2 - z^2}$ is decreasing on the interval (z,∞), $z \le r$ and $\|x-m\|_2 \ge 4r$. We now claim that

$$B_{d_2}(x,r) \cap \mathcal{F}^{(k)} \subset B_{d_2}(x',r). \tag{4.10}$$

Let $x'' \in B_{d_2}(x,r) \cap \mathcal{F}^{(k)}$. We can again use the Pythagorean theorem to deduce

$$\|\tilde{x}-x''\|_2^2 \le r^2 - \|x-\tilde{x}\|_2^2 = r^2 - z^2,$$

which then yields

$$d_2(x',x'') \le d_2(x'',\tilde{x}) + d_2(x',\tilde{x}) \le \sqrt{r^2-z^2} + d_2(\tilde{x},x') \le \sqrt{r^2-z^2} + 4r - \sqrt{(4r)^2-z^2} \le r,$$

because one can verify that $\sqrt{(4r)^2-z^2} - \sqrt{r^2-z^2} \ge 3r$, if $z \in (0,r]$. Using (4.10) yields

$$p_k\big(B_{d_2}(x,r)\big) \subseteq B_{d_{Euc}}\big((\theta_1',\ldots,\theta_k'),r\big) \Longrightarrow \varphi_x^{P_k^{(1)}}(r) \le \varphi_{x'}^{P_k^{(1)}}(r),$$

where $(\theta_1', \dots, \theta_k')$ are the first k Fourier coefficients of x' (the remaining ones being zero). This completes the proof of (4.9). It implies that it suffices to check the condition (P1) (for both $P_k^{(1)}$ and $P_k^{(2)}$) at points $x' \in \mathcal{F}^{(k)}$ because if $x' \notin \mathcal{F}^{(k)}$, then there exists some point in $\mathcal{F}^{(k)}$ that has the same distance towards m as x' and attains a small ball probability that is greater than or equal to the one at x' (both evaluated for that one particular radius). For the sake of brevity we will write

$$\varphi_x^{f_k}(\varepsilon) := \varphi_{\sum_{j=1}^k x_j \nu_j}^{P_k^{(1)}}(\varepsilon) = \int\limits_{B_{d_{Euc}}(x,\varepsilon)} f_k(x)\, \lambda\!\!\lambda^k(dx), \quad \forall x = (x_1, \dots, x_k) \in \mathbb{R}^k, \forall \varepsilon > 0,$$

from now on, and use $d = d_{Euc}$ to denote the k-dimensional Euclidean metric. We will show that (P1) holds for $P_k^{(1)}$ by distinguishing between four cases. Before we start with our case analysis we will fix some $r \in (0, l_k]$ and derive

$$\begin{aligned}
\varphi_{m^{(k)}}^{f_k}(r) &= \int\limits_{B_{d_{Euc}}(m^{(k)},r)} f_k^{(1)}(x)\, \lambda\!\!\lambda^k(dx) + \int\limits_{B_{d_{Euc}}(m^{(k)},r)} f_k^{(2)}(x)\, \lambda\!\!\lambda^k(dx) \\
&= \frac{1}{4\gamma \cdot v_k(l_k)} \int\limits_{B_{d_{Euc}}(0,r)} \left(1 - \left\| \frac{x}{l_k} \right\|^\gamma \right) \lambda\!\!\lambda^k(dx) + \frac{v_k(r)}{v_k(q_k)} \left(1 - \frac{1}{4(k+\gamma)}\right) \cdot \frac{1}{2^k} \\
&= \frac{s_{k-1}}{4\gamma \cdot v_k(l_k)} \int_0^r y^{k-1} \left(1 - \left(\frac{y}{l_k}\right)^\gamma \right) dy + \frac{v_k(r)}{v_k(q_k)} \left(1 - \frac{1}{4(k+\gamma)}\right) \cdot \frac{1}{2^k} \\
&= \frac{1}{4\gamma \cdot (l_k)^k} \left[y^k - \frac{k}{k+\gamma} \frac{y^{k+\gamma}}{(l_k)^\gamma} \right]_0^r + \frac{v_k(r)}{v_k(q_k)} \left(1 - \frac{1}{4(k+\gamma)}\right) \cdot \frac{1}{2^k} \\
&= \frac{1}{4\gamma} \left(\frac{r}{l_k}\right)^k \left(1 - \frac{k}{k+\gamma} \left(\frac{r}{l_k}\right)^\gamma \right) + \left(\frac{r}{q_k}\right)^k \left(1 - \frac{1}{4(k+\gamma)}\right) \cdot \frac{1}{2^k}.
\end{aligned}$$

This yields $\varphi_{m^{(k)}}^{f_k}(l_k) > \frac{1}{4(k+\gamma)}$ and shows that $q_k \nu_1 \in \operatorname{supp}(P_k^{(1)})$. Now let us consider several cases. Firstly, let $r > q_k$ and $x' \in \mathbb{R}^k$ such that $d(m^{(k)}, x') \geq 4r > 4q_k$, which means $\varphi_{x'}^{f_k}(r) = 0$, so there is nothing to show.

Let us now assume that $r \in [l_k, q_k]$ and $d(x', m^{(k)}) \geq 4r \geq 4l_k$. In any of our cases, we can assume that $d(x', m^{(k)})$ is sufficiently small such that $\varphi_{x'}^{f_k}(r) > 0$, because otherwise the inequality is trivial. Let us deduce

$$\varphi_{x'}^{f_k}(r) \leq \frac{v_k(r)}{v_k(q_k)} \cdot \left(1 - \frac{1}{4(k+\gamma)}\right) \cdot \frac{1}{2^k} = \left(\frac{r}{q_k}\right)^k \cdot \left(1 - \frac{1}{4(k+\gamma)}\right) \cdot \frac{1}{2^k}.$$

We can use that to derive

$$\frac{\varphi^{f_k}_{m^{(k)}}(r)}{\varphi^{f_k}_{x'}(r)} \geq \frac{\varphi^{f_k}_{m^{(k)}}(l_k)}{\varphi^{f_k}_{x'}(r)} > \frac{\frac{1}{4(k+\gamma)}}{\left(\frac{r}{q_k}\right)^k \cdot \left(1 - \frac{1}{4(k+\gamma)}\right) \cdot \frac{1}{2^k}} \geq \frac{2^k}{4(k+\gamma)-1},$$

where we used $r \leq q_k$ in the final step. Since $\frac{2^k}{4(k+\gamma)-1} \overset{k\to\infty}{\longrightarrow} \infty$, we will eventually have

$$\frac{\varphi^{f_k}_{m^{(k)}}(r)}{\varphi^{f_k}_{x'}(r)} > \frac{2^k}{4(k+\gamma)-1} > 1 + C_1\Big(d(m^{(k)}, x') - r\Big)^\gamma,$$

because $\left|d(m^{(k)}, x') - r\right| \leq 5q_k \overset{k\to\infty}{\longrightarrow} 0$ holds for every x' and r that fall into this case. The scenario where $r \in [l_k/3, l_k)$ and $d(x', m^{(k)}) \geq 4r \geq 4l_k/3$ can be handled in a very similar way, because we have

$$\frac{\varphi^{f_k}_{m^{(k)}}(r)}{\varphi^{f_k}_{x'}(r)} \geq \frac{\left(\frac{r}{l_k}\right)^k \frac{1}{4\gamma}\left(1 - \frac{k}{k+\gamma} \cdot \left(\frac{r}{l_k}\right)^\gamma\right)}{\left(\frac{r}{q_k}\right)^k \cdot \left(1 - \frac{1}{4(k+\gamma)}\right) \cdot \frac{1}{2^k}} \geq \frac{\left(\frac{r}{l_k}\right)^k \frac{1}{4\gamma} \cdot \frac{\gamma}{k+\gamma}}{\left(\frac{r}{q_k}\right)^k \cdot \left(1 - \frac{1}{4(k+\gamma)}\right) \cdot \frac{1}{2^k}} = \left(\frac{q_k}{l_k}\right)^k \cdot \frac{2^k}{4(k+\gamma)-1},$$

as $q_k/l_k \overset{k\to\infty}{\longrightarrow} \infty$.

Let us finally fix some $r \in (0, l_k/3)$ and $x' \in \mathbb{R}^k$ such that $d(x', m^{(k)}) \geq 4r$. Then we have

$$\begin{aligned} \varphi^{f_k}_{x'}(r) &\leq v_k(r) \cdot f_k\Big((q_k - (d(x', m^{(k)}) - r), 0, \ldots, 0)\Big) \quad \text{and} \\ \varphi^{f_k}_{m^{(k)}}(r) &\geq v_k(r) \cdot f_k\Big((q_k - r, 0, \ldots, 0)\Big), \end{aligned}$$

as, by the construction of f_k, we have

$$\sup_{x \in B_{d_{Euc}}(x', r)} f_k(x) \leq f_k\Big((q_k - (d(x', m^{(k)}) - r), 0, \ldots, 0)\Big)$$

and

$$\inf_{x \in B_{d_{Euc}}(m^{(k)}, r)} f_k(x) = f_k\Big((q_k - r, 0, \ldots, 0)\Big).$$

We will divide this final section into two further cases and now assume that $d(x', m^{(k)}) - r \geq l_k$. Then we have

$$\frac{\varphi^{f_k}_{m^{(k)}}(r)}{\varphi^{f_k}_{x'}(r)} \geq \frac{f_k\Big((q_k - r, 0, \ldots, 0)\Big)}{f_k\Big((q_k - (d(x', m^{(k)}) - r), 0, \ldots, 0)\Big)}$$

$$= \frac{\frac{1}{4\gamma\cdot v_k(l_k)}\left(1-\left(\frac{r}{l_k}\right)^\gamma\right)+\frac{1}{v_k(q_k)}\cdot\left(1-\frac{1}{4(k+\gamma)}\right)\cdot\frac{1}{2^k}}{\frac{1}{v_k(q_k)}\cdot\left(1-\frac{1}{4(k+\gamma)}\right)\cdot\frac{1}{2^k}}$$

$$\geq 1+\frac{1}{4\gamma}\left(\frac{q_k}{l_k}\right)^k\cdot\frac{4(k+\gamma)}{4(k+\gamma)-1}\cdot 2^k\cdot\left(1-\left(\frac{1}{3}\right)^\gamma\right)\overset{k\to\infty}{\longrightarrow}\infty,$$

which again means that, for large k, we have

$$\frac{\varphi^{f_k}_{m^{(k)}}(r)}{\varphi^{f_k}_{x'}(r)}\geq 1+C_1\big(d(x',m^{(k)})-r\big)^\gamma$$

for every $x'\in\mathbb{R}^k$ so that $d(x',m^{(k)})\geq 4r$. Let us now assume that $d(x',m^{(k)})-r<l_k$ and set $a_k:=4\gamma\cdot\frac{v_k(l_k)}{v_k(q_k)}\cdot\left(1-\frac{1}{4(k+\gamma)}\right)\cdot\frac{1}{2^k}$. This yields

$$\begin{aligned}
\frac{\varphi^{f_k}_{m^{(k)}}(r)}{\varphi^{f_k}_{x'}(r)} &\geq \frac{\frac{1}{4\gamma\cdot v_k(l_k)}\left(1-\left(\frac{r}{l_k}\right)^\gamma\right)+\frac{1}{v_k(q_k)}\cdot\left(1-\frac{1}{4(k+\gamma)}\right)\cdot\frac{1}{2^k}}{\frac{1}{4\gamma\cdot v_k(l_k)}\left(1-\left(\frac{d(x',m^{(k)})-r}{l_k}\right)^\gamma\right)+\frac{1}{v_k(q_k)}\cdot\left(1-\frac{1}{4(k+\gamma)}\right)\cdot\frac{1}{2^k}}\\
&= \frac{1-\left(\frac{r}{l_k}\right)^\gamma+a_k}{1-\left(\frac{d(x',m^{(k)})-r}{l_k}\right)^\gamma+a_k}\\
&= \frac{1-\left(\frac{d(x',m^{(k)})-r}{l_k}\right)^\gamma+a_k+\left(\frac{d(x',m^{(k)})-r}{l_k}\right)^\gamma-\left(\frac{r}{l_k}\right)^\gamma}{1-\left(\frac{d(x',m^{(k)})-r}{l_k}\right)^\gamma+a_k}\\
&= 1+\frac{\left(\frac{d(x',m^{(k)})-r}{l_k}\right)^\gamma-\left(\frac{r}{l_k}\right)^\gamma}{1-\left(\frac{d(x',m^{(k)})-r}{l_k}\right)^\gamma+a_k}\\
&> 1+\frac{\left(\frac{d(x',m^{(k)})-r}{l_k}\right)^\gamma-\left(\frac{1}{3}\right)^\gamma\left(\frac{d(x',m^{(k)})-r}{l_k}\right)^\gamma}{1+a_k}\\
&= 1+\frac{1-\frac{1}{3^\gamma}}{1+a_k}\cdot\frac{1}{(l_k)^\gamma}\cdot\Big(d(x',m^{(k)})-r\Big)^\gamma.
\end{aligned}$$

The last inequality is valid due to $\frac{d(x',m^{(k)})-r}{l_k}<1$ and $r\leq\frac{1}{3}\big(d(x',m^{(k)})-r\big)$. Since $a_k\overset{k\to\infty}{\longrightarrow}0$ and $(l_k)^{-\gamma}\overset{k\to\infty}{\longrightarrow}\infty$ we eventually have

$$\frac{1-\frac{1}{3^\gamma}}{1+a_k}\cdot(l_k)^{-\gamma}\geq C_1$$

for k large. Recall that C_1 has always been the constant from our condition (P1) (see

Definition 4.2). Hence, for k large, $P_k^{(1)}$ satisfies (P1) and so does $P_k^{(2)}$ by analogous computations.
To prove that (P2) holds it suffices to find a single cover point located in $E(r^{1+\beta})$ for every $r \in (0, \eta]$ such that the desired inequality and distance relation hold and some probability is positive (see Definition 4.2). We will find out that there always exists such a point in $E(r^{1+\beta}) \cap \mathcal{F}^{(k)}$. Firstly, let $r \in (0, \min(r_{n_k}, \eta)]$ (see properties (1)-(3) at the beginning of the proof). Then, by property (2), we have $q_k\nu_1 \in E(r_{n_k}) \cap W(s, C)$ and, hence, by Proposition 4.10 and condition (C4), $q_k\nu_1 \in E(r)$, which implies $q_k\nu_1 \in E(r^{1+\beta})$, because $\eta < 1$. Hence, we can choose $y(r) = q_k\nu_1$ in (P2), which is the function that is equal to the mode itself. Thus, the quotient of probabilities in (P2) admits the value 1.
If $(r_{n_k}, \eta] = \emptyset$, then there is nothing more to show. In the other case, if $r \in (r_{n_k}, \eta]$, then we can set $y(r) = 0 \in E(r^{1+\beta})$. Since we have that $r_{n_k} \geq 2q_k$ by property (3), we can conclude that $d_2(y(r), q_k\nu_1) = q_k < r$. The desired relation follows as $\varphi_0^{P_k^{(1)}}(r) \geq \varphi_0^{P_k^{(1)}}(2q_k) = 1$, because that means that the quotient on the left side of (P2) attains an upper bound of 1. Due to the symmetry relations between $P_k^{(1)}$ and $P_k^{(2)}$ and the fact that for every $r > 0$ we have $E(r) = -E(r)$, our considerations also serve as a proof of the claim that $P_k^{(2)}$ satisfies (P2).
In an effort to verify (P3) for our distributions we will firstly target the case of $r \in (0, l_k]$. We already know that

$$\begin{aligned}\varphi^{P_k^{(1)}}_{\mathrm{Mod}(P_k^{(1)})}(r) = \varphi^{f_k}_{m^{(k)}}(r) &= \frac{1}{4\gamma}\cdot\left(\frac{r}{l_k}\right)^k\left(1-\frac{k}{k+\gamma}\left(\frac{r}{l_k}\right)^\gamma\right)+\left(\frac{r}{q_k}\right)^k\left(1-\frac{1}{4(k+\gamma)}\right)\cdot\frac{1}{2^k}\\ &\geq \left(\frac{r}{l_k}\right)^k\cdot\frac{1}{4(k+\gamma)}.\end{aligned}$$

Let us consider the function $t\colon(0, l_k] \to (0, \infty), r \mapsto r^k \exp\left(\frac{K}{r^{1/s}}\right)$. We have that

$$\frac{d}{dr}t(r) = r^{k-\frac{1}{s}-1}\exp\left(\frac{K}{r^{1/s}}\right)\left(kr^{1/s}-\frac{K}{s}\right)$$

and deduce that t is falling on its entire domain, because $kr^{1/s} - \frac{K}{s} \leq 0$ holds at every $r \in (0, l_k]$. This means that, for every $r \in (0, l_k]$, we have $t(r) \geq (l_k)^k \exp\left(\frac{K}{(l_k)^{1/s}}\right)$.

We derive that

$$\begin{aligned}
\varphi_{m^{(k)}}^{f_k}(r) &\geq \left(\frac{r}{l_k}\right)^k \frac{1}{C_3} \exp\left(\frac{K}{r^{1/s}}\right) \cdot \frac{1}{4(k+\gamma)} \cdot C_3 \exp\left(-\frac{K}{r^{1/s}}\right) \\
&\geq \frac{1}{C_3} \exp\left(\frac{K}{(l_k)^{1/s}}\right) \cdot \frac{1}{4(k+\gamma)} \cdot C_3 \exp\left(-\frac{K}{r^{1/s}}\right) \\
&= \frac{1}{C_3} \frac{1}{4(k+\gamma)} \cdot \exp(sk) \cdot C_3 \exp\left(-\frac{K}{r^{1/s}}\right).
\end{aligned}$$

Since $\frac{1}{C_3}\frac{1}{4(k+\gamma)} \cdot \exp(sk) \overset{k\to\infty}{\longrightarrow} \infty$, we have just verified that (P3) holds for every $r \in (0, l_k]$ if k is sufficiently large. Note that $l_k \overset{k\to\infty}{\longrightarrow} 0$. Hence, for k large, we have $l_k < \eta$. If $r \in (l_k, \eta]$, then $\varphi_{m^{(k)}}^{f_k}(r) \geq \frac{1}{4(k+\gamma)}$. Since $\varphi_{m^{(k)}}^{f_k}(3q_k) = 1$, we must only handle the case where $r \in (l_k, \min(3q_k, \eta)]$. We deduce that

$$C_3 \exp\left(-K\big(3q_k'\big)^{-1/s}\right) = C_3 \exp\left(-K\frac{\ln(8C_3k)}{K}\right) = \frac{1}{8k} < \frac{1}{4(k+\gamma)}$$

for k sufficiently large. Hence, we can finalise our computations by pointing out that

$$\begin{aligned}
\varphi_{m^{(k)}}^{f_k}(r) \geq \frac{1}{4(k+\gamma)} > \frac{1}{8k} = C_3 \exp\left(-K\big(3q_k'\big)^{-1/s}\right) &\geq C_3 \exp\left(-K\big(3q_k\big)^{-1/s}\right) \\
&\geq C_3 \exp\left(-Kr^{-1/s}\right),
\end{aligned}$$

because by property (1) we have $q_k \leq q_k'$. Thus, the inequality (P3) holds for every $r \in (0, \eta]$. The same computations can be carried out for $P_k^{(2)}$.

This shows that there exists some $k' \in \mathbb{N}$ such that $P_k^{(1)}, P_k^{(2)} \in \mathcal{P}$ for every $k \geq k'$. This completes the proof.

■

We want to note that the distributions $P_k^{(1)}$ and $P_k^{(2)}$ still satisfy our condition (P2) if the covers of the system $\mathcal{E}$ are replaced by some smaller covers, which is due to the fact that the two modes are contained in the truncated ellipsoid $\Theta_1(s, C)$ (see the proof of Proposition 4.10). The following theorem summarises our previous consideration.

Theorem 4.12. *Let $s \in \mathbb{N}, C > 0$ and set $\mathcal{Y} = W(s, C)$ (see Definition 4.8) and $\mathcal{C} = \mathcal{E} = \big(E(r)\big)_{r>0}$ (see (4.7)). Then there exist constants $0 < L' < L$ and some $\rho > 0$ such that the pair $(\mathcal{Y}, \mathcal{C})$ satisfies property (E') for the parameters $\alpha = 1/s, \rho > 0$ and the constants $L, L' > 0$. Let $\beta > 1/s, \gamma > 0$ and $\eta \in \big(0, \min(\rho, 1)\big)$ be parameters and let $K \in (0, L]$ denote a constant. Then let*

$$\mathcal{P} := \mathcal{P}_{\alpha,\beta,\gamma,\eta,K}(\mathcal{Y}, \mathcal{C}) = \mathcal{P}_{1/s,\beta,\gamma,\eta,K}(W(s, C), \mathcal{E})$$

denote the class of probability distributions on $\big(L^2\big([0,1]\big), \mathbb{B}\big(L^2\big([0,1]\big)\big)\big)$ that satisfy the axioms (P0)-(P3) (see Definition 4.2). Then, for $p = 2$, $(\rho_n)_{n\in\mathbb{N}}$, where $\rho_n = \ln(n)^{-2s}$, is the optimal rate of convergence of the mode estimation problem over the class $\mathcal{P}$.

Proof: The properties of the pair $(\mathcal{Y}, \mathcal{C}) = (W(s, C), \mathcal{E})$ can all be derived from Proposition 4.10. Now both (1) and (2) in our Definition 2.18 of the optimal rate of convergence can be checked with the help of our Theorems 4.6 (where $\alpha = 1/s$) and 4.11. Recall that, by our definition, the optimal rate of convergence is unique up to a constant.

■

By Theorem 4.12, the main goal of this thesis has been achieved as we have established the optimal convergence rate. As discussed previously (see the text following Theorem 4.6), the order of the rate merely depends on the parameter $\alpha = 1/s$. E.g., the larger the integer s, the higher the order up to which derivatives of the elements of $\mathcal{Y}$ (which include the mode) exist. Consequently, the greater s the smaller size of the minimum covers of $\mathcal{Y}$ and, hence, the greater the asymptotic lower bound for the small ball probability function of the mode, which has been the crucial quantity in our asymptotic analysis.
The logarithmic order of the rate indicates slow convergence speed such that even for a large number of observations, the (expected) estimation error may still be considerable. However, in nonparametric functional data analysis, it is not unusual to obtain logarithmic rates, e.g. the rates derived by Meister (2016) for the nonparametric estimation of a regression function or the estimation of a classifier (where both are based on functional observations) are also logarithmic. Ferraty and Vieu (2006) also deduce logarithmic bounds for the maximum risk of their functional kernel estimators of the condition mode, conditional quantiles and a regression function (see Proposition 13.5 therein). It is due to the fact that the observations take on values in an infinite-dimensional (function) space that algebraic convergence rates of the order n^{-c}, where $c > 0$, are usually unattainable.

Chapter 5

Discussion

Our approach to estimate the mode of a random function differs from the papers contributed by other authors, e.g. by Ferraty and Vieu (2006) or Dabo-Niang et al. (2010), in the constraints imposed on the probability distribution. The main distinction in our stipulations is the absence of a probability density function, while instead we resort to imposing conditions on the small ball probabilities. One advantage therein is that defining small ball probabilities is always possible, whereas densities of distributions on arbitrary Polish metric spaces only exist with respect to abstract reference measures. By using the Cameron-Martin formula we can precisely define a density of a Gaussian measure with respect to the shifted measure where the shift is an element of the Cameron-Martin space. In a general setting, it is unclear whether there exists a density that additionally fulfills smoothness or boundedness requirements that are often needed to achieve rates of convergence. The small ball problem has been extensively studied in many exemplary settings that have significant relevance in the fields of applied statistics and we can refer to our Section 3.2.2 for a small collection of the results on the problem of Gaussian small balls in Hilbert and Banach spaces. We can also refer to Aurzada and Dereich (2009) for an extension of some of these results to general Lévy-processes (or their induced probability measures, respectively). However, we want to emphasise that in a general setting, accessing the small ball probabilities, by which we mean giving the exact asymptotics or deriving bounds, remains a difficult task.

The definition of the mode is quite general in a sense that regularity of the distribution at the mode (e.g. the continuity of a density) is not necessarily needed. Hence, our Theorem 3.25 and Corollary 3.26 extend the already existing consistency results in the literature that rely on such regularity assumptions, e.g. the smoothness of a density at the mode.

One of the reasons why the second part of this thesis dealing with estimation is divided into two further parts with two separate estimators is that we believe that deducing sharp lower bounds for the minimax risk under the assumption (A2) from Section 3.4.2 is significantly more challenging than under our stipulations (P0)-(P3). For instance, none of the probability measures that were constructed in the proof of Theorem 4.11 satisfy (A2) as they are defined by the mean of a Lebesgue density function on a finite-dimensional subspace. Consequently, the order at which their small ball probabilities at the mode tend to zero is polynomial and not exponential as is required in (A2). It is an open problem whether our approach in Section 3.4.2 can be extended in a way such that the optimal rate of convergence can be deduced. Defining distributions with exponential small ball probabilities at the mode will require techniques that differ from the ones we employed in the proof of Theorem 4.11.
In contrast to (A2), the condition (P3) in Chapter 4 only contains an exponential lower bound for the small ball function at the mode. Additionally, inequalities for the quotient of two probabilities were imposed in (P1) and (P2). In Chapter 4, the dependency of both the estimator and the distribution class on a system of covers was crucial. It is an interesting incentive for further research on the mode estimation problem to extend the convergence results to a setting where $\mathcal{Y}$ is a subset of a non-Hilbert space $\mathcal{F}$ or where the dependence on covers can be entirely eliminated in favour of a more general approach.
The upper bound on the maximum risk which we give in Theorem 4.6 holds under the assumption that the radii (or bandwidth) parameter is asymptotically bounded from both sides. That raises the question of how to select it in a way such that error measures like the mean squared or integrated squared error become small. The bandwidth analysis has been omitted in this thesis and is a promising field for future investigations from both theoretical and practical viewpoints. For the kernel density estimator in finite dimension there exists a variety of techniques to select the bandwidth parameter such as plug-in or cross-validation methods. Some of these become applicable to the functional equivalent of the kernel estimator (given by Ferraty and Vieu (2006) or Dabo-Niang et al. (2010)) if additional hypotheses on the distribution are formulated. It is an open yet interesting question if the (functional) kernel bandwidth leads to small errors when estimating the mode or whether new techniques must be developed.

Implementing our first estimator from Section 3.4.1 to apply our theory to real samples can be easily done, as we only need to compute the distances between the data points and make comparisons of the amounts of data in certain balls. Due to its reliance on a family of covers, applying our second estimator is more complicated. If covers can be explicitly stated, as, for instance, is the case for a Sobolev ellipsoid, then implementations are possible, albeit receiving numerical results will usually be more time-consuming as the cover changes with the radius. We have already elaborated at the end of the preceding chapter that due to the slow, logarithmic convergence rate one should expect a large number of observations is required in order to receive satisfying results. First and foremost, the focus of this thesis has been the study of asymptotic properties of mode estimators, which can be used to explain phenomena that occur in the field of numerical applications.

Bibliography

[1] Abraham, C., G. Biau and B. Cadre: *Simple estimation of the mode of a multivariate density*, The Canadian Journal of Statistics, Volume 31, pp. 23-34, 2003.

[2] Abraham, C., G. Biau and B. Cadre: *On the asymptotic properties of a simple estimate of the mode*, ESAIM: Probability and Statistics, Volume 8, pp. 1-11, 2004.

[3] Aurzada, F. and S. Dereich: *Small deviations of general Lévy processes*, The Annals of Probability, Volume 37, pp. 2066-2092, 2009.

[4] Berghin, L., Y. Nikitin and E. Orsingher: *Exact small ball constants for some Gaussian processes under the L^2-norm*, Journal of Mathematical Sciences, Volume 128, pp. 2493-2502, 2005.

[5] Bogachev, V.I.: *Gaussian measures*, American Mathematical Society, Mathematical Surveys and Monographs, Volume 62, reprinted edition, 2015.

[6] Borovkov, A.A. and A.A. Mogulski: *On probabilities of small deviations for stochastic processes*, Siberian Advances in Mathematics, Volume 1, pp. 39-63, 1991.

[7] Chapados, C. and D.J. Levitin: *Cross-modal interactions in the experience of musical performances: physiological correlates*, Cognition, Volume 108, pp. 639-651, 2008.

[8] Dabo-Niang, S., F. Ferraty and P. Vieu: *Estimation du mode dans un espace vectoriel semi-normé*, Comptes rendus de l'Académie des sciences, Volume 339, pp. 652 - 662, 2004.

[9] Dabo-Niang, S., A. Yao, L. Pischedda, P. Cuny and F. Gilbert: *Spatial mode estimation for functional random fields with application to bioturbation problem*, Stochastic Environmental Research and Risk Assessment, pp. 487-497, 2010.

[10] Deheuvels, P. and G.V. Martynov: *A Karhunen-Loeve decomposition of a Gaussian process generated by independent pairs of exponential random variables*, Journal of Functional Analysis, Volume 255, pp. 2363-2394, 2008.

[11] Delaigle, A. and P. Hall: *Defining probability density for a distribution of random functions*, The Annals of Statistics, Volume 38, pp. 1171-1193, 2010.

[12] Dieudonné, J.: *Grundzüge der modernen Analysis*, VEB Deutscher Verlag der Wissenschaften, Berlin, second edition, 1972.

[13] Donoho, D. L. and R.C. Liu: *Geometrizing rates of convergence III*, The Annals of Statistics, Volume 19, pp. 668-701, 1991.

[14] Dunker, T., M.A. Lifshits and W. Linde: *Small deviation probabilities of sums of independent random variables*, High dimensional probability, Progress in Probability, Volume 43, pp. 59-74, 1998.

[15] Eddy, W.: *Optimal kernel estimators of the mode*, The Annals of Statistics, Volume 8, pp. 870-882, 1980.

[16] Feller, W.: *An introduction to probability theory and its applications, Volume 2*, Wiley Series in Probability and Mathematical Statistics, New York, second edition, 1971.

[17] Ferraty, F. and P. Vieu: *Nonparametric functional data analysis*, Springer Series in Statistics, New York, 2006.

[18] Gao, F., J. Hannig, T. Lee and F. Torcaso: *Laplace transforms via Hadamard factorization*, Electronic Journal of Probability, Volume 8, Paper no. 13, pp. 1-20, 2003a.

[19] Gao, F., J. Hannig and F. Torcaso: *Comparison theorems for small deviations of random series*, Electronic Journal of Probability, Volume 8, Paper no. 21, pp. 1-17, 2003b.

[20] Gao, F. and J.A. Wellner: *Entropy estimate for high dimensional monotonic functions*, Journal of Multivariate Analysis, Volume 98, pp. 1751-1764, 2007.

[21] Gasser, T.: *Nonparametric estimation of the mode of a distribution of random curves*, Journal of the Royal Statistical Society, Series B (Statistical Methodology), Volume 60., pp. 681-691, 1998.

[22] Heinonen, J., P. Koskela, N. Shanmugalingam and J.T. Tyson: *Sobolev spaces on metric measure spaces*, Cambridge University Press, Cambridge, 2015.

[23] Herrmann, E. and K. Ziegler: *Rates of consistency for nonparametric estimation of the mode in absence of smoothness assumptions*, Statistics & Probability Letters, Volume 68, pp. 359-368, 2004.

[24] Hoeffding, W.: *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association, Volume 58, pp. 13-30, 1963.

[25] Hogg, R.V., J.W. McKean and A.T. Craig: *Introduction to mathematical statistics*, Pearson, Boston, seventh edition, 2012.

[26] Hyndman, R.J. and H.L. Shang: *Rainbow plots, bagplots and boxplots for functional data*, Journal of Computational and Graphical Statistics, Volume 19, pp. 29-45, 2010.

[27] Kendall, M.G., A. Stuart and J.K. Ord: *Kendall's advanced theory of statistics, Volume 1, Distribution theory*, Griffin, London, fifth edition 1987.

[28] Klemelä, J.: *Adaptive estimation of the mode of a multivariate density*, Journal of Nonparametric Statistics, Volume 17, pp. 83-105, 2005.

[29] Kolmogorov, A.N. and V.M. Tikhomirov: *ε-entropy and ε-capacity of sets in function spaces*, American Mathematical Society Translations, Volume 17, pp. 277-364, 1961.

[30] Laukaitis, A.: *Functional data analysis for cash flow and transactions intensity continuous-time prediction using Hilbert-valued autoregressive processes*, European Journal of Operational Research, Volume 185, pp. 1607-1614, 2008.

[31] Lewin, J.: *An interactive introduction to mathematical analysis*, Cambridge University Press, Cambridge, 2003.

[32] Li, W. V. and Q. Shao: *Gaussian Processes: Inequalities, small ball probabilities and applications*, Stochastic Processes: Theory and Methods. Handbook of Statistics, Volume 19., pp. 533-598, Elsevier, New York, 2001.

[33] Meister, A.: *On general consistency in deconvolution mode estimation*, Journal of Statistical Planning and Inference, Volume 141, pp. 771-781, 2011.

[34] Meister, A.: *Optimal classification and nonparametric regression for functional data*, Bernoulli, Volume 22, pp. 1729-1744, 2016.

[35] Milbrodt, H.: *Wahrscheinlichkeitstheorie - Eine Einführung mit Anwendungen und Beispielen aus der Versicherungs- und Finanzmathematik*, Verlag Versicherungswirtschaft, Karlsruhe, 2010.

[36] Mood, A.M., D. C. Boes and F. A. Graybill: *Introduction to the theory of statistics*, McGraw-Hill Education, New York, third edition, 1974.

[37] Parzen, E.: *On estimation of probability density function and mode*, The Annals of Mathematical Statistics, Volume 33, pp. 1065-1076, 1962.

[38] Perdices, D., J.E.L. de Vergara and J. Ramos: *Deep FDA: Using functional data analysis and neural networks to characterize network services time series*, IEEE Transactions on Network and Service Management, Volume 18, pp. 986-999, 2021.

[39] Rajput, B.S. and S. Cambanis: *Gaussian processes and Gaussian measures*, The Annals of Mathematical Statistics, Volume 43, pp. 1944-1952, 1972.

[40] Ramsay, J.O. and B.W. Silverman: *Funtional data analysis*, Springer, New York, second edition, 2005.

[41] Rao, A.R., Q. Wang, H. Wang, H. Khorasgani and C. Gupta: *Spatio-temporal functional neural networks*, IEEE 7th International Conference on Data Science and Advanced Analytics (Sydney), pp. 81-89, 2020.

[42] Ross, S.: *A first course in probability*, Pearson, Boston, tenth edition, 2019.

[43] Rudin, W.: *Real and complex analysis*, McGraw-Hill Education, New York, third edition, 1987.

[44] Sager, T.W.: *Estimation of a multivariate mode*, The Annals of Statistics, Volume 6, pp. 802-812, 1978.

[45] Sager, T.W.: *An iterative method for estimating a multivariate mode and isopleth*, Journal of the American Statistical Association, Volume 74, pp. 329-339, 1979.

[46] Shi, X., Y. Wu and B. Miao: *A note on the convergence rate of the kernel density estimator of the mode*, Statistics & Probability Letters, Volume 79, pp. 1866-1871, 2009.

[47] Stromberg, K.R.: *An introduction to classical real analysis*, Chapman and Hall, London, 1981.

[48] Tsybakov, A.B.: *Recursive estimation of the mode of a multivariate density*, Problems of Information Transmission, Volume 26, pp. 31-37, 1990.

[49] Tsybakov, A.B.: *Introduction to nonparametric estimation*, Springer Series in Statistics, second (English) printing, New York, 2008.

[50] Ullah, S. and C.F. Finch: *Applications of functional data analysis: a systematic review*, BMC Medical Research Methodology, Volume 13, Article 43, 2013.

[51] van der Vaart, A.W. and J.A. Wellner: *Weak convergence and empirical processes - With applications to Statistics*, Springer Series in Statistics, New York, 1996.

[52] Vieu, P.: *A note on density mode estimation*, Statistics & Probability Letters, Volume 26, pp. 297-307, 1996.

[53] Wainwright, M.J.: *High-dimensional Statistics: A non-asymptotic viewpoint*, Cambridge Series in Statistical and Probabilistic Mathematics, Volume 48, Cambridge, 2019.

[54] Willard, S.: *General Topology*, Addison-Wesley Publishing Company, Reading, 1970.

[55] Witting, H. and U. Müller-Funk: *Mathematische Statistik II*, B. G. Teubner, Stuttgart, 1995.