DISSERTATION

APPLICATIONS OF LEAST SQUARES PENALIZED SPLINE DENSITY ESTIMATOR

Submitted by

Hanxiao Jing

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2024

Doctoral Committee:

Advisor: Mary Meyer

Daniel Cooley
Piotr Kokoszka
Joshua Berger

ABSTRACT

APPLICATIONS OF LEAST SQUARES PENALIZED SPLINE DENSITY ESTIMATOR

The spline-based method stands as one of the most common nonparametric approaches. The work in this dissertation explores three applications of the least squares penalized spline density estimator. Firstly, we present a novel hypothesis test against the unimodality of density functions, based on unimodal and bimodal estimates of the density function, using penalized splines. The test statistic is the difference in the least-squares criterion, between these fits. The distribution of the test statistics under the null hypothesis is estimated via simulated data sets from the unimodal fit. Large sample theory is derived and simulation studies are conducted to compare its performance with other common methods across various scenarios, alongside a real-world application involving neuro-transmission data from guinea pig brains. Secondly, we tackle the deconvolution density estimation problem, introducing the penalized splines deconvolution estimator. Building upon the results gained from piecewise constant splines, we achieve a cube-root convergence rate for piecewise quadratic splines and uniform errors. Moreover, we derive large sample theories for the penalized spline estimator and the constrained spline estimator. Simulation studies illustrate the competitive performance of our estimators compared to the kernel estimators across diverse scenarios. Lastly, drawing inspiration from the preceding applications, we develop a hypothesis test to discern whether the underlying density is unimodal or multimodal, given data with measurement error. Under the assumption of uniform errors, we introduce the test and derive the test statistic. Simulations are conducted to show the performance of the proposed test under different conditions.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Mary Meyer. Working under her guidance has been an invaluable experience, and I am immensely thankful for her continuous support, guidance, and encouragement.

I extend my heartfelt appreciation to my committee members, Dr. Daniel Cooley, Dr. Piotr Kokoszka and Dr. Joshua Berger, for their valuable insights, constructive feedback, and dedication to guiding me through this academic endeavor.

My deepest gratitude goes to my family - my dad, my mom, and my grandmother - for their endless love and unconditional support. Their belief in my abilities and encouragement have been a constant source of strength and motivation.

I am deeply thankful to my dear friends who have been by my side throughout this journey. Special thanks to Wenqin, whose friendship and support have always been inspiring and comforting during challenging times. I am also grateful for the endless joy and comfort provided by my friends' adorable pets. Their presence has brought warmth and laughter into my life.

I am profoundly grateful to everyone who has contributed to my personal growth and supported my mental well-being, whether through meaningful conversations, acts of kindness, or simply being there for me during difficult times.

I would like to express my appreciation to Fort Collins, my home during my doctoral studies, for its picturesque landscapes, serene surroundings, and vibrant community. The lakes on the mountains, the swings, the benches, and the sprawling lawns in the parks, and the beauty of springtime have been constant sources of peace, joy and inspiration. I am grateful for every moment and every experience that has accompanied me during my PhD journey.

Thank you to everyone who has played a part in shaping my doctoral experience and for enriching my life in countless ways.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Nonparametric Density Estimation

Density estimation is a widely adopted tool in statistics. Consider a set of independent and identically distributed (i.i.d.) observations $X_1, \ldots, X_n$ with an unknown density function f. In order to estimate f, nonparametric approaches are often employed. This technique finds widespread applications across various fields, including statistics, machine learning, econometrics, and signal processing. Over the years, various methods have been developed and refined in this field. One of the earliest and mostly widely used methods is kernel density estimation. Introduced by [Parzen, 1962] and further studied by [Silverman, 1978]. The essence of kernel density estimation lies in smoothing the empirical distribution of the data using a kernel function, typically a symmetric, non-negative function centered at each data point. Commonly used kernel functions include the Gaussian, Epanechnikov, and uniform kernels, each with its own properties and suitability for different types of data. The choice of kernel function and bandwidth significantly impacts the resulting density estimate. A smaller bandwidth results in a more localized and detailed estimate, whereas a larger bandwidth produces a smoother but potentially oversmoothed estimate. Selecting an appropriate bandwidth is crucial for obtaining an accurate and reliable density estimate. Cross-validation techniques are commonly used to tune the bandwidth parameter and assess the performance of the kernel density estimator.

Histogram-based density estimation [Freedman and Diaconis, 1981] is one of the simplest and oldest nonparametric methods. The basic idea is to divide the data range into a set of equally spaced intervals (bins), and the density is estimated based on the frequency of observations within each bin. While histograms are intuitive and computationally efficient, they may produce biased estimates, especially when the number of bins or bin width is not appropriately chosen. Various modifications and adaptions, such as adaptive histograms [Birgé, 1987], kernel density histograms

[Rudemo, 1982], and Bayesian histograms [Leonard, 1973] have been proposed to address these limitations.

The nonparametric maximum likelihood estimation approach was pioneered by [Goodd and Gaskins, 1971] and [de Montricher et al., 1975]. The key idea is to find the density that maximizes the likelihood of observing the data, subject to certain constraints to ensure that the resulting estimate is a valid density function.

Spline-based methods offer a flexible framework for nonparametric density estimation by fitting piecewise polynomial functions to the data, which allows splines to capture complex shapes in the distribution. B-splines [De Boor, 1972], natural splines [Lyche and Schumaker, 1973], and penalized splines [Wegman and Wright, 1983] are among the commonly used spline basis functions. One of the key strengths of spline-based density estimation methods is the ability to easily incorporate shape constraints, for example, unimodality. Once the spline basis functions are constructed, estimating the density involves solving an optimization problem, such as least squares or maximum likelihood problem. The shape constraints can be modeled as either equality or inequality constraints, or both. Solving the optimization problem with these constraints is relatively straightforward.

In this dissertation, we focus on the spline method and its application in nonparametric density estimation. We also compare its performance with kernel density estimation in various scenarios.

1.2 Penalized Spline Density Estimation

Suppose the underlying density f is known to be smooth and we estimate f by constructing a spline density basis $\delta_1, \ldots, \delta_m$. We first determine a support $[S_1, S_2]$ for our estimated density and knots $t_1, \ldots, t_{m_t}$ distributed over the support. The selection of knots can vary depending on the specific requirements of the analysis. We describe two methods for selecting knots in this dissertation, see details of knots based on sample quantiles in Chapter 2 and equality spaced knots in Chapter 3. The number of knots, denoted as m_t, is determined based on the properties of the spline basis functions we use. For example, if the δ_j are B splines, the number of knots $m_t = m-1$.

Figure 1.1 shows an example of B splines with degree 2 on equality spaced knots $0, 1, \ldots, 10$. The choice of spline basis is flexible. For example, if we want to use B splines to estimate a density function that is known to go to zero at the ends of the support, there's no need to include the first and last B spline basis since both are not zero at the endpoint of the support. In the example shown in Figure 1.1, B_1 and B_{12} can be excluded while estimating the density if needed.

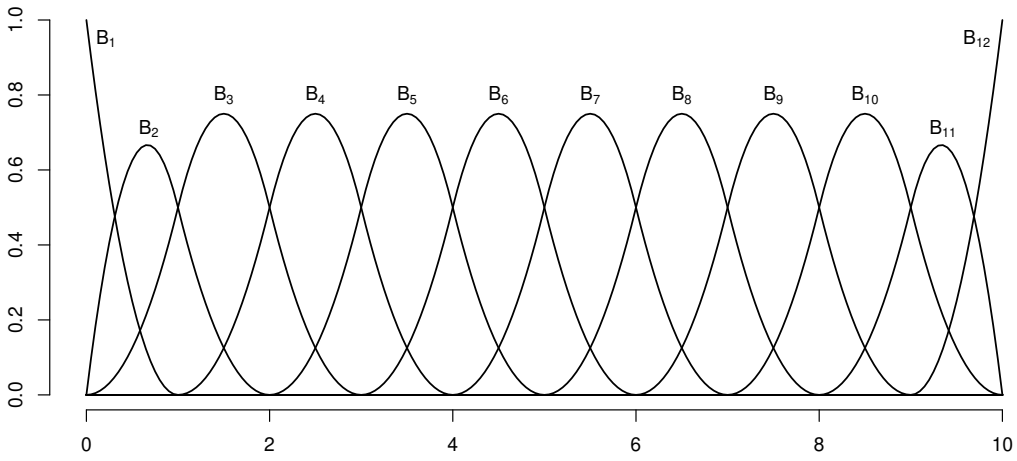

Figure 1.1: An example of B splines with degree 2 on knots 0, 1, ..., 10

We use a linear combination of basis functions $\sum_{j=1}^{m} b_j \delta_j(x)$ to estimate f. The goal is to find an estimator $\tilde{f}$ of the form

$$\tilde{f}(x) = \sum_{j=1}^{m} \tilde{b}_j \delta_j(x), \text{ such that } \tilde{f}(x) \geq 0 \text{ for all } x, \text{ and } \boldsymbol{a}^\top \tilde{\boldsymbol{b}} = 1,$$

where $\boldsymbol{b} = (b_1, \ldots, b_m)^\top$. Ensuring that the area under the estimated density to be one is accomplished by constraining $\boldsymbol{a}^\top \tilde{\boldsymbol{b}} = 1$, where $a_j = \int \delta_j(x) dx$ is the area under $\delta_j(x)$ with $\boldsymbol{a} = (a_1, \ldots, a_m)^\top$. The least-squares criterion of [Groeneboom et al., 2001] is used. We minimize

$$\psi(f; \boldsymbol{X}) = \int f(x)^2 dx - \frac{2}{n} \sum_{i=1}^{n} f(X_i) \tag{1.1}$$

over the set of spline densities f with the desired shape. To accomplish this, define the $m \times m$ matrix $\boldsymbol{H}$ as $H_{j\ell} = \int_{-\infty}^{\infty} \delta_j(x)\delta_\ell(x)dx$, and $\boldsymbol{c}$ as the vector in $\mathbb{R}^m$ with $c_j = \sum_{i=1}^{n} \delta_j(X_i)/n$. The least-squares criterion can be written in terms of the coefficients of the spline basis, resulting in the quadratic programming problem

$$\text{minimize } \boldsymbol{b}^\top \boldsymbol{H}\boldsymbol{b} - 2\boldsymbol{c}^\top \boldsymbol{b}, \text{ subject to } \boldsymbol{a}^\top \boldsymbol{b} = 1 \text{ and } \boldsymbol{A}\boldsymbol{b} \geq \boldsymbol{0},$$

where the constraint matrix $\boldsymbol{A}$ defines the desired shape such as unimodal or bimodal, and constrains the density to be non-negative. Minimizing the criterion function under the equality and inequality constraints is accomplished with the `R` function `solveQP` in the package `quadprog`. The minimizer $\tilde{\boldsymbol{b}}$ provides the constrained density estimator $\tilde{f}(x) = \sum_{j=1}^{m} \tilde{b}_j \delta_j(x)$. Figure 1.2 shows an example of the estimated density with data from a bimodal density with sample size $n = 1000$.

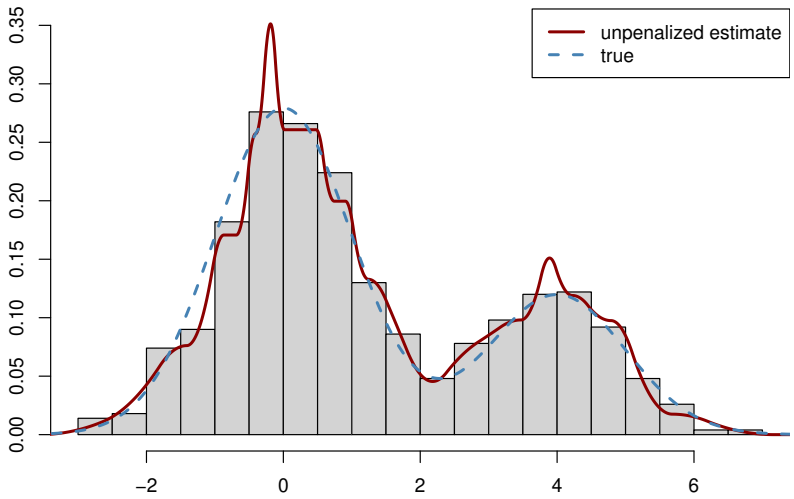

Figure 1.2: Example of data from a bimodal distribution with sample size $n = 1000$, and the corresponding unpenalized B splines density estimate

The estimate in Figure 1.2 appears wiggly and lacks smoothness. To achieve a smooth estimate, a penalty term can be applied to penalize fluctuations in the estimate. This dissertation primarily focuses on scenarios where the spline basis functions are piecewise polynomial with a degree no greater than 2. In the cases described above, the second derivative f'' is piecewise constant, so we employ a penalty that penalizes the consecutive differences in the second derivative. Let $\theta_i = f''(x) = \sum_{j=1}^{m} b_j \delta_j''(x)$ for $x \in (t_i, t_{i+1}), i = 1, \ldots, m_t - 1$, and define

$$\psi_\lambda(f; \boldsymbol{X}) = \int f(x)^2 dx - \frac{2}{n} \sum_{i=1}^{n} f(X_i) + \lambda \sum_{i=1}^{m+1} (\theta_{i+1} - \theta_i)^2,$$

where λ is the penalty parameter.

Let $\boldsymbol{D}^1$ be a $(m-2) \times (m-1)$ matrix such that $D^1_{i,i} = -1, D^1_{i,i+1} = 1$ for $i = 1, \ldots, m_t - 1$, and $\boldsymbol{D}^2$ be the second derivative matrix of the basis function, so $D^2_{i,j} = \delta_j''(x)$ for $x \in (t_i, t_{i+1})$ Then we have

$$\boldsymbol{D}^1 \boldsymbol{D}^2 \boldsymbol{b} = \begin{pmatrix} \theta_2 - \theta_1 \\ \theta_3 - \theta_2 \\ \vdots \\ \theta_{m_t - 1} - \theta_{m_t - 2} \end{pmatrix}.$$

To make sure that the $\boldsymbol{H}$ and the penalty matrix change by the same proportion when the data are re-scaled, for example measured in a different unit, we apply a multiple to $\boldsymbol{D}$. Let $\boldsymbol{D} = d^{5/2} \boldsymbol{D}^1 \boldsymbol{D}^2$ where $d = (t_{m_t} - t_1)/m$; we have a scale-invariant estimator if the penalty matrix is $\boldsymbol{D}^\top \boldsymbol{D}$. Now the criterion with penalty can be written as

$$\text{minimize } \boldsymbol{b}^\top (\boldsymbol{H} + \lambda \boldsymbol{D}^\top \boldsymbol{D}) \boldsymbol{b} - 2\boldsymbol{c}^\top \boldsymbol{b}, \text{ subject to } \boldsymbol{a}^\top \boldsymbol{b} = 1 \text{ and } \boldsymbol{A}\boldsymbol{b} \geq \boldsymbol{0}. \tag{1.2}$$

The penalized estimate $\hat{f}(x) = \sum_{i=1}^{m} \hat{b}_j \delta_j(x)$. [Chen and Meyer, 2023] showed that the optimal convergence rate for $\hat{f}$, in the L_2 norm as n increases, is achieved with m on the order of $n^{1/7}$ and penalty parameter λ on the order of $n^{-1/7}$. λ is chosen by cross-validation, see more details in

Chapter 2 and Chapter 3. Figure 1.3 shows the penalized density estimation using the same data in Figure 1.2.

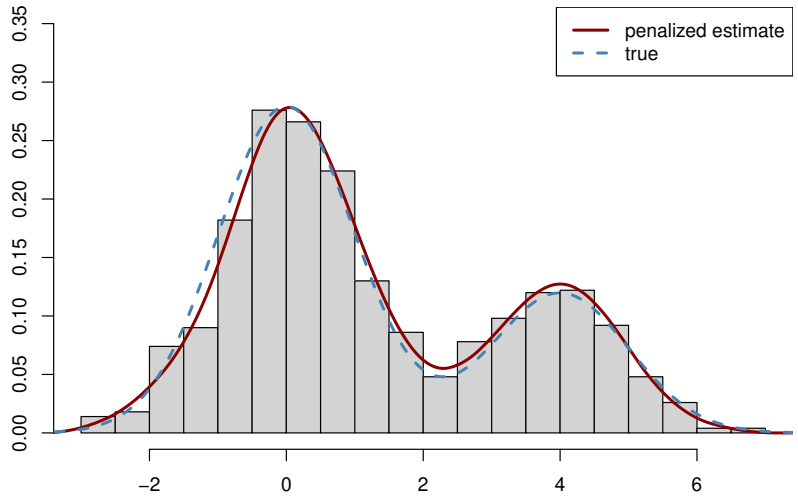

Figure 1.3: The penalized B splines density estimate with the data of 1.2

The remainder of the dissertation unfolds as follows. In Chapter 2, we present a hypothesis test against the unimodality of density functions. Our approach involves getting penalized unimodal and bimodal density estimations, from which a test statistic is computed. The distribution of the test statistics under the null hypothesis is obtained via bootstrapping techniques. The validity of our method is supported by theoretical analyses, and simulation studies are conducted to evaluate its performance across various scenarios. Additionally, we provide a real-world application involving neurotransmission data from guinea pig brains to showcase the practical utility of our methodology. Moving on to Chapter 3, we study the deconvolution problem, introducing a penalized splines deconvolution estimator. Leveraging quadratic splines and assuming uniform errors, we achieve a cube-root convergence rate. Through simulation studies, we demonstrate the competitive performance of our estimators compared to kernel estimators across diverse scenarios. In Chapter 4, we integrate the methodologies developed in Chapters 2 and 3. Specifically, we investigate hypothesis

testing against unimodality for deconvolution problems, using results gained from both chapters to address this novel problem domain.

Chapter 2

Modality Test using Constrained Penalized Splines

This chapter is dedicated to addressing the problem of determining the modality of a distribution given a random sample. Specifically, the interest is in whether the distribution is unimodal or multi-modal. A novel test is proposed, based on unimodal and bimodal estimates of the density function, using penalized splines with a least-squares criterion. The test statistic is the difference in the least-squares criterion, between the unimodal and bimodal fits. The null distribution of the test statistic is estimated with simulated data sets from the unimodal density estimate. We review the constrained spline density estimation in Section 2.2. In Section 2.3, we derive the test statistic and provide relevant large sample theory. Furthermore, the simulations conducted in Section 2.5 demonstrate that the proposed test has higher power than competitors for some simple examples. Finally, we present a real data application in Section 2.6 involving neuro-transmission data from guinea pig brains, thereby showcasing the practical utility of our methodology.

2.1 Introduction and Background

Suppose $X_1, \ldots, X_n$ are independent continuous random variables with common density function f. Interest is in testing the null hypothesis that f is unimodal, versus the alternative that f has more than one mode. [Haldane, 1951] introduced a simple test for detecting bimodality in a distribution. This test relies on the consecutive discrepancies of frequencies for adjacent categories within a sample frequency distribution. In a similar vein, [Larkin, 1979] devised an algorithm tailored to data arranged in an integer array, in a histogram format. [Silverman, 1981] proposed a method to increase the bandwidth of a kernel estimator until the resulting estimator is unimodal. The smallest bandwidth that produces a unimodal fit is the test statistic; its null distribution is determined by simulations. This is one of the most commonly used and studied modality tests. [Mammen et al., 1992] and [Hall and York, 2001] provided more details and large sample theory. The test is implemented in the `R` package `multimode`. [Müller and Sawitzki, 1991] developed a method

based on the excess mass functional which was studied further by [Cheng and Hall, 1998]. The dip test, based on the distance between the empirical distribution and the unimodal distribution closest to it, was proposed by [Hartigan and Hartigan, 1985]. The RUNT test developed by [Hartigan and Mohanty, 1992] used single linkage clusters. The MAP test by [Rozál and Hartigan, 1994] is based on minimal constrained spanning tree. [Minnotte, 1997] proposed a method to test each individual observed modes of the data.

2.2 Penalized Density Estimate with Constrained B-Splines

Suppose we have i.i.d. observations $X_1, \ldots, X_n$ having density f that is known to be smooth and have a shape such as unimodality. Section 1.2 provides a least-squares constrained penalized spline estimate of the density. The estimator proposed in this chapter is similar. We use B-spline basis functions with degree two for this problem. We first determine a support $[S_1, S_2]$ for our estimated density. The chosen support is defined as $(\tilde{q}_{.01} - 0.4(\tilde{q}_{.99} - \tilde{q}_{.01}), \tilde{q}_{.99} + 0.4(\tilde{q}_{.99} - \tilde{q}_{.01}))$, where $\tilde{q}_p$ is the pth quantile of the sample, $p \in (0, 1)$. We use q_p for the population pth quantile. The selection of knots $t_1, \ldots, t_{m-1}$ is based on the sample quantiles, with additional knots placed around the antimode (the local minimum between modes) and at the edges of the support if needed. We strategically place additional knots around the antimode if the low density value leads to sparse quantiles in that region to ensure a more accurate estimation, particularly in intervals where the density undergoes a change in direction. Let $B_1(x), \ldots, B_m(x)$ be the B-spline basis functions using the knots above. We use a linear combination of basis functions $\sum_{j=1}^{m} b_j B_j(x)$ to estimate f. The aim is to find

$$\tilde{f}(x) = \sum_{j=1}^{m} \tilde{b}_j B_j(x), \text{ such that } \tilde{f}(x) \geq 0 \text{ for all } x, \text{ and } \int_{S_1}^{S_2} \tilde{f}(x)dx = 1.$$

Let $a_j = \int_{S_1}^{S_2} B_j(x)dx$ be the area under $B_j(x)$ with $\boldsymbol{a} = (a_1, \ldots, a_m)^\top$. For convenience, we replace $B_j(x)$ with $B_j(x)/a_j$, so that the areas under each $B_j(x)$ are all one. Hence, ensuring that the area under the estimated density to be one is accomplished by constraining $\sum_{j=1}^{m} b_j = 1$, and

shape constraints such as unimodal or bimodal, as well as positivity, are imposed as linear inequality constraints on the spline basis coefficients, in the form of $\boldsymbol{A}\boldsymbol{b} \geq \boldsymbol{0}$. The least-squares criterion introduced in 1.1 is used. Similarly, with the $m \times m$ matrix $\boldsymbol{H}$ as $H_{j\ell} = \int_{-\infty}^{\infty} B_j(x)B_\ell(x)dx$, and $\boldsymbol{c}$ as the vector in $\mathbb{R}^m$ with $c_j = \sum_{i=1}^{n} B_j(X_i)/n$, the least-squares criterion can be written as

$$\text{minimize } \boldsymbol{b}^\top \boldsymbol{H}\boldsymbol{b} - 2\boldsymbol{c}^\top \boldsymbol{b}, \text{ subject to } \sum_{j=1}^{m} b_j = 1 \text{ and } \boldsymbol{A}\boldsymbol{b} \geq \boldsymbol{0}.$$

The constrained density estimator is $\tilde{f}(x) = \sum_{j=1}^{m} \tilde{b}_j B_j(x)$, where $\tilde{\boldsymbol{b}} = (\tilde{b}_1, \ldots, \tilde{b}_m)^\top$ is the minimizer.

Notice f'' is piecewise constant when the B-spline basis functions $B_j(x)$ have degree 2, so we can penalize the consecutive differences in the second derivative as 1.2. We denote the penalized estimate as $\hat{f}(x) = \sum_{i=1}^{m} \hat{b}_j B_j(x)$. The choice of the penalty parameter is introduced in next section.

2.3 Construction of the Test Statistic

Recall the we're interested in the hypothesis

$$H_0 : f \text{ has one mode} \quad \text{vs.} \quad H_A : f \text{ has more than one mode.}$$

With knots $t_1, \ldots, t_{m-1}$, we will obtain a density estimate with unimodal constraints, denoted as $\hat{f}_1$, and another with bimodal constraints, denoted as $\hat{f}_2$. We initially start with $\hat{f}_1$. The unimodal estimation is achieved by constructing constraint matrices $\boldsymbol{A}_1^{(i)}$. Each $\boldsymbol{A}_1^{(i)}$ constrains the density to be unimodal, with the mode between t_i and t_{i+1}. Each element of $\boldsymbol{A}_1^{(i)}$ corresponds to the slopes of the basis functions at the knots, with a positive sign for knots below the mode and a negative sign for knots above the mode. We select the index $i = i_1, \ldots, i_2$ to minimize the criterion function, where $i_1 = \min(i : t_i \geq \tilde{q}_{0.1})$ and $i_2 = \max(i : t_i < \tilde{q}_{0.9}) - 1$. The range of i is to make sure that the mode is between $\tilde{q}_{0.1}$ and $\tilde{q}_{0.9}$. Let $\mathcal{C}^{(i)} = \{\boldsymbol{b} : \boldsymbol{A}_1^{(i)}\boldsymbol{b} \geq 0 \text{ and } \sum_{j=1}^{m} b_j = 1\}$, the unimodal fit

$\hat{f}_1(x) = \sum_{i=1}^{m} \hat{b}_i B_i(x)$ is obtained by minimizing

$$\boldsymbol{b}^\top(\boldsymbol{H} + \lambda \boldsymbol{D}^\top \boldsymbol{D})\boldsymbol{b} - 2\boldsymbol{c}^\top \boldsymbol{b}, \text{ subject to } \boldsymbol{b} \in \bigcup_{i=i_1}^{i_2} \mathcal{C}_1^{(i)}.$$

Similarly, we can get the bimodal density estimation $\hat{f}_2$. To impose bimodal constraints on the density estimation, we identify three knot intervals where the monotonicity constraints change sign, so that the density is increasing to the first knot interval, decreasing after, then increasing, then decreasing again. These intervals are used to construct bimodal constraint matrices $A_2^{(i,j,k)}$. With $i_1 \leq i < j < k \leq i_2$, the first mode is situated between t_i and t_{i+1}, the antimode between t_j and t_{j+1}, and the second mode between t_k and t_{k+1}. Additionally, we ensure that the modes and the antimode are positioned at least one knot apart from each other, specifically setting $j > i + 1$ and $k > j + 1$. The triplet (i, j, k) is chosen to minimize the criterion and obtain the bimodal density estimation. Figure 2.1 shows an example of the unimodal and bimodal fits of a sample that comes from a mixture of two normal distributions $0.4N(0, 1) + 0.6N(3.8, 2.4^2)$ with sample size 800.

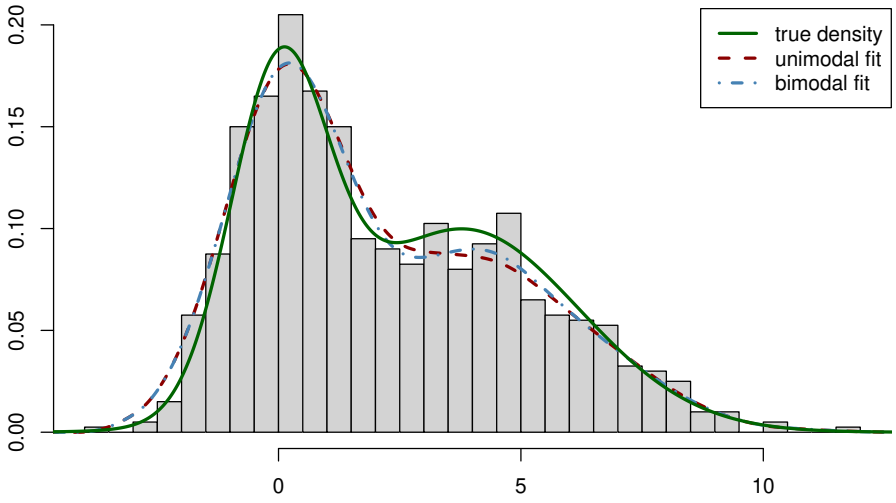

Figure 2.1: Example of data simulated from a mixture of normal densities, with $n = 800$ and the corresponding unimodal and bimodal fits.

The bimodal estimate $\hat{f}_2$ might actually be unimodal with a "flat spot"; in this case we interpret this as supporting evidence for the null hypothesis, so we accept the null hypothesis. We refer this as the "auto-acceptance" rule in the following contents. If $\hat{f}_2$ has two distinct modes, define the test statistic

$$T = \psi_\lambda(\hat{f}_1, \boldsymbol{X}) - \psi_\lambda(\hat{f}_2, \boldsymbol{X}).$$

To find an optimal penalty parameter λ, we think about the concept of kurtosis. Sample kurtosis serves as a statistic to quantify the 'tailedness' of the data. When the true distribution is highly peaked or 'pointy', it suggests that we should opt for a smaller λ, as a large λ in such cases can lead to an over-smoothed density estimation with a shorter peak. Conversely, if the true distribution is not particularly 'pointy,' a small λ may lead to an under-smoothed estimate. To strike a balance between these considerations, we set

$$\lambda = \begin{cases} n^{-1/7}10^2 & K \leq 2 \\ n^{-1/7}10^{4-K} & 2 \leq K < 5 \\ n^{-1/7}10^{3/2-K/2} & 5 \leq K < 9 \\ n^{-1/7}10^{-3} & K \geq 9 \end{cases}.$$

And K is the sample kurtosis of the data. This choice ensures that the penalty is appropriate for the the shape of the data.

To approximate the distribution of the test statistic under H_0, we generate $i = 1, \ldots, B$ bootstrap samples from $\hat{f}_1$. Let $\hat{f}_{1,i}$ and $\hat{f}_{2,i}$ be the estimates with unimodal and bimodal constraints, respectively, for the ith bootstrap sample. For each bootstrap sample, if the bimodal fit $\hat{f}_{2,i}$ turns to be unimodal, we consider this as having less evidence against the null hypothesis than the data set, and we set the corresponding bootstrapped statistic $T_i = -\infty$, otherwise, $T_i = \psi_\lambda(\hat{f}_{1,i}, \boldsymbol{X_i}) - \psi_\lambda(\hat{f}_{2,i}, \boldsymbol{X_i})$.

The null is rejected when

$$\sum_{i=1}^{B} I(T_i > T)/B < \alpha,$$

where α is the level of significance. At times, particularly when the true density is bimodal, the unimodal estimate $\hat{f}_1$ may have a "flat spot" around the mode it fails to detect. In such cases, if we sample bootstrapped samples from this unimodal fit with a "flat spot", the bootstrapped estimates might show more evidence for the null hypothesis. To address this potential bias, we introduce additional constraints for the unimodal fit. Specifically, for knot intervals that are neither around the mode nor at the tails, we constrain the slope to be strictly non-zero by setting its absolute value greater than or equal to σ. The additional constraints ensure that the unimodal fit can only display a 'flat spot' at the mode or the tails. The choice of the bound σ for the slope is determined based on the sample skewness and the support:

$$\sigma = \begin{cases} 2n^{-2/7}/(S_2 - S_1)^2 & |\text{skewness} > 0.7| \\ 5n^{-2/7}/(S_2 - S_1)^2 & \text{o.w.} \end{cases} .$$

The $(S_2 - S_1)^2$ in the demominator ensures that the σ remains the same when the data are re-scaled. Employing a relatively smaller σ for skewed data aims to prevent over-constraining the slope of intervals where the density is 'flatter'. In the above example shown in Figure 2.1, our method results in a p-value of 0.021, while the kernel method [Silverman, 1981] has a p-value of 0.376. Since the true density is bimodal, our method results in the right decision at $\alpha = 0.05$,while the kernel method leads to a type II error.

2.4 Large Sample Theory

In this section, we show that if the true density f_0 is unimodal with mode μ, then the test size converges to (at most) the target test size α as n gets large. Further, we show that if the true f_0 has two distinct modes, then the power increases to one as n increases. We show that if f_0 has no "flat spots" then the test size goes to zero as n increases, for any target α.

Theorem 1. *Suppose f_0 is "strictly unimodal" with mode μ, so that $f_0''(\mu) < 0$, $f_0'(x) > 0$ for $x \in (s_1, \mu)$, $f_0'(x) < 0$ for $x \in (\mu, s_2)$. Further, $\mu \in (q_{0.1}, q_{0.9})$ and for any min$(|f_0'(q_{0.1})|, |f_0'(q_{0.9})|) > \epsilon > 0$, there is a δ such that $f'(x) < -\epsilon$ for $x \in (\mu + \delta, q_{0.9})$, $f'(x) > \epsilon$ for $x \in (q_{0.1}, \mu - \delta)$, and*

$|f'(x)| \leq \epsilon$ *for* $x \in (\mu - \delta, \mu + \delta)$. *Then the probability of rejecting* H_0 *goes to zero as* n *increases.*

Proof: First we show as $\epsilon \to 0$, $\delta \to 0$ as well. For all $x \in (\mu - \delta, \mu + \delta)$, using Taylor's Theorem we have

$$f_0'(x) = f_0'(\mu) + f_0''(\mu)(x - \mu) + h(x)(x - \mu) = f_0''(\mu)(x - \mu) + h(x)(x - \mu),$$

where $\lim_{x \to \mu} h(x) = 0$. So $-\epsilon \leq f_0''(\mu)(x-\mu) + h(x)(x-\mu) \leq \epsilon$ for all $x \in (\mu - \delta, \mu + \delta)$. Then when $\epsilon \to 0$, δ has to converge to zero too. Let $\tilde{f}$ be the estimate without modality constraint. Next we can show that $\tilde{f}$ is unimodal with probability approaching one as the sample size increases, by contradiction. Suppose $\tilde{f}$ is not unimodal. Given that the basis splines are quadratic B-splines, $\tilde{f}$ can change direction at most once between knots. Consequently, the first estimated mode, the estimated antimode, and the second estimated mode are separated by at least one knot. Thus, no matter which knots interval the true mode μ is at, it is at least one knot away from one of the estimated modes. So, there exists an interval $[t_k, t_{k+1}]$ at least one knot away from the mode μ such that f_0 is monotonic while $\tilde{f}$ is increasing on $[t_k, \nu]$ and decreasing on $[\nu, t_{k+1}]$ for some ν in $[t_k, t_{k+1}]$. Without loss of generality, we assume f_0 is decreasing. Now we consider interval $[t_{k-1}, \nu]$, within which f_0 is decreasing and $\tilde{f}$ is non-decreasing. We write $\tilde{f} = g + C$, where $\int_{t_{k-1}}^{\nu} g = 0$ and C is a constant. First we show that for any C, $\int_{t_{k-1}}^{\nu} (g + C - f_0)^2 dx$ is minimized over non-decreasing functions when g is a constant function. We can write the expression as

$$\begin{aligned}\int_{t_{k-1}}^{\nu} (g(x) + C - f_0(x))^2 dx &= \int_{t_{k-1}}^{\nu} (C - f_0(x))^2 dx + \int_{t_{k-1}}^{\nu} g(x)^2 dx + 2\int_{t_{k-1}}^{\nu} g(x)(C - f_0(x))dx \\ &= \int_{t_{k-1}}^{\nu} (C - f_0(x))^2 dx + \int_{t_{k-1}}^{\nu} g(x)^2 dx - 2\int_{t_{k-1}}^{\nu} g(x) f_0(x) dx.\end{aligned}$$

The first term is a constant. Using mean value theorem, we know there is a $c_1 \in [t_{k-1}, \nu]$ such that $g(c_1) = 0$, further $g(x) \leq 0$ on $[t_{k-1}, c_1]$ and $g(x) \geq 0$ on $[c_1, \nu]$. So we have

$$\begin{aligned}\int_{t_{k-1}}^{\nu} g(x)f_0(x)dx &= \int_{t_{k-1}}^{c_1} g(x)f_0(x)dx + \int_{c_1}^{\nu} g(x)f_0(x)dx \\ &\leq \int_{t_{k-1}}^{c_1} g(x)f_0(c_1)dx + \int_{c_1}^{\nu} g(x)f_0(c_1)dx \\ &= 0.\end{aligned}$$

So the last term $-2\int_{t_{k-1}}^{\nu} g(x)f_0(x)dx$ is non-negative. To minimize (2.4), $g(x)$ has to be 0 in $[t_{k-1}, \nu]$. Next, it's easy to find the minimizer $C = \int_{t_{k-1}}^{\nu} f_0(x)/(\nu - t_{k-1})$. Hence the minimized

$$\int_{t_{k-1}}^{\nu} (\tilde{f}(x) - f_0(x))^2 dx = \int_{t_{k-1}}^{\nu} (g(x) + C - f_0(x))^2 dx = \int_{t_{k-1}}^{\nu} (C - f_0(x))^2 dx.$$

To bound the integral from below, notice that there exist $c_2 \in [t_{k-1}, \nu]$ such that $f_0(c_2) = C$. Now consider $f_1(x) = C - \epsilon(x - c_2)$, which is a line passing the point (c_2, C) with slope $-\epsilon$. Then we have

$$\int_{t_{k-1}}^{\nu} (C - f_1(x))^2 dx \geq \int_{t_{k-1}}^{\nu} [C - (C - \epsilon(x - c_2))]^2 dx = \int_{t_{k-1}}^{\nu} [\epsilon(x - c_2)]^2 dx$$

And let $d = m - t_{k-1}$,

$$\begin{aligned}\int_{t_{k-1}}^{\nu} (x - c_2)^2 dx &= \frac{1}{3}[(\nu - c_2)^3 + (c_2 - t_{k-1})^3] = \frac{1}{3}[(\nu - c_2)^3 + (d - (\nu - c_2))^3] \\ &= d(\nu - c_2)^2 - d^2(\nu - c_2) + \frac{1}{3}d^3.\end{aligned}$$

has minimized value $d^3/12$ when $(\nu - c_2) = d/2$. Since $f_0'(x) > -\epsilon$ on $[t_{k-1}, \nu]$, we know that $f_0 \geq f_1$ on $[t_{k-1}, c_2]$ and $f_0 \leq f_1$ on $[c_1, \nu]$. So, $\|\tilde{f} - f_0\|^2 \geq \|\tilde{f} - f_1\|^2 \geq \epsilon d^3/12$ is on the order of $n^{-3/7}$ when d is on the order of $n^{-1/7}$. In other words, if $\tilde{f}$ is bimodal infinitely often,

$\|\tilde{f} - f_0\|^2$ is on the order of $n^{-3/7}$ with. probability approaching one , which is contradictory to $\tilde{f}$ being the unconstrained estimator with $\|\tilde{f} - f_0\|^2 = O_p(n^{-6/7})$. So, $\tilde{f}$ is unimodal with probability approaching one as n increases. Recall that $\tilde{f}$ is the minimizer without modality constraints, since $\tilde{f}$ is unimodal with probability approaching one, it is the minimizer under bimodal constraints as well. Therefore, the bimodal fit $\hat{f}_2$ has to be unimodal with probability approaching one. According to the decision making rule, the null hypothesis is accepted.

Theorem 2. *If f_0 has more than one mode, the probability of rejecting the null hypothesis goes to one as n increases.*

Proof: Suppose f_0 has more than one mode. To obtain the sampling distribution of the test statistics under the null hypothesis, bootstrapped samples are generated from the unimodal estimate $\hat{f}_1$. As per the proof of Theorem 1, given that the data are sampled from a strictly unimodal density $\hat{f}_1$, each bimodal estimate $\hat{f}_{2,i}$ from the bootstrapped samples will turn out to be unimodal. Consequently, this results in a bootstrapped test statistic of $-\infty$, thereby yielding a p-value approaching zero and leading to the rejection of the null hypothesis with a probability approaching one.

2.5 Simulations

2.5.1 Simulation 1

We first design a simulation study to compare the performance of our method, Silverman's kernel method and the Dip test. Figure 2.2 plots the 8 distributions used in this simulation study. The study involves a range of distributions, symmetric or skewed, unimodal or multimodal. For each distribution, we generate 1000 samples of sample size $n = 50, 100, 200, 500, 1000$. For each sample, three methods are applied to get three p-values. Table 2.1 shows the proportion of rejection for each distribution shown in Figure 2.2.

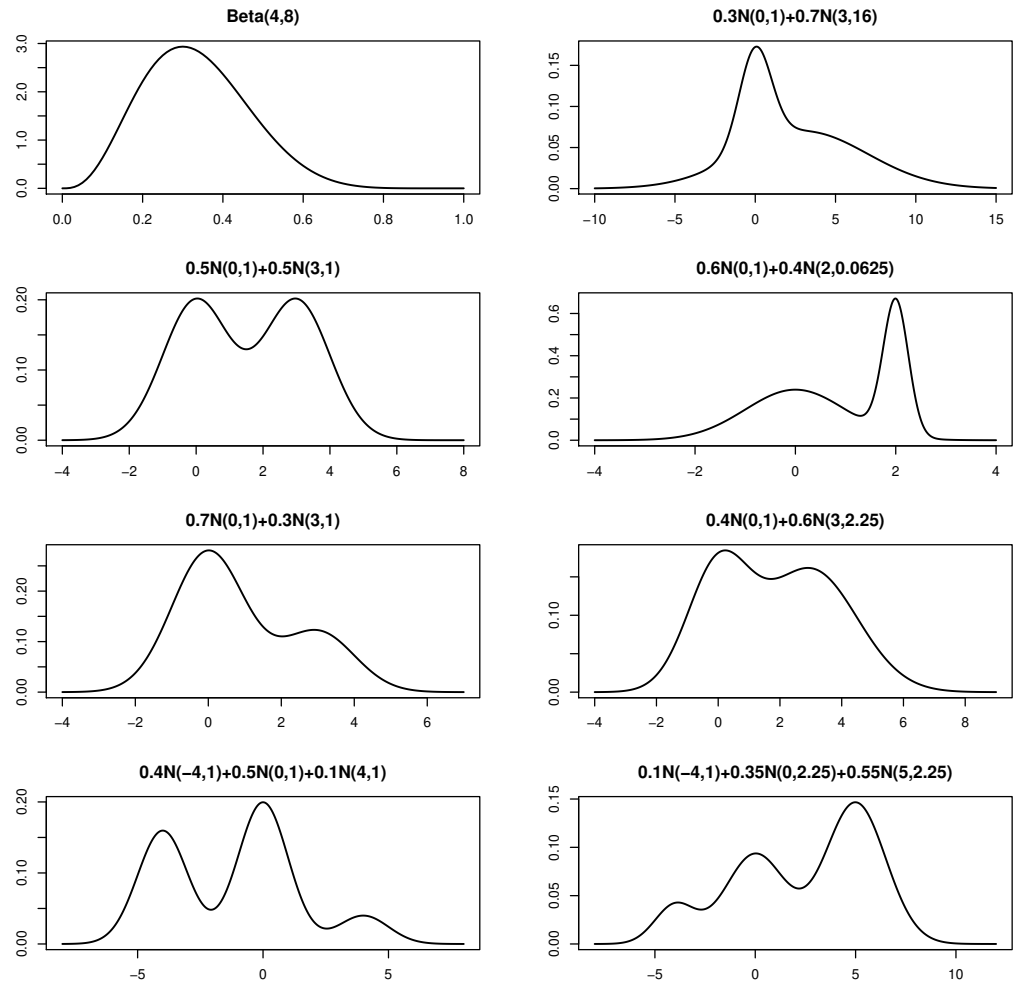

Figure 2.2: Distributions used in the simulation study

The first distribution is a common Beta distribution, which is unimodal. The second distribution is a unimodal mixture of two normal distributions and exhibits a “flatter part". All three methods demonstrate good test size for these two unimodal distributions. Moving on to the next two distributions, they are considered as "obvious" bimodal distributions, which can be clearly observed in the plot. Following these, the next two bimodal distributions are considered as “not that obvious" bimodal distributions, indicating that one of the modes is less prominent. Consequently, the power increases more slowly compared to the first two bimodal distributions. For all four bimodal distributions, our spline method exhibits competitive power for small sample sizes and stronger power for larger sample sizes than the other two methods. Finally, the last two distri-

Table 2.1: The proportion of rejection of tests based on samples from distributions in Figure 2.2 with sample size $n = 50, 100, 200, 500, 1000$

Distribution	method	$n = 50$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
$Beta(4, 8)$	splines	0.011	0.007	0.007	0.003	0.003
	kernel	0.037	0.057	0.032	0.034	0.034
	dip	0.003	0.001	0.001	0	0
$0.3N(0, 1) + 0.7N(3, 16)$	splines	0.024	0.024	0.016	0.019	0.02
	kernel	0.051	0.042	0.038	0.047	0.05
	dip	0.008	0.006	0	0	0
$0.5N(0, 1) + 0.5N(3, 1)$	splines	0.127	0.319	0.576	0.916	1
	kernel	0.177	0.326	0.486	0.797	0.969
	dip	0.063	0.114	0.189	0.475	0.812
$0.6N(0, 1) + 0.4N(2, 0.0625)$	splines	0.265	0.432	0.73	0.981	1
	kernel	0.392	0.537	0.724	0.967	0.998
	dip	0.209	0.299	0.441	0.806	0.977
$0.7N(0, 1) + 0.3N(3, 1)$	splines	0.045	0.061	0.08	0.184	0.428
	kernel	0.059	0.058	0.065	0.123	0.179
	dip	0.014	0.008	0.002	0.003	0.001
$0.4N(0, 1) + 0.6N(3, 2.25)$	splines	0.039	0.087	0.148	0.236	0.405
	kernel	0.095	0.135	0.182	0.231	0.313
	dip	0.025	0.023	0.026	0.025	0.034
$0.4N(-4, 1) + 0.5N(0, 1)$ $+0.1N(4, 1)$	splines	0.169	0.291	0.667	0.978	1
	kernel	0.172	0.244	0.353	0.605	0.837
	dip	0.038	0.045	0.078	0.181	0.366
$0.1N(-4, 1) + 0.35N(0, 2.25)$ $+0.55N(5, 2.25)$	splines	0.133	0.220	0.358	0.689	0.928
	kernel	0.161	0.284	0.389	0.715	0.936
	dip	0.059	0.081	0.1	0.242	0.484

butions are trimodal distributions. In both scenarios, our method demonstrates competitive power in most cases.

2.5.2 Simulation 2

In this simulation study, our focus is to compare the three methods based on the 'distance of the modes'. Three different mixtures of distributions are included:

$$(a) 0.6N(0,1) + 0.4N(d,1)$$

$$(b) 0.4N(0,1) + 0.4N(d,1) + 0.2N(0,9)$$

$$(c) 0.6\chi^2(5) + 0.4\chi^2(d)$$

The first scenario is a mixture of two normal distributions. For the second scenario, a third normal distribution is added to show the performance when there's a 'tail'. The third scenario is about the mixture of chi-squared distributions. With d increasing, the distributions turn bimodal from unimodal. The cut-off values of d are shown by the vertical dashed lines in the figures. For each distribution and each d, we generated 1000 samples with sample size $n = 200$ and $n = 800$. The bootstrap size we use is 1000. We compare our method with Silverman's kernel method and the Dip test. The proportion of p-values below 0.05 was plotted below.

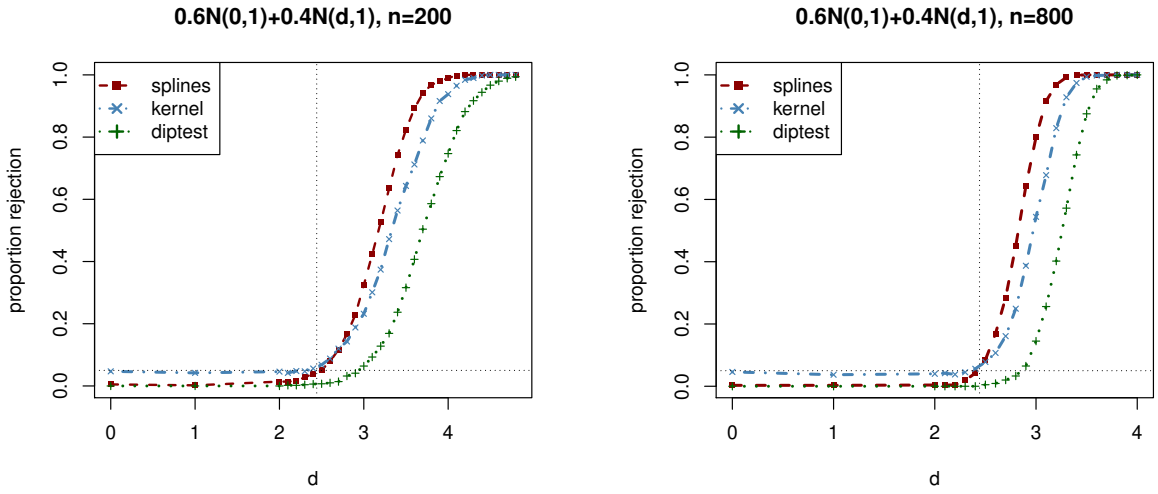

Figure 2.3: The proportion of rejection for samples from 0.6N(0,1)+0.4N(d,1), with sample size $n = 200$ and $n = 800$

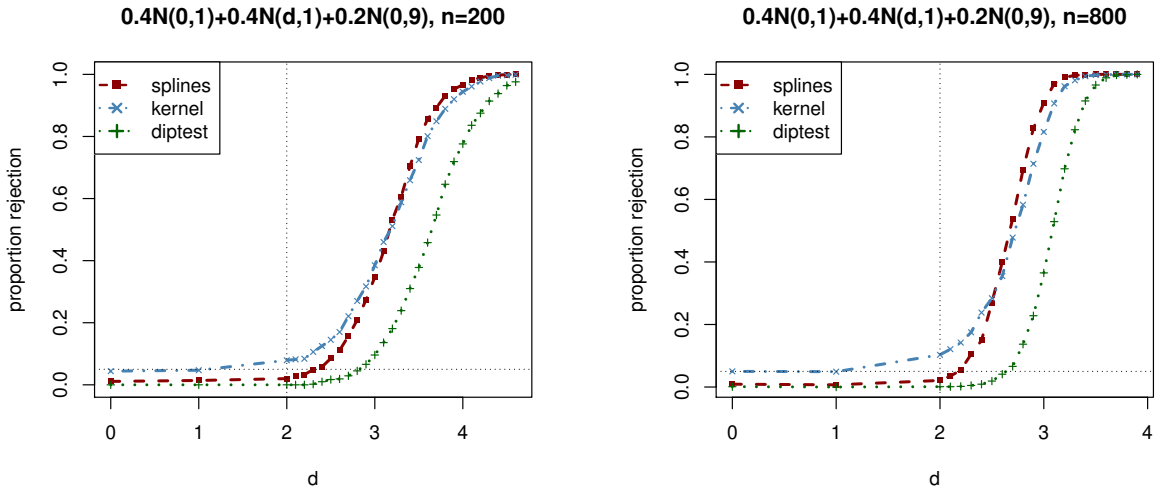

Figure 2.4: The proportion of rejection for samples from 0.4N(0,1)+0.4N(d,1)+0.2N(0,9), with sample size $n = 200$ and $n = 800$

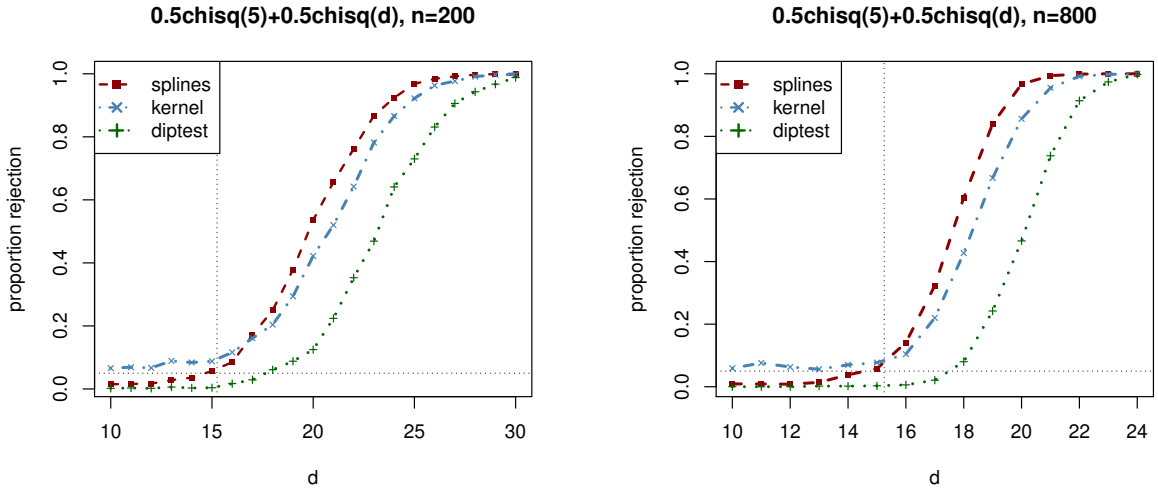

Figure 2.5: The proportion of rejection for samples from 0.5chisq(5)+0.5chisq(d), with sample size $n = 200$ and $n = 800$

Figure 2.3 presents the rejection proportion for $0.6N(0,1) + 0.4N(d,1)$ with sample size $n = 200, 800$. Notice the cut-off value for d in this case is around 2.4. In other words, the mixture distribution remains unimodal until d excesses 2.4, beyond which the simulation indicates a notable increase in the rejection proportion, signifying the transition to a bimodal distribution.

Also our methods show greater power than the other two methods. Notice that since we have the 'auto-acceptance' rule in our method, the rejection proportion is a lot lower than 0.05 when the distribution is unimodal. Figure 2.4 illustrates the rejection proportion for $0.4N(0,1) + 0.4N(d,1) + 0.2N(0,9)$. The addition of a third component with a larger variance is intended to introduce heavier tails to the mixture distribution. The cut-off d value for this case is 2. Furthermore, since the first two major components have equal weights, the unimodal density at $d = 2$ exhibits a 'flat mode'. The simulation results also suggest a potential for an inflated test size with the kernel method. Figure 2.5 shows the rejection proportion for mixures of chi-square distributions $0.6\chi^2(5) + 0.4\chi^2(d)$. The distributions are highly skewed and has the fixed lower support. The cut-off d is 15.25, and when the mixture becomes bimodal, the second mode is a lot flatter and less obvious compared to the first mode. According to the plot, our method method demonstrates competitive performance in terms of both power and test size when compared to the other two methods.

2.6 Real Data Application

We demonstrated our method using Paulsen's data for neurotransmission in guinea pig brains. Measurements were taken on prepared sections from the brains of adult guinea pigs, with recordings taken of spontaneous currents flowing into individual brain cells. The aim of the study was to see whether the current flow exhibited quantal characteristics. If the current was quantal then it would be expected that the distribution of the current amplitudes would be multimodal. We used the data in the `R` package `boot`. Of a total of 346 current were observed and the corresponding histogram is presented in Figure 2.6.

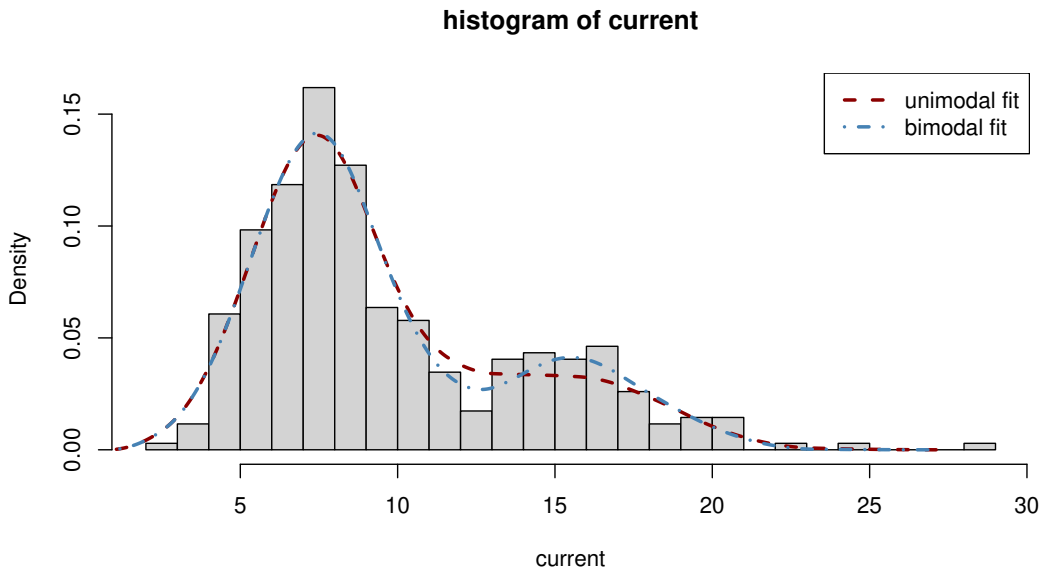

Figure 2.6: The histogram of Paulsen's data along with the unimodal and bimodal fit

The unimodal fit and bimodal fit from our method are also shown in Figure 2.6. Compared to the other two methods used in Section 2.5, our method differed in the ability to distinguish whether the distribution is unimodal or not, p-values are shown in Table 2.2.

method	splines	kernel	dip test
p-value	0.015	0.124	0.699

Table 2.2: P-values for three tests using Paulsen's data.

2.7 Summary

This chapter addresses the challenge of determining the modality of a distribution from a random sample, focusing on distinguishing between unimodal and multi-modal distributions. We propose a novel test using penalized splines and a least-squares criterion. The test statistic compares the least-squares criterion of the unimodal and bimodal fits. We estimate the null distribution of the test statistic using simulated data from the unimodal density estimate. Large sample

theory are derived and simulation studies are conducted to demonstrate the test's superior power compared to competitors in certain scenarios. Additionally, we apply our methodology to real neuro-transmission data from guinea pig brains, illustrating its practical utility.

Chapter 3

Deconvolution Density Estimation

In this Chapter, we present a straightforward solution to the deconvolution density estimation involving penalized splines. The key idea of the estimation is outlined in Section 3.2. The convergence rate for piecewise constant splines is derived in 3.3. Additionally, in Section 3.4, we show that with quadratic splines and uniform errors, a cube-root convergence rate is attained. Furthermore, large sample theories are derived for the penalized spline estimator and the constrained spline estimator. The simulations conducted in Section 3.4.4 show that the estimators perform well compared to kernel estimators in a variety of scenarios.

3.1 Introduction and Background

Suppose X is a random variable with an unknown density f. We are interested in the estimation of f based on observations $Y_1, \ldots, Y_n$ from $Y = X + Z$, where Z is a continuous random variable representing measurement error, with a known density function ϕ. If X and Z are independent, the density function of the random variable Y is the convolution of f and ϕ, i.e.

$$g(y) = f * \phi(y) = \int f(z)\phi(y - z)dz.$$

For this reason, the measurement error problem has been referred to as a density deconvolution problem, which has a variety of applications and has been studied widely.

Most previous approaches used a kernel method. [Carroll and Hall, 1988] and [Stefanski and Carroll, 1990] estimated f using a kernel that is a function of the characteristic functions of the error and a kernel function for estimating g. They showed this estimator is point-wise consistent, and the large sample properties were further studied by [Fan, 1991], [Delaigle and Gijbels, 2004], and [Carroll and Hall, 1988]. They showed that the convergence rate of nonparametric deconvolution methods is quite slow for smooth error distributions such as the normal distribution.

A number of authors considered semi-parametric methods; for example, [Cordy and Thomas, 1997] modeled the distribution of X to be that of a finite mixture of known distributions. [Hazelton and Turlach, 2010] proposed a semi-parametric method, where the ratio of the the unconvoluted and convoluted densities is specified parametrically. [Delaigle and Hall, 2014] proposed a parametrically assisted nonparametric method.

[Mendelsohn and Rice, 1982] used B-splines to estimate f, minimizing the L_2 distance from the empirical distribution, and showed consistency of the estimator. [Koo and Park, 1996] provided a maximum-likelihood B-spline solution, estimating $\log(f)$ by a polynomial spline, and employing an EM algorithm to estimate the spline coefficients. [Hall and Qiu, 2005] described deconvolution methods using a Fourier series expansion. Wavelet methods were studied by [Walter, 1999] and [Pensky, 2002].

Our work focuses on spline density estimation of f where the measurement error distribution is uniform. We define a spline basis $\delta_1, \ldots, \delta_m$ for the estimation of f, and define the spline basis for the estimation of g as $\gamma_j = \delta_j * \phi$. We estimate $\hat{g}(y) = \sum_{j=1}^{m} \hat{b}_j \gamma_j(y)$ with the observed $Y_1, \ldots, Y_n$, then $\hat{f}(x) = \sum_{j=1}^{m} \hat{b}_j \delta_j(x)$. With quadratic splines, a cube-root convergence rate in the L_2 norm is achieved for $\hat{f}$. A main advantage of spline density estimation, compared to kernel density estimation, is that shape constraints such as unimodality are readily imposed on spline estimators.

3.2 The Deconvolution Least-Squares Spline Estimation

We want to estimate f, the density for X, using i.i.d. observations $Y_1, \ldots, Y_n$ having density $g = f * \phi$ with known ϕ. We do this by constructing spline basis functions $\delta_1, \ldots, \delta_m$ for the estimation of f, where the goal is to find an estimator $\tilde{f}$ of the form

$$\tilde{f}(x) = \sum_{j=1}^{m} \tilde{b}_j \delta_j(x), \text{ such that } \tilde{f}(x) \geq 0 \text{ for all } x, \text{ and } \boldsymbol{a}^\top \tilde{\boldsymbol{b}} = 1,$$

where a_j is the area under $\delta_j(x)$ with $\boldsymbol{a} = (a_1, \ldots, a_m)^\top$. For any such $\tilde{f}$, its convolution with the error density is

$$\tilde{g}(y) = \tilde{f} * \phi(y) = \sum_{j=1}^{m} \tilde{b}_j \int_{-\infty}^{\infty} \delta_j(x)\phi(y-x)dx =: \sum_{j=1}^{m} \tilde{b}_j \gamma_j(y),$$

with $\gamma_j(y) = \int_{-\infty}^{\infty} \delta_j(z)\phi(y-z)dz$ for $j = 1, \ldots, m$. The basis functions $\gamma_1, \ldots, \gamma_m$ are used to estimate g, and the coefficients from this estimator are used to construct $\tilde{f}$. Constraints in the form $\boldsymbol{A}\boldsymbol{b} \geq \boldsymbol{0}$ may be imposed, where $\boldsymbol{A}$ is an appropriate constraint matrix. For example, we can constrain f to be unimodal.

Let $\mathcal{F}$ be the space of linear combinations of the δ basis functions and let $\mathcal{G}$ be the space of linear combinations of the γ basis functions. If we choose a set of linearly independent basis functions $\delta_1, \ldots, \delta_m$, then the basis functions $\gamma_1, \ldots, \gamma_m$ will also be linearly independent, and there is a one-to-one mapping between $\mathcal{F}$ and $\mathcal{G}$. To estimate the density g of Y given n independent observations, the least-squares criterion similar as 1.2 is used. We minimize

$$\psi(g; \boldsymbol{Y}) = \int g(y)^2 dy - \frac{2}{n} \sum_{i=1}^{n} g(Y_i)$$

over the subset of $\mathcal{G}$ that corresponds to densities with the desired shape. Define the $m \times m$ matrix $\boldsymbol{H}$ as $H_{j\ell} = \int_{-\infty}^{\infty} \gamma_j(y)\gamma_\ell(y)dy$, and $\boldsymbol{c}$ as the vector in $\mathbb{R}^m$ with $c_j = \sum_{i=1}^{n} \gamma_j(Y_i)/n$. The least-squares criterion can be written in terms of the coefficients of the spline basis, resulting in the quadratic programming problem

$$\text{minimize } \boldsymbol{b}^\top \boldsymbol{H}\boldsymbol{b} - 2\boldsymbol{c}^\top \boldsymbol{b}, \text{ subject to } \boldsymbol{a}^\top \boldsymbol{b} = 1 \text{ and } \boldsymbol{A}\boldsymbol{b} \geq \boldsymbol{0},$$

where the constraint matrix $\boldsymbol{A}$ defines the desired shape and constrains the density to be non-negative. Let $\mathcal{B}$ be the subset of $\boldsymbol{b} \in \mathbb{R}^m$ so that $\boldsymbol{a}^\top \boldsymbol{b} = 1$. We can write $\boldsymbol{b} \in \mathcal{B}$ as $\boldsymbol{b} = \boldsymbol{b}_0 + \boldsymbol{\beta}$

where $\boldsymbol{b}_0 = \boldsymbol{a}/\|\boldsymbol{a}\| \in \mathcal{B}$, and $\boldsymbol{a}^\top\boldsymbol{\beta} = 0$. The equivalent problem is to find $\boldsymbol{\beta}$ to minimize

$$\boldsymbol{\beta}^\top \boldsymbol{H}\boldsymbol{\beta} - 2(\boldsymbol{c} - \boldsymbol{H}\boldsymbol{b}_0)^\top\boldsymbol{\beta}, \text{ subject to } \boldsymbol{a}^\top\boldsymbol{\beta} = 0 \text{ and } \boldsymbol{A}\boldsymbol{\beta} \geq -\boldsymbol{A}\boldsymbol{b}_\mathbf{0}.$$

To deal with the equality constraint, let $\boldsymbol{W}$ be an $m \times (m-1)$ matrix with columns forming a basis for the linear space orthogonal to $\boldsymbol{a}$, so that $\boldsymbol{a}^\top\boldsymbol{\beta} = 0$ if and only if $\boldsymbol{\beta} = \boldsymbol{W}\boldsymbol{\alpha}$ for some $\boldsymbol{\alpha} \in \mathbb{R}^{m-1}$. Hence we have $\boldsymbol{b} = \boldsymbol{b}_0 + \boldsymbol{W}\boldsymbol{\alpha}$ and the criterion function can be written in terms of $\boldsymbol{\alpha}$:

$$\text{minimize } \boldsymbol{\alpha}^\top\boldsymbol{W}^\top\boldsymbol{H}\boldsymbol{W}\boldsymbol{\alpha} - 2(\boldsymbol{c} - \boldsymbol{H}\boldsymbol{b}_0)^\top\boldsymbol{W}\boldsymbol{\alpha} \text{ subject to } \boldsymbol{A}\boldsymbol{W}\boldsymbol{\alpha} \geq -\boldsymbol{A}\boldsymbol{b}_\mathbf{0}.$$

Ignoring the inequality constraints for now, the "unconstrained" estimator is obtained:

$$\tilde{\boldsymbol{\alpha}} = (\boldsymbol{W}^\top\boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top(\boldsymbol{c} - \boldsymbol{H}\boldsymbol{b}_0) \text{ and } \tilde{\boldsymbol{\beta}} = \boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top(\boldsymbol{c} - \boldsymbol{H}\boldsymbol{b}_0).$$

Finally,

$$\tilde{\boldsymbol{b}} = [\boldsymbol{I} - \boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{H}]\boldsymbol{b}_0 + \boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{c},$$

so $\tilde{g}(x) = \sum_{j=1}^m \tilde{b}_j\gamma_j(x)$ is our unconstrained estimator for g and $\tilde{f}(x) = \sum_{j=1}^m \tilde{b}_j\delta_j(x)$ is the corresponding estimator for f. From $\hat{\boldsymbol{b}}$, the constrained density estimators are $\hat{g}(x) = \sum_{j=1}^m \hat{b}_j\gamma_j(x)$ and $\hat{f}(x) = \sum_{j=1}^m \hat{b}_j\delta_j(x)$.

3.3 Piecewise Constant Splines

The simplest case is uniform errors and piecewise constant splines. Suppose the error density is uniform on $(-h, h)$, and we estimate f using piecewise constant splines with equally spaced knots $t_1, \ldots, t_{m+1}$. Let d be $t_{j+1} - t_j$ for $j = 1, \ldots, m$ and for convenience suppose $h = \ell d$ for a positive integer ℓ.

The piecewise constant spline basis for estimating f is $\delta_j(x) = 1$ for $x \in (t_j, t_{j+1})$ and $\delta_j(x) = 0$ for $x \notin (t_j, t_{j+1})$. Then the basis functions for estimating g are trapezoids:

$$\gamma_j(y) = \int_{-\infty}^{\infty} \delta_j(y-x)\phi(x)dx = \begin{cases} 0 & \text{for } y < t_j - h \\ \frac{1}{2h}(y - t_j + h) & \text{for } y \in [t_j - h, t_{j+1} - h) \\ \frac{d}{2h} & \text{for } y \in [t_{j+1} - h, t_j + h] \\ \frac{1}{2h}(t_{j+1} + h - y) & \text{for } y \in (t_j + h, t_{j+1} + h] \\ 0 & \text{for } y > t_{j+1} + h \end{cases}$$

Let $\tilde{m} = m + 2\ell - 1$ and define $s_0 = t_1 - d\ell$ and $s_j = s_0 + jd$, for $j = 1, \ldots, \tilde{m} + 1$ as in the plot below with $m = 6$ and $\ell = 2$. The s_j are an expanded set of knots representing the support for $g \in \mathcal{G}'$.

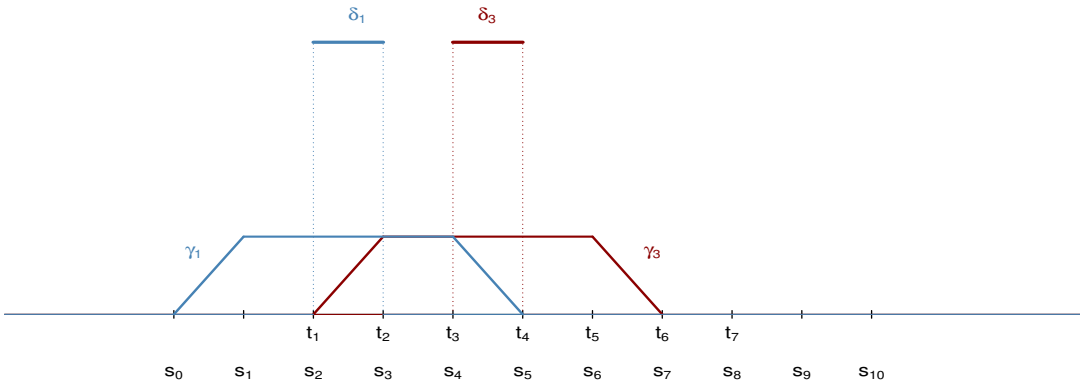

Figure 3.1: Example basis functions for $m = 6$ and $\ell = 2$.

Define $\boldsymbol{\zeta} \in \mathbb{R}^{\tilde{m}}$ so that $\zeta_j = g(s_j)$ for $j = 1, \ldots, \tilde{m}$ and note that if $g(y) = \sum_{i=1}^{m} b_j \gamma_j(y)$, then

$$\boldsymbol{\zeta} = \boldsymbol{M}\boldsymbol{b}$$

and

$$\boldsymbol{b} = (\boldsymbol{M}^\top \boldsymbol{M})^{-1} \boldsymbol{M}^\top \boldsymbol{\zeta}$$

where the $\tilde{m} \times m$ matrix $\boldsymbol{M}$ has nonzero elements $M_{i,j} = \frac{d}{2h}$ for $i = j, \ldots, j + 2\ell - 1$.

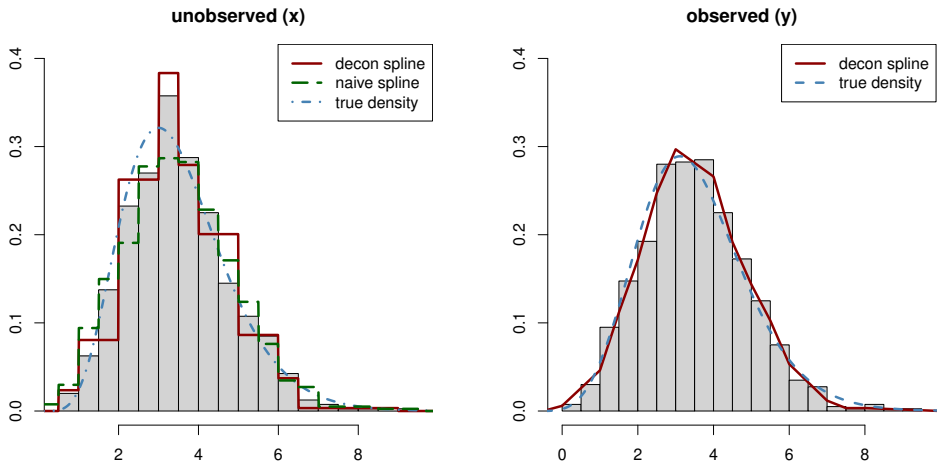

Figure 3.2: Example of simulated data with $n = 800$ and estimated densities constrained to be unimodal. Left: the histogram of the unobserved sample from f. Right: the histogram of the observed sample from g.

Figure 3.2 shows an example of simulated data with $n = 800$ and measurement error from a uniform distribution on the interval $[-1, 1]$. The "naive" spline estimation refers to the estimated density without considering the presence of measurement error. In the piecewise constant basis framework, the naive estimate is the same as the histogram of $Y_1, \ldots, Y_n$.

As the sample size n grows, the number of knots m increases, so that the space d between the knots decreases. In this section our goal is to determine the rate of increase of m that leads to the fastest rate of convergence of $\tilde{f}$. We start with a result that illustrates the difficulty of the deconvolution problem. Functions g_1 and g_2 in $\mathcal{G}$ have to be quite close together to guarantee that the corresponding f_1 and f_2 in $\mathcal{F}$ are near each other.

Theorem 3. *For f_1 and f_2 in $\mathcal{F}$, $\|f_1 - f_2\|^2 = O(m^2\|g_1 - g_2\|^2)$ where g_1 and g_2 are the corresponding functions in $\mathcal{G}$.*

Proof: Because the piecewise constant spline basis functions defined above all have the same area, the nonzero elements of the matrix $\boldsymbol{W}$ used in the solution of the quadratic programming problem in Section 3.2 may be set to $W_{i,i} = -1$,$W_{i+1,i} = 1$, for $i = 1, \ldots, m - 1$.

Let $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ be the corresponding coefficients. Then $\boldsymbol{b}_1 - \boldsymbol{b}_2 = \boldsymbol{W}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)$. First we show that for all $\boldsymbol{b} = \boldsymbol{W}\boldsymbol{\alpha}$, $\|\boldsymbol{b}\|^2 \leq 4\|\boldsymbol{\alpha}\|^2$

$$
\begin{aligned}
\|\boldsymbol{b}\|^2 &= \boldsymbol{\alpha}^\top \boldsymbol{W}^\top \boldsymbol{W}\boldsymbol{\alpha} \\
&= \begin{pmatrix} -\alpha_1 \\ \alpha_1 - \alpha_2 \\ \vdots \\ \alpha_{m-2} - \alpha_{m-1} \\ \alpha_{m-1} \end{pmatrix}^\top \begin{pmatrix} -\alpha_1 \\ \alpha_1 - \alpha_2 \\ \vdots \\ \alpha_{m-2} - \alpha_{m-1} \\ \alpha_{m-1} \end{pmatrix} \\
&= \alpha_1 + \sum_{i=1}^{m-2} (\alpha_i - \alpha_{i+1})^2 + \alpha_{m-1}^2 \\
&\leq \alpha_1 + 2\sum_{i=1}^{m-2} (\alpha_i^2 - \alpha_{i+1}^2) + \alpha_{m-1}^2 \\
&\leq 4\sum_{i=1}^{m-1} \alpha_1^2 = 4\|\alpha\|^2.
\end{aligned}
$$

Hence we have $\|\boldsymbol{b}_1 - \boldsymbol{b}_2\|^2 \leq 4\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|^2$.

Let $\boldsymbol{\zeta}_1 \in \mathbb{R}^{\tilde{m}}$ be such that $\zeta_{1j} = g_1(s_j)$ and $\boldsymbol{\zeta}_2 \in \mathbb{R}^{\tilde{m}}$ be such that $\zeta_{2j} = g_2(s_j)$. Let $r_j = \zeta_{1j} - \zeta_{2j}$. Since $g_1 - g_2$ is piecewise linear, $\|g_1 - g_2\|^2$ can be written as

$$\begin{aligned}
\int [g_1(y) - g_2(y)]^2 dy &= \sum_{i=0}^{\tilde{m}} \int_{s_i}^{s_{i+1}} [g_1(y) - g_2(y)]^2 dy \\
&= \sum_{i=0}^{\tilde{m}} \int_{s_i}^{s_{i+1}} (\frac{r_{i+1} - r_i}{d} y + \frac{s_{i+1} r_i - s_i r_{i+1}}{d})^2 dy \\
&= \sum_{i=0}^{\tilde{m}} \frac{d}{3(r_{i+1} - r_i)} (\frac{r_{i+1} - r_i}{d} y + \frac{s_{i+1} r_i - s_i r_{i+1}}{d})^3 \Big|_{s_i}^{s_{i+1}} \\
&= \sum_{i=0}^{\tilde{m}} \frac{d}{3} (r_{i+1}^2 + r_i^2 + r_{i+1} r_i) \\
&= \sum_{i=1}^{\tilde{m}} \frac{2d}{3} r_i^2 + \sum_{i=1}^{\tilde{m}-1} \frac{d}{3} r_{i+1} r_i \\
&= \frac{d}{6} \left[\sum_{i=1}^{\tilde{m}-1} (r_i + r_{i+1})^2 + 2 \sum_{i=1}^{\tilde{m}} r_i^2 + r_1^2 + r_{\tilde{m}}^2 \right].
\end{aligned}$$

So we have

$$\frac{d}{3} \|\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2\|^2 \leq \int [g_1(y) - g_2(y)]^2 dy \leq d \|\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2\|^2$$

$\|\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2\|^2 = (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha_2})^\top \boldsymbol{W}^\top \boldsymbol{M}^\top \boldsymbol{M} \boldsymbol{W} (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha_2})$, and matrix $\boldsymbol{W}^\top \boldsymbol{M}^\top \boldsymbol{M} \boldsymbol{W} = \frac{d^2}{4h^2} \boldsymbol{Q}$, where $\boldsymbol{Q}$ is a $(m-1) \times (m-1)$ sparse Toepliz matrix with $\boldsymbol{Q}_{ii} = 2$ for $i = 1, \ldots, m-1$ and $\boldsymbol{Q}_{2l+i,i} = \boldsymbol{Q}_{i,2l+i} = -1$ for $i = 1, \ldots, (m-1-2l)$.

Lemma 1. *There exists an $\epsilon > 0$ depending only on S/h (not depending on m or l) such that the lowest eigenvalue of $\boldsymbol{Q}$ is at least ϵ.*

The proof of the Lemma is in Section 3.6. By Lemma 1, we have $\|\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2\|^2 \geq \frac{d^2 \epsilon}{4h^2} \|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|^2$, so $\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|^2 = O(m^3 \|g_1 - g_2\|^2)$. Then

$$\int [f_1(x) - f_2(x)]^2 dx = \sum_{j=1}^{m} \int_{t_j}^{t_{j+1}} [(b_{1j} - b_{2j}) \delta_j(x)]^2 dx = d \sum_{j=1}^{m} (b_{1j} - b_{2j})^2 = O(m^2 \|g_1 - g_2\|^2).$$

□

As an example, consider the densities $f_1(x) = 1/d$ for $x \in (t_1, t_2)$, and $f_2(x) = 1/d$ for $x \in (t_2, t_3)$ and $f_j(x) = 0$ otherwise. The densities $f_1(x)$ and $f_2(x)$ are not overlapping and the distance $\|f_1 - f_2\|^2 = 2/d$ will increase on the order of m with increasing m. The corresponding trapezoids $g_1(y)$ and $g_2(y)$ are

$$g_1(y) = \begin{cases} 0 & \text{for } y < t_1 - h \\ \frac{1}{2hd}(y - t_1 + h) & \text{for } y \in [t_1 - h, t_2 - h) \\ \frac{1}{2h} & \text{for } y \in [t_2 - h, t_1 + h] \\ \frac{1}{2hd}(t_2 + h - y) & \text{for } y \in (t_1 + h, t_2 + h] \\ 0 & \text{for } y > t_2 + h \end{cases}$$

$$g_2(y) = \begin{cases} 0 & \text{for } y < t_2 - h \\ \frac{1}{2hd}(y - t_2 + h) & \text{for } y \in [t_2 - h, t_3 - h) \\ \frac{1}{2h} & \text{for } y \in [t_3 - h, t_2 + h] \\ \frac{1}{2hd}(t_3 + h - y) & \text{for } y \in (t_2 + h, t_3 + h] \\ 0 & \text{for } y > t_3 + h \end{cases}$$

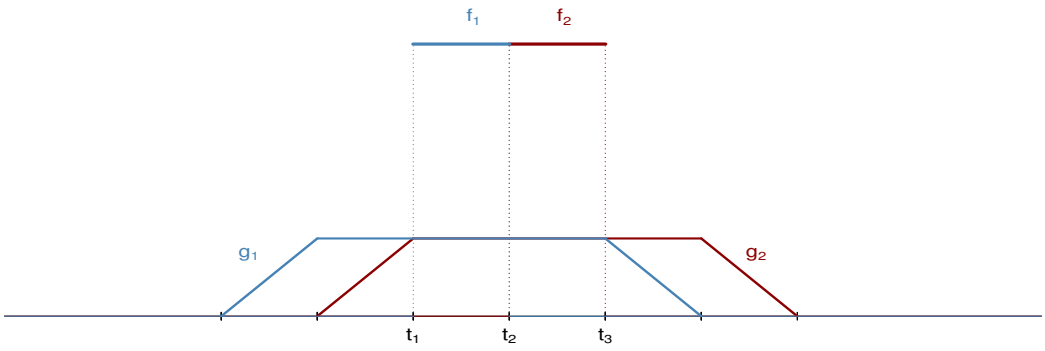

Figure 3.3: Example of $f_1(x)$ and $f_2(x)$ with corresponding $g_1(y)$ and $g_2(y)$.

Apparently $g_1(y)$ and $g_2(y)$ overlap more for larger m, and by symmetry we have the distance

$$\begin{aligned}
\|g_1(y) - g_2(y)\|^2 &= 2\int_{t_1-h}^{t_2-h}\left[\frac{1}{2hd}(y - t_1 + h)\right]^2 dy + 2\int_{t_2-h}^{t_3-h}\left[\frac{1}{2h} - \frac{1}{2hd}(y - t_2 + h)\right]^2 dy \\
&= \frac{1}{2h^2d^2}\frac{d^3}{3} + \frac{1}{2h^2d^2}\frac{d^3}{3} \\
&= \frac{d}{3h^2} \\
&= O(1/m).
\end{aligned}$$

This example shows that even though the f functions are getting farther apart, g functions can get closer together, and the bound above is the closest bound we can get. Therefore, to show that $\|\tilde{f} - f_0\|^2$ is small we must show that $\|\tilde{g} - g_0\|^2$ is very small, where f_0 and g_0 represent the true density functions.

Theorem 4. *If f_0 has a bounded third derivative, then $\|\tilde{f} - f_0\|^2 = O_p(n^{-2/5})$.*

Proof: We first find the convergence rate of $\tilde{g}$, and apply Theorem 3 to get the convergence rate of $\tilde{f}$. Let $\boldsymbol{q}$ be the vector of quantiles of g_0 such that the cdf $G_0(q_i) = \frac{i}{n+1}$ for $i = 1, \ldots, n$, and define $\bar{g}$ as the minimizer $\psi(g, \boldsymbol{q})$ over densities $g \in \mathcal{G}$. We split the error into two pieces: $\|\tilde{g} - g_0\| \leq \|\tilde{g} - \bar{g}\| + \|\bar{g} - g_0\|$, where the first term is called the "estimation error" and the second is the "approximation error."

Starting with the estimation error, we show that $\|\tilde{g} - \bar{g}\|^2 = O_p(m/n)$. Let $\eta_1, \ldots, \eta_m$ be an orthonormal basis of $\mathcal{G}$, and define vector $\boldsymbol{c}_\eta$ as

$$c_{\eta,j} = \frac{1}{n}\sum_{i=1}^{n}\eta_j(y_i)$$

Define the vector $\bar{\boldsymbol{c}}_\eta$ so that

$$\bar{c}_{\eta,j} = \frac{1}{n}\sum_{i=1}^{n}\eta_j(q_i)$$

Then

$$\begin{aligned}
\|\boldsymbol{c}_\eta - \bar{\boldsymbol{c}}_\eta\|^2 &= \sum_{j=1}^{m}\left[\frac{1}{n}\sum_{i=1}^{n}\eta_j(y_i) - \frac{1}{n}\sum_{i=1}^{n}\eta_j(q_i)\right]^2 \\
&= \sum_{j=1}^{m}\left[(\frac{1}{n}\sum_{i=1}^{n}\eta_j(y_i) - E(\eta_j(Y))) + (E(\eta_j(Y)) - \frac{1}{n}\sum_{i=1}^{n}\eta_j(q_i))\right]^2 \\
&\leq \sum_{j=1}^{m}\left[2(\frac{1}{n}\sum_{i=1}^{n}\eta_j(y_i) - E(\eta_j(Y)))^2 + 2(E(\eta_j(Y)) - \frac{1}{n}\sum_{i=1}^{n}\eta_j(q_i))^2\right] \\
&= O_p(m/n).
\end{aligned}$$

where the last equality is obtained by the central limit theorem.

Define $\tilde{\boldsymbol{b}}_\eta$ to be the vector of coefficients of the η basis functions for the estimated density $\tilde{g}$, i.e.$\tilde{g}(y) = \sum_{j=1}^{m}\tilde{b}_\eta\eta_j(y)$.while $\bar{\boldsymbol{b}}_\eta$ is the vector of coefficients for the approximation density $\bar{g}$, i.e. $\bar{g}(y) = \sum_{j=1}^{m}\bar{b}_\eta\eta_j(y)$. Then by the calculations in Section 3.2,

$$\tilde{\boldsymbol{b}}_\eta - \bar{\boldsymbol{b}}_\eta = \boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{H}_\eta\boldsymbol{W})^{-1}\boldsymbol{W}^\top[\boldsymbol{c}_\eta - \bar{\boldsymbol{c}}_\eta]$$

where $\boldsymbol{H}_\eta$ is defined as $H_{jl} = \int_{-\infty}^{\infty}\eta_j(y)\eta_l(y)dy$. Since $\eta_1, \ldots, \eta_m$ are orthonormal, $\boldsymbol{H}_\eta$ is an identity matrix and $\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{H}_\eta\boldsymbol{W})^{-1}\boldsymbol{W}^\top$ is a projection matrix with eigenvalues 0 and 1. So,

$$\|\tilde{g} - \bar{g}\|^2 = (\tilde{\boldsymbol{b}}_\eta - \bar{\boldsymbol{b}}_\eta)^\top\boldsymbol{H}_\eta(\tilde{\boldsymbol{b}}_\eta - \bar{\boldsymbol{b}}_\eta) = \|\tilde{\boldsymbol{b}}_\eta - \bar{\boldsymbol{b}}_\eta\|^2 \leq \|\boldsymbol{c}_\eta - \bar{\boldsymbol{c}}_\eta\|^2 = O_p(m/n).$$

Next we tackle the approximation error $\|\bar{g} - g_0\|^2$. Define $f^*(x) \in \mathcal{F}$ to be

$$f^*(x) = \frac{1}{d}\int_{t_j}^{t_{j+1}} f_0(u)du, \quad \text{for } x \in [t_j, t_{j+1});$$

that is, $b_j^* = \frac{1}{d}\int_{t_j}^{t_{j+1}} f_0(u)du$ and $f^*(x) = \sum_{j=1}^{m} b_j\delta_j(x)$.

We have

$$\|f^* - f_0\|^2 = \int [f^*(x) - f_0(x)]^2 dx \;\; = \;\; \sum_{j=1}^{m} \int_{t_j}^{t_{j+1}} [b_j^* - f_0(x)]^2 dx = O(md^3) = O(m^{-2}).$$

Define $g^*(x) = f^* * \phi(x)$; now we prove $\|g^* - g_0\|^2 = O_p(m^{-4})$. Let F be the cdf of X, then $g(y)$ can be written as

$$g(y) = \frac{1}{2h}[F(y+h) - F(y-h)]. \tag{3.1}$$

So,

$$\begin{aligned} \int [g^*(y) - g_0(y)]^2 dy &= \frac{1}{2h} \int \left\{ [F^*(y+h) - F^*(y-h)] - [F_0(y+h) - F_0(y-h)] \right\}^2 dy \\ &= \frac{1}{2h} \int \left\{ [F^*(y+h) - F_0(y+h)] - [F^*(y-h) - F_0(y-h)] \right\}^2 dy, \end{aligned}$$

where F^* and F_0 are the corresponding cdf with respect to f^* and f_0. Consider $x \in (t_J, t_{J+1}]$,

$$\begin{aligned} F^*(x) = \int_{-\infty}^{x} f^*(u) du &= \sum_{j=1}^{J-1} \int_{t_j}^{t_{j+1}} f^*(u) du + \int_{t_J}^{x} f^*(u) du \\ &= \sum_{j=1}^{J-1} b_j^* d + \int_{t_J}^{x} f^*(u) du \\ &= \sum_{j=1}^{J-1} \int_{t_j}^{t_{j+1}} f_0(u) du + \int_{t_J}^{x} f^*(u) du. \end{aligned}$$

So,

$$F^*(x) - F_0(x) = \int_{t_J}^{x} f^*(u) du - \int_{t_J}^{x} f_0(u) du = O(d^2) = O(m^{-2}).$$

Therefore $\|g^* - g_0\|^2 = O_p(m^{-4})$. Then $\|g^* - g_0\| = O(m^{-2})$. Recall $\boldsymbol{q} = (q_1, \ldots, q_n)$ so that the CDF $G_0(q_i) = i/n$. Let q_0 be the lower bound of the support. Also define

$$g_s(y) = \frac{1}{n(q_i - q_{i-1})} \text{ for } y \in [q_{i-1}, q_i).$$

Also,

$$\frac{1}{2}\frac{d^2}{d\alpha^2}\psi\left(\alpha h + (1-\alpha)g_s; \boldsymbol{q}\right) = \|h - g_s\|^2.$$

Then we have

$$\psi(h; \boldsymbol{q}) - \psi(g_s; \boldsymbol{q}) = 2\|h - g_s\|^2 + \frac{2}{n}\sum_{i=1}^{n}\left[h(c_i) - h(q_i)\right].$$

The second term is $O(1/n)$ if h is of bounded variation. For some M_0 and large enough n,

$$2\|h - g_s\|^2 - \frac{M_0}{n} \leq \psi(h; \boldsymbol{q}) - \psi(g_s; \boldsymbol{q}) \leq 2\|h - g_s\|^2 + \frac{M_0}{n}.$$

Now we show that $\|\bar{g} - g_s\|^2$ is small. First we show that $\psi(g; \boldsymbol{y})$ is strictly convex in g. For any $\boldsymbol{y}$,

$$\begin{aligned}
& [\alpha\psi(g_1; \boldsymbol{y}) + (1-\alpha)\psi(g_2; \boldsymbol{y})] - \psi[\alpha g_1 + (1-\alpha)g_2; \boldsymbol{y}] \\
= & \int \alpha g_1(y)^2 dy + \int (1-\alpha) g_2(y)^2 dy - \int [\alpha g_1(y) + (1-\alpha)g_2(y))]^2 dy \\
= & \alpha(1-\alpha)\int [g_1(y) - g_2(y)]^2 dy,
\end{aligned}$$

which is strictly positive for any g_1, g_2 such that $\int [g_1(y) - g_2(y)]^2 dy > 0$. note that

$$\begin{aligned}
\psi(h; \boldsymbol{q}) - \psi(g_s; \boldsymbol{q}) &= \int_0^1 \frac{d}{d\alpha}\psi\left(\alpha h + (1-\alpha)g_s; \boldsymbol{q}\right) d\alpha \\
&= \frac{d}{d\alpha}\psi\left(\alpha h + (1-\alpha)g_s; \boldsymbol{q}\right)\Big|_{\alpha=0} + \frac{d^2}{d\alpha^2}\psi\left(\alpha h + (1-\alpha)g_s; \boldsymbol{q}\right),
\end{aligned}$$

using integration by parts. For $\alpha \in [0,1]$ and $h = \sum_{j=1}^m b_j \gamma_j$ for $\boldsymbol{b} \in \mathcal{B}$,

$$\psi\left(\alpha h + (1-\alpha) g_s; \boldsymbol{q}\right) \;=\; \int \left[\alpha h(y) + (1-\alpha) g_s(y)\right]^2 dy - \frac{2}{n} \sum_{i=1}^n \left[\alpha h(q_i) + (1-\alpha) g_s(q_i)\right],$$

so that

$$\begin{aligned}
&\frac{d}{d\alpha} \psi\left(\alpha h + (1-\alpha) g_s; \boldsymbol{q}\right) \big|_{\alpha=0} \\
&= \left[\int 2(\alpha h(y) + (1-\alpha) g_s(y))(h(y) - g_s(y)) dy - \frac{2}{n} \sum_{i=1}^n (h(q_i) - g_s(q_i)) \right] \big|_{\alpha=0} \\
&= \int 2 g_s(y)(h(y) - g_s(y)) dy - \frac{2}{n} \sum_{i=1}^n (h(q_i) - g_s(q_i)) \\
&= \sum_{i=1}^n \int_{q_{i-1}}^{q_i} \left[2 g_s(y) h(y) - 2 g_s(y)^2\right] dy - \frac{2}{n} \sum_{i=1}^n (h(q_i) - g_s(q_i)) \\
&= \sum_{i=1}^n \left[\frac{2}{n} h(c_i) + \frac{2}{n^2 (q_i - q_{i-1})} \right] - \frac{2}{n} \sum_{i=1}^n (h(q_i) - g_s(q_i)) = \frac{2}{n} \sum_{i=1}^n \left[h(c_i) - h(q_i)\right]
\end{aligned}$$

where $c_i \in (q_{i-1}, q_i)$. Also,

$$\frac{1}{2} \frac{d^2}{d\alpha^2} \psi\left(\alpha h + (1-\alpha) g_s; \boldsymbol{q}\right) = \|h - g_s\|^2,$$

so that

$$\psi(h; \boldsymbol{q}) - \psi(g_s; \boldsymbol{q}) = 2\|h - g_s\|^2 + \frac{2}{n} \sum_{i=1}^n \left[h(c_i) - h(q_i)\right],$$

and the second term is $O(1/n)$ for any h of bounded variation. For some M_0 and large enough n,

$$2\|h - g_s\|^2 - \frac{M_0}{n} \le \psi(h; \boldsymbol{q}) - \psi(g_s; \boldsymbol{q}) \le 2\|h - g_s\|^2 + \frac{M_0}{n}.$$

Because $|g_0(y) - g_s(y)| = O(n^{-1})$ for $y \in [q_{i-1}, q_i)$, we have $\|g_0 - g_s\|^2 = O(n^{-1})$. Hence $\|g^* - g_s\| \le \|g^* - g_0\| + \|g_0 - g_s\| = O_p(m^{-4}) + O_p(n^{-1})$; then there is M_1 so that for large enough n,

$$\psi(g^*; \boldsymbol{q}) - \psi(g_s; \boldsymbol{q}) \le M_1 m^{-4} + M_0 n^{-1}.$$

Fix $\xi > 0$ and consider all $h \in \mathcal{G}$ such that $\|h - g_s\|^2 = (M_1 + \xi)m^{-4}$. For these h,

$$\psi(h; \boldsymbol{q}) - \psi(g_s; \boldsymbol{q}) \geq (M_1 + \xi)m^{-4} - M_0 n^{-1}.$$

Then $\psi(h; \boldsymbol{q}) > \psi(g^*; \boldsymbol{q})$ when $\|h - g_s\|^2 = (M_1 + \xi)m^{-4}$, for large enough n. Because $\bar{g}$ minimizes $\psi(g; \boldsymbol{q})$ over $g \in \mathcal{G}$, and by convexity of ψ, we must have $\|\bar{g} - g_s\|^2 \leq (M_1 + \xi)m^{-4}$. Hence $\|\bar{g} - g_0\| \leq \|\bar{g} - g_s\| + \|g_s - g^*\| + \|g^* - g_0\| = O_p(m^{-2})$.

To minimize the error $\|\tilde{g} - g_0\|^2$, we set the estimation error and the approximation errors equal. With $m = O_p(n^{1/5})$, $\|\tilde{g} - g_0\|^2 = O_p(n^{-4/5})$, and further by Theorem 3 $\|\tilde{f} - f_0\|^2 = O_p(n^{-2/5})$.

3.4 Piecewise Quadratic Splines

In this section we general the results for piecewise constant spline basis and focus on piecewise quadratic basis functions. We estimate f using piecewise quadratic splines with equally spaced knots $t_1, \ldots, t_{m+3}$. Let d be $t_{j+1} - t_j$ for $j = 1, \ldots, m+2$ and $S = t_{m+3} - t_1$; for convenience we choose the knots so that $h = \ell d$ for a positive integer ℓ. The spline basis functions for estimating f are

$$\delta_j(x) = \begin{cases} 0 & \text{for } x < t_j \\ \frac{2}{3d^2}(x - t_j)^2 & \text{for } x \in [t_j, t_{j+1}) \\ 1 - \frac{4}{3d^2}(x - t_{j+1} - \frac{d}{2})^2 & \text{for } x \in [t_{j+1}, t_{j+2}] \\ \frac{2}{3d^2}(x - t_{j+3})^2 & \text{for } x \in (t_{j+2}, t_{j+3}] \\ 0 & \text{for } x \geq t_{j+3} \end{cases},$$

and the corresponding $\gamma_1(y), \ldots, \gamma_m(y)$ are $\gamma_j(y) = \int_{-\infty}^{\infty} \delta_j(y-x)\phi(x)dx$. Suppose the error density is uniform on $(-h, h)$, we have

$$\gamma_j(y) = \int_{-\infty}^{\infty} \delta_j(y-x)\phi(x)dx$$

$$= \begin{cases} 0 & \text{for } y < t_j - h \\ \frac{1}{9hd^2}(y+h-t_j)^3 & \text{for } y \in [t_j - h, t_{j+1} - h) \\ -\frac{2}{9hd^2}(y+h-t_{j+1}-\frac{1}{2}d)^3 + \frac{1}{2h}(y+h-t_{j+1}) + \frac{d}{12h} & \text{for } y \in [t_{j+1} - h, t_{j+2} - h) \\ \frac{1}{9hd^2}(y+h-t_{j+3})^3 + \frac{2d}{3h} & \text{for } y \in [t_{j+2} - h, t_{j+3} - h) \\ \frac{2d}{3h} & \text{for } y \in [t_{j+3} - h, t_j + h] \\ -\frac{1}{9hd^2}(y-h-t_j)^3 + \frac{2d}{3h} & \text{for } y \in (t_j + h, t_{j+1} + h] \\ \frac{2}{9hd^2}(y-h-t_{j+1}-\frac{1}{2}d)^3 + \frac{1}{2h}(t_{j+2}-y+h) + \frac{d}{12h} & \text{for } y \in (t_{j+1} + h, t_{j+2} + h] \\ \frac{1}{9hd^2}(t_{j+3}-y+h)^3 & \text{for } y \in (t_{j+2} + h, t_{j+3} + h] \\ 0 & \text{for } y > t_{j+3} + h \end{cases}.$$

Define $s_0 = t_1 - ld$ and $\tilde{m} = m + 2l + 1$. The basis functions are illustrated in Figure 3.4 for $m = 17$, where the $\ell = h/d = 4$.

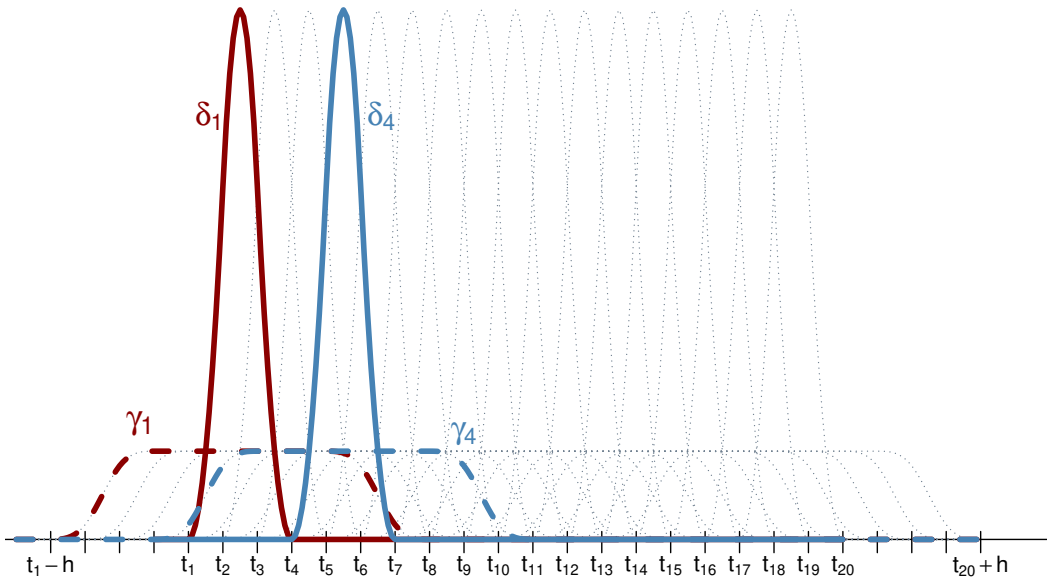

Figure 3.4: Example basis functions for $m = 17$ and $\ell = 4$.

An example data set is shown in Figure 3.5, with a histogram of simulated $X_1, \ldots, X_n$ shown on the left along with the true density shown as a dashed curve. The histogram of $Y_1, \ldots, Y_n$ is shown on right, where $Y_i = X_i + Z_i$ for $Z_i \overset{ind}{\sim} \text{Unif}(-10, 10)$. The solid curve on the right is the least-squares spline density estimate of g, using the γ basis functions, and the solid curve on the left uses the same coefficients with the δ basis functions. The fits are constrained so that $\hat{f}$ is unimodal with mode at zero. The g density can be estimated well, but small variations in $\hat{g}$ can translate to large variations in $\hat{f}$.

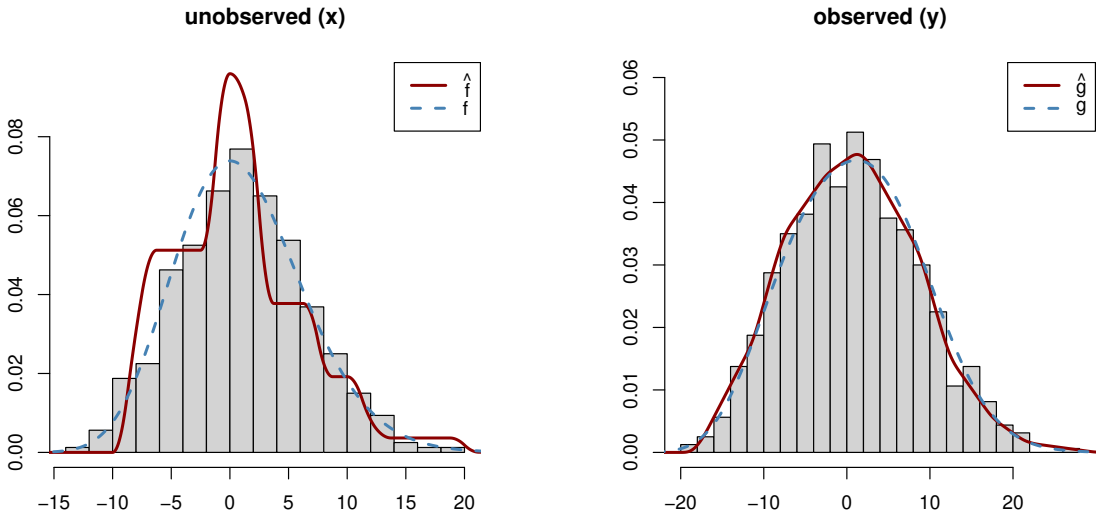

Figure 3.5: Example of simulated data with $n = 800$ and estimated densities constrained to be unimodal with mode (of f) at the origin. Left: the histogram of the unobserved sample from f. Right: the histogram of the observed sample from g.

The choice of the support and the number of knots affects the estimation of the densities. If the support of f is known and finite, a number of knots can be chosen that is sufficiently large for flexibility of the estimate. The over-fitting of f seen in Figure 3.5 can be controlled through penalization as described in Section 3.4.2. If the support of f is unbounded, we expect that with increasing n, the range of observations increases without bound. Especially if the density f is heavy tailed, or is itself "contaminated" so that the observations have "outliers," the range of the data can be large with most of the data near the center. To address the problem of choosing knots

in this case, we actually estimate the density f truncated to a finite support. This support can be chosen by the user, or automatically chosen by the data set. For example, in the simulations in Section 6, we choose the middle 99 percent of the sample, extended by one third of this range upwards and downwards. The support converges to $[(4a-b)/3, (4b-a)/3]$ as the sample size grows, where a and b are the .005th and .995th quantiles of the actual density. For the simulations in Section 3.4.4, this support almost always captures the entire data set.

3.4.1 Large Sample Theory

As the sample size n grows, the number of knots m increases, so that the space d between the knots decreases. In this section our goal is to determine the rate of increase of m that leads to the fastest rate of convergence of $\tilde{f}$. We start with a result that illustrates the difficulty of the deconvolution problem, showing that functions g_1 and g_2 in $\mathcal{G}$ have to be quite close together to guarantee that the corresponding f_1 and f_2 in $\mathcal{F}$ are near each other. The three Lemmas used in the proof are proved in Section 3.7. Define $\|f_1 - f_2\|^2 = \int (f_1(x) - f_2(x))^2 dx$.

Theorem 5. *Theorem 3 holds for piecewise quadratic basis. For f_1 and f_2 in $\mathcal{F}$, $\|f_1 - f_2\| = O(m\|g_1 - g_2\|)$ where g_1 and g_2 are the corresponding functions in $\mathcal{G}$.*

Proof: Let $\boldsymbol{b}_1$ be the vector of spline coefficients for f_1 and g_1, and let $\boldsymbol{b}_2$ be the coefficients for f_2 and g_2. Let $r_j = b_{1j} - b_{2j}$ for $j = 1, \ldots, m$, so that

$$\int (g_1(y) - g_2(y))^2 dy = \int \left[\sum_{j=1}^{m} r_j \gamma_j(y) \right]^2 dy.$$

Lemma 2. *Let $I_j(y) = \frac{2d}{3h}$ for $y \in [t_{j+1} - h + \frac{d}{2}, t_{j+1} + h + \frac{d}{2}]$, $j = 1, \ldots, m$; then*

$$\int (g_1(y) - g_2(y))^2 dy = (1 + O(d)) \int \left[\sum_{i=1}^{m} r_j I_j(y) \right]^2 dy.$$

Because I is a rectangular function, and using $\boldsymbol{r} = \boldsymbol{W}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)$, we can write

$$\begin{aligned}\int (g_1(y) - g_2(y))^2 dy &= (1 + O(d))\frac{4d^3}{9h^2}\boldsymbol{r}^\top \boldsymbol{M}\boldsymbol{r} \\ &= (1 + O(d))\frac{4d^3}{9h^2}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)^\top \boldsymbol{W}^\top \boldsymbol{M}\boldsymbol{W}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2),\end{aligned}$$

where $\boldsymbol{M}$ is an $m \times m$ symmetric Toeplitz matrix with sub-diagonals $2\ell, \ldots, 1$.

Because the spline basis functions all have the same area, the nonzero elements of the matrix $\boldsymbol{W}$ are $\boldsymbol{W}_{i,i} = -1$ and $\boldsymbol{W}_{i+1,i} = 1$. Then $\boldsymbol{Q} = \boldsymbol{W}^\top \boldsymbol{M}\boldsymbol{W}$ is a sparse Toeplitz matrix with $\boldsymbol{Q}_{i,i} = 2$, $\boldsymbol{Q}_{i,i+2\ell} = -1$ for $i = 1, \ldots, m - 2\ell$, and $\boldsymbol{Q}_{i,i-2\ell} = -1$ for $i = 2\ell + 1, \ldots, m$.

By Lemma 1, we have

$$\int (g_1(y) - g_2(y))^2 dy \geq (1 + O(d))\frac{4d^3\epsilon}{9h^2}\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|^2.$$

So $\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|^2 \leq \frac{9h^2}{4d^3\epsilon}\|g_1 - g_2\|^2$ when d is small.

Lemma 3. *The distance between f_1 and f_2 is bounded by the distance between the corresponding $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ as*

$$\frac{29d}{135}\|\boldsymbol{b}_1 - \boldsymbol{b}_2\|^2 \leq \|f_1 - f_2\|^2 \leq \frac{257d}{135}\|\boldsymbol{b}_1 - \boldsymbol{b}_2\|^2.$$

It is straight-forward to show that for $\boldsymbol{b} = \boldsymbol{W}\boldsymbol{\alpha}$, $\|\boldsymbol{b}\|^2 \leq 4\|\boldsymbol{\alpha}\|^2$. Putting everything together, when d is small, we have

$$\|f_1 - f_2\|^2 \leq \frac{257d}{135}\|\boldsymbol{b}_1 - \boldsymbol{b}_2\|^2 \leq \frac{1028d}{135}\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|^2 \leq \frac{257h^2}{15\epsilon d^2}\|g_1 - g_2\|^2 = O(m^2\|g_1 - g_2\|^2)$$

□

To show that the bound in Theorem 5 is attained, consider the densities $f_1(x) = \frac{3}{4d}\delta_1(x)$ and $f_2(x) = \frac{3}{4d}\delta_4(x)$. As can be seen in Figure 3.4, $f_1(x)$ and $f_2(x)$ are not overlapping and the squared

distance between them is

$$\|f_1 - f_2\|^2 = 2\int_{t_1}^{t_4} f_1^2(x)dx = \frac{9}{8d^2}\int_{t_1}^{t_4} \delta_1^2(x)dx = \frac{9}{8d^2}\frac{4d}{5} = \frac{9}{10d},$$

which increases on the order of m. However, $g_1(y) = \frac{3}{4d}\gamma_1(y)$ and $g_2(y) = \frac{3}{4d}\gamma_4(y)$ are more overlapping for larger m, and the squared distance between them is

$$\|g_1(y) - g_2(y)\|^2 \leq \frac{9}{16d^2}\frac{4d^2}{9h^2}6d = \frac{3d}{h^2}.$$

Though g_1 and g_2 get closer together in L_2 distance, the corresponding f_1 and f_2 are getting farther apart. Therefore, to show that $\|\tilde{f} - f_0\|^2$ is small we must show that $\|\tilde{g} - g_0\|^2$ is very small, where f_0 and g_0 represent the true density functions.

Theorem 6. *If f_0 has a bounded third derivative, then $\|\tilde{f} - f_0\|^2 = O_p(n^{-2/3})$.*

Proof: We first find the convergence rate of $\tilde{g}$, and apply Theorem 5 to get the convergence rate of $\tilde{f}$. Let $\boldsymbol{q}$ be the vector of quantiles of g_0 such that the cdf $G_0(q_i) = \frac{i}{n+1}$ for $i = 1, \ldots, n$, and define $\bar{g}$ as the minimizer $\psi(g, \boldsymbol{q})$ over densities $g \in \mathcal{G}$. We split the error into two pieces: $\|\tilde{g} - g_0\| \leq \|\tilde{g} - \bar{g}\| + \|\bar{g} - g_0\|$, where the first term is called the "estimation error" and the second is the "approximation error."

The estimation error is the same as Section 3.3, $\|\tilde{g} - \bar{g}\|^2 = O_p(m/n)$. As for the approximation error, Define $f^*(x) \in \mathcal{F}$ to be a quadratic spline approximation such that $F^*(t_j) = F_0(t_j)$ for $j = 1, \ldots, m+3$. Using Simpson's $1/3$ rule and the bounded derivatives assumption,

$$F_0(t_j) = \frac{d}{3}\left[f_0(t_1) + 4\sum_{i=1}^{\lfloor \frac{j}{2} \rfloor - 1} f_0(t_{2i}) + 2\sum_{i=1}^{\lfloor \frac{j}{2} \rfloor - 1} f_0(t_{2i+1}) + f_0(t_j)\right] + O(d^4),$$

while, because f^* is quadratic between the knots,

$$F^*(t_j) = \frac{d}{3}\left[f^*(t_1) + 4\sum_{i=1}^{\lfloor \frac{j}{2} \rfloor - 1} f^*(t_{2i}) + 2\sum_{i=1}^{\lfloor \frac{j}{2} \rfloor - 1} f^*(t_{2i+1}) + f^*(t_j)\right].$$

Since $F^*(t_j) = F_0(t_j)$, we have $\sum_{i=1}^{j} c_i \left[f_0(t_i) - f^*(t_i)\right] = O(d^3)$ for $j = 1, \ldots, m+3$, where c_j is the corresponding coefficient (1, 2, or 4). So, $|f_0(t_j) - f^*(t_j)| = O(d^3)$ for $j = 1, \ldots, m+3$. Now we use Simpson's $\frac{1}{3}$ rule on $f_0(x)$ and $f^*(x)$, for $j = 1, \ldots, m+3$, so that

$$f_0(t_j) = \frac{d}{3}\left[f_0'(t_1) + 4\sum_{i=1}^{\lfloor \frac{j}{2} \rfloor - 1} f_0'(t_{2i}) + 2\sum_{i=1}^{\lfloor \frac{j}{2} \rfloor - 1} f_0'(t_{2i+1}) + f_0'(t_j)\right] + O(d^4),$$
$$f^*(t_j) = \frac{d}{3}\left[f^{*\prime}(t_1) + 4\sum_{i=1}^{\lfloor \frac{j}{2} \rfloor - 1} f^{*\prime}(t_{2i}) + 2\sum_{i=1}^{\lfloor \frac{j}{2} \rfloor - 1} f^{*\prime}(t_{2i+1}) + f^{*\prime}(t_j)\right] + O(d^4).$$

Similarly we have $|f_0'(t_j) - f^{*\prime}(t_j)| = O(d^2)$ for all $j = 1, \ldots, m+3$. By mean value theorem, there are a_j and $b_j \in (t_j, t_{j+1})$ so that

$$f_0''(a_j) = \frac{f_0'(t_{j+1}) - f_0'(t_j)}{d} \quad \text{and} \quad f^{*\prime\prime}(b_j) = \frac{f^{*\prime}(t_{j+1}) - f^{*\prime}(t_j)}{d}.$$

So for $j = 1, \ldots, m+2$, $|f_0''(a_j) - f^{*\prime\prime}(b_j)| = O(d)$. Using Taylor expansion on f_0' and $f^{*\prime}$,

$$f_0'(t_j) = f_0'(a_j) + f_0''(a_j)(t_j - a_j) + \frac{f_0'''(a_j)}{2}(t_j - a_j)^2 + O(d^3),$$
$$f^{*\prime}(t_j) = f^{*\prime}(b_j) + f^{*\prime\prime}(b_j)(t_j - b_j) + \frac{f^{*\prime\prime\prime}(b_j)}{2}(t_j - b_j)^2 + O(d^3).$$

So we have $|f_0'(a_j) - f^*(b_j)| = O(d^2)$. Now we use expansion on f_0 and f^*,

$$f^*(t_j) = f^*(b_j) + f^{*\prime}(b_j)(t_j - b_j) + \frac{f^{*\prime\prime}(b_j)}{2}(t_j - b_j)^2 + O(d^3).$$

By subtracting we have $|f_0(a_j) - f^*(b_j)| = O(d^3)$ for $j = 1, \ldots, m+2$. From this we can see that $|f_0(x) - f^*(x)| = O(d^3)$ for all x. Using a Taylor expansion at $x \in (t_j, t_{j+1})$,

$$f_0(x) = f_0(a_j) + f_0'(a_j)(x - a_j) + \frac{f_0''(a_j)}{2}(x - a_j)^2 + O(d^3),$$
$$f^*(x) = f^*(b_j) + f^{*\prime}(b_j)(x - b_j) + \frac{f^{*\prime\prime}(b_j)}{2}(x - b_j)^2 + O(d^3).$$

So, $|f_0(x) - f^*(x)| = O(d^3)$.

Now we consider $F^*(x) - F_0(x)$. For $x \in (t_j, t_{j+1}]$,

$$\begin{aligned} F^*(x) - F_0(x) &= F^*(t_j) + \int_{t_j}^{x} f^*(u)du - F_0(t_j) - \int_{t_j}^{x} f_0(u)du \\ &= \int_{t_j}^{x} [f^*(u) - f_0(u)]\, du \\ &= O(d^4). \end{aligned}$$

Then we have, for $g^* \in \mathcal{G}$ with the same coefficients as f^*,

$$\begin{aligned} \int [g^*(y) - g_0(y)]^2 dy &= \frac{1}{4h^2} \int \{[F^*(y+h) - F^*(y-h)] - [F_0(y+h) - F_0(y-h)]\}^2 \, dy \\ &= \frac{1}{4h^2} \int \{[F^*(y+h) - F_0(y+h)] - [F^*(y-h) - F_0(y-h)]\}^2 \, dy \\ &\leq \frac{1}{2h^2} \int \{[F^*(y+h) - F_0(y+h)]^2 + [F^*(y-h) - F_0(y-h)]^2\} dy \\ &= O(d^8). \end{aligned}$$

Recall that $G_0(q_i) = i/(n+1)$. Let q_0 be the lower bound of the support, and define

$$g_s(y) = \frac{1}{n(q_i - q_{i-1})} \text{ for } y \in [q_{i-1}, q_i).$$

Now we show that $\|\bar{g} - g_s\|^2$ is small by first show that $\|g^* - g_s\|^2$ is small using similar approach described in Section 3.3. We have $\|g^* - g_s\| \leq \|g^* - g_0\| + \|g_0 - g_s\| = O(m^{-4}) + O(n^{-1})$. There is M_1 so that for large enough n,

$$\psi(g^*; \boldsymbol{q}) - \psi(g_s; \boldsymbol{q}) \leq M_1 m^{-8} + M_0 n^{-1}.$$

Fix $\xi > 0$ and consider all $h \in \mathcal{G}$ such that $\|h - g_s\|^2 = (M_1 + \xi)m^{-8}$. For these h,

$$\psi(h; \boldsymbol{q}) - \psi(g_s; \boldsymbol{q}) \geq (M_1 + \xi)m^{-8} - M_0 n^{-1}.$$

Then $\psi(h;\boldsymbol{q}) > \psi(g^*;\boldsymbol{q})$ when $\|h - g_s\|^2 = (M_1 + \xi)m^{-8}$, for large enough n. Because $\bar{g}$ minimizes $\psi(g;\boldsymbol{q})$ over $g \in \mathcal{G}$, and by convexity of ψ, we must have $\|\bar{g} - g_s\|^2 \leq (M_1 + \xi)m^{-8}$. Hence $\|\bar{g} - g_0\| \leq \|\bar{g} - g_s\| + \|g_s - g^*\| + \|g^* - g_0\| = O_p(m^{-4})$.

To minimize the error $\|\tilde{g} - g_0\|^2$, we set the estimation error and the approximation error equal to find the optimal m. With m on the order of $n^{1/9}$, $\|\tilde{g} - g_0\|^2 = O_p(n^{-8/9})$, and finally by Theorem 1, $\|\tilde{f} - f_0\|^2 = O_p(n^{-2/3})$.

3.4.2 Penalized Spline Estimator

The example fits in Figure 3.5 illustrate that while the g density can be estimated well, the estimate of the f density can be "wiggly," because small perturbations in g can lead to larger changes in the corresponding f. A penalty to smooth the estimate of f may be used in the estimation of g. Consider the second derivative of $f = \sum_{i=1}^{m} b_j \delta_j$ wtih $\boldsymbol{b} \in \mathcal{B}$; f'' is piecewise constant because the spline basis functions δ_j are piecewise quadratic. Let $\theta_j = f''(y)$ for $y \in (t_j, t_{j+1}), j = 1, \ldots, m+2$, then we have

$$\begin{aligned}
&\theta_1 = \frac{4}{3d^2} b_1, \quad \theta_2 = \frac{4}{3d^2}(-2b_1 + b_2), \\
&\theta_j = \frac{4}{3d^2}(b_{j-2} - 2b_{j-1} + b_j), \text{ for } j = 3, \ldots, m, \\
&\theta_{m+1} = \frac{4}{3d^2}(b_{m-1} - 2b_m), \text{ and } \theta_{m+2} = \frac{4}{3d^2} b_m.
\end{aligned}$$

Define the criterion function $\psi_\lambda(g;\boldsymbol{Y})$ as

$$\psi_\lambda(g;\boldsymbol{Y}) = \int g(y)^2 dy - \frac{2}{n}\sum_{i=1}^{n} g(Y_i) + \lambda \sum_{j=1}^{m+1} (\theta_{j+1} - \theta_j)^2.$$

Let $\boldsymbol{\Delta}$ be a $(m+1) \times m$ matrix such that $\Delta_{i,i} = 3, \Delta_{i+1,i} = -3$ for $i = 1, \ldots, m$ and $\Delta_{i,i+1} = 1, \Delta_{i+2,i} = -1$ for $i = 1, \ldots, m-1$. Then we have

$$\boldsymbol{\Delta b} = \frac{3d^2}{4} \begin{pmatrix} \theta_2 - \theta_1 \\ \theta_3 - \theta_2 \\ \vdots \\ \theta_{m+2} - \theta_{m+1} \end{pmatrix}.$$

To make sure that the $\boldsymbol{H}$ and the penalty matrix change by the same proportion when the data are re-scaled, for example measured in a different unit, we apply a multiple to $\boldsymbol{\Delta}$. Let $\boldsymbol{D} = 4\sqrt{d}\boldsymbol{\Delta}/3$; we have a scale-invariant estimator if the penalty matrix is $\boldsymbol{D}^\top \boldsymbol{D}$. Recall that $\boldsymbol{b} = \boldsymbol{b}_0 + \boldsymbol{W}\boldsymbol{\alpha}$, so the criterion with penalty can be written as

$$\boldsymbol{\alpha}^\top \boldsymbol{W}^\top (\boldsymbol{H} + \lambda \boldsymbol{D}^\top \boldsymbol{D}) \boldsymbol{W}\boldsymbol{\alpha} - 2[\boldsymbol{c} - (\boldsymbol{H} + \lambda \boldsymbol{D}^\top \boldsymbol{D})\boldsymbol{b}_0]^\top \boldsymbol{W}\boldsymbol{\alpha}, \text{ subject to } \boldsymbol{A}\boldsymbol{W}\boldsymbol{\alpha} \geq -\boldsymbol{A}\boldsymbol{b}_0.$$

Figure 3.6 shows the significant impact of different penalty parameters on the estimated density of the underlying density of X. It is also noteworthy that the estimated densities of Y are very close to each other, reinforcing the idea discussed in the previous chapter that estimating f is considerably more challenging than estimating g.

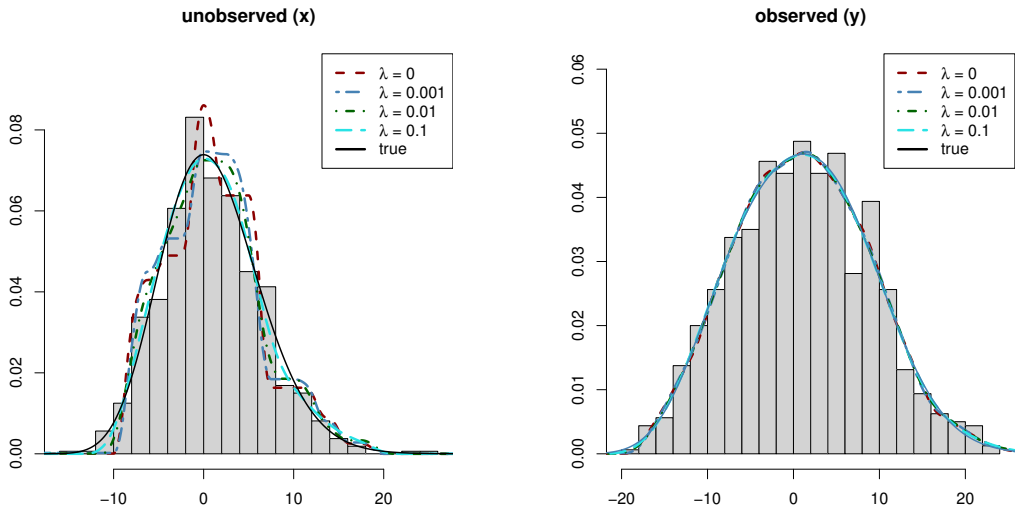

Figure 3.6: Example of the estimated densities with different penalty parameter λ

The proof of the following is in Section 3.8, showing that a small penalty will increase smoothness of the estimate while retaining the convergence rate of the unpenalized estimator.

Theorem 7. *Let $\tilde{g}_\lambda$ minimize $\psi_\lambda(g; Y)$ over densities in $\mathcal{G}$ and let $\tilde{f}_\lambda$ be the corresponding density in $\mathcal{F}$. With $\lambda = O(n^{-5/9})$, $\|\tilde{f}_\lambda - f_0\|^2 = O(n^{-2/3})$.*

The penalty parameter λ is selected by k-fold cross validation as in [Celisse, 2014]. The risk function is defined by

$$E \int [g(y) - \tilde{g}_\lambda(y)]^2 \, dy = E \int g^2(y) dy + E \int \tilde{g}_\lambda^2(y) dy - 2E \int g(y) \tilde{g}_\lambda(y) dy.$$

The first term is constant in λ, so the penalty parameter λ is selected by minimizing the estimated risk

$$\hat{R}_k(\tilde{g}_\lambda) = \int \tilde{g}_\lambda^2(y) dy - \frac{2}{n} \sum_{i=1}^{k} \sum_{j \in n_i} \tilde{g}_{\lambda,(-i)}(y_j) dy,$$

where the data set is partitioned randomly into k subsets, and n_i is a set of indexes of the ith subset. Then $\tilde{g}_{\lambda,(-i)}$ is estimated without the observations in the ith subset.

The effectiveness of the penalty on the "wiggliness" of the estimate of f is illustrated in Figure 3.7, where the simulated data from Figure 3.5 are used. The fit $\hat{f}$ is improved while the fit $\hat{g}$ is similar to the unpenalized estimate.

3.4.3 The Constrained Density Estimator

An advantage of using splines, compared to kernel methods, is that shape constraints are readily imposed using inequality constraints on the spline coefficients. Suppose first that we know the position μ of the mode. A knot is placed at the mode, and a constraint $\boldsymbol{A}\boldsymbol{b} \geq 0$ is used. The elements of the constraint matrix $\boldsymbol{A}$ correspond the slopes of the basis function at the knots, with positive sign for knots below the mode, and negative sign for knots above the mode. To constrain the slope to be zero at the mode, an equality constraint $\boldsymbol{A}_k \boldsymbol{b} = 0$ is imposed, where t_k is the knot

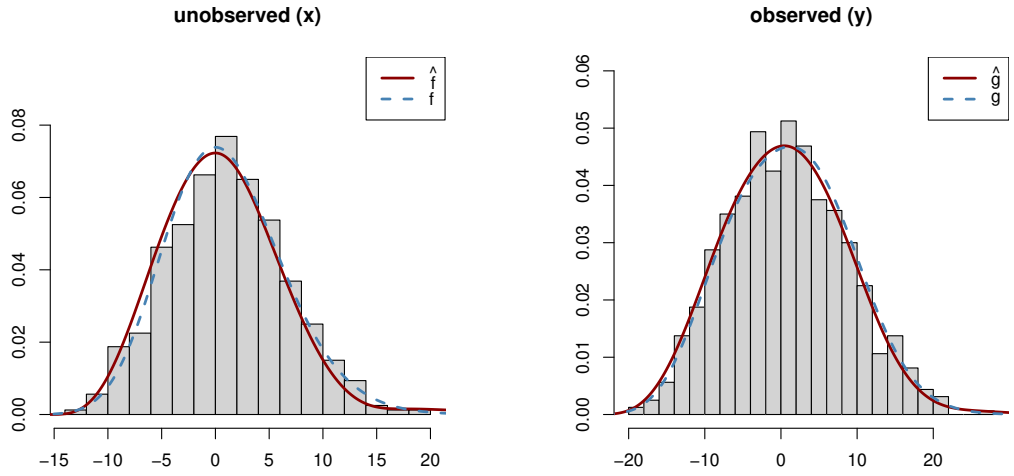

Figure 3.7: The penalized splines estimates with the data of Figure 3.5, showing a smoother $\hat{f}$.

at the mode and $\boldsymbol{A}_k$ is the kth row of $\boldsymbol{A}$. Let $\hat{\boldsymbol{\beta}}_\lambda$ be the solution that minimizes

$$\boldsymbol{\beta}^\top(\boldsymbol{H}+\lambda\boldsymbol{D}^\top\boldsymbol{D})\boldsymbol{\beta}-2(\boldsymbol{c}-\boldsymbol{H}\boldsymbol{b}_0)^\top\boldsymbol{\beta},\ \text{subject to}\ \boldsymbol{a}^\top\boldsymbol{\beta}=0,\ \boldsymbol{A}\boldsymbol{\beta}\geq-\boldsymbol{A}\boldsymbol{b}_\mathbf{0}\ \text{and}\ \boldsymbol{A}_k\boldsymbol{\beta}=0.$$

Notice that

$$\begin{aligned}&\boldsymbol{\beta}^\top(\boldsymbol{H}+\lambda\boldsymbol{D}^\top\boldsymbol{D})\boldsymbol{\beta}-2(\boldsymbol{c}-\boldsymbol{H}\boldsymbol{b}_0)^\top\boldsymbol{\beta}\\&=\|(\boldsymbol{H}+\lambda\boldsymbol{D}^\top\boldsymbol{D})^{-\frac{1}{2}}(\boldsymbol{c}-\boldsymbol{H}\boldsymbol{b}_0)-(\boldsymbol{H}+\lambda\boldsymbol{D}^\top\boldsymbol{D})^{\frac{1}{2}}\boldsymbol{\beta}\|^2-\ \text{constant}.\end{aligned}$$

Then $\hat{\boldsymbol{\beta}}_\lambda$ is the minimizer if ([Silvapulle and Sen, 2011], Chapter 3)

$$[(\boldsymbol{H}+\lambda\boldsymbol{D}^\top\boldsymbol{D})^{-\frac{1}{2}}(\boldsymbol{c}-\boldsymbol{H}\boldsymbol{b}_0)-(\boldsymbol{H}+\lambda\boldsymbol{D}^\top\boldsymbol{D})\hat{\boldsymbol{\beta}}_\lambda]^\top[(\boldsymbol{H}+\lambda\boldsymbol{D}^\top\boldsymbol{D})^{\frac{1}{2}}\boldsymbol{\beta}-(\boldsymbol{H}+\lambda\boldsymbol{D}^\top\boldsymbol{D})^{\frac{1}{2}}\hat{\boldsymbol{\beta}}_\lambda]\leq 0$$
$$\text{for all}\ \boldsymbol{a}^\top\boldsymbol{\beta}=0,\ \boldsymbol{A}\boldsymbol{\beta}\geq-\boldsymbol{A}\boldsymbol{b}_0,\ \text{and}\ \boldsymbol{A}_k\boldsymbol{\beta}=0.$$

With some calculations and plugging in $\boldsymbol{\beta}=\boldsymbol{b}-\boldsymbol{b}_0$,

$$[(\boldsymbol{H}+\lambda\boldsymbol{D}^\top\boldsymbol{D})\hat{\boldsymbol{b}}_\lambda-\lambda\boldsymbol{D}^\top\boldsymbol{D}\boldsymbol{b}_0-\boldsymbol{c}]^\top(\hat{\boldsymbol{b}}_\lambda-\boldsymbol{b})\leq 0,\ \text{for all}\ \boldsymbol{a}^\top\boldsymbol{b}=1,\ \boldsymbol{A}\boldsymbol{b}\geq\mathbf{0},\ \ \text{and}\ \ \boldsymbol{A}_k\boldsymbol{b}=0. \tag{3.2}$$

Similarly, for the unconstrained estimator $\tilde{\boldsymbol{b}}_\lambda$, we have

$$[(\boldsymbol{H}+\lambda \boldsymbol{D}^\top \boldsymbol{D})\tilde{\boldsymbol{b}}_\lambda - \lambda \boldsymbol{D}^\top \boldsymbol{D}\boldsymbol{b}_0 - \boldsymbol{c}]^\top \boldsymbol{b} = 0, \text{ for all } \boldsymbol{b} \in \mathbb{R}^m. \tag{3.3}$$

Now consider

$$\begin{aligned}\|\tilde{g}_\lambda - \bar{g}\|^2 &= (\tilde{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}})^\top \boldsymbol{H}(\tilde{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}}) \\ &= (\tilde{\boldsymbol{b}}_\lambda - \hat{\boldsymbol{b}}_\lambda)^\top \boldsymbol{H}(\tilde{\boldsymbol{b}}_\lambda - \hat{\boldsymbol{b}}_\lambda) + (\hat{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}})^\top \boldsymbol{H}(\hat{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}}) + 2(\tilde{\boldsymbol{b}}_\lambda - \hat{\boldsymbol{b}}_\lambda)^\top \boldsymbol{H}(\hat{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}}).\end{aligned}$$

so

$$\begin{aligned}&\|\tilde{g}_\lambda - \bar{g}\|^2 - \|\hat{g}_\lambda - \bar{g}\|^2 \\ =&\|\tilde{g}_\lambda - \hat{g}_\lambda\|^2 + 2(\tilde{\boldsymbol{b}}_\lambda - \hat{\boldsymbol{b}}_\lambda)^\top \boldsymbol{H}(\hat{\boldsymbol{b}} \\ =&\|\tilde{g}_\lambda - \hat{g}_\lambda\|^2 + 2(\tilde{\boldsymbol{b}}_\lambda - \hat{\boldsymbol{b}}_\lambda)^\top (\boldsymbol{H}+\lambda \boldsymbol{D}^\top \boldsymbol{D})(\hat{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}})\%\& - 2\lambda(\tilde{\boldsymbol{b}}_\lambda - \hat{\boldsymbol{b}}_\lambda)^\top \boldsymbol{D}^\top \boldsymbol{D}(\hat{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}}).\end{aligned}$$

The second term

$$\begin{aligned}&2(\tilde{\boldsymbol{b}}_\lambda - \hat{\boldsymbol{b}}_\lambda)^\top (\boldsymbol{H}+\lambda \boldsymbol{D}^\top \boldsymbol{D})(\hat{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}}) \\ =&2[(\boldsymbol{H}+\lambda \boldsymbol{D}^\top \boldsymbol{D})\tilde{\boldsymbol{b}}_\lambda]^\top (\hat{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}}) - 2[(\boldsymbol{H}+\lambda \boldsymbol{D}^\top \boldsymbol{D})\hat{\boldsymbol{b}}_\lambda]^\top (\hat{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}}) \\ =&2[(\boldsymbol{H}+\lambda \boldsymbol{D}^\top \boldsymbol{D})\tilde{\boldsymbol{b}}_\lambda - \lambda \boldsymbol{D}^\top \boldsymbol{D}\boldsymbol{b}_0 - \boldsymbol{c}]^\top (\hat{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}}) \\ &\quad - 2[(\boldsymbol{H}+\lambda \boldsymbol{D}^\top \boldsymbol{D})\hat{\boldsymbol{b}}_\lambda - \lambda \boldsymbol{D}^\top \boldsymbol{D}\boldsymbol{b}_0 - \boldsymbol{c}]^\top (\hat{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}})\end{aligned}$$

is non-negative by (3.2) and (3.3). Then

$$\begin{aligned}\|\tilde{g}_\lambda - \bar{g}\|^2 - \|\hat{g}_\lambda - \bar{g}\|^2 &\geq 2\lambda(\hat{\boldsymbol{b}}_\lambda - \tilde{\boldsymbol{b}}_\lambda)^\top \boldsymbol{D}^\top \boldsymbol{D}(\hat{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}}) \\ &\geq 2\lambda(\hat{\boldsymbol{b}}_\lambda - \tilde{\boldsymbol{b}}_\lambda)^\top \boldsymbol{D}^\top \boldsymbol{D}(\hat{\boldsymbol{b}}_\lambda - \tilde{\boldsymbol{b}}_\lambda + \tilde{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}}) \\ &\geq 2\lambda\|\boldsymbol{D}(\hat{\boldsymbol{b}}_\lambda - \tilde{\boldsymbol{b}}_\lambda)\| \left[\|\boldsymbol{D}(\hat{\boldsymbol{b}}_\lambda - \tilde{\boldsymbol{b}}_\lambda)\| - \|\boldsymbol{D}(\tilde{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}})\|\right],\end{aligned}$$

i.e.

$$\|\hat{g}_\lambda - \bar{g}\|^2 \leq \|\tilde{g}_\lambda - \bar{g}\|^2 - 2\lambda\|\boldsymbol{D}(\hat{\boldsymbol{b}}_\lambda - \tilde{\boldsymbol{b}}_\lambda)\| \left[\|\boldsymbol{D}(\hat{\boldsymbol{b}}_\lambda - \tilde{\boldsymbol{b}}_\lambda)\| - \|\boldsymbol{D}(\tilde{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}})\|\right].$$

According to the derivations in Sections 3 and 4 we know that $\|\tilde{g}_\lambda - \bar{g}\|^2 = O(n^{-8/9})$ and $\|\tilde{\boldsymbol{b}}_\lambda - \bar{\boldsymbol{b}}\|^2 = O(n^{-5/9})$. For any $\boldsymbol{b}$, $\|\boldsymbol{D}\boldsymbol{b}\|^2 = \boldsymbol{b}^\top \boldsymbol{D}^\top \boldsymbol{D}\boldsymbol{b} = O(n^{-1/9}\|\boldsymbol{b}\|^2)$. So, if $\|\boldsymbol{D}(\hat{\boldsymbol{b}}_\lambda - \tilde{\boldsymbol{b}}_\lambda)\| \leq O(n^{-5/18})$, $\|\hat{g}_\lambda - \bar{g}\|^2 = O(n^{-8/9})$. If $\|\boldsymbol{D}(\hat{\boldsymbol{b}}_\lambda - \tilde{\boldsymbol{b}}_\lambda)\| > O(n^{-5/18})$, $\|\hat{g}_\lambda - \bar{g}\|^2 < \|\tilde{g}_\lambda - \bar{g}\|^2 = O(n^{-8/9})$. Then we have $\|\hat{g}_\lambda - \bar{g}\|^2 = O(n^{-8/9})$ and hence $\|\hat{g}_\lambda - g_0\|^2 = O(n^{-8/9})$.

For the unknown mode case, we first consider that f is symmetric, then g and f have the same mode. This mode can be estimated well from the observed $Y_1, \ldots, Y_n$, using the median. Linear equality constraints can be used to constrain $\hat{f}$ to be symmetric, by putting the estimated mode exactly between two knots, and constraining the $\hat{b}_i = \hat{b}_{m-i}$ for $i = 1, \ldots, m/2$.

If f is not symmetric, we estimate the mode by constructing constraint matrices $\boldsymbol{A}^{(i)}$, $i = 1, \ldots, m-1$, that correspond to the mode of f_0 being between t_i and t_{i+1}, finding the unimodal fit for each $\boldsymbol{A}^{(i)}$, and choosing i to minimize the criterion function. The mode of g may be different from the mode of f, and estimating the mode of g does not result in precise knowledge of the mode of f. From (3.1) we have

$$g'(y) = \frac{1}{2h}\left[f(y+h) - f(y-h)\right] \quad \text{and} \quad g''(y) = \frac{1}{2h}\left[f'(y+h) - f'(y-h)\right],$$

which tells us that the mode of g occurs where $f(y+h) = f(y-h)$, so that if f has no flat spots, we can conclude that g has a unique mode μ_g, that is within $h/2$ units of the mode of f. Let μ_f be

the mode of f_0, and suppose there are $\epsilon > 0$ and $\eta > 0$ so that $f_0''(x) \leq -\epsilon$ on $(\mu_f - \eta, \mu_f + \eta)$. The proof of the following lemma is straight-forward.

Lemma 4. *For such an f_0, if h is any density with mode $\mu_f + \xi$, for $|\xi| < \eta$, then*

$$\int [f_0(x) - h(x)]^2 dx \geq \frac{1}{45}\epsilon^2\xi^5.$$

Let μ_g be the mode of g_0, where $g_0 = f_0 * \phi$. Then $g_0''(\mu_g) < \epsilon\eta/h$, and hence for h with mode $\mu_g + \xi$, we have

$$\int [g_0(x) - h(x)]^2 dx \geq \frac{\epsilon^2\eta^2\xi^5}{45h^2}.$$

Let $\hat{g}_k$ be the minimizer of $\psi(g; \boldsymbol{y})$ under the constraint that the mode of g is in $[t_k, t_{k+1}]$, and suppose that the true mode of g is in $[t_{k_0}, t_{k_0+1}]$. Then

$$\psi(\hat{g}_k; \boldsymbol{y}) - \psi(\hat{g}_{k_0}; \boldsymbol{y}) = \int \hat{g}_k(y)^2 dy - \frac{2}{n}\sum_{i=1}^n \hat{g}_k(y_i) - \int \hat{g}_{k_0}(y)^2 dy + \frac{2}{n}\sum_{i=1}^n \hat{g}_{k_0}(y_i)$$

$$= \int \hat{g}_k(y)^2 dy - 2\int \hat{g}_k(y)g_0(y)dy - \int \hat{g}_{k_0}(y)^2 dy + 2\int \hat{g}_{k_0}(y)g_0(y)dy + O_p(n^{-1/2})$$

$$= \int [\hat{g}_k(y) - g_0(y)]^2 dy - \int [\hat{g}_{k_0}(y) - g_0(y)]^2 dy + O_p(n^{-1/2}).$$

The second integral is of order $n^{-8/9}$. If the distance between the modes of $\hat{g}_k$ and $\hat{g}_{k_0}$ is ξ, then the first integral is at least $\epsilon^2\eta^2\xi^5/(45h)$, by Lemma 4. Therefore if $\hat{\mu}_k$ is the minimizer over all constraint matrices, we must have $\int [\hat{g}_k(y) - g_0(y)]^2 dy$ is at least on the order of $n^{-1/2}$, and ξ is $O_p(n^{-1/10})$. For the unknown mode case, without an assumption of symmetry, we obtain $\|\hat{f} - f_0\| = O_p(n^{-7/36})$. To constrain a density to be bimodal, we choose three knot intervals in which the monotonicity constraints change sign, so that the density is increasing to the first knot interval, decreasing after, then increasing, then decreasing again. We fit the density for many possible choices of the three knot intervals, and choose the one that minimizes the criterion function. The left plot of Figure 3.8 shows a histogram of $n = 1200$ values from the bimodal density f shown as

the dotted curve. These values are not observed; errors uniformly distributed on $(-8, 8)$ are added, to get the observations shown on the right. The density g of the sum is shown as the dotted curve; this is unimodal because of the large error. The penalized spline estimates of the two densities are the solid curves, and the dashed curve on the left is the kernel estimator.

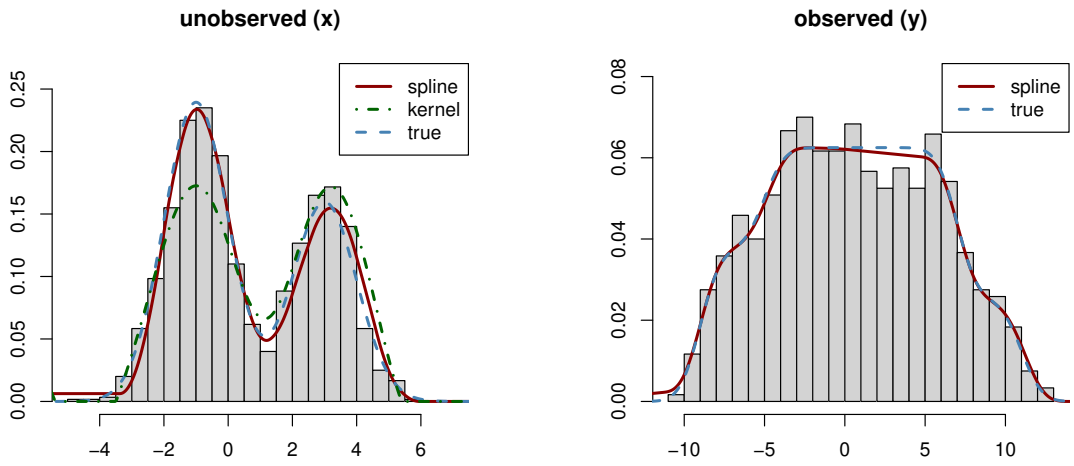

Figure 3.8: Estimating a bimodal density when the observed density is unimodal.

3.4.4 Simulations

Simulations compare the performance of the proposed method with kernel method ([Stefanski and Carroll, 1990]). Samples $x_1, \ldots, x_n$ are generated of varying sizes n as shown in Figure 3.9. We chose four unimodal f distributions with uniform $(-h, h)$ error densities : (a) N(0,1) with $h = 3$, (b) Gamma(4,1) with $h = 4$, (c) .7N(0,1)+.3N(0,2), with $h = 4$, and (d) .7N(0,1)+.3N(0,8) with $h = 4$. For each, the samples $y_1, \ldots, y_n$ are generated by adding a random error that are uniformly distributed on $(-h, h)$. For the four distribution above, we use $h_a = 3, h_b = 4, h_c = 4, h_d = 4$. For (a), (c) and (d), the estimated density $\hat{f}$ is computed using five methods: (i) deconvolution spline methods with known mode and symmetry, (ii) deconvolution spline methods with unknown mode and symmetry, (iii) deconvolution spline methods with known mode and non-symmetry, (iv) deconvolution spline methods with unknown mode and non-symmetry, (v) deconvolution kernel

method. For (b), only methods (iii), (iv) and (v) are applied. The square root of estimated mean integrated squared error (SMISE) is computed as

$$\widehat{\text{SMISE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\int \left[\hat{f}_i(x) - f_0(x)\right]^2 dx},$$

by numerical integration under repeated sampling with $N = 100$. From the results in Figure 3.9 we see that our proposed spline method has smaller SMISE compared with the kernel method, especially when the f density is a mixture that produces "outliers."

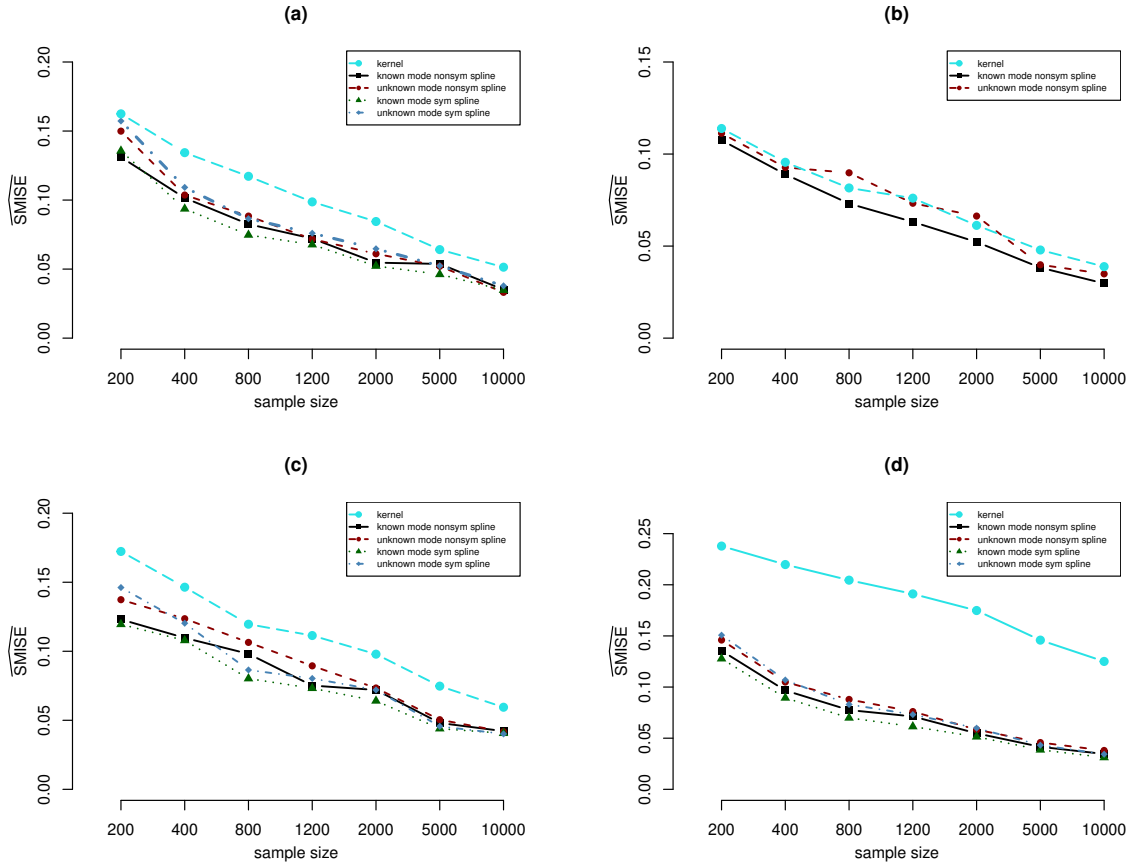

Figure 3.9: The SMISE for (a) N(0,1) and h=3; (b) Gamma(4,1) and h=4; (c) .7N(0,1)+.3N(0,2) and h=4; (d) .7N(0,1)+.3N(0,8) and h=4.

For bimodal densities, we choose (a) .6N(-2,1)+.4N(2,1) and (b) .8N(-2,1)+.2N(2,1), both with $h = 5$. The kernel method has trouble distinguishing the two modes if they are close to each other, or for smaller sample sizes.

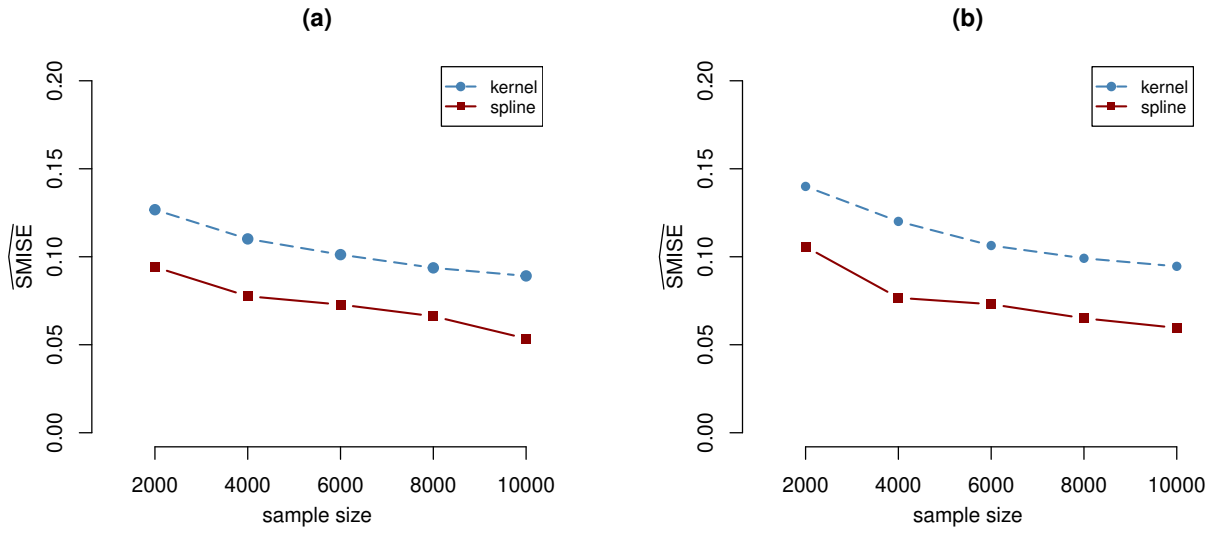

Figure 3.10: The SMISE for (a) .6N(-2,1)+.4N(2,1) with h=5; (b) .8N(-2,1)+.2N(2,1) with h=5.

3.5 Discussion

In this chapter, we propose a deconvolution density estimator using penalized splines. With quadratic splines and uniform errors, our analysis reveals a cube-root convergence rate. The ability to incorporate shape constraints makes the method more predicted and enables the method to outperform kernel method in various scenarios. There are several possible directions for future research. First, our work focuses on uniform error distribution. We would like to generalize our method to other error distributions by starting with the mixtures of uniform distributions. Second, we only demonstrate unimodal and bimodal shape constraints. It would be of interest to study multimodal constraints and propose a criterion to select the number of modes for unknown density f.

3.6 Proof of Lemma 1

For Lemma 1, we note that the matrix $\boldsymbol{Q}$ in Lemma 1 can be characterized in terms of k_1, k_2, and N, where $k_1 = 2l - 3$, N is the integer part of $(m-1)/k_1$, and $k_2 = m - 1 - Nk_1$. Then $k_2 < k_1$, and we define $\boldsymbol{Q}^{N,k_1,k_2}(x)$ to be a $(Nk_1 + k_2) \times (Nk_1 + k_2)$ block matrix with zeroes except $\boldsymbol{Q}^{N,k_1,k_2}_{ii}(x) = x$ for $i = 1, \ldots, Nk_1 + k_2$ and $\boldsymbol{Q}^{N,k_1,k_2}_{k_1+i,i}(x) = \boldsymbol{Q}^{N,k_1,k_2}_{i,k_1+i}(x) = -1$ for $i = 1, \ldots, (N-1)k_1 + k_2$. Then $\boldsymbol{Q}^{N,k_1,k_2}(x)$ can be written as

$$\boldsymbol{Q}^{N,k_1,k_2}(x) = \begin{bmatrix} x\boldsymbol{I}_{k_1\times k_1} & -\boldsymbol{I}_{k_1\times k_1} & & & & \\ -\boldsymbol{I}_{k_1\times k_1} & x\boldsymbol{I}_{k_1\times k_1} & \ddots & & & \\ & \ddots & \ddots & \ddots & & \boldsymbol{B}^{N,k_1,k_2} \\ & & \ddots & \ddots & -\boldsymbol{I}_{k_1\times k_1} & \\ & & & -\boldsymbol{I}_{k_1\times k_1} & x\boldsymbol{I}_{k_1\times k_1} & \\ & & \boldsymbol{C}^{N,k_1,k_2} & & & x\boldsymbol{I}_{k_2\times k_2} \end{bmatrix},$$

where $\boldsymbol{B}^{N,k_1,k_2}$ is a $Nk_1 \times k_2$ sparse matrix with $\boldsymbol{B}^{N,k_1,k_2}_{(N-1)k_1+i,i} = -1$ for $i = 1, \ldots, k_2$ and $\boldsymbol{C}^{N,k_1,k_2} = (\boldsymbol{B}^{N,k_1,k_2})^\top$. Now we study the determinant and eigenvalues of $\boldsymbol{Q}^{N,k_1,k_2}(2)$, which is the $\boldsymbol{Q}$ matrix in the Lemma and in Theorem A1 below. We will show that these depend only on N, not on k_1 or k_2. This will prove the lemma, because N does not increase with n; in fact $N \approx S/(2h)$. As n increases, the sizes of the blocks in $\boldsymbol{Q}^{N,k_1,k_2}$ increase, but the number of blocks does not increase.

We start with a simpler matrix $\boldsymbol{Q}^{N,2,k_2}(x)$, where k_2 can be zero or one. For $N = 5, 6, \ldots,$

$$\psi_N(x) := \det\boldsymbol{Q}^{N,2,k_2}(x) = x\psi_{N-1}(x) - x\psi_{N-3}(x) + \psi_{N-4}(x).$$

To see this we use the method of cofactors:

$$\psi_N(x) = \det \begin{bmatrix} x & 0 & -1 & & & & & \\ 0 & x & 0 & -1 & & & & \\ -1 & 0 & x & 0 & -1 & & & \\ & -1 & 0 & x & 0 & -1 & & \\ & & -1 & 0 & x & 0 & -1 & \\ & & & -1 & 0 & x & 0 & \ddots \\ & & & & -1 & 0 & x & \ddots \\ & & & & & \ddots & \ddots & \ddots \end{bmatrix}$$

$$= x\psi_{N-1}(x) - \det \begin{bmatrix} 0 & x & -1 & & & & \\ -1 & 0 & 0 & -1 & & & \\ & -1 & x & 0 & -1 & & \\ & & 0 & x & 0 & -1 & \\ & & -1 & 0 & x & 0 & \ddots \\ & & & -1 & 0 & x & \ddots \\ & & & & \ddots & \ddots & \ddots \end{bmatrix}$$

$$= x\psi_{N-1}(x) - \det \begin{bmatrix} x & -1 & & & & \\ -1 & x & 0 & -1 & & \\ & 0 & x & 0 & -1 & \\ & -1 & 0 & x & 0 & \ddots \\ & & -1 & 0 & x & \ddots \\ & & & \ddots & \ddots & \ddots \end{bmatrix}$$

$$= x\psi_{N-1}(x) - x\psi_{N-3}(x) - \det \begin{bmatrix} -1 & 0 & & & \\ & x & 0 & -1 & \\ & 0 & x & 0 & \ddots \\ & -1 & 0 & x & \ddots \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

$$= x\psi_{N-1}(x) - x\psi_{N-3}(x) + \psi_{N-4}(x).$$

We have $\psi_1(x) = x, \psi_2(x) = x^2, \psi_3(x) = x^3 - x$, and $\psi_4(x) = x^4 - 2x^2 + 1$. Then we know that $\psi_k(x)$ is a polynomial of x for any $k = 1, 2, \ldots$.

Lemma A1: for each $N > 0$, the smallest eigenvalue of $\boldsymbol{Q}^{N,2,k_2}(2)$ is strictly positive.

Proof: Let e_N be the determinant of $\boldsymbol{Q}^{N,2,k_2}(2)$, i.e. $e_N = \psi_k(2)$. Then we have $e_1 = 2$, $e_2 = 4$, $e_3 = 6$, and $e_4 = 9$, and for $k > 4$, $e_N = 2e_{N-1} - 2e_{N-3} + e_{N-4}$.

Then e_N is and convex in k because for $k > 4$,

$$e_N - 2e_{N-1} + e_{N-2} = [2e_{N-1} - 2e_{N-3} + e_{N-4}] - 2e_{N-1} + e_{N-2} = e_{N-2} - 2e_{N-3} + e_{N-4},$$

so convexity holds by induction. Also by induction, e_N is increasing because

$$e_N - e_{N-1} \geq e_{N-1} - e_{N-2}.$$

Because $\psi_N(x)$ is continuous in x and $\psi_N(2)$ is strictly positive, there is an $\epsilon > 0$ (depending on N) such that $\psi_N(2 - \epsilon)$ is positive for all N. So $\boldsymbol{Q}^{N,2,k_2}(2 - \epsilon)$ is positive definite. Let λ be the eigenvalues of $\boldsymbol{Q}^{N,2,k_2}(2)$, then we have

$$\det(\boldsymbol{Q}^{N,2,k_2}(2) - \lambda \boldsymbol{I}) = \det(\boldsymbol{Q}^{N,2,k_2}(2) - \epsilon \boldsymbol{I} - (\lambda - \epsilon)\boldsymbol{I}) = \det(\boldsymbol{Q}^{N,2,k_2}(2 - \epsilon) - (\lambda - \epsilon)\boldsymbol{I}) = 0$$

So, the eigenvalues of $\boldsymbol{Q}^{N,2,k_2}(2 - \epsilon)$ are $\lambda - \epsilon$ and they are positive. Then the lowest eigenvalue of $\boldsymbol{Q}^{N,2,k_2}(2)$ is at least ϵ, for $k_2 = 0$ or $k_2 = 1$.

Next, we show that the eigenvalues of $\boldsymbol{Q}^{N,k_1,k_2}(2)$, for any $k_1 > 2$ and $0 < k_2 < k_1$, have the same values as the eigenvalues of $\boldsymbol{Q}^{N,2,k_2}(2)$, (with growing multiplicities).

Lemma A2: Define $d_1(x) = x, d_2(x) = x^2 - 1, d_N(x) = x d_{N-1}(x) - d_{N-2}(x)$ for $N = 3, 4, \ldots$. Then we have

$$\det(\boldsymbol{Q}^{N+2,k_1,k_2})(x) = \det(x\boldsymbol{I}_{k_2 \times k_2} - \boldsymbol{C}^{N+1,k_1,k_2}(\boldsymbol{Q}^{N+1,k_1,0}(x))^{-1}\boldsymbol{B}^{N+1,k_1,k_2}) = \left(\frac{d_{N+2}(x)}{d_{N+1}(x)}\right)^{k_2}$$

for $0 < k_2 < k_1$.

Proof: Notice $(\boldsymbol{Q}^{N,k_1,0}(x))^{-1}_{Nk_1,Nk_1} = (\boldsymbol{Q}^{N,1,0}(x))^{-1}_{N,N}$, in other words, the last diagonal element of $(\boldsymbol{Q}^{N,k_1,0}(x))^{-1}$ and $(\boldsymbol{Q}^{N,1,0}(x))^{-1}$ are the same.

$$(\boldsymbol{Q}^{N,1,0}(x))^{-1} = \left[\begin{array}{ccc|c} & & & 0 \\ & \boldsymbol{Q}^{N-1,1,0}(x) & & \vdots \\ & & & 0 \\ & & & -1 \\ \hline & & & \\ 0 & \cdots & 0 & -1 \end{array} \middle| \begin{array}{c} \\ \\ \\ \\ \hline \\ x \end{array}\right]^{-1}.$$

Let $u_N(x) = (\boldsymbol{Q}^{N,1,0}(x))^{-1}_{N,N}$, then

$$u_N(x) = \left[x - \begin{pmatrix} 0 & \cdots & 0 & -1 \end{pmatrix} (\boldsymbol{Q}^{N-1,1,0}(x))^{-1} \begin{pmatrix} 0 & \cdots & 0 & -1 \end{pmatrix}^\top\right]^{-1} = [x - u_{N-1}(x)]^{-1}$$

for $N = 3, 4, \ldots$. Also $u_2(x) = x/(x^2-1) = d_1(x)/d_2(x)$, so we have $u_N(x) = 1/(x - u_{N-1}(x)) = 1/(x - d_{N-2}(x)/d_{N-1}(x)) = d_{N-1}(x)/d_N(x)$ by induction. Then

$$\begin{aligned} &\det(x\boldsymbol{I}_{k_2\times k_2} - \boldsymbol{C}^{N+1,k_1,k_2}(\boldsymbol{Q}^{N+1,k_1,0}(x))^{-1}\boldsymbol{B}^{N+1,k_1,k_2}) \\ &= \det(x\boldsymbol{I}_{k_2\times k_2} - u_{N+1}(x)\boldsymbol{I}_{k_2\times k_2}) \\ &= (x - u_{N+1}(x))^{k_2} \\ &= \left(x - \frac{d_N(x)}{d_{N+1}(x)}\right)^{k_2} \\ &= \left(\frac{d_{N+2}(x)}{d_{N+1}(x)}\right)^{k_2} \end{aligned}$$

In next lemma, we start with $\boldsymbol{Q}^{N+1,k_1,0}(x)$ since it can be written as

$$\boldsymbol{Q}^{N+1,k_1,0}(x) = \begin{pmatrix} \boldsymbol{Q}^{N,k_1,k_2}(x) & \boldsymbol{D}^{N,k_1,k_2} \\ \boldsymbol{E}^{N,k_1,k_2} & x\boldsymbol{I}_{(k_1-k_2)\times(k_1-k_2)} \end{pmatrix},$$

where $\boldsymbol{D}^{N,k_1,k_2}$ is a $(Nk_1 + k_2) \times (k_1 - k_2)$ sparse matrix with $\boldsymbol{D}^{N,k_1,k_2}_{(N-1)k_1+k_2+i,i} = -1$ for $i = 1, \ldots, k_1 - k_2$ and $\boldsymbol{E}^{N,k_1,k_2} = (\boldsymbol{D}^{N,k_1,k_2})^\top$.

Lemma A3: With $\boldsymbol{D}^{N,k_1,k_2}$ and $\boldsymbol{E}^{N,k_1,k_2}$ defined as above, we have

$$\det(x\boldsymbol{I}_{(k_1-k_2)\times(k_1-k_2)} - \boldsymbol{E}^{N,k_1,k_2}(\boldsymbol{Q}^{N,k_1,k_2}(x))^{-1}\boldsymbol{D}^{N,k_1,k_2})) = \left(\frac{d_{N+1}(x)}{d_N(x)}\right)^{k_1-k_2}$$

for $0 < k_2 < k_1$.

Proof:

$$(\boldsymbol{Q}^{N,k_1,k_2}(x))^{-1} = \begin{pmatrix} \boldsymbol{Q}^{N,k_1,0}(x) & \boldsymbol{B}^{N,k_1,k_2} \\ \boldsymbol{C}^{N,k_1,k_2} & x\boldsymbol{I}_{k_2\times k_2} \end{pmatrix}^{-1}.$$

So the bottom right of $(\boldsymbol{Q}^{N,k_1,k_2}(x))^{-1}$ is

$$\begin{aligned} &\left[x\boldsymbol{I}_{k_2\times k_2} - \boldsymbol{C}^{N,k_1,k_2}(\boldsymbol{Q}^{N,k_1,0}(x))^{-1}\boldsymbol{B}^{N,k_1,k_2}\right]^{-1} \\ &= [x\boldsymbol{I}_{k_2\times k_2} - u_{N1}(x)\boldsymbol{I}_{k_2\times k_2}]^{-1} \\ &= u_{N+1}(x)\boldsymbol{I}_{k_2\times k_2} \end{aligned}$$

And the top left block of $(\boldsymbol{Q}^{N,k_1,k_2}(x))^{-1}$ is

$$
\begin{aligned}
&(\boldsymbol{Q}^{N,k_1,0}(x))^{-1} + (\boldsymbol{Q}^{N,k_1,0}(x))^{-1}\boldsymbol{B}^{N,k_1,k_2}\left[u_{N+1}(x)\boldsymbol{I}_{k_2\times k_2}\right]^{-1}\boldsymbol{C}^{N,k_1,k_2}(\boldsymbol{Q}^{N,k_1,0}(x))^{-1} \\
&= (\boldsymbol{Q}^{N,k_1,0}(x))^{-1} + (\boldsymbol{Q}^{N,k_1,0}(x))^{-1}\begin{pmatrix} \boldsymbol{0}_{(N-1)k_1\times(N-1)k_1} & & \\ & \frac{1}{u_{N+1}(x)}\boldsymbol{I}_{k_2\times k_2} & \\ & & \boldsymbol{0}_{(k_1-k_2)\times(k_1-k_2)} \end{pmatrix}(\boldsymbol{Q}^{N,k_1,0}(x))^{-1} \\
&= \begin{pmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} \\ \boldsymbol{T}_{21} & u_N(x)\boldsymbol{I}_{(k_1-k_2)\times(k_1-k_2)} \end{pmatrix}
\end{aligned}
$$

where $\boldsymbol{T}_{11}, \boldsymbol{T}_{12}, \boldsymbol{T}_{21}$ can be computed but we are more interested in the bottom right block.

Since $\boldsymbol{E}^{N,k_1,k_2}(\boldsymbol{Q}^{N,k_1,k_2}(x))^{-1}\boldsymbol{D}^{N,k_1,k_2}$ selects the $((N-1)k_1+k_2+1)$th to the $((N-1)k_1)$th elements in the diagonal, we can get

$$
\begin{aligned}
&\det(x\boldsymbol{I}_{(k_1-k_2)\times(k_1-k_2)} - \boldsymbol{E}^{N,k_1,k_2}(\boldsymbol{Q}^{N,k_1,k_2}(x))^{-1}\boldsymbol{D}^{N,k_1,k_2})) \\
&= \det(x\boldsymbol{I}_{(k_1-k_2)\times(k_1-k_2)} - u_N(x)\boldsymbol{I}_{(k_1-k_2)\times(k_1-k_2)}) \\
&= (x-u_N(x))^{k_1-k_2} \\
&= \left(x - \frac{d_{N-1}(x)}{d_N(x)}\right)^{k_1-k_2} \\
&= \left(\frac{d_{N+1}(x)}{d_N(x)}\right)^{k_1-k_2}
\end{aligned}
$$

Lemma A6: We can write the determinant of $\boldsymbol{Q}^{N,k_1,k_2}(x)$ in terms of N, k_1, and k_2 as follows.

$$\det\boldsymbol{Q}^{N,k_1,k_2}(x) = d_N(x)^{k_1-k_2}d_{N+1}(x)^{k_2}.$$

Proof: We prove this lemma by induction. First we consider the cases when $k_2 = 0$. When $N = 2$,

$$\boldsymbol{Q}^{2,k_1,0}(x) = \begin{pmatrix} x\boldsymbol{I}_{k_1\times k_1} & -\boldsymbol{I}_{k_1\times k_1} \\ -\boldsymbol{I}_{k_1\times k_1} & x\boldsymbol{I}_{k_1\times k_1} \end{pmatrix},$$

so

$$\det\boldsymbol{Q}^{2,k_1,0}(x) = \det(x\boldsymbol{I}_{k_1\times k_1})\det(x\boldsymbol{I}_{k_1\times k_1} - \frac{1}{x}\boldsymbol{I}_{k_1\times k_1}) = (x^2-1)^{k_1} = d_2(x)^{k_1}.$$

Suppose we have $\det\boldsymbol{Q}^{N,k_1,0}(x) = d_N(x)^{k_1}$, now consider $\boldsymbol{Q}^{N+1,k_1,0}(x)$,

$$\begin{aligned}
&\det\boldsymbol{Q}^{N+1,k_1,0}(x) \\
&= \det\boldsymbol{Q}^{N,k_1,0}(x)\det\left[x\boldsymbol{I}_{k_1\times k_1} - \begin{pmatrix} \mathbf{0} & \dots & \mathbf{0} & -\boldsymbol{I} \end{pmatrix}(\boldsymbol{Q}^{N,k_1,0}(x))^{-1}\begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ -\boldsymbol{I} \end{pmatrix}\right] \\
&= d_N(x)^{k_1}\det\left[(x-u_n(x))\boldsymbol{I}_{k_1\times k_1}\right] \\
&= d_N(x)^{k_1}(x-u_N(x))^{k_1} \\
&= d_{N+1}(x)^{k_1}
\end{aligned}$$

Next, consider $0 < k_2 < k_1$. When $N = 2$,

$$\boldsymbol{Q}^{2,k_1,k_2}(x) = \begin{pmatrix} \boldsymbol{Q}^{2,k_1,0} & \boldsymbol{B}^{2,k_1,k_2} \\ \boldsymbol{C}^{2,k_1,k_2} & x\boldsymbol{I}_{k_2\times k_2} \end{pmatrix},$$

so

$$\begin{aligned}
\det\boldsymbol{Q}^{2,k_1,k_2}(x) &= \det\boldsymbol{Q}^{2,k_1,0}(x)\det(x\boldsymbol{I}_{k_2\times k_2} - \boldsymbol{C}^{2,k_1,k_2}(\boldsymbol{Q}^{2,k_1,0}(x))^{-1}\boldsymbol{B}^{2,k_1,k_2}) \\
&= (x^2-1)^{k_1}\det(x\boldsymbol{I}_{k_2\times k_2} - \frac{x}{x^2-1}\boldsymbol{I}_{k_2\times k_2}) \\
&= (x^2-1)^{k_1-k_2}x^{k_2}(x^2-2)^{k_2} \\
&= d_2(x)^{k_1-k_2}d_3(k_2).
\end{aligned}$$

Suppose we have det$\boldsymbol{Q}^{N,k_1,k_2}(x) = d_N(x)^{k_1-k_2} d_{N+1}(x)^{k_2}$, now consider $\boldsymbol{Q}^{N+1,k_1,k_2}(x)$,

$$\begin{aligned}
\det\boldsymbol{Q}^{N+1,k_1,k_2}(x) &= \det\boldsymbol{Q}^{N+1,k_1,0}(x)\det(x\boldsymbol{I}_{k_2\times k_2} - \boldsymbol{C}^{N+1,k_1,k_2}(\boldsymbol{Q}^{N+1,k_1,0}(x))^{-1}\boldsymbol{B}^{N+1,k_1,k_2}) \\
&= \det\boldsymbol{Q}^{N+1,k_1,0}(x)(\frac{d_{N+2}(x)}{d_{N+1}(x)})^{k_2} \\
&= \det\boldsymbol{Q}^{N,k_1,k_2}(x)\det(x\boldsymbol{I}_{(k_1-k_2)\times(k_1-k_2)} \\
&\qquad -\boldsymbol{E}^{N,k_1,k_2}(\boldsymbol{Q}^{N,k_1,k_2}(x))^{-1}\boldsymbol{D}^{N,k_1,k_2})(\frac{d_{N+2}(x)}{d_{N+1}(x)})^{k_2} \\
&= d_N(x)^{k_1-k_2} d_{N+1}(x)^{k_2}(\frac{d_{N+1}(x)}{d_N(x)})^{k_1-k_2}(\frac{d_{N+2}(x)}{d_{N+1}(x)})^{k_2} \\
&= d_{N+1}(x)^{k_1-k_2} d_{N+2}(x)^{k_2},
\end{aligned}$$

So, this lemma holds for all $0 \le k_2 < k_1$.

Theorem A1: For any integer $N > 0, k_1 > 0$ and $0 \le k_2 < k_1$, there exists an $\epsilon > 0$ depending only on N, such that the lowest eigenvalue of $Q^{N,k_1,k_2}(2)$ is at least ϵ.

Proof: From Lemma A6 we know that for any $N > 0, k_1 > 0, 0 < k_2 < k_1$, the lowest eigenvalue of $\boldsymbol{Q}^{N,k_1,k_2}(2)$ is the same as the lowest eigenvalue of $\boldsymbol{Q}^{N,2,1}(2)$. Similarly, for any $N > 0, k_1 > 0, k_2 = 0$, the lowest eigenvalue of $\boldsymbol{Q}^{N,k_1,k_2}(2)$ is the same as the lowest eigenvalue of $\boldsymbol{Q}^{N,2,0}(2)$. So by Lemma A1, there exists an $\epsilon > 0$ depending only on N such that the lowest eigenvalue of $Q^{N,k_1,k_2}(2)$ is at least ϵ.

Proof of Lemma 1: Let $r = S/(2h) = m/(2\ell)$. Define s as the remainder when $m-1$ is divided by $2l-3$. Then the $(m-1)\times(m-1)$ matrix $\boldsymbol{Q}$ can be written as $\boldsymbol{Q}^{\lfloor\frac{m-1}{2\ell-3}\rfloor,2\ell-3,s}(2)$. Since $\frac{m-1}{2\ell-3} = \frac{S/d-3}{2h/d-3}$ decreases to r when d decreases. So by Theorem A1, there exists an $\epsilon > 0$ depending only on r such that the lowest eigenvalue of Q is at least ϵ. □

3.7 Proof of the Lemmas for Theorem 5

To prove Lemma 2, we write

$$\begin{aligned}\int (g_1(y) - g_2(y))^2 dy &= \int \left[\sum_{i=1}^m r_i \gamma_i(y)\right]^2 dy = \int \left[\sum_{i=1}^m r_i \left(I_i(y) + \xi_i(y)\right)\right]^2 dy \\ &= \sum_{i=1}^m r_i^2 \int [I_i(y)^2 + 2I_i(y)\xi_i(y) + \xi_i(y)^2] dy \\ &+ 2\sum_{i=2}^m \sum_{j<i} \int [I_i(y)I_j(y) + I_i(y)\xi_j(y) + I_j(y)\xi_i(y) + \xi_i(y)\xi_j(y)] dy\end{aligned}$$

First consider $\sum_{i=1}^m r_i^2 \int [I_i(y)^2 + 2I_i(y)\xi_i(y) + \xi_i(y)^2] dy$. For each i, we have

$$\int I_i(y)^2 dy = \frac{8d^2}{9h}, \text{ and } \int I_i(y)\xi_i(y) dy = \frac{13d^3}{144h^2}, \text{ and } \int \xi_i(y)^2 dy \leq \int \frac{d^2}{9h^2} dy \leq \frac{2d^3}{3h^2},$$

so that

$$\sum_{i=1}^m r_i^2 \int [I_i(y)^2 + 2I_i(y)\xi_i(y) + \xi_i(y)^2] dy = (1 + O(d)) \sum_{i=1}^m r_i \int I_i(y)^2 dy.$$

Now consider the second term

$$2\sum_{i=2}^m \sum_{j<i} \int [I_i(y)I_j(y) + I_i(y)\xi_j(y) + I_j(y)\xi_i(y) + \xi_i(y)\xi_j(y)] dy$$

For each i, we can show that for $j < i$

$$\int I_i(y)I_j(y)dy = \begin{cases} \frac{4d^2}{9h^2}(2h-(i-j)d) & \text{for } j \geq i+1-2\ell \\ 0 & \text{o.w.} \end{cases}$$

$$\int I_i(y)\xi_j(y)dy = \begin{cases} -\frac{d^3}{432h^2} & \text{for } j = i-1 \\ 0 & \text{for } j = i-2,\ldots,i-2\ell+2 \\ \frac{d^3}{864h^2} & \text{for } j = i-2\ell+1 \\ \frac{13d^3}{144h^2} & \text{for } j = i-2\ell \\ \frac{d^3}{864h^2} & \text{for } j = i-2\ell-1 \\ 0 & \text{for } j = i-2\ell-2,\ldots,1 \end{cases}$$

$$\left|\int \xi_i(y)\xi_j(y)dy\right| \leq \begin{cases} \frac{2d^3}{3h^2} & \text{for } j = i-1, i-2, i-2\ell-2,\ldots,i-2\ell+2 \\ 0 & \text{otherwise} \end{cases}$$

So, for each i we have

$$\begin{aligned} &\sum_{j<i} r_j \int [I_i(y)I_j(y)dy + I_i(y)\xi_j(y) + I_j(y)\xi_i(y) + \xi_i(y)\xi_j(y)]dy \\ =&(1+O(d))\sum_{j<i} r_j \int I_i(y)I_j(y)dy \end{aligned}$$

Hence

$$\begin{aligned} &2\sum_{i=2}^{m}\sum_{j<i} r_i r_j \int [I_i(y)I_j(y) + I_i(y)\xi_j(y) + I_j(y)\xi_i(y) + \xi_i(y)\xi_j(y)]dy \\ =&\ 2(1+O(d))\sum_{i=2}^{m}\sum_{j<i} r_i r_j \int I_i(y)I_j(y)dy \end{aligned}$$

Combining the two terms we have

$$\begin{aligned}
&\int (g_1(y) - g_2(y))^2 dy \\
=&(1 + O(d)) \sum_{i=1}^{m} r_i \int I_i(y)^2 dy + 2(1 + O(d)) \sum_{i=2}^{m} \sum_{j<i} r_i r_j \int I_i(y) I_j(y) dy \\
=&(1 + O(d)) \int \left[\sum_{i=1}^{m} r_i I_i(y) \right]^2 dy.
\end{aligned}$$

□

For Lemma 3:

$$\begin{aligned}
\int (f_1(x) - f_2(x))^2 dx &= \sum_{i=1}^{m+2} \int_{t_j}^{t_{j+1}} \left[\sum_{j=1}^{m} r_j \delta_j(x) \right]^2 dx \\
&= \frac{d}{135} \left[\sum_{i=1}^{m} 132 r_i^2 + \sum_{i=1}^{m-1} 104 r_i r_{i+1} + \sum_{i=1}^{m-2} 4 r_i r_{i+2} \right].
\end{aligned}$$

On one hand,

$$\begin{aligned}
&\sum_{i=1}^{m} 132 r_i^2 + \sum_{i=1}^{m-1} 104 r_i r_{i+1} + \sum_{i=1}^{m-2} 4 r_i r_{i+2} \\
=& \sum_{i=1}^{m} 29 r_i^2 + 54 r_1^2 + 33 r_m^2 + \sum_{i=1}^{m-2} (r_i + 4 r_{i+1} + 2 r_{i+2})^2 + \sum_{i=1}^{m-1} 2 (4 r_i + 5 r_{i+1})^2 \\
&+ (4 r_1 + 2 r_2)^2 + (r_{m-1} + 4 r_m)^2 \\
\geq& \sum_{i=1}^{m} 29 r_i^2.
\end{aligned}$$

On the other hand, we can also find a upper bound

$$\begin{aligned}
&\sum_{i=1}^{m} 132 r_i^2 + \sum_{i=1}^{m-1} 104 r_i r_{i+1} + \sum_{i=1}^{m-2} 4 r_i r_{i+2} \\
=& \sum_{i=1}^{m} 257 r_i^2 - 76 r_1^2 - 33 r_m^2 - \sum_{i=1}^{m-2} (r_i + 4 r_{i+1} - 2 r_{i+2})^2 - \sum_{i=1}^{m-1} 2 (4 r_i - 6 r_{i+1})^2 \\
&- (4 r_1 - 2 r_2)^2 - (r_{m-1} + 4 r_m)^2 \leq \sum_{i=1}^{m} 257 r_i^2.
\end{aligned}$$

So,

$$\frac{29d}{135}\|\boldsymbol{b}_1 - \boldsymbol{b}_2\|^2 \leq \|f_1 - f_2\|^2 \leq \frac{257d}{135}\|\boldsymbol{b}_1 - \boldsymbol{b}_2\|^2. \tag{3.4}$$

3.8 Proof of Theorem 7

We use $\tilde{\boldsymbol{\alpha}}_\lambda$ to denote the unconstrained penalized estimator, i.e.

$$\tilde{\boldsymbol{\alpha}}_\lambda = (\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W} + \lambda \boldsymbol{W}^\top \boldsymbol{D}^\top \boldsymbol{D}\boldsymbol{W})^{-1}\boldsymbol{W}^\top(\boldsymbol{c} - \boldsymbol{H}\boldsymbol{b}_0).$$

By Woodbury matrix identity, we have

$$\begin{aligned}
\tilde{\boldsymbol{\alpha}}_\lambda &= (\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top(\boldsymbol{c} - \boldsymbol{H}\boldsymbol{b}_0) \\
&\quad - (\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top(\frac{1}{\lambda}\boldsymbol{I} + \boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top)^{-1}\boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top(\boldsymbol{c} - \boldsymbol{H}\boldsymbol{b}_0) \\
&= \tilde{\boldsymbol{\alpha}} - (\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top(\frac{1}{\lambda}\boldsymbol{I} + \boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top)^{-1}\boldsymbol{D}\boldsymbol{W}\tilde{\boldsymbol{\alpha}}
\end{aligned}$$

so that the difference in the penalized and unpenalized solution is

$$\tilde{\boldsymbol{\alpha}}_\lambda - \tilde{\boldsymbol{\alpha}} = -\lambda(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top(\boldsymbol{I} + \lambda \boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top)^{-1}\boldsymbol{D}\boldsymbol{W}\tilde{\boldsymbol{\alpha}}.$$

Then we have

$$\tilde{\boldsymbol{\beta}}_\lambda - \tilde{\boldsymbol{\beta}} = -\lambda \boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top(\boldsymbol{I} + \lambda \boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top)^{-1}\boldsymbol{D}\tilde{\boldsymbol{\beta}},$$

and

$$\|\tilde{\boldsymbol{\beta}}_\lambda - \tilde{\boldsymbol{\beta}}\|^2 \leq \lambda^2 \text{maxeval}\left[\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top(\boldsymbol{I} + \lambda \boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top)^{-1}\boldsymbol{D}\right]^2 \|\tilde{\boldsymbol{\beta}}\|^2.$$

By the definition of $\boldsymbol{H}$ and $\boldsymbol{W}$, we know that

$$\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \text{ and } \boldsymbol{D}^\top(\boldsymbol{I}+\lambda \boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top)^{-1}\boldsymbol{D}$$

are both symmetrical, positive semi-definite matrices. Therefore

$$\begin{aligned}\|\tilde{\boldsymbol{\beta}}_\lambda - \tilde{\boldsymbol{\beta}}\|^2 \leq & \lambda^2 \text{maxeval}\left[\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top\right]^2 \times \\ & \text{maxeval}\left[\boldsymbol{D}^\top(\boldsymbol{I}+\lambda \boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top)^{-1}\boldsymbol{D}\right]^2 \|\tilde{\boldsymbol{\beta}}\|^2.\end{aligned}$$

Now we look into the maximum eigenvalues separately. First, by [Higham and Cheng, 1998] we have

$$\text{maxeval}\left[\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top\right] \leq \text{maxeval}\left[(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\right] \text{maxeval}\left[\boldsymbol{W}\boldsymbol{W}^\top\right]$$

By Lemma 1, we have $\|\tilde{g}_\lambda - \tilde{g}\|^2 = (\tilde{\boldsymbol{\alpha}}_\lambda - \tilde{\boldsymbol{\alpha}})^\top \boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W}(\tilde{\boldsymbol{\alpha}}_\lambda - \tilde{\boldsymbol{\alpha}}) \geq \frac{4\epsilon d^3}{9h^2}\|\tilde{\boldsymbol{\alpha}}_\lambda - \tilde{\boldsymbol{\alpha}}\|^2$, so that the lowest eigenvalue of $\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W}$ is at least $\frac{4\epsilon d^3}{9h^2}$. As for $\boldsymbol{W}\boldsymbol{W}^\top$, for any $\boldsymbol{x}$ with $\|\boldsymbol{x}\|^2 = 1$,

$$\boldsymbol{x}^\top \boldsymbol{W}\boldsymbol{W}^\top \boldsymbol{x} = 2\sum_{i=1}^{m} x_i - 2\sum_{i=1}^{m-1} x_i x_{i+1} \leq \sum_{i=1}^{m} x_i + \sum_{i=1}^{m-1}(x_i^2 + x_{i+1}^2) \leq 4.$$

Therefore we have

$$\text{maxeval}\left[\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top\right] \leq \frac{9h^2}{\epsilon d^3}.$$

It is straight-forward to show that the largest eigenvalue $\boldsymbol{D}^\top \boldsymbol{D}$ is $O(d)$. Then,

$$\text{maxeval}\left[\boldsymbol{D}^\top(\boldsymbol{I}+\lambda \boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top)^{-1}\boldsymbol{D}\right]$$

is at most

$$\text{maxeval}\left[\boldsymbol{D}^\top \boldsymbol{D}\right] \times \text{maxeval}\left[(\boldsymbol{I} + \lambda \boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top)^{-1}\right]$$

For any $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|^2 = 1$,

$$\boldsymbol{x}^\top(\boldsymbol{I} + \lambda \boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top)\boldsymbol{x} = \boldsymbol{x}^\top\boldsymbol{x} + \lambda \boldsymbol{x}^\top \boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}\boldsymbol{x} \geq 1.$$

So the maximum eigen value of $(\boldsymbol{I} + \lambda \boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{H}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{D}^\top)^{-1}$ is less than or equal to O(d). Now we have

$$\|\tilde{\boldsymbol{\beta}}_\lambda - \tilde{\boldsymbol{\beta}}\|^2 \leq \lambda^2 \frac{81h^4}{\epsilon^2 d^5}\|\tilde{\boldsymbol{\beta}}\|^2.$$

Then by 3.4,

$$\|\tilde{f}_\lambda - \tilde{f}\|^2 \leq \frac{257d}{135}\|\tilde{\boldsymbol{b}}_\lambda - \tilde{\boldsymbol{b}}\|^2 = O(\lambda^2 m^4).$$

If we retain $m = O(n^{1/9})$, we may choose $\lambda = O(n^{-10/18})$, so that $\|\tilde{f}_\lambda - \tilde{f}\|^2 = O(n^{-2/3})$, and further $\|\tilde{f}_\lambda - f_0\|^2 = O(n^{-2/3})$.

Chapter 4

Multi-Modality Test with Measurement Error

4.1 Motivation

Chapter 2 introduces a hypothesis test against unimodality given data $X_1, \ldots, X_n$. Chapter 3 offers a solution to derive density estimation with shape constraints, given contaminated data $Y_1, \ldots, Y_n$. These two distinct topics naturally converge, prompting the study of a novel inquiry: how do we conduct a hypothesis test for unimodality when the data has measurement error? In this Chapter, we combine the insights from the previous chapters and investigate the modality test for the deconvolution problem.

4.2 Construction of the Test Statistic

Suppose that the observed sample is not from f, but rather $Y_i = X_i + Z_i$, where $X_1, \ldots, X_n$ is a random sample from f, and $Z_1, \ldots, Z_n$ is a random sample from a known error distribution ϕ, and the X_i and Z_i form an independent set. The estimate $\hat{f}$ given only the Y_i is called the deconvolution density estimator. Chapter 3 proposes a penalized spline deconvolution density estimator and shows that if the errors are uniform, a cube-root convergence rate for $\hat{f}$ can be attained.

Let g be the density of Y, then $g(y) = f * \phi(y) = \int f(z)\phi(y-z)dz$. Suppose the error density is uniform on $(-h, h)$. We stick to the spline basis $\delta_1, \ldots, \delta_m$ used in Chapter 3 for the estimation of f, and define $\gamma_i(x) = \delta_i * \phi$ as the spline basis for the estimation of g. Utilizing a similar approach as described in Section 3.4, we can get the estimates $\tilde{g}(y) = \sum_{i=1}^{m} \tilde{b}_i \gamma_i(y)$. The estimate for f can be obtained by $\tilde{f}(x) = \sum_{i=1}^{m} \tilde{b}_i \delta_i(x)$.

To test the hypothesis

$$H_0 : f \text{ has one mode} \quad \text{vs.} \quad H_A : f \text{ has more than one mode}$$

given measurement error data, we extend the decision rule outlined in Section 2.3. If $\hat{f}_2$ has one mode, the null hypothesis is automatically accepted. If $\hat{f}_2$ displays two modes, we proceed generate $i = 1, \ldots, B$ bootstrap samples $\boldsymbol{X}_i$ from $\hat{f}_1$, along with error samples $\boldsymbol{Z}_i$ from ϕ. The test statistics T_i is then computed using $\boldsymbol{Y}_i = \boldsymbol{X}_i + \boldsymbol{Z}_i$.

An example is shown in Figure 4.1, with a histogram of observed $Y_1, \ldots, Y_n$ shown on the left along, and the unobserved $X_1, \ldots, X_n$ shown on the right. The unimodal and bimodal fits are shown, as well as the true densities. The true density is a mixture of two normal distributions $0.6N(0, 1) + 0.4N(4, 1)$, which is clearly bimodal. However, with uniform error $\sim U(-2, 2)$, the contaminated density turns unimodal. Our method leads to a p-value of 0.027, which means we correctly reject the null hypothesis.

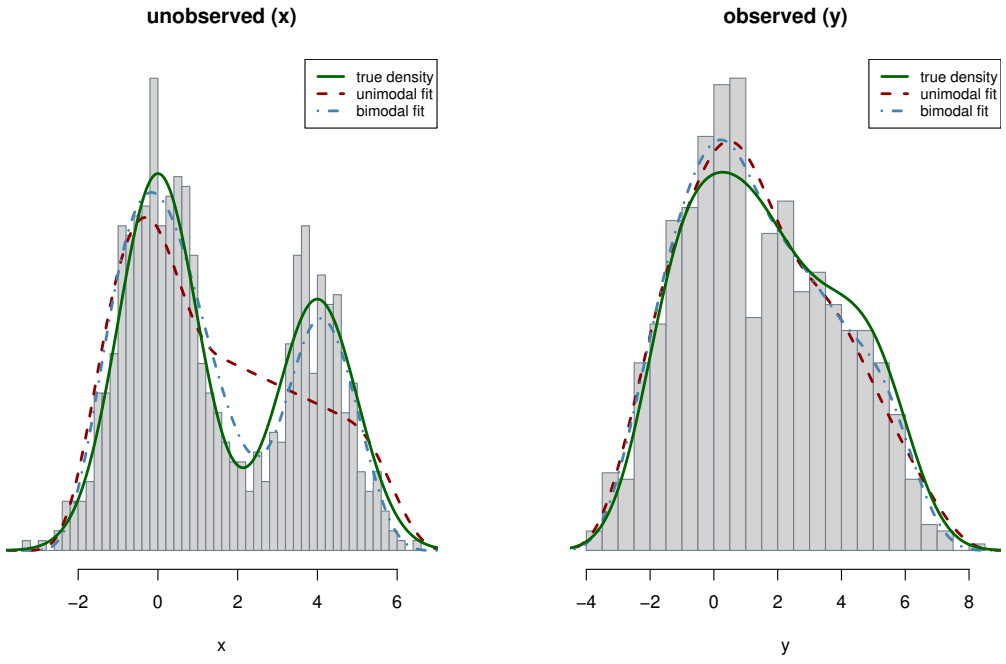

Figure 4.1: Example of simulated data with $n = 800$ and estimated unimodal and bimodal densities.

4.3 Simulations

4.3.1 Simulation 1

We first conducted a simulation study to assess the performance of our method with varying measurement error size h. To illustrate, we considered a simple example distribution: $0.5N(0,1)+0.5N(4,1)$, and introduced different measurement errors. Initially, we generated 1000 uncontaminated samples, each with a sample size of $n = 1000$. The contaminated samples were obtained by adding random errors uniformly distributed on $(-h, h)$, where h took values of $1, 1.5, 2, 2.5, 3$. Figure 4.2 displays the contaminated densities for different h values alongside the underlying true density. It is evident that as h increases, the contaminated density tends towards unimodality. Particularly, when h exceeds 2, it becomes challenging to discern whether the true underlying density is unimodal or not.

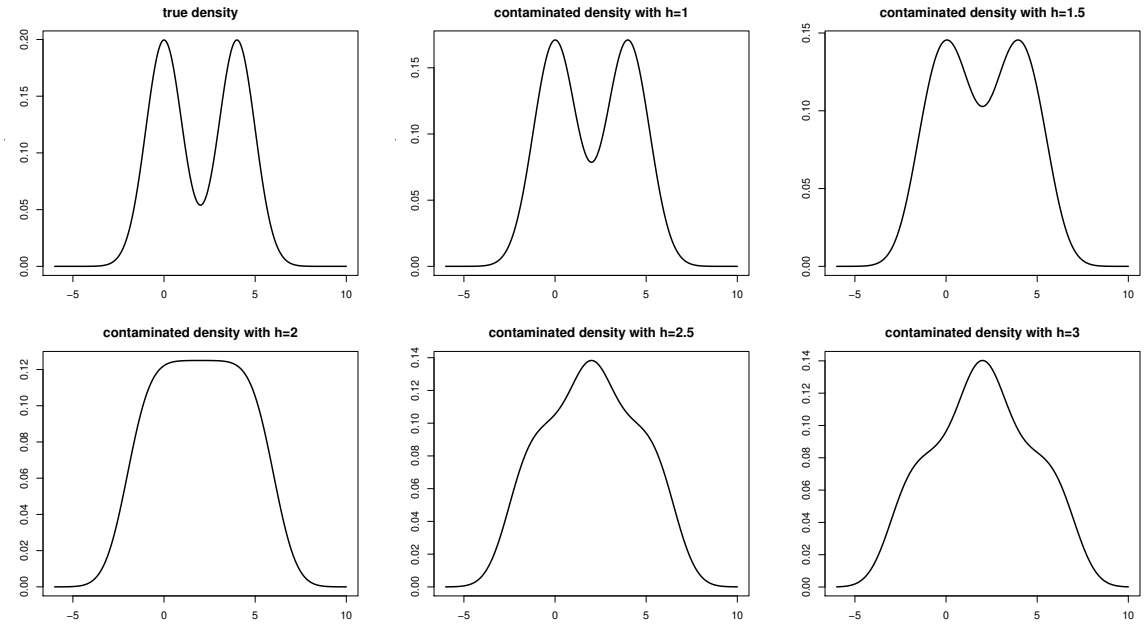

Figure 4.2: The contaminated densities with different measurement error size h, along with the underlying true density.

Table 4.1 presents the proportion of rejections (power) for each h value. The results demonstrate that our method maintains strong power even when the measurement error h is relatively large, leading to contaminated densities that are strictly unimodal.

h	1	1.5	2	2.5	3
power	1	0.997	0.959	0.923	0.918

Table 4.1: The rejection proportion of tests based on 1000 samples from $0.5N(0,1) + 0.5N(4,1)$ with different measurement size h.

4.3.2 Simulation 2

Following similar idea in Section 2.5.2, we run simulation studies according to the mode distance. Uncontaminated data are sampled from $0.6N(0,1) + 0.4N(d,1)$. Contaminated data were simulated with a uniform measurement error with size h, below are the results.

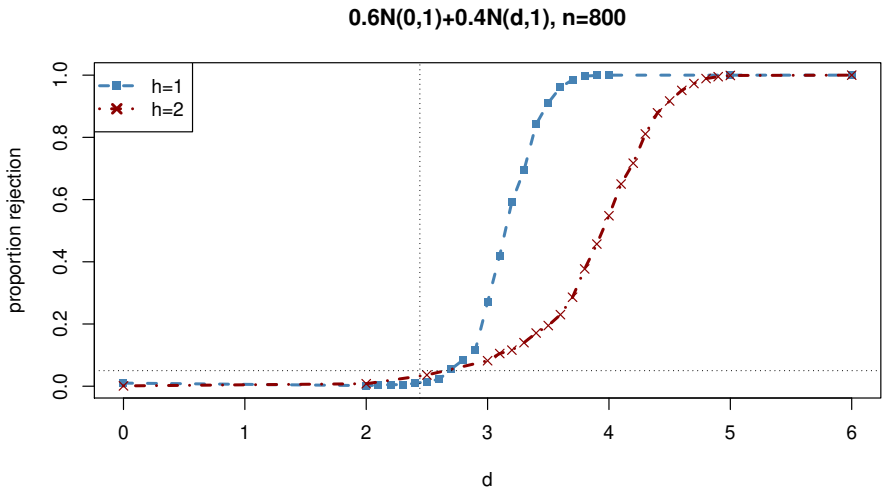

Figure 4.3: The proportion of rejection for samples from 0.6N(0,1)+0.4N(d,1), with sample size $n = 800$, when the measurement error size $h = 1$ and $h = 2$

Notice the cut-off value for d (the vertical dashed line in Figure 4.3) is around 2.4. This implies that the mixture distribution remains unimodal for d values less than 2.4 and becomes bimodal

when d exceeds 2.4. When d is larger than 2.4, as d increases, the power of the test also increases. Additionally, for larger measurement error sizes, the increase in power is slower, which is expected because larger measurement errors make it more challenging to detect bimodality. It's worth noting that due to the "auto-acceptance" rule in our method, the rejection proportion is much lower than 0.05 when the distribution is unimodal.

Chapter 5

Conclusion and Future Directions

The spline-based method stands as one of the most common nonparametric approaches. In terms of density estimation problems, one of the key strengths of spline-based density estimation methods is the ability to easily employ shape constraints. The work in this dissertation explores three applications of the least squares spline density estimator. Represented in matrix form, the least squares spline density estimator can be obtained by solving a quadratic optimization problem. To ensure the smoothness of the estimate, a penalty on the consecutive differences in the second derivative is employed.

In Chapter 2, we present a hypothesis test against the unimodality of density functions. Our approach involves obtaining penalized unimodal and bimodal density estimations. The test statistic is the difference in the least-squares criterion, between these fits. The distribution of the test statistics under the null hypothesis is estimated through simulated data sets from the unimodal fit. The validity of our method is supported by theoretical analyses, and simulation studies are conducted to evaluate its performance across various scenarios. Additionally, a real-world application about neuro-transmission data from guinea pig brains is presented.

The deconvolution density estimation problem is studied in Chapter 3. The penalized splines deconvolution estimator is introduced. Begininng with piecewise constant basis functions with uniform error, we derive the convergence rate and illustrate the difficulty of the deconvolution problem. Building upon the results gained from piecewise constant splines, we achieve a cube-root convergence rate for piecewise quadratic splines. Moreover, we derive large sample theories for the penalized spline estimator and the constrained spline estimator. Through simulation studies, we demonstrate the competitive performance of our estimators compared to kernel estimators across diverse scenarios. Note that our work primarily focuses on the uniform error distribution. We would like to generalize our method to accommodate other error distributions by starting with the mixtures of two uniform distributions. Given that any distribution can be approximated by a

mixture of uniform distributions, our method holds potential for extension to various known error distributions. Another possible future exploration involves studying additional shape constraints beyond the demonstrated unimodal and bimodal constraints in this dissertation.

The work in Chapter 4 is a combination of the findings from Chapters 2 and 3. Specifically, given a random sample with measurement error, the interest is in whether the underlying density is unimodal or multimodal. Under the assumption that the error density is uniform, we introduce a test, with the test statistic similar to that in Chapter 2, which is the difference in the least-squares criterion between the unimodal and bimodal fits. A deconvolution sampling approach is proposed to estimate the null distribution of the test statistic. Simulations are conducted to show the performance of proposed test under different conditions. It would be of interest to study the large number theory of this deconvolution multimodality test. Moreover, drawing inspiration from the future direction of the deconvolution spline estimator, exploring the application of deconvolution multimodality test for various error densities presents an interesting direction for further research.

Bibliography

[Birgé, 1987] Birgé, L. (1987). On the risk of histograms for estimating decreasing densities. *The Annals of Statistics*, pages 1013–1022.

[Carroll and Hall, 1988] Carroll, R. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186.

[Celisse, 2014] Celisse, A. (2014). Optimal cross-validation in density estimation with the $l2$ loss. *The Annals of Statistics*, 42(5):1879–1910.

[Chen and Meyer, 2023] Chen, X. and Meyer, M. C. (2023). Penalized unimodal spline density estimation with applications to m-estimation. *Journal of Statistical Planning and Inference*, 224:84–97.

[Cheng and Hall, 1998] Cheng, M.-Y. and Hall, P. (1998). Calibrating the excess mass and dip tests of modality. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):579–589.

[Cordy and Thomas, 1997] Cordy, C. B. and Thomas, D. R. (1997). Deconvolution of a distribution function. *Journal of the American Statistical Association*, 92(440):1459–1465.

[De Boor, 1972] De Boor, C. (1972). On calculating with b-splines. *Journal of Approximation theory*, 6(1):50–62.

[de Montricher et al., 1975] de Montricher, G. F., Tapia, R. A., and Thompson, J. R. (1975). Nonparametric maximum likelihood estimation of probability densities by penalty function methods. *The Annals of Statistics*, pages 1329–1348.

[Delaigle and Gijbels, 2004] Delaigle, A. and Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics*, 56(1):19–47.

[Delaigle and Hall, 2014] Delaigle, A. and Hall, P. (2014). Parametrically assisted nonparametric estimation of a density in the deconvolution problem. *Journal of the American Statistical Association*, 109(506):717–729.

[Fan, 1991] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19(3):1257–1272.

[Freedman and Diaconis, 1981] Freedman, D. and Diaconis, P. (1981). On the histogram as a density estimator: L 2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4):453–476.

[Goodd and Gaskins, 1971] Goodd, I. and Gaskins, R. A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika*, 58(2):255–277.

[Groeneboom et al., 2001] Groeneboom, P., Jongbloed, G., and Wellner, J. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Annals of Statistics*, 29(6):1653–1698.

[Haldane, 1951] Haldane, J. (1951). Simple tests for bimodality and bitangentiality. *Annals of eugenics*, 16(1):359–364.

[Hall and Qiu, 2005] Hall, P. and Qiu, P. (2005). Discrete-transform approach to deconvolution problems. *Biometrika*, 92(1):135–148.

[Hall and York, 2001] Hall, P. and York, M. (2001). On the calibration of silverman's test for multimodality. *Statistica Sinica*, 11:515–536.

[Hartigan and Hartigan, 1985] Hartigan, J. A. and Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, 13(1):70–84.

[Hartigan and Mohanty, 1992] Hartigan, J. A. and Mohanty, S. (1992). The runt test for multimodality. *Journal of Classification*, 9:63–70.

[Hazelton and Turlach, 2010] Hazelton, M. L. and Turlach, B. A. (2010). Semiparametric density deconvolution. *Scandinavian Journal of Statistics*, 37(1):91–108.

[Higham and Cheng, 1998] Higham, N. J. and Cheng, S. H. (1998). Modifying the inertia of matrices arising in optimization. *Linear Algebra and its Applications*, 275:261–279.

[Koo and Park, 1996] Koo, J.-Y. and Park, B. U. (1996). B-spline deconvolution based on the em algorithm. *Journal of statistical Computation and Simulation*, 54(4):275–288.

[Larkin, 1979] Larkin, R. P. (1979). An algorithm for assessing bimodality vs. unimodality in a univariate distribution. *Behavior Research Methods & Instrumentation*, 11(4):467–468.

[Leonard, 1973] Leonard, T. (1973). A bayesian method for histograms. *Biometrika*, 60(2):297–308.

[Lyche and Schumaker, 1973] Lyche, T. and Schumaker, L. L. (1973). Computation of smoothing and interpolating natural splines via local bases. *SIAM Journal on Numerical Analysis*, 10(6):1027–1038.

[Mammen et al., 1992] Mammen, E., Marron, J. S., and Fisher, N. I. (1992). Some asymptotics for multimodality tests based on kernel density estimates. *Probability Theory and Related Fields*, 91(1):115–132.

[Mendelsohn and Rice, 1982] Mendelsohn, J. and Rice, J. (1982). Deconvolution of microfluorometric histograms with b splines. *Journal of the American Statistical Association*, 77(380):748–753.

[Minnotte, 1997] Minnotte, M. C. (1997). Nonparametric testing of the existence of modes. *The Annals of Statistics*, pages 1646–1660.

[Müller and Sawitzki, 1991] Müller, D. W. and Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *Journal of the American Statistical Association*, 86(415):738–746.

[Parzen, 1962] Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076.

[Pensky, 2002] Pensky, M. (2002). Density deconvolution based on wavelets with bounded supports. *Statistics & probability letters*, 56(3):261–269.

[Rozál and Hartigan, 1994] Rozál, G. P. M. and Hartigan, J. (1994). The map test for multimodality. *Journal of Classification*, 11:5–36.

[Rudemo, 1982] Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, pages 65–78.

[Silvapulle and Sen, 2011] Silvapulle, M. J. and Sen, P. K. (2011). *Constrained statistical inference: Order, inequality, and shape constraints*. John Wiley & Sons.

[Silverman, 1978] Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics*, pages 177–184.

[Silverman, 1981] Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society, Series B*, 43(1):97–99.

[Stefanski and Carroll, 1990] Stefanski, L. and Carroll, R. (1990). Deconvolving kernel density estimators. *Statistics*, 21(2):169–184.

[Walter, 1999] Walter, G. G. (1999). Density estimation in the presence of noise. *Statistics & probability letters*, 41(3):237–246.

[Wegman and Wright, 1983] Wegman, E. J. and Wright, I. W. (1983). Splines in statistics. *Journal of the American Statistical Association*, 78(382):351–365.