Nonparametric estimation of the ability density in the Mixed-Effect Rasch Model

Johanna Kappus, Friedrich Liese* and Alexander Meister

Institut für Mathematik,
Universität Rostock,
18051 Rostock, Germany.
e-mail: johanna_kappus@t-online.de*;* alexander.meister@uni-rostock.de

Abstract: The Rasch model is widely used in the field of psychometrics when n persons under test answer m questions and the score, which describes the correctness of the answers, is given by a binary $n \times m$-matrix. We consider the Mixed-Effect Rasch Model, in which the persons are chosen randomly from a huge population. The goal is to estimate the ability density of this population under nonparametric constraints, which turns out to be a statistical linear inverse problem with an unknown but estimable operator. Based on our previous result on asymptotic equivalence to a two-layer Gaussian model, we construct an estimation procedure and study its asymptotic optimality properties as n tends to infinity, as does m, but moderately with respect to n. Moreover numerical simulations are provided.

1. Introduction

We consider the famous Rasch model which is used to analyse psychometric surveys when n individuals under test answer m questions. The score is given by a realization of a random binary $n \times m$-matrix X. Its (j, k)th entry indicates whether the answer of the jth person to the kth question is correct. In this case we have $X_{j,k} = 1$ and we put $X_{j,k} = 0$ otherwise. In the standard Rasch model, all entries $X_{j,k}$ are assumed to be independent binary random variables which satisfy

$$P(X_{j,k} = 1) = \frac{\exp\{\beta_j - \theta_k\}}{1 + \exp\{\beta_j - \theta_k\}}, \qquad k = 1, \ldots, m; j = 1, \ldots, n,$$

where the parameter θ_k describes the difficulty of the kth question and the parameter β_j represents the ability of the jth person. For fundamental literature on this model we refer to the books of [29]. The Rasch model has also been used in the field of econometrics (e.g. [5], [6], [25], [18]).

---

*Friedrich Liese, who had been engaged in this project, passed away on 28 April 2018.

In this paper we focus on the Mixed-Effect Rasch Model (MRM), in which the individuals under test are assumed to be randomly chosen from a huge population so that the β_j, $j = 1, \ldots, n$, are interpreted as i.i.d. random variables with some distribution F. Thus the underlying statistical experiment $\mathcal{X}_{n,m}$ equals

$$\mathcal{X}_{n,m} := \left(\{0,1\}^{n \times m}, \mathcal{P}(\{0,1\}^{n \times m}), (P_{\theta,F})_{(\theta,F) \in \Theta \times \mathcal{F}}\right),$$

where the probability measure $P_{\theta,F}$ satisfies

$$P_{\theta,F}(\{\omega\}) = \prod_{j=1}^{n} \int \left\{\prod_{k=1}^{m} \frac{\exp\{\omega_{j,k}(\beta - \theta_k)\}}{1 + \exp\{\beta - \theta_k\}}\right\} dF(\beta),$$

for all $\omega \in \{0,1\}^{n \times m}$. The specification of the parameter sets Θ and $\mathcal{F}$ is defered to subsection 2.2.

The MRM is studied e.g. in the books of [11] and [37] as well as [24], [31], [33] and [34]. So far literature on Rasch models has mainly focused on the estimation of the difficulty parameters, consistency and asymptotic normality (mostly for bounded m). Therein maximum likelihood (ML), quasi-ML and conditional ML methods are the most popular procedures, see e.g. [10], [14], [13], [7], [27], [28], [32]. By the conditional ML approach, the difficulty parameters can be estimated without estimating the ability distribution before. Also see [36] for a review on the related literature. [1] consider saddlepoint approximation of the ability parameters. [8] construct confidence intervals for the ability parameters. [33] and [34] study the covariance structure and asymptotic distribution of quasi-ML estimators in the MRM. [24] focus on semiparametric estimation in the MRM where estimation of the ability distribution via ML-related methods (mainly parametrically specified where the number of parameters is allowed to increase in some of the results). [12] considers consistent estimation in the MRM for classes of discrete ability distributions with a finite number of support points. [15] studies identification and estimation of parameters under unconstrained ability distribution; [38] consider log-linear models for the ability distributions; [39] use skew-normal distributions. Still, to our best knowledge, fully nonparametric approaches to the estimation of the ability distribution along with rigorous investigation of the theoretical properties (minimax convergence rates, adaptivity etc.) are apparently missing. Considering large international surveys, estimation of the ability distribution seems very important in order to compare the ability distribution of the students from two different countries, for example.

Therefore the current work focuses on the estimation of the ability density in the MRM. We provide a fully nonparametric procedure for the estimation of this density while the difficulty parameters are unknown. Moreover this nonparametric estimator can be applied to treat parametric submodels and to construct parametric goodness-of-fit tests as well where the investigation of the asymptotic distribution of our estimator remains a challenging task. Our method significantly differs from the existing approaches as we do not use ML-based techniques but, in fact, we exploit asymptotic equivalence of the MRM to a two-layer Gaussian experiment (more precisely, a Gaussian observation and – conditionally on that – another Gaussian observation) in Le Cam's sense, which

has recently been proved in [22]. In that paper the authors use some components of the observation in order to construct an asymptotic confidence region for the difficulty parameters and leave the problem of nonparametric estimation of the ability density open for future research. Thus, in the current work, the same authors tackle exactly this problem. Moreover another asymptotically equivalent Gaussian model with two independent (multivariate) observations will be derived in the current work.

When working with the asymptotically equivalent version of the MRM, estimating the ability density represents a statistical linear inverse problem with Gaussian noise and an unknown but empirically accessible linear operator. Such models occur quite frequently in nonparametric statistics, see e.g. [9], [4], [19] and [35]. Other examples in this class of models concern deconvolution/ image deblurring from an unknown point-spread function (e.g. [17]) or functional linear regression (e.g. [16], [26]). Each of these inverse problems has its own major challenges – also since the procedure and the theoretical results heavily depend on the operator and its properties; and since the type of the empirical information on the operator is various. This holds true for the current model as well.

Two- and three parameter models from logistic item response theory (2-PL IRT and 3-PL IRT) are popular extensions of the Rasch model for the analysis of psychometric surveys (e.g. [2], [3]). Therein the probability that the jth individual provides the correct answer to the kth question is modelled by

$$P(X_{j,k} = 1) = b_k + (1 - b_k) \cdot \exp\{a_k(\beta_j - \theta_k)\} / \left(1 + \exp\{a_k(\beta_j - \theta_k)\}\right),$$

with the difficulty parameters a_k, b_k, θ_k and the person parameter β_j. Note that the standard Rasch model is included for $b_k = 0$ and $a_k = 1$. There is no straight forward extension of our procedures to the estimation of the ability distribution in the corresponding mixed-effect model since the sums of the rows and the columns do apparently not form a sufficient statistic in general. Thus the basic arguments from [22] cannot be taken over. Nonparametric estimation of the ability distribution in 2-PL and 3-PL IRT remains an interesting topic for future research.

The paper is organized as follows: In section 2 we describe the asymptotically equivalent experiments, which motivate our estimation procedures and investigation. In section 3 the estimation procedures are provided and explained. In section 4 the asymptotic properties of these procedures are studied. In section 5 we compute the transforming Markov kernels for the link between the original MRM and the asymptotically equivalent models; and we provide numerical simulations. The proofs are defered to section 6.

2. Preliminaries

2.1. Notation

In the following we introduce some of the frequently used terms where the terminology is partially adopted from [22]. These terms are required to understand the asymptotically equivalent Gaussian models in Subsection 2.2. A shifted ver-

sion of the difficulty parameters is given by $\vartheta_k := \theta_k - \theta_m$ for all $k = 1, \ldots, m$. The set of all vectors $b \in \{0,1\}^m$ with exactly k components equal to 1 is denoted by $B(k,m)$. For any $\vartheta = (\vartheta_1, \ldots, \vartheta_{m-1})$ and $N = (N_1, \ldots, N_m) \in \mathbb{N}_0^m$ we define the function

$$\Psi_N(\vartheta) := \sum_{k=1}^m N_k \cdot \log \Big(\sum_{b \in B(k,m)} \exp \Big\{ - \sum_{l=1}^{m-1} \vartheta_l b_l \Big\} \Big).$$

In order to provide some insight about this function, we mention that, in [22], some statistics containing the sums of the rows and the columns in the MRM have been detected to be sufficient; and their asymptotic distribution has been studied. Therein the conditional expectation of one statistic given the other equals minus the grandient of Ψ_N and the corresponding conditional covariance matrix equals the Hessian matrix of Ψ_N, see [22], p. 1337. Hence this function plays a major role in the asymptotically equivalent Gaussian models, on which our estimation strategy will be based. The function τ maps from $\mathbb{R}^{m+1}$ to $\mathbb{R}^m$ and is defined by

$$\tau \,:\, x = (x_0, \ldots, x_m) \mapsto (x_1, \ldots, x_m) \cdot n / \max \Big\{ n/2, \sum_{j=0}^m x_j \Big\}.$$

Moreover $\Delta\Psi_N(\vartheta)$ stands for the corresponding Hessian matrix, and $[\cdot]$ and $\cdot_+$ denote componentwise rounding and truncation by zero, respectively. Furthermore we write $\mathsf{N}(\mu, \Sigma)$ for the normal distribution with the expectation vector μ and the covariance matrix Σ. The function which maps all elements of the domain to 1 is called $\mathbf{1}$. Finally we define

$$\alpha_{k,m,\theta} := \sum_{b \in B(k,m)} \exp \Big\{ - \sum_{l=1}^m b_l \theta_l \Big\},$$

$$\beta_{k,m,\theta}(x) := \exp\{kx\} / \prod_{l=1}^m \big(1 + \exp\{x - \theta_l\} \big),$$

$$q_k(\theta, F) := \alpha_{k,m,\theta} \int \beta_{k,m,\theta}(x) f(x) dx, \tag{2.1}$$

and the corresponding vector $q(\theta, F)$ and the matrix $Q(\theta, F)$ via

$$q(\theta, F) := \big(q_0(\theta, F), \ldots, q_m(\theta, F) \big)^T,$$

$$Q(\theta, F) := \big\{ 1_{\{j\}}(k) q_k(\theta, F) \big\}_{j,k=0,\ldots,m}.$$

These quantities also occur as the expectation vector and the covariance matrix of one statistic in the asymptotically equivalent Gaussian experiment in the next subsection.

2.2. Asymptotically equivalent models

Let us summarize the conditions on the parameters sets Θ and $\mathcal{F}$ which are assumed in [22] in order to establish asymptotic equivalence of MRM and the

two-layer Gaussian experiment. As our procedure is based on this result we also impose these constraints to deduce all theoretical properties of the current paper while, in practice, the procedure may still be useful when these assumptions are violated. The vector $\theta = (\theta_1, \ldots, \theta_m)$ which contains all difficulty parameters as its components is imposed to lie in the set

$$\Theta = \Theta_m = \left\{ \theta \in [-R, R]^m : \sum_{k=1}^{m} \theta_k = 0 \right\}, \tag{2.2}$$

for some constant $R > 0$. Therein the difficulty parameters are required to sum up to zero, which represents a common calibration in order to guarantee identifiability. The distribution class $\mathcal{F}$ contains exactly those distributions F on $\mathbb{R}$ which have a Lebesgue density f that satisfies the tightness condition

$$\overline{f}(x) \geq f(x) \geq D_0 \exp\left\{ -D_1 |x| \right\}, \qquad \forall x \in \mathbb{R}, \tag{2.3}$$

for some universal envelope function $\overline{f}$ with $\int \overline{f}(x) dx < \infty$ and some universal positive constants D_0 and D_1. Moreover impose the existence of some $\beta > 3D_1 + 29/2$ such that

$$\sup_n m_n^\beta / n < \infty, \tag{2.4}$$

for $m = m_n$; and $m \geq 3$. For simplicity, the index n of m_n will be omitted in the sequel. Condition (2.4) allows m, the number of questions, to increase slowly as n, the number of participants of the survey, tends to infinity. That constraint seems realistic in many psychometric surveys.

Under the conditions (2.3) and (2.4), [22] prove asymptotic equivalence of the MRM $\mathcal{X}_{n,m}$ and the experiment $\mathcal{Y}_{n,m}$, in which the $2m$-dimensional random vector (X, Y) is observed (notation changed with respect to [22]) when n tends to infinity. The distribution of Y equals $\mathcal{L}(Y) = \mathsf{N}(nq(\theta, F), nQ(\theta, F))$. The conditional distribution of X given Y equals $\mathsf{N}(\vartheta, \{\Delta\Psi_{[\tau(Y)]_+}(\vartheta)\}^{-1})$ on the event $\{[\tau(Y)]_{+,1} + \cdots + [\tau(Y)]_{+,m-1} > 0\}$ (put $X := 0$ otherwise by convention).

This consideration is now continued by showing asymptotic equivalence of the experiment $\mathcal{Y}_{n,m}$ and the experiment $\mathcal{Z}_{n,m}$, in which one observes (X^*, Y) where X^* and Y are independent; Y is as in the experiment $\mathcal{Y}_{n,m}$ while $\mathcal{L}(X^*) = \mathsf{N}(\vartheta, \{\Delta\Psi_{nq(\theta,F)}(\vartheta)\}^{-1})$. For this purpose, no transformation of the data by a Markov kernel is required. Note that, in the above Hessian matrix, the partial derivatives are only to be taken with respect to the ϑ_j, not with respect to the θ in the index.

Theorem 2.1. *Under the conditions (2.2)–(2.4) and $m \geq 3$, the experiments $\mathcal{Y}_{n,m}$ and $\mathcal{Z}_{n,m}$ are asymptotically equivalent.*

Experiment $\mathcal{Z}_{n,m}$ is advantageous compared to experiment $\mathcal{Y}_{n,m}$ in the regard that the statistics X^* and Y are independent. On the other hand, the covariance matrix of X^* is unknown to the statistician in the experiment $\mathcal{Z}_{n,m}$; while $\mathcal{Y}_{n,m}$ is preferable for the construction of confidence ellipsoids for the difficulty parameters as in section 6 in [22] since the conditional covariance matrix of X

given Y does not depend on F. In the current work, it makes no difference if one considers model $\mathcal{Y}_{n,m}$ or $\mathcal{Z}_{n,m}$ for the oracle setting (see subsection 4.2); whereas model $\mathcal{Z}_{n,m}$ is used for the estimation of the ability density under unknown difficulty parameter in the subsection 4.3.

For basic literature on Le Cam theory we refer to the books of [20], [21] and [23]. The exact Markov kernels which have been used in [22] to transform the MRM into the experiment $\mathcal{Y}_{n,m}$ are only needed in the simulation section 5 as the original data have to be transformed into data from $\mathcal{Y}_{n,m}$ and $\mathcal{Z}_{n,m}$, respectively, in order to apply our estimation procedure. Therefore those kernels will be presented in that subsection 5.1.

3. Methodology

3.1. Oracle setting

In this subsection we consider estimation of the ability density f under the oracle constraint that the difficulty parameters $\theta = (\theta_1, \ldots, \theta_m)$ are fully known. Writing f for the true density of the ability distribution F, we define

$$h(x) := f(x)/w(x),$$

for some fixed (known) Lebesgue probability density w which has no zeros. We assume that

$$\int f^2(x)/w(x)dx < \infty, \tag{3.1}$$

so that $h \in L_2(\mathbb{R}, w)$, i.e. h lies in the separable Hilbert space which contains all measurable functions g from $\mathbb{R}$ to itself with $\int |g(x)|^2 w(x)dx < \infty$ (or, to be exact, the equivalence classes of those functions which coincide almost everywhere with respect to the LB measure, i.e. the Lebesgue-Borel measure). Accordingly $\langle \cdot, \cdot \rangle_w$ and $\| \cdot \|_w$ stand for the corresponding inner product and norm, respectively. We define the function

$$\gamma_{m,\theta}(x,y) := \sum_{k=0}^{m} \alpha_{k,m,\theta}\beta_{k,m,\theta}(x)\beta_{k,m,\theta}(y)/\int \beta_{k,m,\theta}(z)w(z)dz,$$

and the operator $\Gamma_{m,\theta}$ which satisfies

$$\left[\Gamma_{m,\theta}h\right](x) := \int \gamma_{m,\theta}(x,y)h(y)w(y)dy, \qquad h \in L_2(\mathbb{R}, w),\ x \in \mathbb{R}. \tag{3.2}$$

We consider the observation $Y = (Y_0, \ldots, Y_m)$ from the experiment $\mathcal{Y}_{n,m}$ as defined in subsection 2.2; and realize that

$$E\sum_{k=0}^{m} Y_k \cdot \beta_{k,m,\theta}/\langle\beta_{k,m,\theta}, \mathbf{1}\rangle_w = n \cdot \Gamma_{m,\theta}h. \tag{3.3}$$

Thus, in order to estimate h or f, inversion of the operator $\Gamma_{m,\theta}$ is required. Therefore we study some of its properties, writing $H_{m,\theta}$ for the linear hull of the functions $\beta_{k,m,\theta}$, $k = 0, \ldots, m$.

Lemma 3.1. *(a) The dimension of the linear subspace $H_{m,\theta}$ of $L_2(\mathbb{R}, w)$ equals $m + 1$.*

(b) The mapping $\Gamma_{m,\theta}$ is a compact, self-adjoint and positive semi-definite linear operator from $L_2(\mathbb{R}, w)$ to $H_{m,\theta} \subseteq L_2(\mathbb{R}, w)$. The kernel of $\Gamma_{m,\theta}$ corresponds to the orthogonal complement of $H_{m,\theta}$ with respect to $L_2(\mathbb{R}, w)$.

By Lemma 3.1(b), the unit eigenfunctions of $\Gamma_{m,\theta}$, which are denoted by $\varphi_{j,m,\theta}$, integer $j \geq 0$, form an orthonormal basis of $L_2(\mathbb{R}, w)$. Therein we arrange that the $\varphi_{j,m,\theta}$ are sorted such that the corresponding eigenvalues $\lambda_{j,m,\theta}$ decrease in j. As $\Gamma_{m,\theta}$ is positive semi-definite all $\lambda_{j,m,\theta}$ are non-negative. Moreover Lemma 3.1 yields that the $\lambda_{j,m,\theta}$, $j = 0, \ldots, m$, are positive while $\lambda_{j,m,\theta} = 0$ for all $j > m$. This also implies that the $\varphi_{j,m,\theta}$ are located in $H_{m,\theta}$ for all $j = 0, \ldots, m$ while $\varphi_{j,m,\theta}$ lies in the orthogonal complement of $H_{m,\theta}$ for any $j > m$.

It follows from (3.3) that

$$\hat{h}_{j,m,\theta} := \sum_{k=0}^{m} Y_k \cdot \langle \beta_{k,m,\theta}, \varphi_{j,m,\theta} \rangle_w / \left\{ n \cdot \lambda_{j,m,\theta} \cdot \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w \right\}, \tag{3.4}$$

is an unbiased oracle estimator of the score $\langle h, \varphi_{j,m,\theta} \rangle_w$ for all $j = 0, \ldots, m$. According to the usual structure of orthogonal series estimators we introduce the oracle estimator of h by

$$\hat{h}^{\mathfrak{w}}_{m,\theta} := \mathbf{1} + \sum_{j=1}^{m} \mathfrak{w}_{j,m,\theta} \cdot \hat{h}_{j,m,\theta} \cdot \varphi_{j,m,\theta}, \tag{3.5}$$

for some weights $\mathfrak{w}_{j,m,\theta} \in [0,1]$, $j = 1, \ldots, m$, which remain to be selected. Therein we have exploited that $\varphi_{0,m,\theta} = \mathbf{1}$ (see Proposition 4.1(a)) and, hence, $\langle h, \varphi_{0,m,\theta} \rangle_w = 1$ since $f = h \cdot w$ is a probability density.

3.2. Estimation of h under unknown θ

While, in practice, the vector $\theta = (\theta_1, \ldots, \theta_m)$ is unknown it is empirically available by the statistic X^* from the experiment $\mathcal{Z}_{n,m}$ whereas the oracle density estimator in (3.5) is only based on the independent statistic Y. The expectation of X^* equals $\vartheta = (\vartheta_1, \ldots, \vartheta_{m-1})$. The definition of the ϑ_k along with the fact that the components of θ sum up to 0 yields that $\theta = Z_m \vartheta$ for all $\theta \in \Theta$ where

$$Z_m := \begin{pmatrix} 1 - 1/m & -1/m & \cdots & -1/m \\ -1/m & 1 - 1/m & \cdots & -1/m \\ \vdots & \vdots & \vdots & \vdots \\ -1/m & -1/m & \cdots & 1 - 1/m \\ -1/m & -1/m & \cdots & -1/m \end{pmatrix}.$$

This equation has also been used in [22] in order to construct confidence bands for θ in an intermediate experiment. That motivates us to use $\hat{\theta} := Z_m X^*$ as a plug-in estimator of θ. Inserting $\hat{\theta}$ for θ in (3.5) provides the fully accessible estimator $\hat{h}^{\mathfrak{w}}_{m,\hat{\theta}}$ of h.

3.3. Data-driven selection of the weights

In order to choose the weights $\hat{\mathfrak{w}}_{j,m,\hat{\theta}}$ we apply a keep-or-kill strategy, also known as hard thresholding approach (see e.g. [19]). Concretely we put

$$\hat{\mathfrak{w}}_{j,m,\hat{\theta}} := \begin{cases} 1\,, & \text{if } \hat{h}^2_{j,m,\hat{\theta}} \geq c_W \cdot m^4 \cdot (\log n) \cdot n^{-1}\,\lambda^{-1}_{j,m,\hat{\theta}}\,, \\ 0\,, & \text{otherwise,} \end{cases} \tag{3.6}$$

for some constant $c_W > 0$ to be selected. Writing $\hat{\mathfrak{w}} := (\hat{\mathfrak{w}}_{j,m,\hat{\theta}})_{j=1,\dots,m}$ we introduce our final nonparametric estimator of h by $\hat{h}^{\hat{\mathfrak{w}}}_{m,\hat{\theta}}$. The effect of the plug-in of the estimator $\hat{\theta}$ in Subsection 3.2 and of the weight selection in (3.6) on the estimation error is studied in Theorem 4.4.

4. Theoretical properties

In this section we study the asymptotic properties of the mean integrated squared error (MISE), which is defined by

$$\text{MISE}(\hat{h}, h) := E_h \big\| \hat{h} - h \big\|^2_w\,,$$

for the (oracle and realistic) estimators of h from section 3. For that purpose the spectrum (eigenvalues and eigenfunctions) of the operator $\Gamma_{m,\theta}$ in (3.2) is studied in the following subsection.

4.1. Spectrum of the operator $\Gamma_{m,\theta}$

The following proposition provides important properties of the eigenvalues $\lambda_{j,m,\theta}$ of $\Gamma_{m,\theta}$. The decay of the eigenvalues is crucial to establish convergence rates of the MISE.

Proposition 4.1. *(a) The eigenfunction $\varphi_{0,m,\theta}$ coincides with $\mathbf{1}$; and $\lambda_{0,m,\theta} = 1$. The eigenspace with the eigenvalue 1 is one-dimensional.*

(b) Assume that the density w is bounded and has a continuous and uniformly bounded derivative. Then, for all sequences $(J_m)_m$ and $(J^*_m)_m$ of positive integers which satisfy $\lim_{m \to \infty} J_m / \log^{1/2} m = 0$ and $\lim_{m \to \infty} J^*_m / \log^\nu m = 0$ for some $\nu \in [1/2, 1)$, it holds that

$$\lim_{m \to \infty}\ \sup_{j=0,\dots,J_m}\ \sup_{\theta \in \Theta}\ \big| \lambda_{j,m,\theta} - 1 \big|\ =\ 0\,,$$

$$\liminf_{m \to \infty}\ \inf_{j=0,\dots,J^*_m}\ \inf_{\theta \in \Theta}\ \exp\big\{ c \log^\mu m \big\} \cdot \lambda_{j,m,\theta}\ \geq\ 1\,,$$

for all constants $c > 0$ and $\mu > 2\nu - 1$.

(c) The sequence $(\Gamma_{m,\theta_{(m)}})_m$ with $\theta_{(m)} \in \Theta_m$ has no limit point with respect to the induced operator norm of $L_2(\mathbb{R}, w)$, regardless of the specific sequence $(\theta_{(m)})_m$.

Proposition 4.1(c) shows that no (strong) limit operator of the $\Gamma_{m,\theta}$ exists for large m, which complicates the analysis of the statistical inverse problems compared to the usual setting. Proposition 4.1(b) allows to control the asymptotic behaviour of the first $\lfloor \log^\mu m \rfloor$ eigenvalues for $\mu < 1/2$ and, at least in terms of a lower bound, for $\mu < 1$ as well. Furthermore, an upper bound on the supremum norm of the unit eigenfunctions $\varphi_{j,m,\theta}$ of $\Gamma_{m,\theta}$ will be established in the following lemma under the condition

$$w(x) \geq 2D_0 \exp\{-D_1|x|\}, \qquad \forall x \in \mathbb{R}, \tag{4.1}$$

with the constants D_0 and D_1 from (2.3). We provide

Lemma 4.1. *Under (4.1) there exists some constant $D^* \in (0,\infty)$ which only depends on R, D_0 and D_1 such that*

$$\|\varphi_{j,m,\theta}\|_\infty \leq D^* \, m^{1/2+D_1/2} \cdot \lambda_{j,m,\theta}^{-1/2}.$$

for all $j = 1, \ldots, m$.

Finally, the fact that θ is not exactly known but must be estimated by $\hat{\theta}$, see subsection 3.2, motivates us to study the proximity between the corresponding eigenvalues $\lambda_{j,m,\theta}$ for different values of θ in the following lemma.

Lemma 4.2. *The function $\theta \mapsto \lambda_{j,m,\theta}$ is continuously differentiable and, for any $l = 1, \ldots, m$, it holds that*

$$\left| \frac{\partial}{\partial \theta_l} \lambda_{j,m,\theta} \right| \leq 2\,\lambda_{j,m,\theta}.$$

4.2. Asymptotic optimality in the oracle setting

Let us consider the asymptotic quality of the estimator (3.5) in the oracle setting of subsection 3.1. In the following proposition an upper bound on its MISE is provided.

Proposition 4.2. *It holds that*

$$MISE(\hat{h}_{m,\theta}^{\mathfrak{w}}, h) \leq \frac{1}{n} \cdot \|h\|_\infty \sum_{j=1}^m \mathfrak{w}_{j,m,\theta}^2 / \lambda_{j,m,\theta} + \sum_{j=1}^m |\mathfrak{w}_{j,m,\theta} - 1|^2 \langle \varphi_{j,m,\theta}, h \rangle_w^2$$
$$+ \sum_{j>m} \langle \varphi_{j,m,\theta}, h \rangle_w^2.$$

Now impose that $h = f/w$ lies in the function class $\mathcal{H}$ where $\mathcal{H} = \mathcal{H}(m, \theta, D_0, D_1, D_2, \mathfrak{h}_m)$ collects all functions $h \in L_2(\mathbb{R}, w)$ such that hw is a Lebesgue probability density; we have

$$h(x)w(x) \geq D_0 \exp(-D_1|x|), \text{ for LB-almost all } x \in \mathbb{R},$$

see (2.3); and

$$\|h\|_\infty \leq D_2,$$

(put $\overline{f} := D_2 w$ to satisfy (2.3)); and we impose that the scores of h fulfill

$$\left| \langle h, \varphi_{j,m,\theta} \rangle_w \right| \leq \mathfrak{h}_{m,j} ,$$

for all integer $j = 0, \ldots, m$; and

$$\sum_{j > m} \langle h, \varphi_{j,m,\theta} \rangle_w^2 \leq \mathfrak{h}_{m,m+1}^2 ,$$

for some fixed positive-valued vector $\mathfrak{h}_m := (\mathfrak{h}_{m,j})_{j=0,\ldots,m+1}$. The latter condition replaces the usual Sobolev condition; where Sobolev-like polynomial decay of the $\mathfrak{h}_{m,j}$ but also more restrictive exponential tails are included by appropriate selection of $\mathfrak{h}_m$.

Under oracle choice of the weights $\mathfrak{w}_{j,m,\theta}$ (known $\mathfrak{h}_m$ and θ) the following asymptotic upper bound on the MISE can be established uniformly on $h \in \mathcal{H}$.

Theorem 4.1. *Select* $\mathfrak{w}_{j,m,\theta} := n \cdot \lambda_{j,m,\theta} \cdot \mathfrak{h}_{m,j}^2 / \big(n \cdot \lambda_{j,m,\theta} \cdot \mathfrak{h}_{m,j}^2 + 1 \big)$, $j = 1, \ldots, m$. *Then,*

$$\sup_{h \in \mathcal{H}} MISE(\hat{h}_{m,\theta}^{\mathfrak{w}}, h) \leq \mathfrak{h}_{m,m+1}^2 + \max\{1, D_2\} \cdot \sum_{j=1}^m \frac{\mathfrak{h}_{m,j}^2}{1 + n \cdot \lambda_{j,m,\theta} \cdot \mathfrak{h}_{m,j}^2} .$$

The proof follows from Proposition 4.2 by straight forward calculation. Under additional constraints on $\mathfrak{h}_m$ we will show asymptotic minimax optimality of our estimator (3.5) up to a constant factor.

Theorem 4.2. *Grant the inequality (4.1). Assume that* $\mathfrak{h}_{m,0} \geq 1$, $\mathfrak{h}_{m,m+1} = 0$ *and*

$$\max_{k=1,\ldots,m} \mathfrak{h}_{m,k} / \sqrt{\lambda_{k,m,\theta} + n \cdot \lambda_{k,m,\theta}^2 \cdot \mathfrak{h}_{m,k}^2} \leq \mathfrak{c}^* \cdot m^{-(3+D_1)/2} ,$$

for all integer $m \geq 1$ *and some constant* $\mathfrak{c}^*$. *Moreover impose that* $D_2 \geq 3/2$. *Let* $(\hat{h}_n)_n$ *be an arbitrary estimator sequence, where* $\hat{h}_n$ *is based on the observation from the experiment* $\mathcal{Y}_{n,m}$. *Then there exists some* $z > 0$ *such that*

$$\limsup_{n \to \infty} \sup_{h \in \mathcal{H}} P_h \left[\|\hat{h}_n - h\|_w^2 \geq z \cdot \sum_{j=1}^m \frac{\mathfrak{h}_{m,j}^2}{1 + n \cdot \lambda_{j,m,\theta} \cdot \mathfrak{h}_{m,j}^2} \right] > 0 .$$

Whenever the second term of the upper bound in Theorem 4.1 dominates the first one Theorem 4.1 and 4.2 yield "blind" rate minimaxity as neither $\mathfrak{h}_m$ nor the decay of the eigenvalues $\lambda_{j,m,\theta}$ is specified explicitly. Indeed the condition $m \to \infty$ (as $n \to \infty$) is required in order to identify h under nonparametric constraints, as will be shown in the following theorem.

Theorem 4.3. *Grant (4.1); and fix some constant* m *and* θ. *Then, for any density* $f = h \cdot w$ *such that* $h \in L_2(\mathbb{R}, w)$ *is bounded away from zero, there exists some* $\tilde{h} \neq h$ *such that the parameters* f *and* $\tilde{f} = \tilde{h} \cdot w$ *lead to the same data distribution (under known* θ*), both in the experiment* $\mathcal{Y}_{n,m}$ *and in* $\mathcal{Z}_{n,m}$ *for all integer* $n \geq m$.

4.3. Asymptotic properties of the fully data-driven estimator

Now we study the asymptotic quality of the estimator $\hat{h}^{\hat{\mathfrak{w}}}_{m,\hat{\theta}}$ from subsection 3.3 in the experiment $\mathcal{Z}_{n,m}$, which has been introduced in subsection 2.2 and shown to be asymptotically equivalent to model $\mathcal{Y}_{n,m}$ in Theorem 2.1 and, thus, to the original MRM $\mathcal{X}_{n,m}$ as well.

The smoothness properties contained in the definition of the function class $\mathcal{H}$ are robustified with respect to θ; concretely define the new function set $\mathcal{H}^*$ as the intersection of the classes $\mathcal{H}(m, \theta', D_0, D_1, D_2, \mathfrak{h}_m)$ for $\theta' \in \Theta$ with $|\theta' - \theta| \leq c_\Theta$ for some universal constant $c_\Theta > 0$. Then the following result can be established.

Theorem 4.4. *Consider the experiment $\mathcal{Z}_{n,m}$; and grant the assumptions of Theorem 2.1; and that $m \to \infty$. Write*

$$\mathcal{M}_{n,m} := \mathfrak{h}^2_{m,m+1} + \sum_{j=1}^m \mathfrak{h}^2_{m,j} / \left(1 + n \cdot \lambda_{j,m,\theta} \cdot \mathfrak{h}^2_{m,j}\right),$$

and fix some arbitrary $c_W > 0$. Then the estimator $\hat{h}^{\hat{\mathfrak{w}}}_{m,\hat{\theta}}$ with the weight selection (3.6) satisfies

$$\lim_{d \to \infty} \limsup_{n \to \infty} \sup_{\theta \in \Theta} \sup_{h \in \mathcal{H}^*} P_{\theta,h} \left[\|\hat{h}^{\hat{\mathfrak{w}}}_{m,\hat{\theta}} - h\|^2_w \geq d \cdot m^4 \cdot (\log n) \cdot \mathcal{M}_{n,m}\right] = 0.$$

Theorem 4.4 shows that the convergence rates from Theorem 4.1 can be maintained up to deterioration by the factor $m^4 \log n$. The factor $\log n$ occurs frequently in adaptation by the keep-or-kill strategy (unless appropriate block-wise shrinkage is considered; however, in the current problem, where the exact decay of the eigenvalues $\lambda_{j,m,\theta}$ is unexplored, this seems very hard to handle). The factor m^4 stems from the statistical risk of the plug-in estimator $\hat{\theta}$, see [22]. If we were allowed to use the true θ this factor could be removed.

5. Implementation and simulations

In this section we address the computational aspects of our estimation procedure $\hat{h}^{\hat{\mathfrak{w}}}_{m,\hat{\theta}}$ and investigate its finite-sample performance by numerical simulations.

5.1. Computation of the transforming Markov kernels

The estimator $\hat{h}^{\hat{\mathfrak{w}}}_{m,\hat{\theta}}$ is designed to treat data from the experiment $\mathcal{Y}_{n,m}$ or $\mathcal{Z}_{n,m}$. Therefore, before applying this estimator, the original data from the MRM experiment $\mathcal{X}_{n,m}$ have to be transformed by the Markov kernels provided in [22]. These transformations are listed in the following. While in [22] specific contamination by Gaussian noise is required between step 1 and 2 as well as between step 2 and 3 in order to prove asymptotic equivalence, these blurring steps are omitted here.

1. In the MRM experiment $\mathcal{X}_{n,m}$ one observes a binary $n \times m$-matrix X. Compute the statistic $X_{[1]} := (T_1, \ldots, T_{m-1}, N_1, \ldots, N_m)$ where

$$S_j = \sum_{k=1}^{m} X_{j,k}, \quad N_k = \sum_{j=1}^{n} 1_{\{k\}}(S_j), \quad T_k = \sum_{j=1}^{n} X_{j,k}\,.$$

2. Find y such that $\big| -\nabla\Psi_N(y) - X_{[1]}[1:m-1]\big|$ is minimized over a certain discretized grid (approximate inversion of the function $-\nabla\Psi_N$) where $X_{[1]}[1:m-1]$ denotes the first $m-1$ components of $X_{[1]}$. Replace $X_{[1]}[1:m-1]$ by that y; and keep the latter m components as they are. The outcome is denoted by $X_{[2]}$.

3. Generate the $\mathsf{N}(n,n)$-distributed random variable V independently of $X_{[2]}$. Calculate

$$N^* := \max\{n/2, V\} \cdot X_{[2]}[m:2m-1]\,/\,n\,,$$

and $N_0^* := V - \sum_{j=1}^{m} N_j^*$ where $X_{[2]}[m:2m-1]$ denotes the latter m components of $X_{[2]}$. Replace $X_{[2]}[m:2m-1]$ by the $(m+1)$-dimensional vector (N_0^*, N^*). The outcome is the $2m$-dimensional vector $X_{[3]}$.

Finally the vector $X_{[3]}$ is viewed as (X,Y) or (X^*,Y) from the experiments $\mathcal{Y}_{n,m}$ and $\mathcal{Z}_{n,m}$, respectively. It is inserted in the estimator $\hat{h}_{m,\hat\theta}^{\hat{\mathfrak{w}}}$ from section 3. The simulation results are described in the following subsection.

5.2. Numerical simulations

We generate data from the original MRM experiment in four settings. Therein two different ability densities are considered, namely

$$f_1(x) = \frac{1}{4} \cdot (|x| + 1) \cdot \exp(-|x|)\,,$$
$$f_2(x) = \frac{1}{2}\, f_1(x + 3.5) + \frac{1}{2}\, f_1(x - 3)\,.$$

Three different values of the difficulty parameter vector θ are used: $(-1.5, 1, 1, -0.5)$, $(3, -1, -1, -1)$ and $(0.4, 0.6, -2, 0.2, 0.8)$. In each setting it holds that $n = 1,000$. Moreover we choose the weight function w as $w(x) = 0.5 \cdot f_1(0.5x)$ and $c_W = 0.001$. This small value of c_W takes into account that, in practice, smaller threshold levels are expected than those motivated by the upper bound $m^4 \log n$ in (3.6). We consider the estimated ability densities $\hat{f} := w \cdot \hat{h}_{m,\hat\theta}^{\hat{\mathfrak{w}}}$. In each setting 50 independent replicates have been computed, where five of them (chosen at random) are shown in each of the Figures 1–6, along with the true curves, which are plotted in dashed linestyle. The following table provides the means and the standard deviations of the errors $\left(\int |\hat{f} - f|^2 \right)^{1/2}$, taken over all 50 replicates. The results indicate that the estimator is able to distinguish between the unimodal and the bimodal shape of density f_1 and f_2, respectively, quite well. We realize that, for $m = 5$ as in Figure 5 and 6, the simulation results are
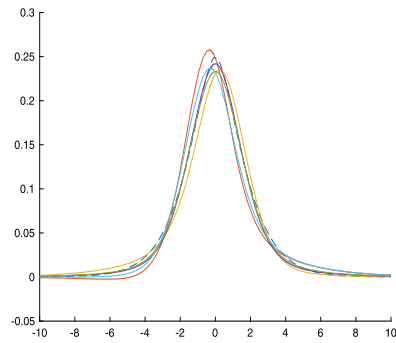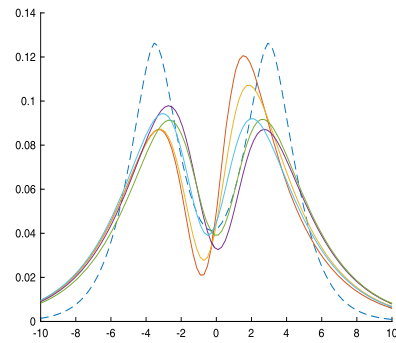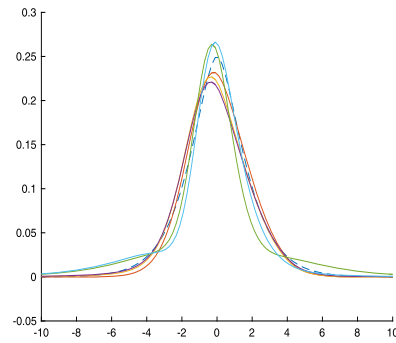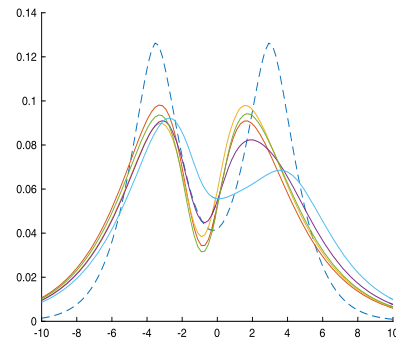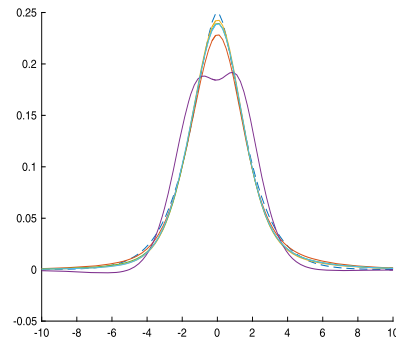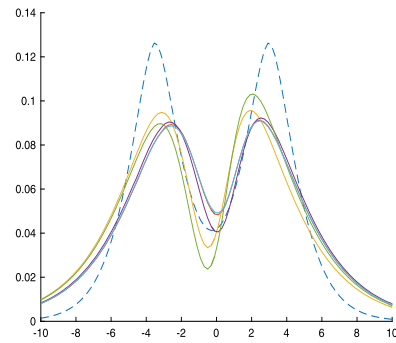
Fig 1. *density f_1, $\theta = (-1.5, 1, 1, -0.5)$.*

Fig 2. *density f_2, $\theta = (-1.5, 1, 1, -0.5)$.*

Fig 3. *density f_1, $\theta = (3, -1, -1, -1)$.*

Fig 4. *density f_2, $\theta = (3, -1, -1, -1)$.*

Fig 5. *density f_1, $\theta = (0.4, 0.6, -2, 0.2, 0.8)$.*

Fig 6. *density f_2, $\theta = (0.4, 0.6, -2, 0.2, 0.8)$.*

Figure	1	2	3	4	5	6
Mean	0.0511	0.0822	0.0593	0.0943	0.0536	0.0794
Standard deviation	0.0394	0.0145	0.0311	0.0138	0.0725	0.0139

not worse than in the case of $m = 4$ in the Figures 1–4; however the numerical effort of the computation increases significantly, in particular due to Step 2 in Subsection 5.1.

6. Proofs

Proof of Theorem 2.1: First we mention that the probability of the event $\{[\tau(Y)]_{+,1} + \cdots + [\tau(Y)]_{+,m-1} > 0\}$ tends to 1 uniformly with respect to $\theta \in \Theta$ and $F \in \mathcal{F}$ thanks to Lemma 5.2 in [22]. Therefore it suffices to prove that

$$\lim_{n \to \infty} \sup_{\theta, F} E_{\theta, F} H^2\big(\mathsf{N}(\vartheta, \{\Delta\Psi_{[\tau(Y)]_+}(\vartheta)\}^{-1}), \mathsf{N}(\vartheta, \{\Delta\Psi_{nq(\theta, F)}(\vartheta)\}^{-1})\big) = 0,$$
(6.1)

where $H^2(f, g) := \int(\sqrt{f} - \sqrt{g})^2/2$ denotes the squared Hellinger distance between two densities f and g. According to equation (A.4) in [30], the squared Hellinger distance between two multivariate normal densities with a joint expectation vector can be bounded from above as follows.

$$E_{\theta, F} H^2\big(\mathsf{N}(\vartheta, \{\Delta\Psi_{[\tau(Y)]_+}(\vartheta)\}^{-1}), \mathsf{N}(\vartheta, \{\Delta\Psi_{nq(\theta, F)}(\vartheta)\}^{-1})\big)$$
$$\leq E_{\theta, F}\big\|\{\Delta\Psi_{[\tau(Y)]_+}(\vartheta)\}^{1/2}\big(\{\Delta\Psi_{[\tau(Y)]_+}(\vartheta)\}^{-1}$$
$$\qquad\qquad - \{\Delta\Psi_{nq(\theta, F)}(\vartheta)\}^{-1}\big)\{\Delta\Psi_{[\tau(Y)]_+}(\vartheta)\}^{1/2}\big\|_F^2$$
$$\leq 4\big\|\{\Delta\Psi_{nq(\theta, F)}(\vartheta)\}^{-1/2}\big\|^2\big(E_{\theta, F}\big\|\{\Delta\Psi_{[\tau(Y)]_+}(\vartheta)\}^{1/2} - \{\Delta\Psi_{nq(\theta, F)}(\vartheta)\}^{1/2}\big\|_F^4\big)^{1/2}$$
$$+ 2\big\|\{\Delta\Psi_{nq(\theta, F)}(\vartheta)\}^{-1}\big\|^2 E_{\theta, F}\big\|\{\Delta\Psi_{[\tau(Y)]_+}(\vartheta)\}^{1/2} - \{\Delta\Psi_{nq(\theta, F)}(\vartheta)\}^{1/2}\big\|_F^4,$$
(6.2)

where, here, $\|\cdot\|_F$ and $\|\cdot\|$ denote the Frobenius norm and the induced matrix norm, respectively. For all symmetric and positive semi-definite $(m-1)\times(m-1)$-matrices A and B, it holds that

$$\|A^{1/2} - B^{1/2}\|_F^2 \leq \sqrt{m-1} \cdot \|A - B\|_F,$$

so that

$$E_{\theta, F}\big\|\{\Delta\Psi_{[\tau(Y)]_+}(\vartheta)\}^{1/2} - \{\Delta\Psi_{nq(\theta, F)}(\vartheta)\}^{1/2}\big\|_F^4$$
$$\leq (m-1)\, E_{\theta, F}\big\|\Delta\Psi_{[\tau(Y)]_+}(\vartheta) - \Delta\Psi_{nq(\theta, F)}(\vartheta)\big\|_F^2$$
$$\leq \Big(m^2(m-1)/2 + 8(m-1)(m+1)\sum_{k=0}^{m} E_{\theta, F}\big|nq_k(\theta, F) - Y_j\big|^2\Big)$$
$$\cdot \max_{k=0,\ldots,m} \Big\|\Delta_\vartheta \log\Big(\sum_{b \in B(k,m)} \exp\Big\{-\sum_{l=1}^{m-1} \vartheta_l b_l\Big\}\Big)\Big\|_F^2$$
$$\leq 4(m-1)^3 \cdot \big(m^2/2 + 8(m+1)n\big).$$

Note that $\left\|\{\Delta\Psi_{nq(\theta,F)}(\vartheta)\}^{-p}\right\|^2$ is bounded from above by the smallest eigenvalue of $\Delta\Psi_{nq(\theta,F)}(\vartheta)$, raised to the $(-2p)$th power, for $p \in \{1/2, 1\}$. The Lemmata 4.3 and 5.1 from [22] provide the lower bound

$$(1 - q_0(\theta, F) - q_m(\theta, F)) \cdot n \cdot (m-1)^{-2} \exp\{-6R\}/(1 + \exp\{2R\}),$$

on that eigenvalue. From the proof of Lemma 5.2 in [22], it follows that

$$\inf_{\theta, F} \inf_{m \geq 3} \left(1 - q_0(\theta, F) - q_m(\theta, F)\right) > 0,$$

and, hence, that the term (6.2) has the asymptotic upper bound $\mathcal{O}\big(m^8/n + m^4/n^{1/2}\big)$ with universal constant factors (only depending on R and $\overline{f}$). Finally the condition (2.4) completes the proof of (6.1). $\square$

Proof of Lemma 3.1: (a) It suffices to show that the $\beta_{k,m,\theta}$, $k = 0, \ldots, m$, are linearly independent elements of $L_2(\mathbb{R}, w)$. As the $\beta_{k,m,\theta}$ are bounded and measurable functions they are located in $L_2(\mathbb{R}, w)$. Now assume that

$$\sum_{k=0}^{m} b_k \cdot \beta_{k,m,\theta}(x)w(x) = 0, \quad \text{for LB-almost all } x \in \mathbb{R},$$

for some $b_0, \ldots, b_m \in \mathbb{R}$. It follows that $P_m\big(\exp(x)\big) = 0$ for LB-almost all $x \in \mathbb{R}$, where

$$P_m(x) := \sum_{k=0}^{m} b_k \cdot x^k,$$

and, hence, that the polynomial P_m has infinitely many zeros in the interval $(0, \infty)$ so that $b_0 = \cdots = b_m = 0$.

(b) Linearity of $\Gamma_{m,\theta}$ is obvious. Also we easily realize that $\Gamma_{m,\theta}h$ lies in $H_{m,\theta}$ for any $h \in L_2(\mathbb{R}, w)$. Therein note that $\int \beta_{k,m,\theta}(z)w(z)dz > 0$ for all $k = 0, \ldots, m$. Therefore the range of $\Gamma_{m,\theta}$ is finite-dimensional so that $\Gamma_{m,\theta}$ is a compact operator. Since $\gamma_{m,\theta}(x,y) = \gamma_{m,\theta}(y,x)$ for all $x, y \in \mathbb{R}$ the operator $\Gamma_{m,\theta}$ is self-adjoint, i.e. we have that

$$\langle \Gamma_{m,\theta}h, h'\rangle_w = \langle h, \Gamma_{m,\theta}h'\rangle_w, \qquad \forall h, h' \in L_2(\mathbb{R}, w),$$

by Fubini's theorem. Moreover we deduce that

$$\langle h, \Gamma_{m,\theta}h\rangle_w = \sum_{k=0}^{m} \alpha_{k,m,\theta}\Big(\int \beta_{k,m,\theta}(x)h(x)w(x)dx\Big)^2 \Big/ \int \beta_{k,m,\theta}(y)w(y)dy \geq 0,$$

for all $h \in L_2(\mathbb{R}, w)$ so that $\Gamma_{k,\theta}$ is positive semi-definite. As the $\beta_{k,m,\theta}$, $k = 0, \ldots, m$, are linearly independent, as shown in part (a), and $\alpha_{k,m,\theta} > 0$, we have $\langle \beta_{k,m,\theta}, h\rangle_w = 0$ for all $k = 0, \ldots, m$ if and only if $\Gamma_{m,\theta}h = 0$. Therefore the kernel of $\Gamma_{m,\theta}$ equals the orthogonal complement of $H_{m,\theta}$. $\square$

Proof of Proposition 4.1: (a) As $\sum_{k=0}^{m} \alpha_{k,m,\theta}\cdot\beta_{k,m,\theta} = \mathbf{1}$ we have that $\Gamma_{m,\theta}\mathbf{1} = \mathbf{1}$ and, hence, $\mathbf{1}$ is an eigenfunction of $\Gamma_{m,\theta}$ with the eigenvalue 1. For any unit

eigenfunction $\varphi_{j,m,\theta}$ of $\Gamma_{m,\theta}$ it holds that

$$\lambda_{j,m,\theta} \cdot \int |\varphi_{j,m,\theta}(x)| w(x) dx = \int \left| \{\Gamma_{m,\theta}\varphi_{j,m,\theta}\}(x) \right| w(x) dx$$

$$\leq \sum_{k=0}^{m} \alpha_{k,m,\theta} \left| \langle \beta_{k,m,\theta}, \varphi_{j,m,\theta} \rangle_w \right| \leq \int |\varphi_{j,m,\theta}(x)| w(x) dx \,,$$

so that $\lambda_{0,m,\theta} = 1$ is the largest eigenvalue of $\Gamma_{m,\theta}$. Moreover, consider that $\|\mathbf{1}\|_w = 1$.

Whenever $\lambda_{j,m,\theta} = 1$ the above inequality is sharp so that

$$\left| \langle \beta_{k,m,\theta}, \varphi_{j,m,\theta} \rangle_w \right| = \langle \beta_{k,m,\theta}, |\varphi_{j,m,\theta}| \rangle_w \,, \qquad \forall k = 0, \ldots, m \,.$$

As the functions $\beta_{k,m,\theta} \cdot w$ take on only positive values the eigenfunction $\varphi_{j,m,\theta}$ coincides with either a non-negative function or a non-positive function Lebesgue-almost everywhere. For all $j \geq 1$, it holds that

$$\int \varphi_{j,m,\theta}(x) w(x) dx = 0 \,,$$

due to the orthogonality of the eigenfunctions, so that $j = 0$ follows. Therefore the eigenspace with the eigenvalue 1 is the linear span of $\mathbf{1}$.

(b) For some arbitrary $h \in L_2(\mathbb{R}, w)$ we have that

$$\|h\|_{\Gamma,m,\theta}^2 := \langle h, \Gamma_{m,\theta} h \rangle_w = \sum_{k=0}^{m} \alpha_{k,m,\theta} \left| \langle \beta_{k,m,\theta}, h \rangle_w \right|^2 / \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w$$

$$= E \left| \int G_{m,\theta}(x) h(x) w(x) \, dx \right|^2 / \int G_{m,\theta}(x) w(x) \, dx \,, \qquad (6.3)$$

where $K_1, \ldots, K_m$ are independent Bernoulli random variables with $P(K_l = 1) =: \eta_l := \exp\{-\theta_l\}/(1 + \exp\{-\theta_l\})$; and

$$G_{m,\theta}(x) := \exp\left\{ x \sum_{l=1}^{m} K_l \right\} \prod_{l=1}^{m} \frac{1 + \exp(-\theta_l)}{1 + \exp(x - \theta_l)} \,.$$

Courant's min-max theorem provides that

$$\lambda_{j,m,\theta} = \sup_{\mathcal{H}} \inf_{h \in \mathcal{H}, \|h\|_w = 1} E \left| \int G_{m,\theta}(x) h(x) w(x) \, dx \right|^2 / \int G_{m,\theta}(x) w(x) \, dx$$

$$= \sup_{\mathcal{F}} \inf_{f \in \mathcal{F}, \|f\|_2 = 1} E \left| \int G_{m,\theta}(x/\sigma_{m,\theta}) w^{1/2}(x/\sigma_{m,\theta}) f(x) \, dx \right|^2$$

$$\Big/ \int G_{m,\theta}(x/\sigma_{m,\theta}) w(x/\sigma_{m,\theta}) \, dx \,, \qquad (6.4)$$

where the suprema are taken over all $(j + 1)$-dimensional linear subspaces $\mathcal{H}$ and $\mathcal{F}$ of $L_2(\mathbb{R}, w)$ and $L_2(\mathbb{R})$, respectively; and

$$\sigma_{m,\theta}^2 := \sum_{l=1}^{m} \eta_l (1 - \eta_l) \,.$$

Therein we have used that the linear operator which maps each $h \in L_2(\mathbb{R}, w)$ to the function

$$x \mapsto \sigma_{m,\theta}^{-1/2} \, h(x/\sigma_{m,\theta}) \, w^{1/2}(x/\sigma_{m,\theta})$$

is an isometric isomorphism from $L_2(\mathbb{R}, w)$ to $L_2(\mathbb{R})$.

Writing $K_l^* := K_l - \eta_l$ for the centered versions of the K_l the function $G_{m,\theta}(\cdot/\sigma_{m,\theta})$ may be represented by

$$G_{m,\theta}(x/\sigma_{m,\theta}) = \exp\{xZ_{m,\theta}\} \Big/ \prod_{l=1}^{m} g_{l,\theta}(x) \,,$$

with

$$g_{l,\theta}(x) := \eta_l \exp\{(1-\eta_l)x/\sigma_{m,\theta}\} + (1-\eta_l)\exp\{-\eta_l x/\sigma_{m,\theta}\} \,,$$

and

$$Z_{m,\theta} := \sum_{l=1}^{m} K_l^*/\sigma_{m,\theta} \,.$$

Taylor approximation yields that

$$g_{l,\theta}(x) = \exp\left\{\frac{1}{2}\eta_l(1-\eta_l)x^2/\sigma_{m,\theta}^2\right\} \cdot \big(1 + \mathcal{R}_{l,m,\theta}(x)\big) \,,$$

where the remainder term satisfies

$$\big|\mathcal{R}_{l,m,\theta}(x)\big| \leq \frac{1}{24}\,|x|^3\,\sigma_{m,\theta}^{-3}\,\exp\{|x|/\sigma_{m,\theta}\} + \frac{1}{32}\,x^4\,\sigma_{m,\theta}^{-4}\,\exp\left\{x^2/(8\sigma_{m,\theta}^2)\right\} \,,$$

for all $x \in \mathbb{R}$. Hence, for all $|x| \leq \sigma_{m,\theta}$, it holds that

$$\left|G_{m,\theta}(x/\sigma_{m,\theta}) - \exp\left\{Z_{m,\theta}^2/2\right\} \cdot \exp\left\{-\frac{1}{2}\big(x - Z_{m,\theta}\big)^2\right\}\right|$$

$$\leq \exp\left\{Z_{m,\theta}^2/2\right\} \cdot \exp\left\{-\frac{1}{2}\big(x - Z_{m,\theta}\big)^2\right\} \cdot \left(\exp\left\{m|x|^3/(4\sigma_{m,\theta}^3)\right\} - 1\right) \,, \tag{6.5}$$

Since $m/4 \geq \sigma_{m,\theta}^2 \geq m/(1 + \exp\{R\})^2$, it follows for $\rho_m := (D \log m)^{1/2}$ and some constant $D > 0$ that

$$\left|\int_{|x| \leq \rho_m} G_{m,\theta}(x/\sigma_{m,\theta}) w(x/\sigma_{m,\theta}) dx \right.$$

$$\left. - \exp\{Z_{m,\theta}^2/2\} w(0) \int_{|x| \leq \rho_m} \exp\left\{-\frac{1}{2}\big(x - Z_{m,\theta}\big)^2\right\} dx\right|$$

$$\leq \exp\{\delta_m^2/2\} \cdot \mathcal{O}\big(m^{-1/2} \log^2 m\big) \,, \tag{6.6}$$

on the event $\mathcal{E}_{m,\theta}(\delta_m) := \{|Z_{m,\theta}| \leq \delta_m\}$, for some $\delta_m \in (1, \rho_m/2)$, where, throughout this proof, the constants contained in $\mathcal{O}(\cdots)$ may only depend on w, R and D. Therein we impose that the function w and its derivative are uniformly bounded; and that $\rho_m < \sigma_{m,\theta}$, which is satisfied for m sufficiently

large with respect to R and D. Under the same constraints which we have used to establish (6.6) we derive that

$$\left| \int_{|x| \le \rho_m} G_{m,\theta}(x/\sigma_{m,\theta}) f(x) w^{1/2}(x/\sigma_{m,\theta}) \, dx \right.$$

$$\left. - \exp\{Z_{m,\theta}^2/2\} w^{1/2}(0) \int_{|x| \le \rho_m} \exp\left\{ -\frac{1}{2}(x - Z_{m,\theta})^2 \right\} f(x) dx \right|$$

$$\le \exp\{\delta_m^2/2\} \cdot \mathcal{O}\big(m^{-1/2} \log^{7/4} m\big), \tag{6.7}$$

for all $f \in L_2(\mathbb{R})$ with $\|f\|_2 = 1$ using the Cauchy-Schwarz inequality in $L_2(\mathbb{R})$.

Moreover, if $\mathcal{E}_{m,\theta}(\delta_m)$ occurs, we have that

$$\left| \exp\{Z_{m,\theta}^2/2\} w(0) \int_{|x| > \rho_m} \exp\left\{ -\frac{1}{2}(x - Z_{m,\theta})^2 \right\} dx \right| \le \exp\{\delta_m^2/2\} \cdot \mathcal{O}\big(m^{-D/8}\big), \tag{6.8}$$

and that

$$\exp\{Z_{m,\theta}^2/2\} w^{1/2}(0) \int_{|x| > \rho_m} \exp\left\{ -\frac{1}{2}(x - Z_{m,\theta})^2 \right\} |f(x)| dx$$

$$\le \exp\{\delta_m^2/2\} \cdot \mathcal{O}\big(m^{-D/8}\big). \tag{6.9}$$

Finally we study the terms

$$\int_{|x| > \rho_m} G_{m,\theta}(x/\sigma_{m,\theta}) w(x/\sigma_{m,\theta}) \, dx, \tag{6.10}$$

$$\int_{|x| > \rho_m} G_{m,\theta}(x/\sigma_{m,\theta}) |f(x)| w^{1/2}(x/\sigma_{m,\theta}) \, dx. \tag{6.11}$$

Taking the derivative of $\log G_{m,\theta}$ we deduce that $G_{m,\theta}$ is a strictly unimodal function on the event $\left\{ \sum_{l=1}^m K_l \in [1, m-1] \right\}$ (and hence on $\mathcal{E}_{m,\theta}(\delta_m)$ for m large enough with respect to R) where the unique mode $x_{m,\theta}$ satisfies

$$\Phi_{m,\theta}(x_{m,\theta}) = Z_{m,\theta}, \tag{6.12}$$

with

$$\Phi_{m,\theta}(x) := \frac{1}{\sigma_{m,\theta}} \sum_{l=1}^m \left(\frac{1}{1 + \exp\{-\theta_l\}} - \frac{1}{1 + \exp\{x - \theta_l\}} \right)$$

$$= (\exp\{x\} - 1) \, \sigma_{m,\theta} - (\exp\{x\} - 1)^2 \frac{1}{\sigma_{m,\theta}} \sum_{l=1}^m \eta_l^2/(1 + \exp\{x - \theta_l\}).$$

The function $\Phi_{m,\theta}$ increases strictly and satisfies

$$\left| \Phi_{m,\theta}(\pm\rho_m/\sigma_{m,\theta}) \mp \rho_m \right| = \mathcal{O}\big(m^{-1/2} \log m\big),$$

so that, on the event $\mathcal{E}_{m,\theta}(\delta_m)$, the mode $x_{m,\theta}$ is located in the interval $[-\rho_m/\sigma_{m,\theta}, \rho_m/\sigma_{m,\theta}]$ for $\rho_m > 2\delta_m$ and m sufficiently large with respect to

w, R and D. Therefore the term (6.10) is bounded from above by

$$\max \left\{ G_{m,\theta}(\rho_m/\sigma_{m,\theta}), G_{m,\theta}(-\rho_m/\sigma_{m,\theta}) \right\} \cdot \sigma_{m,\theta} \leq \exp\{\delta_m^2/2\} \cdot \mathcal{O}\big(m^{(4-D)/8}\big) \,,$$

and the term (6.11) obeys the upper bound

$$\max \left\{ G_{m,\theta}(\rho_m/\sigma_{m,\theta}), G_{m,\theta}(-\rho_m/\sigma_{m,\theta}) \right\} \cdot \sigma_{m,\theta}^{1/2} \leq \exp\{\delta_m^2/2\} \cdot \mathcal{O}\big(m^{(2-D)/8}\big) \,,$$

if $\mathcal{E}_{m,\theta}(\delta_m)$ occurs, thanks to (6.5). Combining these results with (6.6)–(6.9), we deduce that

$$\left| \int G_{m,\theta}(x/\sigma_{m,\theta}) w(x/\sigma_{m,\theta}) dx - \sqrt{2\pi} \exp\{Z_{m,\theta}^2/2\} w(0) \right|$$
$$\leq \exp\{\delta_m^2/2\} \cdot \mathcal{O}\big(m^{-1/2} \log^2 m\big) \,,$$

$$\left| \int G_{m,\theta}(x/\sigma_{m,\theta}) f(x) w^{1/2}(x/\sigma_{m,\theta}) dx \right.$$
$$\left. - \exp\{Z_{m,\theta}^2/2\} \, w^{1/2}(0) \int \exp \left\{ -\frac{1}{2}(x - Z_{m,\theta})^2 \right\} f(x) dx \right|$$
$$\leq \exp\{\delta_m^2/2\} \cdot \mathcal{O}\big(m^{-1/2} \log^{7/4} m\big) \,, \tag{6.13}$$

on $\mathcal{E}_{m,\theta}(\delta_m)$ when $D \geq 8$. Whenever $w(0) > 0$ holds, in addition, then (6.4) and (6.13) provide that

$$\lambda_{j,m,\theta}$$
$$\geq \sup_{\mathcal{F}} \inf_{f \in \mathcal{F}, \|f\|_2 = 1} E \, 1_{\mathcal{E}_{m,\theta}(\delta_m)} \left| \int G_{m,\theta}(x/\sigma_{m,\theta}) w^{1/2}(x/\sigma_{m,\theta}) f(x) dx \right|^2$$
$$\Big/ \int G_{m,\theta}(x/\sigma_{m,\theta}) w(x/\sigma_{m,\theta}) dx$$
$$\geq \sup_{\mathcal{F}} \inf_{f \in \mathcal{F}, \|f\|_2 = 1} E \, 1_{\mathcal{E}_{m,\theta}(\delta_m)} \big| [\mathcal{N} * f](Z_{m,\theta}) \big|^2 / \mathcal{N}(Z_{m,\theta})$$
$$- \exp\{\delta_m^2/2\} \cdot \mathcal{O}\big(m^{-1/2} \log^2 m\big) \,, \tag{6.14}$$

where $\mathcal{N}$ denotes the $\mathsf{N}(0,1)$-density and $*$ denotes convolution.

Now we focus on the distribution of $Z_{m,\theta}$. Writing

$$\mu_{m,\theta} := \sum_{l=1}^{m} \eta_l \,,$$

Fourier inversion yields that

$$P\big(Z_{m,\theta} = (k - \mu_{m,\theta})/\sigma_{m,\theta}\big) = P\Big(\sum_{l=1}^{m} K_l = k \Big)$$
$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\{-ikt\} \prod_{l=1}^{m} \big(\eta_l \exp\{it\} + 1 - \eta_l\big) dt$$
$$= \sigma_{m,\theta}^{-1} \mathcal{N}\big((k - \mu_{m,\theta})/\sigma_{m,\theta}\big) \pm \mathcal{O}\big(m^{-1} \log^2 m\big) \,,$$

for all integer $k = 0, \dots, m$. It follows that

$$
\begin{aligned}
E\, 1_{\mathcal{E}_{m,\theta}(\delta_m)} &\big| \big[\mathcal{N} * f\big](Z_{m,\theta}) \big|^2 / \mathcal{N}(Z_{m,\theta}) \\
&\geq \sum_{k=0}^{m} 1\big\{ |k - \mu_{m,\theta}| \leq \delta_m \cdot \sigma_{m,\theta} \big\} \cdot \big| \big[\mathcal{N} * f\big]\big((k - \mu_{m,\theta})/\sigma_{m,\theta}\big) \big|^2 / \sigma_{m,\theta} \\
&\qquad\qquad\qquad\qquad\qquad\qquad - \exp\{\delta_m^2/2\} \cdot \mathcal{O}\big(\delta_m m^{-1/2} \log^2 m\big),
\end{aligned}
\tag{6.15}
$$

since

$$
\|\mathcal{N} * f\|_\infty \; \leq \; \frac{1}{2\pi} \int \exp\{-t^2/2\} |f^{ft}(t)| dt \; \leq \; 2^{-1/2} \pi^{-1/4},
$$

for all $f \in L_2(\mathbb{R})$ with $\|f\|_2 = 1$ by the Cauchy-Schwarz inequality and Parseval's identity where f^{ft} denotes the Fourier transform of the function f. Analogously we deduce that

$$
\begin{aligned}
\big| \big[\mathcal{N} * f\big](x) &- \big[\mathcal{N} * f\big](y) \big| \\
&= \frac{1}{2\pi} \int \big| \exp\{-ixt\} - \exp\{-iyt\} \big| \exp\{-t^2/2\} |f^{ft}(t)| dt \; \leq \; \pi^{-1/2} \cdot |x - y|,
\end{aligned}
$$

for all $x, y \in \mathbb{R}$, so that

$$
\begin{aligned}
\sum_{k=0}^{m} & 1\big\{ |k - \mu_{m,\theta}| \leq \delta_m \cdot \sigma_{m,\theta} \big\} \cdot \big| \big[\mathcal{N} * f\big]\big((k - \mu_{m,\theta})/\sigma_{m,\theta}\big) \big|^2 / \sigma_{m,\theta} \\
&= \sum_{k=0}^{m} 1\big\{ |k - \mu_{m,\theta}| \leq \delta_m \cdot \sigma_{m,\theta} \big\} \\
&\qquad\qquad\qquad \cdot \int_{x=(k-\mu_{m,\theta}-1/2)/\sigma_{m,\theta}}^{(k-\mu_{m,\theta}+1/2)/\sigma_{m,\theta}} \big| \big[\mathcal{N} * f\big]\big((k - \mu_{m,\theta})/\sigma_{m,\theta}\big) \big|^2 dx \\
&\geq \int_{|x| \leq \delta_m - 1/(2\sigma_{m,\theta})} \big| \big[\mathcal{N} * f\big](x) \big|^2 dx \; - \; \mathcal{O}\big(\delta_m \cdot m^{-1/2}\big).
\end{aligned}
\tag{6.16}
$$

Piecing together (6.14)–(6.16) yields that

$$
\begin{aligned}
\lambda_{j,m,\theta} \geq \sup_{\mathcal{F}} \inf_{f \in \mathcal{F}, \|f\|_2 = 1} & \int_{|x| \leq \delta_m - 1/(2\sigma_{m,\theta})} \big| \big[\mathcal{N} * f\big](x) \big|^2 dx \\
&\qquad - \exp\{\delta_m^2/2\} \cdot \mathcal{O}\big(\delta_m m^{-1/2} \log^2 m\big),
\end{aligned}
\tag{6.17}
$$

for all integer $j \geq 0$.

We define

$$
L_2^S(\mathbb{R}) \; := \; \big\{ f \in L_2(\mathbb{R}) \, : \, \operatorname{supp} f \subseteq [-S, S] \big\},
$$

which forms an infinite-dimensional closed linear subspace of $L_2(\mathbb{R})$ for any $S > 0$. Also we fix that $\delta_m := c \log^{1/2} m$ for some constant $c \in (0, 1)$. By

Jensen's inequality we derive that

$$\int_{|x|>R} \big|[\mathcal{N}*f](x)\big|^2 dx \ \leq \ E \int_{|x|>R} |f(x-Z)|^2 dx \ \leq \ 2\,E \int_{x>R-|Z|} |f(x)|^2 dx$$

$$\leq \ 2 \int_{x>R/2} |f(x)|^2 dx \ + \ 2\,P(|Z|>R/2)\,,$$

for any $R>0$, $f \in L_2(\mathbb{R})$ with $\|f\|_2 = 1$ and $\mathcal{L}(Z) = \mathsf{N}(0,1)$. Putting $R = \delta_m$ and using Parseval's identity we deduce from (6.17) that

$$\lambda_{j,m,\theta} \ \geq \ \inf_{f \in \mathcal{F}, \|f\|_2 = 1} \frac{1}{2\pi} \int \exp\{-t^2\} \big|f^{ft}(t)\big|^2 dt \ - \ \mathcal{O}\big(m^{(c^2-1)/2} \log^{5/2} m\big)$$

$$- \ \mathcal{O}\big(m^{-c^2/2}\big)\,, \quad (6.18)$$

imposing that $\mathcal{F}$ is a $(j+1)$-dimensional linear subspace of $L_2^{\delta_m/4}(\mathbb{R})$, where f^{ft} denotes the Fourier transform of the function f. More precisely, we select any orthonormal system $f_0, \ldots, f_j$ in $L_2^1(\mathbb{R})$ where the f_l are located in the first order Sobolev space, in addition; and put $\mathcal{F}$ equal to the linear hull of the $x \mapsto 2\delta_m^{-1/2} f_l(4x/\delta_m)$, $l = 0, \ldots, j$. Then, for all $f \in \mathcal{F}$ with $\|f\|_2 = 1$, we have that

$$\inf_{f \in \mathcal{F}, \|f\|_2 = 1} \frac{1}{2\pi} \int \exp\{-t^2\} \big|f^{ft}(t)\big|^2 dt$$

$$= \ \inf \Big\{ \frac{1}{2\pi} \int \exp\big\{ -4t^2/\delta_m^2 \big\} \Big| \sum_{l=0}^{j} \alpha_l f_l^{ft}(t) \Big|^2 dt \ : \ \sum_{l=0}^{j} \alpha_l^2 = 1 \Big\}$$

$$\geq \ \exp\{-4d^2/c^2\} \cdot \Big(1 - d^{-2}(\log m)^{-1} \sup \Big\{ \int \Big| \sum_{l=0}^{j} \alpha_l f_l'(t) \Big|^2 dt : \sum_{l=0}^{j} \alpha_l^2 = 1 \Big\} \Big),$$

$$(6.19)$$

for any $d>0$ where we have used integral substitution, the Fourier representation of the Sobolev norm and Parseval's identity again. Let φ be some continuously differentiable function which is supported on $[-1,1]$ and satisfies $\|\varphi\|_2 = 1$. Then put

$$f_l(x) \ := \ \sqrt{2j}\,\varphi\big(2j(x - [1+2l]/[2j])\big)\,, \quad l = 0, \ldots, j\,.$$

The derivatives f_l' form an orthogonal system as well. In particular we have

$$\int \Big| \sum_{l=0}^{j} \alpha_l f_l'(t) \Big|^2 dt \ = \ 4j^2 \|\varphi'\|_2^2\,,$$

whenever $\sum_{l=0}^{j} a_l^2 = 1$ so that (6.18) and (6.19) yield the claimed asymptotic behaviour of the eigenvalues.

(c) Assume convergence of some subsequence $(\Gamma_{\sigma(m),\theta_{(\sigma(m))}})_m$ to some continuous linear operator Γ. As the set of all compact, self-adjoint and positive

semidefinite operators from $L_2(\mathbb{R}, w)$ to itself is closed, Γ is compact, self-adjoint and positive semidefinite as well. Let λ_j denote the $(j+1)$th largest eigenvalue of Γ for any integer $j \geq 0$. Thanks to the imposed convergence and part (b) of the proposition, it holds that $\lambda_j = 1$ since

$$\left|\lambda_{j,\sigma(m),\theta_{(\sigma(m))}} - \lambda_j\right| \leq \left\|\Gamma_{\sigma(m),\theta_{(\sigma(m))}} - \Gamma\right\|.$$

Therefore the eigenspace of Γ with the eigenvalue 1 is infinite-dimensional, which contradicts the compactness of Γ.   □

Proof of Lemma 4.1: We derive that

$$\beta'_{k,m,\theta}(x) = \beta_{k,m,\theta}(x) \cdot \left(k - m + \sum_{l=1}^{m} 1/\big(1 + \exp(x - \theta_l)\big)\right).$$

Thus, for $k \in [1, m-1]$, there exists a unique $x_{k,m,\theta}$ such that $\beta_{k,m,\theta}(x_{k,m,\theta}) = \|\beta_{k,m,\theta}\|_\infty$ where $\beta'_{k,m,\theta}(x_{k,m,\theta}) = 0$ so that

$$x_{k,m,\theta} \in [-R, R] + \log(k/(m-k)) \subseteq \big[-R - \log(m-1), R + \log(m-1)\big].$$

We establish that

$$\langle\beta_{k,m,\theta}, \mathbf{1}\rangle_w = \int \beta_{k,m,\theta}(x)w(x)dx$$

$$= \inf\big\{\beta_{k,m,\theta}(x) : |x - x_{k,m,\theta}| \leq \Delta\big\} \cdot \int_{|x-x_{k,m,\theta}| \leq \Delta} w(x)dx$$

$$\geq \|\beta_{k,m,\theta}\|_\infty\big(1 - m\Delta\big) \cdot \int_{|x-x_{k,m,\theta}| \leq \Delta} w(x)dx,$$

for any $\Delta \in (0, 1/m)$ since $\|\beta'_{k,m,\theta}\|_\infty \leq m \cdot \|\beta_{k,m,\theta}\|_\infty$. Now let us choose $\Delta := 1/(2m)$ so that

$$\langle\beta_{k,m,\theta}, \mathbf{1}\rangle_w \geq \frac{1}{2}\|\beta_{k,m,\theta}\|_\infty \cdot \int_{|x-x_{k,m,\theta}| \leq 1/(2m)} w(x)dx$$

$$\geq D_0 \exp\{-D_1(R + 1/2)\} \cdot m^{-1-D_1} \cdot \|\beta_{k,m,\theta}\|_\infty.$$

If $k = 0$ then $\|\beta_{k,m,\theta}\|_\infty = 1$ and

$$\langle\beta_{k,m,\theta}, \mathbf{1}\rangle_w \geq \int_{-\infty}^{-R-\log m} \beta_{0,m,\theta}(x)w(x)dx \geq (1 + 1/m)^{-m} \int_{-\infty}^{-R-\log m} w(x)dx$$

$$\geq \frac{2D_0}{D_1} \exp\{-D_1 R - 1\} \cdot m^{-D_1}.$$

Finally, in the case of $k = m$, again $\|\beta_{k,m,\theta}\|_\infty = 1$ as $\sum_{l=1}^{m} \theta_l = 0$ and

$$\langle\beta_{k,m,\theta}, \mathbf{1}\rangle_w \geq \int_{R+\log m}^{\infty} \beta_{m,m,\theta}(x)w(x)dx \geq (1 + 1/m)^{-m} \int_{R+\log m}^{\infty} w(x)dx$$

$$\geq \frac{2D_0}{D_1} \exp\{-D_1 R - 1\} \cdot m^{-D_1} \,,$$

so that, for some constant $D^* > 0$, it holds that

$$\max_{k=0,\ldots,m} \|\beta_{k,m,\theta}\|_\infty / \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w \leq (D^*)^2 \cdot m^{1+D_1} \,. \tag{6.20}$$

Then, for any $j \leq m$,

$$\begin{aligned}
\lambda_{j,m,\theta} \cdot \|\varphi_{j,m,\theta}\|_\infty &= \|\Gamma_{m,\theta}\varphi_{j,m,\theta}\|_\infty \\
&\leq \Big\| \sum_{k=0}^m \alpha_{k,m,\theta} \cdot \beta_{k,m,\theta} \cdot \langle \beta_{k,m,\theta}, \varphi_{j,m,\theta} \rangle_w / \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w \Big\|_\infty \\
&\leq \Big\| \sum_{k=0}^m \alpha_{k,m,\theta} \cdot \beta_{k,m,\theta} \cdot \langle \beta_{k,m,\theta}, \varphi_{j,m,\theta} \rangle_w^2 / \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w^2 \Big\|_\infty^{1/2} \\
&\leq \Big(\max_{k=0,\ldots,m} \|\beta_{k,m,\theta}\|_\infty / \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w \Big)^{1/2} \cdot \lambda_{j,m,\theta}^{1/2} \,,
\end{aligned}$$

using Jensen's inequality in the second line as $\sum_{k=0}^m \alpha_{k,m,\theta}\beta_{k,m,\theta} = \mathbf{1}$. Piecing together this inequality with (6.20) completes the proof. $\qquad\square$

Proof of Lemma 4.2: By Taylor approximation we deduce for any fixed m and $\theta', \theta \in \Theta$ that

$$\Big\| \Gamma_{m,\theta'} - \Gamma_{m,\theta} - \sum_{l=1}^m (\theta_l' - \theta_l)\, \Gamma_{l,m,\theta} \Big\| = o(|\theta - \theta'|) \,,$$

where, here, $\|\cdot\|$ stands for the induced operator norm of $L_2(\mathbb{R}, w)$ and $\Gamma_{l,m,\theta}$ denotes the compact linear operator which maps any $h \in L_2(\mathbb{R}, w)$ to the function

$$x \mapsto \int \frac{\partial \gamma_{m,\theta}}{\partial \theta_l}(x, y)\, h(y)\, w(y)\, dy \,.$$

Hence, Courant's min-max theorem yields that

$$\begin{aligned}
\lambda_{j,m,\theta'} \geq \min \Big\{ & \sum_{k=0}^j v_k^2 \lambda_{k,m,\theta} + \sum_{k,k'=0}^j v_k v_{k'} \sum_{l=1}^m \langle \varphi_{k,m,\theta}, \Gamma_{l,m,\theta}\varphi_{k',m,\theta} \rangle_w (\theta_l' - \theta_l) \\
& \hspace{5cm} : v \in \mathbb{R}^{j+1}, |v| = 1 \Big\} \\
& - o(|\theta' - \theta|),
\end{aligned} \tag{6.21}$$

for any $\theta, \theta' \in \Theta$ where the remainder term does not depend on j. Taking the limit $\theta' \to \theta$ that vector v which minimizes the term in (6.21) tends to $(0, \ldots, 0, 1)^\dagger$ for any fixed j and m so that

$$\lambda_{j,m,\theta'} \geq \lambda_{j,m,\theta} + \sum_{l=1}^m \langle \varphi_{j,m,\theta}, \Gamma_{l,m,\theta}\varphi_{j,m,\theta} \rangle_w (\theta_l' - \theta_l) - o(|\theta' - \theta|). \tag{6.22}$$

On the other hand Courant's min-max theorem also provides that

$$
\lambda_{j,m,\theta'} \leq \max \Big\{ \sum_{k \geq j} v_k^2 \, \lambda_{k,m,\theta} \; + \; \sum_{k,k' \geq j} v_k \, v_{k'} \, \sum_{l=1}^{m} \langle \varphi_{k,m,\theta}, \Gamma_{l,m,\theta} \varphi_{k',m,\theta} \rangle_w \, (\theta_l' - \theta_l)
$$
$$
: \sum_{k \geq j} v_k^2 = 1 \Big\}
$$
$$
+ \, o(|\theta' - \theta|). \tag{6.23}
$$

Again, as $\theta' \to \theta$, that sequence $(v_k)_{k \geq j}$ which maximizes the term in (6.23) converges to $(1, 0, 0, \ldots)$ in the Hilbert space $l_2(\{j, j+1, \ldots\})$ so that

$$
\lambda_{j,m,\theta'} \; \leq \; \lambda_{j,m,\theta} \; + \; \sum_{l=1}^{m} \langle \varphi_{j,m,\theta}, \Gamma_{l,m,\theta} \varphi_{j,m,\theta} \rangle_w \, (\theta_l' - \theta_l) \; + \; o(|\theta' - \theta|). \tag{6.24}
$$

Combining (6.22) and (6.24) we conclude that the function $\theta \mapsto \lambda_{j,m,\theta}$ is continuously differentiable and its partial derivatives satisfy

$$
\frac{\partial}{\partial \theta_l} \lambda_{j,m,\theta} \; = \; \langle \varphi_{j,m,\theta}, \Gamma_{l,m,\theta} \varphi_{j,m,\theta} \rangle_w
$$
$$
= \; \sum_{k=0}^{m} \alpha_{k,m,\theta}^{[l]} \, \langle \beta_{k,m,\theta}, \varphi_{j,m,\theta} \rangle_w^2 / \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w \; + \; 2 \lambda_{j,m,\theta} \langle \psi_{l,\theta} \cdot \varphi_{j,m,\theta}, \varphi_{j,m,\theta} \rangle_w \,,
$$

with

$$
\alpha_{k,m,\theta}^{[l]} \; := \; - \sum_{b \in B(k,m)} \exp\{-b^\dagger \theta\} \cdot b_l \; - \; \alpha_{k,m,\theta} \langle \beta_{k,m,\theta} \cdot \psi_{l,\theta}, \mathbf{1} \rangle_w / \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w \,,
$$
$$
\psi_{l,\theta}(x) \; := \; \exp\{-\theta_l\} / \big(\exp\{-x\} + \exp\{-\theta_l\} \big) \,.
$$

As $\| \psi_{l,\theta} \|_\infty \leq 1$ and $0 \geq \alpha_{k,m,\theta}^{[l]} \geq -2\alpha_{k,m,\theta}$ the claim of the lemma has been verified. □

Proof of Proposition 4.2: Orthogonal expansion with respect to $\{\varphi_{j,m,\theta}\}_j$ yields that

$$
E \| \hat{h}_n - h \|_w^2 \; = \; \sum_{j=1}^{m} E \Big| \mathfrak{w}_{j,m,\theta} \sum_{k=0}^{m} \langle \beta_{k,m,\theta}, \varphi_{j,m,\theta} \rangle_w \cdot Y_k
$$
$$
/ \big\{ n \lambda_{j,m,\theta} \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w \big\} \; - \; \langle h, \varphi_{j,m,\theta} \rangle_w \Big|^2 + \sum_{j > m} \langle h, \varphi_{j,m,\theta} \rangle_w^2 \,. \tag{6.25}
$$

Since

$$
E \sum_{k=0}^{m} \langle \beta_{k,m,\theta}, \varphi_{j,m,\theta} \rangle_w \cdot Y_k / \big\{ n \lambda_{j,m,\theta} \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w \big\}
$$

$$= \sum_{k=0}^{m} \alpha_{k,m,\theta} \langle \beta_{k,m,\theta}, \varphi_{j,m,\theta} \rangle_w \cdot \langle \beta_{k,m,\theta}, h \rangle_w / \left\{ \lambda_{j,m,\theta} \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w \right\}$$

$$= \langle \Gamma_{m,\theta} \varphi_{j,m,\theta}, h \rangle_w / \lambda_{j,m,\theta} = \langle \varphi_{j,m,\theta}, h \rangle_w,$$

we have that

$$E \left| \mathfrak{w}_{j,m,\theta} \sum_{k=0}^{m} \langle \beta_{k,m,\theta}, \varphi_{j,m,\theta} \rangle_w \cdot Y_k / \left\{ n \lambda_{j,m,\theta} \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w \right\} - \langle h, \varphi_{j,m,\theta} \rangle_w \right|^2$$

$$= \mathfrak{w}_{j,m,\theta}^2 \operatorname{var} \left(\sum_{k=0}^{m} \langle \beta_{k,m,\theta}, \varphi_{j,m,\theta} \rangle_w \cdot Y_k / \left\{ n \lambda_{j,m,\theta} \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w \right\} \right)$$

$$+ |\mathfrak{w}_{j,m,\theta} - 1|^2 \langle \varphi_{j,m,\theta}, h \rangle_w^2$$

$$= \mathfrak{w}_{j,m,\theta}^2 \sum_{k=0}^{m} \alpha_{k,m,\theta} \langle \beta_{k,m,\theta}, \varphi_{j,m,\theta} \rangle_w^2 \langle \beta_{k,m,\theta}, h \rangle_w / \left\{ n \lambda_{j,m,\theta}^2 \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w^2 \right\}$$

$$+ |\mathfrak{w}_{j,m,\theta} - 1|^2 \langle \varphi_{j,m,\theta}, h \rangle_w^2$$

$$\leq \mathfrak{w}_{j,m,\theta}^2 \langle \Gamma_{m,\theta} \varphi_{j,m,\theta}, \varphi_{j,m,\theta} \rangle_w \cdot \lambda_{j,m,\theta}^{-2} n^{-1} \cdot \max_{k=0,\dots,m} |\langle \beta_{k,m,\theta}, h \rangle_w| / \langle \beta_{k,m,\theta}, \mathbf{1} \rangle_w$$

$$+ |\mathfrak{w}_{j,m,\theta} - 1|^2 \langle \varphi_{j,m,\theta}, h \rangle_w^2$$

$$\leq \mathfrak{w}_{j,m,\theta}^2 \cdot \|h\|_\infty \lambda_{j,m,\theta}^{-1} \cdot n^{-1} + |\mathfrak{w}_{j,m,\theta} - 1|^2 \langle \varphi_{j,m,\theta}, h \rangle_w^2,$$

which completes the proof. $\qquad\square$

Proof of Theorem 4.2: In the experiment $\mathcal{Y}_{n,m}$ one observes (X, Y). As the conditional distribution of X given Y does not depend on f, the statistic Y is sufficient for f (and, thus, h) when θ is known so that $\hat{h}_n$ may be viewed as being based on Y. We introduce the functions

$$h_\nu := \mathbf{1} + \sum_{k=0}^{m} \mathfrak{h}_{m,k}^* \cdot \nu_k \cdot \varphi_{k,m,\theta},$$

with $\nu_k \in \{-1, 1\}$ for all $k = 1, \dots, m$ and

$$\mathfrak{h}_{m,k}^* := c_{\mathfrak{h}} \cdot \mathfrak{h}_{m,k} / \sqrt{1 + n \cdot \lambda_{k,m,\theta} \cdot \mathfrak{h}_{m,k}^2},$$

for some constant $c_{\mathfrak{h}} \in (0, 1/\min\{1, 2D^* \mathfrak{c}^*\})$. Also we define $f_\nu := h_\nu \cdot w$. Thanks to Proposition 4.1(a), f_ν integrates to 1 for all ν; and thanks to Lemma 4.1, $|\mathbf{1} - h_\nu|$ is bounded by $1/2$ so that non-negativity of f_ν is guaranteed. Furthermore h_ν is located in $\mathcal{H}$ for all $\nu \in \{-1, 1\}^m$.

Put $\rho_n := \sum_{k=1}^{m} (\mathfrak{h}_{m,k}^*)^2$ and assume that

$$\limsup_{n \to \infty} \sup_{h \in \mathcal{H}} P_h \left(\|\hat{h}_n - h\|_w^2 > z \cdot \rho_n \right) = 0,$$

for all $z > 0$. Since

$$\sup_{h \in \mathcal{H}} E_h \frac{\|\hat{h}_n - h\|_w^2}{\rho_n + \|\hat{h}_n - h\|_w^2} \leq \int_0^\infty \sup_{h \in \mathcal{H}} P_h \left(\|\hat{h}_n - h\|_w^2 > z \rho_n \right) (1 + z)^{-2} dz,$$

Fatou's lemma implies that

$$\limsup_{n \to \infty} \sup_{h \in \mathcal{H}} E_h \frac{\|\hat{h}_n - h\|_w^2}{\rho_n + \|\hat{h}_n - h\|_w^2} = 0 \,.$$

Then we estimate the maximal risk from below by the Bayesian risk with respect to the a priori measure Q under which all ν_k are i.i.d. random variables with $Q(\nu_1 = 1) = Q(\nu_1 = -1) = 1/2$. Therefore,

$$\limsup_{n \to \infty} E_Q \, E_{h_\nu} \frac{\|\hat{h}_n - h_\nu\|_w^2}{\rho_n + \|\hat{h}_n - h_\nu\|_w^2} = 0 \,. \tag{6.26}$$

Writing $\hat{h}_{n,j} := \langle \hat{h}_n, \varphi_{j,m,\theta} \rangle_w$, Bessel's inequality yields that

$$\|\hat{h}_n - h_\nu\|_w^2 \geq \sum_{j=1}^m \left| \hat{h}_{n,j} - \mathfrak{h}_{m,j}^* \nu_j \right|^2 \,.$$

As we are considering lower bounds we may assume $|\hat{h}_{n,j}| \leq \mathfrak{h}_{m,j}^*$ for all j without any loss of generality so that

$$\sum_{j=1}^m \left| \hat{h}_{n,j} - \mathfrak{h}_{m,j}^* \nu_j \right|^2 \leq 4 \sum_{j=1}^m (\mathfrak{h}_{m,j}^*)^2 = 4\rho_n \,, \qquad Q\text{-a.s.} \,.$$

Thus it follows from (6.26) that

$$\limsup_{n \to \infty} \sum_{j=1}^m E_Q \, E_{h_\nu} \, |\hat{h}_{n,j} - \mathfrak{h}_{m,j}^* \nu_j|^2 / (5\rho_n) = 0 \,. \tag{6.27}$$

By $\eta_{\nu,j,\pm}$ we denote the $\mathsf{N}(nq(\theta, h_{\nu,j,\pm}), nQ(\theta, h_{\nu,j,\pm}))$-density where $h_{\nu,j,\pm}$ stands for the function h_ν when ν_j is replaced by ± 1. We deduce that

$$E_Q \, E_{h_\nu} \, |\hat{h}_{n,j} - \mathfrak{h}_{m,j}^* \nu_j|^2$$

$$= \frac{1}{2} E_Q \int \left(|\hat{h}_{n,j}(x) - \mathfrak{h}_{m,j}^*|^2 \eta_{\nu,j,+}(x) + |\hat{h}_{n,j}(x) + \mathfrak{h}_{m,j}^*|^2 \eta_{\nu,j,-}(x) \right) dx$$

$$\geq (\mathfrak{h}_{m,j}^*)^2 E_Q \int \min \left\{ \eta_{\nu,j,+}, \eta_{\nu,j,-} \right\}(x) dx \,,$$

and, hence,

$$\sum_{j=1}^m E_Q \, E_{h_\nu} \, |\hat{h}_{n,j} - \mathfrak{h}_{m,j}^* \nu_j|^2 / (5\rho_n) \geq \frac{1}{5} \inf_{j=1,\dots,m} E_Q \int \min \left\{ \eta_{\nu,j,+}, \eta_{\nu,j,-} \right\}(x) dx \,.$$

This represents a contradiction to (6.27) if we can show that

$$\liminf_{n \to \infty} \inf_{j=1,\dots,m} E_Q \int \min \left\{ \eta_{\nu,j,+}, \eta_{\nu,j,-} \right\}(x) dx > 0 \,. \tag{6.28}$$

Le Cam's inequality provides that

$$\int \min \left\{ \eta_{\nu,j,+}, \eta_{\nu,j,-} \right\}(x) dx \geq 1 - \left\{ 1 - \left(1 - H^2(\eta_{\nu,j,+}, \eta_{\nu,j,-}) \right)^2 \right\}^{1/2} \,,$$

with the Hellinger distance $H(f,g)$. The squared Hellinger distance between the normal densities $\eta_{\nu,j,+}$ and $\eta_{\nu,j,-}$ equals

$$H^2(\eta_{\nu,j,+},\eta_{\nu,j,-})$$

$$= 1 - \exp\left\{-\frac{n}{4}\sum_{k=0}^{m}(q_k(\theta,h_{\nu,j,+}) - q_k(\theta,h_{\nu,j,-}))^2/(q_k(\theta,h_{\nu,j,+}) + q_k(\theta,h_{\nu,j,-}))\right\}$$

$$\cdot \prod_{k=0}^{m}\left\{1 - \left(\frac{q_k(\theta,h_{\nu,j,+}) - q_k(\theta,h_{\nu,j,+})}{q_k(\theta,h_{\nu,j,+}) + q_k(\theta,h_{\nu,j,+})}\right)^2\right\}^{1/4}.$$

Therein we write $q_k(\theta,h)$ for $q_k(\theta,F)$ when F is the distribution induced by the density hw. Note that

$$\sum_{k=0}^{m}\big(q_k(\theta,h_{\nu,j,+}) - q_k(\theta,h_{\nu,j,-})\big)^2/\big(q_k(\theta,h_{\nu,j,+}) + q_k(\theta,h_{\nu,j,-})\big)$$

$$\leq 4(\mathfrak{h}_{m,j}^*)^2 \cdot \langle\Gamma_{m,\theta}\varphi_{j,m,\theta},\varphi_{j,m,\theta}\rangle_w = 4(\mathfrak{h}_{m,j}^*)^2 \lambda_{j,m,\theta},$$

since $h_{\nu,j,+}(x) + h_{\nu,j,-}(x) \geq 1$ for all $x \in \mathbb{R}$. Moreover,

$$\big|q_k(\theta,h_{\nu,j,+}) - q_k(\theta,h_{\nu,j,-})\big|/\big(q_k(\theta,h_{\nu,j,+}) + q_k(\theta,h_{\nu,j,-})\big)$$

$$\leq 2\mathfrak{h}_{m,j}^*\big|\langle\beta_{k,m,\theta},\varphi_{j,m,\theta}\rangle_w\big|/\big|\langle\beta_{k,m,\theta},\mathbf{1}\rangle_w \leq 2D^*m^{(1+D_1)/2}\lambda_{j,m,\theta}^{-1/2}\,\mathfrak{h}_{m,j}^*,$$

for all $k = 0,\dots,m$, thanks to Lemma 4.1. We conclude that

$$H^2(\eta_{\nu,j,+},\eta_{\nu,j,-}) \leq 1 - \exp\big\{-n(\mathfrak{h}_{m,j}^*)^2\lambda_{j,m,\theta}\big\}$$

$$\cdot \big(1 - 4(D^*)^2m^{1+D_1}(\mathfrak{h}_{m,j}^*)^2\lambda_{j,m,\theta}^{-1}\big)^{(m+1)/4}$$

$$\leq 1 - 2^{-1/4}\cdot\exp\big\{-c_{\mathfrak{h}}^2\big\},$$

if $c_{\mathfrak{h}} \leq 1/(4c^*D^*)$ so that (6.28) has been verified. $\qquad\square$

Proof of Theorem 4.3: Define

$$\tilde{h} := h + c\cdot v,$$

for some constant $c > 0$ and some element v of the orthogonal complement $H_{m,\theta}^{\perp}$ of the linear subspace $H_{m,\theta}$ of $L_2(\mathbb{R},w)$. Without any loss of generality, v may be viewed as essentially bounded, which can be seen as follows: Clearly the set of all essentially bounded functions is dense in $L_2(\mathbb{R},w)$. Thus, for any $b \in H_{m,\theta}^{\perp} \subseteq L_2(\mathbb{R},w)$, there exists some sequence $(b_n)_n$ of essentially bounded $L_2(\mathbb{R},w)$-functions which converges to b. Let u_n and v_n denote the orthogonal projections of b_n onto $H_{m,\theta}$ and $H_{m,\theta}^{\perp}$, respectively. As

$$\|v_n - b\|_w^2 + \|u_n\|_w^2 = \|b_n - b\|_w^2,$$

the sequence $(v_n)_n$ in $H_{m,\theta}^{\perp}$ converges to b as well. Being a linear combination of $\beta_{k,m,\theta}$, $k = 0,\dots,m$, u_n is essentially bounded and so is $v_n = b_n - u_n$. Therefore the set of all essentially bounded elements of $H_{m,\theta}^{\perp}$ is dense in $H_{m,\theta}^{\perp}$.

Choosing $c > 0$ sufficiently small ensures nonnegativity of $\tilde{h}$. As $\langle \varphi_{0,m,\theta}, v \rangle_w = 0$ and $\varphi_{0,m,\theta} = \mathbf{1}$ according to Proposition 4.1, $\tilde{f} = \tilde{h}w$ is a density. As $\langle \beta_{k,m,\theta}, v \rangle_w = 0$ for all $k = 0, \ldots, m$, the parameters f and $\tilde{f}$ lead to identical data distribution, both in experiment $\mathcal{Y}_{n,m}$ and in $\mathcal{Z}_{n,m}$. $\qquad\square$

Proof of Theorem 4.4: Using the arguments from the proof of Theorem 2.1, which involve the Lemmata 4.3 and 5.1 from [22], we deduce that

$$\sup_{\theta \in \Theta} E_\theta \Big(\sum_{l=1}^m |\hat{\theta}_l - \theta_l| \Big)^2 = \mathcal{O}(m^4/n), \tag{6.29}$$

with universal constants which only depend on R, D_0, D_1 and D_2. Writing

$$\tilde{\sigma}_j^2 := c_W \cdot m^4 \cdot (\log n) \cdot n^{-1} \hat{\lambda}_{j,m,\hat{\theta}}^{-1},$$

we consider that

$$\big\| \hat{h}_{m,\hat{\theta}}^{\hat{\mathfrak{w}}} - h \big\|_w^2 - \sum_{j>m} \langle h, \varphi_{j,m,\hat{\theta}} \rangle_w^2 = \sum_{j=1}^m \big| 1\{\hat{h}_{j,m,\hat{\theta}}^2 \geq \tilde{\sigma}_j^2\} \cdot \hat{h}_{j,m,\hat{\theta}} - \langle h, \varphi_{j,m,\hat{\theta}} \rangle_w \big|^2$$

$$= \sum_{j=1}^m 1\{\hat{h}_{j,m,\hat{\theta}}^2 \geq \tilde{\sigma}_j^2\} \cdot \hat{D}_j^2 + 1\{\hat{h}_{j,m,\hat{\theta}}^2 < \tilde{\sigma}_j^2\} \cdot \langle h, \varphi_{j,m,\hat{\theta}} \rangle_w^2$$

$$\leq \sum_{j=1}^m 1\{\langle h, \varphi_{j,m,\hat{\theta}} \rangle_w^2 \geq \tilde{\sigma}_j^2/4\} \cdot \hat{D}_j^2 + 1\{\hat{D}_j^2 \geq \tilde{\sigma}_j^2/4\} \cdot \hat{D}_j^2$$

$$+ 1\{\langle h, \varphi_{j,m,\hat{\theta}} \rangle_w^2 \leq 4\tilde{\sigma}_j^2\} \cdot \langle h, \varphi_{j,m,\hat{\theta}} \rangle_w^2 + 1\{\hat{D}_j^2 \geq \tilde{\sigma}_j^2\} \cdot \langle h, \varphi_{j,m,\hat{\theta}} \rangle_w^2. \tag{6.30}$$

Conditionally on X^*, the random variables $\hat{D}_j := \hat{h}_{j,m,\hat{\theta}} - \langle h, \hat{\varphi}_{j,m,\hat{\theta}} \rangle_w$ are $\mathsf{N}(\hat{\mu}_j, \hat{\sigma}_j^2)$-distributed with

$$\hat{\mu}_j := \lambda_{j,m,\hat{\theta}}^{-1} \sum_{k=0}^m \big(\alpha_{k,m,\theta} \langle \beta_{k,m,\theta}, h \rangle_w - \alpha_{k,m,\hat{\theta}} \langle \beta_{k,m,\hat{\theta}}, h \rangle_w \big) \langle \beta_{k,m,\hat{\theta}}, \varphi_{j,m,\hat{\theta}} \rangle_w$$

$$\big/ \langle \beta_{k,m,\hat{\theta}}, \mathbf{1} \rangle_w,$$

$$\hat{\sigma}_j^2 := n^{-1} \lambda_{j,m,\hat{\theta}}^{-2} \sum_{k=0}^m \alpha_{k,m,\theta} \langle \beta_{k,m,\theta}, h \rangle_w \langle \beta_{k,m,\hat{\theta}}, \varphi_{j,m,\hat{\theta}} \rangle_w^2 / \langle \beta_{k,m,\hat{\theta}}, \mathbf{1} \rangle_w^2.$$

We define

$$r_{j,\theta}(x) := \exp\{x - \theta_j\}/(1 + \exp\{x - \theta_j\}),$$

$$R_{m,\theta,\hat{\theta}}(x) := \prod_{j=1}^m \big(1 + r_{j,\theta}(x) \cdot (\exp\{\theta_j - \hat{\theta}_j\} - 1) \big) - 1,$$

and deduce that

$$\beta_{k,m,\theta}(x) - \beta_{k,m,\hat{\theta}}(x) = \beta_{k,m,\hat{\theta}}(x) \cdot R_{m,\theta,\hat{\theta}}(x).$$

Moreover it holds that

$$\left| \alpha_{k,m,\hat{\theta}} - \alpha_{k,m,\theta} \right| \leq \alpha_{k,m,\hat{\theta}} \cdot \left(\exp\left\{ \sum_{l=1}^{m} |\hat{\theta}_l - \theta_l| \right\} - 1 \right),$$

$$\left\| R_{m,\theta,\hat{\theta}} \right\|_{\infty} \leq \exp\left\{ \exp\{R\} \sum_{l=1}^{m} |\hat{\theta}_l - \theta_l| \right\} - 1,$$

and hence that

$$\hat{\sigma}_j^2 \leq D_2 \cdot \exp\left\{ \left(1 + \exp\{R\} \right) \cdot \sum_{l=1}^{m} |\hat{\theta}_l - \theta_l| \right\} \cdot n^{-1} \lambda_{j,m,\hat{\theta}}^{-1},$$

almost surely. In order to bound the term $\hat{\mu}_j$, consider that, a.s.,

$$\left| \hat{\mu}_j \right| \leq D_2 \cdot \lambda_{j,m,\hat{\theta}}^{-1/2} \cdot \left(\exp\left\{ \sum_{l=1}^{m} |\hat{\theta}_l - \theta_l| \right\} - 1 \right) + \left| \langle R_{m,\theta,\hat{\theta}} \cdot h, \varphi_{j,m,\hat{\theta}} \rangle_w \right|.$$

Taking the conditional expectation given X^* in (6.30) yields the upper bound

$$\sum_{j=1}^{m} \mathbb{1}\{\langle h, \varphi_{j,m,\hat{\theta}} \rangle_w^2 \geq \tilde{\sigma}_j^2/4\} \cdot (\hat{\mu}_j^2 + \hat{\sigma}_j^2) + \mathbb{1}\{\langle h, \varphi_{j,m,\hat{\theta}} \rangle_w^2 \leq 4\tilde{\sigma}_j^2\} \cdot \langle h, \varphi_{j,m,\hat{\theta}} \rangle_w^2$$

$$+ \mathcal{O}\left(1 + \hat{\sigma}_j \cdot \left| \tilde{\sigma}_j/2 - |\hat{\mu}_j| \right| + \hat{\mu}_j^2 \right) \cdot \exp\left\{ - (\tilde{\sigma}_j/2 - |\hat{\mu}_j|)^2/(2\hat{\sigma}_j^2) \right\}.$$

The first line in the above expression is bounded from above by $\mathcal{M}_{n,m}$ multiplied by $(\log n) \, m^4$ and a constant; while the second line obeys the uniform upper bound of $o(1/n)$ and, hence, is negligible. Thanks to Lemma 4.2, the asymptotic behaviour of the eigenvalues $\lambda_{j,m,\hat{\theta}}$ equals that of the $\lambda_{j,m,\theta}$; concretely

$$\mathfrak{h}_{m,j}^2/(1 + \mathfrak{h}_{m,j}^2 n \lambda_{j,m,\hat{\theta}}) \leq \left(1 + \mathcal{O}_P(m^2 n^{-1/2}) \right) \cdot \mathfrak{h}_{m,j}^2/(1 + \mathfrak{h}_{m,j}^2 n \lambda_{j,m,\theta}),$$

again with universal constants. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Acknowledgements

References

[1] Biehler, M., Holling, H. and Doebler, P. (2015). Saddlepoint approximations of the distribution of the person parameter in the two parameter logistic model. *Psychometrika* **80**, 665–688. MR3392024

[2] Bock, R.D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **37**, 29–51. MR0678067

[3] Brown, C., Templin, J. and Cohen, A. (2015). Comparing the two- and three-parameter logistic models via likelihood ratio tests: A commonly misunderstood problem. *Appl. Psychol. Meas.* **39**, 335–348.

[4] Cavalier, L. and Hengartner, N.W. (2005). Adaptive estimation for inverse problems with noisy operators. *Inverse Problems* **21**, 1345–1361. MR2158113

[5] Chamberlain, G. (1984). "Panel data", in *Handbook of Econometrics, Vol. II*, ed. by Z. Griliches and M. Intriligator, Amsterdam: North-Holland, Chapter 22. MR0858544

[6] Chamberlain, G. (2010). Binary response models for panel data: identification and information. *Econometrica* **78**, 159–168. MR2642859

[7] de Leeuw, J. and Verhelst, N. (1986). Maximum likelihood estimation in generalized Rasch models. *J. Educ. Statist.* **11**, 183–196.

[8] Doebler, A., Doebler, P. and Holling, H. (2013). Optimal and most exact confidence intervals for person parameters in item response theory models. *Psychometrika* **78**, 98–115. MR3042821

[9] Efromovich, S. and Koltchinskii, V. (2001). On inverse problems with unknown operators. *IEEE Trans. Inform. Theo.* **47**, 2876–2894. MR1872847

[10] Fischer, G. H., *Einführung in die Theorie psychologischer Tests [Introduction to the theory of psychological tests]*, 1974, Bern: Huber, (in German).

[11] Fischer, G.H. and Molenaar, I.W., *Rasch Models: Foundations, Recent Developments and Applications*, 1995, Springer. MR1367341

[12] Follmann, D. (1988). Consistent estimation in the Rasch model based on nonparametric margins. *Psychometrika* **53**, 553–562. MR0997451

[13] Gustafsson, J.-E. (1980). Testing and obtaining fit of data to the Rasch model. *British J. Math. Statist. Psychology* **33**, 205–233.

[14] Haberman, S.J. (1977). Maximum likelihood estimates in exponential response models. *Ann. Statist.* **5**, 815–841. MR0501540

[15] Haberman, S.J. (2005). Identifiability of parameters in item response models with unconstrained ability distributions. *ETS Research Report Series*, 2005: i-22. doi:10.1002/j.2333-8504.2005.tb02001.x.

[16] Hall, P. and Horowitz, J.L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 70–91. MR2332269

[17] Hall, P. and Qiu, P. (2007). Nonparametric estimation of a point-spread function in multivariate problems. *Ann. Statist.* **35**, 1512–1534. MR2351095

[18] Hoderlein, S., Mammen, E., Yu, K. (2011). Non-parametric models in binary choice fixed effects panel data. *Econometrics Journal* **14**, 351–367. MR2853231

[19] Hoffmann, M. and Reiß, M. (2008). Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Statist.* **36**, 310–336. MR2387973

[20] Le Cam, L., *Asymptotic methods in statistical decision theory*, 1986, Springer. MR0856411

[21] Le Cam, L. and Yang, G.L., *Asymptotics in statistics: some basic concepts,*

2000, Springer. MR1784901

[22] Liese, F., Meister, A. and Kappus, J. (2019). Strong Gaussian approximation of the mixture Rasch model. *Bernoulli* **25**, 1326–1354. MR3920374

[23] Liese, F. and Miescke, K.J., *Statistical decision theory*, 2008, Springer. MR2421720

[24] Lindsay, B., Clogg, C.C. and Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *J. Amer. Statist. Assoc.* **86**, 96–107. MR1137102

[25] Magnac, T. (2004). Panel binary variables and sufficiency: generalizing conditional logit. *Econometrica* **72**, 1859–1876. MR2095535

[26] Meister, A. (2011). Asymptotic equivalence of functional linear regression and a white noise inverse problem. *Ann. Statist.* **39**, 1471–1495. MR2850209

[27] Pfanzagl, J. (1993). On the consistency of conditional maximum likelihood estimators. *Ann. Inst. Statist. Math.* **45**, 703–719. MR1252949

[28] Pfanzagl, J. (1994). On the identifiability of structural parameters in mixtures: applications to psychological tests. *J. Statist. Plann. Inf.* **38**, 309–326. MR1261806

[29] Rasch, G., *Probabilistic models for some intelligence and attainment tests (expanded edition)*, 1960/1980, Chicago: University of Chicago Press.

[30] Reiß, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations. *Ann. Statist.* **39**, 772–802. MR2816338

[31] Rice, K.M. (2004). Equivalence between conditional and mixture approaches to the Rasch model and matched case-control studies, with applications. *J. Amer. Statist. Assoc.* **99**, 510–522. MR2062836

[32] Rost, J. and von Davier, M. (1995). *Mixture distribution Rasch models*. In G. Fischer & I. Molenaar (Eds.), Rasch models: Foundations, recent developments, and applications (pp. 257–268). Springer. MR1367341

[33] Strasser, H. (2012a). The covariance structure of cml-estimates in the Rasch model. *Statistics & Risk Modeling* **29**, 315–326. MR2997038

[34] Strasser, H. (2012b). Asymptotic expansions for conditional moments of Bernoulli trials. *Statistics & Risk Modeling* **29**, 327–343. MR2997039

[35] Trabs, M. (2018). Bayesian inverse problems with unknown operators. *Inverse Problems* **34** (8). MR3817293

[36] von Davier, M. (2016). *The Rasch Model*. Chapter 3 in: van der Linden, W. (ed.) Handbook of Item Response Theory, Vol. 1. Second Edition. CRC Press, p. 31–48. MR3968058

[37] von Davier, M. and Carstensen, C.H., *Multivariate and Mixture Distribution Rasch Models – Extensions and Applications*, 2007, Statistics for Social and Behavioral Sciences, Springer, New York.

[38] Xu, X. and von Davier, M. (2008). Comparing multiple-group multinomial loglinear models for multidimensional skill distributions in the general diagnostic model. *ETS Research Report No. RR-08-35.*

[39] Xu, X. and Jia, Y. (2011). The sensitivity of parameter estimates to the latent ability distribution. *ETS Research Report No. RR-11-40.*