Minimax Risk Bounds in Stochastic Models with Smooth Parameter Function

Dissertation
zur Erlangung des Doktorgrades
der Fakultät für Mathematik, Informatik
und Naturwissenschaften
der Universität Hamburg

vorgelegt
im Department Mathematik
von

Uwe Christian Jönck
aus Cuxhaven

Hamburg
2008

Als Dissertation angenommen vom Department
Mathematik der Universität Hamburg

auf Grund der Gutachten von Prof. Dr. Holger Drees
und Prof. Dr. Natalie Neumeyer

Hamburg, den 02.10.2008

Prof. Dr. Reinhard Diestel
Leiter des Departments Mathematik

Dankeschön

Ich möchte Herrn Prof. Dr. Holger Drees ganz herzlich für die gute und angenehme Betreuung danken. Darüber hinaus möchte ich ihm sowie Frau Prof. Dr. Natalie Neumeyer auch meinen Dank für die tatkräftige Unterstützung meiner Bemühungen um ein Stipendium zur Finanzierung meiner Arbeit aussprechen.
Daneben gilt mein Dank auch meiner Familie, allen voran meinen Eltern, für ihre Unterstützung während der letzten drei Jahre.
Jag vill också tacka min flickvän, Helena Vramming, för hennes tålamod och stöd under den sista tiden. Och för att hon alltid kom med några uppmuntrande ord när det gick inte så bra med doktorarbetet. Tusen tack!

Hamburg, März 2008
Uwe Jönck

Contents

Notation

The following notation (which in many cases is standard in the literature) is used throughout the thesis. Further notation is also introduced in the text. An additional list of special symbols, containing the most important variables that occur in the text, is given in the end of the thesis, on p. 177.

$\sim$	distributed according to
$\simeq$	asymptotic equivalence of sequences, $a_n \simeq b_n \Leftrightarrow a_n/b_n \to 1$
$\subseteq$, $\subsetneq$	subset, possibly including equality, or excluding equality
$\equiv$	equivalent measures; sometimes also used in order to state that a function is constant
$\lesssim$	less or equal up to a positive constant, $a(s) \lesssim b(s)$ means that $a(s) \leq Cb(s)$ with some constant C not depending on the argument s
$\lhd\rhd$	mutually contiguous
$\ll$	absolutely continuous
$\stackrel{\mathscr{L}}{=}$	equality in distribution
$\rightsquigarrow$	convergence in distribution
$\xrightarrow{\mathrm{w}}$	weak convergence of statistical experiments
$\bigotimes$	product of (probability) measures or measurable spaces
$\#$	cardinality of a (finite) set
$\circ$	composition of maps
$\frac{d}{dx}f(x)$	derivative of a function $f\colon \mathbb{R} \to \mathbb{R}$

$\frac{\partial}{\partial x_i} f(x)$	partial derivative of a function $f\colon \mathbb{R}^d \to \mathbb{R}$ with respect to the ith component
∇f	transpose of the Jacobi matrix of a function $f\colon \mathbb{R}^p \to \mathbb{R}^q$; if $q = 1$, ∇f is the gradient
$\nabla^2 f$	Hessian matrix of a function $f\colon \mathbb{R}^d \to \mathbb{R}$
$\mathbb{1}$	unit matrix
$\mathbb{1}\{A\}$, $\mathbb{1}_A$	indicator of the set or statement A
$\overline{A}$	closure of a set A
∂A	boundary of a set A
$\mathbb{B}$, $\mathbb{B}^d$	Borel algebra on $\mathbb{R}$ and $\mathbb{R}^d$, respectively
$\operatorname{Cov}(X, Y)$, $\operatorname{Cov} Z$	covariance of a pair of real random variables X and Y and covariance matrix of a vector-valued random-variable Z, respectively; occasionally, we add an additional subscript, e.g. we write $\operatorname{Cov}_\theta(X, Y)$, to emphasize the underlying measure
$\operatorname{Corr}(X, Y)$	correlation of a pair of real random variables X and Y
δ_0	Dirac measure with mass 1 at the point zero
$\operatorname{diag}(w_1, \ldots, w_d)$	diagonal matrix with entries $w_1, \ldots, w_d$ on the main diagonal
$\mathrm{E}\, X$	expectation of a random variable X; sometimes we also use a more specific notation, e.g. $\mathrm{E}_\theta\, X$ or $\mathrm{E}_P\, X$, in order to emphasize the underlying measure
e_i	ith unit vector in $\mathbb{R}^d$
$H(\cdot, \cdot)$	Hellinger distance
λ	Lebesgue measure on $(\mathbb{R}, \mathbb{B})$
$\mathscr{L}(X \mid P)$	law of a random variable X under some probability measure P
$\mathbb{L}_p(\mu)$	space of p-fold μ-integrable functions
$\mathscr{N}(\mu, \nu^2)$	normal distribution with mean (or mean vector) μ and variance (or covariance matrix) ν^2
$\mathbb{N}$	natural numbers excluding zero
O, o	Landau symbols
O_P, o_P	stochastic Landau symbols
φ, Φ	pdf and cdf, respectively, of the standard normal distribution

$\mathbb{R}$	real numbers
$\mathbb{R}^d$	d-dimensional Euclidean space
$\mathbb{R}^{m\times n}$	space of real $m \times n$-matrices
$\operatorname{sgn}(x)$	sign of the real number, $\operatorname{sgn}(x) := \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$
$A^\top$	transpose of a matrix (or a vector) A
$\operatorname{Var} X$	variance of a real random variable X
x^+	maximum of x and 0
$\\|\cdot\\|$	Euclidean norm
$\\|\cdot\\|_1$	1-norm on $\mathbb{R}^d$, $\\|x\\|_1 := \sum_{i=1}^d \lvert x_i\rvert$
$\vert\vert\vert\cdot\vert\vert\vert$	matrix norm induced by the Euclidean norm
$\\|\cdot\\|_{\mathrm{F}}$	Frobenius norm
$\\|\cdot\\|_\infty$	maximum norm for vectors, $\\|x\\|_\infty := \max\{\lvert x_1\rvert, \dots, \lvert x_d\rvert\}$
$\\|\cdot\\|_u$	uniform norm (supremum norm) for real functions
$\\|\cdot\\|_{u,d}$	uniform norm for d-variate functions, see Section 3.3
$\\|\cdot\\|_{\mu,p}$	p-norm on the space $\mathbb{L}_p(\mu)$
$\\|\cdot\\|_{\mathrm{V}}$	variational norm for probability measures
$\langle\cdot,\cdot\rangle_\mu$	inner product on the Banach space $\mathbb{L}_2(\mu)$
$\langle\cdot,\cdot\rangle$	inner product in Euclidean space
$[x, y]$	closed cuboid in $\mathbb{R}^d$, $[x, y] := [x_1, y_1] \times \ldots \times [x_d, y_d]$
a.e.	almost everywhere
a.s.	almost surely
cdf	cumulative distribution function
i.i.d.	independently and identically distributed
pdf	probability density function
$\blacksquare$	end of a proof

CHAPTER 1

Introduction

In statistical estimation theory and applications one is often confronted with the question of *how good* a certain estimation procedure is in comparison with others. Of course, this "goodness" has to be quantified somehow. One possibility to do so is to use concepts from statistical decision theory and consider upper and lower bounds for the minimax risk. Such an approach is examined by Drees (2001) for the estimation of the extreme value index. We want to follow the ideas of that paper in order to derive minimax risk bounds for the estimation of a smooth parameter function.

1.1 The problem

We consider a stochastic process $\{Y_t : t \in [0,1]\}$ with mutually independent random variables Y_t taking values in a measurable space $(\mathcal{Y}, \mathcal{B})$. The arguments $t \in [0,1]$ are mostly interpreted as time points. We assume that $Y_t \sim P_{\xi(t)}$ for each t. Here $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ (with an open subset $\Theta \subseteq \mathbb{R}^d$) is a parametric family of probability measures and $\xi\colon [0,1] \to \Theta$ is a smooth parameter function from a suitable function space $\mathscr{F}$. Such a modelling approach is e.g. used by Drees and Stărică (2002) and Jönck (2008) for the description and prediction of time series of log-returns. In case of $\mathcal{P}$ being a location family we obtain a nonparametric regression model. It is needless to say that this model plays an important role for many kinds of applications.

We assume that for given time points $0 \leq x_{n1} < \cdots < x_{nn} \leq 1$

the corresponding random variables $Y_{nj} := Y_{x_{nj}}$ can be observed. The joint distribution of these Y_{nj} is then given by the product measure

$$P_\xi^{(n)} := \bigotimes_{j=1}^{n} P_{\xi(x_{nj})}. \tag{1.1}$$

Let $x_0 \in (0,1)$ be a fixed time point and

$$\hat{\xi}_n(x_0) = \hat{\xi}_n(x_0, Y_{n1}, \ldots, Y_{nn})$$

an estimator for $\xi(x_0)$, the true distribution parameter of the process at x_0. In order to evaluate such an estimator we consider the loss function $\mathbb{1}_{[-\delta a_n, \delta a_n]^C}$, where $\delta = (\delta_1, \ldots, \delta_d)^\top$ is a vector with positive components δ_i and $a_n > 0$. Here $[-\delta a_n, \delta a_n] := [-\delta_1 a_n, \delta_1 a_n] \times \ldots \times [-\delta_d a_n, \delta_d a_n]$ denotes a cuboid in $\mathbb{R}^d$; we use this notation throughout the paper. The *minimax risk* with respect to this loss function is then given by

$$\inf_{\hat{\xi}_n(x_0)} \sup_{\xi \in \mathscr{F}} P_\xi^{(n)} \big\{ \hat{\xi}_n(x_0) - \xi(x_0) \in [-\delta a_n, \delta a_n]^C \big\}, \tag{1.2}$$

where the infimum is taken over all sequences of (possibly randomised) estimators $\hat{\xi}_n(x_0)$. The aim of this thesis is to analyse the *asymptotic minimax risk*, i.e. the behaviour of (1.2) for $n \to \infty$. Since an exact analysis of this expression is most often not possible, we have to content ourselves with the attempt of deriving upper and lower bounds. More precisely, we want to derive upper and lower bounds for the limit inferior and the limit superior of the minimax risk, respectively.

Localising the model. The basic idea for our examinations is to first localise the model and to consider the (asymptotic) minimax risk in the resulting local models. From the results within these local models one can then also deduce minimax risk bounds in the *global model*

$$(\mathcal{Y}^n, \mathcal{B}^n, \{P_\xi^{(n)} : \xi \in \mathscr{F}\}).$$

For the localisation we consider a constant function $\gamma \equiv \gamma \mathbb{1}_{[0,1]} \in \mathscr{F}$—also called *centre of localisation*—and we assume that the true function $\xi \in \mathscr{F}$ does not vary significantly from γ around the point

x_0. Concretely, we assume that in a sufficiently small neighbourhood of x_0 the function ξ can be represented by

$$\gamma_{g,n}(x) := \gamma + a_n g\left(b_n(x - x_0)\right), \quad x \in [0,1], \tag{1.3}$$

where $g\colon \mathbb{R} \to \mathbb{R}^d$ is a function with compact support from a certain function space $\mathscr{G}$ and $a_n \to 0$, $b_n \to \infty$. The function $\gamma_{g,n}(\cdot)$ from (1.3) is often also referred to as a *local alternative* to γ. Indeed, the above properties of a_n, b_n and g guarantee that $\gamma_{g,n}(\cdot)$ only differs from γ in a neighbourhood of x_0. In the first instance, we examine the sequence of the resulting *local models*

$$(\mathcal{Y}^n, \mathcal{B}^n, \{P^{(n)}_{\gamma,g} : g \in \mathscr{G}\}),$$

where, in analogy to (1.1),

$$P^{(n)}_{\gamma,g} := \bigotimes_{j=1}^{n} P_{\gamma_{g,n}(x_{nj})} \tag{1.4}$$

In the local model the parameter value $\xi(x_0)$ equals $\gamma_{g,n}(x_0)$. Since

$$\gamma_{g,n}(x_0) = \gamma + a_n g(0)$$

it suffices to consider estimates for $g(0)$. Let now $\hat{g}_n(0)$ be an arbitrary estimator for $g(0)$. Then

$$\hat{\gamma}_{g,n}(x_0) := \gamma + a_n \hat{g}_n(0)$$

yields an estimator for $\gamma_{g,n}(x_0)$, which satisfies

$$(\hat{\gamma}_{g,n}(x_0) - \gamma_{g,n}(x_0))/a_n = \hat{g}_n(0) - g(0).$$

The benefit of first considering the local models is that these can be shown to converge against a Gaussian shift model. These limit experiments can be examined fairly well. It is possible to derive non-trivial minimax risk bounds. We will show that under certain conditions these bounds carry over to the estimation problem in the local model and finally also in the global model $(\mathcal{Y}^n, \mathcal{B}^n, \{P^{(n)}_{\xi} : \xi \in \mathscr{F}\})$.

1.2 Thesis outline

Chapter 2 gives a short overview of some concepts of asymptotic statistics, especially of the Le Cam theory and statistical decision theory. Chapter 3 formulates in detail the model which is to examine and its underlying assumptions. In Chapter 4 we show that the above local experiments converge weakly to a Gaussian experiment, and in Chapter 5 upper and lower bounds for the minimax risk in this limit experiment are derived. The general ideas for the examinations of these two chapters follow those of Drees (2001). In Chapter 6 we then construct an estimator for the parameter $g(0)$ in the local model. The construction principle for this estimator is based upon the proven weak convergence and the results in the Gaussian experiment. We show that—provided some smoothness conditions hold—the asymptotic maximal risk of this estimator converges to the upper minimax risk bound from the Gaussian limit experiment, which thus yields an upper bound for the asymptotic minimax risk in the sequence of the local models. Indeed, these risk bounds carry over to the global model, which is proven in Chapter 7. In Chapter 8 we discuss some concrete models to which our theory applies, such as the nonparametric regression model. On p. 169 we summarise the results of this thesis.

1.3 Related research

There are some other papers in the literature that discuss the minimax risk in models with independent observations, distributed according to some parametric family, with the parameter being driven by a smooth function. The focus of these papers is clearly on nonparametric regression models, and thus more specific than the approach followed here. Brown and Low (1996) show that the classical nonparametric regression model with Gaussian noise is under certain conditions asymptotically equivalent—with respect to the Le Cam deficiency pseudo-distance—to a white noise model with drift. Therefore, the asymptotic risks in both models must be the same. Nussbaum (1996), Grama and Nussbaum (1998), Jähnisch and Nussbaum (2003) as well as Rohde (2004) and Reiß (2008) expand these results. Cai and Low (2004), Donoho (1994) and Donoho and Liu (1991) also investigate the minimax risk for the estimation of functionals in the white noise model.

Apart from this work, Grama and Nussbaum (2002) consider a setup, which is a bit closer to our model from Section 1.1. In their paper the distributions are given by one-parametric exponential families. They are able to show that their model is asymptotically equivalent (again, with respect to the Le Cam deficiency pseudo-distance) to a Gaussian location model. However, their approach to the problem is in some major aspects different from that followed in this thesis. Moreover, they do—as well as the other papers mentioned above—not assess the asymptotic minimax risk directly.

Drees (2001) considers the classical i.i.d. setup. He derives asymptotic minimax risk bounds for the estimation of the extreme value index. Even though Drees' paper deals with a completely different topic, it can be considered a role model for this thesis. Many (yet, not all) techniques used there also apply in the proofs of this thesis.

CHAPTER 2

Preliminaries

This chapter gives an overview of some basic concepts from (asymptotic) statistics and statistical decision theory, which are used throughout the thesis. A list of some general notation is given at the beginning of the thesis.

On the used asymptotics. Throughout this work n denotes an index tending to infinity. Usually, n represents the number of observations. If not stated otherwise (or if it is clear from the context) all asymptotic expressions, e.g. convergence statements, presume that $n \to \infty$.

2.1 Basics of stochastics

We denote the space of p-fold integrable functions on a measurable space $(\Omega, \mathcal{A}, \mu)$ by $\mathbb{L}_p(\mu) = \mathbb{L}_p(\Omega, \mathcal{A}, \mu)$. For $p \geq 1$ this is—as long as we identify equivalence classes of μ-a.e. equivalent functions with their representatives—a Banach space with respect to the norm $\|f\|_{\mu,p} := (\int |f|^p \; d\mu)^{1/p}$. For $p = 2$ the norm is induced by the inner product $\langle f, g\rangle_\mu := \int fg \; d\mu$, whereby $\mathbb{L}_2(\mu)$ becomes a Hilbert space. The Lebesgue measure on $(\mathbb{R}, \mathbb{B})$ is denoted λ. For vector-valued functions $f = (f_1, \ldots, f_d)^\top$ we write $f \in \mathbb{L}_p(\mu)$ if and only if $f_i \in \mathbb{L}_p(\mu)$ for all i. The integral $\int f \; d\mu$ is then to be interpreted component-wise, i.e. as a vector $(\int f_1 \; d\mu, \ldots, \int f_d \; d\mu)^\top$. Integrals of matrix-valued functions are to be interpreted likewise. Let P and Q be probability measures on a space $(\Omega, \mathcal{A})$, dominated by a σ-finite measure μ (e.g. $\mu = P + Q$). The Hellinger distance is then defined by $H(P,Q) := (1/2 \int (\sqrt{p} - \sqrt{q})^2 \; d\mu)^{1/2}$, where $p = dP/d\mu$ and $q = dQ/d\mu$. The variational distance is given by $\|P - Q\|_{\mathrm{V}} = \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$. Note the inequality $\|P - Q\|_{\mathrm{V}} \leq \sqrt{2} H(P,Q)$.

Convergence in distribution (or *weak convergence*) of probability measures or random variables is denoted by the symbol "$\rightsquigarrow$". We also use the convenient stochastic Landau notation often used in the literature. Hence, for a sequence of random variables

$$X_n \colon (\Omega_n, \mathcal{A}_n, \; P_n) \to (\mathbb{R}^k, \mathbb{B}^k)$$

and a real sequence c_n we write $X_n = o_{P_n}(c_n)$ if $X_n/c_n \to 0$ in P_n-probability. We write $X_n = O_{P_n}(c_n)$ if X_n/c_n is stochastically bounded, i.e. if for each $\varepsilon > 0$ there is an $M > 0$ such that $P_n\{\|X_n/c_n\| > M\} < \varepsilon$ for all n.

2.2 Contiguity and differentiability in quadratic mean

In the following, we assume that the reader is familiar to the concept of the likelihood ratio dQ/dP for two probability measures defined on some measurable space $(\Omega, \mathcal{A})$ (cf. van der Vaart, 1998, Section 6.1).

2.1 Definition. Let P_n and Q_n be sequences of probability measures on measurable spaces $(\Omega_n, \mathcal{A}_n)$. These are called *(mutually) contiguous* iff for each sequence of measurable sets $A_n \in \mathcal{A}_n$ holds: $P_n(A_n) \to 0 \Leftrightarrow Q_n(A_n) \to 0$. In that case we also write $Q_n \lhd\rhd P_n$.

The contiguity $Q_n \lhd\rhd P_n$ states that the measures P_n and Q_n can be considered asymptotically mutually continuous to each other. "$\lhd\rhd$" defines an equivalence relation on the space of the sequences of probability measures on $(\Omega_n, \mathcal{A}_n)$. However, $Q_n \equiv P_n$ for each n does not necessarily imply $P_n \lhd\rhd Q_n$. Under certain conditions, asymptotic normality of log-likelihood ratios implies contiguity:

2.2 Lemma (Le Cam's first lemma). *Let P_n and Q_n be probability measures on $(\Omega_n, \mathcal{A}_n)$ such that*

$$\mathscr{L}\left(\log \frac{dQ_n}{dP_n} \,\middle|\, P_n\right) \rightsquigarrow \mathscr{N}(-\kappa^2/2, \kappa^2).$$

Then P_n and Q_n are contiguous, $P_n \lhd\rhd Q_n$, and

$$\mathscr{L}\left(\log \frac{dQ_n}{dP_n} \,\middle|\, Q_n\right) \rightsquigarrow \mathscr{N}(\kappa^2/2, \kappa^2).$$

A proof of the lemma is given in Witting and Müller-Funk (1995), p. 311. Assume we are given a certain statistic, which is observed under P_n, and we are interested in its asymptotic behaviour under Q_n. If $P_n \lhd\rhd Q_n$, then it is sometimes possible to deduce the statistic's behaviour under Q_n from that under P_n. For such a result the reader might confer to Theorem 6.6 in van der Vaart (1998). Here we want to consider a consequence of that theorem, which covers the special case of asymptotically normal statistics (cf. van der Vaart, 1998, Example 6.7).

2.3 Lemma (Le Cam's third lemma). *Let P_n and Q_n be probability measures on $(\Omega_n, \mathcal{A}_n)$, and $X_n\colon (\Omega_n, \mathcal{A}_n) \to (\mathbb{R}^d, \mathbb{B}^d)$ be random variables such that*

$$\mathscr{L}\left(X_n, \log \frac{dQ_n}{dP_n} \,\middle|\, P_n\right) \rightsquigarrow \mathscr{N}\left(\begin{pmatrix} \mu \\ -\frac{1}{2}\sigma^2 \end{pmatrix}, \begin{pmatrix} \Sigma & \nu \\ \nu^\top & \sigma^2 \end{pmatrix}\right).$$

Then $\mathscr{L}(X_n \mid Q_n) \rightsquigarrow \mathscr{N}(\mu + \nu, \Sigma)$.

In the following, let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a family of probability measures on some measurable space. The parameter set Θ is assumed an open subset $\Theta \subseteq \mathbb{R}^d$. Furthermore, we presume $\mathcal{P}$ to be dominated by a σ-finite measure μ, such that we have pdf's $p_\theta = dP_\theta/d\mu$.

2.4 Definition. The family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is *differentiable in quadratic mean* at some point $\theta \in \Theta$, if a measurable function $\dot{\ell}_\theta = (\dot{\ell}_{\theta,1}, \ldots, \dot{\ell}_{\theta,d})^\top$ exists such that

$$\int \left[\sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \tfrac{1}{2} h^\top \dot{\ell}_\theta \sqrt{p_\theta}\right]^2 \, d\mu = o(\|h\|^2), \quad h \to 0. \tag{2.1}$$

The function $\dot{\ell}_\theta$ is then also called the $\mathbb{L}_2$-*derivative* at θ or *score function*. We say that $\mathcal{P}$ is differentiable in quadratic mean on a set $\Theta' \subseteq \Theta$ if this is also true at each point $\theta \in \Theta'$.

Due to the following lemma (which is part of Theorem 7.2 in van der Vaart, 1998, and proved there) the score function has mean zero and existing covariance matrix.

2.5 Lemma. *If $\mathcal{P}$ is differentiable in quadratic mean at θ, then $\int \dot{\ell}_\theta \, dP_\theta = 0$ and the Fisher information matrix $\mathcal{I}_\theta = \int \dot{\ell}_\theta \dot{\ell}_\theta^\top \, dP_\theta$ exists.*

Condition (2.1) states that the map $\tau \mapsto \sqrt{p_\tau}$, interpreted as a map with values in the Banach space $\mathbb{L}_2(\mu)$, is *Fréchet differentiable* at θ (cf. Heuser, 1990, Number 175). The Fréchet derivative—which is not to be mistaken for the $\mathbb{L}_2$-derivative—is then given by $\dot{\ell}_\theta \sqrt{p_\theta}$. According to this we will say that $\mathcal{P}$ is continuously differentiable in quadratic mean, if the map $\tau \mapsto \sqrt{p_\tau}$ is continuously Fréchet differentiable.

2.6 Definition. If $\mathcal{P}$ is differentiable in quadratic mean on an open subset $\Theta' \subseteq \Theta$, and if the $\mathbb{L}_2(\mu)$-valued maps

$$\theta \mapsto \dot{\ell}_{\theta,i} \sqrt{p_\theta}, \quad i = 1, \ldots, d,$$

are continuous on Θ', then $\mathcal{P}$ is called *continuously differentiable in quadratic mean* on Θ'.

2.7 Lemma. *Suppose $\mathcal{P}$ is continuously differentiable in quadratic mean on Θ. Then the following statements hold:*

(i) The information matrix function $\theta \mapsto \mathcal{I}_\theta = \int \dot{\ell}_\theta \dot{\ell}_\theta^\top \, dP_\theta$ is continuous on Θ (i.e. continuous in each component).

(ii) For each $\theta \in \Theta$ there is an open neighbourhood U_θ of θ and a constant $L_\theta > 0$ such that for all $\tau_1, \tau_2 \in U_\theta$ the inequality $H(P_{\tau_1}, P_{\tau_2}) \leq L_\theta \|\tau_1 - \tau_2\|$ holds.

Proof. *Ad (i):* Let $\theta \in \Theta$ and $\theta_n \to \theta$ be given. Then $\dot{\ell}_{\theta_n,i} \sqrt{p_{\theta_n}} \to \dot{\ell}_{\theta,i} \sqrt{p_\theta}$ in $\mathbb{L}_2(\mu)$ for all $i = 1, \ldots, d$. According to Satz 15.9 in Bauer (1992) the products $\dot{\ell}_{\theta_n,i} \dot{\ell}_{\theta_n,j} p_{\theta_n}$ converge in mean with respect to μ, with limits given by $\dot{\ell}_{\theta,i} \dot{\ell}_{\theta,j} p_\theta$. In other words,

$$\|\dot{\ell}_{\theta_n,i} \dot{\ell}_{\theta_n,j} p_{\theta_n} - \dot{\ell}_{\theta,i} \dot{\ell}_{\theta,j} p_\theta\|_{\mu,1} \to 0,$$

and Satz 15.1 in Bauer (1992) implies

$$\begin{aligned} \int \dot{\ell}_{\theta_n,i} \dot{\ell}_{\theta_n,j} \, dP_{\theta_n} &= \int \dot{\ell}_{\theta_n,i} \dot{\ell}_{\theta_n,j} p_{\theta_n} \, d\mu \\ &\to \int \dot{\ell}_{\theta,i} \dot{\ell}_{\theta,j} p_\theta \, d\mu = \int \dot{\ell}_{\theta,i} \dot{\ell}_{\theta,j} \, dP_\theta. \end{aligned}$$

Ad (ii): By assumption the map $\tau \mapsto \sqrt{p_\tau}$ is continuously differentiable on Θ, and thus in particular locally Lipschitz, which follows from the mean value theorem (cf. Heuser, 1990, p. 337). Since $H(P_{\tau_1}, P_{\tau_2}) = (1/\sqrt{2})\|\sqrt{p_{\tau_1}} - \sqrt{p_{\tau_2}}\|_{\mu,2}$ this implies the assertion. ∎

Differentiability in quadratic mean is a suitable framework for proving *local asymptotic normality (LAN)*. In the most simple case of an i.i.d. setting this means that for every sequence $h_n \to h$ the likelihood quotients fulfil

$$\log \frac{dP^n_{\theta+n^{-1/2}h_n}}{dP^n_\theta} = h^\top \Delta_{n,\theta} - \tfrac{1}{2} h^\top I_\theta h + o_{P^n_\theta}(1),$$

with a matrix I_θ and random variables $\Delta_{n,\theta}$ such that

$$\mathscr{L}(\Delta_{n,\theta} \mid P^n_\theta) \rightsquigarrow \mathscr{N}(0, I_\theta).$$

Asymptotic expansions of this type are important for proving weak convergence of experiments, which shall be introduced in the next section. The following theorem provides general conditions under which the log-likelihood ratios of product measures allow for such asymptotic expansions. It is taken from van der Vaart (1988), but can in similar form also be found in Rieder (1994), Theorem 2.3.5.

2.8 Proposition. *For $n = 1, 2, \ldots$ and $j = 1, \ldots, n$ let P_{nj} and Q_{nj} be probability measures on measurable spaces $(\Omega_{nj}, \mathcal{A}_{nj})$ with pdf's p_{nj} and q_{nj} with respect to a σ-finite measure μ_{nj} satisfying $P_{nj} + Q_{nj} \ll \mu_{nj}$. Let $P_n := \bigotimes_{j=1}^n P_{nj}$ and $Q_n := \bigotimes_{j=1}^n Q_{nj}$ denote the product measures on the corresponding product spaces. Suppose that there are measurable functions $U_{nj} \colon (\Omega_{nj}, \mathcal{A}_{nj}) \to (\mathbb{R}, \mathbb{B})$ such that:*

$$\sum_{j=1}^n \int \Big[q_{nj}^{1/2} - p_{nj}^{1/2} - \tfrac{1}{2} U_{nj} p_{nj}^{1/2} \Big]^2 \, d\mu_{nj} = o(1), \tag{2.2}$$

$$\sum_{j=1}^n \int U_{nj} \, dP_{nj} = o(1), \tag{2.3}$$

$$\kappa_n^2 := \sum_{j=1}^n \int U_{nj}^2 \, dP_{nj} = O(1), \tag{2.4}$$

$$\sum_{j=1}^{n} \int U_{nj}^2 \mathbb{1}\{|U_{nj}| \geq \varepsilon\}\, dP_{nj} = o(1) \quad \textit{for all } \varepsilon > 0. \tag{2.5}$$

Then

$$\log \frac{dQ_n}{dP_n} - \sum_{j=1}^{n} U_{nj} + \tfrac{1}{2}\kappa_n^2 = o_{P_n}(1). \tag{2.6}$$

Addendum. *If in addition to the above conditions also*

$$\kappa_n^2 \to \kappa^2 \in [0, \infty), \tag{2.7}$$

$$\sum_{j=1}^{n} \left(\int U_{nj}\, dP_{nj} \right)^2 \to 0, \tag{2.8}$$

then P_n and Q_n are mutually contiguous, and

$$\mathscr{L}\left(\log \frac{dQ_n}{dP_n} \,\middle|\, P_n \right) \rightsquigarrow \mathscr{N}(-\tfrac{1}{2}\kappa^2, \kappa^2).$$

Here $\mathscr{N}(0,0)$ is identified with δ_0, the Dirac measure with probability 1 at the origin.

2.9 Remark. In triangular schemes U_{nj} of the above type we typically interpret each U_{nj} as a random variable on the product space $\bigotimes_{j=1}^n (\Omega_{nj}, \mathcal{A}_{nj}, \mu_{nj})$. This can be done by identifying U_{nj} and the corresponding variable $U_{nj}(\pi_{nj})$, where

$$\pi_{nj} \colon (\Omega_{n1} \times \ldots \times \Omega_{nn}) \to \Omega_{nj}, \quad (\omega_{n1}, \ldots, \omega_{nn}) \mapsto \omega_{nj}$$

denotes the jth canonical projection. For the sake of simplicity the projections will however mostly be suppressed in the notation, e.g. as is the case in (2.6).

Proof of Proposition 2.8. The first part of the proposition corresponds to Proposition A.8 in van der Vaart (1988) if the functions $g_{nj}/\sqrt{n}$ are replaced by the U_{nj}. It remains to prove the addendum. The Lindeberg-Feller theorem (see van der Vaart, 1998, Proposition 2.27) yields $\mathscr{L}(\sum_{j=1}^n U_{nj} \mid P_n) \rightsquigarrow \mathscr{N}(0, \kappa^2)$. Together with (2.6) and Slutsky's lemma this proves the asymptotic normality of the log-likelihood quotients. Contiguity follows from Le Cam's first lemma. ■

2.3 Statistical decision theory

Let $T \neq \emptyset$ be an arbitrary set. A triple $(\Omega, \mathcal{A}, \mathcal{Q})$, consisting of a space Ω, a σ-algebra $\mathcal{A}$ and a family $\mathcal{Q} = \{Q_t : t \in T\}$ of probability measures on $(\Omega, \mathcal{A})$ is called *statistical experiment* or *statistic model* with parameter space T. We often simply call $\mathcal{Q}$ a statistical experiment. An experiment is *dominated*, if the corresponding class $\mathcal{Q}$ is dominated by a σ-finite measure. Suppose we are given a space D endowed with some σ-algebra $\mathcal{D}$. We also call D a *decision space* and interpret it as the set of all possible statements which shall be assessed by means of the observation of the experiment E (cf. Witting, 1985, p. 11). In the context of a testing problem for a simple hypothesis H_0 a suitable decision space would be

$$D = \{\text{“H}_0\text{ is accepted”}, \text{“H}_0\text{ is rejected”}\},$$

with $\mathcal{D}$ being the power set of D. In the context of estimation problems D should be chosen as the space which contains the parameter of interest. In the following, D is always assumed to be a metric space, where $\mathcal{D}$ is the Borel algebra induced by the metric.

2.10 Definition. Let an experiment $E = (\Omega, \mathcal{A}, \mathcal{Q})$ and a decision space $(D, \mathcal{D})$ be given. A Markov kernel ϱ from $(\Omega, \mathcal{A})$ to $(D, \mathcal{D})$ (which is sometimes also written as a map $\varrho\colon \Omega \times \mathcal{D} \to [0, 1]$) is called *decision function*. The set of all such decision functions is denoted $\mathcal{R}(E, D)$. If $\varrho(\cdot, B) \in \{0, 1\}$ Q_t-a.s. for every $t \in T$ and all $B \in \mathcal{D}$, then ϱ is called *non-randomised*.

Suppose we have a decision function ϱ. Then the quantity $\varrho(\omega, B)$ can be interpreted as the probability with which we choose the event B given the fact that the outcome of the observed experiment is ω. In case of a non-randomised ϱ this means that we choose such a B which satisfies $\varrho(\omega, B) = 1$.

2.11 Definition. A *loss function* $W = \{W_t : t \in T\}$ is a family of functions $W_t\colon D \to [0, \infty)$ that are measurable with respect to $\mathcal{D}$. The corresponding *risk function* $R\colon T \times \mathcal{R}(E, D) \to [0, \infty)$ is defined as

$$R(t, \varrho) := R_W(t, \varrho) := \iint W_t(x)\varrho(\omega, dx)Q_t(d\omega),$$

$R(t, \varrho)$ is called the *risk* of ϱ with respect to W. The triple (T,D,W) is called a *decision problem.* We also say that a loss function W possesses a certain property (like continuity, boundedness etc.) iff each W_t possesses this property.

2.12 Example. When a $\mathbb{R}^k$-valued parameter $\gamma = \gamma(t)$ $(t \in T)$ is to be estimated, one often considers a measurable function $\Lambda\colon \mathbb{R}^k \to [0,\infty)$ and then defines a loss function $W_t(a) := \Lambda(a - \gamma(t))$. Often, the function Λ itself is called loss function. Typical examples of such loss functions are the quadratic loss $\Lambda(a) := \|a\|^2$ or a zero-one loss functions of the form $\Lambda(a) := \mathbb{1}\{a \in [-x, x]^C\}$.

The quantity $R(t, \varrho)$ can be interpreted as the loss that we suffer if the underlying distribution is given by Q_t and we decide according to the decision rule ϱ. Thus, it is evident to consider a decision rule to be *good* if the maximal loss that can result from it becomes minimal. This leads to the following concepts for evaluating decision procedures:

2.13 Definition. Let (T, D, W) be a decision problem. For $\varrho \in \mathcal{R}(E, D)$ we then call $\sup_{t\in T} R(t, \varrho)$ the *maximal risk* of ϱ. The *minimax risk* of the decision problem is defined by

$$\inf_{\varrho\in\mathcal{R}(E,D)} \sup_{t\in T} R(t, \varrho).$$

A decision function $\varrho^* \in \mathcal{R}(E, D)$ for which this infimum is attained is called *minimax optimal.*

The aim of this thesis is to investigate the asymptotic behaviour of the minimax risk in a sequence of models. For that purpose we need a concept for convergence of sequences of statistical experiments.

2.14 Definition. Let a non-empty parameter set T and statistical experiments $E_n = (\Omega, \mathcal{A}, \{Q_{t,n} : t \in T\})$ and $E = (\Omega, \mathcal{A}, \{Q_t : t \in T\})$ be given. Then E_n is said to *converge weakly* to E if

$$\mathscr{L}\left(\left\{\frac{dQ_{t,n}}{dQ_{s,n}}\right\}_{t\in T_0} \,\middle|\, Q_{s,n}\right) \rightsquigarrow \mathscr{L}\left(\left\{\frac{dQ_t}{dQ_s}\right\}_{t\in T_0} \,\middle|\, Q_s\right)$$

for all finite subsets $T_0 \subseteq T$ and all $s \in T$. In that case we also write $E_n \xrightarrow{\mathrm{w}} E$.

2.15 Remark. (i) Sometimes one is not interested in the complete experiments E and E_n, but rather in certain subexperiments E' and E'_n indexed by a non-empty subset $T' \subseteq T$. Of course, $E_n \xrightarrow{\mathrm{w}} E$ also implies $E'_n \xrightarrow{\mathrm{w}} E'$.

(ii) Convergence in distribution is also referred to as "weak convergence". However, this should not cause any confusion, because convergence in distribution and weak convergence of experiments are two convergence concepts for different objects. Thus, it will always be clear from the context which kind of "weak convergence" is meant.

The idea behind the concept of weak convergence of experiments is that for large n the *limit experiment* E can be considered an approximation for the true model E_n. From the properties of this limit experiment one might then draw conclusions for the experiments E_n. The following theorem—which can be found in Strasser (1985), Theorem 61.6—yields a sufficient criterion for proving weak convergence.

2.16 Definition. A sequence of experiments E_n is said to be contiguous iff $Q_{s,n} \lhd\rhd Q_{t,n}$ for all $s, t \in T$.

2.17 Theorem. *Let E_n be a contiguous sequence of experiments. Then $E_n \xrightarrow{\mathrm{w}} E$ holds if and only if*

$$\mathscr{L}\left(\left\{\frac{dQ_{t,n}}{dQ_{t_0,n}}\right\}_{t\in T_0} \,\middle|\, Q_{t_0,n}\right) \rightsquigarrow \mathscr{L}\left(\left\{\frac{dQ_t}{dQ_{t_0}}\right\}_{t\in T_0} \,\middle|\, Q_{t_0}\right)$$

for all finite subsets $T_0 \subseteq T$ and some $t_0 \in T$.

Consequently, one may verify $E_n \xrightarrow{\mathrm{w}} E$ by showing that E_n is contiguous and that certain likelihood ratios converge in distribution. For that purpose Theorem 2.8 may be used. Weak convergence of experiments allows us to draw conclusions for the relation between the minimax risk in a sequence of experiments and the minimax risk in a limit experiment. Indeed, if $E_n \xrightarrow{\mathrm{w}} E$, then the minimax risk in the limit experiment is a lower bound for the asymptotic minimax risk in the sequence of experiments E_n.

2.18 Definition. Let D be a metric space. A function $f\colon D \to \mathbb{R}$ is called *level-compact* if $\{x : f(x) \leq y\}$ is compact for all $y <$

$\sup\{f(x) : x \in D\}$. f is called *lower semi-continuous* if it is bounded from below and $\{x : f(x) \leq y\}$ is closed for all $y \in \mathbb{R}$.

2.19 Theorem. *Let E_n and E be dominated experiments with $E_n \xrightarrow{\text{w}} E$. Moreover, let (T, D, W) be a decision problem with level-compact, lower semi-continuous loss function W, and let D be a separable metric space which is locally compact. Then for each sequence $\varrho_n \in \mathcal{R}(E_n, D)$:*

$$\liminf_{n\to\infty} \sup_{t\in T} R(t, \varrho_n) \geq \inf_{\varrho\in\mathcal{R}(E,D)} \sup_{t\in T} R(t, \varrho).$$

Proof. Follows from Corollary 62.6 and Theorem 43.5 in Strasser (1985). ■

2.4 Uniform convergence in distribution

We introduce the concept of *uniform convergence in distribution* (or: *uniform weak convergence*) of probability measures on Euclidean space and discuss some connected results. First, we discuss a criterion for the weak convergence of measures on $(\mathbb{R}^d, \mathbb{B}^d)$. In the following, $\mathfrak{C}^d \subseteq \mathbb{B}^d$ denotes the class of convex Borel sets, ∂C denotes the boundary of a set C.

2.20 Theorem. *Let Q_n and Q be probability measures on $(\mathbb{R}^d, \mathbb{B}^d)$. Suppose that $Q(\partial C) = 0$ for all $C \in \mathfrak{C}^d$. Then: $Q_n \rightsquigarrow Q \Leftrightarrow \sup_{C\in\mathfrak{C}^d} |Q_n(C) - Q(C)| \to 0$.*

The theorem is due to Rao (1962), cf. Theorem 4.2; an alternative proof can be found in Fabian (1970). Throughout the rest of this section we consider statistical experiments $(\mathbb{R}^d, \mathbb{B}^d, \{Q_{s,n} : s \in S\})$ and $(\mathbb{R}^d, \mathbb{B}^d, \{Q_s : s \in S\})$. Here S is always assumed to be a metric space, the Borel algebra induced by the metric is denoted $\mathcal{S}$.

2.21 Definition. The sequence $Q_{s,n}$ is said to converge *uniformly in distribution* to Q_s on a subset $S' \subseteq S$ (we also formulate this as "$Q_{s,n} \rightsquigarrow Q_s$ uniformly on S'") if

$$\sup_{s\in S'} \left| \int f \, dQ_{s,n} - \int f \, dQ_s \right| \to 0$$

for all bounded, continuous functions $f\colon \mathbb{R}^d \to \mathbb{R}$.

If $Q_{s,n} \rightsquigarrow Q_s$ uniformly on S', then of course also $Q_{s,n} \rightsquigarrow Q_s$ uniformly on every subset $S'' \subseteq S'$. In particular, this holds true for all subsets S'' consisting of a single point. The following, equivalent characterisation of uniform convergence in distribution generalises the result from Theorem 2.20.

2.22 Theorem. *Le S be a metric space which is locally compact. Furthermore, assume that the map $s \mapsto Q_s$ is continuous in the sense that $Q_{s_m} \rightsquigarrow Q_s$ as $s_m \to s$. Suppose that $Q_s(\partial C) = 0$ for all $C \in \mathfrak{C}^d$ and all $s \in S$. Then the following assertions are equivalent:*

(i) $Q_{s,n} \rightsquigarrow Q_s$ *uniformly on all compact sets* $K \subseteq S$.

(ii) $\sup\limits_{s \in K} \sup\limits_{C \in \mathfrak{C}^d} |Q_{s,n}(C) - Q_s(C)| \to 0$ *for all compact sets* $K \subseteq S$.

Proof. Every metric space is a Hausdorff space and satisfies the first axiom of countability (cf. Lipschutz, 1977, p. 131). Thus, taking into account Remark 6.7.7 in Pfanzagl (1994), the assertion is a direct consequence of Pfanzagl's Theorem 7.7.10. ■

We recall Slutsky's famous lemma for sequences X_n and Y_n of vector-valued random variables. The lemma states (roughly formulated) that if $X_n \rightsquigarrow X$ and $Y_n \to c$ in probability, where c is a constant, then also $X_n + Y_n \rightsquigarrow X + c$. It turns out that this result also holds true in the generalised context of uniform convergence.

2.23 Definition. Let statistical experiments $(\Omega_n, \mathcal{A}_n, \{P_{s,n} : s \in S\})$ be given, and a sequence of measurable maps

$$f_n \colon (\Omega_n \times S, \mathcal{A}_n \otimes \mathcal{S}) \to (\mathbb{R}^d, \mathbb{B}^d).$$

Then f_n is said to converge *uniformly in probability* to zero with respect to $\{P_{s,n} : s \in S\}$ if

$$\sup_{s \in S} P_{s,n}\{\omega_n \in \Omega_n : \|f_n(\omega_n, s)\| > \varepsilon\} \to 0$$

for all $\varepsilon > 0$.

2.24 Lemma (Slutsky's lemma, generalised). *Let T_n and f_n be measurable maps (in the sense of Definition 2.23), and $Q_{s,n} := \mathscr{L}(T_n(\cdot, s) \mid P_{s,n})$. Let $K \subseteq S$ be a compact subset such that $Q_{s,n} \rightsquigarrow Q_s$ uniformly on K. Furthermore, assume that f_n converges uniformly in probability to zero with respect to $\{P_{s,n} : s \in K\}$. Then $\mathscr{L}(T_n(\cdot, s) + f_n(\cdot, s) \mid P_{s,n}) \rightsquigarrow Q_s$, uniformly on K.*

A proof of this result can be found in Pfanzagl (1994), Lemma 7.7.8.

2.5 Remarks and related literature

The concepts introduced and discussed in this chapter provide only a small insight into the theory of asymptotic statistics and statistical decision theory. Of course, not all connections to other fields from stochastics could be pointed out, which was not the aim of this chapter, either. However, the given survey is sufficient for our purposes. The above examinations were mainly based on the nice monograph by van der Vaart (1998), and on Strasser (1985) and Pfanzagl (1994). Detailed and profound information on the discussed topics can also be found in Witting (1985), Witting and Müller-Funk (1995) and Rieder (1994) as well as in the monographs Le Cam (1986) and Le Cam and Yang (2000), partly also in Bickel et al. (1998).

CHAPTER 3

Model assumptions

In this chapter we specify the model for the process $\{Y_t : t \in [0,1]\}$ from Chapter 1. We formulate exact conditions which the underlying family of distributions and the other quantities involved in the definition of the global and local models are presumed to fulfil. We make use of the notation already introduced in Section 1.1.

3.1 General distributional assumptions

In the following, let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ denote a family of probability measures on some measurable space $(\mathcal{Y}, \mathcal{B})$, with an open parameter space $\Theta \subseteq \mathbb{R}^d$. We assume $\mathcal{P}$ to be dominated by a σ-finite measure μ. Thus, we have densities $p_\theta = dP_\theta/d\mu$. Furthermore, for $\theta \in \mathbb{R}^d \setminus \Theta$ we set $P_\theta := P_{\theta^*}$, with some arbitrarily chosen $\theta^* \in \Theta$.

We presume that $\mathcal{P}$ is *continuously differentiable in quadratic mean* on Θ (in the sense of Definition 2.6), the score function is denoted $\dot{\ell}_\theta$. Then the information matrices $\mathcal{I}_\theta = \int \dot{\ell}_\theta \dot{\ell}_\theta^\top \, dP_\theta$ are defined for all $\theta \in \Theta$ and continuous, which follows from Lemma 2.7. In addition, we assume that $\mathcal{I}_\theta$ is *positive definite* for all $\theta \in \Theta$.

Throughout the rest of this thesis we assume that the above distributional assumptions do hold.

At this stage, the reader may recall the Cholesky decomposition of a positive definite matrix. According to this the symmetric, positive definite matrix $\mathcal{I}_\theta$ can be written as a product $\mathcal{I}_\theta = \mathcal{C}_\theta^\top \mathcal{C}_\theta$ with a unique, invertible upper triangle matrix $\mathcal{C}_\theta \in \mathbb{R}^{d\times d}$. The entries of $\mathcal{C}_\theta$ are linear combinations of that of $\mathcal{I}_\theta$ (cf. Opfer, 2001, p. 175 f.). Consequently, if $\mathcal{I}_\theta$ is continuous in θ, then also $\mathcal{C}_\gamma$, $\mathcal{C}_\theta^{-1}$ and $\|\mathcal{C}_\theta^j\|_1$ ($j = 1, \ldots, d$), with $\mathcal{C}_\theta^j$ denoting the jth row vector of $\mathcal{C}_\theta$. Note that the Cholesky decomposition satisfies $\mathcal{C}_\theta^{-1}(\mathcal{C}_\theta^{-1})^\top = \mathcal{I}_\theta^{-1}$. For further information on the Cholesky decomposition cf. Stoer (1999), p. 207 f.

3.2 Further regularity conditions

Besides the conditions from the previous section, we need to impose some further regularity conditions on the distribution family $\mathcal{P}$. These additional assumptions are essential for the proofs of Chapters 6 and 7.

We assume that there is a set $N \subseteq \mathcal{Y}$ with $\mu(N) = 0$ and $p_\theta(y) > 0$ for all $y \in N^C$ and all $\theta \in \Theta$. In particular, this implies that $P_\theta \equiv \nu$ for all θ. Furthermore, we assume that for all $y \in N^C$ and all $\theta \in \Theta$ the log-density $\log p_\theta(y)$ is three times continuously partially differentiable with respect to the parameter θ, and that the score function satisfies

$$\dot{\ell}_\theta(y) = \frac{\nabla_\theta p_\theta(y)}{p_\theta(y)} \quad \mu\text{-a.e.} \tag{3.1}$$

As a consequence, the score function permits a Taylor expansion in the parameter, which yields

$$\dot{\ell}_{\theta+h}(y) = \dot{\ell}_\theta(y) + \ddot{\ell}_\theta(y)h + O(\|h\|^2), \quad h \to 0, \tag{3.2}$$

where here and in the following

$$\ddot{\ell}_\theta := \nabla_\theta^2 \log p_\theta.$$

(The remainder in (3.2) of course depends on θ and y, which for the moment shall not concern us.) We further require that the third partial derivatives of the log-densities are dominated in some sense. More precisely, we assume that there is a measurable function $J\colon \mathcal{Y} \to \mathbb{R}$ such that

$$\left| \frac{\partial^3}{\partial\theta_i \partial\theta_j \partial\theta_k} \log p_\theta(y) \right| \le J(y) \quad \text{for all } \theta \in \Theta \text{ and all } y, \qquad i, j, k = 1, \ldots, d, \tag{3.3}$$

and we assume that $\mathrm{E}_\theta J$ is uniformly bounded on Θ.

Besides these "point-wise" assumptions, we have to impose some *smoothness conditions* concerning the moments of the score function and its derivatives. We assume that the Fisher information satisfies

$$\mathcal{I}_\theta = \int \dot{\ell}_\theta \dot{\ell}_\theta^\top \, dP_\theta = -\int \ddot{\ell}_\theta \, dP_\theta \tag{3.4}$$

for all $\theta \in \Theta$ and that the functions

$$(\theta, \tau) \mapsto \int (\dot{\ell}_{\theta,i})^{k_1} (\dot{\ell}_{\theta,j})^{k_2} \, dP_\tau, \quad k_1, k_2 \in \{0, 1, 2\}$$

are continuous on $\Theta \times \Theta$ for $k_1, k_2 \in \{0, 1, 2\}$, continuously partially differentiable for $k_1 = 2$ and $k_2 = 0$, and two times continuously partially differentiable for $k_1 = 1$ and $k_2 = 0$. Likewise we presume the maps

$$(\theta, \tau) \mapsto \int (\ddot{\ell}_{\theta,rs})^k \, dP_\tau, \quad k \in \{0, 1, 2\},$$

to be continuously partially differentiable on $\Theta \times \Theta$, where $\ddot{\ell}_{\theta,rs}$ denotes the rs-entry of the matrix $\ddot{\ell}_\theta$. This means in particular that the entries of the information matrix $\mathcal{I}_\theta$ are continuously partially differentiable with respect to θ. Finally, we assume that certain interchangeability properties for integration and differentiation hold true, namely

$$\nabla_\tau \int \dot{\ell}_\theta p_\tau \, d\mu = \int \dot{\ell}_\theta (\nabla_\tau p_\tau)^\top \, d\mu. \tag{3.5}$$

Many of the above assumptions reflect classical regularity conditions used to prove consistency or asymptotic normality of certain estimators, such as maximum likelihood or other likelihood-based estimators. See for example Lehmann and Casella (1998), Sections 6.5 and 6.6, or Aerts and Claeskens (1997). Although the above conditions may appear somewhat restrictive, they are satisfied by a large class of distribution families, e.g. by exponential families (see Chapter 8).

3.3 The global and local function spaces

In Section 1.1 the one-dimensional marginals of the process $\{Y_t : t \in [0, 1]\}$ were modelled according to the approach $Y_t \sim P_{\xi(t)}$, with an unknown parameter function $\mathscr{F} \ni \xi\colon [0,1] \to \Theta$. In the following, we assume this *global function space* $\mathscr{F}$ to be given by

$$\begin{aligned}\mathscr{F} &:= \mathscr{F}(x_0, \Theta^*, u)\\ &:= \big\{\xi\colon [0,1] \to \Theta \mid \xi(x_0) \in \Theta^*,\ \xi \text{ continuous},\\ &\qquad \|\xi(x_0+s) - \xi(x_0)\|_\infty \le u(s),\ s \in [-x_0, 1-x_0]\big\}.\end{aligned} \tag{3.6}$$

Here $\|\cdot\|_\infty$ denotes the maximum norm on $\mathbb{R}^d$; $\Theta^* = \Theta^*(\gamma^*, \varepsilon^*)$ denotes an open cuboid centred around some fixed point $\gamma^* \in \Theta$, more precisely:

$$\Theta^* := (\gamma_1^* - \varepsilon_1^*, \gamma_1^* + \varepsilon_1^*) \times \ldots \times (\gamma_d^* - \varepsilon_d^*, \gamma_d^* + \varepsilon_d^*), \quad \varepsilon_i^* > 0.$$

(More generally, one could also choose Θ^* as an open, bounded neighbourhood of γ^*.) The additional function u indicates the degree of how much the ξ's may vary in a neighbourhood of x_0 relative to $\xi(x_0)$. In the following, we will always assume that u is of the specific form

$$u\colon \mathbb{R} \to [0,\infty), \quad s \mapsto A|s|^\rho,$$

with $A, \rho > 0$. For large values ρ and small values A, respectively, the last condition from (3.6)—in the following often simply referred to as *growth condition*—becomes more restrictive and thus the model, too. To shorten notation we often omit the arguments and simply write $\mathscr{F}$ instead of $\mathscr{F}(x_0, \Theta^*, u)$. From the definition of $\mathscr{F}$ and the above choice of u, one easily concludes that

$$\Theta' := \{\xi(x) \mid x \in [0,1], \xi \in \mathscr{F}\} \subseteq \Theta \tag{3.7}$$

(note that of course $\Theta^* \subseteq \Theta'$, too) is bounded and thus has compact closure $\overline{\Theta'}$. We further presume this set to be included in Θ, in other words, we assume that Θ' is a relatively compact subset of Θ,

$$\overline{\Theta'} \subseteq \Theta. \tag{3.8}$$

For some examinations it will be necessary to further restrict $\mathscr{F}$ to such functions which fulfil a certain growth condition on the whole interval $[0,1]$ as well as at x_0. To this end, we set

$$\begin{aligned}\mathscr{F}_{\mathrm{H}} &:= \mathscr{F}_{\mathrm{H}}(x_0, \Theta^*, u)\\ &:= \big\{\xi \in \mathscr{F} \mid \|\xi(x) - \xi(y)\|_\infty \le u(x-y),\ x, y \in [0,1]\big\}.\end{aligned} \tag{3.9}$$

The restriction on the variation of the parameter functions is in the literature sometimes referred to as *Hölder condition* (or *Hölder continuity*). For $\rho = 1$ this is Lipschitz continuity.

The *time points* x_{nj} at which the model—i.e. the random variables $Y_{nj} = Y_{x_{nj}}$—can be observed, and for which we consider the parameter function ξ, are supposed to be given by equidistant points on the unit interval: we assume that

$$x_{nj} := \frac{j-1}{n-1}, \quad n \in \mathbb{N}, \quad j = 1, \dots, n.$$

The resulting product measures $P_\xi^{(n)}$ are then defined according to (1.1).

The scope of this thesis is to investigate the asymptotic behaviour of estimates for $\xi(x_0)$ under a sequence of loss functions $\mathbb{1}_{[-a_n\delta, a_n\delta]^C}$. Throughout this thesis

$$\delta = (\delta_1, \dots, \delta_d)^\top$$

denotes a fixed vector with entries $\delta_i > 0$. (The sequence a_n is defined below.)

In order to assess the asymptotic minimax risk with respect to this loss function we choose a localisation procedure, as mentioned in Section 1.1. For that purpose we consider the following *local function space* $\mathscr{G}$, which is defined by

$$\begin{aligned} \mathscr{G} &:= \mathscr{G}(K, \rho, d) \\ &:= \big\{ g \colon \mathbb{R} \to \mathbb{R}^d \mid \|g(s) - g(0)\|_\infty \le |s|^\rho,\ s \in \mathbb{R}, \\ &\qquad\qquad g|_{[-K,K]} \text{ is continuous}, \\ &\qquad\qquad g(x) = 0 \text{ for } x \notin [-K, K] \big\} \end{aligned} \tag{3.10}$$

Note that the growth condition in (3.10) holds for all $s \in \mathbb{R}$. The elements $g \in \mathscr{G}$ are also referred to as *local functions.* The exact value of the constant K is not of so great importance. It merely has to be chosen sufficiently large, such that for certain parameters β the maps $s \mapsto (\beta - |s|^\rho)^+$ are included in $\mathscr{G}(K, \rho, 1)$. These maps are essential for the construction of suitable estimators. In Section 3.5 we will state more precisely what "sufficiently large" means. Throughout the thesis we assume that K is chosen according to the criteria presented in that section. Note that

$$\mathscr{G}(K, \rho, d) = \mathscr{G}(K, \rho, 1) \times \dots \times \mathscr{G}(K, \rho, 1) \quad (d \text{ times}).$$

For (bounded) functions $f\colon \mathbb{R} \to \mathbb{R}^d$ we define the uniform norm by

$$\|f\|_{u,d} := \max\{\|f_1\|_u, \dots, \|f_d\|_u\},$$

where $\|f_i\|_u := \sup\{|f_i(x)| : x \in \mathbb{R}\}$ is the usual uniform norm (supremum norm) for real functions. The growth condition in the definition of $\mathscr{G}$ implies that

$$M := \sup\{\|g\|_{u,d} : g \in \mathscr{G}\} < \infty. \tag{3.11}$$

In the same way that the space $\mathscr{F}$ was restricted to those functions which satisfy a Hölder condition, we also restrict $\mathscr{G}$, setting

$$\begin{aligned} \mathscr{G}_{\mathrm{H}} &:= \mathscr{G}_{\mathrm{H}}(K, \rho, d) \\ &:= \big\{g \in \mathscr{G}(K, \rho, d) \mid \|g(x) - g(y)\|_\infty \le |x - y|^\rho, \\ &\qquad\qquad x, y \in [-K, K]\big\}. \end{aligned} \tag{3.12}$$

Moreover, we set

$$\mathscr{G}_{\mathrm{H}}^* := \{g \in \mathscr{G}_{\mathrm{H}} : g(0) = 0\}. \tag{3.13}$$

The relations between the spaces $\mathscr{F}$ and $\mathscr{G}$, and $\mathscr{F}_{\mathrm{H}}$ and $\mathscr{G}_{\mathrm{H}}$, respectively, will become clear in later examinations. The limitations (like the growth conditions and the boundedness of the functions) are mainly necessary in order to take advantage of some compactness arguments, as shall be seen later, too.

For $g \in \mathscr{G}$ and $\gamma \in \Theta$ the local alternatives $\gamma_{g,n}(\cdot)$ with respect to γ may now be defined according to (1.3), where a_n and b_n in the following always are given by

$$\begin{aligned} b_n &:= A^{2/(2\rho+1)} n^{1/(2\rho+1)}, \\ a_n &:= u(1/b_n) = A^{1/(2\rho+1)} n^{-\rho/(2\rho+1)}. \end{aligned} \tag{3.14}$$

As a consequence of this definition we have

$$a_n^2 = \frac{b_n}{n}, \quad n \in \mathbb{N}. \tag{3.15}$$

To shorten notation we further define the *transformed time points* by

$$\tilde{x}_{nj} := b_n(x_{nj} - x_0), \quad j = 1, \dots, n. \tag{3.16}$$

With this notation, the local alternative from (1.3) can also be written as

$$\begin{aligned}\gamma_{g,n}(x_{nj}) &= \gamma + a_n g(b_n(x_{nj} - x_0)) \\ &= \gamma + a_n g(\tilde{x}_{nj}).\end{aligned} \tag{3.17}$$

The product measure $P^{(n)}_{\gamma,g}$ corresponding to this local alternative $\gamma_{g,n}$ may be defined according to (1.4). Moreover, for $g \equiv 0$ we use the convention $P^{(n)}_{\gamma} := P^{(n)}_{\gamma,0}$. (In that case, $P^{(n)}_{\gamma}$ is of course equal to the product of n independent copies of P_γ, i.e. $P^{(n)}_{\gamma} = P^{n}_{\gamma}$.) For some choices of γ and g it might turn out that $\gamma_{g,n}(x_{nj}) \notin \Theta$. This can of course only happen for small indices n. However, by the convention $P_\theta = P_{\theta^*}$ for all $\theta \notin \Theta$ (see Section 3.1) we then have $P_{\gamma_{g,n}(x_{nj})} = P_{\theta^*}$. Hence, the product measures $P^{(n)}_{\gamma,g}$ are always well-defined.

3.4 Existence of a preliminary estimator

Finally, we presume that there is a *uniformly* $(1/a_n)$*-consistent* estimator $\xi^*_n := \xi^*_n(x_0)$ for the true parameter $\xi(x_0)$. This means that for each $\iota > 0$ there is a constant $C = C_\iota > 0$ such that

$$\sup_{\xi \in \mathscr{F}} P^{(n)}_{\xi}\left\{\frac{1}{a_n}\|\xi^*_n(x_0) - \xi(x_0)\| \geq C\right\} < \iota$$

for sufficiently large n. Of course, it is equivalent to require that

$$\limsup_{n\to\infty} \sup_{\xi \in \mathscr{F}} P^{(n)}_{\xi}\left\{\frac{1}{a_n}\|\xi^*_n(x_0) - \xi(x_0)\| \geq C\right\} < \iota.$$

The question of the existence of such preliminary estimators will be discussed Chapter 8 for some models.

3.5 Supplement: specifying K

To completely describe the local function space $\mathscr{G} = \mathscr{G}(K, \rho, d)$, we still have to specify the constant K from the definition of the space $\mathscr{G}(K, \rho, d)$. As was stated before, the *exact* value of this constant is not so important, it is only important that it is sufficiently large to allow for certain local functions to be included in $\mathscr{G}$. In this section—which can be omitted in the first reading—we describe how K can be chosen in order to guarantee that this is the case. The approach may appear somewhat complicated, but it will become clear in the course of Chapter 5 why it precisely satisfies our needs.

Note that in the definition of $\mathscr{G}$, K is the only parameter that is not yet defined through the global function space $\mathscr{F}$. Hence, it is quite clear that K in some way has to depend on the vector δ that defines the loss function. In order to find an answer how large K must at least be, we first consider the simple case of a one-dimensional parameter space $\Theta \subseteq \mathbb{R}$, and hence the function space $\mathscr{G} = \mathscr{G}(K, \rho, 1)$. Subsequently, we consider the general case $\Theta \subseteq \mathbb{R}^d$.

3.5.1 The univariate parameter case

In the one-dimensional case the vector δ that specifies the loss function $\mathbb{1}_{[-\delta,\delta]^C}$ is a real number, $\delta > 0$. We consider the map

$$f\colon (\delta, \infty) \to [0, \infty), \quad \beta \mapsto \frac{\beta^{1+1/\rho}}{\log[(\beta+\delta)/(\beta-\delta)]}.$$

Since the function $\beta \mapsto \log[(\beta+\delta)/(\beta-\delta)]$ is strictly decreasing on (δ, ∞), it is clear that f is strictly increasing. Furthermore, $f(\beta)$ tends to 0 as $\beta \downarrow \delta$ and to ∞ as $\beta \to \infty$. Since the distribution family $\mathcal{P}$ is assumed continuously differentiable in quadratic mean the map $\gamma \mapsto \mathcal{I}_\gamma$ is continuous on Θ, and by assumption it is strictly positive (see Section 3.1). We further assumed in (3.8) that $\overline{\Theta'} \subseteq \Theta$. This implies that there is a compact interval $I \subseteq (0, \infty)$ such that

$$\left\{ \frac{\rho+1}{2\mathcal{I}_\gamma \delta} : \gamma \in \Theta' \right\} \subseteq I.$$

Due to the monotonicity and continuity properties, $f^{-1}(I)$ is also a compact interval, say $[s, t]$, which is included in $(0, \infty)$. We now choose the constant K such that the condition $t^{1/\rho} < K$ is satisfied.

As was stated at the beginning, $f(\beta)$ tends to zero as $\beta \downarrow \delta$, and to infinity as $\beta \to \infty$. Hence, it is clear that for each $\gamma \in \Theta$ the equation

$$\frac{\beta^{1+1/\rho}}{\log[(\beta+\delta)/(\beta-\delta)]} = \frac{\rho+1}{2\mathcal{I}_\gamma \delta}$$

has a unique solution $\beta^* = \beta^*(\gamma)$. If K is chosen according to the above criterion, then we have $\beta^*(\gamma) < K^\rho$ for all $\gamma \in \Theta'$. This assures that each of the corresponding functions $s \mapsto (\beta^*(\gamma) - |s|^\rho)^+$, $\gamma \in \Theta'$ is included in $\mathscr{G}(K, \rho, 1)$.

3.5.2 The multivariate parameter case

We generalise the above procedure in order to derive a criterion for a suitable choice of K for $\mathscr{G} = \mathscr{G}(K, \rho, d)$, with $d \geq 1$. To this end, we first generalise the above function f, setting

$$f\colon (0,\infty)^2 \to [0,\infty), \quad (x,\beta) \mapsto \frac{\beta^{1+1/\rho}}{\log[(\beta+x)/(\beta-x)]}\mathbb{1}\{\beta > x\}.$$

Of course, f has similar properties as the corresponding map from Section 3.5.1. In particular, for fixed x the map $\beta \mapsto f(x,\beta)$ is strictly increasing on (x,∞), and for fixed β, $x \mapsto f(x,\beta)$ is strictly decreasing on $(0,\beta)$. Furthermore, f is continuously differentiable at all points (x,β) for which $f(x,\beta) \neq 0$. In the following, let a family of strictly positive, continuous functions

$$v_j\colon \Theta \to (0,\infty), \quad j = 1,\dots,d,$$

be given. A suitable choice of such functions, which also satisfy certain additional conditions, will be described in detail in Section 3.5.3, and we will in the following always assume that the v_j's are of the specific form described there. As in the one-dimensional case, we exploit that $\mathcal{I}_\gamma$ is positive definite for each $\gamma \in \Theta$ and that the map $\gamma \mapsto \mathcal{I}_\gamma$ is continuous. In particular, each component of $\mathcal{I}_\gamma$ is continuous in γ. The same holds true for the Cholesky decomposition $\mathcal{C}_\gamma$ and for the maps $\gamma \mapsto \|\mathcal{C}_\gamma^j\|_1$ $(j = 1,\dots,d)$, with $\mathcal{C}_\gamma^j$ denoting the jth row vector of $\mathcal{C}_\gamma$ (cf. Section 3.1). Since the Cholesky matrices are invertible, each of the maps $\gamma \mapsto \|\mathcal{C}_\gamma^j\|_1$ is strictly positive on Θ, and the same holds true for the maps $\gamma \mapsto v_j(\gamma)/\|\mathcal{C}_\gamma^j\|_1$. From $\overline{\Theta'} \subseteq \Theta$ (by assumption (3.8)) and an additional continuity argument we conclude that there are compact intervals $I_1, I_2 \subseteq (0,\infty)$ such that

$$\bigcup_{j=1}^{d}\left\{\frac{\rho+1}{2v_j(\gamma)\|\mathcal{C}_\gamma^j\|_1} : \gamma \in \Theta'\right\} \subseteq I_1 \quad \text{and} \quad \bigcup_{j=1}^{d}\left\{\frac{v_j(\gamma)}{\|\mathcal{C}_\gamma^j\|_1} : \gamma \in \Theta'\right\} \subseteq I_2.$$

For each fixed $x \in I_2$ the map $\beta \mapsto f(x,\beta)$ is continuous and strictly increasing on (x,∞), with

$$\lim_{\beta \downarrow x} f(x,\beta) = 0 \quad \text{and} \quad \lim_{\beta \uparrow \infty} f(x,\beta) = \infty.$$

Hence, the set $J_x := \{\beta : f(x,\beta) \in I_1\}$ is a compact interval, say $J_x = [a_x, b_x] \subseteq (x, \infty)$. If we write $I_1 =: [a,b]$, then, by the monotonicity, we have $f(x, b_x) = b > 0$ for all x. An application of the implicit function theorem yields that the map $x \mapsto b_x$ is continuous, and thus in particular bounded on I_2. In other words, $t := \sup_{x\in I_2} b_x$ is finite. We now choose the constant K such that $t^{1/\rho} < K$ holds.

The consequences of this special choice of K are similar to those from the one-dimensional case. Let an arbitrary parameter $\gamma \in \Theta'$ be given. Then $t_j := v_j(\gamma)/\|\mathcal{C}_\gamma^j\|_1 \in I_2$ and $(\rho+1)/(2v_j(\gamma)\|\mathcal{C}_\gamma^j\|_1) \in I_1$ for all $j = 1, \dots, d$. From the monotonicity properties of the functions $\beta \mapsto f(t_j, \beta)$ we conclude that each of the equations

$$f(t_j, \beta) = \frac{\beta^{1+1/\rho}}{\log \frac{\beta + v_j(\gamma)/\|\mathcal{C}_\gamma^j\|_1}{\beta - v_j(\gamma)/\|\mathcal{C}_\gamma^j\|_1}} \stackrel{!}{=} \frac{\rho+1}{2v_j(\gamma)\|\mathcal{C}_\gamma^j\|_1}, \quad j = 1, \dots, d,$$

possesses a unique solution $\beta_j^* = \beta_j^*(\gamma) \in J_{t_j}$, i.e. $\beta_j^* \le t$. These solutions then satisfy $\beta_j^* < K^\rho$ for all j. In particular, for all $\gamma \in \Theta'$ the functions

$$s \mapsto \big((\beta_1^*(\gamma) - |s|^\rho)^+, \dots, (\beta_d^*(\gamma) - |s|^\rho)^+\big)^\top$$

are included in $\mathscr{G}(K, \rho, d)$.

3.5.3 Specifying v

For a precise specification of the constant K in the general case $d \ge 1$ we still have to define the function

$$v = v(\gamma) = (v_1(\gamma), \dots, v_d(\gamma))^\top.$$

Our approach is to define v in such a way that it satisfies

$$[-v(\gamma), v(\gamma)] \subseteq A_\gamma,$$

where A_γ is defined by

$$A_\gamma := \{\mathcal{C}_\gamma x : x \in [-\delta, \delta]\}.$$

In the course of the examinations in Section 5.3 it will become clear why this is a useful condition. Let us first consider an arbitrary $\gamma \in \Theta$, and let $\mathcal{C}_\gamma$ be the corresponding Cholesky matrix. In order to determine an appropriate vector $v = v(\gamma)$ such that $[-v, v] \subseteq A_\gamma$ holds, we denote the 2^d corners of the cuboid $[-\delta, \delta]$ by

$$\delta^{(i)} = (\delta_1^{(i)}, \ldots, \delta_d^{(i)})^\top, \quad \text{where} \quad |\delta_j^{(i)}| = \delta_j \quad \text{for } j = 1, \ldots, d.$$

Furthermore, we set[1)]

$$w^{(i)} := \mathcal{C}_\gamma \delta^{(i)}. \tag{3.18}$$

These points $w^{(i)}$ are the corners of the parallelogram A_γ. Setting

$$w := (w_1, \ldots, w_d)^\top, \quad \text{with} \quad w_i := \max_j |w_i^{(j)}| \tag{3.19}$$

the set $[-w, w]$ is the smallest cuboid which includes A_γ as a subset. In general, of course we have $A_\gamma \subsetneq [-w, w]$. We determine now v as a scalar multiple of w in such a way that the resulting "shrunk" cuboid $[-v, v]$ becomes a subset of A_γ. To this end, denote the 2^d corners of $[-w, w]$ by the vectors

$$\widetilde{w}^{(i)} = (\widetilde{w}_1^{(i)}, \ldots, \widetilde{w}_d^{(i)})^\top, \quad \text{where} \quad |\widetilde{w}_j^{(i)}| = w_j \quad \text{for } j = 1, \ldots, d. \tag{3.20}$$

Setting $z^{(i)} := \mathcal{C}_\gamma^{-1} \widetilde{w}^{(i)}$ and

$$\alpha := \max \bigl\{ |z_j^{(i)}| / \delta_j : i = 1, \ldots, 2^d,\ j = 1, \ldots, d \bigr\}, \tag{3.21}$$

we then define the vector v by

$$v := \frac{1}{\alpha} w.$$

The corners of the cuboid $[-v, v]$ are given by the 2^d vectors $v^{(i)} := \widetilde{w}^{(i)}/\alpha = \mathcal{C}_\gamma z^{(i)}/\alpha$, and because of $z^{(i)}/\alpha \in [-\delta, \delta]$ we have $v^{(i)} \in A_\gamma$. Hence, we achieve $[-v, v] \subseteq A_\gamma$. Note that in the special case of a diagonal matrix $\mathcal{C}_\gamma = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$ (and thus in particular for the case $d = 1$) the above procedure yields $v_j = \lambda_j \delta_j$, and thus $A_\gamma = [-v, v]$.

The Cholesky matrix $\mathcal{C}_\gamma$ is an invertible upper triangle matrix. Therefore, all components of the vector w are strictly positive, and with the same argument we conclude that $\alpha > 0$, too. As a consequence, all the components $v_j(\gamma)$ of the vector $v(\gamma)$ are strictly positive—as required in Section 3.5.2. It remains to prove that $v(\cdot)$ is continuous on Θ.

[1)] Note that all the quantities $w^{(i)}$, w, $\widetilde{w}^{(i)}$, $z^{(i)}$ and α, which are defined in the following are functions of the parameter γ. However, to keep the notation simple this dependence on γ is often suppressed.

3.1 Lemma. *The maps $\gamma \mapsto v_j(\gamma)$ ($j = 1, \ldots, d$) are continuous on Θ.*

Proof. From Section 3.1 we know that $\mathcal{C}_\gamma$ and $\mathcal{C}_\gamma^{-1}$ are continuous in γ, and therefore also all of the quantities defined in equations (3.18)–(3.21). From that follows the continuity of $v = v(\gamma)$. ■

To sum up, with the above choice of the function v the constant K—and thus the space $\mathscr{G} = \mathscr{G}(K, \rho, d)$—can now be precisely specified. We will see in our later examinations that if K is chosen according to the above criterion it is always sufficiently large such that all functions of interest are in included in $\mathscr{G}$. Note that in the special case $d = 1$ the above procedure yields $v(\gamma) = \mathcal{C}_\gamma \delta$. In that case we get (with the notation from Sections 3.5.1 and 3.5.2) $I_1 = I$ and $I_2 = \{\delta\}$. Hence, both, the approach for the specification of K in the univariate parameter case (from Section 3.5.1) and in the multivariate parameter case (from Section 3.5.2), do coincide.

3.2 Example. To better understand what really happens during the above construction procedure, we consider a simple example for the case of a two-dimensional parameter $\gamma \in \mathbb{R}^2$. Consider the Cholesky matrix $\mathcal{C}_\gamma = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, and $\delta = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$. Using the above notation, we have

$$\delta^{(1)} = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right), \quad \delta^{(2)} = \left(\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right), \quad \delta^{(3)} = \left(\begin{smallmatrix} -1 \\ -1 \end{smallmatrix}\right), \quad \delta^{(4)} = \left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right),$$

and therefore, according to (3.18),

$$w^{(1)} = \left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right), \quad w^{(2)} = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right), \quad w^{(3)} = \left(\begin{smallmatrix} -2 \\ -1 \end{smallmatrix}\right), \quad w^{(4)} = \left(\begin{smallmatrix} 0 \\ -1 \end{smallmatrix}\right).$$

From (3.19) we get $w = \left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right)$, and the resulting cuboid $[-w, w]$ has the corners

$$\widetilde{w}^{(1)} = \left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right), \quad \widetilde{w}^{(2)} = \left(\begin{smallmatrix} -2 \\ 1 \end{smallmatrix}\right), \quad \widetilde{w}^{(3)} = \left(\begin{smallmatrix} -2 \\ -1 \end{smallmatrix}\right), \quad \widetilde{w}^{(4)} = \left(\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right).$$

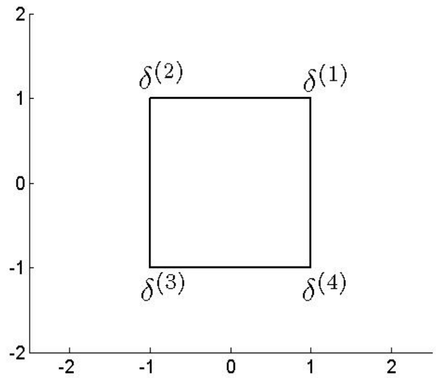
With $\mathcal{C}_\gamma^{-1} = \left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$ this yields

$$z^{(1)} = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right), \quad z^{(2)} = \left(\begin{smallmatrix} -3 \\ 1 \end{smallmatrix}\right), \quad z^{(3)} = \left(\begin{smallmatrix} -1 \\ -1 \end{smallmatrix}\right), \quad z^{(4)} = \left(\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right),$$

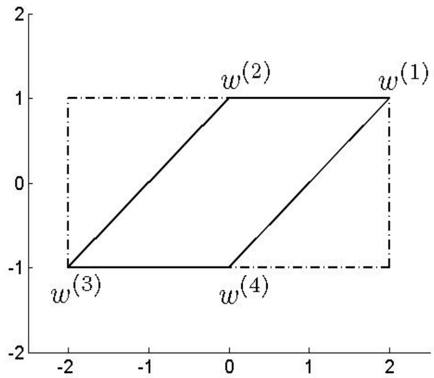
and therefore $\alpha = 3$ (according to (3.21)), which leads to

$$v = w/\alpha = \tfrac{1}{3}\left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right).$$

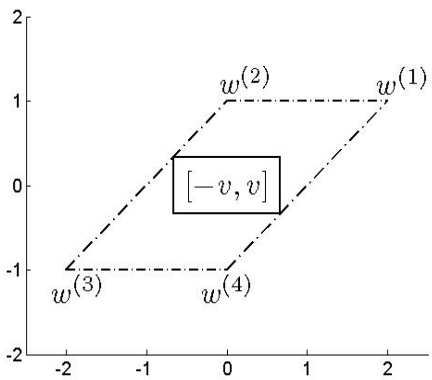
The following plots illustrate what happens during the construction of the cuboid $[-v, v]$.



(a)



(b)



(c)

Figure (a) shows the cuboid $[-\delta, \delta]$, and Figure (b) plots the set A_γ (solid lines) and the cuboid $[-\widetilde{w}, \widetilde{w}]$ (dotted lines). The "shrunk" cuboid $[-v, v]$ is depicted in Figure (c), together with A_γ (dotted lines). This last plot clearly shows that indeed $[-v, v] \subseteq A_\gamma$.

CHAPTER 4

Convergence of the local experiments

In this chapter we prove that the sequence of local experiments

$$(\mathcal{Y}^n, \mathcal{B}^n, \{P_{\gamma,g}^{(n)} : g \in \mathscr{G}\}) \tag{4.1}$$

converges weakly to a Gaussian limit experiment. This Gaussian experiment is introduced in Section 4.1, the weak convergence of the local experiments is proven in Section 4.2. It turns out that differentiability in quadratic mean is a suitable framework for this purpose. Furthermore, we show in Section 4.3 that the minimax risk in this limit experiment yields an asymptotic lower bound for both, the minimax risk in the local model and the minimax risk in the global model.

4.1 Characterising the limit experiment

Throughout this section, let $\gamma \in \Theta$ be an arbitrary, but fixed centre of localisation. We define a special Gaussian shift experiment which will later turn out to be the weak limit of the local models. For intervals $\mathcal{X} \subseteq \mathbb{R}$ we denote the space of continuous functions $f\colon \mathcal{X} \to \mathbb{R}^d$ by $C^d(\mathcal{X})$. The σ-algebra generated by the projections $\pi_t\colon C^d(\mathcal{X}) \to \mathbb{R}^d$, $s \mapsto s(t)$, is denoted $\mathscr{C}^d(\mathcal{X})$. In the following, we always consider the special case of the compact interval $\mathcal{X} = [-K, K]$, in which $\mathscr{C}^1([-K,K])$ coincides with the Borel algebra induced by the uniform norm $\|\cdot\|_u$. As a consequence, $C^1([-K,K])$ is separable, and thus is

$$\mathscr{C}^d([-K,K]) \;=\; \mathscr{C}^1([-K,K]) \otimes \ldots \otimes \mathscr{C}^1([-K,K]) \quad (d \text{ times}).$$

(For the above assertions see e.g. Gänssler and Stute, 1977, p. 21 ff., or Billingsley, 1968, pp. 54 ff. and 220 ff.) In the following, let

$$W = \{W_t = (W_t^{(1)}, \dots, W_t^{(d)})^\top : t \in [-K, K]\}$$

be a (time-shifted) d-dimensional Brownian motion on some abstract probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with $W_{-K} = 0$ $\mathbb{P}$-a.s. On the space $(C^d([-K, K]), \mathscr{C}^d([-K, K]))$ we consider the induced distributions

$$Q_g := Q_{\gamma,g} := \mathscr{L}\left(\int_{-K}^{\cdot} \mathcal{C}_\gamma g(s)\, ds + W\right). \tag{4.2}$$

The integral is of course to be interpreted component-wise. The index γ is sometimes omitted, if—like in this Chapter—a fixed γ is considered. The underlying space $(\Omega, \mathcal{A}, \mathbb{P})$ can be chosen equal to

$$(C^d([-K, K]), \mathscr{C}^d([-K, K]), Q_{\gamma,0}),$$

provided that we identify $W^{(i)}$ with the coordinate mapping

$$\mathrm{x}_i \colon C^d([-K, K]) \to C^1([-K, K]), \quad f = (f_1, \dots, f_d)^\top \mapsto f_i$$

(cf. Karatzas and Shreve, 1988, p. 72 f.). Moreover, we set $\mathrm{x} = (\mathrm{x}_1, \dots, \mathrm{x}_d)^\top$.

In the course of this chapter we want to show that the sequence of local experiments from (4.1) converges weakly to

$$(C^d([-K, K]), \mathscr{C}^d([-K, K]), \{Q_{\gamma,g} : g \in \mathscr{G}\})$$

(in the sense of Definition 2.14). To this end, we have to show that the log-likelihood ratios in the local experiments converge in distribution to the log-likelihood ratios of the Gaussian experiment. We shall first examine the latter. For this purpose we use some concepts from the Itô integration theory, however, only for the simple case of deterministic integrands. Klebaner (1998) provides a very nice presentation of this theory. We introduce some further notation for the stochastic integral calculus: Suppose a function $f = (f_1, \dots, f_d)^\top \in C^d([-K, K])$ is given. With the above considerations of the underlying probability spaces, we simply write the stochastic integrals $\int f_i\, dW^{(i)}$ as $\int f_i\, d\mathrm{x}_i$. Moreover, we set $\int f\, d\mathrm{x} := \sum_{i=1}^d \int f_i\, d\mathrm{x}_i$. This notation for the stochastic integrals corresponds to that used by Strasser (1985), Section 70, and Drees (2001).

4.1 Theorem. *Let $g_1, \ldots, g_m \in \mathscr{G}$ and $\gamma \in \Theta$ be given. Then*

$$\mathscr{L}\left(\left\{ \log \frac{dQ_{\gamma,g_j}}{dQ_{\gamma,0}} \right\}_{1\le j\le m} \,\middle|\, Q_{\gamma,0}\right) = \mathscr{N}(\{-\tfrac{1}{2}k_{jj}\}_{1\le j\le m}, \mathcal{K}),$$

with Q_{γ,g_j} according to (4.2), $k_{ij} = \int g_i(s)^\top \mathcal{I}_\gamma g_j(s)\ ds$ and $\mathcal{K} = (k_{ij}) \in \mathbb{R}^{m\times m}$.

Proof. The proof is based on the Girsanov theorem, which can be found in Karatzas and Shreve (1988), p. 190 ff., or in a simpler, but for our purpose sufficient version in Bingham and Kiesel (1998), Theorem 5.8.1. The Girsanov formula yields

$$\frac{dQ_{\gamma,g}}{dQ_{\gamma,0}}(\mathrm{x}) = \exp\left[\int \mathcal{C}_\gamma g\ d\mathrm{x} - \frac{1}{2}\int_{-K}^{K} \|\mathcal{C}_\gamma g(s)\|^2\ ds\right]. \qquad (4.3)$$

Exploiting the isometry and linearity properties of the Itô integral $(*)$ we conclude that

$$\begin{aligned}\mathscr{L}\left(\log \frac{dQ_{\gamma,g}}{dQ_{\gamma,0}} \,\middle|\, Q_{\gamma,0}\right) &= \mathscr{L}\big(\textstyle\int \mathcal{C}_\gamma g\ d\mathrm{x} - \frac{1}{2}\int \|\mathcal{C}_\gamma g(s)\|^2\ ds \ \big|\ Q_{\gamma,0}\big) \\ &\overset{(*)}{=} \mathscr{N}\big(-\tfrac{1}{2}\textstyle\int \|\mathcal{C}_\gamma g(s)\|^2\ ds, \int \|\mathcal{C}_\gamma g(s)\|^2\ ds\big).\end{aligned}$$

In order to prove that also the complete vector

$$\{\log(dQ_{\gamma,g_j}/dQ_{\gamma,0}) : j = 1, \ldots, m\}$$

is normally distributed, it suffices to show that any linear combination

$$a_1 \log \frac{dQ_{\gamma,g_1}}{dQ_{\gamma,0}} + \ldots + a_m \log \frac{dQ_{\gamma,g_m}}{dQ_{\gamma,0}}$$

is normally distributed. With the stochastic integral notation introduced above, and using the linearity of the Itô integral we conclude that under $Q_{\gamma,0}$

$$\begin{aligned}&\sum_{i=1}^m a_i \log \frac{dQ_{\gamma,g_i}}{dQ_{\gamma,0}} \\ &\quad \overset{\mathscr{L}}{=} \sum_{i=1}^m a_i \left(\int \mathcal{C}_\gamma g_i\ d\mathrm{x} - \tfrac{1}{2}\int \|\mathcal{C}_\gamma g_i(s)\|^2\ ds\right) \\ &\quad = \int \left(\sum_{i=1}^m a_i \mathcal{C}_\gamma g_i\right) d\mathrm{x} - \tfrac{1}{2}\int \left(\sum_{i=1}^m a_i \|\mathcal{C}_\gamma g_i(s)\|^2\right) ds.\end{aligned}$$

The random variable in the last expression is normally distributed. These arguments imply that $\{\log(dQ_{\gamma,g_j}/dQ_{\gamma,0})\}_{1\le j\le m}$ is Gaussian under $Q_{\gamma,0}$, with mean vector $\{-\frac{1}{2}\int \|\mathcal{C}_\gamma g_j(s)\|^2\ ds\}_{1\le j\le m}$. The covariance function can be calculated to

$$\begin{aligned}\operatorname{Cov}(\textstyle\int \mathcal{C}_\gamma g_i\ d\mathrm{x}, \int \mathcal{C}_\gamma g_j\ d\mathrm{x}) &\overset{(\dagger)}{=} \textstyle\int (\mathcal{C}_\gamma g_i(s))^\top \mathcal{C}_\gamma g_j(s)\ ds \\ &= \textstyle\int g_i(s)^\top \mathcal{I}_\gamma g_j(s)\ ds,\end{aligned}$$

where (†) follows from Klebaner (1998), p. 116. This proves the assertion of the theorem. ■

4.2 Proving weak convergence

We prove that the local experiments (4.1) converge weakly to the Gaussian experiment from the preceding section. As before, $\gamma \in \Theta$ denotes an arbitrary, but fixed centre of localisation.

4.2 Theorem. *Consider an arbitrary, but fixed $\gamma \in \Theta$. Under the assumptions given in Sections 3.1 and 3.3 we have*

$$\begin{aligned}&(\mathcal{Y}^n, \mathcal{B}^n, \{P^{(n)}_{\gamma,g} : g \in \mathscr{G}\}) \\ &\qquad \xrightarrow{\mathrm{w}} (C^d([-K,K]), \mathscr{C}^d([-K,K]), \{Q_{\gamma,g} : g \in \mathscr{G}\}).\end{aligned}$$

For the lengthy proof of the theorem we first show that all finite-dimensional vectors of the form

$$\left(\log \frac{dP^{(n)}_{\gamma,g_1}}{dP^{(n)}_\gamma}, \ldots, \log \frac{dP^{(n)}_{\gamma,g_m}}{dP^{(n)}_\gamma}\right)^\top, \quad g_1,\ldots,g_m \in \mathscr{G}, \tag{4.4}$$

converge in distribution under $P^{(n)}_\gamma = P^{(n)}_{\gamma,0}$. From that one can easily conclude the weak convergence of the local experiments. We first address the special case $m = 1$, which is discussed in the following proposition.

4.3 Proposition. *Let arbitrary, but fixed $g \in \mathscr{G}$ and $\gamma \in \Theta$ be given. Under the assumptions from Sections 3.1 and 3.3 we have*

$$\mathscr{L}\left(\log \frac{dP^{(n)}_{\gamma,g}}{dP^{(n)}_\gamma} \,\middle|\, P^{(n)}_\gamma\right) \rightsquigarrow \mathscr{N}(-\tfrac{1}{2}\kappa^2, \kappa^2),$$

where

$$\kappa^2 = \int g(s)^\top \mathcal{I}_\gamma g(s)\ ds \ge 0.$$

In particular, we have $P^{(n)}_\gamma \lhd\rhd P^{(n)}_{\gamma,g}$.

4.4 Remark. For $\kappa^2 = 0$ we identify $\mathscr{N}(0,0)$ with the Dirac measure δ_0, as in the addendum to Proposition 2.8. In that case, the positive definiteness of $\mathcal{I}_\gamma$ and the continuity of g on the compactum $[-K, K]$ imply $g \equiv 0$. Then, of course, the assertion is trivial.

In preparation of the proof, for $g \in \mathscr{G}$ we define a triangular scheme of stochastically independent real random variables U^g_{nj} on $(\mathcal{Y}, \mathcal{B})$, setting

$$U^g_{nj} := a_n g(\tilde{x}_{nj})^\top \dot{\ell}_\gamma, \quad n \in \mathbb{N}, \quad j = 1, \ldots, n. \tag{4.5}$$

In the context of Remark 2.9, this triangular scheme can be interpreted as a family of random variables

$$U^g_{nj} \colon (\mathcal{Y}^n, \mathcal{B}^n) \to (\mathbb{R}, \mathbb{B}),$$

provided that we identify U^g_{nj} with $U^g_{nj}(\pi_{nj})$, where again π_{nj} denotes the jth canonical projection. If we write the elements of the space $\mathcal{Y}^n$ of the form

$$(Y_{n1}, \ldots, Y_{nn})^\top,$$

then we can simply write $U^g_{nj} = U^g_{nj}(\pi_{nj})$ as $U^g_{nj}(Y_{nj})$. We will use this notation throughout the thesis. However, we will often also simply omit the argument Y_{nj} if this does not lead to misinterpretations. In addition, we set

$$U^g_n := \sum_{j=1}^n U^g_{nj}, \tag{4.6}$$

which is to be interpreted as a random variable $U^g_n \colon (\mathcal{Y}^n, \mathcal{B}^n) \to (\mathbb{R}, \mathbb{B})$, more formally written as

$$U^g_n = U^g_n(Y_{n1}, \ldots, Y_{nn}) = \sum_{j=1}^n U^g_{nj}(Y_{nj}).$$

Furthermore, in the following let

$$\begin{aligned} \mathbb{K}_n &:= \big\{ j = 1, \ldots, n : |x_{nj} - x_0| \le \tfrac{K}{b_n} \big\} \\ &= \big\{ j = 1, \ldots, n : |\tilde{x}_{nj}| \le K \big\}, \\ k_n &:= \#\mathbb{K}_n, \end{aligned} \tag{4.7}$$

where K is the constant from (3.10) and the $\tilde{x}_{nj}$ are defined according to (3.16). From (3.14)–(3.16) we conclude that $k_n \to \infty$ and $k_n \simeq 2Kn/b_n$[1], and therefore

$$a_n^2 \simeq 2K/k_n. \tag{4.8}$$

The random variables U_n^g as well as the quantities $\mathbb{K}_n$ and k_n play an important role throughout this thesis.

Proof of Proposition 4.3. The proof is based on an application of Proposition 2.8. First we consider the expressions $a_n g(\tilde{x}_{nj})$. Note that for all vectors $v \in \mathbb{R}^d$ the inequality $\|v\| \le \sqrt{d}\|v\|_\infty$ holds (cf. Königsberger, 2000, p. 11). Thus, we get

$$\begin{aligned} \|a_n g(\tilde{x}_{nj})\|^2 &\le d a_n^2 \|g(\tilde{x}_{nj})\|_\infty^2 \\ &\le d a_n^2 \|g\|_{u,d}^2 \, \mathbb{1}\{j \in \mathbb{K}_n\} \\ &\le a_n^2 d M^2 \, \mathbb{1}\{j \in \mathbb{K}_n\}, \end{aligned}$$

with M as in (3.11). Setting $\Delta_n := a_n M \sqrt{d}$, we thus have

$$\max\{\|a_n g(\tilde{x}_{nj})\| : j = 1, \ldots, n\} \le \Delta_n \to 0. \tag{4.9}$$

In addition, we conclude with (4.8) that

$$\begin{aligned} \Delta &:= \sup\left\{ \sum_{j=1}^n \|a_n g(\tilde{x}_{nj})\|^2 : n \in \mathbb{N} \right\} \\ &\lesssim \sup\{a_n^2 k_n : n \in \mathbb{N}\} < \infty.^{2)} \end{aligned} \tag{4.10}$$

For $n \in \mathbb{N}$ and $j = 1, \ldots, n$ we define

$$U_{nj} := U_{nj}^g = a_n g(\tilde{x}_{nj})^\top \dot{\ell}_\gamma,$$

according to (4.5). In the following, we show that this choice of random variables U_{nj} fits the requirements from Proposition 2.8, where P_γ, $P_{\gamma_{g,n}(x_{nj})}$ and μ take the roles of the measures P_{nj}, Q_{nj} and μ_{nj}, respectively. The assertion of our proposition follows then immediately from that of Proposition 2.8.

[1] For sequences d_n and e_n we write $d_n \simeq e_n$ iff $d_n/e_n \to 1$.

[2] For two expressions $a(s)$ and $b(s)$ we write $a(s) \lesssim b(s)$ iff $a(s) \le kb(s)$ with a positive constant k, which does not depend on the argument s.

Proof of (2.2). From the definitions of U_{nj} and $\gamma_{g,n}(\cdot)$ (see also (3.17)) we get

$$\begin{aligned} R_{nj} &:= \int \left[\sqrt{p_{\gamma_{g,n}(x_{nj})}} - \sqrt{p_\gamma} - \tfrac{1}{2}U_{nj}\sqrt{p_\gamma}\right]^2 d\mu \\ &= \int \left[\sqrt{p_{\gamma+a_n g(\tilde{x}_{nj})}} - \sqrt{p_\gamma} - \tfrac{1}{2}a_n g(\tilde{x}_{nj})^\top \dot{\ell}_\gamma \sqrt{p_\gamma}\right]^2 d\mu. \end{aligned}$$

Let an arbitrary $\varepsilon > 0$ be given. Due to the presumed differentiability in quadratic mean at γ there is a $\delta > 0$, such that (using the convention $0/0 := 0$)

$$\frac{1}{\|s\|^2}\int \left[\sqrt{p_{\gamma+s}} - \sqrt{p_\gamma} - \tfrac{1}{2}s^\top \dot{\ell}_\gamma \sqrt{p_\gamma}\right]^2 d\mu < \frac{\varepsilon}{\Delta},$$

as long as $\|s\| < \delta$. Provided that n is sufficiently large, we know from (4.9) that $\sup_j \|a_n g(\tilde{x}_{nj})\| \leq \Delta_n < \delta$, and thus

$$\frac{R_{nj}}{\|a_n g(\tilde{x}_{nj})\|^2} < \frac{\varepsilon}{\Delta}$$

for all j. Summation of the R_{nj} yields

$$\sum_{j=1}^n R_{nj} < \sum_{j=1}^n \frac{\varepsilon}{\Delta}\|a_n g(\tilde{x}_{nj})\|^2 \overset{(4.10)}{\leq} \varepsilon.$$

Since $\varepsilon > 0$ can be chosen arbitrarily small, this yields (2.2).
Proof of (2.3) and (2.8). Lemma 2.5 implies

$$\int U_{nj}\, dP_\gamma = a_n g(\tilde{x}_{nj})^\top \int \dot{\ell}_\gamma\, dP_\gamma = 0,$$

and hence (2.3) and (2.8) hold.
Proof of (2.4) and (2.7). The positive definiteness of $\mathcal{I}_\gamma$ implies

$$\kappa^2 = \int g^\top(s)\mathcal{I}_\gamma g(s)\, ds \geq 0,$$

as stated in the proposition. We show now that $\sum_{j=1}^n \int U_{nj}^2\, dP_\gamma$ converges to κ^2, i.e. (2.7) holds, and thus of course also (2.4). A simple calculation yields

$$\begin{aligned} \sum_{j=1}^n \int U_{nj}^2\, dP_\gamma &= \sum_{j=1}^n a_n g(\tilde{x}_{nj})^\top \left(\int \dot{\ell}_\gamma \dot{\ell}_\gamma^\top\, dP_\gamma\right) a_n g(\tilde{x}_{nj}) \\ &= \sum_{j=1}^n a_n^2 g(\tilde{x}_{nj})^\top \mathcal{I}_\gamma g(\tilde{x}_{nj}). \end{aligned} \tag{4.11}$$

In order to prove that the second moments converge, we first consider the case of a one-dimensional parameter space $\Theta \subseteq \mathbb{R}$ and then the general case $\Theta \subseteq \mathbb{R}^d$ with $d \geq 1$.
Special case $\Theta \subseteq \mathbb{R}$. In this case (4.11) is reduced to

$$\sum_{j=1}^{n} \int U_{nj}^2 \, dP_\gamma = \mathcal{I}_\gamma \sum_{j=1}^{n} a_n^2 g(\tilde{x}_{nj})^2,$$

where the Fisher information $\mathcal{I}_\gamma$ is a non-negative number. We show that the sum term converges to $\int g(s)^2 \, ds$. To this end, for fixed n we consider the set

$$\mathfrak{Z}_n := \{-K = z_0 < z_1 < \ldots < z_{k_n} = K\},$$

where

$$z_j := -K + \frac{2jK}{k_n}, \quad j = 0, \ldots, k_n.$$

This partition divides the interval $[-K, K]$ into k_n equally large subintervals $[z_{j-1}, z_j)$ of length $2K/k_n$. Obviously, each such subinterval includes one and only one of the points $\tilde{x}_{nj}$ with index $j \in \mathbb{K}_n$. According to (4.8) we have $a_n^2 \simeq 2K/k_n = z_j - z_{j-1}$. Hence, $\sum_{j=1}^{n} a_n^2 g(\tilde{x}_{nj})^2$ can be interpreted as a Riemann sum for $g^2(\cdot)$ with respect to the partition $\mathfrak{Z}_n$. With $a_n \to 0$ the fineness of the partition tends to zero, too. Consequently, since g is continuous and thus integrable (on $[-K, K]$), we have

$$\sum_{j=1}^{n} a_n^2 g(\tilde{x}_{nj})^2 \to \int_{-K}^{K} g(s)^2 \, ds.$$

Applying that $g_{|[-K,K]^C} \equiv 0$, we finally obtain

$$\sum_{j=1}^{n} \int U_{nj}^2 \, dP_\gamma = \mathcal{I}_\gamma \sum_{j=1}^{n} a_n^2 g(\tilde{x}_{nj})^2 \to \mathcal{I}_\gamma \int g(s)^2 \, ds.$$

Hence, we have shown (2.7) holds for the case of a one-dimensional parameter space $\Theta \subseteq \mathbb{R}$.
General case $\Theta \subseteq \mathbb{R}^d$, $d \geq 1$. For the general case $\Theta \subseteq \mathbb{R}^d$ we get

from (4.11) and some simple matrix multiplication

$$\begin{aligned}\sum_{j=1}^{n} \int U_{nj}^2 \, dP_\gamma &= \sum_{j=1}^{n} \sum_{k=1}^{d} \sum_{l=1}^{d} a_n^2 g_k(\tilde{x}_{nj}) g_l(\tilde{x}_{nj}) \mathcal{I}_{\gamma,kl} \\ &= \sum_{k=1}^{d} \sum_{l=1}^{d} \sum_{j=1}^{n} a_n^2 g_k(\tilde{x}_{nj}) g_l(\tilde{x}_{nj}) \mathcal{I}_{\gamma,kl}.\end{aligned}$$

Here $g_k(\cdot)$ and $\mathcal{I}_{\gamma,kl}$ denote the corresponding components of the vector-valued function $g(\cdot)$ and the matrix $\mathcal{I}_\gamma$, respectively. The same arguments as in the above discussed case of a one-dimensional parameter yield

$$\sum_{j=1}^{n} a_n^2 g_k(\tilde{x}_{nj}) g_l(\tilde{x}_{nj}) \mathcal{I}_{\gamma,kl} \to \mathcal{I}_{\gamma,kl} \int g_k(s) g_l(s) \, ds,$$

and thus

$$\begin{aligned}\sum_{j=1}^{n} \int U_{nj}^2 \, dP_\gamma &\to \sum_{k=1}^{d} \sum_{l=1}^{d} \mathcal{I}_{\gamma,kl} \int g_k(s) g_l(s) \, ds \\ &= \int \left(\sum_{k=1}^{d} \sum_{l=1}^{d} \mathcal{I}_{\gamma,kl} g_k(s) g_l(s) \right) ds \\ &= \int g(s)^\top \mathcal{I}_\gamma g(s) \, ds.\end{aligned}$$

Proof of (2.5). Let an arbitrary $\varepsilon > 0$ be given. The Cauchy-Schwarz inequality (applied point-wise) yields

$$U_{nj}^2 = |a_n g(\tilde{x}_{nj})^\top \dot{\ell}_\gamma|^2 \le \|a_n g(\tilde{x}_{nj})\|^2 \, \|\dot{\ell}_\gamma\|^2 \overset{(4.9)}{\le} \Delta_n^2 \|\dot{\ell}_\gamma\|^2,$$

which implies

$$\begin{aligned}\sum_{j=1}^{n} \int U_{nj}^2 \mathbb{1}\{|U_{nj}| > \varepsilon\} \, dP_\gamma &= \sum_{j=1}^{n} \int U_{nj}^2 \mathbb{1}\{|U_{nj}|^2 > \varepsilon^2\} \, dP_\gamma \\ &\le \left(\sum_{j=1}^{n} \|a_n g(\tilde{x}_{nj})\|^2 \right) \int \|\dot{\ell}_\gamma\|^2 \mathbb{1}\{\Delta_n^2 \|\dot{\ell}_\gamma\|^2 > \varepsilon^2\} \, dP_\gamma.\end{aligned}$$

Due to (4.10) the sum in the last expression is bounded by Δ. Since $\Delta_n \to 0$, the dominated convergence theorem implies that the integral converges to zero. ■

4.5 Corollary. *Under the assumptions of Proposition 4.3 we have*

$$\log \frac{dP^{(n)}_{\gamma,g}}{dP^{(n)}_{\gamma}} - \left(\sum_{j=1}^{n} U^{g}_{nj} - \tfrac{1}{2}\kappa^2 \right) = o_{P^{(n)}_{\gamma}}(1),$$

with $\kappa^2 = \int g(s)^\top \mathcal{I}_\gamma g(s)\, ds$.

Proof. The assertion follows directly from the above proof and (2.6) in Proposition 2.8. ■

Next, we want to show that the vector from (4.4) also converges in distribution in the multi-dimensional case. To this end, we consider m arbitrary, but fixed functions $g_1, \ldots, g_m \in \mathscr{G}$. The corresponding local alternatives $\gamma_{g_j,n}$ may be defined according to (1.3). We generalise the result from Proposition 4.3:

4.6 Proposition. *Let m arbitrary, but fixed functions $g_1, \ldots, g_m \in \mathscr{G}$ and some fixed $\gamma \in \Theta$ be given. Moreover, let the assumptions from Sections 3.1 and 3.3 hold. Then*

$$\mathscr{L}\left(\left\{ \log \frac{dP^{(n)}_{\gamma,g_j}}{dP^{(n)}_{\gamma}} \right\}_{1 \le j \le m} \,\middle|\, P^{(n)}_{\gamma} \right) \rightsquigarrow \mathscr{N}(\{-\tfrac{1}{2}k_{jj}\}_{1\le j\le m}, \mathcal{K}),$$

where

$$k_{ij} = \int g_i(s)^\top \mathcal{I}_\gamma g_j(s)\, ds$$

and $\mathcal{K} = (k_{ij}) \in \mathbb{R}^{m\times m}$ *(cf. Theorem 4.1).*

The proposition follows from several auxiliary results, which will be verified separately. Again, we consider random variables

$$U^{g_i}_{nj} := a_n g_i(\tilde{x}_{nj})^\top \dot{\ell}_\gamma \quad \text{and} \quad U^{g_i}_{n} := \sum_{j=1}^{n} U^{g_i}_{nj},$$

defined according to (4.5) and (4.6), respectively, where the variables $U^{g_i}_{nj}$, $j = 1, \ldots, n$, are assumed independent. We examine the asymptotic behaviour of the corresponding vector $(U^{g_1}_n, \ldots, U^{g_m}_n)^\top$. Note that in the further examinations of this section all expectations and covariances are to be built with respect to the product measure $P^{(n)}_{\gamma}$.

4.7 Lemma. *Let the assumptions from Proposition 4.6 hold. Then* $\operatorname{Cov}(U_n^{g_r}, U_n^{g_s}) \to k_{rs}$ *for* $r, s \in \{1, \ldots, m\}$*, and thus in particular*

$$\mathrm{E}[(U_n^{g_1}, \ldots, U_n^{g_m})w]^2 \to w^\top \mathcal{K} w$$

for all $w \in \mathbb{R}^m$*.*

Proof. With Lemma 2.5 we have

$$\mathrm{E}\, U_{nj}^{g_i} = a_n g_i(\tilde{x}_{nj})^\top \int \dot{\ell}_\gamma \, dP_\gamma = 0 \tag{4.12}$$

for all i and all n, j. Furthermore, $U_{nj}^{g_r}$ and $U_{nk}^{g_s}$ are independent for $j \neq k$. We conclude that

$$\begin{aligned}
\operatorname{Cov}(U_n^{g_r}, U_n^{g_s}) &= \sum_{j=1}^n \sum_{k=1}^n \mathrm{E}(U_{nj}^{g_r} U_{nk}^{g_s}) \\
&= \sum_{j=1}^n \mathrm{E}(U_{nj}^{g_r} U_{nj}^{g_s}) \\
&= \sum_{j=1}^n a_n^2 g_r(\tilde{x}_{nj})^\top \left(\int \dot{\ell}_\gamma \dot{\ell}_\gamma^\top \, dP_\gamma \right) g_s(\tilde{x}_{nj}) \\
&= \sum_{j=1}^n a_n^2 g_r(\tilde{x}_{nj})^\top \mathcal{I}_\gamma g_s(\tilde{x}_{nj}).
\end{aligned}$$

The last expression converges to k_{rs}, as in the proof of (2.7) in Proposition 4.3. This implies the assertion. ■

4.8 Lemma. *Under the assumptions from Proposition 4.6*

$$\mathscr{L}((U_n^{g_1}, \ldots, U_n^{g_m})^\top \mid P_\gamma^{(n)}) \rightsquigarrow \mathscr{N}(0, \mathcal{K}).$$

Proof. Using the Cramér-Wold device it suffices to show that

$$\mathscr{L}((U_n^{g_1}, \ldots, U_n^{g_m})w \mid P_\gamma^{(n)}) \rightsquigarrow \mathscr{N}(0, w^\top \mathcal{K} w) \tag{4.13}$$

for all vectors $w \in \mathbb{R}^m$. Let some $w = (w_1, \ldots, w_m)^\top \in \mathbb{R}^m$ be given. Lemma 4.7 yields $\mathrm{E}[(U_n^{g_1}, \ldots, U_n^{g_m})w]^2 \to w^\top \mathcal{K} w$. Moreover, $\mathrm{E}[(U_n^{g_1}, \ldots, U_n^{g_m})w] = 0$, as a consequence of (4.12). In the case $w^\top \mathcal{K} w = 0$ we have $\operatorname{Var}[(U_n^{g_1}, \ldots, U_n^{g_m})w] \to 0$, which implies $(U_n^{g_1}, \ldots, U_n^{g_m})w \rightsquigarrow \delta_0$ (the Dirac measure), and thus the convergence (4.13). It remains to prove (4.13) for the case $w^\top \mathcal{K} w > 0$.

This can be done with an application of the Lindeberg-Feller theorem. We simply have to show that the Lindeberg condition for the functions $\sum_{i=1}^m w_i U_{nj}^{g_i}$ holds, i.e. to verify that

$$L_n(\varepsilon) := \sum_{j=1}^n \int \Big(\sum_{k=1}^m w_k U_{nj}^{g_k} \Big)^2 \mathbb{1}\Big\{ \Big| \sum_{i=1}^m w_i U_{nj}^{g_i} \Big| > \varepsilon \Big\} \, dP_\gamma \to 0$$

for all $\varepsilon > 0$. The Cauchy-Schwarz inequality implies

$$\Big(\sum_{k=1}^m w_k U_{nj}^{g_k} \Big)^2 \le \|w\|^2 \sum_{k=1}^m (U_{nj}^{g_k})^2 \quad P_\gamma\text{-a.s.}$$

Moreover, we have

$$\begin{aligned} \mathbb{1}\Big\{ \Big| \sum_{i=1}^m w_i U_{nj}^{g_i} \Big| > \varepsilon \Big\} &\le \mathbb{1}\Big\{ \sum_{i=1}^m |w_i U_{nj}^{g_i}| > \varepsilon \Big\} \\ &\le \sum_{i=1}^m \mathbb{1}\{ |w_i U_{nj}^{g_i}| > \varepsilon/m \} \quad P_\gamma\text{-a.s.} \end{aligned}$$

Altogether, this yields

$$L_n(\varepsilon) \le \|w\|^2 \sum_{i=1}^m \sum_{k=1}^m \sum_{j=1}^n \int \big(U_{nj}^{g_k} \big)^2 \mathbb{1}\{ |w_i U_{nj}^{g_i}| > \varepsilon/m \} \, dP_\gamma.$$

With some minor modifications of the calculations for the proof of (2.5) in Proposition 4.3 one concludes that

$$\sum_{j=1}^n \int \big(U_{nj}^{g_k} \big)^2 \mathbb{1}\{ |w_i U_{nj}^{g_i}| > \varepsilon/m \} \, dP_\gamma \to 0,$$

and thus $L_n(\varepsilon) \to 0$, too. Now the assertion is a direct consequence of the Lindeberg-Feller theorem (van der Vaart, 1998, Proposition 2.27). ∎

Proof of Proposition 4.6. Since a sequence of vector-valued random variables converges in probability if and only if each of its components converges, Corollary 4.5 yields

$$\Big\{ \log \frac{dP_{\gamma,g_j}^{(n)}}{dP_\gamma^{(n)}} \Big\}_{1 \le j \le m} - \Big((U_n^{g_1}, \dots, U_n^{g_m})^\top - \{ \tfrac{1}{2} k_{jj} \}_{1 \le j \le m} \Big) = o_{P_\gamma^{(n)}}(1).$$

Furthermore, Lemma 4.8 implies

$$\mathscr{L}((U_n^{g_1},\dots,U_n^{g_m})^\top - \{\tfrac{1}{2}k_{jj}\}_{1\le j\le m} \mid P_\gamma^{(n)}) \rightsquigarrow \mathscr{N}(-\{\tfrac{1}{2}k_{jj}\}_{1\le j\le m}, \mathcal{K}).$$

The assertion follows from Slutsky's lemma. ■

With the above results and some additional arguments from the theory of weak convergence of experiments we are now able to prove Theorem 4.2.

Proof of Theorem 4.2. From Proposition 4.3 we obtain $P_{\gamma,g}^{(n)} \lhd\rhd P_\gamma^{(n)}$ for each $g \in \mathscr{G}$. Since "$\lhd\rhd$" is an equivalence relation this implies $P_{\gamma,g}^{(n)} \lhd\rhd P_{\gamma,h}^{(n)}$ for all $g,h \in \mathscr{G}$. Therefore, the local experiments $(\mathcal{Y}^n, \mathcal{B}^n, \{P_{\gamma,g}^{(n)} : g \in \mathscr{G}\})$ are contiguous in the sense of Definition 2.16. Furthermore, for arbitrary $g_1,\dots,g_m \in \mathscr{G}$, and with $P_\gamma^{(n)} = P_{\gamma,0}^{(n)}$, we conclude from Proposition 4.6 and from Theorem 4.1 that

$$\mathscr{L}\left(\left\{\log\frac{dP_{\gamma,g_j}^{(n)}}{dP_\gamma^{(n)}}\right\}_{1\le j\le m}\,\middle|\, P_\gamma^{(n)}\right) \rightsquigarrow \mathscr{L}\left(\left\{\log\frac{dQ_{\gamma,g_j}}{dQ_{\gamma,0}}\right\}_{1\le j\le m}\,\middle|\, Q_{\gamma,0}\right),$$

and the continuous mapping theorem (cf. van der Vaart, 1998, p. 7) yields

$$\mathscr{L}\left(\left\{\frac{dP_{\gamma,g_j}^{(n)}}{dP_\gamma^{(n)}}\right\}_{1\le j\le m}\,\middle|\, P_\gamma^{(n)}\right) \rightsquigarrow \mathscr{L}\left(\left\{\frac{dQ_{\gamma,g_j}}{dQ_{\gamma,0}}\right\}_{1\le j\le m}\,\middle|\, Q_{\gamma,0}\right).$$

The assertion follows now from Theorem 2.17. ■

4.9 Remark. Note that all of the above calculations were done *without* the additional growth condition from (3.10). Indeed, in all of the above proofs we only used the assumption that the functions g are continuous on $[-K,K]$ and equal to zero outside of this compactum.

4.3 Lower asymptotic minimax risk bounds

The above proven weak convergence of the local experiments has some immediate consequences for the asymptotic minimax risk within these experiments. The following theorem states that the maximal risk of any sequence of estimators in the local experiments is asymptotically bounded from below by the minimax risk in the limit experiment.

4.10 Theorem. *Consider an arbitrary, but fixed centre of localisation $\gamma \in \Theta$, and suppose that the assumptions from Sections 3.1 and 3.3 hold (in particular, let $\delta = (\delta_1, \dots, \delta_d)^\top$ be given, with components $\delta_i > 0$). Moreover, let $\hat{\gamma}_{g,n}(x_0)$ be an arbitrary estimator for the parameter $\gamma_{g,n}(x_0)$ (with unknown $g \in \mathscr{G}$). Then*

$$\liminf_{n\to\infty} \sup_{g\in\tilde{\mathscr{G}}} P^{(n)}_{\gamma,g}\big\{\hat{\gamma}_{g,n}(x_0) - \gamma_{g,n}(x_0) \in [-\delta a_n, \delta a_n]^C\big\}$$
$$\geq \inf_{\varrho} \sup_{g\in\tilde{\mathscr{G}}} \int \varrho\left(z, [g(0) - \delta, g(0) + \delta]^C\right) \, dQ_{\gamma,g}(z)$$

holds for arbitrary subspaces $\tilde{\mathscr{G}} \subseteq \mathscr{G}$. The infimum in the above display is built over the class of all randomised estimators ϱ for $g(0)$, i.e. over all Markov kernels ϱ from $(C^d([-K,K]), \mathscr{C}^d([-K,K]))$ to $(\mathbb{R}^d, \mathbb{B}^d)$.

Proof. Consider the loss function $W = \{W_g : g \in \mathscr{G}\}$, where

$$W_g(s) := \mathbb{1}\big\{s - g(0) \in [-\delta, \delta]^C\big\}, \quad s \in \mathbb{R}^d.$$

W is lower semi-continuous and level-compact (cf. Definition 2.18). To simplify notation we write the local experiments and the limit experiment in this proof as

$$E_n := (\mathcal{Y}^n, \mathcal{B}^n, \{P^{(n)}_{\gamma,g} : g \in \mathscr{G}\})$$

and

$$E := (C^d([-K,K]), \mathscr{C}^d([-K,K]), \{Q_{\gamma,g} : g \in \mathscr{G}\}),$$

respectively. It is clear that each of the experiments E_n is dominated, and the application of the Girsanov theorem in the proof of Theorem 4.1 implies that the same holds true for the experiment E. Moreover, we have shown in Theorem 4.2 that $E_n \xrightarrow{\mathrm{w}} E$. Hence, the conditions from Theorem 2.19 apply, and therefore—using the notation from Definition 2.11—every sequence of decision functions $\varrho_n \in \mathcal{R}(E_n, \mathbb{R}^d)$ satisfies the inequality

$$\liminf_{n\to\infty} \sup_{g\in\mathscr{G}} R(g, \varrho_n) \geq \inf_{\varrho\in\mathcal{R}(E,\mathbb{R}^d)} \sup_{g\in\mathscr{G}} R(g, \varrho). \tag{4.14}$$

For $g(0)$ we consider the estimator $\hat{g}_n(0) := (\hat{\gamma}_{g,n}(x_0) - \gamma)/a_n$. Clearly, (4.14) also holds if we specialise ϱ_n to

$$\varrho_n \colon (\mathcal{Y}^n, \mathbb{B}^d) \to [0,1], \quad (y, B) \mapsto \mathbb{1}\{\hat{g}_{n,y}(0) \in B\}.$$

Here we use the notation $\hat{g}_{n,y}(0)$ to clarify that $\hat{g}_n(0)$ is based on the observation $y = (y_1, \dots, y_n)^\top$ from the local experiment E_n. This special choice of ϱ_n fulfils

$$\begin{aligned} R(g, \varrho_n) &= \iint W_g(s) \varrho_n(y, ds)\ P^{(n)}_{\gamma,g}(dy) \\ &= \iint \mathbb{1}\{s - g(0) \in [-\delta, \delta]^C\} \varrho_n(y, ds)\ P^{(n)}_{\gamma,g}(dy) \\ &= \int \mathbb{1}\{\hat{g}_{n,y}(0) - g(0) \in [-\delta, \delta]^C\}\ P^{(n)}_{\gamma,g}(dy) \\ &= P^{(n)}_{\gamma,g}\{\hat{g}_n(0) - g(0) \in [-\delta, \delta]^C\} \\ &= P^{(n)}_{\gamma,g}\{\hat{\gamma}_{g,n}(x_0) - \gamma_{g,n}(x_0) \in [-\delta a_n, \delta a_n]^C\}. \end{aligned}$$

An analogue calculation for $\varrho \in \mathcal{R}(E, \mathbb{R}^d)$ yields

$$R(g, \varrho) = \int \varrho\left(z, [g(0) - \delta, g(0) + \delta]^C\right)\ dQ_{\gamma,g}(z).$$

Plugging these results into (4.14) proves the assertion of the theorem for $\tilde{\mathscr{G}} = \mathscr{G}$. Passing over from the experiments E_n and E to corresponding subexperiments the assertion follows for general subsets $\tilde{\mathscr{G}} \subseteq \mathscr{G}$, too (cf. Remark 2.15). ■

Note that Theorem 4.10 holds for arbitrary sequences of estimators $\hat{\gamma}_{g,n}(x_0)$. Consequently, the minimax risk in the limit experiment (and every lower bound on it) is asymptotically a lower bound for the minimax risk in the sequence of the local models. It can even be shown that this lower bound in a certain way carries over to the global model.

4.11 Theorem. *Let $\hat{\xi}_n(x_0)$ be an arbitrary estimator for $\xi(x_0)$, the value of the parameter function at x_0 in the global model $\mathscr{F} = \mathscr{F}(x_0, \Theta^*, u)$. Presume that the assumptions from Sections 3.1 and 3.3 hold. Then*

$$\begin{aligned} &\liminf_{n\to\infty} \sup_{\xi\in\mathscr{F}} P^{(n)}_{\xi}\{\hat{\xi}_n(x_0) - \xi(x_0) \in [-\delta a_n, \delta a_n]^C\} \\ &\qquad \geq \sup_{\gamma\in\Theta^*} \inf_{\varrho} \sup_{g\in\mathscr{G}} \int \varrho\left(z, [g(0) - \delta, g(0) + \delta]^C\right)\ dQ_{\gamma,g}(z), \end{aligned}$$

with ϱ as in Theorem 4.10.

Proof. Let $B_\varepsilon(v)$ denote the open ball with radius ε, centred around the point v. We set $H := (\Theta^*)^C$, $H_\varepsilon := H \cup (\bigcup_{v\in H} B_\varepsilon(v))$ and $\Theta^*_\varepsilon := (H_\varepsilon)^C$. Then $\Theta^*_\varepsilon \subseteq \Theta^*$ for all $\varepsilon > 0$ and $\Theta^*_\varepsilon \uparrow \Theta^*$ as $\varepsilon \to 0$. From the definition of $\mathscr{G}$ in (3.10) and from (3.11) we conclude that $\sup_{g\in\mathscr{G}} \|a_n g(0)\|_\infty \le a_n M \to 0$. Thus, for every given (small) $\varepsilon > 0$ there is an index N such that, firstly, for all $n \ge N$ holds:

$$\gamma \in \Theta^*_\varepsilon,\ g \in \mathscr{G} \quad \Rightarrow \quad \gamma_{g,n}(x_0) = \gamma + a_n g(0) \in \Theta^*.$$

Secondly, for each $s \in [-x_0, 1-x_0]$ we have

$$\begin{aligned}\|\gamma_{g,n}(x_0+s) - \gamma_{g,n}(x_0)\|_\infty &\overset{(1.3)}{=} a_n \|g(b_n s) - g(0)\|_\infty \\ &\overset{(3.10)}{\le} a_n (b_n |s|)^\rho \\ &\overset{(3.14)}{=} A|s|^\rho = u(s).\end{aligned}$$

Let us presume for a moment that n is fixed and sufficiently large. The product measure $P^{(n)}_{\gamma,g}$ only depends on the values $\gamma_{g,n}(x_{nj})$, which follows from the definition in (1.4). For all other arguments $\gamma_{g,n}(\cdot)$ may be changed without having impact on the resulting product measure. Hence, from the just shown properties we conclude that for each $\gamma_{g,n}$ with $\gamma \in \Theta^*_\varepsilon$ and $g \in \mathscr{G}$ there is a function $\xi \in \mathscr{F}$ (which may vary with n) such that $\gamma_{g,n}(x_0) = \xi(x_0)$ and $\gamma_{g,n}(x_{nj}) = \xi(x_{nj})$ for all j. In addition, the estimator $\hat{\xi}_n(x_0)$ can of course also be interpreted as an estimator for the parameter $\gamma_{g,n}(x_0)$ in the sequence of local models. Thus, we get

$$\begin{aligned}&\liminf_{n\to\infty} \sup_{\xi\in\mathscr{F}} P^{(n)}_\xi\{\hat{\xi}_n(x_0) - \xi(x_0) \in [-\delta a_n, \delta a_n]^C\} \\ &\ge \liminf_{n\to\infty} \sup_{\gamma\in\Theta^*_\varepsilon} \sup_{g\in\mathscr{G}} P^{(n)}_{\gamma,g}\{\hat{\xi}_n(x_0) - \gamma_{g,n}(x_0) \in [-\delta a_n, \delta a_n]^C\} \\ &\overset{(*)}{\ge} \sup_{\gamma\in\Theta^*_\varepsilon} \liminf_{n\to\infty} \sup_{g\in\mathscr{G}} P^{(n)}_{\gamma,g}\{\hat{\xi}_n(x_0) - \gamma_{g,n}(x_0) \in [-\delta a_n, \delta a_n]^C\} \\ &\ge \sup_{\gamma\in\Theta^*_\varepsilon} \inf_{\varrho} \sup_{g\in\mathscr{G}} \int \varrho\left(z, [g(0)-\delta, g(0)+\delta]^C\right)\, dQ_{\gamma,g}(z).\end{aligned} \tag{4.15}$$

The last inequality follows directly from Theorem 4.10. In order to prove $(*)$ we set

$$d_n(\gamma) := \sup_{g\in\mathscr{G}} P^{(n)}_{\gamma,g}\{\hat{\xi}_n(x_0) - \gamma_{g,n}(x_0) \in [-\delta a_n, \delta a_n]^C\}.$$

It is clear that the smallest accumulation point of the sequence $\sup_\gamma d_n(\gamma)$ is not smaller than the smallest accumulation point of any sequence $d_n(\gamma')$ with arbitrary, but fixed γ'. In other words, we have

$$\liminf_{n\to\infty} \sup_{\gamma} \; d_n(\gamma) \geq \liminf_{n\to\infty} \; d_n(\gamma')$$

for all γ'. Consequently,

$$\liminf_{n\to\infty} \sup_{\gamma} \; d_n(\gamma) \geq \sup_{\gamma} \liminf_{n\to\infty} \; d_n(\gamma),$$

and thus $(*)$ holds. Letting $\varepsilon \to 0$ in (4.15) implies the assertion of the theorem. ∎

According to this theorem a lower minimax risk bound in the limit experiment always yields a lower asymptotic bound on the minimax risk in the global model. The question of how to compute reasonable lower bounds in the limit experiment will be discussed in Sections 5.1.4 and 5.3.2. Finally, we derive an assertion on lower asymptotic minimax risk bounds in the restricted global model $\mathscr{F}_{\mathrm{H}}$.

4.12 Corollary. *Consider the restricted global model $\mathscr{F}_{\mathrm{H}}(x_0,\Theta^*,u)$ from (3.9). Let $\hat{\xi}_n(x_0)$ be an arbitrary estimator for $\xi(x_0)$, and let the assumptions from Sections 3.1 and 3.3 hold. Then*

$$\begin{aligned}\liminf_{n\to\infty} \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} \; & P_\xi^{(n)}\{\hat{\xi}_n(x_0) - \xi(x_0) \in [-\delta a_n, \delta a_n]^C\} \\ & \geq \sup_{\gamma\in\Theta^*} \inf_{\varrho} \sup_{g\in\mathscr{K}} \int \varrho\left(z, [g(0)-\delta, g(0)+\delta]^C\right) \; dQ_{\gamma,g}(z),\end{aligned}$$

with $\mathscr{K} := \{g \in \mathscr{G}_{\mathrm{H}} : g$ is continuous on $\mathbb{R}\}$ and ϱ as in Theorem 4.10.

Proof. Suppose that $\gamma \in \Theta^*_\varepsilon$ and $g \in \mathscr{G}_{\mathrm{H}}$ are given. From the definition in (3.12) follows that for $x, y \in [0,1]$

$$\|\gamma_{g,n}(x) - \gamma_{g,n}(y)\|_\infty \leq u(|x-y|).$$

With this additional argument the proof is a simple copy of the proof to Theorem 4.11, replacing there $\mathscr{F}$ by $\mathscr{F}_{\mathrm{H}}$ and $\mathscr{G}$ by $\mathscr{K}$, respectively. ∎

4.13 Remark. The results from Theorems 4.10 and 4.11 correspond to those of Corollaries 2.1 and 2.2 in Drees (2001). For the proofs we mainly used properties of the functions from the spaces $\mathscr{F}$ and $\mathscr{G}$. Moreover, we exploited that the loss function under consideration is lower semi-continuous and level-compact in order to prove the central inequality of Theorem 4.10 (which is based on Theorem 2.19). Thus, similar results should be reached for general lower semi-continuous and level-compact loss functions—such as quadratic loss—by simple modifications of the above proofs.

CHAPTER 5

Assessing the limit experiment

Up to this point, we have proved that the local experiments converge weakly to a Gaussian limit experiment. We now consider the problem of estimating the *local parameter* $g(0)$ within this limit experiment, where $g \in \mathscr{G}$. We want to assess the corresponding minimax risk under zero-one loss. More precisely, we want to derive upper and lower bounds for the minimax risk. In Section 5.1 this is done for the special case of a one-dimensional parameter space, i.e. for the special case $\mathscr{G} = \mathscr{G}(K, \rho, 1)$. In Section 5.3 we expand these results to the general multivariate case in which $\mathscr{G} = \mathscr{G}(K, \rho, d)$ with $d \geq 1$. To distinguish these two cases we shortly speak of the "one-dimensional" (or "univariate") and the "multi-dimensional" (or "multivariate") limit experiment, respectively. Section 5.2 gives some supplementary results on the theory of hardest linear submodels and may be omitted in first reading. We conclude this chapter with a discussion of the results and the used methodology.

5.1 Minimax risk bounds in the univariate limit experiment

Throughout this section we restrict ourselves to the case of a univariate parameter space $\Theta \subseteq \mathbb{R}$. Moreover, in this section $\gamma \in \Theta'$ (defined according to (3.7)) always denotes an arbitrary, but fixed centre of localisation, for which we consider the corresponding limit experiment

$$(C^1([-K,K]), \mathscr{C}^1([-K,K]), \{Q_{\gamma,g} : g \in \mathscr{G}\}),$$

where $\mathscr{G} = \mathscr{G}(K,\rho,1)$ as in (3.10). Within this model we discuss the problem of estimating the local parameter $g(0)$. We derive upper and lower bounds for the minimax risk under zero-one loss $\mathbb{1}_{[-\delta,\delta]^C}$. Note that in the one-dimensional case the information matrix $\mathcal{I}_\theta$ and the Cholesky matrix $\mathcal{C}_\gamma$ are simply positive constants.

5.1.1 Estimating Gaussian means

First we show that the problem of estimating the local parameter within the limit experiment can be reduced to the problem of estimating the mean in certain Gaussian location models. To this end, let us presume for a moment that the *local function*, which is an element of the local function space $\mathscr{G}$, is not completely unknown. Suppose we have the information that the local function is of the form ζg, with known $g \in \mathscr{G}$ but unknown parameter $\zeta \in [-1,1]$ which thus has to be estimated. Hence, for given $g \in \mathscr{G}$ we consider the *(centrosymmetric) linear submodel*

$$\{Q_{\gamma,\zeta g} : \zeta \in [-1,1]\}. \tag{5.1}$$

Within this model we want to assess the minimax risk for the estimation of the parameter $\zeta g(0)$, given the loss function $\mathbb{1}_{[-\delta,\delta]^C}$. If $g(0) = 0$, then each estimator $\hat{\zeta}$ for ζ fulfils $|\hat{\zeta} g(0) - \zeta g(0)| = 0 < \delta$, leading to a minimax risk of zero. Consequently, in the following we can restrict ourselves to the case $g(0) \neq 0$, and because of the symmetry of the space $\mathscr{G}$ we can even assume $g(0) > 0$. Using the notation for the stochastic integrals from Section 4.1—adapted to the here considered univariate case—we set

$$Y_g := \frac{\int \mathcal{C}_\gamma g \, d\mathrm{x}}{\|\mathcal{C}_\gamma g\|_{\lambda,2}}. \tag{5.2}$$

Then $\mathscr{L}(Y_g \mid Q_{\gamma,0}) = \mathscr{N}(0,1)$, and

$$\mathscr{L}(Y_g \mid Q_{\gamma,\zeta g}) = \mathscr{N}(\vartheta,1), \quad \vartheta := \zeta \|\mathcal{C}_\gamma g\|_{\lambda,2}.$$

The fact that $\zeta \in [-1,1]$ leads to the additional constraint

$$|\vartheta| \leq \|\mathcal{C}_\gamma g\|_{\lambda,2} =: \tau_g. \tag{5.3}$$

Given an estimator $\hat{\vartheta}$ for ϑ, then

$$\hat{\zeta} := \hat{\vartheta}/\|\mathcal{C}_\gamma g\|_{\lambda,2} \tag{5.4}$$

yields an estimator for ζ. The resulting estimator $\hat{\zeta}g(0)$ satisfies

$$|\hat{\zeta}g(0) - \zeta g(0)| > \delta \;\Leftrightarrow\; |\hat{\vartheta} - \vartheta| > \delta_g, \tag{5.5}$$

where

$$\delta_g := \frac{\delta\|\mathcal{C}_\gamma g\|_{\lambda,2}}{g(0)} = \frac{\delta}{g(0)}\tau_g. \tag{5.6}$$

The primary task was to find a minimax estimator or an upper minimax risk bound for the estimation of $\zeta g(0)$ in the model (5.1), given the loss function $\mathbb{1}_{[-\delta,\delta]^C}$. As a consequence of (5.5) this is equivalent to finding such an estimate (and an upper bound, respectively) for the estimation of the parameter ϑ in the model

$$\{\mathscr{N}(\vartheta, 1) : \vartheta \in [-\tau_g, \tau_g]\}, \tag{5.7}$$

given the loss function $\mathbb{1}_{[-\delta_g,\delta_g]^C}$. In other words, it suffices to consider the problem of estimating the mean in the Gaussian model (5.7), which we want to assess in the following. All the results found for this model then easily carry over to the model (5.1).

In the case $\vartheta \in \mathbb{R}$, a minimax estimator for ϑ with respect to the loss function $\mathbb{1}_{[-\delta_g,\delta_g]^C}$ is given by Y_g (which follows e.g. from Strasser, 1985, Theorem 38.22). However, the constraint $|\vartheta| \leq \tau_g$—i.e. the additional information on the location of ϑ—somewhat complicates the situation. In the following, we consider so-called *affine* estimators, i.e. estimators which take the form

$$c_g Y_g + d_g.$$

The minimax risk in the model (5.7) is given by

$$\inf_{\hat{\vartheta}} \sup_{|\vartheta|\leq\tau_g} \mathscr{N}_{(\vartheta,1)}\{|\hat{\vartheta} - \vartheta| > \delta_g\},$$

where the infimum is taken over all estimators $\hat{\vartheta}$ for ϑ. Analogously, we define the *minimax affine risk* by

$$\inf_{\hat{\vartheta}\text{ affine}} \sup_{|\vartheta|\leq\tau_g} \mathscr{N}_{(\vartheta,1)}\{|\hat{\vartheta} - \vartheta| > \delta_g\}.$$

An affine estimator $\hat{\vartheta}^*$ whose maximal risk equals the minimax affine risk is called a *minimax affine estimator*.

5.1 Theorem. *Consider the model (5.7) with the loss function* $\mathbb{1}_{[-\delta_g,\delta_g]^C}$. *Then the minimax affine risk is given by*

$$R^*_{\text{aff}}(g) = \left[\Phi\left(-\frac{\tau_g+\delta_g}{c_g}+\tau_g\right)+\Phi\left(\frac{\tau_g-\delta_g}{c_g}-\tau_g\right)\right] \mathbb{1}\{g(0)>\delta\}, \tag{5.8}$$

where

$$\begin{aligned} c_g &= \left(\frac{1}{2}+\left[\frac{1}{4}+\frac{1}{2\delta_g\tau_g}\log\frac{\tau_g+\delta_g}{\tau_g-\delta_g}\right]^{1/2}\right)^{-1} \mathbb{1}\{g(0)>\delta\} \\ &= \left(\frac{1}{2}+\left[\frac{1}{4}+\frac{g(0)}{2\delta\|\mathcal{C}_\gamma g\|^2_{\lambda,2}}\log\frac{g(0)+\delta}{g(0)-\delta}\right]^{1/2}\right)^{-1} \mathbb{1}\{g(0)>\delta\}. \end{aligned}$$

Moreover, c_gY_g *is a minimax affine estimator for* ϑ.

Addendum. *Let the loss function* $\mathbb{1}_{[-\delta,\delta]^C}$ *be given. Then*

$$c_gY_g\,g(0)/\|\mathcal{C}_\gamma g\|_{\lambda,2}$$

is a minimax affine estimator for the parameter $\zeta g(0)$ *within the linear submodel* $\{Q_{\gamma,\zeta g} : \zeta\in[-1,1]\}$, *the minimax affine risk is given by* $R^*_{\text{aff}}(g)$.

Proof. In the case $\tau_g \le \delta_g$—which is equivalent to $\mathbb{1}\{g(0)>\delta\}=0$—the minimax affine risk is equal to zero, as stated in the assertion. Otherwise, if $\tau_g>\delta_g$ the assertion of the theorem coincides with Theorem 1 in Drees (1999). To see this, one simply has to replace there τ with τ_g, x with δ_g, c with c_g and Y with Y_g. Furthermore, the quantity $\Psi_{A,\tau}(x)$ from Drees' paper equals $1-R^*_{\text{aff}}(g)$, and thus

$$\begin{aligned} R^*_{\text{aff}}(g) &= 1-\left[\Phi\left(\frac{\tau_g+\delta_g}{c_g}-\tau_g\right)-\Phi\left(\frac{\tau_g-\delta_g}{c_g}-\tau_g\right)\right] \\ &= \Phi\left(-\frac{\tau_g+\delta_g}{c_g}+\tau_g\right)+\Phi\left(\frac{\tau_g-\delta_g}{c_g}-\tau_g\right). \end{aligned}$$

The assertion of the addendum follows from (5.4) and the equivalence in (5.5). ∎

5.1.2 The theory of hardest linear submodels

Which benefits do we actually get from the above examinations of the minimax affine risk in linear submodels of type (5.1)? To give an answer to this question, we have to do a short excursion into the theory of the estimation of linear functionals. Suppose we are given an observation y that satisfies

$$y = Ax + z, \tag{5.9}$$

where x is an element of a convex subset $\mathscr{X} \subseteq \mathbb{L}_2(\mathbb{R}, \mathbb{B}, \lambda)$, A is a linear operator, and z is a random variable, interpreted as noise. We are interested in the value of $L(x)$, where $L\colon \mathscr{X} \to \mathbb{R}$ is some linear operator. Suppose we are given a loss function Λ as in Example 2.12. Let the corresponding minimax risk with respect to this loss function be denoted $R^*(\mathscr{X})$, i.e. we have

$$R^*(\mathscr{X}) \;=\; \inf_{\hat{L}} \sup_{x \in \mathscr{X}} \mathrm{E}[\Lambda(\hat{L}(y) - L(x))],$$

where the infimum is taken with respect to all possible estimators for $L(x)$. In the following, we focus on affine estimators $\hat{L}$, we denote the corresponding *minimax affine risk* by

$$R^*_{\text{aff}}(\mathscr{X}) \;=\; \inf_{\hat{L}\text{ affine}} \sup_{x \in \mathscr{X}} \mathrm{E}[\Lambda(\hat{L}(y) - L(x))].$$

Let now some elements $x_1, x_2 \in \mathscr{X}$ be given, and let

$$[x_1, x_2] := \{tx_1 + (1-t)x_2 : t \in [0,1]\}$$

denote the set of convex combinations of x_1 and x_2. We consider the minimax affine risk in the corresponding linear submodel

$$R^*_{\text{aff}}([x_1, x_2]) \;:=\; \inf_{\hat{L}\text{ affine}} \sup_{x \in [x_1,x_2]} \mathrm{E}[\Lambda(\hat{L}(y) - L(x))]. \tag{5.10}$$

Trivially, we have $R^*_{\text{aff}}(\mathscr{X}) \geq R^*_{\text{aff}}([x_1, x_2])$ for each choice of x_1 and x_2, which is due to the simple fact that in no case the loss will increase if we are given the additional information that $x \in [x_1, x_2]$, and therefore

$$R^*_{\text{aff}}(\mathscr{X}) \;\geq\; \sup\{R^*_{\text{aff}}([x_1, x_2]) : x_1, x_2 \in \mathscr{X}\}.$$

Donoho (1994) and Donoho and Liu (1991) show that for certain loss functions Λ (such as quadratic loss) and with some additional regularity conditions the above formula is an equality, and that the supremum is attained (and thus a maximum). That is, under certain conditions

$$\begin{aligned} R^*_{\mathrm{aff}}(\mathscr{X}) &= \sup\{R^*_{\mathrm{aff}}([x_1, x_2]) : x_1, x_2 \in \mathscr{X}\} \qquad (5.11) \\ &= R^*_{\mathrm{aff}}([x_1^*, x_2^*]) \end{aligned}$$

holds for some $x_1^*, x_2^* \in \mathscr{X}$. The linear submodel corresponding to the family $[x_1^*, x_2^*]$ is also called a *hardest linear submodel* (also: *least favourable linear submodel*).

5.1.3 An upper minimax risk bound

According to arguments of Donoho (1994) (cf. there Section 9), our white noise model

$$\mathrm{x} = \int_{-K}^{\cdot} \mathcal{C}_\gamma g(s)\, ds + W, \quad g \in \mathscr{G} \qquad (5.12)$$

from Section 4.1 can be identified with a model of the type (5.9). In this case the linear operator A from (5.9) is simply a constant factor. The space $\mathscr{X}$ and the linear map L from the above description correspond in our setting to the space $\mathscr{G}$ and to the linear map $g \mapsto g(0)$, respectively. The affine estimators $\hat{T}$ for $g(0)$ are of the general form $\hat{T}(\mathrm{x}) = e + \int \Psi \, d\mathrm{x}$ (see Donoho and Liu, 1991, for details). As loss function we take $\Lambda = \mathbb{1}_{[-\delta,\delta]^C}$.

For fixed $g \in \mathscr{G}$ let $R^*_{\mathrm{aff}}(g)$ be the minimax affine risk with respect to the loss function $\mathbb{1}_{[-\delta,\delta]^C}$ for the estimation of $g(0)$ in the submodel $\{Q_{\gamma,\zeta g} : \zeta \in [-1,1]\}$. Let $R^*_{\mathrm{aff}}(\mathscr{G})$ denote the minimax affine risk in the complete model $\{Q_{\gamma,g} : g \in \mathscr{G}\}$. It is clear that $\mathscr{G}$ is convex. Hence, it can be concluded from the arguments of Section 9 and Theorem 1 in Donoho (1994) that

$$R^*_{\mathrm{aff}}(\mathscr{G}) \;=\; \sup\{R^*_{\mathrm{aff}}(g) : g \in \mathscr{G}\}, \qquad (5.13)$$

which is a special case of (5.11).

5.2 Remark. The supremum in (5.13) is built over the minimax affine risks corresponding to the linear submodels $\{Q_{\gamma,\zeta g} : \zeta \in [-1,1]\}$, with $g \in \mathscr{G}$. Apparently, this is a smaller class of submodels than that considered in (5.11). Furthermore, the results from Donoho (1994) are formulated for a quadratic loss function (amongst others), yet they do not take into account zero-one loss functions. Therefore, some additional arguments are needed in order to prove that (5.13) indeed holds true. These arguments are assembled in Section 5.2.

The main goal of the remaining part of this Section is to show that there even is a function g^* such that the supremum in (5.13) is attained, i.e.

$$R^*_{\text{aff}}(g^*) \;=\; \sup\{R^*_{\text{aff}}(g) : g \in \mathscr{G}\} \;=\; R^*_{\text{aff}}(\mathscr{G}). \tag{5.14}$$

In other words, we want to find a hardest linear submodel. If we are able to find such a g^*, then

$$R^*(\mathscr{G}) \leq R^*_{\text{aff}}(\mathscr{G}) \overset{(5.13)}{=} R^*_{\text{aff}}(g^*),$$

where $R^*(\mathscr{G})$ denotes the minimax risk for the estimation of $g(0)$ with respect to the loss function $\mathbb{1}_{[-\delta,\delta]^C}$. The minimax affine risk $R^*_{\text{aff}}(g^*)$ from the hardest submodel thus yields an upper bound for $R^*(\mathscr{G})$. Note that the above inequality is quite trivial, because the class of affine estimators for the parameter $g(0)$ in the (complete) limit experiment $\{Q_{\gamma,g} : g \in \mathscr{G}\}$ is a subset of the class of all estimators for $g(0)$. Therefore the minimax *affine* risk can never be smaller than the minimax risk—which takes into account *all* possible estimators.

Consequently, it remains to assign a function g^* which defines a least favourable linear submodel. To this end, we first consider the subspace

$$\mathscr{G}^{(\alpha)} := \{g \in \mathscr{G} : \|g\|_{\lambda,2} = \alpha\} \subseteq \mathscr{G},$$

with some fixed $\alpha > 0$. We investigate the minimax affine risks in the corresponding linear submodels $\{Q_{\gamma,\zeta g} : \zeta \in [-1,1]\}$. For symmetry reasons, we can assume without loss of generality that $g(0) > 0$. According to the arguments from Section 5.1.1 this is equivalent to investigating the minimax risk in the models

$$\{\mathscr{N}(\vartheta, 1) : \vartheta \in [-\tau_g, \tau_g]\},$$

with τ_g and δ_g defined according to (5.3) and (5.6), respectively. As g ranges over $\mathscr{G}^{(\alpha)}$, then clearly

$$\tau_g = \|\mathcal{C}_\gamma g\|_{\lambda,2} = \mathcal{C}_\gamma \alpha,$$

which means that τ_g is constant for whatever choice of $g \in \mathscr{G}^{(\alpha)}$. Furthermore, $\delta_g = \tau_g \delta / g(0)$ is strictly decreasing in $g(0)$. This implies that $R^*_{\text{aff}}(g)$ is strictly increasing in $g(0)$ (for g ranging over $\mathscr{G}^{(\alpha)}$). These arguments lead to the following conclusion:

5.3 Proposition. *Only such functions $g \in \mathscr{G}^{(\alpha)}$ can describe a least favourable submodel for which the value $g(0)$ is maximal.*

With this result and the additional growth condition from (3.10) one can now easily conclude that a function $g \in \mathscr{G}$ to define a hardest submodel must be of the general form

$$g_\beta(s) := (\beta - |s|^\rho)^+. \tag{5.15}$$

Comparing all functions $g \in \mathscr{G}$ that have constant norm it becomes clear that the functions of this specific type (5.15) have the maximal value at the origin. Hence, a hardest linear submodel exists, and thus (5.14) holds, if the function $\beta \mapsto R^*_{\text{aff}}(g_\beta)$ has a maximum, say β^*, for which the corresponding function g_{β^*} lies in $\mathscr{G}$. This is indeed the case, as is shown in the following theorem.

5.4 Theorem. *Consider the loss function $\mathbb{1}_{[-\delta,\delta]^C}$. Let $\beta > 0$ be the unique positive solution of the equation*

$$\frac{\beta^{1+1/\rho}}{\log \frac{\beta+\delta}{\beta-\delta}} = \frac{\rho+1}{2\mathcal{I}_\gamma \delta}, \tag{5.16}$$

and let g_β be defined according to (5.15). Then

$$\hat{g}(0) = \frac{1}{2\mathcal{C}_\gamma}\left(1 + \frac{1}{\rho}\right) \beta^{-(1+1/\rho)} \int g_\beta \, \mathrm{dx}$$

is a minimax affine estimator for $g(0)$ in the model $\{Q_{\gamma,g} : g \in \mathscr{G}\}$. The corresponding minimax affine risk is given by

$$\begin{aligned}\Phi\left(-\frac{\mathcal{C}_\gamma \beta^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}\left[1 + \frac{(2\rho+1)\delta}{\beta}\right]\right)& \\ + \Phi\left(\frac{\mathcal{C}_\gamma \beta^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}\left[1 - \frac{(2\rho+1)\delta}{\beta}\right]\right)&.\end{aligned} \tag{5.17}$$

Proof. The proof mainly follows the lines of the proof of Theorem 3.1 in Drees (2001). With the above examinations it suffices for the definition of a hardest linear submodel to restrict ourselves to functions of the form g_β according to (5.15), with $\beta > \delta$. The minimax affine risk in the corresponding submodels is then a function of β. According to Theorem 5.1 it is given by

$$\begin{aligned} R^*_{\mathrm{aff}}(\beta) &:= R^*_{\mathrm{aff}}(g_\beta) \\ &= \Phi\left(-\frac{\tau_\beta+\delta_\beta}{c_\beta}+\tau_\beta\right)+\Phi\left(\frac{\tau_\beta-\delta_\beta}{c_\beta}-\tau_\beta\right), \end{aligned} \tag{5.18}$$

where $\tau_\beta := \tau_{g_\beta}$, $\delta_\beta := \delta_{g_\beta}$ and $c_\beta := c_{g_\beta}$. In order to determine a β^* and thus a function g_{β^*} for a hardest linear submodel we maximise R^*_{aff} with respect to β. According to Theorem 5.1 c_β is defined by

$$c_\beta = \left(\frac{1}{2}+\left[\frac{1}{4}+\frac{1}{2\tau_\beta\delta_\beta}\log\frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta}\right]^{1/2}\right)^{-1}, \tag{5.19}$$

and thus

$$\begin{aligned} & \frac{1}{c_\beta^2}-\frac{1}{c_\beta} = \frac{1}{2\tau_\beta\delta_\beta}\log\frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta} \\ \Rightarrow\ & \frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta} = \exp\left[2\frac{\tau_\beta\delta_\beta}{c_\beta}\left(\frac{1}{c_\beta}-1\right)\right] \\ \Rightarrow\ & \frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta} = \varphi\left(\frac{\tau_\beta-\delta_\beta}{c_\beta}-\tau_\beta\right)\Big/\varphi\left(\frac{\tau_\beta+\delta_\beta}{c_\beta}-\tau_\beta\right). \end{aligned} \tag{5.20}$$

Differentiation of (5.18) and the symmetry of φ hence yield

$$\begin{aligned} \frac{d}{d\beta}R^*_{\mathrm{aff}}(\beta) &= \varphi\left(-\frac{\tau_\beta+\delta_\beta}{c_\beta}+\tau_\beta\right)\frac{d}{d\beta}\left(-\frac{\tau_\beta+\delta_\beta}{c_\beta}+\tau_\beta\right) \\ &\quad +\varphi\left(\frac{\tau_\beta-\delta_\beta}{c_\beta}-\tau_\beta\right)\frac{d}{d\beta}\left(\frac{\tau_\beta-\delta_\beta}{c_\beta}-\tau_\beta\right) \\ &\overset{(5.20)}{=} \varphi\left(\frac{\tau_\beta+\delta_\beta}{c_\beta}-\tau_\beta\right)\left[\frac{d}{d\beta}\left(-\frac{\tau_\beta+\delta_\beta}{c_\beta}+\tau_\beta\right)\right. \\ &\quad \left.+\frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta}\cdot\frac{d}{d\beta}\left(\frac{\tau_\beta-\delta_\beta}{c_\beta}-\tau_\beta\right)\right]. \end{aligned}$$

Note that

$$\|g_\beta\|^2_{\lambda,2} = \int\left((\beta-|s|^\rho)^+\right)^2 ds = \frac{4\rho^2\beta^{2+1/\rho}}{(\rho+1)(2\rho+1)}, \tag{5.21}$$

and therefore

$$\begin{aligned} \tau_\beta &\overset{(5.3)}{=} \mathcal{C}_\gamma \|g_\beta\|_{\lambda,2} = \mathcal{C}_\gamma \frac{2\rho\beta^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}, \\ \delta_\beta &\overset{(5.6)}{=} \mathcal{C}_\gamma \|g_\beta\|_{\lambda,2} \frac{\delta}{\beta} = \mathcal{C}_\gamma \frac{\delta 2\rho\beta^{1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}. \end{aligned} \tag{5.22}$$

With these expressions one further concludes that

$$\begin{aligned} \frac{d}{d\beta}\tau_\beta &= \tau_\beta \left(\frac{2\rho+1}{2\rho}\right)\frac{1}{\beta}, \\ \frac{d}{d\beta}\left(\frac{\tau_\beta \pm \delta_\beta}{c_\beta}\right) &= \frac{\mp\delta_\beta/\beta + (1 \pm \delta/\beta)[1+1/(2\rho)]\tau_\beta/\beta}{c_\beta} \\ &\quad - \frac{(\tau_\beta \pm \delta_\beta)\frac{d}{d\beta}c_\beta}{c_\beta^2}, \end{aligned} \tag{5.23}$$

and some additional calculations yield

$$\begin{aligned} &\frac{d}{d\beta}\left(-\frac{\tau_\beta+\delta_\beta}{c_\beta} + \tau_\beta\right) + \frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta}\cdot\frac{d}{d\beta}\left(\frac{\tau_\beta-\delta_\beta}{c_\beta} - \tau_\beta\right) \\ &= \left(1 - \frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta}\right)\cdot\frac{d}{d\beta}\tau_\beta - \frac{d}{d\beta}\left(\frac{\tau_\beta+\delta_\beta}{c_\beta}\right) \\ &\quad + \frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta}\cdot\frac{d}{d\beta}\left(\frac{\tau_\beta-\delta_\beta}{c_\beta}\right) \\ &\overset{(5.23)}{=} \left(1 - \frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta}\right)\tau_\beta\left(\frac{2\rho+1}{2\rho}\right)\frac{1}{\beta} \\ &\quad - \left(\frac{-\delta_\beta/\beta + (1+\delta/\beta)[1+1/(2\rho)]\tau_\beta/\beta}{c_\beta} - \frac{(\tau_\beta+\delta_\beta)\frac{d}{d\beta}c_\beta}{c_\beta^2}\right) \\ &\quad + \frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta}\left(\frac{\delta_\beta/\beta + (1-\delta/\beta)[1+1/(2\rho)]\tau_\beta/\beta}{c_\beta}\right. \\ &\qquad\qquad \left. - \frac{(\tau_\beta-\delta_\beta)\frac{d}{d\beta}c_\beta}{c_\beta^2}\right) \dots \end{aligned}$$

$$
\begin{aligned}
&= -\frac{2\tau_\beta\delta_\beta}{\tau_\beta-\delta_\beta}\left(\frac{2\rho+1}{2\rho}\right)\frac{1}{\beta} \\
&\quad + \underbrace{\frac{\tau_\beta-\delta_\beta}{\tau_\beta-\delta_\beta}}_{=1}\left(\frac{\delta_\beta/\beta-(1+\delta/\beta)[1+1/(2\rho)]\tau_\beta/\beta}{c_\beta}\right) \\
&\quad + \frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta}\left(\frac{\delta_\beta/\beta+(1-\delta/\beta)[1+1/(2\rho)]\tau_\beta/\beta}{c_\beta}\right) \\
&= -\frac{2\tau_\beta\delta_\beta}{\tau_\beta-\delta_\beta}\left(\frac{2\rho+1}{2\rho}\right)\frac{1}{\beta}+\frac{1}{\tau_\beta-\delta_\beta}\cdot\frac{2\tau_\beta\delta_\beta/\beta}{c_\beta} \\
&= \frac{2\tau_\beta\delta_\beta}{\beta(\tau_\beta-\delta_\beta)}\left[\frac{1}{c_\beta}-\frac{2\rho+1}{2\rho}\right].
\end{aligned}
$$

Plugging this into the above expression for $\frac{d}{d\beta}R^*_{\mathrm{aff}}(\beta)$ we finally get

$$
\begin{aligned}
&\frac{d}{d\beta}R^*_{\mathrm{aff}}(\beta) \\
&\quad = \varphi\left(\frac{\tau_\beta+\delta_\beta}{c_\beta}-\tau_\beta\right)\frac{2\tau_\beta\delta_\beta}{\beta(\tau_\beta-\delta_\beta)}\left[\frac{1}{c_\beta}-\frac{2\rho+1}{2\rho}\right],
\end{aligned}
\tag{5.24}
$$

and thus:

$$
\begin{aligned}
\frac{d}{d\beta}R^*_{\mathrm{aff}}(\beta)=0 \quad &\Leftrightarrow \quad \frac{1}{c_\beta}=\frac{2\rho+1}{2\rho} \\
&\overset{(5.19)}{\Leftrightarrow} \quad \left[\frac{1}{4}+\frac{1}{2\tau_\beta\delta_\beta}\log\frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta}\right]^{1/2}=\frac{\rho+1}{2\rho} \\
&\Leftrightarrow \quad \frac{1}{2\tau_\beta\delta_\beta}\log\frac{\tau_\beta+\delta_\beta}{\tau_\beta-\delta_\beta}=\frac{2\rho+1}{4\rho^2} \\
&\Leftrightarrow \quad \frac{\beta^{1+1/\rho}}{\log[(\beta+\delta)/(\beta-\delta)]}=\frac{\rho+1}{2\mathcal{I}_\gamma\delta}.
\end{aligned}
$$

The last equivalence follows from inserting the expressions for τ_g, δ_g from (5.22). This last equation possesses a unique solution $\beta^*\in(\delta,\infty)$. This follows from the fact that the map

$$
f(\beta):=\frac{\beta^{1+1/\rho}}{\log[(\beta+\delta)/(\beta-\delta)]},\quad \beta\in(\delta,\infty)
$$

is strictly increasing and tends to 0 as $\beta\downarrow\delta$ and to ∞ as $\beta\to\infty$. The monotonicity of f is a consequence of the observation that $x\mapsto\log[(x+\delta)/(x-\delta)]$ is strictly decreasing on (δ,∞). (Note that this is the map f that we considered in Section 3.5.1 in order to specify the constant K.) Furthermore, if we plug the definitions of τ_β and δ_β into (5.19) we get

$$\frac{1}{c_\beta} = \frac{1}{2} + \left[\frac{1}{4} + \frac{\beta}{2\delta \mathcal{C}_\gamma^2 \|g_\beta\|_{\lambda,2}^2} \log \frac{\beta+\delta}{\beta-\delta}\right]^{1/2}, \quad \text{for} \quad \beta > \delta.$$

Again using the fact that $\beta \mapsto \log[(\beta+\delta)/(\beta-\delta)]$ is strictly decreasing in β, as well as the quotient $\beta/\|g_\beta\|_{\lambda,2}^2$ (which is clear from (5.21)), we see that $1/c_\beta$ is strictly decreasing in β. Thus, the derivative in (5.24) is positive for $\beta < \beta^*$ and negative for $\beta > \beta^*$, and therefore β^* is a unique maximum of $R^*_{\text{aff}}(\beta)$ (see Königsberger, 2001, p. 146). Note that by the specification of the constant K it is assured that the function g_{β^*} is included in $\mathscr{G} = \mathscr{G}(K, \rho, 1)$. Therefore, according to the arguments preceding this theorem, g_{β^*} defines a hardest linear submodel, such that (5.14) holds with $g^* = g_{\beta^*}$. Moreover, from the addendum to Theorem 5.1 we see that

$$\hat{g}(0) := c_{\beta^*} Y_{g_{\beta^*}}\, g_{\beta^*}(0)/\|\mathcal{C}_\gamma g_{\beta^*}\|_{\lambda,2},$$

is a minimax affine estimator for the parameter $\zeta g_{\beta^*}(0)$ in the corresponding submodel, and thus for the parameter $g(0)$ in the complete limit experiment $\{Q_{\gamma,g} : g \in \mathscr{G}\}$, too. Some additional calculations (in which we shortly write β instead of β^*) yield

$$\begin{aligned}
\hat{g}(0) &= \frac{\beta}{\|\mathcal{C}_\gamma g_\beta\|_{\lambda,2}}\, c_\beta Y_{g_\beta} \\
&= \frac{\beta}{\mathcal{C}_\gamma \|g_\beta\|_{\lambda,2}^2} \cdot \frac{2\rho}{2\rho+1} \int g_\beta \, d\mathrm{x} \\
&\overset{(5.21)}{=} \frac{(\rho+1)(2\rho+1)\beta}{4\mathcal{C}_\gamma \rho^2 \beta^{2+1/\rho}} \cdot \frac{2\rho}{2\rho+1} \int g_\beta \, d\mathrm{x} \\
&= \frac{1}{2\mathcal{C}_\gamma}\left(1+\frac{1}{\rho}\right) \beta^{-(1+1/\rho)} \int g_\beta \, d\mathrm{x}.
\end{aligned}$$

Therefore, $\hat{g}(0)$ has the same form as stated in the theorem. Moreover, we obtain

$$\begin{aligned}
\frac{\tau_\beta \pm \delta_\beta}{c_\beta} - \tau_\beta &= \frac{2\rho+1}{2\rho}\left(1 \pm \frac{\delta}{\beta}\right)\tau_\beta - \tau_\beta \\
&= \frac{2\rho+1}{2\rho}\left(1 \pm \frac{\delta}{\beta} - \frac{2\rho}{2\rho+1}\right)\tau_\beta \\
&\overset{(5.22)}{=} \left(\frac{1}{2\rho} \pm \frac{(2\rho+1)\delta}{2\rho\beta}\right)\frac{\mathcal{C}_\gamma 2\rho\beta^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}} \\
&= \left(1 \pm \frac{(2\rho+1)\delta}{\beta}\right)\frac{\mathcal{C}_\gamma \beta^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}.
\end{aligned}$$

Plugging this into (5.8) yields the minimax affine risk for the hardest linear submodel and thus for the complete model. It is easily verified that this coincides with the expression from (5.17). This concludes the proof of the theorem. ■

5.5 Remark. Using the definitions of τ_g and δ_g, the expressions $R^*_{\mathrm{aff}}(\beta)$ and c_β from (5.18) and (5.19), respectively, can also be written as

$$\begin{aligned}
R^*_{\mathrm{aff}}(\beta) = \Bigg[\Phi\bigg(&-\frac{\mathcal{C}_\gamma\|g_\beta\|_{\lambda,2}(1+\delta/\beta)}{c_\beta} + \mathcal{C}_\gamma\|g_\beta\|_{\lambda,2}\bigg) \\
&+ \Phi\bigg(\frac{\mathcal{C}_\gamma\|g_\beta\|_{\lambda,2}(1-\delta/\beta)}{c_\beta} - \mathcal{C}_\gamma\|g_\beta\|_{\lambda,2}\bigg)\Bigg]\mathbb{1}\{\beta > \delta\}
\end{aligned}$$

and

$$c_\beta = \left(\frac{1}{2} + \left[\frac{1}{4} + \frac{\beta}{2\delta\mathcal{C}_\gamma^2\|g_\beta\|^2_{\lambda,2}}\log\frac{\beta+\delta}{\beta-\delta}\right]^{1/2}\right)^{-1}\mathbb{1}\{\beta > \delta\}.$$

This specific representation will turn out helpful in later examinations for the multivariate parameter case.

5.6 Remark. In Section 3.5.1 we provided an approach that should assure that K is sufficiently large such that specific functions of the type g_β are included in $\mathscr{G} = \mathscr{G}(K, \rho, 1)$. With regard to the proof of Theorem 5.4 it is now evident why this effort was made: by choosing K according to the approach from Section 3.5, it is guaranteed that a hardest linear submodel exists.

5.1.4 A lower minimax risk bound

So far we have derived an upper minimax risk bound for the estimation of the parameter $g(0)$ within the univariate limit experiment $(C^1([-K,K]), \mathscr{C}^1([-K,K]), \{Q_{\gamma,g} : g \in \mathscr{G}\})$. We now want to derive a suitable lower bound for the minimax risk. Since we are here considering the case of a univarate parameter the used methods are exactly the same as those in Drees (2001). For the sake of completeness, we repeat the main ideas from that paper.

Assessing the linear submodels. Again, for given $g \in \mathscr{G}$ we first consider the corresponding linear submodel

$$\{Q_{\gamma,\zeta g} : \zeta \in [-1,1]\}$$

from (5.1). The minimax risk with respect to the loss function $\mathbb{1}_{[-\delta,\delta]^C}$ for the estimation of the unknown parameter $\zeta g(0)$ in this submodel is denoted $R^*(g)$. Note that in contrast to Sections 5.1.1 and 5.1.3—where we investigated the minimax *affine* risk $R^*_{\text{aff}}(g)$—we now allow for arbitrary, not necessarily affine estimators for ζ. For symmetry reasons we can (and will) in the following assume without loss of generality that $g(0) > 0$. Following the arguments from Section 5.1.1, $R^*(g)$ equals the minimax risk in the Gaussian model

$$\{\mathscr{N}(\vartheta, 1) : \vartheta \in \tau_g\}$$

from (5.7), with underlying loss function $\mathbb{1}_{[-\delta_g,\delta_g]^C}$, where $\tau_g = \|\mathcal{C}_\gamma g\|_{\lambda,2}$ and $\delta_g = \tau_g \delta / g(0)$ according to (5.3) and (5.6), respectively. In other words,

$$R^*(g) = \inf_{\hat{\vartheta}} \sup_{|\vartheta| \le \tau_g} \mathscr{N}_{(\vartheta,1)}\{|\hat{\vartheta} - \vartheta| > \delta_g\}. \tag{5.25}$$

This quantity is investigated in detail by Zeytinoglu and Mintz (1984), and also by Drees (1999)[1)]: following their results $R^*(g)$ depends on τ_g and δ_g only through the smallest integer l which is greater than or equal to the quotient $\tau_g/\delta_g = g(0)/\delta$. If $l = 1$, then $g(0) \le \delta$, and thus $R^*(g) = 0$. Otherwise, for the case when $l \ge 2$

[1)] Indeed, all of the following results are directly deduced from Drees (1999), p. 400 f., if there—similar to the proof of Theorem 5.1—τ is replaced with τ_g, x is replaced with δ_g and a_j with c_j. The quantity $\Psi_{N,\tau}$ from Drees' paper equals $1 - R^*(g)$.

the minimax risk is then given by

$$R^*(g) = \Phi(c_{\lfloor (l-1)/2 \rfloor} - \delta_g), \tag{5.26}$$

where $\lfloor z \rfloor$ denotes the largest integer less than or equal to z; the factor $c_{\lfloor (l-1)/2 \rfloor}$ is determined through a system of nonlinear equations given by

$$\begin{aligned} \Phi(-c_j - \delta_g) + \Phi(c_{j-1} - \delta_g) &= \Phi(c_{\lfloor (l-1)/2 \rfloor} - \delta_g), \\ \text{for} \quad j &= 1, \ldots, \lfloor (l-1)/2 \rfloor, \end{aligned}$$

with $c_0 := -c_1$ if l is odd and $c_0 := 0$ if l is even.

Computation of a lower bound. Let $R^*(\mathscr{G})$ denote the minimax risk (with respect to the loss function $\mathbb{1}_{[-\delta,\delta]^C}$) in the complete limit experiment $\{Q_{\gamma,g} : g \in \mathscr{G}\}$. Obviously, $R^*(\mathscr{G}) \geq R^*(g)$ for all g, and thus

$$R^*(\mathscr{G}) \geq \sup\{R^*(g) : g \in \mathscr{G}\}. \tag{5.27}$$

To derive a lower bound for $R^*(\mathscr{G})$ we thus only have to compute $R^*(g)$ and then maximise this expression with respect to g. The ideas for this maximisation are exactly the same as in Drees (2001). For symmetry reasons we can restrict ourselves to the case $g(0) > 0$. Furthermore, all of the arguments directly preceding Proposition 5.3 and Theorem 5.4 do also hold if the restriction to affine estimators is dropped. Hence, it is sufficient to consider on the right side of (5.27) only those functions g which are of the specific type $g = g_\beta$, defined according to (5.15). To this end, we first fix an integer $l \geq 2$ and consider all values β such that

$$(l-1)\delta < \beta \leq l\delta.$$

Then equation (5.26) yields

$$R^*(g_\beta) = \Phi(c_{\lfloor (l-1)/2 \rfloor} - \delta_{g_\beta}),$$

with $\delta_{g_\beta} = \delta \|\mathcal{C}_\gamma g_\beta\|_{\lambda,2}/\beta$. Because of (5.22) δ_{g_β} is strictly increasing in β, and in combination with some arguments from the proof of Theorem 2 in Drees (1999), this implies that $R^*(g_\beta)$ is maximised as $\beta \downarrow (l-1)\delta$. Therefore,

$$\sup_{l \geq 2} \lim_{\beta \downarrow (l-1)\delta} R^*(g_\beta)$$

yields a reasonable lower bound for $R^*(\mathscr{G})$, the minimax risk corresponding to the problem of estimating the local parameter $g(0)$ in the univariate limit experiment $\{Q_{\gamma,g} : g \in \mathscr{G}\}$, when the loss function $\mathbb{1}_{[-\delta,\delta]^C}$ is given.

In practice, this lower bound has of course to be calculated numerically. One might do this in two steps, firstly computing $\lim_{\beta\downarrow(l-1)\delta} R^*(g_\beta)$ for each $l = 2, 3, \ldots$, and, secondly, maximising the resulting expressions with respect to l. Combining these arguments with the results from Section 4.3, we are now able to compute lower bounds for the asymptotic minimax risk in the global model $(\mathcal{Y}^n, \mathcal{B}^n, \{P_\xi^{(n)} : \xi \in \mathscr{F}\})$ (provided that the parameter space Θ is one-dimensional).

5.2 On the theory of Donoho and Liu

In Section 5.1.3 we referred to results from the theory of Donoho (1994) and Donoho and Liu (1991) in order to justify (5.13). Doing so, we accepted however two inaccuracies:

Problem 1. In (5.13) the supremum is taken over the minimax affine risks corresponding to submodels indexed by the functions ζg, with fixed g and $\zeta \in [-1, 1]$. However, following the lines of Donoho, the supremum ought to be taken over the minimax affine risks corresponding to a larger class of submodels, namely—as in (5.11)—to the class of all submodels indexed by the sets of convex combinations $\{tg + (1-t)h : t \in [0, 1]\}$ (with $g, h \in \mathscr{G}$). It therefore remains to prove that it is indeed sufficient to take the supremum over to the smaller class of *centrosymmetric* submodels $\{Q_{\gamma,\zeta g} : \zeta \in [-1, 1]\}$.

Problem 2. Donoho and Liu show that the minimax affine risk in the complete model equals the supremum over the minimax affine risks in the linear submodels, which they prove for certain classes of loss functions, such as for quadratic loss and some "fixed-length confidence statements". It remains to prove that such a statement also holds in case of our specific zero-one loss function $\mathbb{1}_{[-\delta,\delta]^C}$.

In the following, we will discuss these two problems and show that (5.13) indeed holds.

Throughout this section let the assumptions made at the beginning of Section 5.1 hold, i.e. we suppose that a fixed $\gamma \in \Theta' \subseteq \mathbb{R}$ is given, for which we consider the corresponding limit experiment $\{Q_{\gamma,g} : g \in \mathscr{G}\}$, with $\mathscr{G} = \mathscr{G}(K, \rho, 1)$. We also use the notation from that section. For arbitrary $g, h \in \mathscr{G}$ let in the following

$$[g, h] := \{tg + (1-t)h : t \in [0, 1]\}$$

denote the set of convex combinations of g and h. The corresponding linear submodel is then defined by

$$Q_{\gamma,[g,h]} := \{Q_{\gamma,f} : f \in [g, h]\},$$

and the minimax affine risk (with respect to the loss $\mathbb{1}_{[-\delta,\delta]^C}$) is denoted $R^*_{\text{aff}}([g, h,])$. With the choice $h = -g$ we obtain the special submodels already considered in (5.1), for which we shortly write $R^*_{\text{aff}}(g)$ for $R^*_{\text{aff}}([-g, g])$.

5.2.1 Restriction to centrosymmetric models

In this section we discuss the first of the two problems described above. We show that the supremum taken over the the minimax affine risks of *all* linear submodels is the same as the supremum taken over the minimax affine risks of the centrosymmetric linear submodels $\{Q_{\gamma,\zeta g} : \zeta \in [-1, 1]\}$ considered in Section 5.1. In other words, we have to prove that

$$\sup\{R^*_{\text{aff}}([g, h]) : g, h \in \mathscr{G}\} = \sup\{R^*_{\text{aff}}(g) : g \in \mathscr{G}\}.$$

To this end, let some arbitrary $g, h \in \mathscr{G}$ be given. We consider the corresponding one-dimensional submodel $Q_{\gamma,[g,h]}$.[2)] Setting $k := (g + h)/2$ and $w := (g - h)/2$ each $f \in [g, h]$ has a unique representation as

$$f = k + \zeta w, \quad \zeta \in [-1, 1].$$

In order to estimate the parameter of interest $f(0) = k(0) + \zeta w(0)$, we have to estimate ζ. For that purpose we consider the random variable

$$Y_w = \frac{\int \mathcal{C}_\gamma w \; d\mathrm{x}}{\|\mathcal{C}_\gamma w\|_{\lambda,2}},$$

[2)] Note that in the case $h = -g$ the notation introduced in the following is completely compatible with that from Section 5.1.

defined according to (5.2). Without loss of generality, we can assume that $w(0) > 0$. Again, we have $\mathscr{L}(Y_w \mid Q_{\gamma,0}) = \mathscr{N}(0,1)$ and

$$\mathscr{L}(Y_w \mid Q_{\gamma,k+\zeta w}) = \mathscr{N}(\vartheta, 1),$$

where

$$\begin{aligned}
\vartheta &= \frac{\langle \mathcal{C}_\gamma w, \mathcal{C}_\gamma (k+\zeta w)\rangle_\lambda}{\|\mathcal{C}_\gamma w\|_{\lambda,2}} \\
&= \frac{\langle \mathcal{C}_\gamma w, \mathcal{C}_\gamma k\rangle_\lambda}{\|\mathcal{C}_\gamma w\|_{\lambda,2}} + \frac{\zeta \|\mathcal{C}_\gamma w\|_{\lambda,2}^2}{\|\mathcal{C}_\gamma w\|_{\lambda,2}} \\
&= \frac{\mathcal{C}_\gamma}{2} \cdot \frac{(\|g\|_{\lambda,2}^2 - \|h\|_{\lambda,2}^2)}{\|g-h\|_{\lambda,2}} + \zeta \|\mathcal{C}_\gamma w\|_{\lambda,2} \\
&=: a_{g,h} + \zeta \tau_w.
\end{aligned}$$

Since $\zeta \in [-1,1]$, the location parameter ϑ underlies the restriction

$$\vartheta \in [a_{g,h} - \tau_w, a_{g,h} + \tau_w].$$

Let an arbitrary estimator $\hat{\vartheta}$ for ϑ be given. We set $\hat{\zeta} := (\hat{\vartheta} - a_{g,h})/\|\mathcal{C}_\gamma w\|_{\lambda,2}$ and $\hat{f}(0) := k(0) + \hat{\zeta} w(0)$, and therefore we get

$$\begin{aligned}
|\hat{f}(0) - f(0)| > \delta \ &\Leftrightarrow\ |\hat{\zeta} - \zeta| > \delta / w(0) \\
&\Leftrightarrow\ |\hat{\vartheta} - \vartheta| > \frac{\delta \|\mathcal{C}_\gamma w\|_{\lambda,2}}{w(0)} =: \delta_w.
\end{aligned}$$

Thus, the problem of finding minimax affine estimators (with respect to the loss function $\mathbb{1}_{[-\delta,\delta]^C}$) for the parameter $f(0)$ in the model $Q_{\gamma,[g,h]}$ is equivalent to finding such estimators in the model

$$\{\mathscr{N}(\vartheta,1) : \vartheta \in [a_{g,h} - \tau_w, a_{g,h} + \tau_w]\}$$

given the loss function $\mathbb{1}_{[-\delta_w,\delta_w]^C}$. For the case $a_{g,h} = 0$ this problem has already been solved in Setion 5.1. As a consequence of the calculations there—using the same notation—the estimator $c_w Y_w$ is minimax affine for estimating ϑ (see Theorem 5.1). For the remaining case $a := a_{g,h} \neq 0$ we consider the affine estimator $c_w(Y_w - a) + a$. It satisfies

$$\begin{aligned}
&\sup_{|\vartheta - a| \le \tau_w} \mathscr{N}_{(\vartheta,1)}\{|c_w(Y_w - a) + a - \vartheta| > \delta_w\} \\
&\qquad\qquad = \sup_{|\tilde{\vartheta}| \le \tau_w} \mathscr{N}_{(\tilde{\vartheta},1)}\{|c_w Y_w - \tilde{\vartheta}| > \delta_w\},
\end{aligned}$$

and from that one concludes $R^*_{\text{aff}}([g,h]) = R^*_{\text{aff}}(w)$. Consequently,

$$\sup\{R^*_{\text{aff}}([g,h]) : g,h \in \mathscr{G}\} = \sup\{R^*_{\text{aff}}(w) : w \in \mathscr{G}\}, \tag{5.28}$$

which was to be shown. Furthermore, the preceding arguments imply that the supremum on the left hand side of (5.28) is attained if and only if it is attained on the right hand side of that equation.

5.2.2 The use of a zero-one loss function

The solution to the second problem is somewhat more complicated and needs some more sophisticated arguments. We start with a generalisation of the notation. For $x \geq 0$ we consider the loss function $\mathbb{1}_{[-x,x]^C}$. Let $R^*_{\text{aff},x}(\mathscr{G})$ denote the corresponding minimax affine risk in the limit experiment $\{Q_{\gamma,g} : g \in \mathscr{G}\}$. Likewise, $R^*_{\text{aff},x}([g,h])$ and $R^*_{\text{aff},x}(g)$ for $g,h \in \mathscr{G}$ may be defined. So far, we have always considered the minimax affine risk for the case of a fixed $x = \delta$, for which we did not use an additional subscript x. First we present some auxiliary results for the functions $x \mapsto R^*_{\text{aff},x}(\mathscr{G})$ and $x \mapsto R^*_{\text{aff},x}([g,h])$.

5.7 Lemma. *Let $g,h \in \mathscr{G}$. Then the following assertions hold:*

(i) The maps $x \mapsto R^*_{\text{aff},x}([g,h])$ and $x \mapsto R^*_{\text{aff},x}(\mathscr{G})$ are strictly decreasing on their support. Moreover, $R^*_{\text{aff},x}([g,h]) \leq R^*_{\text{aff},x}(\mathscr{G})$ for all x and $R^*_{\text{aff},0}(\mathscr{G}) = R^*_{\text{aff},0}([g,h]) = 1$.

(ii) $x \mapsto R^*_{\text{aff},x}([g,h])$ is continuous on $\{x : 0 < R^*_{\text{aff},x}([g,h]) < 1\}$.

(iii) Let $g \in \mathscr{G}$ such that $g(0) > 0$. Then $R^*_{\text{aff},x}(g) \downarrow \Phi(-\tau_g)$ as $x \uparrow g(0)$.

Proof. *Ad (i):* Since most of the assertions are trivial, we only prove that the map $x \mapsto R^*_{\text{aff},x}(\mathscr{G})$ is strictly decreasing on its support. To this end, let $x > 0$ be given such that $R^*_{\text{aff},x}(\mathscr{G}) > 0$, and let $\tilde{x} > x$ be given. We have to show that $R^*_{\text{aff},\tilde{x}}(\mathscr{G}) < R^*_{\text{aff},x}(\mathscr{G})$. By definition, we have

$$R^*_{\text{aff},x}(\mathscr{G}) = \inf_{\hat{T} \text{ affine}} \sup_{g \in \mathscr{G}} Q_{\gamma,g}\{|\hat{T} - g(0)| > x\},$$

where the affine estimators $\hat{T}$ are of the specific form as given on p. 56. We choose a sequence $\hat{T}_n$ of affine estimators for the parameter $g(0)$ such that

$$\inf_{\hat{T}\text{ affine}} \sup_{g\in\mathscr{G}} Q_{\gamma,g}\{|\hat{T}-g(0)|>x\} = \liminf_{n\to\infty} \sup_{g\in\mathscr{G}} Q_{\gamma,g}\{|\hat{T}_n-g(0)|>x\}.$$

Furthermore, we can choose a suitable sequence of functions $g_n \in \mathscr{G}$ such that

$$\begin{aligned}\liminf_{n\to\infty} \sup_{g\in\mathscr{G}} Q_{\gamma,g}\{|\hat{T}_n-g(0)|>\tilde{x}\} \\ = \liminf_{n\to\infty} Q_{\gamma,g_n}\{|\hat{T}_n-g_n(0)|>\tilde{x}\}.\end{aligned} \tag{5.29}$$

From that we conclude that

$$\begin{aligned}
R^*_{\mathrm{aff},x}(\mathscr{G}) &= \liminf_{n\to\infty} \sup_{g\in\mathscr{G}} Q_{\gamma,g}\{|\hat{T}_n-g(0)|>x\} \\
&= \liminf_{n\to\infty} \sup_{g\in\mathscr{G}} \Big(Q_{\gamma,g}\{|\hat{T}_n-g(0)|>\tilde{x}\} \\
&\qquad +Q_{\gamma,g}\{|\hat{T}_n-g(0)|\in(x,\tilde{x}]\}\Big) \\
&\geq \liminf_{n\to\infty} \Big(Q_{\gamma,g_n}\{|\hat{T}_n-g_n(0)|>\tilde{x}\} \\
&\qquad +Q_{\gamma,g_n}\{|\hat{T}_n-g_n(0)|\in(x,\tilde{x}]\}\Big) \\
&\overset{(*)}{\geq} \liminf_{n\to\infty} Q_{\gamma,g_n}\{|\hat{T}_n-g_n(0)|>\tilde{x}\} \\
&\qquad +\liminf_{n\to\infty} Q_{\gamma,g_n}\{|\hat{T}_n-g_n(0)|\in(x,\tilde{x}]\},
\end{aligned}$$

where $(*)$ is a direct consequence of Exercise 14. on p. 58 of Königsberger (2001). Note that

$$\begin{aligned}
\liminf_{n\to\infty} Q_{\gamma,g_n}\{|\hat{T}_n-g_n(0)|>\tilde{x}\} &\overset{(5.29)}{=} \liminf_{n\to\infty} \sup_{g\in\mathscr{G}} Q_{\gamma,g}\{|\hat{T}_n-g(0)|>\tilde{x}\} \\
&\geq \inf_{\hat{T}\text{ affine}} \sup_{g\in\mathscr{G}} Q_{\gamma,g}\{|\hat{T}-g(0)|>\tilde{x}\} \\
&= R^*_{\mathrm{aff},\tilde{x}}(\mathscr{G}),
\end{aligned}$$

and that

$$\liminf_{n\to\infty} Q_{\gamma,g_n}\{|\hat{T}_n-g_n(0)|\in(x,\tilde{x}]\}>0,$$

which follows from the fact that the affine estimators $\hat{T}_n$ are normally distributed under every measure $Q_{\gamma,g}$, and that the sequence of the corresponding maximal risks converges to $R^*_{\mathrm{aff},x}(\mathscr{G})$ which, by assumption, is greater than zero. Putting together the above results, we conclude that $R^*_{\mathrm{aff},x}(\mathscr{G}) > R^*_{\mathrm{aff},\tilde{x}}(\mathscr{G})$, which had to be shown.

Ad (ii): Following the discussion in Section 5.2.1, it is clear that we can without loss of generality assume $h = -g$, and for symmetry reasons we may further assume that $g(0) > 0$. Clearly, $R^*_{\mathrm{aff},x}(g) > 0$ if and only if $x < g(0)$. Let now τ_g be defined as in (5.3), and define x_g according to (5.6), with δ replaced by x, i.e. $x_g = x\|\mathcal{C}_\gamma g\|_{\lambda,2}/g(0)$. Furthermore, let

$$c_g(x) := \left(\frac{1}{2} + \left[\frac{1}{4} + \frac{g(0)}{2x\|\mathcal{C}_\gamma g\|^2_{\lambda,2}} \log\frac{g(0)+x}{g(0)-x}\right]^{1/2}\right)^{-1} \mathbb{1}\{g(0) > x\},$$

which coincides with the constant c_g from Theorem 5.1 (again, with δ replaced by x). Theorem 5.1 yields

$$R^*_{\mathrm{aff},x}(g) = \Phi\left(-\frac{\tau_g + x_g}{c_g(x)} + \tau_g\right) + \Phi\left(\frac{\tau_g - x_g}{c_g(x)} - \tau_g\right), \; x \in (0, g(0)).$$

Obviously, x_g and $c_g(x)$ are continuous functions of x, for x ranging over $(0, g(0))$. This proves the assertion.

Ad (iii): From the definition of $c_g(x)$ it is clear that $c_g(x) \to 0$ as $x \uparrow g(0)$, and thus $\Phi(-[\tau_g + x_g]/c_g(x) + \tau_g) \to 0$. It remains to prove that $(\tau_g - x_g)/c_g(x) \to 0$ as $x \uparrow g(0)$, which implies $\Phi([\tau_g - x_g]/c_g(x) - \tau_g) \to \Phi(-\tau_g)$ and thus the stated convergence. For $x \uparrow g(0)$ the following chain of equivalences holds:

$$\begin{aligned}
&\frac{\tau_g - x_g}{c_g(x)} \to 0 \\
&\Leftrightarrow \tau_g\left(1 - \frac{x}{g(0)}\right)\left(\frac{1}{2} + \left[\frac{1}{4} + \frac{g(0)}{2x\|\mathcal{C}_\gamma g\|^2_{\lambda,2}} \log\frac{g(0)+x}{g(0)-x}\right]^{1/2}\right) \to 0 \\
&\Leftrightarrow (g(0) - x)\left[\frac{1}{4} + \frac{g(0)}{2x\|\mathcal{C}_\gamma g\|^2_{\lambda,2}} \log\frac{g(0)+x}{g(0)-x}\right]^{1/2} \to 0 \\
&\Leftrightarrow (g(0) - x)^2\left[\frac{1}{4} + \frac{g(0)}{2x\|\mathcal{C}_\gamma g\|^2_{\lambda,2}} \log\frac{g(0)+x}{g(0)-x}\right] \to 0 \\
&\Leftrightarrow (g(0) - x)^2 \log\frac{g(0)+x}{g(0)-x} \to 0.
\end{aligned}$$

The last assertion can easily be verified with L'Hospital's rule. ■

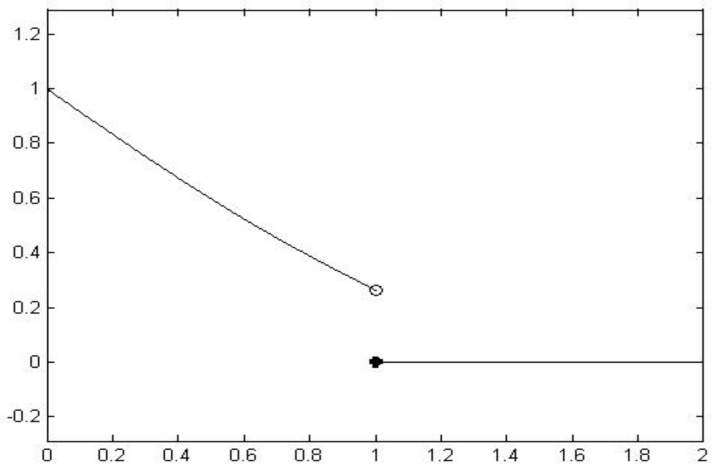


Figure. *Plot of the function* $x \mapsto R^*_{\mathrm{aff},x}(g)$*, where* $g = \mathbb{1}_{[-1,1]}$*, and* $\mathcal{I}_\gamma = 1$*. Note that* $R^*_{\mathrm{aff},1}(g) = 0$*, and* $R^*_{\mathrm{aff},1-}(g) > 0$*, which we tried to indicate by the special marks.*

The third assertion of the lemma says that the map $x \mapsto R^*_{\mathrm{aff},x}(g)$ is discontinuous at the point $x = g(0)$ (provided that $g(0) > 0$). This is not what one might expect—indeed, this fact will somewhat complicate the discussion of the second problem. The figure on the top of this page shows a typical graph of a a function $x \mapsto R^*_{\mathrm{aff},x}(g)$.

Now let a fixed $\alpha \in [0,1]$ be given. We define

$$\chi_{\mathrm{aff},\alpha}(\mathscr{G}) := \inf\{x \geq 0 : R^*_{\mathrm{aff},x}(\mathscr{G}) \leq \alpha\}.$$

One can interpret $2\chi_{\mathrm{aff},\alpha}(\mathscr{G})$ as the length of a minimal symmetric fixed-size confidence interval to the level $1-\alpha$ resulting from the use of an affine estimator (see Stark, 1992). Analogously, $\chi_{\mathrm{aff},\alpha}([g,h])$ is defined for the linear submodel $Q_{\gamma,[g,h]}$. Since the function space $\mathscr{G}$ is convex, Theorem 1 in Donoho (1994), together with the arguments from Section 9 of the same paper, implies the existence of functions $g^*, h^* \in \mathscr{G}$ such that

$$\begin{aligned} \chi_{\mathrm{aff},\alpha}(\mathscr{G}) &= \chi_{\mathrm{aff},\alpha}([g^*,h^*]) \\ &= \sup\{\chi_{\mathrm{aff},\alpha}([g,h]) : g,h \in \mathscr{G}\}. \end{aligned} \tag{5.30}$$

In other words, a hardest linear submodel exists, as far as $\chi_{\mathrm{aff},\alpha}(\cdot)$ is considered as a performance criterion.

We show that a similar statement holds true for the minimax affine risk with respect to our usual loss function $\mathbb{1}_{[-\delta,\delta]^C}$. More precisely, we want to prove that for $\delta > 0$ the minimax affine risk in the complete experiment equals the supremum over all minimax affine risks from the linear submodels, i.e.

$$R^*_{\mathrm{aff},\delta}(\mathscr{G}) = \sup\{R^*_{\mathrm{aff},\delta}([g,h]) : g,h \in \mathscr{G}\}. \tag{5.31}$$

In the following, let $\alpha := R^*_{\mathrm{aff},\delta}(\mathscr{G})$ and

$$\alpha' := \sup\{R^*_{\mathrm{aff},\delta}([g,h]) : g,h \in \mathscr{G}\}.$$

It is clear that $\alpha' \le \alpha$, and for a proof of (5.31) we have to show that even $\alpha' = \alpha$. Since there is nothing to prove if $\alpha = 0$, we can and will in the following assume that $\alpha > 0$. We choose the functions g^* and h^* such that they satisfy (5.30).

Let us assume that the assertion $\alpha' = \alpha$ is false. Then it may be either $\alpha' = 0$ or $0 < \alpha' < \alpha$. We show that each of these two cases leads to a contradiction.

Case 1: $\boldsymbol{\alpha' = 0}$. If $\alpha' = 0$ then also $R^*_{\mathrm{aff},\delta}(g) = 0$ for all $g \in \mathscr{G}$. By Theorem 5.1 this is equivalent to $|g(0)| \le \delta$ for all g. In particular, this would imply $R^*_{\mathrm{aff},\delta}(\mathscr{G}) = 0$, which is a contradiction to the assumption $\alpha > 0$.

Case 2: $\boldsymbol{0 < \alpha' < \alpha}$. By definition, $\chi_{\mathrm{aff},\alpha}(\mathscr{G}) \le \delta$, and because $x \mapsto R^*_{\mathrm{aff},x}(\mathscr{G})$ is strictly decreasing on its support, this relation even holds with equality, and thus

$$\delta = \chi_{\mathrm{aff},\alpha}(\mathscr{G}) \overset{(5.30)}{=} \chi_{\mathrm{aff},\alpha}([g^*,h^*]). \tag{5.32}$$

If $\alpha' \in (0,\alpha)$, then in particular

$$\alpha^* := R^*_{\mathrm{aff},\delta}([g^*,h^*]) \le \alpha' < \alpha.$$

If now $\alpha^* > 0$, then assertion (ii) from Lemma 5.7 implies that there is some $\delta^* < \delta$ for which $R^*_{\mathrm{aff},\delta^*}([g^*,h^*]) \le \alpha$. Therefore, we have

$$\begin{aligned} \delta &= \chi_{\mathrm{aff},\alpha}([g^*,h^*]) \\ &= \inf\{x \ge 0 : R^*_{\mathrm{aff},x}([g^*,h^*]) \le \alpha\} \le \delta^* < \delta, \end{aligned} \tag{5.33}$$

which is contradictory to (5.32).

It remains to consider the case $\alpha^* = 0$. In this case one of the following assertions holds true:

a) there is a $\delta^* < \delta$ such that $R^*_{\mathrm{aff},\delta^*}([g^*,h^*]) < \alpha$, or

b) for all $x < \delta$ we have $R^*_{\mathrm{aff},x}([g^*, h^*]) \geq \alpha$.

Case a) can be discussed with the same arguments as before, which led to the contradiction in (5.33). It thus remains to investigate Case b).
From the discussion in Section 5.2.1 we know that

$$R^*_{\mathrm{aff},x}([g^*, h^*]) = R^*_{\mathrm{aff},x}(w^*)$$

for all x, where $w^* = (g^* - h^*)/2$ (which is an element of $\mathscr{G}$). Furthermore, we have

$$\begin{aligned} \chi_{\mathrm{aff},\alpha}([g^*, h^*]) &= \inf\{x \geq 0 : R^*_{\mathrm{aff},x}([g^*, h^*]) \leq \alpha\} \\ &= \inf\{x \geq 0 : R^*_{\mathrm{aff},x}(w^*) \leq \alpha\} \\ &= \chi_{\mathrm{aff},\alpha}(w^*). \end{aligned}$$

Assertion (iii) in Lemma 5.7 yields

$$R^*_{\mathrm{aff},x}([g^*, h^*]) = R^*_{\mathrm{aff},x}(w^*) \downarrow \Phi(-\tau_{w^*}) \quad \text{as } x \uparrow w^*(0).$$

Since we consider the case that $R^*_{\mathrm{aff},x}(w^*) = R^*_{\mathrm{aff},x}([g^*, h^*]) \geq \alpha$ for all $x < \delta$, this implies $\Phi(-\tau_{w^*}) \geq \alpha$, and thus $w^*(0) = \delta$. Consequently, we have

$$\delta = \chi_{\mathrm{aff},\alpha}([g^*, h^*]) = \chi_{\mathrm{aff},\alpha}(w^*) = w^*(0).$$

Consider now a sequence $g_n \in \mathscr{G}$ that satisfies the following conditions:

$$g_n(0) > w^*(0), \quad g_n(0) \downarrow w^*(0), \quad \text{and} \quad \tau_{g_n} \to \tau_{w^*}.$$

The existence of such a sequence g_n follows with some simple arguments from the definition of the space $\mathscr{G}$. Because of $g_n(0) > w^*(0) = \delta$ and by the definition of α' we have $0 < R^*_{\mathrm{aff},\delta}(g_n) \leq \alpha'$ for all n. Using property (iii) from Lemma 5.7 these arguments yield

$$\alpha > \alpha' \geq \limsup_{n\to\infty} R^*_{\mathrm{aff},\delta}(g_n) \geq \limsup_{n\to\infty} \Phi(-\tau_{g_n}) = \Phi(-\tau_{w^*}) \geq \alpha,$$

and hence the contradiction $\alpha > \alpha$.

Summarising, we have shown that both cases $\alpha' = 0$ and $\alpha' \in (0, \alpha)$ lead to contradictions and therefore must be false. Hence, $\alpha' = \alpha$, and therefore (5.31) holds. Combining (5.31) and (5.28) we have proven the following result:

5.8 Proposition. *Equation (5.13) holds, i.e. the minimax affine risk with respect to the loss function* $\mathbb{1}_{[-\delta,\delta]^C}$ *satisfies* $R^*_{\text{aff}}(\mathscr{G}) = \sup\{R^*_{\text{aff}}(g) : g \in \mathscr{G}\}$.

5.3 Minimax risk bounds in the multivariate limit experiment

So far we have derived upper and lower bounds for the miniax risk in the univariate limit experiment. In this section we want to expand these to the case of the multivariate limit experiment. In the following, let $\gamma \in \Theta'$ be a fixed centre of localisation, where now Θ'—defined according to (3.7)—is a subset of $\Theta \subseteq \mathbb{R}^d$, $d \geq 1$. We consider the corresponding multivariate limit experiment

$$(C^d([-K,K]), \mathscr{C}^d([-K,K]), \{Q_{\gamma,g} : g \in \mathscr{G}\}),$$

with $\mathscr{G} = \mathscr{G}(K,\rho,d)$ according to (3.10). Within this experiment we want to estimate the local parameter $g(0)$. As in Section 5.1 $R^*(\mathscr{G})$ denotes the corresponding minimax risk with respect to the loss function $\mathbb{1}_{[-\delta,\delta]^C}$ (with $\delta = (\delta_1,\ldots,\delta_d)^\top$ as in Section 3.3).

The aim of this section is to derive upper and lower bounds for $R^*(\mathscr{G})$. To this end, we first consider such bounds for the minimax risk corresponding to the estimation of the parameter $\mathcal{C}_\gamma g(0)$. Here, the basic idea consists of splitting the multivariate experiment into several univariate experiments, for which the results from the preceding sections apply. Having derived upper minimax risk bounds for the estimation of $\mathcal{C}_\gamma g(0)$, we can deduce upper bounds for $R^*(\mathscr{G})$. In the following, we write the Cholesky matrix as $\mathcal{C}_\gamma = (\mathcal{C}_\gamma^{ij})$. We set

$$\mathcal{C}_\gamma^i := (\mathcal{C}_\gamma^{i1},\ldots,\mathcal{C}_\gamma^{id}),$$

which is simply the ith row vector of the matrix $\mathcal{C}_\gamma$.

5.3.1 An upper minimax risk bound

In order to derive an upper bound for $R^*(\mathscr{G})$ we first consider the problem of estimating the parameter $\mathcal{C}_\gamma g(0)$ in the limit experiment $\{Q_{\gamma,g} : g \in \mathscr{G}\}$. In the characterisation the limit experiment in Section 4.1 we saw that $(C^d([-K,K]), \mathscr{C}^d([-K,K]))$ is the product of d copies of $(C^1([-K,K]), \mathscr{C}^1([-K,K]))$. Due to the definition in (4.2) we can therefore write $Q_{\gamma,g}$ as a product

$$Q_{\gamma,g} = Q^{(1)}_{\gamma,g} \otimes \ldots \otimes Q^{(d)}_{\gamma,g},$$

with probability measures

$$Q^{(j)}_{\gamma,g} := \mathscr{L}\left(\int_{-K}^{\cdot} \mathcal{C}^j_\gamma g(s)\, ds + W^{(j)}\right) \tag{5.34}$$

defined on the space $(C^1([-K,K]), \mathscr{C}^1([-K,K]))$. Here $W^{(j)}$ denotes the jth component of a d-dimensional Brownian motion (see also Section 4.1 for the notation used here). Since γ is fixed, we will in the following often suppress it in the notation, i.e. we shortly write Q_g and $Q^{(j)}_g$ for $Q_{\gamma,g}$ and $Q^{(j)}_{\gamma,g}$, respectively.

Let us now consider an observation from the limit experiment, i.e. a random variable $\mathrm{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_d)^\top$ with independent components $\mathrm{x}_j \sim Q^{(j)}_g$. Estimators $\hat{\kappa}$ for $\mathcal{C}_\gamma g(0)$ are then in general of the type

$$\hat{\kappa} = \hat{\kappa}(\mathrm{x}) = (\hat{\kappa}_1(\mathrm{x}), \ldots, \hat{\kappa}_d(\mathrm{x}))^\top.$$

In the following, we first examine the minimax risk corresponding to the problem of estimating $\mathcal{C}_\gamma g(0)$. As underlying loss function we consider the function $\mathbb{1}_{[-v,v]^C}$, where the vector

$$v = (v_1, \ldots, v_d)^\top, \quad v_i > 0,$$

is defined according to Section 3.5.3. The minimax risk for the estimation of $\mathcal{C}_\gamma g(0)$ in the experiment $\{Q_g : g \in \mathscr{G}\}$ under this loss function is denoted $\widetilde{R}^*(\mathscr{G})$, i.e. we have

$$\widetilde{R}^*(\mathscr{G}) \;=\; \inf_{\hat{\kappa}} \sup_{g \in \mathscr{G}} Q_g\big\{\hat{\kappa} - \mathcal{C}_\gamma g(0) \in [-v,v]^C\big\}, \tag{5.35}$$

where the infimum is taken over the class of all estimators $\hat{\kappa}$ for the parameter $g(0)$. We want to derive an upper bound for $\widetilde{R}^*(\mathscr{G})$. With our special choice of v from Section 3.5.3 this will yield an upper bound for $R^*(\mathscr{G})$, too, as will be shown later. Note that we here and in the following will suppress the dependence of $v = v(\gamma)$ on the parameter γ. Of course, other choices for a suitable vector v are possible, some alternative approaches will be discussed in Section 5.4.

5.9 Remark. One should distinguish carefully between the just defined risk $\widetilde{R}^*(\mathscr{G})$ and $R^*(\mathscr{G})$. The latter denotes—as in Sections 5.1 and 5.1.4—the minimax risk with respect to the loss function $\mathbb{1}_{[-\delta,\delta]^C}$ for the estimation of $g(0)$. Eventually, it is of course $R^*(\mathscr{G})$, which we are interested in.

For our considerations such estimators $\hat{\kappa}$ are of special interest, that estimate $\mathcal{C}_\gamma^j g(0)$ (the jth component of the vector $\mathcal{C}_\gamma g(0)$) only on the basis of the component x_j, in other words: estimators that fulfil

$$\hat{\kappa}_j(\mathrm{x}) = \hat{\kappa}_j(\pi_j(\mathrm{x})), \quad j = 1, \dots, d, \tag{5.36}$$

where $\pi_j\colon (\mathrm{x}_1, \dots, \mathrm{x}_d)^\top \mapsto (0, \dots, \mathrm{x}_j, \dots, 0)^\top$. The independence of the components x_j implies that for such an estimator

$$\begin{aligned} Q_g\big\{\hat{\kappa} - &\mathcal{C}_\gamma g(0) \in [-v, v]^C\big\} \\ &= 1 - Q_g\{\hat{\kappa} - \mathcal{C}_\gamma g(0) \in [-v, v]\} \\ &= 1 - \prod_{j=1}^{d} Q_g^{(j)}\{|\hat{\kappa}_j - \mathcal{C}_\gamma^j g(0)| \le v_j\}. \end{aligned} \tag{5.37}$$

If we consider the definition of the minimax risk $\widetilde{R}^*(\mathscr{G})$, it is clear that the right hand side of (5.35) will not decrease if the infimum is not taken with respect to all possible estimators, but only with respect to those estimators $\hat{\kappa}$ satisfying (5.36), and therefore,

$$\widetilde{R}^*(\mathscr{G}) \le \inf_{\hat{\kappa} \text{ with } (5.36)} \sup_{g \in \mathscr{G}} Q_g\big\{\hat{\kappa} - \mathcal{C}_\gamma g(0) \in [-v, v]^C\big\}. \tag{5.38}$$

This inequality is presumably in general a strict equality. The following heuristic argument may support this idea—though it does of course not replace a proof. Suppose for a moment that (5.38) always holds with "=". Of course, then the right hand side of the formula equals that of (5.35). Heuristically formulated, this means that the estimate of a component $\mathcal{C}_\gamma^j g(0)$ cannot be improved by incorporating any information that might be obtained for the location of other components $\mathcal{C}_\gamma^i g(0)$ $(i \ne j)$. As a consequence, the set $\{\mathcal{C}_\gamma g(0) : g \in \mathscr{G}\}$ would have to be a cuboid in $\mathbb{R}^d$. However, this is in general not the case. (As a counter-example, consider Example 3.2, in which $\Theta' \subseteq \mathbb{R}^2$ and $\mathcal{C}_\gamma = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. The set $\{\mathcal{C}_\gamma g(0) : g \in \mathscr{G}\}$ would qualitatively look the same as the set A_γ depicted in Figure (b) of that example.)

Let us now examine the right hand side of (5.38). Some elementary calculations yield

$$
\begin{aligned}
&\inf_{\hat\kappa \text{ with } (5.36)} \sup_{g\in\mathscr{G}} Q_g\big\{\hat\kappa - \mathcal{C}_\gamma g(0) \in [-v,v]^C\big\} \\
&\overset{(5.37)}{=} \inf_{\hat\kappa \text{ with } (5.36)} \sup_{g\in\mathscr{G}} \left(1 - \prod_{j=1}^{d} Q_g^{(j)}\{|\hat\kappa_j - \mathcal{C}_\gamma^j g(0)| \le v_j\}\right) \\
&= 1 - \sup_{\hat\kappa \text{ with } (5.36)} \inf_{g\in\mathscr{G}} \prod_{j=1}^{d} Q_g^{(j)}\{|\hat\kappa_j - \mathcal{C}_\gamma^j g(0)| \le v_j\} \\
&\le 1 - \prod_{j=1}^{d} \sup_{\hat\kappa \text{ with } (5.36)} \inf_{g\in\mathscr{G}} Q_g^{(j)}\{|\hat\kappa_j - \mathcal{C}_\gamma^j g(0)| \le v_j\} \\
&= 1 - \prod_{j=1}^{d} \left(1 - \inf_{\hat\kappa \text{ with } (5.36)} \sup_{g\in\mathscr{G}} Q_g^{(j)}\{|\hat\kappa_j - \mathcal{C}_\gamma^j g(0)| > v_j\}\right).
\end{aligned}
$$

In the last product term

$$\inf_{\hat\kappa \text{ with } (5.36)} \sup_{g\in\mathscr{G}} Q_g^{(j)}\{|\hat\kappa_j - \mathcal{C}_\gamma^j g(0)| > v_j\}$$

represents the minimax risk for the estimation of the real parameter $\mathcal{C}_\gamma^j g(0)$ in the univariate, jth subexperiment $\{Q_g^{(j)} : g \in \mathscr{G}\}$, with underlying loss function $\mathbb{1}_{[-v_j,v_j]^C}$. To examine this experiment we consider the function space

$$\mathscr{H}_\gamma^{(j)} := \{\hbar : \hbar = \mathcal{C}_\gamma^j g,\ g \in \mathscr{G}\}.$$

Each $\hbar$ is a continuous function $\hbar\colon [-K,K] \to \mathbb{R}$, which follows from the properties of the functions $g \in \mathscr{G}$. Furthermore, $\hbar(0) = \mathcal{C}_\gamma^j g(0)$ for $\hbar = \mathcal{C}_\gamma^j g$. We set

$$\widetilde{Q}_\hbar^{(j)} := \mathscr{L}\left(\int_{-K}^{\cdot} \hbar(s)\, ds + W^{(j)}\right), \quad \hbar \in \mathscr{H}_\gamma^{(j)}. \tag{5.39}$$

Obviously, $\widetilde{Q}_\hbar^{(j)}$ corresponds to the measure $Q_g^{(j)}$ from (5.34), and we can thus identify the experiments $\{\widetilde{Q}_\hbar^{(j)} : \hbar \in \mathscr{H}_\gamma^{(j)}\}$ and $\{Q_g^{(j)} : g \in \mathscr{G}\}$. Therefore, the minimax risk with respect to the loss function $\mathbb{1}_{[-v_j,v_j]^C}$ is the same in both experiments. In other words,

$$
\begin{aligned}
&\inf_{\hat\kappa_j} \sup_{g\in\mathscr{G}} Q_g^{(j)}\{|\hat\kappa_j - \mathcal{C}_\gamma^j g(0)| > v_j\} \\
&\qquad = \inf_{\hat\kappa_j} \sup_{\hbar\in\mathscr{H}_\gamma^{(j)}} \widetilde{Q}_\hbar^{(j)}\{|\hat\kappa_j - \hbar(0)| > v_j\},
\end{aligned} \tag{5.40}
$$

with the infima taken over all estimators $\hat{\kappa}_j$ satisfying (5.36). We are now going to provide an upper bound for the expression on the right hand side of (5.40), which will eventually yield an upper bound on $R^*(\mathscr{G})$, the minimax risk for the estimation of the parameter $g(0)$.

Note that the random variable

$$\int_{-K}^{\cdot} \hbar(s)\, ds + W^{(j)}$$

from (5.39) is of the same type as the white noise model (5.12) considered in the univariate case. The only difference to that equation is that the factor $\mathcal{C}_\gamma$ from (5.12) is already included in the function $\hbar$ and therefore does not appear any more in the above integral expression. Furthermore, $\mathscr{H}_\gamma^{(j)} \subseteq \mathbb{L}_2(\mathbb{R}, \mathbb{B}, \lambda)$, and $\mathscr{H}_\gamma^{(j)}$ is symmetric and convex. This follows simply from the fact that the space $\mathscr{G}$ has these properties and that each $\hbar \in \mathscr{H}_\gamma^{(j)}$ is a linear combination of the components of a function $g \in \mathscr{G}$. Consequently, Donoho and Liu's theory of hardest linear submodels as well as the results from Section 5.2 also hold true if we consider the function space $\mathscr{H}_\gamma^{(j)}$ instead of $\mathscr{G}$. Consequently,

$$\widetilde{R}^*_{\text{aff}}(\mathscr{H}_\gamma^{(j)}) = \sup\left\{\widetilde{R}^*_{\text{aff}}(\hbar) : \hbar \in \mathscr{H}_\gamma^{(j)}\right\}, \tag{5.41}$$

where $\widetilde{R}^*_{\text{aff}}(\mathscr{H}_\gamma^{(j)})$ denotes the minimax affine risk for the estimation of the parameter $\hbar(0)$ in the model $\{\widetilde{Q}_\hbar^{(j)} : \hbar \in \mathscr{H}_\gamma^{(j)}\}$, with loss function $\mathbb{1}_{[-v_j,v_j]^C}$, and $\widetilde{R}^*_{\text{aff}}(\hbar)$ denotes the minimax affine risk in the corresponding submodel $\{\widetilde{Q}_{\zeta\hbar}^{(j)} : \zeta \in [-1,1]\}$, with given $\hbar \in \mathscr{H}_\gamma^{(j)}$.

The results from Section 5.1 can now easily be transferred to the present setting. Again, we use the notation for the stochastic integrals introduced in Section 4.1. First we compute the minimax affine risk $\widetilde{R}^*_{\text{aff}}(\hbar)$ from the linear submodel corresponding to a fixed $\hbar \in \mathscr{H}_\gamma^{(j)}$. For symmetry reasons we can assume without loss of generality that $\hbar(0) > 0$.

5.10 Theorem. *Let $\hbar \in \mathscr{H}_\gamma^{(j)}$ be given, $\hbar(0) > 0$. We consider the loss function $\mathbb{1}_{[-v_j,v_j]^C}$. Let*

$$c_\hbar := \left(\frac{1}{2} + \left[\frac{1}{4} + \frac{\hbar(0)}{2v_j\|\hbar\|^2_{\lambda,2}} \log\frac{\hbar(0)+v_j}{\hbar(0)-v_j}\right]^{1/2}\right)^{-1} \mathbb{1}\{\hbar(0) > v_j\}.$$

Then $c_\hbar \int \hbar \, d\mathrm{x}_j \, \hbar(0)/\|\hbar\|^2_{\lambda,2}$ *is a minimax affine estimator for the parameter* $\zeta\hbar(0)$ *in the linear submodel* $\{\widetilde{Q}^{(j)}_{\zeta\hbar} : \zeta \in [-1,1]\}$*, the corresponding minimax affine risk is given by*

$$\widetilde{R}^*_{\mathrm{aff}}(\hbar) = \Bigg[\Phi\left(-\frac{\|\hbar\|_{\lambda,2}(1+v_j/\hbar(0))}{c_\hbar} + \|\hbar\|_{\lambda,2}\right) \\ +\Phi\left(\frac{\|\hbar\|_{\lambda,2}(1-v_j/\hbar(0))}{c_\hbar} - \|\hbar\|_{\lambda,2}\right)\Bigg] \mathbb{1}\{\hbar(0) > v_j\}.$$

Proof. The assertion is a direct consequence of the arguments from Section 5.1.1 and Theorem 5.1. There the quantities $Q_{\zeta g}$, g and δ have to be replaced by $\widetilde{Q}^{(j)}_{\zeta\hbar}$, $\hbar$ and v_j, respectively. Furthermore, the Cholesky matrix $\mathcal{C}_\gamma$ in the quantities from Section 5.1.1 has to be replaced by 1, because it is now already included in the function $\hbar$. The expressions for $c_\hbar$ and $\widetilde{R}^*_{\mathrm{aff}}(\hbar)$ then correspond to c_g and $R^*_{\mathrm{aff}}(g)$ from Theorem 5.1, respectively. Therefore, $c_\hbar(\int \hbar \, d\mathrm{x}_j/\|\hbar\|_{\lambda,2})$ is minimax affine for the estimation of $\zeta\|\hbar\|_{\lambda,2}$ (which corresponds to the parameter ϑ in Theorem 5.1), and thus $c_\hbar(\int \hbar \, d\mathrm{x}_j/\|\hbar\|_{\lambda,2}) \cdot (\hbar(0)/\|\hbar\|_{\lambda,2})$ is minimax affine for the estimation of $\zeta\hbar(0)$. ■

We want to show that there is a hardest submodel for the estimation of the parameter $\hbar(0)$ in the complete experiment $\{\widetilde{Q}^{(j)}_{\hbar} : \hbar \in \mathscr{H}^{(j)}_\gamma\}$, in other words, that a function $\hbar^* \in \mathscr{H}^{(j)}_\gamma$ exists for which the supremum on the right hand side of (5.41) is attained. To this end, we have to maximise the minimax affine risk $\widetilde{R}^*_{\mathrm{aff}}(\hbar)$ with respect to $\hbar$, ranging over $\mathscr{H}^{(j)}_\gamma$. For fixed $\alpha > 0$ we define the subspace

$$\mathscr{H}^{(j,\alpha)}_\gamma := \left\{\hbar \in \mathscr{H}^{(j)}_\gamma : \|\hbar\|_{\lambda,2} = \alpha\right\} \subseteq \mathscr{H}^{(j)}_\gamma.$$

For symmetry reasons, it is sufficient to restrict ourselves to the case $\hbar(0) > 0$. It is clear from Theorem 5.10 that $\widetilde{R}^*_{\mathrm{aff}}(\hbar)$ depends only on the value $\hbar(0)$, as $\hbar$ ranges over $\mathscr{H}^{(j,\alpha)}_\gamma$, and with the same arguments as in Section 5.1.3 we conclude that $\widetilde{R}^*_{\mathrm{aff}}(\hbar)$ is strictly increasing in $\hbar(0)$. In complete analogy to Proposition 5.3 this yields:

5.11 Proposition. *Only such functions* $\hbar \in \mathscr{H}^{(j,\alpha)}_\gamma$ *can describe a least favourable submodel for which the value* $\hbar(0)$ *is maximal.*

By definition each $\hbar \in \mathscr{H}_\gamma^{(j)}$ has the form

$$\hbar = \mathcal{C}_\gamma^j g = \sum_{i=1}^d \mathcal{C}_\gamma^{ji} g_i, \quad g = (g_1, \dots, g_d)^\top \in \mathscr{G},$$

and $\hbar(0) = \mathcal{C}_\gamma^j g(0)$. Let us consider the special functions

$$\begin{aligned} \hbar_\beta &:= \sum_{i=1}^d \mathcal{C}_\gamma^{ji} g_i^*, \\ g_i^*(x) &:= \operatorname{sgn}(\mathcal{C}_\gamma^{ji}) \, (\beta - |x|^\rho)^+ \; = \; \operatorname{sgn}(\mathcal{C}_\gamma^{ji}) g_\beta(x), \end{aligned} \tag{5.42}$$

with some $\beta > 0$, and with g_β according to (5.15). Here $\operatorname{sgn}(a) := \mathbb{1}\{a > 0\} - \mathbb{1}\{a < 0\}$ denotes the sign function. If β is chosen appropriately, then of course $\hbar_\beta \in \mathscr{H}_\gamma^{(j,\alpha)}$. From the definitions of the spaces $\mathscr{H}_\gamma^{(j)}$ and $\mathscr{G}$ and the growth condition in (3.10) one concludes that every function $\hbar$ with $\|\hbar\|_{\lambda,2} = \|\hbar_\beta\|_{\lambda,2}$ and $\hbar(0) \geq 0$ must also satisfy $\hbar(0) \leq \hbar_\beta(0)$. In other words, amongst all functions $\hbar$ with constant $\mathbb{L}_2$-norm the functions $\hbar_\beta$ of the type (5.42) have the largest possible value at the origin.

Due to this observation and Proposition 5.11 it therefore suffices to consider the functions $\hbar_\beta$ and the corresponding minimax affine risks $\widetilde{R}^*_{\text{aff}}(\hbar_\beta)$ in order to determine a hardest linear submodel. Note that $\hbar_\beta$ can also be written as

$$\hbar_\beta(x) = \|\mathcal{C}_\gamma^j\|_1 \, g_\beta(x) \tag{5.43}$$

(where $\|\cdot\|_1$ denotes the 1-norm), i.e. the functions $\hbar_\beta$ and g_β have completely the same structure, the only difference lies in a constant factor. Because of this observation, one will of course expect that the calculation of a hardest linear submodel for the problem of estimating the parameter $\hbar(0)$ in the experiment $\{\widetilde{Q}_\hbar^{(j)} : \hbar \in \mathscr{H}_\gamma^{(j)}\}$ can be done with the same techniques applied to the problem of finding a hardest submodel in the univariate case, discussed in Section 5.1.3. This is indeed the case.

5.12 Theorem. *Let $\beta_j > 0$ be the unique positive solution to the equation*

$$\frac{\beta_j^{1+1/\rho}}{\log \frac{\beta_j + v_j/\|\mathcal{C}_\gamma^j\|_1}{\beta_j - v_j/\|\mathcal{C}_\gamma^j\|_1}} = \frac{\rho+1}{2v_j\|\mathcal{C}_\gamma^j\|_1}, \tag{5.44}$$

and let $\hbar_{\beta_j}$ be defined according to (5.43) (with β_j replacing β). Then—using the notation from (5.36)—

$$\hat{\kappa}_j^* := \hat{\kappa}_j^*(\pi_j(\mathrm{x})) := \frac{1}{2\|\mathcal{C}_\gamma^j\|_1}\left(1+\frac{1}{\rho}\right)\beta_j^{-(1+1/\rho)}\int \hbar_{\beta_j}\,d\mathrm{x}_j \tag{5.45}$$

is a minimax affine estimator for $\hbar(0)$ in the model $\{\widetilde{Q}_\hbar^{(j)} : \hbar \in \mathscr{H}_\gamma^{(j)}\}$, given the loss function $\mathbb{1}_{[-v_j,v_j]^C}$, and the minimax affine risk is given by

$$\begin{aligned}\Phi\left(-\frac{\|\mathcal{C}_\gamma^j\|_1\beta_j^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}\left[1+\frac{(2\rho+1)v_j}{\|\mathcal{C}_\gamma^j\|_1\beta_j}\right]\right)\\ +\Phi\left(\frac{\|\mathcal{C}_\gamma^j\|_1\beta_j^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}\left[1-\frac{(2\rho+1)v_j}{\|\mathcal{C}_\gamma^j\|_1\beta_j}\right]\right).\end{aligned} \tag{5.46}$$

Proof. Due to the preceding arguments we only have to maximise $\widetilde{R}^*_{\mathrm{aff}}(\beta_j) := \widetilde{R}^*_{\mathrm{aff}}(\hbar_{\beta_j})$ with respect to β_j. By definition, we have

$$\|\hbar_\beta\|_{\lambda,2} = \|\mathcal{C}_\gamma^j\|_1\ \|g_{\beta_j}\|_{\lambda,2} \quad \text{and} \quad \hbar_{\beta_j}(0) = \|\mathcal{C}_\gamma^j\|_1\ \beta_j.$$

Plugging this into the formulae from Theorem 5.10, we get

$$\begin{aligned}\widetilde{R}^*_{\mathrm{aff}}(\beta_j) = \Bigg[&\Phi\left(-\frac{\|\mathcal{C}_\gamma^j\|_1\ \|g_{\beta_j}\|_{\lambda,2}\left(1+\frac{v_j/\|\mathcal{C}_\gamma^j\|_1}{\beta_j}\right)}{c_{\hbar_{\beta_j}}}+\|\mathcal{C}_\gamma^j\|_1\ \|g_{\beta_j}\|_{\lambda,2}\right)\\ &+\Phi\left(\frac{\|\mathcal{C}_\gamma^j\|_1\ \|g_{\beta_j}\|_{\lambda,2}\left(1-\frac{v_j/\|\mathcal{C}_\gamma^j\|_1}{\beta_j}\right)}{c_{\hbar_{\beta_j}}}-\|\mathcal{C}_\gamma^j\|_1\ \|g_{\beta_j}\|_{\lambda,2}\right)\Bigg]\\ &\mathbb{1}\left\{\beta_j > \frac{v_j}{\|\mathcal{C}_\gamma^j\|_1}\right\},\end{aligned}$$

with

$$\begin{aligned}c_{\hbar_{\beta_j}} = \Bigg(\frac{1}{2}+\Bigg[\frac{1}{4}+&\frac{\beta_j}{2\frac{v_j}{\|\mathcal{C}_\gamma^j\|_1}\|\mathcal{C}_\gamma^j\|_1^2\ \|g_{\beta_j}\|_{\lambda,2}^2}\\ &\log\frac{\beta_j+(v_j/\|\mathcal{C}_\gamma^j\|_1)}{\beta_j-(v_j/\|\mathcal{C}_\gamma^j\|_1)}\Bigg]^{1/2}\Bigg)^{-1}\mathbb{1}\left\{\beta_j > \frac{v_j}{\|\mathcal{C}_\gamma^j\|_1}\right\}.\end{aligned}$$

Note that these expressions for $\widetilde{R}^*_{\mathrm{aff}}(\beta_j)$ and $c_{\hbar_{\beta_j}}$ correspond to those for $R^*_{\mathrm{aff}}(\beta)$ and c_β given in Remark (5.5), if there β, δ and $\mathcal{C}_\gamma$

are replaced by β_j, $v_j/\|\mathcal{C}^j_\gamma\|_1$ and $\|\mathcal{C}^j_\gamma\|_1$, respectively. With these changes of the constants, a copy of the proof of Theorem 5.4 yields that $\widetilde{R}^*_{\text{aff}}(\beta_j)$ is maximal for β_j satisfying (5.44), in which case $c_{\hbar_{\beta_j}} = 2\rho/(2\rho+1)$. Furthermore, the approach from Section 3.5.2 for the specification of the constant K makes sure that the function $\hbar_{\beta_j} = \|\mathcal{C}^j_\gamma\|_1\, g_{\beta_j}$ corresponding to that β_j (as solution to (5.44)) is included in $\mathscr{H}^{(j)}_\gamma$ and therefore defines a hardest linear submodel. Plugging $c_{\hbar_{\beta_j}} = (2\rho)/(2\rho+1)$ and the integral from (5.21) into the above expression for the minimax affine risk completes the proof. ■

5.13 Remark. As in the proof of Theorem 5.4 it now becomes clear why such a complicated approach was chosen in Section 3.5 in order to specify the constant K for the definition of the space $\mathscr{G} = \mathscr{G}(K,\rho,d)$. This approach is adapted to the needs of the above theorem and guarantees that a hardest linear submodel for the estimation of $\hbar(0)$ exists.

We know that the experiments $\{\widetilde{Q}^{(j)}_\hbar : \hbar \in \mathscr{H}^{(j)}_\gamma\}$ and $\{Q^{(j)}_g : g \in \mathscr{G}\}$ can be identified, and therefore the (affine) minimax riks in the corresponding models are the same. Hence, the estimator $\hat\kappa^*_j$, defined according to (5.45), is minimax affine (with respect to the loss function $\mathbb{1}_{[-v_j,v_j]^C}$) for the estimation of the parameter $\mathcal{C}^j_\gamma g(0)$ in the model $\{Q^{(j)}_g : g \in \mathscr{G}\}$. For the estimation of the *complete* vector $\mathcal{C}_\gamma g(0)$ we now consider the estimator

$$\hat\kappa^* := (\hat\kappa^*_1, \ldots, \hat\kappa^*_d)^\top.$$

By definition, $\hat\kappa^*$ satisfies (5.36), i.e. $\hat\kappa^*_j(\mathrm{x}) = \hat\kappa^*_j(\pi_j(\mathrm{x}))$ for all j, and with the above results we can calculate an upper bound on its maximal risk under the loss function $\mathbb{1}_{[-v,v]^C}$. We have

$$
\begin{aligned}
&\sup_{g\in\mathscr{G}} Q_g\big\{\hat{\kappa}^* - \mathcal{C}_\gamma g(0) \in [-v,v]^C\big\} \\
&\overset{(5.37)}{=} \sup_{g\in\mathscr{G}} \left(1 - \prod_{j=1}^{d} Q_g^{(j)}\{|\hat{\kappa}_j^* - \mathcal{C}_\gamma^j g(0)| \le v_j\}\right) \\
&\le\ 1 - \prod_{j=1}^{d} \inf_{g\in\mathscr{G}}\ Q_g^{(j)}\{|\hat{\kappa}_j^* - \mathcal{C}_\gamma^j g(0)| \le v_j\} \\
&=\ 1 - \prod_{j=1}^{d}\left(1 - \sup_{g\in\mathscr{G}}\ Q_g^{(j)}\{|\hat{\kappa}_j^* - \mathcal{C}_\gamma^j g(0)| > v_j\}\right) \\
&\overset{(5.46)}{=} 1 - \prod_{j=1}^{d}\Bigg\{1 - \Phi\left(-\frac{\|\mathcal{C}_\gamma^j\|_1 \beta_j^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}\left[1 + \frac{(2\rho+1)v_j}{\|\mathcal{C}_\gamma^j\|_1\beta_j}\right]\right) \\
&\qquad\qquad - \Phi\left(\frac{\|\mathcal{C}_\gamma^j\|_1 \beta_j^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}\left[1 - \frac{(2\rho+1)v_j}{\|\mathcal{C}_\gamma^j\|_1\beta_j}\right]\right)\Bigg\}. \qquad (5.47)
\end{aligned}
$$

Evidently, the maximal risk of any estimator $\hat{\kappa}$ for the parameter $\mathcal{C}_\gamma g(0)$ is an upper bound for the minimax risk. In other words, the expression from (5.47) is an upper bound for $\widetilde{R}^*(\mathscr{G})$, the minimax risk for the estimation of $\mathcal{C}_\gamma g(0)$ with respect to the loss function $\mathbb{1}_{[-v,v]^C}$. We can now show that this also yields an upper bound for $R^*(\mathscr{G})$, the minimax risk with respect to the loss function $\mathbb{1}_{[-\delta,\delta]^C}$ for the estimation of $g(0)$. To this end, we consider for $g(0)$ the estimator

$$\hat{g}(0) := \mathcal{C}_\gamma^{-1}\hat{\kappa}^*,$$

where $\hat{\kappa}^*$ is the estimator with components $\hat{\kappa}_j^*$ according to Theorem 5.12. We assess the maximal risk of $\hat{g}(0)$. In the following, let

$$A_\gamma = \{\mathcal{C}_\gamma x : x \in [-\delta,\delta]\}$$

be as in Section 3.5.3. Keep in mind, that in that section we chose the vector $v = v(\gamma)$ such that the inclusion

$$[-v,v] \subseteq A_\gamma.$$

holds. From this we get the inequality

$$\begin{aligned} R^*(\mathscr{G}) &\le \sup_{g\in\mathscr{G}} Q_g\big\{\hat g(0) - g(0) \in [-\delta,\delta]^C\big\} \\ &= \sup_{g\in\mathscr{G}} Q_g\big\{\hat\kappa^* - \mathcal{C}_\gamma g(0) \in (A_\gamma)^C\big\} \\ &\le \sup_{g\in\mathscr{G}} Q_g\big\{\hat\kappa^* - \mathcal{C}_\gamma g(0) \in [-v,v]^C\big\}. \end{aligned} \tag{5.48}$$

Therefore, the expression in (5.47) is an upper bound for $R^*(\mathscr{G})$, too. We summarise the main results from the considerations of this section in the following theorem:

5.14 Theorem. *Let the loss function $\mathbb{1}_{[-\delta,\delta]^C}$ be given, and let $v = (v_1,\dots,v_d)^\top$ be constructed according to the procedure provided in Section 3.5.3. Then the estimator $\hat g(0) = \mathcal{C}_\gamma^{-1}\hat\kappa^*$—with $\hat\kappa^* = (\hat\kappa_1^*,\dots,\hat\kappa_d^*)^\top$ according to Theorem 5.12—satisfies*

$$\begin{aligned} R^*(\mathscr{G}) &\le \sup_{g\in\mathscr{G}} Q_{\gamma,g}\big\{\hat g(0) - g(0) \in [-\delta,\delta]^C\big\} \\ &\le 1 - \prod_{j=1}^d \Bigg\{ 1 - \Phi\Bigg(-\frac{\|\mathcal{C}_\gamma^j\|_1 \beta_j^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}} \Bigg[1 + \frac{(2\rho+1)v_j}{\|\mathcal{C}_\gamma^j\|_1\beta_j} \Bigg] \Bigg) \\ &\qquad - \Phi\Bigg(\frac{\|\mathcal{C}_\gamma^j\|_1 \beta_j^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}} \Bigg[1 - \frac{(2\rho+1)v_j}{\|\mathcal{C}_\gamma^j\|_1\beta_j} \Bigg] \Bigg) \Bigg\}. \end{aligned}$$

Addendum. *In the special case $d = 1$ (i.e. for a one-dimensional parameter space) the theorem coincides with Theorem 5.4, since in this case $\|\mathcal{C}_\gamma^1\|_1 = \mathcal{C}_\gamma$ and $v = \delta\mathcal{C}_\gamma$. In particular $\hat g(0)$ is then a minimax affine estimator, and the last inequality from the preceding display holds with equality.*

5.3.2 A lower minimax risk bound

As before, $R^*(\mathscr{G})$ denotes the minimax risk for the estimation of the unknown parameter $g(0)$ with respect to the loss function $\mathbb{1}_{[-\delta,\delta]^C}$. So far, we have derived an upper bound for $R^*(\mathscr{G})$. Now we address the question of a suitable lower bound. We use similar ideas as in Section 5.1.4. Let us first consider the minimax risk in some specific linear submodels of the form

$$\{Q_{\gamma,\zeta g} : \zeta \in [-1,1]\},$$

with fixed $g \in \mathscr{G}$. Of course, the minimax risk in any such submodel is smaller than the minimax risk in the complete model, and this yields

$$R^*(\mathscr{G}) \;\geq\; \sup_{g\in\mathscr{G}} \left(\inf_{\hat\zeta} \;\sup_{\zeta\in[-1,1]} \; Q_{\gamma,\zeta g}\big\{\hat\zeta g(0) - \zeta g(0) \in [-\delta,\delta]^C\big\} \right). \tag{5.49}$$

We now aim to derive lower bounds for the expression in the brackets, which yield a lower bound for $R^*(\mathscr{G})$, too.

To this end, we restrict ourselves to such functions $g \in \mathscr{G}$ which do only in one component, say the kth, differ from zero. In other words, we consider special functions of the form

$$g = (0,\dots,g_k,\dots,0)^\top \in \mathscr{G}, \quad g_k(0) \neq 0. \tag{5.50}$$

Using the notation for the stochastic integrals from Section 4.1, we consider the random variable

$$Z_g := \frac{\int \mathcal{C}_\gamma g \; d\mathrm{x}}{(\sum_{i=1}^d \|\mathcal{C}_\gamma^{ik} g_k\|_{\lambda,2}^2)^{1/2}} = \frac{\sum_{i=1}^d \int \mathcal{C}_\gamma^{ik} g_k \; d\mathrm{x}_i}{(\sum_{i=1}^d \|\mathcal{C}_\gamma^{ik} g_k\|_{\lambda,2}^2)^{1/2}}.$$

The stochastic integral calculus yields

$$\mathscr{L}(\textstyle\int \mathcal{C}_\gamma^{ik} g_k \; d\mathrm{x}_i \mid Q_{\gamma,0}) = \mathscr{N}(0, \|\mathcal{C}_\gamma^{ik} g_k\|_{\lambda,2}^2),$$

which in combination with the independence of the components x_i implies that

$$\mathscr{L}(Z_g \mid Q_{\gamma,0}) = \mathscr{N}(0,1).$$

Furthermore, setting

$$a_g := \left(\sum_{i=1}^d \|\mathcal{C}_\gamma^{ik} g_k\|_{\lambda,2}^2 \right)^{1/2}$$

we conclude that

$$\mathscr{L}(Z_g \mid Q_{\gamma,\zeta g}) = \mathscr{N}(\vartheta, 1), \quad \vartheta := \zeta a_g$$

which follows for example from Problem 5.6 in Karatzas and Shreve (1988). Hence, we obtain again a Gaussian location model

$$\{\mathscr{N}(\vartheta,1) : \vartheta \in [-a_g, a_g]\}.$$

Given an estimator $\hat{\vartheta}$ within this model, then $\hat{\zeta} := \hat{\vartheta}/a_g$ is a an estimator for ζ that satisfies

$$\begin{aligned} \hat{\zeta}g(0) - \zeta g(0) \in [-\delta, \delta]^C \;&\Leftrightarrow\; |\hat{\zeta}g_k(0) - \zeta g_k(0)| > \delta_k \\ &\Leftrightarrow\; |\hat{\vartheta} - \vartheta| > \frac{\delta_k a_g}{|g_k(0)|}. \end{aligned}$$

This equivalence implies that

$$\begin{aligned} &\inf_{\hat{\zeta}} \sup_{\zeta \in [-1,1]} Q_{\gamma,\zeta g}\{\hat{\zeta}g(0) - \zeta g(0) \in [-\delta, \delta]^C\} \\ &\qquad = \inf_{\hat{\zeta}} \sup_{\zeta \in [-1,1]} Q_{\gamma,\zeta g}\{|\hat{\zeta}g_k(0) - \zeta g_k(0)| > \delta_k\} \\ &\qquad = \inf_{\hat{\vartheta}} \sup_{|\vartheta| \le a_g} \mathscr{N}_{(\vartheta,1)}\{|\hat{\vartheta} - \vartheta| > \delta_k a_g/|g_k(0)|\}. \end{aligned}$$

If we take the supremum on the right hand side of (5.49) not with respect to all possible functions $g \in \mathscr{G}$, but only with respect to those g which are of the special form (5.50) (for some $k \in \{1, \ldots, d\}$), we conclude from the above equation that

$$\begin{aligned} &R^*(\mathscr{G}) \\ &\ge \max_k \sup_{g=(0,\ldots,g_k,\ldots,0)^\top} \left(\inf_{\hat{\vartheta}} \sup_{|\vartheta| \le a_g} \mathscr{N}_{(\vartheta,1)}\{|\hat{\vartheta} - \vartheta| > \delta_k a_g/|g_k(0)|\} \right). \end{aligned}$$

The expression in brackets on the right hand side of this inequality corresponds to the quantity $R^*(g)$ from (5.25), if there τ_g and δ_g are replaced with a_g and $\delta_k a_g/|g_k(0)|$, respectively. We can therefore compute this expression with the same ideas that were presented in Section 5.1.4. This yields a lower bound for $R^*(\mathscr{G})$.

In combination with the results from Section 4.3 this also permits the computation of a lower minimax risk bound in the global model $(\mathcal{Y}^n, \mathcal{B}^n, \{P_\xi^{(n)} : \xi \in \mathscr{F}\})$.

5.4 Discussion

Theorems 5.4 and 5.14, respectively, provide devices to calculate concrete upper bounds for the minimax risk $R^*(\mathscr{G})$ in the limit experiment. Of course, the question arises how precise these upper bounds really are.

The univariate parameter case. Again, we first consider the case of a univariate parameter, investigated in Section 5.1, i.e. we consider the limit experiment $\{Q_{\gamma,g} : g \in \mathscr{G}\}$ with $\mathscr{G} = \mathscr{G}(K, \rho, 1)$. In this case the upper minimax risk bound from Theorem 5.4 equals the minimax affine risk $R^*_{\text{aff}}(\mathscr{G})$. Thus, the behaviour of the ratio $R^*_{\text{aff}}(\mathscr{G})/R^*(\mathscr{G})$, in words: the quotient of the minimax affine risk and the minimax risk, is of interest. The closer this quotient is to 1, the more accurate is the upper bound. For the problem of estimating the mean in a Gaussian location model, and using quadratic loss, it is known from the literature that there is a constant $\mu^* \leq 1.25$—sometimes referred to as *Ibragimov-Khas'minskii constant*—which serves as an upper bound for the quotient of the corresponding minimax affine risk and the minimax risk. Similar results do also hold for other performance criteria than quadratic loss. (For more detailed information on this topic, see e.g. Donoho, 1994, Donoho et al., 1990, Drees, 1999 and Ibragimov and Khas'minski, 1985, as well as the references within these papers.) From that one may draw the conclusion that affine estimators are in general only insignificantly less good than other, arbitrary estimation procedures. This suggests that for the estimation of the parameter $g(0)$ in the limit experiment, with loss function $\mathbb{1}_{[-\delta,\delta]^C}$ the restriction to affine estimators should not lead to a great additional loss, and one might expect that the quotient $R^*_{\text{aff}}(\mathscr{G})/R^*(\mathscr{G})$ behaves fairly well.

The multivariate parameter case. In the multivariate case—i.e. for the model $\{Q_{\gamma,g} : g \in \mathscr{G}\}$ with $\mathscr{G} = \mathscr{G}(K, \rho, d)$ and $d \geq 1$—the situation is considerably more complicated. In contrast to the univariate case, the results from the multivariate case use several additional inequalities. The cruder these inequalities are, the bigger, and hence less accurate becomes of course the derived upper bound. The most critical of these inequalities is certainly that in (5.48), where we turn over from the set $(A_\gamma)^C$ to $[-v, v]^C \supseteq (A_\gamma)^C$, which is in general a strict inclusion. Hence, (5.48) is in most cases a strict inequality, which may in the worst case lead to unnecessarily large upper minimax risk bounds. It is therefore reasonable to think about conditions under which (5.48) holds with equality.

To this end, let us for a moment consider the special case that $\mathcal{C}_\gamma$ is a diagonal matrix, say $\mathcal{C}_\gamma = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$, with $\lambda_i > 0$.

With our special choice of $v = v(\gamma)$ according to Section 3.5.3 we then have $v_j = \lambda_j \delta_j$ and thus $A_\gamma = [-v, v]$. In this case (5.48) holds with equality. Consequently, given the underlying distribution family $\{P_\theta : \theta \in \Theta\}$, it appears favourable to find an *orthogonal* parametrisation, i.e. a parametrisation under which the information matrices $\mathcal{I}_\gamma$—and therefore also their Cholesky decompositions $\mathcal{C}_\gamma$—have diagonal form. Huzurbazar (1950) shows that such an orthogonal parametrisation can be found by solving a system of $d(d-1)/2$ differential equations. He shows that for $d = 2$ this problem is usually solvable, whereas for $d > 3$ this is in general not the case. Mitchell (1962) extends these results and discusses certain distribution families with three-dimensional parameter for which the pertaining system of differential equations is also solvable. We conclude: if it is possible to find an orthogonal parametrisation, this will certainly lead to an improvement of the upper minimax risk bound, but unfortunately not every distribution family possesses such an orthogonal parametrisation.

We therefore want to briefly sketch a few ideas for modifications and alternative approaches to derive upper minimax risk bounds.

The most simple modification of the approach from Section 5.3 is of course to use a different $v = (v_1, \dots, v_d)^\top$. So far, we only made use of the fact that the components v_i are strictly positive and—for the specification of the constant K in Section 3.5.2—continuous in γ, and that in addition $[-v, v] \subseteq A_\gamma$. Of course, there are other possibilities to chose an appropriate v that fits these conditions. The aim of any such choice should be to define v in such a way that the resulting upper bound (5.47) is preferably small. Thus, the components v_j should be as large as possible. One approach to determine an optimal vector v is to minimise the expression from (5.47) with respect to the components v_j. This is however a very complex task, because the β_j's, as solutions of (5.44), also depend on the components v_j. Moreover, the specification of a suitable constant K (according to Section 3.5.2), that guarantees the existence of a hardest linear submodel, becomes very complicated, too.

In order to derive upper bounds for $R^*(\mathscr{G})$ we did in Section 5.3 not estimate $g(0)$ directly, but made a long way round, by first estimating $\mathcal{C}_\gamma g(0)$ with respect to the loss function $\mathbb{1}_{[-v,v]^C}$. This contributes to and increases the inaccuracy of the provided upper bounds. One may therefore ask if the idea of first estimating $\mathcal{C}_\gamma g(0)$ is adequate at all. Instead, one might as well try to estimate $g(0)$ directly. A reasonable approach might be to try and boil down this problem to that of estimating a mean vector in some Gaussian model, similar to the ideas from Section 5.1. The work of Donoho et al. (1990) might provide some hints on how to follow such an approach.

To sum up, there are certainly reasonable alternatives to the approach that we chose in Section 5.3, which however are very complex to fully investigate. An examination of these would go far beyond the scope of this thesis, and it is a priori not clear that these alternatives really lead to better (i.e. smaller) upper minimax risk bounds than our approach from Section 5.3 does. Of course, this approach has some drawbacks, and may not yield the best possible upper bounds for the minimax risk. In return, it is however somewhat convenient, because it allows us to reduce the examinations in the multivariate case to those of the univariate case, and the procedure yields expressions for the upper bounds that are rather simple to calculate (e.g. no integrals have to be calculated). This might be a benefit in practice. Note also that Theorem 5.14 always provides an upper bound which is smaller than 1 and thus of at least some worth.

CHAPTER 6

Upper minimax risk bounds in the local model

So far we proved that lower minimax risk bounds in the limit experiment $\{Q_{\gamma,g} : g \in \mathscr{G}\}$ always provide lower asymptotic minimax risk bounds in both the local and the global model. We now want to show that such a statement also holds true for upper bounds. In this chapter we consider the sequence of local models

$$(\mathcal{Y}^n, \mathcal{B}^n, \{P^{(n)}_{\gamma,g} : g \in \mathscr{G}\}), \tag{6.1}$$

in which we want to estimate the local parameter $g(0)$. The aim is to derive an upper bound for the corresponding asymptotic minimax risk. To this end, we first construct an estimator $\hat{\eta}_n$ within the model (6.1). This estimator is essentially based upon and motivated by the results from the previous chapters. Subsequently, we show that the maximal risk of this estimator—and hence, the minimax risk in the local model—is asymptotically bounded from above by the expressions from Theorems 5.4 and 5.14, respectively.

Throughout this chapter we presume that the assumptions from Sections 3.1–3.3 hold. (Note that the additional regularity conditions for the family $\{P_\theta : \theta \in \Theta\}$ have not been needed in the preceding chapters.) As in Chapter 5 we write the Cholesky matrix as $\mathcal{C}_\gamma = (\mathcal{C}^{ij}_\gamma)$, the ith row vector of $\mathcal{C}_\gamma$ is written $\mathcal{C}^i_\gamma = (\mathcal{C}^{i1}_\gamma, \ldots, \mathcal{C}^{id}_\gamma)$.

6.1 Constructing an estimator for the local parameter

Throughout this section let $\gamma \in \Theta'$ denote a fixed centre of localisation ($\Theta' \subseteq \Theta$ as defined in (3.7)). Within the corresponding local model (6.1) we construct an estimator for the parameter $g(0)$. To better motivate our choice of such an estimator, we first consider the more simple case of a univariate parameter space Θ; subsequently, we generalise these ideas to the case $\Theta \subseteq \mathbb{R}^d$. Moreover, first results on the asymptotic behaviour of this estimator are formulated.

6.1.1 A local estimator for the univariate parameter case

For the sake of simplicity let us first consider the special case of a one-dimensional parameter, i.e. we assume that $\gamma \in \Theta' \subseteq \mathbb{R}$ and $\mathscr{G} = \mathscr{G}(K, \rho, 1)$. In this case the limit experiment for the sequence of the local experiments is given by

$$(C^1([-K, K]), \mathscr{C}^1([-K, K]), \{Q_{\gamma,g} : g \in \mathscr{G}\}).$$

According to Theorem 5.4

$$\hat{g}(0) = \frac{1}{2\mathcal{I}_\gamma}\left(1 + \frac{1}{\rho}\right)\beta^{-(1+1/\rho)} \int \mathcal{C}_\gamma g_\beta \, d\mathrm{x}$$

is a minimax affine estimator for the local parameter $g(0)$ within this model, where $\beta = \beta(\gamma)$ is the solution to (5.16) and g_β is defined according to (5.15). By virtue of the Girsanov formula (4.3) and Corollary 4.5 we have

$$\log \frac{dQ_{\gamma,g_\beta}}{dQ_{\gamma,0}} = \int \mathcal{C}_\gamma g_\beta \, d\mathrm{x} - \frac{1}{2}\int g_\beta(s)\mathcal{I}_\gamma g_\beta(s)\, ds,$$

and

$$\log \frac{dP^{(n)}_{\gamma,g_\beta}}{dP^{(n)}_\gamma} = \sum_{j=1}^n U_{nj}^{g_\beta} - \frac{1}{2}\int g_\beta(s)\mathcal{I}_\gamma g_\beta(s)\, ds + o_{P^{(n)}_\gamma}(1),$$

respectively, with $U_{nj}^{g_\beta} = a_n g_\beta(\tilde{x}_{nj})\dot{\ell}_\gamma$ according to (4.5). Obviously, the random variable $\sum_{j=1}^n U_{nj}^{g_\beta}$ from the second display corresponds to the stochastic integral $\int \mathcal{C}_\gamma g_\beta \, d\mathrm{x}$ from the first display. Thus, taking the structure of the above estimator $\hat{g}(0)$ from the limit experiment as a basis, it is somewhat intuitive to choose

$$\frac{1}{2\mathcal{I}_\gamma}\left(1 + \frac{1}{\rho}\right)\beta^{-(1+1/\rho)} \sum_{j=1}^n U_{nj}^{g_\beta} \tag{6.2}$$

as an estimator for the unknown parameter $g(0)$ in the local experiment. Note that the parameter $\beta = \beta(\gamma)$ is a function of the previously fixed centre of localisation γ, and therefore the proposed estimator depends on γ, too.

6.1.2 The general multivariate parameter case

The ideas for the case of a one-dimensional parameter can easily be transferred to the general case of a (possibly) multivariate parameter space $\Theta \subseteq \mathbb{R}^d$, $d \geq 1$. In this case the corresponding Gaussian limit experiment is given by

$$(C^d([-K,K]), \mathscr{C}^d([-K,K]), \{Q_{\gamma,g} : g \in \mathscr{G}\}),$$

with $\mathscr{G} = \mathscr{G}(K,\rho,d)$. In Section 5.3 we considered

$$\hat{g}(0) = \mathcal{C}_\gamma^{-1}\hat{\kappa}^*$$

as an estimator for $g(0)$, with $\hat{\kappa}^* = (\hat{\kappa}_1^*, \dots, \hat{\kappa}_d^*)^\top$ and

$$\begin{aligned} \hat{\kappa}_i^* &:= \hat{\kappa}_i^*(\pi_i(\mathrm{x})) \\ &:= \frac{1}{2\|\mathcal{C}_\gamma^i\|_1}\left(1+\frac{1}{\rho}\right)\beta_i^{-(1+1/\rho)}\int \hbar_{\beta_i}\, d\mathrm{x}_i \\ &\overset{(5.43)}{=} \frac{1}{2}\left(1+\frac{1}{\rho}\right)\beta_i^{-(1+1/\rho)}\int g_{\beta_i}\, d\mathrm{x}_i \end{aligned} \tag{6.3}$$

according to Theorem 5.12. The parameter $\beta_i = \beta_i(\gamma)$ is defined as the solution to equation (5.44), the corresponding functions g_{β_i} are defined according to (5.15). For the following considerations we define

$$G_\gamma^i := \mathcal{C}_\gamma^{-1}(0, \dots, g_{\beta_i(\gamma)}, \dots, 0)^\top, \quad i = 1, \dots, d. \tag{6.4}$$

Using the conventions on the stochastic integral notation from Section 4.1 we thus have

$$\begin{aligned} \int \mathcal{C}_\gamma G_\gamma^i\, d\mathrm{x} &= \int (0, \dots, g_{\beta_i(\gamma)}, \dots, 0)^\top\, d\mathrm{x} \\ &\overset{\mathscr{L}}{=} \int g_{\beta_i(\gamma)}\, d\mathrm{x}_i. \end{aligned}$$

In addition, we define a family of random variables

$$\begin{aligned} U_{nj}^{\gamma,i} &:= a_n G_\gamma^i(\tilde{x}_{nj})^\top \dot{\ell}_\gamma, \\ U_n^{\gamma,i} &:= \sum_{j=1}^n U_{nj}^{\gamma,i}, \end{aligned} \tag{6.5}$$

according to (4.5) and (4.6), respectively. The $U_{nj}^{\gamma,i}$, $j = 1, \ldots, n$, are assumed independent. With similar arguments as in the above case of a one-dimensional parameter one may now replace the stochastic integrals $\int g_{\beta_i(\gamma)}\ d\mathrm{x}_i = \int \mathcal{C}_\gamma G_\gamma^i\ d\mathrm{x}$ in (6.3) by $U_n^{\gamma,i}$ in order to get an estimator within the local models. Setting

$$\begin{aligned} S_i(\gamma) &:= \frac{1}{2}\left(1 + \frac{1}{\rho}\right) \beta_i(\gamma)^{-(1+1/\rho)}, \quad i = 1, \ldots, d, \\ S_\gamma &:= \operatorname{diag}(S_1(\gamma), \ldots, S_d(\gamma)), \end{aligned} \tag{6.6}$$

we define an estimator $\hat{\eta}_n$ for the local parameter $g(0)$ by

$$\hat{\eta}_n := \hat{\eta}_n(\gamma) := \mathcal{C}_\gamma^{-1} S_\gamma (U_n^{\gamma,1}, \ldots, U_n^{\gamma,d})^\top. \tag{6.7}$$

Note that for $d = 1$ this estimator coincides with that proposed in (6.2).

6.1 Remark. As in Section 4.2 we interpret $U_{nj}^{\gamma,i}$ and $U_n^{\gamma,i}$ as random variables on the product space $(\mathcal{Y}^n, \mathcal{B}^n)$. To emphasize this, we occasionally write $U_{nj}^{\gamma,i} = U_{nj}^{\gamma,i}(Y_{nj})$, with the interpretation of Y_{nj} as the jth coordinate projection on the product space $\mathcal{Y}^n$ (see also the comments on p. 37 and Remark 2.9). Then $\hat{\eta}_n$ can be written $\hat{\eta}_n = \hat{\eta}_n(\gamma, Y_{n1}, \ldots, Y_{nn})$.

6.1.3 Asymptotic normality

The aim of this section is to prove that $\hat{\eta}_n$ is asymptotically normal under $P_{\gamma,g}^{(n)}$, the product measure corresponding to the local alternative $\gamma_{g,n}(\cdot)$ (see (1.4) and (1.3), respectively). For that purpose we first introduce some new notation. For $g \in \mathscr{G}$ and $\gamma \in \Theta'$ we define

$$\begin{aligned} e_{\gamma,g} &:= \left(\textstyle\int g(t)^\top \mathcal{I}_\gamma G_\gamma^1(t)\ dt, \ldots, \int g(t)^\top \mathcal{I}_\gamma G_\gamma^d(t)\ dt\right)^\top, \\ \mu_{\gamma,g} &:= (\mathcal{C}_\gamma^{-1} S_\gamma) e_{\gamma,g} - g(0), \\ \mathcal{J}_\gamma &:= \operatorname{diag}(\|g_{\beta_1(\gamma)}\|_{\lambda,2}^2, \ldots, \|g_{\beta_d(\gamma)}\|_{\lambda,2}^2), \\ \Sigma_\gamma &:= (\mathcal{C}_\gamma^{-1} S_\gamma) \mathcal{J}_\gamma (\mathcal{C}_\gamma^{-1} S_\gamma)^\top. \end{aligned} \tag{6.8}$$

In case that a fixed centre of localisation γ is considered (like in this section), we often drop the argument γ in the quantities introduced in (6.8), or in other variables that depend on γ.

6.2 Theorem. *Let a fixed $\gamma \in \Theta'$ be given and let $\hat{\eta}_n = \hat{\eta}_n(\gamma)$ be the local estimator from (6.7). Then*

$$\mathscr{L}(\hat{\eta}_n - g(0) \mid P^{(n)}_{\gamma,g}) \rightsquigarrow \mathscr{N}(\mu_{\gamma,g}, \Sigma_\gamma)$$

for all $g \in \mathscr{G}$.

Proof. Proposition 4.6, in combination with Remark 4.9, implies

$$\mathscr{L}\left(\left(\log \frac{dP^{(n)}_{\gamma,G^1_\gamma}}{dP^{(n)}_\gamma}, \dots, \log \frac{dP^{(n)}_{\gamma,G^d_\gamma}}{dP^{(n)}_\gamma}, \log \frac{dP^{(n)}_{\gamma,g}}{dP^{(n)}_\gamma}\right)^\top \,\middle|\, P^{(n)}_\gamma\right) \rightsquigarrow \mathscr{N}\left(-\frac{1}{2}\begin{pmatrix} V \\ \sigma^2 \end{pmatrix}, \begin{pmatrix} \mathcal{J} & e_g \\ e_g^\top & \sigma^2 \end{pmatrix}\right),$$

where $V = (\|g_{\beta_1}\|^2_{\lambda,2}, \dots, \|g_{\beta_d}\|^2_{\lambda,2})^\top$ and $\sigma^2 = \int g(s)^\top \mathcal{I}_\gamma g(s)\ ds$. With Corollary 4.5 and Slutsky's lemma we further conclude that

$$\mathscr{L}\left(\left(U^{\gamma,1}_n, \dots, U^{\gamma,d}_n, \log \frac{dP^{(n)}_{\gamma,g}}{dP^{(n)}_\gamma}\right)^\top \,\middle|\, P^{(n)}_\gamma\right) \rightsquigarrow \mathscr{N}\left(\begin{pmatrix} 0 \\ -\frac{1}{2}\sigma^2 \end{pmatrix}, \begin{pmatrix} \mathcal{J} & e_g \\ e_g^\top & \sigma^2 \end{pmatrix}\right).$$

The assertion follows now from Le Cam's third lemma (Lemma 2.3) and an application of the delta method (van der Vaart, 1998, Theorem 3.1). ■

In combination with the portmanteau theorem the above theorem implies that

$$P^{(n)}_{\gamma,g}\{\hat{\eta}_n - g(0) \in A\} \to \mathscr{N}_{(\mu_g,\Sigma)}(A)$$

for all sets $A \in \mathbb{B}$ satisfying $\mathscr{N}_{(\mu_g,\Sigma)}(\partial A) = 0$. In particular, this holds for the choice $A = [-\delta, \delta]^C$ (with $\delta = (\delta_1, \dots, \delta_d)^\top$ from Section 3.3), and thus

$$P^{(n)}_{\gamma,g}\big\{\hat{\eta}_n - g(0) \in [-\delta, \delta]^C\big\} \to \mathscr{N}_{(\mu_g,\Sigma)}\big([-\delta, \delta]^C\big). \tag{6.9}$$

In other words: given a fixed γ and a local function $g \in \mathscr{G}$, the risk of $\hat{\eta}_n$ under the loss function $\mathbb{1}_{[-\delta,\delta]^C}$ converges to the limit given in the preceding display. It would be desirable if we were able to show that the convergence from (6.9) holds uniformly on $\mathscr{G}$, or at least on suitable subspaces $\tilde{\mathscr{G}} \subseteq \mathscr{G}$, such that

$$\sup_{g \in \tilde{\mathscr{G}}} P^{(n)}_{\gamma,g}\big\{\hat{\eta}_n - g(0) \in [-\delta, \delta]^C\big\} \to \sup_{g \in \tilde{\mathscr{G}}} \mathscr{N}_{(\mu_g,\Sigma)}\big([-\delta, \delta]^C\big). \tag{6.10}$$

The expression on the right hand side is bounded by the upper minimax risk bound from Theorem 5.14 (see the following Theorem 6.3). Therefore, it is also an upper bound for the asymptotic minimax risk in the sequence of local models, provided that we are able to prove (6.10). The concept of uniform convergence in distribution from Section 2.4 turns out to be an adequate tool to derive a supremum convergence statement of this type. (Actually, uniform convergence in distribution is of course an even stronger result than the simple convergence of the suprema.) We will show in the next section that the convergence result from Theorem 6.2 holds uniformly in g (and even uniformly in γ). Eventually, this implies an assertion of type (6.10).

6.3 Theorem. *Let $\gamma \in \Theta'$ be fixed and let $\delta = (\delta_1, \ldots, \delta_d)^\top$ be as in Section 3.3. Furthermore, let $v = (v_1, \ldots, v_d)^\top$ (which depends on γ) be defined as proposed in Section 3.5.3. Then*

$$\begin{aligned} &\sup_{g \in \mathscr{G}} \mathscr{N}_{(\mu_{\gamma,g}, \Sigma_\gamma)}\big([-\delta,\delta]^C\big) \\ &\leq\ 1 - \prod_{j=1}^{d} \Bigg\{ 1 - \Phi\left(- \frac{\|\mathcal{C}_\gamma^j\|_1 \beta_j^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}} \left[1 + \frac{(2\rho+1)v_j}{\|\mathcal{C}_\gamma^j\|_1 \beta_j} \right] \right) \\ &\qquad\qquad - \Phi\left(\frac{\|\mathcal{C}_\gamma^j\|_1 \beta_j^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}} \left[1 - \frac{(2\rho+1)v_j}{\|\mathcal{C}_\gamma^j\|_1 \beta_j} \right] \right) \Bigg\}, \end{aligned}$$

which is the upper bound given in Theorem 5.14.

Addendum. *In the special case $d = 1$ the above inequality always holds with "=".*

Proof. Clearly, for $g \in \mathscr{G}$ the function $h := g - g(0)\mathbb{1}_{[-K,K]}$ is also an element of $\mathscr{G}$. In a first step we prove that $\mu_g = \mu_h$. With the definitions from (6.8), we have

$$\begin{aligned} \mu_g - \mu_h &= \big(\mathcal{C}_\gamma^{-1} S_\gamma e_g - g(0)\big) - \big(\mathcal{C}_\gamma^{-1} S_\gamma e_h - \underbrace{h(0)}_{=0}\big) \\ &= \mathcal{C}_\gamma^{-1} S_\gamma \begin{pmatrix} \int g(0)^\top \mathcal{I}_\gamma G_\gamma^1(t)\, dt \\ \vdots \\ \int g(0)^\top \mathcal{I}_\gamma G_\gamma^d(t)\, dt \end{pmatrix} - g(0). \qquad (6.11) \end{aligned}$$

Since $\mathcal{I}_\gamma = \mathcal{C}_\gamma^\top \mathcal{C}_\gamma$ (by definition), the entries of the vector from the last expression can be written as

$$\begin{aligned}\int g(0)^\top \mathcal{I}_\gamma G_\gamma^i(t)\, dt &\overset{(6.4)}{=} \int g(0)^\top (\mathcal{C}_\gamma^\top \mathcal{C}_\gamma)\mathcal{C}_\gamma^{-1}(0,\dots,g_{\beta_i}(t),\dots 0)^\top\, dt \\ &= \int g_{\beta_i}(t)\, dt \quad \mathcal{C}_\gamma^i g(0).\end{aligned}$$

Thus, we get for the complete vector

$$\begin{pmatrix} \int g(0)^\top \mathcal{I}_\gamma G_\gamma^1(t)\, dt \\ \vdots \\ \int g(0)^\top \mathcal{I}_\gamma G_\gamma^d(t)\, dt \end{pmatrix} = \operatorname{diag}\big(\textstyle\int g_{\beta_1}(t)\, dt, \dots, \int g_{\beta_d}(t)\, dt\big)\mathcal{C}_\gamma g(0).$$

The entries of the above diagonal matrix can be computed to

$$\int g_{\beta_i}(t)\, dt = 2\frac{\rho}{\rho+1}\beta_i^{1+1/\rho} = \frac{1}{S_i(\gamma)}, \tag{6.12}$$

and therefore

$$\operatorname{diag}\big(\textstyle\int g_{\beta_1}(t)\, dt, \dots, \int g_{\beta_d}(t)\, dt\big) = S_\gamma^{-1}.$$

From the above calculations we conclude that

$$\mathcal{C}_\gamma^{-1} S_\gamma \begin{pmatrix} \int g(0)^\top \mathcal{I}_\gamma G_\gamma^1(t)\, dt \\ \vdots \\ \int g(0)^\top \mathcal{I}_\gamma G_\gamma^d(t)\, dt \end{pmatrix} = \mathcal{C}_\gamma^{-1} S_\gamma S_\gamma^{-1} \mathcal{C}_\gamma g(0) = g(0).$$

Hence, the expression in (6.11) is equal to zero, which implies $\mu_g = \mu_h$. For the assessment of the supremum it is therefore sufficient to consider only such $g \in \mathscr{G}$ with $g(0) = 0$.

Let $A_\gamma = \{\mathcal{C}_\gamma x : x \in [-\delta, \delta]\}$ and $v = v(\gamma)$ be as in Section 3.5.3, and thus in particular $A_\gamma^C \subseteq [-v, v]^C$. If we write $e_g = (e_{g,1}, \dots, e_{g,d})^\top$, then for each $g \in \mathscr{G}$ (with $g(0) = 0$)

$$\begin{aligned}\mathscr{N}_{(\mu_g,\Sigma)}\big([-\delta,\delta]^C\big) &\overset{(6.8)}{=} \mathscr{N}_{(Se_g, S\mathcal{J}S)}\big\{A_\gamma^C\big\} \\ &\leq \mathscr{N}_{(Se_g, S\mathcal{J}S)}\big([-v,v]^C\big) \\ &= 1 - \mathscr{N}_{(Se_g, S\mathcal{J}S)}([-v,v]) \\ &\overset{(*)}{=} 1 - \prod_{j=1}^d \mathscr{N}_{(S_j e_{g,j}, S_j^2 \|g_{\beta_j}\|_{\lambda,2}^2)}([-v_j, v_j]) \\ &= 1 - \prod_{j=1}^d \Big(1 - \mathscr{N}_{(S_j e_{g,j}, S_j^2 \|g_{\beta_j}\|_{\lambda,2}^2)}\big([-v_j, v_j]^C\big)\Big). \end{aligned} \tag{6.13}$$

Passing over to the one-dimensional normal distributions in $(*)$ is unproblematic, because the covariance matrix $S\mathcal{J}S$ has diagonal form, and thus the components of the multivariate normal distribution are independent. According to Anderson's lemma (see e.g. Pfanzagl, 1994, p. 87) the probability

$$\mathscr{N}_{(S_j e_{g,j}, S_j^2 \|g_{\beta_j}\|^2_{\lambda,2})}\big([-v_j, v_j]^C\big)$$

is maximal if

$$|e_{g,j}| = \left|\int g(t)^\top \mathcal{I}_\gamma G_\gamma^j(t)\, dt\right| = \left|\int \mathcal{C}_\gamma^j g(t) g_{\beta_j}(t)\, dt\right|$$

is maximal. Note that we assumed $g(0) = 0$, and therefore

$$\|g(t)\|_\infty = \|g(t) - g(0)\|_\infty \le |t|^\rho$$

(according to (3.10)). In combination with the triangle inequality the integrand can thus be bounded according to

$$\begin{aligned} |\mathcal{C}_\gamma^j g(t)| \ &\le\ \sum_{i=1}^d |\mathcal{C}_\gamma^{ji}|\ |g_i(t)| \\ &\le\ \sum_{i=1}^d |\mathcal{C}_\gamma^{ji}|\ |t|^\rho \quad = \quad \|\mathcal{C}_\gamma^j\|_1\ |t|^\rho, \end{aligned}$$

and we conclude that for every g

$$\begin{aligned} S_j|e_{g,j}| \ &\le\ S_j\|\mathcal{C}_\gamma^j\|_1 \int g_{\beta_j}(t)|t|^\rho\, dt \\ &=\ S_j\|\mathcal{C}_\gamma^j\|_1\ 2\beta_j^{2+1/\rho}\left(\frac{1}{\rho+1} - \frac{1}{2\rho+1}\right) \\ &\overset{(6.6)}{=}\ \frac{\beta_j\|\mathcal{C}_\gamma^j\|_1}{\rho}\left(1 - \frac{\rho+1}{2\rho+1}\right) \\ &=\ \frac{\beta_j\|\mathcal{C}_\gamma^j\|_1}{2\rho+1} \quad =: \quad e_j^*. \end{aligned}$$

Hence, with Anderson's lemma (†) each of the probabilities in (6.13) can be bounded from above in the following way:

$$\begin{aligned}
&\mathscr{N}_{(S_j e_{g,j}, S_j^2 \|g_{\beta_j}\|_{\lambda,2}^2)}\big([-v_j, v_j]^C\big) \\
&\quad \overset{(\dagger)}{\leq} \mathscr{N}_{(e_j^*, S_j^2 \|g_{\beta_j}\|_{\lambda,2}^2)}\big([-v_j, v_j]^C\big) \\
&\quad = \Phi\left(-\frac{v_j - e_j^*}{S_j \|g_{\beta_j}\|_{\lambda,2}}\right) + \Phi\left(-\frac{v_j + e_j^*}{S_j \|g_{\beta_j}\|_{\lambda,2}}\right) \\
&\quad = \Phi\left(-\frac{\|\mathcal{C}_\gamma^j\|_1 \beta_j^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}\left[1 + \frac{(2\rho+1)v_j}{\|\mathcal{C}_\gamma^j\|_1 \beta_j}\right]\right) \\
&\qquad\quad + \Phi\left(\frac{\|\mathcal{C}_\gamma^j\|_1 \beta_j^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}\left[1 - \frac{(2\rho+1)v_j}{\|\mathcal{C}_\gamma^j\|_1 \beta_j}\right]\right).
\end{aligned}$$

In the last equation we simply plugged in e_j^* and the expressions for $\|g_{\beta_j}\|_{\lambda,2}$ and S_j from (5.21) and (6.6), respectively. Together with (6.13) this implies

$$\begin{aligned}
&\mathscr{N}_{(\mu_g, \Sigma)}\big([-\delta, \delta]^C\big) \\
&\leq 1 - \prod_{j=1}^{d}\left\{1 - \Phi\left(-\frac{\|\mathcal{C}_\gamma^j\|_1 \beta_j^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}\left[1 + \frac{(2\rho+1)v_j}{\|\mathcal{C}_\gamma^j\|_1 \beta_j}\right]\right)\right. \\
&\qquad\qquad \left. - \Phi\left(\frac{\|\mathcal{C}_\gamma^j\|_1 \beta_j^{1+1/(2\rho)}}{[(\rho+1)(2\rho+1)]^{1/2}}\left[1 - \frac{(2\rho+1)v_j}{\|\mathcal{C}_\gamma^j\|_1 \beta_j}\right]\right)\right\}.
\end{aligned}$$

Taking the supremum with respect to g yields the assertion of the theorem. The addendum follows from the fact that for $d = 1$ we have $[-v, v] = A_\gamma$. ■

6.2 Uniform convergence of the local estimator

The aim of this section is to prove that the local estimator $\hat{\eta}_n$ converges uniformly in distribution to a family of normal distributions. This would allow us to derive an upper bound for the asymptotic maximal risk of $\hat{\eta}_n$, and thus for the minimax risk in the sequence of the local models (6.1). Throughout this section we use the assumptions and the notation from the preceding section. Note, however, that γ is now no longer fixed. The vector $v = v(\gamma)$, which will often be used in the following, is defined as in Section 3.5.3.

In this section we restrict the local parameter functions g to the space $\mathscr{G}_{\mathrm{H}}$ from (3.12). We consider the product space

$$\Lambda := \overline{\Theta^*} \times \mathscr{G}_{\mathrm{H}} = [\gamma_1^* - \varepsilon_1^*, \gamma_1^* + \varepsilon_1^*] \times \ldots \times [\gamma_d^* - \varepsilon_d^*, \gamma_d^* + \varepsilon_d^*] \times \mathscr{G}_{\mathrm{H}},$$

which we endow with the product metric

$$d_\Lambda((\gamma_1, g_1), (\gamma_2, g_2)) := \max\{\|\gamma_1 - \gamma_2\|, \|g_1 - g_2\|_{u,d}\}.$$

By passing over from $\mathscr{G}$ to $\mathscr{G}_{\mathrm{H}}$ we enforce that the resulting space Λ is compact. Hence, each sequence in Λ possesses a convergent subsequence. This property will turn out to be fundamental for our examinations.

We formulate the main result of this section:

6.4 Theorem. *Presume the assumptions from Sections 3.1–3.3 hold. Let $\hat{\eta}_n = \hat{\eta}_n(\gamma)$ be defined according to (6.7). Then*

$$\mathscr{L}(\hat{\eta}_n(\gamma) - g(0) \mid P^{(n)}_{\gamma,g}) \rightsquigarrow \mathscr{N}(\mu_{\gamma,g}, \Sigma_\gamma)$$

uniformly on $\Lambda = \overline{\Theta^*} \times \mathscr{G}_{\mathrm{H}}$ (in the sense of Definition 2.21).

To prove the theorem we first present several more or less technical auxiliary results. Subsequently, we prove the assertion.

6.5 Proposition. *$\mathscr{G}_{\mathrm{H}}$ is compact with respect to the metric induced by the uniform norm $\|\cdot\|_{u,d}$, and $\Lambda = \overline{\Theta^*} \times \mathscr{G}_{\mathrm{H}}$ is compact with respect to the product metric d_Λ.*

Proof. From the definition of the space $\mathscr{G}_{\mathrm{H}} = \mathscr{G}_{\mathrm{H}}(K, \rho, d)$ in (3.12) follows that this can be written as a product space

$$\mathscr{G}_{\mathrm{H}}(K, \rho, d) = \mathscr{G}_{\mathrm{H}}(K, \rho, 1) \times \ldots \times \mathscr{G}_{\mathrm{H}}(K, \rho, 1) \quad (d \text{ times}).$$

We show that the space $\mathscr{G}_{\mathrm{H},1} := \mathscr{G}_{\mathrm{H}}(K, \rho, 1)$ is compact. By virtue of the Arzelà-Ascoli theorem this is the case if we can prove that $\mathscr{G}_{\mathrm{H},1} \subseteq C([-K, K], \mathbb{R})$ is bounded point-wise and equi-continuous (cf. Heuser, 2000, Satz 106.2). Point-wise boundedness follows trivially from (3.11). In order to prove the equi-continuity let an arbitrary $\varepsilon > 0$ be given. We set $\iota := \varepsilon^{1/\rho}$. The Hölder condition from the definition of $\mathscr{G}_{\mathrm{H}}$ holds of course also for $\mathscr{G}_{\mathrm{H},1}$. Thus, for all $h \in \mathscr{G}_{\mathrm{H},1}$ and all $x, y \in [-K, K]$ with $|x - y| \leq \iota$ we have

$$|h(x) - h(y)| \overset{(3.12)}{\leq} |x - y|^\rho \leq \iota^\rho = \varepsilon.$$

Therefore, $\mathscr{G}_{\mathrm{H},1}$ is equi-continuous and thus compact. From Satz 12.9 in Lipschutz (1977) it follows then that $\mathscr{G}_{\mathrm{H}}$ as a product of the spaces $\mathscr{G}_{\mathrm{H},1}$ is compact with respect to the product norm $\|\cdot\|_{u,d}$ (i.e. compact with respect to the topology induced by the corresponding metric). With the same argument we conclude that Λ is compact. ■

6.6 Lemma. *Let $\beta_j(\gamma)$ be defined as the solution (5.44), and let $S_j(\gamma)$ be according to (6.6). The functions $g_{\beta_j}(\cdot)$ are defined according to (5.15). Then the following statements hold:*

(i) Both of the maps $\gamma \mapsto \beta(\gamma) = (\beta_1(\gamma), \dots, \beta_d(\gamma))^\top$ and $\gamma \mapsto (S_1(\gamma), \dots, S_d(\gamma))^\top$ are continuous on Θ.

(ii) If $\gamma_n \to \gamma$, then $\|g_{\beta_j(\gamma_n)} - g_{\beta_j(\gamma)}\|_u \to 0$.

Proof. *Ad (i):* Let an arbitrary point $\gamma_0 \in \Theta$ be given. We consider the function

$$F\colon (0,\infty)^3 \to \mathbb{R}, \quad F(x,y,z) = \begin{cases} \frac{z^{1+1/\rho}}{\log \frac{z+x/y}{z-x/y}} - \frac{\rho+1}{2xy}, & \text{if } z > x/y, \\ 0, & \text{otherwise.} \end{cases}$$

Of course, this F is closely related to the function f from Section 3.5.2. For the special point

$$(x^*, y^*, z^*) := (v_j(\gamma_0), \|\mathcal{C}^j_{\gamma_0}\|_1, \beta_j(\gamma_0))$$

we have $F(x^*, y^*, z^*) = 0$, which follows directly from the definition of $\beta_j(\gamma_0)$ as the solution to equation (5.44). Obviously, there is a neighbourhood U^* of (x^*, y^*, z^*) such that $F\colon U^* \to \mathbb{R}$ is continuously differentiable. Furthermore, the map $z \mapsto F(x^*, y^*, z)$ is strictly increasing (which follows with the same arguments as the monotonicity of the function f from Section 3.5.1). In particular, the partial derivative $\nabla_z F(x^*, y^*, z)$ is strictly positive, and thus invertible. Hence, the implicit function theorem (Heuser, 1990, p. 292 f.) guarantees the existence of neighbourhoods $U_{(x^*,y^*)}$ and U_{z^*} of (x^*, y^*) and z^*, respectively, as well as a unique continuous (even continuously differentiable) function $f\colon U_{(x^*,y^*)} \to U_{z^*}$ such that $f(x^*, y^*) = z^*$ and

$$F(x, y, f(x,y)) = 0, \quad (x,y) \in U_{(x^*,y^*)}. \tag{6.14}$$

From the arguments in Section 3.1 and Lemma 3.1 we know that the maps $\gamma \mapsto \|\mathcal{C}_\gamma^j\|_1$ and $\gamma \mapsto v_j(\gamma)$ are continuous on Θ. Thus, there is a neighbourhood U_{γ_0} of γ_0, such that $q(\gamma) := (v_j(\gamma), \|\mathcal{C}_\gamma^j\|_1)$ is a continuous map $q\colon U_{\gamma_0} \to U_{(x^*,y^*)}$. Clearly, the composition $f \circ q\colon U_{\gamma_0} \to U_{z^*}$ is continuous, too, as well as the map

$$U_{\gamma_0} \ni \gamma \mapsto F(v_j(\gamma), \|\mathcal{C}_\gamma^j\|_1, f \circ q(\gamma)).$$

From (6.14) and the fact that equation (5.44) has a unique solution we conclude that $\beta_j(\gamma) = f \circ q(\gamma)$ for all $\gamma \in U_{\gamma_0}$. Consequently, $\beta_j(\cdot)$ is continuous on U_{γ_0}. Since γ_0 was chosen arbitrarily, we conclude that $\gamma \mapsto \beta_j(\gamma)$ is continuous on Θ. The continuity of the maps $\gamma \mapsto S_j(\gamma)$ follows directly from the definition in (6.6).
Ad (ii): From (i) it follows that $\beta_j(\gamma_n) \to \beta_j(\gamma)$, as $\gamma_n \to \gamma$, and the definition of the function $g_\beta(\cdot)$ in (5.15) yields $\|g_{\beta_j(\gamma_n)} - g_{\beta_j(\gamma)}\|_u = |\beta_j(\gamma_n) - \beta_j(\gamma)| \to 0$. ■

6.7 Proposition. *Assume the conditions of Theorem 6.4 hold. Furthermore, let a fixed $g \in \mathscr{G}_{\mathrm{H}}$ be given and a convergent sequence $\gamma_n \in \overline{\Theta^*}$, $\gamma_n \to \gamma$. Then*

$$\mathscr{L}(\hat{\eta}_n(\gamma_n) - g(0) \mid P^{(n)}_{\gamma_n,g}) \rightsquigarrow \mathscr{N}(\mu_{\gamma,g}, \Sigma_\gamma).$$

Proof. It suffices to show that

$$\mathscr{L}\big((U_n^{\gamma_n,1}, \ldots, U_n^{\gamma_n,d})^\top \mid P^{(n)}_{\gamma_n,g}\big) \rightsquigarrow \mathscr{N}(e_{\gamma,g}, \mathcal{J}_\gamma). \tag{6.15}$$

With the definition of $\hat{\eta}_n$ in (6.7) and the convergence $\mathcal{C}_{\gamma_n}^{-1} S_{\gamma_n} \to \mathcal{C}_\gamma^{-1} S_\gamma$—as a consequence of the arguments in Section 3.1 and Lemma 6.6—the assertion then follows from an application of the delta method and Slutsky's lemma.

We prove (6.15) with the Lindeberg-Feller theorem (van der Vaart, 1998, Proposition 2.27). According to this, we have to show that the means and the variances of $(U_n^{\gamma_n,1}, \ldots, U_n^{\gamma_n,d})^\top$ converge to the mean and the variance of the normal distribution in (6.15), respectively, and that a Lindeberg condition is satisfied. If not explicitly stated otherwise, in the following calculations all means and variances are to be built with respect to the measure $P^{(n)}_{\gamma_n,g}$.
Convergence of the covariances: We use the notation from Section 6.1.3. By definition, for all $r, s \in \{1, \ldots, d\}$ and $i \neq j$, the

random variables $U_{nj}^{\gamma_n,r}$ and $U_{ni}^{\gamma_n,s}$ introduced in (6.5) are independent. Hence,

$$
\begin{aligned}
&\operatorname{Cov}(U_n^{\gamma_n,r}, U_n^{\gamma_n,s}) \\
&\quad = \sum_{j=1}^{n}\sum_{i=1}^{n} \mathrm{E}\left[\left(U_{nj}^{\gamma_n,r} - \mathrm{E}\,U_{nj}^{\gamma_n,r}\right)\left(U_{ni}^{\gamma_n,s} - \mathrm{E}\,U_{ni}^{\gamma_n,s}\right)\right] \\
&\quad = \sum_{j=1}^{n} \mathrm{E}\left[U_{nj}^{\gamma_n,r} U_{nj}^{\gamma_n,s}\right] - \mathrm{E}\,U_{nj}^{\gamma_n,r}\ \mathrm{E}\,U_{nj}^{\gamma_n,s} \\
&\quad = \sum_{j=1}^{n} a_n^2 \left(G_{\gamma_n}^r(\tilde{x}_{nj})^\top \int \dot{\ell}_{\gamma_n}\dot{\ell}_{\gamma_n}^\top \, dP_{\gamma_n + a_n g(\tilde{x}_{nj})} G_{\gamma_n}^s(\tilde{x}_{nj})\right) \\
&\qquad - \sum_{j=1}^{n} a_n^2 \prod_{q=r,s} \left(G_{\gamma_n}^q(\tilde{x}_{nj})^\top \int \dot{\ell}_{\gamma_n} \, dP_{\gamma_n + a_n g(\tilde{x}_{nj})}\right).
\end{aligned}
\tag{6.16}
$$

In order to examine the first sum in (6.16) we first simplify it, replacing the integrals by the information matrix $\mathcal{I}_\gamma$ and $G_{\gamma_n}^r$ and $G_{\gamma_n}^s$ by G_γ^r and G_γ^s, respectively. The resulting expression

$$
\sum_{j=1}^{n} a_n^2 G_\gamma^r(\tilde{x}_{nj})^\top \mathcal{I}_\gamma G_\gamma^s(\tilde{x}_{nj})
$$

is of the same type as that in (4.11). Hence, with the ideas of the proof of (2.7) in Proposition 4.3 it can be interpreted as a Riemann sum satisfying

$$
\begin{aligned}
\sum_{j=1}^{n} a_n^2 G_\gamma^r(\tilde{x}_{nj})^\top \mathcal{I}_\gamma G_\gamma^s(\tilde{x}_{nj}) \ &\to \ \int G_\gamma^r(w)^\top \mathcal{I}_\gamma G_\gamma^s(w) \, dw \\
&= \ \mathbb{1}_{\{r=s\}} \|g_{\beta_s}\|_{\lambda,2}^2,
\end{aligned}
\tag{6.17}
$$

where the last equality follows from the definition in (6.4). We can now use this observation to examine the asymptotic behaviour of the first sum in (6.16): from the arguments in Section 3.1 we know that the Cholesky matrix is continuous in the parameter, and together with (ii) in Lemma 6.6 we conclude that $\|G_{\gamma_n}^i - G_\gamma^i\|_{u,d} \to 0$ as $n \to \infty$. Furthermore, we postulated in Section 3.2 the maps $(\theta, \tau) \mapsto \int \dot{\ell}_{\theta,k}\dot{\ell}_{\theta,l} \, dP_\tau$ $(k, l = 1, \ldots, d,)$ to be continuous. As a consequence, we have

$$
\int \dot{\ell}_{\gamma_n}\dot{\ell}_{\gamma_n}^\top \, dP_{\gamma_n + a_n g(\tilde{x}_{nj})} \to \int \dot{\ell}_\gamma \dot{\ell}_\gamma^\top \, dP_\gamma = \mathcal{I}_\gamma.
$$

With these observations we conclude that

$$\lim_{n\to\infty}\sum_{j=1}^{n} a_n^2\left(G_{\gamma_n}^r(\tilde{x}_{nj})^\top \int \dot{\ell}_{\gamma_n}\dot{\ell}_{\gamma_n}^\top \, dP_{\gamma_n+a_n g(\tilde{x}_{nj})} G_{\gamma_n}^s(\tilde{x}_{nj})\right)$$

$$= \lim_{n\to\infty}\sum_{j=1}^{n} a_n^2 G_\gamma^r(\tilde{x}_{nj})^\top \mathcal{I}_\gamma G_\gamma^s(\tilde{x}_{nj}) \overset{(6.17)}{=} \mathbb{1}_{\{r=s\}}\|g_{\beta_s}\|_{\lambda,2}^2.$$

The second sum in (6.16) converges to zero. This results from the following calculation:

$$\begin{aligned}
&\left|\sum_{j=1}^{n} a_n^2 \prod_{q=r,s}\left(G_{\gamma_n}^q(\tilde{x}_{nj})^\top \int \dot{\ell}_{\gamma_n}\, dP_{\gamma_n+a_n g(\tilde{x}_{nj})}\right)\right| \\
&\quad = \quad \left|\sum_{j\in\mathbb{K}_n} a_n^2 \prod_{q=r,s}\left(G_{\gamma_n}^q(\tilde{x}_{nj})^\top \int \dot{\ell}_{\gamma_n}\, dP_{\gamma_n+a_n g(\tilde{x}_{nj})}\right)\right| \\
&\quad \leq \quad k_n a_n^2 \max_{q=r,s}\sup_{j\in\mathbb{K}_n} \|G_{\gamma_n}^q(\tilde{x}_{nj})\|^2 \left\|\int \dot{\ell}_{\gamma_n}\, dP_{\gamma_n+a_n g(\tilde{x}_{nj})}\right\|^2,
\end{aligned}$$

with $\mathbb{K}_n$ and k_n as in (4.7). Due to (4.8) we have $k_n a_n^2 = O(1)$. From the boundedness of the maps $G_{\gamma_n}^q(\cdot)$ $(q = r, s)$ and the continuity of the map $(\theta,\tau) \mapsto \int \dot{\ell}_\theta \, dP_\tau$ (cf. Section 3.2) we conclude that the supremum in the last expression—and thus the second sum in (6.16)—tends to zero as $n \to \infty$. Summarising, the above examinations yield $\mathrm{Cov}(U_n^{\gamma_n,r}, U_n^{\gamma_n,s}) \to \mathbb{1}_{\{r=s\}}\|g_{\beta_s}\|_{\lambda,2}^2$, and therefore

$$\mathrm{Cov}((U_n^{\gamma_n,1},\ldots,U_n^{\gamma_n,d})^\top) \to \mathcal{J}_\gamma.$$

Convergence of the means: According to the assumptions from Section 3.2 the map $(\theta,\tau) \mapsto \int \dot{\ell}_\theta \, dP_\tau$ is twice continuously partially differentiable. Hence, for fixed θ a Taylor expansion yields

$$\int \dot{\ell}_\theta \, dP_{\theta+h} = \int \dot{\ell}_\theta \, dP_\theta + h^\top \nabla_\tau\left(\int \dot{\ell}_\theta \, dP_\tau\right)_{|\tau=\theta} + r(\theta,h),$$

where $r(\theta,h)$ is short for a remainder depending on θ and h (see below). The first term on the right hand side of the above display is equal to zero, and for the second term we get

$$\begin{aligned}
\nabla_\tau\left(\int \dot{\ell}_\theta \, dP_\tau\right)_{|\tau=\theta} &\overset{(3.5)}{=} \int \dot{\ell}_\theta (\nabla p_\theta)^\top \frac{p_\theta}{p_\theta}\, d\mu \\
&\overset{(3.1)}{=} \int \dot{\ell}_\theta \dot{\ell}_\theta^\top \, dP_\theta \quad = \quad \mathcal{I}_\theta.
\end{aligned}$$

Hence, we see that

$$\int \dot{\ell}_\theta \, dP_{\theta+h} = \mathcal{I}_\theta h + r(\theta, h). \tag{6.18}$$

With the above Taylor expansion we can now write the expectation $\mathrm{E}\, U_n^{\gamma_n,k}$ as

$$\begin{aligned} \mathrm{E}\, U_n^{\gamma_n,k} &= \sum_{j=1}^n a_n G_{\gamma_n}^k(\tilde{x}_{nj})^\top \int \dot{\ell}_{\gamma_n} \, dP_{\gamma_n + a_n g(\tilde{x}_{nj})} \\ &\overset{(6.18)}{=} \sum_{j=1}^n a_n G_{\gamma_n}^k(\tilde{x}_{nj})^\top \mathcal{I}_{\gamma_n} a_n g(\tilde{x}_{nj}) \\ &\qquad + \sum_{j=1}^n a_n G_{\gamma_n}^k(\tilde{x}_{nj})^\top r(\gamma_n, a_n g(\tilde{x}_{nj})). \end{aligned} \tag{6.19}$$

With the same arguments that were used in the investigation of the first of the two sums in (6.16) we conclude that

$$\sum_{j=1}^n a_n G_{\gamma_n}^k(\tilde{x}_{nj})^\top \mathcal{I}_{\gamma_n} a_n g(\tilde{x}_{nj}) \;\to\; \int G_\gamma^k(w)^\top \mathcal{I}_\gamma g(w) \, dw.$$

The limit equals the kth component of the vector $e_{\gamma,g}$ from (6.8). If we can show that the second summand in (6.19) tends to zero, this implies $\mathrm{E}(U_n^{\gamma_n,1}, \ldots, U_n^{\gamma_n,d})^\top \to e_{\gamma,g}$. To this end, we will now take a closer look at the Taylor expansion from (6.18) and assess the remainder $r(\theta, h)$. The underlying function $(\theta, \tau) \mapsto \int \dot{\ell}_\theta \, dP_\tau$ was in Section 3.2 assumed twice continuously differentiable. In particular, this implies that the second order partial derivatives are continuous on Θ and therefore bounded on the compactum $\overline{\Theta^*} \subseteq \Theta$. Consequently, the remainder $r(\theta, h)$ is also uniformly bounded on $\overline{\Theta^*}$ in the sense that there is a constant C which does not depend on θ and which satisfies

$$\sup\{\|r(\theta, h)\| : \theta \in \overline{\Theta^*}\} \leq C\|h\|^2$$

(cf. Heuser, 1990, p. 285). From that and the Cauchy-Schwarz inequality it follows that

$$
\begin{aligned}
&\Big|\sum_{j=1}^{n} a_n G_{\gamma_n}^k(\tilde{x}_{nj})^\top r(\gamma_n, a_n g(\tilde{x}_{nj}))\Big| \\
&\qquad \leq \quad \sum_{j=1}^{n} a_n \|G_{\gamma_n}^k(\tilde{x}_{nj})\| \; \|r(\gamma_n, a_n g(\tilde{x}_{nj}))\| \\
&\qquad \leq \quad C \underbrace{\sum_{j=1}^{n} a_n^2 \|G_{\gamma_n}^k(\tilde{x}_{nj})\| \; \|g(\tilde{x}_{nj})\|^2}_{=O(1)} \quad a_n \to 0.
\end{aligned}
$$

Proof of the Lindeberg condition: Let $\iota > 0$ be given. For a simplification of notation we set

$$h_{nj} := \gamma_n + a_n g(\tilde{x}_{nj}).$$

We have to prove that the Lindeberg expression

$$
\begin{aligned}
L_n(\iota) = \sum_{j=1}^{n} \int & \|(U_{nj}^{\gamma_n,1}, \ldots, U_{nj}^{\gamma_n,d})^\top\|^2 \\
& \mathbb{1}\big\{\|(U_{nj}^{\gamma_n,1}, \ldots, U_{nj}^{\gamma_n,d})^\top\| > \iota\big\} \; dP_{h_{nj}}
\end{aligned}
$$

converges to zero. As is known, on $\mathbb{R}^d$ all norms are equivalent. In particular, the Euclidean and the maximum norm satisfy $\|z\| \leq \sqrt{d}\|z\|_\infty$ for all vectors z. Hence,

$$
\begin{aligned}
\mathbb{1}\big\{\|(U_{nj}^{\gamma_n,1}, \ldots, U_{nj}^{\gamma_n,d})^\top\| > \iota\big\} \;&\leq\; \mathbb{1}\Big\{\sqrt{d} \max_{k=1,\ldots,d} |U_{nj}^{\gamma_n,k}| > \iota\Big\} \\
&\leq\; \sum_{k=1}^{d} \mathbb{1}\big\{|U_{nj}^{\gamma_n,k}| > \iota/\sqrt{d}\big\}.
\end{aligned}
$$

Furthermore, a simple application of the Cauchy-Schwarz inequality yields

$$|U_{nj}^{\gamma_n,k}| \;=\; |a_n G_{\gamma_n}^k(\tilde{x}_{nj})^\top \dot{\ell}_{\gamma_n}| \;\leq\; a_n \|G_{\gamma_n}^k(\tilde{x}_{nj})\| \; \|\dot{\ell}_{\gamma_n}\|.$$

From the definition in (6.4) one can conclude that the expressions $\|G_{\gamma_n}^k(\tilde{x}_{nj})\|$ are uniformly bounded, more precisely, there is a constant $D > 0$ such that

$$\|G_{\gamma_n}^k(\tilde{x}_{nj})\| \leq D \; \mathbb{1}\{j \in \mathbb{K}_n\}.$$

This follows from the continuity of the inverse of the Cholesky matrix (cf. the remarks in Section 3.1), the fact that $g_{\beta(\gamma_n)} \to g_{\beta(\gamma)}$ (Lemma 6.6) and the assumption that the sequence γ_n lies in the compactum $\overline{\Theta^*}$. With the above arguments we can now bound the Lindeberg expression as follows

$$\begin{aligned} L_n(\iota) &\leq \sum_{l=1}^{d}\sum_{k=1}^{d}\sum_{j=1}^{n}\int |U_{nj}^{\gamma_n,k}|^2 \, \mathbb{1}\big\{|U_{nj}^{\gamma_n,l}| > \iota/\sqrt{d}\big\} \, dP_{h_{nj}} \\ &\leq \sum_{l=1}^{d}\sum_{k=1}^{d}\sum_{j=1}^{n}\int a_n^2 D^2 \, \mathbb{1}\{j \in \mathbb{K}_n\} \, \|\dot{\ell}_{\gamma_n}\|^2 \\ &\qquad\qquad \mathbb{1}\big\{Da_n\|\dot{\ell}_{\gamma_n}\| > \iota/\sqrt{d}\big\} \, dP_{h_{nj}} \\ &\lesssim a_n^2 k_n \left(\sup_{j\in\mathbb{K}_n}\int \|\dot{\ell}_{\gamma_n}\|^2 \, \mathbb{1}\big\{\|\dot{\ell}_{\gamma_n}\| > m_n\big\} \, dP_{h_{nj}}\right), \end{aligned} \tag{6.20}$$

with $m_n := \iota/(\sqrt{d}Da_n)$. According to (4.8) we have $a_n^2 k_n = O(1)$, and therefore it only remains to show that the supremum in (6.20) tends to zero. From Hölder's inequality we get

$$\begin{aligned} &\sup_{j\in\mathbb{K}_n}\int \|\dot{\ell}_{\gamma_n}\|^2 \, \mathbb{1}\big\{\|\dot{\ell}_{\gamma_n}\| > m_n\big\} \, dP_{h_{nj}} \\ &\leq \sup_{j\in\mathbb{K}_n}\left(\int \|\dot{\ell}_{\gamma_n}\|^4 \, dP_{h_{nj}}\right)^{1/2} \sup_{j\in\mathbb{K}_n}\left(\int \mathbb{1}\big\{\|\dot{\ell}_{\gamma_n}\| > m_n\big\} \, dP_{h_{nj}}\right)^{1/2} \\ &= \sup_{j\in\mathbb{K}_n}\left(\int \|\dot{\ell}_{\gamma_n}\|^4 \, dP_{h_{nj}}\right)^{1/2} \sup_{j\in\mathbb{K}_n}\left(P_{h_{nj}}\big\{\|\dot{\ell}_{\gamma_n}\| > m_n\big\}\right)^{1/2}. \end{aligned}$$

In Section 3.2 we assumed the maps $(\theta,\tau) \mapsto \int (\dot{\ell}_{\theta,i})^{k_1}(\dot{\ell}_{\theta,j})^{k_2} \, dP_\tau$ to be continuous on $\Theta\times\Theta$ (for $k_1, k_2 \in \{0,1,2\}$). This implies that the first of the two suprema in the preceding display is bounded. It remains to prove that the second supremum converges to zero: Markov's inequality yields

$$P_{h_{nj}}\big\{\|\dot{\ell}_{\gamma_n}\| > m_n\big\} \leq \frac{1}{m_n^2}\int \|\dot{\ell}_{\gamma_n}\|^2 \, dP_{h_{nj}},$$

and with the same smoothness arguments as before we conclude that

$$\sup_{j\in\mathbb{K}_n}\int \|\dot{\ell}_{\gamma_n}\|^2 \, dP_{h_{nj}} < \infty.$$

Since $m_n \to \infty$, this yields

$$\sup_{j\in\mathbb{K}_n} \Big(P_{h_{nj}}\{\|\dot{\ell}_{\gamma_n}\| > m_n\}\Big)^{1/2} \lesssim \frac{1}{m_n} \to 0.$$

These calculations show that the supremum in (6.20) converges to zero, and thus $L_n(\iota) \to 0$. ∎

In the next step for the proof of the uniform convergence in distribution of the estimator $\hat{\eta}_n$ we generalise the proposition insofar as we replace the single function g with a sequence $g_n \in \mathscr{G}_{\mathrm{H}}$.

6.8 Proposition. *Assume that the conditions of Theorem 6.4 hold. In addition, let a sequence $g_n \in \mathscr{G}_{\mathrm{H}}$ be given such that $g_n \to g$ (with respect to $\|\cdot\|_{u,d}$). Let $\gamma_n \in \overline{\Theta^*}$ be a convergent sequence, $\gamma_n \to \gamma$. Then*

$$\mathscr{L}(\hat{\eta}_n(\gamma_n) - g_n(0) \mid P^{(n)}_{\gamma_n,g_n}) \rightsquigarrow \mathscr{N}(\mu_{\gamma,g}, \Sigma_\gamma).$$

Proof. By assumption we have $\|g_n - g\|_{u,d} \to 0$ and thus in particular $g_n(0) \to g(0)$. Therefore, it is sufficient to show that

$$\mathscr{L}(\hat{\eta}_n(\gamma_n) \mid P^{(n)}_{\gamma_n,g_n}) \rightsquigarrow \mathscr{N}(\mu_{\gamma,g} + g(0), \Sigma_\gamma), \tag{6.21}$$

which, in combination with Slutsky's lemma, implies the assertion. For the proof of (6.21) we show that

$$\sup_{C\in\mathfrak{C}^d} \Big|P^{(n)}_{\gamma_n,g_n}\{\hat{\eta}_n(\gamma_n) \in C\} - \mathscr{N}_{(\mu_{\gamma,g}+g(0),\Sigma_\gamma)}(C)\Big| \to 0, \tag{6.22}$$

where $\mathfrak{C}^d$ denotes the class of all convex Borel sets in $\mathbb{R}^d$. Note that any non-degenerated multivariate normal distribution $\mathscr{N}_{(\nu,\Gamma)}$ is continuous with respect to Lebesgue measure, and therefore satisfies $\mathscr{N}_{(\nu,\Gamma)}(\partial C) = 0$ for all $C \in \mathfrak{C}^d$ (cf. Elstrodt, 1996, p. 68). Therefore, by virtue of Theorem 2.20, (6.22) is equivalent to (6.21). The triangular inequality yields

$$\begin{aligned}
&\sup_{C\in\mathfrak{C}^d} \Big|P^{(n)}_{\gamma_n,g_n}\{\hat{\eta}_n(\gamma_n) \in C\} - \mathscr{N}_{(\mu_{\gamma,g}+g(0),\Sigma_\gamma)}(C)\Big| \\
&\qquad\le \sup_{C\in\mathfrak{C}^d} \Big|P^{(n)}_{\gamma_n,g_n}\{\hat{\eta}_n(\gamma_n) \in C\} - P^{(n)}_{\gamma_n,g}\{\hat{\eta}_n(\gamma_n) \in C\}\Big| \\
&\qquad\quad + \sup_{C\in\mathfrak{C}^d} \Big|P^{(n)}_{\gamma_n,g}\{\hat{\eta}_n(\gamma_n) \in C\} - \mathscr{N}_{(\mu_{\gamma,g}+g(0),\Sigma_\gamma)}(C)\Big|.
\end{aligned} \tag{6.23}$$

In combination with Slutsky's lemma Proposition 6.7 yields

$$\mathscr{L}(\hat{\eta}_n(\gamma_n) \mid P^{(n)}_{\gamma_n,g}) \rightsquigarrow \mathscr{N}(\mu_{\gamma,g} + g(0), \Sigma_\gamma).$$

Therefore, we conclude with Theorem 2.20 that the second supremum on the right hand side of (6.23) tends to zero. It remains to show that the first of the two suprema tends to zero, too. From the definition of the variational distance (cf. Section 2.1) we conclude

$$\begin{aligned}
&\sup_{C\in\mathfrak{C}^d}\left|P^{(n)}_{\gamma_n,g_n}\{\hat{\eta}_n(\gamma_n)\in C\}-P^{(n)}_{\gamma_n,g}\{\hat{\eta}_n(\gamma_n)\in C\}\right|\\
&\qquad\le \|P^{(n)}_{\gamma_n,g_n}-P^{(n)}_{\gamma_n,g}\|_{\mathrm{V}}\\
&\qquad\le \sqrt{2}\,H(P^{(n)}_{\gamma_n,g_n},P^{(n)}_{\gamma_n,g})\\
&\qquad\le \sqrt{2}\Bigg(\sum_{j=1}^{n}H^2(P_{\gamma_n+a_ng_n(\tilde{x}_{nj})},P_{\gamma_n+a_ng(\tilde{x}_{nj})})\Bigg)^{1/2}. \qquad (6.24)
\end{aligned}$$

In the last step we used an inequality for the Hellinger distance of product measures, given in Reiss (1993), p. 23 f. Note that for $j\notin\mathbb{K}_n$ we have

$$\gamma_n+a_ng_n(\tilde{x}_{nj})=\gamma_n+a_ng(\tilde{x}_{nj})=\gamma_n,$$

and thus

$$H(P_{\gamma_n+a_ng_n(\tilde{x}_{nj})},P_{\gamma_n+a_ng(\tilde{x}_{nj})})=0.$$

Furthermore, because of (3.11), we have $\sup_{j=1,\dots,n}\|a_ng_n(\tilde{x}_{nj})\|\to 0$ and, by assumption, $\gamma_n\to\gamma$. Hence, for large n all points $\gamma_n+a_ng_n(\tilde{x}_{nj})$ and $\gamma_n+a_ng(\tilde{x}_{nj})$ are in a sufficiently small neighbourhood of γ, on which assertion (ii) of Lemma 2.7 holds $(*)$. Combining these observations we conclude that the sum in (6.24) can be bounded in the following way:

$$\begin{aligned}
&\sum_{j=1}^{n}H^2(P_{\gamma_n+a_ng_n(\tilde{x}_{nj})},P_{\gamma_n+a_ng(\tilde{x}_{nj})})\\
&\qquad= \sum_{j\in\mathbb{K}_n}H^2(P_{\gamma_n+a_ng_n(\tilde{x}_{nj})},P_{\gamma_n+a_ng(\tilde{x}_{nj})})\\
&\qquad\overset{(*)}{\lesssim} \sum_{j\in\mathbb{K}_n}\left\|(\gamma_n+a_ng_n(\tilde{x}_{nj}))-(\gamma_n+a_ng(\tilde{x}_{nj}))\right\|^2\\
&\qquad= \sum_{j\in\mathbb{K}_n}\|a_ng_n(\tilde{x}_{nj})-a_ng(\tilde{x}_{nj})\|^2\\
&\qquad\le k_na_n^2\|g_n-g\|_{u,d}.
\end{aligned}$$

Because of $\|g_n - g\|_{u,d} \to 0$ and $k_n a_n^2 = O(1)$ (according to (4.8)), this last expression tends to zero, and in combination with (6.24) we conclude that

$$\sup_{C \in \mathfrak{C}^d} \left| P^{(n)}_{\gamma_n, g_n}\{\hat{\eta}_n(\gamma_n) \in C\} - P^{(n)}_{\gamma_n, g}\{\hat{\eta}_n(\gamma_n) \in C\} \right| \to 0.$$

To sum up, we have shown that both of the suprema on the right hand side of (6.23) converge to zero. Hence, (6.22) holds, and according to the above discussion this implies the assertion. ■

Proof of Theorem 6.4. According to Proposition 6.5 $\Lambda = \overline{\Theta^*} \times \mathscr{G}_{\mathrm{H}}$ is compact. Moreover, for every convergent sequence $(\gamma_n, g_n) \to (\gamma, g)$ in Λ we have $\mu_{\gamma_n, g_n} \to \mu_{\gamma, g}$ and $\Sigma_{\gamma_n} \to \Sigma_\gamma$, which is a simple consequence of the definitions in (6.8), Lemma 6.6 and the continuity of the information and Cholesky matrices (see Section 3.1). Consequently, any such sequence satisfies

$$\mathscr{N}(\mu_{\gamma_n, g_n}, \Sigma_{\gamma_n}) \rightsquigarrow \mathscr{N}(\mu_{\gamma, g}, \Sigma_\gamma). \tag{6.25}$$

By virtue of Theorem 2.22 it therefore suffices to prove

$$\sup_{\gamma \in \overline{\Theta^*}} \sup_{g \in \mathscr{G}_{\mathrm{H}}} \sup_{C \in \mathfrak{C}^d} \left| P^{(n)}_{\gamma, g}\{\hat{\eta}_n(\gamma) - g(0) \in C\} - \mathscr{N}_{(\mu_{\gamma,g}, \Sigma_\gamma)}(C) \right| \to 0, \tag{6.26}$$

where, again, $\mathfrak{C}^d = \{C \in \mathbb{B}^d : C \text{ convex}\}$. Let us assume that (6.26) does *not* hold. Then there is an $\varepsilon > 0$ such that

$$\limsup_{n\to\infty} \sup_{\gamma \in \overline{\Theta^*}} \sup_{g \in \mathscr{G}_{\mathrm{H}}} \sup_{C \in \mathfrak{C}^d} \left| P^{(n)}_{\gamma, g}\{\hat{\eta}_n(\gamma) - g(0) \in C\} - \mathscr{N}_{(\mu_{\gamma,g}, \Sigma_\gamma)}(C) \right| > \varepsilon.$$

In particular, there is a sequence $(\gamma_n, g_n) \in \Lambda$ such that

$$\limsup_{n\to\infty} \sup_{C \in \mathfrak{C}^d} \left| P^{(n)}_{\gamma_n, g_n}\{\hat{\eta}_n(\gamma_n) - g_n(0) \in C\} - \mathscr{N}_{(\mu_{\gamma_n,g_n}, \Sigma_{\gamma_n})}(C) \right| > \varepsilon/2. \tag{6.27}$$

Without loss of generality, we can assume that this limit superior is actually a limit. If necessary, one can do so by passing over to some suitable subsequence. The compactness of $\Lambda = \overline{\Theta^*} \times \mathscr{G}_{\mathrm{H}}$ implies that there is a subsequence n', such that $(\gamma_{n'}, g_{n'}) \to (\gamma, g) \in \Lambda$. (In order not to unnecessarily complicate the notation, we further write n for n'.) The Triangle Inequality yields

$$\begin{aligned}\sup_{C\in\mathfrak{C}^d} &\left|P^{(n)}_{\gamma_n,g_n}\{\hat\eta_n(\gamma_n)-g_n(0)\in C\}-\mathscr{N}_{(\mu_{\gamma_n,g_n},\Sigma_{\gamma_n})}(C)\right| \\ &\le \sup_{C\in\mathfrak{C}^d}\left|P^{(n)}_{\gamma_n,g_n}\{\hat\eta_n(\gamma_n)-g_n(0)\in C\}-\mathscr{N}_{(\mu_{\gamma,g},\Sigma_{\gamma})}(C)\right| \\ &\qquad + \sup_{C\in\mathfrak{C}^d}\left|\mathscr{N}_{(\mu_{\gamma,g},\Sigma_{\gamma})}(C)-\mathscr{N}_{(\mu_{\gamma_n,g_n},\Sigma_{\gamma_n})}(C)\right|. \qquad (6.28)\end{aligned}$$

Combining Theorem 2.20 with the results from Proposition 6.8 and (6.25), respectively, we conclude that both of the suprema in (6.28) tend to zero, and thus

$$\sup_{C\in\mathfrak{C}^d}\left|P^{(n)}_{\gamma_n,g_n}\{\hat\eta_n(\gamma_n)-g_n(0)\in C\}-\mathscr{N}_{(\mu_{\gamma_n,g_n},\Sigma_{\gamma})}(C)\right|\to 0.$$

This is a contradiction to (6.27), and therefore the hypothesis that (6.26) does not hold must be false. This implies the assertion of the theorem. ■

As a direct consequence of the above results we obtain the desired upper bound for the asymptotic minimax risk in the sequence of the local models.

6.9 Corollary. *Assume that the conditions from Theorem 6.4 hold. Then*

$$\begin{aligned}\sup_{\gamma\in\overline{\Theta^*}}\sup_{g\in\mathscr{G}_{\mathrm{H}}} &P^{(n)}_{\gamma,g}\big\{\hat\eta_n(\gamma)-g(0)\in[-\delta,\delta]^C\big\} \\ &\longrightarrow \sup_{\gamma\in\overline{\Theta^*}}\sup_{g\in\mathscr{G}_{\mathrm{H}}}\mathscr{N}_{(\mu_{\gamma,g},\Sigma_\gamma)}\big([-\delta,\delta]^C\big),\end{aligned}$$

and the limit is bounded by the supremum (with respect to γ) over the expressions from Theorem 6.3.

Proof. With the same arguments that were used at the end of the proof of Theorem 4.11 we first conclude that

$$\begin{aligned}\liminf_{n\to\infty}\sup_{\gamma\in\overline{\Theta^*}}\sup_{g\in\mathscr{G}_{\mathrm{H}}} &P^{(n)}_{\gamma,g}\big\{\hat\eta_n(\gamma)-g(0)\in[-\delta,\delta]^C\big\} \\ &\ge \sup_{\gamma\in\overline{\Theta^*}}\sup_{g\in\mathscr{G}_{\mathrm{H}}}\liminf_{n\to\infty}P^{(n)}_{\gamma,g}\big\{\hat\eta_n(\gamma)-g(0)\in[-\delta,\delta]^C\big\} \\ &\overset{(6.9)}{=} \sup_{\gamma\in\overline{\Theta^*}}\sup_{g\in\mathscr{G}_{\mathrm{H}}}\mathscr{N}_{(\mu_{\gamma,g},\Sigma_\gamma)}\big([-\delta,\delta]^C\big). \qquad (6.29)\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
&\sup_{\gamma\in\overline{\Theta^*}} \sup_{g\in\mathscr{G}_{\mathrm{H}}} P^{(n)}_{\gamma,g}\big\{\hat{\eta}_n(\gamma) - g(0) \in [-\delta,\delta]^C\big\} \\
&\quad = \sup_{\gamma\in\overline{\Theta^*}} \sup_{g\in\mathscr{G}_{\mathrm{H}}} \Big(1 - P^{(n)}_{\gamma,g}\{\hat{\eta}_n(\gamma) - g(0) \in [-\delta,\delta]\}\Big) \\
&\quad \leq \sup_{\gamma\in\overline{\Theta^*}} \sup_{g\in\mathscr{G}_{\mathrm{H}}} \Big| P^{(n)}_{\gamma,g}\{\hat{\eta}_n(\gamma) - g(0) \in [-\delta,\delta]\} - \mathscr{N}_{(\mu_{\gamma,g},\Sigma_\gamma)}([-\delta,\delta]) \Big| \\
&\qquad + \sup_{\gamma\in\overline{\Theta^*}} \sup_{g\in\mathscr{G}_{\mathrm{H}}} \mathscr{N}_{(\mu_{\gamma,g},\Sigma_\gamma)}\big([-\delta,\delta]^C\big).
\end{aligned}$$

Since $[-\delta,\delta]$ is measurable and convex (i.e. $[-\delta,\delta] \in \mathfrak{C}^d$), it follows from (6.26) that

$$\sup_{\gamma\in\overline{\Theta^*}} \sup_{g\in\mathscr{G}_{\mathrm{H}}} \Big| P^{(n)}_{\gamma,g}\{\hat{\eta}_n(\gamma) - g(0) \in [-\delta,\delta]\} - \mathscr{N}_{(\mu_{\gamma,g},\Sigma_\gamma)}([-\delta,\delta]) \Big| \to 0,$$

and thus

$$\begin{aligned}
\limsup_{n\to\infty} \sup_{\gamma\in\overline{\Theta^*}} \sup_{g\in\mathscr{G}_{\mathrm{H}}} & P^{(n)}_{\gamma,g}\big\{\hat{\eta}_n(\gamma) - g(0) \in [-\delta,\delta]^C\big\} \\
&\leq \sup_{\gamma\in\overline{\Theta^*}} \sup_{g\in\mathscr{G}_{\mathrm{H}}} \mathscr{N}_{(\mu_{\gamma,g},\Sigma_\gamma)}\big([-\delta,\delta]^C\big).
\end{aligned}$$

Together with (6.29) this proves the assertion. ■

6.10 Remark. Passing over from Λ to subsets of the form $\Lambda' = \{\gamma\} \times \mathscr{G}_{\mathrm{H}}$, the above results yield the convergence assertion (6.10) with $\tilde{\mathscr{G}} = \mathscr{G}_{\mathrm{H}}$, which has been the starting point and motivation for the examinations of this section.

CHAPTER 7

Upper minimax risk bounds in the global model

In the preceding chapter we proved that the local estimator $\hat{\eta}_n$ converges uniformly in distribution to a family of normal distributions. This allowed us to derive an upper bound for the asymptotic minimax risk in the local model. We now want to expand these results to the global model. To this end, we use an adaptive estimator, i.e. we combine a preliminary estimator (that converges with appropriate rate) with the local estimator $\hat{\eta}_n$. This yields the desired upper minimax risk bounds. The ideas of this chapter are based on those in Pfanzagl (1994), cf. in particular the proof of Proposition 7.4.13 of that book.

7.1 An upper minimax risk bound in the global model

Throughout this chapter we use the notation introduced in Chapter 6. Furthermore, we assume that the conditions from Sections 3.1–3.3 hold. In addition, we assume that there is a *uniformly* $(1/a_n)$-*consistent* preliminary estimator according to Section 3.4. That is, we have an estimator $\xi_n^* = \xi_n^*(x_0)$ for $\xi(x_0)$ such that for every $\iota > 0$ there is a constant $C = C_\iota > 0$ satisfying

$$\limsup_{n\to\infty} \sup_{\xi\in\mathscr{F}} P_\xi^{(n)} \left\{ \frac{1}{a_n} \|\xi_n^* - \xi(x_0)\| \geq C \right\} < \iota. \tag{7.1}$$

We consider the adaptive estimator

$$\hat{\xi}_n := \xi_n^* + a_n \hat{\eta}_n(\xi_n^*). \tag{7.2}$$

Here the local estimator $\hat{\eta}_n(\xi_n^*)$ is defined according to (6.7), i.e. we have

$$\hat{\eta}_n(\xi_n^*) = \mathcal{C}_{\xi_n^*}^{-1} S_{\xi_n^*}(U_n^{\xi_n^*,1}, \ldots, U_n^{\xi_n^*,d})^\top,$$

with S according to (6.6) and

$$U_n^{\xi_n^*,i} = \sum_{j=1}^{n} U_{nj}^{\xi_n^*,i}, \quad U_{nj}^{\xi_n^*,i} = a_n G_{\xi_n^*}^i(\tilde{x}_{nj})^\top \dot{\ell}_{\xi_n^*}$$

Confer also (6.5) and (6.4) for the definition of the $U_{nj}^{\xi_n^*,i}$ and $G_{\xi_n^*}^i$, respectively.

The construction of this estimator reflects the idea of the localisation of the model. In a first step one estimates roughly the unknown parameter $\xi(x_0)$ by the preliminary estimator ξ_n^*. This corresponds to the localisation idea of the previous chapters of passing over from the *global* to the *local* model, and thus to a fixed centre of localisation. Subsequently, one tries to refine the estimator ξ_n^*, using the results achieved for the estimation of the parameter $g(0)$ in the local models. We will show that $\hat{\xi}_n$ in some sense—namely, as concerns the maximal risk—possesses the same asymptotic properties as $\hat{\eta}_n$ does in the sequence of local models. This is exactly the assertion of the following theorem, which at the same time is one of the main results of this thesis.

7.1 Theorem. *Let the assumptions from Sections 3.1–3.3 hold, i.e. assume that*[1)]

- *the family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is continuously differentiable in quadratic mean,*
- *the information matrices $\mathcal{I}_\theta$ are positive definite for all $\theta \in \Theta$,*
- *the log-densities $\log p_\theta$ are three times continuously partially differentiable with respect to the parameter, and the third partial derivatives of the log-densities are dominated by an integrable function J,*
- *the moments of the score function satisfy certain smoothness conditions.*

[1)] We recap only the main assumptions from Sections 3.1–3.3, for the details see Chapter 3.

Let the function space $\mathscr{F}_{\mathrm{H}}$ be defined according to (3.9). Moreover, let $\xi_n^* = \xi_n^*(x_0)$ be a preliminary $(1/a_n)$-consistent estimator that satisfies the above conditions, and let the resulting estimator $\hat{\xi}_n = \hat{\xi}_n(x_0)$ be defined according to (7.2). Then

$$\sup_{\xi \in \mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\{\hat{\xi}_n(x_0) - \xi(x_0) \in [-a_n\delta, a_n\delta]^C\}$$
$$\to \sup_{\gamma \in \Theta^*} \sup_{g \in \mathscr{G}_{\mathrm{H}}} \mathscr{N}_{(\mu_{\gamma,g},\Sigma_\gamma)}\big([-\delta,\delta]^C\big).$$

In particular, the limit expression is an upper bound for the asymptotic minimax risk in the global model $\{P_\xi^{(n)} : \xi \in \mathscr{F}_{\mathrm{H}}\}$, given by (1.2).

Proof. Throughout the following calculations let $\xi_0 := \xi(x_0)$. The definition of the adaptive estimator yields

$$\begin{aligned} \hat{\xi}_n - \xi_0 &= (\xi_n^* - \xi_0) + a_n\hat{\eta}_n(\xi_n^*) \\ &= (\xi_n^* - \xi_0) + a_n\left(\hat{\eta}_n(\xi_n^*) - \hat{\eta}_n(\xi_0)\right) + a_n\hat{\eta}_n(\xi_0). \end{aligned} \tag{7.3}$$

Using the definition of the local estimator $\hat{\eta}_n$ in (6.7) we further get[2)]

$$\begin{aligned} \hat{\eta}_n(\xi_0) &= \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} (U_n^{\xi_0,1}, \ldots, U_n^{\xi_0,d})^\top \\ &\overset{(6.5)}{=} \sum_{j=1}^n a_n \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} \begin{pmatrix} G_{\xi_0}^1(\tilde{x}_{nj})^\top \\ \vdots \\ G_{\xi_0}^d(\tilde{x}_{nj})^\top \end{pmatrix} \dot{\ell}_{\xi_0}(Y_{nj}) \\ &\overset{(6.4)}{=} \sum_{j=1}^n a_n \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \dot{\ell}_{\xi_0}(Y_{nj}), \end{aligned}$$

where here and in the following

$$B_{\xi_0}^{nj} := \operatorname{diag}\left(g_{\beta_1(\xi_0)}(\tilde{x}_{nj}), \ldots, g_{\beta_d(\xi_0)}(\tilde{x}_{nj})\right). \tag{7.4}$$

The parameters $\beta_j(\xi_0)$ are the solutions of (5.44) (with γ replaced by ξ_0). The above calculations do of course also hold, if ξ_0 is re-

[2)] As in Chapter 6, we interpret Y_{nj} as the jth canonical projection on the product space $\mathcal{Y}^n$; see also Remark 6.1.

placed with ξ_n^*. Hence, we can write

$$\begin{aligned}\hat{\eta}_n(\xi_n^*) &- \hat{\eta}_n(\xi_0)\\ &= \sum_{j=1}^n a_n \left(\mathcal{C}_{\xi_n^*}^{-1} S_{\xi_n^*} B_{\xi_n^*}^{nj} (\mathcal{C}_{\xi_n^*}^{-1})^\top\right) \left\{\dot{\ell}_{\xi_n^*}(Y_{nj}) - \dot{\ell}_{\xi_0}(Y_{nj})\right\} \\ &\qquad + R_1(\xi_0, \xi_n^*),\end{aligned} \tag{7.5}$$

where

$$\begin{aligned}R_1(\xi_0, \xi_n^*) := \sum_{j=1}^n a_n \Big(&\mathcal{C}_{\xi_n^*}^{-1} S_{\xi_n^*} B_{\xi_n^*}^{nj} (\mathcal{C}_{\xi_n^*}^{-1})^\top \\ &- \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top\Big) \dot{\ell}_{\xi_0}(Y_{nj}).\end{aligned} \tag{7.6}$$

In Section 3.2 we required the log-densities to be three times partially differentiable with respect to the parameter and thus in particular to permit a certain Taylor expansion. As a consequence, we may write

$$\dot{\ell}_{\xi_n^*}(Y_{nj}) - \dot{\ell}_{\xi_0}(Y_{nj}) \overset{(3.2)}{=} \ddot{\ell}_{\xi_0}(Y_{nj})(\xi_n^* - \xi_0) + r(\xi_0, \xi_n^*, Y_{nj}),$$

with $\ddot{\ell}_{\xi_0} = \nabla_{\xi_0}^2 \log p_{\xi_0}$ and some remainder term $r(\cdot)$ (which is to be discussed below), and we conclude that (7.5) can be written as

$$\begin{aligned}&\hat{\eta}_n(\xi_n^*) - \hat{\eta}_n(\xi_0)\\ &= \sum_{j=1}^n a_n \left(\mathcal{C}_{\xi_n^*}^{-1} S_{\xi_n^*} B_{\xi_n^*}^{nj} (\mathcal{C}_{\xi_n^*}^{-1})^\top\right) \left\{\dot{\ell}_{\xi_n^*}(Y_{nj}) - \dot{\ell}_{\xi_0}(Y_{nj})\right\} + R_1(\xi_0, \xi_n^*)\\ &= \sum_{j=1}^n a_n \left(\mathcal{C}_{\xi_n^*}^{-1} S_{\xi_n^*} B_{\xi_n^*}^{nj} (\mathcal{C}_{\xi_n^*}^{-1})^\top\right) \ddot{\ell}_{\xi_0}(Y_{nj})(\xi_n^* - \xi_0)\\ &\qquad + R_1(\xi_0, \xi_n^*) + R_2(\xi_0, \xi_n^*),\end{aligned} \tag{7.7}$$

where

$$R_2(\xi_0, \xi_n^*) := \sum_{j=1}^n a_n \left(\mathcal{C}_{\xi_n^*}^{-1} S_{\xi_n^*} B_{\xi_n^*}^{nj} (\mathcal{C}_{\xi_n^*}^{-1})^\top\right) r(\xi_0, \xi_n^*, Y_{nj}). \tag{7.8}$$

Summarising, the above calculations yield

$$
\begin{aligned}
&\frac{1}{a_n}(\hat{\xi}_n - \xi_0) - \hat{\eta}_n(\xi_0) \\
&\quad\overset{(7.3)}{=} \frac{1}{a_n}(\xi_n^* - \xi_0) + (\hat{\eta}_n(\xi_n^*) - \hat{\eta}_n(\xi_0)) \\
&\quad\overset{(7.7)}{=} \left(\frac{1}{a_n}\mathbb{1} + \sum_{j=1}^n a_n \left(\mathcal{C}_{\xi_n^*}^{-1} S_{\xi_n^*} B_{\xi_n^*}^{nj} (\mathcal{C}_{\xi_n^*}^{-1})^\top\right) \ddot{\ell}_{\xi_0}(Y_{nj})\right)(\xi_n^* - \xi_0) \\
&\qquad\quad + \sum_{s=1}^2 R_s(\xi_0, \xi_n^*).
\end{aligned}
$$

Setting

$$
\begin{aligned}
&R_3(\xi_0, \xi_n^*) \\
&:= \left(\mathbb{1} + \sum_{j=1}^n a_n^2 \left(\mathcal{C}_{\xi_n^*}^{-1} S_{\xi_n^*} B_{\xi_n^*}^{nj} (\mathcal{C}_{\xi_n^*}^{-1})^\top\right) \ddot{\ell}_{\xi_0}(Y_{nj})\right)\frac{1}{a_n}(\xi_n^* - \xi_0),
\end{aligned}
\tag{7.9}
$$

and $R(\xi_0, \xi_n^*) := \sum_{s=1}^3 R_s(\xi_0, \xi_n^*)$ the above display can be abbreviated to

$$
\frac{1}{a_n}(\hat{\xi}_n - \xi_0) = \hat{\eta}_n(\xi_0) + R(\xi_0, \xi_n^*). \tag{7.10}
$$

We continue with an examination of the asymptotic behaviour of $\hat{\eta}_n$ under the product measures $P_\xi^{(n)} = \bigotimes_{j=1}^n P_{\xi(x_{nj})}$. From the definition $U_{nj}^{\xi_0,i} = a_n G_{\xi_0}^i(\tilde{x}_{nj})^\top \dot{\ell}_{\xi_0}$, according to (6.5), it follows that $U_{nj}^{\xi_0,i} \equiv 0$ for $j \notin \mathbb{K}_n$. Hence, we conclude for the estimator $\hat{\eta}_n$ from (6.7) that of the total of n values $\xi(x_{nj})$—and thus from the corresponding distributions $P_{\xi(x_{nj})}$—only those have an influence on the estimator's behaviour, for which $|x_{nj} - x_0| \le K/b_n$. In other words: the stochastic behaviour of $\hat{\eta}_n$ only depends on the k_n distributions $P_{\xi(x_{nj})}$ with $j \in \mathbb{K}_n$ (cf. (4.7) for the definition of k_n and $\mathbb{K}_n$).

Let for a moment n be fixed, and consider now some arbitrary, but fixed $\xi \in \mathscr{F}_{\mathrm{H}}$. Without loss of generality we may assume that n is sufficiently large, such that $x_0 \pm K/b_n \in [0,1]$. If we define a map $g = g(\xi, n)$ according to

$$
g(x) := \begin{cases} \left(\xi(x_0 + x/b_n) - \xi(x_0)\right)/a_n, & x \in [-K, K], \\ 0, & x \notin [-K, K]. \end{cases}
$$

then the local alternative $\gamma_{g,n}$ around the point $\gamma := \xi(x_0)$ satisfies

$$\begin{aligned}\gamma_{g,n}(x_{nj}) &\overset{(1.3)}{=} \gamma + a_n g(\tilde{x}_{nj}) \\ &= \gamma + \xi\Big(x_0 + \frac{b_n(x_{nj} - x_0)}{b_n}\Big) - \gamma \;=\; \xi(x_{nj}),\end{aligned}$$

for all $j \in \mathbb{K}_n$, and

$$\gamma_{g,n}(x_0) \;=\; \gamma \;=\; \xi(x_0).$$

Since, by assumption, ξ is continuous on $[0,1]$, g is continuous on $[-K,K]$, and $g(0) = 0$. Furthermore, we conclude from the properties of ξ that

$$\begin{aligned}\|g(x) - g(y)\|_\infty &= \left\| \frac{\xi(x_0 + x/b_n)}{a_n} - \frac{\xi(x_0 + y/b_n)}{a_n} \right\|_\infty \\ &\overset{(3.9)}{\leq} \frac{1}{a_n} A(|x-y|/b_n)^\rho \overset{(3.14)}{=} |x-y|^\rho,\end{aligned}$$

and therefore $g \in \mathscr{G}_{\mathrm{H}}^*$ (i.e. $g \in \mathscr{G}_{\mathrm{H}}$ and $g(0) = 0$, see (3.13)). Consequently, for each $\xi \in \mathscr{F}_{\mathrm{H}}$ and each index n there is a map $g \in \mathscr{G}_{\mathrm{H}}^*$ such that the local alternative $\gamma_{g,n}$ (around the point $\gamma = \xi(x_0)$) and ξ have the same values at the points x_0 and x_{nj}, $j \in \mathbb{K}_n$. The converse of this assertion also holds true. If $\gamma \in \Theta^*$ and $g \in \mathscr{G}_{\mathrm{H}}^*$ are given, then the resulting local alternative $\gamma_{g,n}$ coincides on $[x_0 - K/b_n, x_0 + K/b_n]$—and only this interval is of importance for our examinations—with a function $\xi \in \mathscr{F}_{\mathrm{H}}$ with $\xi(x_0) = \gamma$. Indeed, we have $\gamma_{g,n}(x_0) = \gamma + a_n g(0) = \gamma$ and

$$\begin{aligned}\|\gamma_{g,n}(x) - \gamma_{g,n}(y)\|_\infty &= a_n \|g(b_n(x-x_0)) - g(b_n(y-x_0))\|_\infty \\ &\overset{(3.12)}{\leq} a_n (b_n|x-y|)^\rho \overset{(3.14)}{=} u(|x-y|),\end{aligned}$$

and $\gamma_{g,n}$ is also continuous on $[x_0 - K/b_n, x_0 + K/b_n]$. These notions yield the following equality of families of distributions,

$$\Big\{ \bigotimes_{j\in\mathbb{K}_n} P_{\xi(x_{nj})} : \xi \in \mathscr{F}_{\mathrm{H}} \Big\} = \Big\{ \bigotimes_{j\in\mathbb{K}_n} P_{\gamma_{g,n}(x_{nj})} : \gamma \in \Theta^*,\, g \in \mathscr{G}_{\mathrm{H}}^* \Big\}.$$

As was mentioned before, the distribution of the estimator $\hat{\eta}_n$ depends on the product measure $P_\xi^{(n)} = \bigotimes_{j=1}^n P_{\xi(x_{nj})}$ only through the components $P_{\xi(x_{nj})}$ with $j \in \mathbb{K}_n$. From that we can now conclude that

$$\begin{aligned} &\sup_{\xi \in \mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\big\{\hat{\eta}_n(\xi_0) \in [-\delta,\delta]^C\big\} \\ &\qquad = \sup_{\gamma \in \Theta^*} \sup_{g \in \mathscr{G}_{\mathrm{H}}^*} P_{\gamma,g}^{(n)}\big\{\hat{\eta}_n(\gamma) \in [-\delta,\delta]^C\big\}, \end{aligned} \tag{7.11}$$

where $\xi_0 = \xi(x_0)$ as before.

In Theorem 6.4 it was shown that $\mathscr{L}(\hat{\eta}_n(\gamma) \mid P_{\gamma,g}^{(n)})$ converges uniformly in distribution on $\Lambda = \overline{\Theta^*} \times \mathscr{G}_{\mathrm{H}}$ to a family of normal distributions. Together with the preceding observations this implies that $\mathscr{L}(\hat{\eta}_n(\xi_0) \mid P_\xi^{(n)})$ converges uniformly in distribution on $\mathscr{F}_{\mathrm{H}}$ (to the same family of normal distributions). If in addition we are able to show that the remainder $R(\xi_0, \xi_n^*)$ from (7.10) converges uniformly in probability to zero (in the sense of Definition 2.23), i.e.

$$\sup_{\xi \in \mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\{\|R(\xi_0,\xi_n^*)\| > \varepsilon\} \to 0 \tag{7.12}$$

for all $\varepsilon > 0$, then the generalised version of Slutsky's lemma (Lemma 2.24) implies that

$$\mathscr{L}(\hat{\eta}_n(\xi_0) + R(\xi_0,\xi_n^*) \mid P_\xi^{(n)})$$

converges uniformly in distribution on $\mathscr{F}_{\mathrm{H}}$, too. To sum up, the above arguments then yield

$$\begin{aligned} &\sup_{\xi \in \mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\big\{\hat{\xi}_n - \xi_0 \in [-a_n\delta, a_n\delta]^C\big\} \\ &\qquad \overset{(7.10)}{=} \sup_{\xi \in \mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\big\{\hat{\eta}_n(\xi_0) + R(\xi_0,\xi_n^*) \in [-\delta,\delta]^C\big\} \\ &\qquad \simeq \sup_{\xi \in \mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\big\{\hat{\eta}_n(\xi_0) \in [-\delta,\delta]^C\big\} \\ &\qquad \overset{(7.11)}{=} \sup_{\gamma \in \Theta^*} \sup_{g \in \mathscr{G}_{\mathrm{H}}^*} P_{\gamma,g}^{(n)}\big\{\hat{\eta}_n(\gamma) \in [-\delta,\delta]^C\big\} \\ &\qquad \to \sup_{\gamma \in \Theta^*} \sup_{g \in \mathscr{G}_{\mathrm{H}}} \mathscr{N}_{(\mu_{\gamma,g},\Sigma_\gamma)}\big([-\delta,\delta]^C\big). \end{aligned}$$

The convergence in the last step follows from Corollary 6.9 if $\mathscr{G}_{\mathrm{H}}$ is replaced by $\mathscr{G}_{\mathrm{H}}^*$, and the arguments at the beginning of the proof of Theorem 6.3 imply that it does not play a role, if the supremum is built with respect to $\mathscr{G}_{\mathrm{H}}$ or with respect to $\mathscr{G}_{\mathrm{H}}^*$.

Therefore, the theorem is proved if we can show that (7.12) holds. Following the definition of $R(\xi_0, \xi_n^*)$ we have

$$\begin{aligned}\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\{\|R(\xi_0,\xi_n^*)\| > \varepsilon\} &\leq \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Big\{\sum_{s=1}^{3}\|R_s(\xi_0,\xi_n^*)\| > \varepsilon\Big\} \\ &\leq \sum_{s=1}^{3}\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Big\{\|R_s(\xi_0,\xi_n^*)\| > \frac{\varepsilon}{3}\Big\}.\end{aligned}$$

The assertion (7.12) follows now from Lemmas 7.2–7.4. ■

7.2 Lemma. $\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\{\|R_1(\xi_0,\xi_n^*)\| > \varepsilon\} \to 0$ *for all* $\varepsilon > 0$.

7.3 Lemma. $\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\{\|R_2(\xi_0,\xi_n^*)\| > \varepsilon\} \to 0$ *for all* $\varepsilon > 0$.

7.4 Lemma. $\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\{\|R_3(\xi_0,\xi_n^*)\| > \varepsilon\} \to 0$ *for all* $\varepsilon > 0$.

7.2 Proofs

In this section we give the proofs of Lemmas 7.2–7.4. Throughout this complete section we presume that the assumptions from Theorem 7.1 hold, of course we use the notation used in the previous section.

7.2.1 Preliminary considerations

Before we start with the proofs, we provide some tools and auxiliary results, which will be needed later.

On Lipschitz conditions. From the discussion in Section 3.1 we know that the Cholesky matrix $\mathcal{C}_\gamma$ is continuous in the parameter. Furthermore, we have seen in Lemma 3.1 that the map $\gamma \mapsto v(\gamma)$ is continuous. These results will now be refined: we will show that $v(\cdot)$ and certain other maps even satisfy Lipschitz conditions. To this end, we first provide some basic results on Lipschitz continuous functions.

7.5 Lemma. *(i) Let* $f_1, \ldots, f_m \colon \mathbb{R}^p \to [0,\infty)$ *be Lipschitz continuous on* $U \subseteq \mathbb{R}^p$ *with Lipschitz constants* L_i*. Then* $f := \max_i f_i$ *is Lipschitz continuous on* U*, too, with constant* $L := \max_i L_i$*.*

(ii) *Let $f\colon \mathbb{R}^q \to \mathbb{R}^d$ and $g\colon \mathbb{R}^p \to \mathbb{R}^q$ be Lipschitz continuous on sets $W \subseteq \mathbb{R}^q$ and $V \subseteq \mathbb{R}^p$, respectively, with Lipschitz constants L_f and L_g. Suppose that $g(V) \subseteq W$. Then $f \circ g$ is Lipschitz continuous on V with Lipschitz constant $L_f L_g$.*

(iii) *Let $f, g\colon \mathbb{R}^p \to \mathbb{R}$ be Lipschitz continuous on $U \subseteq \mathbb{R}^p$ with Lipschitz constants L_f and L_g, respectively. Presume that f and g are also bounded on U, i.e. we have $\sup_{x\in U} |f(x)| \le M_f$ and $\sup_{x\in U} |g(x)| \le M_g$. Then fg is Lipschitz continuous on U with constant $L_f M_g + L_g M_f$.*

(iv) *Let $f\colon \mathbb{R}^p \to \mathbb{R}$ be Lipschitz continuous on $U \subseteq \mathbb{R}^p$ with constant L, and suppose that $f(x) \in [a, \infty)$ for all $x \in U$, where $a > 0$. Then $1/f$ is Lipschitz continuous on U with constant L/a^2.*

(v) *Let $U \subseteq \mathbb{R}^p$ be compact and convex. Let $f\colon U \to \mathbb{R}^q$ be locally Lipschitz in the sense that for each $x \in U$ there is an open neigbourhood U_x of x on which f is Lipschitz continuous. Then f is Lipschitz continuous on U.*

Proof. *Ad (i):* Let x and y be arbitrary points in U. Then $|f(x) - f(y)| = |f_k(x) - f_l(y)|$ for certain indices $k, l \in \{1, \ldots, m\}$. In case $k = l$ the Lipschitz continuity of f_k implies $|f(x) - f(y)| \le L\|x - y\|$. In the remaining case $k \ne l$ we can assume without loss of generality that $f_k(x) > f_l(y) \ge 0$. We also have $f_l(y) = \max_i f_i(y) \ge f_k(y)$, and thus

$$\begin{aligned} |f_k(x) - f_l(y)| &= f_k(x) - f_l(y) \\ &\le f_k(x) - f_k(y) \\ &= |f_k(x) - f_k(y)| \ \le \ L_k\|x - y\|. \end{aligned}$$

Ad (ii): Trivial.
Ad (iii): For all $x, y \in U$ we have

$$\begin{aligned} |f(x)g(x) - f(y)g(y)| &\le |f(x)g(x) - f(x)g(y)| \\ &\qquad\qquad + |f(x)g(y) - f(y)g(y)| \\ &\le M_f|g(x) - g(y)| + M_g|f(x) - f(y)| \\ &\le (M_f L_g + M_g L_f)\|x - y\|. \end{aligned}$$

Ad (iv): For all $x, y \in U$ we have

$$\begin{aligned}\left|\frac{1}{f(x)} - \frac{1}{f(y)}\right| &= \left|\frac{f(y)}{f(x)f(y)} - \frac{f(x)}{f(x)f(y)}\right| \\ &= \frac{1}{f(x)f(y)}|f(x) - f(y)| \leq \frac{1}{a^2}L\|x - y\|.\end{aligned}$$

Ad (v): Obviously, the sets U_x build an open cover of U. By compactness, we can select a finite subcover, i.e. we can select $x_1, \ldots, x_m \in U$ such that $U \subseteq \bigcup_{i=1}^m U_{x_i}$. We denote the local Lipschitz constants on the sets U_{x_i} by L_i, and we set $L := m \max_{i=1,\ldots,m} L_i$. Let now arbitrary points $x, y \in U$ be given. If $x, y \in U_{x_i}$ for some index i, then of course $\|f(x) - f(y)\| \leq L_i\|x - y\| \leq L\|x - y\|$. It remains to investigate the case that no such index i exists, i.e. we have $x \in U_{x_i}$, $y \in U_{x_j}$ and $i \neq j$. Since U is convex, the connecting line $V = V(x, y)$ lies in U. Consequently, there are $s \leq m - 1$ points $y_1, \ldots, y_s \in V$ and a sequence of indices $i = i_1, \ldots, i_s = j$ such that $y_r \in U_{x_{i_r}} \cap U_{x_{i_{r+1}}}$. From the triangle inequality and the local Lipschitz property we get

$$\begin{aligned}\|f(x) - f(y)\| &= \|f(x) \mp f(y_1) \mp \cdots \mp f(y_s) - f(y)\| \\ &\leq \|f(x) - f(y_1)\| + \sum_{r=1}^{s-1} \|f(y_r) - f(y_{r+1})\| \\ &\quad + \|f(y_s) - f(y)\| \\ &\leq L_{i_1}\|x - y_1\| + \sum_{r=1}^{s-1} L_{i_{r+1}}\|y_r - y_{r+1}\| + L_s\|y_s - y\| \\ &\leq \max_i L_i \left(\|x - y_1\| + \sum_{r=1}^{s-1} \|y_r - y_{r+1}\| + \|y_s - y\|\right) \\ &\leq m \max_i L_i \|x - y\|.\end{aligned}$$

The last inequality is a consequence of the fact that the distance between any two points $y_r, y_s \in V$ can never be greater than that between x and y. Altogether, the above calculations yield

$$\|f(x) - f(y)\| \leq m \max_i L_i \|x - y\| = L\|x - y\|,$$

and thus the stated global Lipschitz continuity of f. ∎

We will apply this lemma for the proofs of the following results. Moreover, we will use some of the notation that was already used for the construction of $v = v(\gamma)$ in Section 3.5.3. Recall the definition of the set $\Theta' = \{\xi(x) \mid x \in [0,1], \xi \in \mathscr{F}\}$ given in (3.7).

7.6 Proposition. *The maps $\gamma \mapsto \mathcal{C}_\gamma$, $\gamma \mapsto \mathcal{C}_\gamma^{-1}$, $\gamma \mapsto \|\mathcal{C}_\gamma^j\|_1$ and $\gamma \mapsto v_j(\gamma)$ $(j = 1, \ldots, d)$ are Lipschitz continuous on $\overline{\Theta'}$.*

Proof. According to the assumptions from Section 3.2 the entries of the information matrix $\mathcal{I}_\gamma$ are—as functions of γ—continuously partially differentiable on Θ. Since the entries of the corresponding Cholesky decomposition $\mathcal{C}_\gamma$ and its inverse $\mathcal{C}_\gamma^{-1}$ result from linear combinations of the entries of $\mathcal{I}_\gamma$, it follows that the components of the maps $\gamma \mapsto \mathcal{C}_\gamma$ are continuously partially differentiable on Θ, too (cf. Opfer, 2001, p. 175 f.). An analogue result holds for the maps $\gamma \mapsto \mathcal{C}_\gamma^{-1}$ and $\gamma \mapsto \mathcal{I}_\gamma^{-1}$.

As a consequence, the maps $\gamma \mapsto \mathcal{C}_\gamma^j$ $(j = 1, \ldots, d)$ are continuously partially differentiable. In particular, they are Lipschitz continuous on $\overline{\Theta'} \subseteq \Theta$, which follows from the compactness of $\overline{\Theta'}$ and an application of the mean value theorem (cf. Königsberger, 2000, p. 56). An analogue result holds of course for $\gamma \mapsto \|\mathcal{C}_\gamma^j\|_1$. This proves the first part of the assertion.

Furthermore, the just shown partial differentiability of the map $\mathcal{C}_{(\cdot)}$ implies that the vectors $w^{(i)} = w^{(i)}(\gamma) = \mathcal{C}_\gamma \delta^{(i)}$ from (3.18)—interpreted as $\mathbb{R}^d$-valued functions—are continuously partially differentiable on Θ, and thus of course Lipschitz continuous on $\overline{\Theta'}$. From (i) and (ii) in Lemma 7.5 then it follows that also the maps $\gamma \mapsto w(\gamma)$ (since the absolute value is Lipschitz continuous) and $\gamma \mapsto \widetilde{w}^{(i)}(\gamma)$ as in (3.19) and (3.20) are Lipschitz continuous on $\overline{\Theta'}$. The same also holds for the maps $z^{(i)}(\gamma)$. With (i) in Lemma 7.5 it follows, that the function $\alpha = \alpha(\gamma)$ from (3.21) is Lipschitz continuous on $\overline{\Theta'}$. Furthermore, α is on $\overline{\Theta'}$ strictly bounded away from zero, which follows from the continuity of the $z^{(i)}(\gamma)$ and the definition of α. Therefore, assertions (iv) and (iii) in Lemma 7.5 imply that the map $\gamma \mapsto 1/\alpha(\gamma)$, and hence the product $v(\gamma) = w(\gamma)/\alpha(\gamma)$, too, are Lipschitz continuous on $\overline{\Theta'}$. ∎

7.7 Proposition. *The maps $\gamma \mapsto \beta_j(\gamma)$ and $\gamma \mapsto S_j(\gamma)$ $(j = 1, \ldots, d)$ are Lipschitz continuous on $\overline{\Theta'}$.*

Proof. In Lemma 6.6 we already proved that $\beta(\gamma)$ is continuous. As in the proof of that lemma, let an arbitrary $\gamma_0 \in \Theta$ be given and let

$$(x^*, y^*, z^*) = (v_j(\gamma), \|\mathcal{C}^j_{\gamma_0}\|_1, \beta_j(\gamma_0)).$$

The implicit function theorem guaranteed the existence of open neighbourhoods $U_{(x^*,y^*)}$ and U_{z^*} of (x^*, y^*) and z^*, respectively, and of a unique, continuously differentiable function $f\colon U_{(x^*,y^*)} \to U_{z^*}$ satisfying $f(x^*, y^*) = z^*$ and (6.14). We may without loss of generality assume that the partial derivatives of f are bounded on $U_{(x^*,y^*)}$. (Otherwise, one may simply scale down the sets $U_{(x^*,y^*)}$ and U_{z^*} and thus pass over to suitable subsets for which this holds true.) Thus, as a consequence of the mean value theorem f is Lipschitz continuous. In combination with Proposition 7.6 follows the existence of a neighbourhood U_{γ_0} of γ_0 such that $q = (v_j(\cdot), \|\mathcal{C}^j_{(\cdot)}\|_1)\colon U_{\gamma_0} \to U_{(x^*,y^*)}$ is Lipschitz continuous. Then $f \circ q$ is Lipschitz continuous on U_{γ_0}, too. This follows from (ii) in Lemma 7.5. With similar arguments as in the proof of Lemma 6.6 we conclude from the uniqueness of f that the map $\gamma \mapsto \beta_j(\gamma)$ is Lipschitz continuous on U_{γ_0}. Since $\gamma_0 \in \Theta$ was chosen arbitrarily, this implies that β_j and thus β are locally Lipschitz on Θ. In combination with (v) of Lemma 7.5 this implies the assertion. ■

7.8 Remark. In the univariate case (i.e. for $d = 1$) the result from Proposition 7.7 can be proved much more simple. In this special case, the function F from the proof of Lemma 6.6 (which is used to prove continuity of $\beta(\cdot)$) can be simplified to

$$F\colon \Theta \times (\delta, \infty) \to \mathbb{R}, \quad (x, z) \mapsto \frac{z^{1+1/\rho}}{\log \frac{z+\delta}{z-\delta}} - \frac{\rho + 1}{2\mathcal{I}_x \delta}.$$

The implicit function theorem directly implies that for each θ there is a neighbourhood U_θ on which $\gamma \mapsto \beta(\gamma)$ is continuously differentiable and thus locally Lipschitz. Global Lipschitz continuity of $\beta(\cdot)$ is then of course deduced with the same argument as in the proof of Proposition 7.7.

The problem of the general multivariate case $d \geq 1$ in comparison with the univariate one is that the additional functions $\gamma \mapsto v_j(\gamma)$ must be incorporated. In general, these are not differentiable, because in the construction of these functions we often take absolute values (such as in (3.19)), which might destroy differentiability properties. This prohibits us to directly conclude that the function $f \circ q$ from the proof of Lemma 6.6 is continuously differentiable and thus (locally) Lipschitz. Therefore, we are forced to make the above detour to prove the assertion of the theorem.

Matrix norms. As is well known, the Euclidean norm $\|\cdot\|$ on $\mathbb{R}^d$ induces a norm $|||\cdot|||$ on $\mathbb{R}^{d\times d}$, the vector space of $d \times d$-matrices, by the relation

$$|||A||| := \sup_{x\in\mathbb{R}^d, x\neq 0} \frac{\|Ax\|}{\|x\|}, \quad A \in \mathbb{R}^{d\times d}.$$

Since A is a linear map (or *operator*), $|||\cdot|||$ is sometimes also referred to as the *operator norm* with respect to the Euclidean norm. Note that all norms on $\mathbb{R}^{d\times d}$ are equivalent, which means in particular that for matrices $A_n, A \in \mathbb{R}^{d\times d}$ the following holds:

$$A_n \to A \text{ with respect to } |||\cdot||| \;\Leftrightarrow\; A_n \to A \text{ component-wise}$$

(cf. Königsberger, 2000, p. 27). In particular, one may conclude that if $A_n \to A$ component-wise, then also $|||A_n||| \to |||A|||$. (Of course, the inverse statement does in general not hold true.) A different norm on $\mathbb{R}^{d\times d}$ is given by the *Frobenius norm*, which is defined as

$$\|A\|_{\mathrm{F}} := \left(\sum_{i,j=1}^{d} a_{ij}^2\right)^{1/2}, \quad A = (a_{ij}) \in \mathbb{R}^{d\times d}.$$

Of course, one may rearrange the elements of A and write it as a vector of length d^2. In this case, the Frobenius norm coincides with the Euclidean norm on $\mathbb{R}^{d^2}$. Note that we have the inequalities

$$\begin{aligned} &\|Av\| \le |||A|||\,\|v\|, \\ &|||AB||| \le |||A|||\,|||B|||, \quad \text{for all } A, B \in \mathbb{R}^{d\times d},\ v \in \mathbb{R}^d. \end{aligned} \tag{7.13}$$

These will be of great value for the proofs of the lemmas. For further information on matrix norms see Opfer (2001), p. 228 ff., or Königsberger (2000), p. 25 ff.

The role of the preliminary estimator. For the proof of Lemmas 7.2–7.4 we have to show that for $s = 1, 2, 3$ and every $\varepsilon > 0$

holds

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\{\|R_s(\xi_0,\xi_n^*)\| > \varepsilon\} \to 0.$$

To do so, we can often exploit the properties of the preliminary estimator. Besides $\varepsilon > 0$ let some $\iota > 0$ be given, and let $C = C_\iota$ be such that (7.1) is satisfied. Then

$$\begin{aligned}
&\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\{\|R_s(\xi_0,\xi_n^*)\| > \varepsilon\} \\
&\qquad\qquad \le \quad \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\big\{\|R_s(\xi_0,\xi_n^*)\| > \varepsilon,\ \tfrac{1}{a_n}\|\xi_n^* - \xi_0\| > C\big\} \\
&\qquad\qquad\quad\ + \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\big\{\|R_s(\xi_0,\xi_n^*)\| > \varepsilon,\ \tfrac{1}{a_n}\|\xi_n^* - \xi_0\| \le C\big\} \\
&\qquad\qquad \le \quad \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\big\{\tfrac{1}{a_n}\|\xi_n^* - \xi_0\| > C\big\} \\
&\qquad\qquad\quad\ + \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Big\{\sup_{\|z-\xi_0\|\le a_n C} \|R_s(\xi_0,z)\| > \varepsilon\Big\}.
\end{aligned}$$

According to (7.1), the limit superior of the first of these two summands is bounded by ι (as $n \to \infty$). If, in addition, we are able to show that the supremum on the right hand side of the last display converges to zero, we can conclude that

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\{\|R_s(\xi_0,\xi_n^*)\| > \varepsilon\} \\
&\qquad \le \ \iota \ + \ \limsup_{n\to\infty} \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Big\{\sup_{\|z-\xi_0\|\le a_n C} \|R_s(\xi_0,z)\| > \varepsilon\Big\} = \iota.
\end{aligned}$$

Since ι can be chosen arbitrarily small, this implies

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\{\|R_s(\xi_0,\xi_n^*)\| > \varepsilon\} \to 0,$$

and thus one of the Lemmas 7.2–7.4 (depending on whether $s = 1,2,3$). Therefore, it remains to prove that for all $C > 0$, $\varepsilon > 0$ and $s = 1,2,3$ the assertion

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Big\{\sup_{\|z-\xi_0\|\le a_n C} \|R_s(\xi_0,z)\| > \varepsilon\Big\} \to 0 \tag{7.14}$$

holds true.

7.2.2 Proof of Lemma 7.2

Throughout the following calculations let $\xi_0 = \xi(x_0)$. We prove (7.14) for the case $s = 1$. To this end, let some arbitrary, but fixed $C > 0$ and $\varepsilon > 0$ be given. We show that

$$\sup_{\xi \in \mathscr{F}_{\mathrm{H}}} P_\xi^{(n)} \left\{ \sup_{\|z-\xi_0\| \le a_n C} \|R_1(\xi_0, z)\| > \varepsilon \right\} \to 0. \tag{7.15}$$

Together with the considerations from the end of Section 7.2.1 this implies the assertion of Lemma 7.2.

Some notation and auxiliary results. Let us first consider a fixed $\xi \in \mathscr{F}_{\mathrm{H}}$, and let $\xi_0 = \xi(x_0)$. We set

$$\begin{aligned} \beta_n^{\min} &:= \beta_n^{\min}(\xi_0) := \inf\{\beta_i(z) : \|z - \xi_0\| \le a_n C,\ i = 1, \ldots, d\}, \\ \beta_n^{\max} &:= \beta_n^{\max}(\xi_0) := \sup\{\beta_i(z) : \|z - \xi_0\| \le a_n C,\ i = 1, \ldots, d\}. \end{aligned}$$

As usual, $\beta_i(z)$ denotes the solution to equation (5.44) with γ replaced by z. By definition each of the functions $\beta_i(\cdot)$ is strictly positive and furthermore continuous on Θ (Lemma 6.6), and thus

$$\begin{aligned} 0 \;&<\; \inf_{\xi_0 \in \overline{\Theta^*}} \; \liminf_{n\to\infty} \; \beta_n^{\min}(\xi_0) \\ &\le\; \sup_{\xi_0 \in \overline{\Theta^*}} \; \limsup_{n\to\infty} \; \beta_n^{\max}(\xi_0) \;<\; \infty. \end{aligned} \tag{7.16}$$

Moreover, it is clear that for $\xi_0 \in \overline{\Theta^*}$ and all z such that $\|z - \xi_0\| \le a_n C$ holds, we eventually have $z \in \overline{\Theta'}$, provided that n is sufficiently large. Since $\beta(\cdot)$ is even Lipschitz continuous on that set (according to Proposition 7.7) we conclude that

$$\sup_{\xi_0 \in \overline{\Theta^*}} \{\beta_n^{\max}(\xi_0) - \beta_n^{\min}(\xi_0)\} = O(a_n). \tag{7.17}$$

Besides the above quantities we introduce index sets

$$\mathbb{H}_n := \mathbb{H}_n(\xi_0) := \mathbb{K}_n \cap \{j = 1, \ldots, n : |\tilde{x}_{nj}| \le (\beta_n^{\min})^{1/\rho}\},$$

and

$$\begin{aligned} \mathbb{M}_n &:= \mathbb{M}_n(\xi_0) \\ &:= \mathbb{K}_n \cap \{j = 1, \ldots, n : (\beta_n^{\min})^{1/\rho} < |\tilde{x}_{nj}| \le (\beta_n^{\max})^{1/\rho}\}, \end{aligned}$$

where $\mathbb{K}_n$ is defined as in (4.7). Because of $\mathbb{H}_n \subseteq \mathbb{K}_n$ we have $\#\mathbb{H}_n(\xi_0) \le \#\mathbb{K}_n = k_n$ for all ξ_0. Moreover,

$$\begin{aligned}\#\mathbb{M}_n &\lesssim [(\beta_n^{\max})^{1/\rho} - (\beta_n^{\min})^{1/\rho}]n/b_n \\ &\lesssim (\beta_n^{\max} - \beta_n^{\min})n/b_n.\end{aligned}$$

The last inequality follows from (7.16) in combination with the fact that the map $x \mapsto x^{1/\rho}$ is Lipschitz continuous on every compact interval lying in $(0,\infty)$. Together with (7.17) this leads to

$$\sup_{\xi_0\in\overline{\Theta^*}} \#\mathbb{M}_n(\xi_0) \lesssim \frac{a_n n}{b_n}. \tag{7.18}$$

The index set $\mathbb{H}_n$ is motivated by the following observation: let ξ_0 be given and z be such that $\|z-\xi_0\| \le a_n C$. Then, of course, $\beta(\xi_0), \beta(z) > \beta_n^{\min}$, and thus

$$|g_{\beta(z)}(\tilde{x}_{nj}) - g_{\beta(\xi_0)}(\tilde{x}_{nj})| \overset{(5.15)}{=} |\beta(\xi_0) - \beta(z)|, \quad j \in \mathbb{H}_n. \tag{7.19}$$

This relation will turn out helpful in later investigations.

In the following calculations we will often consider the supremum with respect to the set $\{z : \|z-\xi_0\| \le a_n C\}$, for given $\xi_0 \in \Theta^*$. In order to keep the notation short, we will often write this as $\sup_{\|z-\xi_0\|\le a_n C}(\dots)$ or shortly $\sup_z(\dots)$.

Proof of Lemma 7.2. With these additional considerations we can now assess the asymptotic behaviour of the remainder term R_1 from the definition in (7.6). If we replace ξ_n^* with z in that definition, we get

$$R_1(\xi_0, z) = \sum_{j=1}^{n} a_n \Big(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top\Big)\, \dot{\ell}_{\xi_0}(Y_{nj}).$$

From the definition in (7.4) it follows that $B_{\xi_0}^{nj} = B_z^{nj} = 0$ for all indices j with $|\tilde{x}_{nj}| > (\beta_n^{\max})^{1/\rho}$. Therefore, it suffices to build the sum in the above display only with respect to all j for which this is satisfied, namely for $j \in \mathbb{H}_n \cup \mathbb{M}_n$. Hence,

$$
\begin{aligned}
&\sup_{\|z-\xi_0\| \le a_n C} \|R_1(\xi_0, z)\| \\
&\quad \le \sup_z \Bigg\| \sum_{j \in \mathbb{H}_n} a_n \Big(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top \\
&\qquad\qquad - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \Big) \dot{\ell}_{\xi_0}(Y_{nj}) \Bigg\| \\
&\qquad + \sup_z \Bigg\| \sum_{j \in \mathbb{M}_n} a_n \Big(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top \\
&\qquad\qquad - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \Big) \dot{\ell}_{\xi_0}(Y_{nj}) \Bigg\|. \qquad (7.20)
\end{aligned}
$$

In order to prove that $R_1(\xi_0, z)$ converges uniformly in probability to zero we show that both of the suprema on the right hand side of (7.20) do so.

Assessment of the first expression in (7.20). To prove that the first expression in (7.20) converges uniformly in probability to zero, we first derive a (rather crude) upper bound, which does only depend on the previously fixed ξ_0. We then show that this upper bound converges uniformly in probability to zero. Adding and subtracting an additional term $\mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_z^{nj} (\mathcal{C}_z^{-1})^\top$, we obtain

$$
\begin{aligned}
&\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \\
&\quad = \mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top \pm \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_z^{nj} (\mathcal{C}_z^{-1})^\top - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \\
&\quad = \left(\mathcal{C}_z^{-1} S_z - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} \right) B_z^{nj} (\mathcal{C}_z^{-1})^\top \\
&\qquad\qquad + \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} \left(B_z^{nj} (\mathcal{C}_z^{-1})^\top - B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \right). \qquad (7.21)
\end{aligned}
$$

Hence, a multiple application of the triangle inequality and the inequalities for the operator norm yield

$$
\begin{aligned}
&\sup_z \Bigg\| \sum_{j\in\mathbb{H}_n} a_n \left(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \right) \dot\ell_{\xi_0}(Y_{nj}) \Bigg\| \\
&\leq \sup_z \Bigg\| \sum_{j\in\mathbb{H}_n} a_n \left(\mathcal{C}_z^{-1} S_z - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} \right) B_z^{nj} (\mathcal{C}_z^{-1})^\top \dot\ell_{\xi_0}(Y_{nj}) \Bigg\| \\
&\quad + \sup_z \Bigg\| \sum_{j\in\mathbb{H}_n} a_n \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} \left(B_z^{nj} (\mathcal{C}_z^{-1})^\top - B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \right) \dot\ell_{\xi_0}(Y_{nj}) \Bigg\| \\
&\overset{(7.13)}{\leq} \sup_z \|\!| \mathcal{C}_z^{-1} S_z - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} \|\!| \quad \sup_z \Bigg\| \sum_{j\in\mathbb{H}_n} a_n B_z^{nj} (\mathcal{C}_z^{-1})^\top \dot\ell_{\xi_0}(Y_{nj}) \Bigg\| \\
&\quad + \|\!| \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} \|\!| \ \sup_z \Bigg\| \sum_{j\in\mathbb{H}_n} a_n \left(B_z^{nj} (\mathcal{C}_z^{-1})^\top - B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \right) \dot\ell_{\xi_0}(Y_{nj}) \Bigg\| \\
&\overset{(*)}{\lesssim} \sup_z \|\!| \mathcal{C}_z^{-1} S_z - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} \|\!| \sum_{i=1}^d \sup_z \Bigg| \sum_{j\in\mathbb{H}_n} a_n \mathrm{e}_i^\top B_z^{nj} (\mathcal{C}_z^{-1})^\top \dot\ell_{\xi_0}(Y_{nj}) \Bigg| \\
&\quad + \|\!| \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} \|\!| \sum_{i=1}^d \sup_z \Bigg| \sum_{j\in\mathbb{H}_n} a_n \mathrm{e}_i^\top \Bigg(B_z^{nj} (\mathcal{C}_z^{-1})^\top \\
&\qquad\qquad - B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \Bigg) \dot\ell_{\xi_0}(Y_{nj}) \Bigg| .
\end{aligned}
$$

Note that in the last step $(*)$ we simply exploited that $\|w\| \lesssim \|w\|_1 = \sum_{i=1}^d |w_i|$ for all vectors $w \in \mathbb{R}^d$. Here e_i denotes the ith unit vector. Since the matrices $\mathcal{C}_{(\cdot)}$ and $S_{(\cdot)}$ (i.e. their components) are Lipschitz continuous on $\overline{\Theta'}$, the same holds true for the map $\mathcal{C}_{(\cdot)}^{-1} S_{(\cdot)}$ (see the results from Section 7.2.1). Hence,

$$
\sup_z \|\!| \mathcal{C}_z^{-1} S_z - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} \|\!| \lesssim a_n, \tag{7.22}
$$

where the right hand side does not depend on ξ_0. Moreover, because the considered map is continuous, we can bound $\|\!| \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} \|\!|$ uniformly in ξ_0. These properties, combined with the preceding calculations, yield

$$
\begin{aligned}
&\sup_z \left\| \sum_{j\in\mathbb{H}_n} a_n \left(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \right) \dot{\ell}_{\xi_0}(Y_{nj}) \right\| \\
\lesssim\ & a_n \sum_{i=1}^d \sup_z \left| \sum_{j\in\mathbb{H}_n} a_n \mathrm{e}_i^\top B_z^{nj} (\mathcal{C}_z^{-1})^\top \dot{\ell}_{\xi_0}(Y_{nj}) \right| \\
&+ \sum_{i=1}^d \sup_z \left| \sum_{j\in\mathbb{H}_n} a_n \mathrm{e}_i^\top \left(B_z^{nj} (\mathcal{C}_z^{-1})^\top - B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \right) \dot{\ell}_{\xi_0}(Y_{nj}) \right| \\
\overset{(7.4)}{=}\ & a_n \sum_{i=1}^d \sup_z \left| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(z)}(\tilde{x}_{nj}) \mathrm{e}_i^\top (\mathcal{C}_z^{-1})^\top \dot{\ell}_{\xi_0}(Y_{nj}) \right| \\
&+ \sum_{i=1}^d \sup_z \left| \sum_{j\in\mathbb{H}_n} a_n \mathrm{e}_i^\top \Big(g_{\beta_i(z)}(\tilde{x}_{nj}) (\mathcal{C}_z^{-1})^\top \right. \\
&\qquad\qquad \left. - g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) (\mathcal{C}_{\xi_0}^{-1})^\top \Big) \dot{\ell}_{\xi_0}(Y_{nj}) \right|. \qquad (7.23)
\end{aligned}
$$

Using the Cauchy-Schwarz inequality we find an upper bound for the summands of the first sum in (7.23) by

$$
\begin{aligned}
&\left| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(z)}(\tilde{x}_{nj}) \mathrm{e}_i^\top (\mathcal{C}_z^{-1})^\top \dot{\ell}_{\xi_0}(Y_{nj}) \right| \\
&\qquad \le \|\mathcal{C}_z^{-1} \mathrm{e}_i\| \left\| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(z)}(\tilde{x}_{nj}) \dot{\ell}_{\xi_0}(Y_{nj}) \right\| \\
&\qquad \overset{(7.13)}{\le} |||\mathcal{C}_z^{-1}||| \left\| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(z)}(\tilde{x}_{nj}) \dot{\ell}_{\xi_0}(Y_{nj}) \right\| \\
&\qquad \lesssim \left\| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(z)}(\tilde{x}_{nj}) \dot{\ell}_{\xi_0}(Y_{nj}) \right\|. \qquad (7.24)
\end{aligned}
$$

This last expression can further be bounded from above in the following way:

$$
\begin{aligned}
&\Big\| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(z)}(\tilde{x}_{nj})\dot{\ell}_{\xi_0}(Y_{nj})\Big\| \\
&\quad\leq \Big\| \sum_{j\in\mathbb{H}_n} a_n \left(g_{\beta_i(z)}(\tilde{x}_{nj}) - g_{\beta_i(\xi_0)}(\tilde{x}_{nj})\right) \dot{\ell}_{\xi_0}(Y_{nj})\Big\| \\
&\qquad + \Big\| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(\xi_0)}(\tilde{x}_{nj})\dot{\ell}_{\xi_0}(Y_{nj})\Big\| \\
&\quad\overset{(7.19)}{=} |\beta_i(z)-\beta_i(\xi_0)| \quad \Big\| \sum_{j\in\mathbb{H}_n} a_n \dot{\ell}_{\xi_0}(Y_{nj})\Big\| \\
&\qquad + \Big\| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(\xi_0)}(\tilde{x}_{nj})\dot{\ell}_{\xi_0}(Y_{nj})\Big\| \\
&\quad\overset{(\dagger)}{\lesssim} a_n \Big\| \sum_{j\in\mathbb{H}_n} a_n \dot{\ell}_{\xi_0}(Y_{nj})\Big\| + \Big\| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(\xi_0)}(\tilde{x}_{nj})\dot{\ell}_{\xi_0}(Y_{nj})\Big\|, \qquad (7.25)
\end{aligned}
$$

where $(\dagger)$ is a consequence of the Lipschitz continuity of $\beta_i(\cdot)$ (Proposition 7.7). Note that (7.25) also incorporates the inequality

$$
\begin{aligned}
&\Big\| \sum_{j\in\mathbb{H}_n} a_n \left(g_{\beta_i(z)}(\tilde{x}_{nj}) - g_{\beta_i(\xi_0)}(\tilde{x}_{nj})\right) \dot{\ell}_{\xi_0}(Y_{nj})\Big\| \\
&\qquad\qquad\qquad\qquad \lesssim \Big\| \sum_{j\in\mathbb{H}_n} a_n^2 \dot{\ell}_{\xi_0}(Y_{nj})\Big\|, \qquad (7.26)
\end{aligned}
$$

which turns out useful in later examinations. Summarising, the above calculations yield the following upper bound for the first of the two expressions in (7.23):

$$
\begin{aligned}
&a_n \sum_{i=1}^{d} \sup_z \Big| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(z)}(\tilde{x}_{nj}) \mathrm{e}_i^\top (\mathcal{C}_z^{-1})^\top \dot{\ell}_{\xi_0}(Y_{nj})\Big| \\
&\quad\overset{(7.24)}{\lesssim} a_n \sum_{i=1}^{d} \sup_z \Big\| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(z)}(\tilde{x}_{nj})\dot{\ell}_{\xi_0}(Y_{nj})\Big\| \\
&\quad\overset{(7.25)}{\lesssim} a_n \sum_{i=1}^{d} \sup_z \Bigg(\Big\| \sum_{j\in\mathbb{H}_n} a_n^2 \dot{\ell}_{\xi_0}(Y_{nj})\Big\| \dots
\end{aligned}
$$

$$
\begin{aligned}
&\qquad + \Big\| \sum_{j \in \mathbb{H}_n} a_n g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| \Big) \\
\lesssim\ & a_n \Big\| \sum_{j \in \mathbb{H}_n} a_n^2 \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| \\
&\qquad + a_n \sum_{i=1}^{d} \Big\| \sum_{j \in \mathbb{H}_n} a_n g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\|. \qquad (7.27)
\end{aligned}
$$

With similar calculations we may bound the second sum in (7.23). From an application of the triangle and Cauchy-Schwarz inequatl-ities we get

$$
\begin{aligned}
&\Big| \sum_{j \in \mathbb{H}_n} a_n \mathrm{e}_i^\top \left(g_{\beta_i(z)}(\tilde{x}_{nj}) (\mathcal{C}_z^{-1})^\top - g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) (\mathcal{C}_{\xi_0}^{-1})^\top \right) \dot{\ell}_{\xi_0}(Y_{nj}) \Big| \\
\le\ & \Big| \sum_{j \in \mathbb{H}_n} a_n \mathrm{e}_i^\top \left(g_{\beta_i(z)}(\tilde{x}_{nj}) (\mathcal{C}_z^{-1})^\top - g_{\beta_i(z)}(\tilde{x}_{nj}) (\mathcal{C}_{\xi_0}^{-1})^\top \right) \dot{\ell}_{\xi_0}(Y_{nj}) \Big| \\
&\quad + \Big| \sum_{j \in \mathbb{H}_n} a_n \mathrm{e}_i^\top \left(g_{\beta_i(z)}(\tilde{x}_{nj}) (\mathcal{C}_{\xi_0}^{-1})^\top - g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) (\mathcal{C}_{\xi_0}^{-1})^\top \right) \dot{\ell}_{\xi_0}(Y_{nj}) \Big| \\
\le\ & \|(\mathcal{C}_z^{-1} - \mathcal{C}_{\xi_0}^{-1}) \mathrm{e}_i\| \Big\| \sum_{j \in \mathbb{H}_n} a_n g_{\beta_i(z)}(\tilde{x}_{nj}) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| \\
&\quad + \|\mathcal{C}_{\xi_0}^{-1} \mathrm{e}_i\| \Big\| \sum_{j \in \mathbb{H}_n} a_n \left(g_{\beta_i(z)}(\tilde{x}_{nj}) - g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) \right) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| \\
\overset{(7.13)}{\le}\ & |\!|\!| \mathcal{C}_z^{-1} - \mathcal{C}_{\xi_0}^{-1} |\!|\!| \Big\| \sum_{j \in \mathbb{H}_n} a_n g_{\beta_i(z)}(\tilde{x}_{nj}) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| \\
&\quad + |\!|\!| \mathcal{C}_{\xi_0}^{-1} |\!|\!| \Big\| \sum_{j \in \mathbb{H}_n} a_n \left(g_{\beta_i(z)}(\tilde{x}_{nj}) - g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) \right) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\|.
\end{aligned}
$$

The same Lipschitz and continuity arguments that were used before for the proof of (7.22) yield $|\!|\!| \mathcal{C}_z^{-1} - \mathcal{C}_{\xi_0}^{-1} |\!|\!| \lesssim a_n$ and, in addition,

$\|\mathcal{C}_{\xi_0}^{-1}\|$ can be uniformly bounded by some constant. Consequently,

$$\begin{aligned}
&\Big| \sum_{j\in\mathbb{H}_n} a_n \mathrm{e}_i^\top \left(g_{\beta_i(z)}(\tilde{x}_{nj})(\mathcal{C}_z^{-1})^\top - g_{\beta_i(\xi_0)}(\tilde{x}_{nj})(\mathcal{C}_{\xi_0}^{-1})^\top \right) \dot{\ell}_{\xi_0}(Y_{nj}) \Big| \\
&\quad \lesssim a_n \Big\| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(z)}(\tilde{x}_{nj}) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| \\
&\qquad + \Big\| \sum_{j\in\mathbb{H}_n} a_n \left(g_{\beta_i(z)}(\tilde{x}_{nj}) - g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) \right) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| \\
&\quad \overset{(7.25)}{\lesssim} a_n \left(\Big\| \sum_{j\in\mathbb{H}_n} a_n^2 \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| + \Big\| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| \right) \\
&\qquad + \Big\| \sum_{j\in\mathbb{H}_n} a_n \left(g_{\beta_i(z)}(\tilde{x}_{nj}) - g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) \right) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| \\
&\quad \overset{(7.26)}{\lesssim} a_n \left(\Big\| \sum_{j\in\mathbb{H}_n} a_n^2 \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| + \Big\| \sum_{j\in\mathbb{H}_n} a_n g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| \right) \\
&\qquad + \Big\| \sum_{j\in\mathbb{H}_n} a_n^2 \dot{\ell}_{\xi_0}(Y_{nj}) \Big\|.
\end{aligned} \tag{7.28}$$

Therefore, for the second expression in (7.23) we find an upper bound by

$$\begin{aligned}
&\sum_{i=1}^d \sup_z \Big| \sum_{j\in\mathbb{H}_n} a_n \mathrm{e}_i^\top \left(g_{\beta_i(z)}(\tilde{x}_{nj})(\mathcal{C}_z^{-1})^\top - g_{\beta_i(\xi_0)}(\tilde{x}_{nj})(\mathcal{C}_{\xi_0}^{-1})^\top \right) \dot{\ell}_{\xi_0}(Y_{nj}) \Big| \\
&\quad \overset{(7.28)}{\lesssim} a_n \Big\| \sum_{j\in\mathbb{H}_n} a_n^2 \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| + \Big\| \sum_{j\in\mathbb{H}_n} a_n^2 \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| \\
&\qquad\qquad + \sum_{i=1}^d \Big\| \sum_{j\in\mathbb{H}_n} a_n^2 g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| \\
&\quad \lesssim \Big\| \sum_{j\in\mathbb{H}_n} a_n^2 \dot{\ell}_{\xi_0}(Y_{nj}) \Big\| + \sum_{i=1}^d \Big\| \sum_{j\in\mathbb{H}_n} a_n^2 g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) \dot{\ell}_{\xi_0}(Y_{nj}) \Big\|.
\end{aligned}$$

The last step follows simply from the inequality $1 + a_n \leq 2$. Combining this result with (7.27) and (7.23), we get

$$\sup_{\|z-\xi_0\|\le a_n C}\left\|\sum_{j\in\mathbb{H}_n} a_n\left(\mathcal{C}_z^{-1}S_zB_z^{nj}(\mathcal{C}_z^{-1})^\top - \mathcal{C}_{\xi_0}^{-1}S_{\xi_0}B_{\xi_0}^{nj}(\mathcal{C}_{\xi_0}^{-1})^\top\right)\dot{\ell}_{\xi_0}(Y_{nj})\right\|$$
$$\lesssim \left\|\sum_{j\in\mathbb{H}_n} a_n^2\dot{\ell}_{\xi_0}(Y_{nj})\right\| + \sum_{i=1}^d\left\|\sum_{j\in\mathbb{H}_n} a_n^2 g_{\beta_i(\xi_0)}(\tilde{x}_{nj})\dot{\ell}_{\xi_0}(Y_{nj})\right\|. \qquad (7.29)$$

Next we prove that both of the expressions in (7.29) converge uniformly in probability to zero. For arbitrary, but fixed $\xi \in \mathscr{F}_{\mathrm{H}}$ Markov's inequality yields

$$P_\xi^{(n)}\left\{\left\|\sum_{j\in\mathbb{H}_n} a_n^2\dot{\ell}_{\xi_0}(Y_{nj})\right\| > \frac{\varepsilon}{2}\right\} \le \frac{4}{\varepsilon^2}\,\mathrm{E}\left\|\sum_{j\in\mathbb{H}_n} a_n^2\dot{\ell}_{\xi_0}(Y_{nj})\right\|^2, \quad (7.30)$$

and

$$P_\xi^{(n)}\left\{\sum_{i=1}^d\left\|\sum_{j\in\mathbb{H}_n} a_n^2 g_{\beta_i(\xi_0)}(\tilde{x}_{nj})\dot{\ell}_{\xi_0}(Y_{nj})\right\| > \frac{\varepsilon}{2}\right\}$$
$$\le \sum_{i=1}^d P_\xi^{(n)}\left\{\left\|\sum_{j\in\mathbb{H}_n} a_n^2 g_{\beta_i(\xi_0)}(\tilde{x}_{nj})\dot{\ell}_{\xi_0}(Y_{nj})\right\| > \frac{\varepsilon}{2d}\right\}$$
$$\le \sum_{i=1}^d \frac{4d^2}{\varepsilon^2}\,\mathrm{E}\left\|\sum_{j\in\mathbb{H}_n} a_n^2 g_{\beta_i(\xi_0)}(\tilde{x}_{nj})\dot{\ell}_{\xi_0}(Y_{nj})\right\|^2. \qquad (7.31)$$

The expectation in the above displays is to be built with respect to the product measure $P_\xi^{(n)}$. In order to prove that the expressions from (7.29) converge uniformly in probability to zero, it is sufficient to prove that the expectations in (7.30) and (7.31) converge to zero uniformly as $n \to \infty$. We can also write these expectations as

$$\mathrm{E}\left\|\sum_{j\in\mathbb{H}_n} a_n^2 w_{nj}\dot{\ell}_{\xi_0}(Y_{nj})\right\|^2,$$

where either $w_{nj} = g_{\beta_i}(\tilde{x}_{nj})$ for all n, j, or $w_{nj} \equiv 1$ for all n, j. Moreover, we set

$$w := \sup\left\{1 + \beta_i(\xi_0) : \xi_0 \in \overline{\Theta'},\ i = 1, \ldots, d\right\} < \infty,$$

which yields $w_{nj} \le w$ for all n, j and for either choice of w_{nj}. Note that this inequality does not depend on the choice of the point ξ_0 and thus not on the choice of $\xi \in \mathscr{F}_{\mathrm{H}}$ either. Furthermore, we observe that for arbitrary stochastically independent, $\mathbb{R}^d$-valued random variables $X_1, \dots, X_m$ with finite second moments (in vector notation written as $X_j = (X_{j,1}, \dots, X_{j,d})^\top$) the following (in)equalities hold:

$$\begin{aligned}
\mathrm{E}\Big\| \sum_{j=1}^m X_j \Big\|^2 &= \sum_{k=1}^d \mathrm{E}\Big(\sum_{j=1}^m X_{j,k} \Big)^2 \\
&= \sum_{k=1}^d \bigg\{ \operatorname{Var} \sum_{j=1}^m X_{j,k} + \Big(\mathrm{E} \sum_{j=1}^m X_{j,k} \Big)^2 \bigg\} \\
&= \sum_{k=1}^d \bigg\{ \sum_{j=1}^m \operatorname{Var} X_{j,k} + \Big(\sum_{j=1}^m \mathrm{E}\, X_{j,k} \Big)^2 \bigg\} \\
&\le \sum_{k=1}^d \bigg\{ \sum_{j=1}^m \mathrm{E}\, X_{j,k}^2 + \Big(\sum_{j=1}^m \mathrm{E}\, X_{j,k} \Big)^2 \bigg\}. \qquad (7.32)
\end{aligned}$$

Together with the above definition of w (7.32) yields

$$\begin{aligned}
&\mathrm{E}\Big\| \sum_{j \in \mathbb{H}_n} a_n^2 w_{nj} \dot{\ell}_{\xi_0}(Y_{nj}) \Big\|^2 \\
&\le \sum_{k=1}^d \bigg\{ \sum_{j \in \mathbb{H}_n} a_n^4 w_{nj}^2 \int \dot{\ell}_{\xi_0,k}^2 \, dP_{\xi(x_{nj})} \\
&\qquad\qquad + \Big(\sum_{j \in \mathbb{H}_n} a_n^2 w_{nj} \int \dot{\ell}_{\xi_0,k} \, dP_{\xi(x_{nj})} \Big)^2 \bigg\} \\
&\le \sum_{k=1}^d \bigg\{ k_n a_n^4 w^2 \sup_{j \in \mathbb{H}_n} \int \dot{\ell}_{\xi_0,k}^2 \, dP_{\xi(x_{nj})} \\
&\qquad\qquad + \Big(k_n a_n^2 w \sup_{j \in \mathbb{H}_n} \int \dot{\ell}_{\xi_0,k} \, dP_{\xi(x_{nj})} \Big)^2 \bigg\}. \qquad (7.33)
\end{aligned}$$

In the above inequalities we further used that $\#\mathbb{H}_n(\xi_0) \le k_n$ (for each ξ_0). Moreover, from $\mathbb{H}_n \subseteq \mathbb{K}_n$ it follows that $|x_{nj} - x_0| \le K/b_n$ for all indices $j \in \mathbb{H}_n$. According to the smoothness assumptions from Section 3.2, the maps $(\theta, \tau) \mapsto \int \dot{\ell}_{\theta,k}^l \, dP_\tau$ (for $l = 1, 2$) are continuously partially differentiable on Θ and thus—by virtue of the mean value theorem—Lipschitz continuous on the compact and convex set $\overline{\Theta'}$. Since $\xi(x_0) = \xi_0 \in \Theta'$ and $\xi(x_{nj}) \in \Theta'$ (which follows from the definition of Θ') this implies for the case $l = 1$ and all $k = 1, \ldots, d$ that

$$\begin{aligned}\sup_{j\in\mathbb{H}_n}\left|\int \dot\ell_{\xi_0,k}\, dP_{\xi(x_{nj})}\right| &= \sup_{j\in\mathbb{H}_n}\left|\int \dot\ell_{\xi_0,k}\, dP_{\xi(x_{nj})} - \underbrace{\int \dot\ell_{\xi_0,k}\, dP_{\xi_0}}_{=0}\right| \\ &\lesssim \sup_{j\in\mathbb{H}_n}\|\xi(x_{nj}) - \xi_0\| \\ &= \sup_{j\in\mathbb{H}_n}\|\xi(x_{nj}) - \xi(x_0)\|.\end{aligned}$$

As a consequence of the above considerations we have $\sup_{j\in\mathbb{H}_n}|x_{nj} - x_0| \le K/b_n$, and in combination with the growth condition in (3.6) and the definition of a_n in (3.14) it follows that the last expression is of the order $O(a_n)$.

As a further consequence of the differentiability of the maps $(\theta,\tau) \mapsto \int \dot\ell^l_{\theta,k}\, dP_\tau$ we can conclude for the case $l = 2$ that $\int \dot\ell^2_{\theta,k}\, dP_\tau$ is bounded for all $\theta,\tau \in \overline{\Theta'}$, and thus

$$\sup_{j\in\mathbb{H}_n}\int \dot\ell^2_{\xi_0,k}\, dP_{\xi(x_{nj})} \lesssim 1$$

for all $k = 1, \ldots, d$. In addition, one can conclude from the smoothness assumptions in Section 3.2 that the upper bounds just derived do not depend on the choice of ξ and $\xi_0 = \xi(x_0)$, respectively. Consequently, if instead of a single function ξ we consider the whole class $\mathscr{F}_{\mathrm{H}}$, we get

$$\begin{aligned}&\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\sup_{j\in\mathbb{H}_n}\left|\int \dot\ell_{\xi_0,k}\, dP_{\xi(x_{nj})}\right| \lesssim \sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\sup_{j\in\mathbb{H}_n}\|\xi(x_{nj}) - \xi(x_0)\| \lesssim a_n, \\ &\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\sup_{j\in\mathbb{H}_n}\int \dot\ell^2_{\xi_0,k}\, dP_{\xi(x_{nj})} < \infty. \qquad (7.34)\end{aligned}$$

According to (4.8) we have $a_n^2 \simeq 2K/k_n$, and thus in particular $k_n a_n^m = o(1)$, for $m > 2$. From the above calculations we conclude that

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\int\Big\|\sum_{j\in\mathbb{H}_n} a_n^2 w_{nj}\dot{\ell}_{\xi_0}(Y_{nj})\Big\|^2\,dP_\xi^{(n)}$$

$$\overset{(7.33)}{\leq}\sum_{k=1}^{d}\Bigg\{k_n a_n^4 w^2\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\sup_{j\in\mathbb{H}_n}\int\dot{\ell}_{\xi_0,k}^2\,dP_{\xi(x_{nj})} + \Big(k_n a_n^2 w\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\sup_{j\in\mathbb{H}_n}\int\dot{\ell}_{\xi_0,k}\,dP_{\xi(x_{nj})}\Big)^2\Bigg\}$$

$$\overset{(7.34)}{\lesssim} k_n a_n^4+(k_n a_n^3)^2=o(1). \tag{7.35}$$

Of course, the "$\lesssim$" symbol in (7.29) can be replaced by "$\leq$" by multiplying the right hand side of that inequality with some positive constant, say N. Let now some arbitrary $\varepsilon>0$ be given. Putting together the above results, we have shown that

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}P_\xi^{(n)}\Bigg\{\sup_{\|z-\xi_0\|\leq a_nC}\Big\|\sum_{j\in\mathbb{H}_n}a_n\Big(\mathcal{C}_z^{-1}S_zB_z^{nj}(\mathcal{C}_z^{-1})^\top - \mathcal{C}_{\xi_0}^{-1}S_{\xi_0}B_{\xi_0}^{nj}(\mathcal{C}_{\xi_0}^{-1})^\top\Big)\dot{\ell}_{\xi_0}(Y_{nj})\Big\|>\varepsilon\Bigg\}$$

$$\overset{(7.29)}{\leq}\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}P_\xi^{(n)}\Bigg\{N\Big\|\sum_{j\in\mathbb{H}_n}a_n^2\dot{\ell}_{\xi_0}(Y_{nj})\Big\| + N\sum_{i=1}^{d}\Big\|\sum_{j\in\mathbb{H}_n}a_n^2 g_{\beta_i(\xi_0)}(\tilde{x}_{nj})\dot{\ell}_{\xi_0}(Y_{nj})\Big\|>\varepsilon\Bigg\}$$

$$\leq \sup_{\xi\in\mathscr{F}_{\mathrm{H}}}P_\xi^{(n)}\Bigg\{\Big\|\sum_{j\in\mathbb{H}_n}a_n^2\dot{\ell}_{\xi_0}(Y_{nj})\Big\|>\frac{\varepsilon}{N2}\Bigg\} + \sup_{\xi\in\mathscr{F}_{\mathrm{H}}}P_\xi^{(n)}\Bigg\{\sum_{i=1}^{d}\Big\|\sum_{j\in\mathbb{H}_n}a_n^2 g_{\beta_i(\xi_0)}(\tilde{x}_{nj})\dot{\ell}_{\xi_0}(Y_{nj})\Big\|>\frac{\varepsilon}{N2}\Bigg\}$$

$$\overset{(7.30),(7.31)}{\lesssim}\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\int\Big\|\sum_{j\in\mathbb{H}_n}a_n^2\dot{\ell}_{\xi_0}(Y_{nj})\Big\|^2\,dP_\xi^{(n)} + \sum_{i=1}^{d}\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\int\Big\|\sum_{j\in\mathbb{H}_n}a_n^2 g_{\beta_i(\xi_0)}(\tilde{x}_{nj})\dot{\ell}_{\xi_0}(Y_{nj})\Big\|^2\,dP_\xi^{(n)}$$

$$\overset{(7.35)}{\to} 0.$$

To sum up, we have shown that

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_{\xi}^{(n)}\Bigg\{ \sup_{\|z-\xi_0\|\le a_n C} \Bigg\| \sum_{j\in\mathbb{H}_n} a_n \Big(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top\Big)\, \dot{\ell}_{\xi_0}(Y_{nj})\Bigg\| > \varepsilon \Bigg\} \to 0 \tag{7.36}$$

for all $\varepsilon > 0$. In other words, the first of the two expressions in (7.20) converges uniformly in probability to zero. In the next step we show that the same holds true for the second expression in (7.20).

Assessment of the second expression in (7.20). With similar, but much more simple arguments it can be shown that the second expression on the right hand side of (7.20) converges uniformly in probability to zero, too. To this end, we first note that the following inequalities hold:

$$\begin{aligned}
&\sup_{\|z-\xi_0\|\le a_n C} \Bigg\| \sum_{j\in\mathbb{M}_n} a_n \Big(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top\Big) \dot{\ell}_{\xi_0}(Y_{nj})\Bigg\| \\
\le\;& \sup_z \Bigg\| \sum_{j\in\mathbb{M}_n} a_n \mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top \dot{\ell}_{\xi_0}(Y_{nj})\Bigg\| \\
&\quad + \Bigg\| \sum_{j\in\mathbb{M}_n} a_n \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_z^{-1})^\top \dot{\ell}_{\xi_0}(Y_{nj})\Bigg\| \\
\le\;& 2\sup_z \Bigg\| \sum_{j\in\mathbb{M}_n} a_n \mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top \dot{\ell}_{\xi_0}(Y_{nj})\Bigg\| \\
\le\;& 2\sum_{j\in\mathbb{M}_n} a_n \sup_z \Big\| \mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top \dot{\ell}_{\xi_0}(Y_{nj})\Big\| \\
\overset{(7.13)}{\le}\;& 2\sum_{j\in\mathbb{M}_n} a_n \sup_z \Big(|\!|\!| \mathcal{C}_z^{-1} S_z |\!|\!| \quad |\!|\!| B_z^{nj} |\!|\!| \quad |\!|\!| (\mathcal{C}_z^{-1})^\top |\!|\!| \Big) \, \|\dot{\ell}_{\xi_0}(Y_{nj})\| \\
\le\;& 2\sup_z \Big(|\!|\!| \mathcal{C}_z^{-1} S_z |\!|\!| \quad |\!|\!| (\mathcal{C}_z^{-1})^\top |\!|\!| \Big) \sum_{j\in\mathbb{M}_n} a_n \sup_z |\!|\!| B_z^{nj} |\!|\!| \quad \|\dot{\ell}_{\xi_0}(Y_{nj})\| \\
\lesssim\;& \sum_{j\in\mathbb{M}_n} a_n \sup_z |\!|\!| B_z^{nj} |\!|\!| \quad \|\dot{\ell}_{\xi_0}(Y_{nj})\|.
\end{aligned} \tag{7.37}$$

Note that in the last step we exploited the usual continuity argument, according to which

$$\sup_z \left(\| \mathcal{C}_z^{-1} S_z \| \quad \| (\mathcal{C}_z^{-1})^\top \| \right)$$

can be bounded by a constant that does not depend on the parameter $\xi_0 = \xi(x_0)$.

Next we investigate the expressions $\|B_z^{nj}\|$. By the notions on matrix norms from Section 7.2.1 we know that all norms on $\mathbb{R}^{d\times d}$ are equivalent. In particular, $\|A\| \lesssim \|A\|_{\mathrm{F}}$ (the Frobenius norm) for every matrix A. Hence,

$$\begin{aligned}
\|B_z^{nj}\| &\lesssim \|B_z^{nj}\|_{\mathrm{F}} \\
&\overset{(7.4)}{=} \|(g_{\beta_1(z)}(\tilde{x}_{nj}), \ldots, g_{\beta_d(z)}(\tilde{x}_{nj}))^\top\| \\
&\lesssim \|(g_{\beta_1(z)}(\tilde{x}_{nj}), \ldots, g_{\beta_d(z)}(\tilde{x}_{nj}))^\top\|_1 \\
&= \sum_{i=1}^{d} |g_{\beta_i(z)}(\tilde{x}_{nj})|,
\end{aligned}$$

where we used that $\|\cdot\|$ and $\|\cdot\|_1$ are equivalent norms. Due to the definition of the index set $\mathbb{M}_n$ we have $\beta_n^{\min} < |\tilde{x}_{nj}|^\rho \le \beta_n^{\max}$ for all $j \in \mathbb{M}_n$, and thus

$$\begin{aligned}
&\sup_{\|z-\xi_0\|\le a_n C} \left| g_{\beta_i(z)}(\tilde{x}_{nj}) \right| \\
&\quad \le \sup \left\{ (\beta_i - |\tilde{x}_{nj}|^\rho)^+ : \beta_i \in (\beta_n^{\min}(\xi_0), \beta_n^{\max}(\xi_0)] \right\} \\
&\quad \le \beta_n^{\max}(\xi_0) - \beta_n^{\min}(\xi_0)
\end{aligned}$$

for all $j \in \mathbb{M}_n$. By virtue of (7.17) this last expression can be bounded by a term of the order $O(a_n)$, which does not depend on the parameter $\xi_0 = \xi(x_0)$. From these calculations it follows

$$\begin{aligned}
\sup_{\|z-\xi_0\|\le a_n C} \|B_z^{nj}\| &\lesssim \sum_{i=1}^{d} \sup_z |g_{\beta_i(z)}(\tilde{x}_{nj})| \\
&\lesssim \beta_n^{\max}(\xi_0) - \beta_n^{\min}(\xi_0) \qquad (7.38)
\end{aligned}$$

for all $j \in \mathbb{M}_n$. We now combine the above results. Multiplying the right hand side of (7.37) with a suitable constant, say $N > 0$, we may of course replace there "$\lesssim$" with "$\le$". Applying Markov's inequality, we conclude that for all $\varepsilon > 0$

$$\begin{aligned}
&\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Big\{ \sup_{\|z-\xi_0\|\le a_n C} \Big\| \sum_{j\in\mathbb{M}_n} a_n \left(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top \right.\\
&\qquad\qquad \left. -\mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top\right) \dot\ell_{\xi_0}(Y_{nj}) \Big\| > \varepsilon \Big\}\\
&\overset{(7.37)}{\le} \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Big\{ \sum_{j\in\mathbb{M}_n} a_n \sup_{\|z-\xi_0\|\le a_n C} \|\!|B_z^{nj}\|\!| \quad \|\dot\ell_{\xi_0}(Y_{nj})\| > \frac{\varepsilon}{N}\Big\}\\
&\lesssim \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} \int \Big(\sum_{j\in\mathbb{M}_n} a_n \sup_{\|z-\xi_0\|\le a_n C} \|\!|B_z^{nj}\|\!| \quad \|\dot\ell_{\xi_0}(Y_{nj})\|\Big)\, dP_\xi^{(n)}\\
&\overset{(7.38)}{\lesssim} a_n \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} \Big(\big(\beta_n^{\max}(\xi_0) - \beta_n^{\min}(\xi_0)\big) \sum_{j\in\mathbb{M}_n} \int \|\dot\ell_{\xi_0}\|\, dP_{\xi(x_{nj})}\Big)\\
&\le a_n \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} \Big(\#\mathbb{M}_n(\xi_0)\ \big(\beta_n^{\max}(\xi_0) - \beta_n^{\min}(\xi_0)\big)\\
&\qquad\qquad \sup_{j\in\mathbb{M}_n} \int \|\dot\ell_{\xi_0}\|\, dP_{\xi(x_{nj})}\Big)\\
&\le a_n \sup_{\xi_0\in\overline{\Theta^*}} \big(\beta_n^{\max}(\xi_0) - \beta_n^{\min}(\xi_0)\big) \sup_{\xi_0\in\overline{\Theta^*}} \#\mathbb{M}_n(\xi_0)\\
&\qquad\qquad \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} \sup_{j\in\mathbb{M}_n} \int \|\dot\ell_{\xi_0}\|\, dP_{\xi(x_{nj})}.
\end{aligned}\tag{7.39}$$

By virtue of (7.17) we have

$$\sup_{\xi_0\in\overline{\Theta^*}} \big(\beta_n^{\max}(\xi_0) - \beta_n^{\min}(\xi_0)\big) \to 0,$$

and (7.18) in combination with (3.15) implies

$$a_n \sup_{\xi_0\in\overline{\Theta^*}} \#\mathbb{M}_n(\xi_0) \lesssim \frac{a_n^2 n}{b_n} = 1.$$

Moreover, with the same arguments that lead to (7.34) we conclude that

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} \sup_{j\in\mathbb{M}_n} \int \|\dot\ell_{\xi_0}\|\, dP_{\xi(x_{nj})} < \infty.$$

Altogether, these considerations show that the expression in (7.39) converges to zero, and thus

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_{\xi}^{(n)}\Big\{\sup_{\|z-\xi_0\|\le a_nC}\Big\|\sum_{j\in\mathbb{M}_n} a_n\left(\mathcal{C}_z^{-1}S_zB_z^{nj}(\mathcal{C}_z^{-1})^\top\right.$$
$$\left.-\mathcal{C}_{\xi_0}^{-1}S_{\xi_0}B_{\xi_0}^{nj}(\mathcal{C}_{\xi_0}^{-1})^\top\right)\dot{\ell}_{\xi_0}(Y_{nj})\Big\| > \varepsilon\Big\} \to 0, \tag{7.40}$$

i.e. we have proved that the second expression in (7.20) converges uniformly in probability to zero, too.

Completion of the proof. To sum up, a combination of (7.20), (7.36) and (7.40) yields

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_{\xi}^{(n)}\Big\{\sup_{\|\xi_0-z\|\le a_nC}\|R_1(\xi_0,z)\| > \varepsilon\Big\}$$
$$\le \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_{\xi}^{(n)}\Big\{\sup_{\|z-\xi_0\|\le a_nC}\Big\|\sum_{j\in\mathbb{H}_n} a_n\left(\mathcal{C}_z^{-1}S_zB_z^{nj}(\mathcal{C}_z^{-1})^\top\right.$$
$$\left.-\mathcal{C}_{\xi_0}^{-1}S_{\xi_0}B_{\xi_0}^{nj}(\mathcal{C}_{\xi_0}^{-1})^\top\right)\dot{\ell}_{\xi_0}(Y_{nj})\Big\| > \frac{\varepsilon}{2}\Big\}$$
$$+ \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_{\xi}^{(n)}\Big\{\sup_{\|z-\xi_0\|\le a_nC}\Big\|\sum_{j\in\mathbb{M}_n} a_n\left(\mathcal{C}_z^{-1}S_zB_z^{nj}(\mathcal{C}_z^{-1})^\top\right.$$
$$\left.-\mathcal{C}_{\xi_0}^{-1}S_{\xi_0}B_{\xi_0}^{nj}(\mathcal{C}_{\xi_0}^{-1})^\top\right)\dot{\ell}_{\xi_0}(Y_{nj})\Big\| > \frac{\varepsilon}{2}\Big\} \to 0,$$

and thus (7.15) holds. With the preliminary considerations from Section 7.2.1 this proves the assertion of Lemma 7.2. ■

7.2.3 Proof of Lemma 7.4

In order to prove the lemma, we show that (7.14) holds for $s = 3$, i.e. for every fixed $C > 0$ and $\varepsilon > 0$, and with $\xi_0 = \xi(x_0)$

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_{\xi}^{(n)}\Big\{\sup_{\|z-\xi_0\|\le a_nC}\|R_3(\xi_0,z)\| > \varepsilon\Big\} \to 0. \tag{7.41}$$

According to the definition in (7.9) (with ξ_n^* replaced by z) we have

$$R_3(\xi_0,z) = \left(\mathbb{1} + \sum_{j=1}^{n} a_n^2\left(\mathcal{C}_z^{-1}S_zB_z^{nj}(\mathcal{C}_z^{-1})^\top\right)\ddot{\ell}_{\xi_0}(Y_{nj})\right)\frac{1}{a_n}(z-\xi_0).$$

Since $B_z^{nj} = 0$ for $j \notin \mathbb{K}_n$, it is of course sufficient to build the sum with respect to the index set $\mathbb{K}_n$. The multiplicative inequality for the operator norm yields

$$\begin{aligned}
&\sup_{\|z-\xi_0\|\le a_n C} \|R_3(\xi_0, z)\| \\
&\overset{(7.13)}{\le} \sup_z \left\| \mathbb{1} + \sum_{j\in\mathbb{K}_n} a_n^2 \left(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top\right) \ddot{\ell}_{\xi_0}(Y_{nj}) \right\| \sup_z \frac{1}{a_n} \|z - \xi_0\| \\
&\le \sup_z \left\| \mathbb{1} + \sum_{j\in\mathbb{K}_n} a_n^2 \left(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top\right) \ddot{\ell}_{\xi_0}(Y_{nj}) \right\| C. \qquad (7.42)
\end{aligned}$$

In order to shorten notation, we set

$$T_z^{nj} := \mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top, \qquad (7.43)$$

likewise we define $T_{\xi_0}^{nj}$. With this notation and an application of the triangle inequality, we get an upper bound for the supremum in (7.42) by

$$\begin{aligned}
&\sup_{\|z-\xi_0\|\le a_n C} \left\| \mathbb{1} + \sum_{j\in\mathbb{K}_n} a_n^2 \left(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top\right) \ddot{\ell}_{\xi_0}(Y_{nj}) \right\| \\
&= \left\| \mathbb{1} + \sum_{j\in\mathbb{K}_n} a_n^2 T_z^{nj} \ddot{\ell}_{\xi_0}(Y_{nj}) \right\| \\
&\le \left\| \mathbb{1} - \sum_{j\in\mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \mathcal{I}_{\xi_0} \right\| + \sup_z \left\| \sum_{j\in\mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \mathcal{I}_{\xi_0} - \sum_{j\in\mathbb{K}_n} a_n^2 T_z^{nj} \mathcal{I}_{\xi_0} \right\| \\
&\quad + \sup_z \left\| \sum_{j\in\mathbb{K}_n} a_n^2 T_z^{nj} \mathcal{I}_{\xi_0} + \sum_{j\in\mathbb{K}_n} a_n^2 T_z^{nj} \ddot{\ell}_{\xi_0}(Y_{nj}) \right\| \\
&= \left\| \mathbb{1} - \sum_{j\in\mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \mathcal{I}_{\xi_0} \right\| + \sup_z \left\| \sum_{j\in\mathbb{K}_n} a_n^2 \left(T_{\xi_0}^{nj} - T_z^{nj}\right) \mathcal{I}_{\xi_0} \right\| \\
&\quad + \sup_z \left\| \sum_{j\in\mathbb{K}_n} a_n^2 T_z^{nj} \left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\right) \right\| \qquad (7.44)
\end{aligned}$$

Note that only the third expression has a stochastic component, whereas the first and the second one are purely deterministic. We will show that each of these three expressions converges uniformly (in probability) to zero, for ξ ranging over $\mathscr{F}_{\mathrm{H}}$.

Assessment of the first expression in (7.44). In order to prove that

$$\left\lVert\!\left\lvert \mathbb{1} - \sum_{j\in\mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \mathcal{I}_{\xi_0} \right\rvert\!\right\rVert \to 0, \tag{7.45}$$

it is sufficient to show that these matrices converge component-wise (see the remarks in Section 7.2.1). Since $\mathcal{I}_{\xi_0} = \mathcal{C}_{\xi_0}^\top \mathcal{C}_{\xi_0}$, we have

$$\begin{aligned}
\sum_{j\in\mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \mathcal{I}_{\xi_0} &= \sum_{j\in\mathbb{K}_n} a_n^2 \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \mathcal{I}_{\xi_0} \\
&= \sum_{j\in\mathbb{K}_n} a_n^2 \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} \mathcal{C}_{\xi_0} \\
&= \mathcal{C}_{\xi_0}^{-1} \Bigg(\sum_{j\in\mathbb{K}_n} a_n^2 S_{\xi_0} B_{\xi_0}^{nj} \Bigg) \mathcal{C}_{\xi_0}.
\end{aligned}$$

The matrix in parentheses is diagonal, and because of (7.4) the ith entry on the main diagonal is given by

$$\begin{aligned}
\sum_{j\in\mathbb{K}_n} S_i(\xi_0) B_{\xi_0}^{nj} &\overset{(7.4)}{=} S_i(\xi_0) \sum_{j\in\mathbb{K}_n} a_n^2 g_{\beta_i(\xi_0)}(\tilde{x}_{nj}) \\
&\to S_i(\xi_0) \int g_{\beta_i(\xi_0)}(s)\, ds \\
&\overset{(6.12)}{=} 1,
\end{aligned}$$

where the stated convergence follows with the same arguments that were used in the calculations from the proof of Proposition 4.3. Consequently,

$$\sum_{j\in\mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \mathcal{I}_{\xi_0} \;\to\; \mathcal{C}_{\xi_0}^{-1}\, \mathbb{1}\, \mathcal{C}_{\xi_0} \;=\; \mathbb{1},$$

which implies (7.45). This convergence is even uniform for $\xi_0 \in \Theta^*$, i.e.

$$\sup_{\xi_0\in\Theta^*} \left\lVert\!\left\lvert \mathbb{1} - \sum_{j\in\mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \mathcal{I}_{\xi_0} \right\rvert\!\right\rVert \to 0. \tag{7.46}$$

For a proof, let us again take a look at the maps $\beta_i(\cdot)$: According to Lemma 6.6 these are continuous and thus $\{\beta_i(\gamma) : \gamma \in \overline{\Theta'}\}$ is a compact interval, say $[a, b]$. (Note that by definition of the constant

K in Section 3.5.2 we have $\beta_i(\gamma) \leq K^\rho$ for all $\gamma \in \Theta'$, and hence $b \leq K^\rho$.) Let us now consider the sequence of maps

$$f_n\colon [a,b] \to \mathbb{R}, \quad \zeta \mapsto \sum_{j \in \mathbb{K}_n} a_n^2 g_\zeta(\tilde{x}_{nj}). \tag{7.47}$$

Each f_n is continuous, strictly increasing and uniformly bounded. A universal upper bound N such that $f_n(\zeta) \leq N$ for all ζ and all n is for example given by $N := \sup_n f_n(b)$. Note that this N is finite, because $f_n(b) \leq k_n a_n^2 b = O(1)$ by (4.8). For fixed $\varepsilon > 0$ let now $\iota := \varepsilon/L$, where

$$L := \sup_n a_n^2 k_n < \infty,$$

again by (4.8). For all points $\zeta_1, \zeta_2 \in [a,b]$ satisfying $|\zeta_1 - \zeta_2| < \iota$ we then have

$$\begin{aligned} |f_n(\zeta_1) - f_n(\zeta_2)| &= \left| \sum_{j \in \mathbb{K}_n} a_n^2 (g_{\zeta_1}(\tilde{x}_{nj}) - g_{\zeta_2}(\tilde{x}_{nj})) \right| \\ &\leq a_n^2 k_n \|g_{\zeta_1} - g_{\zeta_2}\|_u \\ &\leq L|\zeta_1 - \zeta_2| < \varepsilon \end{aligned}$$

for all $n \in \mathbb{N}$. Hence, we have shown that the sequence f_n is equicontinuous, and thus converges in $C([a,b],\mathbb{R})$, which is a consequence of the Arzelà-Ascoli theorem (cf. Heuser, 2000, Satz 106.2). Consequently, $f_n(\zeta) \to \int g_\zeta(s)\, ds$ uniformly in ζ, and therefore

$$\sum_{j=1}^n a_n^2 B_{\xi_0}^{nj} \overset{\substack{(7.4)\\(7.47)}}{=} \operatorname{diag}\Big(f_n(\beta_1(\xi_0)), \ldots, f_n(\beta_d(\xi_0))\Big) \to \int g_{\beta(\xi_0)}(s)\, ds$$

uniformly on $\overline{\Theta'}$ (and thus in particular on $\overline{\Theta^*} \subseteq \overline{\Theta'}$). From the Lipschitz properties of the maps $\gamma \mapsto \mathcal{C}_\gamma$ and $\gamma \mapsto S(\gamma)$ (see Propositions 7.6 and 7.7) we further conclude that

$$\sum_{j \in \mathbb{K}_n} a_n^2 \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} \mathcal{C}_{\xi_0} \to \mathbb{1},$$

uniformly on $\overline{\Theta^*}$, and thus (7.46) holds.

Assessment of the second expression in (7.44). Obviously,

$$
\begin{aligned}
&\sup_{\xi_0\in\Theta^*}\sup_{\|z-\xi_0\|\le a_nC}\left\|\sum_{j\in\mathbb{K}_n} a_n^2\left(T^{nj}_{\xi_0}-T^{nj}_z\right)\mathcal{I}_{\xi_0}\right\| \\
&\quad\le \sup_{\xi_0\in\Theta^*}\left\|\sum_{j\in\mathbb{K}_n} a_n^2T^{nj}_{\xi_0}\mathcal{I}_{\xi_0}-\mathbb{1}\right\| \\
&\qquad+\sup_{\xi_0\in\Theta^*}\sup_{\|z-\xi_0\|\le a_nC}\left\|\mathbb{1}-\sum_{j\in\mathbb{K}_n} a_n^2T^{nj}_z\mathcal{I}_{\xi_0}\right\| \\
&\quad\le \sup_{\xi_0\in\Theta^*}\left\|\sum_{j\in\mathbb{K}_n} a_n^2T^{nj}_{\xi_0}\mathcal{I}_{\xi_0}-\mathbb{1}\right\| \\
&\qquad+\sup_{\xi_0\in\Theta^*}\sup_{\|z-\xi_0\|\le a_nC}\left\|\mathbb{1}-\sum_{j\in\mathbb{K}_n} a_n^2T^{nj}_z\mathcal{I}_z\right\| \\
&\qquad+\sup_{\xi_0\in\Theta^*}\sup_{\|z-\xi_0\|\le a_nC}\left\|\sum_{j\in\mathbb{K}_n} a_n^2T^{nj}_z\mathcal{I}_z-\sum_{j\in\mathbb{K}_n} a_n^2T^{nj}_z\mathcal{I}_{\xi_0}\right\|. \qquad (7.48)
\end{aligned}
$$

As was just shown in the proof of (7.46), the first of the three suprema in the preceding equation converges to zero. Moreover, for a given $\xi_0\in\Theta^*$ all points z that satisfy $\|z-\xi_0\|\le a_nC$ lie in $\overline{\Theta'}$, provided that n is sufficiently large. We conclude from the proof of (7.46) that for large n

$$
\begin{aligned}
&\sup_{\xi_0\in\Theta^*}\sup_{\|z-\xi_0\|\le a_nC}\left\|\mathbb{1}-\sum_{j\in\mathbb{K}_n} a_n^2T^{nj}_z\mathcal{I}_z\right\| \\
&\qquad\qquad\le \sup_{\xi_0\in\Theta'}\left\|\mathbb{1}-\sum_{j\in\mathbb{K}_n} a_n^2T^{nj}_{\xi_0}\mathcal{I}_{\xi_0}\right\| \;\to\; 0.
\end{aligned}
$$

It remains to investigate the third expression from (7.48). For this we have

$$
\begin{aligned}
&\sup_{\xi_0\in\Theta^*}\sup_{\|z-\xi_0\|\le a_nC}\left\|\sum_{j\in\mathbb{K}_n} a_n^2T^{nj}_z\mathcal{I}_z-\sum_{j\in\mathbb{K}_n} a_n^2T^{nj}_z\mathcal{I}_{\xi_0}\right\| \\
&= \sup_{\xi_0\in\Theta^*}\sup_{\|z-\xi_0\|\le a_nC}\left\|\sum_{j\in\mathbb{K}_n} a_n^2T^{nj}_z\left(\mathcal{I}_z-\mathcal{I}_{\xi_0}\right)\right\| \\
&\overset{(7.13)}{\le} \sup_{\xi_0\in\Theta^*}\sup_{\|z-\xi_0\|\le a_nC}\left(\sum_{j\in\mathbb{K}_n} a_n^2\|T^{nj}_z\| \quad \|\mathcal{I}_z-\mathcal{I}_{\xi_0}\|\right)\dots
\end{aligned}
$$

$$\leq \sup_{\xi_0\in\Theta^*} \sup_{\|z-\xi_0\|\leq a_n C} \left(\sum_{j\in\mathbb{K}_n} a_n^2 \|T_z^{nj}\|\right) \sup_{\xi_0\in\Theta^*} \sup_{\|z-\xi_0\|\leq a_n C} \|\mathcal{I}_z - \mathcal{I}_{\xi_0}\|.$$

The definition of T_z^{nj} in (7.43) and the fact that all the components on the right hand side of (7.43) are bounded implies that $\sup_z \|T_z^{nj}\|$ is bounded by some constant, which does not depend on n, j and ξ_0. Hence, the first of the suprema in the preceding display is of the order $O(1)$. With Lipschitz arguments, that in similar form have been used in the above proof of Lemma 7.2, one further argues that the second of these suprema is of the order $O(a_n)$. Hence,

$$\sup_{\xi_0\in\Theta^*} \sup_{\|z-\xi_0\|\leq a_n C} \left\| \sum_{j\in\mathbb{K}_n} a_n^2 T_z^{nj} \mathcal{I}_z - \sum_{j\in\mathbb{K}_n} a_n^2 T_z^{nj} \mathcal{I}_{\xi_0} \right\| \to 0.$$

To sum up, we have shown that all three suprema in (7.48) converge to zero, and thus

$$\sup_{\xi_0\in\Theta^*} \sup_{\|z-\xi_0\|\leq a_n C} \left\| \sum_{j\in\mathbb{K}_n} a_n^2 \left(T_{\xi_0}^{nj} - T_z^{nj}\right) \mathcal{I}_{\xi_0} \right\| \to 0, \tag{7.49}$$

i.e. the second expression in (7.44) converges to zero.

Assessment of the third expression in (7.44). It remains to prove that the third summand in (7.44) converges uniformly in probability to zero. The triangle inequality in combination with (7.13) yields

$$\begin{aligned}
&\left\| \sum_{j\in\mathbb{K}_n} a_n^2 T_z^{nj} \left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\right) \right\| \\
&\qquad \leq \left\| \sum_{j\in\mathbb{K}_n} a_n^2 \left(T_z^{nj} - T_{\xi_0}^{nj}\right) \left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\right) \right\| \\
&\qquad\quad + \left\| \sum_{j\in\mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\right) \right\| \\
&\qquad \leq \sum_{j\in\mathbb{K}_n} a_n^2 \|T_z^{nj} - T_{\xi_0}^{nj}\| \, \|\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\| \\
&\qquad\quad + \left\| \sum_{j\in\mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\right) \right\|.
\end{aligned} \tag{7.50}$$

In order to examine the first of these two expressions we first note that

$$
\begin{aligned}
\|T_z^{nj} - T_{\xi_0}^{nj}\| &\overset{(7.43)}{=} \|\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top\| \\
&\overset{(7.21)}{\leq} \left\| \left(\mathcal{C}_z^{-1} S_z - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0}\right) B_z^{nj} (\mathcal{C}_z^{-1})^\top \right\| \\
&\qquad + \left\| \mathcal{C}_{\xi_0}^{-1} S_{\xi_0} \left(B_z^{nj} (\mathcal{C}_z^{-1})^\top - B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top \right) \right\| \\
&\overset{(7.13)}{\leq} \|\mathcal{C}_z^{-1} S_z - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0}\| \, \|B_z^{nj} (\mathcal{C}_z^{-1})^\top\| \\
&\qquad + \|\mathcal{C}_{\xi_0}^{-1} S_{\xi_0}\| \, \|B_z^{nj} (\mathcal{C}_z^{-1})^\top - B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top\|.
\end{aligned}
$$

From (7.22) we know that

$$
\sup_{\|\xi_0 - z\| \leq a_n C} \|\mathcal{C}_z^{-1} S_z - \mathcal{C}_{\xi_0}^{-1} S_{\xi_0}\| \lesssim a_n,
$$

which even holds uniformly in ξ_0. From the definition of the matrices $B_\cdot^{nj}$ in (7.4) one further concludes (again, with a Lipschitz argument and Proposition 7.7) that

$$
\sup_{\|\xi_0 - z\| \leq a_n C} \|B_z^{nj} (\mathcal{C}_z^{-1})^\top - B_{\xi_0}^{nj} (\mathcal{C}_{\xi_0}^{-1})^\top\| \lesssim a_n,
$$

which holds for all j, and uniformly in ξ_0, too. Furthermore, $\|B_z^{nj} (\mathcal{C}_z^{-1})^\top\|$ and $\|\mathcal{C}_{\xi_0}^{-1} S_{\xi_0}\|$ are uniformly bounded, because the entries of these matrices are all continuous on Θ (see the results from Section 7.2.1). To sum up, we conclude that there is a constant D such that

$$
\sup_{\xi_0 \in \Theta^*} \sup_{\|z - \xi_0\| \leq a_n C} \|T_z^{nj} - T_{\xi_0}^{nj}\| \;\leq\; D a_n
$$

for all $j \in \mathbb{K}_n$. In combination with (7.50) this yields

$$
\begin{aligned}
&\sup_{\|z - \xi_0\| \leq a_n C} \left\| \sum_{j \in \mathbb{K}_n} a_n^2 T_z^{nj} \left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj}) \right) \right\| \\
&\leq \sum_{j \in \mathbb{K}_n} a_n^2 (D a_n) \|\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\| + \left\| \sum_{j \in \mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj}) \right) \right\|,
\end{aligned}
$$

and thus, for every $\varepsilon > 0$,

$$
\begin{aligned}
P_\xi^{(n)}&\Bigg\{ \sup_{\|z-\xi_0\|\le a_n C} \Bigg\| \sum_{j\in\mathbb{K}_n} a_n^2 T_z^{nj}\left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\right) \Bigg\| > \varepsilon \Bigg\} \\
&\le P_\xi^{(n)}\Bigg\{ \sum_{j\in\mathbb{K}_n} D a_n^3 \|\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\| \\
&\qquad\qquad + \Bigg\| \sum_{j\in\mathbb{K}_n} a_n^2 T_{\xi_0}^{nj}\left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\right) \Bigg\| > \varepsilon \Bigg\} \\
&\le P_\xi^{(n)}\Bigg\{ \sum_{j\in\mathbb{K}_n} D a_n^3 \|\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\| > \frac{\varepsilon}{2} \Bigg\} \\
&\qquad + P_\xi^{(n)}\Bigg\{ \Bigg\| \sum_{j\in\mathbb{K}_n} a_n^2 T_{\xi_0}^{nj}\left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\right) \Bigg\| > \frac{\varepsilon}{2} \Bigg\}. \qquad (7.51)
\end{aligned}
$$

We show that both probabilities in (7.51) converge to zero. An application of Markov's inequality yields

$$
\begin{aligned}
P_\xi^{(n)}&\Bigg\{ \sum_{j\in\mathbb{K}_n} D a_n^3 \|\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\| > \frac{\varepsilon}{2} \Bigg\} \\
&\le \frac{2}{\varepsilon} \mathrm{E} \sum_{j\in\mathbb{K}_n} D a_n^3 \|\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj})\| \\
&\lesssim k_n a_n^3 \sup_{j\in\mathbb{K}_n} \left(\|\mathcal{I}_{\xi_0}\| + \mathrm{E}\, \|\ddot{\ell}_{\xi_0}(Y_{nj})\| \right) \\
&= k_n a_n^3 \sup_{j\in\mathbb{K}_n} \left(\|\mathcal{I}_{\xi_0}\| + \textstyle\int \|\ddot{\ell}_{\xi_0}\|\, dP_{\xi(x_{nj})} \right).
\end{aligned}
$$

Here and in the following the expectation is understood as expectation with respect to the product measure $P_\xi^{(n)}$. Since any two norms on $\mathbb{R}^{d\times d}$ are equivalent we have $\|A\| \lesssim \|A\|_{\mathrm{F}}$ for every matrix $A \in \mathbb{R}^{d\times d}$. Hence, the definition of the Frobenius norm and Hölder's inequality yield

$$
\begin{aligned}
\int \|\ddot{\ell}_{\xi_0}\|\, dP_{\xi(x_{nj})} &\lesssim \int \|\ddot{\ell}_{\xi_0}\|_{\mathrm{F}}\, dP_{\xi(x_{nj})} \\
&\lesssim \left(\int \|\ddot{\ell}_{\xi_0}\|_{\mathrm{F}}^2\, dP_{\xi(x_{nj})} \right)^{1/2} \\
&\lesssim \left(\sum_{r,s=1}^{d} \int (\ddot{\ell}_{\xi_0,rs})^2\, dP_{\xi(x_{nj})} \right)^{1/2}.
\end{aligned}
$$

Recall that in Section 3.2 the maps

$$(\theta,\tau)\mapsto\int(\ddot{\ell}_{\theta,rs})^k\,dP_\tau,\quad k\in\{0,1,2\},\tag{7.52}$$

were presumed continuously partially differentiable on $\Theta\times\Theta$. As a consequence of the above arguments both $\|\mathcal{I}_{\xi_0}\|$ and $\sup_j\int\|\ddot{\ell}_{\xi_0}\|\,dP_{\xi(x_{nj})}$ are bounded, even uniformly in ξ_0, and thus

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\sup_{j\in\mathbb{K}_n}\Big(\|\mathcal{I}_{\xi_0}\|+\textstyle\int\|\ddot{\ell}_{\xi_0}\|\,dP_{\xi(x_{nj})}\Big)=O(1).\tag{7.53}$$

Since $k_na_n^2=O(1)$ (see (4.8)) this implies

$$k_na_n^3\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\sup_{j\in\mathbb{K}_n}\Big(\|\mathcal{I}_{\xi_0}\|+\textstyle\int\|\ddot{\ell}_{\xi_0}\|\,dP_{\xi(x_{nj})}\Big)\to0,$$

and therefore

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}P_\xi^{(n)}\Big\{\sum_{j\in\mathbb{K}_n}Da_n^3\|\mathcal{I}_{\xi_0}+\ddot{\ell}_{\xi_0}(Y_{nj})\|>\frac{\varepsilon}{2}\Big\}\to0.\tag{7.54}$$

It remains to show that the second probability in (7.51) converges to zero: Markov's inequality yields

$$\begin{aligned}P_\xi^{(n)}\Big\{\Big\|\sum_{j\in\mathbb{K}_n}a_n^2T_{\xi_0}^{nj}\Big(\mathcal{I}_{\xi_0}+\ddot{\ell}_{\xi_0}(Y_{nj})\Big)\Big\|>\frac{\varepsilon}{2}\Big\}&\\ \le\ \frac{4}{\varepsilon^2}\,\mathrm{E}\Big\|\sum_{j\in\mathbb{K}_n}a_n^2T_{\xi_0}^{nj}\Big(\mathcal{I}_{\xi_0}+\ddot{\ell}_{\xi_0}(Y_{nj})\Big)\Big\|^2&\\ \lesssim\ \mathrm{E}\Big\|\sum_{j\in\mathbb{K}_n}a_n^2T_{\xi_0}^{nj}\Big(\mathcal{I}_{\xi_0}+\ddot{\ell}_{\xi_0}(Y_{nj})\Big)\Big\|_{\mathrm{F}}^2.&\end{aligned}\tag{7.55}$$

Again, in the last step we exploited that $\|A\|\lesssim\|A\|_{\mathrm{F}}$ for every matrix $A\in\mathbb{R}^{d\times d}$, because any two norms on this space are equivalent. For the following calculations, we set

$$V_{nj}:=T_{\xi_0}^{nj}(\mathcal{I}_{\xi_0}+\ddot{\ell}_{\xi_0}(Y_{nj})),\tag{7.56}$$

and we write the rs-entries of these matrices as $V_{nj}(r,s)$. As we observed in Section 7.2.1, the Frobenius norm $\|\cdot\|_{\mathrm{F}}$ can be identified with the Euclidean norm on $\mathbb{R}^{d^2}$. Therefore,

$$
\begin{aligned}
\mathrm{E}\Big\| \sum_{j\in\mathbb{K}_n} a_n^2 T_{\xi_0}^{nj}\Big(\mathcal{I}_{\xi_0}+\ddot{\ell}_{\xi_0}(Y_{nj})\Big)\Big\|_{\mathrm{F}}^2 &= \mathrm{E}\Big\|\sum_{j\in\mathbb{K}_n} a_n^2 V_{nj}\Big\|_{\mathrm{F}}^2 \\
&\overset{(7.32)}{\leq} \sum_{r,s=1}^{d}\Bigg[\sum_{j\in\mathbb{K}_n} a_n^4\,\mathrm{E}\,V_{nj}(r,s)^2+\Big(\sum_{j\in\mathbb{K}_n} a_n^2\,\mathrm{E}\,V_{nj}(r,s)\Big)^2\Bigg] \\
&= \sum_{r,s=1}^{d}\Bigg[\sum_{j\in\mathbb{K}_n} a_n^4\int V_{nj}(r,s)^2\,dP_{\xi(x_{nj})} \\
&\qquad +\Big(\sum_{j\in\mathbb{K}_n} a_n^2\int V_{nj}(r,s)\,dP_{\xi(x_{nj})}\Big)^2\Bigg]. \qquad (7.57)
\end{aligned}
$$

We investigate the expectations in (7.57) and show that these converge to zero. For the first of the two expectations this proof is rather simple. Obviously,

$$
\begin{aligned}
V_{nj}(r,s)^2 &\leq \|V_{nj}\|_{\mathrm{F}}^2 \\
&\lesssim |||V_{nj}|||^2 \\
&\overset{(7.56)}{=} |||T_{\xi_0}^{nj}(\mathcal{I}_{\xi_0}+\ddot{\ell}_{\xi_0}(Y_{nj}))|||^2 \\
&\overset{(7.13)}{\leq} |||T_{\xi_0}^{nj}|||^2\,\big(|||\mathcal{I}_{\xi_0}|||+|||\ddot{\ell}_{\xi_0}(Y_{nj}))|||\big)^2 \\
&\leq |||T_{\xi_0}^{nj}|||^2\,2\big(|||\mathcal{I}_{\xi_0}|||^2+|||\ddot{\ell}_{\xi_0}(Y_{nj}))|||^2\big). \qquad (7.58)
\end{aligned}
$$

Note that

$$
\sup_{\xi_0\in\Theta^*}\sup_{j\in\mathbb{K}_n} |||T_{\xi_0}^{nj}|||^2<\infty, \qquad (7.59)
$$

which follows from the definition of $T_{\xi_0}^{nj}$ in (7.43) and the fact that all of the components involved in this definition are bounded. In addition, we conclude with the arguments leading to (7.53) that also

$$
\sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\sup_{j\in\mathbb{K}_n}\Big(|||\mathcal{I}_{\xi_0}|||^2+\textstyle\int|||\ddot{\ell}_{\xi_0}|||^2\,dP_{\xi(x_{nj})}\Big)=O(1).
$$

Moreover, we have $a_n^4 k_n \to 0$. These results yield

$$\begin{aligned}
&\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} \Big\{ \sum_{j\in\mathbb{K}_n} a_n^4 \int V_{nj}(r,s)^2 \, dP_{\xi(x_{nj})} \Big\} \\
&\overset{(7.58)}{\lesssim} \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} \Big\{ \sum_{j\in\mathbb{K}_n} a_n^4 \|T_{\xi_0}^{nj}\|^2 \Big(\|\mathcal{I}_{\xi_0}\|^2 + \int \|\ddot{\ell}_{\xi_0}\|^2 \, dP_{\xi(x_{nj})} \Big) \Big\} \\
&\overset{(7.59)}{\lesssim} a_n^4 k_n \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} \sup_{j\in\mathbb{K}_n} \Big(\|\mathcal{I}_{\xi_0}\|^2 + \int \|\ddot{\ell}_{\xi_0}\|^2 \, dP_{\xi(x_{nj})} \Big) \;\to\; 0. \qquad (7.60)
\end{aligned}$$

It remains to prove that also the second expectation in (7.57) converges to zero. From the definition of the matrices V_{nj} it is clear that every $V_{nj}(r,s)$ is a linear combination of the form

$$V_{nj}(r,s) = \sum_{q=1}^{d} T_{\xi_0}^{nj}(r,q)\big(\mathcal{I}_{\xi_0,qs} + \ddot{\ell}_{\xi_0,qs}(Y_{nj})\big), \qquad (7.61)$$

where $\mathcal{I}_{\xi_0,qs}$ and $\ddot{\ell}_{\xi_0,qs}$ denote the qs-entries of the matrices $\mathcal{I}_{\xi_0}$ and $\ddot{\ell}_{\xi_0}$, respectively, and $T_{\xi_0}^{nj}(r,q)$ is the rq-entry of the matrix $T_{\xi_0}^{nj}$. Furthermore, one concludes with (7.59) that

$$\sup_{\xi_0\in\Theta^*} \sup_{j\in\mathbb{K}_n} |T_{\xi_0}^{nj}(q,r)| < \infty, \qquad (7.62)$$

and therefore

$$\begin{aligned}
&\Big| \int V_{nj}(r,s) \, dP_{\xi(x_{nj})} \Big| \\
&\overset{(7.61)}{=} \Big| \int \sum_{q=1}^{d} T_{\xi_0}^{nj}(r,q)\big(\mathcal{I}_{\xi_0,qs} + \ddot{\ell}_{\xi_0,qs}(Y_{nj})\big) \, dP_{\xi(x_{nj})} \Big| \\
&\leq \sum_{q=1}^{d} |T_{\xi_0}^{nj}(r,q)| \, \Big| \mathcal{I}_{\xi_0,qs} + \int \ddot{\ell}_{\xi_0,qs}(Y_{nj}) \, dP_{\xi(x_{nj})} \Big| \\
&\overset{(3.4)}{=} \sum_{q=1}^{d} |T_{\xi_0}^{nj}(r,q)| \, \Big| - \int \ddot{\ell}_{\xi_0,qs}(Y_{nj}) \, dP_{\xi_0} + \int \ddot{\ell}_{\xi_0,qs}(Y_{nj}) \, dP_{\xi(x_{nj})} \Big| \\
&\overset{(7.62)}{\lesssim} \sum_{q=1}^{d} \Big| - \int \ddot{\ell}_{\xi_0,qs}(Y_{nj}) \, dP_{\xi_0} + \int \ddot{\ell}_{\xi_0,qs}(Y_{nj}) \, dP_{\xi(x_{nj})} \Big|.
\end{aligned}$$

Due to the growth conditions of the functions ξ we have $\|\xi(x_{nj}) - \xi_0\| \lesssim a_n$ for all $j \in \mathbb{K}_n$ and all ξ. Moreover, the continuous partial differentiability of the maps in (7.52) implies that these are Lipschitz continuous on the compactum $\overline{\Theta'} \times \overline{\Theta'}$. Combining these two arguments, it follows that

$$\begin{aligned}
&\sup_{\xi \in \mathscr{F}_{\mathrm{H}}} \sup_{j \in \mathbb{K}_n} \left| \int V_{nj}(r,s)\, dP_{\xi(x_{nj})} \right| \\
&\lesssim \sup_{\xi \in \mathscr{F}_{\mathrm{H}}} \sup_{j \in \mathbb{K}_n} \sum_{q=1}^{d} \left| -\int \ddot{\ell}_{\xi_0,qs}(Y_{nj})\, dP_{\xi_0} + \int \ddot{\ell}_{\xi_0,qs}(Y_{nj})\, dP_{\xi(x_{nj})} \right| \\
&\lesssim \sup_{\xi \in \mathscr{F}_{\mathrm{H}}} \sup_{j \in \mathbb{K}_n} \|\xi(x_{nj}) - \xi_0\| \\
&= O(a_n).
\end{aligned}$$

From this calculation, in combination with the property $a_n^2 k_n = O(1)$, it follows that

$$\begin{aligned}
&\sup_{\xi \in \mathscr{F}_{\mathrm{H}}} \left(\sum_{j \in \mathbb{K}_n} a_n^2 \int V_{nj}(r,s)\, dP_{\xi(x_{nj})} \right)^2 \\
&\qquad \lesssim \sup_{\xi \in \mathscr{F}_{\mathrm{H}}} \left(\sup_{j \in \mathbb{K}_n} a_n^2 k_n \left| \int V_{nj}(r,s)\, dP_{\xi(x_{nj})} \right| \right)^2 \\
&\qquad = (a_n^2 k_n)^2 \sup_{\xi \in \mathscr{F}_{\mathrm{H}}} \sup_{j \in \mathbb{K}_n} \left| \int V_{nj}(r,s)\, dP_{\xi(x_{nj})} \right|^2 \to 0. \qquad (7.63)
\end{aligned}$$

Putting together the above results yields

$$\begin{aligned}
&\sup_{\xi \in \mathscr{F}_{\mathrm{H}}} P_\xi^{(n)} \left\{ \left\| \sum_{j \in \mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj}) \right) \right\| > \frac{\varepsilon}{2} \right\} \\
&\qquad \overset{(7.55)}{\lesssim} \sup_{\xi \in \mathscr{F}_{\mathrm{H}}} \mathrm{E} \left\| \sum_{j \in \mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj}) \right) \right\|_{\mathrm{F}}^2 \\
&\qquad \overset{(7.57)}{\lesssim} \sum_{r,s=1}^{d} \sup_{\xi \in \mathscr{F}_{\mathrm{H}}} \sum_{j \in \mathbb{K}_n} a_n^4 \int V_{nj}(r,s)^2\, dP_{\xi(x_{nj})} \\
&\qquad\qquad + \sup_{\xi \in \mathscr{F}_{\mathrm{H}}} \left(\sum_{j \in \mathbb{K}_n} a_n^2 \int V_{nj}(r,s)\, dP_{\xi(x_{nj})} \right)^2.
\end{aligned}$$

According to (7.60) and (7.63) the last expression tends to zero, and therefore

$$\sup_{\xi \in \mathscr{F}_{\mathrm{H}}} P_\xi^{(n)} \left\{ \left\| \sum_{j \in \mathbb{K}_n} a_n^2 T_{\xi_0}^{nj} \left(\mathcal{I}_{\xi_0} + \ddot{\ell}_{\xi_0}(Y_{nj}) \right) \right\| > \frac{\varepsilon}{2} \right\} \to 0.$$

In combination with (7.51) and (7.54) this implies

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Bigg\{\sup_{\|z-\xi_0\|\le a_n C}\Bigg|\Bigg|\Bigg|\sum_{j\in\mathbb{K}_n} a_n^2 T_z^{nj}\left(\mathcal{I}_{\xi_0}+\ddot{\ell}_{\xi_0}(Y_{nj})\right)\Bigg|\Bigg|\Bigg| > \varepsilon\Bigg\}\to 0, \tag{7.64}$$

and thus the third expression in (7.44) converges uniformly in probability to zero.

Completion of the proof. To sum up, we have shown that for all $C>0$ and $\varepsilon>0$

$$\begin{aligned}
&\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Bigg\{\sup_{\|z-\xi_0\|\le a_n C}\|R_3(\xi_0,z)\| > \varepsilon\Bigg\}\\
&\overset{(7.42)}{\le} \sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\Bigg\{\sup_{z}\Bigg|\Bigg|\Bigg|\mathbb{1}+\sum_{j\in\mathbb{K}_n} a_n^2\left(\mathcal{C}_z^{-1}S_zB_z^{nj}(\mathcal{C}_z^{-1})^\top\right)\ddot{\ell}_{\xi_0}(Y_{nj})\Bigg|\Bigg|\Bigg| > \frac{\varepsilon}{C}\Bigg\}\\
&\overset{(7.44)}{\le} \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Bigg\{\Bigg|\Bigg|\Bigg|\mathbb{1}-\sum_{j\in\mathbb{K}_n} a_n^2T_{\xi_0}^{nj}\mathcal{I}_{\xi_0}\Bigg|\Bigg|\Bigg| > \frac{\varepsilon}{C3}\Bigg\}\\
&\qquad+\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Bigg\{\sup_{z}\Bigg|\Bigg|\Bigg|\sum_{j\in\mathbb{K}_n} a_n^2\left(T_{\xi_0}^{nj}-T_z^{nj}\right)\mathcal{I}_{\xi_0}\Bigg|\Bigg|\Bigg| > \frac{\varepsilon}{C3}\Bigg\}\\
&\qquad+\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Bigg\{\sup_{z}\Bigg|\Bigg|\Bigg|\sum_{j\in\mathbb{K}_n} a_n^2T_z^{nj}\left(\mathcal{I}_{\xi_0}+\ddot{\ell}_{\xi_0}(Y_{nj})\right)\Bigg|\Bigg|\Bigg| > \frac{\varepsilon}{C3}\Bigg\}\\
&\lesssim \sup_{\xi_0\in\Theta^*}\Bigg|\Bigg|\Bigg|\mathbb{1}-\sum_{j\in\mathbb{K}_n} a_n^2T_{\xi_0}^{nj}\mathcal{I}_{\xi_0}\Bigg|\Bigg|\Bigg|\\
&\qquad+\sup_{\xi_0\in\Theta^*}\sup_{z}\Bigg|\Bigg|\Bigg|\sum_{j\in\mathbb{K}_n} a_n^2\left(T_{\xi_0}^{nj}-T_z^{nj}\right)\mathcal{I}_{\xi_0}\Bigg|\Bigg|\Bigg|\\
&\qquad+\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Bigg\{\sup_{z}\Bigg|\Bigg|\Bigg|\sum_{j\in\mathbb{K}_n} a_n^2T_z^{nj}\left(\mathcal{I}_{\xi_0}+\ddot{\ell}_{\xi_0}(Y_{nj})\right)\Bigg|\Bigg|\Bigg| > \frac{\varepsilon}{C3}\Bigg\},
\end{aligned}$$

where the last step follows with Markov's inequality. By virtue of (7.46), (7.49) and (7.64), respectively, the expressions from the last inequality all tend to zero. Consequently,

$$\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Bigg\{\sup_{\|z-\xi_0\|\le a_n C}\|R_3(\xi_0,z)\| > \varepsilon\Bigg\}\to 0,$$

i.e. (7.41) holds. This completes the proof of Lemma 7.4. ■

7.2.4 Proof of Lemma 7.3

With the usual argumentation it suffices for the proof of Lemma 7.3 to verify that

$$\sup_{\xi \in \mathscr{F}_{\mathrm{H}}} P_{\xi}^{(n)} \left\{ \sup_{\|z-\xi_0\| \leq a_n C} \|R_2(\xi_0, z)\| > \varepsilon \right\} \to 0 \tag{7.65}$$

for all $\varepsilon > 0$ and $C > 0$. From the definition of R_2 in (7.8) we get

$$R_2(\xi_0, z) = \sum_{j=1}^{n} a_n \left(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top \right) r(\xi_0, z, Y_{nj}).$$

Of course, it suffices to build the sum only with respect to the indices $j \in \mathbb{K}_n$, because otherwise $B_z^{nj} = 0$. Using the variables T_z^{nj} from the proof of Lemma 7.4 we get

$$\begin{aligned}
&\sup_{\|z-\xi_0\| \leq a_n C} \|R_2(\xi_0, z)\| \\
&\quad = \sup_z \left\| \sum_{j \in \mathbb{K}_n} a_n \left(\mathcal{C}_z^{-1} S_z B_z^{nj} (\mathcal{C}_z^{-1})^\top \right) r(\xi_0, z, Y_{nj}) \right\| \\
&\quad \overset{(7.43)}{=} \sup_z \left\| \sum_{j \in \mathbb{K}_n} a_n T_z^{nj} r(\xi_0, z, Y_{nj}) \right\| \\
&\quad \overset{(7.13)}{\leq} \sup_z \sum_{j \in \mathbb{K}_n} a_n \|\!|T_z^{nj}|\!\| \, \|r(\xi_0, z, Y_{nj})\| \\
&\quad \overset{(7.59)}{\lesssim} \sup_z \sum_{j \in \mathbb{K}_n} a_n \|r(\xi_0, z, Y_{nj})\|.
\end{aligned} \tag{7.66}$$

Note that the last inequality does not depend on the value ξ_0. Let us next consider the remainder terms $r(\xi_0, z, Y_{nj})$ in greater detail. These result from the Taylor expansion of the score function in (3.2) and depend on the third partial derivatives of the log-densities $\log p_\theta$ (partial derivative means here with respect to the parameter). The latter are—by the smoothness assumptions from Section 3.2—continuous for all parameters $\theta \in \Theta$, and therefore in particular bounded on the compactum $\overline{\Theta'}$. According to Heuser (1990), p. 285, therefore

$$\begin{aligned}
\|r(\xi_0, z, y)\| &\lesssim \max_{k=1,\dots,d} \sum_{r,s=1}^{d} \left| \frac{\partial^2}{\partial\theta_r \partial\theta_s} \dot{\ell}_{\xi_0,k}(y) \right| \, \|z - \xi_0\|^2 \\
&\lesssim J(y) \|z - \xi_0\|^2,
\end{aligned}$$

where J is the dominating function from (3.3). Plugging this into (7.66) yields

$$\begin{aligned}\sup_{\|z-\xi_0\|\le a_n C}\|R_2(\xi_0,z)\| &\lesssim \sup_z \sum_{j\in\mathbb{K}_n} a_n\|r(\xi_0,z,Y_{nj})\| \\ &\lesssim \sum_{j\in\mathbb{K}_n} a_n^3 J(Y_{nj}).\end{aligned}$$

This can of course also be written as

$$\sup_{\|z-\xi_0\|\le a_n C}\|R_2(\xi_0,z)\| \le N\sum_{j\in\mathbb{K}_n} a_n^3 J(Y_{nj}), \tag{7.67}$$

with some suitable constant $N>0$ (which does not depend on ξ_0). The above arguments and an additional application of Markov's inequality yield

$$\begin{aligned}&\sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Big\{\sup_{\|z-\xi_0\|\le a_n C}\|R_2(\xi_0,z)\|>\varepsilon\Big\}\\ &\overset{(7.67)}{\le} \sup_{\xi\in\mathscr{F}_{\mathrm{H}}} P_\xi^{(n)}\Big\{\sum_{j\in\mathbb{K}_n} a_n^3 J(Y_{nj})>\frac{\varepsilon}{N}\Big\}\\ &\le \sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\frac{N}{\varepsilon}\int\Big|\sum_{j\in\mathbb{K}_n} a_n^3 J(Y_{nj})\Big|\,dP_\xi^{(n)}\\ &= \sup_{\xi\in\mathscr{F}_{\mathrm{H}}}\frac{N}{\varepsilon}\sum_{j\in\mathbb{K}_n} a_n^3\int J(Y_{nj})\,dP_{\xi(x_{nj})}\\ &\lesssim a_n^3 k_n.\end{aligned}$$

In the last step we exploited that, by assumption, $\mathrm{E}_\theta\, J$ is uniformly bounded on Θ (see Section 3.2). Since $k_n a_n^2 = O(1)$ and $a_n = o(1)$ (see (4.8) and (3.14), respectively) the last expression tends to zero. This proves (7.65), and thus the assertion of Lemma 7.3. ■

CHAPTER 8

Applying the theory

In this Chapter we discuss some aspects of an application of the theoretical results presented in the preceding chapters. One of the most important models (maybe *the* most important model) to which our theory applies is clearly the classical nonparametric regression model. This is discussed in Section 8.1. In Section 8.2 possible extensions of these results to more general models are explored. In Section 8.3 we discuss in detail exponential families and prove that these always meet the regularity conditions presented in Sections 3.1 and 3.2.

8.1 The nonparametric regression model

In this section we show that the theory provided in the preceding chapters is in some sense tailor-made to derive asymptotic minimax risk bounds in a nonparametric regression model.

Let independent observations $(x_{n1}, Y_{n1}), \dots, (x_{nn}, Y_{nn})$ be given. The x_{nj} are here interpreted as time points and the (real) variables Y_{nj} describe a certain process which evolves over time. In the classic nonparametric regression setup it is assumed that these observations obey a model of the form

$$Y_{nj} = \xi(x_{nj}) + \varepsilon_{nj},$$

where the unknown *regression function* ξ is assumed to be a smooth function, and the ε_{nj} are interpreted as noise. More precisely, for each n we assume that $\varepsilon_{n1}, \dots, \varepsilon_{nn}$ build a sequence of i.i.d. random variables with mean zero and finite variance. A very popular approach to estimate the value of the regression function at some point x_0 is to use *kernel estimators* like the Nadaraya-Watson estimator. These estimate $\xi(x_0)$ as a weighted average over the observations Y_{nj}, where the observations for which the corresponding x_{nj} lie close to x_0 are usually assigned larger weights than that which are further distant in time. There is a vast, ever-growing literature on nonparametric regression, see for example Fan and Gijbels (1996), Györfi et al. (2002), Härdle (1990) or Loader (1999)—just to mention a few. Rohde (2004) and Reiß (2008) show that the nonparametric regression model is asymptotically equivalent to a Gaussian white noise model (cf. also the references given in Section 1.3).

Assume that the distribution of ε_{nj} has Lebesgue density $p = p_0$. For $\theta \in \mathbb{R}$ we set

$$p_\theta(y) := p(y - \theta), \quad y \in \mathbb{R},$$

and we define

$$P_\theta := p_\theta \, d\lambda.$$

We consider the resulting family $\mathcal{P} := \{P_\theta : \theta \in \Theta\}$, where Θ is an open interval in $\mathbb{R}$ with compact closure. Let us assume that the (unknown) regression function ξ belongs to a function space $\mathscr{F} = \mathscr{F}(x_0, \Theta^*, u)$ of the form (3.6), where $\Theta^* \subseteq \Theta$ is an open interval, and the function u is according to $u\colon \mathbb{R} \to [0, \infty)$, $s \mapsto |s|^\rho$. If, in addition, the time points are given by $x_{nj} = (j-1)/(n-1)$, as in Section 3.3 (which may always be achieved through a rescaling), and $x_0 \in (0, 1)$, we regain our model from Section 1.1. The measurable space $(\mathcal{Y}, \mathcal{B}, \mu)$ from that section is in this case equal to $(\mathbb{R}, \mathbb{B}, \lambda)$. If the distributions P_θ fit the needs of Sections 3.1 and 3.2, and if we are able to prove that a preliminary estimator exists (as in Section 3.4), the theory from the preceding chapters applies, which yields upper and lower asymptotic minimax risk bounds for the estimation of $\xi(x_0)$.

Checking the regularity conditions. First we formulate conditions on the density p which guarantee that the resulting location family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ satisfies the regularity conditions. Assume that p is strictly positive and λ-a.e. three times continuously differentiable (with derivative p'). If the density further satisfies

$$\int \frac{p'(x)^2}{p(x)}\, d\lambda(x) = \int \left(\frac{p'(x)}{p(x)}\right)^2\, dP(x) < \infty,$$

then the location family $\mathcal{P}$ is differentiable in quadratic mean on $\mathbb{R}$, and thus on Θ, too. The score function is given by

$$\dot{\ell}_\theta(y) = -\frac{p'(y-\theta)}{p(y-\theta)} \quad \lambda\text{-a.s.}$$

and therefore satisfies (3.1) (see e.g. Witting, 1985, p. 181). To conclude that $\mathcal{P}$ is even *continuously* differentiable in quadratic mean, we observe that for fixed $\theta_0 \in \Theta$ and a sequence $\theta_n \to \theta_0$ we have $\dot{\ell}_{\theta_n}\sqrt{p_{\theta_n}} \to \dot{\ell}_{\theta_0}\sqrt{p_{\theta_0}}$ λ-a.s. Since we are considering an open interval Θ with compact closure, it is in most cases possible to find a square-integrable function H such that $|\dot{\ell}_\theta(y)\sqrt{p_\theta(y)}| \leq H(y)$ for all y and all θ. In that case the dominated convergence theorem implies

$$\begin{aligned}
\lim_{n\to\infty} \|\dot{\ell}_{\theta_n}\sqrt{p_{\theta_n}} - \dot{\ell}_{\theta_0}\sqrt{p_{\theta_0}}\|^2_{\lambda,2} &= \lim_{n\to\infty} \int \left(\dot{\ell}_{\theta_n}\sqrt{p_{\theta_n}} - \dot{\ell}_{\theta_0}\sqrt{p_{\theta_0}}\right)^2 d\lambda \\
&= \int \lim_{n\to\infty} \left(\dot{\ell}_{\theta_n}\sqrt{p_{\theta_n}} - \dot{\ell}_{\theta_0}\sqrt{p_{\theta_0}}\right)^2 d\lambda \\
&= 0,
\end{aligned}$$

and thus the continuous $\mathbb{L}_2$-differentiability. Note that the Fisher information in location families is constant and thus in particular continuous. The remaining conditions from Section 3.2, i.e. the differentiability assumptions for the moments of the score function can often be proved directly. We do this exemplarily for the case of a Gaussian location model.

The Gaussian location model. Let $p = \varphi$ denote the pdf of a standard normal distribution, i.e. $p(y) = 1/\sqrt{2\pi}\exp(-y^2/2)$. We first compute the derivatives of the (log-)densities. These are given by

$$\begin{aligned}
p_\theta(y) &= \frac{1}{\sqrt{2\pi}} \exp\Big\{ -\frac{(y-\theta)^2}{2} \Big\}, \\
\nabla_\theta p_\theta(y) &= (y-\theta) p_\theta(y), \\
\log p_\theta(y) &= \log \frac{1}{\sqrt{2\pi}} - \frac{(y-\theta)^2}{2}, \\
\dot{\ell}_\theta(y) &= y - \theta, \\
\ddot{\ell}_\theta(y) &= -1.
\end{aligned}$$

It is clear that the third derivative of the log-density is equal to zero, and therefore a function J exists which satisfies the condition from (3.3). Condition (3.4) also follows directly, because $-\int \ddot{\ell}_\theta \, dP_\theta = \int 1 \, dP_\theta = 1$ and $\int \dot{\ell}_\theta^2 \, dP_\theta = \int (y-\theta)^2 \, dP_\theta = 1$ (note that $Y \sim P_\theta$ has variance equal to 1). Evidently, the remaining assumptions from Section 3.2 imposing smoothness conditions for the moments of the score functions are also fulfilled. It remains to prove (3.5). With the above expression we calculate

$$\begin{aligned}
\nabla\tau \textstyle\int \dot{\ell}_\theta p_\tau \, d\lambda &= \nabla_\tau \textstyle\int (y-\theta) p_\tau \, d\lambda(y) \\
&= \nabla_\tau \left(\textstyle\int y p_\tau(y) \, d\lambda(y) - \theta \int p_\tau(y) \, d\lambda(y) \right) \\
&= \nabla_\tau (\tau - \theta) \\
&= 1,
\end{aligned}$$

and

$$\begin{aligned}
\textstyle\int \dot{\ell}_\theta \nabla_\tau p_\tau \, d\lambda &= \textstyle\int (y-\theta)(y-\tau) p_\tau(y) \, d\lambda(y) \\
&= \textstyle\int y^2 p_\tau(y) \, d\lambda(y) - (\theta+\tau) \int y p_\tau(y) \, d\lambda(y) + \theta\tau \int p_\tau \, d\lambda \\
&= 1 + \tau^2 - (\theta+\tau)\tau + \theta\tau \\
&= 1.
\end{aligned}$$

Altogether, these calculations yield $\nabla\tau \int \dot{\ell}_\theta p_\tau \, d\lambda = \int \dot{\ell}_\theta \nabla_\tau p_\tau \, d\lambda$, and thus (3.5). To sum up, we have shown, that the Gaussian location model fits all the conditions from Sections 3.1 and 3.2. Likewise, the regularity conditions for large classes of other distributions can be verified.

The existence of a preliminary estimator. For regression models of the above type, Stone (1980) gives an affirmative answer to the question of the existence of a preliminary estimator.

The following presentation reflects the main ideas and concepts of Stone's article.

Stone discusses optimal rates of convergence in the above regression setting. Assume that $T = T(\xi)$ is a real valued functional, and $\hat{T}_n$ is a sequence of estimators based on a sample of size n. Since we want to estimate the parameter value $\xi(x_0)$, we will in the following always assume that $T(\xi) = \xi(x_0)$. In order to quantify and to compare the speed at which sequences of estimators do converge Stone defines the following concepts (cf. Stone, 1980, p. 1348): A positive constant r is called an *upper bound* to the rate of convergence if for every sequence of estimators $\hat{T}_n$ and every $C > 0$

$$\liminf_{n\to\infty} \sup_{\xi} P_\xi^{(n)}\big\{|\hat{T}_n - T(\xi)| > Cn^{-r}\big\} > 0,$$

and if in addition

$$\lim_{C\to 0} \liminf_{n\to\infty} \sup_{\xi} P_\xi^{(n)}\big\{|\hat{T}_n - T(\xi)| > Cn^{-r}\big\} = 1.$$

Moreover, r is called an *achievable* rate of convergence if a specific sequence of estimators $\hat{T}_n$ exists, for which

$$\lim_{C\to \infty} \limsup_{n\to\infty} \sup_{\xi} P_\xi^{(n)}\big\{|\hat{T}_n - T(\xi)| > Cn^{-r}\big\} = 0.$$

r is called the *optimal* rate of convergence if it is both achievable and an upper bound to the rate of convergence.

Stone (1980) provides results on the optimal rate of convergence in the nonparametric regression model. In his paper he postulates that certain smoothness conditions hold for the parameter functions (more precisely, these are Lipschitz and Hölder conditions on the parameter functions and their derivatives). As a special case this includes models with parameter functions ξ ranging over a function spaces of the type $\mathscr{F}$ (see (3.6), Section 3.3), i.e. for functions that satisfy a Hölder condition with coefficient ρ. For such models Stone shows that—provided that the densities p_θ satisfy some regularity conditions—the optimal rate of convergence is given by

$$r = \frac{\rho}{2\rho + 1}.$$

This means that there is a sequence of estimators $\hat{T}_n =: \xi_n^*$ for $\xi(x_0)$ such that

$$\lim_{C\to \infty} \limsup_{n\to\infty} \sup_{\xi\in\mathscr{F}} P_\xi^{(n)}\big\{|\xi_n^* - \xi(x_0)| > Cn^{-\rho/(2\rho+1)}\big\} = 0. \quad (8.1)$$

By the definition from (3.14) (here we presume $A = 1$), we have

$$a_n = n^{-\rho/(2\rho+1)} = n^{-r}.$$

From (8.1) we conclude that for every $\iota > 0$ there is a constant $C = C_\iota$ for which

$$\limsup_{n\to\infty} \sup_{\xi\in\mathscr{F}} P_\xi^{(n)}\big\{|\xi_n^* - \xi(x_0)| > Cn^{-\rho/(2\rho+1)}\big\} < \iota,$$

and therefore

$$\begin{aligned}\limsup_{n\to\infty} \sup_{\xi\in\mathscr{F}} & P_\xi^{(n)}\big\{|\xi_n^* - \xi(x_0)| > Ca_n\big\} \\ &= \limsup_{n\to\infty} \sup_{\xi\in\mathscr{F}} P_\xi^{(n)}\left\{\frac{1}{a_n}|\xi_n^* - \xi(x_0)| > C\right\} < \iota.\end{aligned}$$

Hence the assumption of a $(1/a_n)$-consistent preliminary estimator in the sense of Section 3.4 is fulfilled in the nonparametric regression model.

Note that the assumptions Stone imposes on the densities p_θ are mainly conditions concerning interchangeability of certain differentiation and integration procedures. In general, these conditions are already included within our conditions from Sections 3.1 and 3.2. In particular, Stone's assumptions hold for the Gaussian location model, which was discussed above (cf. Stone, 1980, for more details).

Conclusion. The results of this thesis on asymptotic upper and lower minimax risk bounds hold for certain nonparametric regression models—in particular they hold for the nonparametric regression model with Gaussian noise.

8.2 Extensions to other models

For the nonparametric regression model discussed in the previous section, we could verify that all the conditions from Chapter 3 hold. In particular, we saw that the existence of a $(1/a_n)$-consistent preliminary estimator, as postulated in Section 3.4, is guaranteed.

However, the results in Stone (1980) also hold for other regression-type models—i.e. for models as described in Section

1.1—in which the parameter function ξ does not necessarily describe a location parameter. For the estimation of the parameter function one may use so-called local likelihood estimators, as discussed by Aerts and Claeskens (1997). As concerns the rate of convergence, Stone shows that also for this more general class of models an optimal rate of convergence

$$r = \frac{\rho}{2\rho + 1}$$

can be established, provided that the parameter functions ξ satisfy a Hölder condition with coefficient $\rho > 0$, and that the densities p_θ satisfy certain regularity conditions. The latter are in general covered by our assumptions from Sections 3.1 and 3.2. As examples for models which satisfy these conditions Stone presents the following distribution families $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ (see Stone, 1980, Examples 2–5):

Exponential distributions. The density of an exponential distribution with parameter $\theta > 0$ is given by

$$p_\theta(y) = \frac{1}{\theta} \exp\{-y/\theta\}, \quad y \geq 0.$$

The parameter space under consideration is assumed to be a relatively compact interval $\Theta \subseteq (0, \infty)$.

Poisson distributions. For $\theta > 0$ and $\mathcal{Y} := \mathbb{N} \cup \{0\}$

$$p_\theta(y) = \frac{\theta^y e^{-\theta}}{y!}, \quad y \in \mathcal{Y}$$

defines the probability function of a Poisson distribution. The parameter space Θ is assumed to be a relatively compact subset of $(0, \infty)$.

Geometric distributions. The probability function of the geometric distribution with parameter $\theta > 0$ is defined by

$$p_\theta(y) = \left(\frac{1}{1+\theta}\right) \left(\frac{\theta}{1+\theta}\right)^y, \quad y \in \mathcal{Y},$$

with $\mathcal{Y}$ as in the preceding example. The parameter space Θ is assumed to be a relatively compact subset of $(0, \infty)$.

Bernoulli distributions. For $\theta \in (0,1)$

$$p_\theta(y) = \theta^y(1-\theta)^{1-y}, \quad y \in \{0,1\}$$

defines the probability function of the Bernoulli distribution with parameter θ. The parameter space Θ is assumed to be a relatively compact subset of $(0,1)$.

Conclusion. In all models in which the distributions are given by either of the above families $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ the results from Stone (1980)—in combination with the arguments from the preceding section—guarantee the existence of a preliminary estimator ξ_n^* for $\xi(x_0)$ that converges with the optimal rate, and thus satisfies the needs of Section 3.4.

For the above distribution families $\{P_\theta : \theta \in \Theta\}$ it remains of course to prove that—besides the conditions from Stone (1980)—also the regularity assumptions from Sections 3.1 and 3.2 hold. Observe that all the above distribution families are one-dimensional exponential families (see the table in Lehmann and Casella, 1998, p. 25). We will prove in the next section that exponential families in most cases satisfy the conditions from Sections 3.1 and 3.2. Consequently, our results on asymptotic upper and lower minimax risk bounds do also hold for these specific models.

8.3 Exponential families

We show that exponential families $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ usually satisfy the regularity conditions from Sections 3.1 and 3.2.

Definition and general properties. In the following, let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ ($\Theta \subseteq \mathbb{R}^d$ open) be a dominated family of probability measures with some dominating, σ-finite measure μ and densities $p_\theta = dP_\theta/d\mu$. The underlying measurable space is denoted $(\mathcal{Y}, \mathcal{B})$. $\mathcal{P}$ is called a *d-dimensional exponential family in ζ and T* if the densities have a—not necessarily unique—representation of the form

$$p_\theta(y) = c(\theta)h(y)\exp\{\langle\zeta(\theta), T(y)\rangle\} \quad \mu\text{-a.e.}, \tag{8.2}$$

with measurable functions

$$\zeta, T\colon (\mathcal{Y}, \mathcal{B}) \to (\mathbb{R}^d, \mathbb{B}^d),$$
$$h\colon (\mathcal{Y}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B}), \quad h \geq 0.$$

The function $T = (T_1, \ldots, T_d)^\top$ is also called the *generating statistic* of the exponential family. Passing over from the dominating measure μ to the also dominating measure $\nu = h\ d\mu$, we can—and will in the following—assume without loss of generality that $h \equiv 1$. For the sake of simplicity, we will further restrict ourselves to the special case $\zeta(\theta) = \theta$ and therefore for the rest of this section presume that the densities under consideration are of the specific *canonical* form

$$p_\theta(y) = c(\theta) \exp\{\langle \theta, T(y) \rangle\} \quad \mu\text{-a.e.}^{1)} \tag{8.3}$$

Checking the regularity conditions. Exponential families in the canonical representation are differentiable in quadratic mean. The score function satisfies

$$\begin{aligned} \dot{\ell}_\theta(y) &= \nabla_\theta \log c(\theta) + T(y) \\ &= T(y) - \mathrm{E}_\theta T, \end{aligned} \tag{8.4}$$

and thus in particular (3.1). Continuous differentiability in quadratic mean can be established with the same ideas as proposed in Section 8.1. Moreover, we have the relations

$$\begin{aligned} \mathrm{E}_\theta T &= -\nabla_\theta \log c(\theta), \\ \mathcal{I}_\theta = \mathrm{Cov}_\theta T &= -\nabla_\theta^2 \log c(\theta). \end{aligned} \tag{8.5}$$

(Cf. van der Vaart, 1998, pp. 38 and 95 f. for proofs of these statements.) According to Satz 1.153 and 1.164 in Witting (1985) $\mathcal{I}_\theta$ is positive definite for all $\theta \in \Theta$ if and only if this is the case for a single point $\theta^* \in \Theta$. This is equivalent to the components $T_1, \ldots, T_d$ of T being $\mathcal{P}$-affinely independent, which means that the implication

$$\sum_{j=1}^{d} b_j T_j(y) = b_0 \quad P_\theta\text{-a.s. for all } \theta \ \Rightarrow \ b_0 = b_1 = \ldots = b_d = 0$$

[1] Although we discuss only exponential families in the canonical representation, most of the following statements also hold for general exponential families with densities according to (8.2). In most relevant cases, the map ζ in (8.2) is one-to-one and sufficiently smooth. The properties of an exponential family in ζ and T can then be deduced with a parameter transformation from that of the exponential family in canonical representation, see e.g. Witting (1985), p. 154., and van der Vaart (1998), Example 7.7.

holds true. Furthermore, Satz 1.164 in Witting (1985) states that the generating statistic T possesses moments of arbitrary order and the maps $\theta \mapsto \mathrm{E}_\theta T_1^{k_1} \dots T_d^{k_d}$ and $\theta \mapsto c(\theta)$ are infinitely often differentiable. Using the representations of the score function in (8.4) one concludes that the functions

$$(\theta, \tau) \mapsto \int (\dot\ell_{\theta,i})^{k_1} (\dot\ell_{\theta,j})^{k_2} \, dP_\tau, \quad k_1, k_2 = 0, 1, 2, 3, \dots$$

and

$$(\theta, \tau) \mapsto \int (\ddot\ell_{\theta,rs})^{k} \, dP_\tau, \quad k = 0, 1, 2, 3, \dots$$

are infinitely often partially differentiable on $\Theta \times \Theta$. In the following we may assume that all of the above functions are also bounded. (This can always be achieved by replacing Θ with a suitable relatively compact subset.)

Evidently, the (log-)densities of an exponential family possess partial derivatives (with respect to the parameter) of up to third order, outside an appropriate set $N \subseteq \mathcal{Y}$ with $\mu(N) = 0$. Moreover, it becomes clear from the expression for the score-function from (8.4) that the second and third order partial derivatives of the log-likelihood function do not depend on y. Consequently, there is a function J which satisfies (3.3), and for which $\mathrm{E}_\theta J$ is uniformly bounded on Θ. The boundedness follows from the property that the map $\theta \mapsto \mathrm{E}_\theta T$ is infinitely often partially differentiable on Θ. Differentiation of (8.4) with respect to θ further yields

$$-\int \ddot\ell_\theta \, dP_\theta = -\int \nabla_\theta^2 \log c(\theta) \, dP_\theta \overset{(8.5)}{=} \mathcal{I}_\theta,$$

and thus (3.4). Finally, we prove (3.5). To this end, we first observe that

$$\begin{aligned} \nabla_\tau \int \dot\ell_\theta p_\tau \, d\mu &\overset{(8.4)}{=} \nabla_\tau \left(\int T p_\tau \, d\mu - \int (\mathrm{E}_\theta T) p_\tau \, d\mu \right) \\ &= \nabla_\tau (\mathrm{E}_\tau T - \mathrm{E}_\theta T) \\ &= \nabla_\tau \mathrm{E}_\tau T \\ &\overset{(8.5)}{=} \mathrm{Cov}_\tau T. \end{aligned}$$

On the other hand, differentiation of (8.3) yields

$$\begin{aligned} \nabla_\tau p_\tau(y) &= \big(\nabla_\tau c(\tau)\big) \exp\{\langle \tau, T(y) \rangle\} + c(\tau) T(y) \exp\{\langle \tau, T(y) \rangle\} \\ &= \big(\nabla_\tau \log c(\tau)\big) p_\tau(y) + T(y) p_\tau(y), \end{aligned}$$

and therefore

$$\begin{aligned}
\int \dot{\ell}_\theta (\nabla_\tau p_\tau)^\top \, d\mu &= \int \dot{\ell}_\theta \big(\nabla_\tau \log c(\tau)\big)^\top p_\tau \, d\mu + \int \dot{\ell}_\theta T^\top p_\tau \, d\mu \\
&\overset{(8.4)}{=} \left(\int (T - \mathrm{E}_\theta\, T) \, dP_\tau \right) \big(\nabla_\tau \log c(\tau)\big)^\top \\
&\qquad + \int (T - \mathrm{E}_\theta\, T) T^\top \, dP_\tau \\
&\overset{(8.5)}{=} (\mathrm{E}_\tau\, T - \mathrm{E}_\theta\, T)(-\mathrm{E}_\tau\, T)^\top + \int T T^\top \, dP_\tau \\
&\qquad - (\mathrm{E}_\theta\, T)(\mathrm{E}_\tau\, T)^\top \\
&= \mathrm{E}_\tau (T T^\top) - (\mathrm{E}_\tau\, T)(\mathrm{E}_\tau\, T)^\top \\
&= \operatorname{Cov}_\tau T.
\end{aligned}$$

These calculations show that (3.5) holds.

Conclusion. To sum up, we have shown that the regularity conditions from Sections 3.1 and 3.2 are in general satisfied if $\mathcal{P}$ is given by an exponential family. Together with the examples from the preceding Section and the results from Stone (1980) this opens diverse opportunities to apply the theoretical results from this thesis.

Summary

To better understand the time-dependent development of certain phenomena—e.g. in economy, meteorology or science—the statistician has to fit a suitable model to a given series of observed data points. This always comes with the task of estimating certain model parameters on the basis of the observed data. Naturally, the question of the quality of the used estimation procedures arises. This problem was tackled in the thesis. A rather abstract stochastic model was considered, in which the distributions of the independent observations were assumed to be given by a parametric distribution family, while the distribution parameter was supposed to be driven by an unknown, smooth function. As an important special case this approach covers the nonparametric regression model. Within this model the minimax risk with respect to a zero-one loss function was investigated. In simple terms, this means that an estimator is judged by the probability with which the estimate takes a value outside of a specific neighbourhood of the true value of the parameter to estimate. The scope of this thesis was to derive asymptotic minimax risk bounds under rather general conditions. The basic idea for the examinations was to use a localisation procedure, which is a common technique in asymptotic statistics. This means that first a localised model was considered in which certain estimators were shown to be asymptotically normally distributed. Both upper and lower bounds for the asymptotic minimax risk within this local model could be established. In a second step these bounds could be transferred to the original model, too. Furthermore, it could be shown that the general nonparametric regression model and also some other models satisfy the regularity conditions that were imposed to derive the asymptotic minimax risk bounds.

References

Aerts, M. and Claeskens, G. (1997). Local polynomial estimation in multiparameter likelihood models. *Journal of the American Statistical Association*, 92:1536–1545. 21, 163

Bauer, H. (1992). *Maß- und Integrationstheorie.* De Gruyter, Berlin. 10

Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. (1998). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer, New York. 18

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York. 34

Bingham, N. and Kiesel, R. (1998). *Risk-Neutral Valuation – Pricing and Hedging of Financial Derivatives.* Springer, London. 35

Brown, L. and Low, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Annals of Statistics*, 24:2384–2398. 4

Cai, T. and Low, M. (2004). Minimax estimation of linear functionals over nonconvex parameter spaces. *Annals of Statistics*, 32:552–576. 4

Donoho, D. (1994). Statistical estimation and optimal recovery. *Annals of Statistics*, 22:238–270. 4, 55, 56, 66, 72, 88

Donoho, D. and Liu, R. (1991). Geometrizing rates of convergence, III. *Annals of Statistics*, 19:668–701. 4, 55, 56, 66, 79

Donoho, D., Liu, R., and MacGibbon, B. (1990). Minimax risk over hyperrectangles, and implications. *Annals of Statistics*, 18:1416–1437. 88, 90

Drees, H. (1999). On fixed-length confidence intervals for a bounded normal mean. *Statistics & Probability Letters*, 44:399–404. 54, 64, 65, 88

Drees, H. (2001). Minimax risk bounds in extreme value theory. *Annals of Statistics*, 29:266–294. 1, 4, 5, 34, 50, 59, 64, 65

Drees, H. and Stărică, C. (2002). A simple non-stationary model for stock returns. Preprint, available at `www.math.chalmers.se/~starica`. 1

Elstrodt, J. (1996). *Maß- und Integrationstheorie*. Springer, Berlin. 108

Fabian, V. (1970). On uniform convergence of measures. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 15:139–143. 16

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall, London. 158

Gänssler, P. and Stute, W. (1977). *Wahrscheinlichkeitstheorie*. Springer, Berlin. 34

Grama, I. and Nussbaum, M. (1998). Asymptotic equivalence for generalized linear models. *Probability Theory and Related Fields*, 111:167–214. 4

Grama, I. and Nussbaum, M. (2002). Asymptotic equivalence for nonparametric regression. *Mathematical Methods of Statistics*, 11:1–36. 4

Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regresssion*. Springer, New York. 158

Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge. 158

Heuser, H. (1990). *Lehrbuch der Analysis, Teil 2.* B.G. Teubner, Stuttgart. 10, 11, 101, 105, 155

Heuser, H. (2000). *Lehrbuch der Analysis, Teil 1.* B.G. Teubner, Stuttgart. 100, 145

Huzurbazar, V. (1950). Probability distributions and orthogonal parameters. *Proceedings of the Cambridge Philosophical Society*, 46:281–284. 89

Ibragimov, I. and Khas'minski, R. (1985). On nonparametric estimation of values of a linear functional in gaussian white noise. *Theory of Probability and its Applications*, 29:18–32. 88

Jähnisch, M. and Nussbaum, M. (2003). Asymptotic equivalence for a model of independent non identically distributed observations. *Statistics & Decisions*, 21:197–218. 4

Jönck, U. (2008). Local likelihood estimators in a regression model for stock returns. *Quantitative Finance*, 8:619–635. 1

Karatzas, I. and Shreve, S. (1988). *Brownian Motion and Stochastic Calculus.* Springer, New York. 34, 35, 86

Klebaner, F. (1998). *Introduction to Stochastic Calculus with Applications.* Imperial College Press, London. 34, 36

Königsberger, K. (2000). *Analysis II.* Springer, Berlin. 38, 123, 125

Königsberger, K. (2001). *Analysis I.* Springer, Berlin. 62, 70

Le Cam, L. (1986). *Asymtptotic Methods in Statistical Decision Theory.* Springer, New York. 18

Le Cam, L. and Yang, G. (2000). *Asymtptotics in Statistics – Some Basic Concepts.* Springer, New York. 18

Lehmann, E. and Casella, G. (1998). *Theory of Point Estimation.* Springer, New York. 21, 164

Lipschutz, S. (1977). *Allgemeine Topologie – Theorie und Anwendung.* McGraw-Hill, Düsseldorf. 17, 101

Loader, C. (1999). *Local Regression and Likelihood.* Springer, New York. 158

Mitchell, A. (1962). Sufficient statistics and orthogonal parameters. *Proceedings of the Cambridge Philosophical Society*, 58:326–337. 89

Nussbaum, M. (1996). Asymptotic equivalence of density estimation and gaussian white noise. *Annals of Statistics*, 24:2399–2430. 4

Opfer, G. (2001). *Numerische Mathematik für Anfänger*. Vieweg, Braunschweig. 20, 123, 125

Pfanzagl, J. (1994). *Parametric Statistical Theory*. De Gruyter, Berlin. 17, 18, 98, 113

Rao, R. (1962). Relations between weak and uniform convergence of measures with applications. *Annals of Mathematical Statistics*, 33:659–680. 16

Reiß, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *Annals of Statistics*, 36:1957–1982. 4, 158

Reiss, R.-D. (1993). *A Course on Point Processes*. Springer, New York. 109

Rieder, H. (1994). *Robust Asymptotic Statistics*. Springer, New York. 11, 18

Rohde, A. (2004). On the asymptotic equivalence and rate of convergence of nonparametric regression and Gaussian white noise. *Statistics & Decisions*, 22:235–243. 4, 158

Stark, P. (1992). Affine minimax confidence intervals for a bounded normal mean. *Statistics & Probability Letters*, 13:39–44. 72

Stoer, J. (1999). *Numerische Mathematik 1*. Springer, Berlin. 20

Stone, C. (1980). Optimal rates of convergence for nonparametric estimators. *Annals of Statistics*, 8:1348–1360. 160, 161, 162, 163, 164, 167

Strasser, H. (1985). *Mathematical Theory of Statistics*. De Gruyter, Berlin. 15, 16, 18, 34, 53 van der Vaart, A. (1988). *Statistical Estimation in Large Parameter Spaces*. CWI Tract 44, Centrum voor Wiskunde en Informatica, Amsterdam. 11, 12

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge. 8, 9, 12, 18, 44, 45, 95, 102, 165

Witting, H. (1985). *Mathematische Statistik I*. B.G. Teubner, Stuttgart. 13, 18, 159, 165, 166

Witting, H. and Müller-Funk, U. (1995). *Mathematische Statistik II*. B.G. Teubner, Stuttgart. 9, 18

Zeytinoglu, M. and Mintz, M. (1984). Optimal fixed size confidence procedures for a restricted parameter space. *Annals of Statistics*, 12:945–957. 64

List of special symbols

Zusammenfassung

In den Wirtschafts- und Naturwissenschaften tritt häufig die Frage auf, wie man bestimmte sich zeitlich entwickelnde Prozesse geeignet beschreiben kann. Aufgabe der Statistik ist es, für solche Prozesse passende Modelle zu entwickeln, um mit diesen die beobachteten Daten beschreiben und erklären zu können. Damit einher geht die Aufgabe, eine bestimmte Anzahl von Parametern innerhalb eines solchen Modells auf Grundlage der vorliegenden Daten zu schätzen. Es stellt sich die Frage nach der Güte der dazu verwendeten Schätzverfahren. Diese Frage wurde im Rahmen der Dissertation behandelt. Grundlage für die Untersuchungen war ein abstraktes stochastisches Modell, bei dem die als unabhängig angenommenen Beobachtungen durch eine parametrische Verteilungsfamilie modelliert werden. Von dem Verteilungsparameter wird angenommen, dass dieser durch eine unbekannte, hinreichend glatte Funktion gesteuert wird. Ein wichtiger Spezialfall dieses Modells ist das klassische nichtparametrische Regressionsmodell. In dem betrachteten Modell wurde das Minimax-Risiko unter Verwendung einer 0-1-Verlustfunktion untersucht. Einfach ausgedrückt bedeutet dies, dass ein Schätzer für den unbekannten Verteilungsparameter danach beurteilt wird, wie groß die Wahrscheinlichkeit ist, dass er einen Wert in einer Umgebung des wahren Parameterwertes annimmt. Ziel der Arbeit war es, unter möglichst allgemeinen Bedingungen asymptotische obere und untere Schranken für das Minimax-Risiko herzuleiten. Eine grundlegende Idee der Untersuchungen hierzu bestand darin, eine Lokalisierung des Modells vorzunehmen, und das Minimax-Risiko zunächst in den resultierenden lokalen Modellen zu untersuchen. In diesen konnte die asymptotische Normalität gewisser Schätzer nachgewiesen und damit obere und untere Schranken für das Minimax-Risiko hergeleitet werden. Durch die Konstruktion von Schätzern mit einer hinreichend guten Konvergenzrate konnten diese Schranken in einem zweiten Schritt auf das ursprüngliche Modell übertragen werden. Es wurden zudem Beispiele für Modelle gebracht, auf die die Theorie dieser Arbeit anwendbar ist.

Lebenslauf

Persönliche Daten

Name	Uwe Christian Jönck
Geburtsdatum	19. Dezember 1979
Geburtsort	Cuxhaven

Schulabschluss

07/1999	Abitur, Gymnasium am Wall, Verden a.d. Aller

Zivildienst

07/1999 – 07/2000	Zivildienst in einer Tagesbildungsstätte

Studium

10/2000 – 03/2005	Studium der Mathematik an der Universität Hamburg

Wissenschaftliche Tätigkeit

05/2005 – 04/2007	Wissenschaftlicher Mitarbeiter am Department Mathematik der Universität Hamburg, gefördert durch ein DFG-Projekt zur Modellierung und Vorhersage von Finanzzeitreihen
05/2007 – 04/2008	Promotionsstipendium der Universität Hamburg