ASYMPTOTIC EQUIVALENCE FOR NONPARAMETRIC REGRESSION

ION GRAMA AND MICHAEL NUSSBAUM

Université de Bretagne-Sud and Cornell University

ABSTRACT. We consider a nonparametric model $\mathcal{E}^n$, generated by independent observations X_i, $i = 1, ..., n$, with densities $p(x, \theta_i)$, $i = 1, ..., n$, the parameters of which $\theta_i = f(i/n) \in \Theta$ are driven by the values of an unknown function $f : [0, 1] \to \Theta$ in a smoothness class. The main result of the paper is that, under regularity assumptions, this model can be approximated, in the sense of the Le Cam deficiency pseudodistance, by a nonparametric Gaussian shift model $Y_i = \Gamma(f(i/n)) + \varepsilon_i$, where $\varepsilon_1, ..., \varepsilon_n$ are i.i.d. standard normal r.v.'s, the function $\Gamma(\theta) : \Theta \to \mathrm{R}$ satisfies $\Gamma'(\theta) = \sqrt{I(\theta)}$ and $I(\theta)$ is the Fisher information corresponding to the density $p(x, \theta)$.

1. Introduction

Consider the nonparametric regression model with non-Gaussian noise

$$X_i = f(i/n) + \xi_i, \quad i = 1, ..., n, \tag{1.1}$$

where $\xi_1, ..., \xi_n$ are i.i.d. r.v.'s of means 0 and finite variances, with density $p(x)$ on the real line, and $f(t)$, $t \in [0, 1]$, is an unknown real valued function. It is well-known that, under some assumptions, this model shares many desirable asymptotic properties of the Gaussian nonparametric regression model

$$X_i = f(i/n) + \varepsilon_i, \quad i = 1, ..., n, \tag{1.2}$$

where $\varepsilon_1, ..., \varepsilon_n$ are i.i.d. standard normal r.v.'s. In a formal way, two sequences of statistical experiments are said to be *asymptotically equivalent* if the Le Cam pseudodistance between them tends to 0 as $n \to \infty$. Such a relation between the model (1.2) and its continuous time analog was first established by Brown and Low [3]. In a paper by Nussbaum [15] the accompanying Gaussian model for the density estimation from an i.i.d. sample was found to be the white noise model (1.2) with the root of the density as a signal. The case of generalized linear models (i.e. a class of nonparametric regression models with non-additive noise) was treated in Grama and Nussbaum [7]. However none of the above results covers observations of the form (1.1). It is the aim of the present paper to develop an asymptotic equivalence theory for a more general class

Date: October, 2001.

1991 *Mathematics Subject Classification.* Primary 62B15; Secondary 62G08, 62G20.

Key words and phrases. Asymptotic equivalence, deficiency distance, Gaussian approximation, nonparametric model.

The first author was supported by the Alexander von Humboldt-Stiftung. Research of the second author supported by the National Science Foundation under Grant DMS 0072162.

of nonparametric models, in particular for the location type regression model (1.1). In Section 4 we shall derive simple sufficient conditions for the models (1.1) and (1.2) to be asymptotically equivalent; a summary can be given as follows. Let f be in a Hölder ball with exponent $\beta > 1/2$. Let $p(x)$ be the density of the noise variables ξ_i, and set $s(x) = \sqrt{p(x)}$. Assume that the function $s'(x)$ satisfies a Hölder condition with exponent α, where $1/2\beta < \alpha < 1$, and, for some $\delta > \frac{2\beta+1}{2\beta-1}$ and $\varepsilon > 0$,

$$\sup_{|u|\leq\varepsilon} \int_{-\infty}^{\infty} \left| \frac{s'(x+u)}{s(x)} \right|^{2\delta} p(x)dx < \infty.$$

Assume also that the Fisher information in the parametric location model $p(x-\theta)$, $\theta \in \mathrm{R}$ is 1. Then the models (1.1) and (1.2) are asymptotically equivalent. The above conditions follow from the results of the paper for a larger class of nonparametric regression models which we now introduce.

Let $p(x,\theta)$ be a parametric family of densities on the measurable space $(X, \mathcal{X}, \mu)$, where μ is a σ-finite measure, $\theta \in \Theta$ and Θ is an interval (possibly infinite) in the real line. Our nonparametrically driven model is such that at time moments $t_1, ..., t_n$ we observe a sequence of independent X-valued r.v.'s $X_1, ..., X_n$ with densities $p(x, f(t_1)), ..., p(x, f(t_n))$, where $f(t)$, $t \in [0,1]$ is an unknown function and $t_i = i/n$, $i = 1, ..., n$. The principal result of the paper is that, under regularity assumptions on the density $p(x,\theta)$, this model is asymptotically equivalent to a sequence of homoscedastic Gaussian shift models, in which we observe

$$Y_i = \Gamma(f(t_i)) + \varepsilon_i, \quad i = 1, ..., n,$$

where $\varepsilon_1, ..., \varepsilon_n$ are i.i.d. standard normal r.v.'s, $\Gamma(\theta) : \Theta \to \mathrm{R}$ is a function such that $\Gamma'(\theta) = \sqrt{I(\theta)}$ and $I(\theta)$ is the Fisher information in the parametric family $p(x,\theta)$, $\theta \in \Theta$.

The function $\Gamma(\theta)$ can be related to the so called variance stabilizing transformation (see Remark 3.3 below). In the case of the location type regression model (1.1), derived from the family $p(x-\theta)$, $\theta \in \Theta$, we have $\Gamma(\theta) = \theta$. For other nontrivial examples we refer the reader to our Section 4, where it is assumed that the density $p(x,\theta)$ is in a fixed exponential family $\mathcal{E}$ (see also Grama and Nussbaum [7]). Notable among these is the binary regression model (cf. [7]): let X_i be Bernoulli 0-1 r.v.'s with unknown probability of success $\theta_i = f(t_i)$, $i = 1, ..., n$, where f is in a Hölder ball with exponent $\beta > 1/2$ and is separated form 0 and 1. Then the accompanying Gaussian model is

$$Y_i = 2 \arcsin \sqrt{f(t_i)} + \varepsilon_i, \quad i = 1, ..., n.$$

The function $\Gamma(\theta) = 2 \arcsin \sqrt{\theta}$ is known to be the variance-stabilizing transformation related to the Bernoulli random variables.

The global result above is derived from the following local result. Define a local experiment to be generated by independent observations $X_1, ..., X_n$ with densities $p(x, g(t_1)), ..., p(x, g(t_n))$, where $g(t)$ is in a certain neighborhood of "nonparametric size" (i.e. whose radius is large relative to $n^{-1/2}$) of a fixed function f. We show that this model is asymptotically equivalent to a heteroscedastic Gaussian model, in which we observe

$$Y_i = g(t_i) + I(f(t_i))^{-1/2}\varepsilon_i, \quad i = 1, ..., n. \tag{1.3}$$

As an example, for binary regression (Bernoulli observations) with probabilities of success $\theta_i = f(t_i)$, $i = 1, ..., n$ we obtain the local Gaussian approximation (1.3) with $I(\theta) = \frac{1}{\theta(1-\theta)}$.

In turn, the local approximation (1.3) is a consequence of a more general local result for nonparametric models satisfying some regularity assumptions, which is of independent interest. Namely, we show that if the log-likelihood of a nonparametric experiment satisfies a certain asymptotic expansion in terms of independent random variables, subject to a Lindeberg type condition, and with a remainder term converging to 0 at some rate, then a version of this experiment can be constructed on the same measurable space with a Gaussian experiment such that the Hellinger distance between the corresponding probability measures converges to 0 at some rate. The proof of the last result is based on obtaining versions of the likelihood processes on a common probability space by means of a functional Hungarian construction for partial sums, established in Grama and Nussbaum [8].

The abovementioned results are part of an effort to extend Le Cam's asymptotic theory (see Le Cam [12] and Le Cam and Yang [13]) to a class of general models with infinite dimensional parameters which *cannot be estimated* at the usual "root-n" rate $n^{-1/2}$. The case of the infinite dimensional parameters which are estimable at this rate was considered for instance in Millar [14]. The approach used in the proofs of the present results is quite different from that in the "root-n" case and was suggested by the papers of Brown and Low [3] and Nussbaum [15] (see also Grama and Nussbaum [7]). An overview of the technique of proof can be found at the end of the Section 3.

A nonparametric regression model with random design, but Gaussian noise was recently treated by Brown, Low and Zhang [5]. We focus here on the nongaussian case, assuming a regular nonrandom design i/n, $i = 1, \ldots, n$: the model is generated by a parametric family of densities $p(x, \theta)$, $\theta \in \Theta$, where θ is assumed to take the values $f(i/n)$ of the regression function f. The term *nonparametrically driven parametric model* shall also be used for this setup.

The paper is organized as follows. Section 2 contains some background on asymptotic equivalence. Our main results are presented in Section 3. In Section 4 we illustrate the scope of our regularity assumptions by considering the case of the location type regression model and the exponential family model (known also as the generalized linear model). In Section 5 we prove our basic local result on asymptotic equivalence for a general class of nonparametric experiments. Then in Section 6 we just apply this general local result to the particular case when the nonparametrically driven parametric model satisfies the regularity assumptions of the Section 3. In Sections 7 and 8 we globalize the previous local result twofold: first over the time interval $[0, 1]$, and then over the parameter set Σ^β. The global form over Σ^β requires a reparametrization of the family $p(x, \theta)$ using $\Gamma(\theta)$. At the end of Section 7, we show that this allows a homoscedastic form of the accompanying local Gaussian experiment, which can readily be globalized. Finally in the Appendix we formulate the functional Komlós-Major-Tusnády approximation proved in [8] and some well-known auxiliary statements.

2. Background on asymptotic equivalence

We follow Le Cam [12] and Le Cam and Yang [13]. Let $\mathcal{E} = (X, \mathcal{X}, \{P_\theta : \theta \in \Theta\})$ and $\mathcal{G} = (Y, \mathcal{Y}, \{Q_\theta : \theta \in \Theta\})$ be two experiments with the same parameter set Θ. Assume that $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ are complete separable (Polish) metric spaces. The deficiency of the experiment $\mathcal{E}$ with respect to $\mathcal{G}$ is defined as

$$\delta(\mathcal{E}, \mathcal{G}) = \inf_K \sup_{\theta \in \Theta} \|K \cdot P_\theta - Q_\theta\|_{\mathrm{Var}},$$

where the infimum is taken over the set of all Markov kernels $K : (X, \mathcal{X}) \to (Y, \mathcal{Y})$ and $\|\cdot\|_{\mathrm{Var}}$ is the total variation norm for measures. Le Cam's distance between $\mathcal{E}$ and $\mathcal{G}$ is defined to be

$$\Delta(\mathcal{E}, \mathcal{G}) = \max\{\delta(\mathcal{E}, \mathcal{G}), \delta(\mathcal{G}, \mathcal{E})\}.$$

An equivalent definition of the Le Cam distance is obtained if we define the one sided deficiency $\delta(\mathcal{E}, \mathcal{G})$ as follows: let $(D, \mathcal{D})$ be a space of possible decisions. Denote by $\Pi(\mathcal{E})$ the set of randomized decision procedures in the experiment $\mathcal{E}$, i.e. the set of Markov kernels $\pi(x, A) : (X, \mathcal{X}) \to (D, \mathcal{D})$. Define $\mathcal{L}(D, \mathcal{D})$ to be the set of all loss functions $L(\theta, z) : \Theta \times D \to [0, \infty)$ satisfying $0 \le L(\theta, z) \le 1$. The risk of the procedure $\pi \in \Pi(\mathcal{E})$ for the loss function $L \in \mathcal{L}(D, \mathcal{D})$ and true value $\theta \in \Theta$ is set to be

$$R(\mathcal{E}, \pi, L, \theta) = \int_X \int_D L(\theta, z) \pi(x, dz) P_\theta(dx).$$

Then the one-sided deficiency can be defined as

$$\delta(\mathcal{E}, \mathcal{G}) = \sup \sup_{L \in \mathcal{L}(D,\mathcal{D})} \inf_{\pi_1 \in \Pi(\mathcal{E})} \sup_{\pi_2 \in \Pi(\mathcal{G})} \sup_{\theta \in \Theta} |R(\mathcal{E}, \pi_1, L, \theta) - R(\mathcal{E}, \pi_2, L, \theta)|.$$

where the first supremum is over all possible decision spaces $(D, \mathcal{D})$.

Following Le Cam [12], we introduce the next definition.

Definition 2.1. Two sequences of statistical experiments $\mathcal{E}^n$, $n = 1, 2, \ldots$ and $\mathcal{G}^n$, $n = 1, 2, \ldots$ are said to be *asymptotically equivalent* if

$$\Delta(\mathcal{E}^n, \mathcal{G}^n) \to 0, \quad \text{as} \quad n \to \infty,$$

where $\Delta(\mathcal{E}^n, \mathcal{G}^n)$ is the Le Cam deficiency pseudo-distance between statistical experiments $\mathcal{E}^n$ and $\mathcal{G}^n$.

From the above definitions it follows that, if in the sequence of models $\mathcal{E}^n$ there is a sequence of procedures π_1^n such that the risks $R(\mathcal{E}^n, \pi_1^n, L_n, \theta)$ converge to the quantity $\rho(\theta)$, for a uniformy bounded sequence of loss functions L_n, then there is a sequence of procedures π_2^n in $\mathcal{G}^n$, such that the risks $R(\mathcal{G}^n, \pi_2^n, L_n, \theta)$ converge to the same quantity $\rho(\theta)$, uniformly in $\theta \in \Theta$. This indicates that the asymptotically minimax risks for bounded loss functions in one model can be transferred to another model. In particular one can compute the asymptotically minimax risk in non-Gaussian models by computing it in the accompanying Gaussian models. This task, however, remains beyond of the scope of the present paper.

3. Formulation of the results

3.1. The parametric model. Let Θ be an interval (possibly infinite) in the real line R and

$$\mathcal{E} = (X, \mathcal{X}, \{P_\theta : \theta \in \Theta\}) \tag{3.1}$$

be a statistical experiment on the measurable space $(X, \mathcal{X})$ with the parameter set Θ and a dominating σ-finite measure μ. The last property means that for each $\theta \in \Theta$ the measure P_θ is absolutely continuous w.r.t. the measure μ. Denote by $p(x, \theta)$ the Radon-Nikodym derivative of P_θ w.r.t. μ :

$$p(x, \theta) = \frac{P_\theta(dx)}{\mu(dx)}, \quad x \in X, \quad \theta \in \Theta. \tag{3.2}$$

For the sake of brevity we set $p(\theta) = p(\cdot, \theta)$. We shall assume in the sequel that for any $\theta \in \Theta$

$$p(\theta) > 0, \quad \mu\text{-a.s. on } X. \tag{3.3}$$

Assumption (3.3) implies that the measures P_θ, $\theta \in \Theta$ are equivalent: $P_\theta \sim P_u$, for $\theta, u \in \Theta$.

3.2. The nonparametric model. Set $T = [0, 1]$. Let $\mathcal{F}_0 = \Theta^T$ be the set of all functions on the unit interval $T = [0, 1]$ with values in the interval Θ :

$$\mathcal{F}_0 = \Theta^T = \{f : [0, 1] \to \Theta\}.$$

Let $\mathcal{H}(\beta, L)$ be a Hölder ball of functions defined on T and with values in R, i.e. the set of functions $f : T \to \mathrm{R}$, which satisfy a Hölder condition with exponent $\beta > 0$ and constant $L > 0$:

$$\left|f^{(\beta_0)}(t) - f^{(\beta_0)}(s)\right| \le L\,|t - s|^{\beta_1}, \quad t, s \in T,$$

where the nonnegative integer β_0 and the real $\beta_1 \in (0, 1]$ are such that $\beta_0 + \beta_1 = \beta$, and which also satisfy

$$\|f\|_\infty \le L, \text{ where } \|f\|_\infty = \sup_{t \in T} |f(t)|.$$

Set for brevity, $\Sigma^\beta = \mathcal{F}_0 \cap \mathcal{H}(\beta, L)$. In the nonparametrically driven model to be treated here it is assumed that we observe a sequence of independent r.v.'s $X_1, ..., X_n$ with values in the measurable space $(X, \mathcal{X})$, such that, for each $i = 1, ..., n$, the observation X_i has density $p(x, f(i/n))$ where the function f is unknown and satisfies the smoothness condition $f \in \Sigma^\beta$. We shall make use of the notation $P_f^n = P_{f(1/n)} \times ... \times P_{f(n/n)}$, where P_θ is the probability measure in the parametric experiment $\mathcal{E}$ and $f \in \Sigma^\beta$.

3.3. **Regularity assumptions.** Assume that $\beta > 1/2$. In the sequel the density $p(x,\theta)$ in the parametric experiment $\mathcal{E}$ shall be subjected to the regularity assumptions $(R1)$, $(R2)$, $(R3)$, which are assumed to hold true with the same $\varepsilon > 0$.

R1: The function $s(\theta) = \sqrt{p(\theta)}$ is smooth in the space $L^2(X, \mathcal{X}, \mu)$: there is a real number $\delta \in (\frac{1}{2\beta}, 1)$ and a map $\overset{\bullet}{s}(\theta) : \Theta \to L^2(X, \mathcal{X}, \mu)$ such that

$$\sup_{(\theta,u)} \frac{1}{|u-\theta|^{1+\delta}} \left(\int_X \left(s(x,u) - s(x,\theta) - (u-\theta)\overset{\bullet}{s}(x,\theta) \right)^2 \mu(dx) \right)^{1/2} < \infty,$$

where sup is taken over all pairs (θ,u) satisfying $\theta, u \in \Theta$, $|u-\theta| \le \varepsilon$.

It is well-known (see Strasser [17]) that there is a map $\overset{\bullet}{l}(\theta) \in L^2(X, \mathcal{X}, \mu)$ such that the function $\overset{\bullet}{s}(\theta)$ in condition $(R1)$ can be written as

$$\overset{\bullet}{s}(\theta) = \frac{1}{2}\overset{\bullet}{l}(\theta)\sqrt{p(\theta)}, \quad \mu\text{-a.s. on } X. \tag{3.4}$$

Moreover, $\overset{\bullet}{l}(\theta) \in L^2(X, \mathcal{X}, P_\theta)$ and

$$\mathrm{E}_\theta \overset{\bullet}{l}(\theta) = \int_X \overset{\bullet}{l}(x,\theta)\, p(x,\theta)\, \mu(dx) = 0, \quad \theta \in \Theta,$$

where E_θ is the expectation under P_θ. The map $\overset{\bullet}{l}(\theta)$ is called *tangent vector* at θ. For any $\theta \in \Theta$, define an *extended tangent vector* $\overset{\bullet}{l}_\theta(u)$, $u \in \Theta$, by setting

$$\overset{\bullet}{l}_\theta(x,u) = \begin{cases} \overset{\bullet}{l}(x,\theta), & \text{if } u = \theta, \\ \frac{2}{u-\theta}\left(\sqrt{\frac{p(x,u)}{p(x,\theta)}} - 1\right), & \text{if } u \ne \theta. \end{cases} \tag{3.5}$$

R2: There is a real number $\delta \in (\frac{2\beta+1}{2\beta-1}, \infty)$ such that

$$\sup_{(\theta,u)} \int_X \left| \overset{\bullet}{l}_\theta(x,u) \right|^{2\delta} p(x,\theta)\, \mu(dx) < \infty,$$

where sup is taken over all pairs (θ,u) satisfying $\theta, u \in \Theta$, $|u-\theta| \le \varepsilon$.

The Fisher information in the local experiment $\mathcal{E}$ is

$$I(\theta) = \int_X \left(\overset{\bullet}{l}(x,\theta) \right)^2 p(x,\theta)\, \mu(dx), \quad \theta \in \Theta. \tag{3.6}$$

R3: There are two real numbers $I_{\min}$ and $I_{\max}$ such that

$$0 < I_{\min} \le I(\theta) \le I_{\max} < \infty, \quad \theta \in \Theta.$$

3.4. **Local result.** First we state a local Gaussian approximation. For any $f \in \Sigma^\beta$, denote by $\Sigma_f^\beta(r)$ the neighborhood of f, shifted to the origin:

$$\Sigma_f^\beta(r) = \left\{ h : \|h\|_\infty \le r,\ f + h \in \Sigma^\beta \right\}.$$

Set

$$\overline{\gamma}_n = c(\beta) \left(\frac{\log n}{n} \right)^{\frac{\beta}{2\beta+1}},$$

where $c(\beta)$ is a constant depending on β. Throughout the paper γ_n will denote a sequence of real numbers satisfying, for some constant $c \ge 0$,

$$\overline{\gamma}_n \le \gamma_n = O(\overline{\gamma}_n \log^c n). \tag{3.7}$$

By definition the *local experiment*

$$\mathcal{E}_f^n = \left(X^n, \mathcal{X}^n, \{ P_{f+h}^n : h \in \Sigma_f^\beta(\gamma_n) \} \right)$$

is generated by the sequence of independent r.v.'s $X_1, ..., X_n$, with values in the measurable space $(X, \mathcal{X})$, such that for each $i = 1, ..., n$, the observation X_i has density $p(x, g(i/n))$, where $g = f + h$, $h \in \Sigma_f^\beta(\gamma_n)$. An equivalent definition is:

$$\mathcal{E}_f^n = \mathcal{E}_f^{(1)} \otimes ... \otimes \mathcal{E}_f^{(n)}, \tag{3.8}$$

where

$$\mathcal{E}_f^{(i)} = \left(X, \mathcal{X}, \{ P_{g(i/n)} : g = f + h,\ h \in \Sigma_f^\beta(\gamma_n) \} \right), \quad i = 1, ..., n. \tag{3.9}$$

Theorem 3.1. *Let $\beta > 1/2$ and $I(\theta)$ be the Fisher information in the parametric experiment $\mathcal{E}$. Assume that the density $p(x, \theta)$ satisfies the regularity conditions $(R1 - R3)$. For any $f \in \Sigma^\beta$, let $\mathcal{G}_f^n$ be the local Gaussian experiment, generated by observations*

$$Y_i^n = h(i/n) + I(f(i/n))^{-1/2} \varepsilon_i, \quad i = 1, ..., n,$$

with $h \in \Sigma_f^\beta(r_n)$, where $\varepsilon_1, ..., \varepsilon_n$ is a sequence of i.i.d. standard normal r.v.'s. Then, uniformly in $f \in \Sigma^\beta$, the sequence of local experiments $\mathcal{E}_f^n$, $n = 1, 2, ...$ is asymptotically equivalent to the sequence of local Gaussian experiments $\mathcal{G}_f^n$, $n = 1, 2, ...$:

$$\sup_{f \in \Sigma^\beta} \Delta\left(\mathcal{E}_f^n, \mathcal{G}_f^n \right) \to 0, \quad \text{as} \quad n \to \infty.$$

3.5. **Global result.** By definition the *global experiment*

$$\mathcal{E}^n = \left(X^n, \mathcal{X}^n, \{ P_f^n : f \in \Sigma^\beta \} \right)$$

is generated by the sequence of independent r.v.'s $X_1, ..., X_n$, with values in the measurable space $(X, \mathcal{X})$, such that for each $i = 1, ..., n$, the observation X_i has density $p(x, f(i/n))$, where $f \in \Sigma^\beta$. In other words $\mathcal{E}^n$ is the product experiment

$$\mathcal{E}^n = \mathcal{E}^{(1)} \otimes ... \otimes \mathcal{E}^{(n)},$$

where

$$\mathcal{E}^{(i)} = \left(X, \mathcal{X}, \{P_{f(i/n)} : f \in \Sigma^\beta\}\right), \quad i = 1, ..., n.$$

We shall make the following assumptions:

G1: For any $\beta > 1/2$, there is an estimator $\widehat{f}_n : X^n \to \Sigma^\beta$, such that

$$\sup_{f \in \Sigma^\beta} P\left(\left\|\widehat{f}_n - f\right\|_\infty \geq \gamma_n\right) \to 0, \quad n \to \infty.$$

G2: The Fisher information $I(\theta) : \Theta \to (0, \infty)$ satisfies a Hölder condition with exponent $\alpha \in (1/2\beta, 1)$.

The main result of the paper is the following theorem, which states a global Gaussian approximation for the sequence of experiments $\mathcal{E}^n$, $n = 1, 2, ...$ in the sense of Definition 2.1.

Theorem 3.2. *Let $\beta > 1/2$ and $I(\theta)$ be the Fisher information in the parametric experiment $\mathcal{E}$. Assume that the density $p(x, \theta)$ satisfies the regularity conditions (R1-R3) and that conditions (G1-G2) are fulfilled. Let $\mathcal{G}^n$ be the Gaussian experiment generated by observations*

$$Y_i^n = \Gamma\left(f\left(i/n\right)\right) + \varepsilon_i, \quad i = 1, ..., n,$$

with $f \in \Sigma^\beta$, where $\varepsilon_1, ..., \varepsilon_n$ is a sequence of i.i.d. standard normal r.v.'s and $\Gamma(\theta) : \Theta \to \mathrm{R}$ is any function satisfying $\Gamma'(\theta) = \sqrt{I(\theta)}$. Then the sequence of experiments $\mathcal{E}^n$, $n = 1, 2, ...$ is asymptotically equivalent to the sequence of Gaussian experiments $\mathcal{G}^n$, $n = 1, 2, ...$:

$$\Delta\left(\mathcal{E}^n, \mathcal{G}^n\right) \to 0, \quad as \quad n \to \infty.$$

Remark 3.1. Examples in Efromovich and Samarov [6], Brown and Zhang [4] [see also Brown and Low [3]] show that asymptotic equivalence, in general, fails to hold true when $\beta \leq 1/2$.

Remark 3.2. Assumption (G1) is related to attainable rates of convergence in the sup-norm $\|\cdot\|_\infty$ for nonparametric regression. It is well known that for a parameter space $\mathcal{H}(\beta, L)$, $\beta \leq 1$ the attainable rate for estimators of f is $(\log n/n)^{\beta/(2\beta+1)}$ in regular cases (cf. Stone [16]). For a choice $\gamma_n = \bar{\gamma}_n \log^c n$ condition (G1) would be implied by this type of result. However for a choice $\gamma_n = \bar{\gamma}_n = c(\beta)\left((\log n/n)^{\beta/(2\beta+1)}\right)$ (G1) is slightly stronger (the classical rate result would also require $c(\beta) \to \infty$ for convergence to 0 in (G1)). In the case of the Gaussian location-type regression ((1.1) for normal ξ_i) this is a consequence of the optimal constant result of Korostelev [10]. The extension to our nongaussian regression models would be of technical nature; for the density estimation model it has been verified in Korostelev and Nussbaum [11] and applied in a similar context to here in Lemma 9.3 of [15] .

Remark 3.3. The function $\Gamma(\theta)$ can be related to so called *variance-stabilizing* transformation, which we proceed to introduce. Let $X_1, ..., X_n$ be a sequence of real valued i.i.d. r.v.'s, with law depending on the parameter $\theta \in \Theta$. Let $\mu(\theta)$ be the common mean and $\sigma(\theta)$ be the common variance. By the central limit theorem,

$$\sqrt{n}\left\{S_n - \mu(\theta)\right\} \xrightarrow{d} N\left(0, \sigma\left(\theta\right)\right),$$

where $S_n = \left(X_1 + ... + X_n\right)/n$. The variance-stabilizing transformation is defined to be a function F on the real line, with the property that the variance of the limiting normal law does not depend on θ, i.e.

$$\sqrt{n}\left\{F\left(S_n\right) - F(\mu(\theta))\right\} \xrightarrow{d} N\left(0, 1\right).$$

The function F pertaining to our nonparametric model can be computed and the relation to the function Γ can be clarified in some particular cases. Let $\sigma(\theta) = I(\theta) > 0$ and $\mu\left(\theta\right)$ satisfy $\mu'\left(\theta\right) = I\left(\theta\right).$ Let $a\left(\lambda\right)$ be the inverse of the strictly increasing function $\mu\left(\theta\right)$ on the interval Θ and $\mu\left(\Theta\right)$ be its range. Assume that the X_i take values in $\mu(\Theta)$. One can easily see that a variance stabilizing transformation is any function $F(\lambda)$, $\lambda \in \mu(\Theta)$ satisfying $F'\left(\lambda\right) = 1/\sqrt{I\left(a\left(\lambda\right)\right)}$, $\lambda \in \mu(\Theta)$. In this case, our transformation of the functional parameter f is actually the transformation of the mean $\mu(\theta)$

$$\Gamma(\theta) = F(\mu(\theta)),$$

corresponding to this variance stabilizing transformation. In the other particular case, when the mean of X_i is $\mu(\theta) = \theta$, both transformations Γ and F just coincide: $\Gamma(\theta) = F(\theta)$.

We follow the line developed in Nussbaum [15] (see also Grama and Nussbaum [7]). The proof of Theorem 3.1 is given in Section 7 and contains three steps:

- Decompose the local experiments $\mathcal{E}_f^n$ and $\mathcal{G}_f^n$ into products of independent experiments:

$$\mathcal{E}_f^n = \mathcal{E}_f^{n,1} \otimes ... \otimes \mathcal{E}_f^{n,m}, \quad \mathcal{E}_f^n = \mathcal{G}_f^{n,1} \otimes ... \otimes \mathcal{G}_f^{n,m}.$$

  Here $m = o(n)$ and $\mathcal{E}_f^{n,k}$ represents the k-th "block" of observations of size (approximately) n/m.
- Show that each component $\mathcal{E}_f^{n,k}$ can be well approximated by its Gaussian counterpart $\mathcal{G}_f^{n,k}$ in the sense that there exist equivalent versions $\widetilde{\mathcal{E}}_f^{n,k}$ and $\widetilde{\mathcal{G}}_f^{n,k}$ on a common measurable space such that

$$H^2\left(\widetilde{P}_{f,h}^{n,k}, \widetilde{Q}_{f,h}^{n,k}\right) = o\left(m^{-1}\right),$$

  where $H^2\left(\cdot, \cdot\right)$ is the Hellinger distance between the probability measures $\widetilde{P}_{f,h}^{n,k}$ and $\widetilde{Q}_{f,h}^{n,k}$ in the local experiments $\widetilde{\mathcal{E}}_f^{n,k}$ and $\widetilde{\mathcal{G}}_f^{n,k}$ respectively.
- Patch together the components $\widetilde{\mathcal{E}}_f^{n,k}$ and $\widetilde{\mathcal{G}}_f^{n,k}$, $k = 1, ..., m$ by means of the convenient property of the Hellinger distance for the product of probability measures

$$H^2\left(\widetilde{P}_{f,h}^{n}, \widetilde{Q}_{f,h}^{n}\right) \leq \sum_{k=1}^{m} H^2\left(\widetilde{P}_{f,h}^{n,k}, \widetilde{Q}_{f,h}^{n,k}\right) = \sum_{i=1}^{m} o\left(m^{-1}\right) = o\left(1\right).$$

The challenge here is the second step. For its proof, in Section 5, we develop a general approach, according to which any experiment $\mathcal{E}^n$ with a certain asymptotic expansion of its log-likelihood ratio (condition LASE) can be constructed on the same measurable space with a Gaussian experiment $\mathcal{G}^n$, such that the Hellinger distance between the corresponding probability measures converges to 0 at a certain rate. Then we are able to check condition LASE for the model under consideration using a strong approximation result (see Theorem 9.1).

Theorem 3.2 is derived from the local result of Theorem 3.1 by means of a globalizing procedure, which is presented in Section 8.

4. Examples

4.1. **Location type regression model.** Consider the regression model with non-Gaussian additive noise

$$X_i = f(i/n) + \xi_i, \quad i = 1, ..., n, \tag{4.1}$$

where $\xi_1, ..., \xi_n$ are i.i.d. r.v.'s of means 0 and finite variances, with density $p(x)$ on the real line, $f \in \Sigma^\beta$ and Σ^β is a Hölder ball on $[0,1]$ with exponent $\beta > 1/2$. This model is a particular case of the nonparametrically driven model, introduced in Section 3, when $p(x,\theta) = p(x-\theta)$ is a shift parameter family and $\theta \in \mathrm{R}$. Assume that the derivative $p'(x)$ exists, for all $x \in \mathrm{R}$. Then the extended tangent vector, defined by (3.4), is computed as follows

$$\dot{l}_\theta(x,u) = \begin{cases} \frac{p'(x-\theta)}{p(x-\theta)}, & \text{if} \quad u = \theta, \\ \frac{2}{u-\theta}\frac{\sqrt{p(x-u)}-\sqrt{p(x-\theta)}}{\sqrt{p(x-\theta)}}, & \text{if} \quad u \neq \theta. \end{cases}$$

Set $s(x) = \sqrt{p(x)}$. Then it is easy to see that conditions (R1-R3) hold true, if we assume that the following items are satisfied:

L1: The function $s'(x)$ satisfies Hölder's condition with exponent α, where $\alpha \in (1/2\beta, 1)$, i.e.

$$|s'(x) - s'(y)| \leq C\,|x-y|^\alpha, \quad x, y \in \mathrm{R}.$$

L2: For some $\delta > \frac{2\beta+1}{2\beta-1}$ and $\varepsilon > 0$,

$$\sup_{|u|\leq\varepsilon} \int_{-\infty}^{\infty} \left| \frac{s'(x+u)}{s(x)} \right|^{2\delta} p(x)dx < \infty.$$

L3: The Fisher informational number is positive:

$$I = 4\int_{-\infty}^{\infty} s'(x)^2 dx = \int_{-\infty}^{\infty} \frac{p'(x)^2}{p(x)} dx > 0.$$

It is well-known that a preliminary estimator satisfying condition (G1) exists in this model. Then, according to Theorem 3.2, under conditions (L1-L3) the model defined by observations (4.1) is asymptotically equivalent to a linear regression with Gaussian noise, in which we observe

$$Y_i = f(i/n) + I^{-1/2}\varepsilon_i,$$

where $\varepsilon_1, ..., \varepsilon_n$ are i.i.d. standard normal r.v.'s.

4.2. **Exponential family model.** Another particular case of the model introduced in Section 3 arises when the parametric experiment $\mathcal{E} = (X, \mathcal{X}, \{P_\theta : \theta \in \Theta\})$ is an one-dimensional linearly indexed exponential family, where Θ is a possibly infinite interval on the real line (see Brown [2]). This means that the measures P_θ are absolutely continuous w.r.t. a σ-finite measure $\mu(dx)$ with densities (in the canonical form)

$$p(x,\theta) = \frac{P_\theta(dx)}{\mu(dx)} = \exp\left(\theta U(x) - V(\theta)\right), \quad x \in X, \quad \theta \in \Theta, \tag{4.2}$$

where the measurable function $U(x) : X \to \mathrm{R}$ is a sufficient statistic in the experiment $\mathcal{E}$ and

$$V(\theta) = \log \int_X \exp\left(\theta U(x)\right) \mu(dx)$$

is the logarithm of the Laplace transformation of $U(x)$. It is easy to see that regularity conditions (R1-R3) and (G1-G2) are satisfied, if we only assume that $0 < c_1 \le V''(\theta) \le c_2 < \infty$ and $\left|V^{(k)}(\theta)\right| \le c_2$, for any $k \ge 3$ and two absolute constants c_1 and c_2 (for other related conditions see Grama and Nussbaum [7]). We point out that, for all these models, the preliminary estimator of condition (G1) above can easily be constructed (see for instance [7]).

Now we shall briefly discuss some examples. Note that the parametrizations in these is different from the canonical one appearing in (4.2). We have chosen the natural parametrizations, where an observation X in the experiment $\mathcal{E}$ has mean $\mu(\theta) = \theta$, since this facilitates computation of the function $\Gamma(\theta)$.

Gaussian scale model. Assume that we are given a sequence of normal observations $X_1, ..., X_n$ with mean 0 and variance $f(i/n)$, where the function $f(t)$, $t \in [0,1]$ satisfies a Hölder condition with exponent $\beta > 1/2$ and is such that $c_1 \le f(t) \le c_2$, for some positive absolute constants c_1 and c_2. In this model the density of the observations has the form $p(x,\theta) = \frac{1}{\sqrt{2\pi}\theta} \exp\left(-\frac{x^2}{2\theta^2}\right)$, $x \in \mathrm{R}$ and the Fisher information is $I(\theta) = 2\theta^{-2}$. This gives us $\Gamma(\theta) = \sqrt{2}\log\theta$. Then, by Theorem 3.2, the model is asymptotically equivalent to the Gaussian model

$$Y_i = \sqrt{2}\log f(i/n) + \varepsilon_i,$$

where $\varepsilon_1, ..., \varepsilon_n$ are i.i.d. standard normal r.v.'s.

Poisson model. Assume that we are given a sequence of Poisson observations $X_1, ..., X_n$ with parameters $f(i/n)$, where the function $f(t)$, $t \in [0,1]$ satisfies a Hölder condition with exponent $\beta > 1/2$ and is such that $c_1 \le f(t) \le c_2$, for some positive absolute constants c_1 and c_2. In this model $p(x,\theta) = \theta^x \exp(-\theta)$, $x \in X = \{0, 1, ...\}$ and $I(\theta) = \theta^{-1}$.

As a consequence $\Gamma(\theta) = 2\sqrt{\theta}$. According to Theorem 3.2, the model is asymptotically equivalent to the Gaussian model

$$Y_i = 2\sqrt{f(i/n)} + \varepsilon_i,$$

where $\varepsilon_1, ..., \varepsilon_n$ are i.i.d. standard normal r.v.'s.

Binary response model. Assume that we are given a sequence of Bernoulli observations $X_1, ..., X_n$ taking values 0 and 1 with probabilities $1 - f(i/n)$ and $f(i/n)$ respectively, where the function $f(t)$, $t \in [0, 1]$ satisfies a Hölder condition with exponent $\beta > 1/2$ and is such that $c_1 \leq f(t) \leq c_2$ for some absolute constants $c_1 > 0$ and $c_2 < 1$. In this model $p(x, \theta) = \theta^x (1 - \theta)^{1-x}$, $x \in X = \{0, 1\}$ and $I(\theta) = \frac{1}{\theta(1-\theta)}$. This yields $\Gamma(\theta) = 2 \arcsin \sqrt{\theta}$. By Theorem 3.2, this model is asymptotically equivalent to the Gaussian model

$$Y_i = 2 \arcsin \sqrt{f(i/n)} + \varepsilon_i,$$

where $\varepsilon_1, ..., \varepsilon_n$ are i.i.d. standard normal r.v.'s.

5. Some nonparametric local asymptotic theory

The aim of this section is to state and discuss a condition on the asymptotic behaviour of the likelihood ratio which can be used for Gaussian approximation of our nonparametric regression models. This condition will resemble an LAN-condition, but will be stronger in the sense that it requires an asymptotic expansion of the log-likelihood ratio in terms of independent random variables. Although LAN conditions for nonparametric experiments have been stated and their consequences been developed extensively, (see for instance Millar [14], Strasser [17] and references therein), the context of this was $n^{-1/2}$-consistent estimation problems, where an $n^{-1/2}$-localization of the experiments is useful and implies weak convergence of the sequence $\mathcal{E}^n$ to a limit Gaussian experiment $\mathcal{G}$. In contrast, we assert the existence of a version of the experiment $\mathcal{E}^n$ on the same sample space with a suitable Gaussian experiment $\mathcal{G}^n$ such that the Hellinger distance between corresponding measures goes to 0 at some rate. The existence of a limit experiment $\mathcal{G}$ is not assumed here: it is replaced by a suitable sequence of approximating experiments $\mathcal{G}^n$.

It should be mentioned that the scope of applicability of this theory is actually larger than exploited here: it can be used to establish asymptotic equivalence results for regression models with i.i.d. observations, such as models with random design, and for models with dependent observations. However these results remain beyond of the scope of the present paper. Here we restrict ourselves to the model introduced in Section 3.

5.1. A bound for the Hellinger distance.

First we shall give sufficient conditions for a rate of convergence to 0 of the Hellinger distance between the corresponding measures of two statistical experiments.

Assume that $\mathcal{F}$ is an arbitrary set and that for any $f \in \mathcal{F}$ we are given a system of sets $\mathcal{F}_f(r)$, $r > 0$, to be regarded as a system of neighborhoods of f. Let r_n, $n = 1, 2, ...$

be a sequence of real numbers satisfying $0 < r_n \leq 1$ and $r_n \to 0$, as $n \to \infty$. Let, for any $n = 1, 2, ...$ and any $f \in \mathcal{F}$,

$$\mathcal{E}_f^n = \left(X^n, \mathcal{X}^n, \left\{ P_{f,h}^n : h \in \mathcal{F}_f \left(r_n \right) \right\} \right)$$

and

$$\mathcal{G}_f^n = \left(X^n, \mathcal{X}^n, \left\{ Q_{f,h}^n : h \in \mathcal{F}_f \left(r_n \right) \right\} \right)$$

be two statistical experiments with the same sample space $(X^n, \mathcal{X}^n)$ and the same parameter space $\mathcal{F}_f \left(r_n \right)$. Assume that the "local" experiments $\mathcal{E}_f^n$ and $\mathcal{G}_f^n$ have a common "central" measure, i.e. that there is an element $h_0 = h_0 \left(f \right) \in \mathcal{F}_f \left(r_n \right)$ such that $P_{f,h_0}^n = Q_{f,h_0}^n = \mathbf{P}_f^n$ and $P_{f,h}^n \ll P_{f,h_0}^n$, $Q_{f,h}^n \ll Q_{f,h_0}^n$, for any $h \in \mathcal{F}_f \left(r_n \right)$.

Theorem 5.1. *Let $\alpha_1 > \alpha \geq 0$. Assume that, for some $c_1 > 0$,*

$$\sup_{f \in \mathcal{F}} \sup_{h \in \mathcal{F}_f(r_n)} \mathbf{P}_f^n \left(\left| \log \frac{dP_{f,h}^n}{dP_{f,h_0}^n} - \log \frac{dQ_{f,h}^n}{dQ_{f,h_0}^n} \right| \geq c_1 r_n^{\alpha_1} \right) = O\left(r_n^{2\alpha_1} \right) \tag{5.1}$$

and, for any $\varepsilon \in (0, 1)$,

$$\sup_{f \in \mathcal{F}} \sup_{h \in \mathcal{F}_f(r_n)} P_{f,h}^n \left(\log \frac{dP_{f,h}^n}{dP_{f,h_0}^n} > -\varepsilon \log r_n \right) = O\left(r_n^{2\alpha_1} \right) \tag{5.2}$$

and

$$\sup_{f \in \mathcal{F}} \sup_{h \in \mathcal{F}_f(r_n)} Q_{f,h}^n \left(\log \frac{dQ_{f,h}^n}{dQ_{f,h_0}^n} > -\varepsilon \log r_n \right) = O\left(r_n^{2\alpha_1} \right). \tag{5.3}$$

Then there is an $\alpha_2 > \alpha$ such that

$$\sup_{f \in \mathcal{F}} \sup_{h \in \mathcal{F}_f(r_n)} H^2 \left(P_{f,h}^n, Q_{f,h}^n \right) = O\left(r_n^{2\alpha_2} \right).$$

Proof. Set, for brevity

$$L_{f,h}^{1,n} = \frac{dP_{f,h}^n}{dP_{f,h_0}^n}, \qquad L_{f,h}^{2,n} = \frac{dQ_{f,h}^n}{dQ_{f,h_0}^n}, \qquad \Psi_n = \sqrt{L_{f,h}^{1,n}} - \sqrt{L_{f,h}^{2,n}}.$$

Consider the set

$$A_n = \left\{ \left| \log L_{f,h}^{1,n} - \log L_{f,h}^{1,n} \right| \leq r_n^{\alpha_1} \right\}.$$

With these notations, by the definition of the Hellinger distance [see (9.2)], we have

$$H^2 \left(P_{f,h}^n, Q_{f,h}^n \right) = \frac{1}{2} \mathbf{E}_f^n \Psi_n^2 = \frac{1}{2} \mathbf{E}_f^n \mathbf{1}_{A_n} \Psi_n^2 + \frac{1}{2} \mathbf{E}_f^n \mathbf{1}_{A_n^c} \Psi_n^2. \tag{5.4}$$

Changing measure in the first expectation in the right hand side of (5.4), we write

$$
\begin{aligned}
\mathbf{E}_f^n \mathbf{1}_{A_n} \Psi_n^2 &= \int_{\Omega^n} \mathbf{1}_{A_n} \left(\sqrt{L_{f,h}^{1,n}/L_{f,h}^{2,n}} - 1 \right)^2 dQ_{f,h}^n \\
&= \int_{\Omega^n} \mathbf{1}_{A_n} \left(\exp\left(\frac{1}{2} \log L_{f,h}^{1,n} - \frac{1}{2} \log L_{f,h}^{2,n} \right) - 1 \right)^2 dQ_{f,h}^n \\
&\leq \left(\exp\left(\frac{1}{2} r_n^{\alpha_1} \right) - 1 \right)^2 = O\left(r_n^{2\alpha_1} \right). \qquad (5.5)
\end{aligned}
$$

An application of the elementary inequality $(a+b)^2 \leq 2a^2 + 2b^2$ gives the bound for the second expectation in the right hand side of (5.4):

$$
\mathbf{E}_f^n \mathbf{1}_{A_n^c} \Psi_n^2 \leq 2\mathbf{E}_f^n \mathbf{1}_{A_n^c} L_{f,h}^{1,n} + 2\mathbf{E}_f^n \mathbf{1}_{A_n^c} L_{f,h}^{2,n}. \qquad (5.6)
$$

We proceed to estimate $\mathbf{E}_f^n \mathbf{1}_{A_n^c} L_{f,h}^{1,n}$. Setting $B_n = \left\{ \log L_{f,h}^{1,n} \leq -\delta \log r_n \right\}$, where $\delta > 0$ will be specified below, one gets

$$
\begin{aligned}
\mathbf{E}_f^n \mathbf{1}_{A_n^c} L_{f,h}^{1,n} &= \mathbf{E}_f^n \mathbf{1}_{A_n^c \cap B_n} L_{f,h}^{1,n} + \mathbf{E}_f^n \mathbf{1}_{A_n^c \cap B_n^c} L_{f,h}^{1,n} \\
&\leq r_n^{-\delta} \mathbf{P}_f^n \left(A_n^c \right) + \mathbf{E}_f^n \mathbf{1}_{B_n^c} L_{f,h}^{1,n}. \qquad (5.7)
\end{aligned}
$$

Choosing δ small such that $\alpha_2 = \alpha_1 - \delta > \alpha$, we get,

$$
\sup_{f,h} n^{\delta} \mathbf{P}_f^n \left(A_n^c \right) = O\left(r_n^{\alpha_1 - \delta} \right) = O\left(r_n^{\alpha_2} \right). \qquad (5.8)
$$

The second term on the right-hand side of (5.7) can be written as

$$
\begin{aligned}
\mathbf{E}_f^n \mathbf{1}_{B_n^c} L_{f,h}^{1,n} &= \mathbf{E}_f^n \left(\log L_{f,h}^{1,n} > -\delta \log r_n \right) L_{f,h}^{1,n} \\
&= P_{f,h}^n \left(\log L_{f,h}^{1,n} > -\delta \log r_n \right) = O\left(r_n^{\alpha_1} \right). \qquad (5.9)
\end{aligned}
$$

Inserting (5.8) and (5.9) in (5.7) we get

$$
\mathbf{E}_f^n \mathbf{1}_{A_n^c} L_{f,h}^{1,n} = O\left(r_n^{\alpha_2} \right). \qquad (5.10)
$$

An estimate for the second term on the right-hand side of (5.6) is proved in exactly the same way. This gives $\mathbf{E}_f^n \mathbf{1}_{A_n^c} \Psi_n^2 = O\left(r_n^{\alpha_2} \right)$, which in turn, in conjunction with (5.5) and (5.4), concludes the proof of Theorem 5.1. ∎

5.2. **Nonparametric experiments which admit a locally asymptotic stochastic expansion.** We shall show that the assumptions in Theorem 5.1 are satisfied if the log-likelihood ratio in the experiment $\mathcal{E}^n$ admits a certain stochastic expansion in terms of independent random variables.

Let $T = [0,1]$ and $\mathcal{F} \subset \Theta^T$ be some set of functions $f(t) : T \to \Theta$. Let $\mathcal{E}^n$, $n \geq 1$ be a sequence of statistical experiments

$$
\mathcal{E}^n = \left(\Omega^n, \mathcal{A}^n, \left\{ P_f^n : f \in \mathcal{F} \right\} \right),
$$

with parameter set $\mathcal{F}$. For simplicity we assume that, for any $n = 1, 2, \ldots$ the measures $P_f^n : f \in \mathcal{F}$ in the experiment $\mathcal{E}^n$ are equivalent, i.e. that $P_f^n \ll P_g^n$, for any $f, g \in \mathcal{F}$. Recall that $\mathcal{H}(\beta, L)$ is a Hölder ball of functions defined on T with values in Θ. The parameters β and L are assumed to be absolute constants satisfying $\beta > 1/2$ and

$0 < L < \infty$. It will be convenient to define the neighborhoods of $f \in \mathcal{F}$ (shifted to the origin) as follows: for any non-negative real number r,

$$\mathcal{F}_f(r) = \{rh : h \in \mathcal{H}(\beta, L),\ f + rh \in \mathcal{F}\}.$$

For stating our definition this we need the following objects:

E1: A sequence of real numbers r_n, $n = 1, 2, ...$ which satisfies $r_n \to 0$, as $n \to \infty$.

E2: A function $I(\theta) : \Theta \to (0, \infty)$, which will play the role of the Fisher information in the experiment $\mathcal{E}^n$.

E3: The triangular array of design points $t_{ni} = i/n$, $i = 1, ..., n$, $n \geq 1$, on the interval $T = [0, 1]$.

Definition 5.1. The sequence of experiments $\mathcal{E}^n$, $n \geq 1$ is said to satisfy condition LASE with rate r_n and local Fisher information function $I(\cdot)$, if, for any fixed $n \geq 1$ and any fixed $f \in \mathcal{F}$, on the probability space $(\Omega^n, \mathcal{A}^n, P_f^n)$ there is a sequence of independent r.v.'s $\xi_{ni}(f)$, $i = 1, ..., n$ of mean 0 and variances

$$E_f^n \xi_{ni}^2(f) = I(f(t_{ni})), \quad i = 1, ..., n,$$

such that the expansion

$$\log \frac{dP_{f+h}^n}{dP_f^n} = \sum_{i=1}^n h(t_{ni}) \xi_{ni}(f) - \frac{1}{2} \sum_{i=1}^n h(t_{ni})^2 I(f(t_{ni})) + \rho_n(f, h),$$

holds true for any $h \in \mathcal{F}_f(r_n)$, where the remainder $\rho_n(f, h)$ satisfies

$$P_{f_n}^n(|\rho_n(f_n, h_n)| > a) \to 0,$$

for any two fixed sequences $(f_n)_{n\geq1}$ and $(h_n)_{n\geq1}$ subject to $f_n \in \mathcal{F}$, $h_n \in \mathcal{F}_{f_n}(r_n)$ and any real $a > 0$, as $n \to \infty$.

In the sequel we shall impose conditions (C1-C4) as formulated below.

C1: The sequence r_n, $n = 1, 2, ...$ has the parametric rate, i.e. is so that

$$r_n = c\frac{1}{\sqrt{n}}.$$

C2: The remainder $\rho_n(f, h)$ in the definition of condition LASE converges to 0 at a certain rate: for some $\alpha \in (\frac{1}{2\beta}, 1)$ and any $\varepsilon > 0$,

$$\sup_f \sup_h P_{f_n}^n\left(|\rho_n(f_n, h_n)| \geq \varepsilon n^{-\alpha/2}\right) = O\left(n^{-\alpha}\right),$$

where sup is taken over all possible $f \in \mathcal{F}$ and $h \in \mathcal{F}_f(r_n)$.

C3: The r.v.'s $\xi_{ni}(f)$, $i = 1, ..., n$ in the definition of condition LASE satisfy a strengthened version of the Lindeberg condition: for some $\alpha \in (\frac{1}{2\beta}, 1)$ and any $\varepsilon > 0$,

$$\sup_{f\in\mathcal{F}} \frac{1}{n} \sum_{i=1}^n E_f^n (n^{\alpha/2}\xi_{ni}(f))^2 \mathbf{1}\left(\left|n^{\alpha/2}\xi_{ni}(f)\right| \geq \varepsilon\sqrt{n}\right) = O\left(n^{-\alpha}\right).$$

C4: For $n = 1, 2, ...$ the local Fisher information function $I(\theta)$ satisfies

$$0 < I_{\min} \leq I(\theta) \leq I_{\max} < \infty, \quad \theta \in \Theta.$$

Let $\mathcal{E}_f^n$ be the local experiment

$$\mathcal{E}_f^n = \left(\Omega^n, \mathcal{A}^n, \left\{P_{f+h}^n : h \in \mathcal{F}_f(r_n)\right\}\right)$$

and let f_n, $n \geq 1$ denote any sequence of functions in $\mathcal{F}$. Let $H(\cdot,\cdot)$ be the Hellinger distance between probability measures, see (9.2). The next theorem states that, under a strengthened version of condition LASE, the sequence of local experiments $\mathcal{E}_{f_n}^n$, $n \geq 1$ can be approximated by a sequence of Gaussian shift experiments uniformly in all sequences f_n, $n \geq 1$, using the distance $H(\cdot,\cdot)$.

Theorem 5.2. *Assume that the sequence of experiments $\mathcal{E}^n$, $n = 1, 2, \ldots$ satisfies condition LASE with rate r_n and local Fisher information function $I(\theta)$ and conditions (C1-C4) hold true. Let ε_i, $i = 1, 2, \ldots$be a sequence of i.i.d. standard normal r.v.'s defined on a probability space $(\Omega^0, \mathcal{A}^0, \mathbf{P})$. Let, for any fixed $n \geq 1$ and fixed $f \in \Sigma$,*

$$\mathcal{G}_f^n = \left(\mathrm{R}^n, \mathcal{B}^n, \left\{Q_{f,h}^n : h \in \mathcal{F}_f(r_n)\right\}\right)$$

be the Gaussian shift experiment generated by n observations

$$Y_i^n = h\left(i/n\right) + \frac{1}{\sqrt{I\left(f\left(i/n\right)\right)}}\varepsilon_i, \quad i = 1, \ldots, n,$$

with $h \in \mathcal{F}_f(r_n)$. Then, for any fixed $n \geq 1$ and $f \in \mathcal{F}$, there are experiments

$$\begin{aligned} \widetilde{\mathcal{G}}_f^n &= \left(\Omega^0, \mathcal{A}^0, \left\{\widetilde{Q}_{f,h}^n : h \in \mathcal{F}_f(r_n)\right\}\right), \\ \widetilde{\mathcal{E}}_f^n &= \left(\Omega^0, \mathcal{A}^0, \left\{\widetilde{P}_{f,h}^n : h \in \mathcal{F}_f(r_n)\right\}\right) \end{aligned}$$

such that

$$\Delta\left(\mathcal{G}_f^n, \widetilde{\mathcal{G}}_f^n\right) = \Delta\left(\mathcal{E}_f^n, \widetilde{\mathcal{E}}_f^n\right) = 0$$

and for some $\alpha \in (1/2\beta, 1)$,

$$\sup_{f \in \mathcal{F}} \sup_{h \in \mathcal{F}_f(r_n)} H^2\left(\widetilde{P}_{f,h}^n, \widetilde{Q}_{f,h}^n\right) = O\left(r_n^{2\alpha}\right),$$

as $n \to \infty$.

We give here some hints how to carry out the proof of Theorem 5.2. Starting with the independent standard Gaussian sequence ε_i, $i = 1, 2, \ldots$, we construct a sequence $\widetilde{\xi}_{n1}, \ldots, \widetilde{\xi}_{nn}, \widetilde{\rho}_n(f,h)$ with the same joint distribution as $\xi_{n1}, \ldots, \xi_{nn}, \rho_n(f,h)$ from the expansion for the likelihood $L_{f,h}^{1,n} = dP_{f+h}^n / dP_f^n$. This will ensure that the "new" likelihood $\widetilde{L}_{f,h}^{1,n}$, as a process indexed by $h \in \mathcal{F}_f(r_n)$, has the same law as $L_{f,h}^{1,n}$, and thus, that the corresponding experiments are exactly equivalent. The key point in this construction is to guarantee that the two length n sequences $I^{1/2}\left(f\left(i/n\right)\right)\varepsilon_i$, $i = 1, \ldots, n$ and $\widetilde{\xi}_{ni}$, $i = 1, \ldots, n$ are as close as possible. For this we make use of a strong approximation result for partial sums of independent r.v.'s indexed by functions, provided by Theorem 9.1 (see the Appendix). We note also that the new remainder $\widetilde{\rho}_n(f,h)$ will satisfy the same requirements as $\rho_n(f,h)$ does, since both are equally distributed.

5.3. **Asymptotic expansion with bounded scores.** Assume that the sequence of experiments $\mathcal{E}^n$, $n = 1, 2, ...$ satisfies condition LASE. This means that, for f, h satisfying $f \in \mathcal{F}$ and $h \in \mathcal{F}_f(r_n)$,

$$\log \frac{dP_{f+h}^n}{dP_f^n} = \sum_{i=1}^n h(t_{ni}) \xi_{ni} - \frac{1}{2} \sum_{i=1}^n h(t_{ni})^2 I(f(t_{ni})) + \rho_n(f, h), \tag{5.11}$$

where $\xi_{ni} = \xi_{ni}(f)$, $i = 1, ..., n$ is a sequence of independent r.v.'s of mean 0 and variances $\mathrm{E}_f^n \xi_{ni}^2 = I(f(t_{ni}))$. If the model is of location type

$$X_i = f(i/n) + \eta_i, \quad i = 1, ..., n,$$

where the noise η_i has density $p(x)$ then ξ_{ni} in (5.11) stands for $l(\eta_i)$ where $l(x) = p'(x)/p(x)$ is often called the score function. In a somewhat loose terminology, the r.v.'s ξ_{ni} may therefore be called "scores". We shall show that, under the conditions (C1-C4), the above expansion can be modified so that the r.v.'s ξ_{ni} are replaced by some bounded r.v.'s ξ_{ni}^* with the same mean and variances. More precisely, we prove the following.

Lemma 5.3. *Let conditions (C1-C4) hold true. Then there is a sequence of independent r.v.* $\xi_{ni}^*(f)$, $i = 1, ..., n$ *of means* 0 *and variances* $E_f^n \xi_{ni}^*(f)^2 = I(f(t_{ni}))$, $i = 1, ..., n$, *satisfying*

$$\left| r_n^{1-\alpha} \xi_{ni}^*(f) \right| \leq c, \quad i = 1, ..., n \tag{5.12}$$

for some real number $\alpha \in (1/2\beta, 1)$, *and such that*

$$\log \frac{dP_{f+h}^n}{dP_f^n} = \sum_{i=1}^n h(t_{ni}) \xi_{ni}^*(f) - \frac{1}{2} \sum_{i=1}^n h(t_{ni})^2 I(f(t_{ni})) + \rho_n^*(f, h), \tag{5.13}$$

where, for any $\varepsilon > 0$,

$$\sup_{f \in \mathcal{F}} \sup_{h \in \mathcal{F}_f(r_n)} P_f^n \left(|\rho_n^*(f, h)| \geq \varepsilon r_n^\alpha \right) = O\left(r_n^{2\alpha} \right). \tag{5.14}$$

Proof. Let $\alpha \in \left(\frac{1}{2\beta}, 1 \right)$ be the real number for which assumptions (C2) and (C3) hold true. Since $E_f^n \xi_{ni} = 0$, we have

$$\xi_{ni} = \xi_{ni}' + \xi_{ni}'' = \eta_{ni}' + \eta_{ni}'', \quad i = 1, ..., n, \tag{5.15}$$

where

$$\xi_{ni}' = \xi_{ni} \mathbf{1}(|r_n \xi_{ni}| \leq r_n^\alpha), \quad \xi_{ni}'' = \xi_{ni} \mathbf{1}(|r_n \xi_{ni}| > r_n^\alpha) \tag{5.16}$$

and

$$\eta_{ni}' = \xi_{ni}' - E_f^n \xi_{ni}', \quad \eta_{ni}'' = \xi_{ni}'' - E_f^n \xi_{ni}''. \tag{5.17}$$

Set

$$v_{ni}^2 = E_f^n \xi_{ni}^2 - E_f^n (\eta_{ni}')^2 \tag{5.18}$$

and

$$p_{ni} = \frac{1}{2} x_n^{-2} v_{ni}^2, \quad x_n = c_1 r_n^{\alpha - 1},$$

Since $r_n \to 0$ as $n \to \infty$ and $v_{ni}^2 \leq E_f^n \xi_{ni}^2 = I(f(t_{ni})) \leq I_{\max}$, the constant c_1 can be chosen large enough so that $p_{ni} \leq 1/2$ for any $n \geq 1$.

Without loss of generality one may assume that on the probability space $(\Omega^n, \mathcal{A}^n, P_f^n)$ there is a sequence of independent r.v.'s η_{ni}''', $i = 1, ..., n$, independent of the sequence ξ_{ni}, $i = 1, ..., n$, which take values $-x_n$, 0, x_n with probabilities p_{ni}, $1 - 2p_{ni}$, p_{ni} respectively. It is clear that the r.v.'s η_{ni}''' are such that

$$\left| r_n^{1-\alpha} \eta_{ni}''' \right| \leq c_1, \quad E_f^n \eta_{ni}''' = 0, \quad E_f^n (\eta_{ni}''')^2 = v_{ni}^2. \tag{5.19}$$

Set

$$\xi_{ni}^* = \eta_{ni}' + \eta_{ni}''', \quad i = 1, ..., n. \tag{5.20}$$

From this definition it is clear that (5.12) holds true with $c = 2 + c_1$. Since ξ_{ni} and η_{ni}''' are independent, taking into account (5.19) and (5.18), we get

$$E_f^n (\xi_{ni}^*)^2 = E_f^n (\eta_{ni}')^2 + E_f^n (\eta_{ni}''')^2 = E_f^n (\eta_{ni}')^2 + v_{ni}^2 = E_f^n \xi_{ni}^2 = I(f(t_{ni})).$$

Set $\rho_n^*(f,h) = \rho_n(f,h) + \rho_n'(f,h)$, where

$$\rho_n'(f,h) = \sum_{i=1}^n h(t_{ni}) (\xi_{ni} - \xi_{ni}^*).$$

The lemma will be proved if we show (5.14). Because of the assumption (C2), it suffices to prove that

$$\sup_{f \in \mathcal{F}} \sup_{h \in \mathcal{F}_f(r_n)} P_f^n \left(|\rho_n'(f,h)| \geq \frac{\epsilon}{2} r_n^\alpha \right) = O\left(r_n^{2\alpha}\right), \tag{5.21}$$

for some $\alpha \in \left(\frac{1}{2\beta}, 1\right)$. To prove (5.21) we note that, by (5.15) and (5.20) we have $\xi_{ni} - \xi_{ni}^* = \eta_{ni}'' - \eta_{ni}'''$ and therefore $\rho_n'(f,h)$ can be represented as follows:

$$\rho_n'(f,h) = \sum_{i=1}^n h(t_{ni}) \eta_{ni}'' - \sum_{i=1}^n h(t_{ni}) \eta_{ni}'''.$$

From the last equality we get

$$P_f^n \left(|\rho_n'(f,h)| \geq \frac{\epsilon}{2} r_n^\alpha \right) \leq J_n^{(1)} + J_n^{(2)}, \tag{5.22}$$

where

$$J_n^{(1)} = P_f^n \left(\left| \sum_{i=1}^n h(t_{ni}) \eta_{ni}'' \right| \geq \frac{\epsilon}{4} r_n^\alpha \right), \quad J_n^{(2)} = P_f^n \left(\left| \sum_{i=1}^n h(t_{ni}) \eta_{ni}''' \right| \geq \frac{\epsilon}{4} r_n^\alpha \right).$$

By Chebyshev's inequality we have

$$J_n^{(1)} \leq c r_n^{-2\alpha} \sum_{i=1}^n h(t_{ni})^2 E_f^n (\eta_{ni}'')^2.$$

Since $E_f^n(\eta_{ni}'')^2 = E_f^n(\xi_{ni}'' - E_f^n\xi_{ni}'')^2 \leq E_f^n(\xi_{ni}'')^2$ and $\|h\|_\infty \leq r_n$, making use of (5.16) and of the strengthened version of the Lindeberg condition (C3), we obtain

$$J^{(1)} \leq c\sum_{i=1}^{n} E_f^n(r_n^{1-\alpha}\xi_{ni})^2\mathbf{1}\left(\left|r_n^{1-\alpha}\xi_{ni}\right| > 1\right) = O\left(r_n^{2\alpha}\right). \tag{5.23}$$

To handle the term $J_n^{(2)}$ on the right-hand side of (5.22) we again invoke the Chebyshev inequality to obtain

$$J_n^{(2)} \leq cr_n^{-2\alpha}\sum_{i=1}^{n} h\left(t_{ni}\right)^2 E_f^n(\eta_{ni}''')^2 = cr_n^{-2\alpha}\sum_{i=1}^{n} h\left(t_{ni}\right)^2 v_{ni}^2. \tag{5.24}$$

Since $\xi_{ni}^2 = (\xi_{ni}')^2 + (\xi_{ni}'')^2$ and $E_f^n\xi_{ni}' = -E_f^n\xi_{ni}''$, we have

$$v_{ni}^2 = E_f^n\xi_{ni}^2 - E_f^n(\xi_{ni}')^2 + (E_f^n\xi_{ni}')^2 = E_f^n(\xi_{ni}'')^2 + (E_f^n\xi_{ni}'')^2 \leq 2E_f^n(\xi_{ni}'')^2,$$

which in turn implies, in the same manner as for $J_n^{(1)}$,

$$J_n^{(2)} \leq cr_n^{-2\alpha}\sum_{i=1}^{n} h\left(t_{ni}\right)^2 E_f^n(\xi_{ni}'')^2 = O\left(r_n^{2\alpha}\right). \tag{5.25}$$

Inserting (5.23) and (5.25) into (5.22) we obtain (5.21). ∎

5.4. **Construction of the likelihoods on the same probability space.** We proceed to construct the local experiment $\mathcal{E}_f^n$ on the same measurable space with a Gaussian experiment. For this let $\varepsilon_1, \varepsilon_2 ...$ be an infinite sequence of i.i.d. standard normal r.v.'s [defined on some probability space $(\Omega^0, \mathcal{A}^0, \mathbf{P})$]. Consider the finite sequence of Gaussian observations

$$Y_i^n = h\left(t_{ni}\right) + I^{-1/2}\left(f\left(t_{ni}\right)\right)\varepsilon_i, \quad i = 1, ..., n, \tag{5.26}$$

with $f \in \mathcal{F}$, $h \in \mathcal{F}_f(r_n)$. The statistical experiment generated by these is

$$\mathcal{G}_f^n = \left(\mathrm{R}^n, \mathcal{B}^n, \left\{Q_{f,h}^n : h \in \mathcal{F}_f(r_n)\right\}\right)$$

and the likelihood $L_{f,h}^{0,n} = dQ_{f,h}^n/dQ_{f,0}^n$ as a r.v. under $Q_{f,0}^n$ has a representation

$$\widetilde{L}_{f,h}^{0,n} = \exp\left(\sum_{i=1}^{n} h\left(t_{ni}\right)\zeta_{ni} - \frac{1}{2}\sum_{i=1}^{n} h^2\left(t_{ni}\right) I\left(f\left(t_{ni}\right)\right)\right). \tag{5.27}$$

where $\zeta_{ni} = I^{1/2}\left(f\left(t_{ni}\right)\right)\varepsilon_i$, $i = 1, ..., n$. It is clear that $\zeta_{n1}, ..., \zeta_{nn}$ is a sequence of independent normal r.v.'s of means 0 and variances $I\left(f\left(t_{ni}\right)\right)$, on the probability space $(\Omega^0, \mathcal{A}^0, \mathbf{P})$.

We shall construct a version of the likelihoods

$$L_{f,h}^{1,n} = \frac{dP_{f+h}^n}{dP_f^n}, \quad h \in \mathcal{F}_f(r_n)$$

of the experiment $\mathcal{E}_f^n$ on the probability space $(\Omega^0, \mathcal{A}^0, \mathbf{P})$, obtaining thus an equivalent experiment $\widetilde{\mathcal{E}}_f^n$. For this we apply Theorem 9.1 with $X_{ni} = \xi_{ni}^*$, $N_{ni} = \zeta_{ni}$, $\lambda_n = r_n^{1-1/(2\beta)}$ and $x = \lambda_n^{-1}\log n$. According to this theorem, there is a sequence of independent r.v.'s

$\widetilde{\xi}_{ni}$, $i=1,...,n$ on the probability space $(\Omega^0, \mathcal{A}^0, \mathbf{P})$, satisfying $\widetilde{\xi}_{ni} \overset{d}{=} \xi^*_{ni}$, for $i=1,...,n$, and such that

$$\sup_h \mathbf{P}\left(\left| \sum_{i=1}^n h(t_{ni}) \left(\widetilde{\xi}_{ni} - \zeta_{ni} \right) \right| \geq c_1 r_n^{1/(2\beta)} (\log n)^2 \right) \leq c_2 \frac{1}{n}, \tag{5.28}$$

where c_1, c_2 are absolute constants and the sup is taken over $h \in \mathcal{F}_f(r_n)$. We recall that $r_n^{-1} h \in \mathcal{H}(\beta, L) \subset \mathcal{H}(1/2, L)$.

In order to construct the log-likelihoods $\log L^{1,n}_{f,h}$ of the experiment $\mathcal{E}^n_f$ it suffices to construct a new "remainder" $\widetilde{\rho}_n = \widetilde{\rho}_n(f,h)$ on the probability space $(\Omega^0, \mathcal{A}^0, \mathbf{P})$, or on some extension of it, such that the joint distribution of the sequence $(\widetilde{\xi}_{n1}, ..., \widetilde{\xi}_{nn}, \widetilde{\rho}_n)$ is the same as that of the original sequence $(\xi^*_{n1}, ..., \xi^*_{nn}, \rho^*_n)$. This can be done using any kind of constructions, since the only property required from the r.v. $\widetilde{\rho}_n$ is to satisfy (5.14), with $\widetilde{\rho}_n$ replacing ρ^*_n, which follows obviously from the fact that (by construction) $\mathcal{L}(\widetilde{\rho}_n) = \mathcal{L}(\rho^*_n)$. We shall describe such a possible construction using some elementary arguments by enlarging the initial probability space, although it is possible to give a more delicate one on the same probability space. Let us consider the probability space $\mathbb{S}^* = (\Omega^0, \mathcal{A}^0, \mathbf{P}) \otimes \left(\mathrm{R}, \mathcal{B}, \mathbf{P}_{\rho^*_n | \xi^*_{n1}, ..., \xi^*_{nn}} \right)$ as an enlargement of the initial space $(\Omega^0, \mathcal{A}^0, \mathbf{P})$, where $\mathbf{P}_{\rho^*_n | \xi^*_{n1}, ..., \xi^*_{nn}}$ is the conditional distribution of ρ^*_n given $\xi^*_{n1}, ..., \xi^*_{nn}$. Now, on the enlarged probability space $\mathbb{S}^*$ we define the r.v. $\widetilde{\rho}_n(\widetilde{\omega}) = y$, for all $\widetilde{\omega} = (x_1, ..., x_n, y) \in \mathbb{S}^*$, which has the desired properties. Without any loss of generality we can assume that the construction is performed on the initial probability space $(\Omega^0, \mathcal{A}^0, \mathbf{P})$. For more complicated constructions we refer to Berkes and Philipp [1]. In any case, the construction is performed in such a way that the new remainder $\widetilde{\rho}_n = \widetilde{\rho}_n(f,h)$ satisfies

$$\sup_{f \in \mathcal{F}} \sup_{h \in \mathcal{F}_f(r_n)} \mathbf{P}\left(|\widetilde{\rho}_n(f,h)| \geq 3 r_n^{\alpha} \right) = O\left(r_n^{2\alpha} \right). \tag{5.29}$$

Define the r.v.'s $\widetilde{L}^n_f(h)$ such that, for any $h \in \mathcal{F}_f(r_n)$,

$$\log \widetilde{L}^n_f(h) = \sum_{i=1}^n h(t_{ni}) \widetilde{\xi}_{ni} - \frac{1}{2} \sum_{i=1}^n h(t_{ni})^2 I(f(t_{ni})) + \widetilde{\rho}_n(f,h). \tag{5.30}$$

On the measurable space $(\Omega^0, \mathcal{A}^0)$ consider the set of laws $\{\widetilde{P}^n_{f,h} : h \in \mathcal{F}_f(r_n)\}$, where $\widetilde{P}^n_{f,0} = \mathbf{P}$ and $\widetilde{P}^n_{f,h}$ is such that

$$\frac{d\widetilde{P}^n_{f,h}}{d\widetilde{P}^n_{f,0}} = \widetilde{L}^{1,n}_{f,h},$$

for any $h \in \mathcal{F}_f(r_n)$. Set

$$\widetilde{\mathcal{E}}^n_f = \left(\Omega^0, \mathcal{A}^0, \{\widetilde{P}^n_{f,h} : h \in \mathcal{F}_f(r_n)\} \right).$$

Since the quadratic terms in the expansions (5.13) and (5.30) are deterministic, the equality in distribution of the two vectors $(\xi^*_{n1}, ..., \xi^*_{ni}, \rho^*_n)$ and $(\widetilde{\xi}_{n1}, ..., \widetilde{\xi}_{ni}, \widetilde{\rho}_n)$ implies for any finite set $S \subset \mathcal{F}_f(r_n)$

$$\mathcal{L}\left((L^{1,n}_{f,h})_{h\in S}|P^n_f\right) = \mathcal{L}\left((\widetilde{L}^{1,n}_{f,h})_{h\in S}|\widetilde{P}^n_{f,0}\right). \tag{5.31}$$

From (5.31) it follows that, for any $n = 1, 2, \ldots$, the experiments $\mathcal{E}^n_f$ and $\widetilde{\mathcal{E}}^n_f$ are (exactly) equivalent, i.e. $\Delta\left(\mathcal{E}^n_f, \widetilde{\mathcal{E}}^n_f\right) = 0$. From the likelihood process $\widetilde{L}^{0,n}_{f,h}$, $h \in \mathcal{F}_f(r_n)$ defined on $(\Omega^0, \mathcal{A}^0, \mathbf{P})$ (cf. (5.29)) we construct an equivalent version

$$\widetilde{\mathcal{G}}^n_f = \left(\Omega^0, \mathcal{A}^0, \left\{\widetilde{Q}^n_{f,h} : h \in \mathcal{F}_f(r_n)\right\}\right)$$

of $\mathcal{G}^n_f$ in the same way.

5.5. **Proof of Theorem 5.2.** To prove Theorem 5.2 we only have to verify the assumptions of Theorem 5.1. In our next lemma it is shown that condition (5.1) is met.

Lemma 5.4. *Assume that the sequence of experiments $\mathcal{E}^n$ satisfies condition LASE and that conditions (C1-C4) hold true. Then the constructed experiments $\widetilde{\mathcal{E}}^n_f$ and $\widetilde{\mathcal{G}}^n_f$ are such that for some $\alpha \in (\frac{1}{2\beta}, 1)$,*

$$\sup_{f\in\mathcal{F}} \sup_{h\in\mathcal{F}_f(r_n)} \mathbf{P}\left(\left|\log\frac{d\widetilde{P}^n_{f,h}}{d\widetilde{P}^n_{f,0}} - \log\frac{d\widetilde{Q}^n_{f,h}}{d\widetilde{Q}^n_{f,0}}\right| > r^\alpha_n\right) = O\left(r^{2\alpha}_n\right).$$

Proof. The proof is based on inequality (5.28) and of the bound (5.14) in Lemma 5.3. Being elementary, it is left to the reader. ∎

Next we need to check condition (5.2) in Theorem 5.1.

Lemma 5.5. *Assume that the sequence of experiments $\mathcal{E}^n$ satisfies condition LASE and that conditions (C1-C4) hold true. Then there is a constant $\alpha \in (1/2\beta, 1)$ such that, for any $\varepsilon \in (0, 1)$,*

$$\sup_{f\in\mathcal{F}} \sup_{h\in\mathcal{F}_f(r_n)} P^n_{f+h}\left(\log\frac{dP^n_{f+h}}{dP^n_f} > -\varepsilon\log r_n\right) = O\left(r^{2\alpha}_n\right), \quad n\to\infty.$$

Proof. Consider the inverse likelihood ratio dP^n_f/dP^n_{f+h} corresponding to the local experiment $\mathcal{E}^n_f$. Setting $g = f + h \in \mathcal{F}$ and using Lemma 5.3, we rewrite it as

(5.32)
$$\log\frac{dP^n_f}{dP^n_{f+h}} = \log\frac{dP^n_{g-h}}{dP^n_g} = -\sum_{i=1}^n h(t_{ni})\,\xi^*_{ni}(g) - \frac{1}{2}\sum_{i=1}^n h(t_{ni})^2 I(g(t_{ni})) + \rho^*_n(g,h),$$

where $h \in \mathcal{F}_g(r_n)$ and $\xi^*_{ni}(g)$, $i = 1, \ldots, n$ are P^n_g-independent r.v.'s of means 0 and variances $E_{P^n_g}\xi^*_{ni}(g)^2 = I(g(t_{ni})) \le I_{\max}$, $i = 1, \ldots, n$. Moreover $|r^{1-\alpha}_n \xi^*_{ni}(f)| \le c$, $\quad i = 1, \ldots, n$. Because of conditions (C1) and (C4), we have

$$\sum_{i=1}^n h(t_{ni})^2 I(g(t_{ni})) = O\left(nr^2_n\right) = O(1). \tag{5.33}$$

Choose $\alpha \in \left(\frac{1}{2\beta}, 1\right)$ such that conditions (C2-C3) hold true. With these notations, for n large enough,

$$P_g^n\left(\log\frac{dP_{g-h}^n}{dP_g^n} \leq \varepsilon \log r_n\right) \leq J_n^{(1)} + J_n^{(2)},$$

where

$$\begin{aligned} J_n^{(1)} &= P_g^n\left(-\sum_{i=1}^n h(t_{ni})\,\xi_{ni}^*(g) < \frac{\varepsilon}{2}\log r_n\right), \\ J_n^{(2)} &= P_g^n\left(\rho_n(g,h) \geq r_n^\alpha\right). \end{aligned} \tag{5.34}$$

Since $\|h\|_\infty \leq r_n$, it follows that the r.v.'s $h(t_{ni})\,\xi_{ni}^*(g)$ are bounded by $r_n^\alpha \leq c_1$, for some absolute constant c_1. By Lemma 9.2 (see the Appendix),

$$E_f^n \exp\left(-\frac{2}{\varepsilon}h(t_{ni})\,\xi_{ni}^*(g)\right) \leq \exp\left(c_2 h(t_{ni})^2\, E_f^n \xi_{ni}^*(g)^2\right),$$

for another absolute constant c_2. Simple calculations yield

$$\begin{aligned} J_n^{(1)} &\leq e^{-2\log r_n}\prod_{i=1}^n E_g^n \exp\left(4\varepsilon^{-1}h(t_{ni})\,\xi_{ni}^*(g)\right) \\ &\leq r_n^{-2}\exp\left(c_2\varepsilon^{-2}\sum_{i=1}^n h(t_{ni})^2\, E_g^n\xi_{ni}^*(g)^2\right). \end{aligned}$$

Taking into account $E_g^n\xi_{ni}^*(g)^2 = I(g(t_{ni}))$ and (5.33) we get $\sup J_n^{(1)} = O(r_n^2) = O(r_n^{2\alpha})$, where the supremum is taken over $f \in \mathcal{F}$ and $h \in \mathcal{F}_f(r_n)$. The bound $\sup J_n^{(2)} = O(r_n^{2\alpha})$, with the supremum over the same f and h, is straightforward, by assumption (C2). Combining the bounds for $J_n^{(1)}$ and $J_n^{(2)}$ we obtain the lemma. ∎

Obviously a similar result holds true for the constructed experiment $\widetilde{\mathcal{E}}^n$.

Remark 5.1. Assume that the sequence of experiments $\mathcal{E}^n$ satisfies condition LASE and that conditions (C1-C4) hold true. Then there is a constant $\alpha \in (1/2\beta, 1)$ such that, for any $\varepsilon \in (0,1)$,

$$\sup_{f\in\mathcal{F}}\sup_{h\in\mathcal{F}_f(r_n)} \widetilde{P}_{f,h}^n\left(\log\frac{d\widetilde{P}_{f,h}^n}{d\widetilde{P}_{f,0}^n} \geq -\varepsilon\log r_n\right) = O\left(r_n^{2\alpha}\right), \quad n\to\infty.$$

We continue with a moderate deviations bound for the log-likelihood ratio of the Gaussian experiment $\mathcal{G}^n$ required by the condition (5.3) of Theorem 5.1, which is proved in the same way as the above Lemma 5.5.

Lemma 5.6. *Assume that the sequence of experiments $\mathcal{E}^n$ satisfies condition LASE and that conditions (C1-C4) hold true. Then there is a constant $\alpha \in (1/2\beta, 1)$ such* *that, for any* $\varepsilon \in (0,1)$,

$$\sup_{f\in\mathcal{F}} \sup_{h\in\mathcal{F}_f(r_n)} Q^n_{f,h}\left(\log \frac{dQ^n_{f,h}}{dQ^n_{f,0}} \geq -\varepsilon \log r_n\right) = O\left(r_n^{2\alpha}\right), \quad n\to\infty.$$

Proof. Consider the likelihood ratio corresponding to the local Gaussian experiment $\mathcal{G}^n_f$:

$$L^{2,n}_{f,h} = \frac{dQ^n_{f+h}}{dQ^n_f} = \exp\left(\sum_{i=1}^n h(t_{ni})\zeta_{ni} - \frac{1}{2}\sum_{i=1}^n h(t_{ni})^2 I(f(t_{ni}))\right),$$

where $f\in\mathcal{F}$, $h\in\mathcal{F}_f(r_n)$ and ζ_{ni}, $i=1,...,n$ are independent normal r.v.'s of means 0 and variances $I(f(t_{ni}))\leq I_{\max}$ respectively. Then, by Chebyshev's inequality,

$$Q^n_f\left(\log L^{2,n}_{f,h} \geq -\varepsilon\log r_n\right) \leq r_n^2 E_{Q^n_{f,0}} \exp\left(2\varepsilon^{-1}\sum_{i=1}^n h(t_{ni})\zeta_{ni} - \varepsilon^{-1}\sum_{i=1}^n h(t_{ni})^2 I(f(t_{ni}))\right).$$

Since $\|h\|_\infty \leq cn^{-1/2}$ and ζ_{ni}, $i=1,...,n$ are independent normal r.v.'s, we get

$$Q^n_f\left(\log L^{2,n}_{f,h} \geq 2\varepsilon\log r_n\right) = O\left(r_n^{-2}\right),$$

uniformly in $f\in\mathcal{F}$ and $h\in\mathcal{F}_f(r_n)$. ∎

6. Application to nonparametrically driven models

We consider a particular case of the general setting of Section 5. Assume that $\mathcal{F}$ is given by $\mathcal{F}=\Sigma^\beta=\mathcal{H}(\beta,L)\cap\Theta^T$, where $T=[0,1]$, and $\mathcal{H}(\beta,L)$ is a Hölder ball on T. Consider the case where the experiment $\mathcal{E}^n$ (appearing in Section 5 in a general form) is generated by a sequence of independent observations $X_1,...,X_n$ where each r.v. X_i has density $p(x,f(t_{ni}))$, $f\in\mathcal{F}$, $t_{ni}=i/n$. The local experiment at $f\in\mathcal{F}$ then is

$$\begin{aligned}\mathcal{E}^n_f &= (X^n,\mathcal{X}^n,\{P^n_{f,h}: h\in\mathcal{F}_f(r_n)\}),\\ P^n_{f,h} &= P_{f(t_{n1})+h(t_{n1})}\times ...\times P_{f(t_{nn})+h(t_{nn})}\end{aligned}$$

and P_θ is the distribution on $(X,\mathcal{X})$ corresponding to the density $p(x,\theta)$, $\theta\in\Theta$. Let $I(\theta)$ be the Fisher information corresponding to the density $p(x,\theta)$, as defined by (3.6).

Theorem 6.1. *Assume that the density* $p(x,\theta)$ *satisfies conditions (R1-R3). Then the sequence of experiments* $\mathcal{E}^n$, $n\geq 1$ *for* $\mathcal{F}=\Sigma^\beta$ *satisfies LASE and conditions (C1-C4) hold true with rate* $r_n=cn^{-1/2}$, *local Fisher information* $I(\theta)$ *and* $\xi_{ni}(f)=\dot{l}(X_i,f(t_{ni}))$, $i=1,...,n$, *where* $\dot{l}(x,\theta)$ *is the tangent vector defined by (3.4).*

The remainder of section 6 will be devoted to the proof of this theorem.

6.1. **Stochastic expansion for the likelihood ratio.** The following preliminary stochastic expansion for the likelihood ratio will lead up to property LASE.

Proposition 6.2. *Assume that the density $p(x,\theta)$ satisfies conditions (R1-R3). Then, for any $f \in \mathcal{F} = \Sigma^\beta$ and $h \in \mathcal{F}_f(n^{-1/2})$,*

$$\log \frac{dP^n_{f+h}}{dP^n_f} = 2X_n(f,h) - 4V_n(f,h) + \rho_n(f,h),$$

where

$$\begin{aligned} X_n(f,h) &= \sum_{i=1}^n \left\{ (\sqrt{z_{ni}} - 1) - E^n_f \left(\sqrt{z_{ni}} - 1\right) \right\}, \\ V_n(f,h) &= \frac{1}{2} \sum_{i=1}^n E^n_f \left(\sqrt{z_{ni}} - 1\right)^2 \end{aligned}$$

and

$$z_{ni} = \frac{p(X_i, f(t_{ni}) + h(t_{ni}))}{p(X_i, f(t_{ni}))}.$$

Moreover, there is an $\alpha \in (1/2\beta, 1)$, such that the remainder $\rho_n(f,h)$ satisfies

$$\sup_{f\in\mathcal{F}} \sup_{h\in\mathcal{F}_f(n^{-1/2})} P^n_f \left(|\rho_n(f,h)| > n^{-\alpha/2} \right) = O\left(n^{-\alpha}\right).$$

Proof. It is easy to see that

$$\log \frac{dP^n_{f+h}}{dP^n_f} = \log \prod_{i=1}^n z_{ni} = \sum_{i=1}^n \log\left(1 + (\sqrt{z_{ni}} - 1)\right),$$

where z_{ni} is defined in Proposition 6.2. Note that, in view of the equalities

$$2(\sqrt{x} - 1) = x - 1 - (\sqrt{x} - 1)^2, \qquad \mathrm{E}^n_f z_{ni} = 1$$

we have

$$2\mathrm{E}^n_f(\sqrt{z_{ni}} - 1) = -\mathrm{E}^n_f(\sqrt{z_{ni}} - 1)^2.$$

By elementary transformations we obtain

$$\log \frac{dP^n_{f+n^{-1/2}h}}{dP^n_f} = X_n(f,h) - Y_n(f,h) - 4V_n(f,h) + \Psi_n(f,h),$$

where $X_n(f,h)$, $V_n(f,h)$ are defined in Proposition 6.2 and

$$Y_n(f,h) = \sum_{i=1}^n \left\{ \left(\sqrt{z_{ni}} - 1\right)^2 - E^n_f \left(\sqrt{z_{ni}} - 1\right)^2 \right\},$$

$$\Psi_n(f,h) = \sum_{i=1}^n \left\{ \log\left(1 + (\sqrt{z_{ni}} - 1)\right) - 2\left(\sqrt{z_{ni}} - 1\right) - \left(\sqrt{z_{ni}} - 1\right)^2 \right\}.$$

Now the result follows from Lemmas 6.3 and 6.4. ∎

Lemma 6.3. *Assume that condition (R2) holds true. Then there is an* $\alpha \in (1/2\beta, 1)$ *such that*

$$\sup_{f\in\mathcal{F}} \sup_{h\in\mathcal{F}_f(n^{-1/2})} P_f^n\left(|Y_n(f,h)| > n^{-\alpha/2}\right) = O\left(n^{-\alpha}\right).$$

Proof. Let $\delta \in (\frac{2\beta+1}{2\beta-1}, \infty)$ be the real number for which condition (R2) holds true (recall that $\beta \geq \frac{1}{2}$). Let $\alpha = \frac{\delta-1}{\delta+1}$, which clearly is in the interval $(\frac{1}{2\beta}, 1)$. Then there is an $\alpha' \in (\frac{1}{2\beta}, 1)$ such that $\alpha' < \alpha$. Set for $i = 1, ..., n$,

$$\xi_{ni} = \sqrt{z_{ni}} - 1, \quad \xi'_{ni} = \xi_{ni}^2 1(|\xi_{ni}| \leq n^{-\alpha/2}), \quad \xi''_{ni} = \xi_{ni}^2 1(|\xi_{ni}| > n^{-\alpha/2}) \tag{6.1}$$

and

$$\eta'_{ni} = \xi'_{ni} - E_f^n \xi'_{ni}, \quad \eta''_{ni} = \xi''_{ni} - E_f^n \xi''_{ni}. \tag{6.2}$$

With the above notations, we write Y_n as follows:

$$Y_n = \sum_{i=1}^n \eta'_{ni} + \sum_{i=1}^n \eta''_{ni}. \tag{6.3}$$

For the first term on the right-hand side of (6.3) we have

$$I_1 \equiv P_f^n\left(\sum_{i=1}^n \eta'_{ni} > \frac{1}{2} n^{-\alpha'/2}\right) \leq \exp\left(-\frac{1}{2} n^{(\alpha-\alpha')/2}\right) \prod_{i=1}^n E_f^n \exp\left(n^{\alpha/2}\eta'_{ni}\right), \tag{6.4}$$

where the r.v.'s $n^{\alpha/2}\eta_{ni}$ are bounded by $2n^{-\alpha/2} \leq 2$. According to Lemma 9.2, with $\lambda = 1$, one obtains

$$E_f^n \exp\left(n^{\alpha/2}\eta'_{ni}\right) \leq \exp\left(cn^\alpha E_f^n(\eta'_{ni})^2\right), \quad i = 1, ..., n. \tag{6.5}$$

Using (6.2) and (6.1),

$$E_f^n(\eta'_{ni})^2 \leq 2n^{-\alpha} E_f^n(\sqrt{z_{ni}} - 1)^2, \quad i = 1, ..., n. \tag{6.6}$$

Set for brevity

$$\dot{l}_{ni}(f,h) = \dot{l}_{f(t_{ni})}(X_i, f(t_{ni}) + h(t_{ni})), \quad i = 1, ..., n,$$

where $\dot{l}_\theta(x,u)$ is the extended tangent vector defined by (3.5) and $f \in \mathcal{F}$, $h \in \mathcal{F}_f(n^{-1/2})$. With these notations,

$$\sqrt{z_{ni}} - 1 = h(t_{ni})\dot{l}_{ni}(f,h), \quad i = 1, ..., n, \tag{6.7}$$

where $n^{1/2}h \in \mathcal{H}(\beta, L)$. Condition (R2) and $\left\|n^{1/2}h\right\|_\infty \leq L$ imply

$$E_f^n(\sqrt{z_{ni}} - 1)^2 \leq cn^{-1}, \quad i = 1, ..., n. \tag{6.8}$$

Inserting these bounds into (6.6) and then invoking the bounds obtained in (6.5), we obtain

$$\prod_{i=1}^n E_f^n \exp\left(n^{\alpha/2}\eta'_{ni}\right) \leq \exp(c_1) \leq c_2.$$

Then, since $\alpha > \alpha'$, from (6.4), we the estimate $I_1 = O(n^{-\alpha'})$ follows. In the same way we establish a bound for the lower tail probability.

Now consider the second term on the right-hand side of (6.3). For this we note that by (6.7),

$$\sum_{i=1}^{n} E_f^n \left|\sqrt{z_{ni}} - 1\right|^{2\delta} \leq cn^{-\delta} \sum_{i=1}^{n} E_f^n |\dot{l}_{ni}(f,h)|^{2\delta} \leq cn^{1-\delta}. \tag{6.9}$$

By virtue of (6.1) and (6.9),

$$\sum_{i=1}^{n} E_f^n \xi''_{ni} \leq cn^{\alpha(\delta-1)} \sum_{i=1}^{n} E_f^n \left|\sqrt{z_{ni}} - 1\right|^{2\delta} \leq cn^{\alpha(\delta-1)} n^{1-\delta} = cn^{-2\alpha}.$$

Then since $\frac{1}{2}n^{-\alpha'/2} - cn^{-2\alpha}$ is positive for n large enough, we get

$$\begin{aligned} I_2 &\equiv P_f^n \left(\left| \sum_{i=1}^{n} \eta''_{ni} \right| > \frac{1}{2} n^{-\alpha'/2} \right) \leq P_f^n \left(\sum_{i=1}^{n} \xi''_{ni} > \frac{1}{2} n^{-\alpha'/2} - cn^{-2\alpha} \right) \\ &\leq P_f^n \left(\max_{1\leq i\leq n} \left|\sqrt{z_{ni}} - 1\right| > n^{-\alpha/2} \right). \end{aligned}$$

The last probability can be bounded, using (6.9), in the following way: for any absolute constant $c > 0$,

(6.10)

$$P_f^n \left(\max_{1\leq i\leq n} \left|\sqrt{z_{ni}} - 1\right| > cn^{-\alpha/2} \right) \leq c_1 n^{\alpha\delta} \sum_{i=1}^{n} E_f^n \left|\sqrt{z_{ni}} - 1\right|^{2\delta} \leq c_2 n^{\alpha\delta} n^{1-\delta} = c_2 n^{-\alpha}.$$

This yields $I_2 = O(n^{-\alpha'})$. The bounds for I_1 (with the corresponding bound of the lower tail) and for I_2, in conjunction with (6.3), obviously imply the assertion. ∎

Lemma 6.4. *Assume that condition (R2) holds true. Then, there is an $\alpha \in (1/2\beta, 1)$, such that*

$$\sup_{f\in\mathcal{F}} \sup_{h\in\mathcal{F}_f(n^{-1/2})} P\left(|\Psi_n(f,h)| > n^{-\alpha/2}\right) = O\left(n^{-\alpha}\right).$$

Proof. We keep the notations from Lemma 6.3. Additionally set for $i = 1, ..., n$,

$$\psi_{ni} = \log(1 + (\sqrt{z_{ni}} - 1)) - 2(\sqrt{z_{ni}} - 1) + (\sqrt{z_{ni}} - 1)^2.$$

Then we can represent $\Psi_n(f,h)$ as follows: $\Psi_n(f,h) = \Psi_1 + \Psi_2$, where

$$\Psi_1 = \sum_{i=1}^{n} \psi_{ni} \mathbf{1}\left(\left|\sqrt{z_{ni}} - 1\right| \leq n^{-\alpha/2}\right), \quad \Psi_2 = \sum_{i=1}^{n} \psi_{ni} \mathbf{1}\left(\left|\sqrt{z_{ni}} - 1\right| > n^{-\alpha/2}\right).$$

Assume that n is large enough so that $n^{-\alpha/2} \leq 1/2$. Then a simple Taylor expansion gives $|\psi_{ni}| \leq c\left|\sqrt{z_{ni}} - 1\right|^3$, provided that $\left|\sqrt{z_{ni}} - 1\right| \leq n^{-\alpha/2}$. This in turn implies

$$|\Psi_1| \leq c \max_{1\leq i\leq n} \left|\sqrt{z_{ni}} - 1\right| \sum_{i=1}^{n} (\sqrt{z_{ni}} - 1)^2.$$

Since by (6.8) one has $E_f^n(\sqrt{z_{ni}}-1)^2 \le cn^{-1}$, we obtain

$$|\Psi_1| \le c_1 \max_{1\le i\le n} |\sqrt{z_{ni}}-1| \left(|Y_n| + c_2\right).$$

Therefore

$$P\left(|\Psi_1| \ge \frac{1}{2} n^{-\alpha'/2}\right) \le P\left(|Y_n| > n^{-\alpha'/2}\right) + P\left(\max_{1\le i\le n} |\sqrt{z_{ni}}-1| > cn^{-\alpha'/2}\right).$$

Now from Lemma 6.3 and (6.10) we obtain the bound

$$P\left(|\Psi_1| \ge n^{-\alpha'/2}\right) = O(n^{-\alpha'}). \tag{6.11}$$

As to Ψ_2, we have

$$\left\{|\Psi_2| > \frac{1}{2} n^{-\alpha'/2}\right\} \subset \left\{\max_{1\le i\le n} |\sqrt{z_{ni}}-1| > cn^{-\alpha'/2}\right\},$$

from which we deduce, by (6.10),

$$P\left(|\Psi_2| > \frac{1}{2} n^{-\alpha'/2}\right) \le P\left(\max_{1\le i\le n} |\sqrt{z_{ni}}-1| > cn^{-\alpha'/2}\right) = O(n^{-\alpha'}). \tag{6.12}$$

The result follows from (6.11) and (6.12). ∎

6.2. **Proof of Theorem 6.1.** We split the proof into two lemmas, in such a way that Theorem 6.1 follows immediately from Proposition 6.2 and these lemmas.

Set

$$M_n(f,h) = \sum_{i=1}^n h(t_{ni}) \dot{l}_{ni}(f),$$

where $\dot{l}_{ni}(f) = \dot{l}(X_i, f(t_{ni}))$ and $\dot{l}(x,\theta)$ is the tangent vector defined by (3.4).

Lemma 6.5. *Assume that conditions (R1-R3) hold true. Then there is an $\alpha \in (1/2\beta, 1)$ such that for $\mathcal{F} = \Sigma^\beta$*

$$\sup_{f\in\mathcal{F}} \sup_{h\in\mathcal{F}_f(n^{-1/2})} P_f^n\left(|2X_n(f,h) - M_n(f,h)| > n^{-\alpha/2}\right) = O\left(n^{-\alpha}\right).$$

Proof. Let $\delta_1 \in (\frac{1}{2\beta}, 1)$ and $\delta_2 \in (\frac{2\beta+1}{2\beta-1}, \infty)$ be respectively the real numbers for which conditions (R1) and (R2) hold true, where $\beta \ge \frac{1}{2}$. Let $\alpha = \min\{\delta_1, \frac{\delta_2-1}{\delta_2+1}\}$, which clearly is in the interval $(\frac{1}{2\beta}, 1)$. Then there is an $\alpha' \in (\frac{1}{2\beta}, 1)$ such that $\alpha' < \alpha$. Denote for $i = 1, ..., n$,

$$\xi_{ni} = 2\left(\sqrt{z_{ni}}-1\right) - 2E_f^n\left(\sqrt{z_{ni}}-1\right) - h(t_{ni})\dot{l}_{ni}(f), \tag{6.13}$$

$$\xi'_{ni} = \xi_{ni} 1(|\xi_{ni}| \le n^{-\alpha/2}), \quad \xi''_{ni} = \xi_{ni} 1(|\xi_{ni}| > n^{-\alpha/2}) \tag{6.14}$$

and

$$\eta'_{ni} = \xi'_{ni} - E_f^n \xi'_{ni}, \quad \eta''_{ni} = \xi''_{ni} - E_f^n \xi''_{ni}. \tag{6.15}$$

With these notations

$$2X_n(f,h) - M_n(f,h) = \sum_{i=1}^n \eta'_{ni} + \sum_{i=1}^n \eta''_{ni}. \tag{6.16}$$

Consider the first term on the right-hand side of (6.16). Since the r.v.'s $n^{\alpha/2}\eta_{ni}$ are bounded by 2, we have by Lemma 9.2 with $\lambda = 1$,

$$E_f^n \exp\left(n^{\alpha/2}\eta'_{ni}\right) \le \exp\left(cn^\alpha E_f^n(\eta'_{ni})^2\right), \quad i = 1, ..., n.$$

Then

$$I_1 \equiv P_f^n\left(\sum_{i=1}^n \eta'_{ni} > \frac{1}{2}n^{-\alpha'/2}\right) \le \exp\left(-\frac{1}{2}n^{(\alpha-\alpha')/2} + cn^\alpha \sum_{i=1}^n E_f^n(\eta'_{ni})^2\right). \tag{6.17}$$

From assumption (R1) we easily obtain

$$\sum_{i=1}^n E_f^n(\eta'_{ni})^2 \le 2\sum_{i=1}^n E_f^n \xi_{ni}^2 \le cn^{-2\delta_1} \le cn^{-2\alpha},$$

which in conjunction with (6.17) implies $I_1 = O(n^{-\alpha'})$. The bound for the lower tail can be established analogously.

For an estimate of the second term in the right-hand side of (6.16), we note that (6.13), (6.14), (6.15) imply

$$\sum_{i=1}^n E_f^n \left|\eta''_{ni}\right|^2 \le cn^{\alpha(\delta-1)}\left(\sum_{i=1}^n E_f^n \left|\sqrt{z_{ni}} - 1\right|^{2\delta} + \frac{c}{n^\delta}\sum_{i=1}^n E_f^n |\dot{l}_{ni}(f)|^{2\delta}\right).$$

Now assumption (R1) and (6.9) imply

$$\sum_{i=1}^n E_f^n \left|\eta''_{ni}\right|^{2\delta} \le cn^{\alpha(\delta-1)-1-\delta} = cn^{-2\alpha}.$$

Thus we obtain

$$I_2 \equiv P_f^n\left(\sum_{i=1}^n \eta''_{ni} > \frac{1}{2}n^{-\alpha'/2}\right) \le cn^{\alpha'}\sum_{i=1}^n E_f^n \left|\eta''_{ni}\right|^2 \le cn^{\alpha'-2\alpha} \le cn^{-\alpha'}.$$

The bounds for I_1 (with the corresponding bound of the lower tail) and for I_2, in conjunction with (6.16) imply the lemma. ∎

Lemma 6.6. *Assume that conditions (R1-R3) hold true. Then, there is an* $\alpha \in (1/2\beta, 1)$, *such that for* $\mathcal{F} = \Sigma^\beta$

$$\sup_{f\in\mathcal{F}} \sup_{h\in\mathcal{F}_f(n^{-1/2})} \left| V_n(f,h) - \frac{1}{8}\sum_{i=1}^n h(t_{ni})^2 I(f(t_{ni})) \right| \le n^{-\alpha/2}.$$

Proof. Let $\delta_1 \in (\frac{1}{2\beta}, 1)$ and $\delta_2 \in (\frac{2\beta+1}{2\beta-1}, \infty)$, be respectively the real numbers for which conditions (R1) and (R2) hold true, where $\beta \ge \frac{1}{2}$. Let $\alpha = \min\{\delta_1, \frac{\delta_2-1}{\delta_2+1}\}$. Set

$$\xi_{ni} = \sqrt{z_{ni} - 1}, \quad \eta_{ni} = \frac{1}{2}h(t_{ni})\dot{l}_{ni}(f), \quad i = 1, ..., n.$$

Then, for $i = 1, ..., n$,

$$\left|E_f^n\left(\xi_{ni}^2 - \eta_{ni}^2\right)\right|^2 \le E_f^n(\xi_{ni} - \eta_{ni})^2 E_f^n(\xi_{ni} + \eta_{ni})^2.$$

In view of assumption (R1), we have $\mathrm{E}_f^n(\xi_{ni} - \eta_{ni})^2 \le cn^{-1-\alpha}$, and assumption (R2) implies $\mathrm{E}_f^n(\xi_{ni} + \eta_{ni})^2 \le cn^{-1}$. Thus,

$$\left|E_f^n\left(\xi_{ni}^2 - \eta_{ni}^2\right)\right| \le cn^{-1-\alpha/2}, \quad i = 1, ..., n.$$

Finally

$$\left|V_n(f,h) - \frac{1}{8}\sum_{i=1}^n h(t_{ni})^2 I(f(t_{ni}))\right| = \frac{1}{2}\sum_{i=1}^n \left|E_f^n\left(\xi_{ni}^2 - \eta_{ni}^2\right)\right| \le cn^{-\alpha/2}.$$

∎

7. Proof of the local result

7.1. **Proof of Theorem 3.1.** Let $\beta > 1/2$ and $f \in \mathcal{F} = \Sigma^\beta$. Recall that γ_n is defined by (3.7). Let $\alpha \in (1/2\beta, 1)$ be the absolute constant in Theorem 6.1 and $d = \alpha - 1/2\beta \le 1$. If we set $\alpha' = 1/2\beta + qd$ where the absolute constant $q \in (0,1)$ will be specified later on, then $\alpha' \in (1/2\beta, 1)$ and $\alpha' < \alpha$. Set $\delta_n = \gamma_n^{2\alpha'}$ and $M_n = [1/\delta_n]$; then clearly $M_n = O\left(\delta_n^{-1}\right)$. Set $t_i = i/n$, $i = 0, ..., n$. Let $a_k = \max\{t_i : t_i \le \frac{k}{M_n}\}$, $k = 0, ..., M_n$. Consider a partition of the unit interval $[0,1]$ into subintervals $A_k = (a_{k-1}, a_k]$ where $k = 1, ..., M_n$. To each interval A_k we attach the affine linear map $a_k(t) : A_k \to [0,1]$ which transforms A_k into the unit interval $[0,1]$. It is clear that $|a_k(t) - a_k(s)| \le c\delta_n^{-1}|t - s|$, for $t, s \in A_k$. Denote by n_k the number of elements in the set $\{i : t_i \in A_k\}$; it obviously satisfies $n\delta_n = O(n_k)$.

Consider the local experiment $\mathcal{E}_f^n$ defined by a sequence of independent r.v.'s $X_1, ..., X_n$, where each X_i has the density $p(x, g(t_i))$ with $g = f + h$, $h \in \Sigma_f^\beta(\gamma_n)$. Since $[0,1] = \sum_{k=1}^{M_n} A_k$, we have in view of the independence of the X_i

$$\mathcal{E}_f^n = \mathcal{E}_f^{n,1} \otimes ... \otimes \mathcal{E}_f^{n,M_n},$$

where the experiment $\mathcal{E}_f^{n,k}$ is generated by those observations X_i for which $t_i \in A_k$. Set for brevity $f_k = f(a_k^{-1}(\cdot))$, $g_k = g(a_k^{-1}(\cdot))$ and $h_k = g_k - f_k$. It is easy to see that $n_k^{1/2} h_k \in \mathcal{H}(\beta, L_1)$, for some positive absolute constant L_1. This means that $h_k \in \mathcal{F}_{f_k}(n_k^{-1/2})$. Consequently

$$\mathcal{E}_f^{n,k} = \left(X^{n_k}, \mathcal{X}^{n_k}, \left\{P_{f_k+h_k}^{n_k} : h \in \Sigma_f^\beta(\gamma_n)\right\}\right),$$

where $P_s^{n_k} = P_{s(1/n_k)} \times ... \times P_{s(1)}$, for any function $s \in \mathcal{F}$, and P_θ is the distribution on $(X, \mathcal{X})$ corresponding to the density $p(x, \theta)$. It is clear that $\mathcal{E}_f^{n,k}$ is just a subexperiment of

$$\mathcal{E}_{f_k}^{n_k} = \left(X^{n_k}, \mathcal{X}^{n_k}, \left\{P_{f_k+h}^{n_k} : h \in \mathcal{F}_{f_k}(n_k^{-1/2})\right\}\right).$$

Exactly in the same way we introduce the Gaussian counterparts of $\mathcal{E}_f^{n,k}$: if $\mathcal{G}_f^n$ denotes the Gaussian experiment introduced in Theorem 3.1 then

$$\mathcal{G}_f^n = \mathcal{G}_f^{n,1} \otimes ... \otimes \mathcal{G}_f^{n,M_n},$$

where the experiment $\mathcal{G}_f^{n,k}$ is generated by those observations Y_i (see Theorem 3.1) for which $t_i \in A_k$:

$$\mathcal{G}_f^{n,k} = \left(R^{n_k}, \mathcal{B}^{n_k}, \left\{ Q_{f,h}^{n_k} : h \in \Sigma_f^{\beta}(\gamma_n) \right\} \right),$$

where $Q_{s,u}^{n_k} = Q_{s(1/n_k),u(1/n_k)} \times ... \times Q_{s(1),u(1)}$, for any functions $s, u \in \mathcal{F}$, and $Q_{\theta,\mu}$ is the normal distribution with mean μ and variance $I(\theta)^{-1}$. It is clear that $\mathcal{G}_f^{n,k}$ is just a subexperiment of

$$\mathcal{G}_{f_k}^{n_k} = \left(R^{n_k}, \mathcal{B}^{n_k}, \left\{ Q_{f_k,h}^{n_k} : h \in \mathcal{F}_{f_k}(n_k^{-1/2}) \right\} \right).$$

According to Theorems 6.1 and 5.2, for any $f \in \mathcal{F}$ there is an experiment

$$\widetilde{\mathcal{E}}_f^{n_k} = \left(\Omega^0, \mathcal{A}^0, \left\{ \widetilde{P}_{f,h}^{n_k} : h \in \mathcal{F}_f(n_k^{-1/2}) \right\} \right),$$

equivalent to $\mathcal{E}_{f_k}^{n_k}$ and an equivalent version $\widetilde{\mathcal{G}}_f^{n_k}$ of $\mathcal{G}_{f_k}^{n_k}$, defined on the same measurable space $(\Omega^0, \mathcal{A}^0)$ with measures $\widetilde{Q}_{f,h}^{n_k}$, such that uniformly in $f \in \mathcal{F}$ and $h \in \mathcal{F}_f(n_k^{-1/2})$,

$$H^2 \left(\widetilde{P}_{f,h}^{n_k}, \widetilde{Q}_{f,h}^{n_k} \right) = O\left(n_k^{-\alpha} \right) \tag{7.1}$$

for some $\alpha \in (1/2\beta, 1)$. Set

$$\begin{aligned} \widetilde{\mathcal{E}}_f^{n,k} &= \left(R^{n_k}, \mathcal{B}^{n_k}, \left\{ P_{f_k,h_k}^{n_k} : h \in \Sigma_f^{\beta}(\gamma_n) \right\} \right). \\ \widetilde{\mathcal{E}}_f^{n} &= \widetilde{\mathcal{E}}_f^{n,1} \otimes ... \otimes \widetilde{\mathcal{E}}_f^{n,M_n}. \end{aligned}$$

and define $\widetilde{\mathcal{G}}_f^n$ analogously. Since $\widetilde{\mathcal{E}}_f^{n,k}$ and $\mathcal{E}_f^{n,k}$ are (exactly) equivalent, it follows that

$$\Delta\left(\mathcal{E}_f^n, \widetilde{\mathcal{E}}_f^n \right) = \Delta\left(\mathcal{G}_f^n, \widetilde{\mathcal{G}}_f^n \right) = 0$$

which in turn implies

$$\Delta\left(\mathcal{E}_f^n, \mathcal{G}_f^n \right) = \Delta\left(\widetilde{\mathcal{E}}_f^n, \widetilde{\mathcal{G}}_f^n \right).$$

In view of (9.1), (9.3), (9.5) and (7.1), we have

$$\begin{aligned} \Delta\left(\widetilde{\mathcal{E}}_f^n, \widetilde{\mathcal{G}}_f^n \right) &\leq 2 \sum_{k=1}^{M_n} \sup_{h \in \Sigma_f^{\beta}(\gamma_n)} H^2 \left(\widetilde{P}_{f_k,h_k}^{n,k}, \widetilde{Q}_{f_k,h_k}^{n,k} \right) \\ &= O\left(M_n n_k^{-\alpha} \right) = O\left(\delta_n^{-1} (n\delta_n)^{-\alpha} \right). \end{aligned}$$

Choosing $q \leq \frac{1}{4}$, by an elementary calculation we obtain $\delta_n^{-1}(n\delta_n)^{-\alpha} = o(1)$. Thus Theorem 3.1 is proved.

7.2. **Homoscedastic form of the local result.** In order to globalize the local result in Theorem 3.1, we need to transform the heteroscedastic Gaussian approximation into a homoscedastic one. This is done by means of any transformation $\Gamma(\theta)$ of the functional parameter f which satisfies $\Gamma'(\theta) = \sqrt{I(\theta)}$, where $I(\theta)$ is the Fisher information in the parametric model $\mathcal{E}$. We shall derive the following corollary of Theorem 3.1.

Corollary 7.1. *Let $\beta > 1/2$ and $I(\theta)$ be the Fisher information in the parametric experiment $\mathcal{E}$. Assume that the density $p(x, \theta)$ satisfies the regularity conditions (R1-R3). Assume also that $I(\theta)$, as a function of θ, satisfies a Hölder condition with exponent $\alpha \in (1/2\beta, 1)$. For any $f \in \Sigma^\beta$, let $\mathcal{G}_f^n$ be the local Gaussian experiment generated by observations*

$$Y_i^n = \Gamma(g(i/n)) + \varepsilon_i, \quad i = 1, ..., n,$$

with $g = f + h$, $h \in \Sigma_f^\beta(\gamma_n)$, where $\varepsilon_1, ..., \varepsilon_n$ is a sequence of i.i.d. standard normal r.v.'s (not depending on f). Then, uniformly in $f \in \Sigma$, the sequence of local experiments $\mathcal{E}_f^n$, $n = 1, 2, ...$ is asymptotically equivalent to the sequence of local Gaussian experiments $\mathcal{G}_f^n$, $n = 1, 2, ...$:

$$\sup_{f \in \Sigma^\beta} \Delta\left(\mathcal{E}_f^n, \mathcal{G}_f^n\right) \to 0, \quad \text{as} \quad n \to \infty.$$

Proof. It will be shown that the Gaussian experiments $\mathcal{G}_f^{1,n}$ and $\mathcal{G}_f^{2,n}$ are asymptotically equivalent, where $\mathcal{G}_f^{1,n}$ is generated by observations

$$Y_i^n = g(i/n) + I(f(i/n))^{-1/2} \varepsilon_i, \quad i = 1, ..., n \tag{7.2}$$

and $\mathcal{G}_f^{2,n}$ is generated by observations

$$Y_i^n = \Gamma(g(i/n)) + \varepsilon_i, \quad i = 1, ..., n, \tag{7.3}$$

with $g = f + h$, $h \in \Sigma_f^\beta(\gamma_n)$ and $\varepsilon_1, ..., \varepsilon_n$ being a sequence of i.i.d. standard normal r.v.'s. Since $\Gamma'(\theta) = \sqrt{I(\theta)}$ and $I(\theta)$ satisfies a Hölder condition with exponent $\alpha \in (\frac{1}{2\beta}, 1)$, a Taylor expansion yields

$$\begin{aligned} \Gamma(\theta + u) - \Gamma(\theta) &= u\sqrt{I(\theta)} + u\left(\sqrt{I(\theta + u)} - \sqrt{I(\theta)}\right) \\ &= u\sqrt{I(\theta)} + o(|u|^{1+\alpha}). \end{aligned}$$

Then, taking into account (3.7), we arrive at

$$\Gamma(g(i/n)) - \Gamma(f(i/n)) = h(i/n)\sqrt{I(f(i/n))} + o(n^{-1/2}).$$

Set for brevity $m_i^1 = \Gamma(g(i/n)) - \Gamma(f(i/n))$ and $m_i^2 = h(i/n)\sqrt{I(f(i/n))}$. Let $Q_{f,h}^{1,n}$ and $Q_{f+h}^{2,n}$ be the probability measures induced by (7.2) and (7.3). Then, using (9.5) and (9.6), the Hellinger distance between $Q_{f,h}^{1,n}$ and $Q_{f+h}^{2,n}$ can easily be seen to satisfy

$$\frac{1}{2} H^2\left(Q_{f,h}^{1,n}, Q_{f+h}^{2,n}\right) = 1 - \exp\left(-\frac{1}{8}\sum_{i=1}^n \left(m_i^1 - m_i^2\right)^2\right) = o(1), \quad n \to \infty.$$

The claim on Le Cam distance convergence now follows from (9.1) and (9.3). ∎

8. Proof of the global result

In this section we shall prove Theorem 3.2.

Let $\mathcal{E}^n$ and $\mathcal{G}^n$ be the global experiments defined in Theorem 3.2. Let $f \in \Sigma$ (we shall omit the superscript β from notation Σ^β and $\Sigma_f^\beta(\gamma_n)$) Denote by J' and J'' the sets of odd and even numbers, respectively, in $J = \{1, ..., n\}$. Put

$$X'^{,n} = \prod_{i \in J'} X^{(i)}, \quad X''^{,n} = \prod_{i \in J''} X^{(i)}, \quad R''^{,n} = \prod_{i \in J''} R^{(i)}, \quad \mathbf{S}^n = \prod_{i=1}^{n} \mathbf{S}_i,$$

where $X^{(i)} = X$, $R^{(i)} = \mathrm{R}$, $\mathbf{S}_i = X$ if i is odd and $\mathbf{S}_i = \mathrm{R}$ if i is even, $i \in J$. Consider the following product (local) experiments corresponding to observations with even indices $i \in J$:

$$\mathcal{E}_f''^{,n} = \bigotimes_{i \in J''} \mathcal{E}_f^{(i)}, \quad \mathcal{G}_f''^{,n} = \bigotimes_{i \in J''} \mathcal{G}_f^{(i)},$$

where

$$\begin{aligned} \mathcal{E}_f^{(i)} &= \left(X, \mathcal{X}, \left\{P_{g(i/n)} : g = f + h, \ h \in \Sigma_f(\gamma_n)\right\}\right), \\ \mathcal{G}_f^{(i)} &= \left(\ \mathrm{R}, \mathcal{B}, \left\{Q_{g(i/n)} : g = f + h, \ h \in \Sigma_f(\gamma_n)\right\}\right). \end{aligned}$$

Along with this introduce the global experiments

$$\mathcal{E}'^{,n} = \bigotimes_{i \in J'} \mathcal{E}^{(i)}, \quad \mathcal{F}^n = \bigotimes_{i=1}^{n} \mathcal{F}^{(i)},$$

where,

$$\mathcal{F}^{(i)} = \begin{cases} \mathcal{E}^{(i)}, & \text{if} \quad i \quad \text{is odd}, \\ \mathcal{G}^{(i)}, & \text{if} \quad i \quad \text{is even}, \end{cases}$$

and

$$\mathcal{E}^{(i)} = \left(X, \mathcal{X}, \left\{P_{f(t_i)} : f \in \Sigma\right\}\right), \quad \mathcal{G}^{(i)} = \left(\mathrm{R}, \mathcal{B}, \left\{Q_{f(t_i)} : f \in \Sigma\right\}\right).$$

It is clear that

$$\mathcal{F}^n = \left(\mathbf{S}^n, \mathcal{B}(\mathbf{S}^n), \left\{F_f^n : f \in \Sigma\right\}\right),$$

where $F_f^n = F_f^{(1)} \times ... \times F_f^{(n)}$,

$$F_f^{(i)} = \begin{cases} P_{f(i/n)}, & \text{if} \quad i \quad \text{is odd}, \\ Q_{f(i/n)}, & \text{if} \quad i \quad \text{is even}, \end{cases}$$

for $i \in J$. We will show that the global experiments $\mathcal{E}^n$ and $\mathcal{F}^n$ are asymptotically equivalent. Toward this end, we note that by Corollary 7.1 the local experiments $\mathcal{E}_f''^{,n}$ and $\mathcal{G}_f''^{,n}$ are asymptotically equivalent uniformly in $f \in \Sigma$:

$$\sup_{f \in \Sigma^\beta} \Delta\left(\mathcal{E}_f''^{,n}, \mathcal{G}_f''^{,n}\right) = o(1). \tag{8.1}$$

Let $\|\cdot\|_{\mathrm{Var}}$denote the total variation norm for measures and let $P_g^{\prime\prime,n}$, $Q_g^{\prime\prime,n}$ be the product measures corresponding to the local experiments $\mathcal{E}_f^{\prime\prime,n}$ and $\mathcal{G}_f^{\prime\prime,n}$:

$$P_g^{\prime\prime,n} = \bigotimes_{i\in J''} P_{g(t_i)}, \quad Q_g^{\prime\prime,n} = \bigotimes_{i\in J''} Q_{g(t_i)}. \tag{8.2}$$

Then (8.1) implies that for any $f \in \Sigma$ there is a Markov kernel K_f^n such that

$$\sup_{f\in\Sigma} \sup_{h\in\Sigma_f(\gamma_n)} \left\| K_f^n \cdot P_{f+h}^{\prime\prime,n} - Q_{f+h}^{\prime\prime,n} \right\|_{\mathrm{Var}} = o\left(1\right). \tag{8.3}$$

Let us establish that there is a Markov kernel M^n (not depending on f) such that

$$\sup_{f\in\Sigma} \left\| M^n \cdot P_f^n - F_f^n \right\|_{\mathrm{Var}} = o\left(1\right). \tag{8.4}$$

First note that any vector $x \in X^n$ can be represented as (x', x'') where x' and x'' are the corresponding vectors in $X^{\prime,n}$ and $X^{\prime\prime,n}$. The same applies for $s \in \mathbf{S}^n : s = (x', y'')$, where $x' \in X^{\prime,n}$ and $y'' \in R^{\prime\prime,n}$. For any $x = (x', x'') \in X^n$ and $B \in \mathcal{B}(\mathbf{S}^n)$ set

$$M^n(x, B) = \int_{R^{\prime\prime,n}} \mathbf{1}_B\left((x', y'')\right) K^n_{\widehat{f}_n(x')}(x'', dy''),$$

where $\widehat{f}_n(x')$ is the preliminary estimator provided by Assumption (G1). It is easy to see that

$$\begin{aligned} \left(M^n \cdot P_f^n\right)(B) &= \int_{X^{\prime,n}} \int_{X^{\prime\prime,n}} M^n\left((x', x''), B\right) P_f^{\prime\prime,n}(dx'') P_f^{\prime,n}(dx') \\ &= \int_{X^{\prime,n}} \int_{R^{\prime\prime,n}} \mathbf{1}_B\left((x', y'')\right) \left(K^n_{\widehat{f}_n(x')} \cdot P_f^{\prime\prime,n}\right)(dy'') P_f^{\prime,n}(dx') \end{aligned} \tag{8.5}$$

and

$$F_f^n(B) = \int_{X^{\prime,n}} \int_{X^{\prime\prime,n}} \mathbf{1}_B\left((x', y'')\right) Q_f^{\prime\prime,n}(dy'') P_f^{\prime,n}(dx'), \tag{8.6}$$

where $P_f^{\prime,n}$ is the measure in the experiment $\mathcal{E}^{\prime,n}$ defined by the analogy with $P_f^{\prime\prime,n}$ in (8.2), but with J' replacing J''. By Assumption (G1),

$$\sup_{f\in\Sigma} P_f^{\prime,n}(A_f^c) = o\left(1\right), \tag{8.7}$$

where $A_f = \left\{ x' \in X^{\prime,n} : \left\| \widehat{f}_n(x') - f \right\|_\infty \le c\gamma_n \right\}$. Then (8.5) and (8.6) imply

$$\begin{aligned} &\left| \left(M^n \cdot P_f^n\right)(B) - F_f^n(B) \right| \le 2P_f^{\prime,n}(A_f^c) \\ &+ \int_{A_f} \sup_{f\in\Sigma} \sup_{h\in\Sigma_f^\beta(\gamma_n)} \left\| K_f^n \cdot P_{f+h}^{\prime\prime,n} - Q_{f+h}^{\prime\prime,n} \right\|_{\mathrm{Var}} P_f^{\prime,n}(dx'). \end{aligned}$$

Using (8.3) and (8.7) we obtain (8.4). This implies that the one-sided deficiency $\delta\left(\mathcal{E}^n, \mathcal{F}^n\right)$ is less that $c_2\varepsilon_n$. The bound for $\delta\left(\mathcal{F}^n, \mathcal{E}^n\right)$ can be obtained in the same way; for this we need a result analogous to condition (G1) in the Gaussian experiment $\mathcal{G}^n$. Since the function Γ is smooth and strictly monotone, the existence of a such preliminary estimator in $\mathcal{G}^n$ follows from Korostelev ([10]). This proves that the Le Cam distance between $\mathcal{E}^n$ and $\mathcal{F}^n$ goes to 0. In the same way we can show that $\mathcal{F}^n$

and $\mathcal{G}^n$ are asymptotically equivalent. As a result we obtain asymptotic equivalence of the experiments $\mathcal{E}^n$ and $\mathcal{G}^n$. Theorem 3.2 is proved.

9. Appendix

9.1. **Komlós-Major-Tusnády approximation.** Assume that on the probability space $(\Omega', \mathcal{F}', P')$ we are given a sequence of independent r.v.'s $X_1, ..., X_n$ such that for all $i = 1, ..., n$

$$E'X_i = 0, \quad c_{\min} \leq E'X_i^2 \leq c_{\max},$$

for some positive absolute constants $c_{\min}$ and $c_{\max}$. Hereafter E' denotes the expectation under the measure P'. Assume that the following condition, due to Sakhanenko, is satisfied: there is a sequence λ_n, $n = 1, 2, ...$ of real numbers, $0 < \lambda_n \leq \lambda_{\max}$, where $\lambda_{\max}$ is an absolute constant, such that for all $i = 1, ..., n$,

$$\lambda_n E' \left| X_i \right|^3 \exp\left(\lambda_n \left| X_i \right| \right) \leq E'X_i^2.$$

Along with this, assume that on another probability space $(\Omega, \mathcal{F}, P)$ we are given a sequence of independent normal r.v.'s $N_1, ..., N_n$, such that, for any $i = 1, ..., n$,

$$EN_i = 0, \quad EN_i^2 = E'X_i^2.$$

Let $\mathcal{H}(1/2, L)$ be a Hölder ball with exponent $1/2$ on the unit interval $[0, 1]$, i.e. the set of functions satisfying:

$$|f(x) - f(y)| \leq L \left| x - y \right|^{1/2}, \quad |f(x)| \leq L, \quad x, y \in [0, 1],$$

where L is an absolute constant.

In the sequel, the notation $\mathcal{L}(X) = \mathcal{L}(Y)$ for r.v.'s means equality of their distributions.

The following assertion is proved in Grama and Nussbaum [8].

Theorem 9.1. *A sequence of independent r.v.'s $\widetilde{X}_1, ..., \widetilde{X}_n$ can be constructed on the probability space $(\Omega, \mathcal{F}, P)$ such that $\mathcal{L}(\widetilde{X}_i) = \mathcal{L}(X_i)$, $i = 1, ..., n$, and*

$$\sup_{f \in \mathcal{H}(1/2, L)} P\left(\left| \sum_{i=1}^{n} f(i/n) \left(\widetilde{X}_i - N_i \right) \right| > x (\log n)^2 \right) \leq c_0 \exp\left(-c_1 \lambda_n x \right), \quad x \geq 0,$$

where c_0 and c_1 are absolute constants.

This result is an analog of the functional strong approximation established by Koltchinskii [9] for the empirical processes.

Remark 9.1. Note that the r.v.'s $X_1, ..., X_n$ are not assumed to be identically distributed. We also do not assume any additional richness of the probability space $(\Omega, \mathcal{F}, P)$: the only assumption is that the normal r.v.'s $N_1, ..., N_n$ exist.

9.2. **Le Cam and Hellinger distances.** We made use of the following facts.

a. Let $\mathcal{E} = (X, \mathcal{X}, \{P_\theta : \theta \in \Theta\})$ and $\mathcal{G} = (Y, \mathcal{Y}, \{Q_\theta : \theta \in \Theta\})$ be two experiments with the same parameter space Θ. Assume that there is some point $\theta_0 \in \Theta$ such that $P_\theta \ll P_{\theta_0}$ and $Q_\theta \ll Q_{\theta_0}$, $\theta \in \Theta$ and that there are versions Λ_θ^1 and Λ_θ^2 of the likelihoods dP_θ/dP_{θ_0} and dQ_θ/dQ_{θ_0} on a common probability space $(\Omega, \mathcal{A}, P)$. Then the Le Cam deficiency distance between $\mathcal{E}$ and $\mathcal{G}$ satisfies

$$\Delta(\mathcal{E}, \mathcal{G}) \leq \sup_{\theta \in \Theta} \frac{1}{2} E_P \left| \Lambda_\theta^1 - \Lambda_\theta^2 \right|. \tag{9.1}$$

The proof of this assertion can be found in Le Cam and Yang [13], p. 16 (see also Nussbaum [15]).

Let $H(\cdot,\cdot)$ denotes the Hellinger distance between probability measures: if P and Q are probability measures on the measurable space $(\Omega, \mathcal{A})$ and $P \ll \nu$ and $Q \ll \nu$, where ν is a σ-finite measure on $(\Omega, \mathcal{A})$, then

$$H^2(P,Q) = \frac{1}{2} \int_\Omega \left(\left(\frac{dP}{d\nu} \right)^{1/2} - \left(\frac{dQ}{d\nu} \right)^{1/2} \right)^2 d\nu. \tag{9.2}$$

Define the measures $\widetilde{P}_\theta$ and $\widetilde{Q}_\theta$ by setting $d\widetilde{P}_\theta = \Lambda_\theta^1 dP$ and $d\widetilde{Q}_\theta = \Lambda_\theta^2 dP$. Using the well-known relation of the L_1 norm to $H(\cdot,\cdot)$ (see Strasser [17], 2.15)

$$\frac{1}{2} E_P \left| \Lambda_\theta^1 - \Lambda_\theta^2 \right| \leq \sqrt{2} H(\widetilde{P}_\theta, \widetilde{Q}_\theta). \tag{9.3}$$

b. Let $P_1, ..., P_n$ and $Q_1, ..., Q_n$ be probability measures on $(\Omega, \mathcal{A})$. Set $P^n = P_1 \times ... \times P_n$ and $Q^n = Q_1 \times ... \times Q_n$. Then

$$1 - H^2(P^n, Q^n) = \prod_{i=1}^n \left(1 - H^2(P_i, Q_i)\right) \tag{9.4}$$

and (cf. Strasser [17], 2.17)

$$H^2(P^n, Q^n) \leq \sum_{i=1}^n H^2(P_i, Q_i). \tag{9.5}$$

c. Let Φ_μ be the normal distribution with mean μ and variance 1. Then

$$H^2(\Phi_{\mu_1}, \Phi_{\mu_2}) = 1 - \exp\left(-\frac{1}{8}(\mu_1 - \mu_2)^2 \right). \tag{9.6}$$

9.3. **An exponential inequality.** We made use of the following well-known inequality.

Lemma 9.2. *Let ξ be a r.v. such that $E\xi = 0$ and $|\xi| \leq a$, for some positive constant a. Then*

$$E \exp(\lambda \xi) \leq \exp\left(c\lambda^2 E\xi^2\right), \quad |\lambda| \leq 1,$$

where $c = e^a/2$.

Proof. Set $\mu(\lambda) = E\exp(\lambda\xi)$ and $\psi(\lambda) = \log\mu(\lambda)$. A simple Taylor expansion yields

$$\psi(\lambda) = \psi(0) + \lambda\psi'(0) + \frac{\lambda^2}{2}\psi''(t), \quad |\lambda| \leq 1, \tag{9.7}$$

where $|t| \leq 1$, $\psi(0) = 0$, $\psi'(0) = 0$,

$$\psi''(t) = \frac{\mu''(t)}{\mu(t)} - \frac{\mu'(t)^2}{\mu(t)^2} \leq \mu''(t) \leq e^a E\xi^2. \tag{9.8}$$

Inserting (9.8) in (9.7) we get $\psi(\lambda) \leq \frac{1}{2}e^a \mathrm{E}\xi^2$. ∎

References

[1] Berkes, I. and Philipp, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Prob.* **7**, 29-54.

[2] Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory.* IMS Lecture Notes - Monograph Series, Vol. 9, Institute of Mathematical Statistics, Hayward, California.

[3] Brown, L. D. and Low, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24**, 2384-2398.

[4] Brown, L. D. and Zhang, C.-H. (1998). Asymptotic nonequivalence of nonparametric experiments when the smoothness index is 1/2. *Ann. Statist.* **26**, 279-287.

[5] Brown, L. D., Low M. and Zhang, C.-H. (2000). Asymptotic equivalence theory for nonparametric regression with random design. Unpublished manuscript (to appear, *Ann. Statist.*)

[6] Efromovich, S. and Samarov, A. (1996). Asymptotic equivalence of nonparametric regression and white noise has its limits. *Statist. Probab. Lett.* **28**, 143-145.

[7] Grama I. and Nussbaum M. (1998). Asymptotic equivalence for generalized linear models, *Probability Theory and Related Fields,* Vol. **111** 167-214.

[8] Grama, I. and Nussbaum, M. (2001). A functional Hungarian construction for sums of independent random variables. Unpublished manuscript.

[9] Koltchinskii, V. I. (1994). Komlós-Major-Tusnády approximation for the general empirical process and Haar expansions of classes of functions. *Journal of Theoretical Probability.* **7**, 73-118.

[10] Korostelev, A. (1993). Exact asymptotically minimax estimator for nonparametric regression in the uniform norm. *Theory Probab. Appl.* **38** 775–782

[11] Korostelev, A. and Nussbaum, M. (1999). The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli* **5** 1099-1118

[12] Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New-York etc.

[13] Le Cam, L. and Yang, G. L. (2000). *Asymptotics in Statistics: Some Basic Concepts. 2nd Ed.* Springer-Verlag, New York

[14] Millar, P. W. (1983). The minimax principle in asymptotic statistical theory. *Ecole d'Eté de probabilités de Saint-Flour XI-1981. Lecture Notes in Mathematics.* **976.** 76-267.

[15] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and white noise. *Ann. Statist.* **24** 2399-2430.

[16] Stone, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040-1053

[17] Strasser, H. (1985). *Mathematical Theory of Statistics.* Walter de Gruyter, Berlin-New York.

(Grama, I.), Université de Bretagne-Sud, Laboratoire SABRES, Rue Ives Mainguy, 56000 Vannes, France

E-mail address: ion.grama@univ-ubs.fr

(Nussbaum, M.), Department of Mathematics, Mallot Hall, Cornell University, Ithaca, NY 14853-4201, USA
E-mail address: nussbaum@math.cornell.edu