Asymptotic Equivalence for Nonparametric Generalized Linear Models

Ion Grama[1] and Michael Nussbaum
Institute of Mathematics, Chisinau, Moldova and
Weierstrass Institute, Berlin, Germany

January 1997

Abstract

We establish that a non-Gaussian nonparametric regression model is asymptotically equivalent to a regression model with Gaussian noise. The approximation is in the sense of Le Cam's deficiency distance Δ; the models are then asymptotically equivalent for all purposes of statistical decision with bounded loss. Our result concerns a sequence of independent but not identically distributed observations with each distribution in the same real-indexed exponential family. The canonical parameter is a value $f(t_i)$ of a regression function f at a grid point t_i (nonparametric GLM). When f is in a Hölder ball with exponent $\beta > \frac{1}{2}$, we establish global asymptotic equivalence to observations of a signal $\Gamma(f(t))$ in Gaussian white noise, where Γ is related to a variance stabilizing transformation in the exponential family. The result is a regression analog of the recently established Gaussian approximation for the i. i. d. model. The proof is based on a functional version of the Hungarian construction for the partial sum process.

1 Introduction

The remarkable success of the Le Cam's asymptotic theory is mostly due to the power of the concept of *weak convergence* of statistical experiments, which can be established via LAN conditions. Weak convergence takes place for experiments localized at the normalizer rate for the underlying central limit theorem, i. e. the usual $n^{-1/2}$. However this is useless if estimators have a slower rate of convergence, i. e. when the problem becomes "ill posed" as in many nonparametric estimation problems. It is therefore natural to abandon the localization concept in this case and to replace limits of experiments by approximations in the sense of Le Cam's deficiency pseudodistance Δ. The Δ-distance can be accessed via coupling of likelihood processes and new results on strong approximation for sums of random variables. We refer to Koltchinskii [22] for a result on the empirical processes in the i. i. d. case.

---

[1] Research supported by the Deutsche Akademische Austauschdienst and by the Deutsche Forschungsgemeinschaft.
1990 *Mathematics Subject Classification.* Primary 62B15; Secondary 62G07, 62G20
Key words and phrases. Nonparametric regression, statistical experiment, deficiency distance, global white noise approximation, exponential family, variance stabilizing transformation.

The global Δ-distance for nonparametric Gaussian experiments was first studied by Brown and Low [5], who showed that a normal nonparametric regression is asymptotically equivalent to its continuous version - the signal recovery problem in Gaussian white noise. Then in Nussbaum [31] it was established that density estimation from i. i. d. data on an interval is asymptotically equivalent to the signal recovery problem, where the signal is the root density. The two sequences of experiments are then *accompanying* in the sense that their deficiency pseudodistance Δ tends to zero. This can be regarded as the natural generalization of the classical local asymptotic normality theory to "ill posed" problems. The implication for decision theory is "automatic" transfer of risk bounds from one sequence to another.

The purpose of the present paper is to accomplish a logical next step in these developments, i. e. to treat the case of *non-Gaussian nonparametric regression.* Our model is such that at points $t_i = i/n$, $i = 1, ..., n$ we observe independent r. v.'s X_i which follow a distribution from an exponential family $\mathcal{P}$ with parameters $\theta_i = f(t_i) \in \Theta$, where $f : [0,1] \to \Theta$ is an "unknown" function to be estimated. The function f is assumed to belong to the smoothness class Σ. The main result of the paper is asymptotic equivalence of this model to a Gaussian experiment of the *homoscedastic* form

$$dY_t^n = \Gamma\left(f(t)\right) dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0,1], \tag{1.1}$$

with $f \in \Sigma$, where the one-to-one transformation $\Gamma(\theta) : \Theta \to R$ is entirely determined by the local exponential family $\mathcal{P}$ (see Section 3.3 for a precise formula for Γ). Here W is the standard Wiener process, and f runs a set of functions in a Hölder ball with exponent $\beta > \frac{1}{2}$.

Note that our function f is tied to the canonical parametrization of the exponential family, while the "natural" parameter (the intensity for the Poisson case etc.) is generally different. But there is a smooth parameter transformation $\lambda = b(\theta)$ (defined in Section 2.2 below) which permits to formulate global results of the type (1.1) in "natural" regression models. Some examples are:

[1] Poisson case: X_i is Poisson$(g(t_i))$, where g is a function on $[0,1]$ in a Hölder ball with exponent $\beta > \frac{1}{2}$, with values in $[\epsilon, \epsilon^{-1}]$ for some $\epsilon > 0$. The Poisson intensity is $\lambda = b(\theta)$, where b is a strictly increasing smooth function. Defining the function $F(\lambda) = \Gamma(b^{-1}(\lambda))$, we obtain (see Section 4 below) $F(\lambda) = 2\sqrt{\lambda}$. The accompanying Gaussian experiment is

$$dY_t^n = 2\sqrt{g(t)}\, dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0,1]. \tag{1.2}$$

[2] Bernoulli case: X_i is Binomial$(1, g(t_i))$, g as above but with values in $[\epsilon, 1-\epsilon]$ for some $\epsilon > 0$. The natural parameter is $\lambda = b(\theta)$ for some function b with properties as above. We have $F(\lambda) = 2\arcsin(\sqrt{\lambda})$, and the accompanying Gaussian experiment is

$$dY_t^n = 2\arcsin\sqrt{g(t)}\, dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0,1].$$

[3] Gaussian variance case: X_i is $N(0, g(t_i))$, g as in example 1. We have $F(\lambda) = 2^{-1/2}\log\lambda$, and

$$dY_t^n = \frac{1}{\sqrt{2}} \log g(t)\, dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0,1]. \tag{1.3}$$

For more details see Section 4. We chose to give here the continuous versions of the accompanying experiments, but the discrete versions are also available.

The motivation for this paper can be concisely expressed in three points.

[A] The first example essentially recovers the result of [31] on i. i. d. data with density f on the unit interval, since the proof in [31] used a Poisson approximation. In the second example, $F(\lambda) = 2\arcsin(\sqrt{\lambda})$ is the well known variance stabilizing transformation for the binomial distribution. For scale parameters like the Gaussian variance, the logarithm is also known to be a stabilizing transformation, and the same applies to $2\lambda^{1/2}$ in the Poisson case. Thus, in the evolving theory of asymptotic equivalence of experiments, we have achieved a better understanding of where global closed form approximations like (1.1) arise from. The formal connection to differential geometric theory in statistics seems very interesting and remains to be explored (cp. Čencov [8], Amari et al. [1]).

[B] The case where the function f is in a linear parametric class $\{f(x) = \beta x,\ \beta \in R\}$ is known as a *generalized linear model* (GLM). The inverse of the transformation $b(\theta)$ would then be the canonical link function. Accordingly, our model is of the type *nonparametric GLM*, cp. Green and Silverman [19], sec. 5.1.2. Models like these, which offer tremendous flexibility, have received much attention in the recent literature, see also Fan and Gijbels [14]. It would be beyond the scope of this paper to treat the many (semiparametric) variants and extensions; we refer to [19] and [14]. In particular we do not discuss logit and probit analysis in our context (cp. example 3, p. 92 in [19]). Empirical process theory has also been applied by Mammen and van de Geer [28] in our model (in a more general variant), for constructing estimators.

[C] There are implications for time series models. Example 3 leads on to the white noise equivalence for the *spectral density model* for a Gaussian stationary sequence; cf. Golubev and Nussbaum [17]. Furthermore, Example 3 is related to discrete observations of a diffusion process:

$$dY_t^n = g^{1/2}(Y_t^n, t)dW_t,\ Y_0^n = y_0,\ t \in [0,1].$$

Suppose observations in points $t_i = i/n$, $i = 1, \ldots, n$, where g is unknown. Nonparametric estimation of g has recently been considered by Genon-Catalot, Laredo and Picard [16], Florens-Zmirou [15]. Example 3 might be seen as a possible pilot result for those models, where the distributions of processes on $[0,1]$ are mutually orthogonal and the asymptotics is given by grid refinement.

The standard method of proof is to establish first a local version of (1.1) and then to globalize it by means of a preliminary estimator. We obtain our initial local approximation in the *heteroscedastic* form

$$dY_t^n = f(t)\,dt + \frac{1}{\sqrt{n}} I\left(f_0(t)\right)^{-1/2} dW_t, \quad t \in [0,1], \tag{1.4}$$

where f is in a shrinking neighborhood of a function f_0 and $I(\theta)$ is the Fisher information in the local exponential family $\mathcal{P}$ (given in its canonical form). To obtain a global asymptotic equivalence, the function f_0 which was technically assumed "known" has to be replaced by a preliminary estimator. However, the homoscedastic form (1.1) can be obtained only if the function f_0 does not show up explicitly in the local approximation. The problem arises to find a transformation $\Gamma(\theta)$ on the parameter space of the local exponential family $\mathcal{P}$ such that an asymptotically equivalent form of (1.4) would be (1.1), with f in a neighborhood of f_0. It is known that such a transformation exists for any exponential family $\mathcal{P}$. The problem is related to that of a variance-stabilizing transformation. Indeed, for the Poisson observations of example [1], one can prove easily that an accompanying local experiment, besides (1.4), is also given by

$$dY_t^n = g(t)\,dt + \frac{1}{\sqrt{n}}\sqrt{g_0(t)}dW_t, \quad t \in [0,1], \tag{1.5}$$

with g in a neighborhood of g_0 (for an analogy with the i. i. d. model see also Nussbaum [31]). The observations in (1.5) have roughly a Poisson character - the expectation is $g(t)$ and the variance is approximately also $g(t)$ (since g is in a neighborhood of g_0). For the Poisson distribution, the square root is a variance stabilizing transformation, which agrees with (1.2).

For the proof of the local heteroscedastic approximation (1.4) we establish a functional Hungarian construction for the partial sums of independent but nonidentically distributed random variables. The result is similar to that of Koltchinskii [22] for the empirical process, but the assumption of non-identity and non-smoothness of distributions of the summands substantially complicate the problem. This is treated separately in [18] based on methods developed in Sakhanenko [35]. Due to the particularly simple structure of our nonparametric exponential model, we can straightforwardly apply our strong approximation result to couple the likelihood process with that of an appropriately chosen Gaussian experiment. Again *coupling of likelihood processes* is the key idea for proving asymptotic equivalence, as in [31]. However we would like to mention that our KMT result is useful in more general situations also.

An essential step in proving a local approximation result like (1.4) is to study the local experiments generated by the fragments of observations X_i over shrinking time intervals. These experiments we call *doubly local.* With an appropriate choice of the length of the shrinking interval and after rescaling it to the unit interval, a doubly local experiment can be viewed as local experiment on the interval $[0,1]$ of the usual type, but now with a neighborhood of the "almost root-n" size $(n/\log n)^{-1/2}$. A similar renormalization technique is known to be effective for pointwise estimation in nonparametrics, cf. Donoho and Liu [10] and Low [27]. We also refer to Millar [30] for $n^{-1/2}$ -shrinking neighborhoods in the context of nonparametric estimation. Were it not for the log-factor in $(n/\log n)^{-1/2}$, these rescaled experiments on the interval $[0,1]$ would converge to a Gaussian limit in the sense of Δ. The motivation for applying the Hungarian construction at this stage is, roughly speaking, to obtain a good rate for this convergence, i. e. $n^{-1/2}$ up to some logarithmic factor. We thus implicitly address a question of Le Cam on rates for convergence of experiments (cp. the remark on p. 509 of [25]). The use of the Hungarian construction for this purpose is in line with its original motivation, i. e. optimal rates in the functional central limit theorem for the partial sum process.

Our results are formulated in terms of Δ-distance approximations; we do not exhibit the recipes (Markov kernels) which transfer a decision function in one experiment to the other. The problem of constructive equivalence is an important issue and some promising research in this direction is going on (Brown and Low [6]). Markov kernels in the present case can be extracted from the Hungarian construction, but this is beyond of the scope of the present paper.

The method of the proof is similar to that in Nussbaum [31]. We utilize the natural independence structure of our regression model for decomposing it into fragments on the appropriate shrinking intervals (or equivalently into doubly local experiments). We then apply the functional version of the KMT construction for establishing a Gaussian local approximation for the fragments. It is important is to have this approximation with a rate of convergence which is enough to "beat" the number of shrinking intervals into which the whole interval $[0,1]$ has been split. Having obtained the above rate, we still have to ensure the global approximation on $[0,1]$. This we achieve by passing from the Le Cam pseudo-distance between statistical experiments to the Hellinger distance between the corresponding probability measures. The passage is made possible by a construction of the likelihood processes of the local experiments on the same probability space with a Gaussian likelihood process. The reason to work with the Hellinger distance rather than with total variation distance is the convenient behavior of the former under multiplication of probability measures (see (2.13)). This allows to patch together the doubly local experiments for obtaining the Gaussian approximation (1.4) valid globally on the interval of observations $[0,1]$, but still around a specified regression function f_0.

After that, we choose the variance-stable version of the Gaussian local approximation as a starting point for globalization over the parameter space Σ. The result is the global approximation (1.1). We trace the rates of convergence throughout, so that the rate of the deficiency distance approximation can be made explicit.

2 Background

2.1 Exponential families of distributions

We will consider a one-dimensional linearly indexed exponential family (see Brown [4] or Le Cam [25], p. 144), which is described by means of the following objects:

- A measurable space $(X, \mathcal{B}(X), \mu)$ equipped with the positive measure $\mu(dx)$, where X is a Borel measurable subset in the real line R and $\mathcal{B}(X)$ is the Borel σ-field on X;

- a measurable map $U(x) : X \to R$;

- an open (possibly infinite) interval Θ in R where the Laplace transformation

  $$L(\theta) = \int_X \exp\{\theta U(x)\}\mu(dx)$$

  is finite.

Put $V(\theta) = \log L(\theta)$. Denote by $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ the set of probability measures P_θ on the space $(X, \mathcal{B}(X), \mu)$ of the form

$$P_\theta(dx) = \exp\{\theta U(x) - V(\theta)\}\mu(dx). \tag{2.1}$$

We call $\mathcal{P}$ an *exponential family* on the space $(X, \mathcal{B}(X), \mu)$ with parameter set Θ. The family of measures $\mathcal{P}$ defines the *exponential experiment* $\mathcal{E} = (X, \mathcal{B}(X), \mathcal{P})$, which will be the background object for constructing our nonparametric model. In the case when the measures P_θ are defined by (2.1) we also say that the exponential family $\mathcal{P}$ (or the exponential experiment $\mathcal{E}$) is given in its *canonical form.*

It follows from the definition of an exponential family that the function $V(\theta)$ is analytical on Θ. Also note that, for a fixed $\theta \in \Theta$, the function $V(\theta+t) - V(\theta)$ coincides with the cumulant generating function

$$G_\theta(t) = \log \int_X \exp\{tU(x)\} P_\theta(dx)$$

of the r. v. $Y(\theta) = U(X(\theta))$, where the r. v. $X(\theta)$ has the distribution in the exponential family $\mathcal{P}$ with parameter θ. This implies that $\frac{d^m}{d\theta^m}V(\theta)$ is the cumulant of order m of the r. v. $Y(\theta)$. In particular

$$V'(\theta) = E_\theta Y(\theta), \tag{2.2}$$
$$V''(\theta) = E_\theta Y(\theta)^2 - (E_\theta Y(\theta))^2, \tag{2.3}$$
$$V'''(\theta) = E_\theta Y(\theta)^3 - 3E_\theta Y(\theta))^2 E_\theta Y(\theta)) + 2(E_\theta Y(\theta))^3. \tag{2.4}$$

For this reason we will call the function $V(\theta)$ the *cumulant generating function* associated with the exponential family $\mathcal{P}$ or with the exponential experiment $\mathcal{E}$.

Consider an exponential experiment $\mathcal{E}$. The minimal sufficient statistic in this experiment is the function $U(x)$, $x \in X$. It is easy to see that the corresponding Fisher information $I(\theta)$ is

$$I(\theta) = V''(\theta) = \int_X \frac{(p'_\theta(x))^2}{p_\theta(x)} \mu(dx), \tag{2.5}$$

where $p_\theta(x) = P_\theta(dx)/\mu(dx) = \exp\{\theta U(x) - V(\theta)\}$. For any $\theta_0 \in \Theta$ and $\varepsilon_0 > 0$ denote

$$B(\theta_0, \varepsilon_0) = \{\theta \in \Theta : |\theta - \theta_0| \le \varepsilon_0\}.$$

Throughout the paper we assume that the following conditions hold true.

- The Fisher information is positive on Θ, i. e.

$$I(\theta) > 0, \quad \theta \in \Theta. \tag{2.6}$$

- There exists a (possibly infinite) interval Θ_0 in the parameter set Θ such that

$$I_{\min} \le \inf_{\theta \in \Theta_0} I(\theta), \quad \sup_{\theta_0 \in \Theta_0} \sup_{\theta \in B(\theta_0, \varepsilon_0)} I(\theta) \le I_{\max}, \tag{2.7}$$

  where $I_{\max}$, $I_{\min}$ and ε_0 are positive constants depending only on the family $\mathcal{P}$.

We will see that condition (2.7) can easily be checked for all examples in Section 4.

Let us denote by $\overline{Y}(\theta)$ the sufficient statistic $Y(\theta) = U(X(\theta))$ centered under the measure $P_\theta : \overline{Y}(\theta) = Y(\theta) - E_\theta Y(\theta)$. The following assertions are almost trivial.

Proposition 2.1 *Assume that condition (2.7) holds true. Then for any* $|t| \le \varepsilon_0$

$$\sup_{\theta \in \Theta_0} E_\theta \exp\{t\overline{Y}(\theta)\} \le \exp\{t^2 I_{\max}/2\}.$$

Proof. For the proof it is enough to remark that for $|t| \leq \varepsilon_0$

$$\begin{aligned} E_\theta \exp\{t\overline{Y}(\theta)\} &= \exp\{V(\theta+t) - V(\theta) - tV'(\theta)\} \\ &= \exp\{\frac{1}{2}t^2 I(\theta + \lambda t)\} \\ &\leq \exp\{\frac{1}{2}t^2 I_{\max}\}, \end{aligned}$$

where $0 \leq \lambda \leq 1$. ■

Proposition 2.2 *Assume that condition (2.7) holds true. Then*

$$\sup_{\theta_0 \in \Theta_0} \sup_{\theta \in B(\theta, \varepsilon_0/2)} V'''(\theta) \leq c,$$

where c is a constant depending only on $I_{\max}$ and ε_0.

Proof. This assertion is an easy consequence of Proposition 2.1 and (2.2), (2.3), (2.4). ■

2.2 Variance stabilizing transformation

Let $X_i, i = 1, ..., n$ be a sequence of i. i. d. r. v.'s each with distribution in the exponential family $\mathcal{P}$ for the same parameter $\theta \in \Theta$. Let $V(\theta)$ be the cumulant generating function associated with the exponential family $\mathcal{P}$. Put for brevity $S_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $b(\theta) = V'(\theta) = E_\theta X_1$, $I(\theta) = V''(\theta) = Var_\theta X_1$. According to the central limit theorem the sequence $\sqrt{n}\,(S_n - b(\theta))$ converges weakly under the measure P_θ to the normal r. v. with zero mean and variance $I(\theta)$. We are interested in finding a function $F : R \to R$, the variance-stabilizing transformation, such that under the same measure P_θ

$$\sqrt{n}\,(F\,(S_n) - F\,(b(\theta))) \xrightarrow{d} N(0,1), \tag{2.8}$$

for all $\theta \in \Theta$. Such a transformation exists and is given by the equation

$$F'\,(b(\theta)) = \frac{1}{\sqrt{I(\theta)}}. \tag{2.9}$$

The straightforward arguments are similar to those in Barndorff-Nielsen and Cox [3] (p. 37) or in Andersen et al. [2] (p. 109). Indeed, it is easy to see by standard reasoning that two sequences of r. v.'s $\sqrt{n}\,(F\,(S_n) - F\,(b(\theta)))$ and $\sqrt{n}F'(b(\theta))\,(S_n - b(\theta))$ are asymptotically equivalent in the sense that under the probability P_θ

$$\sqrt{n}\,(F\,(S_n) - F(\,b(\theta)\,)) - \sqrt{n}F'(b(\theta))\,(S_n - b(\theta)) \xrightarrow{d} 0, \tag{2.10}$$

as $n \to \infty$. By the central limit theorem, under the measure P_θ

$$\sqrt{n}F'(b(\theta))\,(S_n - b(\theta)) \xrightarrow{d} N(0,1), \tag{2.11}$$

as $n \to \infty$, if the function F is chosen such that (2.9) holds true. From (2.10) and (2.11) we immediately infer the claim (2.8).

Our proof of the variance-stable form of the asymptotically equivalent Gaussian experiments follows a similar pattern, see Section 5.5.

2.3 Basic facts on statistical equivalence

Let P and Q be two probability measures on the measurable space $(\Omega, \mathcal{F})$. The Hellinger distance between the probability measures P and Q is defined as

$$H^2(P,Q) = \frac{1}{2}\mathbf{E}_\mu \left((dP/d\mu)^{1/2} - (dQ/d\mu)^{1/2} \right)^2, \tag{2.12}$$

where P and Q are absolutely continuous w.r.t. the probability measure μ.

Let $P_1, ..., P_n$ and $Q_1, ..., Q_n$, be probability measures on $(\Omega, \mathcal{F})$. Put $P^n = P_1 \times ... \times P_n$ and $Q^n = Q_1 \times ... \times Q_n$. Then (see Strasser [36])

$$H^2(P^n, Q^n) \leq \sum_{i=1}^{n} H^2(P_i, Q_i). \tag{2.13}$$

Let $\mathcal{E} = \left(\Omega^1, \mathcal{F}^1, \{P_\theta : \theta \in \Theta\}\right)$ and $\mathcal{G} = \left(\Omega^2, \mathcal{F}^2, \{Q_\theta : \theta \in \Theta\}\right)$ be two statistical experiments with the same parameter set Θ. Assume that $(\Omega^1, \mathcal{F}^1)$ and $(\Omega^2, \mathcal{F}^2)$ are complete separable (Polish) metric spaces. The deficiency of the experiment $\mathcal{E}$ with respect to the experiment $\mathcal{G}$ is defined as

$$\delta(\mathcal{E}, \mathcal{G}) = \inf \sup_{\theta \in \Theta} \|K P_\theta - Q_\theta\|,$$

where the sup is taken over the set $\mathcal{M}(\Omega^1, \mathcal{F}^2)$ of all Markov kernels K from $(\Omega^1, \mathcal{F}^1)$ to $(\Omega^2, \mathcal{F}^2)$. Le Cam's Δ- distance between the experiments $\mathcal{E}$ and $\mathcal{G}$ is defined by

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

Let $\mathcal{E}^n$ and $\mathcal{G}^n$, $n = 1, 2, ...$ be two sequences of statistical experiments. We say that $\mathcal{E}^n$ and $\mathcal{G}^n$ are *asymptotically equivalent* if

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

We will need a relation between Le Cam and Hellinger distances. Let $\mathcal{E} = (\Omega^1, \mathcal{F}^1, \{P_\theta : \theta \in \Theta\})$ and $\mathcal{G} = (\Omega^2, \mathcal{F}^2, \{Q_\theta : \theta \in \Theta\})$ be two experiments with the same parameter set Θ. Assume that there is some point $\theta_0 \in \Theta$ such that $P_\theta \ll P_{\theta_0}$ and $Q_\theta \ll Q_{\theta_0}$. Suppose that there are versions of the likelihood ratios $\Lambda^1(\theta) = P_\theta/dP_{\theta_0}$ and $\Lambda^2(\theta) = dQ_\theta/dQ_{\theta_0}$ (as processes indexed by θ) on a common probability space $(\overline{\Omega}, \overline{\mathcal{F}}, \mathbf{P})$. Then the Δ-distance satisfies the inequality

$$\Delta^2(\mathcal{E}, \mathcal{G}) \leq \sup_{\theta \in \Theta} H^2\left(\Lambda^1(\theta), \Lambda^2(\theta)\right), \tag{2.14}$$

where the Hellinger distance between likelihood ratios $\Lambda^1(\theta)$ and $\Lambda^2(\theta)$ is defined in analogy to the case of probability measures:

$$H^2\left(\Lambda^1(\theta), \Lambda^2(\theta)\right) = \frac{1}{2}\mathbf{E}\left(\sqrt{\Lambda^1(\theta)} - \sqrt{\Lambda^2(\theta)}\right)^2. \tag{2.15}$$

In particular it follows that for two experiments we construct likelihood ratios on a common probability space which *coincide* as random variables, then these experiments are equivalent. For more details we refer to Nussbaum [31], Proposition 2.2.

Denote by $\left(C_{[0,1]}, \mathcal{B}(C_{[0,1]})\right)$ the measurable space of all continuous functions on the unit interval $[0,1]$ endowed with the uniform metric and by Q_W the Wiener measure on $\left(C_{[0,1]}, \mathcal{B}(C_{[0,1]})\right)$.

Let $P^{(i)}$ $i = 1, 2$ be the Gaussian shift measures on $\left(C_{[0,1]}, \mathcal{B}(C_{[0,1]}), Q_W\right)$ induced by the following observations

$$dX_t^{(i)} = f^{(i)}(t)dt + \frac{1}{\sqrt{\sigma}}dW_t, \quad 0 \le t \le 1,$$

where $\sigma > 0$ and W is a Wiener process on $\left(C_{[0,1]}, \mathcal{B}(C_{[0,1]}), Q_W\right)$. Then the Hellinger distance between the measures $P^{(1)}$ and $P^{(2)}$ satisfies the inequality (see for instance Jacod and Shiryaev [21])

$$H^2\left(P^{(1)}, P^{(2)}\right) \le \frac{1}{8}\sigma \int_0^1 \left(f^{(1)}(t) - f^{(2)}(t)\right)^2 dt. \tag{2.16}$$

2.4 A Komlós-Major-Tusnády approximation for independent r. v.'s

Suppose 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 any $i = 1, ..., n$

$$EX_i = 0$$

and

$$C_{\min} \le EX_i^2 \le C_{\max}$$

for some constants $0 < C_{\min} < C_{\max} < \infty$. Assume also that the following Cramér condition

$$E\exp\{C_0|X_i|\} \le C_1$$

holds for $i = 1, ..., n$ with some constants $C_0 > 0$ and $1 < C_1 < \infty$. Along with this consider that on another probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ we are given a sequence of independent normal r. v.'s $N_1, ..., N_n$ with

$$\widetilde{E}N_i = 0, \quad \widetilde{E}N_i^2 = EX_i^2,$$

for $i = 1, ..., n$. Let $\mathcal{H}(\frac{1}{2}, L)$ be the Hölder ball with exponent $\frac{1}{2}$, i. e. the set of all real valued functions f defined on the unit interval $[0, 1]$ and satisfying the following conditions

$$|f(x) - f(y)| \le L|x - y|^{1/2},$$

where $L > 0$ and

$$||f||_\infty \le L/2.$$

Let $t_i = \frac{i}{n}$, $i = 1, ..., n$ be a uniform grid on the interval $[0, 1]$.

The following theorem is crucial for our results. The proof can be found in the paper by Grama and Nussbaum [18].

Theorem 2.1 *If the probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ is rich enough (cf. [18]), a sequence of independent r. v.'s $\widetilde{X}_1, ..., \widetilde{X}_n$ can be constructed on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ such that $\widetilde{X}_i \overset{d}{=} X_i$, $i = 1, ..., n$, and such that for $S_n(f)$ defined by*

$$S_n(f) = \sum_{i=1}^n f(t_i)(\widetilde{X}_i - N_i)$$

we have

$$\sup_{f \in \mathcal{H}(\frac{1}{2}, L)} \widetilde{P}(|S_n(f)| > x\log^2 n) \le c_1 \exp\{-c_2 x\}, \quad x \ge 0,$$

where and c_1, c_2 are constants depending only on $C_{\min}$, $C_{\max}$, C_0, C_1, L.

Remark 2.1 *It should be pointed out that in the above theorem the r. v.'s X_i, $i = 1, ..., n$ are not supposed to be identically distributed nor to have smooth distributions.*

Note that under these circumstances the construction of the r. v.'s $\widetilde{X}_i$, $i = 1, ..., n$ in Theorem 2.1 appears extremely difficult. That is why our construction of the asymptotically equivalent Gaussian experiments can be viewed as an existence theorem rather than a prescription for transforming the initial sample into "asymptotically equivalent Gaussian data". This is the price to pay for the fairly general setting of the model as well as for the optimal result. One can expect that a simpler construction can be performed in case of a stronger smoothness assumption on the parameter $f(t) \in \Sigma$ and/or on the distributions of the observed data, i. e. restricting somewhat the class of functions Σ and/or the class of allowed distributions. This is the path taken in Brown and Low [6]. The gain is that one can plug in the "asymptotically equivalent Gaussian data" into an optimal estimator in the accompanying Gaussian model and use it as an estimator for $f(t)$ in the initial model.

A related approach is to use the Hungarian construction for constructive purposes via closeness of sample paths of the processes, along with some "continuity" properties of estimators. As examples for the case of kernel density estimators we mention Rio [34] or Einmahl and Mason [13]. The precise relation to asymptotic equivalence theory remains to be investigated.

For more details on related subjects as well as for various versions of KMT results we refer the reader to the papers of Komlós et al. [23], [24], Csörgő and Révész [9], Sakhanenko [35], Einmahl [12], Massart [29], Rio [32], [33], [34], Koltchinskii [22], Einmahl and Mason [13], Grama and Nussbaum [18] and to the references therein.

3 Main results

3.1 Notations and formulation of the problem

Let $(X, \mathcal{B}(X), \mu)$ a measurable space equipped with the σ-finite measure $\mu(dx)$, where X is a subset of the real line R and $\mathcal{B}(X)$ is the Borel σ-field on X. Let Θ be an open (possibly infinite) interval in R and $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be an exponential family of distributions on $(X, \mathcal{B}(X), \mu)$ with parameter set Θ. The corresponding exponential experiment we denote by $\mathcal{E} = (X, \mathcal{B}(X), \{P_\theta : \theta \in \Theta\})$. Assume that condition (2.7) holds true. Let us introduce some further notations.

Let $\mathcal{J}(\Theta_0)$ be the set of all functions f, defined on the unit interval $[0,1]$ with values in the parameter set Θ_0

$$\mathcal{J}(\Theta_0) = \{f : [0,1] \to \Theta_0\}.$$

Let $\mathcal{H}(\beta, L)$ be a Hölder ball, i. e. the set of functions $f : [0,1] \to R$ which satisfy the conditions

$$|f(x)| \leq L, \quad |f^{(m)}(x) - f^{(m)}(x)| \leq L|x-y|^\alpha,$$

for $x, y \in [0,1]$, where $\beta = m + \alpha$, $0 < \alpha \leq 1$. Later we shall require $\beta \geq \frac{1}{2}$. Consider also the following set of functions:

$$\Sigma = \Sigma(\beta, L) = \mathcal{J}(\Theta_0) \cap \mathcal{H}(\beta, L);$$

this will be the basic parameter set in our model.

Let $X(\theta)$ stand for a r. v. whose distribution is in the exponential family $\mathcal{P}$, with parameter $\theta \in \Theta$. On the unit interval $[0,1]$ consider the time points $t_i = \frac{i}{n}$, $i = 1, ..., n$. Assume that we observe the sequence of r. v.'s

$$X_i = X(\theta_i), \quad i = 1, .., n, \tag{3.1}$$

with $\theta_i = f(t_i)$, where the "unknown" function f is in the set Σ. We shall prove that the statistical experiment generated by these observations is asymptotically equivalent to an experiment of observing the function f in white noise.

Let us give another formal definition of the statistical experiments related to the observed data X_i, $i = 1, ..., n$. To each time t_i we associate an exponential experiment $\mathcal{E}_{t_i}$ indexed by functions $f \in \Sigma$ as follows:

$$\mathcal{E}_{t_i} = \left(X, \mathcal{B}(X), \{ P_{f(t_i)} : f \in \Sigma \} \right).$$

Define $\mathcal{E}^n$ to be the product experiment $\mathcal{E}^n = \mathcal{E}_{t_1} \otimes ... \otimes \mathcal{E}_{t_n}$. In other words the experiment corresponding to the sequence of observations X_i, $i = 1, ..., n$ defined by (3.1) is

$$\mathcal{E}^n = \mathcal{E}(X^n) = (X^n, \mathcal{B}(X^n), \{P_f^n : f \in \Sigma\}), \tag{3.2}$$

where P_f^n is the product measure

$$P_f^n = P_{f(t_1)} \otimes ... \otimes P_{f(t_n)}, \quad f \in \Sigma. \tag{3.3}$$

The experiment defined by (3.2) (or equivalently by (3.1)) will be also called *global* to distinguish it from the *local* experiment to be introduced now.

For any fixed function f_0 in the parameter set Σ define a neighborhood by

$$\Sigma_{f_0}(\gamma_n) = \{ f \in \Sigma : ||f - f_0||_\infty \le \gamma_n \}$$

where $\gamma_n \to 0$ as $n \to \infty$. In accordance with rate of convergence results in nonparametric statistics, the shrinking rate γ_n of the neighborhood $\Sigma_{f_0}(\gamma_n)$ should be slower than $n^{-\frac{1}{2}}$. We will study the case where

$$\gamma_n = \kappa_0 (n / \log n)^{-\frac{\beta}{2\beta+1}}, \tag{3.4}$$

where β is the exponent in the Hölder ball $\mathcal{H}(\beta, L)$ and $\kappa_0 = \kappa_0(\beta)$ is some constant depending on β. This choice can be explained in the following way. The neighborhood $\Sigma_{f_0}(\gamma_n)$ with shrinking rate γ_n given by (3.4), is such that in the experiment $\mathcal{E}^n$ there exists a preliminary estimator $\widehat{f}_n$ satisfying

$$\sup_{f_0 \in \mathcal{H}(\beta, L)} P_{f_0}(\widehat{f}_n \in \Sigma_{f_0}(\gamma_n)) \to 1, \quad n \to \infty. \tag{3.5}$$

This property will be of use later when we globalize our local results.

The local experiment, which we will denote by $\mathcal{E}_{f_0}^n$ is defined as

$$\mathcal{E}_{f_0}^n = \left(X^n, \mathcal{B}(X^n), \{ P_f^n : f \in \Sigma_{f_0}(\gamma_n) \} \right). \tag{3.6}$$

Let us remark that generally speaking for the sequence of nonparametric global experiments $\mathcal{E}^n$ there is no Gaussian limit experiment in the usual weak sense, since the corresponding likelihood ratios are asymptotically degenerate. Instead it is appropriate to consider a sequence of *accompanying* Gaussian experiments $\mathcal{G}^n$, which is asymptotically equivalent to the initial sequence $\mathcal{E}^n$:

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

The same applies to the local experiments $\mathcal{E}^n_{f_0}$.

The corresponding accompanying Gaussian experiments will be introduced in subsequent sections as the results are formulated. We now describe the likelihood ratios for the experiments $\mathcal{E}^n$ and $\mathcal{E}^n_{f_0}$. Note that since the measure P_θ is absolutely continuous w.r.t. the measure $\mu(dx)$,

$$\frac{P_\theta(dx)}{\mu(dx)} = \exp\{\theta U(x) - V(\theta)\},$$

thus for any $f \in \Sigma$

$$\frac{dP^n_f}{d\mu^n} = \prod_{i=1}^{n} \exp\{f(t_i)U(X_i) - V(f(t_i))\}.$$

From this we derive that for any $f,\ f_0 \in \Sigma$ the likelihood ratio of measures P^n_f and $P^n_{f_0}$ corresponding to the experiment $\mathcal{E}^n = \mathcal{E}(X^n)$ has the form

$$\frac{dP^n_f}{dP^n_{f_0}} = \exp\{\sum_{i=0}^{n}[f(t_i) - f_0(t_i)]U(X_i) - \sum_{i=0}^{n}[V(f(t_i)) - V(f_0(t_i))]\}. \tag{3.7}$$

For the local experiment $\mathcal{E}^n_{f_0}$ the likelihood ratio has the same form (3.7) but with $f \in \Sigma_{f_0}(\gamma_n)$.

3.2 Local experiments: nonparametric neighborhoods

We start with the local framework since our global results are essentially based upon the results for local experiments. For this let $f_0 \in \Sigma$ be fixed. The corresponding local Gaussian experiment $\mathcal{G}^n_{f_0}$ is generated by the following Gaussian observations in continuous time

$$dY^n_t = f(t)dt + \frac{1}{\sqrt{n}} I\left(f_0(t)\right)^{-1/2} dW_t, \quad t \in [0,1], \quad f \in \Sigma_{f_0}(\gamma_n), \tag{3.8}$$

where W is the standard Wiener process on the probability space $\left(C_{[0,1]}, \mathcal{B}(C_{[0,1]}), Q_W\right)$. Denote by $Q^n_{f_0,f}$ the Gaussian shift measure on $\left(C_{[0,1]}, \mathcal{B}(C_{[0,1]})\right)$ induced by the observations $(Y^n_t)_{0 \le t \le 1}$ determined by (3.8). Then $\mathcal{G}^n_{f_0}$ can be defined as

$$\mathcal{G}^n_{f_0} = \left(C_{[0,1]}, \mathcal{B}(C_{[0,1]}), \left\{Q^n_{f_0,f} : f \in \Sigma_{f_0}(\gamma_n)\right\}\right). \tag{3.9}$$

Theorem 3.1 *Assume that* $\beta\frac{1}{2}$. *Then the experiments* $\mathcal{E}^n_{f_0}$ *and* $\mathcal{G}^n_{f_0}$ *are asymptotically equivalent uniformly over* f_0 *in* Σ

$$\sup_{f_0 \in \Sigma} \Delta(\mathcal{E}^n_{f_0}, \mathcal{G}^n_{f_0}) \to 0, \quad n \to \infty.$$

Moreover

$$\sup_{f_0 \in \Sigma} \Delta^2(\mathcal{E}^n_{f_0}, \mathcal{G}^n_{f_0}) \le c_1 n^{-\frac{2\beta-1}{2\beta+1}} (\log n)^{\frac{14\beta+5}{2\beta+1}},$$

where c_1 *is a constant depending only on* $I_{\min}$, $I_{\max}$, ε_0, L, β.

The limit $\frac{1}{2}$ for the smoothness index β is exact in view of an example by Brown and Zhang [7] showing that the above asymptotic equivalence fails when $\beta < \frac{1}{2}$ even for the case of Gaussian observations. We can also refer the reader to the arguments in Efromovich and Samarov [11].

We will now present a discrete version of the asymptotically equivalent Gaussian experiment. The corresponding local experiment $\mathcal{G}_{f_0}^{\prime,n}$ is generated by the Gaussian observations in discrete time

$$Y_i = f(t_i) + I\left(f_0(t_i)\right)^{-\frac{1}{2}} \varepsilon_i, \quad i = 1, ..., n, \quad f \in \Sigma_{f_0}(\gamma_n), \tag{3.10}$$

where ε_i are standard normal r. v.'s. If we denote by $Q_{f_0(t_i),f(t_i)}$ the Gaussian measure corresponding to one observation Y_i of the form (3.10), i. e. the Gaussian measure on real line with mean $f(t_i)$ and variance $I\left(f_0(t_i)\right)^{-1}$, then $\mathcal{G}_{f_0}^{\prime,n}$ can be defined as

$$\mathcal{G}_{f_0}^{\prime,n} = \left(R^n, \mathcal{B}(R^n), \left\{Q_{f_0,f}^{\prime,n} : f \in \Sigma_{f_0}(\gamma_n)\right\}\right),$$

where

$$Q_{f_0,f}^{\prime,n} = Q_{f_0(t_1),f(t_1)} \otimes ... \otimes Q_{f_0(t_n),f(t_n)}, \quad f \in \Sigma_{f_0}(\gamma_n). \tag{3.11}$$

Theorem 3.2 *Assume that $\beta > \frac{1}{2}$. Then the experiments $\mathcal{E}_{f_0}^n$ and $\mathcal{G}_{f_0}^{\prime,n}$ are asymptotically equivalent uniformly over f_0 in Σ. Moreover*

$$\sup_{f_0 \in \Sigma} \Delta^2\left(\mathcal{E}_{f_0}^n, \mathcal{G}_{f_0}^{\prime,n}\right) \le c_1 n^{-\frac{2\beta-1}{2\beta+1}} (\log n)^{\frac{14\beta+5}{2\beta+1}},$$

where c_1 is a constant depending only on $I_{\min}$, $I_{\max}$, ε_0, L, β.

It is easy to see that Theorem 3.1 is a consequence of Theorem 3.2 in view of results of Brown and Low [5]. Although the rate argument is not developed there, it can easily be made explicit.

3.3 Variance-stable form of the local approximation

Generally speaking there are no reasons to assume that the center of the neighborhood $\Sigma_{f_0}(\gamma_n)$ is known to a statistician doing nonparametric inference. That is why sometimes we would prefer to have another form of the asymptotically equivalent Gaussian experiment $\mathcal{G}_{f_0}^n$, in which the expression $I(f_0(t))$ does not appear. It turns out that such a form of the Gaussian accompanying experiment does exist. We will call it *variance-stable form*, since it involves the variance-stabilizing transformation pertaining to the exponential family. To introduce it we need some notations.

Let as before $V(\theta)$ be the cumulant generating function associated with the exponential experiment $\mathcal{E} = (X, \mathcal{B}(X), \{P_\theta : \theta \in \Theta\})$. Put for brevity $b(\theta) = V'(\theta)$, $\theta \in \Theta$. It follows from the assumption (2.6) that $b(\theta)$ is an increasing differentiable function on the open interval Θ. Denote by Λ the range of $b(\theta)$, i. e. $\Lambda = \{b(\theta) : \theta \in \Theta\}$. It is clear that Λ is also an open interval in R. Let $a(\lambda)$, $\lambda \in \Lambda$ be the inverse of $b(\theta)$, $\theta \in \Theta$, i. e.

$$a(\lambda) = \inf\{\theta \in \Theta : b(\theta) > \lambda\}, \quad \lambda \in \Lambda,$$

which obviously is an increasing differentiable function on Λ.

Another equivalent way to define $a(\lambda)$ would be to put $a(\lambda) = T'(\lambda)$, where $T(\lambda)$ is the Legendre transformation (or the rate function) of the function $V(\theta)$:

$$T(\lambda) = \inf \{\lambda\theta - V(\theta) : \lambda \in \Lambda\}.$$

It is also easy to see that $a(\lambda)$ satisfies the equation

$$a'(\lambda) = I(a(\lambda))^{-1}, \quad \lambda \in \Lambda. \tag{3.12}$$

Let $F(\lambda)$ be any function on Λ, having the property

$$F'(\lambda) = \sqrt{a'(\lambda)}, \quad \lambda \in \Lambda. \tag{3.13}$$

The relations (3.12) and (3.13) show that $F(\lambda)$, $\lambda \in \Lambda$ coincides with the variance-stabilizing transformation defined in Section 2.2. The functions $b(\theta) : \Theta \to \Lambda$ and $F(\lambda) : \Lambda \to R$ define a transformation of Θ into the real line as follows

$$\Gamma(\theta) = F(b(\theta)) : \Theta \to R. \tag{3.14}$$

Let $\widehat{\mathcal{G}}^n_{f_0}$ be the Gaussian experiment generated by observations

$$d\widehat{Y}^n_t = \Gamma(f(t))\, dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0,1], \quad f \in \Sigma_{f_0}(\gamma_n), \tag{3.15}$$

i. e.

$$\widehat{\mathcal{G}}^n_{f_0} = \left(C_{[0,1]}, \mathcal{B}(C_{[0,1]}), \left\{ \widehat{Q}^n_f : f \in \Sigma_{f_0}(\gamma_n) \right\} \right),$$

where $\widehat{Q}^n_f$ is the Gaussian shift measure on $\left(C_{[0,1]}, \mathcal{B}(C_{[0,1]})\right)$ induced by the observations $(\widehat{Y}^n_t)_{0 \le t \le 1}$ determined by (3.15). Let $\mathcal{G}^n_{f_0}$ be the Gaussian experiment defined in (3.8) and (3.9).

Theorem 3.3 *Let $\beta > \frac{1}{2}$. Then the experiments $\mathcal{G}^n_{f_0}$ and $\widehat{\mathcal{G}}^n_{f_0}$ are asymptotically equivalent uniformly in $f_0 \in \Sigma$. Moreover the Le Cam distance between $\mathcal{G}^n_{f_0}$ and $\widehat{\mathcal{G}}^n_{f_0}$ satisfies*

$$\sup_{f_0 \in \Sigma} \Delta^2 \left(\mathcal{G}^n_{f_0}, \widehat{\mathcal{G}}^n_{f_0} \right) \le c_1 n^{-\frac{2\beta-1}{2\beta+1}} (\log n)^{\frac{4\beta}{2\beta+1}},$$

where c_1 is a constant depending only on $I_{\max}$ and ε_0.

As an immediate consequence of Theorems 3.1 and 3.3 we get the following result.

Theorem 3.4 *Let $\beta > \frac{1}{2}$. Then the experiments $\mathcal{E}^n_{f_0}$ and $\widehat{\mathcal{G}}^n_{f_0}$ are asymptotically equivalent uniformly in $f_0 \in \Sigma$ and the Le Cam distance between $\mathcal{E}^n_{f_0}$ and $\widehat{\mathcal{G}}^n_{f_0}$ satisfies*

$$\sup_{f_0 \in \Sigma} \Delta^2 \left(\mathcal{E}^n_{f_0}, \widehat{\mathcal{G}}^n_{f_0} \right) \le c_1 n^{-\frac{2\beta-1}{2\beta+1}} (\log n)^{\frac{14\beta+5}{2\beta+1}},$$

where c_1 is a constant depending only on $I_{\max}$, $I_{\min}$, ε_0, L, β.

We turn to the discrete version of the asymptotically equivalent Gaussian experiment. The local experiment $\widehat{\mathcal{G}}_{f_0}^{\prime,n}$ is generated by the Gaussian observations in discrete time

$$Y_i = \Gamma\left(f(t_i)\right) + \varepsilon_i, \quad f \in \Sigma_{f_0}(\gamma_n), \quad i = 1, ..., n, \tag{3.16}$$

where ε_i are standard normal r. v.'s. If we denote by $\widehat{Q}_{f(t_i)}$ the Gaussian measure corresponding to one observation Y_i of the form (3.16), i. e. the Gaussian measure on the real line with mean $\Gamma\left(f(t_i)\right)$ and variance 1, then $\widehat{\mathcal{G}}_{f_0}^{\prime,n}$ can be defined as

$$\widehat{\mathcal{G}}_{f_0}^{\prime,n} = \left(R^n, \mathcal{B}(R^n), \left\{\widehat{Q}_f^{\prime,n} : f \in \Sigma_{f_0}(\gamma_n)\right\}\right),$$

where

$$\widehat{Q}_f^{\prime,n} = \widehat{Q}_{f(t_1)} \otimes ... \otimes \widehat{Q}_{f(t_n)}, \quad f \in \Sigma_{f_0}(\gamma_n).$$

Theorem 3.5 *Assume that $\beta > \frac{1}{2}$. Then the experiments $\mathcal{E}_{f_0}^n$ and $\widehat{\mathcal{G}}_{f_0}^{\prime,n}$ are asymptotically equivalent uniformly over f_0 in Σ. Moreover*

$$\sup_{f_0 \in \Sigma} \Delta^2\left(\mathcal{E}_{f_0}^n, \widehat{\mathcal{G}}_{f_0}^{\prime,n}\right) \leq c_1 n^{-\frac{2\beta-1}{2\beta+1}} (\log n)^{\frac{14\beta+5}{2\beta+1}},$$

where c_1 is a constant depending only on $I_{\min}$, $I_{\max}$, ε_0, L, β.

In the above theorems the initial exponential experiment $\mathcal{E}$ which generates $\mathcal{E}_{f_0}^n$, is assumed to be in its canonical form. The variance-stable Gaussian approximation appears in an equivalent but a little more pleasant form (as we will see in Section 4) if the experiment $\mathcal{E}$ is *naturally parametrized,* i. e. if $\mathcal{E}$ given in its canonical form by (2.1) is reparametrized by means of the one-to-one map $\lambda = b(\theta) : \Theta_0 \to \Lambda_0$. Introduce the set of functions

$$\overline{\Sigma} = \{g = b \circ f : f \in \Sigma\}$$

and for any $g_0 \in \overline{\Sigma}$ the neighborhoods

$$\overline{\Sigma}_{g_0}(\gamma_n) = \left\{g \in \overline{\Sigma} : \|g - g_0\|_\infty \leq \overline{c}_0 \gamma_n\right\},$$

with some constant $\overline{c}_0$ depending $I_{\max}$, ε_0. Let $\overline{\mathcal{E}}_{g_0}^n$ be the corresponding nonparametric product experiment, defined analogously to (3.6) and (3.3):

$$\overline{\mathcal{E}}_{g_0}^n = \left(X^n, \mathcal{B}(X^n), \left\{\overline{P}_g^n : g \in \overline{\Sigma}_{g_0}(\gamma_n)\right\}\right),$$

with $\overline{P}_g^n = P_{b \circ f}^n$, where P_f^n, with $f = a \circ g$ is the product measure defined by (3.3). The accompanying Gaussian experiment $\overline{\mathcal{G}}_{g_0}^n$ is defined by the observations

$$d\overline{Y}_t^n = F(g(t)) + \frac{1}{\sqrt{n}} W_t, \quad t \in [0,1], \quad g \in \overline{\Sigma}_{g_0}(\gamma_n).$$

Theorem 3.6 *Let $\beta > \frac{1}{2}$. Then the experiments $\overline{\mathcal{E}}_{g_0}^n$ and $\overline{\mathcal{G}}_{g_0}^n$ are asymptotically equivalent uniformly in $g_0 \in \overline{\Sigma}$ and the Le Cam distance between $\overline{\mathcal{E}}_{g_0}^n$ and $\overline{\mathcal{G}}_{g_0}^n$ satisfies*

$$\sup_{g_0 \in \overline{\Sigma}} \Delta^2\left(\overline{\mathcal{E}}_{g_0}^n, \overline{\mathcal{G}}_{g_0}^n\right) \leq c_1 n^{-\frac{2\beta-1}{2\beta+1}} (\log n)^{\frac{14\beta+5}{2\beta+1}},$$

where c_1 is a constant depending only on $I_{\max}$, $I_{\min}$, ε_0, L, β, $\overline{c}_0$.

The discrete version of this theorem is straightforward.

3.4 Local experiments: almost $n^{-1/2}$- neighborhoods

It turns out that the key point in the study of asymptotic equivalence of local experiments is the behavior of its fragments over shrinking time intervals of length

$$\delta_n = \gamma_n^{1/\beta} = \kappa_0^{1/\beta}(n/\log n)^{-\frac{1}{2\beta+1}}.$$

After a rescaling we arrive at local experiments parametrized by functions f of the form $f = f_0 + \gamma_n^* g$, where $g \in \Sigma$ and

$$\gamma_n^* = \kappa_0^*(n/\log n)^{-\frac{1}{2}}, \tag{3.17}$$

with some $\kappa_0^* > 0$. We will present the corresponding results since they are of independent interest. Before stating these results we introduce the necessary notations.

Let $f_0 \in \Sigma$. For any $g \in \Sigma$ put

$$P_{f_0,g}^n = P_{f_0+\gamma_n^* g}^n$$

and consider the local experiment

$$\mathcal{E}_{f_0}^{*,n} = \left(X^n, \mathcal{B}(X^n), \{P_{f_0,g}^n : g \in \Sigma\}\right). \tag{3.18}$$

In the same way we introduce the accompanying sequence of Gaussian experiments:

$$\mathcal{G}_{f_0}^{*,n} = \left(C_{[0,1]}^n, \mathcal{B}(C_{[0,1]}^n), \left\{Q_{f_0,g}^{*,n} : g \in \Sigma\right\}\right),$$

where

$$Q_{f_0,g}^{*,n} = Q_{f_0,f_0+\gamma_n^* g}^n,$$

with $Q_{f_0,f}^n$ from (3.9).

Theorem 3.7 *Assume that $\beta > \frac{1}{2}$. Then the experiments $\mathcal{E}_{f_0}^{*,n}$ and $\mathcal{G}_{f_0}^{*,n}$ are asymptotically equivalent uniformly over f_0 in Σ. Moreover*

$$\sup_{f_0 \in \Sigma} \Delta^2\left(\mathcal{E}_{f_0}^{*,n}, \mathcal{G}_{f_0}^{*,n}\right) \leq c_1 n^{-1}(\log n)^7,$$

where c_1 is a constant depending only on $I_{\min}$, $I_{\max}$, ε_0, L, β.

Let $\mathcal{G}_{f_0}^{\prime,*,n}$ be the corresponding local experiment generated by the Gaussian observations in discrete time Y_i, $i = 1, ..., n$ defined by (3.10) with $f = f_0 + \gamma_n^* g$, $g \in \Sigma$. More precisely, let

$$Q_{f_0,g}^{\prime,*,n} = Q_{f_0,f_0+\gamma_n^* g}^{\prime,n}$$

with $Q_{f_0,f}^{\prime,n}$ from (3.11) and set

$$\mathcal{G}_{f_0}^{\prime,*,n} = \left(R^n, \mathcal{B}(R^n), \left\{Q_{f_0,g}^{\prime,*,n} : g \in \Sigma\right\}\right). \tag{3.19}$$

Theorem 3.8 *Assume that $\beta > \frac{1}{2}$. Then the experiments $\mathcal{E}_{f_0}^{*,n}$ and $\mathcal{G}_{f_0}^{\prime,*,n}$ are asymptotically equivalent uniformly over f_0 in Σ. Moreover*

$$\sup_{f_0 \in \Sigma} \Delta^2(\mathcal{E}_{f_0}^{*,n}, \mathcal{G}_{f_0}^{\prime,*,n}) \leq c_1 n^{-1}(\log n)^7,$$

where c_1 is a constant depending only on $I_{\min}$, $I_{\max}$, ε_0, L, β.

3.5 Global experiments

The variance-stable form of the accompanying *local* Gaussian experiments allows to construct an accompanying *global* Gaussian experiment. The main idea is to substitute a preliminary estimator satisfying (3.5) for the unknown function f_0 around which the local experiment is built. Such an estimator is provided by Lemma 6.1 (see Section 6.1). In the variance-stable form of the local experiment given by (3.15) (or (3.16) in the discrete case) the unknown function f_0 does not show up in the distributions themselves, but appears only as a center of the parametric neighborhood. This will imply in the sequel that the globalized experiment does not depend on the specific preliminary estimator used; thus a convenient closed form global approximation for the original regression experiment $\mathcal{E}^n$ can be obtained.

Let $\Gamma(\theta) : \Theta_0 \to R$ be the transformation given by (3.14). Let the global Gaussian experiment $\mathcal{G}^n$ be defined as

$$\mathcal{G}^n = \left(R^n, \mathcal{B}(R^n), \left\{\widehat{Q}_f^n : f \in \Sigma\right\}\right),$$

where $\widehat{Q}_f^n$ is the Gaussian shift measure induced by the observations

$$dY_t^n = \Gamma\left(f(t)\right) dt + \frac{1}{\sqrt{n}} dW_t, \quad f \in \Sigma, \quad t \in [0,1].$$

Theorem 3.9 *Let $\beta > \frac{1}{2}$. Then the global experiments $\mathcal{E}^n$ and $\mathcal{G}^n$ are asymptotically equivalent: $\Delta(\mathcal{E}^n, \mathcal{G}^n) \to 0$ as $n \to \infty$. Moreover for $\beta \in (\frac{1}{2}, 1)$*

$$\Delta^2(\mathcal{E}^n, \mathcal{G}^n) \le cn^{-\frac{2\beta-1}{2\beta+1}} (\log n)^{\frac{14\beta+5}{2\beta+1}},$$

where c is a constant depending only on $I_{\max}$, $I_{\min}$, κ_0, L, β.

As a particular case this theorem gives us the main result in Brown and Low [5]. We will discuss the case of normal observations and other examples of interest in the next section.

We present also a discrete version of the asymptotically equivalent global Gaussian experiment in its variance-stable form. The Gaussian experiment $\mathcal{G}^{\prime,n}$ is generated by the Gaussian observations in discrete time

$$Y_i = \Gamma\left(f(t_i)\right) + \varepsilon_i, \quad f \in \Sigma, \quad i = 1, ..., n \tag{3.20}$$

where ε_i are standard normal r.v's. If we denote by $\widehat{Q}_{f(t_i)}$ the Gaussian measure corresponding to one observation Y_i of the form (3.20), $i = 1, ..., n$, then $\mathcal{G}^{\prime,n}$ can be defined as

$$\mathcal{G}^{\prime,n} = \left(R^n, \mathcal{B}(R^n), \left\{\widehat{Q}_f^{\prime,n} : f \in \Sigma\right\}\right), \tag{3.21}$$

where

$$\widehat{Q}_f^{\prime,n} = \widehat{Q}_{f(t_1)} \otimes ... \otimes \widehat{Q}_{f(t_n)}, \quad f \in \Sigma. \tag{3.22}$$

Theorem 3.10 *Let $\beta > \frac{1}{2}$. Then the global experiments $\mathcal{E}^n$ and $\mathcal{G}^{\prime,n}$ are asymptotically equivalent: $\Delta(\mathcal{E}^n, \mathcal{G}^{\prime,n}) \to 0$ as $n \to \infty$. Moreover for $\beta \in (\frac{1}{2}, 1)$*

$$\Delta^2(\mathcal{E}^n, \mathcal{G}^n) \le cn^{-\frac{2\beta-1}{2\beta+1}} (\log n)^{\frac{14\beta+5}{2\beta+1}},$$

where c is a constant depending only on $I_{\max}$, $I_{\min}$, κ_0, L, β.

The case when $\mathcal{E}$ is naturally parametrized is similar to Theorem 3.6.

The proofs for the passage from local to global are in Section 6.

4 Examples and applications

The most striking form for the asymptotically equivalent Gaussian approximation in the following examples will be obtained if the initial exponential experiment $\mathcal{E}$ is taken under its *natural parametrization.* By a natural parametrization we mean the following. Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be an exponential family on $(X, \mathcal{B}(X), \mu)$ in the canonical form, i. e. whose Radon-Nikodym derivatives $dP_\theta/d\mu$ are defined by (2.1). Let $V(\theta)$ be its cumulant generating function. It is clear (see also Section 3.3) that $b(\theta) = V'(\theta)$ is a one-to-one map from Θ_0 to Λ_0. If we reparametrize the family $\mathcal{P}$ by means of the map $b(\theta)$, we will call the family $\mathcal{P}$ *naturally parametrized.* Indeed the parameter $\lambda = b(\theta)$ is the "natural" parameter in many specific families: mean or variance for normal distributions, intensity for exponential, gamma, or Poisson distributions, probability of success for Bernoully and binomial distributions etc..

4.1 Gaussian observations: unknown mean

Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be the family of normal distributions on the real line $X = R$ with mean $\theta \in \Theta = R$ and variance 1. The normal distribution $P_\theta(dx)$ can be written

$$P_\theta(dx) = e^{\theta x - V(\theta)} \mu(dx),$$

$\mu(dx)$ being the standard normal distribution

$$\mu(dx) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

and $U(x) = x$, $V(\theta) = \theta^2/2$. Then the corresponding Fisher information $I(\theta) = V''(\theta) \equiv 1$, so the condition (2.7) holds true with $\Theta_0 = R$, $I_{\max} = I_{\min} = 1$, $\varepsilon_0 > 0$. Hence the parameter set Σ coincides with the Hölder ball $\mathcal{H}(\beta, L)$.

Assume that our observations $X_i = X(\theta_i)$, $i = 1, ..., n$ are normal with mean $\theta_i = f(t_i)$, $t_i = i/n$ and variance 1, where the function f is in the Hölder ball $\Sigma = \mathcal{H}(\beta, L)$. Let us remark that these observations correspond to the regression model

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

with standard normal r. v.'s ε_i, $i = 1, ..., n$.

Let $\mathcal{E}^n_{f_0}$ be the local experiment generated by the sequence of observations X_i, $i = 1, ..., n$, with $f \in \Sigma_{f_0}(\gamma_n)$, for some $f_0 \in \Sigma$. According to Theorem 3.1 for any $\beta > \frac{1}{2}$ the experiment $\mathcal{E}^n_{f_0}$ is asymptotically equivalent to the local experiment $\mathcal{G}^n_{f_0}$ generated by the observations

$$dY^n_t = f(t)dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0, 1], \tag{4.1}$$

where $f \in \Sigma_{f_0}(\gamma_n)$.

The global form of the asymptotically equivalent Gaussian experiment is given also by (4.1) but with $f \in \Sigma$. Thus we recover the main result of Brown and Low [5].

4.2 Gaussian observations: unknown variance

Let $\mathcal{P} = \{P_\lambda : \lambda \in \Lambda\}$ be the family of normal distributions on the real line $X = R$ with mean 0 and variance $\lambda \in \Lambda \equiv [0, \infty)$. The normal distribution $P_\lambda(dx)$ has the form

$$P_\lambda(dx) = \frac{1}{\sqrt{2\pi\lambda}} \exp\{-\frac{x^2}{2\lambda}\} dx. \tag{4.2}$$

After the reparametrization $\theta = -1/\lambda$, $\theta \in \Theta = (-\infty, 0]$, we obtain the linearly indexed exponential model

$$P_\theta(dx) = \exp\{\theta U(x) - V(\theta)\}\mu(dx), \tag{4.3}$$

where $\mu(dx)$ is the Lebesgue measure on the real line and $U(x) = x^2/2$, $V(\theta) = -\frac{1}{2}\log(-\frac{\theta}{2\pi})$. The corresponding Fisher information is $I(\theta) = V''(\theta) = \frac{1}{2}\theta^{-2}$, so condition (2.7) holds true with $\Theta_0 = [\theta_{\min}, \theta_{\max}]$ for some constants $-\infty < \theta_{\min} < \theta_{\max} < 0$ and ε_0 small enough. Then the parameter set Σ contains all the functions $f(t)$ from the Hölder ball $\mathcal{H}(\beta, L)$, which satisfy $\theta_{\min} \leq f(t) \leq \theta_{\max}$.

Consider a sequence of normal observations

$$X_i = X(\theta_i), \quad \theta_i = f(t_i), \quad i = 1, ..., n, \tag{4.4}$$

with $t_i = i/n$, where the unknown function f is in the set Σ. Let $f_0(t) \in \Sigma$ and let $\mathcal{E}^n_{f_0}$ be the local experiment generated by the observations (4.4). Then by Theorem 3.1 the experiment $\mathcal{E}^n_{f_0}$ is asymptotically equivalent to the Gaussian experiment $\mathcal{G}^n_{f_0}$ generated by the observations

$$dY^n_t = f(t)dt - \frac{\sqrt{2}}{\sqrt{n}} f_0(t) dW_t, \quad t \in [0, 1],$$

for $f \in \Sigma_{f_0}(\gamma_n)$, W being the standard Wiener process.

For the variance-stable form we easily compute $b(\theta) = V'(\theta) = -\frac{1}{2\theta}$ and $F(\lambda) = 2^{-1/2}\log\lambda$. Thus by Theorem 3.3 the variance-stable accompanying Gaussian experiment is given by

$$dY^n_t = \frac{1}{\sqrt{2}} \log\left(-\frac{1}{2f(t)}\right) dt - \frac{1}{\sqrt{n}} dW_t, \quad t \in [0, 1]. \tag{4.5}$$

Note that in the above formula $f(t)$ and$f_0(t)$ are less than $\theta_{\max} < 0$.

A more compact form for the accompanying Gaussian experiments is obtained using the natural parametrization. If we reparametrize the exponential family (4.3) by means of the map $b(\theta) = -\frac{1}{2\theta}$, we recover its original form given by (4.2). Let $\overline{\Sigma} = \overline{\Sigma}(\beta, L)$ be the set of all functions g from the Hölder ball $\mathcal{H}(\beta, L)$, which satisfy $\lambda_{\min} \leq g(t) \leq \lambda_{\max}$, where $0 < \lambda_{\min} < \lambda_{\max} < \infty$. For any function $g_0 \in \overline{\Sigma}$ let $\overline{\Sigma}_{g_0}(\gamma_n)$ be its neighborhood of radius γ_n (see (3.4)) in $\overline{\Sigma}$. Denote by $\overline{\mathcal{E}}^n_{g_0}$ the local experiment generated by the observations (4.4) with $f(t) = -1/g(t)$, $g \in \overline{\Sigma}_{g_0}(\gamma_n)$. This experiment can be regarded as generated by the observations

$$X_i = \sqrt{g(t_i)}\varepsilon_i, \quad t \in [0, 1],$$

with $g \in \overline{\Sigma}_{g_0}(\gamma_n)$, where ε_i are standard normal r. v.'s. Then for any $\beta > \frac{1}{2}$ the experiment $\overline{\mathcal{E}}^n_{g_0}$ is asymptotically equivalent to the local Gaussian experiment generated by observations

$$dY^n_t = -\frac{1}{g(t)} dt + \frac{\sqrt{2}}{\sqrt{n} g_0(t)} dW_t, \quad 0 \leq t \leq 1,$$

with $g \in \overline{\Sigma}_{g_0}(\gamma_n)$, W being standard Wiener process on $(C_{[0,1]}, \mathcal{B}(C_{[0,1]}), Q_W)$.

The corresponding variance-stable form is determined by the equation (cf. Theorem 3.6)

$$dY^n_t = \frac{1}{\sqrt{2}} \log g(t) dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0, 1]. \tag{4.6}$$

Global variants of the accompanying Gaussian experiments are also given by (4.5) and (4.6), with extended parameter space.

4.3 Poisson observations

Let $X = \{0, 1, ...\}$ and $\mu(dx)$ be the σ-finite measure on X assigning $1/x!$ to each point $x \in X$. Let us consider the case when $\mathcal{P}$ is the family of Poisson distributions $P_\lambda(x)$, $x \in X$ with intensity $\lambda \in (0, \infty)$. After the reparametrization $\theta = \log \lambda$, $\theta \in \Theta = (-\infty, \infty)$, we get the canonical form

$$P_\theta(dx) = e^{\theta x - V(\theta)} \mu(dx),$$

with $V(\theta) = e^\theta$. The corresponding Fisher information is $I(\theta) = e^\theta$. It is clear that condition (2.7) holds true with $\Theta_0 = [\theta_{\min}, \theta_{\max}]$, where $-\infty < \theta_{\min} < \theta_{\max} < \infty$. Then the parameter set Σ contains all the functions f from the Hölder ball $\mathcal{H}(\beta, L)$, which satisfy $\theta_{\min} \le f(t) \le \theta_{\max}$.

Consider a sequence of Poisson observations

$$X_i = X(\theta_i), \quad \theta_i = f(t_i), \quad i = 1, ..., n, \tag{4.7}$$

with $t_i = i/n$, where the unknown function f is assumed to be in the set Σ. Let $f_0(t) \in \Sigma$ and $\mathcal{E}^n_{f_0}$ be the local experiment generated by the observations (4.7) with $f \in \Sigma_{f_0}(\gamma_n)$. Then, according to theorem 3.1, the experiment $\mathcal{E}^n_{f_0}$ is asymptotically equivalent to the local Gaussian experiment $\mathcal{G}^n_{f_0}$ generated by the observations

$$dY^n_t = f(t)dt + \frac{1}{\sqrt{n}} e^{-f_0(t)/2} dW_t, \quad t \in [0, 1], \tag{4.8}$$

where $f \in \Sigma_{f_0}(\gamma_n)$.

A variance-stable form of the observations (4.8) can be obtained if we note that $b(\theta) = e^\theta$ and $F(\lambda) = 2\sqrt{\lambda}$. In view of theorem 3.3 the experiment $\mathcal{G}^n_{f_0}$ is asymptotically equivalent to the experiment $\widehat{\mathcal{G}}^n_{f_0}$ given by observations

$$d\widehat{Y}^n_t = 2\sqrt{\log f(t)} dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0, 1], \tag{4.9}$$

where $f \in \Sigma_{f_0}(\gamma_n)$.

In terms of the original parameter $\lambda = e^\theta$ these results can be formulated as follows. Let $\overline{\Sigma} = \overline{\Sigma}(\beta, L)$ be the set of all functions g from the Hölder ball $\mathcal{H}(\beta, L)$, which satisfy $\lambda_{\min} \le g(t) \le \lambda_{\max}$, where $0 < \lambda_{\min} < \lambda_{\max} < \infty$. For any function $g_0 \in \overline{\Sigma}$ let $\overline{\Sigma}_{g_0}(\gamma_n)$ be its neighborhood of radius γ_n (see (3.4)) in $\overline{\Sigma}$. Denote by $\overline{\mathcal{E}}^n_{g_0}$ the local experiment generated by Poisson observations

$$X_i = X(\lambda_i), \quad i = 1, ..., n,$$

with unknown intensities $\lambda_i = g(t_i)$, where $g \in \overline{\Sigma}^n_{g_0}(\gamma_n)$. Then for any $\beta > \frac{1}{2}$ the experiment $\overline{\mathcal{E}}^n_{g_0}$ is asymptotically equivalent to the local Gaussian experiment generated by the observations

$$dY^n_t = \log g(t) dt + \frac{1}{\sqrt{n g_0(t)}} dW_t, \quad t \in [0, 1],$$

where $g \in \overline{\Sigma}_{g_0}(\gamma_n)$, W being the standard Wiener process on $(C_{[0,1]}, \mathcal{B}(C_{[0,1]}), Q_W)$.

The variance-stable result is furnished by theorem 3.6: an accompanying Gaussian experiment (local or global) is also given by the equation

$$dY^n_t = 2\sqrt{g(t)} dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0, 1].$$

4.4 Bernoulli observations

Let $\mathcal{P}$ be the family of Bernoulli distributions $P_\lambda(x)$, $x \in X = \{0,1\}$ with $P_\lambda(1) = \lambda$, $P_\lambda(0) = 1 - \lambda$, $\lambda \in (0,1)$. After the reparametrization $\theta = \log \frac{\lambda}{1-\lambda}$, $\theta \in \Theta = (-\infty, \infty)$ we arrive at the following canonical form

$$P_\theta(x) = e^{\theta x - V(\theta)}, \quad x \in X,$$

where $V(\theta) = \log(1 + e^\theta)$. The corresponding Fisher information is $I(\theta) = e^\theta/(1+e^\theta)^2$. One can easily check that the condition (2.7) holds true with $\Theta_0 = [\theta_{\min}, \theta_{\max}]$, where $-\infty < \theta_{\min} < \theta_{\max} < \infty$. Then the parameter set Σ contains all the functions $f(t)$ from the Hölder ball $\mathcal{H}(\beta, L)$, which satisfy $\theta_{\min} \leq f(t) \leq \theta_{\max}$.

Consider a sequence of Bernoulli observations

$$X_i = X(\theta_i), \quad \theta_i = f(t_i), \quad i = 1, ..., n \tag{4.10}$$

with $t_i = i/n$, where the unknown function f is in the parameter set Σ. Let $f_0(t) \in \Sigma$ and $\mathcal{E}^n_{f_0}$ be the local experiment generated by the observations (4.10), with $f \in \Sigma_{f_0}(\gamma_n)$. Then, according to Theorem 3.1, the experiment $\mathcal{E}^n_{f_0}$ is asymptotically equivalent to the local Gaussian experiment $\mathcal{G}^n_{f_0}$ generated by the observations

$$dY^n_t = f(t)dt + \frac{1}{\sqrt{n}} \frac{1 + e^{f_0(t)}}{e^{f_0(t)/2}} dW_t, \quad t \in [0,1],$$

with $f \in \Sigma_{f_0}(\gamma_n)$.

For the variance-stable form we compute $b(\theta) = e^\theta/(1+e^\theta)$ and $F(\lambda) = 2 \arcsin \sqrt{\lambda}$. Then by Theorem 3.3 the variance-stable accompanying Gaussian experiment is associated with the equation

$$dY^n_t = 2 \arcsin \sqrt{\frac{e^{f(t)}}{1 + e^{f(t)}}}\, dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0,1].$$

In term of the original parameter $\lambda = e^\theta/(1+e^\theta)$ this result can be formulated in the following way. Let $\overline{\Sigma} = \overline{\Sigma}(\beta, L)$ be the set of all functions g from the Hölder ball $\mathcal{H}(\beta, L)$, which satisfy $\lambda_{\min} \leq g(t) \leq \lambda_{\max}$, where $0 < \lambda_{\min} < \lambda_{\max} < 1$. For any function $g_0 \in \overline{\Sigma}$ let $\overline{\Sigma}_{g_0}(\gamma_n)$ be its neighborhood of radius γ_n (see (3.4)) in $\overline{\Sigma}$. Denote by $\overline{\mathcal{E}}^n_{g_0}$ the local experiment generated by the observations (4.10) with $f(t) = \log \frac{g(t)}{1-g(t)}$, $g \in \overline{\Sigma}_{g_0}(\gamma_n)$. Then for any $\beta > \frac{1}{2}$ the experiment $\overline{\mathcal{E}}^n_{g_0}$ is asymptotically equivalent to the local Gaussian experiment generated by the observations

$$dY^n_t = \log \frac{g(t)}{1 - g(t)} dt + \frac{1}{\sqrt{n g_0(t)(1 - g_0(t))}} dW_t, \quad t \in [0,1],$$

where $g \in \overline{\Sigma}_{g_0}(\gamma_n)$, W being the standard Wiener process on $(C_{[0,1]}, \mathcal{B}(C_{[0,1]}), Q_W)$.

The variance-stable form of the Gaussian accompanying experiment parametrized by g is (according to Theorem 3.6)

$$dY^n_t = 2 \arcsin \sqrt{g(t)}\, dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0,1].$$

The global variants are straightforward.

4.5 Exponential observations

Let $X = (0, \infty)$ and $\mu(dx)$ be the Lebesgue measure on X. Let $\mathcal{P}$ be the family of exponential distributions $P_\theta(dx)$, on X with parameter $\theta \in \Theta = (-\infty, 0)$

$$P_\theta(x) = e^{\theta x - V(\theta)} \mu(dx),$$

where $V(\theta) = \log \theta$. The corresponding Fisher information is $I(\theta) = \theta^{-2}$, so condition (2.7) holds true with $\Theta_0 = [\theta_{\min}, \theta_{\max}]$ for some constants $-\infty < \theta_{\min} < \theta_{\max} < 0$ and ε_0 small enough. Then the parameter set Σ contains all the functions $f(t)$ from the Hölder ball $\mathcal{H}(\beta, L)$, which satisfy $\theta_{\min} \le f(t) \le \theta_{\max}$.

Consider a sequence of exponential observations

$$X_i = X(\theta_i), \quad \theta_i = f(t_i), \quad i = 1, ..., n \tag{4.11}$$

with $t_i = i/n$, where unknown function f is in the set Σ. Let $f_0 \in \Sigma$ and $\mathcal{E}^n_{f_0}$ be the local experiment generated by the observations (4.11). Then, by Theorem 3.1, the experiment $\mathcal{E}^n_{f_0}$ is asymptotically equivalent to the local Gaussian experiment $\mathcal{G}^n_{f_0}$ generated by observations

$$dY^n_t = f(t)dt + \frac{1}{\sqrt{n}} f_0(t) dW_t, \quad t \in [0, 1], \quad f \in \Sigma_{f_0}(\gamma_n).$$

A variance-stable form for the global experiment can be also obtained. For this we remark that $b(\theta) = -\theta$ and $F(\lambda) = \log \lambda$. Thus by Theorem 3.3 a variance-stable form (local or global) is given by the equation

$$dY^n_t = \log(-f(t))\, dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0, 1].$$

4.6 Application to the density model

Assume that we observe a sequence of i. i. d. r. v.'s X_i, $i = 1, ..., n$, each with density $f \in \Sigma$, where the set Σ is as in the previous section. By a poissonization technique one can show that this experiment is asymptotically equivalent to observing a sequence of Poisson r. v.'s X_i, $i = 1, ..., n$ with intensities $f(t_i)$, $i = 1, ..., n$ respectively, where $t_i = \frac{i}{n}$, $i = 1, ..., n$ is the uniform grid on the unit interval $[0, 1]$. We skip this technical step, since its proof is similar to that given in Nussbaum [31], Section 4.

The conclusion we draw from this fact and from the example in Section 4.3 is that estimating a density $f(t)$ from i. i. d. data is asymptotically equivalent to estimating 2 times the square root of $f(t)$ in white noise. Thus we recover the main result of [31].

5 Local approximation

5.1 Bounds for the Hellinger distance

Let $f_0 \in \Sigma$ and let γ_n be the nonparametric shrinking rate defined by (3.4). Recall briefly the setting from Section 3.1. Put $t_i = \frac{i}{n}$, $i = 1, ..., n$. Consider the local experiment

$$\mathcal{E}^n_{f_0} = \left(X^n, \mathcal{B}(X^n), \{ P^n_f : f \in \Sigma_{f_0}(\gamma_n) \} \right), \tag{5.1}$$

generated by the discrete observations

$$X_i = X(f(t_i)), \quad i = 1, ..., n, \quad f \in \Sigma_{f_0}(\gamma_n), \tag{5.2}$$

with distributions in the exponential family $\mathcal{P}$. Its accompanying local Gaussian experiment

$$\mathcal{G}_{f_0}^{\prime,n} = \left(R^n, \mathcal{B}(R^n), \left\{Q_{f_0,f}^{\prime,n} : f \in \Sigma_{f_0}(\gamma_n)\right\}\right), \tag{5.3}$$

is generated by the observations in discrete time

$$Y_i = f(t_i) + I\left(f_0(t_i)\right)^{-\frac{1}{2}} \varepsilon_i, \quad i = 1, ..., n, \quad f \in \Sigma_{f_0}(\gamma_n), \tag{5.4}$$

with standard normal r. v.'s ε_i.

Theorem 5.1 *Assume that $\beta > \frac{1}{2}$. For any $f_0 \in \Sigma$ and any $n = 1, 2, ...$ experiments $\mathcal{E}_{f_0}^n$ and $\mathcal{G}_{f_0}^{\prime,n}$ can be constructed on the measurable space $(R^n, \mathcal{B}(R^n))$ such that*

$$\sup_{f \in \Sigma_{f_0}(\gamma_n)} H^2(P_f^n, Q_{f_0,f}^{\prime,n}) \leq c n^{-\frac{2\beta-1}{2\beta+1}} (\log n)^{\frac{14\beta+5}{2\beta+1}},$$

where c is a constant depending only on $I_{\max}$, $I_{\min}$, ε_0, L, β.

For the proof we make use of the following assertion, which is also of independent interest. This theorem corresponds to the local experiments obtained by looking only at observations from a shrinking time interval, which after a rescaling leads to neighborhoods of the "almost $n^{-1/2}$" size. Let γ_n^* be the shrinking rate defined by (3.17), with constant κ_0^* being arbitrary positive. Consider the local experiment $\mathcal{E}_{f_0}^{*,n}$ defined in Section 3.4 obtained by localizing: $f = f_0 + \gamma_n^* g$ and indexing by $g \in \Sigma$ (cf. (3.18)). Consider also the discrete Gaussian accompanying experiment $\mathcal{G}_{f_0}^{\prime,*,n}$ from Section 3.4, relation (3.19).

Theorem 5.2 *Let $\beta > \frac{1}{2}$. For any $f_0 \in \Sigma$ local experiments $\mathcal{E}_{f_0}^{*,n}$ and $\mathcal{G}_{f_0}^{\prime,*,n}$ can be constructed on the same measurable space $(R^n, \mathcal{B}(R^n))$ such that*

$$\sup_{g \in \Sigma} H^2(P_{f_0,g}^{*,n}, Q_{f_0,g}^{\prime,*,n}) \leq c n^{-1} (\log n)^7,$$

where c is a constant depending only on $I_{\max}$, $I_{\min}$, ε_0, κ_0^*, L.

Remark 5.1 *Theorems* 3.2 *and* 3.8 *follow immediately from Theorems* 5.1 *and* 5.2 *respectively, by (*2.14*).*

The construction for Theorem 5.2 heavily relies upon the results on strong approximation in Section 2.4 and is given in the next Section 5.2. The proof of Theorem 5.2 is presented in Section 5.3.

Theorem 5.1 is a consequence of Theorem 5.2. Its proof is presented in Section 5.4.

5.2 Construction of a local experiment

The construction on the same probability space of a Gaussian experiment and of an exponential one is particularly simple in view of the convenient form of the likelihood ratio of the latter (cf. (3.7)). This form allows to employ the strong approximation result of Section 2.4. We proceed to describe formally this construction.

Let f_0 be a fixed function in the parameter set Σ. Consider the local experiments $\mathcal{E}^{*,n}_{f_0}$ and $\mathcal{G}'^{,*,n}_{f_0}$. Recall that the shrinking rate γ^*_n is defined by

$$\gamma^*_n = \kappa^*_0 (n/\log n)^{-\frac{1}{2}},$$

where κ^*_0 is arbitrary positive. Since f_0 is considered fixed, for the sake of brevity we will drop the index f_0 from the notations for measures $P^n_{f_0,g}$ and $Q'^{,*,n}_{f_0,g}$, so we will write $\tilde{P}^n_g = P^n_{f_0,g}$ and $\tilde{Q}^n_g = Q'^{,*,n}_{f_0,g}$. The corresponding expectations are denoted by $\mathbf{E}_{\tilde{P}^n_g}$ and $\mathbf{E}_{\tilde{Q}^n_g}$ respectively. Given a function $f = f_0 + \gamma^*_n g$, $g \in \Sigma$, we consider the likelihood ratio of the local experiment $\mathcal{E}^{*,n}_{f_0}$, which according to (3.7) is

$$\frac{d\tilde{P}^n_g}{d\tilde{P}^n_0} = \exp\left\{ \gamma^*_n \sum_{i=1}^n g(t_i) U(X_i) - \sum_{i=1}^n \left(V(f(t_i)) - V(f_0(t_i)) \right) \right\}.$$

Denote by $\overline{U}(X_i)$ the r. v. $U(X_i)$ centered under the measure $\tilde{P}^n_0$, i. e.

$$\overline{U}(X_i) = U(X_i) - V'(f_0),$$

so that by (2.2) and (2.3) we have

$$\mathbf{E}_{\tilde{P}^n_0} \overline{U}(X_i) = 0, \quad \mathbf{E}_{\tilde{P}^n_0} \overline{U}(X_i)^2 = I(f_0(t_i)) = V''(f_0(t_i)).$$

The corresponding local Gaussian experiment $\mathcal{G}'^{,*,n}_{f_0}$ has likelihood ratio

$$\frac{d\tilde{Q}^n_g}{d\tilde{Q}^n_0} = \exp\left\{ \gamma^*_n \sum_{i=1}^n g(t_i) I(f(t_i))^{1/2} \varepsilon_i - \frac{1}{2} (\gamma^*_n)^2 \sum_{i=1}^n g(t_i)^2 I(f_0(t_i)) \right\}, \tag{5.5}$$

with $f = f_0 + \gamma^*_n g$, $g \in \Sigma$. Put $\mathbf{P} = \tilde{Q}^n_0$ and consider the probability space $(R^n, \mathcal{B}(R^n), \mathbf{P})$ on which ε_i, $i = 1, ..., n$ is a sequence of i. i. d. standard normal r. v.'s. Put $N_i = I(f(t_i))^{1/2} \varepsilon_i$. Thus we are given a sequence of independent normal r. v.'s N_i, $i = 1, ..., n$ with zero means and variances $\mathbf{E} N_i^2 = I(f_0(t_i))$, for $i = 1, ..., n$. Because of condition (2.7) and Proposition 2.1 we can apply Theorem 2.1, according to which on this probability space there is a sequence of independent r. v.'s $\widetilde{U}_i$ $i = 1, ..., n$ such that $\widetilde{U} \overset{d}{=} \overline{U}(X_i)$ for any $i = 1, ..., n$ and for any function $g \in \mathcal{H}(\frac{1}{2}, L)$

$$\mathbf{P}\left(|S_n(g)| \geq x (\log n)^2 \right) \leq c_1 \exp\{-c_2 x\}, \quad x \geq 0, \tag{5.6}$$

where

$$S_n(g) = \sum_{i=1}^n g(t_i) \left(\widetilde{U}_i - N_i \right),$$

and c_1, c_2 are constants depending only on $I_{\min}$, $I_{\max}$, ε_0, L.

Now we proceed to construct a version of the likelihood process for $\mathcal{E}_{f_0}^{*,n}$ on the probability space $(R^n, \mathcal{B}(R^n), \mathbf{P})$. For this define the experiment $\mathcal{F}_{f_0}^{*,n}$ as follows:

$$\mathcal{F}_{f_0}^{*,n} = \left(R^n, \mathcal{B}(R^n), \{F_g^n : \ g \in \Sigma\}\right),$$

where F_g^n is the probability measure on the measurable space $(R^n, \mathcal{B}(R^n))$ defined for any $g \in \Sigma$ by the equality $f = f_0 + \gamma_n^* g$

$$\frac{dF_g^n}{d\mathbf{P}} = \exp\left\{\gamma_n^* \sum_{i=1}^n g(t_i)\widetilde{U}_i - \sum_{i=1}^n \left(V(f(t_i)) - V(f_0(t_i)) - \gamma_n^* g(t_i) V'(f_0(t_i))\right)\right\} \tag{5.7}$$

where $f = f_0 + \gamma_n^* g$; then $F_0^n = \mathbf{P}$.

Let us remark that since the sequences $\widetilde{U}_i$, $i = 1, ..., n$ and $\overline{U}(X_i)$, $i = 1, ..., n$ have the same joint distributions and the sufficient statistic in the experiment $\mathcal{E}_{f_0}^{*,n}$ is

$$\sum_{i=1}^n g(t_i) U(X_i),$$

the experiments $\mathcal{E}_{f_0}^{*,n}$ and $\mathcal{F}_{f_0}^{*,n}$ are equivalent. Therefore we will assume in the sequel that $\widetilde{U}_i = \overline{U}(X_i)$, for $i = 1, ..., n$ and $\mathcal{E}_{f_0}^{*,n} = \mathcal{F}_{f_0}^{*,n}$. In particular $\tilde{P}_g^n = F_g^n$ for any $g \in \Sigma$ and $\tilde{P}_0^n = \mathbf{P} = \tilde{Q}_0^n$.

5.3 Proof of local equivalence: almost $n^{-1/2}$- neighborhoods

In this section we present a proof of Theorem 5.2. First recall that according to the last remark in the previous section we consider the experiment $\mathcal{E}_{f_0}^{*,n}$ to be a version of the original one constructed such that its likelihood process is on a common probability space $(R^n, \mathcal{B}(R^n), \mathbf{P})$ with the Gaussian likelihood process for $\mathcal{G}_{f_0}^{\prime,*,n}$. Recall also that the "central" measures (i. e. those with $g = 0$) in the local experiments $\mathcal{E}_{f_0}^{*,n}$ and $\mathcal{G}_{f_0}^{\prime,*,n}$ coincide with the measure $\mathbf{P}$.

Let $f_0 \in \Sigma$ and $f = f_0 + \gamma_n^* g$, $g \in \Sigma$. Note that the function g belongs also to the Hölder ball $\mathcal{H}(\frac{1}{2}, L)$ since $\beta \geq \frac{1}{2}$. Then by taking $x = \frac{c_3}{c_2} \log n$ in (5.6) we arrive at

$$\mathbf{P}(|S_n(g)| \geq \frac{c_3}{c_2} \gamma_n (\log n)^3) \leq c_1 \exp\{-c_3 \log n\}, \tag{5.8}$$

where c_1, c_2 are the same as in (5.6) and c_3 is a "free" constant whose value will be chosen later. Recall that according to our agreement $\widetilde{U}_i = \overline{U}(X_i)$, $i = 1, ..., n$ and therefore

$$S_n(g) = \sum_{i=1}^n g(t_i)(\overline{U}(X_i) - N_i).$$

What we have to prove is that the Hellinger distance between the measures $\tilde{P}_g^n$ and $\tilde{Q}_g^n$ satisfies

$$H^2\left(\tilde{P}_g^n, \tilde{Q}_g^n\right) \leq c_4 n^{-1} (\log n)^7, \tag{5.9}$$

for some constant c_4 depending only on $I_{\max}$, $I_{\min}$, ε_0, κ_0^*, L.

Well known properties of the Hellinger distance (see (2.12) and (2.15)) imply

$$H^2\left(\tilde{P}_g^n, \tilde{Q}_g^n\right) = \frac{1}{2} \mathbf{E_P} \left(\Lambda^1(g)^{\frac{1}{2}} - \Lambda^2(g)^{\frac{1}{2}}\right)^2,$$

where we denote for brevity

$$\Lambda^1(g) = \frac{d\tilde{P}_g^n}{d\mathbf{P}}, \quad \Lambda^2(g) = \frac{d\tilde{Q}_g^n}{d\mathbf{P}}.$$

Put $u_n = \frac{c_3}{c_2}\gamma_n(\log n)^3$. Then obviously

$$H^2\left(\tilde{P}_g^n, \tilde{Q}_g^n\right) \leq J_1 + J_2,$$

where ($\mathbf{1}(A)$ being the indicator for event A)

$$\begin{aligned} J_1 &= \frac{1}{2}\mathbf{E}_{\mathbf{P}}\mathbf{1}\left(|S_n(g)| < u_n\right)\left(\Lambda^1(g)^{\frac{1}{2}} - \Lambda^2(g)^{\frac{1}{2}}\right)^2, \\ J_2 &= \frac{1}{2}\mathbf{E}_{\mathbf{P}}\mathbf{1}\left(|S_n(g)| \geq u_n\right)\left(\Lambda^1(g)^{\frac{1}{2}} - \Lambda^2(g)^{\frac{1}{2}}\right)^2. \end{aligned}$$

First we give an estimate for J_1. Changing the probability measure we obtain

$$J_1 = \frac{1}{2}\mathbf{E}_{\tilde{Q}_g^n}\mathbf{1}\left(|S_n(g)| \leq u_n\right)\left(\Lambda^2(g)^{\frac{1}{2}}\Lambda^1(g)^{-\frac{1}{2}} - 1\right)^2.$$

According to (5.7) and (5.5)

$$\Lambda^1(g) = \exp\left\{\gamma_n^*\sum_{i=1}^n g(t_i)U(X_i) - \sum_{i=1}^n \left(V(f(t_i)) - V(f_0(t_i))\right)\right\} \tag{5.10}$$

and

$$\Lambda^2(g) = \exp\left\{\gamma_n^*\sum_{i=1}^n g(t_i)N_i - \frac{1}{2}(\gamma_n^*)^2\sum_{i=1}^n g(t_i)^2 I(f_0(t_i))\right\}.$$

This gives

$$\frac{\Lambda^2(g)}{\Lambda^1(g)} = \exp\left\{-S_n(f) + R(f_0, f)\right\},$$

where

$$\begin{aligned} R(f_0, f) &= \sum_{i=1}^n \{V(f(t_i)) - V(f_0(t_i)) - \gamma_n^* g(t_i)V'(f_0(t_i)) \\ &\qquad - \frac{1}{2}(\gamma_n^*)^2 g(t_i)^2 V''(f_0(t_i))\}. \end{aligned}$$

A three term Taylor expansion yields

$$|R(f_0, f)| \leq \frac{1}{6}(\gamma_n^*)^3\sum_{i=1}^n |g(t_i)|^3\left|V'''(\widetilde{f}_i)\right|,$$

with $\widetilde{f}_i = f_0(t_i) + \nu_i\gamma_n^* g(t_i)$, $0 \leq \nu_i \leq 1$. Since $||g||_\infty \leq L$ and $\left|V'''(\widetilde{f}_i)\right| \leq c_5$, with c_5 depending only on $I_{\max}$, ε_0 (see Proposition 2.2), it is clear that

$$|R(f_0, f)| \leq \frac{1}{6}L^3 c_5(\gamma_n^*)^3 n \leq c_6 n^{-\frac{1}{2}}(\log n)^{\frac{3}{2}},$$

with c_6 depending on c_5, L, κ_0^* and for n large enough. This yields the following estimate:

$$\begin{aligned} J_1 &= \frac{1}{2}\mathbf{E}_{Q_f^n}\mathbf{1}(|S_n(f)|\le u_n)\left(\exp\left\{-\frac{1}{2}S_n(f)+\frac{1}{2}R(f_0,f)\right\}-1\right)^2 \\ &\le c_7\left(|S_n(f)|+|R(f_0,f)|\right)^2 \\ &\le c_7\left(\frac{c_2}{c_1}n^{-\frac{1}{2}}(\log n)^{\frac{7}{2}}+c_6 n^{-\frac{1}{2}}(\log n)^{\frac{3}{2}}\right)^2, \end{aligned}$$

with c_7 depending on c_2, c_3, c_6. Hence with some c_8 depending on c_2, c_3, c_6, c_7

$$J_1 \le c_7 n^{-1}(\log n)^7. \tag{5.11}$$

Now we proceed to estimate J_2. The Hölder inequality implies

$$J_2 \le \frac{1}{2}J_3^{\frac{1}{2}}J_4^{\frac{1}{2}}, \tag{5.12}$$

where by (5.8)

$$J_3 = \mathbf{P}(|S_n(f)|\ge u_n)\le c_1\exp\{-c_3\log n\}, \tag{5.13}$$

for n large enough, and

$$J_4 = \mathbf{E_P}\left(\Lambda^1(g)^{\frac{1}{2}}+\Lambda^2(g)^{\frac{1}{2}}\right)^4. \tag{5.14}$$

Note that the constant c_3 in (5.13) is "free". We will show that J_4 is bounded from above by $32\exp\{c_9\log n\}$, with some constant c_9 which we calculate below. Indeed from (5.14) we get

$$J_4 \le 16\left(\mathbf{E_P}(\Lambda^1(g))^2+\mathbf{E_P}(\Lambda^2(g))^2\right). \tag{5.15}$$

First we give a bound for $\mathbf{E_P}(\Lambda^1(g))^2$. It follows from (5.10) that

$$\mathbf{E_P}(\Lambda^1(g))^2 = \mathbf{E}\exp\{2\gamma_n^*\sum_{i=1}^n g(t_i)\overline{U}(X_i)-2R_0(f_0,f)\},$$

where

$$R_0(f_0,f)=\sum_{i=1}^n\left\{V(f(t_i))-V(f_0(t_i))-\gamma_n^* g(t_i)V'(f_0(t_i))\right\}.$$

The estimate for the remainder $R_0(f_0,f)$ is straightforward by Taylor's formula:

$$|R_0(f_0,f)|\le\frac{1}{2}(\gamma_n^*)^2\sum_{i=1}^n|g(t_i)|^2\left|V''(\widetilde{f}_i)\right|\le\frac{1}{2}L^2 I_{\max}(\gamma_n^*)^2 n,$$

where $\widetilde{f}_i = f(t_i)+\nu_i\gamma_n^* g(t_i)$, $0\le\nu_i\le 1$. Since $\overline{U}(X_i)$, $i=1,...,n$ are independent r. v.'s, by using Proposition 2.1

$$\begin{aligned} \mathbf{E_P}(\Lambda^1(g))^2 &= \exp\{-2R_0(f_0,f)\}\prod_{i=1}^n\mathbf{E_P}\exp\left\{2\gamma_n^* g(t_i)\overline{U}(X_i)\right\} \\ &\le \exp\{2|R_0(f_0,f)|\}\prod_{i=1}^n\exp\left\{2(\gamma_n^*)^2 g(t_i)^2 I_{\max}\right\} \\ &\le \exp\left\{3L^2 I_{\max}(\gamma_n^*)^2 n\right\}\le\exp\{c_9\log n\}, \end{aligned}$$

where c_9 depends on $I_{\max}$, L, κ_0^*. The bound $\mathbf{E_P}(\Lambda^1(g))^2 \leq \exp\{c_9 \log n\}$ can be proved similarly. These bounds and (5.15) yield

$$J_4 \leq 32 \exp\{c_9 \log n\}. \tag{5.16}$$

Using the bounds for J_3 and J_4 given by (5.13) and (5.16) in (5.12) we obtain

$$J_2 \leq 2\sqrt{2c_1} \exp\left\{-\frac{1}{2}(c_3 - c_9)\log n\right\}$$

from which, taking $c_3 = c_9 + 2$, we get

$$J_2 \leq 2\sqrt{2c_1}/n \leq 2\sqrt{2c_1} n^{-1}(\log n)^7. \tag{5.17}$$

The desired inequality (5.9) follows from (5.11) and (5.17).

5.4 Proof of local equivalence: nonparametric neighborhoods

We present here a proof for Theorem 5.1. Before the rigorous argument let us briefly expound the main idea. We start by splitting the original local experiment $\mathcal{E}_{f_0}^n$ into m parts which correspond to fractions of the observations over shrinking time intervals having length of order $\delta_n = \gamma_n^{1/\beta}$, where $\gamma_n$is the shrinking rate of the neighborhood $\Sigma_{f_0}(\gamma_n)$. One may call the corresponding experiments doubly local. Denote by $n_k = O(n\delta_n)$ the number of observations (i. e. number of design points t_i) in the k-th doubly local experiment. After rescaling the latter can be viewed as an experiment on the whole interval $[0,1]$ over a shrinking neighborhood of size $O(n_k/\log n_k)^{-\frac{1}{2}}$. By Theorem 5.2 we can "approximate" this experiment by a corresponding Gaussian one with a bound for the *squared* Hellinger distance between corresponding measures of order $O(n_k^{-1}(\log n_k)^7)$. Further arguments are based on the crucial inequality (2.13) which is applied to the original (parameter-local) experiment on the unit interval $[0,1]$, construed as a product of the m local doubly local experiments on the intervals of size δ_n. Since the Gaussian experiment $\mathcal{G}_{f_0}^n$ can be decomposed similarly, we obtain a bound for the squared Hellinger distance between $\mathcal{E}_{f_0}^n$ and $\mathcal{G}_{f_0}^n$

$$O(mn_k^{-1}(\log n_k)^7) = O(n\delta_n^{-2}(\log n_k)^7) = o(1),$$

as $n \to \infty$, for $\beta > \frac{1}{2}$, which proves our theorem.

Now we turn to the argument in detail. Let $\beta > \frac{1}{2}$ and $f_0 \in \Sigma$. Assume that the shrinking rate of the neighborhood $\Sigma_{f_0}(\gamma_n)$ is given by

$$\gamma_n = \kappa_0(n/\log n)^{-\frac{\beta}{2\beta+1}}, \tag{5.18}$$

with some constant κ_0 depending on β. Put $m = [1/\delta_n]$, where

$$\delta_n = (\gamma_n)^{\frac{1}{\beta}} = \kappa_0^{1/\beta}(n/\log n)^{-\frac{1}{2\beta+1}}. \tag{5.19}$$

Consider a partition $\mathfrak{A}$ of the unit interval $[0,1]$ into intervals $A_k = (a_k, b_k]$, $k = 1, ..., m$ of length $1/m$. It is easy to see that for n large enough

$$\frac{1}{2\delta_n} \leq m \leq \frac{1}{\delta_n}. \tag{5.20}$$

Put $I_k = \{i : t_i \in A_k\}$. Denote by n_k the cardinality of I_k, i. e. the number of design points which fall into the interval A_k, $k = 1, ..., m$. It is clear that for n large enough

$$n\delta_n \leq n_k \leq 2n\delta_n. \tag{5.21}$$

We particularly point out that (5.18), (5.19), (5.20) and (5.21) imply

$$\gamma_n \leq \gamma_n^* \equiv \kappa_0^*(n_k/\log n_k)^{-1/2}, \tag{5.22}$$

for some constant κ_0^* depending on κ_0 and β. Let $a_k(t)$ be the linear function which maps the unit interval $[0,1]$ onto the interval $[a_k, b_k]$, i. e. $a_k(t) = t/m + a_k$, $t \in [0,1]$. For any $f \in \Sigma_{f_0}(\gamma_n)$ and $k = 1, ..., m$ consider the function f_k defined on the interval $[0,1]$ as follows

$$f_k(t) = \frac{(f - f_0)(a_k(t))}{\gamma_n^*}. \tag{5.23}$$

We will prove that $f_k \in \Sigma^* = \Sigma(\beta, L^*)$, with some L^* depending on L and β. Indeed, since $f \in \Sigma_{f_0}(\gamma_n)$, then $||f_k||_\infty \leq ||f - f_0||_\infty/\gamma_n^* \leq \gamma_n/\gamma_n^* \leq 1$. On the other hand since f, f_0 are in the Hölder ball $\mathcal{H}(\beta, L)$, the function $\psi = f - f_0$ is also in the Hölder ball. Taking into account (5.19), (5.20) and (5.22) we obtain for any $x, y \in [0,1]$

$$\begin{aligned}
\left| f_k^{([\beta])}(x) - f_k{}^{([\beta])}(y) \right| &= m^{-[\beta]} \left| \psi^{([\beta])}(a_k(x)) - \psi^{([\beta])}(a_k(y)) \right| / \gamma_n^* \\
&\leq m^{-[\beta]} L \left| a_k(x) - a_k(y) \right|^{\beta - [\beta]} / \gamma_n^* \\
&\leq 2^\beta \delta_n^\beta L \left| x - y \right|^{\beta - [\beta]} / \gamma_n^* \\
&= 2^\beta \gamma_n L \left| x - y \right|^{\beta - [\beta]} / \gamma_n^* \\
&\leq 2^\beta L \left| x - y \right|^{\beta - [\beta]},
\end{aligned}$$

proving that f_k is in the Hölder ball $\Sigma(\beta, L^*)$ with $L^* = 2^\beta L$.

Let $X^{k,n} = \{X_i,\ i \in I_k\}$ be the fragment of observations $\{X_i,\ i = 1, ..., n\}$ (defined by (5.2)) associated to the time interval A_k for some $k \in \{1, ..., m\}$. After a rescaling with the linear function $a_k(t)$ these observations can be associated to design points $t_i^k = \frac{i}{n_k}$, $i = 1, ..., n_k$ on the unit interval $[0,1]$. Let $P_f^{n,k}$ be the measure on $(R^{n_k}, \mathcal{B}(R^{n_k}))$ induced by the set of r. v.'s $X^{n,k}$ and set $P_{f_0,g}^{k,n} = P_f^{k,n}$ for $f = f_0 + \gamma_n^* g$. For each $k \in \{1, ..., m\}$ consider the local experiment

$$\mathcal{E}_{f_0}^{k,n} = \left(R^{n_k}, \mathcal{B}(R^{n_k}), \left\{ P_{f_0,g}^{k,n} : g \in \Sigma^* \right\} \right), \tag{5.24}$$

In the same way introduce the local Gaussian experiment

$$\mathcal{G}_{f_0}^{\prime,k,n} = \left(R^{n_k}, \mathcal{B}(R^{n_k}), \left\{ Q_{f_0,g}^{\prime,k,n} : g \in \Sigma^* \right\} \right), \tag{5.25}$$

generated by the observation fragment $Y^{k,n} = \{Y_i,\ i \in I_k\}$, with Y_i defined by (3.10). Here $Q_{f_0,g}^{\prime,k,n}$ is the Gaussian shift measure on $(R^{n_k}, \mathcal{B}(R^{n_k}))$ induced by the observations $Y^{n,k}$ under $f = f_0 + \gamma_n^* g$.

According to Theorem 5.2 the experiments $\mathcal{E}_{f_0}^{k,n}$ and $\mathcal{G}_{f_0}^{\prime,k,n}$ can be constructed on the measurable space $(R^{n_k}, \mathcal{B}(R^{n_k}))$ such that

$$\sup_{g \in \Sigma^*} H^2 \left(P_{f_0,g}^{k,n}, Q_{f_0,g}^{\prime,k,n} \right) \leq c_1 n_k^{-1} (\log n_k)^7, \tag{5.26}$$

with a constant c_1 depending on $I_{\max}$, $I_{\min}$, ε_0, κ_0^*, L. Now consider the subexperiments of $\mathcal{E}_{f_0}^{k,n}$ and $\mathcal{G}_{f_0}^{\prime,k,n}$ obtained by setting $g = f_k$, $f \in \Sigma_{f_0}(\gamma_n)$ in (5.24) and (5.25), where f_k is defined by (5.23). Reindex those subexperiments by $f \in \Sigma_{f_0}(\gamma_n)$ and call them $\widetilde{\mathcal{E}}_{f_0}^{k,n}$, $\widetilde{\mathcal{G}}_{f_0}^{k,n}$ respectively. For the respective reindexed measures $\tilde{P}_f^{k,n} = P_{f_0,f_k}^{k,n}$ and $\tilde{Q}_{f_0,f}^{\prime,k,n} = Q_{f_0,f_k}^{\prime,k,n}$ we have as a consequence of (5.26)

$$\sup_{f\in\Sigma_{f_0}(\gamma_n)} H^2\left(\tilde{P}_f^{k,n}, \tilde{Q}_{f_0,f}^{\prime,k,n}\right) \le c_1 n_k^{-1}(\log n_k)^7, \tag{5.27}$$

since $f_k \in \Sigma^*$ for any $f \in \Sigma_{f_0}(\gamma_n)$.

Define the experiment $\mathbb{E}_{f_0}^n$ as

$$\mathbb{E}_{f_0}^n = \widetilde{\mathcal{E}}_{f_0}^{1,n} \otimes \ldots \otimes \widetilde{\mathcal{E}}_{f_0}^{m,n},$$

which obviously is (exactly) equivalent to $\mathcal{E}_{f_0}^n$ defined by (5.1), (5.2). The corresponding local Gaussian experiment $\mathcal{G}_{f_0}^n$ defined by (5.3), (5.4) is (exactly) equivalent to the experiment

$$\mathbb{G}_{f_0}^n = \widetilde{\mathcal{G}}_{f_0}^{1,n} \otimes \ldots \otimes \widetilde{\mathcal{G}}_{f_0}^{m,n}.$$

It remains only to note that $P_f^n = \tilde{P}_f^{1,n} \otimes \ldots \otimes \tilde{P}_f^{m,n}$ and $Q_{f_0,f}^{\prime,n} = \tilde{Q}_{f_0,f}^{\prime,1,n} \otimes \ldots \otimes \tilde{Q}_{f_0,f}^{\prime,m,n}$ and thus, according to (2.13),

$$\begin{aligned} H^2\left(P_f^n, Q_{f_0,f}^{\prime,n}\right) &\le \sum_{i=1}^m H^2\left(\tilde{P}_f^{k,n}, \tilde{Q}_{f_0,f}^{\prime,k,n}\right) \\ &\le m n_k^{-1}(\log n_k)^7 \\ &\le n^{-1}\delta_n^{-2}(\log n)^7 \\ &\le c_2 n^{-\frac{2\beta-1}{2\beta+1}}(\log n)^{\frac{14\beta+5}{2\beta+1}}, \end{aligned}$$

where c_2 is a constant depending on $I_{\max}$, $I_{\min}$, ε_0, β, L. This completes the proof of Theorem 5.1.

5.5 Proof of variance-stable form

In this section we present a proof of Theorem 3.3.

As in Section 2.2, set $b(t) = V'(\theta)$, $\theta \in \Theta_0$. Consider the following experiments: $\mathcal{F}_{f_0}^{1,n}$ generated by observations

$$dY_t^{1,n} = (f(t) - f_0(t))\,dt + \frac{1}{\sqrt{n}} I(f_0(t))^{-1/2} dW_t, \quad t \in [0,1], \tag{5.28}$$

where $f \in \Sigma_{f_0}(\gamma_n)$, and $\mathcal{F}_{f_0}^{2,n}$ generated by the observations

$$dY_t^{2,n} = (b(f(t)) - b(f_0(t)))\, I(f_0(t))^{-1} dt + \frac{1}{\sqrt{n}} I(f_0(t))^{-1/2} dW_t, \quad t \in [0,1], \tag{5.29}$$

where $f \in \Sigma_{f_0}(\gamma_n)$.

Proposition 5.1 *The experiments $\mathcal{F}_{f_0}^{1,n}$ and $\mathcal{F}_{f_0}^{2,n}$ are asymptotically equivalent. Moreover*

$$\Delta^2\left(\mathcal{F}_{f_0}^{1,n},\mathcal{F}_{f_0}^{2,n}\right) \le c_1 n^{-\frac{2\beta-1}{2\beta+1}}(\log n)^{\frac{4\beta}{2\beta+1}},$$

where c_1 is a constant depending only on $I_{\max}$, $I_{\min}$ and ε_0.

Proof. By Taylor expansion we have for any $\theta \in B(\theta_0, \varepsilon_0/2)$ and $\theta \in \Theta_0$

$$b(\theta) - b(\theta_0) = (\theta-\theta_0)b'(\theta_0) + \frac{1}{2}(\theta-\theta_0)^2 b''(\widetilde{\theta}_0),$$

where $\widetilde{\theta}_0 = \theta_0 + \nu(\theta-\theta_0)$, $0 \le \nu \le 1$. According to (2.5) and Proposition 2.2

$$b'(\theta_0) = V''(\theta_0) = I(\theta_0), \qquad \left|b''(\widetilde{\theta}_0)\right| = \left|V'''(\widetilde{\theta}_0)\right| \le c_2, \tag{5.30}$$

with c_2 depending only on $I_{\max}$ and ε_0. Hence, for any $f \in \Sigma_{f_0}(\gamma_n)$ we obtain with some $|\nu| \le 1$

$$b(f(t)) - b(f_0(t)) = (f(t) - f_0(t))\, b'(f_0(t)) + \nu \frac{c_2}{2}\gamma_n^2.$$

Put for brevity

$$\begin{aligned} m_1(t) &= f(t) - f_0(t), \\ m_2(t) &= (b(f(t)) - b(f_0(t)))\, I(f_0(t))^{-1}. \end{aligned}$$

Since $I(\theta) \ge I_{\min}$, we have

$$|m_1(t) - m_2(t)| \le \gamma_n^2 \frac{c_2}{2} I_{\min}^{-1}.$$

Let $Q_{f_0,f}^{1,n}$ and $Q_{f_0,f}^{2,n}$ be the measures induced by observations (5.28) and (5.29) respectively. Then by formula (2.16) we get for any $f \in \Sigma_{f_0}(\gamma_n)$

$$\begin{aligned} H^2\left(Q_{f_0,f}^{1,n}, Q_{f_0,f}^{2,n}\right) &\le \frac{n}{8}\int_0^1 (m_1(t) - m_2(t))^2\, dt \\ &\le \frac{n}{32}\gamma_n^4 c_2^2 I_{\min}^{-2} = c_1 n^{-\frac{2\beta-1}{2\beta+1}}(\log n)^{\frac{4\beta}{2\beta+1}}. \end{aligned}$$

To complete the proof we refer to inequality (2.14). ■

Introduce the following experiments: $\mathcal{F}_{f_0}^{3,n}$ generated by observations

$$dY_t^{3,n} = (b(f(t)) - b(f_0(t)))\, I(f_0(t))^{-1/2} dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0,1], \tag{5.31}$$

where $f \in \Sigma_{f_0}(\gamma_n)$, and $\mathcal{F}_{f_0}^{4,n}$ generated by observations

$$dY_t^{2,n} = (\Gamma(f(t)) - \Gamma(f_0(t)))\, dt + \frac{1}{\sqrt{n}} dW_t, \quad t \in [0,1], \tag{5.32}$$

where $f \in \Sigma_{f_0}(\gamma_n)$.

Proposition 5.2 *The experiments $\mathcal{F}_{f_0}^{3,n}$ and $\mathcal{F}_{f_0}^{4,n}$ are asymptotically equivalent. Moreover*

$$\Delta^2\left(\mathcal{F}_{f_0}^{3,n},\mathcal{F}_{f_0}^{4,n}\right) \le c_1 n^{-\frac{2\beta-1}{2\beta+1}}(\log n)^{\frac{4\beta}{2\beta+1}},$$

where c_1 is a constant depending only on $I_{\max}$, $I_{\min}$ and ε_0.

Proof. By Taylor expansion we have for any $\lambda, \lambda_0 \in \Lambda_0 = b(\Theta_0)$

$$F(\lambda) - F(\lambda_0) = (\lambda - \lambda_0)F'(\lambda_0) + \frac{1}{2}(\lambda - \lambda_0)^2 F''(\widetilde{\lambda}_0), \tag{5.33}$$

where $\widetilde{\lambda}_0 = \lambda_0 + \nu(\lambda - \lambda_0)$, $0 \le \nu \le 1$. Now it follows from (3.13) and (3.12) that

$$F'(\lambda_0) = 1/\sqrt{I(a(\lambda_0))}, \tag{5.34}$$

while, using (5.30), the second derivative of $F(\lambda)$ can easily be seen to satisfy

$$\left| F''(\widetilde{\lambda}_0) \right| \le c_2, \tag{5.35}$$

where c_2 is a constant depending only on $I_{\max}$, $I_{\min}$, ε_0. Put for brevity $g_0(t) = b(f_0(t))$ and $g(t) = b(f(t))$ for $f \in \Sigma_{f_0}(\gamma_n)$. It is easy to see that

$$|g(t) - g_0(t)| \le |f(t) - f_0(t)|\, I_{\max} \le \gamma_n I_{\max}. \tag{5.36}$$

Then by (5.33), (5.34), (5.35) and (5.36) we get

$$F(g(t)) - F(g_0(t)) = (g(t) - g_0(t))\, I(f_0(t))^{-1/2} + \nu c_3 \gamma_n^2,$$

with $c_3 = c_2 I_{\max}^2$ and $|\nu| \le 1$.

Put for brevity

$$\begin{aligned} m_3(t) &= (g(t) - g_0(t))\, I(f_0(t))^{-1/2}, \\ m_4(t) &= F(g(t)) - F(g_0(t)). \end{aligned}$$

Let $Q^{3,n}_{f_0,f}$ and $Q^{4,n}_{f_0,f}$ be the measures induced by observations (5.31) and (5.32) respectively. Then by formula (2.16) we get

$$\begin{aligned} H^2\left(Q^{3,n}_{f_0,f}, Q^{4,n}_{f_0,f}\right) &= 1 - \exp\left\{ -\frac{n}{8}\int_0^1 (m_3(t) - m_4(t))^2\, dt \right\} \\ &\le \frac{n}{8}\gamma_n^4 c_3^2 = c_1 n^{-\frac{2\beta-1}{2\beta+1}} (\log n)^{\frac{4\beta}{2\beta+1}}. \quad \blacksquare \end{aligned}$$

Finally, Theorem 3.3 can be obtained easily from the above propositions if we note that $\Delta(\mathcal{G}^n_{f_0}, \mathcal{F}^{1,n}_{f_0}) = 0$ and also $\Delta(\mathcal{F}^{2,n}_{f_0}, \mathcal{F}^{3,n}_{f_0}) = 0$, $\Delta(\mathcal{F}^{4,n}_{f_0}, \widehat{\mathcal{G}}^n_{f_0}) = 0$, i. e. these experiments are (exactly) equivalent by the remark following immediately after formula (2.15).

6 Global approximation

6.1 The preliminary estimator

The following lemma provides the preliminary estimator which is necessary for the globalization procedure over the parameter set Σ.

Lemma 6.1 *Let $\beta \in (\frac{1}{2}, 1)$. In the experiment $\mathcal{E}^n$ there is an estimator $\widehat{f}_n : X^n \to \Sigma$ taking a finite number of values and such that*

$$\sup_{f\in\Sigma} P^n_f \left(\left\| \widehat{f}_n - f \right\|_\infty > c_1\gamma_n \right) \le c_2 \frac{1}{\sqrt{n}},$$

where c_1 and c_2 are constants depending only on $I_{\max}$, $I_{\min}$, κ_0, L, β.

Proof. Let γ_n be given by (3.4) and $\delta_n = \gamma_n^{1/\beta}$. Introduce the kernel $K(u)$ as a bounded function of finite support such that

$$0 \le K(u) \le k_{\max}, \quad K(u) = 0, \ u \notin (-\tau, \tau), \quad \int_{-\tau}^{\tau} K(u) du = 1, \tag{6.1}$$

where $k_{\max}$ and τ are some absolute constants. We will assume also that $K(u)$ satisfies a Hölder condition with exponent β. Let

$$\rho_n(t) = \frac{1}{n\delta_n} \sum_{i=1}^{n} K\left(\frac{t_i - t}{\delta_n}\right), \tag{6.2}$$

where $t_i = i/n$, $i = 1, ..., n$. It is easy to see that there are two constants $\rho_{\min}$ and $\rho_{\max}$ such that for n large enough

$$0 < \rho_{\min} \le \rho_n(t) \le \rho_{\max} < \infty, \tag{6.3}$$

for any $t \in [0,1]$.

Consider the functions $f \in \Sigma$ and $g(t) = b\,(f(t))\,,\, t \in [0,1]$, where $b(\theta) = V'(\theta)$, $\theta \in \Theta_0$ (see Section 3.3). Define an estimator g_n^* of g as follows: for any $t \in [0,1]$ put

$$g_n^*(t) = \frac{1}{n\delta_n\rho_n(t)} \sum_{i=1}^{n} K\left(\frac{t_i - t}{\delta_n}\right) U(X_i),$$

where $U(x)$ is the sufficient statistic in the exponential experiment $\mathcal{E}$. The estimator g_n^* is known as Nadaraya-Watson estimator. We will show that there are two constants c_3 and c_4 depending only on $I_{\max}$, κ_0, L, β, $k_{\max}$, $\rho_{\min}$, τ such that

$$\sup_{f \in \Sigma} P_f^n \left(\|g_n^* - g\|_\infty > c_3\gamma_n\right) \le c_4 \frac{1}{n}. \tag{6.4}$$

First we note that by (6.2) for any $t \in [0,1]$

$$E_f^n g_n^*(t) - g(t) = \frac{1}{n\delta_n\rho_n(t)} \sum_{i \in J_n(t)} K\left(\frac{t_i - t}{\delta_n}\right) \left(g(t_i) - g(t)\right), \tag{6.5}$$

where $J_n(t) = \{i : t_i \in (t - \tau\delta_n, t + \tau\delta_n)\}$ and $\# J_n(t) \le 2\tau n\delta_n$. It is easy to see that since $f \in \Sigma$, we have for $i \in J_n(t)$

$$|g(t_i) - g(t)| \le I_{\max} L (2\tau\delta_n)^\beta = (2\tau)^\beta I_{\max} L \gamma_n. \tag{6.6}$$

From (6.1), (6.5) and (6.6) we have

$$\left\| E_f^n g_n^* - g \right\|_\infty \le c_5\gamma_n, \tag{6.7}$$

with some constant c_5 depending on $I_{\max}$, L, β, $k_{\max}$, $\rho_{\min}$, τ. To handle the difference $g_n^* - E_f^n g_n^*$ we remark that

$$g_n^*(t) - E_f^n g_n^*(t) = \frac{1}{n\delta_n\rho_n(t)} \sum_{i \in J_n(t)} K\left(\frac{t_i - t}{\delta_n}\right) \overline{U}(X_i), \tag{6.8}$$

where $\overline{U}(X_i) = U(X_i) - E_f^n U(X_i) = U(X_i) - g(t_i)$. Put for brevity $L_i(t) = K\left(\frac{t_i - t}{\delta_n}\right)$. Define a piecewise constant approximation of $L_i(t)$ as follows: put $\widetilde{L}_i(t) = L_i(s_k)$ for $t \in A_k$, where $A_1 = [0, s_1]$, $A_k = (s_{k-1}, s_k]$, $k = 2, ..., n^2$, $s_k = k/n^2$, $k = 0, ..., n^2$. Since the function $K(u)$ satisfies a Hölder condition with exponent β, there is a constant c_6 such that

$$\left\| L_i - \widetilde{L}_i \right\|_\infty \leq c_6 \left(\frac{1}{n}\right)^{2\beta}. \tag{6.9}$$

Then (6.8) and (6.3) imply

$$\left\| g_n^* - E_f^n g_n^* \right\|_\infty \leq Q_1 + Q_2,$$

where

$$\begin{aligned} Q_1 &\leq \frac{1}{n\delta_n \rho_{\min}} \sup_{t\in[0,1]} \left| \sum_{i=1}^n \left(L_i(t) - \widetilde{L}_i(t) \right) \overline{U}(X_i) \right|, \\ Q_2 &\leq \frac{1}{n\delta_n \rho_{\min}} \sup_{t\in[0,1]} \left| \sum_{i=1}^n \widetilde{L}_i(t) \overline{U}(X_i) \right|. \end{aligned}$$

Using (6.9) we get

$$Q_1 \leq \frac{c_6}{n^{2\beta+1}\delta_n \rho_{\min}} \sum_{i=1}^n \left| \overline{U}(X_i) \right|.$$

Put for brevity $u_n = c_6 \left(\rho_{\min} n^{2\beta+1} \delta_n \gamma_n\right)^{-1} \log n$. Then with a "free" constant $c_7 \geq 1$

$$P_f^n(Q_1 > c_7\gamma_n) \leq e^{-c_7 \log n} \prod_{i=1}^n E_f^n \exp\left\{u_n \left|\overline{U}(X_i)\right|\right\}.$$

Since $u_n \leq \varepsilon_0$ for n large enough, it is easy to see using Proposition 2.1 that

$$E_f^n \exp\left\{u_n \left|\overline{U}(X_i)\right|\right\} \leq \exp\left\{u_n c_8\right\},$$

where c_8 is a constant depending only on $I_{\max}$, ε_0. As $nu_n \to 0$ for $n \to \infty$, we have for sufficiently large n

$$P_f^n(Q_1 > c_7\gamma_n) \leq \exp\left\{-c_7 \log n + c_8 n u_n\right\} \leq 2\exp\left\{-c_7 \log n\right\} \leq 2\frac{1}{n}. \tag{6.10}$$

To obtain a bound for Q_2 we remark that $\widetilde{L}_i(t)$ is piecewise constant and $\widetilde{L}_i(s_k) = 0$ if $i \notin J_n(s_k)$. With $c_9 > 0$ being a "free" constant we obtain

$$P_f^n\left(Q_2 > c_9\gamma_n\right) \leq \sum_{k=1}^{n^2} P_f^n\left(\frac{1}{n\delta_n\rho_{\min}} \left| \sum_{i\in J_n(s_k)} \widetilde{L}_i(s_k)\overline{U}(X_i) \right| > c_9\gamma_n\right). \tag{6.11}$$

Put for brevity $v_n = \widetilde{L}_i(s_k) \left(\rho_{\min} n \delta_n \gamma_n\right)^{-1} \log n$. Then Chebyshev's inequality and the independence of the r. v.'s $\overline{U}(X_i)$, $i \in I_k$ imply

$$P_f^n\left(\frac{1}{n\delta_n\rho_{\min}} \sum_{i\in J_n(s_k)} \widetilde{L}_i(s_k)\overline{U}(X_i) > c_9\gamma_n\right) \leq e^{-c_9 \log n} \prod_{i\in J_n(s_k)} E_f^n \exp\left\{v_n \overline{U}(X_i)\right\}. \tag{6.12}$$

Since $v_n \leq k_{\max} \rho_{\min}^{-1} \kappa_0^{-(2\beta+1)/\beta} \gamma_n \leq \varepsilon_0$ (for n large enough), by Proposition 2.1 we obtain

$$\prod_{i \in J_n(s_k)} E_f^n \exp\left\{ v_n \overline{U}(X_i) \right\} \leq \exp\left\{ \frac{I_{\max}}{2} v_n^2 2\tau n \delta_n \right\} \leq \exp\left\{ c_{10} \log n \right\}, \tag{6.13}$$

for some constant c_{10} depending on $I_{\max}$, κ_0, β, $k_{\max}$, $\rho_{\min}$, τ. Choosing c_9 to be $c_{10} + 3$ we get from (6.12) and (6.13)

$$P_f^n \left(\frac{1}{n \delta_n \rho_{\min}} \sum_{i \in J_n(s_k)} \widetilde{L}_i(s_k) \overline{U}(X_i) > c_9 \gamma_n \right) \leq \exp\left\{ -3 \log n \right\}. \tag{6.14}$$

In the same way we establish that

$$P_f^n \left(\frac{1}{n \delta_n \rho_{\min}} \sum_{i \in J_n(s_k)} \widetilde{L}_i(s_k) \overline{U}(X_i) < -c_9 \gamma_n \right) \leq \exp\left\{ -3 \log n \right\}. \tag{6.15}$$

From (6.14), (6.15) and (6.11) we get

$$P_f^n \left(Q_2 > c_9 \gamma_n \right) \leq 2n^2 \exp\left\{ -3 \log n \right\} = 2\frac{1}{n}. \tag{6.16}$$

Now (6.10) and (6.16) give us

$$P_f^n \left(\left\| g_n^* - E_f^n g_n^* \right\|_\infty \geq c_{11} \gamma_n \right) \leq 4\frac{1}{n}, \tag{6.17}$$

for an appropriate constant c_{11}. Finally (6.7) and (6.17) imply (6.4).

Generally speaking g_n^* is not bounded. But it is easy to define another estimator on its basis which satisfies this requirement. For this it is enough to put

$$g_n^{**}(t) = \max\left\{ \min\left\{ g_n^*(t), \Lambda_{\max} \right\}, \Lambda_{\min} \right\},$$

where $\Lambda_{\min}$ and $\Lambda_{\max}$ are the ends of the interval Λ_0. The estimator g_n^{**} satisfies (6.4), since for any $f \in \Sigma$ we have $\Lambda_{\min} \leq g(t) = b(f(t)) \leq \Lambda_{\max}$ which in turn implies

$$\left\{ \left\| g_n^{**} - g \right\|_\infty > c_3 \gamma_n \right\} = \left\{ \left\| g_n^* - g \right\|_\infty > c_3 \gamma_n \right\}.$$

An estimator for f can be defined by putting $f_n^*(t) = a(g_n^{**}(t))$, where $a(\lambda)$ is the inverse of the function $b(\theta)$ (see also Section 3.3). Since the function $a(\lambda)$ is Lipschitz, we obtain from (3.12)

$$\left| f_n^*(t) - f(t) \right| \leq I_{\min}^{-1} \left| g_n^{**}(t) - g(t) \right|.$$

This implies that (6.4) is also satisfied with f_n^* and f replacing g_n^* and g, but with other constants (depending also on $I_{\min}$). Now we will define an estimator taking a finite number of values in Σ. Since the set Σ is compact in the uniform metric (it is equicontinuous and bounded, hence compact by the Arzelà-Ascoli theorem), we can cover it by a finite number of balls of radius γ_n and with centers $f_i \in \Sigma$, $i = 1, ..., M$. The estimate $\widehat{f}_n$ can be defined to be the f_i closest to the estimate f_n^*. In case of nonuniqueness take the f_i with lowest index. The estimator constructed has the properties claimed. ■

In particular, if we take $\mathcal{E}$ to be the Gaussian shift experiment, then from Lemma 6.1 we get the following.

Lemma 6.2 *Let $\beta \in (\frac{1}{2}, 1)$. In the experiment $\mathcal{G}'^{,n}$ there is an estimator $\widehat{f}_n : X^n \to \Sigma$ taking a finite number of values and such that*

$$\sup_{f\in\Sigma} P_f^n \left(\left\| \widehat{f}_n - f \right\|_\infty > c_1 \gamma_n \right) \leq c_2 \frac{1}{\sqrt{n}},$$

where c_1 and c_2 are constants depending only on $I_{\max}$, $I_{\min}$, κ_0, L, β.

6.2 Proof of global equivalence

In this section we prove Theorem 3.10.

Let $\mathcal{E}^n$ and $\mathcal{G}'^{,n}$ be the experiments defined by (3.2), (3.3) and (3.21), (3.22). Let $f_0 \in \Sigma$. 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 $X_{(i)} = X$, $R_i = R$, $\mathbf{S}_i = X$ if i is odd and $\mathbf{S}_i = R$ if i is even, $i \in J$. Consider the following product (local) experiments corresponding to observations at points t_i with even indexes $i \in J$:

$$\mathcal{E}_{f_0}''^{,n} = \bigotimes_{i\in J''} \mathcal{E}_{f_0,t_i}, \quad \mathcal{G}_{f_0}''^{,n} = \bigotimes_{i\in J''} \mathcal{G}_{f_0,t_i},$$

where

$$\begin{aligned} \mathcal{E}_{f_0,t_i} &= \left(X, \mathcal{B}(X), \{P_{f(t_i)} : f \in \Sigma_{f_0}(\gamma_n)\}\right), \\ \mathcal{G}_{f_0,t_i} &= \left(R, \mathcal{B}(R), \left\{\widehat{Q}_{f(t_i)} : f \in \Sigma_{f_0}(\gamma_n)\right\}\right). \end{aligned}$$

(cf. (3.20)-(3.21) for the definition of $\widehat{Q}_{f(t_i)}$). Along with this introduce global experiments

$$\mathcal{E}'^{,n} = \bigotimes_{i\in J'} \mathcal{E}_{t_i}, \quad \mathcal{F}^n = \bigotimes_{i=1}^{n} \mathcal{F}_i^n,$$

where $\mathcal{F}_i^n = \mathcal{E}_{t_i}$ if i is odd and $\mathcal{F}_i^n = \mathcal{G}_{t_i}$ if i is even, $i \in J$ and where

$$\begin{aligned} \mathcal{E}_{t_i} &= \left(X, \mathcal{B}(X), \{P_{f(t_i)} : f \in \Sigma\}\right), \\ \mathcal{G}_{t_i} &= \left(R, \mathcal{B}(R), \left\{\widehat{Q}_{f(t_i)} : f \in \Sigma\right\}\right). \end{aligned}$$

We will show that the global experiments $\mathcal{E}^n$ and $\mathcal{F}^n$ are asymptotically equivalent. Toward this end we note that by Theorem 3.5 the experiments $\mathcal{E}_{f_0}''^{,n}$ and $\mathcal{G}_{f_0}''^{,n}$ are asymptotically equivalent uniformly in $f_0 \in \Sigma$. Theorem 3.5 in particular implies that the one-sided deficiency $\delta\left(\mathcal{E}_{f_0}''^{,n}, \mathcal{G}_{f_0}''^{,n}\right)$ satisfies for any $f_0 \in \Sigma$

$$\delta\left(\mathcal{E}_{f_0}''^{,n}, \mathcal{G}_{f_0}''^{,n}\right) \leq \varepsilon_n \equiv \left(c_1 n^{-\frac{2\beta-1}{2\beta+1}} (\log n)^{\frac{14\beta+5}{2\beta+1}}\right)^{1/2},$$

where c_1 is a constant depending only on $I_{\min}$, $I_{\max}$, κ_0, L, β. Let $\|\cdot\|$ denote the total variation norm for measures and let $P_f''^{,n}$, $Q_f''^{,n}$ be the product measures corresponding to the experiments $\mathcal{E}_{f_0}''^{,n}$ and $\mathcal{G}_{f_0}''^{,n}$:

$$P_f''^{,n} = \bigotimes_{i\in J''} P_{f(t_i)}, \quad Q_f''^{,n} = \bigotimes_{i\in J''} \widehat{Q}_{f(t_i)}. \tag{6.18}$$

By lemma 9.2 in [31] for any $f_0 \in \Sigma$ there is a Markov kernel $K^n_{f_0} : (X^{\prime\prime,n}, \mathcal{B}(R^{\prime\prime,n})) \to [0,1]$ such that

$$\sup_{f_0 \in \Sigma} \sup_{f \in \Sigma_{f_0}(\gamma_n)} \left\| K^n_{f_0} \cdot P^{\prime\prime,n}_f - Q^{\prime\prime,n}_f \right\| \leq \varepsilon_n, \tag{6.19}$$

Let us establish that there is a Markov kernel $M^n : (X^n, \mathcal{B}(\mathbf{S}^n)) \to [0,1]$ such that

$$\sup_{f \in \Sigma} \left\| M^n \cdot P^n_f - F^n_f \right\| \leq c_2 \varepsilon_n, \tag{6.20}$$

for some constant c_2. 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 of lemma 6.1 in the experiment $\mathcal{E}^{\prime,n}$. It is easy to see that

$$\begin{aligned} \left(M^n \cdot P^n_f \right)(B) &= \int_{X^{\prime,n}} \int_{X^{\prime\prime,n}} M^n \left((x'; x''), B \right) P^{\prime\prime,n}_f(dx'') P^{\prime,n}_f(dx') \\ &= \int_{X^{\prime,n}} \int_{R^{\prime\prime,n}} \mathbf{1}_B \left((x'; y'') \right) \left(K^n_{\widehat{f}_n(x')} \cdot P^{\prime\prime,n}_f \right)(dy'') P^{\prime,n}_f(dx') \end{aligned} \tag{6.21}$$

and

$$F^n_f(B) = \int_{X^{\prime,n}} \int_{R^{\prime\prime,n}} \mathbf{1}_B \left((x'; y'') \right) Q^{\prime\prime,n}_f(dy'') P^{\prime,n}_f(dx'), \tag{6.22}$$

where $P^{\prime,n}_f$ is the measure in the experiment $\mathcal{E}^{\prime,n}$ defined by the analogy with $P^{\prime\prime,n}_f$ in (6.18), but with J' replacing J''. By Lemma 6.1 there are two constants c_3 and c_4 depending only on $I_{\max}$, $I_{\min}$, κ_0, L, β such that

$$P^{\prime,n}_f(A^c_f) \leq c_4 \varepsilon_n \tag{6.23}$$

where $A_f = \left\{ x' \in X^{\prime,n} : \left\| \widehat{f}_n(x') - f \right\|_\infty \leq c_3 \gamma_n \right\}$. Then (6.21) and (6.22) imply

$$\begin{aligned} \left| \left(M^n \cdot P^n_f \right)(B) - F^n_f(B) \right| \leq\ & 2 P^{\prime,n}_f(A^c_f) \\ & + \int_{A_f} \sup_{f_0 \in \Sigma} \sup_{f \in \Sigma_{f_0}(\gamma_n)} \left\| K^n_{f_0} \cdot P^{\prime\prime,n}_f - Q^{\prime\prime,n}_f \right\| P^{\prime,n}_f(dx'). \end{aligned}$$

Using (6.19) and (6.23) we obtain (6.20). This implies that the one-sided deficiency $\delta(\mathcal{E}^n, \mathcal{F}^n)$ is less that $c_2 \varepsilon_n$. The bound for $\delta(\mathcal{F}^n, \mathcal{E}^n)$ can be obtained in the same way, using Lemma 6.2. This proves that the Le Cam distance between $\mathcal{E}^n$ and $\mathcal{F}^n$ is less that $c_2 \varepsilon_n$. 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 $\mathcal{G}^n$. As to the rate of convergence, it is straightforward from the above inequality (6.20) and an analogous one for the pair $\mathcal{F}^n$ and $\mathcal{G}^n$. Theorem 3.10 is proved.

Acknowledgement. The authors wish to thank Iain Johnstone and Enno Mammen for pointing out important theoretical connections.

References

[1] Amari, S.-I., Barndorff-Nielsen, R. E., Lauritzen, S. L., Rao, C. R. (1987). *Differential Geometry in Statistical Inference.* IMS Lecture Notes-Monograph Series, Vol. 10, Institute of Mathematical Statistics, Hayward, California.

[2] Andersen, P. A., Borgan, Ø., Gill, R. D., Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York etc.

[3] Barndorff-Nielsen, O. E. and Cox D. R. (1989). *Asymptotic Techniques for use in Statistics.* Monographs on Statistics and Applied Probability, 31. Chapman and Hall, London.

[4] 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.

[5] Brown, L. D. and Low, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** (6) (to appear)

[6] Brown, L. D. and Low, M. (1996). Personal communication, "Asymptotic Methods in Stochastic Dynamics and Nonparametric Statistics", Berlin

[7] Brown, L. D. and Zhang, C.-H. (1996). Asymptotic nonequivalence of nonparametric experiments when the smoothness index is 1/2. *Unpublished Manuscript.*

[8] Čencov, N. N. (1982). *Statistical Decision Rules and Optimal Inference.* Translations of Mathematical Monographs 53, Amer. Math. Soc., Providence

[9] Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

[10] Donoho, D. L. and Liu, R. C. (1991). Geometrizing rates of convergence, III. *Ann. Statist.* **19**, 668-701.

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

[12] Einmahl, U. (1989). Extensions on results of Komlós, Major and Tusnády to the multivariate Case. *Journal of Multivariate Analysis* 28, 20-68.

[13] Einmahl, U. and Mason, D. M. (1995). Gaussian approximation of local empirical processes indexed by functions. *Unpublished manuscript.*

[14] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Monographs on Statistics and Applied Probability, 66. Chapman & Hall, London.

[15] Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *J. Appl. Prob.* **30** 790-804

[16] Genon-Catalot, V., Laredo, C. and Picard, D. (1992). Nonparametric estimation of the diffusion coefficient by wavelet methods. *Scand. J. Statist.* **19,** 317-335

[17] Golubev, Yu., and Nussbaum, M. (1995). Asymptotic equivalence of spectral density estimation and white noise. *Unpublished manuscript.*

[18] Grama, I. G. and Nussbaum, M. (1996). A functional Komlós-Major-Tusnády approximation for partial sums of independent random variables. *Unpublished manuscript.*

[19] Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models.* Monographs on Statistics and Applied Probability, 58, Chapman & Hall, London.

[20] Ibragimov, I. A. and Khasminskii, R. Z. (1981). *Statistical Estimation. Asymptotic Theory.* Springer-Verlag, New-York etc.

[21] Jacod, J. and Shiryaev A. N. (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag, New York etc.

[22] 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.

[23] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent rv's and the sample df. I *Z. Wahrsch. verw. Gebiete* **32**, 111-131.

[24] Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent rv's and the sample df. II *Z. Wahrsch. verw. Gebiete* **34**, 33-58.

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

[26] Le Cam, L. and Yang, G. (1990). *Asymptotics in Statistics: Some Basic Concepts.* Springer-Verlag, New-York etc.

[27] Low, M. G. (1992). Renormalization and white noise approximation for nonparametric functional estimation problems. *Ann. Statist.* **20**, 545-554.

[28] Mammen, E. and van de Geer, S. (1995). Penalized quasi-likelihood estimation in partial linear models. *Ann. Statist., to appear.*

[29] Massart, P. (1989). Strong approximation for multivariate empirical and related processes, via K.M.T. constructions. *Ann. Probab.* **17**, 266-291.

[30] 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.

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

[32] Rio, E. (1993). Strong approximation for set-indexed partial sum Processes, via K.M.T. constructions I. *Ann. Probab.* **21**, 759-790.

[33] Rio, E. (1993). Strong approximation for set-indexed partial sum Processes, via K.M.T. constructions I. *Ann. Probab.* **21**, 1706-1727.

[34] Rio, E. (1994). Local invariance principles and their applications to density estimation. *Probability Theory and Related Fields.* **98**, 21-45.

[35] Sakhanenko, A. (1984). The rate of convergence in the invariance principle for non-identically distributed variables with exponential moments. *Limit theorems for sums of random variables: Trudy Inst. Matem., Sibirsk. Otdel. AN SSSR.* Vol. 3, 3-49 (In Russian).

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

Institute of Mathematics
Academy of Sciences
Academiei str. 5
Chisinau 277028, Moldova
e-mail: 16grama@mathem.moldova.su

Weierstrass Institute
Mohrenstr. 39
D-10117 Berlin, Germany
e-mail: nussbaum@wias-berlin.de