Statistics & Decisions 21, 197–218 (2003)

Asymptotic equivalence for a model of independent non identically distributed observations

Michael Jähnisch, Michael Nussbaum

Received: March 12, 2003; Accepted: October 8, 2003

Summary: It is shown that a nonparametric model of independent non identically distributed observations on the unit interval can be approximated, in the sense of Le Cam's Δ-distance, by a bivariate Gaussian white noise model. The parameter space is a smoothness class of conditional densities uniformly bounded away from zero on the unit square. The proof is based on coupling of likelihood processes via a functional Hungarian construction of the sequential empirical process and the Kiefer–Müller process.

1 Introduction and main result

In the asymptotic theory of experiments, regression models have served as a prime example, along with i. i. d. models. With regard to local Gaussian limits involving an $n^{-1/2}$ renormalization rate, the most general theory for nonparametric regression has been worked out by Millar in 1982 [11]. Millar's neighborhoods and limit experiments for regression are an analog of the theory for the empirical distribution function in the i. i. d. case. It is well-known however that local Gaussian limits over $n^{-1/2}$-sized parametric neighborhoods are insufficient to treat many nonparametric function estimation problems, and should be replaced by global approximations in terms of Le Cam's Δ-distance. The first such approximation, or asymptotic equivalence, in a truly nonparametric model (after a result by Le Cam about Poisson experiments in [10]) was obtained by Brown and Low [1] for a Gaussian regression model. This regression result gave rise to the corresponding conjecture for the nonparametric i. i. d. case which was subsequently confirmed in [12]. The methodology of the latter paper, i. e. coupling of likelihood processes, then allowed to treat nonparametric regression also in the non-Gaussian case (Grama and Nussbaum [5, 6]). The purpose of the present paper is to extend the result of [12] to the case of independent but non identically distributed (i. n. i. d.) observations on the

AMS 1991 subject classifications: Primary 62G07; Secondary 62B15, 62G20
Key words and phrases: nonparametric experiments, deficiency distance, global Gaussian approximation, likelihood process, Hungarian construction, sequential empirical process, regression models

unit interval. Assume that $h_s(\cdot)$ is a family of probability densities on $[0,1]$ indexed by $s \in [0,1]$. We observe independent data X_{in}, $i = 1,\dots,n$ such that X_{in} has density $h_{i/n}$. If h_s does not depend on the index s, i. e. $h_s = h_0$ then we recover the identically distributed case of [12] where under smoothness and boundedness conditions on h_0 an approximation in Δ-distance by the white noise model

$$dy(t) = h_0^{1/2}(t)dt + \frac{1}{2}n^{-1/2}dW(t),\ t \in [0,1]$$

was established. Below we will show (cf. Theorem 1.1) that this result extends to the nonindentically distributed case, in the sense that if $h_s(t)$ fulfills a joint smoothness condition in s and t and is bounded away from 0 then the accompanying Gaussian experiment is

$$dy(s,t) = h_s^{1/2}(t)dsdt + \frac{1}{2}n^{-1/2}dW(s,t), \quad (s,t) \in [0,1]^2$$

where W is a two-dimensional Brownian sheet.

In Grama and Nussbaum [6] asymptotic equivalence to a Gaussian experiment is proved for a related model of i. n. i. d. observations: for a fixed (known) parametric family of densities f_ϑ, $\vartheta \in \Theta$, observations X_{in} have density $f_{g(i/n)}$ where g is an unknown smooth regression function on $[0,1]$. Thus formally we also have a family of densities $h_s(\cdot) = f_{g(s)}(\cdot)$, but this is a narrower class since it is not given solely in terms of smoothness conditions in the two variables (s,t). The model in [6] is closer to a semiparametric one since f_ϑ, $\vartheta \in \Theta$ is a known parametric family. On the other hand, in [6] the X_{in} need not take values in the unit interval and can be discrete.

Let us also briefly discuss the general setup of Millar [11] who aims at local limit experiments. The starting point there is also a parametric family of densities f_ϑ, $\vartheta \in \Theta$ combined with a regression function g on $[0,1]$ such that X_{in} has density $f_{g(i/n)}$. However the perturbation neighborhoods for the LAN-result there are not just in terms of the regression function g but they also account for nonparametric deviations from the model $\{f_\vartheta, \vartheta \in \Theta\}$. This general limit experiment theory for regression indeed suggests that smoothness conditions on the function of two variables $h_s(t)$ enable an asymptotic equivalence result.

To be precise, we write $h(s,t)$ for $h_s(t)$ where $h_s(\cdot)$, $s \in [0,1]$ is a family of Lebesgue densities on $[0,1]$. For functions g on $[0,1]^2$ and for any $\sigma \in (0,1)$ define a Hölder seminorm by

$$\left|g|C^\sigma\right| := \sup \frac{|g(x)-g(y)|}{|x-y|^\sigma}$$

where the supremum is taken over all $x \in [0,1]^2$, $y \in [0,1]^2$ such that $x \neq y$. We also denote $\left|g|C^0\right|$ with $\|g\|_\infty$, where $\|g\|_\infty = \sup_{x\in[0,1]^2}|g(x)|$. Let $z = (z_1,z_2) \in \mathbb{N}_0^2$ be a multiindex where $\mathbb{N}_0 = \mathbb{N}\cup\{0\}$, let $|z| = z_1 + z_2$ and let D^z be the differential operator $D^z = \frac{\partial^{|z|}}{\partial^{z_1}\partial^{z_2}}$. For any positive smoothness index α, define a Hölder norm

$$\left\|g|C^\alpha\right\| := \sum_{|z|\le\alpha}\left\|D^z g\right\|_\infty + \sum_{|z|=[\alpha]}\left|D^z g|C^{\alpha-[\alpha]}\right|$$

and for any $M > 0$ consider a Hölder smoothness class on $[0, 1]^2$

$$\mathcal{C}^\alpha(M) := \left\{g : \|g|\mathcal{C}^\alpha\| \le M\right\}.$$

Furthermore, we will require that densities $h(s, \cdot)$ are uniformly bounded away from 0. For $\epsilon > 0$ consider a set of continuous functions on $[0, 1]^2$

$$\mathcal{F}_{\ge\epsilon} := \left\{ h : h(s,t) \ge \epsilon, \forall\ (s,t) \in [0,1]^2, \int_0^1 h(s,t)dt = 1, \forall\ s \in [0,1] \right\}.$$

We write $X \sim f$ if the random variable X has a Lebesgue density f. For a given parameter space $\Sigma \subset \mathcal{F}_{\ge\epsilon}$, consider the two experiments indexed by $h \in \Sigma$

$$E_n : \quad X_{1n}, \ldots, X_{nn} \text{ independent, } X_{in} \sim h(i/n, \cdot),$$

$$F_n : \quad dy(s,t) = h^{1/2}(s,t)dsdt + \tfrac{1}{2}n^{-1/2}dW(s,t), \quad (s,t) \in [0,1]^2.$$

Theorem 1.1 *For $M > 0$, $0 < \epsilon < 1$ and $\alpha > 3$, let $\Sigma \subset \mathcal{C}^\alpha(M) \cap \mathcal{F}_{\ge\epsilon}$. Then we have*

$$\lim_{n\to\infty} \Delta(E_n, F_n) = 0,$$

i. e. the experiments E_n and F_n are asymptotically equivalent.

Remark 1.2 Similarly to the approach in [12], the proof is based on coupling of likelihood processes via a functional Hungarian construction. In this construction (Theorem 4.3 below) the sequential empirical process and the Kiefer–Müller process replace the empirical process and the Brownian bridge which figure in the i. i. d. case; cf. Koltchinski [9] for the respective coupling result. Theorem 4.3 below also generalizes one aspect of the functional Hungarian construction for the partial sum process (Grama and Nussbaum [7]) which was used for the likelihood processes in [6].

Remark 1.3 The smoothness condition $\alpha > 3$ appears not to be optimal; since $\alpha > 1/2$ is a sharp condition in the i. i. d. case for Hölder smoothness on the unit interval (Brown and Zhang [3]), it can be conjectured that $\alpha > 1$ is a minimal condition on the unit square. A proof that $\alpha > 1$ is necessary for the present smoothness classes $\mathcal{C}^\alpha(M)$ can be found in [8]. The reason for our gap in terms of α is that in the i. n. i. d. case, there seems to be no obvious analog of the technique applied in [12] where the empirical process was first approximated by a Poisson process with a convenient independence structure on the sample space In the present paper, we are only using the a priori independence of the data X_{in}, $i = 1, \ldots, n$. Note that in the i. i. d. case, an alternative to Poissonization has recently been found by Carter [4] who applied a multiresolution scheme for Gaussian approximation of multinomial experiments.

Remark 1.4 Consider i. i. d. observations $Z_1, \ldots, Z_n$ on the unit square distributed as $Z = (X, S)$ where S is uniform on $[0, 1]$ and conditionally on S, $X \sim h(S, \cdot)$. This represents the "random design" analog of the present model; the result of Brown et al. [2] for Gaussian regression with random design suggests that the above white noise experiment F_n is still asymptotically equivalent. More generally, it can be conjectured that the white noise approximation with signal $h^{1/2}$ is valid for i. i. d. bivariate observations having smooth density h.

Remark 1.5 The Gaussian approximations of Millar [11] and Grama and Nussbaum [6] apply to the important case of the location type regression model where $X_{in} \sim f(\cdot - g(i/n))$, f is a density on the real line and g is a regression function. This model is not covered by our results since the conditional densities $h(s, \cdot)$ all have support [0, 1] and are bounded away from 0. Indeed the corresponding problem for the i. i. d. case (to include location families in the nonparametric density class in [12]) is still open.

2 The local approximation

We will prove Theorem 1.1 by localizing the parameter space. Let $\gamma_n := n^{-1/4}(\log n)^{-5}$ and for some fixed $h_0 \in \Sigma$, let $\Sigma_n(h_0)$ be defined as follows:

$$\Sigma_n(h_0) := \left\{h \in \Sigma : \|D^z h - D^z h_0\|_\infty \le \gamma_n, \forall\, |z| \le 1\right\}$$

(where $z \in \mathbb{N}_0^2$ and $|z| = z_1 + z_2$).

Remark 2.1 The quantity γ_n is the minimax rate of convergence (up to logarithmic factors) for an estimator of a first order derivative of the function h, provided the function is Hölder continuous with smoothness 3. To that extent, $\Sigma_n(h_0)$ is of minimal size such that an estimator for h_0 (for a true h_0) lies in that vicinity, with high probability as $n \to \infty$.

Consider now the following localized experiments:

$$E_{0,n}(h_0) : \; X_{1n}, \ldots, X_{nn} \text{ independent, } X_{in} \sim h(i/n, \cdot), \quad h \in \Sigma_n(h_0)$$

$$E_{1,n}(h_0) : \; dy_i(t) = (\lambda_{h,h_0}(i/n, t) - \textstyle\int_0^1 \lambda_{h,h_0}(i/n, s)ds)dt + dW_i(t), \quad h \in \Sigma_n(h_0),$$

$$t \in [0, 1], i = 1, \ldots, n$$

where $W_1, \ldots, W_n$ are independent Brownian motions and

$$\lambda_{h,h_0}(s, t) := \log\left(\frac{h}{h_0}\left(s, P_0^{-1}(s, t)\right)\right),$$

where $P_0^{-1}(s, t)$ is the inverse mapping of

$$t \longmapsto P_0(s, t) = \int_0^t h_0(s, v)dv.$$

Theorem 2.2 *For the previously defined experiments we have*

$$\lim_{n\to\infty} \Delta(E_{0,n}(h_0), E_{1,n}(h_0)) = 0$$

uniformly over $h_0 \in \Sigma$.

The main tool for proving this result will be the following well know inequality, where we assume to have versions of the likelihood processes $\Lambda_i^*(\theta)$ of experiments E_i on some common probability space:

Fact 2.3

$$\Delta(E_0, E_1) \leq \sup_{\theta\in\Theta} E_P\left|\Lambda_0^*(\theta) - \Lambda_1^*(\theta)\right|$$

(Cf. [12, p. 2404]).

Remark 2.4 The likelihood processes have the properties of densities since

$$\text{i) } \Lambda_i^*(\theta) \geq 0 \qquad \text{ii) } E_P\Lambda_i^*(\theta) = 1.$$

Thus the squared Hellinger distance between likelihood processes can be defined by

$$H^2\left(\Lambda_0^*(\theta), \Lambda_1^*(\theta)\right) := E_P\left((\Lambda_0^*(\theta))^{1/2} - (\Lambda_1^*(\theta))^{1/2}\right)^2$$

and since the L^1-distance $E_P\left|\Lambda_0^*(\theta) - \Lambda_1^*(\theta)\right|$ is bounded by the Hellinger distance from above, we obtain

$$\begin{aligned}\Delta^2(E_0, E_1) &\leq \sup_{\theta\in\Theta}\left(E_P\left|\Lambda_0^*(\theta) - \Lambda_1^*(\theta)\right|\right)^2 \\ &\leq \sup_{\theta\in\Theta} H^2\left(\Lambda_0^*(\theta), \Lambda_1^*(\theta)\right).\end{aligned}$$

Furthermore we need the following well-known equalities for computing the Hellinger distance.

Fact 2.5 *Let P_i be the probability measures on $(\mathcal{C}[0,1], \mathcal{B}_{\mathcal{C}[0,1]})$ that are induced by the distributions of the stochastic processes y_i where*

$$dy_i(t) = g_i(t)dt + \epsilon dW(t) \qquad t \in [0,1], i = 1, 2$$

(W is a Brownian motion, $\epsilon > 0$ and $g_1, g_2 \in L^2([0,1], \lambda)$). Then

$$H^2(P_1, P_2) = 2\left(1 - \exp\left(-\frac{1}{8\epsilon^2}\|g_1 - g_2\|_2^2\right)\right).$$

Furthermore, a similar statement holds for the Brownian sheet:

Fact 2.6 *Let P_i be the probability measures on $(\mathcal{C}[0,1]^2, \mathcal{B}_{\mathcal{C}[0,1]^2})$ that are induced by the distributions of the stochastic processes y_i where*

$$dy_i(s,t) = g_i(s,t)dsdt + \epsilon dW(s,t) \qquad (s,t) \in [0,1]^2, i = 1, 2$$

(W is a Brownian sheet, $\epsilon > 0$ and $g_1, g_2 \in L^2([0,1]^2, \lambda^2)$). Then

$$H^2(P_1, P_2) = 2\left(1 - \exp\left(-\frac{1}{8\epsilon^2}\|g_1 - g_2\|_2^2\right)\right).$$

The likelihood processes of factorized experiments bound the Hellinger distance of the product processes due to the following lemma:

Lemma 2.7 *Let $P_1, \dots, P_n$ and $Q_1, \dots, Q_n$ be probability measures. Then, for the product measures $P^n := \otimes_{i=1}^n P_i$ and $Q^n := \otimes_{i=1}^n Q_i$:*

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

Proof: [13, Lemma 2.19]. □

Due to Fact 2.3 we can now bound the Δ-distance of experiments $E_{0,n}(h_0)$ and $E_{1,n}(h_0)$ from above by constructing versions of their likelihood processes on some common probability space such that these processes are close to each other. The likelihood processes can be easily computed. Let $\Lambda_{i,n}(h, h_0)$ be the likelihood process of $E_{i,n}(h_0)$, then

$$\Lambda_{0,n}(h, h_0) = \exp\left[n^{1/2}\widehat{G}_n(\lambda_{h,h_0}) + \sum_{i=1}^n \int_0^1 \lambda_{h,h_0}(i/n, t)dt \right] \tag{2.1}$$

where

$$\widehat{G}_n(s, t) := \frac{1}{\sqrt{n}} \sum_{i=1}^{[ns]} (1_{[0,t]}(z_i) - t)$$

is the sequential empirical process. The z_is are i. i. d. uniform [0, 1] random variables and 1_A is the indicator function of set A; furthermore, for $f \in L^2([0,1]^2, \lambda^2)$, define

$$\widehat{G}_n(f) := \int_{[0,1]^2} f d\widehat{G}_n.$$

For the Gaussian experiment $E_{1,n}(h_0)$ the likelihood process is the following:

$$\Lambda_{1,n}(h, h_0) = \exp\left(K_n(\lambda_{h,h_0}) - \frac{1}{2} \sum_{i=1}^n \mathrm{Var}(\lambda_{h,h_0}(i/n, z_i)) \right). \tag{2.2}$$

Again, $z_1, \dots, z_n$: i. i. d. $\sim U[0, 1]$ and

$$K_n(s, t) := \sum_{i=1}^{[ns]} B_i(t)$$

is a discretized version of the Kiefer–Müller process, the B_i are independent Brownian bridges and $K_n(f)$ is the stochastic integral over f with respect to the process K_n.

Remark 2.8 We could now use Theorem 4.3 in order to define versions of the likelihood processes $\Lambda_{0,n}$ and $\Lambda_{1,n}$ for which we could estimate the Hellinger distance. However, this would not lead to the desired result. The reason is that the coupling result 4.3 is applied in an optimal way only if the rate of convergence of the parameter space $\Sigma_n(h_0)$ is the inverse of the square root of the number of observations in each experiment.

We will therefore split the experiments $E_{0,n}(h_0)$ and $E_{1,n}(h_0)$ into smaller factors for which this assumption is fulfilled. These doubly localized experiments will be defined by taking only a fraction of the observations of the original experiment (cp. Definition 4.4). This idea was already applied in Nussbaum [12] and Grama and Nussbaum [5]. As a consequence, the likelihood processes of experiments $E_{j,n}(h_0)$ $(j = 0, 1)$ are the products of the independent likelihood processes of the doubly localized experiments respectively. Theorem 4.3 will then be used to construct versions of the likelihood processes of these doubly localized experiments. Thus we constructed the processes by decomposing them into independent factors and applying the Hungarian construction to each factor.

We finally obtain the following estimate of the Hellinger distance:

Lemma 2.9 *There are versions $\Lambda^*_{0,n}(h, h_0)$ and $\Lambda^*_{1,n}(h, h_0)$ of the likelihood processes of experiments $E_{0,n}(h_0)$ and $E_{1,n}(h_0)$ on some common probability space, such that*

$$\sup_{h_0 \in \Sigma} \sup_{h \in \Sigma_n(h_0)} H^2\big(\Lambda^*_{0,n}(h, h_0), \Lambda^*_{1,n}(h, h_0)\big) \longrightarrow 0$$

holds for $n \to \infty$, where $H^2(\cdot, \cdot)$ is the Hellinger distance.

This lemma leads directly to the proof of Theorem 2.2 by invoking Remark 2.4. In order to prove Theorem 1.1, we need asymptotic equivalence of $E_{0,n}(h_0)$ to a Gaussian experiment which does not depend explicitly on the center h_0 of the parameter space. Otherwise it will not be possible to globalize the equivalence result. The next theorem states local asymptotic equivalence of $E_{1,n}(h_0)$ to four other Gaussian experiments, one of which does not depend on h_0.

Theorem 2.10 *Consider the following experiments:*

$$E_{2,n}(h_0): \quad dy_i(t) = (h - h_0)(i/n, t)dt + h_0^{1/2}(i/n, t)dW_i(t)$$

$$E_{3,n}(h_0): \quad dy_i(t) = \big(h^{1/2} - h_0^{1/2}\big)(i/n, t)dt + \tfrac{1}{2}dW_i(t)$$

(Let $t \in [0, 1]$, $i = 1, \ldots, n$ and W_i are independent Brownian motions and $h \in \Sigma_n(h_0)$.)

$$E_{4,n}(h_0): \quad dy(s, t) = \log\big(h\big(s, P_0^{-1}(s, t)\big)\big)dsdt + n^{-1/2}dW(s, t)$$

$$F_n(h_0): \quad dy(s, t) = h^{1/2}(s, t)dsdt + \tfrac{1}{2}n^{-1/2}dW(s, t).$$

(Let $(s, t) \in [0, 1]^2$, $h \in \Sigma_n(h_0)$ and W be a two-dimensional Brownian sheet.) Then, each of these experiments is asymptotically equivalent to $E_{1,n}(h_0)$, uniformly over $h_0 \in \Sigma$.

Together with Theorem 2.2 and the triangle inequality for the Δ-distance, we proved a local version of Theorem 1.1:

Corollary 2.11 *Uniformly over $h_0 \in \Sigma$, we have*

$$\lim_{n\to\infty} \Delta(E_{0,n}(h_0), F_n(h_0)) = 0.$$

Remark 2.12 Experiment $F_n(h_0)$ no longer depends on h_0. The reason for this is that in experiment $E_{3,n}(h_0)$, one can omit the term $h_0^{1/2}$ in the observations without changing the equivalence class of the experiment. Function h_0 is an a priori known parameter and therefore we can transform the observed data by adding the integral over $h_0^{1/2}(i/n, \cdot)$ to the ith observation.

3 Globalization of the results

Corollary 2.11 is somewhat unsatisfactory since in practice one can not assume to have such prior knowledge on the function h. However, we can globalize this result following the ideas that are described in detail in Nussbaum [12, sec. 9, p. 2425]. The proofs of the equivalence results (e. g. Equations (3.1) and (3.2)) work exactly as in [12, p. 2425] using the properties of the estimators of Lemmas 3.1 and 3.2.

We will proceed as follows: first of all, split the observations of E_n into two sets of the same size. With the first set of observations we will construct an estimator $\widehat{h}_n$ for h. Then we define a new experiment $F_n^{\#}$, which is almost the same as E_n but where the second set of observations in E_n is replaced by its "locally asymptotic equivalent" set of observations from F_n. If the estimator $\widehat{h}_n$ fulfills a certain optimality criterion (namely Lemma 3.1), then one can show that E_n and $F_n^{\#}$ are asymptotically equivalent. Lemma 3.1 states that the estimator $\widehat{h}_n$ is asymptotically, with probability tending to one, an element of the set $\Sigma_n(h)$ – uniformly over $h \in \Sigma$. For that reason the radius γ_n of the set $\Sigma_n(h)$ should not tend to zero faster than the minimax rate of convergence for an estimator of h. We will then apply this procedure again to $F_n^{\#}$ in order to replace the first set of observations by its "asymptotically equivalent set". Again, we need an estimator for h which is derived from the second part of the observations in $F_n^{\#}$ and which has to fulfill the same optimality criterion (Lemma 3.2).

More precisely, we have:

Lemma 3.1 *Let $N_n = [n/2]$. Then there exists a sequence of estimators $\widehat{h}_n$ in $E_{0,n}(h_0)$ that depend only on the observations $\{y_{2i}; i = 1, \dots, N_n\}$ and for which*

$$\inf_{h\in\Sigma} P_{n,h}(\widehat{h}_n \in \Sigma_n(h)) \longrightarrow 1$$

(as $n \to \infty$) holds. Without loss of generality we can assume that this estimator takes values only in the finite set $\Sigma_{0,n} \subset \Sigma$.

We will now define a compound experiment $F_n^{\#}$ with the following independent observations:

$$\left\{\{y_{2i};\, i = 1, \ldots, N_n\},\, (y(s,t))_{(s,t)\in[0,1]^2}\right\}$$

where the y_i are independent with densities $y_j \sim h(j/n, \cdot)$; $y(s,t)$ is given by

$$dy(s,t) = h^{1/2}(s,t)dsdt + \frac{1}{2}(n - N_n)^{-1/2}dW(s,t),$$

and $h \in \Sigma$. For the previously defined experiments we have:

$$\lim_{n\to\infty} \Delta\left(E_n, F_n^{\#}\right) = 0. \tag{3.1}$$

Lemma 3.2 *In $F_n^{\#}$ there exists an estimator $\check{h}_n$, depending only on the observations $(y(s,t))_{(s,t)\in[0,1]^2}$ such that*

$$\inf_{h\in\Sigma} P_{n,h}(\check{h}_n \in \Sigma_n(h)) \longrightarrow 1$$

as $n \to \infty$. We may assume again that $\check{h}_n$ takes only finitely many values in the set Σ.

Remark 3.3 Lemmas 3.1 and 3.2 can be proved via standard wavelet estimators. Details of the proof of these lemmas can be found in [8, p. 51]. As already mentioned, the diameter of the localized parameter space equals the minimax rate of convergence (up to some logarithmic factor) for the estimation of a first order derivative of a Hölder continuous function on the unit square with smoothness $\alpha = 3$ (which is exactly the lower bound of the smoothness for which our result (Theorem 1.1) holds). In general, this rate is (up to logarithmic factors)

$$n^{-(\alpha/2-d/2)/(\alpha+1)}$$

where α is the smoothness of the functions and d is the order of the derivative that one wants to estimate. Comparing this to the minimax rate for functions on the unit interval we see, that the "effective smoothness" for functions on the unit square is $\alpha/2$. This leads to the conjecture that a sharp bound for the smoothness parameter in Theorem 1.1 is $\alpha > 1$, since for densities on the unit interval $\alpha > 1/2$ is a sharp bound for a similar equivalence result (cp. [12], Theorem 1.1).

With this lemma we can prove Theorem 1.1 similarly to the last theorem. Consider the compound experiment $F_n^{\#\#}$ where one observes

$$\left[(y_1(s,t)), (y_2(s,t))\right]_{(s,t)\in[0,1]^2},$$

where $h \in \Sigma$ and

$$dy_1(s,t) = h^{1/2}(s,t)dsdt + \frac{1}{2}(n - N_n)^{-1/2}dW_1(s,t)$$

$$dy_2(s,t) = h^{1/2}(s,t)dsdt + \frac{1}{2}N_n^{1/2}dW_2(s,t)$$

holds. (W_1 and W_2 are independent Brownian sheets.) Similarly to Equation (3.1) we have

$$\lim_{n\to\infty} \Delta\big(F_n^{\#}, F_n^{\#\#}\big) = 0. \tag{3.2}$$

By applying a sufficiency argument we see that $F_n^{\#\#}$ is equivalent to a model where we observe n i. i. d. stochastic processes, each of which is distributed according to

$$dy(s,t) = h^{1/2}(s,t)dsdt + \frac{1}{2}dW(s,t).$$

By the same sufficiency argument as before, this experiment is equivalent to F_n which finally proves Theorem 1.1.

4 Proofs of Lemma 2.9 and Theorem 2.10

4.1 Coupling of likelihood processes

Remark 4.1 In this section we will often prove existence of absolute, positive constants. By this we mean that these constants depend only on the quantities M, ϵ and α from Theorem 1.1 and thus hold uniformly over the set Σ.

The main tool of the proof of asymptotic equivalence of the previously described doubly localized experiments is the following theorem on a coupling of the Kiefer–Müller process and the sequential empirical process. Its proof is rather long and technical; we refer to the thesis [8] for more details.

Definition 4.2 *For $g \in L^2([0,1],\lambda)$, $\delta > 0$ let*

$$\omega_1^2(g,\delta) := \sup_{0\le|\xi|\le\delta} \int_{[0,1]\cap[0,1]-\xi} (g(t) - g(t+\xi))^2 dt.$$

For $f \in L^2([0,1]^2, \lambda^2)$, $\delta_1, \delta_2 > 0$ let

$$\omega_2^2(f,\delta_1,\delta_2) := \sup_{\substack{0\le|\xi_1|\le\delta_1\\ 0\le|\xi_2|\le\delta_2}} \int_{\substack{u\in[0,1]\cap[0,1]-\xi_1\\ v\in[0,1]\cap[0,1]-\xi_2}} [f(u,v) - f(u+\xi_1,v) - f(u,v+\xi_2) + f(u+\xi_1,v+\xi_2)]^2 dudv$$

$$\omega_3^2(f,\delta_1,\delta_2) := \sup_{\substack{0\le|\xi_1|\le\delta_1\\ 0\le|\xi_2|\le\delta_2}} \sup_{\substack{s\in[0,1]\\ \cap[0,1]-\xi_1}} \int_{v\in[0,1]\cap[0,1]-\xi_2} (f(s,v) - f(s+\xi_1, v+\xi_2))^2 dv.$$

For the next definition we assume that a sum equals zero if its index runs from 0 to −1. (This convention is in force throughout the paper.)

$$
\begin{aligned}
R_M^2(f) \quad := \quad & \sum_{r=1}^{m} \sum_{i=0}^{M_r-1} 2^i 2 \sup_{u\in[0,1]} \omega_1^2(f(u,\cdot), 2^{-i}) \\
& + \sum_{r=1}^{m} \sum_{i=0}^{M_r-1} \sum_{j=0}^{M_r-1} \frac{n}{n_r} 2^{j+i} \omega_2^2\left(f, \frac{n_r}{n} 2^{-(j+1)}, 2^{-(i+1)}\right) \\
& + \sum_{r=1}^{m} 2^{2M_r} \omega_3^2(f, 2^{-M_r}, 2^{-M_r}) \\
& + \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left[f\left(\frac{i}{n}, t\right) - \int_0^1 f\left(\frac{i}{n}, w\right) dw \right]^2 dt
\end{aligned}
$$

where

$$
\begin{aligned}
n_1 &= 2^{k_1} & k_1 &= [\log_2(n)] \\
n_2 &= 2^{k_2} & k_2 &= [\log_2(n - n_1)] \\
& & & \dots \\
n_m &= 2^{k_m} & k_m &= \left[\log_2 \left(n - \textstyle\sum_{i=1}^{m-1} n_i \right) \right]
\end{aligned}
$$

and $n = n_1 + \cdots + n_m$.

For $r \in \{1, \ldots, m\}$ *let*

$$
M_r(y) := \max \left\{ \left[\frac{\log_2 \left(\frac{n_r}{y} \right)}{2} \right], 0 \right\}
$$

and $M = M(y) = (M_1, \ldots, M_m)(y)$.

The following result holds:

Theorem 4.3 *There is a probability space and for all* $n \in \mathbb{N}$ *there exist versions of the processes* $\widehat{G}_n$ *and* K_n *on that space such that for all* $x \geq 0$, $y \geq 0$ *and* $\mathcal{F} \subset L^2([0,1]^2, \lambda^2)$ *where* $\|f\|_\infty \leq 1$ *for all* $f \in \mathcal{F}$ *and* $\#\mathcal{F} < \infty$ *holds, we have*

$$
\begin{aligned}
& P\Big(n^{1/2} \| \widehat{G}_n(f) - n^{-1/2} K_n(f) \|_{\mathcal{F}} \\
& \qquad \geq (\log n)^2 \left(Ax + Bx^{1/2}(y^{1/2} + C)(\log n)^3 R_M(\mathcal{F}) \right) \Big) \\
\leq \quad & D[\#\mathcal{F} \exp(-Gx) + n \exp(-Gy)].
\end{aligned}
$$

(A, B, C, D and G are positive, absolute constants and $R_M(\mathcal{F}) = \|R_M(f)\|_{\mathcal{F}}$ *and* $\|\cdot\|_{\mathcal{F}} = \max_{f\in\mathcal{F}} |\cdot|$.*)*

In this section we will prove Lemma 2.9. Because of Remark 2.8 we will split the experiments $E_{0,n}(h_0)$ and $E_{1,n}(h_0)$ into products of experiments, in which we have less observations than in the original ones.

More precisely, let $k_n := n^{1/2}(\log n)^{-10}$ and let $n_0 \in \mathbb{N}$ be such that $k_{n_0} \geq 1$ holds. For $n \geq n_0$ consider

$$\Delta_k := \left]\frac{k-1}{[k_n]}, \frac{k}{[k_n]}\right], \quad k = 1, \ldots, [k_n]$$

$$n_k := \# A_k \quad where \quad A_k := \{i/n | i/n \in \Delta_k\}.$$

Then it follows $\gamma_n = (n/k_n)^{-1/2}$ and for each $k \in \{1, \ldots, [k_n]\}$ we have $n_k^{-1/2} \geq \frac{1}{\sqrt{2}}\gamma_n$. Consider the doubly localized experiments

Definition 4.4

$$E_{0,k,n}(h_0) : \{y_i | i/n \in A_k\}, \qquad y_i \quad i.n.i.d., \quad y_i \sim h(i/n, \cdot),$$
$$E_{1,k,n}(h_0) : \{y_i(t) | t \in [0,1], i/n \in A_k\},$$

$$dy_i(t) = \left(\lambda_{h,h_0}(i/n, t) - \int_0^1 \lambda_{h,h_0}(i/n, s) ds\right) dt + dW_i(t),$$

$W_1, \ldots, W_n$ are again independent Brownian motions, $k \in \{1, \ldots, [k_n]\}$ and $h \in \Sigma_n(h_0)$.

Let $(\Lambda_{jk,n}(h, h_0))_{h \in \Sigma_n(h_0)}$ $(j = 0, 1)$ be the likelihood processes of these experiments.

Because of the independence structure, we have $\bigotimes_{k=1}^{[k_n]} E_{j,k,n}(h_0) = E_{j,n}(h_0)$, $j = 0, 1$ where for experiments $E_i = (\Omega_i, \mathcal{A}_i, (P_{i,\theta}, \theta \in \Theta))$ $(i = 1, \ldots, k)$, the product is defined as follows:

$$\bigotimes_{i=1}^{k} E_i := \left(\bigotimes_{i=1}^{k} \Omega_i, \bigotimes_{i=1}^{k} \mathcal{A}_i, \left(\bigotimes_{i=1}^{k} P_{i,\theta}, \theta \in \Theta\right)\right).$$

As already mentioned, we will use Theorem 4.3 to construct the likelihood processes of $E_{0,k,n}(h_0)$ and $E_{1,k,n}(h_0)$ on a common probability space. In this way we obtain a construction of the likelihood processes of $E_{0,n}(h_0)$ and $E_{1,n}(h_0)$, since these are the products of the independent processes of the previous experiments.

The coupling has the desired approximation quality so we can estimate the Hellinger distance of $E_{0,n}(h_0)$ and $E_{1,n}(h_0)$ in order to prove Lemma 2.9.

The following lemma is the crucial step of the proof of asymptotic equivalence. The Hungarian construction (i. e. Theorem 4.3) will be used here.

Lemma 4.5 *There exists a constant $K > 0$ and on the probability space of Theorem 4.3 there exist versions $\Lambda^*_{jk,n}$ of the likelihood processes of experiments $E_{j,k,n}(h_0)$ such that for all $n \in \mathbb{N}$ and all $k \in \{1, \ldots, [k_n]\}$ we have*

$$\sup_{h_0 \in \Sigma} \sup_{h \in \Sigma_n(h_0)} H^2\left(\Lambda^*_{0,k,n}(h, h_0), \Lambda^*_{1,k,n}(h, h_0)\right) \leq K(n_k)^{-1}(\log n)^{18}.$$

Proof: With the mapping

$$\begin{array}{rcc} a_k: \quad \Delta_k & \longmapsto & [0,1] \\ t & \longmapsto & \frac{\sum_{j=1}^{k-1} n_j}{n} + t\frac{n_k}{n} \end{array}$$

and the definition $\lambda_{h,h_0,k}(s,t) := \lambda_{h,h_0}(a_k(s),t)$ we can write the likelihood processes of experiments $E_{0,k,n}(h_0)$ and $E_{1,k,n}(h_0)$ as follows (cp. Equations (2.1) and (2.2)):

$$\Lambda_{0,k,n}(h_0) = \exp\left(n_k^{1/2}\widehat{G}_{n_k}(\lambda_{h,h_0,k}) + \sum_{i=1}^{n_k}\int_0^1 \lambda_{h,h_0,k}(i/n_k,t)dt\right) \tag{4.1}$$

and

$$\Lambda_{1,k,n}(h_0) = \exp\left(K_{n_k}(\lambda_{h,h_0,k}) - \frac{1}{2}\sum_{i=1}^{n_k}\mathrm{Var}(\lambda_{h,h_0,k}(i/n_k,z_i))\right) \tag{4.2}$$

where the z_is are i. i. d. random variables on the unit interval.

For the construction of these processes we now use our coupling result. Therefore, we set in Theorem 4.3

$$\mathcal{F} := \left\{\frac{\lambda_{h,h_0,k}}{\|\lambda_{h,h_0,k}\|_\infty}\right\}.$$

The first task is to estimate the term $R_M(\lambda_{h,h_0,k})$ (cp. Definition 4.2).

The following lemma holds for functions $\lambda_{h,h_0,k}$ as well as λ_{h,h_0}. Indeed they differ only by the linear transformation a_k and, therefore, they have the same smoothness properties. Thus we write λ^*_{h,h_0} meaning $\lambda_{h,h_0,k}$ as well as λ_{h,h_0}. The proof follows easily from the smoothness of the functions.

Lemma 4.6 *There is a constant $K_1 > 0$ such that*

$$\sup_{h_0\in\Sigma}\sup_{h\in\Sigma_n(h_0)} R_M^2(\lambda^*_{h,h_0}) \le K_1\gamma_n^2(\log n)^3.$$

Because of $n_k^{-1/2} \ge \frac{\gamma_n}{\sqrt{2}}$ we can increase the diameter of the localized parameter space to $\widetilde{\gamma}_n := \sqrt{2}n^{-1/2}$. By substituting n_k by n we end up with the experiments $\widetilde{E}_{0,n}(h_0)$ and $\widetilde{E}_{1,n}(h_0)$ where the same objects as in $E_{0,n}(h_0)$ and $E_{1,n}(h_0)$ respectively are observed. A bound for the Hellinger distance of the likelihood processes of these experiments is obviously (in view of the enlargement of the parameter space and the substitution) also a bound for the likelihood processes of $E_{0,k,n}(h_0)$ and $E_{1,k,n}(h_0)$ for all k. Therefore, in order to prove Lemma 4.5 we show that for versions $\widetilde{\Lambda}_{0,n}(h,h_0)$ and $\widetilde{\Lambda}_{1,n}(h,h_0)$ of the likelihood processes of experiments $\widetilde{E}_{0,n}(h_0)$ and $\widetilde{E}_{1,n}(h_0)$ it holds

$$\sup_{h_0\in\Sigma}\sup_{h\in\widetilde{\Sigma}_n(h_0)} H^2(\widetilde{\Lambda}_{0,n}(h,h_0),\widetilde{\Lambda}_{1,n}(h,h_0)) \le Kn^{-1}(\log n)^{18} \quad \forall n\in\mathbb{N} \tag{4.3}$$

for a constant $K > 0$ and for all $n \in \mathbb{N}$. Obviously, Lemma 4.5 follows directly from this inequality.

Proof of Inequality (4.3): For simplicity of notation, we write λ instead of $\lambda_{h_0,h}$. From Theorem 4.3 with the choice $\mathcal{F} = \left\{\frac{\lambda}{\|\lambda\|_\infty}\right\}$ we get for all $x, y \geq 0$

$$\begin{aligned}
&D[\exp(-Gx) + n\exp(-Gy)] \\
&\quad\geq \; P\Big(n^{1/2}\big|\widehat{G}_n\big(\tfrac{\lambda}{\|\lambda\|_\infty}\big) - n^{-1/2}K_n\big(\tfrac{\lambda}{\|\lambda\|_\infty}\big)\big| \\
&\qquad\qquad \geq (\log n)^2\big[Ax + Bx^{1/2}(y^{1/2} + C)(\log n)^3 R_M\big(\tfrac{\lambda}{\|\lambda\|_\infty}\big)\big]\Big) \\
&\quad= \; P\Big(n^{1/2}\big|\widehat{G}_n(\lambda) - n^{-1/2}K_n(\lambda)\big| \\
&\qquad\qquad \geq (\log n)^2\big[Ax\|\lambda\|_\infty + Bx^{1/2}(y^{1/2} + C)(\log n)^3 R_M(\lambda)\big]\Big) \\
&\quad\geq \; P\Big(n^{1/2}\big|\widehat{G}_n(\lambda) - n^{-1/2}K_n(\lambda)\big| \\
&\qquad\qquad \geq (\log n)^2\big[Ax\widetilde{\gamma}_n + Bx^{1/2}(y^{1/2} + C)(\log n)^5 D\widetilde{\gamma}_n\big]\Big).
\end{aligned}$$

(Because $R_M(\lambda) \leq D\widetilde{\gamma}_n(\log n)^{3/2}$ and $\|\lambda\|_\infty \leq D\widetilde{\gamma}_n$.)

Now we choose

$$y = G^{-1}(x + \log n), \tag{4.4}$$

thus we get i) $\exp(-Gy) = \exp(-x)n^{-1}$ and ii) $x^{1/2}y^{1/2} \leq Ex\log n$ (for an absolute constant $E > 0$ and only for $x \geq 1$ and $n \geq 3$).

For all $x \geq 1$, $n \geq 3$ and $G_0 := \min\{1, G\}$ we therefore have

$$2D\exp(-G_0x) \geq P\left(n^{1/2}\big|\widehat{G}_n(\lambda) - n^{-1/2}K_n(\lambda)\big| \geq (\log n)^8 Fx\widetilde{\gamma}_n\right) \tag{4.5}$$

(where $F = A + BED + BCD$). Recall the likelihood processes of experiments $\widetilde{E}_{0,n}(h_0)$ and $\widetilde{E}_{1,n}(h_0)$:

$$\widetilde{\Lambda}_{0,n}(h, h_0) = \exp\left(n^{1/2}\widehat{G}_n(\lambda) + \sum_{i=1}^n \int_0^1 \lambda(i/n, t)dt\right)$$

$$\widetilde{\Lambda}_{1,n}(h, h_0) = \exp\left(K_n(\lambda) - \frac{1}{2}\sum_{i=1}^n \mathrm{Var}(\lambda(i/n, z_i))\right).$$

(Where the z_is are i. i. d. random variables on the unit interval.) Furthermore, we consider the following process

$$\widetilde{\Lambda}_{\#,n}(h, h_0) := \exp\left(K_n(\lambda) + \sum_{i=1}^n \int_0^1 \lambda(i/n, t)dt\right).$$

Although this is not a likelihood process of any experiment (as its expectation is not equal to 1), we can compute its Hellinger distance to the previous processes. We will proceed as follows: There exists a constant $K_1 > 0$, such that uniformly over h and h_0 we have

$$H^2(\widetilde{\Lambda}_{0,n}(h, h_0), \widetilde{\Lambda}_{\#,n}(h, h_0)) \leq K_1 n^{-1}(\log n)^{18} \quad \forall n \in \mathbb{N}. \tag{4.6}$$

As we also have $H(\widetilde{\Lambda}_{\#,n}, \widetilde{\Lambda}_{1,n}) \leq H(\widetilde{\Lambda}_{\#,n}, \widetilde{\Lambda}_{0,n})$, inequality (4.3) follows immediately from the triangle inequality. □

Consider the space $L^2(\Omega, \mathcal{A}, P)$ of real-valued random variables. Let $H(\widetilde{\Lambda}_{\#,n}, \widetilde{\Lambda}_{1,n})$ be the distance between $(\widetilde{\Lambda}_{\#,n})^{1/2}$ and $(\widetilde{\Lambda}_{1,n})^{1/2}$ in that space.

Because of

$$(\widetilde{\Lambda}_{1,n})^{1/2} = (\widetilde{\Lambda}_{\#,n})^{1/2}(E\widetilde{\Lambda}_{\#,n})^{-1/2},$$

$\widetilde{\Lambda}_{1,n}$ is the orthogonal projection of $\widetilde{\Lambda}_{\#,n}$ on the unit sphere. Therefore, $(\widetilde{\Lambda}_{1,n})^{1/2}$ is the element on the unit sphere that has the smallest distance to $(\widetilde{\Lambda}_{\#,n})^{1/2}$ and thus we get

$$H(\widetilde{\Lambda}_{\#,n}, \widetilde{\Lambda}_{1,n}) \leq H(\widetilde{\Lambda}_{\#,n}, \widetilde{\Lambda}_{0,n}).$$

Proof of Inequality (4.6): Let $u_n = 2(\log n)^9\left(\frac{F}{G_0}\widetilde{\gamma}_n\right)$ (with the constants from Inequality (4.5)). We define the event

$$A := \left\{\omega : \left|n^{1/2}\widehat{G}_n(\lambda) - K_n(\lambda)\right| < u_n\right\}$$

and we split the expectation

$$\begin{aligned} & H^2(\widetilde{\Lambda}_{0,n}, \widetilde{\Lambda}_{\#,n}) \\ = \; & E_P\left([\widetilde{\Lambda}_{0,n}]^{1/2} - [\widetilde{\Lambda}_{\#,n}]^{1/2}\right)^2 \\ = \; & E_P 1_A\left([\widetilde{\Lambda}_{0,n}]^{1/2} - [\widetilde{\Lambda}_{\#,n}]^{1/2}\right)^2 + E_P 1_{A^C}\left([\widetilde{\Lambda}_{0,n}]^{1/2} - [\widetilde{\Lambda}_{\#,n}]^{1/2}\right)^2 \\ =: \; & I_1 + I_2. \end{aligned}$$

(From now on, we will omit the parameters h and h_0 in the notation.)

Estimation of I_1: By a change of measure, we get

$$I_1 = E_P 1_A\left(\left(\frac{\widetilde{\Lambda}_{\#,n}}{\widetilde{\Lambda}_{0,n}}\right)^{1/2} - 1\right)^2 \widetilde{\Lambda}_{0,n} = E_{P_{0,n}} 1_A\left(\left(\frac{\widetilde{\Lambda}_{\#,n}}{\widetilde{\Lambda}_{0,n}}\right)^{1/2} - 1\right)^2$$

(where $dP_{0,n} = \widetilde{\Lambda}_{0,n} dP$). Now we have

$$\left(\frac{\widetilde{\Lambda}_{\#,n}}{\widetilde{\Lambda}_{0,n}}\right)^{1/2} - 1 = \exp\left(\frac{K_n(\lambda) - n^{1/2}\widehat{G}_n(\lambda)}{2}\right) - 1.$$

On the event A it holds

$$\left|n^{1/2}\widehat{G}_n(\lambda) - K_n(\lambda)\right| < u_n.$$

For all $n \in \mathbb{N}$ we have therefore on the event A

$$\left(\left(\frac{\widetilde{\Lambda}_{\#,n}}{\widetilde{\Lambda}_{0,n}}\right)^{1/2} - 1\right)^2 \leq C_1 u_n^2,$$

and thus

$$I_1 \leq C_1 u_n^2 = K_0(\widetilde{\gamma}_n)^2(\log n)^{18} = 2K_0 n^{-1}(\log n)^{18}.$$

Estimation of I_2:

$$\begin{aligned} I_2 &= E1_{A^C}\big([\widetilde{\Lambda}_{0,n}]^{1/2} - [\widetilde{\Lambda}_{\#,n}]^{1/2}\big)^2 \le E1_{A^C}(\widetilde{\Lambda}_{0,n} + \widetilde{\Lambda}_{\#,n}) \\ &\le \big(P(A^C)E(\widetilde{\Lambda}_{0,n})^2\big)^{1/2} + \big(P(A^C)E(\widetilde{\Lambda}_{\#,n})^2\big)^{1/2} \end{aligned}$$

(from the Cauchy–Schwarz inequality). There is a constant $K > 0$ such that

a) $E\big[(\widetilde{\Lambda}_{0,n})^2\big] \le K$

b) $E\big[(\widetilde{\Lambda}_{\#,n})^2\big] \le K.$

a) For $z_i : [i.i.d.] \sim U[0,1]$:

$$E(\widetilde{\Lambda}_{0,n})^2 = E\exp\left(2\sum_{i=1}^{n} \lambda(i/n, z_i)\right) = \prod_{i=1}^{n} E\exp(2\lambda(i/n, z_i)),$$

and

$$E\exp(2\lambda(i/n, z_i)) = \int_0^1 (\varphi(t))^2 dt = 1 + \int_0^1 (\varphi(t) - 1)^2 dt,$$

(for $\varphi(t) = \exp(\lambda(i/n, t))$, since $\int_0^1 \varphi(t)dt = 1$). Because of $|\lambda(i/n, t)| \le K_0\widetilde{\gamma}_n$ we get directly

$$(\varphi(t) - 1)^2 \le K_1(\widetilde{\gamma}_n)^2 \tag{4.7}$$

and thus we get $E(\widetilde{\Lambda}_{0,n})^2 \le \big(1 + \frac{K_1}{n}\big)^n \le 2\exp(K_1) =: K$.

b) Because of $K_n(\lambda) \sim \mathcal{N}(0, \sum_{i=1}^n \mathrm{Var}(\lambda(i/n, z_i)))$ it follows:

$$\begin{aligned} E\exp(2K_n(\lambda)) &= \exp\left(2\sum_{i=1}^{n} \mathrm{Var}(\lambda(i/n, z_i))\right) \\ \Rightarrow \quad E(\widetilde{\Lambda}_{1,n})^2 &= E\exp\left(2\left[K_n(\lambda) - \frac{1}{2}\sum_{i=1}^{n} \mathrm{Var}(\lambda(i/n, z_i))\right]\right) \\ &= \exp\left(\sum_{i=1}^{n} \mathrm{Var}(\lambda(i/n, z_i))\right) = \prod_{i=1}^{n} \exp\left(\mathrm{Var}(\lambda(i/n, z_i))\right). \end{aligned}$$

Furthermore we have $\mathrm{Var}(\lambda(i/n, z_i)) \le \|\lambda\|_\infty^2 \le K_0(\widetilde{\gamma}_n)^2 = K_1 n^{-1}$, such that

$$E(\widetilde{\Lambda}_{1,n})^2 \le \exp\big(K_1 n^{-1}\big)^n = \exp(K_1) =: K_2.$$

From $\widetilde{\Lambda}_{\#,n} = \widetilde{\Lambda}_{1,n} E\widetilde{\Lambda}_{\#,n}$ we get

$$E(\widetilde{\Lambda}_{\#,n})^2 = \underbrace{E(\widetilde{\Lambda}_{1,n})^2}_{\le K_2}(E\widetilde{\Lambda}_{\#,n})^2.$$

The proof of b) follows from

$$\begin{aligned} E\widetilde{\Lambda}_{\#,n} &= E\exp(K_n(\lambda))\exp\left(\sum_{i=1}^{n}\int_0^1 \lambda(i/n,t)dt\right) \\ &\leq \underbrace{\exp\left(\frac{1}{2}\sum_{i=1}^{n}\mathrm{Var}(\lambda(i/n,z_i))\right)}_{\underbrace{\leq K_0}_{\text{(as above)}}} \underbrace{\prod_{i=1}^{n}\int_0^1 \exp(2\lambda(i/n,t)dt))^{1/2}}_{\underbrace{\leq K_1}_{\text{(as in a)}}} \end{aligned}$$

And i) follows easily:

$$I_2 \leq \big(P(A^C)\underbrace{E(\widetilde{\Lambda}_{1,n})^2}_{\leq K}\big)^{1/2} + \big(P(A^C)\underbrace{E(\widetilde{\Lambda}_{\#,n})^2}_{\leq K}\big)^{1/2}.$$

In inequality (4.5) we set $x = x_n = \frac{2\log n}{G_0}$ and we get $P(A^C) \leq 2Dn^{-2}$ and thus, (4.6) follows. □

Note that in inequality (4.5) $n \geq 3$ and $x \geq 1$ are assumed and fulfilled here. Finally, we proved Lemma 4.5. □

Lemma 4.5 leads us directly to Lemma 2.9:

Proof: From Lemma 2.7 it follows that for all $h \in \Sigma$ and $n \in \mathbb{N}$ we have

$$\begin{aligned} &\sup_{h_0\in\Sigma}\sup_{h\in\Sigma_n(h_0)} H^2\big(\Lambda^*_{0,n}(h,h_0),\Lambda^*_{1,n}(h,h_0)\big) \\ &= \sup_{h_0\in\Sigma}\sup_{h\in\Sigma_n(h_0)} H^2\left(\prod_{k=1}^{[k_n]}\Lambda^*_{0,k,n}(h,h_0),\prod_{k=1}^{[k_n]}\Lambda^*_{1,k,n}(h,h_0)\right) \\ &\leq 2\sup_{h_0\in\Sigma}\sup_{h\in\Sigma_n(h_0)}\sum_{k=1}^{[k_n]} H^2\left(\Lambda^*_{0,k,n}(h,h_0),\Lambda^*_{1,k,n}(h,h_0)\right) \\ &\leq K[k_n]n_{\widetilde{k}}^{-1}(\log n)^{18} \end{aligned}$$

(from Lemma 4.5 for a $\widetilde{k} \in \{1,\dots,[k_n]\}$). We had

$$n_k = \#\left\{i/n : i/n \in \left]\frac{k-1}{[k_n]},\frac{k}{[k_n]}\right]\right\}$$

and thus we get for all $k \in \{1,\dots,[k_n]\}$ and n large enough:

$$n_k \geq \frac{1}{2}\frac{n}{[k_n]} \geq \frac{n}{2k_n} = \frac{1}{2}\gamma_n^{-2}.$$

Thus we have $n_k^{-1} \leq 2\gamma_n^2$ and

$$\begin{aligned}\sup_{h_0\in\Sigma}\sup_{h\in\Sigma_n(h_0)} H^2\big(\Lambda_{0,n}^*(h,h_0),\Lambda_{1,n}^*(h,h_0)\big) &\leq 2Kk_n\gamma_n^2(\log n)^{18}\\ = \quad 2Kn\gamma_n^4(\log n)^{18} = 2K(\log n)^{-20}(\log n)^{18} &= 2K(\log n)^{-2} \longrightarrow 0\end{aligned}$$

for $n \to \infty$, which finally proves Lemma 2.9. □

4.2 Further local approximations for the experiment $E_{1,n}(h_0)$

In this section we will prove Theorem 2.10. In addition to the experiments already introduced, let

$$\begin{aligned} E_{2,n}^{\#}(h_0): &\quad dy_i(t) = \big(\tfrac{h}{h_0}-1\big)\big(i/n,\,P_0^{-1}(i/n,t)\big)dt + dW_i(t)\\ E_{3,n}^{\#}(h_0): &\quad dy_i(t) = 2\big(\big(\tfrac{h}{h_0}\big)^{1/2}-1\big)\big(i/n,\,P_0^{-1}(i/n,t)\big)dt + dW_i(t)\\ E_{4,n}^{\#}(h_0): &\quad dy(s,t) = 2\Big[\big(\tfrac{h}{h_0}\big)^{1/2}-1\Big]\big(s,\,P_0^{-1}(s,t)\big)dsdt + n^{-1/2}dW(s,t)\end{aligned}$$

where $t \in [0,1]$, $(s,t) \in [0,1]^2$, $i = 1,\dots,n$; furthermore, $W_1,\dots,W_n$ are independent Brownian motions, W is a two-dimensional Brownian sheet and for each experiment let $h \in \Sigma_n(h_0)$.

Lemma 4.7 *For $i = 2,3$ we have*

$$\Delta\big(E_{i,n}^{\#}(h_0),\,E_{i,n}(h_0)\big) = 0.$$

This can be easily shown as the transformations

$$W_i^*(t) := \int_0^t h_0^{-1/2}(i/n,s)d(W_i\circ P_0(i/n,s))$$

are also independent Brownian motions.

Proof of Theorem 2.10: In view of Lemma 4.7 and the triangle inequality it suffices to prove that uniformly over $h_0 \in \Sigma$ the following holds:

$$\Delta\big(E_{1,n}(h_0),\,E_{2,n}^{\#}(h_0)\big) \longrightarrow 0 \qquad (4.8)$$

$$\Delta\big(E_{1,n}(h_0),\,E_{3,n}^{\#}(h_0)\big) \longrightarrow 0 \qquad (4.9)$$

$$\Delta(F_n(h_0),\,E_{4,n}(h_0)) \longrightarrow 0 \qquad (4.10)$$

$$\Delta(E_{3,n}(h_0),\,F_n(h_0)) \longrightarrow 0 \qquad (4.11)$$

for $n \to \infty$.

Proof of Relation (4.8): In the sequel we omit the subscript of the function λ_{h,h_0}. With Facts 2.3 and 2.5 and Lemma 2.7 we have

$$\begin{aligned}
&\Delta\big(E_{1,n}(h_0), E^{\#}_{2,n}(h_0)\big) \\
&\quad \le \sup_{h\in\Sigma_n(h_0)} H^2\big(\Lambda_{1,n}(h,h_0), \Lambda^{\#}_{2,n}(h,h_0)\big) \\
&\quad \le \sup_{h\in\Sigma_n(h_0)} 2\sum_{i=1}^{n} 2\Bigg(1-\exp\Bigg[-\frac{1}{8}\Bigg\|\lambda(i/n,\cdot)-\int_0^1 \lambda(i/n,s)ds \\
&\qquad\qquad -\left(\frac{h}{h_0}-1\right)\big(i/n, P_0^{-1}(i/n,\cdot)\big)\Bigg\|_2^2\Bigg]\Bigg).
\end{aligned}$$

Consider the Taylor expansion of the logarithm

$$\log(1+(x-1)) = x-1-\frac{1}{2}(\theta(x-1))^2$$

for some $\theta \in [0,1]$. For $x := \frac{h}{h_0}\big(i/n, P_0^{-1}(i/n,t)\big)$, we get

$$\begin{aligned}
\lambda(i/n,t) &= \log\frac{h}{h_0}\big(i/n, P_0^{-1}(i/n,t)\big) \\
&= \left(\frac{h}{h_0}-1\right)\big(i/n, P_0^{-1}(i/n,t)\big) - \frac{1}{2}\left(\theta\left[\frac{h}{h_0}-1\right]\big(i/n, P_0^{-1}(i/n,t)\big)\right)^2
\end{aligned}$$

hence

$$\begin{aligned}
&\left\|\lambda(i/n,\cdot)-\int_0^1 \lambda(i/n,s)ds-\left(\frac{h}{h_0}-1\right)\big(i/n, P_0^{-1}(i/n,\cdot)\big)\right\|_2 \\
&\quad \le \underbrace{\left\|\lambda(i/n,\cdot)-\left(\frac{h}{h_0}-1\right)\big(i/n, P_0^{-1}(i/n,\cdot)\big)\right\|_2}_{=:A} + \underbrace{\left|\int_0^1 \lambda(i/n,s)ds\right|}_{=:B}.
\end{aligned}$$

$$\begin{aligned}
A^2 &= \int_0^1 \left[\lambda(i/n,t)-\left(\frac{h}{h_0}-1\right)\big(i/n, P_0^{-1}(i/n,t)\big)\right]^2 dt \\
&= \int_0^1 \left(\frac{1}{2}\left[\theta_t\left(\frac{h}{h_0}-1\right)\big(i/n, P_0^{-1}(i/n,t)\big)\right]^2\right)^2 dt \\
&\qquad\qquad (\text{where } \theta_t \in [0,1] \text{ for all } t\in[0,1]) \\
&\le \frac{1}{4}\int_0^1 \underbrace{\left(\frac{h}{h_0}-1\right)^4}_{=(h-h_0)^4/h_0^4}\big(i/n, P_0^{-1}(i/n,t)\big)dt \le K\gamma_n^4
\end{aligned}$$

(since $h_0 \ge \epsilon$ for all $h_0 \in \Sigma$). For proving $B \le K\gamma_n^2$ we set $\varphi(t) := \frac{h}{h_0}\big(i/n, P_0^{-1}\ (i/n,t)\big)$. We get $\lambda(i/n,t) = \log(\varphi(t))$ and $\int_0^1 \varphi(t)dt = 1$. Since due to inequality (4.7) there exists a constant $K_0 > 0$, such that $|\varphi(t)-1| \le K_0\gamma_n$ holds, we get from the Taylor expansion (for $\theta_t \in [0,1]$)

$$\log(\varphi(t)) = \varphi(t)-1-\frac{1}{2}(\theta_t(\varphi(t)-1))^2.$$

Thus

$$\left|\int_0^1 \log(\varphi(t))dt\right| = \left|\underbrace{\int_0^1 (\varphi(t)-1)dt}_{=0} - \frac{1}{2}\int_0^1 (\theta_t(\varphi(t)-1))^2 dt\right| \le K\gamma_n^2.$$

Now, the proof of (4.8) is easy:

$$\Delta\big(E_{1,n}(h_0), E^{\#}_{2,n}(h_0)\big) \le 4n\left(1-\exp\left(-\frac{K\gamma_n^4}{8}\right)\right) \le K_1 n\gamma_n^4 = K_1(\log n)^{-20} \longrightarrow 0$$

(by the Taylor expansion of $\exp(x)$ and for $n \to \infty$). □

Proof of Relation (4.9): Again, according to Facts 2.3, 2.5 and Lemma 2.7 we have to show that

$$\left\|\lambda(i/n,\cdot) - 2\left[\left(\frac{h}{h_0}\right)^{1/2} - 1\right]\big(i/n, P_0^{-1}(i/n,\cdot)\big)\right\|_2^2 \le K\gamma_n^4 \tag{4.12}$$

holds.

Again, we use the Taylor expansion of the logarithm and get

$$\begin{aligned}
&\lambda_{h,h_0}(i/n,t)\\
&= 2\log\left[\left(\frac{h}{h_0}\right)^{1/2}\big(i/n, P_0^{-1}(i/n,t)\big)\right]\\
&= 2\left[\left(\frac{h}{h_0}\right)^{1/2} - 1\right]\big(i/n, P_0^{-1}(i/n,t)\big) - \theta^2\left[\left(\frac{h}{h_0}\right)^{1/2} - 1\right]^2\big(i/n, P_0^{-1}(i/n,t)\big)
\end{aligned}$$

for some $\theta \in [0,1]$. Because of $h, h_0 \ge \epsilon$ we have $\left|\left(\frac{h}{h_0}\right)^{1/2} - 1^{1/2}\right| \le \frac{1}{\sqrt{\epsilon}}\left|\frac{h}{h_0} - 1\right|$ and thus

$$\begin{aligned}
&\left\|\lambda(i/n,\cdot) - 2\left(\frac{h}{h_0}\right)^{1/2}\big(i/n, P_0^{-1}(i/n,\cdot)\big)\right\|_2^2\\
&\le \left\|\left(\left(\frac{h}{h_0}\right)^{1/2} - 1\right)^2\big(i/n, P_0^{-1}(i/n,\cdot)\big)\right\|_2^2\\
&\le \frac{1}{\epsilon}\left\|\left(\frac{h}{h_0} - 1\right)^2\big(i/n, P_0^{-1}(i/n,\cdot)\big)\right\|_2^2 \le K_0\gamma_n^4
\end{aligned}$$

which proves (4.9). □

Proof of Relation (4.10): Recall that

$$E_{4,n}(h_0): \qquad dy(s,t) = \log\big(h\big(s, P_0^{-1}(s,t)\big)\big)dsdt + n^{-1/2}dW(s,t)$$

where $(s,t) \in [0,1]^2$ and $h \in \Sigma_n(h_0)$. We proceed as before, i. e. let

$$E^{\#}_{4,n}: \qquad dy(s,t) = 2\left[\left(\frac{h}{h_0}\right)^{1/2} - 1\right]\left(s, P_0^{-1}(s,t)\right)dsdt + n^{-1/2}dW(s,t).$$

According to Fact 2.6, in order to prove

$$\sup_{h_0\in\Sigma}\lim_{n\to\infty}\Delta\left(E_{4,n}(h_0), E^{\#}_{4,n}(h_0)\right) = 0,$$

it suffices to show that

$$n\left\| \log\left(\frac{h}{h_0}\left(s, P_0^{-1}(s,t)\right)\right) - 2\left[\left(\frac{h}{h_0}\right)^{1/2} - 1\right]\left(s, P_0^{-1}(s,t)\right)\right\|_2^2 \longrightarrow 0$$

holds uniformly over $h_0 \in \Sigma$ and for $n \to \infty$. The proof is exactly the same as the one for Inequality (4.12).

For the experiment

$$\widetilde{F}_n(h_0): \qquad dy(s,t) = \left(h^{1/2} - h_0^{1/2}\right)(s,t)dsdt + \frac{1}{2}n^{-1/2}dW(s,t)$$

we obviously have $\Delta(F_n(h_0), \widetilde{F}_n(h_0)) = 0$.

As

$$W^*(s,t) := \int_0^s\int_0^t h_0^{-1/2}(u,v)dW(u, P_0(u,v))$$

is again a Brownian sheet, the likelihood processes of $E^{\#}_{4,n}(h_0)$ and $\widetilde{F}_n(h,h_0)$ have the same distribution which proves (4.10). □

Proof of Relation (4.11): It remains to show that $E_{3,n}(h_0)$ and $F_n(h_0)$ are asymptotically equivalent uniformly over $h_0 \in \Sigma$. This result holds even globally and thus we omit the localizing notation (h_0) for all experiments. The idea is to discretize experiment $F_n(h_0)$ and it was first applied by Brown and Low (cp. [1]):

Consider another experiment

$$\bar{F}_n: \quad dy_n(s,t) = h_n^{1/2}(s,t)dsdt + \frac{1}{2}dW(s,t)$$

where

$$h_n(s,t) := h(i/n, t) \quad \text{for} \quad s \in \left]\frac{i-1}{n}, \frac{i}{n}\right].$$

Then we have $\lim_{n\to\infty}\Delta(F_n, \bar{F}_n) = 0$ as can be shown by computing the L^2 distance of h and h_n. As

$$T(y) := (n[y(1/n,t) - y(0,t)]_{t\in[0,1]}, \ldots, n[y(n/n,t) - y(n-1/n,t)]_{t\in[0,1]})$$

is a sufficient statistic and as

$$T\bar{F}_n = \bar{E}_{3,n}: \quad dy_i(t) = h^{1/2}\left(\frac{i}{n}, t\right)dt + \frac{1}{2}dW_i(t)$$

we have shown (4.11). □

Now the proof of Theorem 2.10 is finished. □

References

[1] Brown, L. D. and Low, M. Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, 24:2384–2398, 1996

[2] Brown, L. D., Cai, T. T., Low, M. and Zhang, C.-H. Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.*, 30:688–707, 2002

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

[4] Carter, A. V. Deficiency distance between multinomial and multivariate normal experiments. *Ann. Statist.*, 30:708–730, 2002

[5] Grama, I. and Nussbaum, M. Asymptotic equivalence for nonparametric generalized linear models. *Probab. Theory Related Fields*, 111(2):167–214, 1996

[6] Grama, I. and Nussbaum, M. Asymptotic equivalence for nonparametric regression. *Math. Meth. Statist.*, 11(1):1–36, 2002

[7] Grama, I. and Nussbaum, M. A functional Hungarian construction for sums of independent random variables. *Ann. Inst. H. Poincaré, Probab. Statist.*, 38(6):923–957, 2002

[8] Jähnisch, M. *Asymptotic Equivalence for a Model of Independent Non Identically Distributed Data.* Doctoral Thesis, Humboldt University Berlin, 1999

[9] Koltchinskii, V. I. Komlós–Major–Tusnády approximation for the general empirical process and Haar expansions of classes of functions. *J. Theoret. Probab.*, 7:73–118, 1994

[10] Le Cam, L. *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York-Berlin, 1986

[11] Millar, P. W. Optimal estimation of a general regression function. *Ann. Statist.*, 10(3):717–740, 1982

[12] Nussbaum, M. Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.*, 24:2399–2430, 1996

[13] Strasser, H. *Mathematical Theory of Statistics.* De Gruyter, Berlin, 1985

Michael Jähnisch
SAP AG
Neurottstrasse 16
69190 Walldorf, Germany
michael.jaehnisch@sap.com

Michael Nussbaum
Department of Mathematics
Cornell University
Ithaca, NY 14853-4201, USA
nussbaum@math.cornell.edu