The Annals of Statistics
1996, Vol. 24, No. 6, 2399–2430

ASYMPTOTIC EQUIVALENCE OF DENSITY ESTIMATION AND GAUSSIAN WHITE NOISE

BY MICHAEL NUSSBAUM

Weierstrass Institute, Berlin

Signal recovery in Gaussian white noise with variance tending to zero has served for some time as a representative model for nonparametric curve estimation, having all the essential traits in a pure form. The equivalence has mostly been stated informally, but an approximation in the sense of Le Cam's deficiency distance Δ would make it precise. The models are then asymptotically equivalent for all purposes of statistical decision with bounded loss. In nonparametrics, a first result of this kind has recently been established for Gaussian regression. We consider the analogous problem for the experiment given by n i.i.d. observations having density f on the unit interval. Our basic result concerns the parameter space of densities which are in a Hölder ball with exponent $\alpha > \frac{1}{2}$ and which are uniformly bounded away from zero. We show that an i. i. d. sample of size n with density f is globally asymptotically equivalent to a white noise experiment with drift $f^{1/2}$ and variance $\frac{1}{4}n^{-1}$. This represents a nonparametric analog of Le Cam's heteroscedastic Gaussian approximation in the finite dimensional case. The proof utilizes empirical process techniques related to the Hungarian construction. White noise models on f and $\log f$ are also considered, allowing for various "automatic" asymptotic risk bounds in the i.i.d. model from white noise.

1. Introduction and main result. One of the basic principles of Le Cam's (1986) asymptotic decision theory is to approximate general experiments by simple ones. In particular, *weak convergence to Gaussian shift experiments* has now become a standard tool for establishing asymptotic risk bounds. The risk bounds implied by weak convergence are generally estimates from below, and in most of the literature the efficiency of procedures is more or less shown on an ad hoc basis. However, a systematic approach to the attainment problem is also made possible by Le Cam's theory, based on the notion of *strong convergence of experiments*, which means proximity in the sense of the full deficiency distance. However, due to the inherent technical difficulties of handling the deficiency concept, this possibility is rarely used, even in root-n consistent parametric problems.

In nonparametric curve estimation models of the "ill posed" class where there is no root-n consistency, research has focused for a long time on optimal rates of convergence. In these problems, limits of experiments for $n^{-1/2}$-localized parameter are not directly useful for risk bounds. But now a theory of *exact asymptotic risk constants* is also developing in the context of slower

Received March 1993; revised April 1996.
AMS 1991 *subject classifications*. Primary 62G07; secondary 62B15, 62G20.
Key words and phrases. Nonparametric experiments, deficiency distance, likelihood process, Hungarian construction, asymptotic minimax risk, curve estimation.

rates of convergence. Such an exact risk bound was first discovered by Pinsker (1980) in the problem of *signal recovery in Gaussian white noise*, which is by now recognized as the basic or "typical" nonparametric curve estimation problem. The cognitive value of this model had already been realized by Ibragimov and Khasminski (1977). These risk bounds have been established since then in a variety of other problems, for example, density, nonparametric regression and spectral density [see Efroimovich and Pinsker (1982), Golubev (1984) and Nussbaum (1985)], and they have also been substantially extended conceptually [(Korostelev (1993) and Donoho and Johnstone (1992)]. The theory is now at a stage where the approximation of the various particular curve estimation problems by the white noise model could be made formal. An important step in this direction has been made by Brown and Low (1996) by relating Gaussian regression to the signal recovery problem. These models are essentially the continuous and discrete versions of each other. The aim of this paper is to establish the *formal approximation by the white noise model* for the problem of density estimation from an i.i.d. sample.

To formulate our main result, define a basic parameter space Σ of densities as follows. For $\alpha \in (0,1]$ and $M > 0$, let

$$\Lambda^\alpha(M) = \left\{f\colon\ \left|f(x) - f(y)\right| \le M\ |x-y|^\alpha,\ x, y \in [0,1]\right\}$$

be a Hölder ball of functions with exponent α. For $\varepsilon > 0$ define a set $\mathscr{F}_{\ge\varepsilon}$ as the set of densities on $[0,1]$ bounded below by ε:

$$\mathscr{F}_{\ge\varepsilon} = \left\{f\colon\ \int_0^1 f = 1,\ f(x) \ge \varepsilon, x \in [0,1]\right\}. \tag{1}$$

Define an a priori set, for given $\alpha \in (0,1]$, $M > 0$, $\varepsilon > 0$,

$$\Sigma_{\alpha, M, \varepsilon} = \Lambda^\alpha(M) \cap \mathscr{F}_{\ge\varepsilon}. \tag{2}$$

Let Δ be Le Cam's deficiency pseudodistance between experiments having the same parameter space. For the convenience of the reader a formal definition is given in Section 9. For two sequences of experiments $\mathbb{E}_n$ and $\mathbb{F}_n$ we shall say that they are *asymptotically equivalent* if $\Delta(\mathbb{E}_n, \mathbb{F}_n) \to 0$ as $n \to \infty$. Let dW denote the standard Gaussian white noise process on the unit interval.

THEOREM 1.1. *Let Σ be a set of densities contained in $\Sigma_{\alpha, M, \varepsilon}$ for some $\varepsilon > 0$, $M > 0$ and $\alpha > \frac{1}{2}$. Then the experiments given by observations*

$$y_i, \qquad i = 1, \dots, n,\ \textit{i.i.d. with density } f, \tag{3}$$

$$dy(t) = f^{1/2}(t)\, dt + \tfrac{1}{2} n^{-1/2}\, dW(t), \qquad t \in [0,1], \tag{4}$$

with $f \in \Sigma$ are asymptotically equivalent.

This result is closely related to Le Cam's global asymptotic normality for parametric models. In the i.i.d. model let f be in a parametric family $(f_\vartheta,\ \vartheta \in \Theta)$, where $\Theta \subset \mathbb{R}^k$, which is sufficiently regular and has Fisher information matrix $I(\vartheta)$ at point ϑ. Then the i.i.d. model may be approximated by a heteroscedastic Gaussian experiment

$$y = \vartheta + n^{-1/2} I(\vartheta)^{-1/2} \eta, \tag{5}$$

where η is a standard normal vector and $\vartheta \in \Theta$. We see that (4) is a nonparametric analog of (5) when ϑ is identified with $f^{1/2}$. Indeed, consider the identity for the Fisher information matrix in the parametric case

$$\left\| f_{\vartheta'}^{1/2} - f_{\vartheta}^{1/2} \right\|_2 = \tfrac{1}{2} \left\| I^{1/2}(\vartheta)(\vartheta' - \vartheta) \right\| + o\left(\|\vartheta' - \vartheta\| \right).$$

Formally regarding $f^{1/2}$ itself as a parameter, we find the corresponding Fisher information to be 4 times the unit operator. However, even for parametric families (4) seems to be an interesting form of a global approximation: if $f_{\vartheta}^{1/2}$ is taken as parameter, then the resulting Gaussian model has a simple form. One recognizes that the heteroscedastic nature of (5) derives only from the "curved" nature of a general parametric family within the space of roots of densities.

This observation was in fact made earlier by Le Cam (1985). In his Theorem 4.3 he established the homoscedastic global Gaussian approximation for i.i.d. models in the *finite dimensional case*. We give a paraphrase of that result in a specialized form. A set Θ' in $L_2(0,1)$ is said to have finite metric dimension if there is a number D such that every subset of Θ' which can be covered by an ε-ball can be covered by no more than 2^D balls of radius $\varepsilon/2$, where D does not depend on ε. A set of densities f has this property in Hellinger metric if the corresponding set of $f^{1/2}$ has it in $L_2(0,1)$.

PROPOSITION 1.2 [Le Cam (1985)]. *Let Σ be a set of densities on* $[0,1]$ *having finite dimension in Hellinger metric and fulfilling a further regularity condition* (*see Section* 10). *Then the experiments given by observations* (3) *and* (4) *with $f \in \Sigma$ are asymptotically equivalent.*

The actual formulation in Le Cam (1985) is more abstract and general, giving a global asymptotic normality in the i.i.d. case for arbitrary random variables, in particular without assumed existence of densities; but finite dimensionality is essential. This result in its conceptual clarity and potential impact seems not to have been well appreciated by researchers; the heteroscedastic form (5) under classical regularity conditions is somewhat better known [cf. Mammen (1986)].

Our main result can thus be viewed as an extension of Le Cam's Proposition 1.2 to a nonparametric setting. The value $\frac{1}{2}$ of the Hölder exponent α seems to be a critical one; for discretization of Gaussian white noise, this has been shown by Brown and Low (1996).

White noise models with *fixed* variance do occur as local limits of experiments in root-n consistent nonparametric problems [Millar (1979)], and, via specific renormalizations, also in non-root-n consistent curve estimation [Low

(1992) and Donoho and Low (1992)]. Thus various central limit theorems for i.i.d. experiments can be embedded in a relatively simple and closed-form approximation by (4). Moreover, for the density f itself and for $\log f$ we also give Gaussian approximations which are "heteroscedastic" in analogy to (5); see Remark 2.8 and Corollary 3.3.

The paper is organized as follows. The basic results are developed in an overview fashion in Sections 2 and 3, which may suffice for a first reading. By default, proofs or technical comments for all statements are to be found in Sections 4–10.

In Section 2 we develop the basic approximation of likelihood ratios over shrinking neighborhoods of a given density f_0. These neighborhoods $\Sigma_n(f_0)$ are already "nonparametric," in the sense of shrinking slower than $n^{-1/2}$. For proving this, we partition the sample space $[0, 1]$ into small intervals and obtain a product experiment structure via Poissonization. The Gaussian approximation is then argued via the "space-local" empirical process on the small intervals; piecing this together on $[0, 1]$ yields the basic parameter-local Gaussian approximation over $f \in \Sigma_n(f_0)$. Once in a Gaussian framework, we manipulate likelihood ratios to obtain other approximations, in particular the one with trend $f^{1/2}$. For these experiments, which are all Gaussian, we use the methodology of Brown and Low (1996), who compared the white noise model with its discrete version (the Gaussian regression model).

It remains to piece together the parameter-local approximations using a preliminary estimator; this is the subject of Section 3. Our method of globalization is somewhat different from Le Cam's, which works in the parametric case; the concept of metric entropy or dimension and related theory are not utilized. But obviously these methods, which already have proved fruitful in nonparametrics, have a potential application here as well.

Our results are nonconstructive in spirit; that is, they are estimates of the Δ-distance which imply asymptotic risk bounds. The question of a constructive recipe for procedures, as obtained by Brown and Low (1996), is more complex to treat in the present case. Nevertheless, application to the theory of asymptotic minimax constants in nonparametrics is possible. We do not develop this here, but a possible first exercise would be to derive the result of Efroimovich and Pinsker (1982) on density estimation with L_2-loss over Sobolev classes from Pinsker's (1980) result in the white noise model. Since the deficiency distance refers to loss functions which are uniformly bounded, one would use the version of Pinsker's result for bounded L_2-related loss [cf. Tsybakov (1994)]. For a recent account of exact constants for L_2-loss over ellipsoids see Belitser and Levit (1995).

This application is limited in scope by the fact that our critical smoothness $1/2$ is for Hölder classes, so that for Sobolev classes it comes out as 1. For Gaussian nonparametric regression [Brown and Low (1996)], the limit is actually $1/2$ in the Sobolev (or Besov $B^{1/2}_{2,2}$) sense, but it is not clear if this holds in the i.i.d. model. Another result which can be carried over to density estimation is the L_∞-analog of Pinsker's constant, found by Korostelev (1993). The details for this case, where the Hölder classes are natural, are developed in Korostelev and Nussbaum (1996).

As a basic text for the asymptotic theory of experiments we refer to Strasser (1985). We use C as a generic notation for positive constants; for sequences the symbol $a_n \asymp b_n$ means the usual equivalence in rate, while $a_n \sim b_n$ means $a_n = b_n(1 + o(1))$.

2. The local approximation. Our first Gaussian approximation will be established in a parameter-local framework. Suppose we have i.i.d. observations y_i, $i = 1, \ldots, n$, with distribution P_f having Lebesgue density f on the interval $[0, 1]$, and it is known a priori that f belongs to a set of densities Σ. Henceforth in the paper we will set $\Sigma = \Sigma_{\alpha, M, \varepsilon}$ for some $\varepsilon > 0$, $M > 0$ and $\alpha > 1/2$.

Let $\| \cdot \|_p$ denote the norm in the space $L_p(0, 1)$, $1 \le p \le \infty$. Let γ_n be the sequence

$$\gamma_n = n^{-1/4}(\log n)^{-1}, \tag{6}$$

and for any $f_0 \in \Sigma$ define a class $\Sigma_n(f_0)$ by

$$\Sigma_n(f_0) = \left\{ f \in \Sigma \colon \left\| \frac{f}{f_0} - 1 \right\|_\infty \le \gamma_n \right\}. \tag{7}$$

For given $f_0 \in \Sigma$ we define a local (around f_0) product experiment

$$\mathbb{E}_{0,n}(f_0) = \left([0, 1]^n, \mathscr{B}^n_{[0,1]}, (P_f^{\otimes n},\ f \in \Sigma_n(f_0)) \right). \tag{8}$$

Let F_0 be the distribution function corresponding to f_0, and let

$$K(f_0 \| f) = -\int \log \frac{f}{f_0} dF_0$$

be the Kullback–Leibler relative entropy. Let W be the standard Wiener process on $[0, 1]$, and consider an observed process

$$y(t) = \int_0^t \log \frac{f}{f_0}(F_0^{-1}(u))du + tK(f_0 \| f) + n^{-1/2}W(t), \qquad t \in [0, 1]. \tag{9}$$

Let Q_{n,f,f_0} be the distribution of this process on the function space $C_{[0,1]}$ equipped with its Borel σ-algebra $\mathscr{B}_{C_{[0,1]}}$, and let

$$\mathbb{E}_{1,n}(f_0) = \left(C_{[0,1]}, \mathscr{B}_{C_{[0,1]}}, (Q_{n,f,f_0}, f \in \Sigma_n(f_0)) \right) \tag{10}$$

be the corresponding experiment when f varies in the neighborhood $\Sigma_n(f_0)$.

THEOREM 2.1. *Define $\Sigma_n(f_0)$ as in* (7) *and* (6). *Then*

$$\Delta(\mathbb{E}_{0,n}(f_0), \mathbb{E}_{1,n}(f_0)) \to 0 \quad \text{as } n \to \infty$$

uniformly over $f_0 \in \Sigma$.

The proof is based upon the following principle, described in Le Cam and Yang [(1990), page 16]. Consider two experiments $\mathbb{E}_i = (\Omega_i, \mathscr{A}_i, (P_{i,\vartheta}, \vartheta \in \Theta))$, $i = 0, 1$, having the same parameter set Θ. Assume there is some point $\vartheta_0 \in \Theta$ such that all the $P_{i,\vartheta}$ are dominated by P_{i,ϑ_0}, $i = 0, 1$, and form $\Lambda_i(\vartheta) = dP_{i,\vartheta}/dP_{i,\vartheta_0}$. Consider $\Lambda_i = (\Lambda_i(\vartheta), \vartheta \in \Theta)$ as stochastic processes indexed by ϑ given on the probability space $(\Omega_i, \mathscr{A}_i, P_{i,\vartheta_0})$. By a slight abuse of language, we call these the *likelihood processes* of the experiments $\mathbb{E}_i$ (note that the distribution is taken under P_{i,ϑ_0} here). Suppose also that there are versions Λ_i^* of these likelihood processes defined on a common probability space $(\Omega, \mathscr{A}, \mathbb{P})$.

PROPOSITION 2.2. *The deficiency distance* $\Delta(\mathbb{E}_0, \mathbb{E}_1)$ *satisfies*

$$\Delta(\mathbb{E}_0, \mathbb{E}_1) \le \sup_{\vartheta\in\Theta} E_{\mathbb{P}}|\Lambda_0^*(\vartheta) - \Lambda_1^*(\vartheta)|.$$

PROOF. It is one of the basic facts of Le Cam's theory that, for dominated experiments, the equivalence class is determined by the distribution of the likelihood processes under P_{i,ϑ_0} when ϑ_0 is assumed fixed. This means that in the above framework we have $\Delta(\mathbb{E}_0, \mathbb{E}_1) = 0$ iff $\mathscr{L}(\Lambda_0|P_{0,\vartheta_0}) = \mathscr{L}(\Lambda_1|P_{1,\vartheta_0})$. Thus, if we construct an experiment $\mathbb{E}_i^*$ with likelihood process Λ_i^*, we obtain equivalence: $\Delta(\mathbb{E}_i, \mathbb{E}_i^*) = 0$. The random variables $\Lambda_i^*(\vartheta)$ on $(\Omega, \mathscr{A}, \mathbb{P})$ have the same distributions as $\Lambda_i(\vartheta)$ on $(\Omega_i, \mathscr{A}_i, P_{i,\vartheta_0})$, for all $\vartheta \in \Theta$; hence they are positive and integrate to 1. They may hence be considered as $\mathbb{P}$-densities on $(\Omega, \mathscr{A})$, indexed by ϑ. These densities define measures $P_{i,\vartheta}^*$ on $(\Omega, \mathscr{A})$ and experiments $\mathbb{E}_i^* = (\Omega, \mathscr{A}, (P_{i,\vartheta}^*, \vartheta \in \Theta))$, $i = 0, 1$. By construction, the likelihood process for $\mathbb{E}_i^*$ is $\Lambda_i^*(\vartheta)$, so $\Delta(\mathbb{E}_i, \mathbb{E}_i^*) = 0$, $i = 0, 1$. Hence $\Delta(\mathbb{E}_0, \mathbb{E}_1) = \Delta(\mathbb{E}_0^*, \mathbb{E}_1^*)$, and $\mathbb{E}_0^*$ and $\mathbb{E}_1^*$ are given *on the same measurable space* $(\Omega, \mathscr{A})$. In this case, an upper bound for the deficiency distance is

$$\Delta(\mathbb{E}_0^*, \mathbb{E}_1^*) \le \sup_{\vartheta\in\Theta} \|P_{0,\vartheta}^* - P_{1,\vartheta}^*\|,$$

where $\|\cdot\|$ is the total variation distance between measures [in (68), Section 9, take the identity map as a transition M]. But $\|P_{0,\vartheta}^* - P_{1,\vartheta}^*\|$ coincides with $E_{\mathbb{P}}|\Lambda_0^*(\vartheta) - \Lambda_1^*(\vartheta)|$, which is just an L_1-distance between densities. □

The argument may be summarized as follows: versions Λ_i^* of the likelihood processes on a common probability space generate (equivalent) versions of the experiments on a common measurable space for which $\Lambda_i^*(\vartheta)$ are densities. Their L_1-distance bounds the deficiency.

When $\Lambda_i^*(\vartheta)$ are considered as densities it is natural to employ also their Hellinger distance $H(\cdot,\cdot)$; extending the notation we write

$$H^2(\Lambda_0^*(\vartheta), \Lambda_1^*(\vartheta)) = E_{\mathbb{P}}\left((\Lambda_0^*(\vartheta))^{1/2} - (\Lambda_1^*(\vartheta))^{1/2}\right)^2. \tag{11}$$

Making use of the general relation of Hellinger distance to L_1-distance, we obtain

$$\Delta^2(\mathbb{E}_0^*, \mathbb{E}_1^*) \le 4 \sup_{\vartheta\in\Theta} H^2(\Lambda_0^*(\vartheta), \Lambda_1^*(\vartheta)). \tag{12}$$

In the sequel we will work basically with this relation to establish asymptotic equivalence. For our problem, we identify $\vartheta = f$, $\vartheta_0 = f_0$, $\Theta = \Sigma_n(f_0)$, $P_{0,\vartheta} = P_f^{\otimes n}$ and $P_{1,\vartheta} = Q_{n,f,f_0}$. Furthermore, we represent the observations y_i as $y_i = F^{-1}(z_i)$, where z_i are i.i.d. uniform (0, 1) random variables and F is the distribution function for the density f (note that F is strictly monotone for $f \in \Sigma$). Let U_n be the empirical process of $z_1, \dots, z_n$, that is,

$$U_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\chi_{[0,t]}(z_i) - t), \qquad t \in [0, 1].$$

Note that $\mathbb{E}_{0,n}(f_0)$ is dominated by $P_{f_0}^{\otimes n}$; then the likelihood process is

$$\Lambda_{0,n}(f, f_0) = \exp \sum_{i=1}^{n} \log \left\{ \frac{f}{f_0}(F_0^{-1}(z_i)) \right\}.$$

Defining

$$\lambda_{f,f_0}(t) = \log \left\{ \frac{f}{f_0}(F_0^{-1}(t)) \right\} \tag{13}$$

and observing that

$$\int \lambda_{f,f_0}(t) dt = -K(f_0 \| f),$$

we then have the following representation:

$$\Lambda_{0,n}(f, f_0) = \exp \left\{ n \int \lambda_{f,f_0}(t) \frac{1}{\sqrt{n}}\, dU_n(t) - nK(f_0 \| f) \right\}. \tag{14}$$

This suggests a corresponding Gaussian likelihood process: substitute U_n by a Brownian bridge B and renormalize to obtain integral 1. We thus form, for a uniform(0, 1) random variable Z,

$$\Lambda_{1,n}(f, f_0) = \exp \left\{ n \int \lambda_{f,f_0}(t) \frac{1}{\sqrt{n}}\, dB(t) - \frac{n}{2} \operatorname{Var}(\lambda_{f,f_0}(Z)) \right\}. \tag{15}$$

For an appropriate standard Wiener process W we have

$$\int \lambda_{f,f_0}(t)\, dB(t) = \int (\lambda_{f,f_0}(t) + K(f_0 \parallel f))\, dW(t).$$

By rewriting the likelihood process $\Lambda_{1,n}(f, f_0)$ accordingly, we see that it corresponds to observations (9) or, equivalently, to

$$dy(t) = (\lambda_{f,f_0}(t) + K(f_0 \| f))dt + n^{-1/2}\, dW(t), \qquad t \in [0, 1], \tag{16}$$

at least when the parameter space is Σ. Thus $\Lambda_{1,n}(f, f_0)$ is in fact the likelihood process for $\mathbb{E}_{1,n}(f_0)$ in (10).

To find nearby versions of these likelihood processes, fulfilling

$$\sup_{f \in \Sigma_n(f_0)} H^2 \left(\Lambda_{0,n}^*(f, f_0), \Lambda_{1,n}^*(f, f_0) \right) \to 0, \tag{17}$$

it would be natural to look for versions of U_n and B on a common probability space ($\mathbb{U}_n$ and $\mathbb{B}_n$, say) which are close, such as in the classical *Hungarian construction* [see Shorack and Wellner (1986), Chapter 12, Section 1, Theorem 2]. However, the classical Hungarian construction [Komlós–Major–Tusnady (KMT) inequality] gives an estimate of the uniform distance $\|\mathbb{U}_n - \mathbb{B}_n\|_\infty$ which for our purpose is not optimal. The reason is that the uniform distance may be construed as

$$\|\mathbb{U}_n - \mathbb{B}_n\|_\infty = \sup_{g\in\mathscr{G}} \left|\mathbb{U}_n(g) - \mathbb{B}_n(g)\right|,$$

where $\mathscr{G}$ is a class of indicators of subintervals of $[0, 1]$. Considering more general classes of functions $\mathscr{G}$ leads to *functional* KMT type results [see Koltchinskii (1994) and Rio (1994)]. However, for an estimate (17) we need to control the random difference $\mathbb{U}_n(g) - \mathbb{B}_n(g)$ only for one given function (λ_{f,f_0} in this case), with a supremum over a function class only after taking expectations [cf. the remark of Le Cam and Yang (1990), page 16]. Thus for our purpose we ought to use a functional KMT type inequality for a one-element function class $\mathscr{G} = \{g\}$, but where the same constants and *one Brownian bridge* are still available over a class of smooth g. Such a result is provided by Koltchinskii [(1994), Theorem 3.5]. We present a version slightly adapted for our purpose. Let $\mathscr{L}_2[0, 1]$ be the space of all square integrable measurable functions on $[0, 1]$, and let $\|\cdot\|_{H_2^{1/2}}$ be the seminorm associated with a Hölder condition with exponent $1/2$ in the L_2-sense (see Section 5 for details).

PROPOSITION 2.3. *There are a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ and a number C such that for all n there are versions of the uniform empirical process $\mathbb{U}_n(g)$ and of the Brownian bridge $\mathbb{B}_n(g)$, $g \in \mathscr{L}_2[0, 1]$, such that for all g with $\|g\|_\infty < \infty$, $\|g\|_{H_2^{1/2}} < \infty$ and, for all $t \geq 0$,*

$$\mathbb{P}\big(n^{1/2}|\mathbb{U}_n(g) - \mathbb{B}_n(g)| \geq C(\|g\|_\infty + \|g\|_{H_2^{1/2}})(t + \log n)\log^{1/2} n\big) \leq C\exp(-t).$$

Specializing $g = \lambda_{f,f_0} - \int \lambda_{f,f_0}$, we come close to establishing relation (17) for the likelihood processes, but we need an assumption that the neighborhoods $\Sigma_n(f_0)$ shrink with rate $o(n^{-1/3})$. Comparing this with the usual nonparametric rates of convergence, we see that such a result is useful only for smoothness $\alpha > 1$. To treat the case $\alpha > 1/2$, however, we need neighborhoods of size $o(n^{-1/4})$.

To obtain such a result, it is convenient, rather than using the Hungarian construction globally on $[0, 1]$, to subdivide the interval and use a corresponding independence structure (approximate or exact) of both experiments. In this connection the following result is useful [see Strasser (1985), Lemma 2.19].

LEMMA 2.4. *Suppose that P_i and Q_i are probability measures on a measurable space $(\Omega_i, \mathscr{A}_i)$, for $i = 1, \ldots, k$. Then*

$$H^2\left(\bigotimes_{i=1}^k P_i, \bigotimes_{i=1}^k Q_i\right) \leq 2\sum_{i=1}^k H^2(P_i, Q_i).$$

Consider a partition of $[0,1]$ into subintervals D_j. The Gaussian experiment $\mathbb{E}_{1,n}(f_0)$ has a convenient independence structure: in the representation (16), observations on the signal $\lambda_{f,f_0}(t) + K(f_0\|f)$ are independent on different pieces D_j. A corresponding approximate product structure for the i.i.d. experiment $\mathbb{E}_{0,n}(f_0)$ will be established by Poissonization. Let $\mathbb{E}_{0,j,n}(f_0)$ be the experiment given by observing "interval censored" observations

$$y_i \chi_{D_j}(y_i), \qquad y_i \text{ i. i. d. with density } f,\ i = 1, \dots, n, \tag{18}$$

with $f \in \Sigma_n(f_0)$. We use the symbol $\bigotimes$ for products of experiments having the same parameter space.

PROPOSITION 2.5. *Let k_n be a sequence with $k_n \to \infty$, and consider a partition $D_j = [(j-1)/k_n,\ j/k_n)$, $j = 1, \dots, k_n$. Then*

$$\Delta\left(\mathbb{E}_{0,n}(f_0), \bigotimes_{j=1}^{k_n} \mathbb{E}_{0,j,n}(f_0)\right) \to 0$$

uniformly over $f_0 \in \Sigma$.

Our choice of k_n will be

$$k_n \sim n^{1/2}/\log^{3/2} n. \tag{19}$$

For each D_j we form a local likelihood process $\Lambda_{0,j,n}(f, f_0)$, as the likelihood process for observations in (18) for given j, and establish a Gaussian approximation like (17) with a rate. Let $A_j = F_0(D_j)$ and let $\mathbb{E}_{1,j,n}(f_0)$ be the Gaussian experiment

$$dy(t) = \chi_{A_j}(t)\,(\lambda_{f,f_0}(t) + K(f_0\|f))\,dt + n^{-1/2}\,dW(t), \qquad t \in [0,1], \tag{20}$$

with parameter space $\Sigma_n(f_0)$. Let $\Lambda_{1,j,n}(f, f_0)$ be the corresponding likelihood process.

PROPOSITION 2.6. *On the probability space $(\Omega, \mathscr{A}, \mathbb{P})$ of Proposition 2.3, there are versions $\Lambda^*_{i,j,n}(f, f_0)$, $i = 0, 1$, such that*

$$\sup_{f \in \Sigma_n(f_0)} H^2\big(\Lambda^*_{0,j,n}(f, f_0), \Lambda^*_{1,j,n}(f, f_0)\big) = O(\gamma_n^2 (\log n)^3) \tag{21}$$

uniformly over $j = 1, \dots, k_n$ and $f_0 \in \Sigma$.

This admits the following interpretation. Define $m_n = n/k_n$; in our setting this is the stochastic order of magnitude of the number of observations y_i falling into D_j. Thus for the local likelihood process $\Lambda_{0,j,n}(f, f_0)$ the number m_n represents an "effective sample size" in a rate sense. In view of (6) and (19) we have $\gamma_n \sim m_n^{-1/2} (\log n)^{-1/4}$, and since this is the shrinking rate of $\Sigma_n(f_0)$ in the uniform norm, it is also the shrinking rate of this set of densities restricted to D_j, and of the corresponding set of conditional densities. Thus in a sense we are "almost" in a classical setting with sample size m_n and a root-m_n shrinking neighborhood. Result (21) implies

$$\Delta(\mathbb{E}_{0,\,j,\,n}(f_0), \mathbb{E}_{1,\,j,\,n}(f_0)) = O(m_n^{-1/2}(\log n)^{5/4}), \tag{22}$$

that is, we have a root-m_n rate up to a log term. Note that here we have introduced a "space-local" aspect in addition to the already present parameter-local one. In piecing together these space-local approximations, we will crucially use the product measure estimate of Lemma 2.4. This motivates our choice to work with the Hellinger distance for the likelihood processes construed as densities.

PROOF OF THEOREM 2.1. The Gaussian experiment $\mathbb{E}_{1,\,n}(f_0)$ decomposes exactly:

$$\Delta\left(\mathbb{E}_{1,\,n}(f_0), \bigotimes_{j=1}^{k_n} \mathbb{E}_{1,\,j,\,n}(f_0)\right) = 0.$$

According to (12) and Lemma 2.4 we have

$$\begin{aligned}\Delta^2&\left(\bigotimes_{j=1}^{k_n} \mathbb{E}_{0,\,j,\,n}(f_0), \bigotimes_{j=1}^{k_n} \mathbb{E}_{1,\,j,\,n}(f_0)\right)\\ &\le 4 \sup\nolimits_{f\in\Sigma_n(f_0)} \textstyle\sum_{j=1}^{k_n} H^2\big(\Lambda^*_{0,j,n}(f, f_0), \Lambda^*_{1,\,j,\,n}(f, f_0)\big).\end{aligned}$$

By Proposition 2.6 this is bounded by

$$O\left(k_n \gamma_n^2 (\log n)^3\right) = O\left((\log n)^{-1/2}\right) = o(1),$$

and these estimates hold uniformly over $f_0 \in \Sigma$. $\square$

Low (1992) considered experiments given by local (on D_j) perturbations of a fixed density f_0 and applied a local asymptotic normality argument to obtain strong convergence to a Gaussian experiment. This amounts to having (22) without a rate, and it is already useful for a number of nonparametric decision problems, like estimating the density at a point. Golubev (1991) used a similar argument for treating estimation with L_2-loss.

We are now able to identify several more asymptotically equivalent models. This is based on the following reasoning, applied by Brown and Low (1996) to compare Gaussian white noise models. Consider the measure of the process $n^{-1/2}\,W(t)$, $t \in [0, 1]$, shifted by a function $\int_0^t g_i$, $i = 1, 2$, where $g_i \in \mathscr{L}_2[0, 1]$; call these measures P_i. Then

$$H^2(P_1, P_2) = 2\left(1 - \exp\left\{-\frac{n}{8}\, \|g_1 - g_2\|_2^2\right\}\right). \tag{23}$$

If $(g_{i,\vartheta},\ \vartheta \in \Theta)$, $i = 1, 2$, are two parametric families, then the respective experiments are asymptotically equivalent if $\|g_{1,\vartheta} - g_{2,\vartheta}\|_2 = o(n^{-1/2})$ uniformly over $\vartheta \in \Theta$. In the Gaussian experiment $\mathbb{E}_{1,n}(f_0)$ of (16), the shift is essentially a log-density ratio. We know that $\log(f/f_0)$ is small over $f \in \Sigma_n(f_0)$; expanding the logarithm, we get asymptotically equivalent experiments with parameter space $\Sigma_n(f_0)$.

Accordingly, let $\mathbb{E}_{2,n}(f_0)$ be the experiment given by observations

$$dy(t) = (f(t) - f_0(t))\,dt + n^{-1/2} f_0^{1/2}(t)\,dW(t), \qquad t \in [0,1], \tag{24}$$

with parameter space $\Sigma_n(f_0)$, and let $\mathbb{E}_{3,n}(f_0)$ correspondingly be given by

$$dy(t) = (f^{1/2}(t) - f_0^{1/2}(t))\,dt + \tfrac{1}{2} n^{-1/2}\,dW(t), \qquad t \in [0,1]. \tag{25}$$

THEOREM 2.7. *The experiments* $\mathbb{E}_{i,n}(f_0)$, $i = 1, 2, 3$, *are asymptotically equivalent*, *uniformly over* $f_0 \in \Sigma$.

REMARK 2.8. The equivalence class of $\mathbb{E}_{2,n}(f_0)$ is not changed when the additive term $-f_0(t)\,dt$ in (24) is omitted, since this term does not depend on the parameter f, and omitting it amounts to a translation of the observed process y by a known quantity. Moreover, in the proof below it will be seen that in the representation (16) of $\mathbb{E}_{1,n}(f_0)$ the term $K(f_0\|f)\,dt$ is asymptotically negligible. Analogous statements are true for the other variants; hence locally asymptotically equivalent experiments for $f \in \Sigma_n(f_0)$ (with uniformity over $f_0 \in \Sigma$) are also given by the following:

$$y_i, \quad i = 1, \dots, n \qquad \text{i.i.d. with density } f; \tag{26}$$

$$dy(t) = \log f(F_0^{-1}(t))\,dt + n^{-1/2}\,dW(t), \qquad t \in [0,1]; \tag{27}$$

$$dy(t) = f(t)\,dt + n^{-1/2} f_0^{1/2}(t)\,dW(t), \qquad t \in [0,1]; \tag{28}$$

$$dy(t) = f^{1/2}(t)\,dt + \tfrac{1}{2} n^{-1/2}\,dW(t), \qquad t \in [0,1]. \tag{29}$$

Note that (28) is related to the weak convergence of the empirical distribution function $\bar{F}_n$,

$$n^{1/2}\left(\bar{F}_n - F\right) \Rightarrow B \circ F.$$

Indeed, arguing heuristically, when F is in a shrinking neighborhood of F_0 we have $B \circ F \approx B \circ F_0$, while $\bar{F}_n$ is a sufficient statistic. We obtain

$$\bar{F}_n \approx F + n^{-1/2} B \circ F_0,$$

which suggests a Gaussian accompanying experiment (28). This reasoning is familiar as a heuristic introduction to limiting Gaussian shift experiments, when neighborhoods are shrinking with rate $n^{-1/2}$. However, our neighborhoods $f \in \Sigma_n(f_0)$ are larger [recall $\gamma_n = n^{-1/4}(\log n)^{-1}$].

3. From local to global results. The local result concerning a shrinking neighborhood of some f_0 is of limited value for statistical inference since usually such prior information is not available. Following Le Cam's general principles, we shall construct an experiment where the prior information is furnished by a preliminary estimator, and subsequently the local Gaussian approximation is built around the estimated parameter value.

To formalize this approach, let N_n define a "fraction of the sample size;" that is, N_n is a sequence $N_n \to \infty$, $N_n < n$, and consider the corresponding fraction of the sample $y_1, \dots, y_{N_n}$. Then let $\hat{f}_n$ be an estimator of f based on this fraction, fulfilling (with $P_{n,f}$ the pertinent measure)

$$\inf_{f \in \Sigma} P_{n,f}(\hat{f}_n \in \Sigma_n(f)) \to 1. \tag{30}$$

The set Σ must be such that the shrinking rate of $\Sigma_n(f)$ is an attainable rate for estimators. If f has a bounded derivative of order α, we have for f an attainable rate in sup norm $(n/\log n)^{-\alpha/(2\alpha+1)}$ [see Woodroofe (1967)]. The required sup norm rate is $\gamma_n = o(n^{-1/4})$; this corresponds to $\alpha > 1/2$. Thus we may expect for the Hölder smoothness classes assumed here that the rate γ_n is attainable if the size N_n of the fraction is sufficiently large. We will allow for a range of choices:

$$n/\log n \le N_n \le n/2. \tag{31}$$

Define $\mathbb{E}_{0,n}$ to be the original i.i.d. experiment (3) with global parameter space Σ.

LEMMA 3.1. *Suppose* (31) *holds. Then in* $\mathbb{E}_{0,n}$ *there exists a sequence of estimators* $\hat{f}_n$ *depending only on* $y_1, \dots, y_{N_n}$ *fulfilling* (30). *One may assume that for each* n *the estimator takes values in a finite subset of* Σ.

The proof is in Section 8. The following construction of a global approximating experiment assumes such an estimator sequence fixed. The idea is to substitute $\hat{f}_n$ for f_0 in the local Gaussian approximation and to retain the first fraction of the i.i.d. sample. Recall that our local Gaussian approximations were given by families $(Q_{n,f,f_0}, f \in \Sigma_n(f_0))$ [cf. (10)]. Note that $f \in \Sigma_n(f_0)$ is essentially the same as $f_0 \in \Sigma_n(f)$. Accordingly we now consider the event $\hat{f}_n \in \Sigma_n(f)$ and let f range in the unrestricted parameter space Σ. We look at the second sample part, of size $n - N_n$, with its initial i.i.d family $(P_f^{\otimes(n-N_n)}, f \in \Sigma)$. Based on the results of the previous section, we can hope that this family will be close, in the experiment sense, to the conditionally Gaussian family $(Q_{n-N_n,f,\hat{f}_n}, f \in \Sigma)$, on the event $\hat{f}_n \in \Sigma_n(f)$. The measures $Q_{n,f,\hat{f}_n}$, which now depend on $\hat{f}_n$, have to be interpreted as conditional measures, and we form a joint distribution with the first sample fraction.

This idea is especially appealing when the locally approximating Gaussian measure Q_{n,f,f_0} does not depend on the "center" f_0. In this case the resulting global experiment will have a convenient product structure, as we shall see.

This is the case with the variant (29) in Remark 2.8, when we parametrize with $f^{1/2}$.

To be more precise, define Q_{i,n,f,f_0}, $i = 1, 2, 3$, to be the distributions of $(y(t), t \in [0,1])$ in (27)–(29). Consider a "compound experiment" given by joint observations $y_1, \dots, y_{N_n}$ and $y = (y(t),\ t \in [0,1])$, where

$$y_1, \dots, y_{N_n} \text{ are i.i.d. with density } f, \tag{32}$$

$$\mathscr{L}(y|y_1, \dots, y_{N_n}) = Q_{i,n-N_n,f,\hat{f}_n}. \tag{33}$$

Here (33) describes the conditional distribution of y given $y_1, \dots, y_{N_n}$. Define $R_{i,n,f}(\hat{f})$ to be the joint distribution of $y_1, \dots, y_{N_n}$ and y in this setup, for $i = 1, 2, 3$; the notation signifies dependence on the sequence of decision functions $\hat{f} = \{\hat{f}_n\}_{n\geq 1}$ (not dependence on the estimator value). Then the compound experiment is

$$\mathbb{E}_{i,n}(\hat{f}) = \big([0,1]^{N_n} \times C_{[0,1]}, \mathscr{B}_{[0,1]}^{N_n} \otimes \mathscr{B}_{C_{[0,1]}}, (R_{i,n,f}(\hat{f}),\ f \in \Sigma)\big).$$

Since $Q_{3,n,f,f_0} = Q_{3,n,f}$ does not depend on f_0, the measure $R_{3,n,f}(\hat{f}) = R_{3,n,f}$ does not depend on $\hat{f}$ either and is just the product measure $P_f^{\otimes N_n} \otimes Q_{3,n-N_n,f}$. We also write $\mathbb{E}_{3,n}(\hat{f}) = \mathbb{E}_{3,n}$. The technical implementation of the above heuristic reasoning (see Section 9) gives the following result.

THEOREM 3.2. *Suppose* (31) *holds and let $\hat{f}_n$ be a sequence of estimators as in Lemma* 3.1. *Then*, *for $i = 1, 2, 3$*,

$$\Delta(\mathbb{E}_{0,n}, \mathbb{E}_{i,n}(\hat{f})) \to 0.$$

To restate this in a more transparent fashion, we refer to $y_1, \dots, y_{N_n}$ and $y = (y(t),\ t \in [0,1])$ in (32) and (33) as the first and second parts of the compound experiment, respectively. Let $\hat{F}_n$ be the distribution function corresponding to the realized density estimator $\hat{f}_n$.

COROLLARY 3.3. *Under the conditions of Theorem* 3.2, *the compound experiments with first part*

$$y_i, \qquad i = 1, \dots, N_n,\ \textit{i.i.d. with density } f \tag{34}$$

and respective second parts

$$y_i, \qquad i = N_n + 1, \dots, n,\ \textit{i.i.d.with density } f, \tag{35}$$

$$dy(t) = \log f(\hat{F}_n^{-1}(t)) + (n - N_n)^{-1/2}\, dW(t), \qquad t \in [0,1], \tag{36}$$

$$dy(t) = f(t)\, dt + (n - N_n)^{-1/2} \hat{f}_n^{1/2}(t)\, dW(t), \qquad t \in [0,1], \tag{37}$$

$$dy(t) = f^{1/2}(t)\, dt + \tfrac{1}{2}(n - N_n)^{-1/2} dW(t), \qquad t \in [0,1], \tag{38}$$

with $f \in \Sigma$ are all asymptotically equivalent.

For obtaining a closed-form global approximation, the compound experiment $\mathbb{E}_{3,n}$ [i.e., (34), (38)] is the most interesting one, in view of its product structure and independence of $\hat{f}$. Here the estimator sequence $\hat{f}$ only serves to show asymptotic equivalence to $\mathbb{E}_{0,n}$; it does not show up in the target experiment $\mathbb{E}_{3,n}$ itself. This structure of $\mathbb{E}_{3,n}$ suggests employing an estimator based on the second part for a next step.

LEMMA 3.4. *Suppose* (31) *holds. Then in* $\mathbb{E}_{3,n}$ *there exists a sequence of estimators* $\check{f}_n$ *depending only on* y *in* (38) *fulfilling* (30). *The second statement of Lemma* 3.1 *also applies.*

Note the similarity to Lemma 3.1. Here we exploit the well-known parallelism of density estimation and white noise on the rate of convergence level.

PROOF OF THEOREM 1.1. We choose $N_n = [n/2]$. On the resulting compound experiment $\mathbb{E}_{3,n}$ we may then operate again, reversing the roles of first and second part. We may in turn substitute $y_1, \dots, y_{N_n}$ by a white noise model, using a preliminary estimator based on (38). The existence of such an estimator is guaranteed by the previous lemma. Thus substituting $y_1, \dots, y_{N_n}$ by white noise leads to an experiment with joint observations

$$dy_1(t) = f^{1/2}(t)\,dt + \tfrac{1}{2}N_n^{-1/2}dW_1(t), \qquad t \in [0, 1],$$
$$dy_2(t) = f^{1/2}(t)\,dt + \tfrac{1}{2}(n - N_n)^{-1/2}\,dW_2(t), \qquad t \in [0, 1],$$

where W_1 and W_2 are independent Wiener processes. A sufficiency argument shows this is equivalent to observing n i.i.d. processes, each distributed as

$$dy(t) = f^{1/2}(t)\,dt + \tfrac{1}{2}\,dW(t), \qquad t \in [0, 1],$$

which in turn is equivalent to (4). □

4. Poissonization and product structure. For the proof of Proposition 2.5 we need some basic concepts from the theory of point processes [see Reiss (1993)]. A point measure on $(\mathbb{R}, \mathscr{B})$ is a measure μ: $\mathscr{B} \mapsto [0, \infty]$ of form $\mu = \sum_{i \in I} \mu_{x_i}$, where $I \subset \mathbb{N}$, x_i are points in $\mathbb{R}$ and μ_x is Dirac measure at x. A point process is a random variable on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with values in the space of point measures $\mathbb{M}$ equipped with the appropriate σ-algebra $\mathscr{M}$ [see Reiss (1993), page 6]. If $Y = \{y_i,\ i = 1, 2, \dots\}$ is a sequence of i.i.d. r.v.'s, then the random measure $\mu_{0,n} = \sum_{i=1}^{n} \mu_{y_i}$ is called an empirical point process. More generally if ν is a random natural number independent of Y, then $\mu = \sum_{i=1}^{\nu} \mu_{y_i}$ is a mixed empirical point process. In particular, if $\nu = \pi_n$ is Poisson(n), then $\mu_{*,n} = \sum_{i=1}^{\pi_n} \mu_{y_i}$ is a Poisson process which has intensity function nf if y_1 has density f. If f and f_0 are two densities for y_1 such that $P_f \ll P_{f_0}$ and the law of ν is given, then it is possible to write down densities for the distributions $\Pi_f := \mathscr{L}(\mu \mid P_f)$ of the mixed empirical point process μ. For the case of the empirical and the Poisson point process ($\nu = n$ or $\nu = \pi_n$) we shall denote these distributions, respectively, by $\Pi_{0,n,f}$ and $\Pi_{*,n,f}$.

For observations $(\nu,\ y_i,\ i = 1,\dots,\nu)$ write the likelihood ratio for hypotheses $(P_f,\ \mathscr{L}(\nu))$ versus $(P_{f_0},\ \mathscr{L}(\nu))$,

$$\prod_{i=1}^{\nu}(f/f_0)(y_i) = \exp\int \log(f/f_0)\,d\mu. \tag{39}$$

This is a function of μ which can be construed as a density of the point process law Π_f on $(\mathbb{M}, \mathscr{M}, \Pi_{f_0})$ or as a likelihood process when f varies. Note that for different $\mathscr{L}(\nu)$ these densities are defined on different probability spaces, since the respective laws Π_{f_0} differ. However, let $(\Omega, \mathscr{A}, \mathbb{P}) = ([0,1]^\infty, \mathscr{B}^\infty_{[0,1]}, \lambda^{\otimes\infty})$, where λ is Lebesgue measure on $[0,1]$, and let Y and ν be defined on that space (as independent r.v.'s). Then (39) also describes versions on the probability space $(\Omega, \mathscr{A}, \mathbb{P})$, which is common for different $\mathscr{L}(\nu)$. For the case of the empirical and the Poisson point process ($\nu = n$ or $\nu = \pi_n$) we shall denote these likelihood processes, respectively, by $\Lambda_{0,n}(f, f_0)(\omega)$ and $\Lambda_{*,n}(f, f_0)(\omega)$. The experiments defined by these versions construed as $\mathbb{P}$-densities are then equivalent to the respective point process experiments, for any parameter space. In particular, the empirical point process experiment (with laws $\Pi_{0,n,f}$) is equivalent to the original i.i.d. experiment with n observations; $\mu_{0,n} = \sum_{i=1}^n \mu_{y_i}$ is a sufficient statistic.

For our particular parameter space $\Sigma_n(f_0)$ define the Poisson process experiment

$$\mathbb{E}_{*,n}(f_0) = (\mathbb{M}, \mathscr{M}, (\Pi_{*,n,f},\ f \in \Sigma_n(f_0)))$$

and recall the definition (8) of the i.i.d. experiment $\mathbb{E}_{0,n}(f_0)$.

PROPOSITION 4.1. *We have*

$$\Delta(\mathbb{E}_{0,n}(f_0), \mathbb{E}_{*,n}(f_0)) \to 0$$

uniformly over $f_0 \in \Sigma$.

PROOF. We use an argument adapted from Le Cam (1985). It suffices to establish that

$$H^2(\Lambda_{0,n}(f, f_0), \Lambda_{*,n}(f, f_0)) = O(n^{1/2}\gamma_n^2)$$

uniformly over $f \in \Sigma_n(f_0)$, $f_0 \in \Sigma$. With $\nu_{\min} = \min(\pi_n, n)$ and $\nu_{\max} = \max(\pi_n, n)$ we get

$$\begin{aligned} H^2(\Lambda_{0,n}(f, f_0), \Lambda_{*,n}(f, f_0)) &= E_{\mathbb{P}}\left(\prod_{i=1}^{n}(f/f_0)^{1/2}(y_i) - \prod_{i=1}^{\pi_n}(f/f_0)^{1/2}(y_i)\right)^2 \\ &= E_{\mathbb{P}}\prod_{i=1}^{\nu_{\min}}(f/f_0)(y_i)\left(\prod_{i=\nu_{\min}+1}^{\nu_{\max}}(f/f_0)^{1/2}(y_i) - 1\right)^2. \end{aligned}$$

Consider first the conditional expectation when π_n is given; since y_i are independent it is

$$E_{\mathbb{P}}\biggl(\biggl(\prod_{i=\nu_{\min}+1}^{\nu_{\max}}(f/f_0)^{1/2}(y_i)-1\biggr)^2\biggm|\pi_n\biggr).$$

This can be construed as the squared Hellinger distance of two product densities, one of which has $\nu_{\max}-\nu_{\min}=|\pi_n-n|$ factors and the other has as many factors equal to unity. Applying Lemma 2.4, we get an upper bound

$$2\sum_{i=\nu_{\min}+1}^{\nu_{\max}}E_{\mathbb{P}}\bigl(((f/f_0)^{1/2}(y_i)-1)^2\bigm|\pi_n\bigr)\le 2\;|\pi_n-n|\;\gamma_n^2.$$

Taking an expectation and observing $E|\pi_n-n|\le Cn^{1/2}$ completes the proof. □

If μ is a point process and D a measurable set, then define the truncated point process

$$\mu_D(B)=\mu(B\cap D),\qquad B\in\mathscr{B}.$$

Let $\mu_{0,n,D}$ and $\mu_{*,n,D}$ be truncated empirical and Poisson point processes on $[0,1]$, respectively. The following Hellinger distance estimate is due to Falk and Reiss (1992); see also Reiss [(1993), Theorem 1.4.2]:

$$H(\mathscr{L}(\mu_{0,n,D}\,|\,f),\mathscr{L}(\mu_{*,n,D}\,|\,f))\le\sqrt{3}P_f(D).\tag{40}$$

PROOF OF PROPOSITION 2.5. By the previous proposition it suffices to establish that

$$\Delta\biggl(\mathbb{E}_{*,n}(f_0),\bigotimes_{j=1}^{k_n}\mathbb{E}_{0,j,n}(f_0)\biggr)\to 0\tag{41}$$

uniformly over $f_0\in\Sigma$. In $\mathbb{E}_{0,j,n}(f_0)$ we observe n i.i.d. truncated random variables (18); their empirical point process is a sufficient statistic. Hence μ_{0,n,D_j} (the truncated empirical point process for the original y_i) is a sufficient statistic also; let $\Pi_{0,j,n,f}=\mathscr{L}(\mu_{0,n,D_j}\,|\,f)$ be the corresponding law. It follows that each $\mathbb{E}_{0,j,n}(f_0)$ is equivalent to an experiment

$$\mathbb{E}^*_{0,j,n}(f_0)=(\mathbb{M},\mathscr{M},(\Pi_{0,j,n,f},f\in\Sigma_n(f_0))).$$

Let $\Pi_{*,j,n,f}=\mathscr{L}(\mu_{*,n,D_j}|f)$ be the law of the truncated Poisson point process and

$$\mathbb{E}_{*,j,n}(f_0)=(\mathbb{M},\mathscr{M},(\Pi_{*,j,n,f},f\in\Sigma_n(f_0))).$$

Then by the properties of the Poisson process, $\mathbb{E}_{*,n}(f_0)$ is equivalent to $\otimes_{j=1}^{k_n}\mathbb{E}_{*,j,n}(f_0)$. It now suffices to show that

$$\Delta\biggl(\bigotimes_{j=1}^{k_n}\mathbb{E}_{*,j,n}(f_0),\bigotimes_{j=1}^{k_n}\mathbb{E}^*_{0,j,n}(f_0)\biggr)\to 0$$

uniformly over $f_0 \in \Sigma$. From Lemma 2.4 and (40) we obtain

$$H^2\left(\bigotimes_{j=1}^{k_n} \Pi_{*,\,j,\,n,\,f}, \bigotimes_{j=1}^{k_n} \Pi_{0,\,j,\,n,\,f}\right) \le 2\sum_{j=1}^{k_n} H^2(\Pi_{*,\,j,\,n,\,f}, \Pi_{0,\,j,\,n,\,f}) \le 6\sum_{j=1}^{k_n} P_f^2(D_j)$$
$$\le 6 \sup_{1\le j\le k_n} P_f(D_j).$$

The functions $f \in \Sigma$ are uniformly bounded, in view of the uniform Hölder condition and $\int f = 1$. Hence $P_f(D_j) \to 0$ uniformly in $f \in \Sigma$ and j. □

5. Empirical processes and function classes. From the point process framework we now return to the traditional notion of the empirical process as a normalized and centered random function. However, we consider processes indexed by functions. Let z_i, $i = 1, \ldots, n$, be i.i.d. uniform random variables on $[0, 1]$. Then

$$U_n(f) = n^{1/2}\left(n^{-1}\sum_{i=1}^{n} f(z_i) - \int f\right), \qquad f \in \mathscr{L}_2[0, 1],$$

is the uniform empirical process. The corresponding Brownian bridge process is defined as a centered Gaussian random function $B(f)$, $f \in \mathscr{L}_2[0, 1]$, with covariance

$$EB(f)B(g) = \int fg - \left(\int f\right)\left(\int g\right), \qquad f, g \in \mathscr{L}_2[0, 1],$$

For any natural i, consider the subspace of $\mathscr{L}_2[0, 1]$ consisting of piecewise constant functions on $[0, 1]$ for a partition $[(j-1)2^{-i},\ j2^{-i})$, $j = 1, \ldots, 2^i$. Let $g_{\langle i\rangle}$ be the projection of a function g onto that subspace, and define, for natural K,

$$q_K(g) = \left(\sum_{i=0}^{K} 2^i \left\| g - g_{\langle i\rangle}\right\|_2^2\right)^{1/2}.$$

The following version of a KMT inequality is due to Koltchinskii [(1994), Theorem 3.5] (specialized to a single element function class $\mathscr{F}$ there and to $K = \log_2 n$).

PROPOSITION 5.1. *There are a probability space* $(\Omega, \mathscr{A}, \mathbb{P})$ *and numbers* C_1, C_2 *such that for all* n *there are versions* $\mathbb{U}_n$ *and* $\mathbb{B}_n$ *of the empirical process and of the Brownian bridge such that, for all* $g \in \mathscr{L}_2[0, 1]$ *with* $\|g\|_\infty \le 1$ *and for all* x, $y \ge 0$,

$$\mathbb{P}\big(n^{1/2}\left|\mathbb{U}_n(g) - \mathbb{B}_n(g)\right| \ge x + x^{1/2}y^{1/2}(q_{\log_2 n}(g) + 1)\big) \le C_1(\exp(-C_2 x) + n\exp(-C_2 y)). \tag{42}$$

To set $q_K(g)$ in relation to a smoothness measure, consider functions $g \in \mathscr{L}_2[0,1]$ satisfying, for some C,

$$\int_h^{1-h} (g(u+h) - g(u))^2 du \le Ch \quad \text{for all } h > 0. \tag{43}$$

For a given g, define $\|g\|^2_{H_2^{1/2}}$ as the infimum of all numbers C such that (43) holds; it is easy to see that $\|\cdot\|_{H_2^{1/2}}$ is a seminorm. The corresponding space $H_2^{1/2}$ with norm $\|\cdot\|_2 + \|\cdot\|_{H_2^{1/2}}$ coincides with the Besov space $B^{1/2}_{2,\infty}$ on $[0,1]$ [see Nikolskij (1975), 4.3.3, 6.2]. Furthermore [cf. Koltchinskii (1994), relation (4.5)],

$$q_K^2(g) \le 4K \, \|g\|^2_{H_2^{1/2}}.$$

PROOF OF PROPOSITION 2.3. If g fulfills $\|g\|_\infty < \infty$, we divide by $\|g\|_\infty$, and apply (42); furthermore, we set $y = x + C_2^{-1}\log n$, $x = C_2^{-1}t$ and obtain, from (42),

$$\begin{aligned}
&2C_1 \exp(-t)\\
&\quad \ge \mathbb{P}\big(n^{1/2}\left|\mathbb{U}_n(g) - \mathbb{B}_n(g)\right| \ge \|g\|_\infty\, x + x^{1/2}y^{1/2}(q_{\log_2 n}(g) + \|g\|_\infty)\big)\\
&\quad \ge \mathbb{P}(n^{1/2}\left|\mathbb{U}_n(g) - \mathbb{B}_n(g)\right| \ge \|g\|_\infty\, x + x^{1/2}y^{1/2}(2\,\|g\|_{H_2^{1/2}}(\log_2 n)^{1/2} + \|g\|_\infty))\\
&\quad \ge \mathbb{P}(n^{1/2}\left|\mathbb{U}_n(g) - \mathbb{B}_n(g)\right| \ge C(\|g\|_\infty + \|g\|_{H_2^{1/2}})(t + \log n)(\log n)^{1/2}). \qquad \Box
\end{aligned}$$

LEMMA 5.2. *There is a C such that, for all $f \in \Sigma_n(f_0)$, $f_0 \in \Sigma$,*

$$\left\|\lambda_{f,f_0}\right\|_\infty \le C\gamma_n, \qquad \lambda_{f,f_0} \in \Lambda^\alpha(C).$$

PROOF. The first relation is obvious. For the second, note that F_0^{-1} has derivative $1/f(F_0^{-1}(\cdot))$, and since $f \ge \varepsilon$, we have $F_0^{-1} \in \Lambda^1(C)$. Now write λ_{f,f_0} as a difference of logarithms and again invoke $f \ge \varepsilon$. $\Box$

Next we have to consider the likelihood ratio for interval censored observations (18). We shall do this for a generic interval $D \subset [0,1]$ of length k_n^{-1}. We wish to represent the observations via the quantile function F_0^{-1} in the usual fashion; we therefore assume $D = F_0^{-1}(A)$, where $A \subset [0,1]$. Consider a class of intervals, for given $C_1, C_2 > 0$,

$$\mathfrak{A}_n = \{A\colon A = [a_1, a_2) \subset [0,1],\ C_1 \le k_n \operatorname{mes}(A) \le C_2\}. \tag{44}$$

The assumption $f_0 \in \Sigma$ implies that f_0 is uniformly bounded and bounded away from zero. Hence $\operatorname{mes}(D) = k_n^{-1}$ implies that $A = F_0(D) \in \mathfrak{A}_n$ for all $f_0 \in \Sigma$ and appropriately chosen C_1 and C_2. The technical development will now be carried out uniformly over all intervals $A \in \mathfrak{A}_n$. We shall set $P_f(F_0^{-1}(A)) = p$, $P_{f_0}(F_0^{-1}(A)) = p_0$. The corresponding log-likelihood ratio under f_0, expressed as a function of a uniform$[0,1]$ variable z, is then $\lambda_{f,f_0,A}(z)$, where

$$\lambda_{f,f_0,A}(t) = \chi_A(t)\log\frac{f}{f_0}(F_0^{-1}(t)) + (1 - \chi_A(t))\log\frac{1-p}{1-p_0}. \tag{45}$$

Since $\lambda_{f,f_0,A}$ has jumps at the endpoints of A, it is not in a Hölder class $\Lambda^\alpha(M)$; but it is in an L_2-Hölder class, so that we can ultimately estimate $\|\lambda_{f,f_0,A}\|_{H_2^{1/2}}$ and apply the KMT inequality of Proposition 2.3. We first need some technical lemmas.

LEMMA 5.3. *There is a C such that, for all $f \in \Sigma_n(f_0)$, $f_0 \in \Sigma$, $A \in \mathfrak{A}_n$,*

$$\sup_{t\in A} |\lambda_{f,f_0,A}(t)| \le C\gamma_n, \qquad \sup_{t\in A^c} |\lambda_{f,f_0,A}(t)| \le Ck_n^{-1}\gamma_n.$$

PROOF. For $t \in A$ we invoke the previous lemma. For $t \in A^c$ we estimate

$$\left|1 - \frac{p}{p_0}\right| \le \frac{\int_D |f - f_0|}{\int_D f_0} \le \frac{\int_D |f/f_0 - 1| f_0}{\int_D f_0} \le \gamma_n.$$

In view of (44) we also have $p_0 \asymp k_n^{-1} \le 1/2$; hence

$$\left|1 - \frac{1-p}{1-p_0}\right| = \frac{p_0}{1-p_0}\left|1 - \frac{p}{p_0}\right| \le Ck_n^{-1}\gamma_n. \tag{46}$$

This implies a similar estimate for $|\log((1-p)/(1-p_0))|$ and thus yields the estimate for $t \in A^c$. $\square$

LEMMA 5.4. *There is a constant C such that, for all $f \in \Sigma_n(f_0)$, $f_0 \in \Sigma$, $A \in \mathfrak{A}_n$,*

$$\int \lambda^2_{f,f_0,A} \le C\,(n\log^{1/2} n)^{-1}, \qquad -\int \lambda_{f,f_0,A} \le C\,(n\log^{1/2} n)^{-1}.$$

PROOF. From the previous lemma and (44) we obtain

$$\int \lambda^2_{f,f_0,A} = \int_A \lambda^2_{f,f_0,A} + \int_{A^c} \lambda^2_{f,f_0,A} \le Ck_n^{-1}\gamma_n^2 + Ck_n^{-2}\gamma_n^2 \le Ck_n^{-1}\gamma_n^2; \tag{47}$$

hence, in view of (6) and (19),

$$n\int \lambda^2_{f,f_0,A} \le Cnk_n^{-1}\gamma_n^2 \le C(\log n)^{-1/2}.$$

To prove the second relation, define $\varphi(t) = \exp\lambda_{f,f_0,A}(t)$; then $\int\varphi = 1$, and Lemma 5.3 implies $|\varphi(t) - 1| \le C\gamma_n$ uniformly. Hence

$$-n\int \lambda_{f,f_0,A} = -n\int \log\varphi \le n\int(1 - \varphi + C(\varphi-1)^2) = Cn\int(\varphi-1)^2.$$

Here we treat the r.h.s. analogously to (47), using the fact that Lemma 5.3 remains true with $\varphi - 1$ in place of λ, so that

$$n\int(\varphi-1)^2 \le C. \tag{48}$$

$\square$

LEMMA 5.5. *There is a C such that, for all $f \in \Sigma_n(f_0)$, $f_0 \in \Sigma$, $A \in \mathfrak{A}_n$,*

$$\|\lambda_{f,f_0,A}\|_{H_2^{1/2}} \le C\gamma_n.$$

PROOF. It suffices to show

$$\int_h^{1-h} (\lambda_{f,\,f_0,\,A}(x+h) - \lambda_{f,\,f_0,\,A}(x))^2\,dx \le C\gamma_n^2 h \quad \text{for } 0 < h < \tfrac{1}{2}. \tag{49}$$

Let $A = [a_1, a_2)$ and define $A_{1,h} = [a_1+h, a_2-h)$ and $A_{2,\,h} = [a_1-h, a_2+h)\cap [h, 1-h]$ (here $A_{1,\,h}$ is empty for $h > k_n/2$). The integral in (49) over $[h, 1-h]$ will be split into integrals over $A_{1,h}$, $A_{2,\,h}\setminus A_{1,\,h}$ and $[h, 1-h]\setminus A_{2,\,h}$. According to Lemma 5.2, $\lambda_{f,\,f_0,\,A}$ fulfills a Hölder condition on A and is bounded by $C\gamma_n$, so that

$$\int_{A_{1,\,h}} (\lambda_{f,\,f_0,\,A}(x+h) - \lambda_{f,\,f_0,,A}(x))^2 dx \le C\min(h^{2\alpha}, \gamma_n^2)k_n^{-1}.$$

We have $k_n^{-1} \sim \gamma_n^2 \log^{7/2} n$ in view of (6) and (19), so that for the above we obtain a bound $Ch\gamma_n^2\{\min(h^{2\alpha-1}, h^{-1}\gamma_n^2)\log^{7/2} n\}$. Since $\alpha > 1/2$, the expression in the curly brackets tends to zero uniformly in $0 < h < 1/2$, as can be seen by distinguishing the cases $h \le \gamma_n$ and $h \ge \gamma_n$. For the second integral, we use the estimate $\|\lambda_{f,\,f_0,\,A}\|_\infty \le C\gamma_n$ implied by Lemma 5.3 and obtain

$$\int_{A_{2,\,h}\setminus A_{1,\,h}} (\lambda_{f,\,f_0,\,A}(x+h) - \lambda_{f,\,f_0,\,A}(x))^2\,dx \le C\gamma_n^2 h.$$

Finally, note that $\lambda_{f,f_0,A}$ is constant on $[0,1]\setminus A$, so that

$$\int_{[h,\,1-h]\setminus A_{2,\,h}} (\lambda_{f,\,f_0,\,A}(x+h) - \lambda_{f,\,f_0,\,A}(x))^2 dx = 0.$$

Thus (49) is established. □

6. The local likelihood processes. Consider now the likelihood process for n observations (18) when D_j is replaced by the generic subinterval $D = F_0^{-1}(A)$ with $A \in \mathfrak{A}_n$ from (44). With n i.i.d. uniform(0, 1) variables z_i we get an expression for the likelihood process

$$\Lambda_{0,n}(f, f_0, A) = \exp\left\{\sum_{i=1}^n \lambda_{f,f_0,A}(z_i)\right\}; \tag{50}$$

for $A = F_0(D_j)$ this is the same as $\Lambda_{0,\,j,\,n}(f, f_0)$ as defined after (19). Let

$$K(f_0\|f, A) = -\int \lambda_{f,\,f_0,\,A}(t)\,dt$$

denote the pertaining Kullback information number. We assume that $\mathbb{U}_n$ and $\mathbb{B}_n$ are sequences of uniform empirical processes and Brownian bridges which both come from the Hungarian construction of Proposition 2.3. We obtain the representation [cf. (14) and Proposition 2.6, suppressing the notational distinction of versions]

$$\Lambda_{0,n}(f, f_0, A) = \exp\{n^{1/2}\,\mathbb{U}_n(\lambda_{f,\,f_0,\,A}) - n\;K(f_0\|f, A)\}. \tag{51}$$

The corresponding Gaussian likelihood ratio is [cf. (15)]

$$\Lambda_{1,n}(f, f_0, A) = \exp\left\{n^{1/2}\,\mathbb{B}_n(\lambda_{f,f_0,A}) - \frac{n}{2}\operatorname{Var}(\lambda_{f,f_0,A}(Z))\right\}. \tag{52}$$

Consider also an intermediary expression,

$$\Lambda_{\#,n}(f, f_0, A) = \exp\{n^{1/2}\,\mathbb{B}_n(\lambda_{f,f_0,A}) - n\ K(f_0\|f, A)\}.$$

The expression $\Lambda_{\#,n}(f, f_0, A)$ is not normalized to expectation 1, but we consider it as the density of a positive measure on the probability space $(\Omega, \mathscr{A}, \mathbb{P})$. The Hellinger distance $H^2(\cdot,\cdot)$ is then naturally extended to these positive measures.

LEMMA 6.1. *There is a C such that, for all $f \in \Sigma_n(f_0)$, $f_0 \in \Sigma$, $A \in \mathfrak{A}_n$,*

$$E_{\mathbb{P}}(\Lambda_{i,n}(f, f_0, A))^2 \le C, \qquad i = 0, 1, \qquad E_{\mathbb{P}}(\Lambda_{\#,n}(f, f_0, A))^2 \le C.$$

PROOF. Define [for a uniform $(0,1)$ variable Z]

$$T_{10} = n\ K(f_0\|f, A), \qquad T_{11} = \frac{n}{2}\operatorname{Var}(\lambda_{f,f_0,A}(Z)), \tag{53}$$

$$T_{20} = n^{1/2}\,\mathbb{U}_n(\lambda_{f,f_0,A}), \qquad T_{21} = n^{1/2}\,\mathbb{B}_n(\lambda_{f,f_0,A}). \tag{54}$$

Since T_{21} is a zero-mean Gaussian r.v., we have

$$E_{\mathbb{P}}\exp(2T_{21}) = \exp(4T_{11}).$$

Hence

$$E_{\mathbb{P}}\Lambda_{1,n}^2 = E_{\mathbb{P}}\exp(2(T_{21} - T_{11})) = \exp(2T_{11}) \le \exp\Big(n\int \lambda_{f,f_0,A}^2\Big).$$

Now from Lemma 5.4 we obtain the assertion for $i = 1$. For the case $i = 0$, we get, from (50),

$$E_{\mathbb{P}}\Lambda_{0,n}^2 = E_{\mathbb{P}}\exp\left\{2\sum_{i=1}^n \lambda_{f,f_0,A}(z_i)\right\} = (E\exp(2\lambda_{f,f_0,A}(Z)))^n.$$

Now we have for $\varphi(t) = \exp(\lambda_{f,f_0,A}(t))$

$$E\exp(2\lambda_{f,f_0,A}(Z)) = \int(\varphi(t))^2\,dt = 1 + \int(\varphi(t) - 1)^2\,dt \le 1 + Cn^{-1}$$

as a consequence of (48). Hence

$$E_{\mathbb{P}}\Lambda_{0,n}^2 \le (1 + Cn^{-1})^n \le 2\exp C,$$

so that the lemma is established for $i = 0$. Finally, to treat $E_{\mathbb{P}}\Lambda_{\#,n}^2$, observe that Lemma 5.4 implies that T_{10} and T_{11} are uniformly bounded. Hence

$$E_{\mathbb{P}}\Lambda_{\#,n}^2 = E_{\mathbb{P}}\Lambda_{1,n}^2\exp(2(T_{12} - T_{11})) \le C. \qquad \square$$

The next lemma is the key technical step, bringing in the Hungarian construction estimate of Proposition 2.3.

LEMMA 6.2. *There is a C such that, for all $f \in \Sigma_n(f_0)$, $f_0 \in \Sigma$, $A \in \mathfrak{A}_n$,*

$$H(\Lambda_{0,n}(f, f_0, A), \Lambda_{\#,n}(f, f_0, A)) \le C\,\gamma_n (\log n)^{3/2}.$$

PROOF. Define

$$T_0 = n^{1/2}(\mathbb{B}_n - \mathbb{U}_n)(\lambda_{f, f_0, A}).$$

Combining Proposition 2.3 with Lemmas 5.3 and 5.5, we obtain

$$\mathbb{P}(|T_0| \ge C\gamma_n (t + \log n)\log^{1/2} n) \le C\exp(-t).$$

Set $t = t_n = 4\log n$ and, for the above C,

$$u_n = 5C\gamma_n \log^{3/2} n.$$

For an event

$$B = B_{f, f_0, A} = \{\omega\colon\ |T_0| \le u_n\},$$

we obtain an estimate

$$\mathbb{P}(B^c) \le Cn^{-4}. \tag{55}$$

To treat $H^2(\Lambda_{0,n}, \Lambda_{\#,n})$, split the expectation there into $E_{\mathbb{P}}\chi_B(\cdot)$ and $E_{\mathbb{P}}\chi_{B^c}(\cdot)$, and observe

$$\begin{aligned} E_{\mathbb{P}}\chi_{B^c}(\Lambda_{0,n}^{1/2} - \Lambda_{\#,n}^{1/2})^2 &\le 2E_{\mathbb{P}}\chi_{B^c}(\Lambda_{0,n} + \Lambda_{\#,n}) \\ &\le 2\left(\mathbb{P}(B^c) 2E_{\mathbb{P}}(\Lambda_{0,n}^2 + \Lambda_{\#,n}^2)\right)^{1/2}. \end{aligned}$$

According to the previous lemma $E_{\mathbb{P}}(\Lambda_{0,n}^2 + \Lambda_{\#,n}^2)$ is uniformly bounded, so that (55) implies

$$E_{\mathbb{P}}\chi_{B^c}(\Lambda_{0,n}^{1/2} - \Lambda_{\#,n}^{1/2})^2 \le Cn^{-2}. \tag{56}$$

For the other part, observe that on $\omega \in B$, in view of $u_n = o(1)$,

$$\left|1 - \exp(T_0/2)\right| \le Cu_n,$$

so that, on $\omega \in B$,

$$(\Lambda_{0,n}^{1/2} - \Lambda_{\#,n}^{1/2})^2 = (1 - \exp(T_0/2))^2 \Lambda_{0,n} \le Cu_n^2 \Lambda_{0,n}.$$

Since $E_{\mathbb{P}}\Lambda_{0,n} = 1$, we obtain

$$E_{\mathbb{P}}\chi_B(\Lambda_{0,n}^{1/2} - \Lambda_{\#,n}^{1/2})^2 \le Cu_n^2.$$

This completes the proof in view of (56) and $n^{-2} = o(u_n^2)$. □

LEMMA 6.3. *For all $f \in \Sigma_n(f_0)$, $f_0 \in \Sigma$, $A \in \mathfrak{A}_n$,*

$$H(\Lambda_{0,n}(f, f_0, A), \Lambda_{1,n}(f, f_0, A)) \le 2H(\Lambda_{0,n}(f, f_0, A), \Lambda_{\#,n}(f, f_0, A)).$$

PROOF. Consider the space of random variables $L_2(\Omega, \mathscr{A}, \mathbb{P})$ and note that $H(\Lambda_{\#, n}, \Lambda_{1, n})$ is the distance of $\Lambda_{\#, n}^{1/2}$ and $\Lambda_{1, n}^{1/2}$ in that space. Furthermore,

$$\Lambda_{1, n}^{1/2} = \Lambda_{\#, n}^{1/2}(E_{\mathbb{P}}\Lambda_{\#, n})^{-1/2}$$

is the element of the unit sphere of $L_2(\Omega, \mathscr{A}, \mathbb{P})$ closest to $\Lambda_{\#, n}^{1/2}$. Since $\Lambda_{0, n}^{1/2}$ is on the unit sphere, we have

$$H(\Lambda_{\#, n}, \Lambda_{1, n}) \leq H(\Lambda_{\#, n}, \Lambda_{0, n})$$

and therefore

$$H(\Lambda_{0, n}, \Lambda_{1, n}) \leq H(\Lambda_{0, n}, \Lambda_{\#, n}) + H(\Lambda_{\#, n}, \Lambda_{1, n}) \leq 2H(\Lambda_{0, n}, \Lambda_{\#, n}). \qquad \square$$

Now let $A = A_j = F_0(D_j)$ and consider also the likelihood process $\Lambda_{1, j, n}(f, f_0)$ of the Gaussian experiment $\mathbb{E}_{1, j, n}(f_0)$ of (20). Notice that this differs from $\Lambda_{1, n}(f, f_0, A_j)$ [cf. (52) and (45)]. We consider versions of both likelihood processes which are functions of the Brownian bridge version $\mathbb{B}$.

LEMMA 6.4. *There is a C such that, for all $f \in \Sigma_n(f_0)$, $f_0 \in \Sigma$ and $j = 1, \dots, k_n$,*

$$H(\Lambda_{1, n}(f, f_0, A_j), \Lambda_{1, j, n}(f, f_0)) \leq C\gamma_n \log^{3/2} n.$$

PROOF. The likelihood process $\Lambda_{1, n}(f, f_0, A_j)$ is $\Lambda_{1, n}(f, f_0)$ from (15) with λ_{f, f_0} replaced by λ_{f, f_0, A_j}, so it corresponds to a Gaussian model

$$dy(t) = (\lambda_{f, f_0, A_j}(t) + K(f_0 \| f, A_j))\, dt + n^{-1/2}\, dW(t), \qquad t \in [0, 1]$$

with $f \in \Sigma_n(f_0)$ [cf. (16)]. Moreover, $\Lambda_{1, j, n}(f, f_0)$ corresponds to the Gaussian model (20). Hence the distance $H(\cdot, \cdot)$ between the likelihood processes on $(\Omega, \mathscr{A}, \mathbb{P})$ equals the Hellinger distance between the two respective shifted Wiener measures. We may apply (23), setting

$$g_1 = \lambda_{f, f_0, A_j} - \int \lambda_{f, f_0, A_j}, \qquad g_2 = \chi_{A_j}\left(\lambda_{f, f_0} - \int \lambda_{f, f_0}\right).$$

To estimate $g_1 - g_2$, we note that, in view of (13) and (45),

$$g_1 - g_2 = \chi_{A_j^c}\lambda_{f, f_0, A_j} - \tilde{g}, \quad \text{where } \tilde{g} = \int \lambda_{f, f_0, A_j} + \chi_{A_j}\int \lambda_{f, f_0}.$$

We claim

$$\|\tilde{g}\|_2^2 \leq Cn^{-1}\gamma_n^2 \tag{57}$$

uniformly. Indeed, using the expansion

$$\log x = \log(1 + x - 1) = x - 1 - \tfrac{1}{2}(x-1)^2 + o((x-1)^2) \tag{58}$$

and setting $x = (f/f_0) \circ F_0^{-1}(t)$ and $\lambda_{2, f, f_0} = (f/f_0) \circ F_0^{-1}(t) - 1$, we note that, for $f \in \Sigma_n(f_0)$,

$$\lambda_{f, f_0}(t) = \lambda_{2, f, f_0}(t) + O(\gamma_n^2) \tag{59}$$

uniformly. Since $\int \lambda_{2,f,f_0} = 0$, we obtain

$$(60)\quad \left|\int \lambda_{f,f_0}\right| = K(f_0\|f) = \int(\lambda_{2,f,f_0} - \lambda_{f,f_0}) \le \left\|\lambda_{2,f,f_0} - \lambda_{f,f_0}\right\|_2 = O(\gamma_n^2).$$

Furthermore, according to Lemma 5.4,

$$\left|\int \lambda_{f,f_0,A_j}\right|^2 \le Cn^{-2} = n^{-1}\gamma_n^4 \log^4 n.$$

In conjunction with (60) this implies

$$\|\tilde{g}\|_2^2 \le C(n^{-1}\gamma_n^4 \log^4 n + \operatorname{mes}(A_j)\gamma_n^4),$$

where $\operatorname{mes}(A_j) = p_0 = P_{f_0}(D_j)$. Using $p_0 \le Ck_n^{-1}$ and

$$k_n^{-1}\gamma_n^4 \sim n^{-1}(\log n)^{-1/2}\gamma_n^2,$$

we obtain (57). Furthermore,

$$\left\|g_1 - g_2 + \tilde{g}\right\|_2^2 = \left\|\chi_{A_j^c}\lambda_{f,f_0,A_j}\right\|_2^2 = (1-p_0)\log^2\frac{1-p}{1-p_0},$$

where $p = P_f(D_j)$. Using (46) we find

$$(61)\qquad \|g_1 - g_2 + \tilde{g}\|_2^2 \le Ck_n^{-2}\gamma_n^2 = Cn^{-1}\gamma_n^2 \log^3 n.$$

By (23) the squared Hellinger distance is

$$2\left(1 - \exp\left\{-\frac{n}{8}\|g_1 - g_2\|_2^2\right\}\right),$$

and the lemma follows from (57) and (61). $\square$

PROOF OF PROPOSITION 2.6. Consider $\Lambda_{0,n}(f, f_0, A)$ for $A = A_j$ and identify this with $\Lambda^*_{0,j,n}(f, f_0)$. Identify $\Lambda_{1,j,n}(f, f_0)$ of Lemma 6.4 to $\Lambda^*_{1,j,n}(f, f_0)$. The result then follows from Lemmas 6.2–6.4. $\square$

7. Further local approximations.

Define functions

$$\lambda_{1,f,f_0} = \lambda_{f,f_0} + K(f_0\|f), \qquad \lambda_{2,f,f_0} = (f/f_0 - 1)\circ F_0^{-1},$$
$$\lambda_{3,f,f_0} = 2((f/f_0)^{1/2} - 1)\circ F_0^{-1}$$

and experiments $\mathbb{E}^{\#}_{i,n}(f_0)$ given by observations

$$(62)\qquad dy(t) = \lambda_{i,f,f_0}(t)dt + n^{-1/2}\,dW(t), \qquad t \in [0,1],$$

and parameter space $f \in \Sigma_n(f_0)$, for $i = 1, 2, 3$. We have seen that $\mathbb{E}^{\#}_{1,n}(f_0) = \mathbb{E}_{1,n}(f_0)$ [cf. (16)].

LEMMA 7.1. *We have*

$$\Delta\left(\mathbb{E}^{\#}_{i,n}(f_0), \mathbb{E}_{i,n}(f_0)\right) = 0, \qquad i = 1, 2, 3.$$

PROOF. The likelihood process for $\mathbb{E}^{\#}_{i,n}(f_0)$ is

$$\Lambda_{i,n}(f, f_0) = \exp\left\{n \int \lambda_{i,f,f_0} \frac{1}{\sqrt{n}}\, dW - \frac{n}{2}\|\lambda_{i,f,f_0}\|_2^2\right\}, \qquad i = 1, 2, 3.$$

Define a process

$$W^*(t) = \int_0^t f_0^{-1/2}\, d(W \circ F_0).$$

This is a centered Gaussian process with independent increments and variance at t given by $\int_0^t f_0^{-1} dF_0 = t$. Hence W^* is a Wiener process, and we have, for every continuous g on $[0, 1]$,

$$\int g f_0^{1/2}\, dW^* = \int g\, d(W \circ F_0).$$

Utilizing W^* in (24), we get a likelihood process for $\mathbb{E}_{2,n}(f_0)$,

$$\begin{aligned}
&\exp\left\{n \int (f - f_0) f_0^{-1} n^{-1/2} f_0^{1/2}\, dW^* - \frac{n}{2}\int (f - f_0)^2 f_0^{-1}\right\}\\
&\quad = \exp\left\{n \int \left(\frac{f}{f_0} - 1\right) n^{-1/2}\, d(W \circ F_0) - \frac{n}{2}\int \left(\frac{f}{f_0} - 1\right)^2 dF_0\right\}\\
&\quad = \Lambda_{2,n}(f, f_0).
\end{aligned}$$

Similarly for $\mathbb{E}_{3,n}(f_0)$ we obtain a likelihood process,

$$\begin{aligned}
&\exp\left\{4n \int (f^{1/2} - f_0^{1/2}) \frac{1}{2} n^{-1/2}\, dW^* - \frac{4n}{2}\int (f^{1/2} - f_0^{1/2})^2\right\}\\
&\quad = \exp\left\{2n \int \left(\left(\frac{f}{f_0}\right)^{1/2} - 1\right) n^{-1/2}\, d(W \circ F_0) - \frac{4n}{2}\int \left(\left(\frac{f}{f_0}\right)^{1/2} - 1\right)^2 dF_0\right\}\\
&\quad = \Lambda_{3,n}(f, f_0). \qquad \square
\end{aligned}$$

PROOF OF THEOREM 2.7. It now remains to apply (23) to the measures given by (62) when $f \in \Sigma_n(f_0)$. We have to prove

$$\sup_{f \in \Sigma_n(f_0)} \|\lambda_{1,f,f_0} - \lambda_{i,f,f_0}\|_2^2 = o(n^{-1}) \tag{63}$$

for $i = 2, 3$, uniformly over $f_0 \in \Sigma$. Now (59) and (60) in the proof of Lemma 6.4 imply

$$\|\lambda_{f,f_0} + K(f_0\|f) - \lambda_{2,f,f_0}\|_2^2 = O(\gamma_n^4) = O(n^{-1}(\log n)^{-4}),$$

which proves (63) for $i = 2$. For $i = 3$, note first that for $f \in \Sigma_n(f_0)$ we have

$$\left\|(f/f_0)^{1/2} - 1\right\|_\infty = O(\gamma_n),$$

and use (58) with $x = (f/f_0)^{1/2} \circ F_0^{-1}(t)$ to obtain

$$\lambda_{f,f_0}(t) = 2\log(f/f_0)^{1/2} \circ F_0^{-1}(t) = \lambda_{3,f,f_0}(t) + O(\gamma_n^2) \tag{64}$$

uniformly. Now (64) and (60) imply (63) for $i = 3$. $\square$

8. The preliminary estimator. Consider first an estimator based on the whole sample. For attainable rates in the uniform norm we can employ the kernel-type estimators used in the more delicate exact constant theory. Let $\psi_n = (\log n/n)^{\alpha/(2\alpha+1)}$. The following result is shown in Korostelev and Nussbaum [(1996), Section 4].

LEMMA 8.1. *In the experiment* $\mathbb{E}_{0,n}$ *there is an estimator* $\tilde{f}_n$ *and a* $\kappa > 0$ *such that*

$$\sup_{f\in\Sigma} P_{n,f}\left(\left\|\tilde{f}_n - f\right\|_\infty \geq \kappa\, \psi_n\right) \to 0.$$

PROOF OF LEMMA 3.1. Consider the estimator applied to a sample fraction y_i, $i = 1, \ldots, N_n$; call it $\tilde{f}_{N_n}$. Then, since $\alpha > 1/2$,

$$\psi_{N_n} = (N_n^{-1} \log N_n)^{\alpha/(2\alpha+1)} \leq (n^{-1} \log(n/2) \log n)^{\alpha/(2\alpha+1)} = o(\gamma_n).$$

This immediately implies

$$\sup_{f\in\Sigma} P_{n,f}\left(\sup_{t\in[0,1]} \left|f(t) - \tilde{f}_{N_n}(t)\right| > c\gamma_n\right) \to 0 \quad \text{for all } c > 0. \tag{65}$$

It is easy to verify that the quantity

$$\mu_\Sigma = \sup_{f\in\Sigma} \|f\|_\infty \tag{66}$$

is finite; this is a consequence of Hölder continuity in conjunction with $\int |f| = 1$. Note that the set Σ is compact in the uniform metric: indeed it is equicontinuous and uniformly bounded according to (66), so compactness is implied by the Arzela–Ascoli theorem. Now cover Σ, by a finite set of uniform γ_n-balls with centers in Σ, and define $\Sigma_{0,n}$ to be the set of the centers. Define $\hat{f}_n$ as the element in $\Sigma_{0,n}$ closest to $\tilde{f}_{N_n}$ (or in case of nonuniqueness, select an element measurably). Analogously, for $f \in \Sigma$ select a closest element $g_f \in \Sigma_{0,n}$. Then we have

$$\begin{aligned}
\left\|\hat{f}_n - f\right\|_\infty &\leq \left\|\hat{f}_n - \tilde{f}_{N_n}\right\|_\infty + \left\|\tilde{f}_{N_n} - f\right\|_\infty \\
&\leq \left\|g_f - \tilde{f}_{N_n}\right\|_\infty + \left\|\tilde{f}_{N_n} - f\right\|_\infty \\
&\leq \left\|g_f - f\right\|_\infty + 2\left\|\tilde{f}_{N_n} - f\right\|_\infty \leq 2\left\|\tilde{f}_{N_n} - f\right\|_\infty + \gamma_n.
\end{aligned}$$

Hence $\hat{f}_n$ also satisfies (65), and it takes values in the finite set $\Sigma_{0,n} \subset \Sigma$. From this we obtain immediately

$$\sup_{f\in\Sigma} P_{n,f}\left(\sup_{t\in[0,1]} \left|f(t)/\tilde{f}_{N_n}(t) - 1\right| > \gamma_n\right) \to 0$$

in view of the uniform bound $f(t) \geq \varepsilon$ for $f \in \Sigma$. $\square$

For Lemma 3.4, we first consider estimation of the signal (rather than its root) in the white noise model. Again let $\psi_n = (\log n/n)^{\alpha/(2\alpha+1)}$.

LEMMA 8.2. *Consider an experiment given by observations*

$$dy(t) = g(t)\,dt + n^{-1/2}\,dW(t), \qquad t \in [0,1], \tag{67}$$

with $g \in \Lambda^\alpha(M)$. There one can find an estimator $\tilde{g}_n$ and a κ such that

$$\sup_{g \in \Lambda^\alpha(M)} P_{n,g}(\|\tilde{g}_n - g\|_\infty \ge \kappa\,\psi_n) \to 0.$$

The proof could be analogous to Lemma 8.1, with simplifications due to Gaussianity. Alternatively, we may refer to Korostelev (1993) or Theorem C in Donoho (1994), where sharper results (optimal constants) are obtained.

PROOF OF LEMMA 3.4. If $g = f^{1/2}$ with $f \in \Sigma$, then since $f \in \mathscr{F}_{\ge\varepsilon}$,

$$\left|f^{1/2}(t) - f^{1/2}(u)\right| \le \varepsilon^{-1/2}\left|f(t) - f(u)\right|,$$

so we obtain $g \in \Lambda^\alpha(\varepsilon^{-1/2}M)$. Also, by the previous argument we may assume that $\tilde{g}_n$ takes values in a finite subset of $\{f^{1/2}\colon f \in \Sigma\}$. On the other hand, if $\check{f}_n = \tilde{g}_n^2$, then

$$\left|\check{f}_n(t) - f(t)\right| \le \left|\tilde{g}_n(t) + g(t)\right|\left|\tilde{g}_n(t) - g(t)\right|.$$

Since both $\tilde{g}_n$ and g are in $\{f^{1/2}\colon f \in \Sigma\}$ they are uniformly bounded by $\mu_\Sigma^{1/2}$ [cf. (66)], so that, for some κ,

$$\sup_{f \in \Sigma} P_{n,f}\left(\left\|\check{f}_n - f\right\|_\infty \ge \kappa\,\psi_n\right) \to 0.$$

Finally, assume that $\check{f}_n$ is based on observations with noise intensity $(n - N_n)^{-1/2}$ instead of $n^{-1/2}$ [i.e., based on (38)]. Then $(n - N_n)^{-1/2} \le (n/2)^{-1/2}$ so that attainable rates are not worse. As in Lemma 3.1 we now infer that the estimator $\check{f}_n$ based on (38) fulfills (30). $\square$

9. Experiments and globalization. We collect some basic facts about experiments and deficiencies following Strasser (1985) ([S] henceforth). Let $\mathbb{E}_1 = (\Omega_1, \mathscr{A}_1, (P_{1,\vartheta}, \vartheta \in \Theta))$ be an experiment and let $L(\mathbb{E}_1)$ be the corresponding L-space (see [S], 41.4); $L(\mathbb{E}_1)$ is a certain subspace of the set of signed measures on $(\Omega_1, \mathscr{A}_1)$ which is a Banach lattice under the variation norm $\|\cdot\|$. Let $\mathbb{E}_2 = (\Omega_2, \mathscr{A}_2, (P_{2,\vartheta}, \vartheta \in \Theta))$ be another experiment with the same parameter set Θ with L-space $L(\mathbb{E}_2)$. A transition from $L(\mathbb{E}_1)$ to $L(\mathbb{E}_2)$ is a positive linear map with norm 1 [i.e., a linear map $M\colon L(\mathbb{E}_1) \mapsto L(\mathbb{E}_2)$ such that for $\sigma \in \mathbb{E}_1$, $\sigma \ge 0$, one has $M\sigma \ge 0$ and $\|M\sigma\| = \|\sigma\|$, cf. [S], 55.2]. Every Markov kernel $K\colon \Omega_1 \times \mathscr{A}_2 \mapsto [0,1]$ defines a transition. For the definition of the deficiency $\delta(\mathbb{E}_1, \mathbb{E}_2)$ of $\mathbb{E}_1$ with respect to $\mathbb{E}_2$ via decision problems, see [S], Section 59. An equivalent characterization is ([S], 59.6)

$$\delta(\mathbb{E}_1, \mathbb{E}_2) = \inf_M \sup_{\vartheta \in \Theta} \left\| MP_{1,\vartheta} - P_{2,\vartheta} \right\|, \tag{68}$$

where the infimum extends over all transitions from $L(\mathbb{E}_1)$ to $L(\mathbb{E}_2)$. The two-sided deficiency is

$$\Delta(\mathbb{E}_1, \mathbb{E}_2) = \max(\delta(\mathbb{E}_1, \mathbb{E}_2), \delta(\mathbb{E}_2, \mathbb{E}_1)).$$

This defines a pseudodistance on the set of all experiments with parameter space Θ; in particular, the triangle inequality holds ([S], 59.2). Experiments $\mathbb{E}_1$ and $\mathbb{E}_2$ are called equivalent (or of the same type) if $\Delta(\mathbb{E}_1, \mathbb{E}_2) = 0$.

We are interested in conditions under which every transition is given by a Markov kernel; [S], 55.6, (3), gives it for the case that $\mathbb{E}_1$ is dominated and Ω_2 is a locally compact space with countable base and $\mathscr{A}_2$ is its Borel σ-algebra. However, spaces like $C[0, 1]$ are not locally compact, so we would like to have the result for a complete separable metric (Polish) space instead. We briefly complete the argument.

DEFINITION 9.1. An experiment $\mathbb{E} = (\Omega, \mathscr{A}, (P_\vartheta, \vartheta \in \Theta))$ is called *Polish* if Ω is a Polish space and $\mathscr{A}$ is the pertaining Borel σ-algebra.

PROPOSITION 9.2. *Suppose that $\mathbb{E}_1$ is a dominated experiment and $\mathbb{E}_2$ is Polish. Then every transition from $L(\mathbb{E}_1)$ to $L(\mathbb{E}_2)$ is given by a Markov kernel.*

PROOF. It is well known that $(\Omega_2, \mathscr{A}_2)$ is Borel isomorphic to a subset of the unit interval [Dudley (1989), Lemma 13.1.3, and Parthasarathy (1978), Proposition 25.6]. This means that there is a one-to-one function φ from Ω_2 onto a Borel subset S of the unit interval such that φ and φ^{-1} are both measurable. It is clear that $\mathbb{E}_2$ is then equivalent to an experiment $\mathbb{E}_2^*$ given on the measurable space $(S, \mathscr{B}_S)$, and this equivalence is realized by Markov kernel transitions given by the mappings φ and φ^{-1}. Thus it suffices to prove the theorem for $\mathbb{E}_2 = \mathbb{E}_2^*$. We now refer to Remark 5.5.6(3) in [S]. $\square$

For the proof of Theorem 3.2 we formulate a lemma in an abstract framework. Let $\mathbb{X} = (X, \mathscr{X}, (P_\vartheta, \vartheta \in \Theta))$ be an experiment. Suppose also that there are a system of subsets $\Theta(\phi) \subset \Theta$, $\phi \in \Theta$, and experiments

$$\mathbb{F}_i(\phi) = (\Omega_i, \mathscr{A}_i, (Q_{i,\vartheta,\phi},\ \vartheta \in \Theta(\phi))), \qquad i = 1, 2,\ \phi \in \Theta.$$

Suppose further that there are a finite subset of $\Theta_0 \subset \Theta$ and an estimator $\hat{\phi}$: $(X, \mathscr{X}) \mapsto (\Theta_0, 2^{\Theta_0})$, and form Markov kernels

$$Q_{i,\vartheta}(x, A') = Q_{i,\vartheta,\hat{\phi}(x)}(A'), \qquad x \in X,\ A' \in \mathscr{A}_i,\ i = 1, 2.$$

Let $(\bar{X}_i, \bar{\mathscr{X}_i}) = (X \times \Omega_i, \mathscr{X} \times \mathscr{A}_i)$ be a product measurable space. For any Markov kernel K: $X \times \mathscr{A}_i \mapsto [0, 1]$ and a measure $\mu \mid \mathscr{X}$ we shall form the usual composed measure $\mu \otimes K \mid \bar{\mathscr{X}}_i$. Define measures $P_{i,\vartheta}|\bar{\mathscr{X}}_i = P_\vartheta \otimes Q_{i,\vartheta}|\bar{\mathscr{X}}_i$ and experiments $\mathbb{F}_i = (\bar{X}_i, \bar{\mathscr{X}}_i, (P_{i,\vartheta}, \vartheta \in \Theta))$, $i = 1, 2$.

LEMMA 9.3. *Suppose that for all $\phi \in \Theta$ the experiments $\mathbb{F}_i(\phi)$, $i = 1, 2$, are Polish and dominated, and*

$$\sup_{\phi \in \Theta} \Delta(\mathbb{F}_1(\phi), \mathbb{F}_2(\phi)) \leq \varepsilon. \tag{69}$$

Suppose also that the estimator $\hat{\phi}$ with values in Θ_0 fulfills

$$\inf_{\vartheta \in \Theta} P_\vartheta(\vartheta \in \Theta(\hat{\phi})) \geq 1 - \varepsilon. \tag{70}$$

Then

$$\Delta(\mathbb{F}_1, \mathbb{F}_2) \leq 4\varepsilon.$$

PROOF. Observe that since Θ_0 is finite and $\hat{\phi}$ is 2^{Θ_0}-measurable, the set $V_\vartheta = \{x\colon \vartheta \in \Theta(\hat{\phi}(x))\}$ is in $\mathscr{X}$. In accordance with Proposition 9.2, let $K_\phi(\omega_1, \cdot)$ be a Markov kernel realizing

$$\sup_{\vartheta \in \Theta(\phi)} \left\| Q_{2,\vartheta,\phi} - K_\phi Q_{1,\vartheta,\phi} \right\| \leq \delta(\mathbb{F}_1(\phi), \mathbb{F}_2(\phi)) + \varepsilon \leq 2\varepsilon$$

and define

$$M(\bar{x}, A) = \int_{\Omega_2} \chi_A(x, \omega_2) K_{\hat{\phi}(x)}(\omega_1, d\omega_2), \qquad \bar{x} = (x, \omega_1) \in \bar{X}_1,\ A \in \bar{\mathscr{X}}_2.$$

It is easy to see that M is a Markov kernel. Indeed by standard arguments this claim is reduced to the measurability of $K_{\hat{\phi}(x)}(\omega_1, A')$ in $\bar{x} = (x, \omega_1)$ for given $A' \in \mathscr{A}_2$, which again follows from the properties of $\hat{\phi}$. Now we have, for $A \in \bar{\mathscr{X}}_2$,

$$\begin{aligned} MP_{1,\vartheta}(A) &= \int_X \int_{\Omega_1} M(x, \omega_1, A) Q_{1,\vartheta}(x, d\omega_1) P_\vartheta(dx) \\ &= \int_X \int_{\Omega_2} \chi_A(x, \omega_2)(K_{\hat{\phi}(x)} Q_{1,\vartheta,\hat{\phi}(x)})(d\omega_2) P_\vartheta(dx). \end{aligned}$$

Hence

$$\begin{aligned} &\left|P_{2,\vartheta}(A) - MP_{1,\vartheta}(A)\right| \\ &\quad \leq 2P_\vartheta(V_\vartheta^c) + \int_{V_\vartheta} \left| \int_{\Omega_2} \chi_A(x, \omega_2)(K_{\hat{\phi}(x)} Q_{1,\vartheta,\hat{\phi}(x)} - Q_{2,\vartheta,\hat{\phi}(x)})(d\omega_2) \right| P_\vartheta(dx) \\ &\quad \leq 2P_\vartheta(V_\vartheta^c) + \sup_{\phi \in \Theta_0} \sup_{\vartheta \in \Theta(\phi)} \left\| K_\phi Q_{1,\vartheta,\phi} - Q_{2,\vartheta,\phi} \right\| \leq 4\varepsilon \end{aligned}$$

and we obtain

$$\delta(\mathbb{F}_1, \mathbb{F}_2) \leq \sup_{\vartheta \in \Theta} \left\| P_{2,\vartheta} - MP_{1,\vartheta} \right\| \leq 4\varepsilon.$$

The argument for $\delta(\mathbb{F}_2, \mathbb{F}_1)$ is similar. □

PROOF OF THEOREM 3.2. In the previous lemma we set $\vartheta = f$, $\phi = f_0$, $\Theta = \Sigma$, $\Theta(\phi) = \Sigma_n(f_0)$ and identify the experiment $\mathbb{X}$ with the one given by the sample fraction $y_1, \ldots, y_{N_n}$ (which may be written $\mathbb{E}_{0, N_n}$). Furthermore, $\mathbb{F}_1(\phi)$ is given by the second sample fraction with f restricted to a neighborhood $\Sigma_n(f_0)$ [which may be written $\mathbb{E}_{0, n-N_n}(f_0)$, cf. (8)]. Experiment $\mathbb{F}_2(\phi)$ is given by one of the three local experiments (27)–(29) in Remark 2.8 [we have seen that those are asymptotically or exactly equivalent to the respective $\mathbb{E}_{j,n}(f_0)$, $j = 1, 2, 3$ from Theorem 2.7]. Note that both $\mathbb{F}_i(\phi)$, $i = 1, 2$, are then Polish and dominated; in particular, $C_{[0,1]}$ is a Polish space [see Dudley (1989), Corollary 11.2.5]. The estimator $\hat{\phi}$ is taken to be $\hat{f}_n$ according to Lemma 3.1, and the finite set Θ_0 is the range of this estimator. To identify the global experiments $\mathbb{F}_i$ of the lemma, note that the measures in $\mathbb{F}_1(\phi)$ do not depend on ϕ [indeed, $\mathbb{F}_1(\phi) = \mathbb{E}_{0, n-N_n}(f_0)$ is obtained by just restricting the parameter space in $\mathbb{E}_{0, n-N_n}$]. Therefore $\mathbb{F}_1$ coincides with the experiment given by product measures $P_f^{\otimes N_n} \otimes P_f^{\otimes(n-N_n)}$, $f \in \Sigma$ (i.e., with $\mathbb{E}_{0,n}$). Experiment $\mathbb{F}_2$ coincides with $\mathbb{E}_{j,n}(\hat{f})$ as constructed; for $j = 3$, again $\mathbb{F}_2$ is given by a set of product measures $P_f^{\otimes N_n} \otimes Q_{3, n-N_n, f}$. Take ε arbitrary; then for sufficiently large n we achieve (69) by Theorems 2.1 and 2.7 (they were shown for sample size n, but since $n - N_n$ is of order n the argument remains valid for the now relevant diminished sample size). We achieve (70) by Lemma 3.1. We have shown $\Delta(\mathbb{E}_{0,n}, \mathbb{E}_{j,n}(\hat{f})) \leq 4\varepsilon$ for sufficiently large n, which proves the theorem. □

10. Addendum for Proposition 1.2. Let Σ' denote an arbitrary set of probability measures on $[0, 1]$. Define

$$S_n(\Sigma') = \{(P, Q) \in \Sigma' \times \Sigma' \colon H^2(P, Q) \leq n^{-1},\ P, Q \in \Sigma'\}.$$

and let dP/dQ be the R.N. derivative of the Q-continuous part of P. Le Cam's second regularity condition for Proposition 1.2 on the set of densities Σ is as follows: if Σ' is the associated set of probability measures, then

$$\sup_{(P,Q) \in S_n(\Sigma')} n(P+Q)\left(\left|\frac{dP}{dQ} - 1\right| \leq \varepsilon\right) \to 0.$$

This is fulfilled in the case $\Sigma' = (P_\vartheta, \vartheta \in K)$, where K is a compact subset of an open set $\Theta \subset \mathbb{R}^k$ and the family $(P_\vartheta, \vartheta \in \Theta)$ is differentiable in quadratic mean uniformly on compacts $K \subset \Theta$ [see Le Cam (1986), Proposition 1, Chapter 17.3].

Acknowledgments. The author wishes to thank David Donoho for encouraging discussions. Vladimir Koltchinskii, Mark Low and Enno Mammen suggested important improvements at various stages.

REFERENCES

BELITSER, E. and LEVIT, B. (1995). On minimax filtering over ellipsoids. *Math. Methods Statist.* **4** 259–273.

BROWN, L. D. and LOW, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398.

DONOHO, D. (1994). Asymptotic minimax risk (for sup-norm loss): solution via optimal recovery. *Probab. Theory Related Fields* **99** 145–170.

DONOHO, D. L. and JOHNSTONE, I. (1992). Minimax estimation via wavelet shrinkage. Unpublished manuscript.

DONOHO, D. L. and LOW, M. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20** 944–970.

DUDLEY, R. (1989). *Real Analysis and Probability*. Wadsworth & Brooks/Cole, Pacific Grove, CA.

EFROIMOVICH, S. YU. and PINSKER, M. S. (1982). Estimating a square integrable probability density of a random variable. *Problems Inform. Transmission* **18** 172–189.

FALK, M. and REISS, R.-D. (1992). Poisson approximation of empirical processes. *Statist. Probab. Lett.* **14** 39–48.

GOLUBEV, G. K. (1984). On minimax estimation of regression. *Problems Inform. Transmission* **20** 56–64. (In Russian.)

GOLUBEV, G. K. (1991). LAN in problems of nonparametric estimation of functions and lower bounds for quadratic risks. *Theory Probab. Appl.* **36** 152–157.

IBRAGIMOV, I. A. and KHASMINSKI, R. Z. (1977). On the estimation of an infinite dimensional parameter in Gaussian white noise. *Soviet Math. Dokl*. **236** 1053–1055.

KOLTCHINSKII, V. (1994). Komlos–Major–Tusnady approximation for the general empirical process and Haar expansions of classes of functions. *J. Theoret. Probab.* **7** 73–118.

KOROSTELEV, A. P. (1993). An asymptotically minimax regression estimate in the uniform norm up to an exact constant. *Theory Probab. Appl*. **38** 737–743.

KOROSTELEV, A. P. and NUSSBAUM, M. (1996). The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. Discussion paper, SFB 373, Humboldt Univ., Berlin.

LE CAM, L. (1985). Sur l'approximation de familles de mesures par des familles gaussiennes. *Ann. Inst. H. Poincaré* **21** 225–287.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

LE CAM, L. and YANG, G. (1990). *Asymptotics in Statistics*. Springer, New York.

LOW, M. (1992). Renormalization and white noise approximation for nonparametric functional estimation problems. *Ann. Statist.* **20** 545–554.

MAMMEN, E. (1986). The statistical information contained in additional observations. *Ann. Statist.* **14** 665–678.

MILLAR, P. W. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. Verw. Gebiete* **48** 233–252.

NIKOLSKIJ, S. M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems.* Springer, Berlin.

NUSSBAUM, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13** 984–997.

PARTHASARATHY, K. R. (1978). *Introduction to Probability and Measure*. Springer, New York.

PINSKER, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* **16** 120–133.

REISS, R.-D. (1993). *A Course on Point Processes.* Springer, New York.

RIO, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields* **98** 21–45.

SHORACK, G. and WELLNER, J. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

STRASSER, H. (1985). *Mathematical Theory of Statistics*. de Gruyter, Berlin.

TSYBAKOV, A. B. (1994). Efficient nonparametric estimation in L_2 with general loss. Unpublished manuscript.

WOODROOFE, M. (1967). On the maximum deviation of the sample density. *Ann. Math. Statist.* **38** 475–481.

WEIERSTRASS INSTITUTE
MOHRENSTRASSE 39
D-10117 BERLIN
GERMANY
E-MAIL: nussbaum@wias-berlin.de