The Annals of Statistics
1991, Vol. 19, No. 4, 1681–1724

THE 1989 WALD MEMORIAL LECTURES

ASYMPTOTICALLY NORMAL FAMILIES OF DISTRIBUTIONS AND EFFICIENT ESTIMATION

BY I. A. IBRAGIMOV AND R. Z. KHAS'MINSKII

Leningrad Branch of the Steklov Mathematical Institute and Institute of Information Transmission

1. Introduction. In parametric estimation theory a very important role is played by the notion of local asymptotic normality (LAN), an idea introduced by Le Cam [Le Cam (1953, 1956, 1960)]. In particular, this theory establishes very general lower bounds on the accuracy of estimates [Le Cam (1953, 1972) and Hájek (1972, 1970), theorems]. Below in all references to the LAN theory, we follow our treatment of the theory [Ibragimov and Has'minskii (1981)]. A different treatment can be found in Le Cam (1986), Chapters 7 and 8.

It seems that Levit was the first to understand the importance of the LAN concept for nonparametric estimation theory [see Levit (1974, 1975b)]. He also showed that the corresponding lower bounds can be attained in some infinite-dimensional estimation problems [Levit (1978)]. Further advances and generalizations of these results were obtained by Millar (1983). The first part of this chapter looks at the investigations of Levit and Millar from a new point of view and may be considered as an infinite-dimensional variant of Ibragimov and Has'minskii (1981), Chapter 2. We suggest a new [different from Levit (1978) or Millar (1983)] definition of LAN for families of distributions $\{P_\theta^{(\varepsilon)}, \theta \in \Theta\}$, where the parametric set Θ is a subset of a normed space (Section 2) or a smooth infinite-dimensional manifold (Section 5).

For families $\{P_\theta^{(\varepsilon)}, \theta \in \Theta\}$ satisfying the LAN condition with an infinite-dimensional parametric set Θ, we consider the following estimation problem. We would like to estimate the value $\phi(\theta)$ of a known (Euclid- or) Hilbert-valued function $\phi(\cdot)$ at an unknown point $\theta \in \Theta$ on the basis of observations $X^{(\varepsilon)}$ corresponding to the family $\{P_\theta^{(\varepsilon)}, \theta \in \Theta\}$. Although this is a rather nonparametric estimation problem, we may also consider it as a problem of specifying a plausible value for the parameter ϕ in the presence of an infinite-dimensional nuisance parameter θ. Instead, we may treat the problem as a semiparametric estimation problem [see Begun, Hall, Huang and Wellner (1983) and Wellner (1985)]. We prove under our LAN conditions a variant of Hájek's convolution theorem (Sections 3 and 5) and a variant of the Hájek–Le Cam asymptotic minimax bound. The latter result enables us to define the

Received May 1989; revised January 1991.
AMS 1980 *subject classifications*. Primary 62G20; secondary 62M10.
Key words and phrases. LAN, asymptotic efficiency, non- and semiparametric estimation.

notion of an asymptotically efficient estimator in the spirit of Ibragimov and Has'minskii (1981).

It is well known that it is difficult to construct asymptotically efficient estimators in situations where the parameter set is infinite dimensional [see, e.g., Geman and Hwang (1982), Gill (1986), Grenander (1987) and Kiefer and Wolfowitz (1956)]. The authors suggested a method of constructing such estimators in Has'minskii and Ibragimov (1979, 1980, 1986) and Ibragimov and Has'minskii (1977). We demonstrate in Sections 6 to 8 how this method works for a Gaussian white-noise estimation problem. These sections are a further development of the paper by Ibragimov, Nemirovskii and Has'minskii (1986).

We hope the notation used is rather standard and familiar to the reader. The symbols $\|\cdot\|$ and $(\cdot,\cdot)$ are used for norms and scalar products in different Banach or Hilbert spaces. If it is necessary, we denote the norms in the spaces B and H by $\|\cdot\|_B$ and $\|\cdot\|_H$. The projection operator which projects onto the space L is denoted by P_L. The letters C, C_i, c and c_i are used to denote constants or generic bounded quantities.

Part I.

2. Local asymptotic normality (LAN).

2.1. *Definition.* Consider a family $E_\varepsilon = \{X^{(\varepsilon)}, \Omega^{(\varepsilon)}, P_\theta^{(\varepsilon)}, \theta \in \Theta\}$ of statistical experiments and corresponding observations X_ε. The variables X_ε take their values in the measurable space $(X^{(\varepsilon)}, \Omega^{(\varepsilon)})$ and have distribution $P_\theta^{(\varepsilon)}$. We suppose that the parameter set Θ is a subset of a normed space $\mathbf{L}$. Denote by $dP_t^{(\varepsilon)}/dP_\theta^{(\varepsilon)}$ the derivative of the absolutely continuous component of the measure $P_t^{(\varepsilon)}$ with respect to the measure $P_\theta^{(\varepsilon)}$. For the sake of convenience we study asymptotic estimation problems through experiments indexed by ε and let $\varepsilon \to 0$. In fact, one may instead suppose that a filter is defined on the set $\{\varepsilon\}$ of indexes and consider limits with respect to this filter.

DEFINITION 2.1. A family $\{P_\theta^{(\varepsilon)}, \theta \in \Theta\}$ is called locally asymptotically normal at a point $\theta \in \Theta$ in the direction $H(\theta) = H$ with norming factors $A_\varepsilon(\theta) = A_\varepsilon$ if there exists a Hilbert space $\mathbf{H}$ with norm $\|\cdot\|$, a linear manifold $H_0 \subseteq \mathbf{H}$ with closure $\overline{H}_0 = H$ and a family $\{A_\varepsilon\}$ of linear operators $A_\varepsilon\colon H \to \mathbf{L}$ such that:

1. for any $h \in H_0$, $\lim\|A_\varepsilon h\| = 0$;
2. for any $h_1, \dots, h_n \in H_0$ and $c > 0$, there exists $\varepsilon(h_1, \dots, h_n, c) > 0$ such that, if $\varepsilon < \varepsilon(h_1, \dots, h_n, c)$ and $|x_i| < c$, all points
$$\theta + A_\varepsilon\left(\sum_1^n x_i h_i\right) \in \Theta;$$
3. for any $h \in H_0$ and $\varepsilon < \varepsilon(h)$, the representation
$$\frac{dP^{(\varepsilon)}_{\theta + A_\varepsilon(h)}}{dP_\theta}(X_\varepsilon) = \exp\left\{\Delta_\varepsilon(h) - \tfrac{1}{2}\|h\|_H^2 + \psi(\varepsilon, h)\right\}$$

is valid, where $\Delta_\varepsilon(h)$ is a linear random function on H_0 [i.e., $\Delta_\varepsilon(\alpha h_1 + \beta h_2) = \alpha\Delta_\varepsilon(h_1) + \beta\Delta_\varepsilon(h_2)$], for any $h \in H_0$ the variables $\Delta_\varepsilon(h)$ are asymptotically $N(0, \|h\|_H^2)$ and $\psi(\varepsilon, h) \to 0$ as $\varepsilon \to 0$ in $P_\theta^{(\varepsilon)}$-probability.

Note that in the definition, the objects H, A_ε and $\psi(\varepsilon, h)$ all depend on θ. Note also that the LAN conditions are determined by θ, H and $\{A_\varepsilon\}$. The choice of H_0 may be made rather arbitrarily and depends on the problem under investigation. We need only that the closure $\overline{H}_0 = H$.

It is important to realize that the structure of the space H depends on how large the set Θ is near the point θ. (A ball of very small radius can be very massive in a Hilbert space.) In fact, the space H describes (linear) restrictions imposed on Θ. It may happen that the family $\{P_\theta,\ \theta \in \Theta\}$ has a natural embedding into a larger family $\{P_\theta, \theta \in \Theta\}$ and satisfies LAN conditions in the direction $H' \supset H$ with norming operators $A'_\varepsilon = A_\varepsilon$ on H. The reader should not forget this possibility and not confuse H and A_ε with H' and A'_ε (see Example 2.1 below and Sections 3 and 4).

2.2. *Examples.* It is well known that the LAN conditions are fulfilled for many classical estimation problems with $\Theta \subseteq R^k$ [see, e.g., Ibragimov and Has'minskii (1981)] and so in the examples given below we consider only infinite-dimensional cases.

EXAMPLE 2.1. Let the experiment E_ε be generated by the observation

$$X_\varepsilon(t) = \int_0^t \theta(u)\,du + \varepsilon w(t), \qquad 0 \le t \le 1, \tag{2.1}$$

where w is a standard Wiener process and the parameter set $\Theta \subseteq L_2(0,1) = \mathbf{L}$ [see Ibragimov and Has'minskii (1981), page 345]. Let $P_\theta^{(\varepsilon)}$ be a probability measure on $C(0,1)$ generated by the observation (2.1). For any two points $\xi, \eta \in L_2(0,1)$ the measures $P_\xi^{(\varepsilon)}, P_\eta^{(\varepsilon)}$ are equivalent and the density is

$$\frac{dP_\xi^{(\varepsilon)}}{dP_\eta^{(\varepsilon)}}(X_\varepsilon) = \exp\left\{\frac{1}{\varepsilon^2}\int_0^1(\xi(u) - \eta(u))\,dX_\varepsilon(u) - \frac{1}{2\varepsilon^2}\left(\int_0^1|\xi(u)|^2\,du - \int_0^1|\eta(u)|^2\,du\right)\right\}$$

[see, e.g., Ibragimov and Has'minskii (1981), page 385].

Let H be a subspace of $L_2(0,1)$. Suppose that a point $\theta \in \Theta$ is such that for any $h \in H$ and all sufficiently small ε the points $\theta + \varepsilon h \in \Theta$. We then have

$$\frac{dP_{\theta+\varepsilon h}^{(\varepsilon)}}{dP_\theta^{(\varepsilon)}}(X_\varepsilon) = \exp\left\{\Delta_\varepsilon(h) - \frac{1}{2}\|h\|^2\right\},$$

where $\Delta_\varepsilon(h) = \int_0^1 h(u)\,dw(u)$. Hence the LAN conditions are fulfilled in the direction H with norming factors $A_\varepsilon = \varepsilon I$. (Here and below, I denotes an identity operator.) In this example the remainder term $\psi(\varepsilon, h)$ is equal to 0.

In this example, it turned out that the measures $P_\theta^{(\varepsilon)}$ may be defined for all $\theta \in L_2(0,1)$ and, for such an enlarged family $\{P_\theta^{(\varepsilon)},\ \theta \in L_2(0,1)\}$, the LAN conditions are fulfilled in the direction of the maximal possible space $L_2(0,1)$. The space $H \subseteq L_2(0,1)$ reflects the restrictions imposed on Θ. For example, if Θ is an open subset of $L_2(0,1)$, then $H = L_2(0,1)$. If Θ consists of functions $\theta(u)$ which satisfy the conditions

$$\int_0^1 \theta(u)\,du = \int_0^1 u\theta(u)\,du = \cdots = \int_0^1 u^{k-1}\theta(u)\,du = 0,$$

then H is a proper subspace of $L_2(0,1)$ of codimension k. If Θ consists of functions $\theta(u) = \sum_0^{k-1}\theta_j \cos(2\pi ju)$, then H has dimension k.

EXAMPLE 2.2. Let the observation $X_\varepsilon = (X_1,\dots,X_n)$ consist of n iid random variables. Suppose that the X_j have a common probability density $\theta(x)$ with respect to a σ-finite measure μ defined on the measurable space $(\mathbf{X},\Omega)$. Take $\mathbf{L}$ to be the space $L(\mathbf{X},\mu)$ of all functions g integrable with respect to μ with norm $\|g\|_{\mathbf{L}} = \int |g|\,d\mu$. Let the parameter set Θ consist of all densities θ and the Hilbert space $\mathbf{H}$ be the space $L_2(\mathbf{X},\mu)$. Define a linear manifold H_0 as the set of all bounded functions $h \in \mathbf{L}$ such that:

(i) $\int_{\mathbf{X}} h(x)\sqrt{\theta(x)}\,\mu(dx) = 0$;
(ii) there exists an integer $k = k(h)$ such that $h(x) = 0$ if $\theta(x) < 1/k$.

The space H is the closure of H_0 in $\mathbf{H}$. Define operators A_ε by $A_\varepsilon = n^{-1/2}A$, where A is the operator of multiplication by the function $\sqrt{\theta}$ and $\varepsilon = n^{-1/2}$.

We now prove that the LAN conditions are satisfied in the direction H with norming factors A_ε. Indeed, if $h \in H_0$ and ε is sufficiently small (i.e., n is sufficiently large), then

$$\theta + A_\varepsilon h = \theta + n^{-1/2}\theta h \ge 0, \qquad \int_{\mathbf{X}} (\theta + A_\varepsilon h)\,d\mu = 1.$$

Hence, for such ε, all functions $\theta + A_\varepsilon h \in \Theta$. Furthermore,

$$\begin{aligned}\frac{dP^{(\varepsilon)}_{\theta + A_\varepsilon h}}{dP^{(\varepsilon)}_\theta}(X_\varepsilon) &= \prod_1^n \left(1 + n^{-1/2}h(X_j)\theta^{-1/2}(X_j)\right)\\ &= \exp\left\{ n^{-1/2}\sum_1^n \frac{h(X_j)}{\sqrt{\theta(X_j)}} - \frac{1}{2n}\sum_1^n \frac{h^2(X_j)}{\theta(X_j)} + r_n \right\}.\end{aligned}$$

Here the function

$$\Delta_\varepsilon(h) = n^{-1/2}\sum_1^n \frac{h(X_j)}{\sqrt{\theta(X_j)}}$$

is linear and asymptotically normal with mean $\int h(x)\sqrt{\theta(x)}\,d\mu = 0$ and variance $\|h\|^2$. By the law of large numbers, the sum $n^{-1}\sum_1^n h^2(X_j)/\theta(X_j)$ converges to $\|h\|^2$ in $P_\theta^{(\varepsilon)}$-probability. The proof is completed by noting that the remainder r_n goes to 0 in $P_\theta^{(\varepsilon)}$-probability.

Note that the space H consists of all $h \in L_2(\mathbf{X}, \mu)$ such that

$$\int_{\mathbf{X}} h(x)\sqrt{\theta(x)}\,d\mu = 0, \qquad \operatorname{supp} h = \operatorname{supp} \theta.$$

EXAMPLE 2.3. Now suppose that the observation $X_\varepsilon = (X_1, \dots, X_n)$ is a segment of length n from a real stationary Gaussian sequence with unknown mean value a and spectral density $f(\lambda)$. Elements of the parametric set Θ are points of the form $\theta = (a, f)$. We consider Θ as a subset of the Hilbert space $\mathbf{L}$ consisting of the points (x, g), $x \in R^1$, $g \in L_2(-\pi, \pi)$, with the norm $(|x|^2 + \int_{-\pi}^{\pi}|g|^2\,d\lambda)^{1/2}$. [In other words, $\mathbf{L} = R^1 \oplus L_2(-\pi, \pi)$.] Moreover, suppose that if $\theta = (a, f) \in \Theta$, then (i) $a \in (\alpha, \beta) \subset R^1$ and (ii) the function $f(\lambda)$ is strictly positive and continuous. Take the Hilbert space $\mathbf{H} = \mathbf{L}$. Define H_0 to be the linear manifold in $\mathbf{H}$ consisting of all points (x, g), where $x \in R^1$ and g is continuous on $[-\pi, \pi]$. Evidently, the closure $\overline{H}_0 = H = \mathbf{H} = \mathbf{L}$. Define the operators $A_\varepsilon(\theta)$: $H \to H$ by the equation $[\theta = (a, f)]$:

$$A_\varepsilon(x, g) = \varepsilon\left(x\sqrt{f(0)}\,, gf\right), \qquad \varepsilon = \frac{\sqrt{2\pi}}{\sqrt{n}}.$$

We now give an outline of a proof that the family $\{P_\theta^{(\varepsilon)},\ \theta \in \Theta\}$ satisfies the LAN condition at any point $\theta \in \Theta$ in the direction H with norming factors A_ε. A more complete proof will be published elsewhere. First note that, for all $h \in H_0$ and for all sufficiently small ε (large n), $\theta + A_\varepsilon h \in \Theta$.

Denote by $R_n(f)$ the correlation matrix of the vector X_ε which has spectral density f. Then for $\theta = (a, f)$, $h = (\alpha, g)$,

$$\begin{aligned}
&\frac{dP_{\theta+A_\varepsilon h}^{(\varepsilon)}}{dP_\theta^{(\varepsilon)}}(X_\varepsilon)\\
&\quad= \exp\left\{\frac{1}{2}\ln\frac{\det R(f+\varepsilon gf)}{\det R(f)} - \frac{1}{2}\left(\left(R^{-1}(f+\varepsilon gf) - R^{-1}(f)\right)Y_\varepsilon, Y_\varepsilon\right)\right.\\
&\qquad\left. + \varepsilon\alpha\sqrt{f(0)}\left(R^{-1}(f+\varepsilon gf)Y_\varepsilon, \mathbf{1}\right) - \frac{\varepsilon^2\alpha^2}{2}f(0)\left(R^{-1}(f+\varepsilon gf)\mathbf{1}, \mathbf{1}\right)\right\}.
\end{aligned}$$

Here $\mathbf{1}$ denotes the vector $(1, 1, \dots, 1)$ and $Y_\varepsilon = X_\varepsilon - a\mathbf{1}$.

Now note that the summand under the exp sign does not contain α and may be written in the form [see Hannan (1970)]:

$$\xi_n = \frac{1}{2\pi\sqrt{n}}\int_{-\pi}^{\pi}\frac{I_n(\lambda) - f(\lambda)}{f(\lambda)}g(\lambda)\,d\lambda - \frac{1}{2\pi n}\int_{-\pi}^{\pi}\frac{I_n(\lambda)}{f(\lambda)}g^2(\lambda)\,d\lambda + r_n,$$

$$I_n(\lambda) = \left|\sum_1^n e^{ij\lambda}Y_j\right|^2,$$

and so when $n \to \infty$ the first term in the summand which defines ξ_n is asymptotically $N(0, \|g\|^2)$, the second one converges in probability to $\|g\|^2/2$ and $r_n \to 0$ in probability.

Furthermore, note that $\varepsilon(R_n^{-1}(f + \varepsilon g f)Y_\varepsilon, \mathbf{1})$ is Gaussian with mean 0 and variance asymptotically equal to $\varepsilon^2(R_n^{-1}(f + \varepsilon g f)\mathbf{1}, \mathbf{1})$.

It is also possible to prove that

$$\varepsilon^2\left|\left(R_n^{-1}(f + \varepsilon g f)\mathbf{1}, \mathbf{1}\right) - \frac{1}{4\pi^2}\left(R_n\left((f + \varepsilon g f)^{-1}\right)\mathbf{1}, \mathbf{1}\right)\right| = o(1).$$

Hence

$$\begin{aligned}\varepsilon^2\left(R_n^{-1}(f + \varepsilon g f)\mathbf{1}, \mathbf{1}\right) &= \frac{1}{2\pi n}\int_{-\pi}^{\pi}\frac{\sin^2(n\lambda/2)}{\sin^2(\lambda/2)}\frac{d\lambda}{f(\lambda)(1 + \varepsilon g(\lambda))}\\ &= (f(0))^{-1} + o(1).\end{aligned}$$

Finally, to complete the proof, we need only note that the linear part and the quadratic part are asymptotically independent.

EXAMPLE 2.4 (Regression problem). Suppose we observe

$$X_i = \theta(t_i) + \xi_i(t_i), \qquad i = 1, \dots, n,$$

where the design $t^n = (t_1, \dots, t_n)$ is a sequence of independent random variables with a common density function $\rho(t)$ on $[0, 1]$. When the plan t^n is fixed the random variables $\xi_i(t_i)$ are iid with common density function f which has finite Fisher information $\mathbf{T}(t)$, $t \in [0, 1]$. Suppose that the unknown parameter $\theta \in \Theta = \mathbf{L} = L_2(0, 1)$. Then the LAN conditions are satisfied in the direction $H = L_2(0, 1)$ with norming operators $A_\varepsilon = \varepsilon A$, $\varepsilon = n^{-1/2}$, where A is the operator of multiplication by the function $(\mathbf{T}(t)\rho(t))^{-1/2}$.

A formal proof is based on an analysis of

$$\frac{dP^{(\varepsilon)}_{\theta + A_\varepsilon h}}{dP^{(\varepsilon)}_\theta}(X^\varepsilon) = \prod_1^n \frac{f\left(\theta(t_i) + \varepsilon\left(\rho(t_i)\mathbf{T}(t_i)\right)^{-1.2} h(t_i) + \xi_i(t_i), t_i\right)}{f(\theta(t_i) + \xi_i(t_i), t_i)}.$$

We omit the details since the necessary calculations coincide with those which have been done in the finite parametric case [see Ibragimov and Has'minskii (1981), Chapter 2].

EXAMPLE 2.5. Let the observation $X_\varepsilon = \theta + \varepsilon Z$, where $\theta \in \Theta \subseteq \mathbf{L}$, $\mathbf{L}$ is a Hilbert space and Z is a Gaussian $\mathbf{L}$-valued random variable with mean 0 and correlation operator R. Suppose that $\Theta \subseteq D(R^{-1/2})$ and that, for all $\xi \in R^{-1/2}\Theta$, all $h \in H \subseteq \mathbf{H} = \mathbf{L}$ and all sufficiently small ε, the vector $\xi + \varepsilon h \in R^{-1/2}\Theta$. We claim that in this case the LAN condition is satisfied in the direction H with norming factors $A_\varepsilon = \varepsilon R^{1/2}$. Indeed, the condition $\Theta \subseteq D(R^{-1/2})$ guarantees that the measures $P^{(\varepsilon)}_\theta$, $\theta \in \Theta$, are absolutely continuous with respect to each other [see Skorokhod (1974), Chapter 3].

Simple calculations show that the derivative $(dP^{(\varepsilon)}_{\theta+A_\varepsilon h}/dP^{(\varepsilon)}_\theta)(X_\varepsilon)$ may be written as $\exp\{(h, R^{-1/2}Z) - \frac{1}{2}\|h\|^2\}$ and that the random function $\Delta(h) = (h, R^{-1/2}Z)$ is $N(0, \|h\|^2)$ and linear with respect to h [for an accurate definition of $(h, R^{-1/2}Z)$, see Skorokhod (1974), Chapter 3]. One can prove that this result is also valid for generalized Gaussian variables Z whose correlation operator is supposed only symmetric and positive. In this case the observation is a collection of linear functionals $(X^\varepsilon, \phi) = (\theta, \phi) + \varepsilon(Z, \phi)$, $\phi \in \Phi$, such that for $\phi_i \in D(R)$ the vectors $(Z, \phi_1), \dots, (Z, \phi_K)$ are joint normal with mean 0 and covariance matrix $\|(R\phi_i, \phi_j)\|$. Example 2.1 is the special case when R is the identity operator [cf. Ibragimov and Has'minskii (1987)].

EXAMPLE 2.6 (Diffusion process). Let the observation X_ε be a diffusion process defined by

$$X_\varepsilon(t) = x_0 + \int_0^t \theta(u)\,du + \varepsilon \int_0^t \sigma(X_\varepsilon(u), u)\,dw(u), \qquad 0 \le t \le j, \tag{2.2}$$

with known x_0 and σ and an unknown parameter $\theta \in \Theta \subseteq \mathbf{L} = L_2(0,1)$. We suppose that $\sigma > 0$ and that σ satisfies conditions which guarantee that a unique solution of equation (2.2) exists. Let $H_0 \subseteq \mathbf{L}$ be a linear manifold. Suppose that for any $h \in H_0$ and all sufficiently small ε the point $\theta + \varepsilon h \in \Theta$. We claim that the LAN conditions are satisfied in the direction H which is the closure of H_0 in $\mathbf{L}$ with norming factors $A_\varepsilon = \varepsilon A$, where A is the operator of multiplication by the function $\sigma(x_0(t,\theta), t)$, $x_0(t,\theta) = x_0 + \int_0^t \theta(u)\,du$.

Indeed [see Gikhman and Skorokhod (1979)],

$$\frac{dP^{(\varepsilon)}_{\theta+A_\varepsilon h}}{dP^{(\varepsilon)}_\theta}(X_\varepsilon) = \exp\left\{ \int_0^1 \frac{\sigma(x_0(t,\theta),t)}{\sigma(X_\varepsilon(t),t)} h(t)\,dw(t) \right.$$
$$\left. - \frac{1}{2}\int_0^1 \left| \frac{\sigma(x_0(t,\theta),t)}{\sigma(X_\varepsilon(t),t)} h(t) \right|^2 dt \right\}.$$

When $\varepsilon \to 0$, the first summand under the exp sign goes to $\int_0^1 h\,dw$, and the second one converges to $\|h\|^2/2$. The special case of $x_0 = 0$ and $\sigma \equiv 1$ was given in Example 2.1.

One can easily restate the results of this example for the case of m-dimensional diffusions.

2.3. *Some properties of distributions satisfying the LAN conditions.* The properties we are going to describe here coincide with those of one-dimensional parametric sets Θ and can be proved in the same way. For this reason we refer the reader to the proofs given in Ibragimov and Has'minskii (1981).

Let $P^{(\varepsilon)}_{u,s}$ denote the singular component of the measure $P^{(\varepsilon)}_u$ with respect to the measure $P^{(\varepsilon)}_\theta$. Define the function $Z_\varepsilon(h)$ by

$$Z_\varepsilon(h) = \frac{dP^{(\varepsilon)}_{\theta+A_\varepsilon h}}{dP^{(\varepsilon)}_\theta}(X_\varepsilon)$$

(as we have mentioned earlier we denote by $dP_u^{(\varepsilon)}/dP_\theta^{(\varepsilon)}$ the derivative of the absolutely continuous component of the measure $P_u^{(\varepsilon)}$ with respect to $P_\theta^{(\varepsilon)}$).

LEMMA 2.1. *If the family $\{P_\theta^{(\varepsilon)},\ \theta \in \Theta\}$ satisfies the LAN condition at a point θ, then for all $h \in H_0$,*

$$\lim_{\varepsilon \to 0} E_\theta^{(\varepsilon)} Z_\varepsilon(h) = 1, \qquad \lim_{\varepsilon \to 0} \operatorname{Var} P_{\theta+A_\varepsilon h,s}^{(\varepsilon)} = 0.$$

The proof coincides with that of Lemma 8.1 from Ibragimov and Has'minskii (1981), page 147.

We set for an arbitrary random variable ξ and positive constant a,

$$\hat{\xi} = \begin{cases} \xi, & |\xi| \le a, \\ 0, & |\xi| > a. \end{cases}$$

We refer to $\hat{\xi}$ as the a-truncation of ξ.

THEOREM 2.1. *If a family $\{P_\theta^{(\varepsilon)},\ \theta \in \Theta\}$ satisfies the LAN condition at a point θ, then for any $h \in H_0$ there exists an a_ε-truncation $\hat{\Delta}_\varepsilon(h)$ of $\Delta_\varepsilon(h)$ such that*

$$\hat{\Delta}_\varepsilon(h) - \Delta_\varepsilon(h) \to 0 \quad \textit{in } P_\theta^{(\varepsilon)}\textit{-probability as } \varepsilon \to 0,$$

and the random field

$$\hat{Z}_\varepsilon(h) = \exp\{\hat{\Delta}_\varepsilon(h) - \tfrac{1}{2}\|h\|^2\}$$

possesses the following properties: For any $h \in H_0$,

$$\begin{aligned} E_\theta^{(\varepsilon)} \hat{Z}_\varepsilon(h) &\to 1, \\ E_\theta^{(\varepsilon)} \big|\hat{Z}_\varepsilon(h) - Z_\varepsilon(h)\big| &\to 0 \quad \textit{as } \varepsilon \to 0. \end{aligned} \tag{2.3}$$

The proof coincides with the proof of Theorem 8.1 from Ibragimov and Has'minskii (1981), page 149.

Note that as $\varepsilon \to 0$,

$$P_\theta^{(\varepsilon)}\{\hat{\Delta}_\varepsilon(h_1 + h_2) = \hat{\Delta}_\varepsilon(h_1) + \hat{\Delta}_\varepsilon(h_2)\} \to 1,$$
$$P_\theta^{(\varepsilon)}\{\hat{\Delta}_\varepsilon(\alpha h) = \alpha \hat{\Delta}_\varepsilon(h)\} \to 1.$$

Taking into account the last relation we shall not always differentiate between $\hat{\Delta}_\varepsilon(\alpha h)$ and $\alpha \hat{\Delta}_\varepsilon(h)$.

3. Characterization of limiting distributions of estimators under the LAN condition.

3.1. We shall now study the problem of estimating the value $\phi(\theta)$ of a known function $\phi(\cdot)$ at an unknown point $\theta \in \Theta$ on the basis of an observation X_ε which has distribution $P_\theta^{(\varepsilon)}$. We shall assume that the family $\{P_\theta^{(\varepsilon)}, \theta \in \Theta\}$ satisfies the LAN conditions at the point θ in the direction H with norming factors A_ε. The function ϕ: $\mathbf{L} \to U$ takes its values in a Hilbert space U. We shall also assume that this function is Fréchet differentiable with derivative $\phi'(\theta)$.

The main purpose of the present and the next section is to study the limiting behavior of estimators T_ε of $\phi(\theta)$. Since $T_\varepsilon - \phi(\theta) \to 0$, as $\varepsilon \to 0$ for any reasonable estimator T_ε we have to norm the difference $T_\varepsilon - \phi(\theta)$ in a proper way. We shall norm these differences by bounded linear operators B_ε: $U \to U$ with norms $\|B_\varepsilon\| \to \infty$ as $\varepsilon \to 0$.

Let T_ε be an estimator of $\phi(\theta)$. Suppose that one can find norming operators B_ε in such a way that the normed difference has a proper limiting distribution as $\varepsilon \to 0$. We show that under some regularity conditions this limit distribution is necessarily a convolution of a normal distribution and some other probability distribution. This phenomenon was first discovered by Hájek. He proved the corresponding theorems for finite-dimensional Θ and $\phi(\theta) = \theta$ [Hájek (1970); see also Ibragimov and Has'minskii (1981), Section 2.9]. The first infinite-dimensional variant of this characterization theorem was proved by Beran (1977) who studied the classical problem of estimating a distribution function. General theorems for infinite-dimensional Θ were proved by Millar (1979, 1983).

We call *an estimator T_ε-regular at the point θ with respect to the triple* $(H_0, A_\varepsilon, B_\varepsilon)$, or $(H_0, A_\varepsilon, B_\varepsilon)$*-regular, if for any $h \in H_0$ there exists a proper limit distribution F of the differences $\Lambda_\varepsilon = B_\varepsilon(T_\varepsilon - \phi(\theta + A_\varepsilon h))$ as $\varepsilon \to 0$ and this limit distribution does not depend on h*:

$$\mathbf{L}\left\{B_\varepsilon\left(T_\varepsilon - \phi(\theta + A_\varepsilon h)\right)\middle|P_{t+A_\varepsilon h}^{(\varepsilon)}\right\} \to F.$$

Denote by P_L the projection in $\mathbf{H}$ onto a subspace L. Define the operators K_ε by

$$K_\varepsilon\colon h \to K_\varepsilon h = B_\varepsilon \int_0^1 \phi'(\theta + tA_\varepsilon P_H h)\,dt\,A_\varepsilon P_H h.$$

Since by the definition of the LAN conditions the operators A_ε are defined on H, the presence of P_H in the definition of K_ε may seem unnecessary. However, as we have mentioned the operators A_ε may often be defined on the whole space $\mathbf{H}$ allowing different restriction spaces H to be treated at the same time.

THEOREM 3.1. *Suppose the family $\{P_\theta^{(\varepsilon)}, \theta \in \Theta\}$ satisfies the LAN conditions at a point θ in a direction $H = \overline{H}_0$ with norming factors A_ε and that T_ε is a $(B_\varepsilon, H_0, A_\varepsilon)$-regular estimate of $\phi(\theta)$. Then if the operators K_ε have a weak limit $K = \lim_{\varepsilon \to 0} K_\varepsilon$ and if the operator K is a Hilbert–Schmidt operator, the limiting distribution F of the difference $B_\varepsilon(T_\varepsilon - \phi(\theta))$ is a convolution $\mathbf{N}(0, K)*G$. Here $\mathbf{N}(0, K)$ is a normal distribution on U with mean 0 and correlation operator KK^* and G is a distribution on U.*

PROOF. Let $f(u)$ be the characteristic function of the distribution F. Since the estimator T_ε is regular, it follows that for any $h \in H_0$,

$$\begin{aligned} f(u) &= \lim_{\varepsilon \to 0} E^{(\varepsilon)}_{\theta + A_\varepsilon h} \exp\{i(u, B_\varepsilon(T_\varepsilon - \phi(\theta + A_\varepsilon h)))\} \\ &= e^{-i(u, Kh)} \lim_{\varepsilon \to 0} E^{(\varepsilon)}_{\theta + A_\varepsilon h} \exp\{i(u, B_\varepsilon(T_\varepsilon - \phi(\theta)))\}. \end{aligned}$$

The LAN conditions, Lemma 2.1 and Theorem 2.1 imply that

$$\begin{aligned} &f(u)\exp\{i(u, Ku) + \tfrac{1}{2}\|h\|^2\} \\ &\quad = \lim_{\varepsilon \to 0} E^{(\varepsilon)}_{\theta} \exp\{i(u, B_\varepsilon(T_\varepsilon - \phi(\theta))) + \hat{\Delta}_\varepsilon(h)\}. \end{aligned}$$

Define for a complex number Z,

$$\zeta_\varepsilon(u, Z) = E^{(\varepsilon)}_{\theta}\{\exp[i(u, B_\varepsilon(T_\varepsilon - \phi(\theta))) + Z\hat{\Delta}_\varepsilon(h)]\}.$$

It follows from Section 3.2 that for real Z,

$$\lim_{\varepsilon \to 0} \zeta_\varepsilon(u, Z) = f(u)\exp\left\{iZ(u, Kh) + \frac{Z^2}{2}\|h\|^2\right\}. \tag{3.1}$$

The functions $\zeta_\varepsilon(u, Z)$ are a family of analytic functions of Z. This family is bounded in every region $|\mathrm{Re}\, Z| \le a$. Indeed, in view of (2.3), we have for all sufficiently small $\varepsilon > 0$,

$$\sup_{|\mathrm{Re}\, Z| < a} |\zeta_\varepsilon(u, Z)| = \sup_{|\mathrm{Re}\, Z| < a} E \exp\{\mathrm{Re}\, Z\hat{\Delta}_\varepsilon(h)\} \le 2\exp\{a^2(\|h\|^2 + 1)\}.$$

Consequently, by Montel's theorem the family $\{\zeta_\varepsilon(u, Z)\}$ is compact and, since the right-hand side of (3.1) is analytic for all Z, we have for all complex Z,

$$\lim_{\varepsilon \to 0} \zeta_\varepsilon(u, Z) = \zeta_0(u, Z) = f(u)\exp\left\{iZ(u, Kh) + \frac{z^2}{2}\|h\|^2\right\}.$$

Suppose for a moment that $H_0 = H$. We may then take $h = K^*u$, $Z = -i$ and obtain

$$\begin{aligned} f(u) &= \exp\{-\tfrac{1}{2}(u, KK^*u)\}\phi_0(u, -i) \\ &= \exp\{-\tfrac{1}{2}(KK^*u, u)\}g(u). \end{aligned}$$

Evidently, the first factor on the left is the characteristic function of the normal distribution $\mathbf{N}(0, K)$. We now prove that $g(u)$ is a characteristic function. First note that the equality

$$g(u) = \lim_{\varepsilon \to 0} E^{(\varepsilon)}_{\theta} \exp\{i(u, B_\varepsilon(T_\varepsilon - \phi(\theta))) - i\hat{\Delta}_\varepsilon(K^*u)\}$$

shows that the function $g(u)$ is positive definite. Indeed, with probability going to 1 as $\varepsilon \to 0$,

$$\hat{\Delta}_\varepsilon(K^*(u - v)) = \hat{\Delta}_\varepsilon(K^*u) - \hat{\Delta}_\varepsilon(K^*v)$$

and hence

$$\sum_{k,l} \lambda_k \lambda_l g(u_k - u_l)$$
$$= E_\theta^{(\varepsilon)} \left| \sum_k \lambda_k \exp\{i(u_k, B_\varepsilon(T_\varepsilon - \phi(\theta))) - i\hat{\Delta}_\varepsilon(K^* u_k)\} \right|^2 + o(1).$$

Now by the Minlos–Sazonov theorem [see Skorokhod (1974), Section 1.4], the positive-definite function $g(u)$ is a characteristic function if and only if one can find nuclear operators $\{R_\delta, \delta > 0\}$ such that $|g(u) - g(0)| < \delta$ if $(R_\delta u, u) \leq 1$. We now construct such operators.

We have

$$|g(u) - g(0)| \leq |f(u)\exp\{\tfrac{1}{2}(KK^* u, u)\} - 1|$$
$$\leq |f(u) - 1| + |\exp\{\tfrac{1}{2}(KK^* u, u)\} - 1|.$$

By virtue of the Minlos–Sazonov theorem, one can find a family $\{\tilde{R}_\delta, \delta > 0\}$ of nuclear operators $\tilde{R}_\delta$ such that for $(\tilde{R}_\delta u, u) \leq 1$,

$$|f(u) - 1| \leq \frac{\delta}{3}.$$

Set $R_\delta = \tilde{R}_\delta + (1/\delta)KK^*$. If $(R_\delta u, u) < 1$ and $\delta < 1/2$, then

$$|g(u) - 1| \leq \frac{\delta}{3} + (e^{\delta/2} - 1) \leq \delta.$$

Now let $H_0 \neq H$ (but $\overline{H}_0 = H$). It is possible that $K^* u \notin H_0$, but one can always find a sequence $h_n \in H_0$, $h_n \to K^* u$ in such a way that

$$f(u)\exp\left\{\frac{1}{2}(KK^* u, u)\right\}$$
$$= \lim_n f(u)\exp\{-\tfrac{1}{2}\|h_n\|^2 + (u, h_n)\}$$
$$= \lim_n \lim_\varepsilon E_\theta^{(\varepsilon)} \exp\{i(u, B_\varepsilon(T_\varepsilon - \phi(\theta))) - i\Delta h_\varepsilon(h_n)\} = g(u).$$

The theorem is proved. □

REMARK. It is easy to see that in many cases the operators K_ε and K may be defined in the simple way

$$K_\varepsilon = B_\varepsilon \phi'(\theta) A_\varepsilon P_H, \qquad (3.2)$$
$$K = \lim_\varepsilon B_\varepsilon \phi'(\theta) A_\varepsilon P_H.$$

3.2. *Example.*

EXAMPLE 3.1. Consider one of the simplest situations: Example 2.1. We observe X_ε where

$$dX_\varepsilon(t) = \theta(t)\,dt + \varepsilon\,dw(t), \qquad 0 \leq t \leq 1.$$

Let $l_1, l_2, \dots$ be orthonormal vectors in $L_2(0,1)$. Let $H \subseteq L_2$ be a subspace orthogonal to all l_j. Suppose that the parametric set $\Theta \subseteq H$. If the set Θ is sufficiently large and if for $\theta \in \Theta$ we have $\theta + \varepsilon h \in \Theta$, for any $h \in H$ and $\varepsilon < \varepsilon(h)$, then as we have seen the LAN conditions are fulfilled in the direction H with $A_\varepsilon = \varepsilon I$. Setting $B_\varepsilon = \varepsilon^{-1} I$, where I is the identity operator, we find that

$$K = \lim_{\varepsilon} K_\varepsilon = \phi'(\theta) P_H$$

and that the conditions of the theorem will be fulfilled if $\phi'(\theta)$ is a Hilbert–Schmidt operator.

For example, let $\phi(\theta) = \int_0^1 |\theta(t)|^2\, dt$. Then $\phi'(\theta) = 2\theta$ and $\mathbf{N}(0, K)$ is the distribution of the normal random variable

$$\xi = 2\int_0^1 \theta(t)\, dw(t) - 2\sum_j \int_0^1 l_j(t)\theta(t)\, dt \int_0^1 l_j(t)\, dw(t).$$

We show later (see Sections 6 to 8) that if the set Θ is not too large one can construct an estimator T_ε such that the limiting distribution of $\varepsilon^{-1}(T_\varepsilon - \phi(\theta))$ as $\varepsilon \to 0$ coincides with the distribution of ξ.

Now let $\phi(\theta) = \int_0^t \theta(s)\, ds$. We then have $\phi'(\theta)h = \int_0^t h(s)\, ds$ and the random function

$$\xi = w(t) - \sum_j \int_0^t l_j(u)\, du \int_0^1 l_j(u)\, dw(u) \text{ has an } \mathbf{N}(0, K) \text{ distribution}.$$

Setting

$$T_\varepsilon = X_\varepsilon(t) - \sum_j \int_0^t l_j(u)\, du \int_0^1 l_j(t)\, dX_\varepsilon(t),$$

we find that the distribution of $\xi^{-1}(T_\varepsilon - \phi(\theta))$ coincides with the distribution of ξ.

We shall return to this and other examples in the next section where we consider them from a slightly different point of view.

4. A lower bound for the asymptotic minimax risk.

4.1. In this section we prove a variant of the Le Cam–Hájek theorem which was mentioned in Section 1 [Le Cam (1953, 1972) and Hájek (1972); see also Ibragimov and Has'minskii (1981)].

We consider here the same estimation problem as in the previous section. If the function $\phi(\theta)$ takes its values in a Euclidean or Hilbert space $\mathbf{U}$, we measure the closeness of an estimator $\mathbf{T}$ to $\phi(\theta)$ by a loss function $l\colon \mathbf{U} \to R^1$. We assume that l is subconvex, that is, (1) $l(u) \ge 0$, (2) $l(u) = l(-u)$, (3) for all $\lambda > 0$ the sets $U_\lambda = \{u\colon l(u) < \lambda\}$ are convex. We denote this class of loss functions by Λ.

We suppose (see Section 2) that the parameter set $\Theta \subset \mathbf{L}$, where $\mathbf{L}$ is a normed space. Let $\mathbf{S}$ be a topology on $\mathbf{L}$ compatible with the linear structure of $\mathbf{L}$. *We call the topology* $\mathbf{S}$ *compatible with the LAN conditions if for any* $h \in H_0$ *the vectors* $A_\varepsilon h \to 0$ *as* $\varepsilon \to 0$ *in the* $\mathbf{S}$*-topology.*

As before let us denote by K_ε the operator for $h \in H$ defined by

$$K_\varepsilon h = B_\varepsilon\left(\int_0^1 \phi'(\theta + tA_\varepsilon P_H h)\, dt\right) A_\varepsilon P_H h.$$

THEOREM 4.1. *Suppose the family* $\{P_\theta^{(\varepsilon)},\ \theta \in \Theta\}$ *satisfies the LAN conditions in the direction* H *with the norming factors* A_ε, $\mathbf{S}$ *is a topology compatible with the LAN conditions and* $\{V\}$ *is a family of vicinities of* θ *in the* $\mathbf{S}$*-topology. Suppose that* $K_\varepsilon h \to_{\varepsilon \to 0} Kh$ *in* $\mathbf{U}$ *for any* $h \in H$, *where* $K\colon \mathbf{U} \to \mathbf{U}$ *is a Hilbert–Schmidt operator. Then for any* $l \in \Lambda$ *and any family of estimators* $\{\phi_\varepsilon\}$ *for* $\phi(\theta)$,

$$\inf_{\{V\}} \liminf_{\varepsilon \to 0} \sup_{u \in V} E_u^{(\varepsilon)} l\big(B_\varepsilon(\phi_\varepsilon - \phi(u))\big) \geq El(\xi), \tag{4.1}$$

where ξ *is a Gaussian random vector in* $\mathbf{U}$ *with distribution* $\mathbf{N}(0, K)$.

REMARK. It is often sufficient to consider the simpler operators $B_\varepsilon \phi'(\theta) A_\varepsilon P_H$ instead of the operators K_ε. For example, we can do so if $\|A_\varepsilon\|\,\|B_\varepsilon\| = O(1)$ as $\varepsilon \to 0$.

PROOF. Fix a positive number a and set

$$l_a(u) = \begin{cases} l(u), & |l(u)| \leq a, \\ a, & |l(u)| > a. \end{cases}$$

Fix further an integer n and choose n orthonormal vectors $h_i \in H_0$, $i = 1, \dots, n$. Let $Q(b) = Q = [-b, b]^n$ be a cube in R^n. If ε is sufficiently small, then for all $x = (x_1, \dots, x_n) \in Q$ the points $\theta + \Sigma x_i A_\varepsilon h_i \in V$. Hence, setting $\Sigma x_i h_i = h$ and taking into account the relation (2.3), we find that for sufficiently small ε,

$$\begin{aligned}
&\sup_{u \in V} E_u^{(\varepsilon)} l\big(B_\varepsilon(\phi_\varepsilon - \phi(u))\big) \\
&\quad \geq \sup_{x \in Q} E_{\theta + A_\varepsilon h} l_a\big(B_\varepsilon(\phi_\varepsilon - \phi(\theta + A_\varepsilon h))\big) \\
&\quad \geq (\operatorname{mes} Q)^{-1} \int_Q E_{\theta + A_\varepsilon h}^{(\varepsilon)} l_a\big(B_\varepsilon(\phi_\varepsilon - \phi(\theta + A_\varepsilon h))\big)\, dx \\
&\quad = (\operatorname{mes} Q)^{-1} \int_Q E_\theta^{(\varepsilon)} \Big\{ l_a\big(B_\varepsilon(\phi_\varepsilon - \phi(\theta + A_\varepsilon h))\big) \\
&\qquad\qquad \times \exp\Big(\hat{\Delta}_\varepsilon(h) - \frac{1}{2}\|h\|^2 + \psi(\varepsilon, h)\Big)\Big\}\, dx.
\end{aligned}$$

The random variable $\hat{\Delta}_\varepsilon(h) = \sum_1^n x_i \hat{\Delta}_\varepsilon(h_i)$ with probability going to 1 as $\varepsilon \to 0$. Hence we can represent $\hat{\Delta}_\varepsilon(h)$ as a scalar product (ξ_ε, x), where ξ_ε denotes an n-dimensional bounded random vector with distribution converging to $\mathbf{N}(0, I)$ as $\varepsilon \to 0$. Thus as $\varepsilon \to 0$,

$$\begin{aligned}
(\operatorname{mes} Q)^{-1} \int_Q E_\theta^{(\varepsilon)} \bigg\{ & l_a(B_\varepsilon(\phi_\varepsilon - \phi(\theta + A_\varepsilon h))) \\
& \times \exp\left(\hat{\Delta}_\varepsilon(h) - \frac{1}{2}\|h\|^2 + \psi(\varepsilon, h) \right) \bigg\} \, dx \\
\le (\operatorname{mes} Q)^{-1} \int_Q E_\theta^{(\varepsilon)} \bigg\{ & l_a(B_\varepsilon(\phi_\varepsilon - \phi(\theta)) - Kh + o(1)) \\
& \times \exp\left((\hat{\xi}_\varepsilon, x) - \frac{1}{2}|x|^2 \right) \bigg\} \, dx + o(1) \\
= (\operatorname{mes} Q)^{-1} \int_Q E_\theta^{(\varepsilon)} \bigg\{ & l_a(\Gamma_\varepsilon - Kh + o(1)) \exp\left((\xi_\varepsilon, x) - \frac{1}{2}|x|^2 \right) \bigg\} \, dx + o(1).
\end{aligned}$$

Consider the function $\zeta \colon R^n \to R^1$ defined by

$$\zeta(x) = l_a(Kh) = l_a\left(\sum_1^n x_i K h_i \right).$$

The function ζ is symmetric and has the following property: For all $\lambda > 0$ the sets $\{x \colon \zeta(x) < \lambda\}$ are convex. Hence, using the arguments which prove the original Le Cam–Hájek theorem [see Ibragimov and Has'minskii (1981), pages 162–168], we find that

$$\tag{4.2} \begin{aligned}
\liminf_{\varepsilon \to 0} (\operatorname{mes} Q)^{-1} \int_Q E_\theta^{(\varepsilon)} & \left\{ l_a(\Gamma_\varepsilon - Kh) \exp\left((\hat{\xi}_\varepsilon, x) - \frac{1}{2}|x|^2 \right) \right\} dx \\
& \ge J(a, b)(2\pi)^{-n/2} \left(1 - \frac{1}{\sqrt{b}} \right)^n,
\end{aligned}$$

$$J(a, b) = \int_{Q(\sqrt{b})} l_a(Kh) \exp(-|x|^2/2) \, dx.$$

It follows from (4.2) that if the function ζ is continuous, then

$$\tag{4.3} \begin{aligned}
\liminf_{\varepsilon \to 0} (\operatorname{mes} Q)^{-1} \int_Q E_\theta^{(\varepsilon)} & \left\{ l_a(\Gamma_\varepsilon - Kh + o(1)) \exp\left((\hat{\xi}_\varepsilon, x) - \frac{1}{2}|x|^2 \right) \right\} dx \\
& \ge J(a, b)(2\pi)^{-n/2} \left(1 - \frac{1}{\sqrt{b}} \right)^n.
\end{aligned}$$

The last inequality turns out to be true also for functions ζ which take on only a finite number of values. Indeed, in that case the function $\zeta(x)$ is continuous everywhere except on a finite number of convex surfaces. Next, if l is an arbitrary function from Λ, we can approximate the function $\zeta(x) = l_a(Kh)$ by functions $\zeta_N(x)$, where

$$\zeta_N(x) = \frac{k}{N} \quad \text{if} \quad \frac{k}{N} \le \zeta(x) < \frac{k+1}{N}.$$

Hence the inequality (4.3) is valid for all $l \in \Lambda$.

It follows from (4.3) that for all integers n,

$$\begin{aligned} &\inf_{\{V\}} \liminf_{\varepsilon \to 0} \sup_{u \in V} E_u^{(\varepsilon)} l\big(B_\varepsilon(\phi_\varepsilon - \phi(\theta))\big) \\ &\qquad \ge \lim_{b \to \infty} J(a, b)(2\pi)^{-n/2}\left(1 - \frac{1}{\sqrt{b}}\right)^n = El_a(\xi_n), \end{aligned} \tag{4.4}$$

where ξ_n denotes a Gaussian vector in U with distribution $\mathbf{N}(0, KP_n)$, P_n denotes the projection in H on the subspace spanned by the vectors $h_1, \dots, h_n$.

Let $\xi_1, \xi_2, \dots$ be a sequence of Gaussian iid variables with $E\xi_j = 0$, $\operatorname{Var} \xi_j = 1$. One can write

$$El_a(\xi^n) = El_a\left(\sum_1^n \xi_j K h_j\right)$$

and to finish the proof, we need only check that

$$\liminf_{n \to \infty} El_a\left(\sum_1^n \xi_j K h_j\right) \ge El_a\left(\sum_1^\infty \xi_j K h_j\right).$$

Set $U_x = \{h \in U: l_a(h) < x\}$. Suppose that we have proved the inequality

$$P\left\{\sum_1^n \xi_j K h_j \in U_x\right\} \ge P\left\{\sum_1^\infty \xi_j K h_j \in U_x\right\}. \tag{4.5}$$

Then

$$\begin{aligned} \liminf_{n \to \infty} El_a\left(\sum_1^n \xi_j K h_j\right) &= \liminf_{n \to \infty} \int_0^\infty \left(1 - P\left\{l_a\left(\sum_1^n \xi_j K h_j\right) > x\right\}\right) dx \\ &\ge \int_0^\infty \left(1 - P\left\{l_a\left(\sum_1^\infty \xi_j K h_j\right) > x\right\}\right) dx \\ &= El_a\left(\sum_1^\infty \xi_j K h_j\right). \end{aligned}$$

We must prove (4.5). This inequality follows from the following infinite-dimensional analog of Anderson's lemma.

Let μ be a centered Gaussian measure in a Hilbert space U. Let A be a symmetric convex Borel subset of U. Then for any $u \in U$,

$$\mu(A + u) \le \mu(A). \tag{4.6}$$

In its turn, this inequality is an immediate consequence of the following general result of Ehrhard.

THEOREM [Ehrhard (1983)]. *Let γ be a Gaussian Radonian measure in a locally convex Hausdorff space E. The function $\Phi(x) = (2\pi)^{-1} \int_{-\infty}^{x} \exp\{-v^2/2\}\, dv$. Then for any two convex Borel sets A_1 and A_2 and for any number $\lambda \in (0, 1)$,*

$$(4.7)\quad \Phi^{-1}(\gamma(\lambda A_1 + (1-\lambda)A_2)) \geq \lambda\Phi^{-1}(\gamma(A_1)) + (1-\lambda)\Phi^{-1}(\gamma(A_2)).$$

To deduce (4.6) from (4.7), choose in (4.7) $\lambda = 1/2$, $A_1 = A + u$, $A_2 = -A_1$. This completes the proof of Theorem 4.1. □

4.2. *Examples.*

EXAMPLE 4.1. Let us return to Examples 2.1 and 3.1. In this case (see Example 3.1) the operator $K = \phi'(\theta)P_H$. Hence, for any estimator ϕ_ε of $\phi(\theta) = \int_0^1 \phi^2(t)\, dt$, we have

$$\inf_V \lim_{\varepsilon\to 0} \sup_{u\in V} \varepsilon^{-2} E_u^{(\varepsilon)} |\phi_\varepsilon - \phi(\theta)|^2 \geq 4\left(\int_0^1 \theta^2(t)\, dt - \sum_j \left(\int_0^1 l_j(\theta)\theta(t)\, dt\right)^2\right)$$

$$= 4\|\theta - P_H\theta\|^2.$$

We show in Part II that if the set Θ is sufficiently small (a long one-dimensional segment of a Hilbert space is very meager), one can construct an estimator T_ε such that

$$\varepsilon^{-2} E_\theta |T_\varepsilon - \phi(\theta)|^2 = 4\|\theta - P_H\theta\|^2.$$

If $\phi(\theta) = \int_0^t \theta(s)\, ds$, then

$$\inf_V \lim_{\varepsilon\to 0} \sup_{u\in V} \varepsilon^{-2} E_u^{(\varepsilon)} \|\phi_\varepsilon - \phi(u)\|_{L_2}^2 \geq \frac{1}{2} - \sum_j \int_0^1 \left(\int_0^t l_j(u)\, du\right)^2 dt.$$

EXAMPLE 4.2. We now consider several estimation problems in the context of Example 2.1. Let $\xi_1, \dots, \xi_n$ be real-valued random variables. Therefore, $\mathbf{X}$ is the real line R^1, Ω is the collection of Borel sets and μ is a measure on Ω. If we choose $B_\varepsilon = n^{-1/2}I$, then the operator K is equal to $\phi'(\theta)AP_H$, A is the operator of multiplication by $\sqrt{\theta}$. Let $w_\mu(x)$ be a Gaussian random function on R^1 with independent increments, $Ew_\mu(x) = 0$ and $E(dw_\mu(x))^2 = d\mu(x)$.

EXAMPLE 4.2.1. Suppose we would like to estimate $\phi(\theta) = \int_{-\infty}^{+\infty} x^2\theta(x)\mu(dx)$. This functional is differentiable if we consider only functions θ with $\int_{-\infty}^{+\infty} x^2\theta(x)\mu(dx) < \infty$ and $\phi'(\theta)h = \int_{-\infty}^{+\infty} x^2 h(x)\mu(dx)$. [Of course, we have to choose H in such a way that, for $h \in H$, $\int x^2|h|\sqrt{\theta}\, d\mu < \infty$; in the sequel we omit such evident explanations.] It follows from Theorem 4.1 that for any estimator ϕ_ε and for any compatible topology S,

$$\begin{aligned}
&\inf_{\theta\in V} \liminf_{n\to\infty} n \sup_{u\in V} E_u^{(\varepsilon)}(\phi_\varepsilon - \phi(\theta))^2 \\
&\geq E\left|\int_{-\infty}^{+\infty} x^2\sqrt{\theta}\, dw_\mu(x) - \int_{-\infty}^{+\infty} x^2\theta(x)\, d\mu(x) \int_{-\infty}^{+\infty} \sqrt{\theta}\, dw_\mu(x)\right|^2 \\
&= E_\theta\xi_1^4 - \left(E_\theta\xi_1^2\right)^2 \\
&= \lim_{n\to\infty} nE_\theta\left|\frac{1}{n}\sum_1^n \xi_j^2 - \phi(\theta)\right|^2.
\end{aligned}$$

EXAMPLE 4.2.2. Suppose now that we wish to estimate the same $\phi(\theta)$ but under the additional restriction that $\int_{-\infty}^{+\infty} x\theta(x)\mu(dx) = 0$. This restriction imposes the condition $\int h\sqrt{\theta}\, d\mu = 0$ but also the condition $\int xh\sqrt{\theta}\, d\mu = 0$. Hence

$$\begin{aligned}
&\inf_V \lim_{n\to\infty} \sup_{u\in V} hE_u^{(\varepsilon)}|\phi_\varepsilon - \phi(\theta)|^2 \\
&\geq E\Bigg|\int_{-\infty}^{+\infty} x^2\sqrt{\theta}\, dw_\mu(x) - \int_{+\infty}^{+\infty} x^2\theta(x)\mu(dx) \int_{-\infty}^{+\infty} \sqrt{\theta(x)}\, dw_\mu(x) \\
&\qquad - \frac{\int_{-\infty}^{+\infty} x^3\theta(x)\mu(dx)}{\int_{-\infty}^{+\infty} x^2\theta(x)\mu(dx)} \int_{-\infty}^{+\infty} x\sqrt{\theta(x)}\, dw_\mu(x)\Bigg|^2 \\
&= E_\theta\xi_1^4 - \left(E_\theta\xi_1^2\right)^2 - \frac{\left(E_\theta\xi_1^3\right)^2}{E_\theta\xi_1^2}.
\end{aligned}$$

EXAMPLE 4.2.3. Take $\phi(\theta) = \int_{-\infty}^{x} \theta(y)\mu(dy)$. [There are many ways to represent $\phi(\theta)$ as an element of a Hilbert space U.] The derivative of ϕ is the integral operator $\phi'(\theta)h = \int_{-\infty}^{x} h(y)\mu(dy)$.

$$l_t(u) = \begin{cases} 0, & \sup|u(y)| \leq t, \\ 1, & \sup|u(y)| > t, \end{cases}$$

belongs to the class Λ. Hence, for any compatible topology,

$$\begin{aligned}
&\inf_V \liminf_{n\to\infty} \sup_{n\in V} P_u\left\{\sqrt{n}\, \sup_n |\phi_\varepsilon - \phi(u)| > t\right\} \\
&\geq P\left\{\sup_x \left|\int_{-\infty}^{x} \sqrt{\theta(y)}\, dw_\mu(y) - \int_{-\infty}^{x} \theta(y)\, d\mu(y) \int_{-\infty}^{+\infty} \sqrt{\theta(y)}\, dw_\mu(y)\right| > t\right\}
\end{aligned}$$

[see Levit (1978)].

5. Nonlinear restrictions.

5.1. We have mentioned many times that the linear space H of the LAN conditions reflects the restrictions imposed on the parametric set Θ near the point θ. By definition these restrictions are linear and the set Θ has to be linear in a neighborhood of θ. Since in the end the characterization of the limit distributions of estimators under the LAN conditions depends on the behavior of $\theta + A_\varepsilon h$ as $A_\varepsilon h \to 0$, we can expect that these results continue to hold if Θ is linear only infinitesimally, that is, if Θ is locally a differentiable manifold. In this section we try to outline the corresponding results without any attempt to give them in utmost generality.

Consider a family of distributions $\{P_\theta^{(\varepsilon)}, \theta \in \Theta\}$ where Θ is a subset of a normed linear space $\mathbf{L}$. Suppose further that Θ is a differentiable manifold with tangent space T_θ at the point $\theta \in \Theta$.

DEFINITION. A family $\{P_\theta^{(\varepsilon)}, \theta \in \Theta\}$ is called locally asymptotically normal (LAN) at a point $\theta \in \Theta$ as $\varepsilon \to 0$, if there exists a Hilbert space $\mathbf{H}$ with norm $\|\cdot\|$, a linear manifold $H_0 \subseteq \mathbf{H}$ and a family $\{A_\varepsilon\}$ of linear operators A_ε: $H_0 \to T_\theta$ such that:

1. $\lim\|A_\varepsilon h\|_{\mathbf{L}} = 0$ as $\varepsilon \to 0$ for any $h \in H_0$.
2. For any $h \in H_0$ and all $\varepsilon < \varepsilon(h)$, the following representations are true: Let $m(\varepsilon)$ be the point of Θ closest to the point $\theta + A_\varepsilon h$ (in $\mathbf{L}$). Then

$$\frac{dP_{m(\varepsilon)}^{(\varepsilon)}}{dP_\theta^{(\varepsilon)}}(X_\varepsilon) = \exp\left\{\Delta_\varepsilon(h) - \frac{1}{2}\|h\|^2 + \psi(\varepsilon, h)\right\},$$

where $\Delta_\varepsilon(h)$ is a linear random function on H_0, the random variables $\Delta_\varepsilon(h)$ are asymptotically $\mathbf{N}(0, \|h\|^2)$ as $\varepsilon \to 0$ and $\psi(\varepsilon, h) \to 0$ as $\varepsilon \to 0$ in $P_\theta^{(\varepsilon)}$-probability.

As before we put $H = \overline{H}_0$ and speak about the LAN conditions in the direction H with the norming factor A_ε.

Suppose further that ϕ: $\Theta \to U$ is a Fréchet differentiable function and consider the problem of estimating $\phi(\theta)$ on the basis of observations X_ε. As before we shall consider normed differences $B_\varepsilon(\phi_\varepsilon - \phi(\theta))$, where ϕ_ε is an estimator of $\phi(\theta)$. For the sake of simplicity, we shall consider only such normed factors B_ε for which $\|A_\varepsilon P_H\|\|B_\varepsilon\|$ is bounded as $\varepsilon \to 0$.

An estimator ϕ_ε (more precisely a sequence of estimators $\{\phi_\varepsilon\}$) will be called $(H_0, A_\varepsilon, B_\varepsilon)$-regular if for any $m(\varepsilon)$ defined by the LAN conditions the normed difference $B_\varepsilon(\phi_\varepsilon - \phi(m(\varepsilon)))$ has with respect to the measures $P_{m(\varepsilon)}^{(\varepsilon)}$ (as $\varepsilon \to 0$), the same limit distribution as $B_\varepsilon(\phi_\varepsilon - \phi(\theta))$ does with respect to the measures $P_\theta^{(\varepsilon)}$.

THEOREM 5.1. *Suppose the family $\{P_\theta^{(\varepsilon)}, \theta \in \Theta\}$ satisfies the LAN conditions and that the estimator ϕ_ε is $(H_0, A_\varepsilon, B_\varepsilon)$-regular. Then if the operators $K_\varepsilon = B_\varepsilon \phi'(\theta) A_\varepsilon P_H$ converge weakly (as $\varepsilon \to 0$) to a Hilbert–Schmidt operator*

K, the limit distribution F of the difference $B_\varepsilon(\phi_\varepsilon - \phi(\theta))$, if it exists, has the form $F = \mathbf{N}(0, K) * G$.

THEOREM 5.2. *Suppose the family $\{P_\theta^{(\varepsilon)}, \theta \in \Theta\}$ satisfies the LAN conditions that S is a topology compatible with the LAN conditions and that $\{V\}$ is a system of neighborhoods of θ. If the operators $K_\varepsilon = B_\varepsilon \phi'(\theta) A_\varepsilon P_H$ converge strongly (as $\varepsilon \to 0$) to a Hilbert–Schmidt operator K, then for any loss function $l \in \Lambda$ and any estimator ϕ_ε,*

$$\inf_{\{V\}} \liminf_{\varepsilon \to 0} \sup_{u \in V} E_u^{(\varepsilon)} l\big(B_\varepsilon(\phi_\varepsilon - \phi(u))\big) \geq El(\xi), \tag{5.1}$$

where ξ is a Gaussian vector in U with distribution $\mathbf{N}(0, K)$.

REMARK. As before we could define the operators K_ε in a simpler way because of the condition $\|A_\varepsilon P_H\| \, \|B_\varepsilon\| = O(1)$.

We omit the proofs of Theorems 5.1 and 5.2 since they coincide with the basic results of Sections 3 and 4.

5.2. *Examples.*

EXAMPLE 5.1. Suppose that as in Examples 2.1, 3.1 and 4.1 we observe the process X_ε:

$$dX_\varepsilon(t) = \theta(t)\, dt + \varepsilon\, dw(t), \qquad 0 \leq t \leq 1. \tag{5.2}$$

EXAMPLE 5.1.1. Let $\theta(t)$ in (5.2) be an element of a one-parameter family of t-functions $S(t, \theta)$. In this case the set Θ is a curve in $L_2(0,1)$. Define the functional ϕ by the equation $\phi(s(\cdot, \theta)) = \theta$ and the norming factors $A_\varepsilon = \varepsilon I$. Putting $B_\varepsilon = \varepsilon^{-1}$, we find that $K_\varepsilon = K = \phi' P_H$. The operator P_H is a projection onto the one-dimensional subspace spanned by $S_\theta^{(1)}(\cdot, \theta)$. Differentiating $\phi(s(\cdot, \theta)) = \theta$, we find that $\phi' = (S_\theta')^{-1}$. It follows that the limiting Gaussian distribution $\mathbf{N}(0, K)$ is the distribution of the random variable

$$\xi = \| S_\theta' \|^{-2} \int_0^1 S_\theta'(t, \theta)\, dw(t).$$

In particular [see Ibragimov and Has'minskii (1981), Sections 2.7 and 3.5], for a real-valued parameter θ,

$$\lim_{\delta \to 0} \liminf_{\varepsilon \to 0} \sup_{|\theta - u| \leq \delta} E_u^{(\varepsilon)} l\Big(\varepsilon^{-1}\big(\hat{\theta}_\varepsilon - \theta\big)\Big) \geq El(\xi).$$

One can obtain an analogous result for $S(t, \theta_1, \theta_2, \dots, \theta_n)$ as an element of an n-dimensional parametric family. In that case the set Θ is an n-dimensional surface in $L_2(0,1)$. It is easy to show that the limiting Gaussian

n-dimensional vector ξ has the correlation matrix

$$\left\| \int \frac{\partial S}{\partial \theta_i} \frac{\partial S}{\partial \theta_j} dt \right\|^{-1}$$

[see Ibragimov and Has'minskii (1981)].

EXAMPLE 5.1.2. Now let the estimand $\phi(\theta) = \int_0^1 l(u)\theta(u)\,du$ and $\Theta\{\theta: \|\theta\| = 1\}$. We then have $\phi'(\theta) = l$ and the tangent space at the point θ consists of all vectors orthogonal to θ. Hence

$$Kh = \phi'(\theta) P_H h = \int_0^1 l(t) h(t)\,dt - \phi(\theta) \int_0^1 h(t)\theta(t)\,dt.$$

The distribution of the limit Gaussian variable ξ is then the distribution of

$$\xi = \int_0^1 l(t)\,dw(t) - \phi(\theta) \int_0^1 \theta(t)\,dw(t).$$

In particular,

$$\inf_{\{V\}} \liminf_{\varepsilon \to 0} \sup_{u \in V} E_u^{(\varepsilon)} \varepsilon^{-2} |\phi_\varepsilon - \phi(\theta)|^2 \geq \|l\|^2 - \phi^2(\theta).$$

Note that without the restriction $\|\theta\| = 1$ the bound will be different, namely $\|l\|^2$.

EXAMPLE 5.1.3. Consider the differential equation

$$y'(t) = f(y(t), \theta(t), t), \qquad y(0) = y_0. \tag{5.3}$$

We assume that f is sufficiently smooth and raise the question of estimating the solution $y = \phi(\theta, t)$ on the basis of the observation (5.2). The solution ϕ of the problem (5.3) is (as function of θ) a nonlinear operator ϕ: $L_2(0,1) \to L_2(0,1)$. Its derivative is an integral operator with kernel

$$\begin{aligned} H_\theta(s,t) &= \chi_t(s) \exp\left\{ \int_s^t \frac{\partial f}{\partial y}(\phi(\theta, v); \theta(v); v)\,dv \right\} \\ &\quad \times \frac{\partial f}{\partial \theta}(\phi(\theta, s); \theta(s); s), \end{aligned} \tag{5.4}$$

where

$$\chi_t(s) = \begin{cases} 1, & s < t, \\ 0, & s \geq t. \end{cases}$$

Hence, if there are no restrictions on θ, the limiting Gaussian distribution $\mathbf{N}(0, K)$ is the distribution in $L_2(0,1)$ generated by the random function

$$\xi(s) = \int_0^1 H_\theta(s,t)\,dw(t).$$

If $\theta \in \Theta = \{\theta: \|\theta\| = 1\}$, then

$$\xi(s) = \int_0^1 H_\theta(s,t)\,dw(t) - \int_0^1 H_\theta(s,t)\theta(t)\,dt \int_0^1 \theta(t)\,dw(t).$$

We shall return to this in Example 8.4.

EXAMPLE 5.2. Let $X_1, \dots, X_n$ be a sample of size n where the X_j take their values in the measurable space $(\mathbf{X}, \Omega)$. Suppose that the X_j have density function $p(x)$ with respect to a σ-finite measure μ on Ω. As in Example 3.2 and 4.2, we consider the problem of estimating $\phi(p)$ but treat it from a different point of view.

Let $\mathbf{L}$ be the Hilbert space $L_2(\mu)$ of all functions $g: \mathbf{X} \to R^1$ with $\int |g(x)|^2 \mu(dx) < \infty$. Represent the density function $p(x)$ as $p(x) = |\theta(x)|^2$. Then θ is a point in the unit sphere $\Sigma = \{\theta: \|\theta\| = 1\}$ in $\mathbf{L}$. We take $\mathbf{H} = \mathbf{L}$. If there are no restrictions imposed on p, the space H consists of all vectors orthogonal to θ. Define the operators $A_\varepsilon = (2\sqrt{n})^{-1} I$. Then $[m(\varepsilon)$ is the point of Θ closest to $\theta + A_\varepsilon h]$

$$\frac{dP_{m(\varepsilon)}^{(\varepsilon)}}{dP_\theta^{(\varepsilon)}}(X_\varepsilon) = \prod_1^n \left(1 + \frac{1}{2\sqrt{n}} \frac{h(X_i)}{\theta(X_i)} + \frac{\psi(X_i)}{\theta(X_i)}\right)^2,$$

where $\|\psi\| = O(\|h\|^2/n)$. It is easy to see that as $\varepsilon \to 0$,

$$\frac{dP_{m(\varepsilon)}^{(\varepsilon)}}{dP_\theta^{(\varepsilon)}}(X_\varepsilon) = \exp\left\{\Delta_\varepsilon(h) - \frac{1}{2}\|h\|^2 + \psi_\varepsilon\right\},$$

where the random variables

$$\Delta_\varepsilon(h) = \frac{1}{\sqrt{n}} \sum_1^n \frac{h(X_i)}{\theta(X_i)}$$

are asymptotically $\mathbf{N}(0, \|h\|^2)$ and $\psi_\varepsilon \to 0$ in probability. Thus the LAN conditions are fulfilled.

If in Theorems 5.1 and 5.2 we choose the operators $B_\varepsilon = (1/\sqrt{n})I$, then the operators $K_\varepsilon = K = \phi'(\theta) P_H$.

EXAMPLE 5.2.1. Let the density function $p(x)$ be an element of the one-parameter family $\{p(x,t)\}$. In this case we can treat the parametric set Θ as a curve on the unit sphere Σ. The tangent space T_t is one dimensional and consists of the vector proportional to the vector $p_t'(\cdot,t)/\sqrt{p(\cdot,t)}$. Keeping in mind that we wish to estimate t, consider the functional ϕ defined by the equality $\phi(\sqrt{p(\cdot,t)}) = t$. Arguing as before (see Example 5.1.1), we find that the limiting Gaussian variable ξ of Theorems 5.1 and 5.2 may be written as

$$\xi = \left\| \frac{p_t'}{\sqrt{p}} \right\|^{-1} \int_{\mathbf{X}} \frac{p_t'(x,t)}{\sqrt{p(x,'t)}}\, dw_\mu(x).$$

Here w_μ is the Gaussian measure on Ω with $Ew_\mu(A) = 0$ and $Ew_\mu(A) \times w_\mu(A') = \mu(A \wedge A')$. In particular,

$$\lim_{\delta \to 0} \liminf_{\varepsilon \to 0} \sup_{|u-t| \le \delta} E_u^{(\varepsilon)} l\big(\sqrt{n}\,|\hat{t}_n - u|\big) \ge \frac{\sqrt{I(t)}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-I(t)\frac{x^2}{2}\right) dx,$$

where $I(t)$ is the Fisher information

$$I(t) = \int_{\mathbf{X}} \frac{|p_t'(x,t)|^2}{p(x,t)} \mu(dx).$$

To avoid any possible misunderstanding, note that to check the LAN conditions under the minimal restrictions given in this example is in essence the same as repeating the proof of Le Cam's theorem [see Ibragimov and Has'minskii (1981), Section 2.1].

EXAMPLE 5.2.2. Suppose now that the variables X_j take real values and that the estimand

$$F(p) = \phi(\theta) = \int_{-\infty}^{x} p(y)\mu(dy).$$

If there are no additional restrictions imposed on p, the space H consists of all vectors orthogonal to $\theta = \sqrt{p}$. The derivative $\phi'(\theta)$ is the integral operator

$$\phi'(\theta)h(x) = 2\int_{-\infty}^{x} h(y)\theta(y)\mu(dy)$$

and the limiting Gaussian distribution $\mathbf{N}(0, K)$ is the distribution of the random function

$$\xi(x) = \int_{-\infty}^{x} \sqrt{p(y)}\, dw_\mu(y) - \int_{-\infty}^{x} p(y)\mu(dy) \int_{-\infty}^{\infty} \sqrt{p(y)}\, dw_\mu(y).$$

Part II.

Here and below we shall suppose that the conditions under which Theorems 4.1 and 5.2 have been proved are fulfilled. Estimates T_ε of $\phi(\theta)$ for which equality holds in the inequalities (4.1) or (5.1) are called asymptotically efficient [cf. Ibragimov and Has'minskii (1981), Section 1.9]. In this part we are interested in how one can construct efficient estimators. This problem has a relatively simple solution in the finite-dimensional case: If $\hat{\theta}_\varepsilon$ is an asymptotically efficient estimator for θ (e.g., the maximum likelihood estimator), then $\phi(\hat{\theta}_\varepsilon)$ is asymptotically efficient for $\phi(\theta)$. If the parametric set Θ is infinite dimensional, asymptotically efficient estimators will not usually exist (under a natural normalization).

One of the possible approaches to the construction of asymptotically efficient estimators is expounded below following the example of estimation in Gaussian white noise (see Examples 2.1, 3.1, 4.1 and 5.1):

$$(*) \qquad dX_\varepsilon(t) = \theta(t)\,dt + \varepsilon\,dW(t), \qquad 0 \le t \le 1.$$

If the functional ϕ to be estimated is linear or polynomial, we can replace the argument θ of ϕ by the "maximum likelihood estimator" X (see Section 6). In general, one can first replace $\phi(\theta)$ by a polynomial (by a part of a Taylor's series, for example) and estimate this polynomial (see Sections 7 and 8). We discuss this construction in detail for the problem $(*)$.

IMPORTANT REMARK. We denote by $\|A\|_2$ the Hilbert–Schmidt norm of a linear or multilinear operator A. The norm in $L_2(0,1)$ is denoted by $\|\cdot\|$ or $\|\ \|_{L_2}$.

For the statistics generated by $(*)$, we often write $E_\theta(\cdot)$ instead of $E_\theta^{(\varepsilon)}(\cdot)$. The reader can find results for other problems in the papers of Levit (1974, 1975b), Pfanzagl (1982) and Has'minskii and Ibragimov (1979, 1986).

6. Asymptotically efficient estimation for linear and multilinear functions.

6.1. *Linear functions.*

6.1.1. Suppose that in Example 2.1 (see also Example 3.1, 4.1 and 5.1) the estimated ϕ: $L_2 \to \mathbf{U}$ has the form

$$\phi(\theta) = \int_0^1 l(t)\theta(t)\,dt, \tag{6.1}$$

where $l(t)$ is a U-valued function and $\int_0^1 \|l(t)\|_U^2\,dt < \infty$. Then the estimator

$$\hat{\phi}_\varepsilon = \int_0^1 l(t)\,dX_\varepsilon(t) \tag{6.2}$$

is asymptotically efficient.

6.1.2. Return to Examples 2.2, 3.2, 4.2 and 5.2. Let X_j be real-valued and $d\mu = dx$. If the estimand

$$\phi(\theta) = \int_{-\infty}^{+\infty} \phi(x)\theta(x)\,dx,$$

$\phi(x) \in U$, $\int \|\phi(x)\|_U^2 \theta(x)\,dx < \infty$ and the loss function $l(u) = \|u\|^2$, then the estimator $\hat{\phi} = (1/n)\sum_1^n \phi(X_i)$ will be asymptotically efficient if the set Θ is sufficiently large [see Levit (1975b)].

6.1.3. Reconsider Example 2.3. Let the estimand be the linear functional

$$\phi(f) = \int_{-\pi}^{\pi} \phi(\lambda) f(\lambda)\,d\lambda,$$

where $f(\lambda)$ is the spectral density of $\{X_j\}$, $\phi \in L_2(-\pi,\pi)$. It is possible to prove arguing as in Has'minskii and Ibragimov (1986) that under some additional conditions the estimator

$$\hat{\phi}_n = \frac{1}{2\pi n}\int_{-\pi}^{\pi}\phi(\lambda)\left|\sum_1^n e^{i\lambda j}(X_j - X)\right|^2 d\lambda, \qquad X = \frac{1}{n}\sum_1^n X_j,$$

will be asymptotically efficient.

6.1.4. In Example 2.4 suppose the estimand is

$$\phi(\theta) = \int_0^1 \theta(t)\, d\mu.$$

Consider the estimator

$$\hat{\phi} = \sum_{K=1}^{N}\int_{A_K} p\, d\mu\, \nu_K^{-1}\sum_{t_i\in A_K}\frac{X_i}{p(t_i)},$$

where A_K is a decomposition of the interval $[0,1]$, $\mathbf{N} = [\gamma n/\ln n]$, ν_K denotes the number of $t_i \in A_K$. One can prove that $\hat{\phi}$ is asymptotically efficient [for details see Pastuchova and Has'minskii (1988)].

6.2. *Polynomial functions* ϕ. We have more or less complete results only for the models 2.1 (3.1, 4.1, 5.1) and 2.2 (3.2, 4.2, 5.2). In Example 2.2 let the random variables X_j be real-valued, $d\mu = dx$ and the estimand

$$\phi_k(\theta) = \int\cdots\int_{R^k}\phi(x_1, x_2, \dots, x_k)\theta(x_1)\cdots\theta(x_k)\,dx_1\cdots dx_k,$$

where ϕ: $R^k \to U$. In this case the U-statistics

$$\hat{\phi}_k = \binom{n}{k}^{-1}\sum\phi(X_{i_1}, \dots, X_{i_k}) \tag{6.3}$$

will be unbiased and asymptotically efficient estimators under fairly weak assumptions, see Levit (1974). It is easy to check that under the same assumption the statistics

$$\hat{\phi} = \sum_{k=0}^{r}\hat{\phi}_k$$

will be unbiased and asymptotically efficient estimators of the polynomials

$$\phi(\theta) = \sum_{k=0}^{r}\phi_k(\theta).$$

We now consider in detail the problem of estimating polynomial functions under the assumptions of Examples 2.1, 3.1, 4.1 and 5.1. This problem has been considered in Ibragimov, Nemirovskii and Has'minskii (1986). As distinct from the paper we consider here a different and more natural construction of estimators and a wider class of estimands.

6.2.1. Let $A_P(t_1, \dots, t_P)$ be a symmetric function from R^P to U such that

$$\int_0^1 \cdots \int_0^1 \| A_P(t_1, \dots, t_P) \|_U^2 \, dt_1 \cdots dt_P = \| A_P \|_2^2 < \infty. \tag{6.4}$$

($\|A\|_2$ is the Hilbert–Schmidt norm of the operator A. In this case the norm of the integral operator generated by the kernel A_P.) Let the function to be estimated A_P: $L_2(0,1) \to U$ be defined by

$$A_P(\theta) = \int_0^1 \cdots \int_0^1 A_P(t_1, \dots, t_P) \theta(t_1) \cdots \theta(t_P) \, dt_1 \cdots dt_P.$$

We shall estimate $A_P(\theta)$ by multiple stochastic integrals, for example, by the integrals

$$J_P(A_P) = \int_0^1 \cdots \int_0^1 A_P(t_1, \dots, t_P) \, dX_\varepsilon(t_1) \cdots dX_\varepsilon(t_P). \tag{6.5}$$

Under the assumptions (6.4) the integrals (6.5) are defined as the usual stochastic multiple integrals with respect to a Wiener process [see McKean (1969) and Meyer (1976)]. First we define in the usual way the integral (6.5) for indicators of parallelepipeds lying inside one of the sets $B_i = \{0 \le t_{i_1} \le t_{i_2} \le \cdots \le t_{i_P} \le 1\}$. The integral is then extended to all linear combinations of such indicator functions. The integral for a symmetric step function is then defined as a sum of the integrals over all $B_i = B_{i_1 i_2, \dots, i_P}$. It is then easy to check that for step functions

$$\begin{aligned} &E_\theta \int_0^1 \cdots \int_0^1 A_P(t_1, \dots, t_P) \, dX(t_1) \cdots dX(t_P) \\ &\qquad = \int_0^1 \cdots \int_0^1 A_P(t_1, \dots, t_P) \, dt_1 \cdots dt_P \triangleq A_P[\theta]_P, \\ &E_\theta \left\| \int_0^1 \cdots \int_0^1 A_P(t_1, \dots, t_P) \, dX(t_1) \cdots dX(t_P) - A_P[\theta]_P \right\|_U^2 \\ &\qquad = \sum_{j=1}^{P} \| D^j A_P[\theta]_P \|_2^2 \frac{\varepsilon^{2j}}{j!}. \end{aligned} \tag{6.6}$$

Here $\|D^j A_P[\theta]_P\|_2$ denotes the Hilbert–Schmidt norm of the jth derivative of the function $A_P[\theta]_P$ [all these norms are finite because of (6.4)].

The integral (6.5) for a symmetric function satisfying (6.4) is the limit of integrals of step functions with respect to the norm $(E_\theta^{(\varepsilon)} \| \cdot \|_U)^{1/2}$. The basic relations remain valid for such functions. It is important to note that the integrals (6.5) are statistics; in other words, they are functions of the observation X_ε.

Note also the following formula:

$$
\begin{aligned}
J_P(A_P) &= A_P[\theta]_P + \varepsilon \int_0^1 DA_P[\theta]_P(t)\,dw(t) \\
&\quad + \frac{\varepsilon^2}{2!}\int_0^1\int_0^1 D^2A_P[\theta]_P(t_1,t_2)\,dw(t_1)\,dw(t_2) + \cdots \\
&\quad + \frac{\varepsilon^P}{P!}\int_0^1\int_0^1 D^PA_P[\theta]_P(t_1,\dots,t_P)\,dw(t_1)\cdots dw(t_P).
\end{aligned} \tag{6.7}
$$

For step functions formula (6.7) is an immediate corollary of the definition of the integral (6.5) and the Itô stochastic multiple integral.

6.2.2. Let

$$T(\theta) = \sum_1^k A_P[\theta]_P$$

be a polynomial of degree k defined on $L_2[0,1]$. The next theorem is a generalization of Theorem 1.1 from Ibragimov, Nemirovskii and Has'minskii (1986).

THEOREM 6.1. *Let the functions $A_P[\theta]_P$, $P = 1,\dots,k$, satisfy conditions* (6.4). *Then the statistic*

$$\hat{T} = \sum_{P=1}^k J_P(A_P) = \sum_{P=1}^k \int_0^1 \cdots \int_0^1 A_P\,dX(t_1)\cdots dX(t_P)$$

is an unbiased estimator of $T(\theta)$. Moreover,

$$
\begin{aligned}
E_\theta\|\hat{T} - T(\theta)\|_U^2 &= \sum_{j=1}^k \varepsilon^{2j}(j!)^{-1}\|D^jT(\theta)\|_2^2, \\
E_\theta\left\|\hat{T} - T(\theta) - \varepsilon\int_0^1 [DT(\theta)](t)\,dw(t)\right\|_U^2 &= \sum_{j=2}^k \frac{\varepsilon^{2j}}{j!}\|D^jT(\theta)\|_2^2.
\end{aligned} \tag{6.8}
$$

The proof follows from (6.7). We need only note that the multiple Wiener–Itô integrals of different multiplicity are orthogonal and that

$$E_\theta\left\|\int_0^1\cdots\int_0^1 A_P(t_1,\dots,t_P)\,dw(t_1)\cdots dw(t_P)\right\|_U^2 = P!\|A_P\|_2^2.$$

COROLLARY 6.1. *Let $T(\theta)$ be a polynomial of degree k. If conditions* (6.4) *are satisfied, then*

$$E_\theta\left\|\hat{T} - T(\theta) - \varepsilon\int_0^1 DT(\theta)(t)\,dw(t)\right\|_U^2 \le C(k)\varepsilon^4(1 + \|\theta\|^{2k-4}).$$

The statistic $\hat{T}$ is unbiased, asymptotically normal and an asymptotically efficient estimator of $T(\theta)$ in any ball $\|\theta\| < R$.

PROOF. The corollary follows immediately from (6.8) and the inequality

$$\|D^j A_P[\theta]_P\|_2^2 \le C(P)\|\theta\|^{2(P-j)}, \qquad j = 1, 2, \dots, P. \qquad \square$$

Let $\phi_1, \phi_2, \dots$ be an orthonormal basis in $L_2[0,1]$. Let $\theta = \sum \theta(i)\phi_i$ be the expansion of θ with respect to $\{\phi_i\}$. It follows from (5.2) that

$$X(i) = \int_0^1 \phi_i(t)\, dX_\varepsilon(t) = \theta(i) + \varepsilon\xi(i), \qquad i = 1, 2, \dots,$$

where $\xi(i)$ are independent normal variables with mean 0 and variance 1. Let

$$\dot{X}_n = \sum_1^n X(i)\phi_i, \qquad \theta_n = \sum_1^n \theta(i)\phi_i, \qquad X^n(t) = X(t) - \int_0^t \dot{X}_n(u)\, du,$$

$$\theta^n = \theta - \theta_n, \qquad \dot{\xi}_n = \sum_1^n \xi(i)\phi_i, \qquad w^n(t) = w(t) - \int_0^t \dot{\xi}_n(u)\, du.$$

The process $\dot{X}_n(t)$ is measurable with respect to the σ-algebra D_n generated by $\xi(1), \dots, \xi(n)$. Since the Gaussian process $w^n(t)$ is orthogonal to all $\xi(i)$, $i \le n$, the process

$$X^n(t) = \int_0^t \theta^n(u)\, du + \varepsilon w^n(t)$$

does not depend on D_n.

Our next interest centers on the estimation of polynomials

$$T(\theta^n) = \sum_{P=0}^k A_P[\theta^n]_P. \tag{6.9}$$

We begin with a definition of the stochastic integrals

$$J_P^{(n)}(A_P) = \int_0^1 \cdots \int_0^1 A_P(t_1, \dots, t_P)\, dX^n(t_1) \cdots dX^n(t_P) \tag{6.10}$$

when the conditions (6.4) are satisfied. For $P = 1$ we define

$$\int_0^1 A_1(t)\, dX^n(t) \triangleq \int_0^1 A_1(t)\, dX(t) - \int_0^1 A_1(t)\dot{X}_n(t)\, dt.$$

It is obvious that

$$\int_0^1 A_1(t)\, dX^n(t) = \int_0^1 A_1(t)\theta^n(t)\, dt + \varepsilon\int_0^1 A_1(t)\, dw^n(t) = \int_0^1 A_1^n(t)\, dX(t),$$

where A_1^n is the projection in $L_2[0,1]$ of A_1 onto the subspace spanned by $\{\phi_{n+1}, \phi_{n+2}, \dots\}$.

If $P > 1$ we can consider the kernel $\dot{A}_P(t_1, \dots, t_P)$ as an element of the space $L_2([0,1]^P)$. The set of functions $\{\phi_{i_1}(t_1) \cdots \phi_{i_P}(t_P)\}$ is an orthonormal basis in this space. Denote by H_P^n the subspace of $L_2([0,1]^P)$ spanned by

$\phi_{i_1}, \phi_{i_2} \cdots \phi_{i_P}$, $i_1, \ldots, i_P \geq n+1$, and let A_P^n be the projection of A_P onto H_P^n. Define

$$
\begin{aligned}
(6.11) \qquad & \int_0^1 \cdots \int_0^1 A_P(t_1, \ldots, t_P)\, dw^n(t_1) \cdots dw^n(t_P) \\
& \triangleq \int_0^1 \cdots \int_0^1 A_P^n(t_1, \ldots, t_P)\, dw(t_1) \cdots dw(t_P).
\end{aligned}
$$

Define $J_P^n(A_P)$ by putting into the right-hand side of (6.10) $\theta^n(t)\,dt + \varepsilon\, dw^n(t)$ instead of $dX^n(t)$ and multiplying by $\prod_1^n(\theta^n(t_j)\,dt_j + \varepsilon\, dw^n(t_j))$. Obviously, the integral $J_P^n(A_P)$ does not depend on $\mathbf{L}_n$. Moreover, the following generalization of (6.11) holds:

$$
(6.12) \qquad J_P^n(A_P) = \int_0^1 \cdots \int_0^1 A_P^n(t_1, \ldots, t_P)\, dX(t_1) \cdots dX(t_P).
$$

It follows from (6.5), (6.6) and (6.12) that

$$
(6.13) \qquad E_\theta J^n(A_P) = A_P^n[\theta]_P = A_P[\theta^n]_P,
$$

$$
\begin{aligned}
(6.14) \qquad J_P^n(A_P) &= A_P[\theta^n]_P + \varepsilon \int_0^1 DA_P^n[\theta]_P(t)\, dw(t) + \cdots \\
& + \frac{\varepsilon^P}{P!} \int_0^1 \cdots \int_0^1 D^P A_P^n[\theta]_P(t_1, \ldots, t_P)\, dw(t_1) \cdots dw(t_P).
\end{aligned}
$$

Hence, if conditions (6.4) are satisfied for $p = 1, 2, \ldots, k$, we may consider as an estimator for the function (6.9) the expression

$$
(6.15) \qquad \tilde{T} = \sum_1^k J_P^n(A_P) = \sum_{p=1}^k \int_0^1 \cdots \int_0^1 A_P^n(t_1, \ldots, t_P)\, dX(t_1) \cdots dX(t_P).
$$

It follows from (6.13) and (6.14) that this estimator is unbiased and that

$$
\begin{aligned}
(6.16) \qquad & E_\theta \left\| \tilde{T} - T(S^n) - \varepsilon \int_0^1 P_{H_1^n} DT(S^n)\, dw(t) \right\|_U^2 \\
& = \sum_{j=2}^k \frac{\varepsilon^{2j}}{j!} \left\| P_{H_j^n} D^j T(S^n) \right\|_2^2.
\end{aligned}
$$

Remark 6.1. Since X^n does not depend on $\mathbf{L}_n$, the stochastic integrals $J_P^n(A_P)$ may be defined in the same way for random $\mathbf{L}_n$-measurable kernels A_P. Formula (6.16) continues to be true for the conditional expectation $E\{\cdot|\mathbf{L}_n\}$.

Remark 6.2. Let A be Hilbert–Schmidt integral operator with the kernel $A(t, s)$. Sometimes we shall write the stochastic integrals $\int_0^1 A(t, s)\, dw(s)$, $\int_0^1 A(t, s)\, dx(s)$, and so on, in the form $(A, \dot{W}), (A, \dot{X})$, or $A(\dot{W})$, $A(\dot{X})$ as the result of the operator A acting on the "function" $\dot{W}$ or $\dot{X}$.

Note that a Hilbert–Schmidt operator A can be realized as an integral operator. Usually we denote the kernel of this operator with the same letter A.

6.2.3. Suppose now that conditions (6.4) are satisfied for $0 \le P \le k-1$ but not for $P = k$. In this case there are no unbiased estimators of T defined by (6.9). Consider instead of the kth condition (6.4) the weaker one:

$$(6.17) \qquad \int_0^1 \cdots \int_0^1 dt_2 \cdots dt_k \left\| \int_0^1 A_k(t_1, \dots, t_k)\theta(t_1)\, dt_1 \right\|_U^2 \le L^2 \|\theta\|^2.$$

Following Ibragimov, Nemirovskii and Has'minskii (1986), take an integer $N > 0$ and substitute $A_k[\theta]_k$ by

$$A_k^N[\theta]_k \triangleq \int_0^1 \cdots \int_0^1 A_k(t_1, \dots, t_k)\theta_N(t_1)\theta(t_2) \cdots \theta(t_k)\, dt_1\, dt_2 \cdots dt_k$$
$$= \int_0^1 \cdots \int_0^1 A_{kN}(t_1, \dots, t_k)\theta(t_1) \cdots \theta(t_k)\, dt_1 \cdots dt_k$$

with symmetric kernel A_{kN}. We have [cf. Ibragimov, Nemirovskii and Has'minskii (1986)]

$$\| A_{kN} \|_2^2 \le L^2 N.$$

Define

$$T_N(\theta) = \sum_{p=1}^{k-1} A_p[\theta]_p + A_{kN}[\theta]_k.$$

Then the statistic

$$\tilde{T}_n = \sum_{p=1}^{k-1} J(A_p) + J(A_{kN})$$

is an unbiased estimator for $T_N(\theta)$. It follows from (6.17) that

$$(6.18) \qquad |T_N(\theta) - T(\theta)| \le |A_{kN}[\theta]_k - A_k[\theta]_k| \le L\|\theta^N\|\,\|\theta\|^{k-1}.$$

Using Theorem 6.1 and the last inequality, we deduce that

$$(6.19) \qquad \begin{aligned} &E_\theta \left\| \tilde{T}_N - T(\theta) - \varepsilon \int_0^1 DT_N(\theta)\, dw(t) \right\|_U^2 \\ &\quad \le c(k)\left(\varepsilon^4 + \left(N + \|\theta\|^{2k}\right)\varepsilon^{2k} + \|\theta^N\|^2 \|\theta\|^{2k-2}\right). \end{aligned}$$

7. Asymptotically efficient estimation of smooth functions. Let $X_\varepsilon(t)$ be the observation process of Examples 2.1, 3.1, 4.1 and 5.1. In other words,

$$dX_\varepsilon(t) = \theta(t)\, dt + \varepsilon\, dw(t), \qquad 0 \le t \le 1.$$

Suppose that the parameter set Θ is a compact subset of $L_2(0,1)$. For the sake of simplicity, we suppose that $\Theta \subseteq \Theta_1 = \{\theta\colon \|\theta\| \le 1\}$. Kolmogorov's diameters $d_n(\Theta)$, $n = 1, 2, 3, \dots$, of the set Θ are defined as follows:

$$d_n(\Theta) = \inf_M \sup_{x \in \Theta} \inf_{y \in M} \|x - y\|_{L_2},$$

where the infimum is taken over all n-dimensional linear manifolds in $L_2(0,1)$. We would like to estimate the value $\phi(\theta)$ of a known function ϕ at an unknown point $\theta \in \Theta$. Suppose that the function ϕ satisfies the following conditions:

1. The function $\phi\colon \Theta_1 \to U$ is $k \geq 1$ times Fréchet differentiable the derivatives $D^j\phi$, $j = 1, \dots, k-1$, are continuous functions of θ for $\|\theta\| \leq 1$ with

$$\sup_{\|\theta\| \leq 1} \| D^j \phi(\theta) \|_2 \leq L < \infty, \qquad j = 1, \dots, k-1.$$

The last derivative $D^k\phi$ satisfies the condition

$$\| D^k\phi(\theta)[g]_1 \|_2 \leq L\|g\| \tag{7.1}$$

for all $\|\theta\| < 1$.

2. The derivative $D^k\phi(\theta)$ satisfies Hölder's condition of order γ with respect to the usual operator norm in the ball $\Theta_1 = \{\theta\colon \|\theta\| \leq 1\}$:

$$\| D^k\phi(\theta_2) - D^k\phi(\theta_1) \| \leq L\|\theta_2 - \theta_1\|^\gamma, \qquad \theta_1, \theta_2 \in \Theta_1.$$

The following theorem is a modification and generalization of Ibragimov, Nemirovskii and Has'minskii (1986).

THEOREM 7.1. *Suppose that the parameter set* $\Theta \leq \operatorname{int} \Theta_1$ *and that the Kolmogorov diameters of* Θ *satisfy*

$$d_n(\Theta) \leq cn^{-\beta}.$$

Suppose further that a function $\phi\colon \Theta_1 \to U$ *satisfies conditions* 1 *and* 2, *the constants* γ *and* β *being connected by the relations*

$$k + \gamma > (2\beta)^{-1}, \qquad k \geq 3,$$

$$k + \gamma > 1 + (2\beta)^{-1}, \qquad k = 1, 2.$$

Then there exists an estimator ϕ_ε *of* $\phi(\theta)$ *such that*

$$\sup_{\theta \in \Theta} E_\theta \left\| \phi_\varepsilon - \phi(\theta) - \varepsilon \int_0^1 \phi'(\theta)(t)\, dw(t) \right\|_U^2 \varepsilon^{-2} \to 0. \tag{7.2}$$

PROOF. We will only sketch the proof because it is mainly a repeat of the proof of the afore-mentioned results from Ibragimov, Nemirovskii and Has'minskii (1986). We restrict attention to the case $k \geq 3$. The case $k = 1$ has been considered in Has'minskii and Ibragimov (1980) and the case $k = 2$ may be treated in the same way as the case $k \geq 3$. For the sake of simplicity, suppose that Θ is symmetric.

Choose two integers $n = [\varepsilon^{-2}(\ln(1/\varepsilon))^{-1}]$ and $N = [\varepsilon^{-2(k-1)}(\ln 1/\varepsilon)^{-1}]$. Pick an orthonormal basis $\{\phi_{jn}\} = \{\phi_j\}$ in $L_2(0,1)$ in such a way that

$$\sup_{x \in \Theta} \inf_{y \in E_n} \|x - y\| \leq 2d_n(\Theta),$$

where E_n is the span of $\phi_1, \dots, \phi_n$. By integrals of the type $\int_0^1 \cdots \int_0^1 a(t_1, \dots, t_p)\, dX^n(t_1) \cdots dX^n(t_p)$ we mean the basis $\{\phi_j\}$ (see Section 6).

Fix a small positive number δ and let

$$\overline{X}_n = \begin{cases} \dot{X}_n, & \text{if } \|\dot{X}_n\| < 1 - \delta, \\ \dot{X}_n \|\dot{X}_n\|^{-1}(1-\delta), & \text{if } \|\dot{X}_n\| \ge 1 - \delta. \end{cases}$$

Define the estimator

$$\begin{aligned} \phi_\varepsilon &= \phi(\overline{X}_n) \\ &\quad + \sum_{j=1}^{k-1} \frac{1}{j!} \int_0^1 \cdots \int_0^1 D^j \phi(\overline{X}_n)(t_1, \dots, t_j)\, dX^n(t_1) \cdots dX^n(t_j) \\ &\quad + \frac{1}{k!} \int_0^1 \cdots \int_0^1 \left(D^k \phi(\overline{X}_n)\right)^N (t_1, \dots, t_k)\, dX^n(t_1) \cdots dX^n(t_k). \end{aligned} \tag{7.3}$$

Here the symmetric kernel $(D^k \phi(\overline{X}_n))^N$ is constructed in the same way as the kernel A_k^N in (6.18). We will show that ϕ_ε satisfies (7.2).

We begin with the identity

$$\begin{aligned} &\phi_\varepsilon - \phi(\theta) - \varepsilon \int_0^1 D\phi(\theta)\, dw(t) \\ &\quad = \left[\phi_\varepsilon - \phi(\overline{X}_n + \theta^n) - \varepsilon \int_0^1 D\phi(\overline{X}_n + \theta^n)(t)\, dw^n(t)\right] \\ &\qquad + \left[\phi(\overline{X}_n + \theta^n) - \phi(\theta) - \varepsilon\left(D\phi(\theta)(t), \dot{\xi}_n(t)\right)\right] \\ &\qquad + \varepsilon\left[\int_0^1 \left(D\phi(\dot{X}_n + \theta^n)(t) - D\phi(\theta)(t)\right) dw^n(t)\right] \\ &\quad = I_1 + I_2 + I_3. \end{aligned}$$

The relation (7.2) will be proved if we check that

$$\lim_{\varepsilon \to 0} \varepsilon^{-2} \sup_{\theta \in \Theta} E I_j^2 = 0, \qquad j = 1, 2, 3. \tag{7.4}$$

For sufficiently small δ,

$$P\left\{\overline{X}_n \ne \dot{X}_n\right\} \le c_1 \exp\left\{c_2 n \delta^2 \ln \varepsilon\right\} = o(\varepsilon^2).$$

Since $\dot{X}_n + \theta^n = \theta + \varepsilon \dot{\xi}_n$, $\varepsilon^2 n \to 0$ and $D\phi(\theta)$ is a smooth function of θ, it follows that the equality (7.4) is satisfied for $j = 3$. For $j = 1$ the proof of (7.4) is given as in Ibragimov, Nemirovskii and Has'minskii (1986) and so we omit it. The proof of (7.4) for $j = 2$ is based on two lemmas.

LEMMA 7.1. *Let ϕ: $\Theta_1 \to U$ be a twice Fréchet differentiable function such that*

$$\|\phi'(\theta)\|_2 \le c < \infty, \qquad \|\phi''(\theta)\|_2 \le c < \infty, \qquad \theta \in \Theta_1.$$

Let $E(n)$ be an n-dimensional subspace of $L_2[0,1]$. Denote by $\dot{\xi}(n)$ an

$E(n)$-valued Gaussian random vector with mean 0 and unit covariance matrix. Let $n = n(\varepsilon)$ be chosen in such a way that $n(\varepsilon) \to \infty$, $\varepsilon\sqrt{n} \to 0$ as $\varepsilon \to 0$. Then for any $\delta > 0$,

$$\sup_{\|\theta\| < 1-\delta} E\Big\{\big\|\phi\big(\theta + \varepsilon\dot{\xi}_n\big) - \phi(\theta)\big\|_U^2 1_{(\varepsilon\|\dot{\xi}_n\| < \delta)} - \varepsilon^2\big\|\phi'(\theta)P_{E(n)}\big\|_2^2\Big\} \le \alpha(\varepsilon)\varepsilon^2,$$

where $P_{E(n)}$ denotes the projection on $E(n)$ and $\alpha(\varepsilon) \to 0$ as $\varepsilon \to 0$. Moreover, $\alpha(\varepsilon)$ depends on ε, $n(\varepsilon)$ and C but not on ϕ.

Proof. The proof of this lemma is essentially the proof of Lemma 2.1.1 of Ibragimov, Nemirovskii and Has'minskii (1986) and so we only give a sketch of it. Let the function

$$g(h) = \|\phi(\theta + h) - \phi(\theta)\|_U^2.$$

The function g is defined and twice continuously differentiable in the ball $V = \{h \in E(n): \|h\| < \delta\}$. Denote by $I_\rho(f)$ the mean value of a function f on the sphere $\{h \in E(n), \|h\| \le \rho\}$. Let Δ denote the Laplace operator on functions $E(n) \to U$ and let ∇ denote, for $\rho < R < \delta$, the gradient of such functions. Arguing as in Ibragimov, Nemirovskii and Has'minskii (1986), we find that for $\rho < R < \delta$,

$$I_R(g) = (2n)^{-1}R^2\int_0^R \theta_R^n(p)I_\rho(\Delta g)\,d\rho, \tag{7.5}$$

where

$$\theta_R^n(p) = \frac{2n}{R^2(n-2)}(\rho - \rho^{n-1}R^{2-n}), \qquad \theta_R^n(\rho) \ge 0, \qquad \int_0^R \theta_R^n(\rho)\,dp = 1.$$

Setting $\phi(\theta + h) - \phi(\theta) = l(h)$, we have

$$\tfrac{1}{2}\Delta g = \tfrac{1}{2}\Delta(l(h), l(h))_U = (\Delta l(h), l(h))_U + \|\nabla l(h)\|_U^2. \tag{7.6}$$

Set

$$A(R) = \max_{0\le\rho\le R} I_\rho(g), \qquad B(R) = \max_{0\le\rho\le R} I_\rho\big(\|\Delta l\|_U^2\big),$$

$$C(R) = \max_{0\le\rho\le R} I_\rho\big(\|\nabla l\|_U^2\big).$$

It follows from (7.5) and (7.6) that, for $0 < \rho \le R$,

$$I_\rho(g) \le \frac{\rho^2}{n}(A^{1/2}(\rho)B^{1/2}(\rho) + c(\rho)) \le \frac{R^2}{n}(A^{1/2}(R)B^{1/2}(R) + c(R)).$$

Hence for $\alpha > 0$,

$$A(R) \le \frac{R^2}{n}C(R)(1+\alpha) + \frac{R^4}{4n^2}B(R)\left(1 + \alpha + \frac{1}{\alpha}\right).$$

Let Q denote the distribution of $\varepsilon\|\dot{\xi}_n\|$. The last inequality gives

$$\begin{aligned}
&E\left\|\phi\left(\theta+\varepsilon\dot{\xi}_n\right)-\phi(\theta)\right\|_U^2 1_{\{\varepsilon\|\dot{\xi}_n\|<\delta\}}\\
&\quad=\int_0^\delta I_\rho(g)Q(d\rho)\\
&\quad\le(1+\alpha)\int_0^\delta\frac{R^2}{n}C(R)Q(dR)+\frac{1}{4}\left(1+\alpha+\frac{1}{\alpha}\right)\int_0^\delta\frac{R^4}{n^4}B(R)Q(dR).
\end{aligned}$$

Now $C(R)\le\|\phi'(\theta)P_{E(n)}\|_2^2+c_1R$ and $B(R)\le c_2n$ (all c_i depend on C only). Hence

$$\begin{aligned}
&E_\theta\left\|\phi\left(\theta+\varepsilon\dot{\xi}_n\right)-\phi(\theta)\right\|_U^2 1_{\{\varepsilon\|\dot{\xi}_n\|<\delta\}}\\
&\quad\le(1+\alpha)\left\{\varepsilon^2\left\|\phi'(\theta)P_{E(n)}\right\|_2^2+c_3\varepsilon^2\left(\varepsilon\sqrt{n}\right)\right\}+c_4\left(1+\alpha+\frac{1}{\alpha}\right)\varepsilon^2\left(\varepsilon\sqrt{n}\right)^2
\end{aligned}$$

and it is sufficient to choose $\alpha=\alpha(\varepsilon)\to0$ as $\varepsilon\to0$. □

LEMMA 7.2. *Let $\phi(\theta)$ be a twice Fréchet differentiable function in the ball $\Theta_1=\{\theta\colon\|\theta\|<1\}$. Suppose that for all $\theta\in\Theta$,*

$$\|\phi'(\theta)\|_2\le C<\infty,\qquad\|\phi''(\theta)\|_2\le c<\infty.$$

Let $E(n)$ be an n-dimensional subspace of $L_2(0,1)$ and let $\dot{\xi}_n$ denote an $E(n)$-valued Gaussian vector with mean 0 and unit covariance matrix. Finally, suppose that $\varepsilon\sqrt{n}\to0$ as $\varepsilon\to0$. Then for any $\delta>0$,

$$\lim_{\varepsilon\to0}\varepsilon^{-2}E\left\{\left\|\phi\left(\theta+\varepsilon\dot{\xi}_n\right)-\phi(\theta)-\varepsilon\phi'(\theta)\dot{\xi}_n\right\|_U^2 1_{\{\varepsilon\|\dot{\xi}_n\|<\infty\}}\right\}=0.\tag{7.7}$$

This last relation is uniform with respect to all $\theta\colon\|\theta\|\le1-\delta$.

PROOF. It follows from Lemma 7.1 that

$$\begin{aligned}
&\lim_{\varepsilon\to0}\sup_{\|\theta\|<1-\delta}\left\{\varepsilon^{-2}E\left\{\left\|\phi\left(\theta+\varepsilon\dot{\xi}_n\right)-\phi(\theta)-\varepsilon\phi'(\theta)\dot{\xi}_n\right\|_U^2 1_{\{\varepsilon\|\dot{\xi}_n\|<\delta\}}\right\}\right\}\\
&\quad\le2\lim_{\varepsilon\to0}\sup_{\|\theta\|\le1-\delta}\Big\{E\left\|\phi'(\theta)\dot{\xi}_n\right\|_U^2\\
&\qquad\qquad-\varepsilon^{-1}E\left[\left(\phi\left(\theta+\varepsilon\dot{\xi}_n\right)-\phi(\theta),\phi'(\theta)\dot{\xi}_n\right)_U 1_{\{\varepsilon\|\dot{\xi}_n\|<\delta\}}\right]\Big\}
\end{aligned}$$

and hence we need only prove that the right-hand side of the last inequality is 0.

Let $\psi_1,\dots,\psi_n$ be an orthonormal basis in $E(n)$ and

$$\dot{\xi}_n=\sum_1^n\xi(i)\psi_i,\qquad\xi_n^k=\dot{\xi}_n-\xi(k)\psi_k.$$

Put $\phi'_k(\theta)=\psi'(\theta)\phi_k$. Since $P\{\varepsilon\|\dot{\xi}_n\|>\delta\}=o(\varepsilon^2)$, we can discard the factor $1_{\{\varepsilon\|\dot{\xi}_n\|>\delta\}}$ and suppose that the function $\phi(\theta)$ is defined for all $\theta\in L_2(0,1)$ and that $\|\phi'(\theta)\|_2 \le c$ for all $\theta \in L_2(0,1)$.

Then

$$E\Big\{\big(\phi(\theta + \varepsilon\dot{\xi}_n) - \phi(\theta), \phi'(\theta)\dot{\xi}_n\big)_U\Big\} \tag{7.8}$$
$$= \sum_{k=1}^{n} E\Big\{\xi(k)\big(\phi'_k(\theta), \phi(\theta + \varepsilon\dot{\xi}_n) - \phi(\theta)\big)_U\Big\}.$$

Since

$$\phi(\theta + \varepsilon\dot{\xi}_n) - \phi(\theta) = \big(\phi(\theta + \varepsilon\dot{\xi}_n) - \phi(\theta + \varepsilon\xi_n^k)\big) + \big(\phi(\theta + \varepsilon\xi_n^k) - \phi(\theta)\big)$$

and $\xi(k)$ does not depend on ξ_n^k, we can write (7.8) as

$$E\Big\{\big(\phi(\theta + \varepsilon\dot{\xi}_n) - \phi(\theta), \phi'(\theta)\dot{\xi}_n\big)_U\Big\}$$
$$= \sum_{k=1}^{n} \Big(\phi'_k(\theta), E\big[\xi(k)\big(\phi(\theta + \varepsilon\dot{\xi}_n) - \phi(\theta + \varepsilon\xi_n^k)\big)\big]\Big)_U.$$

By Taylor's formula,

$$\phi(\theta + \varepsilon\dot{\xi}_n) - \phi(\theta + \varepsilon\xi_n^k) = \varepsilon\xi(k)\phi'_k(\theta + \varepsilon\dot{\xi}_n) + R_k\varepsilon^2|\xi(k)|^2$$

and $\|R_k\|_U \le C$, $C < \infty$. Hence

$$\left|E\Big\{\big(\phi(\theta + \varepsilon\dot{\xi}_n) - \phi(\theta), \phi'(\theta)\dot{\xi}_n\big)_U\Big\} - \varepsilon\sum_{k=1}^{n}\Big(\phi'_k(\theta), E\big[\xi^2(k)\phi'_k(\theta + \varepsilon\dot{\xi}_n)\big]\Big)_U\right|$$
$$\le c\varepsilon^2\sqrt{n} = o(\varepsilon).$$

Finally,

$$\left|\sum_{k=1}^{n} E\Big(\phi'_k(\theta), \xi^2(k)\big(\phi'_k(\theta + \varepsilon\dot{\xi}_n) - \phi'_k(\theta)\big)\Big)_U\right|$$
$$\le E^{1/2}\left\{\sum_{k=1}^{n}\xi^4(k)\|\phi'_k(\theta)\|_U^2\right\}E^{1/2}\left\{\sum_{k=1}^{n}\big\|\phi'_k(\theta + \varepsilon\dot{\xi}_n) - \phi'_k(\theta)\big\|_U^2\right\}$$
$$\le c_1\|\phi'(\theta)\|_2 E^{1/2}\big\|\phi'(\theta + \varepsilon\dot{\xi}_n) - \phi'(\theta)\big\|_2^2$$
$$\le c_2 E^{1/2}\left\|\int_0^{\varepsilon}\phi''(\theta + t\dot{\xi}_n)\dot{\xi}_n\,dt\right\|_2^2 \le c_3 \quad \text{as } \varepsilon\sqrt{n} \to 0,\ \varepsilon \to 0.$$

Hence, uniformly in $\|\theta\| < 1 - \delta$,

$$\varepsilon^{-1}E\Big\{\big(\phi(\theta + \varepsilon\dot{\xi}_n) - \phi(\theta), \phi'(\theta)\dot{\xi}_n\big)_U\Big\}$$
$$= E\big\|\phi'(\theta)\dot{\xi}_n\big\|_U^2 + o(1) = \big\|\phi'(\theta)P_{E(u)}\big\|_2^2 + o(1), \qquad \varepsilon \to 0.$$

This completes the proof of the lemma and hence also the proof of Theorem 7.1. □

8. Asymptotically efficient estimation of smooth functions on smooth manifolds.

8.1. Once again consider the problem of estimating a function $\phi(\theta)$ on the basis of the observation X_ε:

$$dX_\varepsilon(t) = \theta(t)\,dt + \varepsilon\,dw(t), \qquad 0 \le t \le 1.$$

We shall now suppose that the parameter set $\Theta = \Sigma \cap \Gamma$, where as before Σ is a compact subset of the unit ball $\Theta_1 = \{\theta\colon \|\theta\| \le 1\}$ and Γ is a smooth manifold (cf. Section 5). We show that under some restrictions imposed on Γ a small modification of the estimator given in Section 7 will be asymptotically efficient.

Let the manifold Γ be defined by the parametric equation

$$\theta = F(u), \qquad v \in V,$$

where V is a subset of an Euclidean or Hilbert space H_1. We suppose that the map $F\colon V \to L_2$ satisfies the following properties:

A_γ. The map F is Fréchet differentiable in V and

$$\|F'(v_2) - F'(v_1)\| \le L\|v_2 - v_1\|^\gamma$$

for any $v_1, v_2 \in V$.

B. The linear operators $F'(v)^*F'(v)\colon H_1 \to H_1$ are bounded and strictly positive defined uniformly in V.

Let $T(\theta)$ denote the tangent space to Γ at the point θ. Denote by $P_{T(\theta)}$ the projector in L_2 onto $T(\theta)$. It follows from the conditions A_γ and B that

$$\text{(8.1)} \qquad \begin{aligned} \|P_{T(\theta_1)}(\theta_2 - \theta_1)\| &\le \|\theta_2 - \theta_1\|, \\ \|P_{T(\theta_2)} - P_{T(\theta_1)}\| &\le c\|\theta_2 - \theta_1\|^\gamma. \end{aligned}$$

Only the second formula needs a proof. The tangent space $T(\theta)$ consists of the vectors $\{F'(\theta)v,\ v \in V\}$. Hence

$$\text{(8.2)} \qquad P_{T(\theta)} = F'(\theta)((F'(\theta))^* F'(\theta))^{-1}(F'(\theta))^*.$$

Indeed, $P^2_{T(\theta)} = P_{T(\theta)}$ and $P_{T(\theta)}F'(\theta)u = F'(\theta)u$. It follows from A_γ, B and (8.2),

$$\|P_{T(\theta_2)} - P_{T(\theta_1)}\| < c\|\theta_2 - \theta_1\|^\gamma.$$

Now we present a modification of Theorem 7.1. For the sake of simplicity, we consider only the case $k = 1$. There are no principal difficulties in treating the general case.

THEOREM 8.1. *Let the parametric set $\Theta = \Sigma \cap \Gamma$ where Kolmogorov's diameters $d_n(\Sigma)$ of the set Σ satisfy the condition*

$$d_n(\Sigma) \le cn^{-\beta}, \qquad \beta > 0,$$

and the manifold Γ satisfy the conditions A_γ and B. Let the function $\phi(\theta)$, ϕ: $\Theta_1 \to U$, be defined and Fréchet differentiable in the ball $\Theta_1 = \{\theta: \|\theta\| \le 1\}$. Let

$$\|\phi'(\theta)\|_2 \le c; \quad \|\phi'(\theta') - \phi'(\theta)\|_2 \to 0, \quad \textit{for } \theta' \to \theta,$$
$$\|\phi'(\theta) - \phi'(\theta_1)\| \le c\|\theta_2 - \theta_1\|^\gamma$$

for all $\theta, \theta', \theta_i \in \Theta_1$. Suppose that γ and β satisfy

$$\gamma > (2\beta)^{-1}. \tag{8.3}$$

Then there exists an estimator ϕ_ε such that

$$\sup_{\theta\in\Theta} E_\theta \left\| \phi_\varepsilon - \phi(\theta) - \varepsilon\big(\phi'(\theta) P_{T(\theta)}, \dot{w}\big)\right\|_U^2 = o(\varepsilon^2), \qquad \varepsilon \to 0. \tag{8.4}$$

REMARK. Formula (8.4) means, in particular, that the normed difference $\varepsilon^{-1}(\phi_\varepsilon - \phi(\theta))$ is asymptotically normal with the limit distribution equal to the distribution of $(\phi'(\theta)P_{T(\theta)}, \dot{w})$ and

$$\lim_{\varepsilon\to 0} \varepsilon^{-2} E_\theta \|\phi_\varepsilon - \phi(\theta)\|_U^2 = \|\phi'(\theta)P_{T(\theta)}\|_2^2.$$

Hence the estimator ϕ_ε is asymptotically efficient.

PROOF. Take $n = [\varepsilon^{-2/(2\beta+1)}]$. Let $\dot{X}_n$ and $\overline{X}_n$ be defined as in Section 7 and let d_n be the distance from $\overline{X}_n$ to Γ. Let $\tilde{X}_n$ denote a point $\tilde{X}_n \in \Gamma$ such that

$$\left|\overline{X}_n - \tilde{X}_n\right| < d_n + \varepsilon^3. \tag{8.5}$$

We show that (8.4) holds for the estimator ϕ_ε defined by

$$\phi_\varepsilon = \phi(\tilde{X}_n) + \big(F'(\tilde{X}_n) P_{T(\tilde{X}_n)}, \dot{X}_n\big). \tag{8.6}$$

As in Section 7, it is sufficient to prove that

$$\sup_{\theta\in\Theta} E_\theta \Big\{ \left\| \phi_\varepsilon - \phi(\theta) - \varepsilon\big(\phi'(\theta) P_{T(\theta)}, \dot{w}\big)\right\|_U 1_{\{\varepsilon\|\dot{\xi}_n\| < \delta\}} \Big\} = o(\varepsilon^2) \tag{8.7}$$

as $\varepsilon \to 0$.

Put $\phi' = P_{T(\tilde{X}_n)}\theta$. By the conditions of the theorem,

$$\left\|\phi(\tilde{X}_n) - \phi(\theta') - \phi'(\tilde{X}_n)(\tilde{X}_n - \theta')\right\|_U \le L\left\|\tilde{X}_n - \theta'\right\|^{1+\gamma}.$$

It follows from this inequality and (8.6) that

$$\left\|\phi_\varepsilon - \phi(\theta') - \phi'(\tilde{X}_n) P_{T(\tilde{X}_n)}\big(\dot{X}^n + \tilde{X}_n - \theta'\big)\right\|_U \le L\left\|\tilde{X}_n - \theta'\right\|^{1+\gamma}.$$

The inequality (8.5) implies that

$$E_\theta \left\| P_{T(\tilde{X}_n)} \overline{X}_n - \tilde{X}_n \right\| = o(\varepsilon), \tag{8.8}$$

where $o(\varepsilon)$ is uniform in $\theta \in \Theta$. If $\varepsilon\|\dot{\xi}_n\| < \delta$, then $\overline{X}_n = \dot{X}_n$. Hence

$$\left\| P_{T(\tilde{X}_n)}\left(\dot{X}^n + \tilde{X}_n - \theta'\right) - P_{T(\tilde{X}_n)}\left(\dot{X}_n + \dot{X}^n - \theta\right)\right\| = o(\varepsilon).$$

Therefore, if $\varepsilon\|\dot{\xi}_n\| < \delta$,

$$\begin{aligned}
&\left\|\phi_\varepsilon - \phi(\theta') - \phi'(\tilde{X}_n)P_{T(\tilde{X}_n)}\left(\dot{X}_n + \dot{X}^n - \theta\right)\right\|_U \\
&\quad \le L\left\|\tilde{X}_n - \theta'\right\|^{1+\gamma} + o(\varepsilon) \\
&\quad \le L\left\|\dot{X}_n - \theta\right\|^{1+\gamma} + o(\varepsilon), \qquad \varepsilon \to 0.
\end{aligned} \tag{8.9}$$

Since $\dot{X}_n + \dot{X}^n - \theta = \varepsilon\dot{w}$, formula (8.7) follows from (8.9) and the relations:

$$E_\theta\left\|\dot{X}_n - \theta\right\|^{2(1+\gamma)} = o(\varepsilon^2),$$

$$E_\theta\left\{\|\phi(\theta) - \phi(\theta')\|_U^2 1_{\{\varepsilon\|\dot{\xi}_n\| < \delta\}}\right\} = o(\varepsilon^2),$$

$$E_\theta\left\|\left(\phi'(\tilde{X}_n)P_{T(\tilde{X}_n)} - \phi'(\theta)P_{T(\theta)}\right)\varepsilon\dot{w}\right\|^2 = o(\varepsilon^2).$$

The first relation has been established in Section 7. The second one follows from the equality $E_\theta\|\theta - \theta'\|^2 = o(\varepsilon^2)$. We now prove the last one.

Let A be a random $\mathbf{L}_n$-measurable linear operator $A\colon L_2 \to U$ with $\|A\|_2 < \infty$. Then

$$\begin{aligned}
E\|A\dot{w}\|_U^2 &= E\left\|A\dot{\xi}_n + A\dot{w}^n\right\|_U^2 \\
&\le 2E\left\|A\dot{\xi}_n\right\|_U^2 + 2E\|A\dot{w}^n\|_U^2 \\
&\le 2E\left\{\left\|AP_{E(n)}\right\|^2\left\|\dot{\xi}_n\right\|^2\right\} + 2E\|A\dot{w}^n\|_U^2 \\
&\le 4nE^{1/2}\left\|AP_{E(n)}\right\|^4 + 2E\|A\|_2^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\varepsilon^2 E\left\|\left(\phi'(\tilde{X}_n)P_{T(\tilde{X}_n)} - \phi'(\theta)P_{T(\tilde{X}_n)}\right)\dot{w}\right\|_U^2 \\
&\quad \le n\varepsilon^2E^{1/2}\left\|\phi'(\tilde{X}_n) - \phi'(\theta)\right\|^4 + 2\varepsilon^2E^2\left\|\phi'(\tilde{X}_n) - \phi'(\theta)\right\|_2^2 \\
&\quad \le o(\varepsilon^2) + cn\varepsilon^2E^{1/2}\left\|\tilde{X}_n' - \theta\right\|^{4\gamma} \\
&\quad = o(\varepsilon^2) + O\left((\varepsilon^2 n)^{1+\gamma}\right) = o(\varepsilon^2).
\end{aligned} \tag{8.10}$$

Using (8.2), we have in the same way

$$E\left\|\left(\phi'(\theta)P_{T(\tilde{X}_n)} - \phi'(\theta)P_{T(\theta)}\right)\varepsilon\dot{w}\right\|_U^2 = o(\varepsilon^2)$$

and the theorem is proved. □

8.2. The estimator (8.6) can be relatively easily constructed if the manifold Γ has finite dimension. We now show that there is also a rather simple way to construct efficient estimators if Γ has finite codimension. Suppose that the surface Γ is determined by the equations

$$F_1(\theta) = 0, \dots, F_r(\theta) = 0. \tag{8.11}$$

Here the real-valued functions $F_j(\theta)$ are supposed to be defined and Fréchet differentiable in the unit ball Θ_1. We also suppose that the vectors $F_j'(\theta)$ are linearly independent for all θ and that the Gram determinant

$$G(\theta) = \det\left\|\left(F_i'(\theta), F_j'(\theta)\right)\right\| \geq m > 0 \tag{8.12}$$

for all $\theta \in \Theta$. Denote by $E(\theta)$ the subspace in L_2 spanned by $F_j'(\theta)$, $j = 1, \dots, r$. Evidently, the tangent space $T(\theta)$ at the point θ consists of vectors orthogonal to $E(\theta)$ and so

$$\phi'(\theta) P_{T(\theta)} = \phi'(\theta) - \phi'(\theta) P_{E(\theta)}.$$

Hence we can plug into (8.6) an estimator of $\phi'(\theta) - \phi'(\theta)P_{E(\theta)}$ instead of an estimator of $\phi'(\theta)P_{T(\theta)}$ and which is easier to analyze since the projection P_E is finite dimensional. The situation is especially simple if the function ϕ takes real values. Suppose as before that the Kolmogorov diameters satisfy

$$d_n(\Sigma) \leq cn^{-\beta}.$$

Also suppose that the functions $\phi, F_1, \dots, F_r$ satisfy conditions 1 and 2 of Section 7 with constants $k, \gamma, k_1, \gamma_1, \dots, k_r, \gamma_r$, respectively. Let

$$\mathbf{x} = \min\{k + \gamma, k_1 + \gamma_1, \dots, k_r + \gamma_r\}.$$

Consider the conditions (cf. Theorem 7.1)

$$\begin{aligned} &\mathbf{x} > (2\beta)^{-1}, && \text{if } \mathbf{x} > 3, \\ &\mathbf{x} > (2\beta)^{-1} + 1, && \text{if } \mathbf{x} \leq 3. \end{aligned} \tag{8.13}$$

Under these conditions denote by $\phi_\varepsilon, F_{1\varepsilon}, \dots, F_{r\varepsilon}$ the same efficient estimators for $\phi(\theta), F_1(\theta), \dots, F_r(\theta)$, which have been constructed in Theorem 7.1. Let $\lambda_j(\theta)$, $j = 1, 2, \dots, r$, be determined from the expansion

$$P_{E(\theta)}\phi'(\theta) = \sum_{j=1}^{r} \lambda_j(\theta) F_j'(\theta).$$

Finally, define the estimator ϕ_ε^* by

$$\phi_\varepsilon^* = \phi_\varepsilon - \sum_{j=1}^{r} \bar{\lambda}_j(\bar{X}_n) F_{j\varepsilon}, \tag{8.14}$$

where $\bar{X}_n$ and n are the same as in Theorem 7.1, and

$$\bar{\lambda}_j(\theta) = \begin{cases} \lambda_j(\theta), & \text{if } G(\theta) \geq m/2, \\ 0, & \text{if } G(\theta) < m/2. \end{cases}$$

THEOREM 8.2 [Cf. Levit (1975a)]. *Under conditions* (8.11) *to* (8.13),

$$\sup_{\theta\in\Theta} E_\theta \left\| \phi_\varepsilon^* - \phi(\theta) - \varepsilon\phi'(\theta)P_{T(\theta)}\dot{w} \right\|^2 = o(\varepsilon^2)$$

and so the estimator ϕ_ε^* is asymptotically efficient (*see Section* 5).

PROOF. One can deduce Theorem 8.2 from Theorem 8.1, but we will instead deduce it from Theorem 7.1.

It follows from Theorem 7.1 and (8.11) that

$$\begin{aligned} \phi_\varepsilon &= \phi(\theta) + \varepsilon\phi'(\theta)\dot{w} + \alpha_0(\varepsilon)\varepsilon, \\ F_{j\varepsilon} &= \varepsilon F_j'(\theta)\dot{w} + \alpha_j(\varepsilon)\varepsilon, \qquad j = 1,\dots,r, \end{aligned} \tag{8.15}$$

and the random variables α_j satisfy the condition

$$\sum_{j=0}^{r} \sup_\theta E_\theta \alpha_j^2(\varepsilon) = o(1), \qquad \varepsilon \to 0. \tag{8.16}$$

Put $\delta_j(\varepsilon) = \lambda_j(\overline{X}_n) - \lambda_j(\theta)$. Formulas (8.14) and (8.15) imply that

$$\begin{aligned} \phi_\varepsilon^* &= \phi(\theta) + \varepsilon\phi'(\theta)\dot{w} + \alpha_0(\varepsilon)\varepsilon - \varepsilon\sum_{j=1}^{r}\big(\lambda_j(\theta) + \delta_j(\varepsilon)\big)\big(F_j'(\theta)\dot{w} + \alpha_j(\varepsilon)\big) \\ &= \phi(\theta) + \varepsilon\phi'(\theta)P_{T(\theta)}\dot{w} + v(\varepsilon), \end{aligned}$$

where

$$\varepsilon^{-2}E_\theta v^2(\varepsilon) \le c\left(\sum_0^r E_\theta\alpha_j^2(\varepsilon) + \sum_1^r E_\theta\delta_j^2(\varepsilon)\right).$$

Because of (8.16) we need only prove that

$$\lim_{\varepsilon\to 0}\sum_{j=1}^{r} E_\theta\delta_j^2(\varepsilon) = 0. \tag{8.17}$$

Now note that the coefficients $\lambda_j(\theta)$ are determined by the linear equations

$$\sum_{i=1}^{r}\lambda_i(\theta)\big(F_i'(\theta), F_j'(\theta)\big) = \big(\phi'(\theta), F_j'(\theta)\big), \qquad j = 1,\dots,r, \tag{8.18}$$

and that the conditions of the theorem imply that there exist $\alpha \in (0,1)$, $c > 0$ such that

$$\left\| F_j'(\theta) - F_j'(\theta') \right\| \le c\|\theta - \theta'\|^\alpha, \qquad \theta, \theta' \in \Theta. \tag{8.19}$$

It follows from (8.18) and (8.19) that if the Gram determinants $G(\theta) > m/2$, $G(\theta') > m/2$, then

$$\left|\lambda_j(\theta) - \lambda_j(\theta')\right| \le c_1\|\theta - \theta'\|^\alpha, \tag{8.20}$$

where the constant c_1 depends only on c, m and r.

Hence

$$\begin{aligned} E_\theta \delta_i^2(\varepsilon) &\le E_\theta\left\{\delta_i^2(\varepsilon) 1_{\{G(\bar{X}_n) > m/2\}}\right\} + \lambda_i^2(\theta) P_\theta\left\{G(\bar{X}_n) < \frac{m}{2}\right\} \\ &\le c_1 E_\theta \left\| \bar{X}_n - \theta \right\|^\alpha + c_2 P_\theta\left\{\left|G(\theta) - G(\bar{X}_n)\right| > \frac{m}{2}\right\} \\ &\le c_3 E_\theta \left\| \bar{X}_n - \theta \right\|^\alpha. \end{aligned}$$

This inequality together with the inequalities

$$E_\theta \left\| \bar{X}_n - \theta \right\|^2 \le E_\theta \left\| \dot{X}_n - \theta \right\|^2 + o(\varepsilon^2) \le C(n^{-2\beta} + \varepsilon^2 n)$$

and the relation $\varepsilon^2 n \to 0$ imply (8.17), which as we noted before completes the proof. □

8.3. *Examples.*

EXAMPLE 8.1. Suppose that conditions (8.11) are satisfied, that the functionals $F_j(\theta)$ are linear and that the restrictions (8.11) have the form

$$\int_0^1 l_i(t)\theta(t)\,dt - \alpha_j = 0. \tag{8.21}$$

Then condition (8.12) takes the form

$$\det\left\|(l_i, l_j)\right\| > 0, \tag{8.22}$$

and the $\lambda_i(\theta)$ are the solutions of the equations

$$\sum_{i=1}^r \lambda_i(\theta)(l_i, l_j) = \left(\phi'(\theta), l_j\right), \qquad j = 1, \dots, r. \tag{8.23}$$

Then if the function ϕ satisfies the conditions of Theorem 8.2, the estimator

$$\phi_\varepsilon^* = \phi_\varepsilon - \sum_{i=1}^r \lambda_i(\bar{X}_n)\left(\int_0^1 l_i(t)\,dX_\varepsilon(t) - \alpha_i\right)$$

is asymptotically efficient (and asymptotically normal), and

$$\lim_{\varepsilon\to 0} \varepsilon^{-2} E_\theta \left|\phi_\varepsilon^* - \phi(\theta)\right|^2 = \left\|\phi'(\theta) - \sum_1^r \lambda_i(\theta) l_i\right\|^2 = \left\|P_{T(\theta)}\phi'(\theta)\right\|^2.$$

EXAMPLE 8.2. Now suppose that under conditions (8.11) we have only one restriction:

$$F_1(\theta) = F(\theta) = (K\theta, \theta) - \alpha = 0, \qquad \alpha \ne 0.$$

Here we assume K is a bounded symmetric operator. Then

$$F'(\theta) = 2K\theta, \qquad \lambda(\theta) = \frac{(\phi'(\theta), K\theta)}{2\|K\theta\|}.$$

Since $\alpha \neq 0$ and $\Theta \subset O_1$,

$$\|K\theta\| \geq |\alpha| > 0, \qquad \theta \in \Theta,$$

and condition (8.12) is satisfied. The estimator (8.14) then has the form

$$\phi_\varepsilon^* = \phi_\varepsilon - \frac{\left(\phi'(\overline{X}_n), K\overline{X}_n\right)}{2\|K\overline{X}_n\|^2} 1_{\{\|K\dot{X}_n\| > \alpha/2\}} F_\varepsilon .$$

If K is a Hilbert–Schmidt operator with the kernel $K(s,t)$,

$$F_\varepsilon = \int_0^1 \int_0^1 K(s,t)\, dX_\varepsilon(s)\, dX_\varepsilon(t),$$

and the estimator ϕ_ε^* is efficient if the function ϕ satisfies the conditions of Theorem 7.1.

If the operator K is only bounded [e.g., $F(\theta) = \|\theta\|^2 - \alpha$], then the estimator ϕ_ε^* will be asymptotically efficient if $d_n(\Sigma) < cn^{-\beta}$, $\beta > 1/4$, and ϕ satisfies the condition $k + \gamma > 1 + (2\beta)^{-1}$. The quadratic error is

$$E_\theta |\phi_\varepsilon^* - \phi(\theta)|^2 = \varepsilon^2 \left(\|\phi'(\theta)\|^2 - \frac{(\phi'(\theta), K\theta)}{\|K\theta\|^2} \right) + o(\varepsilon^2).$$

EXAMPLE 8.3 (Cf. Example 4.1). Let $\phi(\theta) = \int_0^t \theta(u)\, du$. If there are no restrictions imposed on θ, $X_\varepsilon(t)$ will be an unbiased and asymptotically efficient estimator of $\phi(\theta)$. Now impose on θ, r linear restrictions

$$F_j(\theta) = \int_0^1 l_j(t)\theta(t)\, dt = 0, \qquad j = 1, \dots, r,$$

and suppose for the sake of simplicity that the vectors $l_1, \dots, l_r$ are orthonormal. The derivative $\phi'(\theta)$ is then the integral operator with the kernel

$$\chi_t(s) = \begin{cases} 1, & s < t, \\ 0, & s > t. \end{cases} \tag{8.24}$$

Hence the operator

$$\phi'(\theta) P_{E(\theta)} = \sum_{j=1}^r (l_j, \cdot) \int_0^t l_j(u)\, du$$

and the estimator

$$\phi_\varepsilon^* = X_\varepsilon(t) - \sum_{j=1}^r \int_0^t l_j(u)\, du \int_0^1 l_j(t)\, dX_\varepsilon(t)$$

will be asymptotically efficient. For example,

$$E_\theta \|\phi_\varepsilon^* - \phi(\theta)\|^2 = \varepsilon^2 \left(\frac{1}{2} - \sum_{j=1}^r \int_0^1 dt \left(\int_0^t l_j(u)\, du \right)^2 \right) + o(\varepsilon^2).$$

EXAMPLE 8.4. Let us return to Example 5.1.3. We wish to estimate the solution $\phi(\theta)$ of the differential equation:

$$y' = f(y, \theta, t), \qquad y(0) = y_0,$$

where $f(\cdot, \cdot, \cdot)$ is a known function. In Section 5, we saw that $F'(\theta)$ is the integral operator with the kernel $H_\theta(s, t)$ given by (5.4).

Suppose that the Kolmogorov diameters $d_n(\Theta) \le cn^{-\beta}$, $\beta > 1/2$. Also suppose that the functions f, $\partial f/\partial y$, $\partial^2 f/\partial y^2$, $\partial^2 f/\partial y\, \partial\theta$ *and* $\partial^2 f/\partial \theta^2$ are uniformly bounded in $\theta \in \Theta$. Then

$$\| F'(\theta) \|_2^2 < c < \infty, \qquad \| F'(\theta^2) - F'(\theta_1) \| < c \| \theta_2 - \theta_1 \|$$

and the conditions of Theorem 7.1 are satisfied with $k = 1$, $\gamma = 1$.

Denote by Y_0 the solution of the problem

$$y' = f(y, \overline{X}_n(t)), \qquad y(0) = y_0,$$

and let $Y_1(t)$ be the solution of the following stochastic equation:

$$\begin{aligned} dY_1(t) &= \frac{\partial f}{\partial y}\big(Y_0(t), \dot{X}_n(t), t\big) Y_1(t)\, dt \\ &\quad + \frac{\partial f}{\partial \theta}\big(Y_0(t), \dot{X}_n(t), t\big)\big(dX_\varepsilon(t) - \dot{X}_n(t)\, dt\big), \qquad Y_1(t_0) = 0. \end{aligned}$$

Then the statistic $\phi_\varepsilon(t) = Y_0(t) + Y_1(t)$ is an asymptotically efficient estimator for $\phi(\theta)$. Moreover, the normed differences $\varepsilon^{-1}(\phi_\varepsilon(t) - \phi(\theta, t))$ converge in distribution as $\varepsilon \to 0$ to the solution $U(t)$ of the stochastic differential equation

$$dU(t) = \frac{\partial f}{\partial y}(\phi(\theta, t), \theta(t), t) U(t)\, dt + \frac{\partial f}{\partial \theta}(\phi(\theta, t), \theta(t), t)\, dw(t).$$

An efficient estimator for an estimation problem closely connected with the problem of this example is constructed in Korostelev (1988).

Acknowledgments. W. R. van Zwet suggested we write a paper for the *Annals of Statistics*, based on a lecture we gave at the First Congress of the Bernoulli Society, Tashkent, 1986. We are grateful to him both for this suggestion and also for the patience he showed while he waited for the promised paper. This paper also contains a large part of Ibragimov's Wald Lecture of 1989. We would also like to thank V. N. Sudakov and M. A. Lifschitz for some helpful suggestions and useful discussions while this paper was being written. Special thanks are due to A. N. Borodin and A. Yu. Zaitsev for their valuable help. With great pleasure we thank Professor Mark Low of the University of California, Berkeley, who provided extensive sets of valuable suggestions and enormously improved our English.

REFERENCES

BEGUN, J. M., HALL, W. J., HUANG, W. M. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.

BERAN, R. J. (1977). Estimating a distribution function. *Ann. Statist.* **5** 400–404.

EHRHARD, A. (1983). Symmetrization dans l'espace de Gauss. *Math. Scand.* **53** 281–301.

GEMAN, S. and HWANG, C. R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10** 410–414.

GIKHMAN, I. I. and SKOROKHOD, A. V. (1979). *The Theory of Stochastic Processes* **8** (translated from the Russian). Springer, New York.

GILL, G. D. (1986). Non- and semi-parametric maximum likelihood estimators and the von Mises method. I. Technical Report MS-R8604, Centrum voor Wiskunde en Informatica, Amsterdam.

GRENANDER, U. (1981). *Abstract Inference*. Wiley, New York.

HÁJEK, J. (1970). A characterisation of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* **14** 323–330.

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194. Univ. California Press, Berkeley.

HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.

HAS'MINSKII, R. Z. and IBRAGIMOV, I. A. (1979). On the nonparametrical estimation of functionals. In *Proceedings of the Second Prague Symposium on Asymptotic Statistics* (P. Mandl and M. Huskova, eds.) 41–51. North-Holland, Amsterdam.

HAS'MINSKII, R. Z. and IBRAGIMOV, I. A. (1980). Some estimation problems for stochastic differential equations. *Lecture Notes Control Inform. Sci.* **25** 1–12. Springer, New York.

HAS'MINSKII, R. Z. and IBRAGIMOV, I. A. (1986). Asymptotically efficient nonparametric estimation of functionals of a spectral density function. *Probab. Theory Related Fields* **73** 447–461.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1977). A problem of statistical estimation in Gaussian noise. *Dokl. Akad. Nauk SSSR* **236** 1053–1055.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, New York.

IBRAGIMOV, I. A., NEMIROVSKII, A. S. and HAS'MINSKII, R. Z. (1986). Some problems on nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.* **31** 391–406.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1987). Estimation of linear functionals in Gaussian noise. *Theory Probab. Appl.* **32** 30–39.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

KOROSTELEV, A. P. (1988). Minimax filtration of the trajectory of a dynamic system depending on a nonparametric signal. *Automat. i Teleme.* **2** 252–264. (In Russian.)

LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications in Statistics* **1** 277–330.

LE CAM, L. (1956). On the asymptotic theory of estimation and testing hypotheses. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 129–156. Univ. California Press, Berkeley.

LE CAM, L. (1960). Locally asymptotically normal families of distributions. *University of California Publications in Statistics* **3** 37–98.

LE CAM, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 245–261. Univ. California Press, Berkeley.

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

LEVIT, B. (1974). On optimality of some statistical estimates. In *Proceedings of the Prague Symposium on Asymptotic Statistics* (J. Hájek, ed.) **2** 215–238. Univ. Karlova, Prague.

LEVIT, B. YA. (1975a). Conditional estimation of linear functionals. *Problems Inform. Transmission* **11** 291–302.

LEVIT, B. YA. (1975b). Efficiency of a class of nonparametric estimates. *Theory Probab. Appl.* **20** 738–754.

LEVIT, B. YA. (1978). Infinite-dimensional informational inequalities. *Theory Probab. Appl.* **23** 371–377.

McKEAN, H. P. (1969). *Stochastic Integrals*. Academic, New York.

MEYER, P. A. (1976). Un cours sur les intégrales stochastique. *Séminaire de Probabilités X. Lecture Notes in Math.* **511** 332–353. Springer, New York.

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

MILLAR, P. W. (1983). The minimax principle in asymptotic statistical theory. *Sèminaire de Probabilitès. Lecture Notes in Math.* **976**. Springer, New York.

PASTUCHOVA, YU. I. and HAS'MINSKII, R. Z. (1988). Nonparametric estimation of a linear functional of the regression function for a given observation design. *Problems Inform. Transmission* **24** 215–223.

PFANZAGL, J. (1982). *Contributions to a General Asymptotic Theory*. Springer, New York.

SKOROKHOD, A. V. (1974). Integration in Hilbert space. *Ergebnisse des Math.* **79**. Springer, New York.

WELLNER, J. A. (1985). Semiparametric models: Progress and problems. *CWI Newslett.* **9** 2–24.

STEKLOV MATHEMATICAL INSTITUTE
USSR ACADEMY OF SCIENCE
27 FONTANKA
191011 LENINGRAD
USSR

INSTITUTE OF INFORMATION TRANSMISSION
USSR ACADEMY OF SCIENCE
ERMOLOVOY 19
MOSCOW 101447
USSR