*The Annals of Statistics*
1993, Vol. 21, No. 2, 798–829

# GOODNESS OF FIT PROBLEM AND SCANNING INNOVATION MARTINGALES

By E. V. Khmaladze

*Razmadze Mathematical Institute and Mathematical Institute V. A. Steklova*

This paper is mainly devoted to the following statistical problem: in the case of random variables of any finite dimension and both simple or parametric hypotheses, how to construct convenient "empirical" processes which could provide the basis for goodness of fit tests—more or less in the same way as the uniform empirical process does in the case of simple hypothesis and scalar random variables.

The solution of this problem is connected here with the theory of multiparameter martingales and the theory of function-parametric processes. Namely, for the limiting Gaussian processes some kind of filtration is introduced and so-called scanning innovation processes are constructed—the adapted standard Wiener processes in one-to-one correspondence with initial Gaussian processes. This is done for the function-parametric versions of the processes.

**1. Introduction.** This paper deals with three topics that usually are not very much associated: goodness of fit theory; innovation martingales for Gaussian processes with $m$-dimensional time parameter; theory of function-parametric empirical processes.

Namely, consider i.i.d. random vectors $X_1, \dots, X_n$ taking values in $m$-dimensional Euclidean space $\mathbb{R}^m$, and denote by $\mathbb{F} = \{F(\cdot, \theta),\ \theta \in \Theta\}$ a parametric family of distributions in $\mathbb{R}^m$. If $\Theta$ contains only one point $\theta_0$, let us write $F_0$ instead of $\mathbb{F}$. Denote by $F$ the unknown distribution of each $X_i$. It is well known that the so-called uniform empirical process (1.1) plays a fundamental role in the theory of goodness of fit tests for testing hypotheses concerning $F$. However, it does so only in the case of testing a simple hypothesis $F = F_0$ for scalar random variables ($m = 1$). The first and main aim of this paper is to introduce an empirical process of some kind, which can play a role similar to that of the uniform empirical process but for both simple ($F = F_0$) and parametric ($F \in \mathbb{F}$) hypotheses and for any finite-dimensional random vectors ($1 \le m < \infty$).

This empirical process is derived on the basis of some "innovation" reasoning for the "usual" empirical process $v_n$,

$$v_n(x) = \sqrt{n}\,[F_n(x) - F_0(x)]$$

Received May 1990; revised September 1992.

*AMS* 1991 *subject classifications*. 62G10, 62F03.

*Key words and phrases*. Empirical processes, parametric empirical process, goodness of fit tests, asymptotically distribution-free processes, contiguous alternatives, innovation process, increasing family of projectors, multiparameter martingales, function-parametric martingales.

and for the parametric empirical process

$$v_n(x, \hat{\theta}) = \sqrt{n}\left[F_n(x) - F(x, \hat{\theta})\right].$$

In connection with this the second aim of the paper is to discuss what could be understood as innovation martingale processes with multidimensional time parameter. We will see that these innovation martingales—we call them scanning innovations—can be introduced even in the case of an infinite-dimensional time parameter, that is, for function-parametric processes.

The formal setting of the problem will be given in Section 3. Here in the introduction we will continue with an informal discussion of both aims.

*Goodness of fit theory.* Somewhere in the beginning of the thirties, Kolmogorov realized that if the scalar random variable $X$ has continuous distribution function $F$, then the random variable $F(X)$ has the uniform distribution on [0, 1]. He used this observation in the lemma of his well-known 1933 paper: Let $\phi_n(\lambda)$ denote the probability of the inequality

$$\sup|F_n(x) - F(x)| < \lambda/\sqrt{n}.$$

LEMMA [Kolmogorov (1933) or (1986)]. *The distribution function $\phi_n(\lambda)$ does not depend on $F$ if $F$ is continuous.*

The eventual logical mastering of the transformation $U = F(X)$ is connected with Doob (1949), where the uniform empirical process $u_n$ appeared to everyone's sight:

$$u_n(t) = v_n(x), \qquad t = F_0(x). \tag{1.1}$$

Since the process $u_n$ can be viewed as an empirical process based on independent uniformly distributed random variables $U_i = F(X_i)$, $i = 1, \dots, n$, the distribution of $u_n$ does not depend on $F_0$. Therefore if one chooses as the test statistic a functional $\psi[v_n, F_0]$ of $v_n$ and $F_0$, which could be represented as a functional $\phi[u_n]$ of $u_n$ only,

$$\psi[v_n, F_0] = \phi[u_n],$$

the distribution of such a statistic is free from $F_0$. In the whole subsequent development of the theory of goodness of fit tests, such a choice of test statistic became the universal principle.

Why is it so important to use distribution-free—hence, asymptotically distribution free—statistics? To clarify this let us remark that there are two different kinds of tests. The tests of the first kind are based on one or a "few" linear functionals of $v_n$. Examples are the Neyman–Pearson statistic

$$\frac{1}{\sqrt{n}}\sum_1^n\left[\ln\frac{dA}{dF_0}(X_i) - E\ln\frac{dA}{dF_0}(X_i)\right] = \int \ln\frac{dA}{dF_0}(x)v_n(dx)$$

(where $A$ denotes the alternative distribution of $X_i$), Student's statistic

$$\frac{\sqrt{n}\left(\bar{X} - EX_i\right)}{S_n} \approx \frac{1}{\sigma}\int x v_n(dx),$$

statistics of the $F$-test, statistics of $C_\alpha$-tests and so on. The asymptotic distribution of a linear statistic is "usually" the normal distribution and the calculation of asymptotic levels of such tests is simple. Therefore it is completely unimportant whether we represent these statistics as functionals of $u_n$ or not.

Tests based on one or a "few" linear functionals are particularly sensitive to deviations from $F_0$ in one or a "few" directions, but they are very insensitive to deviations in all other directions (see Section 2 for a precise statement for contiguous alternatives). Tests of the second kind—the goodness of fit tests—are of different behavior. These tests are usually not most sensitive to any particular deviation from $F_0$ but they have at least "some" sensitivity to "all" deviations from $F_0$.

Statistics of these tests are essentially nonlinear functionals of $v_n$. The calculation of the limit distribution of these functionals is a serious and complicated mathematical problem. Examples like the (weighted) Kolmogorov–Smirnov statistic or (weighted) Cramér–von Mises statistic are well known. Recall that it was quite a difficult task to derive and to calculate the limit distribution of each of these statistics. It is hard even now to calculate the limit distributions of weighted Kolmogorov–Smirnov or weighted $\omega^2$ statistics except for a few special weight functions.

Because of this it is of prime practical importance that we must calculate the limit distribution of each functional $\psi[v_n, F_0] = \phi[u_n]$ only once for all continuous distribution functions $F_0$.

However, since Simpson (1951) and Rosenblatt (1952) it became clear that the transformation (1.1) does not lead to distribution-free processes if the $X_i$'s are $m$-dimensional random vectors with $m \geq 2$. After the work of Gikhman (1953, 1954) and Kac, Kiefer and Wolfowitz (1955) it became clear that, in the case of a parametric hypothesis $F \in \mathbb{F}$ and $m = 1$, if we consider the natural analogue of (1.1)

$$\hat{u}_n(t) = v_n(x, \hat{\theta}), \qquad t = F(x, \hat{\theta}),$$

where $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ is an estimator of the unknown value of the parameter $\theta$, it does not lead to distribution-free or asymptotically distribution-free processes as well (see Section 2). As a consequence, the classical statistics like

$$\sup_x |v_n(x, \hat{\theta})| \quad \text{or} \quad \int v_n^2(x, \hat{\theta}) F(dx, \hat{\theta})$$

have limit distributions depending on $\mathbb{F}$ (and even on the true parameter value $\theta$, in general).

Because of these difficulties there were few, if any, attempts to develop systematically asymptotically distribution-free goodness of fit tests for testing a parametric hypothesis in $\mathbb{R}^m$, $m \geq 2$.

The main purposes of this paper are (a) to formulate the mathematical problem of finding "proper" asymptotically distribution-free processes which can play a role similar to that of the uniform empirical process $u_n$ (see Section 3), and then (b) to propose one solution of this problem for all four cases: $m = 1$, $F = F_0$ (simple hypothesis); $m = 1$, $F \in \mathbb{F}$ (parametric hypothesis); $m \geq 2$, $F = F_0$; $m \geq 2$, $F \in \mathbb{F}$.

*Innovation for function-parametric processes.* Under the hypothesis $F \in \mathbb{F}$, the limit distribution of the parametric empirical process $v_n(\cdot, \hat{\theta})$ is some 0-mean Gaussian process $\hat{v}$ (see Section 2). Put $m = 1$ and transform the process $\hat{v}$ to its innovation martingale $\hat{w}$ [see definitions in, e.g., Lipzer and Shiryayev (1977)], which is a Gaussian process with independent increments and covariance function $F(x \wedge y, \theta)$, where $\theta$ denotes the "true value" of the parameter. Now transform $\hat{w}$ to the standard Wiener process $w$, which is an easy step. In the resulting transformation of $\hat{v}$ to $w$ substitute $v_n(\cdot, \hat{\theta})$ instead of $\hat{v}$. What we get will be a process $w_n$ which converges in distribution under the hypothesis to a standard Wiener process $w$. Hence $w_n$ is an asymptotically distribution-free process (and possesses other desired properties). Just this was the solution described in Khmaladze (1981) for the case $m = 1$, $F \in \mathbb{F}$.

However, attempts to develop a similar approach in the case $m \geq 2$, even for the simple hypothesis $F = F_0$, did not have success for quite a long time [until as late as Khmaladze (1987) and Nikabadze and Khmaladze (1987)]. The problem is that it is not clear how to construct and even what to call an innovation process for processes with multidimensional time parameter $x$.

This problem was illustrated to some extent in Khmaladze [(1988), Example 3]. In the present paper we want to construct a scanning innovation process not only for finite-dimensional $x$, but also when we use functions $f$ or sets $A$ in place of a time parameter. We do this not for the sake of formal generality, but to see better the true nature of scanning innovations and to serve some practical needs explained in Section 3. Our starting point could be illustrated even in the one-dimensional case and very simple limiting Gaussian process—let it be just Brownian bridge $u = \{u(t),\ t \in [0,1]\}$. Let us equip it with filtration $\mathscr{F} = \{\mathscr{F}(t),\ t \in [0,1]\}$ which is formed by $\sigma$-algebras $\mathscr{F}_t = \sigma\{u(s),\ s \leq t\}$. Now the process $\{u, \mathscr{F}\}$ is $\mathscr{F}$-adapted [i.e., each $u(t)$ is a $\mathscr{F}_t$-measurable random variable] and possesses the innovation Wiener process $\{w, \mathscr{F}\}$

$$w(t) = u(t) + \int_0^t \frac{u(s)}{1-s}\, ds.$$

But along with $\{u, \mathscr{F}\}$ one may consider function-parametric or set-parametric versions of Brownian bridge

$$u(f) = \int_0^1 f(t)\, du(t), \qquad f \in \Phi \subset L_2[0,1], \tag{1.2}$$

$$u(A) = \int_0^1 I\{t \in A\}\, du(t), \qquad A \in \mathscr{A}. \tag{1.3}$$

Although to a great extent the processes $u$, (1.2) and (1.3) represent equivalent objects, the adapted process $\{u, \mathscr{F}\}$ differs, say, from $\{u(A),\ A \in \mathscr{A}\}$ in two points: the range $\mathscr{A}$ of time parameter $A$ should not be necessarily a linearly ordered class of sets, and $\{u(A),\ A \in \mathscr{A}\}$ is not $\mathscr{F}$-adapted.

Nevertheless the natural desire is to call the process

$$w(A) = \int_0^1 I\{t \in A\}\, dw(t), \qquad A \in \mathscr{A},$$

the innovation of $\{u(A),\ A \in \mathscr{A}, \mathscr{F}\}$. Hence we have process $\{u(A),\ A \in \mathscr{A}\}$ with more or less general range of time parameter which is equipped with linearly ordered filtration $\mathscr{F}$ and which, presumably, possesses sensible innovation process $\{w(A),\ A \in \mathscr{A}\}$.

In Section 3 we will see how to transform $\{u(A),\ A \in \mathscr{A}, \mathscr{F}\}$ to $\{w(A),\ A \in \mathscr{A}, \mathscr{F}\}$ without intermediate mention of the adapted process $\{u(t),\ t \in [0,1], \mathscr{F}\}$ (see Example 2). More generally, we consider the function-parametric process

$$\hat{v}(f) = \int_{R^m} f(x)\hat{v}(dx), \qquad f \in \Phi,$$

along with linearly ordered filtration $\mathscr{F}$ and construct for it a scanning innovation.

Two more things should be noted.

First, for the convenience of a reader familiar with existing theory of martingales with multidimensional time parameter, one should remark that the important condition (F4) of the basic paper by Cairoli and Walsh (1975) [see also Wong and Zakai (1974)] is satisfied neither for empirical processes $v_n(\cdot)$ and $v_n(\cdot, \hat{\theta})$ nor for limiting Gaussian processes $v$ and $\hat{v}$.

Second, starting with Lévy (1948) one of the main questions in innovation theory of Gaussian processes was this: What are the conditions on the covariance function $R$ of a Gaussian process which guarantee the existence of innovation of this process? In well-known papers [Shepp (1966) and Hitsuda (1968)] one can find necessary and sufficient condition for the one-dimensional time case:

$$R(t,s) = t \wedge s - \int_0^t \int_0^s k(\tau, \sigma)\, d\tau\, d\sigma, \tag{1.4}$$

with $k(\cdot, \cdot)$ being a Hilbert–Schmidt kernel which has no eigenvalue equal to 1. Our processes $v$ and $\hat{v}$ have covariance functions of just this form for any $m \geq 1$, and they do possess scanning innovation martingales as we shall see in Section 3. However, the author believes that, for $m \geq 2$, condition (1.4) is *sufficient* for existence of scanning innovation for a Gaussian process, but it is *not necessary* any more. This was demonstrated recently by McKeague, Nikabadze and Sun (1992)—they constructed scanning innovations for processes with covariance function

$$(t \wedge s)F(x' \wedge y') - (t \wedge s)\int_{-\infty}^{x'}\int_{-\infty}^{y'} k(\xi,\eta)\,d\xi\,d\eta,$$

$$t, s \in [0,1],\ x', y' \in \mathbb{R}^{m-1},$$

and to similar processes.

Several useful references in innovation theory are Cramér (1964), Rozanov (1974), Lipzer and Shiryayev (1977) and Gohberg and Krein (1967).

**2. Convergence in distribution of the parametric empirical processes $\hat{v}_n(\cdot,\hat{\theta})$; the description of limiting process $\hat{v}$ as a projection; consequences and remarks.** In this section we collect some of the definitions, assumptions and statements used throughout the paper. We avoid all proofs—partly because many of these statements are known, partly to make the paper shorter. The longer version of this section with proofs is given in Section 2 of Khmaladze (1989).

*The parametric family* $\mathbb{F}$. Suppose the range $\Theta$ of $\theta$ to be an open subset of $\mathbb{R}^k$. Assume that each $F(\cdot,\theta)$ is absolutely continuous w.r.t. Lebesgue measure and the corresponding densities $f(\cdot,\theta)$ have the following regularity properties.

CONDITION 1. The $k$-dimensional vector function

$$q(x,\theta) = \frac{\partial}{\partial\theta}\ln f(x,\theta)$$

is square integrable:

$$\int q^T(x,\theta)q(x,\theta)F(dx,\theta) < \infty.$$

As a consequence, the Fisher information matrix

$$B(\theta) = \int q(x,\theta)q^T(x,\theta)F(dx,\theta)$$

is finite. Here and throughout this paper $\alpha^T$ means the transpose of the column vector $\alpha$.

CONDITION 2. If a $k$-dimensional vector function $\xi$ has coordinates

$$\xi_m(x,\theta,\varepsilon) = \sup_{\mathscr{O}:\,\|\mathscr{O}-\theta\|<\varepsilon} |q_m(x,\mathscr{O}) - q_m(x,\theta)|,$$

where $q_m$ is the $m$th coordinate of the vector function $q$, then

$$\int \xi^T(x,\theta,\varepsilon)\xi(x,\theta,\varepsilon)F(dx,\theta) \to 0 \quad \text{as } \varepsilon \to 0.$$

A family $\mathbb{F}$ with these properties (Conditions 1 and 2) we will call *regular*.

Condition 2 is more or less traditionally used in asymptotic statistics [cf. condition c of Ibragimov and Has'minskii (1981), Chapter 1, Section 7, or Definition 2 of Pollard (1984), Chapter VII]. We need it to estimate the remainder in Lemma 2.1.

As test statistics for testing the parametric hypothesis $F \in \mathbb{F}$, let us consider functionals of the so-called parametric empirical process

$$v_n(x,\hat{\theta}) = \sqrt{n}\left[F_n(x) - F(x,\hat{\theta})\right], \qquad F_n(x) = \frac{1}{n}\sum_1^n I\{X_i \le x\},$$

where $\hat{\theta} = \hat{\theta}(X_1,\dots,X_n)$ is an estimate of the unknown parameter value. Let us clarify the asymptotic behavior of $v_n(\cdot,\hat{\theta})$ as $n \to \infty$. To do this we need some assumption on the asymptotic behavior of $\hat{\theta}$.

*The estimator* $\hat{\theta}$. Suppose the following condition holds:

CONDITION 3. There exists a $k$-dimensional vector function $l(\cdot,\theta)$ such that for each $\theta \in \Theta$

$$\int l^T(x,\theta)l(x,\theta)F(dx,\theta) < \infty,$$

and

$$\int g(x,\theta)l^T(x,\theta)F(dx,\theta) = I_k, \qquad \int l(x,\theta)F(dx,\theta) = 0, \tag{2.1}$$

where $I_k$ is the $k \times k$ identity matrix and

$$\sqrt{n}(\hat{\theta} - \theta) = \int l(x,\theta)v_n(dx,\theta) + o_p(1), \qquad n \to \infty.$$

An estimator $\hat{\theta}$, which satisfies Condition 3 we call *projective* [cf. Khmaladze (1979)]. The reason for this definition is explained by Lemma 2.2.

We will formulate limit theorems for $v_n$ both under the hypothesis and under contiguous alternatives. Let us describe the alternative sequences of distributions precisely.

*The alternatives* $A_n$. Under the alternative assume that for each $n = 1,2,\dots$ the random vectors $X_1,\dots,X_n$ are again i.i.d. with distribution $A_n$, which has the following properties.

CONDITION 4. There exists $F(\cdot,\theta) \in \mathbb{F}$ such that if $A_n = A_n^c + A_n^s$ is the Lebesgue decomposition of $A_n$ into absolutely continuous and singular parts w.r.t. $F(\cdot, \theta)$, then

$$n \operatorname{var}(A_n^s) \to 0, \qquad n \to \infty \tag{2.2}$$

$$\left[\frac{dA_n^c(\cdot)}{dF(\cdot, \theta)}\right]^{1/2} = 1 + \frac{1}{2\sqrt{n}} h_n(\cdot) \tag{2.3}$$

and

$$\int [h_n(x) - h(x)]^2 F(dx, \theta) \to 0 \tag{2.4}$$

for some function $h(\cdot)$, such that

$$\int h^2(x) F(dx, \theta) < \infty, \tag{2.5}$$

$$\int h(x) F(dx, \theta) = 0 \tag{2.6}$$

and

$$\int h(x) q(x, \theta) F(dx, \theta) = 0. \tag{2.7}$$

Hence, under the hypothesis, the distribution of the sample $X_1, \dots, X_n$ is the $n$-fold direct product $\mathbb{P}_{n\theta} = F(\cdot, \theta) \times \cdots \times F(\cdot, \theta)$ with some $F(\cdot, \theta) \in \mathbb{F}$, while under each particular sequence of alternatives the distribution of this sample is $\tilde{\mathbb{P}}_{n\theta}(h) = A_n \times \cdots \times A_n$. Condition 4 guarantees that the sequence $\{\tilde{\mathbb{P}}_{n\theta}(h)\}$ is contiguous w.r.t. the sequence $\{\mathbb{P}_{n\theta}\}$: $\tilde{\mathbb{P}}_{n\theta} \lhd \mathbb{P}_{n\theta}$ [cf. Oosterhoff and van Zwet (1979)]. The function $h$ which participates in these conditions can be viewed as a function which determines from what "direction" the alternative distribution $A_n$ approaches some hypothetical distribution $F(\cdot, \theta)$.

REMARK. Any function $h$ which satisfies (2.4) and (2.5) must satisfy condition (2.6), but the orthogonality condition (2.7) is an additional requirement on $h$. This requirement is convenient and natural, as can be seen later, but is not necessary for further development.

*Function-parametric processes.* From now on we will systematically adopt function-parametric versions of the processes involved, which we introduce here formally. Denote by $L_2(\theta)$ the space of functions with norm

$$\|f\| = \|f\|_\theta = \left[\int f^2(x) F(dx, \theta)\right]^{1/2},$$

and scalar product

$$(f, \phi) = (f, \phi)_\theta = \int f(x) \phi(x) F(dx, \theta).$$

For any $f \in L_2(\theta)$ let

$$(2.8)\quad v_n(f) = \int f(x) v_n(dx, \theta) = \frac{1}{\sqrt{n}} \sum_1^n \left[ f(X_i) - \int f(x) F(dx, \theta) \right].$$

Clearly $v_n(f)$ for each $f \in L_2(\theta)$ is a random variable with finite variance,

$$Ev_n^2(f) = (f, f) - (f, 1)^2.$$

Moreover,

$$Ev_n(f) v_n(\phi) = (f, \phi) - (f, 1)(\phi, 1).$$

If $f$ is a vector function, then clearly $v_n(f)$ is a random vector. In particular, Condition 3 says that

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) = v_n(l(\cdot, \theta)) + o_p(1).$$

Similarly for any $f \in \bigcap_{\theta \in \Theta} L_2(\theta)$ let

$$\hat{v}_n(f) = \int f(x) v_n(dx, \hat{\theta}).$$

Notice that $\bigcap_{\theta \in \Theta} L_2(\theta)$ is frequently quite a rich set. If, for example, $\mathbb{F}$ is the family of normal distributions with shift parameter $\theta$, this set contains all functions with finite variance under normal distribution with any mean $\theta$. Also, all indicator functions as well as all bounded functions belong to $L_2$ for any $\mathbb{F}$.

Now let

$$(2.9)\qquad \Pi_1 v_n(f) = v_n(f) - (f, q^T(\cdot, \theta)) v_n(l(\cdot, \theta)),$$

which is the expansion of $\hat{v}_n(f)$ in $\hat{\theta}$ up to the linear term.

LEMMA 2.1. *Let $\mathbb{F}$ be a regular parametric family (of distributions). Then*

$$|\hat{v}_n(f) - \Pi_1 v_n(f)| \le \varepsilon_n \|f\|_\theta, \quad \textit{with } \varepsilon_n \to_{\mathbb{P}} 0,$$

*both under $\mathbb{P}_{n\theta}$ and $\tilde{\mathbb{P}}_{n\theta}(h)$.*

It is not difficult to observe that

$$(2.10)\qquad \Pi_1 v_n(f) = v_n(\Pi_1^* f)$$

where

$$\Pi_1^* f = f - (f, q^T(\cdot, \theta)) l(\cdot, \theta).$$

Notice that $v_n(f)$ is a bilinear functional—for each $f$ it is a linear functional of the trajectories of the empirical process $v_n(\cdot, \theta)$, and for each trajectory of $v_n(\cdot, \theta)$ it is a linear functional of $f$. Then the equality (2.10) simply says that $\Pi_1^*$ is the adjoint projector of the projector $\Pi_1$ in the bilinear functional (2.8).

*The limiting Gaussian processes $v$ and $\hat{v}$.* Denote by $b(\cdot, \theta)$ the Gaussian process with mean 0 and covariance function $F(x \wedge y, \theta)$, and, for each

$f \in L_2(\theta)$, let

$$(2.11) \qquad b(f) = \int f(x) b(dx, \theta),$$

$$(2.12) \qquad v(f) = b(f) - (f, 1) b(1),$$

$$(2.13) \qquad \hat{v}(f) = v(f) - (f, q^T(\cdot, \theta)) v(l(\cdot, \theta)) = \Pi_1 v(f).$$

It is convenient to introduce extended vector functions

$$(2.14) \qquad q(x) = \begin{pmatrix} 1 \\ q(x, \theta) \end{pmatrix}, \qquad l(x) = \begin{pmatrix} 1 \\ l(x, \theta) \end{pmatrix}$$

and to substitute (2.12) in (2.13)—if (2.1) is satisfied, then

$$(2.15) \qquad \hat{v}(f) = \Pi b(f) = b(f) - (f, q^T) b(l) = b(\Pi^* f)$$

with

$$(2.16) \qquad \Pi^* f = f - (f, q^T) l.$$

LEMMA 2.2. *The transformation* (2.12) *of $b$ to $v$ is a projection. If* (2.1) *is satisfied, then the transformation $\Pi_1$ defined by* (2.9) *is a projection*: $\Pi_1 \Pi_1 = \Pi_1$. *Consequently, if* (2.1) *is satisfied, then the transformation* (2.15) *of $b$ to $\hat{v}$ is a projection*: $\Pi\Pi = \Pi$.

REMARK. The study of $\hat{v}$ as a projection of $b$ does not lie in the main stream of this paper. That is why we avoid here a more rigorous description of $\Pi$. More precise discussion can be found, for example, in Khmaladze (1979). Earlier the description of $\hat{v}$ as a projection of $v$ in the case of the maximum likelihood estimator $\hat{\theta}$ was mentioned in Tyurin (1970).

Denote by $C$ the extended Fisher information matrix

$$C = (q, q^T) = \begin{pmatrix} 1 & 0 \\ 0 & B(\theta) \end{pmatrix}.$$

A condition given later (4, Section 3) will guarantee that $C$ has the unique inverse $C^{-1}$. Let us consider then the special choice of the function $l$:

$$(2.17) \qquad \bar{l} = C^{-1} q$$

and denote

$$(2.18) \qquad \bar{v}(f) = b(f) - (f, q^T) C^{-1} b(q).$$

Remark that $\bar{v}$ is the orthogonal projection of $b$ while $\hat{v}$ is, in general, a skew projection. The choice of $\bar{l}$ corresponds to the case when $\hat{\theta}$ is the maximum likelihood estimator.

We are ready now to formulate the statement concerning convergence in distribution of $v_n$ and $\hat{v}_n$. Let $\mathscr{F}$ be some subset of $L_2(\theta)$, and denote by $\mathscr{X}(\mathscr{F})$ the space of bounded functions $x(f)$, $f \in \mathscr{F}$, with the norm

$\sup_{f \in \mathscr{F}}|x(f)|$ [cf. Section VII.5 of Pollard (1984)]. Let $H(f) = (f, h)$ and

$$\hat{H}(f) = \Pi H(f) = H(f) - (f, q^T)H(l) = H(\Pi^* f).$$

THEOREM 2.3. *Suppose $\mathscr{F}$ is a compact subset in $L_2(\theta)$ such that $|f(x)| < c$ for the same constant c for all $f \in \mathscr{F}$, and that in the space $\mathscr{X}$ as $n \to \infty$,*

$$v_n \to_{\mathscr{D}(\mathbb{P}_{n\theta})} v \quad and \quad v_n \to_{\mathscr{D}(\tilde{\mathbb{P}}_{n\theta}(h))} v + H.$$

*If $\mathbb{F}$ is regular and $\hat{\theta}$ is projective, then*

$$\hat{v}_n \to_{\mathscr{D}(\mathbb{P}_{n\theta}(h))} \hat{v} \quad and \quad \hat{v}_n \to_{\mathscr{D}(\tilde{\mathbb{P}}_{n\theta}(h))} \hat{v} + \hat{H}.$$

REMARK. Convergence in distribution of the parametric empirical process $\hat{v}_n$, as well as of the scanning innovation process $w_n$ of Section 3, is the consequence of the convergence of $v_n$ and regularity of $\mathbb{F}$. To make it clearer, we kept convergence of $v_n$ as a condition of the theorem. [The convergence is true, e.g., if $\mathscr{F} = \{I\{\cdot \le x\}, x \in \mathbb{R}^k\}$ and $\{A_n\}$ satisfies conditions (2.1)–(2.6).]

The assumption of the pointwise boundedness of functions is rather specific. But for the purposes of the present paper, this is sufficient and will help in the proof of the convergence theorem of Section 3.

REMARK. If the function $f$ is fixed (the condition of boundedness is then not necessary), we get from Theorem 2.3 some support for the informal reasoning in the introduction: $v_n(f)$ and $\hat{v}_n(f)$ are asymptotically Gaussian indeed, and for all sequences of alternatives $\tilde{\mathbb{P}}_{n\theta}(h)$ such that

$$(f, h) = 0 \quad [\text{and } (f, q) = 0],$$

the limiting distribution of these linear statistics is the same as under the hypothesis. Hence, these statistics cannot distinguish between $\mathbb{P}_{n\theta}$ and all such $\tilde{\mathbb{P}}_{n\theta}(h)$, although they are asymptotically most "sensitive" to specific $h =$ const. $f$ (cf. next subsection).

*Distance in variation.* We will need also statements to judge how "sensitive" the processes $v_n$ and $\hat{v}_n$ are to the alternatives $\tilde{\mathbb{P}}_{n\theta}(h)$. First let us see "how far" the sequences $\{\mathbb{P}_{n\theta}\}$ and $\{\tilde{\mathbb{P}}_{n\theta}(h)\}$ are from each other. Denote $d(P, Q)$ the distance in variation between distributions $P$ and $Q$:

$$d(P, Q) = \sup_{B \in \mathscr{B}} |P(B) - Q(B)|,$$

where $\mathscr{B}$ is the $\sigma$-algebra, on which $P$ and $Q$ are defined. Let $\Phi$ be the standard normal distribution function and

$$\lambda(h) = 2\Phi(\|h\|/2) - 1, \qquad \|h\| = (h, h)^{1/2}.$$

LEMMA 2.4. *If the sequence $\{A_n\}$ satisfies conditions* (2.2)–(2.6), *then*

$$d\big(\tilde{\mathbb{P}}_{n\theta}(h), \mathbb{P}_{n\theta}\big) \to \lambda(h), \qquad n \to \infty.$$

Now turn to the processes $v_n(\cdot, \theta)$ and $v_n(\cdot, \hat{\theta})$. Let $P^{\xi}$ denote the distribution of a process $\xi$ (a random variable $\xi$).

For two Gaussian processes $\xi = \{\xi(f),\ f \in \mathscr{I}\}$ and $\eta = \{\eta(f),\ f \in \mathscr{I}\}$, let us define the distance in variation between $P^{\xi}$ and $P^{\eta}$ as

$$d(P^{\xi}, P^{\eta}) = \max\{d(P^{\xi(f)}, P^{\eta(f)}),\ f \in \Lambda(\mathscr{I})\}$$

where $\Lambda(\mathscr{I})$ denotes the closed linear span of $\mathscr{I}$ and $P^{\xi(f)}$ and $P^{\eta(f)}$ are Gaussian distributions of random variables $\xi(f)$ and $\eta(f)$, respectively.

LEMMA 2.5. *$d(P^{v}, P^{v+H}) = \lambda(h)$. If (2.7) is satisfied, then $d(P^{\hat{v}}, P^{\hat{v}+H}) = \lambda(h)$.*

Now we have prepared everything we will need in Section 3.

**3. Formulation of the problem; scanning innovations; function-parametric version.** Let us consider again the classical transformation of the empirical process $v_n$, based on scalar random variables (i.e., $m = 1$), to the uniform empirical process $u_n$:

$$u_n(t) = \mathscr{K}[v_n, F_0](t) = v_n\big(F_0^{-1}(t)\big). \tag{3.1}$$

It is common knowledge that $\mathscr{K}$ transforms $v_n$ to a distribution-free—hence, asymptotically distribution-free—process, but this cannot be the only important property of the transformation $\mathscr{K}$. An alternative property of $\mathscr{K}$ is that in the process $u_n$ "the whole information is preserved" that helps "to distinguish" between the hypothesis and alternatives. If we focus on contiguous alternatives, this property formally can be expressed by Lemmas 2.4 and 3.1.

LEMMA 3.1. *Let $u$ be a standard Brownian bridge. Then*

$$u_n \to_{\mathscr{D}(\mathbb{P}_n)} u, \qquad u_n \to_{\mathscr{D}(\tilde{\mathbb{P}}_n(h))} u + H \circ F^{-1},$$

*and*

$$v(P^{u}, P^{u+H\circ F^{-1}}) = \lambda(h).$$

PROOF. The convergence in distribution of $u_n$ is an old and well-known fact [see Gaenssler and Stute (1979) or Shorack and Wellner (1986)]. The last equality follows from Lemma 2.5 and the fact that the transformation $\mathscr{K}(v, F_0) = u$ is one-to-one. □

*Formulation of the problem.* The transformation $\mathscr{K}$ cannot be extended directly to the case of a parametric hypothesis and of a simple hypothesis for random vectors ($m \geq 2$), but one does not have to copy $\mathscr{K}$ in all cases. Instead, one can try to find another transformation which may differ from $\mathscr{K}$ in form but will lead to the same goal.

Let us now formalize this goal for the case of simple hypothesis [cf. Khmaladze (1988)]: To find a transformation $w[v_n, F_0]$ which may depend also on the hypothetical distribution $F_0 = F(\cdot, \theta_0)$, with the following properties:

(a1) $w[v_n, F_0] \to_{\mathscr{D}(\mathbb{P}_n)} w$ and the distribution $P^w$ of $w$ does not depend on $F_0$ for any absolutely continuous $F_0$.

(a2) For any sequence of alternatives $\{A_n\}$ satisfying conditions (2.2)–(2.6), $w[v_n, F_0] \to_{\mathscr{D}(\tilde{\mathbb{P}}_n(h))} w'$ such that $d(P^w, P^{w'}) = \lambda(h)$.

As the test statistics one can choose now functionals $\phi[w[v_n, F]]$ of the process $w[v_n, F]$ in the same way as they choose functionals of $u_n$ in classical goodness of fit theory.

For practical convenience we find it proper to add two additional heuristic requirements:

(b1) The transformation $w[v_n, F]$ must be simple enough to make the calculation of test statistics simple.

(b2) The distribution $P^w$ must be convenient to make the simple calculation of the null distribution of test statistics feasible.

In the case of parametric hypotheses one can formulate a similar problem. Now, we want to find a transformation $w[\hat{v}_n, \mathbb{F}]$ which may depend on hypothetical parametric family $\mathbb{F}$ with the following properties:

(a1) For each $\theta$, $w[\hat{v}_n, \mathbb{F}] \to_{\mathscr{D}(\mathbb{P}_{n\theta})} w$ and $P^w$ does not depend on $\mathbb{F}$ if $\mathbb{F}$ is regular.

(a2) For any sequence of alternatives $\{A_n\}$ satisfying (2.2)–(2.7), $w[\hat{v}_n, \mathbb{F}] \to_{\mathscr{D}(\tilde{\mathbb{P}}_{n\theta}(h))} w'$ such that $d(P^w, P^{w'}) = \lambda(h)$.

Notice that now condition (2.7) is required—this seems natural in view of Lemma 2.5.

Conditions (b1) and (b2) are exactly the same as above and we will not write them down anew.

Our plan in what follows is this: We construct the one-to-one correspondence between the limiting Gaussian process $\hat{v}$ and some Gaussian process $\hat{w}$ with independent increments—the scanning innovation of $\hat{v}$. This is the first and the main step. Then we normalize $\hat{w}$ and get the standard Wiener process. In the resulting transformation of $\hat{v}$ to $w$ we will substitute $\hat{v}_n$ (and even simply $\sqrt{n}\,F_n$) instead of $\hat{v}$ and prove that this is a transformation with desirable properties.

*Function-parametric innovation process* $w_n(f)$. For any two orthogonal projectors $\pi'$ and $\pi''$ we call $\pi''$ larger than $\pi'$, and denote this $\pi' \prec \pi''$, if $\pi'\pi'' = \pi'$. Let $\{\pi_\lambda\}$, $0 \le \lambda \le 1$, be a family of orthogonal projectors, defined on each $L_2(\theta)$. Assume that $\{\pi_\lambda\}$ has the following properties:

1. $\lambda \le \lambda' \Rightarrow \pi_\lambda \prec \pi_{\lambda'}$.
2. $\pi_0 = 0$, $\pi_1 = I$, $I$ denotes the identity operator.
3. For any $f$, $\phi \in L_2(\theta)$, the function $(f, \pi_\lambda \phi)$ is absolutely continuous in $\lambda$.

We recall for the reader's convenience some identities, which we will use later without comment: for orthoprojectors $\pi, \pi', \pi'', \pi' \prec \pi''$, we have

$$(\pi f, \pi\phi) = (f, \pi\phi), \qquad (\pi' f, \pi''\phi) = (\pi' f, \phi).$$

One can imagine the family $\{\pi_\lambda\}$ to be constructed as follows. Let $\{A_\lambda\}$, $0 \le \lambda \le 1$, be a family of measurable subsets of $\mathbb{R}^m$ with the following properties:

1′. $\lambda \le \lambda' \Rightarrow A_\lambda \subset A_{\lambda'}$;
2′. $\mu(A_0) = 0$, $\mu(A_1) = 1$;
3′. $\mu(A_{\lambda'} \setminus A_\lambda) \to 0$ if $\lambda' \downarrow \lambda$;

where $\mu(A)$ denotes Lebesgue measure of a set $A$. Then put

$$\pi_\lambda f(x) = I\{x \in A_\lambda\} f(x).$$

If $\pi_\lambda$ are defined in this way then a projector $\pi_\lambda^\perp = I - \pi_\lambda$ is, obviously, defined as

$$\pi_\lambda^\perp f(x) = I\{x \notin A_\lambda\} f(x).$$

Now consider the following specific condition on the function $q$ and the family $\{\pi_\lambda\}$:

4. For any $\lambda \in [0,1]$, the matrix

$$C_\lambda = \left(\pi_\lambda^\perp q, \pi_\lambda^\perp q^T\right)$$

is nondegenerate, that is, for any $\lambda \in [0,1)$ the inverse matrix $C_\lambda^{-1}$ exists.

Obviously $C_0 = C$. Condition 4 here is convenient rather then necessary, but we will use it for simplicity.

Now we are going to construct the process $w(f)$ which could be viewed as an innovation process for $\bar{v}(f)$. Associate with each $\lambda$ the $\sigma$-algebra

$$\mathscr{F}_\lambda^v = \sigma\{\bar{v}(\pi_\lambda f),\ f \in L_2(\theta)\}.$$

Let us understand this $\sigma$-algebra as the one containing "the past" of $\bar{v}(f)$ up to "the moment" $\lambda$. Let us understand $\bar{v}(f)$ as an increment forward at $\lambda$ if $\pi_\lambda f = 0$, so that for any $f \in L_2(\theta)$ the random variable $\bar{v}(\Delta\pi_\lambda f)$ with $\Delta\pi_\lambda = \pi_{\lambda+\Delta} - \pi_\lambda$ is "a small increment forward" if $\Delta\lambda$ is "small." What we want to do is to construct the innovation of $\{\bar{v}(\pi_\lambda f), \mathscr{F}_\lambda^v\}$. Let us replace this (still uncertain) problem by another: Consider the $\sigma$-algebras

$$\mathscr{F}_\lambda^b = \sigma\{b(\pi_\lambda f),\ f \in L_2(\theta)\}$$

and

$$\mathscr{F}_\lambda = \mathscr{F}_\lambda^b \vee \sigma\{b(q)\} = \mathscr{F}_\lambda^b \vee \sigma\{b(\pi_\lambda^\perp q)\}$$

and consider what could be called an innovation of $\{b(\pi_\lambda f), \mathscr{F}_\lambda\}$, $0 \le \lambda \le 1$. A "small" increment of an innovation process should be defined as

$$\hat{w}(\Delta\pi_\lambda f) = b(\Delta\pi_\lambda f) - E[b(\Delta\pi_\lambda f)|\mathscr{F}_\lambda]. \tag{3.2}$$

Since

$$Eb(\Delta\pi_\lambda f)b(\pi_\lambda f) = (\Delta\pi_\lambda f, \pi_\lambda f) = 0,$$

the Gaussian random variable $b(\Delta\pi_\lambda f)$ is independent of $\mathscr{F}_\lambda^b$. Hence

$$\begin{aligned} E[b(\Delta\pi_\lambda f)|\mathscr{F}_\lambda] &= E[b(\Delta\pi_\lambda f)|b(\pi_\lambda^\perp q)] \\ &= (\Delta\pi_\lambda f, \pi_\lambda^\perp q^T)C_\lambda^{-1}b(\pi_\lambda^\perp q) \\ &= (f, \Delta\pi_\lambda q^T)C_\lambda^{-1}b(\pi_\lambda^\perp q). \end{aligned} \tag{3.3}$$

Expressions (3.2) and (3.3) lead to the following expression

$$\hat{w}(f) = b(f) - \int (f, d\pi_\lambda q^T)C_\lambda^{-1}b(\pi_\lambda^\perp q), \tag{3.4}$$

which still needs precise definition.

LEMMA 3.2. *If* 1–3 *are satisfied, then almost all trajectories of the process* $b(\pi_\lambda^\perp f)$, $\lambda \in [0,1]$, *are continuous in* $\lambda$.

Let $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_N = 1$ be a partition of $[0,1]$ and let

$$c_N(f) = \sum_{i=0}^{N-1} (f, \Delta\pi_{\lambda_i} q^T)C_{\lambda_i}^{-1}b(\pi_{\lambda_i}^\perp q), \qquad \Delta\pi_{\lambda_i} = \pi_{\lambda_{i+1}} - \pi_{\lambda_i}.$$

LEMMA 3.3. *If* 1–4 *are satisfied, then for any* $f$ *such that* $\pi_{1-\varepsilon} f = f$ *for some* $\varepsilon > 0$, *the sequences of random variables* $c_N(f)$ *converges with probability* 1 *as* $N \to \infty$ *and* $\max_i(\lambda_{i+1} - \lambda_i) = \delta_N \to 0$.

Let us denote this limit as

$$c(f) = \int (f, d\pi_\lambda q^T)C_\lambda^{-1}b(\pi_\lambda^\perp q). \tag{3.4$'$}$$

LEMMA 3.4. *If* 1–4 *are satisfied and* $f = \pi_{1-\varepsilon} f$ *for some* $\varepsilon > 0$, *then*

$$E[b(f) - c(f)]^2 = (f, f).$$

Hence, for any $f \in L_2(\theta)$ the random variables $c(\pi_{1-\varepsilon} f)$ converge in mean square as $\varepsilon \to 0$.

Let us denote this limit again as the integral (3.4′). Now the right-hand side of (3.4) exists for all $f \in L_2(\theta)$.

PROOF OF LEMMA 3.2. The $(k+1)$-dimensional Gaussian process $b(\pi_\lambda^\perp q)$ has covariance function $C_{\lambda \vee \mu}$. Consequently, for any $\alpha \in \mathbb{R}^{k+1}$ the process $\alpha^T b(\pi_\lambda^\perp q)$ is a Wiener process w.r.t. the time $t = \alpha^T C_\lambda \alpha$. Therefore, for any $\alpha \in \mathbb{R}^{k+1}$ almost all trajectories of $\alpha^T b(\pi_\lambda^\perp q)$ are continuous in $\lambda$, which proves the lemma. □

REMARK. Another proof of this lemma follows from Theorem 13 of Pollard [(1984), Chapter VII.3]. Indeed the set of functions $\{\alpha^T \pi_\lambda^\perp q\}$, $0 \le \lambda \le 1$, forms a subset in $L_2(\theta)$ with $\varepsilon$-net containing no more than $1 + (\alpha^T C \alpha)^{1/2}/\varepsilon$ points and hence the covering integral for this subset is finite. According to the theorem mentioned above the process $b(\alpha^T \pi_\lambda^\perp q)$ indexed by functions $\alpha^T \pi_\lambda^\perp q$ is continuous w.r.t. the $L_2(\theta)$-norm. However, according to condition 3 the norm $\|\alpha^T \pi_\lambda^\perp q\| = (\alpha^T C_\lambda \alpha)^{1/2}$ is continuous in $\lambda$. Therefore, $b(\alpha^T \pi_\lambda^\perp q)$ and, hence, $b(\pi_\lambda^\perp q)$ is continuous in $\lambda$.

PROOF OF LEMMA 3.3. Let $\{\lambda_i\}_{i=0}^N$ and $\{\mu_j\}_{j=0}^M$ be two partitions of $[0, 1-\varepsilon]$. Assume for simplicity that each $\lambda_i \in \{\mu_j\}_{j=0}^M$. Consider points $\mu_j$ which are contained between $\lambda_i$ and $\lambda_{i+1}$. The corresponding sums in expression of $c_N(f)$ and $c_M(f)$ are, respectively,

$$\sum_{\lambda_i \le \mu_j < \lambda_{i+1}} \left(f, \Delta\pi_{\mu_j} q^T\right) C_{\lambda_i}^{-1} b\left(\pi_{\lambda_i}^\perp q\right)$$

and

$$\sum_{\lambda_i \le \mu_j < \lambda_{i+1}} \left(f, \Delta\pi_{\mu_j} q^T\right) C_{\mu_j}^{-1} b\left(\pi_{\mu_j}^\perp q\right).$$

Consider the difference

$$\Delta_i = \sum_{\lambda_i \le \mu_j < \lambda_{i+1}} \left(f, \Delta\pi_{\mu_j} q^T\right)\left[C_{\mu_j}^{-1} b\left(\pi_{\mu_j}^\perp q\right) - C_{\lambda_i}^{-1} b\left(\pi_{\lambda_i}^\perp q\right)\right].$$

For any vector $\xi = (\xi_1, \dots, \xi_{k+1})^T \in \mathbb{R}^{k+1}$ let $\rho_1(\xi) = |\xi_1| + \cdots + |\xi_{k+1}|$ and $\rho_\infty(\xi) = \max_i |\xi_i|$. Then clearly

$$|\xi^T \eta| \le \rho_1(\xi) \rho_\infty(\eta).$$

Apply this inequality to $\xi = (f, \Delta\pi_\mu' q)$ and $\eta = \eta(\mu, \lambda) = C_\mu^{-1} b(\pi_\mu^\perp q) - C_\lambda^{-1} b(\pi_\lambda^\perp q)$. The matrix $C_\lambda^{-1}$ is continuous on $[0, 1-\varepsilon]$ for any $\varepsilon > 0$. Since $b(\pi_\lambda^\perp q)$ is also continuous in $\lambda$ with probability 1, we can get

$$\rho_\delta = \sup_{\substack{|\lambda - \mu| \le \delta \\ 0 \le \lambda, \mu \le 1-\varepsilon}} \rho_\infty(\eta(\mu, \lambda)) \to 0$$

with probability 1 for any fixed $\varepsilon > 0$ and $\delta \to 0$. Since $\rho_\infty(\eta(\lambda_i, \mu_j)) \le \rho_\delta$ with

$\delta = \delta_N$, we get

$$\Delta_i \leq \sum_{\lambda_i \leq \mu_j < \lambda_{i+1}} \rho_1(f, \Delta\pi_{\mu_j} q) \rho_{\delta_N},$$

and consequently

$$|c_N(f) - c_M(f)| \leq \sum_{j=0}^{M-1} \rho_1(f, \Delta\pi_{\mu_j} q) \rho_{\delta_N}. \tag{3.5}$$

From (3.5) the statement of the lemma will follow if we can prove that

$$\sum_{j=0}^{M-1} \rho_1(f, \Delta\pi_{\mu_j} q) \leq \text{const}. \tag{3.6}$$

Denote by $q_r$ the $r$th coordinate of the vector function $q$. Then

$$|(f, \Delta\pi_{\mu j} q_r)| \leq (\Delta\pi_{\mu j} f, f)^{1/2} (\Delta\pi_{\mu_j} q_r, q_r)^{1/2}$$

and as a consequence

$$\begin{aligned} \sum_{j=0}^{M-1} \rho_1((f, \Delta\pi_{\mu_j} q)) &\leq \sum_{r=1}^{k+1} \sum_{j=0}^{M-1} (\Delta\pi_{\mu_j} f, f)^{1/2} (\Delta\pi_{\mu_j} q_r, q_r)^{1/2} \\ &\leq \sum_{r=1}^{k+1} \left[ \sum_{j=0}^{M-1} (\Delta\pi_{\mu_j} f, f) \right]^{1/2} \left[ \sum_{j=0}^{M-1} (\Delta\pi_{\mu_j} q_r, q_r) \right]^{1/2}. \end{aligned}$$

However,

$$\sum_{j=0}^{M-1} (\Delta\pi_{\mu_j} f, f) = (f, f), \qquad \sum_{j=0}^{M-1} (\Delta\pi_{\mu_j} q_r, q_r) = (q_r, q_r).$$

Therefore (3.6) is correct with

$$\text{const.} = \|f\| \sum_{r=1}^{k+1} \|q_r\|. \qquad \square$$

PROOF OF LEMMA 3.4. Using the formula

$$Eb(f)b(\phi) = (f, \phi),$$

we can obtain by direct calculation: If $f = \pi_{1-\varepsilon} f$, then

$$\begin{aligned} E[b(f) - c(f)]^2 = (f, f) &- 2\int_0^{1-\varepsilon} (f, d\pi_\lambda q^T) C_\lambda^{-1} (\pi_\lambda^\perp q, f) \\ &+ \int_0^{1-\varepsilon} \int_0^{1-\varepsilon} (f, d\pi_\lambda q^T) C_\lambda^{-1} C_{\lambda \vee \mu} C_\mu^{-1} (d\pi_\mu q, f). \end{aligned}$$

Under conditions 3 and 4 both integrals exist and are the usual Stieltjes integrals. The function under the double integral sign is symmetric in $\lambda$ and

$\mu$; therefore the integral is equal to

$$2\int_0^{1-\varepsilon}(f, d\pi_\lambda q)C_\lambda^{-1}\int_\lambda^{1-\varepsilon}(d\pi_\mu q, f) = 2\int_0^{1-\varepsilon}(f, d\pi_\lambda q)C_\lambda^{-1}(\pi_\lambda^\perp q, f).$$

This equality and the previous one lead to

$$E[b(f) - c(f)]^2 = (f, f). \qquad \square$$

Now we can turn back to the process $\{\bar{v}(\pi_\lambda f), \mathscr{F}_\lambda^v\}$. Since the process $(f, q^T)C^{-1}b(q)$ is $\mathscr{F}_\lambda$-measurable for all $\lambda$ we can subtract the identity

$$0 = (\Delta\pi_\lambda f, q^T)C^{-1}b(q) - E\big[(\Delta\pi_\lambda f, q^T)C^{-1}b(q)|\mathscr{F}_\lambda\big]$$

from (3.2) and get

$$\hat{w}(\Delta\pi_\lambda f) = \bar{v}(\Delta\pi_\lambda f) - E[\bar{v}(\Delta\pi_\lambda f)|\mathscr{F}_\lambda]. \tag{3.7}$$

What we finally get from (3.4) and (3.7) is the expression

$$\hat{w}(f) = \bar{v}(f) - \int(f, d\pi_\lambda q^T)C_\lambda^{-1}\bar{v}(\pi_\lambda^\perp q). \tag{3.8}$$

Let us call $\hat{w}$ the *scanning innovation* of $\bar{v}$ and let us call the integral term in (3.8) the *compensator* of $\{\bar{v}(f), \mathscr{F}_\lambda^v\}$. The adjective "scanning" is clarified by Example 1.

REMARK. Since $\bar{v}(q) = 0$, we have $\bar{v}(\pi_\lambda q) = -\bar{v}(\pi_\lambda^\perp q)$. Hence $\bar{v}(\pi_\lambda^\perp q)$ is $\mathscr{F}$-measurable.

Let us call, following Pollard (1984), the function-parametric process $\{b(f), f \in \mathscr{I}\}$ a *Wiener process w.r.t.* $(f, f)^{1/2}$ if for any finite number $r$, the random variables $b(f_1), \dots, b(f_r)$, $f_i \in \mathscr{I}$, have a joint normal distribution with mean 0 and covariance matrix $((f_i, f_j))$, $i, j = 1, \dots, r$, and if almost all trajectories of $\{b(f), f \in \mathscr{I}\}$ are bounded and uniformly continuous on $\mathscr{I}$. Consider also another scalar product:

$$\langle f, \phi\rangle = \int_{x\in[0,1]^m} f(x)\phi(x)\,dx.$$

Let us call the Wiener process w.r.t. $\langle f, f\rangle^{1/2}$ the *standard Wiener process*.

The following very simple lemma shows the transformation of a Wiener process to the standard Wiener process.

LEMMA 3.5. *Suppose* $F([0,1]^m, \theta) = 1$ *for all* $\theta$. *If the density* $f(\cdot, \theta)$ *of the distribution* $F(\cdot, \theta)$ *is positive a.e. on* $[0,1]^m$ *and* $\{\hat{w}(f), f \in \mathscr{I}\}$ *is a Wiener process w.r.t.* $(f, f)^{1/2}$, *then* $\{w(\phi), \phi \in \mathscr{I}'\}$ *with* $w(\phi) = \hat{w}(\phi/f^{1/2}(\cdot, \theta))$ *and* $\mathscr{I}' = \{\phi: \phi/f^{1/2}(\cdot\theta) \in \mathscr{I}\}$ *is a standard Wiener process.*

The condition that $F(\cdot, \theta)$ is nested on $[0,1]^m$ could be satisfied in many ways. If, for example, distributions $F(\cdot, \theta)$ are absolutely continuous w.r.t.

some distribution $\Phi$ for all $\theta \in \Theta$, and $\Phi_1, \dots, \Phi_m$ denote marginal distributions of $\Phi$ we could transform original random vectors $X_i = (X_{i1}, \dots, X_{im})$ into $Y_i = (Y_{i1}, \dots, Y_{im})$ with $Y_{ij} = \Phi_j(X_{ij})$. However, one has to remark that this condition is of no direct need in Lemma 3.5. It will play a role only for convergence (3.27). That is, if we want to have factual convergence under normalization $f^{-1/2}(\cdot, \theta)$ we need our observation to be distributed on a compact set and not to be "smeared" over a whole space. From now on we will *always assume* that conditions 1–4 are satisfied. The notion of covering integral used below can be found in Pollard [(1984), Chapter VII].

THEOREM 3.6. *Let $\mathscr{I}$ be a subset of $L_2(\theta)$ with a finite covering integral. Then the process $\{\hat{w}(f),\ f \in \mathscr{I}\}$, defined by* (3.8), *is a Wiener process w.r.t. $(f,f)^{1/2}$. For any subset $\mathscr{I}$ such that the closed linear span of $\mathscr{I}$ is $L_2(\theta)$, the relation between the processes $\{\hat{w}(f),\ f \in \mathscr{I}\}$ and $\{\bar{v}(f),\ f \in \mathscr{I}\}$ is one-to-one.*

REMARK 1. Since $\{\hat{w}(f),\ f \in \mathscr{I}\}$ and $\{\bar{v}(f),\ f \in \mathscr{I}\}$ can be extended in a one-to-one way to corresponding processes with $\mathscr{I}$ replaced by its closed linear span, the one-to-one correspondence between $\{\hat{w}(f),\ f \in L_2(\theta)\}$ and $\{\bar{v}(f),\ f \in L_2(\theta)\}$ is equivalent to the one-to-one correspondence stated in Theorem 3.6.

REMARK 2. This statement of the theorem can be refined as follows: For any subset $\mathscr{I}$ the relation between the processes $\{\hat{w}(f),\ f \in \mathscr{I},\ \hat{w}(\pi_\lambda^\perp q),\ \lambda \in [0,1]\}$ and $\{\bar{v}(f),\ f \in \mathscr{I},\ \bar{v}(\pi_\lambda q),\ \lambda \in [0,1]\}$ is one-to-one.

REMARK 3. The set $\mathscr{I} = \{I\{\cdot \le x\},\ x \in [0,1]^m\}$ of functions $f(y) = I\{y \le x\}$ satisfies the conditions of Theorem 3.6 for any finite $m$.

PROOF OF THEOREM 3.6. Since $\hat{w}$ is the linear transformation of the Gaussian process $\bar{v}$, it is a Gaussian process as well. The equality

$$E\hat{w}(f)\hat{w}(\phi) = (f, \phi)$$

can be derived from

$$E\hat{w}^2(f + \phi) - E\hat{w}^2(f) - E\hat{w}^2(\phi) = 2E\hat{w}(f)\hat{w}(\phi)$$

and from the equality

$$E\hat{w}^2(f) = (f, f),$$

already proved in Lemma 3.4.

The boundedness and uniform continuity of trajectories of $\hat{w}(f)$ on $\mathscr{I}$ is proved in Theorem 13 of Pollard (1984), Chapter VII [For the reader not quite involved in the theory of function-parametric processes, let us remark that for $w(f)$ the modulus of continuity is derived in exactly the same way as it is for the Wiener process on $[0, 1]$—see, e.g., Itô and McKean (1965).]

What remains is to prove the one-to-one correspondence between $\hat{w}$ and $\bar{v}$. We will prove it through the following lemmas. Reformulate first Lemma 3.4.

LEMMA 3.4′. *The linear operator* $Z$,

$$Zf = f - \int (f, d\pi_\lambda q^T) C_\lambda^{-1} \pi_\lambda^\perp q, \tag{3.9}$$

*is a norm-preserving operator on* $L_2(\theta)$.

Let us now rewrite (3.8) as

$$\hat{w}(f) = \bar{v}(Zf). \tag{3.10}$$

Consider the adjoint [in the scalar product $(f, \phi)$] operator $Z'$ of the operator $Z$:

$$Z'\phi = \phi - \int d\pi_\lambda q^T C_\lambda^{-1} (\pi_\lambda^\perp q, \phi). \tag{3.11}$$

Since $Z$ is norm-preserving, $Z'$ is its unique inverse on the subspace

$$\operatorname{Im} Z = \{\phi: Zf = \phi \text{ for some } f \in L_2(\theta)\}.$$

One can expect now that the inverse of (3.11) is

$$\hat{w}(Z'f) = \bar{v}(f). \tag{3.12}$$

The next lemma proves that this is true on the whole $L_2(\theta)$.

LEMMA 3.7. $\operatorname{Im} Z = \{\phi \in L_2(\theta): (\phi, q) = 0\}$. *Besides*, $Z'q = 0$.

Now, (3.12) is correct on the subspace $\{\phi \in L_2(\theta): (\phi, q) = 0\}$ since $Z'$ is the inverse of $Z$, and (3.12) is correct for $f = \alpha^T q$ as well, since $\bar{v}(q) = 0 = \hat{w}(Z'q)$. Theorem 3.6 is proved. □

PROOF OF LEMMA 3.7. Let us prove first that (3.11) can be determined for all $\phi \in L_2(\theta)$. Clearly the right-hand side of (3.11) exists for all $\phi$ such that $\phi = \pi_{1-\varepsilon}\phi$ for some $\varepsilon > 0$. For all such $\phi$ let us prove that

$$(Z'\phi, Z'\phi) = (\phi - (\phi, q^T) C^{-1} q, \phi) \tag{3.13}$$

and then let $\varepsilon \to 0$. However,

$$\begin{aligned}(Z'\phi, Z'\phi) &= (\phi, \phi) - 2\int (\phi, d\pi_\lambda q^T) C_\lambda^{-1} (\pi_\lambda^\perp q, \phi) \\ &\quad + \int (\phi, \pi_\lambda^\perp q^T) C_\lambda^{-1}\, dC_\lambda\, C_\lambda^{-1} (\pi_\lambda^\perp q, \phi) \\ &= (\phi, \phi) - (\phi, \pi_\lambda^\perp q^T) C_\lambda^{-1} (\pi_\lambda^\perp q, \phi)\Big|_{\lambda=0}^{1}.\end{aligned} \tag{3.14}$$

The last equality is true because of the following ones:

$$dC_\lambda^{-1} = -C_\lambda^{-1}\, dC_\lambda\, C_\lambda^{-1}$$

[consider the identity $C_{\lambda+s}^{-1} - C_\lambda^{-1} = C_{\lambda+s}^{-1}(C_\lambda - C_{\lambda+s})C_\lambda^{-1}$] and

$$d\big[(\phi, \pi_\lambda^\perp q^T)C_\lambda^{-1}(\pi_\lambda^\perp q, \phi)\big] = -2(\phi, d\pi_\lambda q^T)C_\lambda^{-1}(\pi_\lambda^\perp q, \phi)$$
$$-(\phi, \pi_\lambda^\perp q^T)\, dC_\lambda^{-1}(\pi_\lambda^\perp q, \phi)$$

(which is correct because of condition 3). Finally, from Lemma 3.8 it follows that $(\phi, \pi_\lambda^\perp q)C_\lambda^{-1}(\pi_\lambda^\perp q, \phi) = 0$ at $\lambda = 1$. Hence (3.14) gives (3.13).

Now, it is clear that $(q, Zf) = 0$, which implies $\operatorname{Im} Z \subset \{\phi \in L_2(\theta): (\phi, q) = 0\}$. Now let $\phi \neq 0$ belong to the last subspace. Then (3.13) implies $(Z'\phi, Z'\phi) = (\phi, \phi)$ and, hence, there exists $f \neq 0$ such that $Z'\phi = f$ and clearly $Zf = \phi$. This implies $\operatorname{Im} Z \supset \{\phi \in L_2(\theta): (\phi, q) = 0\}$. □

LEMMA 3.8. *The following inequality is true*:

$$(\phi, \pi_\lambda^\perp q^T)C_\lambda^{-1}(\pi_\lambda^\perp q, \phi) \le (\pi_\lambda^\perp \phi, \phi).$$

PROOF. $\xi = \pi_\lambda^\perp q^T C_\lambda^{-1}(\pi_\lambda^\perp q, \phi)$ is the projection of $\pi_\lambda^\perp \phi$ on the subspace spanned on $\pi_\lambda^\perp q$. Consequently

$$(\pi_\lambda^\perp\phi, \pi_\lambda^\perp\phi) = (\pi_\lambda^\perp\phi - \xi, \pi_\lambda^\perp\phi - \xi) + (\xi, \xi)$$
$$\ge (\xi, \xi) = (\pi_\lambda^\perp\phi, \xi) = (\phi, \xi).$$ □

It might seem natural and unavoidable that the possible transformation of $\hat{v}$ to its scanning innovation for arbitrary choice of the function $l$ (which corresponds to arbitrary choice of the projective estimator $\hat{\theta}$) should depend on this $l$. If so, it will be a bit inconvenient and somewhat tiring. Fortunately, the transformation (3.8) is valid for any process $\hat{v}$ and the choice of $\bar{v}$ was simply a convenient way to derive (3.8).

THEOREM 3.9. *Let $\mathscr{I}$ be a subset of $L_2(\theta)$, and let the process $\hat{v}(f)$ be defined by* (2.13). *Then the processes $\hat{w}(f)$ defined by* (3.8) *and by*

$$\hat{w}(f) = \hat{v}(f) - \int (f, d\pi_\lambda^\perp q^T)C_\lambda^{-1}\hat{v}(\pi_\lambda^\perp q)$$
(3.15)
$$= v(f) - \int (f, d\pi_\lambda^\perp q^T)C_\lambda^{-1}v(\pi_\lambda^\perp q)$$

*coincide*.

PROOF. Rewrite (3.15) as $\hat{w}(f) = \hat{v}(Zf) = v(Zf)$. According to definitions (2.15) and (2.18)

$$\hat{v}(f) - \bar{v}(f) = (f, q^T)\big[b(l) - b(\bar{l})\big].$$

However, since $(Zf, q) = 0$ this implies

$$\hat{v}(Zf) - \bar{v}(Zf) = 0.$$

Similarly the difference $v(f) - \hat{v}(f)$ can be written as

$$v(f) - \hat{v}(f) = (f, q^T)\left[\begin{pmatrix} b(1) \\ 0 \end{pmatrix} - b(l)\right],$$

where $b(1)$ stands for $b(f)$ with the function $f$ identically equal to 1 and 0 is the $k$-dimensional vector 0. Hence, again

$$v(Zf) = \hat{v}(Zf).$$

That is, the equality in (3.15) is correct, and the processes (3.8) and (3.15) coincide. □

Now it is quite clear that Theorem 3.9 jointly with Lemma 3.5 gives the transformation of $\hat{v}$ to the standard Wiener process:

$$w(\phi) = \hat{v}\left(Z\left(\frac{\phi}{f^{1/2}(\cdot, \theta)}\right)\right). \tag{3.16}$$

*Examples*; *comparison with some of previous results.* Consider one example which shows the origin of the term *scanning* innovation.

EXAMPLE 1. Let $x, y, (s, t) \in [0, 1]^2$ and $\pi(x) f(y) = I\{y \le x\} f(y)$. Consider a partition of the range of $t$ by points $0 = t_0 < t_1 < \cdots < t_N = 1$ and introduce the $\sigma$-algebras

$$\mathscr{F}_{(1, t_i)} = \sigma\{\bar{v}(\pi(1, t_i) f),\ f \in L_2(\theta)\},$$

$$\mathscr{C}_{(s, t_i)} = \sigma\{\bar{v}(\pi(s, \Delta t_i) f),\ f \in L_2(\theta)\},$$

where

$$\pi(s, \Delta t_i) = \pi(s, t_{i+1}) - \pi(s, t_i),$$

and

$$\mathscr{H}_{(s, t_i)} = \mathscr{F}_{(1, t_i)} \vee \mathscr{C}_{(s, t_i)}.$$

Clearly the family $\{\mathscr{H}_{(s, t_i)},\ s \in [0, 1],\ t_i \in \{t_j\}_1^N\}$— the row-wise scanning family for $\bar{v}$—is linearly ordered w.r.t. inclusion: for any two $(s, t_i)$ and $(s', t_j)$ either $\mathscr{H}_{(s, t_i)} \subseteq \mathscr{H}_{(s', t_j)}$ or $\mathscr{H}_{(s', t_j)} \subseteq \mathscr{H}_{(s, t_i)}$. Because of this the increments

$$w(\pi(\Delta s, \Delta t_i) f) = \bar{v}(\pi(\Delta s, \Delta t_i) f) - E\left[\bar{v}(\pi(\Delta s, \Delta t_i) f) | \mathscr{H}_{(s, t_i)}\right], \tag{3.2'}$$

where $\pi(\Delta s, \Delta t_i) = \pi(s + \Delta, \Delta t_i) - \pi(s, \Delta t_i)$, are independent of previous increments. Hence, if the partition becomes finer, that is, if $\max|t_{i+1} - t_i| \to 0$ as $N \to \infty$, one can hope to glue these increments and get as a limit a Gaussian process with independent increments. The only delicate question is whether one can neglect the $\sigma$-algebras $\mathscr{C}_{(s, t_i)}$ and consider only $\mathscr{F}_{(1, t_i)}$. If yes, then $\{\pi(1, t),\ t \in [0, 1]\}$ is just an example of a family$\{\pi_\lambda\}$ with properties 1–3 and the rectangles $\{[0, (1, t)],\ t \in [0, 1]\}$ are an example of the family $\{A_\lambda\}$ with properties 1′–3′. Theorem 3.6 proves that in the case of $\bar{v}$ the $\sigma$-algebras $\mathscr{C}_{(s, t_i)}$ really can be neglected.

However, the recent research of McKeague, Nikabadze and Sun (1992) shows that $\mathscr{E}_{(s,t_i)}$ cannot always be neglected although a scanning innovation as the limit of (3.2)′ exists.

Consider particular cases of (3.8) and (3.15).

EXAMPLE 2. Suppose $f(y) = I\{y \le x\}$. Let $x, y \in [0,1]$, that is, let $m = 1$. Then

$$Zf(y) = I\{y \le x\} - \int I\{z \le x\} q^T(z) C_z^{-1} F(dz, \theta) q(y) I\{y > z\}$$

$$= I\{y \le x\} - \int_{z \le (x \wedge y)} q^T(z) C_z^{-1} F(dz, \theta) q(y).$$

Consequently,

$$\hat{w}(x) = \bar{v}(x) - \int_{z \le x} q^T(z) C_z^{-1} \int_z^1 q(y) \bar{v}(dy) F(dz, \theta). \tag{3.17}$$

The Wiener process $\{\hat{w}(x),\ x \in [0,1]\}$ defined by (3.17) is just what was considered in Khmaladze (1981), and (3.17) is simply the Doob–Meyer decomposition of $\{\bar{v}(x), \mathscr{F}_x^{v}\}$.

EXAMPLE 3. Again for one-dimensional time parameter consider the case of simple hypothesis $F = F_0$. In this case vector function $q$ [see (2.14)] is equal to 1. Then $C_\lambda = 1 - F_0(\lambda)$ and $v(\pi_\lambda^\perp q) = -v(\pi_\lambda q) = -v(I\{\cdot \le \lambda\})$. Hence

$$\hat{w}(f) = v(f) + \int_0^1 (d\pi_\lambda^\perp f, 1)[1 - F_0(\lambda)]^{-1} v(I\{\cdot \le \lambda\})$$

$$= v(f) + \int_0^1 f(\lambda) F_0(d\lambda)[1 - F_0(\lambda)]^{-1} v(I\{\cdot \le \lambda\}).$$

In the set-parametric version, that is, for $f(y) = I\{y \in A\}$, we get

$$\hat{w}(A) = v(A) + \int_A F_0(d\lambda)[1 - F_0(\lambda)]^{-1} v((0, \lambda])$$

and for the "usual" time parameter, that is, for $f(y) = I\{y \le x\}$, we get

$$\hat{w}(x) = v(x) + \int^x F_0(d\lambda)[1 - F_0(\lambda)]^{-1} v(\lambda),$$

which is the well-known Doob–Meyer decomposition of Brownian bridge $v(x)$.

EXAMPLE 4. Now let $m \ge 2$ and consider still the case of simple hypothesis. Let $\{A_\lambda\}$ be just some family of sets, satisfying conditions 1′–3′. Then

$$(d\pi_\lambda^\perp f, 1) = Ef(X) I\{X \in A_{d\lambda}\}$$

and for $f(y) = I\{y \in A\}$

$$(d\pi_\lambda^\perp f, 1) = F_0(A \cap A_{d\lambda}).$$

Hence

$$\hat{w}(f) = v(f) + \int_0^1 Ef(X) I\{X \in A_{d\lambda}\} [1 - F_0(A_\lambda)]^{-1} v(A_\lambda)$$

and in set-parametric version

$$\hat{w}(A) = v(A) + \int_0^1 F_0(A \cap A_{d\lambda}) [1 - F_0(A_\lambda)]^{-1} v(A_\lambda). \tag{3.18}$$

This last equality gives the transformation of the Gaussian measure $v(A)$ into its scanning innovation measure $\hat{w}(A)$, and this without much respect to dimension $m$ of space, where the $A$'s lie.

EXAMPLE 5. Many choices of $\{A_\lambda\}$ are possible, as is especially clear in the case $m \geq 2$. Consider one particular choice. Let $x, y \in [0,1]^m$, $\mathbb{1} = (1, \dots, 1) \in \mathbb{R}^{m-1}$, and put $A_\lambda = [0, (\lambda, \mathbb{1})]$. For this choice of $A_\lambda$ and for $f(y) = I\{y \leq x\}$,

$$\left(d\pi_\lambda^\perp f, 1\right) = Ef(X) I\{X_{(1)} \in d\lambda\} = F_0\left(x | X_{(1)} = \lambda\right) f_{01}(\lambda)\, d\lambda,$$

where $X_{(1)}$ denotes the first coordinate of random vector $X$ and $f_{01}$ is its density. Hence

$$\hat{w}(x) = v(x) + \int_0^1 F_0\left(x | X_{(1)} = \lambda\right) f_{01}(\lambda) [1 - F_{01}(\lambda)]^{-1} v(\lambda, \mathbb{1})\, d\lambda. \tag{3.19}$$

If we assume that the marginal distribution of $X_{(1)}$ is uniform (on $[0,1]$), we get an even simpler expression,

$$\hat{w}(x) = v(x) + \int_0^1 F_0\left(x | X_{(1)} = \lambda\right) (1 - \lambda)^{-1} v(\lambda, \mathbb{1})\, d\lambda.$$

In the case of a parametric hypothesis we get

$$\hat{w}(x) = \bar{v}(x) - \int_{z \leq x} q^T(z) C_{(\sigma, \mathbb{1})} F(dz, \theta) \int_{(\sigma, \mathbb{1}) \leq y \leq (1,1)} q(y) \bar{v}(dy), \tag{3.20}$$

where $\sigma$ is the first coordinate of $z$, and

$$C_{(\sigma, \mathbb{1})} = \int_{y \in [0,1]^m \setminus [0, ((\sigma, \mathbb{1})} q(y) q^T(y) F(dy, \theta).$$

REMARK. The processes (3.19) and (3.20), suggested in earlier papers, Khmaladze (1988) and Nikabadze and Khmaladze (1987), respectively, left the impression of an essentially nonsymmetric solution—the choice of the rectangles $[0, (1, t)]$ paid, by some arbitrary reason, too much attention to one of the coordinates. Unsatisfied with this we looked for a more symmetric construction. Now, first, we are practically free in our choice of $\{A_\lambda\}$ and, second, it is now obvious that for $m = 1$ (on the real line) we have the same variety of choices.

*Distance in variation; condition a2.* Our further program is clear; in the next subsection we will consider the empirical analogue of $\{w(\phi),\ \phi \in \mathscr{I}'\}$, the process $\{w_n(\phi),\ \phi \in \mathscr{I}'\}$ with $w_n(\phi) = \hat{v}_n(Z(\phi/f^{1/2}(\cdot,\theta)))$ and will prove that this process gives a solution of the problem, stated at the beginning of this section. In the present subsection we will prove that the provisional limiting processes of $\{w_n(\phi),\ \phi \in \mathscr{I}'\}$ under the hypothesis and alternative satisfy condition a2.

Recall that $H = H(f)$, $f \in \mathscr{I}$, is the function on $\mathscr{I}$ defined by $H(f) = (h, f)$ and denote by $J = J(\phi)$, $\phi \in \mathscr{I}'$, the transformation of the function $H$ similar to (3.16):

$$J(\phi) = H\left(Z\frac{\phi}{f^{1/2}(\cdot,\theta)}\right).$$

LEMMA 3.10. *Let $\mathscr{I}$ be a subset of $L_2(\theta)$ with the following properties: $\Lambda(\mathscr{I}) = L_2(\theta)$, where $\Lambda(\mathscr{F})$ is the closed linear span of $\mathscr{F}$, and the set $\mathscr{I}' = \{\phi\colon \phi/f^{1/2}(\cdot,\theta) \in \mathscr{I}\}$ has finite covering integral in the norm $\langle\phi,\phi\rangle^{1/2}$. If the function $h$ satisfies conditions* (2.5)–(2.7), *then*

$$d(P^{w}, P^{w+J}) = d(P^{\bar{v}}, P^{\bar{v}+H}) = \lambda(h).$$

The process $w$ has a standard distribution not depending on $\mathbb{F}$, hence it is a good candidate for the limiting process of condition a1. Lemma 3.10 says that the process $w + J$ is a good candidate for the limiting process $w'$ of the condition a2.

PROOF OF LEMMA 3.10. The second equality in the assertion of the lemma is already proved in Lemma 2.5. The first equality follows from the one-to-one correspondence between $w$ and $\bar{v}$ and, similarly, between $J$ and $H$. It could be also easily seen directly:

$$\begin{aligned} d(P^{w(\phi)}, P^{w(\phi)+J(\phi)}) &= 2\Phi\left(\frac{(h, Z\phi/[f^{1/2}(\cdot,\theta)])}{\langle\phi,\phi\rangle^{1/2}}\right) - 1 \\ &= 2\Phi\left(\frac{\langle f^{1/2}(\cdot,\theta)Z'h, \phi\rangle}{2\langle\phi,\phi\rangle^{1/2}}\right) - 1 \end{aligned}$$

and the maximum of the argument of $\Phi$ is reached at $\phi = f^{1/2}(\cdot,\theta)Z'h$ and is equal to

$$\tfrac{1}{2}\langle f^{1/2}(\cdot,\theta)Z'h, f^{1/2}(\cdot,\theta)Z'h\rangle^{1/2} = \tfrac{1}{2}(Z'h, Z'h)^{1/2} = \tfrac{1}{2}(h,h)^{1/2},$$

where the equality follows from Lemma 3.4′. Hence

$$\max d(P^{w(\phi)}, P^{w(\phi)+J(\phi)}) = 2\Phi\left(\frac{(h,h)^{1/2}}{2}\right) - 1 = \lambda(h). \qquad \square$$

*Convergence in distribution.* Let us turn now to the empirical analogues of $\hat{w}$ and $w$ and consider the problem of convergence in distribution. For any $f \in L_2 = \bigcap_{\theta \in \Theta} L_2(\theta)$ introduce the random variable

$$\hat{w}_n(f) = \hat{v}_n(f) - \int (f, d\pi_\lambda q^T) C_\lambda^{-1} \hat{v}_n(\pi_\lambda^\perp q), \tag{3.21}$$

or, in short,

$$\hat{w}_n(f) = \hat{v}_n(Zf).$$

It is not difficult to prove that $\hat{w}_n(f)$ exists for all $n = 1, 2, \dots$ and all $f \in L_2$ [cf. formula (3.26)] and we will not dwell on this problem. It is technically convenient to get a little simpler approximation of $\hat{w}_n$.

LEMMA 3.11. *Let $\mathbb{F}$ be a regular parametric family. Let $S$ be any bounded subset of $L_2$, that is, for some $c$ and for all $f \in S$, $\|f\|_\theta < c$ for all $\theta$. Then both under $\mathbb{P}_{n\theta}$ and under $\tilde{\mathbb{P}}_{n\theta}$*

$$\sup_{f \in S} |\hat{w}_n(f) - v_n(Zf)| \to_P 0, \qquad n \to \infty.$$

Hence without loss of generality we can replace $\hat{w}_n(f)$ by $v_n(Zf)$ if we are about to study the convergence in distribution of $\hat{w}_n$.

PROOF OF LEMMA 3.11. Using the property $Zq = 0$, one can write

$$\hat{v}_n(Zf) - v_n(Zf) = (Zf, \xi_n),$$

where $r_n = v_n(\cdot, \hat{\theta}) - \Pi_1 v_n(\cdot, \theta)$, and $\xi_n(x) = \partial r_n(x)/\partial F(x, \theta)$. Now Lemma 3.4′ leads to

$$|(Zf, \xi_n)| \le \|Zf\|_\theta \|\xi_n\|_\theta = \|f\|_\theta \|\xi_n\|_\theta$$

and, hence,

$$\sup_{f \in S} |(Zf, \xi_n)| \le c\|\xi_n\|_\theta \to_P 0$$

under $\mathbb{P}_{n\theta}$. However, $\|\xi_n\|_\theta \to_P 0$ under $\tilde{\mathbb{P}}_{n\theta}$ as well because of contiguity of $\{\tilde{\mathbb{P}}_{n\theta}\}$ to $\{\mathbb{P}_{n\theta}\}$. □

Denote

$$c_n(f) = \int (f, d\pi_\lambda q^T) C_\lambda^{-1} v_n(\pi_\lambda^\perp q).$$

LEMMA 3.12. *Let $S_\varepsilon$ be a bounded subset of $L_2$ of functions $f$ such that $\pi_{1-\varepsilon} f = f$. Then for each $\eta > 0$ and $\Delta > 0$, there exists $\delta > 0$ for which*

$$\limsup_{n \to \infty} \mathbb{P}_{n\theta}\left\{ \sup_{f, g \in S_\varepsilon : \|f - g\|_\theta < \delta} |c_n(f - g)| > \Delta \right\} < \eta.$$

The same statement is true with $\mathbb{P}_{n\theta}$ replaced by $\tilde{\mathbb{P}}_{n\theta}$.

PROOF OF LEMMA 3.12. Apply to the scalar product

$$c_n(f) = \left(f, \int d\pi_\lambda\, q^T C_\lambda^{-1} v_n(\pi_\lambda^\perp q)\right)$$

Schwarz's inequality:

$$|c_n(f)|^2 \le (f, f)(z_n, z_n),$$

where

$$(z_n, z_n) = \int_0^{1-\varepsilon} v_n(\pi_\lambda^\perp q^T) C_\lambda^{-1}(d\pi_\lambda\, q, d\pi_\lambda\, q^T) C_\lambda^{-1} v_n(\pi_\lambda^\perp q).$$

However, $(z_n, z_n)$ is bounded in probability because it has finite expectation. Hence, it is bounded in $\tilde{\mathbb{P}}_{n\theta}$-probability as well. Now Lemma 3.12 follows from the inclusion

$$\left\{ \sup_{f, g \in S_\varepsilon, \|f-g\|<\delta} |c_n(f-g)| > \Delta \right\} \subseteq \left\{ (z_n, z_n) > \frac{\Delta}{\delta} \right\}. \qquad \square$$

The reader can guess now that the convergence in distribution of $c_n(f)$ on a set of functions $f = \pi_{1-\varepsilon} f$ is an easy matter: Convergence in distribution at each $f$ is easy to prove and tightness is granted by Lemma 3.12.

Denote by $\mathscr{X}(\mathscr{I})$ the space of bounded functions $x(f)$, $f \in \mathscr{I}$, with the norm $\sup_{f \in \mathscr{I}} |x(f)|$ [cf. Pollard (1984), Section VII.5].

THEOREM 3.13. *Suppose $\mathscr{I}$ is a subset of $L_2$ such that $|f| < c$ for the same constant $c$ for all $f$ and that in $\mathscr{X}(\mathscr{I})$*

$$v_n \to_{\mathscr{D}(\mathbb{P}_{n\theta})} v, \qquad v_n \to_{\mathscr{D}(\tilde{\mathbb{P}}_{n\theta})} v + H \tag{3.22}$$

*with $H(f) = (f, h)$. Suppose also that the family $\mathbb{F}$ is regular and that the function*

$$\alpha_\lambda = \left[(1, d\pi_\lambda q^T) C_\lambda^{-1}(d\pi_\lambda q, 1)\right]^{1/2} \Big/ d\lambda$$

(*where* 1 *stands for the function which is identically equal to the number* 1) *is integrable. Then in $\mathscr{X}(\mathscr{I})$*

$$\hat{w}_n \to_{\mathscr{D}(\mathbb{P}_{n\theta})} \hat{w}, \qquad \hat{w}_n \to_{\mathscr{D}(\tilde{\mathbb{P}}_{n\theta})} \hat{w} + H(Z).$$

REMARK. The condition of integrability of $\alpha_\lambda$ is mild but nevertheless an additional restriction on $q$—it is not satisfied for all $q \in L_2(\theta)$. If, in particular, $q$ is a one-dimensional function of the scalar variable $x \in [0, 1]$, then

$$\alpha_\lambda = \frac{|q(\lambda)|}{\left(\int_\lambda^1 q^2(x)\, dx\right)^{1/2}}$$

is not integrable for $q_\lambda \to 0$ "too quickly" as $\lambda \to 1$. If, for example, $q(x) = \exp[-1/(1-x)]$, then $\alpha_\lambda \sim 1/(1-\lambda)$. However, if $q(x) = \exp[-1/(1-x)^\beta]$, with $\beta < 1$, then $\alpha_\lambda \sim 1/(1-\lambda)^\beta$ is integrable. Obviously $\alpha_\lambda$ is integrable for

$q(x) \sim (1-x)^{\beta}$, $\beta < \infty$, and for any $q(x)$ bounded away from 0 at a neighborhood of $x = 1$: If $|q(x)| > \delta$ for $x > 1 - \varepsilon$, then

$$\alpha_\lambda \leq \frac{q^2(\lambda)}{\delta\left(\int_\lambda^1 q^2(x)\,dx\right)^{1/2}}, \qquad \lambda > 1 - \varepsilon,$$

and the right-hand side is integrable. The condition of integrability of $\alpha$, which we did not need in previous papers [Khmaladze (1981, 1986)], is the price we pay for the extension to "very large" $\mathscr{I}$—as will be clear from the proof of Theorem 3.13 [see (3.20)]: If $\alpha_\lambda$ is integrable in a neighborhood of 1, then $c_n(\cdot)$ converges in distribution in $\mathscr{X}(\mathscr{I})$ for $\mathscr{I}$ being the set of all pointwise bounded functions, for example, the indicator functions of all measurable subsets of $[0, 1]^m$.

Proof of Theorem 3.13. Replace $\hat{w}_n(f)$ by $v_n(Zf)$. One can do this because $\mathbb{F}$ is regular (Lemma 3.11). It is clear that for any square-integrable function $f$ the sequence $\{v_n(Zf)\}$ converges in distribution under $\mathbb{P}_{n\theta}$ to $v(Zf)$ [under $\mathbb{P}_{n\theta}$ to $v(Zf) + H(Zf)$] simply as a consequence of the CLT. Let us verify tightness. Since $v_n = \{v_n(f), f \in \mathscr{I}\}$ converges in distribution the sequence $\{v_n\}$ is tight, and Lemma 3.12 asserts that the sequence $\{c_n(\pi_{1-\varepsilon})\}$ is also tight. Since addition is a continuous operation in $\mathscr{X}(\mathscr{I})$ the sequence $\{v_n - c_n(\pi_{1-\varepsilon})\}$ is also tight. What remains is to consider the difference

$$v_n(Zf) - v_n(f) + c_n(\pi_{1-\varepsilon} f) = c_n\left(\pi_{1-\varepsilon}^{\perp} f\right).$$

Let us show that for any $\Delta > 0$ and $\eta > 0$ there exists $\varepsilon > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}_{n\theta}\left\{\sup_{f\in\mathscr{I}} \left|c_n\left(\pi_{1-\varepsilon}^{\perp} f\right)\right| > \Delta\right\} < \eta, \tag{3.23}$$

but

$$\sup_{f\in\mathscr{I}} \left|c_n\left(\pi_{1-\varepsilon}^{\perp} f\right)\right| \leq c\int_{1-\varepsilon}^{1} \left|\left(1, d\pi_\lambda q^T\right) C_\lambda^{-1} v_n\left(\pi_\lambda^{\perp} q\right)\right| \tag{3.24}$$

and

$$\begin{aligned} E\int_{1-\varepsilon}^{1} \left|\left(1, d\pi_\lambda q^T\right) C_\lambda^{-1} v_n\left(\pi_\lambda^{\perp} q\right)\right| &\leq \int_{1-\varepsilon}^{1} \left[E\left|\left(1, d\pi_\lambda q^T\right) C_\lambda^{-1} v_n\left(\pi_\lambda^{\perp} q\right)\right|^2\right]^{1/2} \\ &= \int_{1-\varepsilon}^{1} \left[\left(1, d\pi_\lambda q^T\right) C_\lambda^{-1} \left(d\pi_\lambda q, 1\right)\right]^{1/2} \qquad (3.25) \\ &= \int_{1-\varepsilon}^{1} \alpha_\lambda \, d\lambda. \end{aligned}$$

Since $\alpha_\lambda$ is integrable, the last integral can be made arbitrarily small if $\varepsilon$ is small. Hence, the random variable in the right-hand side of (3.24) is small in probability for $\varepsilon$ small, and (3.23) is proved. □

REMARK. The approximation $v_n(Zf)$ of the empirical scanning innovation $\hat{w}_n(f)$ can be replaced by a simpler expression since

$$v_n(Zf) = \sqrt{n}\,F_n(Zf),$$

namely, the $\hat{w}_n(f)$ can be approximated by

$$\sqrt{n}\,F_n(Zf) = \sqrt{n}\left[F_n(f) - \int (f, d\pi_\lambda q^T) C_\lambda^{-1} F_n(\pi_\lambda^\perp q)\right]. \tag{3.26}$$

Clearly $F_n(f)$ denotes the sum

$$F_n(f) = \frac{1}{n}\sum_1^n f(X_i) = \int f(x) F_n(dx).$$

By the way, for the function-parametric point process $F_n(f)$ expression (3.26) gives in a good sense its Doob–Meyer decomposition w.r.t. the filtration $\{\mathscr{F}_\lambda{}^n\}$,

$$\mathscr{F}_\lambda{}^n = \sigma\{F_n(\pi_\lambda f),\ f \in L_2,\ F_n(q)\}.$$

The increments of $F_n(Zf)$ are not independent, of course, but they are uncorrelated [cf. the definition of innovation processes in Rozanov (1974)].

Theorem 3.13 is adjusted to the possibility of choosing as $f$ the indicator functions $f(x) = I\{x \le z\}$ and, hence, to prove convergence in distribution for $\hat{w}_n$ regarded as a process with the "usual" time parameter $z \in [0,1]^m$. A more schematic formulation of the theorem is as follows: *Let* $\mathscr{I}$ *be such that conditions* (3.22) *and* (3.23) *are fulfilled. Then the assertion of Theorem* 3.13 *is correct.*

One can adopt this formulation and state the following theorem concerning $w_n$: Let $\mathscr{I}' = \{\phi\colon \phi/f_\theta^{1/2}(\cdot) \in \mathscr{I}\}$ and let for $\mathscr{I}$ conditions (3.22) and (3.23) be fulfilled. Then in $\mathscr{X}(\mathscr{I}')$

$$w_n \to_{\mathscr{D}(\mathbb{P}_{n\theta})} w, \qquad w_n \to_{\mathscr{D}(\tilde{\mathbb{P}}_{n\theta})} w + J. \tag{3.27}$$

This formulation can help in a search for sets $\mathscr{I}$ different from the one described in Theorem 3.13, but we prefer to formulate our last theorem in the same fashion as Theorem 3.13.

THEOREM 3.14. *Let* $F(\cdot,\theta)$ *be nested on* $[0,1]^m$. *Suppose* $\mathscr{I}'$ *is a subset of* $L_2[0,1]^m$ *such that the following hold*:

(i) $|\phi| < c$ *for the same constant* $c$ *for all* $\phi \in \mathscr{I}'$;
(ii) $\Lambda(\mathscr{I}') = L_2[0,1]^m$;
(iii) *in the space* $\mathscr{X}(\mathscr{I})$, *where* $\mathscr{I} = \{f\colon f = \phi/f_\theta^{1/2}(\cdot), \phi \in \mathscr{I}'\}$ *convergence* (3.22) *holds.*

*Suppose also that* $\mathbb{F}$ *is a regular family and the function*

$$\alpha'_\lambda = \left[\langle f_\theta^{1/2}, d\pi_\lambda q^T\rangle C_\lambda^{-1}\langle d\pi_\lambda q, f_\theta^{1/2}\rangle\right]^{1/2}/d\lambda$$

*is integrable. Then in* $\mathscr{X}(\mathscr{I}')$ (3.27) *holds, where* $w$ *is the standard Wiener* *process and the shift function $J$ is defined just before Lemma* 3.10. *The assertion of Lemma* 3.10 *is correct.*

REMARK. Since the matrix $C_\lambda$ can be defined as $C_\lambda = \langle f_\theta^{1/2}\pi_\lambda^\perp q, f_\theta^{1/2}\pi_\lambda^\perp q^T\rangle$ for one-dimensional $q$ the function $\alpha'$ has the form

$$\alpha'_\lambda = \frac{f_\theta^{1/2}(\lambda)|q(\lambda)|}{\left(\int_\lambda^1 f_\theta(x)q^2(x)\,dx\right)^{1/2}}.$$

Hence the previous remark can be applied to the function $f_\theta^{1/2}(\cdot)q(\cdot)$.

According to Theorem 3.14 the process $w_n$ is the desirable transformation $w[\hat{v}_n, \mathbb{F}]$ with properties a1 and a2. In our view this transformation possesses properties b1 and b2 as well.

PROOF OF THEOREM 3.14. This proof is in fact contained in the proofs of Theorem 3.13 and Lemma 3.10, but let us repeat it for the reader's convenience. Since $\mathbb{F}$ is regular, one can replace $w_n(\phi)$ by $v_n(Zf)$ with $f = \phi/f_\theta^{1/2}$ (cf. Lemma 3.11). Since $v_n(Zf)$ is a normalized sum of i.i.d. random variables, $Ev_n^2(Zf) = \langle\phi, \phi\rangle$, its convergence in distribution under $\mathbb{P}_{n\theta}$ to $w(\phi) = v(Zf)$ and under $\tilde{\mathbb{P}}_{n\theta}$ to $w(\phi) + J(\phi)$ for each given $\phi$ is a consequence of CLT. According to condition 3 and Lemma 3.12, the sequences $\{v_n(\cdot/f_\theta^{1/2})\}$ and $\{c_n(\pi_{1-\varepsilon}(\cdot/f_\theta^{1/2})\}$ are tight in $\mathscr{X}(\mathscr{I}')$. Consider the difference

$$v_n\left(Z\frac{\phi}{f_\theta^{1/2}}\right) - v_n\left(\frac{\phi}{f_\theta^{1/2}}\right) + c_n\left(\pi_{1-\varepsilon}\frac{\phi}{f_\theta^{1/2}}\right) = c_n\left(\pi_{1-\varepsilon}^\perp\frac{\phi}{f_\theta^{1/2}}\right).$$

However,

$$\sup_{\phi\in\mathscr{I}'}\left|c_n\left(\pi_{1-\varepsilon}^\perp\frac{\phi}{f_\theta^{1/2}}\right)\right| \le c\int_{1-\varepsilon}^1 \left|\langle f_\theta^{1/2}, d\pi_\lambda q^T\rangle C_\lambda^{-1}v_n(\pi_\lambda^\perp q)\right| \tag{3.28}$$

and

$$E\int_{1-\varepsilon}^1 \left|\langle f_\theta^{1/2}, d\pi_\lambda q^T\rangle C_\lambda^{-1}v_n(\pi_\lambda^\perp q)\right| \le \int_{1-\varepsilon}^t \left(E\left|\langle f_\theta^{1/2}, d\pi_\lambda q^T\rangle C_\lambda^{-1}v_n(\pi_\lambda^\perp q)\right|^2\right)^{1/2}$$
$$= \int_{1-\varepsilon}^1 \alpha'_\lambda\,d\lambda.$$

Hence, the upper bound in the left-hand side of (3.28) can be made arbitrarily small in probability for sufficiently small $\varepsilon > 0$. This means that the sequence $\{v_n(Z(\cdot/f_\theta^{1/2})\}$ is tight in $\mathscr{X}(\mathscr{I}')$. Convergence (3.27) follows. Since $\langle\phi, \phi\rangle = (f, f)$ we get $\Lambda(\mathscr{I}') = L_2[0,1]^m$ iff $\Lambda(\mathscr{I}) = L_2(\theta)$. Hence condition 2 allows us to apply Lemma 3.10 and to conclude the proof. $\square$

**Acknowledgments.** The author would like to thank Professors R. D. Gill and K. O. Dzhaparidze as well as Dr. S. van de Geer for their interest and helpful comments.

The author also is thankful to the referees who read the text very carefully and made several good suggestions.

# REFERENCES

BICKEL, P. and WICHURA, M. (1971). Convergence criteria for multiparameter stochastic processes. *Ann. Math. Statist.* **42** 1656–1670.

BILLINGSLEY, D. (1968). *Convergence of Probability Measures*. Wiley, New York.

CAIROLI, R. and WALSH, J. B. (1975). Stochastic integrals in the plane. *Acta Math.* **134** 111–183.

CRAMÉR, H. (1964). Stochastic process as curves in Hilbert space. *Teor. Veroyatnost. i Primenen.* **9** 193–204.

DOOB, J. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **20** 393–403.

DURBIN, J., KNOTT, M. and TAYLOR, C. C. (1975). Components of Cramér–von Mises statistics, II. *J. Roy. Statist. Soc. Ser. B* **37** 216–237.

GAENSSLER, P. and STUTE, W. (1979). Empirical processes: A survey of results for independent and identically distributed random variables. *Ann. Probab.* **7** 193–243.

GIKHMAN, J. I. (1953). Some remarks on A. Kolmogorov's goodness of fit test. *Dokl. Acad. Nauk* **91** 715–718. (In Russian.)

GIKHMAN, J. I. (1954). On the theory of $\omega^2$-tests. *Mat. Zb. Kiev Gos. Univ.* **5** 51–59. (In Ukrainian.)

GIKHMAN, J. I. (1982). Biparametric martingales. *Uspekhi Mat. Nauk* **37** 215–237. (English translation in *Russian Math. Surveys.*)

GOHBERG, I. and KREIN, M. (1967). *Theory of Volterra Operators in Hilbert Space*. Nauka, Moscow.

GREENWOOD, P. and SHIRYAYEV, A. N. (1985). *Contiguity and the Statistical Invariance Principle*. Gordon and Breach, London.

HITSUDA, M. (1968). Representation of Gaussian process equivalent to Wiener process. *Osaka J. Math.* **5** 299–312.

IBRAGIMOV, I. A. and HAS'MINSKII, R. (1981). *Asymptotic Theory of Estimation*. Springer, Berlin.

ITÔ, K. and MCKEAN, H. P., JR. (1965). *Diffusion Processes and Their Sample Paths*. Springer, Berlin.

KAC, M., KIEFER, J. and WOLFOWITZ, J. (1955). On tests of normality and other tests of goodness of fit based on distance methods. *Ann. Math. Statist.* **26** 189–211.

KHMALADZE, E. V. (1975). On estimation of necessary sample size for testing simple contiguous alternatives. *Teor. Veroyatnost. i Primenen.* **20** 115–125. (In Russian.)

KHMALADZE, E. V. (1979). The use of $w^2$ tests for testing parametric hypotheses. *Teor. Veroyatnost. i Primenen.* (English translation in *Theory Probab. Appl.* **24** 283–301.)

KHMALADZE, E. V. (1981). Martingale approach to the theory of goodness of fit tests. *Teor. Veroyatnost. i Primenen.* (English translation in *Theory Probab. Appl.* **26** 240–257.)

KHMALADZE, E. V. (1987). Innovation fields and the problem of testing simple statistical hypotheses in $R^m$. *Dokl. Acad. Nauk* **290** 2. (English translation in *Soviet Math. Dokl.* **34** 293–295)

KHMALADZE, E. V. (1988). An innovation approach in goodness-of-fit tests in $\mathbb{R}^m$. *Ann. Statist.* **16** 1503–1516.

KHMALADZE, E. M. (1989). Theory of goodness of fit tests and scanning innovation martingales. Report BS-R890-4, Center for Mathematics and Computer Science, Amsterdam.

KOLMOGOROV, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Instit. Ital. Attuari* **4** 83–91.

KOLMOGOROV, A. N. (1986). *Selected Papers* **2**: *Probability Theory and Mathematical Statistics*. Nauka, Moscow. (In Russian.)

KUO, H.-H. (1975). *Gaussian Measures in Banach Spaces. Lecture Notes in Math.* **463**. Springer, Berlin.

LÉVY, P. (1948). *Processus Stochastiques et Mouvement Brownien*. Gauthiers-Villars, Paris.

LIPZER, R. S. and SHIRYAYEV, A. N. (1977). *Statistics of Random Processes*. Springer, Berlin.

MCKEAGUE, I. W., NIKABADZE A. M. and SUN, Y. (1992). Transformation of Gaussian random fields and a test for independence of a survival time from covariate. Technical Report, Dept. Statistics, Florida State Univ.

NEUHAUSS, G. (1971). On weak convergence of stochastic processes with multi-dimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295.

NIKABADZE, A. M. and KHMALADZE, E. V. (1987). On goodness-of-fit tests for parametric hypotheses in $\mathbb{R}_m$. *Dokl. Acad. Nauk* **294**. (English translation in *Soviet Math. Dokl.* **35** 627–629.)

OOSTERHOFF, J. and VAN ZWET, W. (1979). A note on contiguity and Hellinger distance. In *Contributions to Statistics: Jaroslav Hájek Memorial Volume* (J. Jurečhková, ed.). Reidel, Dordrecht.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

ROZANOV, YU. (1974). *Theory of Innovation Processes*. Nauka, Moscow.

ROSENBLATT, M. (1952). Remark on multivariate transformation. *Ann. Math. Statist.* **23** 470–472.

SEN, P. K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference*. Wiley, New York.

SHEPP, L. A. (1966). Radon–Nikodym derivatives of Gaussian measures. *Ann. Math. Statist.* **37** 321–354.

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

SIMPSON, P. B. (1951). Note on the estimation of a bivariate distribution function. *Ann. Math. Statist.* **22** 476–478.

SMIRNOV, N. V. (1937). On the distribution of $\omega^2$-tests of Mises. *Mat. Sb.* **2** 973–993. (In Russian.)

TYURIN, YU. N. (1970). On testing parametric hypotheses by nonparametric tests. *Teor. Veroyatnost. i Primenen.* **15** 745–749. (In Russian.)

WALD, A. and WOLFOWITZ, J. (1939). Confidence limits for continuous distribution functions. *Ann. Math. Statist.* **10** 105–118.

WONG, E. and ZAKAI, M. (1974). Martingales and stochastic integrals for processes with multidimensional parameters. *Z. Wahrsch. Verw. Gebiete* **29** 109–122.

A. M. RAZMADZE MATHEMATICAL INSTITUTE
GEORGIAN ACADEMY OF SCIENCE
Z. RUKHADZE STREET 1
380093 TBILISI
GEORGIA

V. A. STEKLOV MATHEMATICAL INSTITUTE
RUSSIAN ACADEMY OF SCIENCE
VAVILOV ST. 42
117966 GSP-1 MOSCOW
RUSSIA