Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

E.V. Khmaladze

Theory of goodness of fit tests and scanning innovation martingales

Department of Operations Research, Statistics, and System Theory Report BS-R8904 January

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (N.W.O.).

Theory of Goodness of Fit Tests and Scanning Innovation Martingales

E.V. Khmaladze
Steklov Mathematical Institute, Moscow, and
Razmadze Mathematical Institute, Tbilisi

Mathematical Institute V.A. Steklova AN SSSR
ul. Vavilova 42
117966 GSP-1 Moscow

This paper is mainly devoted to the following statistical problem: in 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 on-to-one correspondence with initial Gaussian processes. This is done for the function-parametric versions of the processes.

Keywords and phrases: empirical processes, parametric empirical process, regular parametric hypothesis, asymptotically distribution free processes, contiguous alternatives, innovation process, increasing family of projectors, goodness of fit tests, multiparameter martingales.
AMS 1980 subject classifications: 62G10, 62F03.
Acknowledgement: The author started the work contained in this report in the course of preparation for a visit to the Department of Mathematical Statistics of CWI. During the visit, Summer 1987, the work was continued and finished in its main features. The research was very essentially stimulated by the interest shown by the members of the department and, in particular, by discussions with prof. R.D. Gill and prof. K. Dzaparidze. (Stimulating interest was shown also by prof. C. Klaassen from University of Leiden.)
It is a true honour for the author to publish this research now as a report of CWI.

1. Introduction

Consider i.i.d. random vectors $X_1, \cdots, X_n$ 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 Θ contains only one point θ_0 let us write F_0 instead of $\mathbb{F}$. Denote by F the unknown distribution of each X_i. The fundamental role of the uniform empirical process u_n in the theory of goodness of fit tests for testing the simple hypothesis $F=F_0$ for scalar random variables ($m=1$) is well known. The first and main aim of the present 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 ($1\leqslant m<\infty$).

This empirical process is derived on the basis of some 'innovation' reasoning for the 'usual' empirical process ν_n,

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

and for the parametric empirical process

$$\nu_n(x,\hat{\theta}) = \sqrt{n}[F_n(x)-F(x,\hat{\theta})]$$

In connection with this the second but more modest 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.

Report BS-R8904
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The formal setting of the problem will be given in § 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 in the seminar hold in Moskow State University listening to the lecture of V. Glivenko, A. 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 idea of this lemma was quickly grasped by N. Smirnov, who sugested replacing the original form of Cramér-von Mises statistic

$$n\int[F_n(x)-F_0(x)]^2dx$$

by its present form

$$n\int[F_n(x)-F_0(x)]^2F_0(dx)$$

(SMIRNOV 1937), and by A. Wald and J. Wolfowitz who sugested considering confidence bounds for F based on statistics

$$\sup_x|F_n(x)-F(x)|\alpha[F(x)]$$

with a weight function α being essentially a function of F (WALD, WOLFOWITZ 1939). 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:

$$u_n(t) = \nu_n(x), \quad t = F_0(x). \tag{1}$$

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

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

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 statistics 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 "few" linear functionals of ν_n. Examples are the Neymann-Pearson statistic

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

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

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

statistics of C_α-tests etc. 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 u_n or not.

Tests based on one or "few" linear functionals are particularly sensitive to deviations from F_0 in one or "few" directions, but they are very insensitive to deviations in all other directions (see a precise statement for contiguous alternatives in § 2). Tests of the second kind—the goodness of fit tests—are of different behaviour. 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 Kolmogorov-Smirnov statistic

$$\sup|v_n(x)|,$$

the Anderson-Darling statistic

$$\int \frac{v_n^2(x)}{F_0(x)[1-F_0(x)]} F_0(dx),$$

ω^2-; or Cramer-von Mises statistic

$$\int v_n^2(x) F_0(dx)$$

are well known. Recall that it was quite difficult to derive and to calculate the limit distribution of each of these statistics. It is hard to calculate the limit distributions of weighted Kolmogorov-Smirnov or weighted ω^2 statistics except for a few special weight functions.

Because of this it is of prime practical importance that we have to 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) does not lead to distribution-free processes if the X_i's are m-dimensional random vectors with $m \geq 2$. Since (GIHMAN 1953, 1954) and (KAC, KIEFER, WOLFOVITZ 1955) it became clear that in the case of a parametric hypothesis $F \in \mathbb{F}$ if we consider the natural analogue of (1)

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

where $\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$ is an estimator of the unknown value of the parameter θ and $v_n(\cdot, \hat{\theta})$ stands for the parametric empirical process,

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

it does not lead to distribution free or asymptotically distribution free process as well (see § 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 θ, 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 purpose of this paper is

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 § 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), $m \geq 2$, $F=F_0$ and $m \geq 2$, $F \in \mathbb{F}$.

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 § 2). Put $m=1$ and transform the process $\hat{v}$ to its innovation martingale $\hat{w}$ (see definitions, e.g., in (LIPTSER, SHIRYAYEV 1977), which is a Gaussian process with independent increments and covariance unction $F(x \wedge y, \theta)$

where θ denotes the 'true value' of 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 $\hat{v}_n(\cdot, \hat{\theta})$ 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 '$m=1$, $F \in \mathbb{F}$'.

But attempts to develop a similar approach in the case $m \geqslant 2$ even for the simple hypothesis $F=F_0$ did not bring a success for quite a long time (until as late as (KHMALADZE, 1986) and (NIKABADZE, KHMALADZE 1987)). The problem is that it is not clear how to construct and even what to call an innovation processes for processes with multidimensional time parameter x. Let us explain this difficulty by example.

Let v be Brownian bridge on $[0,1]^m$ with respect to (w.r.t.) $F(\cdot)$, that is, the Gaussian process with mean 0 and covariance function $F(x \wedge y) - F(x)F(y)$, and let $m=1$. Consider the partition of $[0,1]$ by N points $\{i/N\}$, $i \leqslant N$. The Gaussian vector of increments $\{\Delta v(i/N)\}$, $i \leqslant N$, where $\Delta v(i/N) = v((i+1)/N) - v(i/N)$, has dependent coordinates, $E\Delta v(i/N)\Delta v(j/N) = -\Delta F(i/N)\Delta F(j/N)$ for $i \neq j$. Consider now the transformation of $\{\Delta v(i/N)\}$ into the Gaussian vector $\{\Delta w(i/N)\}$ defined as follows:

$$\Delta w_N(i/N) = \Delta v(\frac{i}{N}) - E[\Delta v(\frac{i}{N})|\mathfrak{F}_i^N] = \Delta v(\frac{i}{n}) + \frac{v(\frac{i}{N})}{1-F(\frac{i}{N})}\Delta F(\frac{i}{N}) \tag{1}$$

where the σ-algebra $\mathfrak{F}_i^N$ is generated by $v(j/N)$, $j \leqslant i$. It is clear that $\{\Delta w_N(i/N)\}$ is a Gaussian vector with independent coordinates. Now let

$$w_N(x) = \sum_{i<Nx} \Delta w_N(\frac{i}{N}).$$

The limit of the process w_N as $N \to \infty$ is a Gaussian process with independent increments

$$w(x) = v(x) + \int_0^x \frac{v(y)}{1-F(y)} F(dy)$$

(with covariance function $F(x \wedge y)$). The relation between v and w is one-to-one. The process w is called the innovation process for the process v.

But if we let $m=2$ consider a similar partition of $[0,1]^2$ by points $\{\mathbf{i}/N\}$, $i=(i_1, i_2)$, $i_1, i_2 \leqslant N$, and the σ-algebras $\mathfrak{F}_{\mathbf{i}}^N$ generated by $v(\mathbf{j}/N)$, $\mathbf{j} \leqslant \mathbf{i}$, the increments

$$\Delta M_N(\frac{\mathbf{i}}{N}) = \Delta v(\frac{\mathbf{i}}{N}) - E[\Delta v(\frac{\mathbf{i}}{N})|\mathfrak{F}_{\mathbf{i}}^N] = \Delta v(\frac{\mathbf{i}}{N}) + \frac{v(\frac{\mathbf{i}}{N})}{1-F(\frac{\mathbf{i}}{N})}\Delta F(\frac{\mathbf{i}}{N})$$

are not any more independent random variables (obviously, $\Delta v(\mathbf{i}/N)$ denotes an increment on the rectangle $[\mathbf{i}/N, (\mathbf{i}+\mathbb{1})/N]$, $\mathbb{1}=(1,1)$, that is

$$\Delta v(\mathbf{i}/N) = v(\frac{\mathbf{i}+\mathbb{1}}{N}) - v(\frac{i_1+1}{N}, \frac{i_2}{N}) - v(\frac{i_1}{N}, \frac{i_2+1}{N}) + v(\frac{i_1}{N}, \frac{i_2}{N}).$$

The reason for this phenomenon lies in the fact that if $m \geqslant 2$ the family of σ-algebras $\{\mathfrak{F}_i^N\}$ is not linearly ordered any more w.r.t. inclusion: it is not true that for any $\mathbf{i}, \mathbf{j}$ either $\mathfrak{F}_{\mathbf{i}}^N \subseteq \mathfrak{F}_{\mathbf{j}}^N$ or $\mathfrak{F}_{\mathbf{j}}^n \subseteq \mathfrak{F}_{\mathbf{i}}^N$. As a result, the process

$$M_N(x) = \sum_{\mathbf{i}<Nx} \Delta M_N(\frac{\mathbf{i}}{N})$$

and its limit

$$M(x) = v(x) + \int_{y \leqslant x} \frac{v(y)}{1-F(y)} F(dy)$$

does not have sufficiently 'simple' covariance function and to transform M to a process with a standard distribution does not seem easier then to transform v itself.

If we replace the σ-algebras $\mathcal{F}_{\mathrm{i}}^{N}$ by σ-algebras

$$\mathcal{H}_{\mathrm{i}}^{N} = \mathcal{F}_{(1,i_2)}^{N} \vee \mathcal{F}_{(i_1,1)}^{N}, \quad \mathrm{i}=(i_1, i_2),$$

which frequently appear in the theory of biparametric martingales (see, in particular, (CAIROLI, WALSH 1975)), what we get will be the process

$$M'(x) = v(x) + \int_{(\tau,\sigma)\leqslant x} \frac{v(1,\tau)+v(\sigma,1)-v(\tau,\sigma)}{1-F(1,\tau)-F(\sigma,1)+F(\tau,\sigma)} F(d\tau, d\sigma)$$

with dependent increments—again not what we seek for.

The reader interested in the theory of biparametric martingales (see e.g. the fundamental papers of (CAIROLI, WALSH 1975) and WONG, ZAKAI (1974) and the review paper (GIHMAN 1982)) may wish to remark, that the process $\{M(x), \mathcal{F}_x\}$, $\mathcal{F}_x = \sigma\{v(y), y \leqslant x\}$, is a weak martingale but not a strong martingale. The process $\{M'(x), \mathcal{F}_x\}$ is also not a strong martingale since $M'(x)$ is not $\mathcal{F}_x$-measurable.

In § 3 the 'scanning innovation' processes are introduced (see (KHMALADZE 1986)), which are one-to-one transformations of v (and $\hat{v}$) and which are Gaussian processes with independent increments. These scanning innovations can be introduced for x of any finite dimension $m \geqslant 1$ but it was interesting to find out whether or not they can be introduced for an infinite-dimensional time parameter. In § 3 we consider the function parametric version of $\hat{v}$,

$$\hat{v}(f) = \int_{x \in [0,1]^m} f(x)\hat{v}(dx)$$

and introduce scanning innovations for $\hat{v}(f)$.

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.

Let $X_1, \cdots, X_n$ be i.i.d. random vectors taking values in m-dimensional Euclidean space $\mathbb{R}^m$, and F denote the unknown distribution of each X_i. Let $\mathbb{F} = \{F(\cdot, \theta), \theta \in \Theta\}$ be some family of distributions depending on a k-dimensional parameter θ. We are interested in the problem of testing the parametric hypothesis $F \in \mathbb{F}$.

The parametric family $\mathbb{F}$. Suppose the range Θ of θ to be an open subset of $\mathbb{R}^k$. Let us assume from now on that all $F(\cdot, \theta)$ have support in $[0,1]^m$.(*) Assume also the following regularity conditions on $\mathbb{F}$—each $F(\cdot, \theta)$ is absolutely continuous w.r.t. Lebesque measure and the corresponding densities $f(\cdot, \theta)$ have the following regularity properties:

1) the k-dimensional vector-function

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

is square integrable:

$$\int g'^T(x, \theta) g'(x, \theta) F(dx, \theta) < \infty.$$

*) During the final preparation of the manuscript the author realizes that this condition is only a reminicence of some earlier versions. Though it brings some notational convenience it is not necessary in § 2 or in § 3. But at this stage it is too late to work a careful elimination of this condition throughout the whole text.

As a consequence the Fisher information matrix

$$B(\theta) = \int g'(x, \theta) g'^T(x, \theta) F(dx, \theta)$$

is finite. Here and everywhere below α^T means the transpose of the column-vector α;

2) if

$$\xi(x, \theta, \epsilon) = \sup_{\Theta:\, \|\Theta - \theta\| < \epsilon} |g'(x, \Theta) - g'(x, \theta)|$$

then

$$\int \xi^T(x, \theta, \epsilon) \xi(x, \theta, \epsilon) F(dx, \theta) \to 0$$

as $\epsilon \to 0$.

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

The condition 2) is more or less traditionally used in asymptotic statistics—cf. condition c) of § 7 Ch. I of (IBRAGIMOV, HAS'MINSKII 1981) or 2 Definition in Ch. VII of (POLLARD 1984). We will need it in estimation of the remainder in Lemma 2.

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

$$\nu_n(x, \hat{\theta}) = \sqrt{n}[F_n(x) - F(x, \hat{\theta})], \quad F_n(x) = \frac{1}{n} \sum_1^n I\{X_i \leq x\}$$

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

The estimator $\hat{\theta}$. Suppose

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

$$\int g'(x, \theta) l^T(x, \theta) F(dx, \theta) = I_k, \tag{1}$$

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

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

An estimator $\hat{\theta}$, which satisfies condition 3) we call projective (cf. (KHMALADZE 1979)). The reason for this definition is explained by Lemma 1 below.

Suppose, for example, that θ is one-dimensional and $\hat{\theta}$ is an M-estimator, that is $\hat{\theta}$ is the root of the equation

$$\sum_1^n \psi(X_i, \theta) = 0$$

for some function $\psi(x, \theta)$ such that

$$\int \psi(x, \theta) F(dx, \theta) = 0, \quad \int \psi^2(x, \theta) F(dx, \theta) < \infty.$$

If $\mathbb{F}$ is regular and ψ possesses some regularity as well, that is, if $\int \psi(x, \theta) F(dx, \theta)$ can be differentiated w.r.t. θ under the integral sign,

$$\int \frac{\partial}{\partial \theta} \psi(x, \theta) F(dx, \theta) + \int \psi(x, \theta) g'(x, \theta) F(dx, \theta) = 0,$$

and if

$$\sqrt{n}(\hat{\theta}-\theta) = -\frac{1}{\int \frac{\partial}{\partial \theta} \psi(x, \theta) F(dx, \theta)} \frac{1}{\sqrt{n}} \sum_1^n \psi(X_i, \theta) + o_P(1)$$

then the condition (1) is satisfied with

$$l(x, \theta) = -\psi(x, \theta) / \int \frac{\partial}{\partial \theta} \psi(x, \theta) F(dx, \theta).$$

We will formulate limit theorems for ν_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 $n = 1,2, \cdots$ the random vectors $X_1, \cdots, X_n$ are again i.i.d. with distribution A_n, which has the following properties:

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

$$n \operatorname{var}(A_n^s) \to 0, \quad n \to \infty \tag{2}$$

$$\left[\frac{dA_n^c(\cdot)}{dF(\cdot, \theta)}\right]^{1/2} = 1 + \frac{1}{2\sqrt{n}} h_n(\cdot) \tag{3}$$

and

$$\int [h_n(x) - h(x)]^2 F(dx, \theta) \to 0 \tag{4}$$

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

$$\int h^2(x) F(dx, \theta) < \infty \tag{5}$$

$$\int h(x) F(dx, \theta) = 0 \tag{6}$$

$$\int h(x) g'(x, \theta) F(dx, \theta) = 0. \tag{7}$$

Hence, under the hypothesis, the distribution of the sample $X_1, \cdots, X_n$ 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 $\mathbb{P}_{n\theta}(h) = A_n \times \cdots \times A_n$. According to (OOSTERHOFF, VAN ZWET 1979) condition 4) guarantees that the sequence $\{\mathbb{P}_{n\theta}(h)\}$ is contiguous w.r.t. the sequence $\{\mathbb{P}_{n\theta}\}$: $\mathbb{P}_{n\theta} \lhd \mathbb{P}_{n\theta}$. In fact, according to (OOSTERHOOF, VAN ZWET 1979) the necessary and sufficient conditions for $\mathbb{P}_{n\theta}(h) \lhd \mathbb{P}_{n\theta}$ are conditions (2), (3) and

$$\limsup_{n \to \infty} \int h_n^2(x) F(dx, \theta) < \infty. \tag{8}$$

Conditions (4) and (5) slightly streng then condition (8). The function h which participates in these conditions can be viewed as a function which determines from what 'direction' the alternative distribution A_n approach some hypothetical distribution $F(\cdot, \theta)$.

Let us remark that any function h, which satisfies (4) and (5) must satisfy condition (6). But the orthogonality condition (7) is an additional requirement on h. This requirement is convenient and natural as can be seen later, but not necessary for the further development.

We are ready now to formulate the statements concerning convergence in distribution of $\nu_n = \nu_n(\cdot, \theta)$ and $\nu_n(\cdot, \hat{\theta})$. But it seems convenient to describe first their limits in distribution—the Gaussian processes ν and $\hat{\nu}$.

The Gaussian processes v and $\hat{v}$. Denote by $b(\cdot,\theta)$ the Gaussian process on $[0,1]^m$ with mean 0 and covariance function $F(x\wedge y,\theta)$. Let

$$v(x,\theta) = b(x,\theta)-F(x,\theta)b(\mathbb{1},\theta), \quad \mathbb{1}=(1,\cdots,1)\in\mathbb{R}^m. \tag{9}$$

The covariance function of $v(\cdot,\theta)$ is $F(x\wedge y,\theta)-F(x,\theta)F(y,\theta)$. Let us call $b(\cdot,\theta)$ and $v(\cdot,\theta)$ the Wiener process w.r.t. $F(\cdot,\theta)$ and Brownian bridge w.r.t. $F(\cdot,\theta)$ respectively, and let us omit θ in $b(\cdot,\theta)$, $v(\cdot,\theta)$ etc., when it does not lead to misunderstanding. Now, let

$$\hat{v}(x,\theta) = v(x,\theta)-g^T(x,\theta)\int_{0\leq y\leq 1} l(y,\theta)v(dy,\theta)=[\Pi_1 v(\cdot,\theta)](x), \tag{10}$$

where $g(x,\theta)=\int_{y\leq x} g'(y,\theta)F(dy,\theta)$. It is convenient to introduce extended vector-functions

$$g'(x) = \begin{bmatrix}1\\ g'(x,\theta)\end{bmatrix},\quad g(x) = \begin{bmatrix}F(x,\theta)\\ g(x,\theta)\end{bmatrix},\quad l(x)=\begin{bmatrix}1\\ l(x,\theta)\end{bmatrix} \tag{11}$$

and to substitute (9) in (10)—if (1) is satisfied then

$$\hat{v}(x,\theta) = b(x,\theta)-g^T(x)\int_{0\leq y\leq 1} l(y)b(dy,\theta)=[\Pi b(\cdot,\theta)](x). \tag{12}$$

LEMMA 1. *The transformation (9) of b to v is a projection. If (1) is satisfied, then the transformation of v to $\hat{v}$ is a projection. Consequently, if (1) is satisfied, then the transformation (12) of b to $\hat{v}$ is a projection.*

PROOF. Let us prove the last statement only. What we need is to show for the linear transformation Π that $\Pi\Pi=\Pi$ is true. But the definition of $g'(\cdot,\theta)$ and condition (1) lead to the biorthogonality condition

$$\int g'(x)l^T(x)F(dx,\theta) = I_{k+1} \tag{13}$$

and equality $\Pi\Pi b=\Pi b$ is the direct consequence of (13). □

It is easy to notice that the kernel of the projection Π does not depend on the choice of l. It is always the linear subspace $Ker\Pi=\{\alpha^T g,\ \alpha\in\mathbb{R}^{k+1}\}$.

REMARK. The study of $\hat{v}$ as a projection of b does not lie in the main stream of the present text. That is why we avoid here a more rigorous description of Π. More precise text can be found, e.g., 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).

As we see below, the asymptotic shifts of $v_n(\cdot,\theta)$ and $v_n(\cdot,\hat{\theta})$ under $\tilde{\mathbb{P}}_{n\theta}(h)$ are the functions

$$H(x) = \int_{y\leq x} h(y)F(dy,\theta)$$

and

$$\hat{H}(x) = \Pi H(x)=H(x)-g^T(x)\int_{0\leq y\leq 1} l(y)h(y)F(dy,\theta) \tag{14}$$

respectively.

Now we turn to convergence in distribution of $v_n(\cdot,\hat{\theta})$ and $v_n(\cdot,\theta)$, which we will prove in the space $D[0,1]^m$ introduced by (BICKEL, WICHURA 1971) and (NEUHAUS 1971). This space is conveniently described, e.g. in Ch. 2 of (SEN 1981).

THEOREM. *As* $n \to \infty$

$$\nu_n(\cdot, \theta) \xrightarrow{\mathfrak{D}(\mathbb{P}_{n\theta})} \nu.$$

If the sequence $\{A_n\}$ *satisfies conditions (2)—(6), then*

$$\nu_n(\cdot, \theta) \xrightarrow{\mathfrak{D}(\tilde{\mathbb{P}}_{n\theta}(h))} \nu + H.$$

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

$$\nu_n(\cdot, \hat{\theta}) \xrightarrow{\mathfrak{D}(\mathbb{P}_{n\theta})} \hat{\nu}.$$

If $\mathbb{F}$ *is regular,* $\hat{\theta}$ *is projective and* $\{A_n\}$ *satisfies conditions (2)—(6), then*

$$\nu_n(\cdot, \hat{\theta}) \xrightarrow{\mathfrak{D}(\tilde{\mathbb{P}}_{n\theta}(h))} \hat{\nu} + \hat{H}.$$

According to this theorem, the substitution of a projective estimator instead of the true parameter value is asymptotically equivalent to projection of ν_n parallel to the function $g'(\cdot, \theta)$.

Define by $(\phi, \xi) = \int \phi(x)\xi(x)F(dx, \theta)$.

COROLLARY. *Under the conditions of the theorem, for any function* ϕ*, such that* $(\phi, \phi) < \infty$*, we get as* $n \to \infty$

$$\nu_n(\phi) = \int \phi(x)\nu_n(dx, \theta) \xrightarrow{\mathfrak{D}(\mathbb{P}_{n\theta})} \int \phi(x)\nu(dx) = \nu(\phi)$$

$$\nu_n(\phi) \xrightarrow{\mathfrak{D}(\tilde{\mathbb{P}}_{n\theta})} \nu(\phi) + (\phi, h)$$

$$\hat{\nu}_n(\phi) = \int \phi(x)\nu_n(dx, \hat{\theta}) \xrightarrow{\mathfrak{D}(\mathbb{P}_{n\theta})} \hat{\nu}(\phi)$$

$$\hat{\nu}_n(\phi) \xrightarrow{\mathfrak{D}(\tilde{\mathbb{P}}_{n}, \theta)} \hat{\nu}(\phi) + (\phi, h) - (\phi, g'^T(\cdot, \theta))(l, h).$$

This Corollary is included to give some support to the informal reasoning in the introduction: being linear functionals of Gaussian processes, $\nu(\phi)$ and $\hat{\nu}(\phi)$ are Gaussian random variables, hence $\nu_n(\phi)$ and $\hat{\nu}_n(\phi)$ are asymptotically Gaussian indeed, and for all sequences of alternatives $\tilde{\mathbb{P}}_{n\theta}(h)$ such that

$$(\phi, h) = 0 \quad (\text{and } (\phi, g'(\cdot, \theta)) = 0) \tag{15}$$

the limit distributions of $\nu_n(\phi)$ (and $\hat{\nu}_n(\phi)$) are the same as the limit distributions under the hypothesis $\mathbb{P}_{n\theta}$, hence tests based on these linear functionals asymptotically cannot distinguish between $\mathbb{P}_{n\theta}$ and all $\tilde{\mathbb{P}}_{n\theta}(h)$ with the orthogonality property (15).

PROOF OF THEOREM. The statement concerning the empirical process $\nu_n(\cdot, \theta)$ is well-known and there is no need to prove it here (see, e.g., GAENSSLER and STUTE (1979)). The statement concerning $\nu_n(\cdot, \hat{\theta})$ is a consequence of the following

LEMMA 2. *Let*

$$r_n = \nu_n(\cdot, \hat{\theta}) - \Pi_1 \nu_n(\cdot, \theta).$$

If $\mathbb{F}$ *is regular, then under* $\mathbb{P}_{n\theta}$ *and, hence, under* $\tilde{\mathbb{P}}_{n\theta}(h)$

$$\int \left[\frac{\partial r_n(x)}{\partial F(x, \theta)} \right]^2 F(dx, \theta) \xrightarrow{P} 0, \quad n \to \infty.$$

PROOF. Let us remark that

$$\frac{\partial g(x, \theta)}{\partial F(x, \theta)} = g'(x, \theta).$$

Now using a Taylor expansion for $\nu_n(\cdot, \hat{\theta})$ in $\hat{\theta} - \theta$ and condition (3) one can write down

$$\frac{\partial r_n(x)}{\partial F(x, \theta)} = [g'(x, \tilde{\theta}) - g'(x, \theta)]^T \sqrt{n}(\hat{\theta} - \theta) + g'^T(x, \theta)[\sqrt{n}(\hat{\theta} - \theta) - \int l(y) \nu_n(dy, \theta)].$$

But the first summand in the right-hand side is small because of condition (2) and the second summand is small because of conditions (1) and (3). □

Now for the proof of the statement concerning $\nu_n(\cdot, \hat{\theta})$ it is sufficient to prove that $\Pi_1 \nu_n(\cdot, \theta)$ has prescribed limits in distribution. But this can be done easily by the standard arguments—approximate $\int l d\nu_n$ by a continuous linear functional $\int \tilde{l} d\nu_n$ of $\nu_n(\cdot, \theta)$ and then use for $\nu_n - g \int \tilde{l} d\nu_n$ the continuous mapping theorem and Theorem 4.2 of (BILLINGSLEY 1968). □

We will need also some statements to judge how 'sensitive' the processes $\nu_n(\cdot, \theta)$ and $\nu_n(\cdot, \hat{\theta})$ are to the alternatives $\tilde{\mathbb{P}}_{n\theta}(h)$. First let us see 'how far' are the sequences $\{\mathbb{P}_{n\theta}\}$ and $\{\tilde{\mathbb{P}}_{n\theta}(h)$ from each other. Denote $\nu(P, Q)$ the distance in variation between distributions P and Q:

$$\nu(P, Q) = \sup_{B \in \mathfrak{B}} |P(B) - Q(B)|,$$

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

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

LEMMA 3. *If the sequence* $\{A_n\}$ *satisfies conditions (2)-(6), then*

$$\nu(\tilde{\mathbb{P}}_{n\theta}(h), \mathbb{P}_{n\theta}) \to \lambda(h), \quad n \to \infty.$$

PROOF. This will not be given in full details: the first and main point is that under $\mathbb{P}_{n\theta}$

$$\sum_{i=1}^{n} \ln \frac{dA_n(X_i)}{dF(X_i, \theta)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(X_i) - \frac{1}{2}(h, h) + o_P(1) \tag{16}$$

(see (OOSTERHOFF, VAN ZWET 1979) or (GREENWOOD, SHIRYAYEV 1985)). Now, according to the central limit theorem

$$\frac{1}{\sqrt{n}} \sum_{1}^{n} h(X_i) \xrightarrow{\mathfrak{D}(\mathbb{P}_{n\theta})} \mathfrak{N}(0, (h, h))$$

where $\mathfrak{N}(\mu, \sigma^2)$ denotes a normal random variable with mean μ and variance σ^2. Hence

$$\sum_{1}^{n} \ln \frac{dA_n(X_i)}{dF(X_i, \theta)} \xrightarrow{\mathfrak{D}(\mathbb{P}_{n\theta})} \mathfrak{N}(-\tfrac{1}{2}(h, h), (h, h)). \tag{17}$$

As a consequence (see, e.g., Theorems I, VII in (KHMALADZE 1975))

$$\sum_1^n \ln\frac{dA_n(X_i)}{dF(X_i,\theta)} \xrightarrow{\mathcal{L}(\tilde{\mathbb{P}}_{n\theta}(h))} \mathcal{N}(\frac{1}{2}(h,h),(h,h)). \tag{18}$$

Since

$$\nu(\tilde{\mathbb{P}}_{n\theta}(h),\mathbb{P}_{n\theta}) = \tilde{\mathbb{P}}_{n\theta}(h,B_n)-\mathbb{P}_{n\theta}(B_n)$$

where the event B_n is

$$B_n = \{\sum_1^n \ln\frac{dA_n(X_i)}{dF(X_i,\theta)} > 0\},$$

(17) and (18) imply (cf. VII in (KHMALADZE 1975))

$$\nu(\tilde{\mathbb{P}}_{n\theta}(h),\mathbb{P}_{n\theta}) \to \Phi(\frac{1}{2}\|h\|)-\Phi(-\frac{1}{2}\|h\|) = \lambda(h).$$

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

LEMMA 4. *$\nu(P^\nu, P^{\nu+H})=\lambda(h)$. If (7) is satisfied then $\nu(P^{\hat{\nu}}, P^{\hat{\nu}+H})=\lambda(h)$.*

PROOF. Consider first the linear functionals

$$\nu(\phi) = \int\phi(x)\nu(dx),\quad \nu(\phi)+(\phi,h) = \int\phi(x)\nu(dx)+\int\phi(x)H(dx)$$

generated from ν and $\nu+H$ by the function ϕ such that $\|\phi\|<\infty$. Without loss of generality one can assume that

$$E\phi = \int\phi(x)F(dx,\theta) = 0.$$

Indeed, $\nu(\mathbb{1})=0$ implies

$$\int\phi(x)\nu(dx) = \int(\phi(x)-E\phi)\nu(dx) \tag{19}$$

and condition (6) implies $(\phi-E\phi,h)=(\phi,h)$. But then from (9) it follows

$$\nu(\phi)=b(\phi),\quad \nu(\phi)+(\phi,h) = b(\phi)+(\phi,h).$$

The distance in variation between the distributions of these Gaussian random variables, i.e. between the distributions $\Phi(\cdot/\|\phi\|)$ and $\Phi((\cdot-(\phi,h)/\|\phi\|)$ is equal to

$$2\Phi\left[\frac{|(\phi,h)|}{2\|\phi\|}\right] - 1. \tag{20}$$

The function ϕ, which maximizes (20) is $\phi=const.\ h$:

$$\max_\phi 2\Phi\left[\frac{|(\phi,h)|}{2\|\phi\|}\right]-1 = 2\Phi\left[\frac{\|h\|}{2}\right]-1.$$

But recall now, that since ν and $\nu+H$ are Gaussian processes

$$\nu(P^\nu,P^{\nu+H}) = \max_\phi \nu(P^{\nu(\phi)},P^{\nu(\phi)+(\phi,h)})$$

(see, e.g., (KUO 1975)). Hence, the first statement is proved. Consider now the linear functionals

$$\hat{\nu}(\phi) = \int\phi(x)\hat{\nu}(dx),\quad \hat{\nu}(\phi)+(\phi,h).$$

Again without loss of generality one can assume $\phi-(\phi,q^T)l=\phi$, since $\hat{\nu}(\phi)=\hat{\nu}(\phi-(\phi,q^T)l)$ and

$(\phi, \hat{h})=(\phi-(\phi, q^T)l, \hat{h})$. But then (12) implies $\hat{v}(\phi)=b(\phi)$ and condition (7) implies $h=h-(h, q^T)l$. Consequent we can proceed as above and choose $\phi=h$ that gives

$$\nu(P^{\hat{v}}, P^{\hat{v}+H}) = \max_{\phi} 2\Phi\left[\frac{|\phi, h|}{2\|\phi\|}\right]-1 = \lambda(h). \quad \square$$

Now we have prepared everything we will need in § 3. But before we conclude this present § 2 we would like to consider:

Asymptotically distribution free transformation of $v_n(\cdot, \hat{\theta})$ based on 'components' of $v_n(\cdot, \hat{\theta})$. Although this subsection looks like a deviation from the main line of this paper, it may help to notice better the difference between the formal problem a1) - a2) of the next § 3 and the less formal problem a1) - a2), b1) - b2). Besides the content of this subsection is of some practical value (cf. e.g., with (DURBIN, KNOTT, TAYLOR 1975)).

Consider the Karhunen-Loève expansion of the Gaussian process $\hat{v}$:

$$\hat{v}(x) = \sum \lambda_k V_k a_k(x) \tag{21}$$

where $\{\lambda_k^2\}$ is the sequence of eigenvalues and $\{a_k(\cdot)\}$ is the orthonormal sequence of eigenfunctions of the covariance function

$$R(v, y) = F(x \wedge y, \theta)-g^T(x)e(y)-e^T(x)g(y)+g^T(x)Dg(y) \tag{22}$$

of $\hat{v}$ (in (22) we use notations

$$e(x) = \int_{0 \leq y \leq x} l(y)F(dy, \theta), \quad D = \int l(y)l^T(y)F(dy, \theta).\ )$$

Expansion (21) establishes a one-to-one linear relation between $\hat{v}$ and a sequence $\{V_k\}$ of independent $\mathfrak{N}(0,1)$-random variables

$$V_k = \lambda_k^{-1}\int a_k(x)\hat{v}(x)dx.$$

Having this sequence of random variables with standard distribution one can choose some finite or infinite sequence $\{\sigma_k\}_1^N$, $\Sigma_1^N\sigma_k^2<\infty$, of coefficients and some orthonormal sequence of functions $\{\psi_k(\cdot)\}_1^N$ and 'construct' a process

$$\xi(x) = \sum \sigma_k V_k \psi_k(x). \tag{22}$$

Clearly ξ is a linear transformation of $\hat{v}$ and its distribution is free from $\mathbb{F}$ and at one's disposal, since the choice of $\{\sigma_k\}_1^N$ and $\{\psi_k(\cdot)\}_1^N$ is at one's disposal. One could now replace $\hat{v}$ by $v_n(\cdot, \hat{\theta})$ which will lead to a transformation of $v_n(\cdot, \theta)$ to an asymptotically distribution free process.

But this sugestion is very inconvenient practically, because it involves the spectral decomposition of $R(x, y)$—a too difficult problem to be used in a testing procedure for each parametric hypothesis.

Luckily one can exploit the fact that $\hat{v}$ is not just any Gaussian process with some covariance function, but it is the projection of the Wiener process b (see (12)).

LEMMA 5. (See (KHMALADZE 1979)). *If $\{\phi_k\}$ is any orthonormal sequence, i.e. if*

$$(\phi_k, \phi_n) = \delta_{kn}$$

and if

$$(\phi_k, q) = 0, \quad k = 1,2, \cdots, \tag{23}$$

then

$$Z_k = \int \phi_k(x)\hat{v}(dx), \quad k = 1,2, \cdots \tag{24}$$

are independent $\mathfrak{N}(0,1)$*-random variables.*

Hence, in (22) instead of V_k one can use Z_k. If in (24) $\hat{v}$ is replaced by $v_n(\cdot,\hat{\theta})$ we get random variables Z_{kn}, and the process

$$\xi_n(x) = \sum_{k=1}^{N} \sigma_k Z_{kn} \psi_k(x) \tag{25}$$

is a linear transformation of $v_n(\cdot,\hat{\theta})$ with prescribed, independent of $\mathbb{F}$, asymptotic distribution under the hypothesis.

The orthonormal sequence $\{\phi_k\}$ can be obtained from any orthogonal system by the Gram-Schmidt orthogonalization process to satisfy condition (23)—much more easily then the sequence $\{a_k\}$.

Hence, the process ξ_n defined by (25) and functionals of ξ_n can be used in practice for N not too large. As it's another advantage one can remark, that the process $v_n(\cdot,\hat{\theta})$ with multidimensional time-parameter can be transformed to ξ_n with one-dimensional time-parameter, which can simplfy the calculation of test statistics and their limit distributions. But the transformation (25) can satisfy conditions a1) - a2) only if $N=\infty$ and all $\sigma_k \neq 0$. A transformation that requires infinite series cannot be viewed as a 'simple' one, that is, (25) does not satisfy condition b1) of § 3. That was why after (KHMALADZE 1979) we were still looking for something else.

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:

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

It is common knowledge that $\mathfrak{K}$ transforms v_n to a distribution free, hence, asymptotically distribution free process. But this cannot be the only important property of the transformation $\mathfrak{K}$—for example, the transformation of v_n to the process which is identically 0, also leads, of course, to an asymptotically distribution free but useless, process. An alternative property of $\mathfrak{K}$ 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 Lemma 3 of § 2 and the next

LEMMA 1. *Let v be the process defined by (9) of § 2 and let $u = v \circ F^{-1}$ be a standard Brouwnian bridge. Then*

$$u_n \xrightarrow{\mathfrak{D}(\mathbb{P}_n)} u$$

$$u_n \xrightarrow{\mathfrak{D}(\tilde{\mathbb{P}}_n(h))} u + H \circ F^{-1}$$

and

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

PROOF. The convergence in distribution of u_n is the old and wellknown fact (see, e.g., (GAENSSLER, STUTE 1979) or (SHORACK, WELLNER 1986)). The last equality follows from Lemma 4 of § 2 and the fact that the transformation $\mathfrak{K}$ is one-to-one. □

Formulation of the problem. As it was described in the introduction, the transformation $\mathfrak{K}$ cannot be extended directly to the case of a parametric hypothesis and of a simple hypothesis for random vectors ($m \geqslant 2$). But why should we copy the transformation $\mathfrak{K}$ in all cases? Why cannot we find another transformation, which may differ from $\mathcal{K}$ in form, but which will lead to the same goal?

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

a1) $w[\nu_n, F_0] \xrightarrow{\mathfrak{D}(\mathbb{P}_n)} w$ 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)-(6) of § 2 $w[\nu_n, F_0] \xrightarrow{\mathfrak{D}(\tilde{\mathbb{P}}_n(h))} w'$ such that $\nu(P^w, P^{w'}) = \lambda(h)$.

As the test statistics one can choose now functionals $\phi[w[\nu_n, F]]$ of the process $w[\nu_n, F]$.

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

b1) the transformation $w[\nu_n, F]$ 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 feasable.

In the case of parametric hypotheses one can formulate a similar problem. In doing this let us preserve the notations of the conditions and use the shorter notation $\hat{\nu}_n$, $\hat{\nu}_n(x) = \nu_n(x, \hat{\theta})$. Now, we want: to find a transformation $w[\hat{\nu}_n, \mathbb{F}]$ which may depend on hypothetical parametric family $\mathbb{F}$ with the following properties:

a1) for each θ, $w[\hat{\nu}_n, \mathbb{F}] \xrightarrow{\mathfrak{D}(\mathbb{P}_{n\theta})} w$ and P^w does not depend on $\mathbb{F}$ if $\mathbb{F}$ is regular,

a2) for any sequence of alternatives $\{A_n\}$ satisfying conditions (2)-(7) of § 2 $w[\hat{\nu}_n, \mathbb{F}] \xrightarrow{\mathfrak{D}(\tilde{\mathbb{P}}_{n\theta}(h))} w'$ such that $\nu(P^w, P^{w'}) = \lambda(h)$.

Notice that now condition (7) of § 2 is required—this seems natural in view of Lemma 4 of § 2.

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

Our plan in what follows in this: we construct the one-to-one correspondence between the limiting Gaussian process $\hat{\nu}$ and some Gaussian process $\hat{w}$ with independent increments—the scanning innovation of $\hat{\nu}$. This is the first and the main step. Then we normalize $\hat{w}$ and get the standard Wiener process on $[0,1]^m$. In the resulting transformation of $\hat{\nu}$ to w we will substitute $\hat{\nu}_n$ instead of $\hat{\nu}$ and prove that this is a transformation with desirable properties. All this will be done for function-parametric versions of the processes involved, which versions we now introduce formally.

Function parametric processes. Denote by $L_2(\theta)$ the space of functions with the norm

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

In accordance with the notation already used in § 2 denote by (f, ϕ) the scalar product in $L_2(\theta)$:

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

Later we will need also another scalar product:

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

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

$$\nu_n(f) = \int f(x) \nu_n(dx, \theta) = \frac{1}{\sqrt{n}} \sum_1^h [f(X_i) - \int f(x) F(dx, \theta)] \tag{2}$$

(cf. (POLLARD 1984)). Clearly $\nu_n(f)$ for each $f \in L_2(\theta)$ is a random variable with finite variance,

$$E\nu_n^2(f) = (f, f) - (f, 1)^2.$$

Moreover

$$E\nu_n(f)\nu_n(\phi) = (f, \phi) - (f, 1)(\phi, 1).$$

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

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

According to Lemma 2 of § 2, if $\mathbb{F}$ is regular then

$$\hat{\nu}_n(f) = \Pi_1 \nu_n(f) + o_P(1), \quad n \to \infty. \tag{3}$$

It is not difficult to observe that

$$\Pi_1 \nu_n(f) = \nu_n(\Pi_1^* f) \tag{4}$$

where

$$\Pi_1^* f = f - (f, g'^T(\cdot, \theta)) l(\cdot, \theta)$$

(the functions $g'(\cdot, \theta)$ and $l(\cdot, \theta)$ are defined in conditions 1) and 3) of § 2). Notice that $\nu_n(f)$ is a bilinear functional — for each f it is a linear functional of the trajectories of the empirical process $\nu_n(\cdot, \theta)$ and for each trajectory of $\nu_n(\cdot, \theta)$ it is linear functional of f. Then the equality (3) simply says that Π_1^* is the adjoint projector of the projector Π_1 in the bilinear functional (2).

Notice that $\cap_{\theta \in \Theta} L_2(\theta)$ is frequently quite a rich set. If, e.g., $\mathbb{F}$ is the family of normal distributions with shift parameter θ, this set contains all functions with finite variance under any normal distribution with any mean θ.

Now let

$$b(f) = \int f(x) b(dx, \theta) \tag{5}$$

where the Wiener process $b(\cdot, \theta)$ is introduced just before (9) of § 2 , and let

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

According to the representation (12) of § 2

$$\hat{v}(f) = \Pi b(f) = b(f) - (f, q^T) b(l) = b(\Pi^* f), \tag{6}$$

where

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

is the adjoint projector of Π in the bilinear functional (5) (functions q and l are defined by (11) of § 2).

Denote by C the extended Fisher information matrix

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

Condition 4) below will guarantee that C has the unique inverse C^{-1}. Let us consider then the special choice of the function l:

$$\bar{l} = C^{-1} q \tag{7}$$

and denote

$$\bar{v}(f) = b(f) - (f, q^T) C^{-1} b(q). \tag{8}$$

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.

Now, for any two orthogonal projectors π' and π'' we call π'' larger then π', and denote this $\pi' \prec \pi''$, if $\pi'\pi'' = \pi'$. Let $\{\pi_\lambda\}$, $0 \leqslant \lambda \leqslant 1$, be a family of orthogonal projectors, defined on each $L_2(\theta)$. Assume that $\{\pi_\lambda\}$ has the following properties:

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

Recall for the reader's convenience some identities, which we will use later without comments: for orthoprojectors π, π', π'', $\pi' \prec \pi''$, we have

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

One can imagine the family $\{\pi_\lambda\}$ to be constructed as follows: let $\{A_\lambda\}$, $0 \leqslant \lambda \leqslant 1$, be a family of measurable subsets of $[0,1]^m$ with the following properties

1′) $\lambda \leqslant \lambda' \Rightarrow A_\lambda \subset A_{\lambda'}$

2′) $\mu(A_0) = 0, \quad \mu(A_1) = 1$

3′) $\mu(A_{\lambda'} \setminus A_\lambda) \to 0 \quad \text{if} \quad \lambda' \downarrow \lambda,$

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

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

If π_λ 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 specific condition on the function q and the family $\{\pi_\lambda\}$:

4) for any $\lambda \in [0,1]$ the matrix

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

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

Obviously $C_0 = C$. Condition 4) is convenient rather then necessary (cf. (Nikabadze 1990)), 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 $\bar{v}(f)$.

Innovation process for $\bar{v}(f)$. Associate with each λ the σ-algebra

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

Let us understand this σ-algebra as the one containing 'the past' of $\bar{v}(f)$ up to 'the moment' λ. Let us understand $\bar{v}(f)$ as an increment forward at λ 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 $\{\bar{v}(\pi_\lambda f), \mathfrak{F}_\lambda^v\}$. Let us replace this, still uncertain problem by another one: consider the σ-algebras

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

and

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

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

$$\hat{w}(\Delta\pi_\lambda f) = b(\Delta\pi_\lambda f) - E[b(\Delta\pi_\lambda f) | \mathfrak{F}_\lambda] \tag{10}$$

(cf. (1) of § 1). 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 $\mathfrak{F}_\lambda$. Hence

$$E[b(\Delta\pi_\lambda f)|\mathfrak{F}_\lambda] = E[b(\Delta\pi_\lambda f)|b(\pi_\lambda^\perp q)]$$
$$= (\Delta\pi_\lambda f, \pi_\lambda^\perp q^T)C_\lambda^{-1}b(\pi_\lambda^\perp q) = (f, \Delta\pi_\lambda q^T)C_\lambda^{-1}b(\pi_\lambda^\perp q). \tag{11}$$

Expressions (10) and (11) lead to the following expression

$$\hat{w}(f) = b(f) - \int (f, d\pi_\lambda q^T)C_\lambda^{-1}b(\pi_\lambda^\perp q), \tag{12}$$

which still needs precise definition.

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

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

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

LEMMA 2. *If 1)-4) are satisfied, then for any f such that $\pi_{1-\epsilon}f = f$ for some $\epsilon>0$ 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

$$c(f) = \int (f, d\pi_\lambda q^T)C_\lambda^{-1}b(\pi_\lambda^\perp q). \tag{12'}$$

LEMMA 3. *If 1)-4) are satisfied and $f=\pi_{1-\epsilon}f$ for some $\epsilon>0$ then*

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

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

Let us denote this limit again as the integral (12'). Now the right-hand side of (12) exists for all $f\in L_2(\theta)$.

PROOF OF LEMMA 1. 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 $t=1-\alpha^T C_\lambda\alpha$. Therefore, for any $\alpha\in\mathbb{R}^{k+1}$ almost all trajectories of $\alpha^T b(\pi_\lambda^\perp q)$ are continuous in λ, which proves the lemma. □

REMARK. Another proof of this lemma follows from Theorem 13, Ch. VII.3 of (POLLARD 1984). Indeed the set of functions $\{\alpha^T\pi_\lambda^\perp q\}$, $0\leq\lambda\leq 1$, forms a subset in $L_2(\theta)$ with ϵ-net containing no more then $1+(\alpha^T C\alpha)^{1/2}/\epsilon$ 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. But according to condition 3) the norm $\|\alpha^T\pi_\lambda^\perp q\| = (\alpha^T C_\lambda\alpha)^{1/2}$ is continuous in λ. Therefore $b(\alpha^T\pi_\lambda^\perp q)$ and, hence $b(\pi_\lambda^\perp q)$ is continuous in λ.

PROOF OF LEMMA 2. Let $\{\lambda_i\}_{i=0}^N$ and $\{\mu_j\}_{j=0}^M$ be two partitions of $[0,1-\epsilon]$. Assume for simplicity that each $\lambda_i\in\{\mu_j\}_{j=0}^M$. Consider points μ_j which are contained between λ_i and λ_{i+1}. The corresponding sums in expression of $c_N(f)$ and $c_M(f)$ are respectively

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

and

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

Consider the difference

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

For any vector $\xi = (\xi_1, \cdots, \xi_{k+1})^T \in \mathbb{R}^{k+1}$ 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_λ^{-1} is continuous on $[0, 1-\epsilon]$ for any $\epsilon > 0$. Since $b(\pi_\lambda^{\perp} q)$ is also continuous in λ we can get

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

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

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

and consequently

$$|c_N(f) - c_M(f)| \le \sum_{j=0}^{M-1} \rho_1(f, \Delta\pi_{\mu_j} q)) \rho_{\delta_N}. \tag{13}$$

From (13) the statement of lemma will follow if we can prove that

$$\sum_{j=0}^{M-1} \rho_1(f, \Delta\pi_{\mu_j} q)) \le \text{const.} \tag{14}$$

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

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

and as a consequence

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

But

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

Therefore (14) is correct with

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

PROOF OF LEMMA 3. Using the formula

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

we can get by direct calculation: if $f = \pi_{1-\epsilon} f$ then

$$E[b(f)-c(f)]^2 = (f,f)-2\int_0^{1-\epsilon}(f,d\pi_\lambda q^T)C_\lambda^{-1}(\pi_\lambda^\perp q,f) +$$

$$+\int_0^{1-\epsilon}\int_0^{1-\epsilon}(f,d\pi_\lambda q^T)C_\lambda^{-1}C_{\lambda\vee\mu}C_\mu^{-1}(d\pi_\mu q,f).$$

Under conditions 3) and 4) both integrals exist and are the usual Stieltjes integrals. The function under the double integral sign is symmetric in λ and μ, therefore the integral is equal to

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

This equality and the previous one lead to

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

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

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

from (10) and get

$$\hat{w}(\Delta\pi_\lambda f) = \overline{v}(\Delta\pi_\lambda f)-E[\overline{v}(\Delta\pi_\lambda f)|\mathfrak{F}_\lambda]. \tag{15}$$

What we finally get from (12) and (15) is the expression

$$\hat{w}(f) = \overline{v}(f)-\int(f,d\pi_\lambda q^T)C_\lambda^{-1}\overline{v}(\pi_\lambda^\perp q). \tag{16}$$

Let us call $\hat{w}$ the *scanning innovation* of $\overline{v}$ and let us call the integral term in (16) the *compensator* of $\{\overline{v}(f),\mathfrak{F}_\lambda^v\}$. The adjective 'scanning' is clarified by Example 1 below. The term 'innovation' is motivated by Theorem 1 below.

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

Let us call following (POLLARD 1984) the function-parametric process $\{b(f),f\in\mathcal{G}\}$ a *Wiener process w.r.t* $(f,f)^{1/2}$ if for any finite number r, the random variables $b(f_1),\cdots,b(f_r)$, $f_i\in\mathcal{G}$, have a joint normal distribution with mean 0 and covariance matrix $((f_i,f_j))$, $i,j = 1,\cdots,2$, and if almost all trajectories of $\{b(f),f\in\mathcal{G}\}$ are bounded and uniformly continuous on $\mathcal{G}$. Let us call the Wiener process w.r.t. $<f,f>^{1/2}$ the *standard Wiener process.*

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

LEMMA 4. *If the density $f(\cdot,\theta)$ of the distribution $F(\cdot,\theta)$ is positive a.e. on $[0,1]^m$ and $\{\hat{w}(f),f\in\mathcal{G}\}$ is a Wiener process w.z.t. $(f,f)^{1/2}$, then $\{w(\phi),\phi\in\mathcal{G}'\}$ with $w(\phi) = \hat{w}(\phi/f^{1/2}(\cdot,\theta))$ and $\mathcal{G}'=\{\phi\colon \phi/f^{1/2}(\cdot\theta)\in\mathcal{G}\}$ is a standard Wiener process.*

PROOF OF LEMMA 4 is left to reader.

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), Ch. VII. $\square$

THEOREM 1. *Let $\mathcal{G}$ be a subset of $L_2(\theta)$ with a finite covering integral. Then the process $\{\hat{w}(f),f\in\mathcal{G}\}$, defined by (16) is a Wiener process w.r.t. $(f,f)^{1/2}$. For any subset $\mathcal{G}$ such that the closed linear span of $\mathcal{G}$ is $L_2(\theta)$ the relation between the processes $\{\hat{w}(f),f\in\mathcal{G}\}$ and $\{\overline{v}(f),f\in\mathcal{G}\}$ is one-to-one.*

REMARK.

1) Since $\{\hat{w}(f), f\in\mathcal{G}\}$ and $\{\bar{v}(f), f\in\mathcal{G}\}$ can be extended in a one-to-one way to corresponding processes with $\mathcal{G}$ 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 1.
2) This statement of the theorem can be refined as follows: for any subset $\mathcal{G}$ the relation between the processes $\{\hat{w}(f), f\in\mathcal{G}, \hat{w}(\pi_\lambda^\perp q), \lambda\in[0,1]\}$ and $\{\bar{v}(f), f\in\mathcal{G}, \bar{v}(\pi_\lambda q), \lambda\in[0,1]\}$ is one-to-one.

PROOF. 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.

The boundedness and uniform continuity of trajectories of $\hat{w}(f)$ on $\mathcal{G}$ is proved in Theorem 13 Ch.VII of (POLLARD 1984). (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 done for the Wiener process on [0,1]—see, e.g., (ITO, 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.

LEMMA 3'. *The linear operator Z,*

$$Zf = f - \int (f, d\pi_\lambda q^T) C_\lambda^{-1} \pi_\lambda^\perp q, \tag{17}$$

is a unitary, i.e. norm-preserving, operator on $L_2(\theta)$.

Let us now rewrite (16) as

$$\hat{w}(f) = \bar{v}(Zf). \tag{18}$$

Consider the adjoint (in the scalar product (f,ϕ)) operator Z' of the operator Z:

$$Z'\phi = \phi - \int d\pi_\lambda q^T C_\lambda^{-1} (\pi_\lambda^\perp q, \phi). \tag{19}$$

Since Z is unitary, 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 (19) is

$$\hat{w}(Z'f) = \bar{v}(f). \tag{20}$$

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

LEMMA 5. $\operatorname{Im} Z = \{\phi\in L_2(\theta) : (\phi, q) = 0\}$. *Besides* $Z'q = 0$.

Now, (20) is correct on the subspace $\{\phi\in L_2(\theta): (\phi,q)=0\}$ since Z' is the inverse of Z, and (20) is correct for $f=\alpha^T q$ as well, since $\bar{v}(q) = 0 = \hat{w}(Z'q)$. Theorem 1 is proved.

PROOF OF LEMMA 5. Let us prove first that (19) can be determined for all $\phi \in L_2(\theta)$. Clearly the right hand side of (19) exists for all ϕ such that $\phi = \pi_{1-\epsilon}\phi$ for some $\epsilon>0$. For all such ϕ let us prove that

$$(Z'\phi, Z'\phi) = (\phi-(\phi,q^T)C^{-1}q,\phi) \tag{21}$$

and then let $\epsilon \to 0$. But

$$(Z'\phi,Z'\phi) = (\phi,\phi)-2\int(\phi,d\pi_\lambda q^T)C_\lambda^{-1}(\pi_\lambda^\perp q,\phi)+\int(\phi,\pi_\lambda^\perp q^T)C_\lambda^{-1}dC_\lambda C_\lambda^{-1}(\pi_\lambda^\perp q,\phi) =$$

$$= (\phi,\phi)-(\phi,\pi_\lambda^\perp q^T)C_\lambda^{-1}(\pi_\lambda^\perp q,\phi)|_{\lambda=0}^{1}. \tag{22}$$

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

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

(which is correct because of condition (3)). Finally from Lemma 6 below it follows that $(\phi,\pi_\lambda^\perp q)C_\lambda^{-1}(\pi_\lambda^\perp q,\phi) = 0$ at $\lambda = 1$. Hence (22) gives (21).

Now, it is clear that $(q,Zf)=0$, which implies Im $Z \subset \{\phi \in L_2(\theta):(\phi,q)=0\}$. Let now $\phi \neq 0$ belong to the last subspace. Then (21) 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 Im $Z \supset \{\phi \in L_2(\theta):(\phi,q) = 0\}$. □

LEMMA 6. *The following inequality is true:*

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

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

$$(\pi_\lambda^\perp\phi,\pi_\lambda^\perp\phi) = (\pi_\lambda^\perp\phi-\xi,\pi_\lambda^\perp\phi-\xi)+(\xi,\xi) \geq (\xi,\xi) = (\pi_\lambda^\perp\phi,\xi) = (\phi,\xi). \ \square$$

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 (16) is valid for any process $\hat{v}$ and the choice of $\bar{v}$ was simply a convenient way to derive (16).

THEOREM 2. *Let $\mathcal{G}$ be a subset of $L_2(\theta)$, and let the process $\hat{v}(f)$ be defined by (6). Then the processes $\hat{w}(f)$ defined by (16) and by*

$$\hat{w}(f) = \hat{v}(f)-\int(f,d\pi_\lambda^\perp q^T)C_\lambda^{-1}\hat{v}(\pi_\lambda^\perp q)=v(f)-\int(f,d\pi_\lambda^\perp q^T)C_\lambda^{-1}v(\pi_\lambda^\perp q) \tag{23}$$

coincide.

PROOF. Rewrite (23) as $\hat{w}(f) = \hat{v}(Zf) = v(Zf)$. According to definitions (6) and (9)

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

But 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

$$v(f)-\hat{v}(f) = (f,q^T)[\binom{b(1)}{0}-b(l)]$$

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 (23) is correct, and the processes (16) and (23) coincide. □

Now it is quite clear that Theorem 2 jointly with Lemma 4 gives the transformation of $\hat{v}$ to the standard Wiener process:

$$w(\phi) = \hat{v}(Z(\frac{\phi}{f^{1/2}(\cdot,\theta)})). \tag{24}$$

EXAMPLES. *Comparison with 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 $\pi(x)f(y) = I\{y \leq x\}f(y)$. Consider a partition of the range of t by points $0 = t_0 < t_1 < \cdots < t_N = 1$ and introduce the σ-algebras

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

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

where

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

and

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

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

$$w(\pi(\Delta s,\Delta t_i)f) = \overline{v}(\pi(\Delta s,\Delta t_i)f) - E[\overline{v}(\pi(\Delta s,\Delta t_i)f)|H_{(s,t_i)}]$$

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 more fine, i.e. if $\max|t_{i+1}-t_i| \to 0$ as $N \to \infty$, one can hope to glue these incrementss and get as a limit a gaussian process with independent increments. The only delicate question is whether one can neglect the σ-algebras $\mathcal{C}_{(s,t_i)}$ and consider only $\mathcal{F}_{(1,t_i)}$. If yes, then $\{\pi(1,t), t \in [0,1]\}$ is just an example of the 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 1 proves that in the case of $\overline{v}$ the σ-algebras $\mathcal{C}(s,t_i)$ really are negligible.

EXAMPLE 2. Consider particular cases of (16), (23). Suppose $f(y) = I\{y \leq x\}$. Let $x,y \in [0,1]$, i.e. let $m = 1$. Then

$$\begin{aligned} Zf(y) &= I\{y \leq x\} - \int I\{z \leq x\} q^T(z) C_z^{-1} F(dz,\theta) q(y) I\{y > z\} \\ &= I\{y \leq x\} - \int_{z \leq (x \wedge y)} q^T(z) C_z^{-1} F(dz,\theta) q(y). \end{aligned}$$

Consequently

$$\hat{w}(x) = \overline{v}(x) - \int_{z \leq x} q^T(z) C_z^{-1} \int_z^1 q(y) \overline{v}(dy) F(dz,\theta). \tag{25}$$

The Wiener process $\{\hat{w}(x), x \in [0,1]\}$ defined by (25) is just what was considered in (KHMALADZE 1981) and (25) is simply the Doob-Meyer decomposition of $\{\bar{v}(x), \mathcal{F}_x^v\}$. If a simple hypothesis is considered then $q=1$ and (25) gives

$$\hat{w}(x) = v(x) + \int_{z \leqslant x} \frac{v(z)}{1-F_0(z)} F(dz) \tag{26}$$

which is the Doob-Meyer decomposition of the Brownian bridge v.

If $x, y \in [0,1]^2$ and $\pi_\lambda f(y) = I\{y \leqslant (1,\lambda)\} f(y)$ then for $f(y) = I\{y \leqslant x\}$ from (23) in the case of a simple hypothesis we get

$$\hat{w}(x) = v(x) + \int_{(\tau,\sigma) \leqslant x} \frac{v(1,\sigma)}{1-F(1,\sigma)} F(d\tau, d\sigma) \tag{27}$$

—the scanning innovation of the Brownian bridge v (on $[0,1]^2$) as it was introduced in (KHMALADZE 1986). In the case of a parametric hypothesis we get

$$\hat{w}(x) = \bar{v}(x) - \int_{z \leqslant x} q^T(z) C_{(1,\sigma)}^{-1} F(dz, \theta) \int_{(1,\sigma) \leqslant y \leqslant (1,1)} q(y) \bar{v}(dy) \tag{28}$$

where σ is the second coordinate of z, and

$$C_{(1,\sigma)} = \int_{y \in [0,1]^2 \setminus [0,(1,\sigma)]} q(y) q^T(y) F(dy, \theta).$$

The Wiener process defined by (28) was considered in (NIKABADZE, KHMALADZE 1987).

Remark, that the set $\mathcal{G} = \{I\{\cdot \leqslant x\}, x \in [0,1]^m\}$ of functions $f(y) = I\{y \leqslant x\}$ satisfies the conditions of Theorem 1 for any finite m.

The processes (27) and (28) suggested in earlier papers left the impression of an essentially nonsymmetric solution—the choice of rectangles $[0,(1,t)]$ payed by some arbitrary reason too much attention to one of coordinates. Unsatisfied with this we looked for a more symmetric construction. Now, first, we are practically free in choice of $\{A_\lambda\}$ and, second, it is now obvious that for $m=1$ we have (on the real line) 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 $\{w(\phi), \phi \in \mathcal{G}'\}$, the process $\{w_n(\phi), \phi \in \mathcal{G}'\}$ 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 §3. In the present subsection we will prove that the provisional limiting processes of $\{w_n(\phi), \phi \in \mathcal{G}'\}$ under the hypothesis and alternative satisfy condition a2).

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

$$\nu(P^\xi, P^\eta) = \max\{\nu(P^{\xi(f)}, P^{\eta(f)}), f \in \Lambda(\mathcal{G})\}$$

where $\Lambda(\mathcal{G})$ denotes the closed linear span of $\mathcal{G}$ and $P^{\xi(f)}$ and $P^{\eta(f)}$ are Gaussian distributions of random variables $\xi(f)$ and $\eta(f)$ respectively (cf. this definition with proof of Lemma 4 in §2).

Denote by $H = H(f), f \in \mathcal{G}$, the function on $\mathcal{G}$ defined by $H(f) = (h, f)$ and denote by $J = J(\phi)$, $\phi \in \mathcal{G}$, the transformation of the function H similar to (24):

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

LEMMA 7. *Let $\mathcal{G}$ be a subset of $L_2(\theta)$ with the following properties: $\Lambda(\mathcal{G}) = L_2(\theta)$, and the set $\mathcal{G}' = \{\phi : \phi / f^{1/2}(\cdot, \theta) \in \mathcal{G}\}$ has finite covering integral in the norm $<\phi, \phi>^{1/2}$. If the function h satisfies conditions (5) - (7) of § 2 then*

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

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 7 says that the process $w+J$ is a good candidate for the limiting process w' of the condition a2).

PROOF. The second equality in the assertion of the lemma is already proved in Lemma 4 of § 2. 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:

$$\nu(P^{w(\phi)}, P^{w(\phi)+J(\phi)}) = 2\Phi\left[\frac{(h, Z\frac{\phi}{f^{1/2}(\cdot,\theta)})}{<\phi,\phi>^{1/2}}\right] - 1 = 2\Phi\left[\frac{<f^{1/2}(\cdot,\theta)Z'h,\phi>}{2<\phi,\phi>^{1/2}}\right] - 1$$

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

$$\frac{1}{2}<f^{1/2}(\cdot,\theta)Z'h, f^{1/2}(\cdot,\theta)Z'h>^{1/2} = \frac{1}{2}(Z'h, Z'h)^{1/2} = \frac{1}{2}(h,h)^{1/2}$$

where the equality follows from Lemma 3′. Hence

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

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 $f \in L_2$ introduce the random variable

$$\hat{w}_n(f) = \hat{v}_n(f) - \int (f, d\pi_\lambda q^T) C_\lambda^{-1} \hat{v}_n(\pi_\lambda^\perp q) \tag{29}$$

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 $n = 1,2,...$ and all $f \in L_2$ (cf. formula (34) below) and we will not dwell on this problem. It is technically convenient to get a little simpler approximation of $\hat{w}_n$.

LEMMA 8. *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 θ. Then both under $\mathbb{P}_{n\theta}$ and under $\tilde{\mathbb{P}}_{n\theta}$*

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

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. Using Lemma 2 of §2 and the property $Zq = 0$ one can write

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

where $\xi_n(x) = \partial r_n(x)/\partial F(x,\theta)$. Now Lemma 3′ leads to

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

and, hence,

$$\sup_{f \in S} |(Zf,\xi_n)| \leqslant c\|\xi_n\|_\theta \xrightarrow{P} 0$$

under $\mathbb{P}_{n\theta}$. But $\|\xi_n\|_\theta \xrightarrow{P} 0$ under $\tilde{\mathbb{P}}_{n\theta}$ as well because of contiguity of $\{\tilde{\mathbb{P}}_{n\theta}\}$ to $\{\mathbb{P}_{n\theta}\}$. $\square$

Denote

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

LEMMA 9. *Let S_ϵ be a bounded subset of L_2 of functions f such that $\pi_{1-\epsilon}f = f$. Then for each $\eta>0$ and $\Delta>0$ there exists $\delta>0$ for which*

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

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

PROOF. Apply to the scalar product

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

Schwarz's inequality:

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

where

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

But (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 9 follows from the inclusion

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

The reader can guess now that the convergence in distribution of $c_n(f)$ on a set of functions $f = \pi_{1-\epsilon}f$ is an easy matter: convergence in distribution at each f is easy to prove and tightness is granted by Lemma 9.

Denote by $\mathcal{X}(\mathcal{G})$ the space of bounded functions $x(f), f\in\mathcal{G}$, with the norm $\sup_{f\in\mathcal{G}}|x(f)|$ (cf. Sec. VII.5 of (POLLARD 1984).

THEOREM 3. *Suppose $\mathcal{G}$ is the subset of L_2 such that $|f|<c$ for the same constant c for all f and that in $\mathcal{X}(\mathcal{G})$*

$$v_n \xrightarrow{\mathcal{D}(\mathbb{P}_{n\theta})} v,\ v_n \xrightarrow{\mathcal{D}(\tilde{\mathbb{P}}_{n\theta})} v+H \tag{30}$$

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

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

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

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

REMARK. The condition of integrability of α_λ is mild but never the less 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

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

is not integrable for $q_\lambda\to 0$ 'too fast' as $\lambda\to 1$. If, e.g., $q(x) = \exp(-\frac{1}{1-x})$ then $\alpha_\lambda \sim 1/(1-\lambda)$. But if $q(x) = \exp(-\frac{1}{(1-x)^\beta})$, with $\beta<1$ then $\alpha_\lambda\sim 1/(1-\lambda)^\beta$ is integrable. Obviously α_λ is integrable for

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

$$\alpha_\lambda \leqslant q^2(\lambda)/\delta\left(\int_\lambda^1 q^2(x)dx\right)^{1/2}, \ \lambda > 1-\epsilon$$

and the right hand side is integrable. The condition of integrability of α, which we did not need in previous papers (KHMALADZE 1981, 1986), is the price we pay for the extension to 'very large' $\mathcal{G}$: as it will be clear from the proof of Theorem 3 (see (32)): if α_λ is integrable in a neighbourhood of 1, then $c_n(\cdot)$ converges in distribution in $\mathfrak{X}(\mathcal{G})$ for $\mathcal{G}$ being the set of all pointwise bounded functions, e.g. - of indicator functions of all measurable subsets of $[0,1]^m$.

PROOF. Replace $\hat{w}_n(f)$ by $\nu_n(Zf)$. One can do this because $\mathbb{F}$ is regular (Lemma 8). It is clear that for any square integrable function f the sequence $\{\nu_n(Zf)\}$ converges in distribution under $\mathbb{P}_{n\theta}$ to $\nu(Zf)$ (under $\tilde{\mathbb{P}}_{n\theta}$ to $\nu(Zf)+H(Zf)$) simply as a consequence of CLT. Let us verify tightness. Since $\nu_n = \{\nu_n(f), f\in\mathcal{G}\}$ converges in distribution the sequence $\{\nu_n\}$ is tight, and Lemma 9 asserts that the sequence $\{c_n(\pi_{1-\epsilon})\}$ is also tight. Since addition is a continuous operation in $\mathfrak{X}(\mathcal{G})$ the sequence $\{\nu_n - c_n(\pi_{1-\epsilon})\}$ is also tight. What remains is to consider the difference

$$\nu_n(Zf) - \nu_n(f) + c_n(\pi_{1-\epsilon}f) = c_n(\pi_{1-\epsilon}^{\perp}f).$$

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

$$\limsup_{n\to\infty} \mathbb{P}_{n\theta}\{\sup_{f\in\mathcal{G}}|c_n(\pi_{1-\epsilon}^{\perp}f)| > \Delta\} < \eta. \tag{31}$$

But

$$\sup_{f\in\mathcal{G}}|c_n(\pi_{1-\epsilon}^{\perp}f)| \leqslant c\int_{1-\epsilon}^1 |(1, d\pi_\lambda q^T)C_\lambda^{-1}\nu_n(\pi_\lambda^{\perp}q)| \tag{32}$$

and

$$E\int_{1-\epsilon}^1 |(1,d\pi_\lambda q^T)C_\lambda^{-1}\nu_n(\pi_\lambda^{\perp}q)| \leqslant \int_{1-\epsilon}^1 [E|(1,d\pi_\lambda q^T)C_\lambda^{-1}\nu_n(\pi_\lambda^{\perp}q)|^2]^{1/2}$$

$$= \int_{1-\epsilon}^1 [(1,d\pi_\lambda q^T)C_\lambda^{-1}(d\pi_\lambda q, 1)]^{1/2} = \int_{1-\epsilon}^1 \alpha_\lambda d\lambda. \tag{33}$$

Since α_λ is integrable the last integral can be made arbitrarily small if ϵ is small. Hence, the random variable in the right-hand side of (32) is small in probability for ϵ small and (31) is proved. □

REMARK. The approximation $\nu_n(Zf)$ of the 'empirical scanning innovation' $\hat{w}_n(f)$ can be replaced by an even more simple expression since

$$\nu_n(Zf) = \sqrt{n}F_n(Zf)$$

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

$$\sqrt{n}F_n(Zf) = \sqrt{n}[F_n(f) - \int (f, d\pi_\lambda q^T)C_\lambda^{-1}F_n(\pi_\lambda^{\perp}q)]. \tag{34}$$

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 (34) gives in a good sense its Doob-Meyer decomposition w.r.t. the filtration $\{\mathfrak{F}_\lambda^n\}$,

$$\mathcal{F}^n_\lambda = \sigma\{F_n(\pi_\lambda f), f \in L_2, F_n(q)\}.$$

The increments of $F_n(Zf)$ are not independent, of course, but they are uncorrelated (cf. the definition of innovation processes in (ROZANOV 19?)).

Theorem 3 is adjusted to the possibility of choosing as f the indicator functions $f(x) = I\{x \leqslant z\}$ 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 sounds as follows: let $\mathcal{G}$ be such that conditions (30 and (31) are fullfilled, then the assertion of Theorem 3 is correct.

One can adopt this formulation and state the following theorem concerning w_n: let $\mathcal{G}' = \{\phi : \phi / f_\theta^{1/2}(\cdot) \in \mathcal{G}\}$ and let for $\mathcal{G}$ conditions (30) and (31) be fulfilled, then in $\mathcal{X}(\mathcal{G}')$

$$w_n \xrightarrow{\mathcal{D}(\mathbb{P}_{n\theta})} w, \quad w_n \xrightarrow{\mathcal{D}(\tilde{\mathbb{P}}_{n\theta})} w + J. \tag{35}$$

This formulation can help in a search for sets $\mathcal{G}$ different from the one described in Theorem 3. But we prefer to formulate our last theorem in the same fashion as Theorem 3.

THEOREM 4. *Suppose $\mathcal{G}'$ is a subset of $L_2[0,1]^m$ such that*
1) $|\phi| < c$ *for the same constant c for all $\phi \in \mathcal{G}'$,*
2) $\Lambda(\mathcal{G}') = L_2[0,1]^m$
3) *in the space $\mathcal{X}(\mathcal{G})$, where $\mathcal{G} = \{f : f = \phi / f_\theta^{1/2}(\cdot), \phi \in \mathcal{G}'\}$ convergence (30) holds.*

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

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

is integrable. Then in $\mathcal{X}(\mathcal{G}')$ (35) holds where w is the standard Wiener process and the shift function J is defined just before Lemma 7. The assertion of Lemma 7 is correct.

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

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

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

According to Theorem 4 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 4 is in fact contained in the proof of Theorem 3 and Lemma 7 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 8). Since $v_n(Zf)$ is a normalized sum of i.i.d. random variables, $Ev_n^2(Zf) = <\phi, \phi>$, 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 ϕ is a consequence of CLT. According to condition 3) and Lemma 9 the sequences $\{v_n(\cdot / f_\theta^{1/2})\}$ and $\{c_n(\pi_{1-\epsilon}(\cdot / f_\theta^{1/2}))\}$ are tight in $\mathcal{X}(\mathcal{G}')$. Consider the difference

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

But

$$\sup_{\phi \in \mathcal{G}'} |c_n(\pi_{1-\epsilon}^\perp \frac{\phi}{f_\theta^{1/2}})| \leqslant c \int_{1-\epsilon}^1 |<f_\theta^{1/2}, d\pi_\lambda q^T > C_\lambda^{-1} v_n(\pi_\lambda^\perp q)| \tag{36}$$

and

$$E\int_{1-\epsilon}^{1}|<f_\theta^{1/2},d\pi_\lambda q^T>C_\lambda^{-1}\nu_n(\pi_\lambda^\perp q)| \leqslant \int_{1-\epsilon}^{t}(E|<f_\theta^{1/2},d\pi_\lambda q^T>C_\lambda^{-1}\nu_n(\pi_\lambda^\perp q)|^2)^{1/2} = \int_{1-\epsilon}^{1}\alpha'_\lambda d\lambda.$$

Hence, the upper bound in the left hand side of (36) can be made arbitrarily small in probability for sufficiently small $\epsilon>0$. This means that the sequence $\{\nu_n(Z(\cdot/f_\theta^{1/2})\}$ is tight in $\mathcal{X}(\mathcal{G}')$. Convergence (35) follows. Since $<\phi,\phi> = (f,f)$ we get: $\Lambda(\mathcal{G}') = L_2[0,1]^m$ iff $\Lambda(\mathcal{G}) = L_2(\theta)$. Hence condition 2) allows to apply Lemma 7 and to conclude the proof. □

REFERENCES

1. P. BICKEL, M. WICHURA (1971). Convergence criteria for multiparameter stochastic processes. *Ann. Math. Statist. 42,* 1656-1670.
2. D. BILLINGSLEY (1968). *Convergence of probability measures.* J. Wiley, New York.
3. R. CAIROLI, J.B. WALSH (1975). Stochastic integrals in plane. *Acta Mathemat., 134, 1-2,* 111-183.
4. J. DOOB (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist. 20, 3,* 393-403.
5. J. DURBIN, M. KNOTT, C.C. TAYLOR (1975). Components of Cramér-von Mises statistics, II. *J. Roy. Statist. Soc. 37; 2,* 216-236.
6. P. GAENSSLER, W. STUTE (1979). Empirical processes: a survey of results for independent and identically distributed random variables. *Ann. Probab. 7,* 193-243.
7. J.I. GIHMAN (1953). Some remarks on A. Kolmogorov's goodness of fit test. *Doklady Acad. Nauk, 91, 4,* 715-718 (Russian).
8. J.I. GIHMAN (1954). On the theory of ω^2-tests. *Math. Zbornik of Kiev State University, 5,* 51-59 (Ukrainian).
9. J.I. GIHMAN (1982). Biparametric martingales. *Uspechi Math. Nauk.* (English transl. — *Russian Math. Surveys) 37, 6,* 215-237.
10. P. GREENWOOD, A.N. SHIRYAYEV (1985). *Contiguity and Statistical Invariance Principle,* Gordon and Breach, London.
11. I.A. IBRAGIMOV, R. HAŠMINSKI (1981). *Asymptotic Theory of Estimation,* Springer Verlag, Berlin.
12. K. ITO, H.R. MCKEAN (1965). *Diffusion Processes and Their Sample Pathes.* Springer Verlag, Berlin.
13. M. KAC, J. KIEFER, S. WOLFOVITZ (1955). On test of goodness of fit based on distance method. *Ann. Math. Statist. 26, 2,* 189-211.
14. E.V. KHMALADZE (1975). On estimation of necessary sample size for testing simple contiguous alternatives. *Teorya Verohatn. Primen. XX, 2,* 115-125 (Russian).
15. E.V. KHMALADZE (1979). On the use of w^2-type tests for testing parametric hypothesis. *Teorya Verojatn. Primen.* (English transl. — *Theory Probab. Appl.*).
16. E.V. KHMALADZE (1981). Martingale approach to the theory of goodness of fit tests. *Teorya Verojatn. Primen.* (English transl.— *Theory Probab. Appl.*).
17. E.V. KHMALADZE (1986). Innovation fields and the problem of testing simple statistical hypothesis in R^m. *Doklady Acad. Nauk* (English transl. — *Soviet Math. Doklady*).
18. A.N. KOLMOGOROV (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell Instituto degli Ital. Attuadi 4, 1,* 83-91:
19. A.N. KOLMOGOROV (1986). *Theory of Probability and Mathematical Statistics,* (Vol. 2 of Selected Papers) Nauka, Moskow (Russian).
20. H.-H. KUO (1975). *Gaussian Measures in Banach Spaces,* Lecture Notes in Math. 463.
21. R.S. LIPZER, A.N. SHIRYAYEV (1977). Statistics of random processes. Springer Verlag, Berlin.
22. G. NEUHAUSS (1971). On weak convergence of stochastic processes with multi-dimensional time parameter. *Ann. Math. Statist. 42, 5,* 1285-1295.
23. A.M. NIKABADZE (1990). Scanning innovations and goodness of fit tests for vector-valued random variables against general alternatives. *Teorya Verojatn. Primen.* (English transl. - *Theory Probab. Appl.*)

24. A.M. NIKABADZE, E.V. KHMALADZE (1987). On goodness-of-fit tests for parametric hypotheses in $\mathbb{R}_m$. *Soviet Math. Dokl. 35, 3,* 627-629.
25. J. OOSTERHOFF, W. VAN ZWET (1979). A note on contiguity and Hellinger distance. *Contributions to Statistics.* (Ed. J. Jurechkova) Dordrecht, Reidel.
26. D. POLLARD (1984). *Convergence of Stochastic Processes,* Springer Verlag, New York.
27. YU. ROZANOV (19?). *Theory of Innovation Processes.* Nauka. Moskou
28. M. ROSENBLATT (1952). Remark on multivariate transformation. *Ann. Math. Statist. 23, 3,* 470-472.
29. P.K. SEN (1981). *Sequential Nonparametrics: Innovation Principles and Statistical Inference,* J. Wiley, New York.
30. G. SHORACK, J. WELLNER (1986). *Empirical Processes with Application to Statistics,* J. Wiley, New York.
31. P. SIMPSON (1951). Note on the estimation of bivariation distribution function. *Ann. Math. Statist. 22, 3,* 476-478.
32. N.V. SMIRNOV (1937). On the distribution of ω^2-tests of Mises. *Math. Sbornik, 2, 1,* 973-993 (Russian).
33. YU.N. TYURIN (1970). On testing parametric hypotheses by nonparametric tests. *Teorya Verojatn. Primen. XV, 4,* 745-749 (Russian).
34. A. WALD, J. WOLFOWITZ (1939). Confidence limits for continuous distribution function. *Ann. Math. Statist. 10, 2,* 105-118.
35. E. WONG, M. ZAKAI (1974). Martingales and stochastic integrals for processes with multidimensional parameters. *Z. Wahrscheinlichkeitstheorie Geb. 291,* 109-122.