The Annals of Statistics
1988, Vol. 16, No. 4, 1503–1516

AN INNOVATION APPROACH TO GOODNESS-OF-FIT TESTS IN R^m

BY E. V. KHMALADZE

Steklov Mathematical Institute and Razmadze Mathematical Institute

We present a solution to the goodness-of-fit problem for mulivariate observations, using the innovation process for the (sequential) empirical distribution function with respect to a conveniently chosen linear ordering or scanning system in R^m.

1. Introduction. The aim of this paper is to introduce in m-dimensional Euclidean space R^m empirical processes, which would, for arbitrary $m < \infty$, play an analogous role to that of the uniform empirical process and the uniform sequential empirical process in R^1.

To be specific let $X_1, \dots, X_n$ be independent random vectors taking values in R^m. Consider the problem of testing a simple hypothesis that these random vectors are identically distributed and the distribution function (d.f.) of each X_i is some specified absolutely continuous distribution function F. Let v_n and z_n be the empirical and the sequential empirical processes, respectively,

$$z_n(s,x) = n^{-1/2} \sum_{i \le sn} [I\{X_i \le x\} - F(x)], \; v_n(x) = z_n(1,x).$$

For $m = 1$ let z_n^0 and v_n^0 denote the corresponding uniform sequential empirical and uniform empirical processes, respectively,

$$z_n^0(s,t) = z_n(s,x), \qquad v_n^0(t) = v_n(x) \quad \text{for } t = F(x). \tag{1.1}$$

It is common knowledge that if the hypothesis holds, i.e., if the distribution of the sequence $X_1, \dots, X_n$ is the direct product $\mathbb{P}_n = F \times \cdots \times F$, then

$$z_n^0 \to_{\mathscr{D}} z^0 \quad \text{and} \quad v_n^0 \to_{\mathscr{D}} v^0$$

in the spaces $D[0,1]^2$ and $D[0,1]$, respectively. Here z^0 and v^0 are Gaussian processes with mean 0 and covariance functions $(s \wedge s')(t \wedge t' - tt')$ and $t \wedge t' - tt'$, respectively. The main point is that the distributions of z^0 and v^0 do not depend on the d.f. F, that is, the transformation (1.1) maps z_n and v_n into asymptotically distribution-free processes.

Since the work of Simpson (1951) and Rosenblatt (1952) it is understood that the process v_n^0 loses its key property if $m \ge 2$—it is no longer asymptotically distribution free [if F is the d.f. of the m-dimensional random vector X and $m \ge 2$, then the d.f. of $U = F(X)$ depends on F even when F is absolutely continuous]. Simpson and Rosenblatt suggested how to avoid the difficulty but the problem is still alive as is demonstrated by the papers of Bickel and Breiman

Received September 1985; revised February 1988.
AMS 1980 *subject classifications*. 62G10, 62F03.
Key words and phrases. Empirical processes, multivariate innovation process, Doob transformation.

(1983) and of Schilling (1983a, b). Bickel and Breiman considered an empirical process based on the m-dimensional analogue of uniform spacings and showed that this process is asymptotically distribution free. Schilling considered some "weighted" modification of this process to make it more sensitive to local alternatives.

The opinion of the present author is that the problem of finding a proper substitute for the uniform sequential empirical process and the uniform empirical process is still open. Let us consider what should be understood as this "proper substitute." Rather let us remark what makes the uniform empirical process v_n^0 and the uniform sequential empirical process z_n^0 important in goodness-of-fit theory.

One important property is, of course, that v_n^0 and z_n^0 are asymptotically distribution free. But that cannot be the only necessary property—for example, processes which are identically constant for all x and n are asymptotically distribution free but useless. Another important property of v_n^0 and z_n^0 is that they are very sensitive to "all" deviations from the hypothesis, that is, to "all" alternatives to F (see the following discussion). To formulate this property precisely, that is, to formulate necessary conditions on the processes we are seeking, let us first describe the class of alternatives formally.

Under an alternative hypothesis it is supposed that the X_i's are independent and that the d.f. of each X_i is A_{in}, $i = 1, \dots, n$, with the following properties:

Let $A_{in} = A_{in}^c + A_{in}^s$ be the Lebesgue decomposition of A_{in} into its continuous (with respect to F) and singular parts. Then

$$\text{(1)} \qquad \text{as } n \to 0, \qquad \sum_{i=1}^{n} \nu(A_{in}^s) \to 0,$$

where $\nu(P)$ denotes the total variation of P, and

(2) for the functions h_n defined by the equality

$$\left[\frac{dA_{in}^c}{dF}(x)\right]^{1/2} = 1 + \frac{1}{2n^{1/2}} h_n(s,x), \qquad \frac{i-1}{n} \le s < \frac{i}{n},$$

there is a function h such that

$$\text{(1.2)} \qquad \begin{gathered} \int_{(s,x)\in[0,1]\times R^m} |h_n(s,x) - h(s,x)|^2\, ds\, F(dx) \to 0, \qquad n \to \infty, \\ \int_{(s,x)\in[0,1]\times R^m} h^2(s,x)\, ds\, F(dx) < \infty. \end{gathered}$$

An important special case of the alternatives considered is that when $A_{1n} = \cdots = A_{nn}$, i.e., when $X_1, \dots, X_n$ are still assumed to be identically distributed and

$$\text{(1.3)} \qquad \left[\frac{dA_n}{dF}(x)\right]^{1/2} = 1 + \frac{1}{2n^{1/2}} h_n(x),$$

where

$$\int [h_n(x) - h(x)]^2 F(dx) \to 0, \qquad \int h^2(x) F(dx) < \infty.$$

Another special case is given by the so-called change-point alternatives when (1.2) is satisfied with $h_n(s,x) = I\{s \geq s_0\} h_n(x)$ for some "change-point" $s_0 \in (0,1)$.

Denote the alternative distributions of the sample $X_1, \dots, X_n$ by $\tilde{\mathbb{P}} = \tilde{\mathbb{P}}_n(h) = A_{1n} \times \cdots \times A_{nn}$ and let us use the notation $\overline{\mathbb{P}}_n$ in the case of (1.3), that is $\overline{\mathbb{P}}_n = \overline{\mathbb{P}}_n(h) = A_n \times \cdots \times A_n$.

It is well known that under conditions (1) and (2) the sequence $\{\tilde{\mathbb{P}}_n\}$ is contiguous to the sequence $\{\mathbb{P}_n\}$ [see Oosterhoff and van Zwet (1975) and also Greenwood and Shiryayev (1985)]. In this sense the alternatives considered here are the "most difficult" to distinguish from the hypothesis. It is also known [cf. Khmaladze (1975)] that under conditions (1) and (2) the following limit exists:

$$\lim_{n \to \infty} \nu\big(\tilde{\mathbb{P}}_n(h) - \mathbb{P}_n\big) = \lambda(h) \tag{1.4}$$

and, in particular,

$$\lim_{n \to \infty} \nu\big(\overline{\mathbb{P}}_n(h) - \mathbb{P}_n\big) = \bar{\lambda}(h), \tag{1.5}$$

where λ and $\bar{\lambda}$ stand for the functional of the functions used in conditions (1.2) and (1.3), respectively (the precise form of λ is simple but we will not need it).

As the last preliminary step recall some known weak convergence results which we will need in the sequel [see Gaenssler and Stute (1979) and Shorack and Wellner (1986) for references]. Let z and v be Gaussian processes with zero mean and covariance functions $(s \wedge s')[F(x \wedge x') - F(x)F(x')]$ and $F(x \wedge x') - F(x)F(x')$, respectively, and in the case $m = 1$ let z^0 and v^0 be the Kiefer field and Brownian bridge, respectively, that is, Gaussian processes with mean and covariance functions $(s \wedge s')(t \wedge t' - tt')$ and $t \wedge t' - tt'$.

Then

$$z_n \to_{\mathscr{D}(\tilde{\mathbb{P}}_n)} z + H, \qquad v_n \to_{\mathscr{D}(\tilde{\mathbb{P}}_n)} v + H(1, \cdot),$$

$$H(s,x) = \int_{(\sigma, y) < (s,x)} h(\sigma, y)\, d\sigma\, F(dy)$$

in $D[0,1]^{m+1}$ and $D[0,1]^m$, respectively, and

$$z_n^0 \to_{\mathscr{D}(\tilde{\mathbb{P}}_n)} z^0 + H^0, \qquad v_n^0 \to_{\mathscr{D}(\tilde{\mathbb{P}}_n)} v + H^0(1, \cdot),$$

$$H^0(s,t) = H(s,x), \qquad t = F(x)$$

in $D[0,1]^2$ and $D[0,1]$, respectively. Under the null hypothesis, i.e., under $\mathbb{P}_n$, these statements hold with H replaced by 0.

Now the following lemma states "sensitivity" properties of z_n^0 and v_n^0. Let P_ξ denote the distribution of a process ξ (in the corresponding functional space).

LEMMA. *The following equalities hold*:

$$\nu(P_z - P_{z+H}) = \nu(P_{z^0} - P_{z^0+H^0}) = \lambda(h) \tag{1.6}$$

and if $h(s, x) = h(x)$,

$$\nu(P_v - P_{v+H(1,\cdot)}) = \nu(P_{v^0} - P_{v^0+H^0(1,\cdot)}) = \bar{\lambda}(h). \tag{1.7}$$

The proof of the lemma is left to the reader.

The second equality in (1.6) shows that in z_n^0 asymptotically nothing is lost that allows one to distinguish between the null hypothesis and the alternatives (1) and (2). The second equality in (1.7) shows the same for the uniform empirical process v_n^0 in the case of alternatives (1.3).

Now we are in a position to formulate the formal mathematical problem we are concerned with in the present paper.

We intend to look for transformations $w[z_n, F]$ and $w[v_n, F]$ of z_n and v_n, which could depend on F [as in (1.1)], and which possess the following properties:

(1°) Under the null hypothesis $\mathbb{P}_n$ the processes $w[z_n, F]$ and $w[v_n, F]$ should have limit distributions, P and Q, respectively, which are independent of F for any absolutely continuous F.

(2°) (a) Under any sequence of alternatives $\tilde{\mathbb{P}}_n(h)$, satisfying conditions (1) and (2), the process $w[z_n, F]$ should have a limit distribution P', and $\nu(P - P') = \lambda(h)$.

(b) Under any sequence of alternatives $\bar{\mathbb{P}}_n(h)$, satisfying conditions (1) and (2), the process $w[v_n, F]$ should have a limit distribution Q', and $\nu(Q - Q') = \bar{\lambda}(h)$.

Condition (2°) means that asymptotically nothing is lost in $w[z_n, F]$ or in $w[v_n, F]$, which would allow us to distinguish between $\mathbb{P}_n$ and $\tilde{\mathbb{P}}_n(h)$ or between $\mathbb{P}_n$ and $\bar{\mathbb{P}}_n(y)$, respectively.

As test statistics one can now use various functionals of the transformed processes $w[z_n, F]$ and $w[v_n, F]$ such as the Kolmogorov–Smirnov or ω^2 (Cramér–von Mises–Smirnov) statistics. The question which particular functional should be used is obviously a separate question and should be treated separately in other work. But for any choice of these functionals, that is, test statistics, it seems reasonable from a practical point of view to place the additional heuristic requirements on the transformations $w[z_n, F]$ and $w[v_n, F]$:

(3°) The limit distributions P and Q of $w[z_n, F]$ and $w[v_n, F]$, respectively, should be "simple enough" to allow calculation of limit distributions of statistics based on these processes.

(4°) The transformations $w[z_n, F]$ and $w[v_n, F]$ themselves should be simple so that the test statistics can be easily calculated.

2. The scanning innovation process: Examples. Call a Gaussian process with mean 0 and covariance function $F(x \wedge x')$ a Wiener process w.r.t. F. Put $\mathbb{1} = \mathbb{1}_{m-1}$ and put $(t, \mathbb{1}) = t$ if $m = 1$. Let $G(s, x) = sF(x)$ and let us assume that $F(t, \mathbb{1}) = t$, though it is only a matter of notational convenience for Theorems 1 and 2.

Consider the processes

$$b(s,t,y) = z(s,t,y) + \int_0^t \frac{z(s,\tau,\mathbf{1})}{1-\tau} F(d\tau, y), \tag{2.1}$$

$$w(t,y) = v(t,y) + \int_0^t \frac{v(\tau,\mathbf{1})}{1-\tau} F(d\tau, y), \tag{2.2}$$

where z and v are the Gaussian processes defined in the Introduction.

THEOREM 1. *The process b is a Wiener process w.r.t. G. The process w is a Wiener process w.r.t. F. The relations* (2.1) *and* (2.2) *between b and z and between w and v are one-to-one.*

For reasons that will be clarified by Example 4, we will call w a scanning innovation process for v.

Theorem 1, particularly the part concerning w, expresses the basic point of this paper. This part relates to certain innovation arguments for the process v, and we find it necessary to clarify its statistical meaning by some examples.

The first two examples show that when $m = 1$ the process w is not new in goodness-of-fit theory. For $m = 1$ formula (2.2) takes the form

$$w(t) = v(t) + \int_0^t \frac{v(\tau)}{1-\tau} \, d\tau, \tag{2.3}$$

which is the well-known Doob–Meyer decomposition of the Brownian bridge v [see, e.g., Liptser and Shiryayev (1977); for some statistical discussion see, e.g., Khmaladze (1981)]. The Wiener process w is the innovation process of v, that is, for any t the random variable $w(t)$ is measurable w.r.t. the σ-algebra $\mathscr{F}(t) = \sigma\{v(\tau),\ \tau \le t\}$ and the inverse of (2.3) is

$$v(t) = (1-t) \int_0^t \frac{1}{1-\tau} w(d\tau). \tag{2.4}$$

EXAMPLE 1. One of the basic purposes of Doob (1949) was to show that the limiting d.f. of the Kolmogorov test is nothing more than the d.f. of $\sup_t |v(t)|$. Doob's approach was to observe that

$$P\Big\{\sup_t |v(t)| < \lambda\Big\} = P\bigg\{\forall\, t \in [0,1]: \left|\frac{v(t)}{1-t}\right| < \frac{\lambda}{1-t}\bigg\},$$

and then to calculate the probability appearing on the right-hand side. This was convenient because in contrast to v the process $v(t)/(1-t)$, $t \in [0,1]$, is Gaussian with *independent increments*. But just this is properly explained by the representation (2.4) of v by its innovation process w—it is clear that the integral

$$\int_0^t \frac{1}{1-\tau} w(d\tau)$$

is a Wiener process w.r.t. θ, where

$$\theta(t) = \int_0^t \frac{1}{(1-\tau)^2}\, d\tau = \frac{t}{1-t}.$$

Rényi (1953) introduced a goodness-of-fit test based on the statistic

$$R_n(\varepsilon) = \sup_{t \le 1-\varepsilon} \frac{v_n(t)}{1-t}.$$

Under the null hypothesis the limit distribution of $R_n(\varepsilon)$ for fixed ε is that of

$$R(\varepsilon) = \sup_{t \le 1-\varepsilon} \frac{v(t)}{1-t}$$

and according to (2.4), $R(\varepsilon)$ is the supremum over $[0, 1-\varepsilon]$ of a Wiener process w.r.t. θ. Therefore the d.f. of $R(\varepsilon)$ is $2\Phi(x/\sigma) - 1$, $x \ge 0$, where $\sigma^2 = \theta(1-\varepsilon)$ and Φ is a standard normal d.f., as obtained by Rényi.

EXAMPLE 2. Let $0 = t_0 < t_1 \cdots < t_N = 1$ be some partition of $[0,1]$ and consider the Gaussian vector $\{\Delta v(t_j)\}$ of increments $\Delta v(t_j) = v(t_{j+1}) - v(t_j)$. Associate with this vector an increasing sequence $\{\mathscr{F}_N(t_j)\}$ of σ-algebras $\mathscr{F}_N(t_j) = \sigma\{v(t_1), \dots, v(t_j)\}$. Consider the vector $\{\Delta w(t_j)\}$, where

$$\begin{aligned} \Delta w(t_j) &= \Delta v(t_j) - E\left[\Delta v(t_j) | \mathscr{F}_N(t_j)\right] \\ &= \Delta v(t_j) + \frac{v(t_j)}{1-t_j}\Delta t_j. \end{aligned} \tag{2.5}$$

In contrast to $\{\Delta v(t_j)\}$ the Gaussian vector $\{\Delta w(t_j)\}$ has independent coordinates, and (2.5) is a discrete time analogue of (2.3). Define increments $\Delta w_n(t_j)$ using (2.5) with $\Delta v(t_j)$ replaced by $\Delta v_n(t_j)$, the increments of the empirical process v_n. After a simple rearrangement one can easily verify that

$$X_{N,n}^2 = \sum_{j=0}^{N-1} \frac{\left[\Delta w_n(t_j)\right]^2}{E\left[\Delta w_n(t_j)\right]^2}$$

coincides with the classical Pearson χ^2 statistic.

The next example shows that some care must be exercised in extending (2.3) to the multidimensional case.

For $x \le y$ let $[x, y)$ be the rectangle $\{x' \colon x \le x' < y\}$. For simplicity, let $m = 2$ and let $0 = t_0 < t_1 < \cdots < t_N = 1$, $0 = u_0 < u_1 < \cdots < u_N = 1$ be partitions of $[0,1]$. Consider the partition of $[0,1]^2$ by the rectangles $[x_{ij}, x_{i+1,j+1})$, where $x_{ij} = (t_i, u_j)$. Let $\Delta v(x_{ij})$ be the increment of v on $[x_{ij}, x_{i+1,j+1})$, i.e.,

$$\Delta v(x_{ij}) = v(x_{i+1,j+1}) - v(x_{i+1,j}) - v(x_{i,j+1}) + v(x_{ij})$$

and denote by $\Delta F(x_{ij})$ a similar increment of the d.f. F. As opposed to the one-dimensional case there are several natural choices of increasing families of σ-algebras, which one can associate with $\{\Delta v(x_{ij})\}$.

EXAMPLE 3. Put $\mathscr{F}_N(x_{ij}) = \sigma\{v(x_{lm}),\ x_{lm} \le x_{ij}\}$. Obviously, $\mathscr{F}_N = \{\mathscr{F}_N(x_{ij})\}$ is an increasing but not linearly ordered family of σ-algebras. Because of this the increments $\{\Delta M^1(x_{ij})\}$, where

$$\begin{aligned}\Delta M^1(x_{ij}) &= \Delta v(x_{ij}) - E\left[\Delta v(x_{ij}) | \mathscr{F}_N(x_{ij})\right] \\ &= \Delta v(x_{ij}) + \frac{v(x_{ij})}{1 - F(x_{ij})} \Delta F(x_{ij}),\end{aligned} \tag{2.6}$$

are not independent random variables, in contrast to the $m = 1$ case. Consequently, the simple equality (2.6) is not the proper analogue of (2.3), which we seek.

Consider another natural family of σ-algebras $\mathscr{H}_N = \{\mathscr{H}_N(x_{ij})\}$, where $\mathscr{H}_N(x_{ij}) = \mathscr{F}_N(t_i, 1) \vee \mathscr{F}_N(1, u_j)$. The increments $\{\Delta M^2(x_{ij})\}$, where

$$\begin{aligned}\Delta M^2(x_{ij}) &= \Delta v(x_{ij}) - E\left[\Delta v(x_{ij}) | \mathscr{H}_N(x_{ij})\right] \\ &= \Delta v(x_{ij}) - \frac{v(1,1) - v(1, u_j) - v(t_i, 1) + v(t_i, u_j)}{1 - F(1, u_j) - F(t_i, 1) + F(t_i, u_j)} \Delta F(x_{ij}),\end{aligned} \tag{2.7}$$

are also not independent random variables and therefore (2.7) is still not what is needed.

REMARK. For readers familiar with the theory of martingales in two-dimensional time [see Cairoli and Walsh (1975) and Wong and Zakai (1976); see also Gihman (1982) for references] Example 3 shows that if $\mathscr{F} = \{\mathscr{F}(x)\}$, $\mathscr{F}(x) = \sigma\{v(y),\ y \le x\}$ and

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

then the process $\{M^1, \mathscr{F}\}$ is only a weak martingale, and not a strong martingale. The process $\{M^2, \mathscr{F}\}$,

$$M^2(x) = v(x) + \int_{y \le x} \frac{\Delta_{(1,1)} v(y)}{\Delta_{(1,1)} F(y)} F(dy),$$

where, say, $\Delta_{(1,1)} v(y)$ is an increment of v on $[y, (1,1))$, is also not a strong martingale.

The last example explains the nature of (2.2).

EXAMPLE 4. Consider the σ-algebras

$$\mathscr{C}_N(x_{ij}) = \sigma\left\{\Delta v(x_{mj}),\ m \le i - 1\right\},$$
$$\mathscr{J}_N(x_{ij}) = \mathscr{F}_N(x_{Nj}) \vee \mathscr{C}_N(x_{ij}).$$

In contrast to $\mathscr{F}_N$ and $\mathscr{H}_N$, the family $\mathscr{J}_N = \{\mathscr{J}_N(x_{ij})\}$ is not only increasing but also linearly ordered—for any two x_{ij} and x_{lm} either $\mathscr{J}(x_{ij}) \subseteq \mathscr{J}(x_{lm})$ or

$\mathscr{I}(x_{lm}) \subseteq \mathscr{I}(x_{ij})$. This implies that the increments $\{\Delta w(x_{ij})\}$, where

$$\begin{aligned}\Delta w(x_{ij}) &= \Delta v(x_{ij}) - E\left[\Delta v(x_{ij}) | \mathscr{I}_N(x_{ij})\right] \\ &= \Delta v(x_{ij}) + \frac{v(t_i,1) + v(x_{i,j+1}) - v(x_{ij})}{1 - F(t_i,1) + F(x_{i,j+1}) - F(x_{ij})} \Delta F(x_{ij})\end{aligned} \tag{2.8}$$

are independent Gaussian random variables. Equality (2.2) is nothing more than a continuous time version of (2.8), and Theorem 1 states that the σ-algebras $\mathscr{C}_N(x_{ij})$ can be neglected as N increases.

Let $\{\Delta w_n(x_{ij})\}$ be the increments obtained by replacing v by v_n in (2.8). Then the process

$$w_{N,n}^*(x) = \sum_{i,j=1}^{N} \frac{\Delta w_n(x_{ij})}{\left(E\left[\Delta w_n(x_{ij})\right]^2\right)^{1/2}} I\{x_{ij} < x\}$$

is a discrete time analogue of the process w_n^* in Theorem 3, and

$$X_{N,n}^2 = \sum_{i,j=1}^{N} \frac{\left[\Delta w_n(x_{ij})\right]^2}{E\left[\Delta w_n(x_{ij})\right]^2}$$

is again the classical χ^2 statistic.

PROOF OF THEOREM 1. Since z is a Gaussian process with mean 0 and b is a linear transformation of z, the process b is also Gaussian with mean 0. By direct calculation the covariance function of b can be shown to be $(s \wedge s')F(x \wedge x')$. Therefore b is a Wiener process w.r.t. G, and, consequently, w is a Wiener process w.r.t. F. It can easily be shown that

$$z(s,t,y) = b(s,t,y) - \int_0^t \int_0^\tau \frac{1}{1-u} b(s,du,\mathbb{I}) F(d\tau, y)$$

is the inverse of (2.1) and all that remains is an argument showing that this inverse is unique. But the equation

$$0 = \phi(t,y) + \int_0^t \frac{\phi(\tau,\mathbb{I})}{1-\tau} F(d\tau, y) \tag{2.9}$$

has the unique solution $\phi = 0$. Indeed, if we put $y = \mathbb{I}$ we get an equation for $\phi(\cdot,\mathbb{I})$, which obviously has the unique solution $\phi(\cdot,\mathbb{I}) = 0$. Therefore the integral term in (2.9) is 0 and this implies that $\phi = 0$. □

3. Convergence in distribution. Define the process b_n, a sequential empirical scanning process, and w_n, an empirical scanning process, by

$$\begin{aligned}b_n(s,t,y) &= z_n(s,t,y) + \int_0^t \frac{z_n(s,\tau,\mathbb{I})}{1-\tau} F(d\tau, y), \\ w_n(t,y) &= v_n(t,y) + \int_0^t \frac{v_n(\tau,\mathbb{I})}{1-\tau} F(d\tau, y).\end{aligned} \tag{3.1}$$

THEOREM 2. *Let b and w be Wiener processes w.r.t. G and F, respectively. Then as $n \to \infty$,*

$$b_n \to_{\mathscr{D}(\mathrm{P}_n)} b, \qquad w_n \to_{\mathscr{D}(\mathrm{P}_n)} w$$

in the spaces $D[0,1]^{m+1}$ and $D[0,1]^m$, respectively. Under any sequence of alternatives $\tilde{\mathrm{P}}_n(h)$, $n = 1, 2, \dots,$ satisfying (1) and (2),

$$b_n \to_{\mathscr{D}(\tilde{\mathrm{P}}_n)} b + \mu, \qquad w_n \to_{\mathscr{D}(\tilde{\mathrm{P}}_n)} w + \mu(1, \cdot),$$

where the shift function μ is

$$\mu(s, t, y) = H(s, t, y) + \int_0^t \frac{H(s, \tau, \mathbb{1})}{1 - \tau} F(d\tau, y)$$

and H is as defined in the Introduction.

REMARK. The process b_n (and w_n) can be written in the simple form

$$b_n(s, x) = n^{-1/2} \sum_{i \le sn} \left[I\{X_i \le x\} - \int_0^t \frac{1 - I\{X_i \le (\tau, \mathbb{1})\}}{1 - \tau} F(d\tau, y) \right], \quad x = (t, y), \tag{3.2}$$

$$w_n(x) = b_n(1, x).$$

By the way, this allows us to view (3.2) as Doob–Meyer decomposition of a multiparameter point process

$$\xi_n(s, x) = \sum_{i \le sn} I\{X_i \le x\}.$$

PROOF OF THEOREM 2. The mapping $\phi(s, x) \to \phi(1, x)$ from $D[0,1]^{m+1}$ to $D[0,1]^m$ is continuous and hence the statement for w_n follows from that for b_n. Consider b_n. The mapping $\phi(s, t, y) \to \phi(s, t, \mathbb{1})$ from $D[0,1]^{m+1}$ to $D[0,1]^2$ is continuous and the convergence of ϕ_n to continuous ϕ in $D[0,1]^{m+1}$ implies convergence of $\phi_n(s, t, \mathbb{1})$ to $\phi(s, t, \mathbb{1})$ for all s and t. Besides,

$$\sup_{s,t} |\phi(s, t, \mathbb{1})| < \infty$$

and

$$\sup_{s,t} |\phi_n(s, t, \mathbb{1})| \to \sup_{s,t} |\phi(s, t, \mathbb{1})|$$

as $n \to \infty$. Hence for any $T < 1$ the operator

$$K_T\phi(s, t, y) = \phi(s, t, y) + \int_0^{t \wedge T} \frac{\phi(s, \tau, \mathbb{1})}{1 - \tau} F(d\tau, y)$$

is continuous in a neighborhood of any continuous $\phi \in D[0,1]^{m+1}$. Since almost all paths of $z + H$ are continuous and $z_n \to_{\mathscr{D}(\tilde{\mathrm{P}}_n)} z + H$ we have that

$$K_T z_n \to_{\mathscr{D}(\tilde{\mathrm{P}}_n)} K_T(z + H) \tag{3.3}$$

in $D[0,1]^{m+1}$. Now we prove that, under the hypothesis,

$$P\Big\{\sup_{s,x}|K_T z_n(s,x) - K_1 z_n(s,x)| > \varepsilon\Big\} \to 0, \qquad T \to 1, \tag{3.4}$$

uniformly in n and

$$P\Big\{\sup_{s,x}|K_T z(s,x) - K_1 z(s,x)| > \varepsilon\Big\} \to 0, \qquad T \to 1. \tag{3.5}$$

But

$$\begin{aligned} |K_T z_n - K_1 z_n| &\le \sup_s \int_T^1 \frac{|z_n(s,\tau,\mathbb{1})|}{1-\tau} F(d\tau, y) \\ &\le \sup_s \int_T^1 \frac{|z_n(s,\tau,\mathbb{1})}{1-\tau}\, d\tau, \end{aligned} \tag{3.6}$$

where the last inequality is true because $\tau - F(\tau, y) = F(\tau, \mathbb{1}) - F(\tau, y) \ge 0$. Since the process $z_n(s, \tau, \mathbb{1})$ for each τ is a process with independent increments the process $|z_n(s, \tau, \mathbb{1})|$ is a submartingale in s. Therefore the integral on the right-hand side of (3.6) is a submartingale, and hence

$$\begin{aligned} &P\Big\{\sup_{0\le s\le 1} \int_T^1 \frac{|z_n(s,\tau,\mathbb{1})|}{1-\tau}\, d\tau > \varepsilon\Big\} \\ &\qquad \le \frac{1}{\varepsilon} E\int_T^1 \frac{|v_n(\tau,\mathbb{1})|}{1-\tau}\, d\tau \le \frac{1}{\varepsilon} 2(1-T)^{1/2}. \end{aligned} \tag{3.7}$$

This proves (3.4). Since for all τ the process $z(s, \tau, \mathbb{1})$ in s is a process with independent increments we can use the inequalities (3.7) and (3.6) with z_n replaced by z to prove (3.5). But now since our sequence of alternatives $\{\tilde{\mathbb{P}}_n\}$ is contiguous to $\{\mathbb{P}_n\}$, (3.4) holds under the alternatives too. Since the distribution of $z + H$ is absolutely continuous w.r.t. the distribution of z the relation (3.5) holds for $z + H$ as well. An application of Theorem 4.2 in Chapter 1 of Billingsley (1968) finishes the proof. □

The processes b_n and w_n satisfy conditions (2°), (3°) and (4°) but not yet condition (1°). What is needed is a transformation of a Wiener process w.r.t. F to a standard Wiener process. For $m = 1$ we have the simple transformation $w^0(t) = w(x)$, $t = F(x)$. For $m \ge 2$ one can use the following lemma.

Let

$$f^{(-1/2)} = \begin{cases} 0, & f = 0, \infty, \\ f^{(-1/2)}, & 0 < f < \infty. \end{cases}$$

LEMMA. *Let w be a Wiener process w.r.t. the absolutely continuous d.f. F and let $A \subseteq [0,1]^m$ be a support of the density f of F. Then*

$$w^*(x) = \int_{y \le x} f^{(-1/2)}(y) w(dy)$$

is a Wiener process w.r.t. the uniform d.f. on A. In particular, if $A = [0,1]^m$, then w^* is a standard Wiener process.

For a proof of the lemma simply calculate the covariance function of the Gaussian process w^* by applying the equality

$$E\int_{y\le x} g(y)w(dy)\int_{y\le x'} g(y)w(dy) = \int_{y\le x\wedge x'} g^2(y)F(dy),$$

which holds for any function g which is square integrable w.r.t. F.

Now consider the transformations

$$(3.8)\qquad \begin{aligned} w[z_n, F](s,x) &= b_n^*(s,x) = \int_{y\le x} f^{(-1/2)}(y)b_n(s,dy),\\ w[v_n, F](s,x) &= w_n^*(x) = \int_{y\le x} f^{(-1/2)}(y)w_n(dy). \end{aligned}$$

THEOREM 3. *Let b^* and w^* denote the standard Wiener process on $[0,1]^{m+1}$ and $[0,1]^m$, respectively. Suppose that the support A of the density f in (3.8) is the whole of $[0,1]^m$. Then*

$$b_n^* \to_{\mathscr{D}(\mathbb{P}_n)} b^*, \qquad w_n^* \to_{\mathscr{D}(\mathbb{P}_n)} w^*$$

in the spaces $D[0,1]^{m+1}$ and $D[0,1]^m$, respectively. Under any sequence of alternatives $\{\tilde{\mathbb{P}}_n(h)\}$,

$$b_n^* \to_{\mathscr{D}(\tilde{\mathbb{P}}_n)} b^* + \mu^*, \qquad w_n^* \to_{\mathscr{D}(\tilde{\mathbb{P}}_n)} w^* + \mu^*(1,\cdot),$$

where the function μ^* is given by

$$\mu^*(s,x) = \int_{y\le x} f^{(-1/2)}(y)\mu(s,dy)$$

and μ is defined as in Theorem 2.

PROOF. It is sufficient to prove the convergence of b_n^* (cf. the proof of Theorem 2). Let

$$\mathscr{L}_T\phi(s,t,y) = \int_{(\tau, y')\le(t\wedge T, y)} \phi(s,\tau,\mathbb{1})\frac{f^{1/2}(\tau,y')}{1-\tau}\,d\tau\,dy'$$

and

$$z_n^*(s,x) = \int_{y\le x} f^{(-1/2)}(y)z_n(s,dy), \qquad z(s,x) = \int_{y\le x} f^{(-1/2)}(y)z(s,dy).$$

Then

$$b_n^* = z_n^* + \mathscr{L}_1 z_n.$$

The proof will follow from the next four statements.

(1) $z_n^* \to_{\mathscr{D}(\tilde{\mathbb{P}}_n)} z^* + H^*$, $H^*(s,x) = \int_{y\le x} f^{(-1/2)}(y)H(s,dy)$.

(2) $\{z_n^*, z_n\} \to_{\mathscr{D}(\tilde{\mathbb{P}}_n)} \{z^* + H^*, z + H\}$ in the space $D[0,1]^{m+1}\times D[0,1]^{m+1}$.

(3) For any $T<1$ the operator $\mathscr{L}_T$ is continuous in a neighborhood of any continuous function.

(4) Under the hypothesis,

$$P\Big\{\sup_{s,x}|\mathscr{L}_T z_n(s,x) - \mathscr{L}_1 z_n(s,x)| > \varepsilon\Big\} \to 0, \qquad T\to 1,$$

uniformly in n, and

$$P\Big\{\sup_{s,x}|\mathscr{L}_T z(s,x) - \mathscr{L}_1 z(s,x)| > \varepsilon\Big\} \to 0, \qquad T \to 1.$$

Indeed since almost all paths of $z + H$ are continuous it follows from (2) and (3) that for all $T < 1$,

$$z_n^* + \mathscr{L}_T z_n \to_{\mathscr{D}(\tilde{\mathbf{P}}_n)} z^* + H^* + \mathscr{L}_T(z + H).$$

From (4) [cf. the reasoning immediately after (3.8)] it follows that

$$z_n^* + \mathscr{L}_1 z_n \to_{\mathscr{D}(\tilde{\mathbf{P}}_n)} z^* + H^* + \mathscr{L}_1(z + H) = b^* + \mu^*.$$

Therefore we need only prove statements (1)–(4). But z_n^* is a sum of independent random functions

$$z_n^*(s,x) = n^{-1/2} \sum_{i \le sn} \left[f^{(-1/2)}(X_i) I\{X_i \le x\} - U(x)\right],$$

where U denotes a uniform d.f. on $[0,1]^m$. Thus the proof of (1) follows the same line as that of $z_n \to_{\mathscr{D}(\tilde{\mathbf{P}}_n)} z + H$, so there is no need to repeat it here. To prove (2), recall that $z_n^*(s,x)$ and $z_n(s,x)$ for all s and x are sums of independent random variables. Consequently, the convergence of the finite-dimensional distributions of $\{z_n^*, z_n\}$ easily follows from the central limit theorem. Since z_n^* and z_n separately are convergent in distribution the sequence of distributions of the pair $\{z_n^*, z_n\}$ is tight in $D[0,1]^{m+1} \times D[0,1]^{m+1}$ and this finishes the proof of (2). Statement (3) holds because the mapping $\phi(s,t,y) \to \phi(s,t,\mathbf{1})$ is continuous and the function $f^{1/2}(t,y)/(1-t)$ is integrable on $[0,T] \times [0,1]^{m-1}$. To prove (4), observe that

$$\begin{aligned}(3.9)\qquad &\sup_{s,x}|\mathscr{L}_T z_n(s,x) - \mathscr{L}_1 z_n(s,x)| \\ &\quad\le \sup_s \int_T^1 \frac{|z_n(s,\tau,\mathbf{1})|}{1-\tau} g(\tau)\,d\tau \le \sup_s \int_T^1 \frac{|z_n(s,\tau,\mathbf{1})|}{1-\tau}\,d\tau,\end{aligned}$$

where the function g is given by

$$g(\tau) = \int_{y \le \mathbf{1}} f^{1/2}(\tau,y)\,dy \le \left(\int_{y \le \mathbf{1}} f(\tau,y)\,dy\right)^{1/2} \le 1.$$

An application of (3.7) establishes the first relation in (4). The second relation can be proved in the same way. □

For computational convenience the processes b_n^* and w_n^* can be rewritten in the simple form (cf. the remark following Theorem 2)

$$\begin{aligned} b_n^*(s,n) &= n^{-1/2} \sum_{i \le sn} \Bigg[f^{(-1/2)}(X_i) I\{X_i \le x\} \\ &\qquad - \int_{(\tau,y) \le x} \frac{1 - I\{X_i \le (\tau,\mathbf{1})\}}{1-\tau} f^{1/2}(\tau,y)\,d\tau\,dy \Bigg], \\ w_n^*(x) &= b_n^*(1,x). \end{aligned}$$

4. Concluding remarks. According to Theorem 3 the transformations (3.8) satisfy condition (1°). According to Theorem 1 the transformations (3.8) are one-to-one and hence condition (2°) is satisfied. One can say that these transformations satisfy conditions (3°) and (4°) too. Regarding, in particular, the use of Kolmogorov–Smirnov and ω^2 statistics, an approximation for the probability

$$P\left\{\sup_{x \leq \mathbb{1}_m} |w^*(x)| > \lambda\right\},$$

when λ is large can be found in Piterbarg and Fatalov (1983). The distribution

$$P\left\{\int_{x \leq \mathbb{1}_m} [w^*(x)]^2 U(dx) < \lambda\right\}$$

can be easily calculated as described in Martinov (1978). However although we *have* found a solution to the problem we originally posed, it is clear that it is not the unique solution. In particular, the scanning process defined by scanning row-wise as before leads to a different transformation and hence to different test statistics from when one scans column-wise. Put another way, we have an unpleasant dependence on the choice of first, second,... coordinate of our vector observations X_i. We are currently looking at other solutions, e.g., scanning in concentric and increasing circles or ellipses. It will be important that, for instance, a system of ellipses can be determined by the data through preliminary estimates of multivariate location and dispersion. This is similar to using estimated rather then given cell boundaries in a χ^2 test.

In practical situations, goodness-of-fit problems nearly always involve estimated parameters, i.e., a composite null hypothesis, whereas we have only dealt with the simple null hypothesis case. For the case $m = 1$ we already have a solution to this problem in Khmaladze (1981), and it is now quite easy to combine this approach with the present one, since both are based on innovation processes. This synthesis will be the topic of a future paper.

Acknowledgment. The present revision of an earlier paper was written while I was a guest at CWI (the Centre for Mathematics and Computer Science), Amsterdam. I am especially thankful to Professor R. D. Gill whose help made the manuscript truly better and easier to read.

REFERENCES

BICKEL, P. J. and BREIMAN, L. (1983). Sums of functions of nearest neighbour distance, moment bounds, limit theorems and a goodness of fit test. *Ann. Probab.* **11** 185–214.

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

CAIROLI, R. and WALSH, J. (1975). Stochastic integrals in the plane. *Acta Math.* **134** 393–403.

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

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

GIHMAN, J. (1982). Biparametric martingales. *Russian Math. Surveys* **37** (6) 1–29.

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

KHMALADZE, E. V. (1975). An estimation of necessary sample size for testing simple hypothesis. *Theory Probab. Appl.* **20** 115–125.

KHMALADZE, E. V. (1981). Martingale approach to the goodness of fit tests. *Theory Probab. Appl.* **26** 246–265.

KOLMOGOROV, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell Instituto degli Attuari* **4** 83–91.

LIPTSER, R. S. and SHIRYAYEV, A. N. (1977). *Statistics of Random Processes I, General Theory.* Springer, Berlin.

MARTINOV, G. V. (1978). *Omega-square Tests.* Nauka, Moscow.

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

PITERBARG, V. I. and FATALOV, V. R. (1983). Exact asymptotics for probabilities of large deviations for certain Gaussian fields of use in statistics. In *Probabilistic Methods of Research* (A. N. Kolmogorov, ed.). Moscow Univ. Press.

RÉNYI, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hungar.* **4** 191–232.

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

SCHILLING, M. E. (1983a). Goodness of fit testing based on the weighted empirical distribution of certain nearest neighbour statistics. *Ann. Statist.* **11** 1–12.

SCHILLING, M. E. (1983b). An infinite-dimensional approximation to nearest neighbour goodness of fit tests. *Ann. Statist.* **11** 13–24.

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

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

WONG, E. and ZAKAI, M. (1976). Weak martingales and stochastic integrals in the plane. *Ann. Probab.* **4** 570–586.

STEKLOV MATHEMATICAL INSTITUTE
VAVILOVA 42
117966 GSP-1 MOSCOW
USSR