The Annals of Mathematical Statistics
1969, Vol. 40, No. 2, 517–526

AN APPLICATION OF THE SOBOLEV IMBEDDING THEOREMS TO CRITERIA FOR THE CONTINUITY OF PROCESSES WITH A VECTOR PARAMETER[1]

BY HAROLD J. KUSHNER

1. Introduction. Let $f(x)$ be a real valued random process with parameter x, and whose parameter set $\bar{R}$ is a set in E^n (Euclidean n-space). Following the usual terminology, a version of $f(x)$ is any process $\tilde{f}(x)$, defined for $x \varepsilon \bar{R}$, which satisfies $P\{f(x) = \tilde{f}(x)\} = 1$ for $x \varepsilon \bar{R}$. It is sometimes useful to know that there exists a version of $f(x)$ which is wp 1. (with probability one) continuous, Holder continuous, or perhaps differentiable in some component on $\bar{R}$. In the sequel, we give some criteria for these properties. The criteria are somewhat analogous to the criteria, depending on integrability of certain 'weak' derivatives, for the continuity of a sure function (see Smirnov [4], Sec. 114–118). The work was motivated by some problems in stochastic control theory of which one is very briefly discussed in the example of Section 5.

The results involve notions of separability and measurability for vector parameter processes. The applicability of Doobs arguments concerning the existence of separable or measurable versions was noted in Doob [1] (his remark preceding Lemma 2.1, Chapter 2). Although the exact form of the required results does not appear to have been stated, the proof of our Theorem 1 is almost identical in form to that of Neveu [3], p. 91–92.

2. Separability and measurability. N, N_i or $N(y)$ denote null ω-sets, where ω is the generic variable of the sample space. Let R be a bounded open set with closure $\bar{R}$. Let $f(x)$ be a family of real random variables whose parameter x is defined on some domain $\bar{R} \subset E^n$ (Euclidean n-space). Let A be an arbitrary open set in E^n with closure $\bar{A}$ and write $f(A) = \bigcup_{x \varepsilon A} f(x)$. If there is an N and a dense (in $\bar{R}$) denumerable set $\mathfrak{J} \subset \bar{R}$ so that, for each $x \varepsilon \bar{R}$ and each $\omega \not\varepsilon N$,

$$f(x) \ \varepsilon \bigcap_{x \varepsilon A} \overline{f(\mathfrak{J} \cap A)}$$

where the intersection is taken over all open A containing x, then the process $f(\cdot)$ is said to be *separable*, with *separability set* $\mathfrak{J}$.

This (vector parameter) definition is a natural analog of the scalar parameter definition of separability of Neveu [3]. If $f(\cdot)$ is separable then (the analog of the definition of separability of Doob [1]) there is some null set N so that for any compact real interval Γ, and any open set A, the ω sets

$$\{f(x) \ \varepsilon \ \Gamma, x \ \varepsilon \ A \cap \bar{R}\}, \qquad \{f(x) \ \varepsilon \ \Gamma, x \ \varepsilon \ A \cap \bar{R} \cap \mathfrak{J}\}$$

Received 21 September 1968; revised 12 July 1968.
[1] This research has been supported in part by the National Science Foundation under Contract No. GF-289 and in part by the National Science Foundation under Contract No. GK-967.

differ by at most a subset of N. Furthermore, separability implies that, for $\omega \not\varepsilon N$, and any open set A,

$$\sup_{\mathfrak{J}\cap\bar{R}\cap A} f(x) = \sup_{\bar{R}\cap A} f(x) \qquad \inf_{\mathfrak{J}\cap\bar{R}\cap A} f(x) = \inf_{\bar{R}\cap A} f(x).$$

Also, if $f(\cdot)$ is continuous[2] on $\bar{R} \cap \mathfrak{J}$, for $\omega \not\varepsilon N_1$, then $f(\cdot)$ is continuous on $\bar{R}$, for $\omega \not\varepsilon N_1 \cup N$.

It is not hard to show that every process with parameter set R has a separable version. The proofs are almost identical to those in the scalar parameter case. The rest of the paper is concerned only with processes which are continuous in probability and only the proof in this simpler case will be given.

THEOREM 1. *Let $\bar{R}$ be the closure of a bounded open Borel measurable set. Let $f(\cdot)$ be continuous in probability*[3] *on $\bar{R}$. Then there is a version of $f(\cdot)$ which is separable and measurable (as a function of (x, ω)). Any set $\mathfrak{J}$ dense in $\bar{R}$ is a separability set.*

PROOF. Write $x = (x(1), \cdots, x(n))$, where the $x(i)$ are the scalar components of x. By continuity in probability, for each integer $m \geqq 1$, there is some $0 \leqq \Delta_m \leqq 2^{-m}$ so that

$$\sup_i \sup_{|x(i)-x'(i)|\leqq\Delta_m} P\{|f(x) - f(x')| \geqq 1/m\} \leqq 2^{-m}.$$

Suppose that $\bar{R}$ is contained in the rectangle $S = \{x: -\infty < l \leqq x(i) \leqq u < \infty, i = 1, \cdots, n\}$. For each m, let the points $l \equiv x(i; m, 0) < \cdots < x(i; m, k) < x(i; m, k+1) < \cdots < x(i; m, M_m) \equiv u$ divide the interval $[l, u]$ into a finite number $(M_m - 1)$ of segments, with $|x(i; m, k+1) - x(i; m, k)| \leqq \Delta_m$, for each $k = 0, \cdots, M_m - 1$. Define the half open rectangles

$$I^m(k_1^m, \cdots, k_n^m) = \{x: x(i; m, k_i^m) \leqq x(i) < x(i; m, k_i^m + 1), i = 1, \cdots, n\},$$

where the k_i^m ranges over $0, 1, \cdots, M_m - 1$, for each fixed m.

Define the 'lower left' vertex $x[k_1^m, \cdots, k_n^m]$ with components $x(i, m, k_i^m)$, $i = 1, \cdots, n$. Define $\mathfrak{J}_m = [\bigcup_{k_1^m,\cdots,k_n^m} x[k_1^m, \cdots, k_n^m]] \cap \bar{R}$ and suppose that $\mathfrak{J}_{m+1} \supset \mathfrak{J}_m$. Define the function $f^m(\cdot)$ on $\bar{R}$ as follows: If $x[k_1^m, \cdots, k_n^m] \varepsilon \bar{R}$, set $f^m(x)$ equal to $f(x[k_1^m, \cdots, k_n^m])$ for x in $I^m(k_1^m, \cdots, k_n^m)$. Otherwise, if $I^m(k_1^m, \cdots, k_n^m) \cap \bar{R}$ is not empty, let $f^m(x)$ for x in $I^m(k_1^m, \cdots, k_n^m)$ equal the value of $f(x)$ at any point in the set. Write $\hat{f}(\cdot) = \lim_m \sup f^m(\cdot)$. $\hat{f}(\cdot)$ is measurable, as a function of (x, ω). Also $P\{|f(x) - f^m(x)| \geqq 1/m\} \leqq 2^{-m}$, for $x \varepsilon \bar{R}$, implies that $f^m(x) \rightarrow f(x)$ wp 1. for each fixed x. Hence $\hat{f}(\cdot)$ is a version of $f(\cdot)$. Now write

$$\mathfrak{J} = \bigcup_m \mathfrak{J}_m$$

Since $\Delta_m \rightarrow 0$, $\mathfrak{J}$ is dense in $\bar{R}$. Define $\tilde{f}(x) = \hat{f}(x)$ for x in R (the interior of $\bar{R}$). For $x \varepsilon \mathfrak{J} \cap R$, $f(x) = \hat{f}(x) = \tilde{f}(x)$ since $\mathfrak{J}_{m+1} \supset \mathfrak{J}_m$. The value of $\tilde{f}(\cdot)$ on ∂R (the boundary of R) does not effect measurability since ∂R has Lebesgue measure zero. Define $\tilde{f}(x)$ for $x \varepsilon \partial R$, as follows: for each $x \varepsilon \partial R$, there is a sequence $x_m \varepsilon \mathfrak{J} \cap R$

[2] By continuity on $\bar{R} \cap \mathfrak{J}$, we mean continuity with respect to the topology on $\bar{R} \cap \mathfrak{J}$ induced by that on E^n.

[3] $P\{|f(x) - f(x')| > \epsilon\} \rightarrow 0$ as $|x - x'| \rightarrow 0$, where $|\ |$ is the Euclidean norm.

so that $P\{|f(x) - f(x_m)| \geqq 1/m\} < 2^{-m}$. Let $\tilde{f}(x) = \lim_m \sup f(x_m) = \lim_m \sup \tilde{f}(x_m)$. Then $\tilde{f}(x) = f(x)$ wp 1 for $x \varepsilon \partial R$ and, hence $\tilde{f}(x)$ is a version of $f(x)$, for all $x \varepsilon \bar{R}$.

By construction, for each $x \varepsilon \bar{R}$ there is a sequence $x_m \to x$ so that $x_m \varepsilon \mathfrak{I}$ and $\tilde{f}(x) = \lim_m \sup f^m(x_m) = \lim_m \sup \tilde{f}(x_m)$. Hence

$$\tilde{f}(x) \; \varepsilon \; \bigcap_{\varepsilon x A} \overline{\tilde{f}(\mathfrak{I} \cap A)}$$

and $\tilde{f}(\cdot)$ is separable, where N, the null set in the definition of separability, is the empty set.

Suppose $\mathfrak{I}' = \{x_m'\}$ and $\mathfrak{I} = \{x_m\}$ are countable dense sets in $\bar{R}$. Then continuity in probability implies that for each $x_m \varepsilon \mathfrak{I}$, there is a null set N_m so that, for $\omega \not\varepsilon N_m$.

$$f(x_m) \; \varepsilon \; \bigcap_{x_m \varepsilon A} \overline{f(\mathfrak{I}' \cap A)}. \tag{1}$$

Thus, (1) holds for any $\omega \not\varepsilon \bigcup_m N_m$ and any $x_m \epsilon \mathfrak{I}$. This implies that, if $\mathfrak{I}$ is a separability set, so is $\mathfrak{I}'$. Q.E.D.

The following Corollary will be useful in Section 4.

COROLLARY. *Let $f(x, y)$ be a real valued random function of the n-vector x and scalar variable y, where the latter takes values in a compact interval I. Let $f(\cdot, y)$ and R satisfy the conditions of Theorem* 1 *and, in addition let $f(\cdot, \cdot)$ be continuous in probability on $\bar{R} \times I$. Let $f(\cdot, y)$ be continuous* wp 1 *for each fixed y in I, and let $f(x, \cdot)$ be continuous, uniformly for x on $\bar{R}$,* wp 1. *Then there is a separable version of $f(\cdot, \cdot)$ which is continuous* wp 1 *on $\bar{R} \times I$.*

Proof. Let $\mathfrak{I} \subset \bar{R}$ be countable and dense in $\bar{R}$, and $\mathfrak{F} \subset I$ countable and dense in I. Then $\mathfrak{I} \times \mathfrak{F}$ is a separability set for $f(\cdot, \cdot)$ by Theorem 1. Since $\mathfrak{F}$ is countable, there is a null set N_1 so that, for $\omega \not\varepsilon N_1$, $f(\cdot, y)$ is continuous for any $y \varepsilon \mathfrak{F}$. Let x and x' be in $\mathfrak{I}$ and y and y' in $\mathfrak{F}$, and write

$$|f(x, y) - f(x', y')| \leqq |f(x, y) - f(x', y)| + |f(x', y) - f(x', y')|. \tag{2}$$

There is a null set N_2, not depending on x, x', y or y' so that, for $\omega \not\varepsilon N_2$, the second term on the right of (2) goes to zero (uniformly in $x \varepsilon \mathfrak{I}$) as $y \to y'$ along any sequence in $\mathfrak{F}$. Also the first term goes to zero as $x' \to x$, along any sequence in $\mathfrak{I}$. Thus, for $\omega \not\varepsilon N_2$, $f(\cdot, \cdot)$ is continuous at each point of $\mathfrak{I} \times \mathfrak{F}$ in the induced topology from E^n. Thus, $f(\cdot, \cdot)$ is continuous wp 1 on $\bar{R} \times I$, by separability. Q.E.D.

3. Continuity and differentiability. Preliminaries. Write D_i^j or $D_{x_i}^j$ for the differential operator[4] $\partial^j/\partial x_i^j$ and write D_i, D_{x_i} for D_i^1, $D_{x_i}^1$ respectively. Let R be a bounded open set with closure $\bar{R}$. Let $C_l(\bar{R})$ be the Banach space of functions $\varphi(\cdot)$ which are continuous, together with all of their mixed derivatives up to order l, and with norm (the sup is over $x \varepsilon \bar{R}$)

$$\|\varphi\|_{C_l(\bar{R})} = \sup |\varphi(x)| + \sum_{k=1}^{l} \sum_{l_1+\cdots+l_n=k} \sup |D_1^{l_1} \cdots D_n^{l_n} \varphi(x)|.$$

[4] In the sequel x_i, $i = 1, \cdots, n$, will denote the scalar components of the vector x.

$C(\bar{R})$ denotes $C_0(\bar{R})$. $C_l^0(\bar{R})$ denotes the sub (Banach) space of functions in $C_l(\bar{R})$ which, together with all their mixed derivatives up to order l, tend to zero as $x \to \partial R$, the boundary of $\bar{R}$. A measurable random function, denoted by $(D_i f(\cdot))$ and satisfying $\int E\,|(D_i f(x))|^p\,dx < \infty$, $p \geqq 1$, is to be called a *weak stochastic derivative*[5] of the process $f(\cdot)$, if for each fixed $\varphi(\cdot)\ \varepsilon\ C_1^0(\bar{R})$,

$$\int \varphi(x)(D_i f(x))\,dx = -\int D_i\varphi(x)\cdot f(x)\,dx \tag{3}$$

wp 1. Similarly, a measurable random function, denoted by $(D_1^{l_1} \cdots D_n^{l_n} f(\cdot))$ and satisfying $\int E\,|(D_1^{l_1} \cdots D_n^{l_n} f(x)|^p\,dx < \infty$ for some $p \geqq 1$, is called a weak stochastic derivative of order $(l_1, \cdots, l_n)$ if, for all $\varphi(\cdot)\ \varepsilon\ C_l^0(\bar{R})$,

$$(-1)^l \int (D_1^{l_1} \cdots D_n^{l_n} f(x))\varphi(x)\,dx = \int f(x)\cdot D_1^{l_1} \cdots D_n^{l_n}\varphi(x)\,dx \tag{4}$$

wp 1, where $l = l_1 + \cdots + l_n$.

Now, suppose that (4) holds wp 1 for each fixed $\varphi(\cdot)\ \varepsilon\ C_l^0(\bar{R})$. Since there is a *countable dense set* A_l in $C_l^0(\bar{R})$, there is a null set N_l so that $\omega \not\varepsilon N_l$ implies that (4) holds for all ω and $\varphi(\cdot)\ \varepsilon\ A_l$. We may suppose that $(D_1^{l_1} \cdots D_n^{l_n} f(\cdot))$ is integrable for $\omega \not\varepsilon N_l$. Thus, for $\omega \not\varepsilon N_l$, (4) holds for all $\varphi(\cdot)$ in $C_l^0(\bar{R})$. Thus, wp 1, the weak stochastic derivative is a weak derivative in the ordinary sense.

A 'mean square' derivative is a weak stochastic derivative as seen by the following: Let $f(\cdot)$ be a process with parameter set $\bar{R}$, and suppose that $Ef(x)f(y) = S(x, y)$ where the first and second mixed derivatives (with respect to $x_i, y_i, i = 1, \cdots, n$), of $S(x, y)$ are continuous on $\bar{R} \times \bar{R}$. Thus the mean square derivative of $f(\cdot)$ is continuous in probability. Let $(\tilde{D}_i f(\cdot))$ denote a measurable version of the mean square derivative of $f(\cdot)$ with respect to x_i. If, for each $\varphi(\cdot)\ \varepsilon\ C_1^0(\bar{R})$

$$\int \varphi(x)(\tilde{D}_i f(x))\,dx = -\int D_i\varphi(x)\cdot f(x)\,dx \tag{5}$$

wp 1, then $(\tilde{D}_i f(\cdot))$ is a version of $(D_i f(\cdot))$. But (5) follows from the evaluation

$$\begin{aligned} E\{\int [\varphi(x)(\tilde{D}_i(f(x)) + D_i\varphi(x)f(x)]\,dx\}^2 = \int\int [&\varphi(x)D_{y_i}\varphi(y)D_{x_i}S(x, y) \\ &+ \varphi(y)D_{x_i}\varphi(x)D_{y_i}S(x, y) + \varphi(x)\varphi(y)D_{y_i}D_{x_i}S(x, y) \\ &+ D_{x_i}\varphi(x)D_{y_i}\varphi(y)\cdot S(x, y)]\,dx\,dy = 0. \end{aligned}$$

The properties of the weak stochastic derivative and separability can be used to infer the existence of continuous or differentiable versions of $f(\cdot)$ on $\bar{R}$. However, since the weak stochastic derivatives are weak derivatives in the ordinary sense wp 1, it is far more convenient to refer directly to the so-called imbedding theorems of Sobolev (Smirnov [4], Section [114–118]). The proofs of these theorems implicitly involve the construction of continuous or differentiable $f(\cdot)$ in terms of integrals of the weak derivatives.

The Sobolev Embedding Theorems. Suppose R is bounded and open and there is a sphere S interior to R with the property that any radius vector drawn from any point on the surface of S intersects ∂R only once. Then R is asid to be *star shaped*

[5] If $f(\cdot)$ were a sure function, then a function satisfying (3) for each $\varphi(\cdot)$ in $C_1^0(\bar{R})$ is called a weak derivative.

with respect to the sphere S. Let $(D_1^{l_1} \cdots D_n^{l_n}\psi(\cdot))$ denote the weak $(l_1, \cdots, l_n)$ derivative of a function $\psi(\cdot)$ defined on $\bar{R}$. Let $W_{l,p}(\bar{R})$, $p \geqq 1$, denote the Banach space of functions with norm

$$\|\psi\|_{W_{l,p}(\bar{R})} = \sum_{k=0}^{l} \sum_{l_1+\cdots+l_n=k} \|(D_1^{l_1} \cdots D_n^{l_n}\psi)\|_{L_p(\bar{R})}$$

where

$$\|\Psi\|_{L_p(\bar{R})}^p = \int |\Psi(x)|^p \, dx.$$

The proof of the following theorem of Sobolev can be found in Smirnov [4], Section [114–118].

THEOREM S. *Let $pl > n$ and $p > 1$. Let the bounded open set R be the sum of bounded open sets A_i, which are connected by bounded differentiable $(n - 1)$ manifolds, and let each A_i be star shaped with respect to some sphere $S_i \subset A_i$. Then any $\psi(\cdot) \in W_{l,p}(\bar{R})$ is continuous on $\bar{R}$ and every set of $\psi(\cdot)$ which is bounded in $W_{l,p}(\bar{R})$ is pre-compact in $C(\bar{R})$. Also*

$$\|\psi\|_{C(\bar{R})} \leqq K \|\psi\|_{W_{l,p}(\bar{R})},$$

where K depends only on R, l, n and p. Let $0 < m < l - n/p$. Then all the weak derivatives or order m are continuous in $\bar{R}$ and bounded sets in $W_{l,p}(\bar{R})$ are pre-compact in $C_m(\bar{R})$.

REMARK. The definition of a weak derivative involves equivalence classes of functions. The members of an equivalence class differ on at most a set of Lebesgue measure zero. The Sobolev theorem states essentially that, if the weak derivatives satisfy certain integrability properties, then $\psi(\cdot)$ (and perhaps its lower order weak derivatives) can be chosen to be continuous.

4. Continuity and differentiability.

THEOREM 2. *Let R satisfy the conditions of Theorem S. Let $(D_1^{l_1} \cdots D_n^{l_n}f(\cdot))$ be a version of the weak stochastic derivative of order $l_1, \cdots, l_n$ of a process $f(\cdot)$ which is continuous in probability. Let the weak stochastic derivatives satisfy*

$$\int E\,|(D_1^{l_1} \cdots D_r^{l_n}f(x))|^p \, dx < \infty \tag{6}$$

for all l_j and k for which $0 \leqq l_1 + \cdots + l_n \leqq k \leqq l$. Let $pl > n$ and $p > 1$. Then there is a version of $f(\cdot)$ which is continuous on $\bar{R}$ wp 1. *For each $\epsilon > 0$, there is an ω set B, $P\{B\} > 1 - \epsilon$, such that the set of $f(\cdot)$, for $\omega \in B$, is pre-compact in $C(\bar{R})$.*

Let $0 < m < l - n/p$. Then there are versions of the weak stochastic derivatives of order $\leqq m$ which are continuous on $\bar{R}$ wp 1. *For each $\epsilon > 0$, there is a set B, $P\{B\} > 1 - \epsilon$, such that the set of $f(\cdot)$, for $\omega \in B$, is pre-compact in $C_l(\bar{R})$.*

PROOF. The proofs are almost (but not quite) immediate consequences of Theorem S since the weak stochastic derivatives are ordinary weak derivatives for each sample function not in some null set N. By (6), we may suppose that the $(D_1^{l_1} \cdots D_n^{l_n}f(\cdot))$ are in $L_p(\bar{R})$ for $\omega \not\in N$. By applying Theorem S to each sample function $f(\cdot)$, for ω not in the null set N, we obtain that each sample function is equivalent to a continuous function $\tilde{f}(\cdot)$ in the sense that both have the same weak derivatives. We must show that $\tilde{f}(\cdot)$ is measurable as a function of (x, ω) and is a continuous *version of* $f(\cdot)$.[6]

For any measurable version $f(\cdot)$, and ω not in some null set N, and for any $\varphi(\cdot) \varepsilon C_l^0(\bar{R})$, we now have

$$\int D_i\varphi(x)(f(x) - \tilde{f}(x))\, dx = 0. \tag{7}$$

We will choose a suitable sequence of $\varphi_n(\cdot)$ which will yield for each pair $x, y \varepsilon R$ that

$$\tilde{f}(x) - \tilde{f}(y) = f(x) - f(y) \tag{8}$$

for ω not in some null set $N(x, y)$. Suppose, *for the moment*, that (8) is true. Let $\mathfrak{J}$ be a separability set for $f(\cdot)$. Then on $\mathfrak{J} \cap R$, $f(x) = \tilde{f}(x) + z(\omega)$ for $\omega \not\varepsilon \bigcup_{x\varepsilon\mathfrak{J}\cap R} N(x, y)$, and where $z(\omega) = f(y) - \tilde{f}(y)$. Then, wp 1, $f(\cdot)$ is continuous on $\mathfrak{J} \cap R$. Now, by separability, some version of $f(\cdot)$ is continuous wp 1 on R. Thus $z(\omega) = 0$, since $\tilde{f}(\cdot)$ is obtained by changing $f(\cdot)$ on a set of Lebesgue measure zero wp 1, and $f(\cdot)$ and $\tilde{f}(\cdot)$ are continuous wp 1; and this version of $f(\cdot)$ equals $\tilde{f}(\cdot)$ on R wp 1. Finally, since $\tilde{f}(\cdot)$ is continuous on $\bar{R}$ wp 1., and $\tilde{f}(\cdot)$ equals $f(\cdot)$ on R, and $f(\cdot)$ is continuous in probability on $\bar{R}$, there is a wp 1 continuous and measurable version of $f(\cdot)$ on $\bar{R}$. Thus, we only need prove (8).

Define the vectors $x' = (c_1, \cdots, c_{i-1}, b_i, c_{i+1}, \cdots, c_n)$ and

$$x'' = (c_1, \cdots, c_{i-1}, c_i, c_{i+1}, \cdots, c_n).$$

Let c_j^m and b_j^m be sequences tending to c_j from below and above, respectively, for $j \neq i$. Define the rectangles $S_m(x', x'') = \{x: b_i \leqq x_i \leqq c_i\,;\, c_j^m \leqq x_j \leqq b_j^m, j \neq i\}$. Let $x' \varepsilon R$ and suppose that $|b_i - c_i|$ is small enough so that the line connecting x' to x'' lies entirely in R. Define the functions of x (with all integrals to be interpreted in the Lebesgue sense).

$$\begin{aligned} \rho_\epsilon(x) &= \exp -[\epsilon^2/(\epsilon^2 - |x|^2)], \qquad && |x| \leqq \epsilon \\ &= 0, && |x| > \epsilon; \\ \varphi_{m,\epsilon}(x', x'', x) &= K_\epsilon \int Y_m(x', x'', y)\rho_\epsilon(y - x)\, dy, \end{aligned}$$

where $Y_m(x', x'', \cdot)$ is the characteristic function of $S_m(x', x'')$, and K_ϵ is the constant for which $\int \varphi_{m,\epsilon}(x', x'', x)\, dx = 1$. Define the sphere $B_\epsilon(x) = \{\tilde{x}: |x - \tilde{x}| < \epsilon\}$, and the ϵ neighborhood of $S_m(x', x'')$ as the union of the spheres $B_\epsilon(x)$, as x ranges over $S_m(x', x'')$. For each sufficiently small $\epsilon > 0$, $S_{m,\epsilon}(x', x'')$ is in R for all sufficiently large m and $\varphi_{m,\epsilon}(x', x'', x)$ is zero for x outside of $S_{m,\epsilon}(x', x'')$. Then, for sufficiently small $\epsilon > 0$, $\varphi_{m,\epsilon}(x', x'', \cdot) \varepsilon C_l^0(\bar{R})$ for all $l < \infty$, and, as $\epsilon \to 0$, its values tend to the inverse of the volume of $S_m(x', x'')$ interior to $S_m(x', x'')$, and to zero outside $S_m(x', x'')$.

---

[6] This is not obvious, since applying Theorem S to the function $f(\cdot) + \epsilon(\cdot)$, where $\epsilon(x) = 1$ at $x = x^0$ and is zero elsewhere and $f(\cdot)$ is differentiable, gives a continuous equivalent $\tilde{f}(\cdot) = f(\cdot)$. Yet, the continuous equivalent would not have the same distribution as $f(\cdot) + \epsilon(\cdot)$ at $x = x^0$. The assumption of stochastic continuity is used to eliminate this possibility.

Furthermore, for all sufficiently small $\epsilon > 0$ and all large $m < \infty$, there are disjoint neighborhoods $A'_{m,\epsilon}$, $A''_{m,\epsilon}$ of x', x'', respectively, contained in R and with the properties: $D_i\varphi_{m,\epsilon}(x', x'', x) > 0$ in $A'_{m,\epsilon}$, $D_i\varphi_{m,\epsilon}(x', x'', x) < 0$ in $A''_{m,\epsilon}$, and is zero elsewhere. As $m \to \infty$ and $\epsilon \to 0$, $A'_{m,\epsilon}$ and $A''_{m,\epsilon}$ tend to x' and x'', respectively. Since $\varphi_{m,\epsilon}(x', x'', \cdot)$ is symmetric, we have, for large m and small ϵ,

$$\int_{A'_{m,\epsilon}} D_i\varphi_{m,\epsilon}(x', x'', x)\,dx = -\int_{A''_{m,\epsilon}} D_i\varphi_{m,\epsilon}(x', x'', x)\,dx \equiv d_{m,\epsilon} > 0.$$

Now, using $\varphi_{m,\epsilon}(x', x'', \cdot)$ for the $\varphi(\cdot)$ in (7), together with the continuity wp 1 of $\tilde{f}(\cdot)$, yields the limiting relation

$$(9)\quad \tilde{f}(x') - \tilde{f}(x'') = \lim_{m\to\infty}\lim_{\epsilon\to 0}\Big[\int_{A'_{m,\epsilon}} D_i\varphi_{m,\epsilon}(x', x'', x)\cdot f(x)\,dx + \int_{A''_{m,\epsilon}} D_i\varphi_{m,\epsilon}(x', x'', x)\cdot f(x)\,dx\Big]/d_{m,\epsilon}$$

wp 1 for each fixed x', x''. Owing to the continuity in probability of $f(\cdot)$, the right hand integrals converge to random variables with the distributions of (hence, versions of) $f(x')$ and $-f(x'')$, respectively. Then (8) holds for any x, y lying on a line segment contained in R and parallel to some coordinate axis.

Let α denote the direction $\sum e_i\alpha_i$, where e_i is the unit vector in the ith coordinate direction and $\sum \alpha_i^2 = 1$. Then we identify the 'weak stochastic derivative in the direction α' $(D_\alpha f(\cdot))$ with $\sum \alpha_i(D_i f(\cdot))$ (where $D_\alpha = \sum \alpha_i D_i$) in the sense that, wp 1 for each $\varphi(\cdot) \in C_1^0(\bar{R})$,

$$\int D_\alpha\varphi(x)\cdot f(x)\,dx = -\int \varphi(x)(D_\alpha f(x))\,dx.$$

By repeating the derivation of (9) with D_α replacing D_i and $S_m(x', x'')$ oriented in the α direction, we obtain that (8) holds wp 1 for any x, y on any connected line segment in R. Since any two points in R can be connected by a finite chain of line segments in R, (8) holds as stated.

The compactness statement of the first paragraph of the theorem follows from Theorem S since, by (6) for each $\epsilon > 0$, there is a constant $M_\epsilon < \infty$ and a set B_ϵ, with $P\{B_\epsilon\} > 1 - \epsilon$ so that $K\,\|f\|_{W_{l,p}(\bar{R})} \leqq M_\epsilon$ for $\omega \in B_\epsilon$.

The proof of the statement of the last paragraph of the theorem is almost identical to the proof of the first paragraph, using the fact that the 'higher order' weak stochastic derivatives are weak stochastic derivatives of the 'lower order' weak stochastic derivatives, and we omit the details. Q.E.D.

The proof of Theorem 3 is rather similar to the proof of Theorem 2 and we omit the details.

THEOREM 3. *Let R satisfy the conditions of Theorem 2. Let $f(\cdot)$ have continuous sample functions* wp 1, *and let there exist continuous* wp 1 *versions of the weak stochastic derivatives $(D_i f(\cdot))$, $i = 1, \cdots, s$. Then, $f(\cdot)$ has continuous derivatives* wp 1 *with respect to x_i, $i = 1, \cdots, s$. These derivatives can be identified with the $(D_i f(\cdot))$, $i = 1, \cdots, s$.*

Theorem 4 is useful in some cases where some parameter is 'time' and no weak stochastic derivative with respect to 'time' exists; e.g., where, as a function of time, the process behaves as a Wiener process. See example in Section 5.

THEOREM 4. *Let $f(x, y)$ be a separable process with parameter (x, y) and which is* *continuous in probability. Let x vary over $\bar{R}$, where R obeys the conditions of Theorem 2, and let y be a scalar parameter varying over the compact interval I. For each fixed pair y, y' let $f(\cdot, y) - f(\cdot, y') \equiv F(\cdot, y, y')$ have the properties of the $f(\cdot)$ of Theorem 2. For some real $K_2 < \infty$ and $\infty > q \geqq 1$, let there be a real $\alpha > 0$ so that*

$$E\,\|f(\cdot, y) - f(\cdot, y')\|^q_{W_{l,p}(\bar{R})} \leqq K_2\,|y - y'|^{1+\alpha}.$$

then there is a wp 1 *continuous version of $f(\cdot, \cdot)$ on $\bar{R} \times I$. Also, for any $\beta < \alpha/q$, there is a version of $f(\cdot, \cdot)$, and a $K(\omega) < \infty$* wp 1, *so that* wp 1,

$$g(y, y') \equiv \sup_{x\varepsilon\bar{R}} |f(x, y) - f(x, y')| \leqq K(\omega)|y - y'|^{\beta} \tag{10}$$

i.e., *$f(\cdot, \cdot)$ is Holder continuous in y, uniformly in x,* wp 1.

PROOF. For simplicities sake, let $I = [0, 1]$, and denote

$$\mathfrak{F} = \{k2^{-n}; k = 0, 1, \cdots, 2^n; n = 1, 2, \cdots\}.$$

If $\mathfrak{I}$ is dense in $\bar{R}$, then $\mathfrak{I} \times \mathfrak{F}$ is a separability set for $f(\cdot, \cdot)$ by Theorem 1. Suppose that there is some $K(\omega) < \infty$ wp 1 so that, for some version of $f(\cdot, \cdot)$,

$$g(y, y') \leqq K(\omega)|y - y'|^{\beta} \tag{11}$$

for all y, y' in $\mathfrak{F}$ with $y' > y$, and $\omega \not\varepsilon N$, a null set. Then, for $\omega \not\varepsilon N$,

$$|f(x, y) - f(x, y')| \leqq K(\omega)|y - y'|^{\beta} \tag{12}$$

for x, y, y' in $\mathfrak{I} \times \mathfrak{F} \times \mathfrak{F}$, and there is a separable version of $f(\cdot, \cdot)$ so that (12) holds for all x, $y \,\varepsilon\, \bar{R} \times I$ and for all ω not in some null set N_1. Then, since there is a version of each $f(\cdot, y)$ which is continuous wp 1., and a version of $f(\cdot, \cdot)$ so that $f(x, \cdot)$ is Hölder continuous, uniformly in x, wp 1., the Corollary to Theorem 1 implies that there is a version of $f(\cdot, \cdot)$ which is continuous wp 1. Also, this version satisfies (11) wp 1 for all y, y' in I. Thus, we need only demonstrate (11) for y, $y' \,\varepsilon\, \mathfrak{F}$, $y' > y$ and $\omega \not\varepsilon N$. The proof of this is essentially that given by Neveu [3], Proposition III. 5.2, for the continuity of scalar parameter processes.

By Theorem 2, for each $y \,\varepsilon\, I$, there is a null set $N(y)$ so that a version of $f(\cdot, y)$ is continuous for $\omega \not\varepsilon N(y)$. Let $N_1 = \bigcup_{y\varepsilon\mathfrak{F}} N(y)$. Then for any y, $y' \,\varepsilon\, \mathfrak{F}$, $y' > y$, and $\omega \not\varepsilon N_1$, Theorems 2 and S imply that these versions satisfy

$$g(y, y') \leqq K\,\|f(\cdot, y) - f(\cdot, y')\|_{W_{l,p}(\bar{R})}\,.$$

By hypothesis

$$Eg(y, y')^q \leqq K_2\,|y - y'|^{1+\alpha}. \tag{13}$$

Write $Z_m = \sup_{2^m \geqq k \geqq 0} g(k2^{-m}, (k + 1)2^{-m})$. Then, for any $\gamma > 0$ for which $0 < \alpha - \gamma q \equiv \rho$, an application of Chebychevs' inequality yields

$$P_m \equiv P\{Z_m \geqq \epsilon 2^{-m\gamma}\} \leqq 2^m P\{g(k2^{-m}, (k + 1)2^{-m}) \geqq \epsilon 2^{-m\gamma}\}$$

$$\leqq K_2 2^m \cdot 2^{-m(1+\alpha)}(\epsilon 2^{-m\gamma})^{-q} = K_2 2^{-m\rho}\epsilon^{-q}.$$

Since $\sum_{m=1}^{\infty} P_m < \infty$, there is some $m(\omega) < \infty$ wp 1 so that $m \geqq m(\omega)$ implies that $\sup_{2^m \geqq k \geqq 0} g(k2^{-m}, (k + 1)2^{-m}) < \epsilon 2^{-m\gamma}$. Choose any random $h > 0$ so that

$h < 2^{-m(\omega)}$. Let m be the unique random integer ($\geqq m(\omega)$) satisfying $h/2 \leqq 2^{-m} \leqq h$.

Let k be any random integer (perhaps depending on y) so that $|y - k2^{-m}| < 2^{-m}$. Then $g(y, k2^{-m}) \leqq 2 \sum_m^\infty Z_r$. If $|y - y'| < 2^{-m}$, there is some random k so that $|y - k2^{-m}| < 2^{-m}$ and $|y' - k2^{-m}| < 2^{-m}$. Using this and $g(y, y') \leqq g(y, k2^{-m}) + g(k2^{-m}, y')$ we have, for $K_3 = 2\epsilon \sum_0^\infty 2^{-r\gamma}$,

$$(14) \quad \sup_{|y-y'|<h/2, y'>y; y,y' \varepsilon \mathfrak{T}} g(y, y') \leqq \sup_{|y-y'|<2^{-m}, y'>y; y,y' \varepsilon \mathfrak{T}} g(y, y') \leqq 4 \sum_m^\infty Z_r \leqq 2 K_3 2^{-m\gamma}.$$

(14) implies (11) for, $\beta = \gamma$, $K(\omega) = K_3 2^{1+\beta}$ and $|y - y'| < 2^{-m(\omega)}/2$. Since $m(\omega) < \infty$ wp 1., (14) implies (11) as stated for some $K(\omega) < \infty$ wp 1. Q.E.D.

5. Example. An application of Theorem 4, which has been found useful in a problem in stochastic control theory [2] will be given. Let R satisfy the conditions of Theorem 2, and let z_t be a Wiener process (see Doob [1] for the definition). If $\int_0^t g^2(x, t, s)\, ds < M < \infty$ for $(x, t) \in \bar{R} \times [0, T]$, then, using the Itö or Wiener definition of the stochastic integral ([1], IV, Section 2), the integral

$$(15) \qquad \psi(t) = \int_0^t g(x, t, s)\, dz_s$$

is well defined wp 1 for each fixed $(x, t) \in \bar{R} \times [0, T]$. Let $\mathfrak{L}$ be an elliptic (partial differential) operator, let ξ_t denote the *formal* expression dz_t/dt, and consider the *formal* equation

$$(16) \qquad \partial u(x, t)/\partial t = \mathfrak{L}u(x, t) + \sigma(x, t)\xi_t .$$

Equations such as (16) appear in a natural way in certain problems in stochastic control theory, where the object to be controlled is 'distributed' in space, and in certain statistical estimation problems concerning the estimation of the solution of $\partial u(x, t)/\partial t = \mathfrak{L}u(x, t) + k(x, t)$, when noise corrupted observations of $u(\cdot, \cdot)$ are taken. See Kushner [2] for some more details.

In any case, it would be useful to define a precise solution, $u(x, t)$, to (16), which is consistent with its intuitive meaning and which is continuous wp 1 on $\bar{R} \times [0, T]$, and, perhaps, which is even a Markov process (with parameter t) and values in a suitable space of continuous functions. (Some of these questions are treated in [2].) Write

$$(17) \qquad u(x, t) = \int_0^t dz_s \{\int G(x, x'; t, s)\sigma(x', s)\, dx'\}$$

where $G(\cdot, \cdot; \cdot, \cdot)$ is the Greens function for $(\partial/\partial t - \mathfrak{L})$ with the necessary boundary conditions. Instead of treating (17) here, we will only consider the derived form

$$(18) \qquad \psi_0(x, t) = \int_0^t dz_s \alpha_0(x, t, s)$$

and give conditions on $\alpha_0(\cdot, \cdot, \cdot)$ which guarantee that there is a version of each $\psi_0(x, t)$ so that $\psi_0(\cdot, \cdot)$ is continuous wp 1 on $\bar{R} \times [0, T]$.

Assume that (A-1): $\alpha_0(\cdot, \cdot, \cdot)$ and $D_i\alpha_0(\cdot, \cdot, \cdot) \equiv \alpha_i(\cdot, \cdot, \cdot)$, $i = 1, \cdots, n$, are bounded, uniformly in (x, t) by a square integrable function of s, $s \in [0, T]$.

(A-2): let

$$\int_0^T [\alpha_i(x + \delta, t, s) - \alpha_i(x, t, s)]^2 \, ds \to 0$$

as $|\delta| \to 0$. (A-3): for some real K_1 and $\gamma > 0$

$$\int_0^t [\alpha_i(x, t', s) - \alpha_i(x, t, s)]^2 \, ds + \int_t^{t'} \alpha_i^2(x, t', s) \, ds \leqq K_1 |t - t'|^\gamma.$$

By (A-1), each $\psi_i(x, t) \equiv \int_0^t \alpha_i(x, t, s) \, dz_s$ is well defined wp 1 for each $(x, t) \in \bar{R} \times [0, T]$. By (A-2) and (A-3), for $\Delta \geqq 0$,

$$\begin{aligned} &E[\psi_i(x + \delta, t + \Delta) - \psi_i(x, t)]^2 \\ &\quad = E\{\int_0^{t+\Delta} \alpha_i(x + \delta, t + \Delta, s) \, dz_s - \int_0^t \alpha_i(x, t, s) \, dz_s\}^2 \\ &\quad = \int_0^t [\alpha_i(x + \delta, t + \Delta, s) - \alpha_i(x, t, s)]^2 \, ds + \int_t^{t+\Delta} \alpha_i^2(x + \delta, t + \Delta, s) \, ds \\ &\quad \to 0 \end{aligned}$$

as $|\delta| \to 0$ and $\Delta \to 0$. Thus the $\psi_i(\cdot, \cdot)$ are continuous in probability on $\bar{R} \times [0, T]$ and, by Theorem 1, there are separable and measurable versions and any dense set in $\bar{R} \times [0, T]$ is a separability set.

Now, for any even integer $r \geqq 1$, there is a real number B_r for which (for $t' \geqq t$)

$$\begin{aligned} &\textstyle\sum_{i=0}^n E\, |\psi_i(x, t) - \psi_i(x, t')|^r \\ &\quad = B_r \textstyle\sum_{i=0}^n \{\int_0^t [\alpha_i(x, t', s) - \alpha_i(x, t, s)]^2 \, ds + \int_t^{t'} \alpha_i^2(x, t', s) \, ds\}^{r/2} \qquad (19) \\ &\quad \leqq (n + 1) B_r K_1 |t - t'|^{\gamma r/2}. \end{aligned}$$

Now, for each fixed t, t', $\psi_i(\cdot, t) - \psi_i(\cdot, t')$ is the derivative in mean square (with respect to x_i) of $\psi_0(\cdot, t) - \psi_0(\cdot, t')$. Hence, as noted in Section 3, $\psi_i(\cdot, t) - \psi_i(\cdot, t')$ is a version of the weak stochastic derivative $(D_i(\psi_0(\cdot, t) - \psi_0(\cdot, t'))$. Now, let $l = 1$ and $r \geqq n + 1$. Then (19) implies that $\psi_0(\cdot, t) - \psi_0(\cdot, t')$ satisfies the conditions of Theorem 2 on $f(\cdot)$. Furthermore, to each r there is a real K_{2r} so that

$$\begin{aligned} E\, \|\psi_0(\cdot, t) - \psi_0(\cdot, t')\|^r_{W_{1,r}(\bar{R})} &= \int_{\bar{R}} \textstyle\sum_{i=0}^n E\, |\psi_i(x, t) - \psi_i(x, t')|^r \, dx \\ &\leqq K_{2r} |t - t'|^{\gamma r/2} = K_{2r} |t - t'|^{1+\alpha_r} \end{aligned}$$

where, for sufficiently large even r, $\alpha_r > 0$. Then, since we can assume that $r \geqq n + 1$, Theorem 4 is applicable, and there is a version of $\psi_0(\cdot, \cdot)$ which is continuous on $\bar{R} \times [0, T]$ and Holder continuous in t, uniformly in x, wp 1.

REFERENCES

[1] Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York.

[2] Kushner, H. J. (1968). On the optimal control of a system governed by a parabolic equation and 'white noise' inputs. *SIAM J. on Control* 6 No. 4.

[3] Neveu, J. (1965). *Mathematical Foundations of the Calculus of Probability* (English translation). Holden-Day, San Francisco.

[4] Smirnov, V. I. (1964). *Integration and Functional Analysis* (English translation). Pergamon Press, New York.