JOURNAL OF APPROXIMATION THEORY **66**, 225–228 (1991)

Approximation by Mutually Completely Dependent Processes

RICHARD A. VITALE*

Department of Statistics, U-120, University of Connecticut, Storrs, Connecticut 06269, U.S.A.

Communicated by Allan Pinkus

Received August 23, 1990; revised September 22, 1990

Any stochastic process indexed on the integers can be approximated arbitrarily well in the L_∞ sense by a stochastic process in which each value is (essentially) a 1–1 image of any other value. If the original process has continuous one dimensional marginal distributions, then these can be replicated in the approximating process.

1

In the literature on probability distributions with "fixed marginals," attention has been given to questions of approximation by random variables which have prescribed marginal distributions and which exhibit not merely statistical but *functional* dependence and even *invertible* functional dependence. The first result in this vein is apparently Theorem 1 of Brown [1] which states (in different language) that if U and V are each uniformly distributed variables on $[0, 1]$ then there is a sequence $\{\tau_n\}$ of invertible, (Borel) measure-preserving maps of the interval and pairs of uniforms (U_n, V_n), $V_n = \tau_n U_n$, such that (U_n, V_n) converges in distribution to (U, V). Variants and elaborations of this result appear in Kimeldorf and Sampson [2], Mikusínski, Sherwood, and Taylor [5], and Vitale [6]. Each of these discusses approximation in the weak sense and apart from the last reference where a finite collection of variables is considered, attention has always been restricted to a pair of variables. Recently it has become apparent that one can formulate a substantially stronger result which deals with a countable collection of random variables and in which approximation is in the mode of a.s. uniform convergence. It asserts in

* Supported in part by NSF Grants DMS-8603944, DMS-900265, and ONR Grant N00014-90-J-1641.

0021-9045/91 $3.00

effect that rather general stochastic processes on the integers can be quite precisely approximated by "nearly deterministic" ones.

For convenience, we will call any collection $\mathbf{X} = \{X_k\}_{k=1}^{\infty}$ of random variables a *process*. If, for each $j \neq k$, there is a function $\varphi_{jk}: R \to R$ such that $X_k = \varphi_{jk}(X_j)$ a.s., then the process is *mutually completely dependent* (*mcd*) (cf. Lancaster, [3]). Observe that in this case all the randomness in the process can be regarded as loaded into the initial datum via the representation $X_k = \varphi_{1k}(X_1)$.

THEOREM 1. *Suppose that $\mathbf{X} = \{X_k\}_{k=1}^{\infty}$ is a process with continuous (one-dimensional) marginals $X_k \sim F_k$. Then there is a sequence of mcd processes $\mathbf{X}^{(n)} = \{X_k^{(n)}\}_{k=1}^{\infty}$, $n = 1, 2, \ldots$, such that*

$$X_k^{(n)} \sim F_k \qquad k, n = 1, 2, \ldots \tag{1}$$

and

$$\|\mathbf{X} - \mathbf{X}^{(n)}\|_{\infty} = \sup_k |X_k - X_k^{(n)}| \to 0 \qquad \text{a.s. as } n \to \infty. \tag{2}$$

To establish this, we first recall some familiar facts.

PROPOSITION 1. *If X has a continuous distribution function F, then $U = F(X)$ is uniformly distributed on the interval and $X = F^{-1}(U)$ a.s., where $F^{-1}(u) \equiv \inf\{x \mid u \leqslant F(x)\}$.*

PROPOSITION 2 (Kuratowski; Royden, [5]). *Every uncountable, complete, separable metric space is Borel equivalent to* $[0, 1]$.

In applying Proposition 2, we have in mind the metric space $S = (R^{\infty}, d)$ where

$$d((x_1, x_2, \ldots), (y_1, y_2, \ldots)) = \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{|x_k - y_k|}{1 + |x_k - y_k|}.$$

In referring to maps of the interval, measure-preserving will be always be in the Borel sense and τ invertible will mean $\exists \tau^{-1}$ such that $\tau \circ \tau^{-1}(U) = \tau^{-1} \circ \tau(U) = U$ a.s. for a uniform U. We will also need the following technical result, whose proof is spelled out for completeness.

LEMMA. *Suppose that $X = L \circ \tau(U)$ a.s. where U is uniform variable, τ is measure-preserving on the interval, and L is non-decreasing on* $(0, 1)$. *Then given $\varepsilon > 0$, there is an invertible measure-preserving map of the interval $\tilde{\tau}$ such that*

$$P[|X - \tilde{X}| > \varepsilon] < \varepsilon, \qquad \text{where} \quad \tilde{X} = L \circ \tilde{\tau}(U). \tag{3}$$

Proof. Since L has only a countable number of jumps, there is a partition of the interval $0=u_0<u_1<\cdots<u_N=1$ such that (i) $u_1+1-u_{N-1}<\varepsilon$ and (ii) if $u_i<u<v<u_{i+1}$, then $0\leqslant L(v)-L(u)<\varepsilon/2$. We define $\tilde{\tau}$ in a similar manner on each of the sets $A_i=\tau^{-1}((u_i, u_{i+1}))$ as follows (on $[0,1]/\bigcup_{i=0}^{N-1} A_i$, $\tilde{\tau}$ can be defined arbitrarily): Let Y be U conditioned to lie in A_i (observe that A_i has positive Lebesgue measure $m(A_i)=u_{i+1}-u_i$). It has a continuous distribution function G and by Proposition 1, $G(Y)$ is uniform on $(0, 1)$. It follows that we may take $u_i+(1-G(\cdot))(u_{i+1}-u_i)$ to be the restriction of $\tilde{\tau}$ to A_i.

By (ii) and the fact that τ and $\tilde{\tau}$ simultaneously map A_i into (u_i, u_{i+1}), $i=1,\ldots,N-2$, it follows that $|L\circ\tau(u)-L\circ\tilde{\tau}(u)|<\varepsilon$ for $u_1<u<u_{N-1}$. Hence $P[|X-\tilde{X}|>\varepsilon]\leqslant P[U\leqslant u_1 \text{ or } U\geqslant u_{N-1}]<\varepsilon$ by (i).

Proof of Theorem 1. Let $\varepsilon_n \searrow 0$ and observe that

$$\sum_n P(\sup_k |X_k-X_k^{(n)}|>\varepsilon_n)\leqslant\sum_n\sum_k P(|X_k-X_k^{(n)}|>\varepsilon_n).$$

If it can be shown, within the required conditions, that $P(|X_k-X_k^{(n)}|>\varepsilon_n)$ can be made arbitrarily small (so as to ensure convergence of the double series), the theorem will follow by the Borel–Cantelli lemma.

By Proposition 2 and the subsequent remark, there is a Borel equivalence $\psi\colon S\to[0,1]$ such that $\psi(X)$ is a random variable on $[0, 1]$. Owing to the continuous marginals of $\mathbf{X}$, $\psi(\mathbf{X})$ can have no atom. Applying its distribution function H, we have $U=H\circ\psi(\mathbf{X})$ uniform on $[0, 1]$ and $\mathbf{X}=(H\circ\psi)^{-1}(U)$. Let π_k be the kth coordinate projection of the $\mathbf{X}$-process; i.e., $X_k=\pi_k\mathbf{X}$. Then

$$F_k(X_k)=F_k\circ\pi_k\circ(H\circ\psi)^{-1}(U)\equiv\tau_k(U)$$

is a uniform variable and $X_k=F_k^{-1}\circ\tau_k(U)$ a.s. By the lemma, we can find an invertible $\tau_k^{(n)}$ such that with $X_k^{(n)}\equiv F_k^{-1}\circ\tau_k^{(n)}(U)$, $P[|X_k-X_k^{(n)}|>\varepsilon]$ is as small as desired.

It remains to be noted that for each $j\neq k$, $X_k^{(n)}=\varphi_{jk}^{(n)}(X_j^{(n)})$ a.s., where $\varphi_{jk}^{(n)}=F_k^{-1}\circ\tau_k^{(n)}\circ\tau_j^{(n)^{-1}}\circ F_j$.

2

It is easy to see (e.g., iid Bernoulli variables) that (1) generally fails in the absence of continuous marginals. We show next, however, that by supplementing our previous argument (2) can always be achieved.

THEOREM 2. *For any process* $\mathbf{X}$, (2) *holds.*

Proof. We consider first the case when there is no realization of $\mathbf{X}$ with positive probability: that is, for each $\vec{x} = (x_1, x_2, ...)$

$$P(\mathbf{X} = \vec{x}) = 0. \tag{4}$$

We proceed as in the proof of Theorem 1 to obtain $\psi(\mathbf{X})$, which now by (4) has a continuous distribution. We obtain U to be uniform in the same manner and have $X_k = \pi_k \circ (H \circ \psi)^{-1}(U) \equiv M_k(U)$. By methods similar to those used for the lemma, one can find an invertible $M_k^{(n)}$ such that $P[|X_k - X_k^{(n)}| > \varepsilon]$ is as small as desired. Then $X_k^{(n)} = M_k^{(n)} \circ M_j^{(n)-1}(X_j)$.

In the general case, suppose that (4) fails for $\vec{x}_1, \vec{x}_2, ...$. Condition $\mathbf{X} \notin (\vec{x}_1, \vec{x}_2, ...\}$. Then by our previous argument we have the existence of coordinate variables (conditionally) mcd which are close to those of the conditional $\mathbf{X}$. They will remain close under the following transformation of the line into "half of itself": for a large N and integers $-\infty < j < \infty$, map all values in $[2j/2^N, (2j+2)/2^N)$ into $[2j/2^N, (2j+1)/2^N)$ via $x \to 2j/2^N + \frac{1}{2}(x - 2j/2^N)$. It remains in the definition of $X_k^{(n)}$ to replace $\pi_k \vec{x}_1$, $\pi_k \vec{x}_2, ...$ by distinct values lying in the other half of the line $\bigcup_{j=-\infty}^{+\infty} [(2j+1)/2^N, (2j+2)/2^N)$.

References

1. J. R. Brown, Approximation theorems for Markov operators, *Pacific J. Math.* **16** (1966), 13–23.
2. G. Kimeldorf and A. R. Sampson, Monotone dependence, *Ann. Statist.* **6** (1978), 895–903.
3. H. O. Lancaster, Correlation and complete dependence of random variables, *Ann. Statist.* **34** (1963), 1315–1321.
4. P. Mikusinski, H. Sherwood, and M. D. Taylor, Shuffles of Min, Preprint, University of Central Florida, 1989.
5. H. L. Royden, "Real Analysis" (3rd ed.), Macmillan Co., New York, 1988.
6. R. A. Vitale, Stochastic dependence and a class of degenerate distributions, *in* "Topics in Statistical Dependence" (H. Block, A. R. Sampson, and T. Savits, Eds.), IMS Lecture Notes— Monograph Series, to appear.

Printed in Belgium