# ON STOCHASTIC DEPENDENCE AND A CLASS OF DEGENERATE DISTRIBUTIONS

By RICHARD A. VITALE[1]

*University of Connecticut*

We investigate the approximation of stochastic dependence by functional relationships involving so-called cyclic permutations of the interval.

**1. Introduction.** Dependence between two random variables can take, of course, a variety of forms, of which stochastic independence and functional dependence can be argued to be most opposite in character. In the one case, neither variable provides any information about the other, whereas in the second case there is complete determination (or *complete dependence*: Lancaster, 1963). By means of a direct construction for uniform variables, Kimeldorf and Sampson (1978) showed, however, that one can pass continuously from one to the other of these situations in the natural sense of weak convergence. This obviously weakens complete dependence as a foil for independence (and led Kimeldorf and Sampson, 1978, to the fruitful concept of monotone dependence). Indeed, couched in somewhat different language, Theorem 1 of Brown (1966) can be read to state that *any* form of dependence between uniform random variables can be approximated in the weak sense by functionally related random variables.

On the other hand, this raises the question, of theoretical and obvious computational interest, of the extent to which complete dependence can be used to approximate forms of stochastic dependence. We pursue this in several directions. First we show that functional dependence can be specified to a highly stylized class of invertible functions, the so-called cyclic permutations of the interval. Second, we show that it is possible to move from two to an arbitrary finite number of random variables. Finally we extend to arbitrary (continuous) marginals. Regarding the last point, we systematically take the viewpoint of fixing marginals and consider dependence within this constraint; for the narrower question of regression in this context, see Vitale and Pipkin (1976) and Vitale (1979). We make extensive use of the *uniform representation* of random variables (Kimeldorf and Sampson, 1975).

[1]Research supported in part by NSF Grant DMS 8603944.

*AMS 1980 subject classifications* (1985 revision). Primary 60E05; secondary 28D05, 60B10, 62E10, 62H05.

*Key words and phrases.* Complete dependence, functional dependence, maximally dependent collection, measure-preserving transformation, permutation, stochastic independence.

For various references I am indebted to P. Shields and R. Sine. M. Klass offered stimulating comments.

In the next section, we set out some definitions and make some preliminary comments. Theorem 1 of Section 3 states that any pair of uniform random variables can be approximated in distribution by a second pair which exhibits (invertible) functional dependence. Although Theorem 2 of Section 5 strictly includes this result, we present a proof in detail to give an idea of what the hands-on analysis looks like and, in particular, to present a version of the construction of Brown (1966). The interested reader may like to compare Theorem 1 to results of Garsia (1976) and Holbrook (1981) which give exact, but generally non-invertible, relationships in the case of an independent pair of random variables. Section 4 contains remarks on Theorem 1. Using an approach that differs from those of both Brown (1966) and Kimeldorf and Sampson (1978), Section 5 takes up the general case of approximating a collection of random variables with continuous marginals but otherwise arbitrary joint distribution. Section 6 relates the foregoing to the extremal distributions of Hoeffding. Discrete approximating distributions occupy Section 7, and a result of Fairley, Pearl, and Verducci (1987) is sharpened in Section 8. Section 9 concludes with a construction of a canonical sequence (Lai and Robbins, 1976, 1978) which is degenerate in a stronger sense than previous examples.

**2. Notation and Preliminaries.** We shall deal with Borel maps of the line (or the interval) equipped with Lebesgue measure $m(\cdot)$. A Borel map $T$ from the interval to itself such that $m(T^{-1}(B)) = m(B)$ for any Borel $B$ will be called *measure-preserving* and the entire collection denoted $\mathrm{T}$. Within $\mathrm{T}$, we consider the class $\mathrm{T}_{\mathrm{inv}}$ of measure-preserving maps $T$ which are invertible, i.e., $T$ is $1-1$ and $T^{-1}$ is Borel and measure-preserving as well.

An interval of the form $((j-1)/n, j/n)$ for some $n \geq 1$ and $1 \leq j \leq n$ will be called a *dyadic interval of rank n*. Among invertible measure-preserving maps, we call $T$ a *permutation of rank n* if it maps by a translation each dyadic interval of rank $n$ onto a dyadic interval of rank $n$. This specifies $T$ except for its values at the end-points of the subintervals; we allow these values to be assigned in any way that makes $T$ one-to-one. We shall also refer to usual permutations $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$, but there should be no confusion. Recall that $\pi$ is *cyclic* if it has a single closed cycle (i.e., of length $n$, which need not be traversed in the natural order). We call a permutation $T$ *cyclic*, and denote the entire class by $\mathrm{T}_{\mathrm{cyc}}$, if as a map of dyadic intervals it similarly has a single cycle. Our aim in focusing on the last class of maps is two-fold. One is to produce a simple universal model for (approximate) dependence among random variables; the second is to prepare the way for future work of a computational nature (see Section 7). These definitions are drawn from Halmos (1956) to which we refer later.

The following is standard.

PROPOSITION 2.1. *Suppose that $\{E_1, \ldots, E_k\}$ is a Borel partition of $[0,1]$ and that $\delta > 0$ is given. There is a partition $\{F_1, \ldots, F_k\}$ of $[0,1]$ in which each $F_i$ is a finite union of intervals, $m(F_i) = m(E_i)$, and $m(E_i \triangle F_i) < \delta$, $i = 1, \ldots, k$. (Here $\triangle$ denotes symmetric difference.)*

**3. Dependence in the Square.**

THEOREM 1. *Let $U$ and $V$ be uniformly distributed variables. There is a sequence of cyclic permutations $T_1, T_2, \ldots$ such that $(U, T_nU)$ converges in distribution to $(U, V)$ as $n \to \infty$.*

PROOF. The proof consists of two parts. First we divide $[0,1] \times [0,1]$ into subsquares and adapt an argument of Brown (1966) to find a permutation $T$ such that the distributions of $(U, V)$ and $(U, TU)$ coincide on subsquares. If $T$ is not a cyclic permutation, then we proceed to find such an approximation to it; for this, Halmos (1956, p. 56) would suffice, but we supply a direct proof for completeness. Finally, we note that reducing the size of the sub-squares finishes the proof.

PART ONE. Let $n$, a power of 2, be fixed, and let $I_j = ((j-1)/n, j/n)$, $j = 1, \ldots, n$. We first produce a $T \varepsilon \mathrm{T_{inv}}$ such that

$$P(U \varepsilon I_j, TU \varepsilon I_k) = P(U \varepsilon I_j, V \varepsilon I_k) \equiv p_{jk}, \quad j, k = 1, \ldots, n. \tag{1}$$

Define two systems of subintervals

$$\begin{aligned}
I_{j1} &= \left(\frac{(j-1)}{n}, \frac{j-1}{n} + p_{j1}\right) \\
I_{j2} &= \left(\frac{j-1}{n} + p_{j1}, \frac{j-1}{n} + p_{j1} + p_{j2}\right) \quad 1 \le j \le n \\
&\ \ \vdots \\
I_{jn} &= \left(\frac{j-1}{n} + p_{j1} + \ldots + p_{j,n-1}, j/n\right)
\end{aligned}$$

and

$$\begin{aligned}
\tilde{I}_{j1} &= \left(\frac{j-1}{n}, \frac{j-1}{n} + p_{1j}\right) \\
\tilde{I}_{j2} &= \left(\frac{j-1}{n} + p_{1j}, \frac{j-1}{n} + p_{1j} + p_{2j}\right) \quad 1 \le j \le n \\
&\ \ \vdots \\
\tilde{I}_{jn} &= \left(\frac{j-1}{n} + p_{1j} + \ldots + p_{n-1,j}, j/n\right)
\end{aligned}$$

Note that there is a coincidence of Lebesgue measure, $m(I_{jk}) = m(\tilde{I}_{kj})$. The invertible map $T$ which sends each $I_{jk}$ onto $\tilde{I}_{jk}$, $j, k = 1, \ldots, N$ by a translation is what we need (a pencil sketch will help to see this).

Observe that if each $p_{jk}$ is a dyadic rational, then $T$ is a permutation of the interval. If not, we approximate: given $\varepsilon > 0$, find dyadic rationals $\tilde{p}_{jk}$ such that

$$| p_{jk} - \tilde{p}_{jk} | < \varepsilon, \quad j, k = 1, \ldots, n \tag{2}$$

under the constraints (for eventual uniform marginals)

$$\sum_{j=1}^{n} \tilde{p}_{jk} = 1/n, \quad k = 1, \ldots, n \tag{3}$$

and

$$\sum_{k=1}^{n} \tilde{p}_{jk} = 1/n, \quad j = 1, \ldots, n. \tag{4}$$

This is a tricky problem to solve directly, but it yields to an appeal to Birkhoff's (1946) theorem (see, for example, Roberts and Varberg, 1973): if $P$ is the $n \times n$ matrix with $(j,k)$ entry $p_{j,k}$, then it can be written as a convex combination of permutation matrices

$$P = \sum_{i=1}^{n!} \theta_i \Pi^{(i)}, p_{jk} = [P]_{jk} = \sum_{i=1}^{n!} \theta_i [\Pi^{(i)}]_{jk} \ .$$

Now the $\theta_i$'s can be varied at will (subject to $\theta_i \geq 0$ and $\sum_{i=1}^{n!} \theta_i = 1$) and the constraints (3), (4) will still hold. Accordingly, find dyadic rationals $\tilde{\theta}_i$, $i = 1, \ldots, n$, so that $\tilde{\theta}_i \geq 0$ and $| \theta_i - \tilde{\theta}_i | < \varepsilon/n!$, $i = 1, \ldots, n!$. It follows that for each $j, k$

$$\tilde{p}_{jk} = \sum_{i=1}^{n!} \tilde{\theta}_i [\Pi^{(i)}]_{jk}$$

is within $\varepsilon$ of $p_{jk}$, and the marginal constraints are satisfied. With this done, we proceed as before to get a permutation $\tilde{T}$ which satisfies

$$| P(U \varepsilon I_j, V \varepsilon I_k) - P(U \varepsilon I_j, \tilde{T} U \varepsilon I_k) | < \varepsilon, \quad j, k = 1, \ldots, n. \tag{5}$$

PART TWO. We next want to arrange for a *cyclic* permutation. Note that if $\tilde{T}$ is a permutation of rank $N$, then it is also a permutation of rank $\hat{N}$ for any $\hat{N} > N$. We exploit this by choosing $\hat{N}$ very large and modifying $\tilde{T}$ on a small number of dyadic intervals of rank $\hat{N}$. More precisely, given $\varepsilon > 0$, we shall find a cyclic permutation $\tilde{\tilde{T}}$ such that

$$P(\tilde{T} U \varepsilon I_j, \tilde{\tilde{T}} \not\varepsilon I_j) + P(\tilde{T} U \not\varepsilon I_j, \tilde{\tilde{T}}\, U \varepsilon I_j) < \varepsilon, \quad j = 1, \ldots, n. \tag{6}$$

The details follow, but first we must modify $\tilde{T}$ so as to leave no dyadic interval invariant. Suppose that $\tilde{T}$ leaves $(0, k/2^N)$ invariant (i.e., it acts as the identity). Then redefine $\tilde{T}$ on this interval to be $\tilde{T}x = x + r \bmod k/2^N$ for some small dyadic rational $r$. A similar procedure can be done for other invariant intervals. It is clear that the new $\tilde{T}$ can be constructed so as to satisfy (6) in place of $\tilde{\tilde{T}}$ and with $\varepsilon$ replaced by, say, $2\varepsilon$.

Assume that this adjustment has been done and the modified function is $\tilde{T}$. For ease of notation, we retain $N$ as the rank of $\tilde{T}$. Suppose that $\tilde{T}$ has $p$ cycles of length $h_1, h_2, \ldots, h_p$ ($p \leq 2^{N-1}$ since each cycle contains at least two intervals). If we instead consider $\tilde{T}$ as of rank $\hat{N} > N$, then it has $p$ cycles of length $h_1 2^{\hat{N}-N}, h_2 2^{\hat{N}-N}, \ldots, h_p 2^{\hat{N}-N}$.

We snip these cycles and patch them together. In cycle #1, choose one of the $h_1 2^{\hat{N}-N}$ intervals to be the "out-interval" and its image under $\tilde{T}$ to be the "in-interval." Now define $\tilde{T}$ on in- and out-intervals so that it maps the out-interval of cycle $\#j$ onto the in-interval of cycle $\#j+1$, $j = 1, \ldots, p-1$, and the out-interval of cycle $\#p$ onto the in-interval of cycle #1. The resulting permutation $\approx{T}$ is cyclic and differs from $\tilde{T}$ on at most a set of measure $p2^{-\hat{N}}$. It follows that, with $\hat{N}$ suitably large, (6) holds as does (5) with $\tilde{T}$, $\varepsilon$ replaced by $\overset{\approx}{T}$, $2\varepsilon$ respectively.

Finally, we note that the required sequence $\{T_n\}$ is obtained by effecting the construction for $\overset{\approx}{T}$ for successive values $n = 1, 2, 4, 8, \ldots, 2^j, \ldots$ so that (5) and (6) are satisfied with $\varepsilon = \varepsilon_n = o(1/n^2)$ at each step.

**4. Discussion.** Especially in the case when $U$ and $V$ are independent, Theorem 1 calls for some explanation. How is it possible, after all, to arrive at a limiting pair of independent variables when at each stage the components of each approximating pair stand in an invertible, functional relationship to one another? We refer the reader to the discussion of Kimeldorf and Sampson (1978) and provide some other remarks here.

One ingredient which might be questioned is the mode of convergence. Convergence in distribution may be insufficiently stringent. If, for instance, convergence in the variation metric is substituted then the theorem obviously fails. On the other hand, so important a mode as convergence in distribution ought not to be easily dismissed. It is after all central to questions of sampling. In this context, given any sample $(u_i, v_i)$, $i = 1, \ldots, N$ in the square with the $u_i$'s distinct there is a $T \varepsilon \ \mathrm{T_{inv}}$ which fully "explains" the sample in the sense that $Tu_i = v_i$, $i = 1, \ldots, N$.

This leads to a related comment. Granted that the components of each approximating pair stand in a functional relationship to one another, the function itself may be so wild that in practical terms it is not feasible as a predictive device. That is, small errors in $U$ may lead to large errors in $T_n U$. To combat this, one may try to refine the process. But this seems to butt up against an inevitable problem of computational complexity. In this sense, $T_n$ cannot be used for prediction in any meaningful way. A possible rejoinder to these comments is that rather than using $T_n$ pointwise one should smooth it slightly so as to get, in effect, a smoothed regression function. We take up an aspect of regression in a later section.

A different, and provocative, comment has been offered by M. Klass. As we have said, the puzzling note is that the theorem makes an asymptotic statement in which there is an abrupt change of behavior *at* the limit. This type of phenomenon often occurs when one is dealing with a "large" space where there is flexibility for discontinuous behavior to occur. The key here is the roominess of $[0, 1]$. A smaller domain, or what is the same, random variables with "fewer" values illustrates the point. Take, for instance, $U$ and $V$ to be independent, $p = 1/2$, Bernoulli variables. Then the analogous theorem fails. It is clear that there is no sequence, convergent in distribution to $(U, V)$, of pairs $(U, T_n U)$ where $T_n : \{0, 1\} \to \{0, 1\}$. This is true more generally for any pair of discrete random variables which are not already in a relationship of functional dependence. The case of random variables $U$ and $V$, each uniformly distributed on $\{1, 2, \ldots, N\}$, but with otherwise arbitrary joint distribution, is interesting to consider. There is generally *no* approximating $(U, T_n U)$, as we have said, but we are always within a randomization of matching the joint distribution exactly. Consider that Birkhoff's theorem provides a set $\{\Pi_1, \ldots, \Pi_M\}$ of permutations of $\{1, \ldots, N\}$ and probabilities $\theta_1, \ldots, \theta_M$, $\theta_i > 0$, $\Sigma\theta_i = 1$ such that if $X$ is chosen at random in $\{1, 2, \ldots, N\}$ and if $J$ is chosen equal to $j$ with probability $\theta_j$, $j = 1, \ldots, M$, then the pair $(U, \Pi_J U)$ is distributed like $(U, V)$.

**5. General Form.** Theorem 1 can be generalized in two ways: relaxation of the condition of uniform marginals and, what we take up first, treatment of an arbitrary number of uniform random variables. This generalizes Theorem 1; now we take some shortcuts in the proof.

THEOREM 2. *Let $U_1, \ldots, U_d$ be a collection of uniform random variables. There is a sequence of vectors converging in distribution to $(U_1, \ldots, U_d)$ of the form $(T_1 U, T_2 U, \ldots, T_d U)$ where $U$ is uniform and $T_1, \ldots, T_d$ are cyclic permutations.*

PROOF. It is well-known that there are $S_1, S_2, \ldots, S_d \varepsilon$ T such that $(S_1 U, S_2 U, \ldots, S_d U)$ is distributed like $(U_1, \ldots, U_d)$ (see, for example, Billingsley, 1971, Theorem 3.2).

We proceed to approximate each $S_i$ by an invertible map. Consider $S_1$. Let $E_j = S_1^{-1}((j-1)/n, j/n)$, $j = 1, \ldots, n$. Proposition 2.1 provides a partition $\{F_1, \ldots, F_n\}$ (each $F_j$ a finite union of intervals) such that $m(E_j \triangle F_j) < \delta$. Create $\tilde{S}_1$ by taking a $1-1$ piecewise linear map of the interior of $F_j$ onto $((j-1)/n, j/n)$, $j = 1, \ldots, n$. If $\max\{\delta, 1/n\}$ is small, then $\tilde{S}_1 U$ is close to $S_1 U$ in probability. Since $\tilde{S}_1$ is invertible, Halmos (1956, p. 65) applies and ensures a cyclic permutation $T_1$ such that $T_1 U$ is close to $\tilde{S}_1 U$, and, from our construction, to $S_1 U$ in probability. Proceeding similarly for $S_2, S_3, \ldots, S_n$ completes the argument.

COROLLARY. *The approximating random vectors in the theorem may also be taken of the form $(U, \tilde{T}_2 U, \tilde{T}_3 U, \ldots, \tilde{T}_d U)$ where $\tilde{T}_2, \tilde{T}_3, \ldots, \tilde{T}_d$ are cyclic permutations.*

PROOF. $(T_1 U, T_2 U, \ldots, T_d U) \stackrel{d}{=} (U, T_2 T_1^{-1} U, \ldots, T_d T_1^{-1} U)$ and define $\hat{T}_j = T_j T_1^{-1}$, $j = 2, \ldots, d$. If $\hat{T}_j$ is a cyclic permutation, then set $\tilde{T}_j = \hat{T}_j$. Otherwise, let $\tilde{T}_j$ be a cyclic permutation such that $\tilde{T}_j U$ and $\hat{T}_j U$ are close in probability as argued above.

We turn next to our central result, which treats more general marginals. In view of the discussion in Section 4, we fix attention on continuous marginals. We recall a standard definition.

DEFINITION. The INVERSE of a distribution function $F$ is given by $F^{-1}(u) = \inf\{x \mid u \leq F(x)\}$.

THEOREM 3. *Let $X_1, \ldots, X_d$ be random variables with continuous marginal distributions $X_j \sim F_j$, $j = 1, \ldots, d$. There is a sequence of random vectors, converging in distribution to $(X_1, \ldots, X_d)$, of the form*

$$(F_1^{-1}(T_1 U), F_2^{-1}(T_2 U), \ldots, F_d^{-1}(T_d U))$$

*where $T_1, \ldots, T_d$ are cyclic permutations.*

PROOF. Set $U_1 = F_1(X_1), \ldots, U_d = F_d(X_d)$ as uniform random variables and note that

$$(X_1, \ldots, X_d) = (F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d)) \quad \text{a.s.}$$

The event $(F_1^{-1}(U_1) \leq x_1, \ldots, F_d^{-1}(U_d) \leq x_d)$ is the same (up to an event of probability zero) as $(U_1 \leq F_1(x_1), \ldots, U_d \leq F_d(x_d))$ whose image in $[0,1]^d$ is a continuity set of the joint measure of $(U_1, \ldots, U_d)$ (e.g., Billingsley, 1971, p. 3). It follows that if a sequence of random vectors of the form $(T_1 U, \ldots, T_d U)$ converges in distribution to $(U_1, \ldots, U_d)$, then the same holds for $(F_1^{-1}(T_1 U), \ldots, F_d^{-1}(T_d U))$ and $(X_1, \ldots, X_d) = (F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d))$. Together with Theorem 2 this concludes the proof.

The variable $U$ serves to parameterize each approximating vector. It can be removed by observing that if we set $Y_j = F_j^{-1}(T_j U)$, then $U = T_j F_j(Y_j)$ a.s. and hence $Y_k = F_k^{-1} \circ T_k \circ T_j F_j(Y_j)$ a.s. As in the corollary to Theorem 2, $T_k \circ T_j$ may be adjusted to be a cyclic permutation.

COROLLARY. *The approximating vectors in the theorem may be taken to be of the form*

$$(X_1, F_2^{-1} \circ T_2 \circ F_1(X_1), \ldots, F_d^{-1} \circ T_d \circ F_1(X_1))$$

*where $T_2, \ldots, T_d$ are cyclic permutations.*

**6. Remarks.** It is interesting to link up Theorem 3 with the extremal distributions of Hoeffding (1940) (see also, Frechet, 1951; Whitt, 1976; Tchen, 1980). They relate to the following question: Among all random vectors $(X, Y)$ with $X \sim F$ and $Y \sim G$, when is $\mathrm{Cov}(X, Y)$ smallest and largest? ($F$ and $G$ are assumed to yield finite second moments). The answers use $T_{\min} \varepsilon$ $\mathrm{T_{inv}}$ and $T_{\max} \varepsilon$ $\mathrm{T_{inv}}$ respectively:

$$(F^{-1}(U), G^{-1}(T_{\min} U)), \; T_{\min} u = 1 - u,$$

and

$$(F^{-1}(U), G^{-1}(T_{\max}U)),\ T_{\max}u = u,\ \ 0 \le u \le 1.$$

Thus Theorem 3 can be thought of as embedding the (degenerate) distributions of Hoeffding in a class dense in the collection of all distributions with the specified marginals.

An amusing note regarding Theorem 3 is a quick answer to the ancient classroom problem of displaying a random vector $(X, Y)$ which has normal marginals but is not bivariate normal (Vitale, 1978). It is enough to take $X \sim N(0,1)$ and $Y = N^{-1} \circ T \circ N(X)$ where $T \varepsilon$ T is *anything but* $T_{\min}$ or $T_{\max}$, say, $Tu = u + 1/2$ mod 1.

**7. Discretization.** While it is true, as discussed before, that Theorem 1 (and hence Theorem 3) has no general analogue for discrete marginal distributions, the result itself can be discretized.

THEOREM 4. *Suppose that $(X_1, \ldots, X_d)$ is a random vector with continuous marginals $F_1, \ldots, F_d$ respectively. Then there is a sequence of random vectors converging in distribution to $(X_1, \ldots, X_d)$ of the form*

$$(F_1^{-1}(\Pi_1(J)/n), \ldots, F_d^{-1}(\Pi_d(J)/n)$$

*where $n \ge 1$, $J$ is uniform on $\{1, 2, \ldots, n\}$ and $\Pi_1, \ldots, \Pi_d$ are cyclic permutations of $\{1, 2, \ldots, n\}$.*

PROOF. We adapt the argument of Theorem 3. With the uniform variable $U$ given, define $J$ via $J = j$ if $U \varepsilon ((j-1)/n, j/n)$. Also, let $\Pi_\ell(r) = s$ if $T_\ell$ takes the $r^{\text{th}}$ dyadic subinterval onto the $s^{\text{th}}$ dyadic subinterval. It follows that $((\Pi_1(J)/n), \ldots, (\Pi_d(J)/n))$ approximates $(U_1, \ldots, U_d)$ in distribution and this suffices.

COROLLARY. *The approximating vectors may be taken of the form*

$$(F_1^{-1}(J/n),\ F_2^{-1}(\Pi_2(J)/n), \ldots, F_d^{-1}(\Pi_d(J)/n)$$

*with $n$, $J$, $\Pi_2, \ldots, \Pi_d$ as given in the theorem.*

An interesting special case is that of a Markov chain.

COROLLARY. *Let $X_1, X_2, \ldots$ be a discrete time Markov chain in equilibrium with stationary continuous density $F$. Then, for any $d$, $(X_1, \ldots, X_d)$ can be approximated in distribution by*

$$(F^{-1}(J/n),\ F^{-1}(\Pi(J)/n),\ F^{-1}(\Pi^2(J)/n), \ldots, F^{-1}(\Pi^{d-1}(J)/n))$$

*with $n$, $J$ as in the theorem and $\Pi$ a cyclic permutation of $\{1, 2, \ldots, n\}$.*

These results suggest a practical method for generating random vectors with arbitrary dependence structure among components. Aside from evaluation of inverse distribution functions and a single uniform variate, all that is needed are the appropriate permutations of $\{1, 2, \ldots, n\}$. How these permutations are determined seems to be an interesting question.

**8. The Penalty for Using Linear Regression.** In an interesting study, Fairley, Pearl, and Verducci (1987) look at the penalty incurred using various forms of constrained regression. In particular, the expected unexplained variation left from linear prediction can be partitioned into an "intrinsic variation" component and an "extra-linear variation" component

$$(7) \quad E(Y - \rho X)^2 = E(Y - \phi(X))^2 + E(\phi(X) - \rho X)^2, \; \phi(X) = E[Y \mid X].$$

They point out that it is of interest to bound the first, $\eta = E(Y - \phi(X))^2$, and note that it can be seen arbitrarily close to 0 under certain conditions. Their technique uses data (i.e., point-mass) distributions as approximants. The machinery we have developed provides a stronger form of their result.

THEOREM 5. *Suppose that a bivariate distribution is given with marginals $F$ and $G$ and correlation $\rho$. Then there is another bivariate distribution with marginals uniformly close to those of the first, correlation arbitrarily close to $\rho$, and for which $\eta = 0$.*

PROOF. First slightly smooth the given distribution so as to have continuous marginals and then truncate it to a large rectangle. By the corollary to Theorem 3, the resulting distribution has a degenerate distribution close to it and for which $\eta = 0$.

**9. Maximally Dependent Random Variables.** Suppose that $X_1, X_2, \ldots, X_n$ are independent, identically distributed random variables with common distribution $F$. They are said to be *maximally dependent* if

$$(8) \qquad P(M_n > x) = \min\{1, n(1 - F(x))\}$$

where $M_n = \max\{X_1, \ldots, X_n\}$. The expression on the right in (8) is clearly the largest conceivable value for $P(M_n > x)$. The study of maximally dependent random variables was inaugurated in Lai and Robbins (1976) and continued in Lai and Robbins (1978). Elaborations and generalizations appear in Tchen (1980) and Rüschendorf (1981).

An important question is the existence of a *canonical* sequence; that is, $X_1, X_2, \ldots, (X_i \sim F \forall i)$ such that, for each $n$, $X_1, \ldots, X_n$ are maximally dependent. Lai and Robbins (1978) establish existence (and non-uniqueness) in the case when $F$ is continuous (see Tchen, 1980, for the case $F$ discontinuous). They argue that an application of the inverse distribution function renders it sufficient to consider the case of uniform variables and then they provide a direct construction. We show that it is possible to construct a different canonical sequence in which the joint distributions completely degenerate in the sense of our foregoing discussion.

PROPOSITION 9.1. *There is a canonical sequence $U_1, U_2, \ldots$ of uniform random variables such that for each $i \geq 2$*

$$U_i = T_i U_1 \text{ where } T_i \varepsilon \ \mathrm{T_{inv}} \text{ is piecewise linear.} \tag{9}$$

PROOF. We use Lai and Robbins' (1978) observation that a collection $U_1, \ldots, U_n$ of uniform random variables is maximally dependent if and only if $M_n$ is uniformly distributed on $(1 - 1/n, 1)$.

We proceed by induction. The case $n = 1$ is trivial. Assume that $U_1, \ldots, U_n$ have been constructed in the form (9). Note that

$$B = \{u \varepsilon [0,1] \mid \max\{u, T_2 u, \ldots, T_n u\} < 1 - 1/(n+1)\}$$

is a union of intervals and that, by Lai and Robbins' criterion, $m(B) = \frac{1}{n+1}$. Construct $U_{n+1}$ by mapping the interior of $B$ in a piecewise linear manner onto $\left(1 - \frac{1}{n+1}, 1\right)$. Elsewhere define $U_{n+1}$ to be piecewise linear so as to be finally $1-1$ and measure-preserving. It follows that $M_{n+1}$ is uniform on $\left(1 - \frac{1}{n+1}, 1\right)$ and hence $U_1, \ldots, U_{n+1}$ is a maximally dependent collection.

## REFERENCES


BILLINGSLEY, P. (1971). Weak Convergence of Measures: Applications in Probability. Regional Conference Series in Applied Mathematics No. 5, SIAM, Philadelphia.

BIRKHOFF, G. (1946). Tres observaciones sobre et lineal. *Univ. Nac. Tucumán Rev. Ser. A* **5** 147–150.

BROWN, J.R. (1966). Approximation theorems for Markov operators. *Pac. J. Math.* **16** 13–23.

FAIRLEY, D., PEARL, D.K., and VERDUCCI, J.S. (1987). The penalty for assuming that a monotone regression is linear. *Ann. Statist.* **15** 443–448.

FRÉCHET, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon Sect. A* **14** 53–77.

GARSIA, A.M. (1976). Combinatorial inequalities and smoothness of functions. *Bull. Amer. Math. Soc.* **82** 157–170.

HALMOS, P. (1956). *Lectures on Ergodic Theory.* Chelsea, New York.

HARDY, G.H., LITTLEWOOD, J.E., and PÓLYA, G. (1952). *Inequalities*, 2nd ed. Cambridge University Press, Cambridge.

HOEFFDING, W. (1940). Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts fur Angewandte Mathematik der Universität Berlin* **5** 179–233.

HOLBROOK, J.A.R. (1981). Stochastic independence and space-filling curves. *Amer. Math. Monthly* **88** 426–432.

KIMELDORF, G. and SAMPSON, A.R. (1975). Uniform representation of bivariate distributions. *Comm. Statist.* **4** 617–627.

KIMELDORF, G. and SAMPSON, A.R. (1978). Monotone dependence. *Ann. Statist.* **6** 895–903.

LAI, T.L. and ROBBINS, H. (1976). Maximally dependent random variables. *Proc. Nat. Acad. Sci. USA* **73** 286–288.

LAI, T.L. and ROBBINS, H. (1978). A class of dependent random variables and their maxima. *Z. Wahrsch. verw. Geb.* **42** 89–111.

LANCASTER, H.O. (1963). Correlation and complete dependence of random variables. *Ann. Math. Stat.* **34** 1315–1321.

ROBERTS, A.W. and VARBERG, D.E. (1973). *Convex Functions.* Academic Press, New York.

RÜSCHENDORF, L. (1981). Sharpness of Fréchet-bounds. *Z. Warsch. verw. Geb.* **57** 293–302.

TCHEN, A. (1980). Inequalities for distributions with given marginals. *Ann. Statist.* **8** 814–827.

VITALE, R.A. (1978). Joint vs. individual normality. *Math. Mag.* **51** 123.

VITALE, R.A. (1979). Regression with given marginals. *Ann. Statist.* **7** 653–658.

VITALE, R.A. and PIPKIN, A.C. (1976). Conditions on the regression function when both variables are uniformly distributed. *Ann. Prob.* **4** 869–873.

WHITT, W. (1976). Bivariate distributions with given marginals. *Ann. Statist.* **4** 1280–1289.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
STORRS, CT 06269