The Annals of Statistics
1976, Vol. 4, No. 6, 1280-1289

BIVARIATE DISTRIBUTIONS WITH GIVEN MARGINALS[1]

BY WARD WHITT

Yale University

Bivariate distributions with minimum and maximum correlations for given marginal distributions are characterized. Such extremal distributions were first introduced by Hoeffding (1940) and Fréchet (1951). Several proofs are outlined including ones based on rearrangement theorems. The effect of convolution on correlation is also studied. Convolution makes arbitrary correlations less extreme while convolution of identical measures on R^2 makes extreme correlations more extreme. Extreme correlations have applications in data analysis and variance reduction in Monte Carlo studies, especially in the technique of antithetic variates.

1. Introduction. Let $\Pi \equiv \Pi(F, G)$ be the set of all cdf's (cumulative distribution functions) H on R^2 having F and G as marginal cdf's, where F and G have finite positive variances. Within Π there are cdf's H^* and H_* discovered by Hoeffding (1940) and Fréchet (1951) which have maximum and minimum correlation. These extremal cdf's and the associated extreme correlations are of interest in data analysis to place the sample correlation in perspective and in variance reduction to deliberately create positive or negative correlation. This paper characterizes these extremal distributions. Section 2 briefly reviews a proof due to Hoeffding and presents two other proofs based on two versions of the rearrangement theorem of Hardy, Littlewood, and Pólya (1952) which predates Hoeffding (1940) and Fréchet (1951). One advantage of the rearrangement theorem in the case of bivariate data is that it clearly indicates a simple algorithm for constructing extremal distributions; see Lemma 2.6. The situation with data is also very instructive because sets of data obviously correspond to very simple probability measures, namely, probability measures which attach masses of size n^{-1} to each of n points. As is the case here, proofs are often much easier in this setting and so are of pedagogical value. Furthermore, such data distributions are easily seen to be dense in the space of all probability measures in the usual topology of weak convergence, so results for data distributions often extend to arbitrary measures by continuity; see the first proof of Theorems 2.1 and 2.5. Section 2 is concluded by showing the connection to work by Strassen (1965). Section 3 contains examples of marginal distributions with extreme correlations arbitrarily close to zero, while Section 4 investigates convolution and correlation. Convolution makes arbitrary correlations less extreme while convolution

Received September 1972; revised March 1976.

[1] Partially supported by National Science Foundation Grants GK-27866 and GK-38149.

AMS 1970 *subject classifications.* Primary 62E10; Secondary 62E25, 62H05, 62H20.

Key words and phrases. Bivariate distributions, bivariate distributions with given marginals, extreme correlation, nearest random variables, rearrangement theorems, variance reduction, Monte Carlo, antithetic variates, generating random variables.

of identical measures on R^2 makes extreme correlations more extreme. Finally, Section 5 gives a few additional results for the set of all n-dimensional distributions with given marginals which should be useful for more complicated variance reduction problems.

2. The extremal distributions. Let H^* and H_* be the cdf's in $\Pi(F, G)$ with maximum and minimum correlations. Existence of such cdf's is contained in Theorem 2.1. Let $[x]^+ = \max\{0, x\}$ and $x \wedge y = \min\{x, y\}$.

THEOREM 2.1 (Hoeffding). *In* $\Pi(F, G)$

$$H^*(x, y) = F(x) \wedge G(y) \qquad \textit{and} \qquad H_*(x, y) = [F(x) + G(y) - 1]^+ \qquad \textit{for all} \quad (x, y) \in R^2 .$$

PROOF. The theorem is an immediate consequence of Lemmas 2.2 and 2.3 below, each of which is easily proved.

LEMMA 2.2 (Hoeffding). *For any* $H \in \Pi(F, G)$ *and all* $(x, y) \in R^2$, $H_*(x, y) \leqq H(x, y) \leqq H^*(x, y)$.

PROOF. See page 31 of Mardia (1970).

LEMMA 2.3 (Hoeffding). *Let the random vector* (X, Y) *have* cdf H *with marginals* F *and* G. *Then*

$$EXY - EXEY = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [H(x, y) - F(x)G(y)]\, dx\, dy .$$

PROOF. See Lemma 2, page 1139 of Lehmann (1966).

The extremal distributions can also be characterized in another way, based on the following familiar lemma. Let $F^{-1}(y) = \inf\{x : F(x) > y\}$.

LEMMA 2.4. *Let* X *be real-valued random variable with* cdf F *and let* U *be a uniform random variable with* cdf G. *Then*

(i) $F^{-1}(U)$ *has* cdf F;

and

(ii) *if* F *is continuous*, $F(X)$ *has* cdf G.

THEOREM 2.5. *For any* F, G *with finite positive variances,*

$$[F^{-1}(U), G^{-1}(U)] \quad \textit{has cdf} \quad H^*$$

and

$$[F^{-1}(U), G^{-1}(1 - U)] \quad \textit{has cdf} \quad H_* .$$

In addition to previously mentioned sources, discussion and references related to extremal cdf's appears on pages 355 and 359 of Hall (1969) and page 22 of Johnson and Kotz (1972). Theorems 2.1 and 2.5 are equivalent; it is easy to derive each from the other. We give two more proofs based on rearrangement theorems. First consider the subset of probability measures on R^2 in which each measure assigns masses of n^{-1} to each of n points, (x_i, y_i), $1 \leqq i \leqq n$, $n \geqq 1$. We call such probability measures *data distributions* because they naturally arise when considering a set of n data points. The possible bivariate data distributions with given marginals obviously corresponds to the permutations of $\{y_i\}$ for a given ordering of $\{x_i\}$. In this setting the problem is to pair the x_i with the y_j in order to maximize or minimize the sum of products $\sum_{i=1}^n x_i y_i$. We can apply the classical rearrangement theorem for finite sets, Theorem 368 of Hardy, Littlewood and Pólya (1952).

LEMMA 2.6 (Hardy, Littlewood and Pólya). *The sum of products* $\sum_{i=1}^n x_i y_i$ *is a maximum when* $\{x_i\}$ *and* $\{y_i\}$ *are both increasing and a minimum when one is increasing and the other is decreasing.*

Let $\Delta_n(F, G)$ be the set of all data distributions of size n on R^2 with marginal cdf's F and G. Obviously, F and G must then be cdf's of one-dimensional data distributions of size n. It is well known that $\Delta_n(F, G)$ coincides with the set of extreme points in $\Pi(F, G)$, cf. Hammersley and Mauldon (1956). If the x_i and y_j are all distinct, then $\Delta_n(F, G)$ corresponds exactly to the set of $n \times n$ permutation matrices, while $\Pi(F, G)$ corresponds to the set of all $n \times n$ doubly stochastic matrices. Consequently, Lemma 2.6 implies Theorems 2.1 and 2.5 when F and G are cdf's of data distributions.

For the second proof below of Theorems 2.1 and 2.5, we will want

LEMMA 2.7. *For any* cdf H *on* R^n *and uniform random variable* U, *there exists a measurable function* $\mathbf{X} = (X_1, \cdots, X_n): [0, 1] \to R^n$ *such that* $\mathbf{X}(U)$ *has* cdf H.

PROOF. One early source is Section 23 of Lévy (1937). It is also possible to let ϕ be a Borel isomorphic map of R^n onto R, page 7 of Parthasarathy (1967). Let P_H be the measure on R^n associated with H. Let F be the cdf on R associated with the measure $P_H \phi^{-1}$. Then let $\mathbf{X} = \phi^{-1} \circ F^{-1}$. By Lemma 2.4, $F^{-1}(U)$ has cdf F. Consequently, $\phi^{-1}(F^{-1}(U))$ has cdf H. A still different approach is via Theorem 3.1.1 of Skorohod (1956) cf. Theorem 3.2 of Billingsley (1971).

FIRST PROOF OF THEOREMS 2.1 AND 2.5. The idea is to consider arbitrary distributions as limits of data distributions. For arbitrary F and G with finite second moments γ_F^2 and γ_G^2, it is easy to construct one-dimensional data distributions F_n and G_n of size n with finite second moments $\gamma_{F_n}^2$ and $\gamma_{G_n}^2$ such that $\gamma_{F_n}^2 \uparrow \gamma_F^2$, $\gamma_{G_n}^2 \uparrow \gamma_G^2$, $|F_n(t) - F(t)| \leqq n^{-1}$, and $|G_n(t) - G(t)| \leqq n^{-1}$, $-\infty < t < \infty$. For these data distributions, by Lemma 2.6, $\rho(H_{*n}) \leqq \rho(H_n) \leqq \rho(H_n^*)$ for all $H_n \in \Delta_n(F_n, G_n)$, where ρ is the correlation coefficient. Obviously $H_{*n} \to H_*$ and $H_n^* \to H^*$ as $n \to \infty$ because $F_n \to F$ and $G_n \to G$. An arbitrary sequence $\{H_n\}$ with $H_n \in \Delta_n(F_n, G_n)$ need not converge, but it is not difficult to show that any $H \in \Pi(F, G)$ is the uniform limit of some sequence $\{H_n\}$ with $H_n \in \Delta_n(F_n, G_n)$, where F_n and G_n are the one-dimensional data distributions converging to F and G. More generally, data distributions are dense in the space of all probability measures. The proof is completed by noting that $\rho(H_n) \to \rho(H)$ if $H_n \to H$ in R^2 with $H_n \in \Pi(F_n, G_n)$ and $H \in \Pi(F, G)$. (Additional details are contained in an earlier version of this paper.)

SECOND PROOF OF THEOREMS 2.1 AND 2.5. The idea is to apply the rearrangement theorem for functions on [0, 1], Theorem 378 of Hardy, Littlewood and Pólya (1952), instead of the rearrangement theorem for finite sets. Let f' and g' denote increasing rearrangements and f'' and g'' decreasing rearrangements of f and g on [0, 1] as defined on page 276 of Hardy, Littlewood and Pólya (1952). Then

$$\int_0^1 f'(x)g''(x)\,dx \leqq \int_0^1 f(x)g(x)\,dx \leqq \int_0^1 f'(x)g'(x)\,dx\,.$$

For arbitrary cdf $H \in \Pi(F, G)$ there exists $(X, Y): [0, 1] \to R^2$ such that $[X(U), Y(U)]$ has cdf H. Thus, we can let $f(x) = X(x)$ and $g(x) = Y(x)$ so that $EXY = \int_0^1 f(x)g(x)$. The increasing and decreasing rearrangements of f and g are obviously just $f'(x) = F^{-1}(x)$, $f''(x) = F^{-1}(1 - x)$, $g'(x) = G^{-1}(x)$ and $g''(x) = G^{-1}(1 - x)$. This completes the proof.

It is worth noting that there is a more general rearrangement theorem from which both Theorem 2.1 and Lemma 2.2 follow as corollaries. In the following version due to Lorentz (1953) there are compactness and continuity conditions which can be relaxed. Extensions, where these conditions are relaxed, and applications are discussed by Day (1972) and Tchen (1975). Superadditivity has been studied extensively by A. F. Veinott, Jr. and D. Topkis; see Topkis (1968) and forthcoming papers.

THEOREM 2.8 (Lorentz). *If F and G have compact support and $\phi: R^2 \to R$ is continuous and superadditive*, i.e.,

$$\phi(x_1, y_1) + \phi(x_2, y_2) \leqq \phi(x_1, y_2) + \phi(x_2, y_1)$$

whenever $x_1 \leqq x_2$ and $y_1 \leqq y_2$, then

$$\int \phi\,dH_* \gtreqqless \int \phi\,dH \leqq \int \phi\,dH^*$$

for all $H \in \Pi(F, G)$.

PROOF. The standard proof of Lemma 2.6 applies with each superadditive ϕ. For the general case, follow the first proof of Theorems 2.1 and 2.5.

Theorem 2.5 and Lemma 2.7 clearly indicate an alternate way to express the extreme correlation result. Let $L^2 \equiv L^2[0, 1]$ be the usual Hilbert space of square integrable real-valued functions on [0, 1] with Lebesgue measure, page 111 of Royden (1968). Let $L^2(F, G)$ be the subset of ordered pairs (X, Y) of functions in L^2 such that X has cdf F and Y has cdf G. Note that $(F^{-1}, G^{-1}) \in L^2(F, G)$. Let $\lambda(t) = 1 - t$, $0 \leqq t \leqq 1$.

COROLLARY 2.9. *In the setting above,*

$$\inf\{\|X - Y\|_2 : (X, Y) \in L^2(F, G)\} = \|F^{-1} - G^{-1}\|_2$$

and

$$\sup\{\|X - Y\|_2 : (X, Y) \in L^2(F, G)\} = \|F^{-1} - G^{-1} \circ \lambda\|_2\,.$$

We can also consider one of the random variables in L^2 as fixed. Let $L^2(F)$ be the subset of X in L^2 such that X has cdf F. From Corollary 2.9 and Lemma 2.4(ii), we also have

COROLLARY 2.10. *Let $X \in L^2(F)$, where F is continuous. Then*

$$\inf\{||X - Y||_2 : Y \in L^2(G)\} = ||X - G^{-1}(F(X))||_2$$

and

$$\sup\{||X - Y||_2 : Y \in L^2(G)\} = ||X - G^{-1}(\lambda(F(X)))||_2 .$$

Corollary 2.9 closely parallels Section 6 of Strassen (1965), Section 2 of Dudley (1968), and Schay (1974); there the distance is the usual distance associated with convergence in probability instead of the L^2 distance. For comparison, we state a special case here. Let $V \equiv V[0, 1]$ be the space of all real-valued random variables on $[0, 1]$ with Lebesgue measure μ. Let d be the usual metric on V associated with convergence in probability, defined by

$$d(X, Y) = \inf\{\varepsilon > 0 : \mu(|X - Y| > \varepsilon) < \varepsilon\} .$$

Let m be the Prohorov metric on the space of all probability measures on R, defined by

$$m(F, G) = \inf\{\varepsilon > 0 : F(A) \leqq G(A^\varepsilon) + \varepsilon, \text{ for all closed } A \subseteqq R\} ,$$

where $A^\varepsilon = \{x \in R : |x - y| < \varepsilon \text{ for some } y \in A\}$ and F, G are regarded as the measures associated with the cdf's, cf. page 1564 of Dudley (1968). Let $V(F, G)$ be the subset of ordered pairs of functions (X, Y) in V such that X has cdf F and Y has cdf G.

THEOREM 2.11 (Strassen). *In the setting above,*

$$\sup\{d(X, Y) : (X, Y) \in V(F, G)] = m(F, G) .$$

It is easy to see, in the spirit of Lemma 2.6, that there is a simple proof of Theorem 2.11 for data distributions.

3. Examples. If X is a random variable with a distribution symmetric about its mean μ, then (X, X) and $(X, 2\mu - X)$ have bivariate distributions with common marginals and correlations ± 1. On the other hand, extreme correlations arbitrarily close to 0 can be achieved. For example, let

$$\begin{aligned} P\{X = k\} = P\{Y = k\} &= (n-1)n^{-1} , \qquad & k &= -1 \\ &= n^{-1} , & k &= (n-1) . \end{aligned}$$

Then $\rho^* = 1$ and $\rho_* = -(n-1)^{-1}$, where ρ^* and ρ_* are the maximum and minimum correlation coefficients for bivariate distributions with these marginals. Another example shows that ρ^* and ρ_* can both be arbitrarily close to 0. (The independent case shows that $-1 \leqq \rho_* < 0 < \rho^* \leqq 1$.) Let $P\{X = +1\} = P\{X = -1\} = 2^{-1}$, $P\{Y = +1\} = P\{Y = -1\} = (2n)^{-1}$, and $P\{Y = 0\} = 1 - n^{-1}$. Then $\rho^* = -\rho_* = n^{-\frac{1}{2}}$.

4. Correlation and convolution.

THEOREM 4.1. *If $S_2 = X_1 + X_2$, where X_1 and X_2 are independent random vectors* *in* R^2, *then*

$$\rho(S_2) = a\rho(X_1) + b\rho(X_2)\,,$$

where $0 \leqq a, b \leqq 1$ *and* $a + b \leqq 1$. *Moreover, if* $\operatorname{Var}(X_{11}) \operatorname{Var}(X_{22}) = \operatorname{Var}(X_{12}) \operatorname{Var}(X_{21})$, *then* $a + b = 1$.

PROOF. Using the independence, we obtain the desired relation with $c = [\operatorname{Var}(X_{11}) \operatorname{Var}(X_{12}) + \operatorname{Var}(X_{11}) \operatorname{Var}(X_{22}) + \operatorname{Var}(X_{21}) \operatorname{Var}(X_{12}) + \operatorname{Var}(X_{21}) \operatorname{Var}(X_{22})]$, $a = (\operatorname{Var}(X_{11}) \operatorname{Var}(X_{12})/c)^{\frac{1}{2}}$ and $b = (\operatorname{Var}(X_{21}) \operatorname{Var}(X_{22})/c)^{\frac{1}{2}}$. Obviously $0 \leqq a, b \leqq 1$. We obtain $a + b \leqq 1$ because $(x_1 x_2)^{\frac{1}{2}} \leqq (x_1 + x_2)/2$ for all $x_1, x_2 > 0$, with equality holding if and only if $x_1 = x_2$. The extra condition on the variances provides equality.

COROLLARY 4.2. *Let* $S_n = X_1 + \cdots + X_n$ *where* $X_1, \cdots, X_n$ *are mutually independent random vectors in* R^2.

(i) *If* $\operatorname{Var}(X_{11}) = \operatorname{Var}(X_{21}) = \cdots = \operatorname{Var}(X_{n1})$ *and* $\operatorname{Var}(X_{12}) = \cdots = \operatorname{Var}(X_{n2})$, *then*

$$\rho(S_n) = n^{-1} \textstyle\sum_{i=1}^{n} \rho(X_i)\,.$$

(ii) *If* $X_1, \cdots, X_n$ *have a common distribution, then*

$$\rho(S_n) = \rho(X_1)\,.$$

THEOREM 4.3. *Let* $S_n = X_1 + \cdots + X_n$ *where* $X_1, \cdots, X_n$ *are* i.i.d. *random vectors in* R^2, *then* $\rho^*(S_n) \geqq \rho^*(X_1)$ *and* $\rho_*(S_n) \leqq \rho_*(X_1)$.

PROOF. Let X_1' have the same marginals as X_1 but an extreme distribution so that $\rho(X_1') = \rho^*(X_1)$. Then

$$\rho^*(S_n) = \rho^*(S_n') \geqq \rho(S_n') = \rho(X_1') = \rho^*(X_1)$$

by Corollary 4.2(ii). Similarly for ρ_*.

Theorem 4.3 shows for example that the minimum correlation for random vectors with identical gamma marginals of parameter α is nonincreasing as α runs through values of $t2^n$, $n \geqq 1$, where the gamma density of parameter α is $f_\alpha(x) = x^{\alpha-1}e^{-x}/\Gamma(\alpha)$, $x \geqq 0$. Is the minimum correlation strictly decreasing from 0 to -1 as α goes from 0 to infinity?

5. Variance reduction. Since $\operatorname{Var}(X_1 + X_2) = \operatorname{Var}(X_1) + \operatorname{Var}(X_2) + 2\operatorname{Cov}(X_1, X_2)$, $\operatorname{Var}(X_1 + X_2)$ can be minimized without changing the marginal distributions of X_1 and X_2 and thus without changing $E(X_1 + X_2)$ by making $\operatorname{Cov}(X_1, X_2)$ as small as possible while leaving the marginal distributions unchanged. More generally, $\operatorname{Var}[f(X_1 + \cdots + X_n)]$ can usually be reduced without changing the marginal distributions of $X_1, \cdots, X_n$ by introducing negative or positive correlations between X_i and X_j, $1 \leqq i, j \leqq n$, while leaving the marginal distributions unchanged. This technique of variance reduction has been very useful in Monte Carlo studies, cf. Fishman (1972), Hammersley and Handscomb (1964), Hammersley and Mauldon (1956), Handscomb (1958), and Page (1965). The optimal variance reduction scheme for the sum of two variables is of course contained in Section 2, with Theorem 2.5 being especially important because it provides a natural way to actually generate random variables with the desired joint distribution. However, general optimal variance reduction schemes for the sum of more than two variables or for more complicated functions than the sum have not yet been discovered. There is a general principle developed by Hammersley and Mauldon (1956) and Handscomb (1958), page 61 of Hammersley and Handscomb (1964), which says that in order to minimize Var $(X_1 + \cdots + X_n)$, it suffices to generate all the random variables from a single uniform random variable in a rather special way. This principle deserves some comment. First, Lemma 2.7 shows that the distinction between stochastic dependencies and functional dependencies made there is unnecessary; all stochastic dependencies among the marginal distributions can be represented as functional dependencies. Second, representations in terms of a single uniform random variable are not limited to the particular optimization for the variance of the sum considered there. In addition to Lemma 2.7, we have

THEOREM 5.1. *Let $F_1, \cdots, F_n$ be n real* cdf's *and let U be a random variable uniformly distributed on* $[0, 1]$. *If $H \in \Pi(F_1, \cdots, F_n)$, then there exist measurable measure-preserving functions $\phi_1, \cdots, \phi_n$ mapping* $[0, 1]$ *into itself such that* $[F_1^{-1}(\phi_1(U)), \cdots, F_n^{-1}(\phi_n(U))]$ *has* cdf *H. If F_j is discrete, then ϕ_j can be a Borel isomorphism of* $[0, 1]$. *If H is a data distribution of size n, then each ϕ_j can have derivative* 1 *everywhere except at the* $(n + 1)$ *points* kn^{-1}, $0 \leqq k \leqq n$.

The key is the following result, which is Theorem 1 of Sklar (1973) and Theorem 2 of Schweizer and Sklar (1973). Let U also denote the cdf of the uniform distribution on $[0, 1]$.

THEOREM 5.2. *For any $H \in \Pi(F_1, \cdots, F_n)$, there is a $C \in \Pi(U_1, \cdots, U_n)$ such that $C(F_1(x), \cdots, F_n(x)) = H(x_1, \cdots, x_n)$ for all $(x_1, \cdots, x_n) \in R^n$.*

PROOF OF THEOREM 5.2. Let C be defined on Ran $(F_1) \times \cdots \times$ Ran (F_n) by $C(s_1, \cdots, s_n) = H(F_1^{-1}(s_1), \cdots, F_n^{-1}(s_n))$, where Ran (F) is the range of F. Let C be defined on the closure by continuity. Since $x_i \leqq F_i^{-1}(s_i)$ if and only if $F_i(x) \leqq s_i$ for s_i in $\overline{\text{Ran}(F_i)} = \sup(F_i^{-1})$, the support of F_i^{-1}, $C(F_1(x_1), \cdots, F_n(x_n)) = H(x_1, \cdots, x_n)$ for all $(x_1, \cdots, x_n) \in R^2$. The proof is completed by extending C to $[0, 1]^n$ so that C has uniform marginals. Consider the bivariate case. For any $(a, b) \in [0, 1]^2$, let a_1 and a_2 (b_1 and b_2) be the greatest and least elements of $\overline{\text{Ran } F_1}$ ($\overline{\text{Ran } F_2}$) satisfying $a_1 \leqq a \leqq a_2$ ($b_1 \leqq b \leqq b_2$). Set

$$\begin{aligned} \lambda_1 &= (a - a_1)/(a_2 - a_1)\,, && a_1 < a_2 \\ &= 1\,, && a_1 = a_2 \end{aligned}$$

and

$$\begin{aligned} \mu_1 &= (b - b_1)/(b_2 - b_1)\,, && b_1 < b_2 \\ &= 1\,, && b_1 = b_2\,. \end{aligned}$$

The extension is obtained by letting

$$C(a, b) = (1 - \lambda_1)(1 - \mu_1)C(a_1, b_1) + (1 - \lambda_1)\mu_1 C(a_1, b_2) + \lambda_1(1 - \mu_1)C(a_2, b_1) + \lambda_1\mu_1 C(a_2, b_2) .$$

For additional details, see Schweizer and Sklar (1973).

The conclusion for discrete cdf's is based on

LEMMA 5.3. *Let X and Y be two countably-valued random variables defined on* $[0, 1]$ *with Lebesgue measure. If X has the same distribution as Y, then there exists a measurable isomorphism ϕ of* $[0, 1]$ *such that* $Y(t) = X(\phi(t))$, $0 \leqq t \leqq 1$.

PROOF OF LEMMA 5.3. Let N be an arbitrary uncountable measurable subset of $[0, 1]$ of Lebesgue measure 0. Let $\{a_n\}$ be the sequence of values assumed by X and Y with positive measure. Let $A_n = \{t \in [0, 1] - N: X(t) = a_n\}$ and $B_n = \{t \in [0, 1] - N: Y(t) = a_n\}$. Let ϕ be defined on $[0, 1]$ so that $\phi(B_n) = A_n$, $n \geqq 1$, and $\phi([0, 1] - \bigcup_{n=1}^\infty B_n) = [0, 1] - \bigcup_{n=1}^\infty A_n$. Each of these sets is an uncountable measurable subset of $[0, 1]$. Hence, a one-to-one map ϕ of $[0, 1]$ onto itself can be defined so that the appropriate subsets are related and both ϕ and its inverse are measurable, cf. page 7 of Parthasarathy (1967).

PROOF OF THEOREM 5.1. For any $H \in \Pi(F_1, \cdots, F_n)$, let C be the cdf determined by Theorem 5.2. Suppose $(X_1, \cdots, X_n)$ is a random vector with cdf C. Then $P(X_1 \leqq F_1(x_1), \cdots, X_n \leqq F_n(x_n)) = H(x_1, \cdots, x_n)$ for all $(x_1, \cdots, x_n) \in R^n$ by Lemma 5.2, but also $P(X_1 \leqq F_1(x_1), \cdots, X_n \leqq F_n(x_n)) = P(F_1^{-1}(X_1) \leqq x_1, \cdots, F_n^{-1}(X_n) \leqq x_n)$ for all $(x_1, \cdots, x_n) \in R^n$ by the usual argument used to prove Lemma 2.4(i). In particular, when $(x_1, \cdots, x_n)$ is in the support of H, $u_i \leqq F_i(x_i)$ for all i if and only if $F_i^{-1}(u_i) \leqq x_i$ for all i, which implies equality for $(x_1, \cdots, x_n)$ in the support of H and thus everywhere. Now note that C is the cdf of a nonatomic measure on $[0, 1]^n$. Hence, there is a measurable measure-preserving map $\mathbf{X} = (X_1, \cdots, X_n)$ of $[0, 1]$ into $[0, 1]^n$ such that $\mathbf{X}(U)$ has cdf C, cf. page 327 of Royden (1968). Let the desired ϕ_j be X_j, $1 \leqq j \leqq n$. If F_j is discrete, Lemma 5.3 can be used because $F_j^{-1}(U)$ has the same distribution as $F_j^{-1}(X_j(U))$. This means there is a Borel isomorphism ϕ_j of $[0, 1]$ such that $F_j^{-1}(\phi_j(t)) = F_j^{-1}(X_j(t))$, $0 \leqq t \leqq 1$, so we can replace X_j with ϕ_j. Multivariate data distributiohs are easy because the ϕ_j can correspond to permutations of the subintervals $[kn^{-1}, (k + 1)n^{-1})$ in $[0, 1)$.

REMARKS. Lemma 5.3 does not extend to arbitrary random variables as can be seen by considering the following simple example (courtesy of S. Kakutani). Let $X(t) = t$, $0 \leqq t \leqq 1$, and

$$Y(t) = 2t , \qquad 0 \leqq t < \tfrac{1}{2}$$
$$= 2t - 1 , \qquad \tfrac{1}{2} \leqq t \leqq 1 .$$

Acknowledgments. My interest in this topic stems from discussions with G. Fishman. I am grateful to A. A. Balkema, W. J. Hall, L. Rafsky, A. Sklar, A. Tchen and the referee for providing references.

REFERENCES

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

[2] CHONG, K. M. and RICE, N. M. (1971). Equimeasurable rearrangements of functions. Queen's Papers in Pure and Applied Mathematics No. 28, Queen's University, Kingston, Ontario.

[3] DAY, P. W. (1972). Rearrangement inequalities. *Canad. J. Math.* **24** 930–943.

[4] DUDLEY, R. M. (1968). Distances of probability measures and random variables. *Ann. Math. Statist.* **39** 1563–1572.

[5] FISHMAN, G. S. (1972). Variance reduction in simulation studies. *J. Statist. Comp. and Simulation* **1** 173–182.

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

[7] HALL, W. J. (1969). On characterizing dependence in joint distributions. *Essays in Probability and Statistics*, Bose et al., ed. Univ. of North Carolina Press, Chapel Hill.

[8] HALTON, J. H. (1970). A retrospective and prospective survey of the Monte Carlo method. *SIAM Rev.* **12** 1–63.

[9] HAMMERSLEY, J. M. and HANDSCOMB, D. C. (1964). *Monte Carlo Methods.* Methuen, London.

[10] HAMMERSLEY, J. M. and MAULDON, J. G. (1956). General principles of antithetic variates. *Proc. Camb. Philos. Soc.* **52** 476–481.

[11] HANDSCOMB, D. C. (1958). Proof of the antithetic-variates theorem for $n \geqq 2$. *Proc. Camb. Philos. Soc.* **54** 300–301.

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

[13] HOEFFDING, W. (1940). Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universitat Berlin* **5** 179–233.

[14] JOHNSON, N. L. and KOTZ, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* Wiley, New York.

[15] LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–1153.

[16] LÉVY, P. (1973). *Théorie de l'Addition des Variables Aleatoires.* Gauthier-Villars, Paris.

[17] LORENTZ, G. G. (1953). An inequality for rearrangements. *Amer. Math. Monthly* **60** 176–179.

[18] MARDIA, K. V. (1970). *Families of Bivariate Distributions.* Griffin's Statistical Monographs, Hofner, Darien, Conn.

[19] PAGE, E. S. (1965). On Monte Carlo methods in congestion problems: II. Simulation of queueing systems. *Operations Res.* **13** 300–305.

[20] PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces.* Academic Press, New York.

[21] ROYDEN, H. L. (1968). *Real Analysis*, 2nd ed. Macmillan, London.

[22] SCHAY, G. (1974). Nearest random variables with given distributions. *Ann. Probability* **2** 163–166.

[23] SCHWEIZER, B. and SKLAR, A. (1974). Operations on distribution functions not derivable from operations on random variables. *Studia Math.* **52** 43–52.

[24] SKLAR, A. (1973). Random variables, joint distributions, and copulas. *Kybernetika* **9** 449–460.

[25] SKOROHOD, A. V. (1956). Limit theorems for stochastic processes. *Theor. Probability Appl.* **1** 261–290.

[26] STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

[27] TCHEN, A. H. T. (1975). Exact inequalities for $\int \phi(x, y)\, dH(x, y)$ when H has given marginals. Stanford Univ.

[28] Topkis, D. (1968). *Ordered Optimal Solutions.* Ph. D. dissertation, Dept. of Operations Research, Stanford Univ.

School of Organization and Management
Yale University
56 Hillhouse Avenue
New Haven, Connecticut 06520