MUTUAL INFORMATION AND MAXIMAL CORRELATION AS MEASURES OF DEPENDENCE[1]

By C. B. Bell

San Diego State College and University of Washington

1. Introduction and summary. Rényi [19] gives a set of seven postulates which a measure of dependence for a pair of random variables should satisfy. Of the dependence measures considered by Rényi, only Gebelein's [5] maximal correlation, S_P, satisfies all seven postulates. Kramer [10] in considering the uncertainty principle in Fourier analysis [11] generalizes the Gebelein maximal correlation to the case of arbitrary pairs of σ-algebras; and asks whether this generalization is equivalent to Shannon's mutual information, C_P, [4, 9, 21] for pairs of σ-algebras—equivalent in the sense of preserving order.

The object of this note is to compare S_P and the two normalizations C'_P and C''_P, of C_P, as dependence measures for strictly positive probability spaces (which are necessarily generated by random variables). It is found that for such spaces with the proper finiteness restrictions

(a) (Thm 5.1) $0 \leqq S_P, C'_P, C''_P \leqq 1$;

(b) (Thm 5.2) $S_P = 0$ iff $C'_P = 0$ iff $C''_P = 0$ iff the random variables are independent;

(c) (Thm 5.4) $S_P = 1$ if the two generated algebras have a nontrivial intersection (the conditions are equivalent for finite algebras); $C'_P = 1$ iff one of the random variables is a function of the other; and $C''_P = 1$ iff the random variables are functions of each other; and, consequently,

(d) (Thm 5.5) there exist probability spaces for which the dependence measures are not equivalent.

The paper is divided into six sections. Section 1 contains the introduction and summary. Section 2 introduces the terminology, notation and preliminaries. Section 3 treats S_P and the Rényi postulates.

In Section 4, the basic Shannon-Feinstein-Khinchin mutual information is extended to strictly positive measure spaces, not necessarily finite. The comparison of the dependence measures and postulate modifications are given in Section 5. Finally, in Section 6 some extensions and open problems are mentioned.

2. Terminology, notation and preliminaries. Throughout the paper $(Z, \mathcal{S}, P)$ will be an arbitrary but fixed probability space unless there is a statement to the contrary; $\emptyset$ will denote the empty subset of Z; $\mathfrak{D}$, $\mathcal{A}$ and $\mathcal{B}$ will always denote sub-σ-algebras of $\mathcal{S}$; $\mathcal{S}(K)$ will denote the least σ-algebra containing the class K of subsets of Z; and X, Y, U and V will denote nontrivial real-valued

Received July 12, 1961; revised October 30, 1961.

[1] The work was supported by the National Science Foundation through grant NSF G12973. Communication with H. P. Kramer and A. Rényi was helpful.

random variables, i.e., functions measurable with respect to (w.r.t.) $\mathcal{S}$ and not constant with probability 1.

Extending the notation of Kramer [10], one defines

$$\mathfrak{A}_Y = \text{the } \sigma\text{-algebra generated by } Y;$$

$$\mathfrak{L}_2(\mathfrak{D}) = \{X : X \text{ is measurable w.r.t. } \mathfrak{D} \text{ and } \int X^2\, dP < \infty\};$$

$$\mathfrak{M}_n(\mathfrak{D}) = \{X \varepsilon \mathfrak{L}_2(\mathfrak{D}) : \int X\, dP = 0 \text{ and } \int X^2\, dP = 1\};$$

and

$$S_P(\mathfrak{A}, \mathfrak{B}) = \sup |R(X, Y)|,$$

the Kramer-Gebelein maximal correlation of $\mathfrak{A}$ and $\mathfrak{B}$ w.r.t. P, where $R(X, Y)$ is the correlation coefficient, and the supremum is taken over $X \varepsilon \mathfrak{L}_2(\mathfrak{A})$ and $Y \varepsilon \mathfrak{L}_2(\mathfrak{B})$.

Following Marczewski [17], one defines (for fixed $(Z, \mathcal{S}, P)$), an atom, α, of σ-algebra $\mathfrak{A} \subset \mathcal{S}$, to be an element of $\mathfrak{A}$ such that $P(\alpha) > 0$; and such that whenever $A \varepsilon \mathfrak{A}$ and $A \subset \alpha$, then $P(A) = 0$ or $P(A) = P(\alpha)$. $\{\alpha_i\}$, $\{\beta_j\}$ and $\{\gamma_k\}$ will denote, respectively, the atoms of the σ-algebras $\mathfrak{A}$, $\mathfrak{B}$ and $\mathfrak{D}$.

A σ-algebra $\mathfrak{A}$ is called "purely atomic w.r.t. P" if $\sum_i P(\alpha_i) = 1$. It is called "strictly positive w.r.t. P" if, whenever $A \varepsilon \mathfrak{A}$ and $P(A) = 0$, then $A = \emptyset$.

It is known, e.g., [17] that

LEMMA 2.1. *If* $\mathfrak{D}$ *is strictly positive* w.r.t. P, *then* $\mathfrak{D}$ *is purely atomic* w.r.t. P.

Since strictly positive probability spaces are generated by random variables, the atoms of $\mathfrak{A}_X$ are sets of the form $\{\alpha_i\} = \{X = x_i\}$. Consequently, treating the random variables is equivalent to treating the generated σ-algebras and their atoms. This will be done in the sequel.

Consistent with the notation of Feinstein, one defines for purely atomic σ-algebras $\mathfrak{A}$ and $\mathfrak{B}$,

$$H_P(\mathfrak{A}) = -\sum_i P(\alpha_i) \log P(\alpha_i)$$ (where through the paper 2 is the base of the logarithms);

$$H_P(\mathfrak{A}, \mathfrak{B}) = -\sum_{i,j} P(\alpha_i \cap \beta_j) \log P(\alpha_i \cap \beta_j);$$ and

$$C_P(\mathfrak{A}, \mathfrak{B}) = -\sum P(\alpha_i \cap \beta_j) \log \{P(\alpha_i \cap \beta_j)/P(\alpha_i) \cdot P(\beta_j)\},$$

the mutual information of $\mathfrak{A}$ and $\mathfrak{B}$ w.r.t. P.

Since, as is easily seen, $0 \leqq S_P \leqq 1$, it will facilitate the comparison of S_P and C_P if one normalizes C_P to give it the same bounds. To this end one defines

$$C'_P(\mathfrak{A}, \mathfrak{B}) = C_P(\mathfrak{A}, \mathfrak{B})/\min [H_P(\mathfrak{A}), H_P(\mathfrak{B})];$$ and

$$C''_P(\mathfrak{A}, \mathfrak{B}) = C_P(\mathfrak{A}, \mathfrak{B})/\max [H_P(\mathfrak{A}), H_P(\mathfrak{B})],$$

whenever $H_P(\mathfrak{A})$ and $H_P(\mathfrak{B})$ are finite. (For other normalizations see [22, 23].)

Finally, one needs to introduce two different types of independence. A pair $\mathfrak{A}$, $\mathfrak{B}$ of σ-algebras is said to be set independent if $A \cap B \neq 0$ for all $\emptyset \neq A \varepsilon \mathfrak{A}$

and $\emptyset \neq B \varepsilon \mathcal{B}$. $\mathcal{A}$ and $\mathcal{B}$ are independent w.r.t. P if $P(A \cap B) = P(A) \cdot P(B)$ for all $A \varepsilon \mathcal{A}$ and $B \varepsilon \mathcal{B}$.

An immediate extension of the results of several authors, e.g., [1, 2, 15, 16], yields

LEMMA 2.2.

(i) *If* $\mathcal{A}$ *and* $\mathcal{B}$ *are set independent, then* $\mathcal{A} \cap \mathcal{B} = \{\emptyset, Z\}$, *the trivial algebra;*

(ii) $\mathcal{A}$ *and* $\mathcal{B}$ *are set independent* iff *there exists* P_0 *on* $\mathcal{S}(\mathcal{A} \cup \mathcal{B})$, *the least* σ-*algebra containing* $\mathcal{A}$ *and* $\mathcal{B}$, *such that* $\mathcal{A}$ *and* $\mathcal{B}$ *are independent* w.r.t. P_0.

With these preliminaries one can proceed to investigate the basic properties of S_P, C'_P and C''_P.

3. Maximal correlation and the Rényi postulates. Rényi [19] gives the following set of seven properties which should be satisfied by a measure, $\delta(\cdot, \cdot)$, of dependence of two random variables X, Y on a given probability space.

(A) $\delta(X, Y)$ is defined for any pair X, Y neither of which is constant with probability 1.

(B) $\delta(X, Y) = \delta(Y, X)$.

(C) $0 \leqq \delta(X, Y) \leqq 1$.

(D) $\delta(X, Y) = 0$ iff X and Y are independent.

(E) $\delta(X, Y) = 1$ if either $X = g(Y)$ or $Y = f(X)$, where $g(\cdot)$ and $f(\cdot)$ are Borel-measurable functions.

(F) If the Borel-measurable functions $f(\cdot)$ and $g(\cdot)$ map the real axis in a one-to-one way into itself, then $\delta(f(X), g(Y)) = \delta(X, Y)$.

(G) If the joint distribution of X and Y is normal, then $\delta(X, Y) = |R(X, Y)|$, where $R(X, Y)$ is the correlation coefficient of X and Y.

[Note: It may seem natural to replace the "if" in (E) by "if and only if." However, Rényi views this requirement as too restrictive. This point will be discussed further at the end of Section 5.]

In [19] Rényi considers six dependence measures—the correlation coefficient; three correlation ratios; the mean square contingency; and Gebelein's maximal correlation. Of these dependence measures, only the last satisfies all seven postulates.

The object of this paper is to compare the generalization, S_P, of Gebelein's maximal correlation with two normalizations of Shannon's mutual information for the purpose of establishing a lack of equivalence, and of suggesting some revisions of the set of postulates. To this end one easily proves the following two lemmas.

LEMMA 3.1.

(i) $S_P(\mathcal{A}, \mathcal{B}) = \sup R(X, Y)$, *where the supremum is taken over all* $X \varepsilon \mathfrak{M}_n(\mathcal{A})$ *and* $Y \varepsilon \mathfrak{M}_n(\mathcal{B})$; *and*

(ii) $0 \leqq S_P(\mathcal{A}, \mathcal{B}) \leqq 1$.

LEMMA 3.2. *If* $\mathcal{A}$ *and* $\mathcal{B}$ *are both finite, then* $S_P(\mathcal{A}, \mathcal{B}) = r_0$ iff *there exists* $X_0 \varepsilon \mathfrak{M}_n(\mathcal{A})$ *and* $Y_0 \varepsilon \mathfrak{M}_n(\mathcal{B})$ *such that* $R(X_0, Y_0) = r_0$.

This latter result will be used in showing that for finite algebras $S_P = 1$ iff $\mathcal{A}$

and $\mathcal{B}$ have a nontrivial intersection; and hence that S_P is not equivalent to the normalized measures of mutual information considered below.

4. Mutual information. Shannon [21], Feinstein [4], and Khinchin [9], among others, have demonstrated the basic properties of $H_P(\cdot)$, $H_P(\cdot\,,\cdot)$ and $C_P(\cdot\,,\cdot)$ for finite algebras. In general, their results can be immediately extended to arbitrary strictly positive measure spaces, if one excludes cases in which H_P is not finite. (See Carleson [3].) Those results needed in the sequel are as follows:

LEMMA 4.1. *If* $\mathcal{S}(\mathcal{A} \cup \mathcal{B})$ *is strictly positive* w.r.t. P, *then*

(i) $H_P(\mathcal{A}) \geqq 0$, *with equality* iff $\mathcal{A}$ *is the trivial* σ*-algebra*;

(ii) *if* $\mathcal{A}$ *is finite, then* $H_P(\mathcal{A}) < \infty$;

(iii) $H_P(\mathcal{A}, \mathcal{B}) = H_P(\mathcal{B}, \mathcal{A}) = H_P(\mathcal{S}(\mathcal{A} \cup \mathcal{B}))$;

(iv) *if* $H_P(\mathcal{A}) < \infty$, *then* $H_P(\mathcal{A}) \leqq H_P(\mathcal{A}, \mathcal{B})$ *with equality* iff $\mathcal{A}$ *contains* $\mathcal{B}$;

(v) *if* $H_P(\mathcal{A}), H_P(\mathcal{B}) < \infty$, *then* $H_P(\mathcal{A}) + H_P(\mathcal{B}) \geqq H_P(\mathcal{A}, \mathcal{B})$, *with equality* iff $\mathcal{A}$ *and* $\mathcal{B}$ *are independent* w.r.t. P; *and*

(vi) $H_P(\mathcal{A}), H_P(\mathcal{B}) < \infty$ iff $H_P(\mathcal{A}, \mathcal{B}) < \infty$.

Using these results and the definitions of Section 2, one deduces the validity of the following for the mutual information, C_P.

LEMMA 4.2. *If* $\mathcal{S}(\mathcal{A} \cup \mathcal{B})$ *is strictly positive* w.r.t. P *and* $H_P(\mathcal{A}, \mathcal{B}) < \infty$, *then*

(i) $C_P(\mathcal{A}, \mathcal{B}) = H_P(\mathcal{A}) + H_P(\mathcal{B}) - H_P(\mathcal{A}, \mathcal{B})$;

(ii) $C_P(\mathcal{A}, \mathcal{B}) = 0$ iff $\mathcal{A}$ *and* $\mathcal{B}$ *are independent* w.r.t. P;

(iii) $C_P(\mathcal{A}, \mathcal{B}) \leqq \min\,[H_P(\mathcal{A}), H_P(\mathcal{B})]$, *with equality* iff $\mathcal{A} \subset \mathcal{B}$ *or* $\mathcal{B} \subset \mathcal{A}$; *and*

(iv) $0 \leqq C''_P(\mathcal{A}, \mathcal{B}) \leqq C'_P(\mathcal{A}, \mathcal{B}) \leqq 1$, *with* $C''_P(\mathcal{A}, \mathcal{B}) = C'_P(\mathcal{A}, \mathcal{B})$ iff $H_P(\mathcal{A}) = H_P(\mathcal{B})$.

Finally for the set independence case, one can employ Lemmas 2.2, 4.1 and 4.2 to conclude,

LEMMA 4.3. *If* $\mathcal{S}(\mathcal{A} \cup \mathcal{B})$ *is strictly positive* w.r.t. P; $H_P(\mathcal{A}, \mathcal{B}) < \infty$; *and* $\mathcal{A}$ *and* $\mathcal{B}$ *are set independent, then*

(i) $\max\,[H_P(\mathcal{A}), H_P(\mathcal{B})] < H_P(\mathcal{A}, \mathcal{B})$;

(ii) $\min\,[H_P(\mathcal{A}), H_P(\mathcal{B})] > C_P(\mathcal{A}, \mathcal{B})$; *and*

(iii) *there exists a probability measure* P_0 *on* $\mathcal{S}(\mathcal{A}, \mathcal{B})$ *such that* $H_{P_0}(\mathcal{A}, \mathcal{B}) = H_{P_0}(\mathcal{A}) + H_{P_0}(\mathcal{B})$, *and, hence,* $C_{P_0}(\mathcal{A}, \mathcal{B}) = 0$.

It is now possible to demonstrate the similarities and dissimilarities of S_P, C'_P and C''_P.

5. Comparison of the dependence measures. From Lemmas 3.1 and 4.2 one concludes

THEOREM 5.1. *If* $\mathcal{S}(\mathcal{A} \cup \mathcal{B})$ *is strictly positive* w.r.t. P *and* $H_P(\mathcal{A}, \mathcal{B}) < \infty$, *then* $0 \leqq S_P(\mathcal{A}, \mathcal{B}), C'_P(\mathcal{A}, \mathcal{B}), C''_P(\mathcal{A}, \mathcal{B}) \leqq 1$.

Further, from Lemma 4.2 and the definitions one can establish

THEOREM 5.2. *If* $(\mathcal{A} \cup \mathcal{B})$ *is strictly positive* w.r.t. P *and* $H_P(\mathcal{A}, \mathcal{B}) < \infty$, *then the following four conditions are equivalent*:

(i) $\mathcal{A}$ *and* $\mathcal{B}$ *are independent* w.r.t. P,

(ii) $S_P(\mathcal{A}, \mathcal{B}) = 0$;

(iii) $C'_P(\mathfrak{A}, \mathfrak{B}) = 0$;
(iv) $C''_P(\mathfrak{A}, \mathfrak{B}) = 0$.

PROOF. The equivalence of conditions (i), (iii) and (iv) follows immediately from the definitions and Lemma 4.2 (ii).

Further, if $\mathfrak{A}$ and $\mathfrak{B}$ are independent, then whenever $X \varepsilon \mathfrak{L}_2(\mathfrak{A})$ and $Y \varepsilon \mathfrak{L}_2(\mathfrak{B})$, $R(X, Y) = 0$. Hence, $S_P(\mathfrak{A}, \mathfrak{B}) = 0$. It remains, therefore, to prove that (ii) implies (i). Clearly, if $S_P(\mathfrak{A}, \mathfrak{B}) = 0$, then $R(X, Y) = 0$ for all $X \varepsilon \mathfrak{L}_2(\mathfrak{A})$ and $Y \varepsilon \mathfrak{L}_2(\mathfrak{B})$. If $A \neq Z$ and $B \neq Z$ are arbitrary non void elements of $\mathfrak{A}$ and $\mathfrak{B}$, respectively, let $I_A(\cdot)$ be the indicator function of A, i.e., I_A assumes the value 1 for $z \varepsilon A$ and 0 otherwise, and let $I_B(\cdot)$ be the indicator function of B. Now $I_A \varepsilon \mathfrak{L}_2(\mathfrak{A})$, $I_B \varepsilon \mathfrak{L}_2(\mathfrak{B})$, and $R(I_A, I_B) = K^{-1}[P(A \cap B) - P(A)P(B)]$, where $K = \{[P(A)][1 - P(A)][P(B)][1 - P(B)]\}^{\frac{1}{2}} > 0$. Consequently, $0 = R(I_A, I_B)$ and $0 = P(A \cap B) - P(A) \cdot P(B)$, or, equivalently, $P(A \cap B) = P(A) \cdot P(B)$ for arbitrary nontrivial $A \varepsilon \mathfrak{A}$ and $B \varepsilon \mathfrak{B}$. Therefore, $\mathfrak{A}$ and $\mathfrak{B}$ are independent w.r.t. P.

A corresponding but weaker result is available in the case of set independence.

THEOREM 5.3. *It is assumed that* $\mathcal{S}(\mathfrak{A} \cup \mathfrak{B})$ *is strictly positive* w.r.t. P *and* $H_P(\mathfrak{A}, \mathfrak{B}) < \infty$.

(i) *if* $\mathfrak{A}$ *and* $\mathfrak{B}$ *are set independent, then* $C'_P(\mathfrak{A}, \mathfrak{B})$ *and* $C''_P(\mathfrak{A}, \mathfrak{B}) < 1$;
(ii) *if* $\mathfrak{A}$ *and* $\mathfrak{B}$ *are set independent and finite, then* $\mathcal{S}_P(\mathfrak{A}, \mathfrak{B}) < 1$;
(iii) $\mathfrak{A}$ *and* $\mathfrak{B}$ *are set independent* iff *there exists a probability measure* P_0 *and* $\mathcal{S}(\mathfrak{A}, \mathfrak{B})$ *such that* $S_{P_0}(\mathfrak{A}, \mathfrak{B}) = C_{P_0}(\mathfrak{A}, \mathfrak{B}) = C''_{P_0}(\mathfrak{A}, \mathfrak{B}) = 0$.

PROOF. (i) follows immediately from Lemma 4.3 and the definitions. (ii) follows from Theorem 5.4. (iv) which is proved below. (iii) is an immediate consequence of Lemma 2.2 and Theorem 5.2.

The three theorems above illustrate the similarities of S_P, C'_P and C''_P. The next theorem demonstrates that the three dependence measures have fundamental differences at the maximum value.

THEOREM 5.4. *If* $\mathcal{S}(\mathfrak{A} \cup \mathfrak{B})$ *is strictly positive* w.r.t. P *and* $H(\mathfrak{A}, \mathfrak{B}) < \infty$, *then*

(i) $C'_P(\mathfrak{A}, \mathfrak{B}) = 1$ iff $\mathfrak{A} \subset \mathfrak{B}$ *or* $\mathfrak{B} \subset \mathfrak{A}$;
(ii) $C''_P(\mathfrak{A}, \mathfrak{B}) = 1$ iff $\mathfrak{A} = \mathfrak{B}$;
(iii) $S_P(\mathfrak{A}, \mathfrak{B}) = 1$ *if* $\mathfrak{A} \cap \mathfrak{B} \neq \{\emptyset, Z\}$, *the trivial algebra*;
(iv) *if* $\mathfrak{A}$ *and* $\mathfrak{B}$ *are finite*, $S_P(\mathfrak{A}, \mathfrak{B}) = 1$ iff $\mathfrak{A} \cap \mathfrak{B} \neq \{\emptyset, Z\}$; *and always*
(v) $C''_P(\mathfrak{A}, \mathfrak{B}) = 1 \Rightarrow C'_P(\mathfrak{A}, \mathfrak{B}) = 1 \Rightarrow S_P(\mathfrak{A}, \mathfrak{B}) = 1$.

PROOF.

(i) follows from Lemma 4.2 (iii) and the definition of C'_P.

(ii) In view of Lemma 4.1 (iv) and Lemma 4.2 (iii), (iv), $\mathfrak{A} = \mathfrak{B}$ iff both of the following conditions are satisfied:

$$H_P(\mathfrak{A}) = H_P(\mathfrak{B}) = H_P(\mathfrak{A}, \mathfrak{B}) \qquad \text{and}$$

$$C_P(\mathfrak{A}, \mathfrak{B}) = \max [H_P(\mathfrak{A}), H_P(\mathfrak{B})] = \min [H_P(\mathfrak{A}), H_P(\mathfrak{B})].$$

These latter two conditions are equivalent to the assertion $C''_P(\mathfrak{A}, \mathfrak{B}) = 1$.

(iii) If $D \varepsilon \mathfrak{A} \cap \mathfrak{B}$ and $D \neq \emptyset$ or Z, then $I_D \varepsilon \mathfrak{L}_2(\mathfrak{A}) \cap \mathfrak{L}_2(\mathfrak{B})$ and $R(I_D, I_D) = 1$. Therefore, $1 \geqq S_P(\mathfrak{A}, \mathfrak{B}) \geqq R(I_D, I_D) = 1$.

(iv) If $\mathcal{A}$ and $\mathcal{B}$ are finite, then Lemma 3.2 guarantees that whenever $S_P(\mathcal{A}, \mathcal{B}) = 1$, there exist $X_0 \varepsilon \mathfrak{M}_n(\mathcal{A})$ and $Y_0 \varepsilon \mathfrak{M}_n(\mathcal{B})$ such that $R(X_0, Y_0) = 1$. However, if $R(X_0, Y_0) = 1$, there exist constants a and b such that $Y_0 = aX_0 + b$. But since $X_0 \varepsilon \mathfrak{M}_n(\mathcal{A})$ and $Y_0 \varepsilon \mathfrak{M}_n(\mathcal{B})$, $X_0 = Y_0$ and neither is a constant function. Consequently, there exists nontrivial $D \varepsilon \mathcal{A} \cap \mathcal{B}$, i.e., $0 < P(D) < 1$. This result and (iii) above establish the validity of (iv).

(v) follows immediately from (i), (ii) and (iii).

The preceding theorem indicates the method of constructing three examples, which demonstrate that the measures are not equivalent in the sense of preserving order. These examples are given below.

THEOREM 5.5. *There exist probability spaces* $(Z_i, \mathcal{S}_i, P_i)$, *and* σ-*algebras* $\mathcal{A}_i$, $\mathcal{B}_i$, *and* $\mathcal{G}_i$, *contained in* $S_i (i = 1, 2, 3)$ *such that*

(i) $S_{P_1}(\mathcal{A}_1, \mathcal{B}_1) > S_{P_1}(\mathcal{A}_1, \mathcal{G}_1)$, *but* $C'_{P_1}(\mathcal{A}_1, \mathcal{B}_1) < C'_{P_1}(\mathcal{A}_1, \mathcal{G}_1)$;

(ii) $S_{P_2}(\mathcal{A}_2, \mathcal{B}_2) > S_{P_2}(\mathcal{A}_2, \mathcal{G}_2)$, *but* $C''_{P_2}(\mathcal{A}_2, \mathcal{B}_2) < C''_{P_2}(\mathcal{A}_2, \mathcal{G}_2)$; *and*

(iii) $C''_{P_3}(\mathcal{A}_3, \mathcal{B}_3) > C''_{P_3}(\mathcal{A}_3, \mathcal{G}_3)$, *but* $C'_{P_3}(\mathcal{A}_3, \mathcal{B}_3) < C'_{P_3}(\mathcal{A}_3, \mathcal{G}_3)$.

Consequently, no two of the measures are equivalent.

PROOF.

(iii) Let Z_3 be the unit square; $\mathcal{S}_3$, the Borel sets on the unit square;

$$A_1 = \{0 \leqq x < \tfrac{1}{2}\}; \quad A_2 = \{\tfrac{1}{2} \leqq x \leqq 1\}; \quad B_1 = \{\tfrac{1}{2} \leqq y \leqq 1\}; \quad B_2 = \{0 \leqq y < \tfrac{1}{2}\};$$

$$G_i = \{(i-1)/2n \leqq x < i/2n\} \quad \text{for} \quad 1 \leqq i \leqq 2n - 1; \quad G_{2n} = \{(2n-1)/2n \leqq x \leqq 1\};$$

$\mathcal{A}_3 = \mathcal{S}(\{A_i\})$; $\mathcal{B}_3 = \mathcal{S}(\{B_j\})$; $\mathcal{G}_3 = \mathcal{S}(\{G_k\})$; and define P_3 such that $P_3(A_1) = P_3(A_2) = \frac{1}{2}$; $P_3(B_1) = \frac{3}{8}$; $P_3(B_2) = \frac{5}{8}$; $P_3(G_i) = 1/2n$ for $1 \leqq i \leqq 2n$.

(a) Clearly, $H_{P_3}(\mathcal{A}_3) = \log 2 = 1 > H_{P_3}(\mathcal{B}_3) = (\frac{3}{8}) \log (\frac{8}{3}) + (\frac{5}{8}) \log (\frac{8}{5})$; and $H_{P_3}(\mathcal{G}_3) = \log 2n = H_{P_3}(\mathcal{A}_3, \mathcal{G}_3)$, since $\mathcal{A}_3 \subset \mathcal{G}_3$.

(b) Further, $H_{P_3}(\mathcal{A}_3, \mathcal{B}_3) = (\frac{1}{8}) \log 8 + 2(\frac{1}{4}) \log (\frac{8}{3}) = \frac{5}{2} - (\frac{3}{8}) \log 3$.

(c) Also, since $H_{P_3}(\mathcal{A}_3, \mathcal{G}_3) = H_{P_3}(\mathcal{G}_3)$, $C'_{P_3}(\mathcal{A}_3, \mathcal{G}_3) = 1$; while $C'_{P_3}(\mathcal{A}_3, \mathcal{B}_3) < 1$, since $H_{P_3}(\mathcal{A}_3) < H_{P_3}(\mathcal{A}_3, \mathcal{B}_3) < H_{P_3}(\mathcal{A}_3) + H_{P_3}(\mathcal{B}_3)$.

(d) From Lemma 4.2 one concludes that $C''_{P_3}(\mathcal{A}_3, \mathcal{B}_3) > 0$.

(e) Finally, since $C''_{P_3}(\mathcal{A}_3, \mathcal{G}_3) = \{H_{P_3}(\mathcal{A}_3)\}/\{H_{P_3}(\mathcal{G}_3)\} = 1/(1 + \log n)$, there exists an n (and, hence, a σ-algebra $\mathcal{G}_3$) such that $C''_{P_3}(\mathcal{A}_3, \mathcal{G}_3) < C''_{P_3}(\mathcal{A}_3, \mathcal{B}_3)$.

(i)–(ii). To complete the proof of the theorem it will be sufficient to exhibit an example in which $C'_P(\mathcal{A}, \mathcal{B}) = C''_P(\mathcal{A}, \mathcal{B})$; $C'_P(\mathcal{A}, \mathcal{G}) = C''_P(\mathcal{A}, \mathcal{G})$; and (i) holds. To this end, let $(Z, \mathcal{S}, P)$ be the Lebesgue measure space of the unit square; $B_1 = \{0 \leqq x < \frac{1}{10}\}$; $B_2 = \{\frac{1}{10} \leqq x < \frac{4}{10}\}$; $B_3 = \{\frac{4}{10} \leqq x < \frac{7}{10}\}$; $B_4 = \{\frac{7}{10} \leqq x < 1\}$; $A_1 = B_1$; $A_2 = \{\frac{1}{10} \leqq x \leqq 1; 0 \leqq y < \frac{1}{3}\}$; $A_3 = \{\frac{1}{10} \leqq x \leqq 1; \frac{1}{3} < y \leqq \frac{2}{3}\}$; $A_4 = \{\frac{1}{10} \leqq x \leqq 1; \frac{2}{3} < y \leqq 1\}$; $G_1 = \{\frac{9}{10} \leqq x \leqq 1\}$; $G_2 = \{\frac{6}{10} \leqq x < \frac{9}{10}\}$; $G_3 = \{\frac{3}{10} \leqq x < \frac{6}{10}\}$; $G_4 = \{0 \leqq x < \frac{3}{10}\}$; $\mathcal{A} = \mathcal{S}(\{A_i\})$; $\mathcal{B} = \mathcal{S}(\{B_j\})$, and $\mathcal{G} = \mathcal{S}(\{G_k\})$.

(a) Clearly, $P(A_1) = P(G_1) = P(B_1) = \frac{1}{10}$; and $P(B_i) = P(A_i) = P(G_i)$, for $i = 2, 3, 4$, and consequently, $H_P(\mathfrak{A}) = H_P(\mathfrak{B}) = H_P(\mathfrak{G}) = (\frac{1}{10}) \log 10 + 3(\frac{3}{10}) \log (\frac{10}{3}) = \log 10 - (\frac{9}{10}) \log 3$.

(b) Since $\mathfrak{S}(\mathfrak{A} \cup \mathfrak{B})$ has ten atoms each of probability $\frac{1}{10}$, $H_P(\mathfrak{A}, \mathfrak{B}) = \log 10$.

(c) Examining the atoms of $\mathfrak{S}(\mathfrak{B} \cup \mathfrak{G})$, one finds that $H_P(\mathfrak{B}, \mathfrak{G}) = 4\ (\frac{1}{10}) \log 10 + 3\ (\frac{2}{10}) \log (\frac{10}{2}) = \log 10 - (\frac{3}{5}) < H_P(\mathfrak{A}, \mathfrak{B})$.

(d) From the definitions and (a), it is clear that $C'_P(\mathfrak{A}, \mathfrak{B}) = C''_P(\mathfrak{A}, \mathfrak{B})$; $C'_P(\mathfrak{A}, \mathfrak{G}) = C''_P(\mathfrak{A}, \mathfrak{G})$; and (i) holds.

(e) From Theorem 5.4, one concludes, that $S_P(\mathfrak{A}, \mathfrak{B}) = 1 > S_P(\mathfrak{A}, \mathfrak{G})$.

(f) Further, from (c) above it is seen that,

$$C'_P(\mathfrak{A}, \mathfrak{B}) = \frac{H_P(\mathfrak{A}) + H_P(\mathfrak{B}) - H_P(\mathfrak{A}, \mathfrak{B})}{\min\,[H_P(\mathfrak{A}), H_P(\mathfrak{B})]}$$

$$< \frac{H_P(\mathfrak{A}) + H_P(\mathfrak{G}) - H_P(\mathfrak{A}, \mathfrak{G})}{[\min H_P(\mathfrak{A}), H_P(\mathfrak{B})]} = C'_P(\mathfrak{A}, \mathfrak{G})$$

Although the dependence measures are quite different, as is illustrated by Theorems 5.4 and 5.5, all three satisfy the postulates in the following sense.

THEOREM 5.6.

(i) *S_P satisfies all seven postulates. For strictly positive probability spaces (which are necessarily generated by random variables),*

(ii) *C'_P and C''_P satisfy* (B), (C), (D), (E), *and* (F); *and* (G) *vacuously*; *and, further,*

(iii) *C'_P and C''_P satisfy* (A) *whenever $H_P(\mathfrak{S}) < \infty$; and in, particular, when $\mathfrak{S}$ is finite.*

Despite this result, it seems that the postulates should be modified for at least two reasons. First, S_P has the serious shortcoming demonstrated in Theorem 5.4, e.g., if for any $0 < \epsilon < 1$, there exists a set $D \in \mathfrak{A} \cap \mathfrak{B}$ and with $0 < P(D) < \epsilon$, then $S_P(\mathfrak{A}, \mathfrak{B}) = 1$, although the σ-algebras $\mathfrak{A}$ and $\mathfrak{B}$ might otherwise be quite different in structure. Secondly, the Gel'fand-Yaglom result [6] that $C_P(\mathfrak{A}_X, \mathfrak{A}_Y) = -(\frac{1}{2}) \log (1 - \rho^2)$ for joint normal random variables indicates that for nonatomic σ-algebras C'_P and C''_P do not satisfy (G).

It is, therefore, suggested that a more desirable set of postulates is obtained if one replaces (E) by one of the following two postulates, and (G) by the third postulate below.

(E′): $\delta(X, Y) = 1$ iff there is strict dependence, i.e., there exist Borel measurable $f(\cdot)$ and $g(\cdot)$ such that either $X = g(Y)$ or $Y = f(X)$.

(E″): $\delta(X, Y) = 1$ iff $\mathfrak{A}_X = \mathfrak{A}_Y$, i.e., X and Y are functions of each other.

(G′): $\delta(X, Y)$ is a strictly monotone function of $|R(X, Y)|$, if the joint distribution of X and Y is normal.

Clearly,

THEOREM 5.7. *For strictly positive $(Z, \mathfrak{S}, P)$ with $H_P(\mathfrak{S}) < \infty$,*

(i) *C'_P satisfies* (A), (B), (C), (D), (E′), (F), *and* (G′);

(ii) *C''_P satisfies* (A), (B), (C), (D), (E″), (F), *and* (G′);

(iii) *S_P does not satisfy* (E′) *or* (E″).

The preference of (E′) or (E″) depends on what features are desirable in a dependence measure. For example, $C'_P(\mathfrak{A}_X, \mathfrak{A}_U) = 1$ and $C''_P(\mathfrak{A}_X, \mathfrak{A}_U) < 1$; while $C'_P(\mathfrak{A}_X, \mathfrak{A}_Y) = 1 = C''_P(\mathfrak{A}_X, \mathfrak{A}_Y)$ for the related random variables X, $Y = e^{-X}$ and $U = X^2 = X$. Under what circumstances should a dependence measure assume its maximum value 1? (See [22] and [23].)

Unfortunately, there may also be some difficulty with (C) for the case of σ-algebras, which are not generated by random variables. For example, the Lloyd generalization of C_P, which is defined in Section 6, sometimes is infinite.

There is, then, the problem of modifying the definitions and the postulates still further, if the spaces are not strictly postiive.

6. Extensions and open problems. Since the preceding discussion was primarily restricted to strictly positive probability spaces, which are necessarily generated by random variables, the full generality of the work of Kramer [10] and Lloyd [13] with arbitrary σ-algebras, has not been employed; nor has been the work of Gel'fand and Yaglom [6] for nonatomic σ-algebras generated by random variables.

For algebras generated by random variables, Gel'fand and Yaglom [6] have defined $C_P = \sup\, [H(\mathfrak{A}') + H(\mathfrak{B}') - H(\mathfrak{A}', \mathfrak{B}')]$ where the supremum is taken over finite subalgebras $\mathfrak{A}'$ and $\mathfrak{B}'$ of $\mathfrak{A}$ and $\mathfrak{B}$, respectively. Further, Rényi [20] has essentially proved that for such algebras, $\sup H(\mathfrak{A}') = +\infty$.

These results suggest that one define

$$C'_P(\mathfrak{A}, \mathfrak{B}) = \sup \{C_P(\mathfrak{A}', \mathfrak{B}')/\min\, [H_P(\mathfrak{A}'), H_P(\mathfrak{B}')]\} \quad \text{and}$$

$$C''_P(\mathfrak{A}, \mathfrak{B}) = \sup \{C_P(\mathfrak{A}', \mathfrak{B}')/\max\, [H_P(\mathfrak{A}'), H_P(\mathfrak{B}')]\},$$

where the suprema are taken as above. In this case it is easily demonstrated that the immediate extension of Theorem 5.2 is valid, i.e.,

THEOREM 6.1. *If $\mathfrak{A}$ and $\mathfrak{B}$ are generated by random variables and $H_P(\mathfrak{A})$, $H_P(\mathfrak{B})$ and $H_P(\mathfrak{A}, \mathfrak{B})$ are finite, then the following conditions are equivalent*:

(i) $S_P(\mathfrak{A}, \mathfrak{B}) = 0$;

(ii) $C'_P(\mathfrak{A}, \mathfrak{B}) = 0$;

(iii) $C''_P(\mathfrak{A}, \mathfrak{B}) = 0$;

(iv) $\mathfrak{A}$ *and* $\mathfrak{B}$ *are independent* w.r.t. P.

For arbitrary σ-algebras, Lloyd [13] defines $C_P(\mathfrak{A}, \mathfrak{B}) = \int \log\, (d\nu_0/d\lambda_0)\, d\nu$ if $\nu_0 \ll \lambda_0$, $= +\infty$ otherwise, where $\lambda = P \times P$ is the product measure on the product σ-algebra $\mathcal{S} \times \mathcal{S}$; ν is the measure on $\mathfrak{D} = \{D \subset Z \times Z: [z: (z, z) \varepsilon D] \varepsilon \mathcal{S}\}$ such that $\nu(A \times B) = P(A \cap B)$ for each $A \times B \varepsilon \mathcal{S} \times \mathcal{S}$; and λ_0 and ν_0 are, respectively, their restrictions to the product σ-algebra $\mathfrak{A} \times \mathfrak{B}$. For σ-algebras generated by random variables, the Lloyd definition coincides with the Shannon definition.

OPEN PROBLEM 1. For arbitrary σ-algebras, what are the natural definitions of C'_P and C''_P, and which of Theorems 5.3–5.5 are valid? [Rényi, in a personal communication, suggests that perhaps there are no natural extensions of C'_P and C''_P in the continuous case.]

OPEN PROBLEM 2. Rényi [19] has shown that, for random variables X and Y, if the transformation A defined by $Af = E\{E[f(X) \mid Y]Y\}$, is completely continuous on $\mathcal{L}_2(\mathcal{A}_X)$, then there exist $X_0 \varepsilon \mathfrak{M}_n(\mathcal{A}_X)$ and $Y_0 \varepsilon \mathfrak{M}_n(\mathcal{A}_Y)$, such that $S_P(\mathcal{A}, \mathcal{B}) = R(X_0, Y_0)$. For arbitrary σ-algebras, what is a necessary and sufficient condition that there exist random variables X_0, Y_0 measurable respectively w.r.t. $\mathcal{A}$ and $\mathcal{B}$ and such that $S_P(\mathcal{A}, \mathcal{B}) = R(X_0, Y_0)$?

REFERENCES

[1] C. B. Bell, "On the structure of stochastic independence," *Ill. J. Math.*, Vol. 2 (1958), pp. 415–424.

[2] C. B. Bell, "Maximal independent stochastic processes," *Ann. Math. Stat.*, Vol. 32 (1961), pp. 704–708.

[3] Lennart Carleson, "Two remarks on basic theorems of information theory," *Math. Scand.* Vol. 6 (1958), pp. 175–180.

[4] Amiel Feinstein, *Foundations of Information Theory*, McGraw Hill, New York, 1958.

[5] H. Gebelein, "Das statistische Problem der Correlation also variations und eigenwert Problem," *Zeitschrift fur Angewandte Mathematik und Mechanik*, Vol. 21 (1941), pp. 364–379.

[6] I. M. Gel'fand and A. M. Yaglom, "Computation of the amount of information about a stochastic function contained in another such function," (Russian) *Upsehi Mat. Nank* (N.S.), Vol. 12 (1957), pp. 3–52; [MR 18, 980].

[7] Paul R. Halmos, *Introduction to Hilbert Space*, Chelsea, New York, 1951.

[8] Paul R. Halmos, *Entropy in Ergodic Theory*, University of Chicago Press, 1959.

[9] A. I. Khinchin, *Mathematical Foundations of Information Theory*, Dover, New York, 1957.

[10] H. P. Kramer, "Maximal correlation and uncertainty," Abstract 576-213, Notices *Amer. Math. Soc.*, Vol. 7, No. 7, Issue No. 50 (1960), pp. 977–978.

[11] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty-I, II," *Bell System Tech. J.*, Vol. 50 (1961), pp. 43–84.

[12] E. H. Linfoot, "An informational measure of correlation," *Information and Control*, Vol. 1, (1957) pp. 85–89.

[13] S. P. Lloyd, "On a measure of stochastic independence," submitted for publication.

[14] Michel Loève, *Probability Theory*, D. Van Nostrand, New York, 1960.

[15] E. Marczewski, "Independence d'ensembles et prolongement de mesures," *Colloq. Math.*, Vol. 1 (1947–48), pp. 122–132.

[16] Edward Marczewski, "Measures in almost independent fields," *Fund Math.*, Vol. 38 (1951), pp. 217–229.

[17] Edward Marczewski, "Remarks on the convergence of measurable sets and measurable functions," *Colloq. Math.*, Vol. 3 (1955), pp. 118–124.

[18] A. Rényi, "New version of the probabilistic generalization of the large sieve," *Acta Math. Acad. Sci. Hung.*, Vol. 10 (1959), pp. 218–226.

[19] A. Rényi, "On measures of dependence," *Acta Math. Acad. Sci. Hung.*, Vol. 10, (1959), pp. 441–451.

[20] A. Rényi, "On the dimension and entropy of probability distributions," *Acta Math. Acad. Sci. Hung.*, Vol. 10 (1959), pp. 193–215.

[21] Claude E. Shannon and Warren Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.

[22] Manuel Castans Camargo, "A theory of certitute I. Definitions and properties," (Spanish) *An. Real Soc. Espan. Fis. Quim.*, Ser. A, Vol. 51 (1955), pp. 215–232.

[23] Manuel Castans Camargo and Mariano Medina e Isabel, "The logarithm correlation," *An. Real Soc. Espan. Fis. Quim.*, Ser. A, Vol. 52 (1956), pp. 117–136.