R E P O R T   R E S U M E S

ED 016 289 EA 001 061

ON DEPARTURES FROM INDEPENDENCE IN CROSS-CLASSIFICATIONS.

BY- CASE, C. MARSTON
NATIONAL CENTER FOR EDUCATIONAL STATISTICS (DHEW)

REPORT NUMBER TN-10 PUB DATE 18 NOV 66
EDRS PRICE MF-$0.25 HC-$1.52 36P.

DESCRIPTORS- *CLASSIFICATION, *STATISTICAL DATA, *PROBABILITY, *STATISTICAL ANALYSIS, *STATISTICS, MODELS, CHARTS, DISTRICT OF COLUMBIA, Z-MEASURE,

THIS NOTE IS CONCERNED WITH IDEAS AND PROBLEMS INVOLVED IN CROSS-CLASSIFICATION OF OBSERVATIONS ON A GIVEN POPULATION, ESPECIALLY TWO-DIMENSIONAL CROSS-CLASSIFICATIONS. MAIN OBJECTIVES OF THE NOTE INCLUDE--(1) ESTABLISHMENT OF A CONCEPTUAL FRAMEWORK FOR CHARACTERIZATION AND COMPARISON OF CROSS-CLASSIFICATIONS, (2) DISCUSSION OF EXISTING METHODS FOR CHARACTERIZING CROSS-CLASSIFICATIONS, (3) PROPOSAL OF A NEW APPROACH TO AND A NEW METHOD FOR CHARACTERIZING AND MAKING INFERENCES FROM CROSS-CLASSIFICATIONS, AND (4) INDICATION OF HOW MARKOV PROCESSES CAN BE TREATED AS CROSS-CLASSIFICATIONS. THREE KINDS OF PROBABILITIES (UNCONDITIONAL, CONDITIONAL, AND JOINT) ARE DISCUSSED IN TERMS OF STATISTICAL INDEPENDENCE. MEASURES OF ASSOCIATION, ESPECIALLY THE Z-MEASURE OF ASSOCIATION AND ITS RELATION TO THE PHI COEFFICIENT IN VARIOUS SITUATIONS, ARE DISCUSSED AS MEANS OF INVESTIGATING NON-INDEPENDENCE. AN EXAMPLE OF THE Z-MEASURE APPLICATION IS PRESENTED. THE PROCESS OF DERIVING JOINT PROBABILITIES FROM Z-MEASURES AND THE INTERPRETATION OF MARKOV PROCESSES AS CROSS-CLASSIFICATIONS ARE ILLUSTRATED. (HW)

NATIONAL CENTER FOR EDUCATIONAL STATISTICS
Division of Operations Analysis

ON DEPARTURES FROM INDEPENDENCE IN CROSS-CLASSIFICATIONS

by

C. Marston Case

Technical Note
Number 10

November 18, 1966

OFFICE OF EDUCATION/U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

NATIONAL CENTER FOR EDUCATIONAL STATISTICS
Alexander M. Mood, Assistant Commissioner

DIVISION OF OPERATIONS ANALYSIS
David S. Stoller, Director

TABLE OF CONTENTS

I. Introduction

This note*is concerned with ideas and problems involved in cross-classification of observations on a given population. Most of the note will be confined to the discussion of two-dimensional cross-classifications. An example of this is the two-dimensional cross-classification of a portion of a deck (population) of playing cards resulting from the classification of the cards according to suit and according to whether or not the card is a face card. Such a cross-classification would consist of a tabulation of the numbers in each of the ($4 \times 2 = 8$) possible combinations of suit and face-or-no-face characteristics.

The main objectives of this note are:

1) to establish a conceptual framework for characterization and comparison of cross-classifications;

2) to discuss existing methods for characterization of cross-classifications;

3) to propose a new approach and a new method for characterizing and making inferences from cross-classifications;

4) to indicate how Markov processes can be treated as cross-classifications.

* The author wishes to thank Stephen Clark, George Mayeske, Richard O'Brien, and Frederic Weinfeld for suggestions made during the writing of this note.

II. Terminology

The word "event" is the conventional probabilistic term used to indicate what is observed; we will speak of observations on a given population with the assumption that one or more events are observed with each observation. The generic term "event type" will be used to specify the characteristic being classified in one dimension of a cross-classification. When we observe two or more events in a single observation we are observing the joint occurrence of the given events. The number of (joint) events so observed is the dimension of the cross-classification and is also the number of event types in the cross-classification. The suit of a card classifies it according to one event type while the face-or-no-face characteristic classifies it according to another event type. Within each event type there are two or more "event classes"; these classes are mutually exclusive and exhaustive, i.e., each observation belongs to exactly one event class within each event type. Within the suit event type the four event classes are club, diamond, heart, and spade. See figure 1.

Event types → ↓	Event classes → ↓	Suit			
		diamond	heart	club	spade
Face-or-no-face	face				
	no face				

Figure 1

What are called "event classes within event types" here are called "subcategories within attributes" by Guttman (1), p. 258 and "classifications within criteria" by Mood (2), p. 274, and "classes within polytomies" by Goodman and Kruskal (3), and "categories within variables" by Kendall and Stuart (6), p. 256. An extensive treatment of cross-classifications is given by Kendall and Stuart (6) in a 55-page chapter entitled "Categorized Data." The present paper deals exclusively with categorized data, which are observations which identify events (event classes within event types) in a qualitative, non-numerical, non-ordered manner. For instance, the suits and colors of playing cards are event types which are based on categorized data.

Three kinds of probabilities are distinguished in this note:

a) unconditional (or marginal) probabilities

b) conditional probabilities

c) joint probabilities.

We will be assuming that we are observing elements of a well-defined basic population and at each observation two or more events occur. The unconditional probability of an event is the probability of the occurrence of that event without regard to the occurrence of any other event. It is the fraction of the observations in which that event has occurred if observations have been taken on the entire basic population.

The idea of joint probability is the same as unconditional probability except that we are concerned with two (or more) events occurring in a single observation instead of just one. The joint probability of two events is the fraction of the observations in which both events would occur were the whole basic population observed. The ampersand will be used to indicate joint occurrences in this note; "A&B" means the joint occurrence of events A and B and Pr(A&B)" means the joint probability of their occurrence.

The idea of conditional probability also involves at least two events, say A and B. The conditional probability that A occurs given the condition that B also occurs means that we are evaluating a probability within a population restricted by some condition which was not part of the original definition of the basic population. The occurrence of event A given that event B also occurs is symbolized by "A|B" and "Pr(A|B)" means the conditional probability of event A given event B. The conditional probability of A given B is often defined as the ratio of their joint probability to the unconditional probability of B:

$$Pr(A|B) = \frac{Pr(A\&B)}{Pr(B)} \quad .$$

This means: of all the times that B occurs, Pr(A|B) is that fraction of the times wherein A also occurs.

When we speak of the occurrence of event A, we imply the idea of the nonoccurrence of A. In other words, in the back of our minds we are considering an A-type event which includes two event classes: class A and class non-A. Not-A will be symbolized by "$\bar{A}$" in this note. This A-type event is also known as a "binary variable" or a variable which can take on two values, A and $\bar{A}$.* "Event type" is equivalent to the word "variable" here, and the language of this note could have been based on "qualitative variables" instead of "categorized event types."

---

*Other words which are used for this kind of variable are: counting, indicator, dichotomous, two-state, two-point, and zero-one variable or distribution.

III. Statistical Independence

The standard definition of the statistical independence of two events, A and B, is that the probability of their joint occurrence is the product of their probabilities:

$$Pr(A\&B) = Pr(A) \times Pr(B) .$$

Using this joint probability definition of independence and the definition of conditional probability we obtain the definition of independence in terms of conditional probability: events A and B are independent if and only if

$$Pr(A|B) = \frac{Pr(A)Pr(B)}{Pr(B)} = Pr(A).$$

The Venn diagram which is Figure 2 can be used to show the (probability domain) relationship of the two events, A and B. We see that the domains of the two events must overlap a certain amount in order to be independent. The amount required for their independence is the product of the probabilities of the events involved; *e.g.*, if Pr(A)=.6 and Pr(B)=.8 then in order for A and B to be independent the intersection of their probability domains must be Pr(A) x Pr(B), *i.e.*, .6 times .8 or .48.

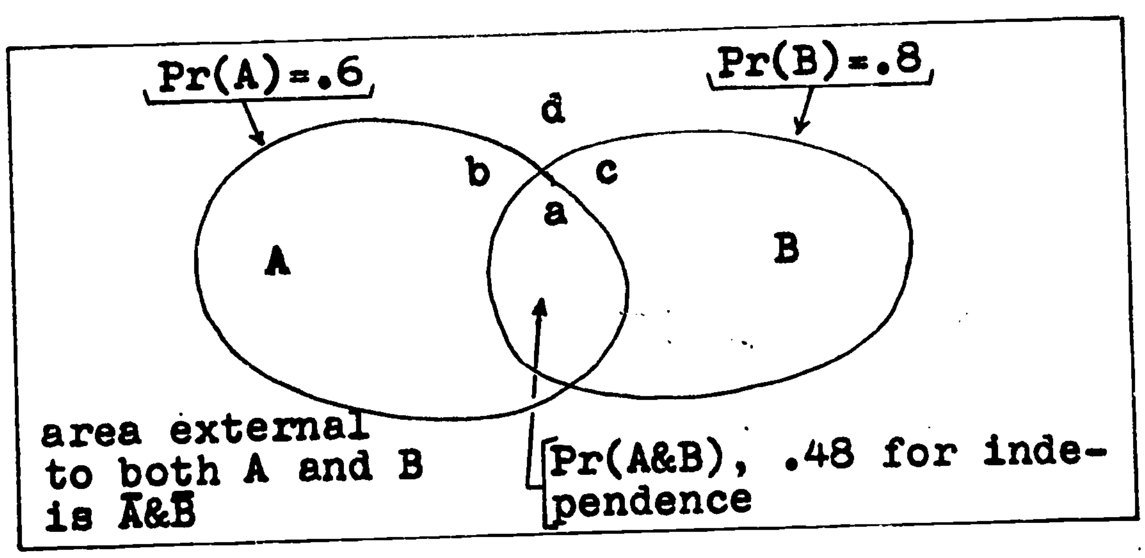

Figure 2

If the amount of overlap is less than the product required for independence, the occurrence of either event has a tendency to preclude the occurrence of the other. If the two probability domains do not overlap, then each event completely precludes the occurrence of the other and they are mutually exclusive. On the other hand, if the amount of overlap is greater than the amount required for independence then the occurrence of either event enhances the probability of the occurrence of the other. If there is complete overlap with one probability domain covering that of the other, then, of course, the occurrence of the second is completely dependent on the occurrence of the first.

To put the Venn diagram of Figure 2 into tabular form, a 2 by 2 joint probability table with marginal (unconditional) probabilities is drawn (Figure 3). In the body of the table a, b, c, and d are joint probabilities and the sums of the joint probabilities in rows and columns are marginal probabilities.

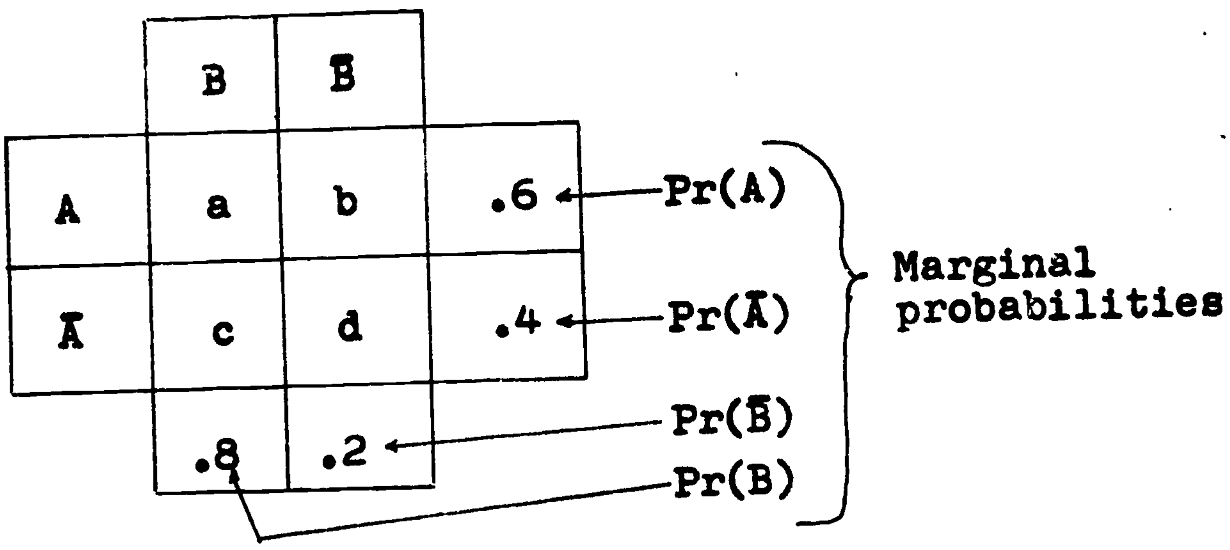

Figure 3

The probability domains (a, b, c, d) of the Venn diagram correspond to those in the tabulated form, Figure 3. These are, respectively, the joint probabilities: Pr(A&B), Pr(A&$\bar{B}$), Pr($\bar{A}$&B), Pr($\bar{A}$&$\bar{B}$). Note that area d is the area external to the A&B domains. Given two marginal probabilities in Figure 3 (one of the A type marginals and one of the B type) and an entry in any of the four cells of the body of the table (a, b, c, or d), the remaining cells are easily deduced. For instance, if a is .5 then since a+b=.6, b=.1 and since a+c=.8, c=.3 and finally, since c+d=.4, d=.1. Note also that if a is .5, A and B are not independent; they enhance each other's occurrence.

The tabular representation has the advantages of explicitly identifying $\bar{A}$ and $\bar{B}$ and showing A and $\bar{A}$ in a uniform manner. The tabular form also lends itself more easily to event types which are classified into several (not just two) classes. Suppose that the type A event has three classes A_1, A_2 and A_o where A_o are events which belong to neither A_1 nor A_2. Similarly, suppose the type B events are categorized into classes B_1, B_2 and B_o. The Venn diagram of this with some probabilities put in as an example is Figure 4.

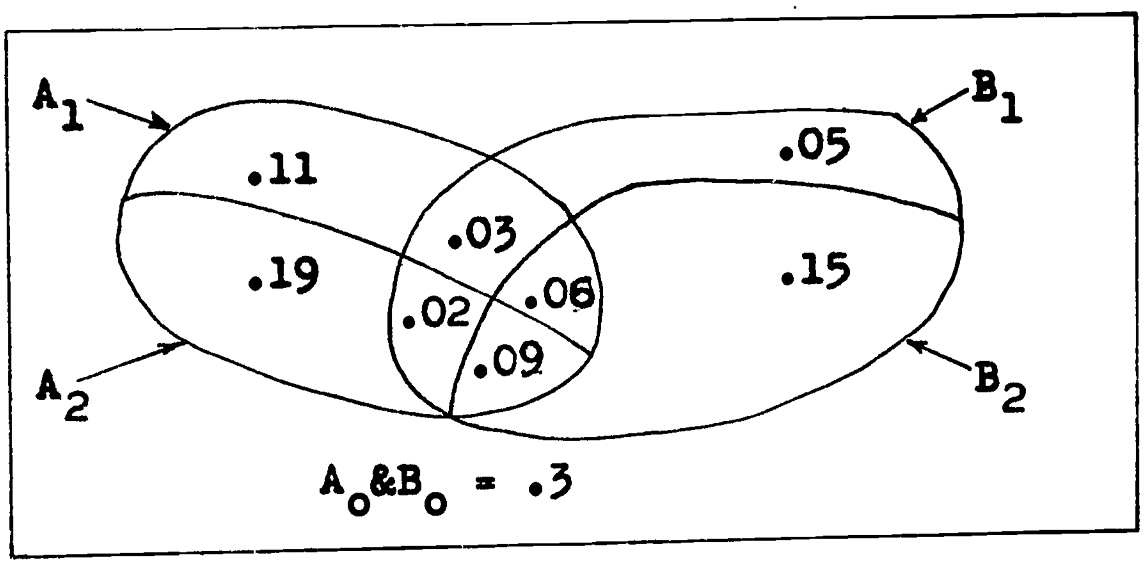

Figure 4

The same joint probabilities information in tabular form appears in Figure 5.

	B_1	B_2	B_o	A marginals
A_1	.03	.06	.11	.2
A_2	.02	.09	.19	.3
A_o	.05	.15	.3	.5
B marginals	.1	.3	.6	

Figure 5

Consider the independence-dependence characteristics of the nine joint events whose probabilities are in the body of the table. We make the following four observations:

1) A_o is independent of all the B_i
2) B_2 is independent of all the A_i
3) A_1 enhances the probability of B_1 while diminishing that of B_o
4) A_2 enhances the probability of B_o while diminishing that of B_1.

This example shows that instances of dependence can be scattered about among various joint events when the event types each include several classes.

In this example with three classes in each of two event types, there are (3-1)(3-1)=4 degrees of freedom in the determination of the probabilities. In other words, given the marginal probabilities, knowledge of certain sets of four of the nine joint probabilities in the body of the table determines all of them. These sets of probabilities are sets of four such that (after entering the set) at least one cell in each row and column remains empty and none of the remaining probabilities in the body of the table can be determined immediately from both its row and its column. Another way of stating the second requirement is that if an entry can be determined from the other entries in its row, it must not be possible to determine it from the other entries in its column. This can be generalized to m event types with n_1, n_2. . . . n_m classes, respectively. Then, given the marginal probabilities, there are $(n_1-1)(n_2-1) \ldots (n_m-1)$ degrees of freedom in determining the joint probabilities.

Testing for independence in a contingency table is a classical statistical problem discussed in many statistical books (see, for instance, Mood (2), pp. 273-81). A contingency table tabulates a set of observations according to two event types (criteria). Consider, for example, the classification of a group of people according to the two event types of vision and weight. The vision of each person in the group belongs to one of the three classes: near-sighted, normal sighted, and far-sighted; and the weight of each belongs to one of the three classes: underweight, normal weight, and overweight. The contingency table for these two event types would be a 3 x 3 table indicating the number of observations in each of the nine possible combinations of vision and weight.

Testing the independence of the two event types in a contingency table can be done in the following three steps.

1) Change the number in each of the cells to the fraction it is of the total number of observations, *i.e.*, divide each cell number by the total number.

2) Enter the row and column sums of these fractions as marginals. (The $ij^{\underline{th}}$ cell now is an estimate of the probability that the $i^{\underline{th}}$ row event and the $j^{\underline{th}}$ column event occur togehter (jointly). If the two events are independent, another estimate of their joint probability is the product of the $i^{\underline{th}}$ row marginal and the $j^{\underline{th}}$ column marginal.)

3) Let O_{ij} be the observed fraction of the total in the $ij^{\underline{th}}$ cell and E_{ij} be the expected fraction assuming independence, i.e., the product of the $i^{\underline{th}}$ row and $j^{\underline{th}}$ column marginals; then compute the test statistic for the (n x m) contingency table,

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(O_{ij}-E_{ij})^2}{E_{ij}} .$$

Under the assumption of independence this statistic is approximately chi-square distributed with (m-1)(n-1) degrees of freedom. The approximation to chi-square improves with larger numbers of observations in the cells.

IV. Measures of Association

The test for independence of two types of events can be a basis for deciding whether or not the two event types are independent; if the test leads one to reject the independence hypothesis, one may want a more complete explanation of the non-independence. One way to investigate non-independence is by employing a "measure of association" which, in the statistical literature, has meant a measure of the direction and amount of departure from independence of a given cross-classification. Goodman and Kruskal have written three extensive (32, 40, and 54 pages) papers on measures of association (3), (4), (5). They favor a measure originally proposed by Guttman in (1), but they feel that the uses of measures of association are varied enough so that there should be several from which to choose. None of the studies seen by the author, however, considers analysis of departures from independence in detail within a cross-classification (i.e., in the body of a joint probability table). This is to say that the measures of association thus far developed are meant to show (dependence) relationships between event types and not those between the event classes of one event type and those of another event type.

If we were working with numerical variables instead of categorized event types and we assumed that the variables were linearly related, we would probably consider the correlation coefficient as the first candidate for measuring their association. A set of three characteristics of the correlation coefficient traditionally has been sought in the appraisal of measures of association for categorized data:

1) the measure is zero when the event types are independent;

2) the measure is minus one when they have the maximum disassociation;

3) the measure is plus one when they have the maximum association.

We will consider a measure of association which, in its immediate application, will be applied only to a 2 x 2 cross classification although in its ultimate application it will appear in the assessment of general n x m cross classifications. In addition to the above set of characteristics we shall require our measure to have another set of characteristics related to the correlation coefficient. This set has to do with the symmetry of the measure between events (classes) and their complements. Specifically, if Z is the value of a measure of the association between events A and B, then it must also be the value between $\bar{A}$ and $\bar{B}$. Also, the value of the measure between A and $\bar{B}$ and between $\bar{A}$ and B must be -Z.

V. The Z-measure of Association

A method will now be shown that develops a measure of association conforming to the requirements specified above and applies to individual cells, allowing them to be compared with one another. Since each cell in the given (n x m) joint probability table is to be treated individually, we treat each cell as the upper left corner entry in a 2 by 2 table as in Figure 6.
This is done by letting

a = the cell being studied

b = sum of the remainder of the row entries

c = sum of the remainder of the column entries and

d = sum of the remainder of the table (= 1-a-b-c);

	B	$\bar{B}$	
A	a	b	y
$\bar{A}$	c	d	1-y
	x	1-x	

Figure 6

x and y are marginal probabilities in both the n x m and the 2 x 2 tables; x = a+c and y = a+b. All the information about the 2 x 2 table is contained in the three quantities, a, x, and y.

A measure of association between two binary variables is their correlation coefficient. The definition of the correlation coefficient for two variables A and B is:

$$\frac{E(AB) - E(A)E(B)}{\sqrt{V(A)V(B)}}$$

where E(A) is the expected value of A and V(A) is the variance of A which is $E(A^2) - E(A)^2$.

We are concerned here with binary variables, specifically the joint distribution of A and B, a bivariate binary probability distribution. In this case we can define the distribution of the A-or-not-A type of event on the integers zero and one:

$$\Pr(\text{A-or-not-A event type} = 0) = \Pr(\bar{A}) = 1-\Pr(A)$$

and

$$\Pr(\text{A-or-not-A event type} = 1) = \Pr(A).$$

A similar definition of the binary distribution of B applies. Then the following equations hold:

$$E(AB) = \Pr(A\&B) = a$$

$$E(A) = E(A^2) = \Pr(A) = y$$

$$E(B) = E(B^2) = \Pr(B) = x$$

$$V(A) = y(1-y)$$

$$V(B) = x(1-x).$$

This leads to a formula for the correlation coefficient associated with the 2 x 2 table in Figure 6,

$$\text{Phi}(a,x,y) = \frac{a-xy}{\sqrt{y(1-y)x(1-x)}} .$$

"Phi coefficient" is the usual name for this statistic, especially in psychological statistics.* It is zero when the events are independent ($a = xy$). It takes on the value plus one only when a is at its maximum and when $x = y$. The maximum of a is the lesser of x and y. The phi coefficient takes on the value minus one only when a is at its minimum and when $x + y = 1$. The minimum of a is zero when $x + y \leq 1$ and $x + y - 1$ when $x + y \geq 1$.

We want a measure of association which always takes on the value minus one when the least possible association prevails and always takes on the value plus one when the greatest possible association prevails. Such a measure is obtained by dividing the phi coefficient by its maximum possible value when positive association prevails and by its minimum possible value when negative association prevails. We now define such a Z measure of association

---

* For a discussion of the phi-coefficient see Guilford (7), pp. 333-36.

for four possible situations which cover all the possibilities of a 2 x 2 table.

Situation I. $a > xy$, $x \geq y$, the binary variables are positively associated. Z is the ratio of phi coefficient to its maximum:

$$Z(a,x,y) = \frac{a-xy}{\sqrt{x(1-x)y(1-y)}} \Bigg/ \frac{y-xy}{\sqrt{x(1-x)y(1-y)}}$$

$$= \frac{a-xy}{y(1-x)} \quad \left[= \frac{a-xy}{y-xy} \right]$$

Situation II. $a < xy$ and $x+y \leq 1$, the binary variables are negatively associated (disassociated); Z is the negative of the ratio of phi coefficient to its minimum:

$$Z(a,x,y) = (-1) \frac{a-xy}{\sqrt{x(1-x)y(1-y)}} \Bigg/ \frac{-xy}{\sqrt{x(1-x)y(1-y)}}$$

$$= \frac{a-xy}{xy}$$

Situation III. $a < xy$ and $x+y \geq 1$, the binary variables are negatively associated (as in II). Z is the negative of the ratio of phi coefficient to its minimum:

$$Z(a,x,y) = (-1) \frac{a-xy}{\sqrt{x(1-x)y(1-y)}} \Bigg/ \frac{x+y-1-xy}{\sqrt{x(1-x)y(1-y)}}$$

$$= \frac{a-xy}{1-x-y+xy} = \frac{a-xy}{(1-x)(1-y)}$$

Situation IV. $a = xy$, the binary variables are independent; $Z = 0$.

In situations II and III, the sign is made negative to indicate disassociation.

Horst (8), pp. 238-39, points out that this Z-measure is the ratio of two covariances, an observed covariance over a maximum (Sit. I) or a minimum (Sits. II and III) covariance. He uses it as a measure of the homogeneity of test items. This is an interesting case in which the event classes are not mutually exclusive.

The 2 x 2 table which is Figure 6 can be written solely in terms of a, x, and y. Such a table equivalence is shown in Figure 7.

	B	$\bar{B}$	
A	a	y-a	y
$\bar{A}$	x-a	1+a -x-y	1-y
	x	1-x	

=

a	b	y
c	d	1-y
x	1-x	

Figure 7

Assuming $a > xy$ and $x \geqslant y$, we now find the Z-measure for each of the four cells in the body of the table. An arrow ($\longrightarrow$) is used to mean implication.

$$Z(a,x,y) = \frac{a-xy}{y(1-x)}$$

$a > xy \longrightarrow 1+a-x-y > 1+xy-x-y = (1-x)(1-y)$. Thus since

$1+a-x-y > (1-x)(1-y)$, the computation for the lower right cell, d, is done using the Situation I formula.

$$Z(d,\ 1-x,\ 1-y) = Z(1+a-x-y,\ 1-x,\ 1-y)$$

$$= \frac{1+a-x-y-(1-x)(1-y)}{(1-x)\left[1-(1-y)\right]} = \frac{a-xy}{y(1-x)} .$$

For the upper right cell, b,

$a > xy \rightarrow y-a < y(1-x)$

$x > y \rightarrow 1-x < 1-y \rightarrow (1-x) + y < 1$.

Using Situation II,

$$Z(b,\ 1-x,\ y) = Z(y-a,\ 1-x,\ y)$$

$$= \frac{y-a-(1-x)y}{y(1-x)} = - \frac{a-xy}{y(1-x)} .$$

For the lower left cell, c,

$a > xy \rightarrow x-a < x(1-y)$

$x > y \rightarrow 1-x < 1-y \rightarrow (1-y) + x > 1$.

Using Situation III,

$$Z(c,\ x,\ 1-y) = Z(x-a,\ x,\ 1-y)$$

$$= \frac{x-a-x(1-y)}{(1-x)\ \ 1-(1-y)} = - \frac{a-xy}{y(1-x)} .$$

Thus $Z(a,\ x,\ y) = Z(d,\ 1-x,\ 1-y) =$

$$-Z(b,\ 1-x,\ y) = -Z(c,\ x,\ 1-y)$$

or in terms of the binary variables A and B,

$$Z(A,B) = Z(\bar{A},\ \bar{B}) = -Z(A,\ \bar{B}) = -Z(\bar{A},B) .$$

Thus we have a measure of association for a pair of binary variables which has the desired characteristics:

1. zero when variables are independent (Situation IV)

2. minus one when variables are as completely disassociated as possible (in Situations II and III)

3. plus one when the variables are as completely associated as possible (in situation I)

4. $Z(A,B) = Z(\bar{A},\bar{B}) = -Z(\bar{A},B) = -Z(A,\bar{B})$

Z(a,x,y) is a function which maps points in a three-dimensional space to points in a one-dimensional space. Figure 8 shows one aspect of this mapping: Z as a function of a with x and y fixed.

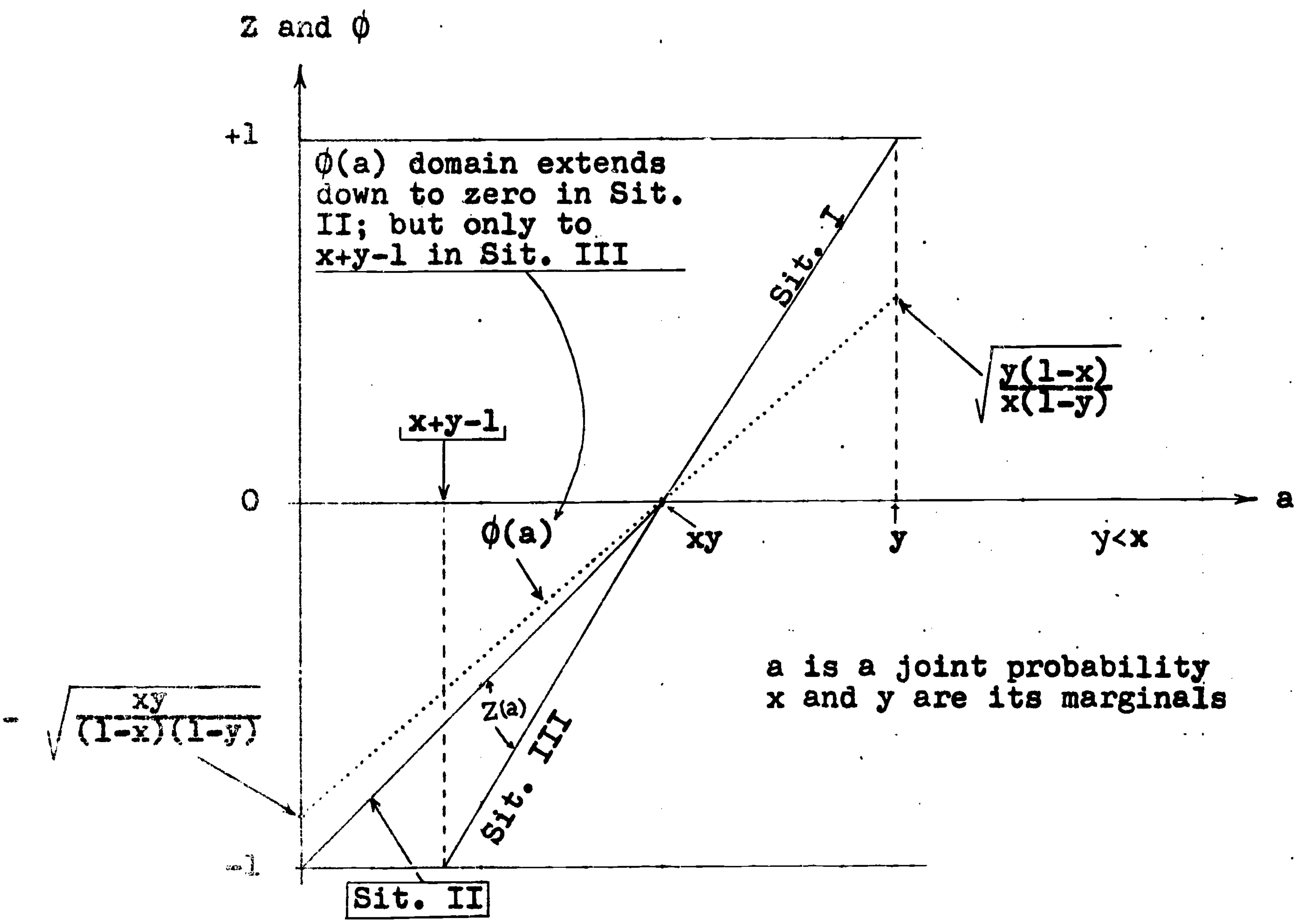

Figure 8

The bivariate binary correlation coefficient, ϕ(a) or phi(a), is plotted as a dotted line. It has the same domain (set of arguments) as Z but is linear throughout its domain.

One should keep in mind that, although the Z-measure is related to the correlation coefficient it differs from the correlation coefficient in fundamental ways. The Z-measure is not linear at Z=0 except when one or both of the marginals is .5. It is a two-part linear function of the correlation coefficient such that the coefficients of the function are non-linear functions of x and y. The Z-measure should be judged, however, in its own right, and not as a replacement for the correlation coefficient which fails to behave for binary variables as it does for linearly related continuous numerical variables. It is hard to find a common situation with less favorable conditions for establishing linear relations than that of the bivariate binary distribution with point-masses of probabilities at the four corners of a square. The perfect correlation condition there is when all the mass is on diagonally opposite corners. As we saw earlier, this can occur only when x = y (for perfect correlation) or when x + y = 1 (for perfect negative correlation). In all other cases the correlation coefficient cannot attain 1.

The advantage of Z over the correlation coefficient is that it has the range of values minus one to plus one for all x and y combinations. This allows one to compare Z measures based on different x and y combinations and say that the amount of interaction between two event classes is greater or less than that between another pair.

Additional insight into the nature of the Z-measure is gained when considering it from the set-theoretic point of view using the Venn diagram approach of Section III. Z-measure essentially shows the relationship between the amount of observed overlap and the maximum or minimum potential overlap of two probability domains. It may be thought of as the attraction or repulsion force existing between the probability domains of two events. In this sense it is considered to be measuring a symmetric force such as gravity or magnetic forces.

There have been several discussions of the Z-measure in psychological statistics. The Z-measure is usually called "ϕ / ϕ max" with little mention of "ϕ / ϕ min". Carroll (9), pp. 363-64 discusses this measure in terms of 2 x 2 tables derived from the division of continuous bivariate distributions into four parts; observations are classified as being above or below a point on the scale of one marginal distribution and above or below a point on the scale of the other marginal distribution. Carroll particularly refers to the "tetrachoric correlation coefficient" (which assumes a bivariate normal underlying distribution) as the ideal one and finds the Z-measure does not approximate the tetrachoric correlation coefficient very well. He rather emphatically dismisses the Z-measure on this basis. Guilford (7), pp. 337-38, also warns against the use of the Z-measure as an "indicator of intrinsic correlation." In general, the psychometricians discount the use of the Z-measure in lieu of a (Pearson) correlation coefficient. The mathematical treatment of the Z-measure, however, needs more development than is given in the psychometric literature in order to understand enough about the measure to make proper use of it. Cureton (10), p. 89, for instance, states that Z can be +1 only when $x = y = .5$ and Z can be -1 only when $x = 1-y = .5$; whereas $Z = +1$ if and only if $x = y = a$ and $Z = -1$ if and only if $x = 1-y$ and $a = 0$.

Comparing the joint probability (a) with the Z-measure for a given joint event, and generalizing over the four situations we see that

$$a = xy + cZ$$

where

$$\begin{aligned} c &= \min(x,y) - xy && \text{if } Z \geq 0 \\ &= xy && \text{if } Z \leq 0 \text{ and } x+y \leq 1 \\ &= (1-x)(1-y) && \text{if } Z \leq 0 \text{ and } x+y \geq 1. \end{aligned}$$

Z is thus seen to show the relationship between the joint probability and the associated marginal probabilities, x and y. The nature of c is further elaborated in Section VII.

There are also two conditional probabilities associated with a joint event and these show, respectively, the relationship between one marginal and the joint probability and the other marginal and the joint probability. Conditional probabilities are used to improve one's ability to predict the occurrence or non-occurrence of an event by obtaining information about the occurrence or non-occurrence of another event. Such a prediction operation sometimes tends to lead to the imputation of causality between the two events. The Z-measure is symmetric in the marginals and thus avoids such an imputation. A conditional probability without the appropriate unconditional probability does not tell one whether the event being conditioned is more or less likely when it occurs jointly with the conditioning event. Thus the conditional probability by itself is not a measure of association whereas the Z-measure is.

VI. An Application of the Z-measure

A contingency table which has often appeared as an example in textbooks and articles involves the cross-classification of 6800 males according to their hair and eye color.* Four classes of hair color were observed: fair, brown, black, and red. Three classes of eye color were observed: blue, hazel or green, and brown. Figure 9 shows the original observations (contingency table); Figure 10 shows the corresponding joint probabilities table; Figure 11 shows the corresponding values of the Z-measure of association; Figure 12 shows conditional probabilities of hair color given eye color and eye color given hair color.

This cross-classification fails the classical chi-square test of independence at the .000001 level. Some of the details brought out by the Z-measure are that red hair is essentially independent of eye color for this population while a general correlation of eye and hair pigment holds. The disassociation of fair hair and brown eyes (-.678) and of black hair and blue eyes (-.626) are, however, much more pronounced than the associations of fair hair and blue eyes (+.365) and of black hair and brown eyes (+.190). The weakness of the black hair and brown eye association is surprisingly less than that of both black hair and hazel eyes (.277) and that of brown hair and brown eyes (.201).

---

* _E.g._, see Goodman and Kruskal (3).

		HAIR		COLOR		
		fair	brown	black	red	
EYE COLOR	blue	1768	807	189	47	2811
	hazel	946	1387	746	53	3132
	brown	115	438	288	16	857
		2829	2632	1223	116	

FIGURE 9

JOINT PROBABILITY TABLE

	fair	brown	black	red	
blue	.26	.119	.028	.0069	.4134
hazel	.139	.204	.110	.0078	.4606
brown	.017	.064	.042	.0024	.1260
	.416	.387	.1799	.0171	

Figure 10

Z MEASURE OF ASSOCIATION

	fair	brown	black	red
blue	.365	-.258	-.626	-.02
hazel	-.274	.084	.277	-.008
brown	-.678	.201	.190	.014

Figure 11

CONDITIONAL PROBABILITY TABLES

	EYE GIVEN HAIR				HAIR GIVEN EYE				
	fair	brown	black	red	fair	brown	black	red	
blue	.625	.326	.154	.405	.629	.287	.0672	.0167	1.0
hazel	.334	.525	.610	.457	.302	.443	.238	.0169	1.0
brown	.041	.166	.235	.138	.134	.511	.336	.0187	1.0
	1.0	1.0	1.0	1.0					

Figure 12

VII. The Reverse Inference: Joint Probabilities From Z-measures

Heretofore we have been considering the derivation of the Z-measure from a joint probability table. We now consider the derivation of a joint probability table given a set of Z-measures and marginal probabilities. We may be given the marginal distributions of a cross-classification, for instance, and (perhaps subjective) estimates of some of the associations between various states. We can specify no more Z-measures than the degrees of freedom involved. The joint probability table and the Z-measure table of a three by three cross-classification are shown in Figure 13.

a_{11}	a_{12}	a_{13}	y_1
a_{21}	a_{22}	a_{23}	y_2
a_{31}	a_{32}	a_{33}	y_3
x_1	x_2	x_3	

z_{11}	z_{12}	z_{13}
z_{21}	z_{22}	z_{23}
z_{31}	z_{32}	z_{33}

Figure 13

We derive definitions of the a_{ij} from the specifications for Z's on page 16.

$$a_{ij} = x_j y_i + c_{ij} Z_{ij}$$

$$\text{where } c_{ij} = \min(x_j, y_i) - x_j y_i \quad \text{if } Z \geq 0$$

$$= x_j y_i \quad \text{if } Z \leq 0 \text{ and } x_j + y_i \leq 1$$

$$= (1-x_j)(1-y_i) \quad \text{if } Z \leq 0 \text{ and } x_j + y_i \geq 1$$

Theorem: $c_{ij} \leq .25$

c_{ij} is composed of two factors, say f_1 and f_2. One of the factors, say f_1, is no greater than .5 so $f_1 = .5-u$, $u \geq 0$. Then $f_2 \leq .5+u$. Suppose (at worst) $f_2 = .5+u$. Then $f_1f_2 = (.5-u)(.5+u) = .25-u^2 \leq .25$.

c_{ij} is the maximum possible difference between a_{ij} and x_jy_i given the information about whether a_{ij} is larger or smaller than x_jy_i; Z_{ij} indicates how much of this maximum deviation is attained by a_{ij}. One possibility is that $a_{ij} = x_jy_i$; then for each row and each column $\sum a_{ij} = \sum x_jy_i$. However, the row and column sums are the same (i.e., the marginal probabilities are fixed) for all permissible sets of a_{ij}. Thus for any permissible configuration of the a_{ij} for each row and column, $\sum a_{ij} = \sum x_jy_i$ and $\sum a_{ij} = \sum (x_jy_i + c_{ij}Z_{ij})$. Therefore, for each row and column $\sum c_{ij}Z_{ij} = 0$. For convenience, let $d_{ij} = c_{ij}Z_{ij}$. We now have the following set of six equations defining the relations among the d_{ij} for a 3 x 3 joint probability table:

1) $d_{11} + d_{12} + d_{13} = 0$

2) $d_{21} + d_{22} + d_{23} = 0$ row sums

3) $d_{31} + d_{32} + d_{33} = 0$

4) $d_{11} + d_{21} + d_{31} = 0$

5) $d_{12} + d_{22} + d_{32} = 0$ column sums

d) $d_{13} + d_{23} + d_{33} = 0$

$0 < c_{ij} \leq .25$ and $-1 \leq Z_{ij} \leq +1$ means that $-.25 \leq d_{ij} \leq +.25$.

Earlier the number of degrees of freedom associated with a cross-classification table were discussed (pp.9,10). If we decide certain sets of four d_{ij}, the remainder of a 3 x 3 table of d_{ij}'s are determined. There are 126 ways of selecting four cells from among nine; only 81 of these ways conform to the degrees of freedom requirements stated earlier. Each of those 81 provides us with a different configuration of four cell choices in a 3 x 3 table. If we specify one of these configurations, the d_{ij} for the remaining five cells can be stated in terms of our four initial d_{ij}. For instance, suppose we are given d_{11}, d_{22}, d_{33} and d_{31} (see Figure 14). Using the row and column sum equations we obtain the remaining d_{ij} in terms of these four:

$d_{32} = -d_{31} - d_{33}$

$d_{21} = -d_{11} - d_{31}$

$d_{12} = -d_{22} + d_{31} + d_{33}$

$d_{13} = -d_{11} + d_{22} - d_{31} - d_{33}$

$d_{23} = d_{11} - d_{22} + d_{31}$

d_{11}	d_{12}	d_{13}
d_{21}	d_{22}	d_{23}
d_{31}	d_{32}	d_{33}

Figure 14

VIII The Interpretation of Markov Processes as Cross-Classifications

A Markov chain (or process) is generally defined by its probability transition matrix and its initial distribution. Figure 15 is an example of a transition matrix of a 4-state Markov chain; it shows the probability of the occurrence of event E_j at time k+1 given that event E_i occurred at time k (i and j are the row and column indices respectively).

		Time k+1			
		E_1	E_2	E_3	E_4
Time k	E_1	.6	.4	0	0
	E_2	0	.6	.4	0
	E_3	0	0	.6	.4
	E_4	.2	0	0	.8

Figure 15

This transition matrix is a table similar to that associated with a cross-classification. It is, however, a table of conditional probabilities whereas for the complete specification of a cross-classification a joint probability table is needed. A conditional probability table can be deduced from an underlying joint probability table but a joint probability table cannot be deduced from either or both of its corresponding conditional probability tables. A joint probability table can, however, be deduced from one of its conditional tables and one of its marginal probability distributions. In a Markov chain, an initial probability distribution vector (hereafter PDV) is given; this PDV is one of the two marginal probability sets in the underlying joint probability table. The initial PDV completes the specifications needed for a cross-classification interpretation of a Markov chain. We will often subscript the PDV's with their associated time index, e.g., the initial PDV will be PDV_k.

Continuing with the example in Figure 15, suppose we are given a PDV of (.1, .2, .3, .4) for events E_1 to E_4. Then the joint probability table becomes that shown in Figure 16.

	E_1	E_2	E_3	E_4		
E_1	.06	.04	0	0	.1	Time k
E_2	0	.12	.08	0	.2	
E_3	0	0	.18	.12	.3	
E_4	.08	0	0	.32	.4	
	.14	.16	.26	.44		
	Time k+1					

Figure 16

The joint probabilities in the body of the table are obtained from the conditional probabilities in Figure 15 by multiplying the probabilities in each row by the corresponding probabilities of PDV_k. The event types are events observed at the two times, k and k+1 while the event classes are E_1 to E_4 for each type.

The present discussion tends to be in terms of the transformation of PDV's in contrast to the usual emphasis on probabilities of going from one state to another in a certain number of transitions (e.g., from E_1 to E_4 during time k to k+m). Thus the usual approach is based on powers of the transition matrix whereas this is based on sequences of PDV's. A large portion of the problems couched in terms of Markov chain theory are problems which assume that the process is ergodic. A Markov process is ergodic* if it converges to a steady state as k gets larger. A Markov process is in a steady state at time k only if $PDV_k = PDV_{k+1}$. For practical purposes we could say that a process is in the neighborhood of a steady state at time k+1 if PDV_k differs from PDV_{k+1} by less than a specified small amount. This neighborhood might be defined in terms of the accuracy of the estimates of the probabilities in the PDV's.

The bottom marginal (or PDV) in Figure 16 could be obtained by multiplying together the right marginal probabilities and the transition matrix of Figure 15 thus:

$$(.1, .2, .3, .4) \begin{matrix} .6 & .4 & 0 & 0 \\ 0 & .6 & .4 & 0 \\ 0 & 0 & .6 & .4 \\ .2 & 0 & 0 & .8 \end{matrix} = (.14, .16, .26, .44)$$

Such a multiplication indicates the transformation of the PDV between time k and k+1. We can obtain PDV_{k+2} by multiplying the result by the same transition matrix. In repeated multiplication of the previous result, the difference between successive PDV's is seen to diminish; the process is approaching the steady state. The process is in a steady state if the input PDV is the same as the output PDV, i.e.

$$(x_1, x_2, x_3, x_4) \begin{matrix} .6 & .4 & 0 & 0 \\ 0 & .6 & .4 & 0 \\ 0 & 0 & .6 & .4 \\ .2 & 0 & 0 & .8 \end{matrix} = (x_1, x_2, x_3, x_4)$$

*"Ergodic" is defined by Feller (11), p. 353. A different definition is given by Kemeny & Snell (12), p. 99; they include periodic or cyclic Markov chains in ergodic chains, Feller does not.

Figure 17 is the joint probability table of the steady state of the process, the transition matrix of which is shown in Figure 15.

	E_1	E_2	E_3	E_4	
E_1	.12	.08	0	0	.2
E_2	0	.12	.08	0	.2
E_3	0	0	.12	.08	.2
E_4	.08	0	0	.32	.4
	.2	.2	.2	.4	

Figure 17

Figure 18 is the corresponding Z-measure table.

.5	.25	-1	-1
-1	.5	.25	-1
-1	-1	.5	0
0	-1	-1	.666

Figure 18

As an ergodic Markov process progresses toward the steady state, the transition matrix remains invariant. The following items generally change as the time index (k) changes:

1. Both sets of marginal probabilities (the PDV's)

2. The joint probability table

3. The Z-measure

4. The conditional probability table which reverses the effect of the transition matrix, i.e., that which would take PDV_{k+1} to PDV_k.

The table in (4.) is that obtained by dividing the joint probabilities associated with the k to k+1 transition by the marginal probabilities in the margin below. The transformation from PDV_{k+1} to PDV_k could also be accomplished with the inverse of the transition matrix if the transition matrix is non-singular; such an inverse is invariant with respect to k.

A special kind of Markov chain which is an exception to the above characterization is that in which the transition to each state is independent of the state from which the transition began. Then the following conditions prevail:

1. All rows of the transition matrix are the same

2. The joint probabilities are products of their marginals

3. All Z-measures are zero

4. Given any initial PDV the convergence to the steady state is immediate and complete at time k+1 and all further transitions do not change the PDV's.

This special kind of Markov process suggests a potential basis for the comparison or characterization of Markov processes: the rapidity of convergence to a neighborhood of the steady state. Apparently the farther the Z-measures of the joint probability table of the steady state are from zero, the slower is the convergence. Notice, however, that zeroes and ones in the transition matrix indicate cells whose Z-measures are invariant during convergence so that such cells should be treated differently (perhaps excluded) from the cells which change during convergence. The theory needs further development in this area.

The steady state joint probability table and Z-measures are fixed for a given transition matrix and are independent of the initial PDV. Thus they are obvious choices for characterizing a Markov process. An infinite number of transition matrices converge to each steady state PDV; they all have different joint probability tables and different Z-measures.

If one is altering a Markov process in order to have it converge to a target steady state one method would be the comparison of the Z-measure matrix of the process at present with that of the target steady state. This will show which transitions must become more likely and which ones must become less likely in order to attain the target. It may well be, however, that the transition matrix which maintains the steady state PDV does not matter, i.e., only the PDV itself matters. Then many alternative Z-measure configurations could be compared to see which would be the preferred target based perhaps on time, cost, and other criteria.

The Markov process defined in Figures 15 to 18 has a 4 x 4 joint probability table so it has 3 x 3 = 9 degrees of freedom. There are eight zeroes in the joint table which remain there throughout all transitions; these lead to corresponding fixed Z-measures of minus one. By choosing any one of the non-negative Z's one can determine the entire process. This is a special (flow process) case and illustrates the method of reverse inference discussed in the previous section.

REFERENCES

1. Guttman, Louis. "An Outline of the Statistical theory of Prediction," Supplementary Study B-1 (pp. 253-311) in Paul Horst and others, The Prediction of Personal Adjustment. New York: Social Science Research Council, 1941. Bulletin 4B.

2. Mood, A.M. Introduction to the Theory of Statistics. New York: McGraw Hill, 1950.

3. Goodman, L.A. and W.H. Kruskal. "Measures of Association for Cross Classifications," Journal of American Statistical Association: 49 (1954), 723-64.

4. Ibid. Vol. 54 (1959), 123-63.

5. Ibid. Vol. 58 (1963), 310-64.

6. Kendall, M.G. and A. Stuart. The Advanced Theory of Statistics. Vol. II. New York: McGraw Hill, 1961.

7. Guilford, J.P. Fundamental Statistics in Psychology and Education. 4th Ed. New York: McGraw Hill, 1965.

8. Horst, Paul. Psychological Measurement and Prediction. Belmont, Calif.: Wadsworth Publishing Co., 1966.

9. Carroll, John B. "The Nature of the Data, or How to Choose a Correlation Coefficient," Psychometrika: 26 (Dec. 1961), pp. 347-72.

10. Cureton, Edward E. "Note on ϕ / ϕmax," Psychometrika:24 (March 1959), pp. 89-91.

11. Feller, William. An Introduction to Probability Theory and Its Applications. New York: John Wiley, 1957.

12. Kemeny, J.G. and J.L. Snell. Finite Markov Chains. Princeton: Van Nostrand, 1960.